



The application of wavelet transform of Raman spectra to facilitate transfer learning for gasoline detection and classification



Ting-Yu Huang ^{a,d}, Jianzhong Wang ^b, Qingzhong Liu ^c, Jorn Yu ^{a,*}

^a Department of Forensic Science, College of Criminal Justice, Sam Houston State University, Huntsville, TX 77341, USA

^b Department of Mathematics and Statistics, Sam Houston State University, Huntsville, TX 77341, USA

^c Department of Computer Science, Sam Houston State University, Huntsville, TX 77341, USA

^d Department of Criminal Justice, Ming Chuan University, Taiwan

ARTICLE INFO

Keywords:

Hand-held Raman spectrometer
Convolutional neural network (CNN)
Continuous wavelet transform (CWT)
Transfer learning
Ignitable liquid
Artificial intelligence (AI)

ABSTRACT

This study proposed an artificial intelligence (AI) approach for rapid, automatic detection of gasoline samples from different ignitable liquid samples based on their Raman spectra. The system integrated Raman spectroscopy, Raman signal transform, and transfer learning with a convolutional neural network (CNN). A hand-held Raman spectrometer was utilized to collect the learning dataset, including 180 gasoline spectra and 170 non-gasoline spectra from 17 various types of liquid samples. Continuous wavelet transform (CWT) was adopted to transform Raman spectra into image representations to facilitate transfer learning using a pre-trained CNN for image recognition. This approach streamlined AI model training, performance verification, and assessment. Experimental results showed that the CNN model could achieve 100% classification performance in precision, sensitivity, specificity, F1 score, and accuracy. The established AI model could be adopted for the Raman spectrum collected by different accessories, such as using the point-and-shoot attachment of liquid samples in glass vials. The experimental results suggested that CWT processing of Raman scattering signals enabled effective CNN transfer learning. The study demonstrated that CWT could be applied for Raman spectra processing to develop edge AI in forensic field testing of physical evidence.

1. Introduction

In the United States, more than 16,000 arson cases were reported in 2020 in residential buildings that resulted in more than 280 deaths and \$590 million in property, according to the U.S. Fire Administration [1]. In the report, gasoline was one of the most encountered ignitable liquids in intentionally set fires. The detection and identification of gasoline in arson investigation is crucial based on evidence rules [2]. It helps to establish the associations between the commitment of the suspect and the cause of the fire. However, modern standards for identifying forensic gasoline evidence require sample preparation and gas chromatography–mass spectrometry (GC-MS) analysis [3,4], which could be time-consuming [5], and not field-deployable. Even though those methods can provide detailed analytical information, the interpretation of the data can be a challenge because it relies heavily on the analyst's experience and skill [6].

Raman spectroscopy is an analytical technique based on detecting molecular vibrational energy from inelastic light scattering. During the

analysis, a laser light source irradiates the sample and is scattered by the sample molecule that is characteristic of a specific bond. A Raman spectrum provides precise spectral fingerprint information that allows molecular structure identification [7]. Raman spectroscopy has been considered a straightforward analytic approach with minimal sample preparation and short analysis time [8–10]. The advantages have made Raman spectroscopy be widely explored for its application in forensic science. For example, Dies et al. demonstrated illicit drugs (cocaine, heroin, THC, and oxycodone) in aqueous samples, and saliva could be differentiated by Raman spectroscopy coupled with principal component analysis (PCA) and support vector machine (SVM) [11]. Doty and Lednev used a Raman spectrometer and partial least squares discriminant analysis (PLSDA) to distinguish human and animal blood [12]. Widjaja and Lim classified human gender based on fingernail clippings analyzed by Raman spectroscopy and PCA combined with SVM [13]. Wiktelius et al. employed Raman spectroscopy and random forest method to model the synthesis route attribution of sulfur mustard, a chemical warfare agent, to determine the substance's origin of

* Corresponding author.

E-mail address: jornyu@shsu.edu (J. Yu).

Table 1
Raman database of the proposed study.

| Dataset | Number of Raman spectra acquired | Total |
|--------------------------|----------------------------------|------------|
| Training | Gasoline | 180 |
| | Non-gasoline ILs | 170 |
| Performance verification | Gasoline | 45 |
| | Non-gasoline ILs | 50 |
| Performance assessment | Weathered samples | 70% |
| | Ambient environment | 50% 25% |
| | 0 mm | 3 |
| | 1 mm | 3 |
| | 2 mm | 3 |
| | 3 mm | 3 |
| | 4 mm | 3 |
| | 5 mm | 3 |
| Total | | 553 |

Table 2
Parameters for CWT and fine-tuning the weights of the pre-trained CNN model.

| Phase | Parameter | Input |
|-----------------------|--------------------------|--------|
| CWT | Signal length | 1000 |
| | Sampling frequency | 128 |
| CNN transfer learning | Voices per octave | 12 |
| | Weight learn rate factor | 5 |
| | Bias learn rate factor | 5 |
| | Mini batch size | 15 |
| | Max epochs | 5 |
| | Initial learn rate | 0.0001 |
| | Validation frequency | 10 |
| | Execution environment | cpu |

production [14]. Gasser et al. developed a stand-off hyperspectral Raman imaging system with random forest modeling to identify explosives remotely and rapidly [15]. These studies showed that chemometrics is an essential method for providing accurate classification or regression results when optical-based signals such as Raman data are analyzed. Typically, a multivariate data analysis algorithm is exploited in machine learning to recognize features in Raman spectra so that the algorithm can acquire knowledge from raw spectra to make future data predictions [16]. To make the decision precisely and accurately, those machine learning approaches usually require advanced and sophisticated data pre-processing techniques to extract high-quality spectral features [17]. To some degree, this can be a complicated or even

overwhelming data analysis process for analysts. Deep learning is a sub-branch of machine learning which improves the limitations of traditional learning approaches. A deep learning algorithm employs a neural network with multiple layers to learn complex relationships and generalize learned features among input data. Increasing attention to neural networks from research domains is a convolutional neural network (CNN).

CNN is one of the most popular deep learning algorithms inspired by biological visual mechanisms [18]. A CNN architecture comprises convolutional layers, pooling layers, and fully-connected layers responsible for specific functions, including feature extraction, dimension reduction, or data classification [19]. Each layer can be viewed as a filter and operates in a hierarchical manner, where earlier layers recognize common features and deeper layers focus on more detailed features and make a final decision on classification [20]. This modular architecture simplifies model development but achieves even higher performance on classification tasks than traditional methods [21,22]. CNN models have been applied in speech [23,24], face [25,26], and image [27–30] recognition. In particular, it is reported that a CNN has a powerful and superior capability to classify image-based data [31–33]. Because of the data-hungry nature, developing a CNN model from scratch depends heavily on large-scale datasets [34]. However, a huge amount of data is often unavailable in practical applications. An alternative solution for this issue is to apply transfer learning, which is to repurpose and fine-tune a pre-trained CNN model for a new task.

This study aims to develop artificial intelligence (AI) system for rapid, accurate, and automatic gasoline detection that can be deployed to fields. The design combined a hand-held Raman spectrometer and deep CNN transfer learning. One new approach of this study was to apply wavelet transform to convert Raman scattering signals into image

Table 3
Performance indicators used in the proposed study.

| Performance Indicator | Formulation | |
|-----------------------|-------------------------------------|---|
| Precision | $\frac{TP}{TP + FP}$ | TP: True positive FP: False positive |
| Sensitivity | $\frac{TP}{TP + FN}$ | TN: True negative FN: False negative |
| Specificity | $\frac{TN}{TN + FP}$ | |
| F1 score | $\frac{2TP}{2TP + FN + FP}$ | |
| Accuracy | $\frac{TP + TN}{TP + FP + TN + FN}$ | |

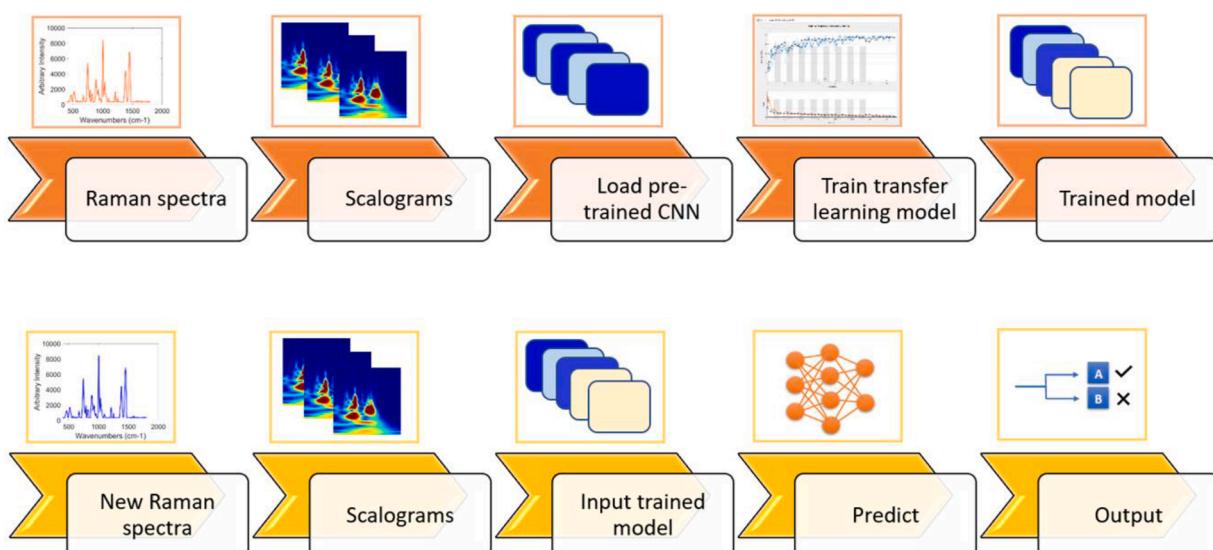


Fig. 1. The workflow of the proposed study.

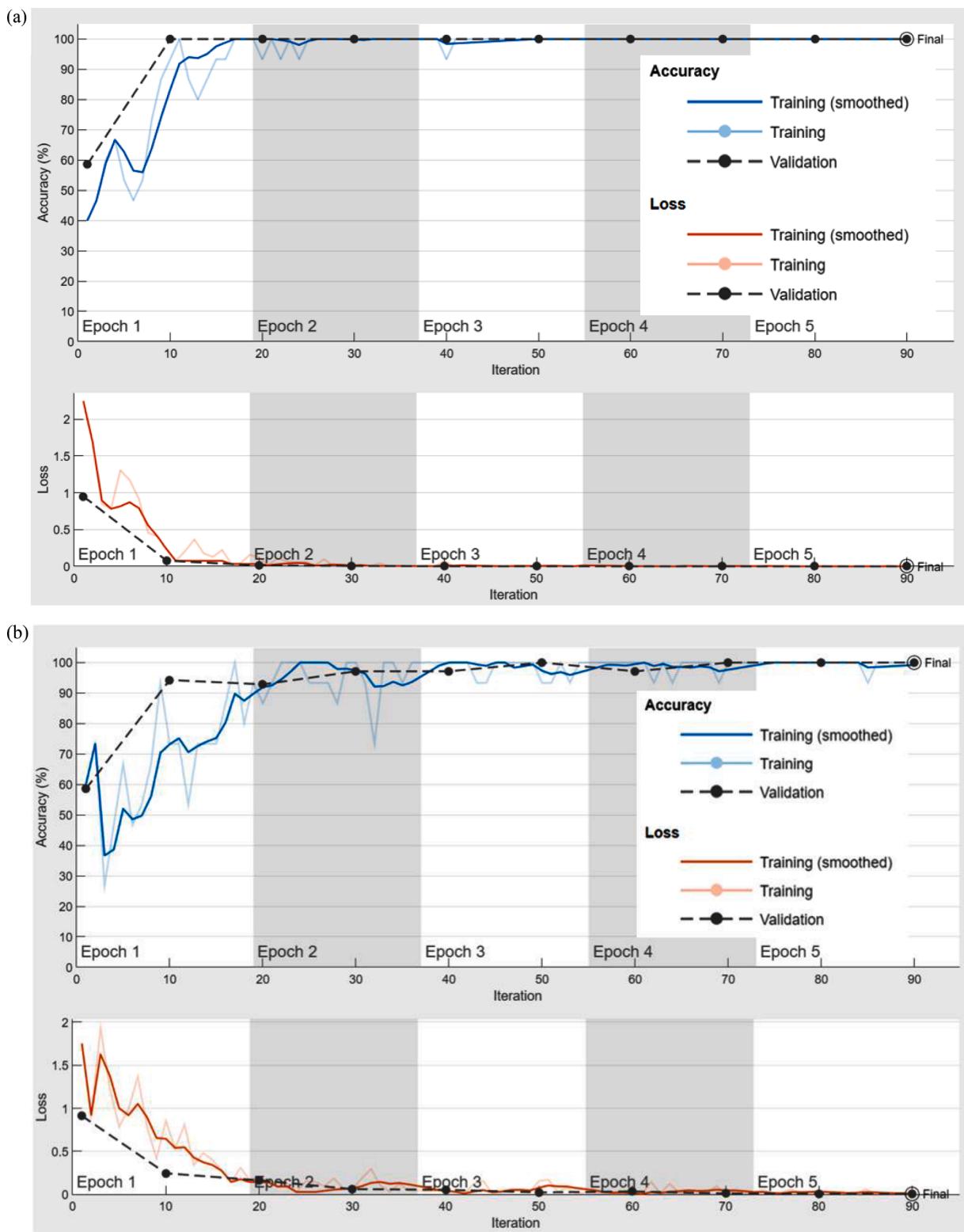


Fig. 2. The training progress of the re-trained GoogLeNet model (a) trained on CWT transformed images, and (b) trained on raw Raman spectra.

representations to facilitate transfer learning. Wavelet transform has been introduced to Raman spectroscopy to eliminate noises and background signals or detect peaks for chemical identification [35–37]. Our research group also demonstrated the feasibility of using continuous wavelet transform (CWT) to process Raman signals in crime scene intelligence [38]. Different from our previous work [38], in this study, the impact of the Raman measurement environment (i.e., signal acquisition

modes) on the classification performance of the AI model was investigated. The performance of the AI model related to sample weathering was also reported to discuss the sensitivity and limitation of CWT transformation in transfer learning. We used a hand-held Raman spectrometer to collect 180 gasoline Raman spectra and 170 non-gasoline Raman spectra from 17 ignitable liquids (ILs) as the training data. After spectra collection, each spectrum was transformed into an

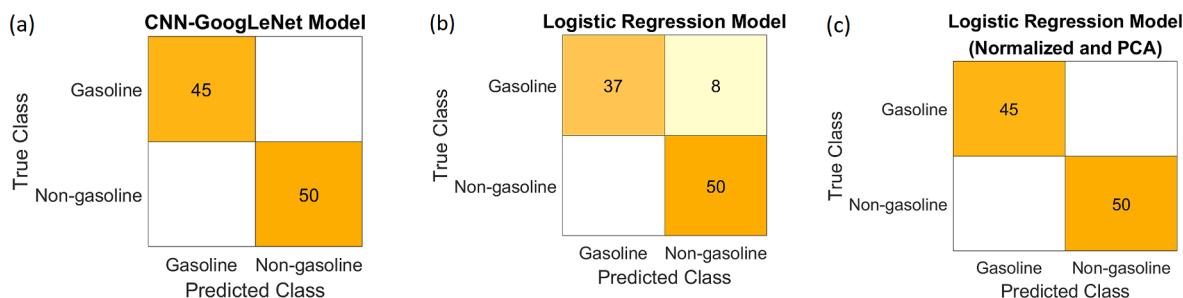


Fig. 3. Confusion matrices of performance verification dataset (a) GoogLeNet model, (b) logistic regression model (raw Raman spectra), and (c) logistic regression model (pre-processed Raman spectra).

Table 4

Comparison of performance verification between the proposed method, logistic regression model using raw Raman spectra, and logistic regression model using pre-processed Raman spectra.

| Indicators | Logistic regression model | Logistic regression model (Normalized and PCA) | GoogLeNet Model |
|-------------|---------------------------|--|-----------------|
| Precision | 1 | 1 | 1 |
| Sensitivity | 0.82 | 1 | 1 |
| Specificity | 1 | 1 | 1 |
| F1 score | 0.90 | 1 | 1 |
| Accuracy | 0.92 | 1 | 1 |

imaginary datum compatible with the pre-trained CNN model chosen for the investigation in this work. Upon completing model training, separate gasoline and non-gasoline ILs datasets were utilized to verify the training result. A conventional machine learning model was developed for classification performance compared with the proposed method. To assess the effects of sample nature and testing conditions on the performance of the proposed method, weathered gasoline samples were measured under ambient environment by the hand-held Raman spectrometer for evaluation. The work is a demonstration to show that AI has great potential for field analysis of forensic evidence. The developed system can be a promising tool to assist and support crime scene investigations.

2. Materials and methods

2.1. Sample collection and preparation

Gasoline samples included 87 regular, 89 mid-grade, and 93 premium grades were purchased from a local gas station in Houston, Texas. Aliquots of gasoline were transferred into 2 mL standard glass vials (Glass Vials, Maryland, USA) that could be fitted inside the measuring chamber of the hand-held Raman spectrometer. To train the model and verify the performance of the model, three 1 mL aliquots were sampled and measured from each grade of gasoline. To assess the effect of the weathering of gasoline on the performance of the model, three levels of weathered gasoline (70, 50, and 25%) were prepared from each gasoline grade in the lab. Three 1 mL aliquots of weathered gasoline were sampled and measured. To assess the ambient environment's effect during Raman measurements on the performance of the model, the Raman spectra of each gasoline grade were obtained using the point-and-shot method.

Seventeen non-gasoline ILs, including kerosene, pure gum spirits turpentine, charcoal lighter, odorless mineral spirit, VM&P naphtha, petroleum ether, ultra-grade performance vacuum oil, toluene, xylene, benzene, pentane, hexane, heptane, dodecane, ethanol, acetone, and nail polish were purchased from the local stores. Three 1-mL aliquots of each IL were sampled and measured to train the model. Among all the investigated ILs, another dataset was created from five ILs (kerosene, petroleum ether, toluene, xylene, and nail polish) to verify the model's

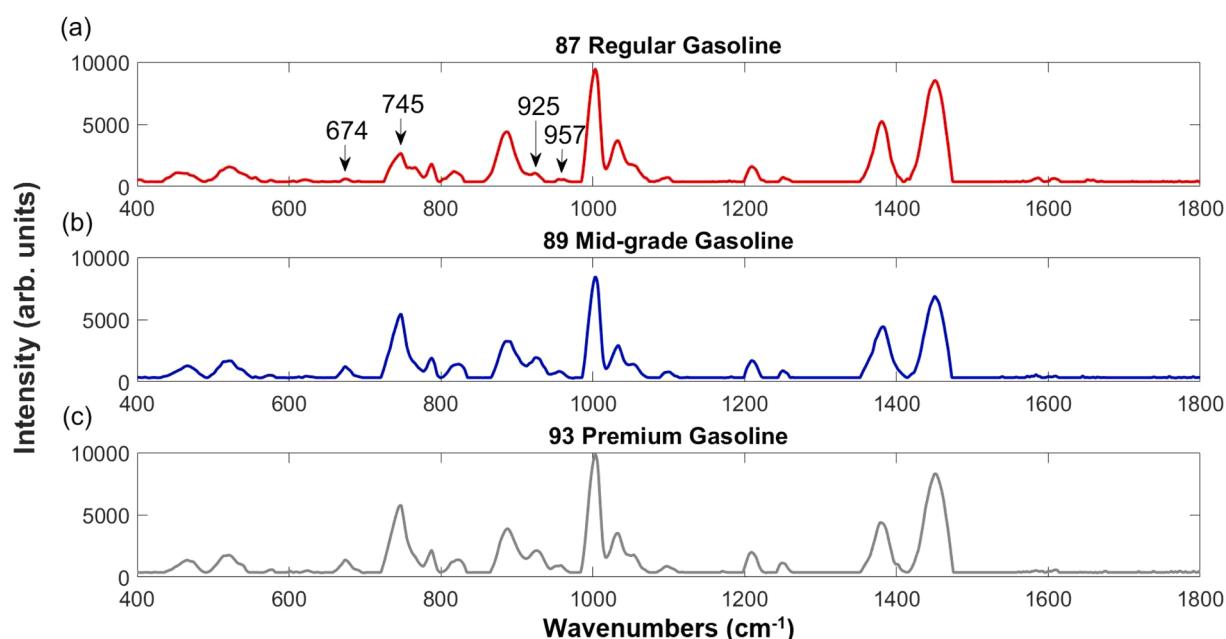


Fig. 4. Comparison of raw gasoline Raman spectra of 87 regular, 89 mid-grade, and 93 premium grades.

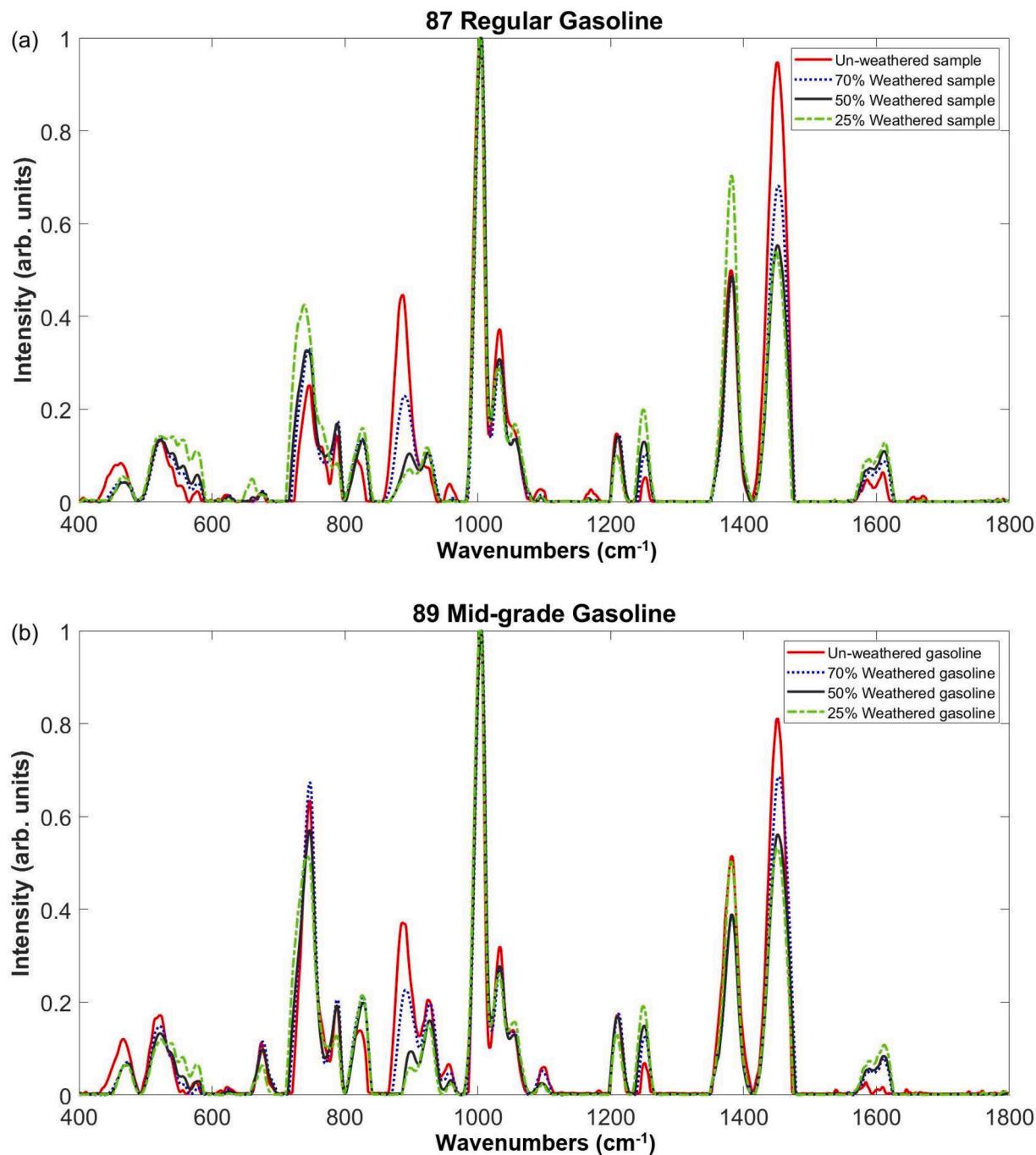


Fig. 5. Comparison of normalized gasoline Raman spectra of un-weathered state and three levels of weathered states (70%, 50%, and 25%) (a) 87 regular grade, (b) 89 mid-grade, and (c) 93 premium grade.

performance.

2.2. Hand-held Raman spectrometer measurements

Raman measurements were performed on a portable Raman spectrometer with a 785 nm laser wavelength (HandyRamTM, Field Forensic Inc., Florida, USA). The laser power of the spectrometer was 100 mW with 12 cm^{-1} spectral resolution. All samples except those used for evaluating ambient environment effect were placed in the vial compartment of the hand-held Raman spectrometer for spectra collection. For assessing the ambient environment effect, samples were measured by the point and shoot method. The distances between the

side of each vial and the point-and-shoot attachment were set as 0, 1, 2, 3, 4, and 5 mm. Each spectrum was acquired by employing the hand-held Raman's autointegration function to automatically determine the integration time and baseline-subtracted to correct the background signal. The spectral range was recorded from 400 to 2300 cm^{-1} for analysis. Table 1 depicts that a total of 553 Raman spectra were acquired for training the model (350 Raman spectra), verifying the performance of the model (95 Raman spectra), and assessing the effect of weathered gasoline (90 Raman spectra) and ambient environment (18 Raman spectra) on the performance of the model.

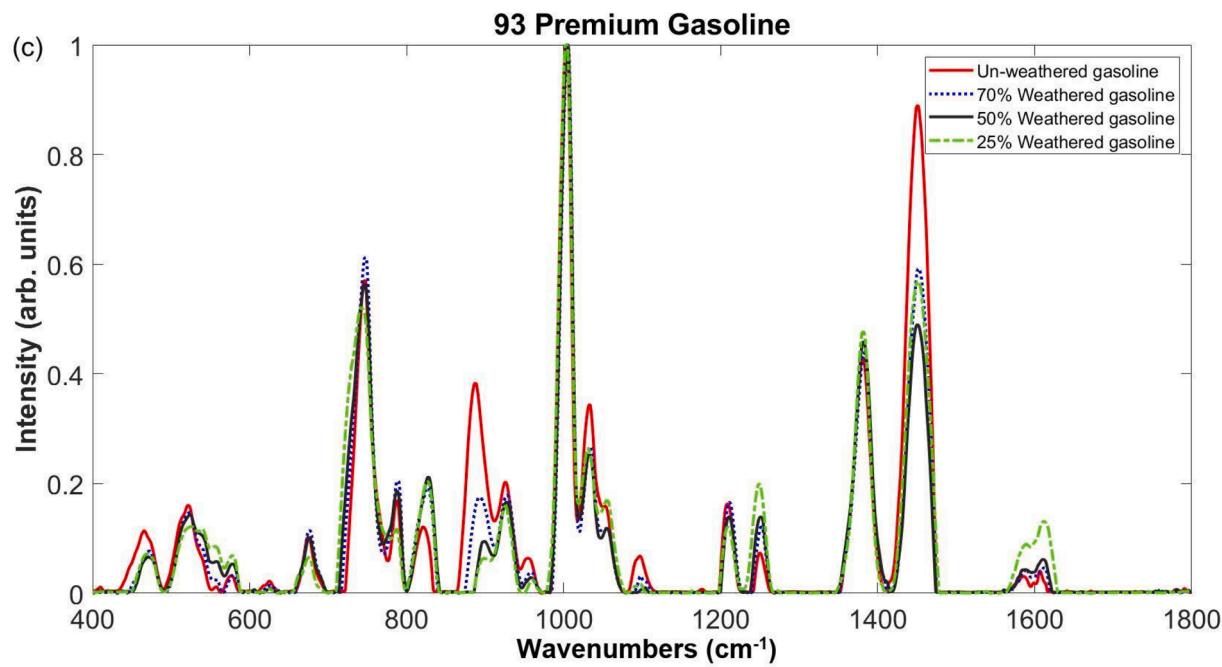


Fig. 5. (continued).

Table 5
Performance assessment prediction results.

| (a) Weathered gasoline | | | | | |
|--------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|--|
| | Un-weathered Sample | 70% Weathered sample | 50% Weathered sample | 25% Weathered sample | |
| Predictive power | 1 | 1 | 0 | 0.17 | |
| Average prediction probability | 1 | 0.87 | 0.07 | 0.29 | |
| Standard deviation | 2.68×10^{-5} | 1.92×10^{-2} | 9.22×10^{-4} | 3.39×10^{-2} | |

| (b) Ambient environment (point and shoot) | | | | | | |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | Close contact | 1 mm | 2 mm | 3 mm | 4 mm | 5 mm |
| Predictive power | 0 | 0.67 | 1 | 0.67 | 0 | 0 |
| Average prediction probability | 0.14 | 0.68 | 1 | 0.73 | 5.67×10^{-4} | 3.33×10^{-4} |
| Standard deviation | 1.45×10^{-3} | 4.65×10^{-2} | 4.33×10^{-8} | 6.33×10^{-2} | 1.63×10^{-7} | 1.03×10^{-7} |

2.3. Raman spectra processing and deep CNN training

After Raman spectra collection, the next step was to process the spectra by continuous wavelet transform (CWT) [39]. In our work, the purpose of adopting CWT was to transform Raman spectra into imaginary data called scalograms. We used a CWT filter bank based on an analytic Morse wavelet for processing Raman spectra. Morse wavelets were developed and applied in modulated signals whose amplitude and frequency vary with time [40–43]. There are two parameters in Morse wavelets that control the visual pattern and the behavior of wavelets, which are symmetry and time-bandwidth products [41,42]. In our study, symmetry and time-bandwidth products were set as a value of 3 and 60, respectively. Other parameters related to the CWT filter bank are listed in Table 2. After specifying the parameters, the filter bank was precomputed so that both the CWT of Raman signals in all spectra and the corresponding scalograms could be taken and generated automatically. Raman scattering signals in each spectrum were transformed into

a time and frequency representation with the proposed settings. We hypothesized that Raman signals being analogous to modulated signals for image transformation could benefit feature extraction by subsequent deep CNN training to discriminate species. To be compatible with the pre-trained CNN model, each scalogram was generated in an RGB image format with an array of 224 by 224 by 3.

The pre-trained CNN model for the transfer learning task in this study was GoogLeNet. The GoogLeNet model was initially designed for classifying images into 1000 categories. The architecture of the model comprises 22 layers deep and 144 layers. The network is famous for the inception module in the architecture, enabling lower computational cost with a high classification accuracy by reducing the input image's dimensionality and the number of the training parameters [44]. The first step to re-train the GoogLeNet model was replacing the final two layers, loss3-classifier, and output. In the 144 layers of the GoogLeNet CNN structure, the “loss3-classifier” is the 142nd layer trained as the last learnable layer to extract features from images [46]. The output layer is the 144th layer designed as the final classification layer to classify objects. The two layers aimed to integrate the extracted features by the network to produce predicted labels and probabilities of labels. The reasons for replacing the two layers with new layers were to adapt to the data labels and the number of classes investigated in our work. In addition, it helped increase the learning efficiency. The second step was to tune the training options. We used a stochastic gradient descent with a momentum optimizer (sgdm) to minimize the loss function involved in the iterative training process. Options revised for training also included the size of a sub-training set in each iteration, total epochs used for training, the first step size in the orientation of the negative gradient of the loss function, and so on. The specification of each parameter for the deep CNN training is provided in Table 2.

Once the network had been fine-tuned, a personal laptop CPU was used to run the training. During the run, the CNN algorithm validated its accuracy according to the frequency we designated in the training options, as shown in Table 2. The scalograms in the training dataset were randomly divided into two groups. 80% of the images (288 scalograms) were used to train the model, and 20% of the images (72 scalograms) were used to validate the model. After completing training, scalograms in the verification and assessment datasets were inputted into the model to verify and assess the model's performance. The outputs were a label of

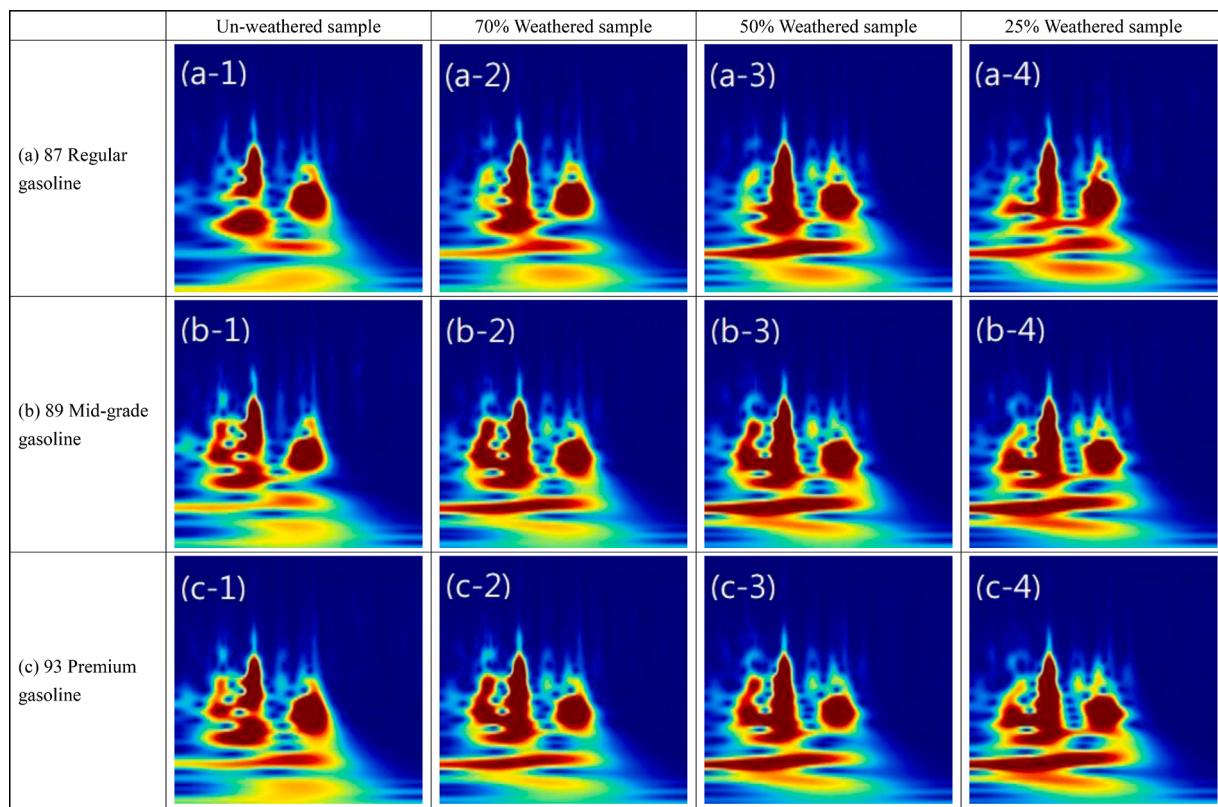


Fig. 6. Comparison of scalograms of un-weathered state and weathered states (a) 87 regular grade, (b) 89 mid-grade, and (c) 93 premium grade: (a-1/b-1/c-1) un-weathered sample, (a-2/b-2/c-2) 70% weathered sample, (a-3/b-3/c-3) 50% weathered sample, (a-4/b-4/c-4) 25% weathered sample.

Table 6
Performance assessment ANOVA results.

| ANOVA | Weathered gasoline | Ambient environment |
|---------------------------|------------------------|------------------------|
| Group sum of squares | 20.71 (df = 3) | 6.97 (df = 6) |
| Error sum of squares | 1.57 (df = 131) | 0.22 (df = 56) |
| Total | 22.28 (df = 134) | 7.19 (df = 62) |
| Means square of the group | 6.90 | 1.16 |
| Means square of the error | 0.01 | 4×10^{-3} |
| F -value | 575.96 | 290.78 |
| p -value | 3.11×10^{-75} | 2.56×10^{-40} |

the gasoline or non-gasoline category and the probabilities of each of the categories.

Fig. 1 depicts the overall workflow of transfer learning performed in this study. The entire process was conducted on MATLAB R2020b (MathWorks, Natick, Massachusetts, USA).

2.4. Data analysis

To gain a deeper insight into the proposed method's performance, we designed a stepwise evaluation work to analyze the data derived from verification and assessment experimental results. First, we compared the classification performance of the proposed method with the conventional machine learning approach. The same training dataset and verification dataset containing raw Raman spectra in the database (Table 1) were utilized to develop and verify a logistic regression model. In addition, raw Raman spectra in the two datasets were further pre-processed by exporting to the Raman Processing program (SSIM / CARES research group at Wayne State University, USA) [45] to perform normalization for eliminating intensity differences and MATLAB to perform PCA for feature extraction and transformation. The training and verification procedures were then implemented again for the logistic

regression model using the pre-processed spectra. The entire development and verification process of the logistic regression model was conducted on the Classification Learner App in MATLAB. To quantitatively analyze the classification performance between the proposed method and the conventional method, performance indicators including precision, sensitivity, specificity, and F1 score were calculated to evaluate the models' performance on predicting each label (gasoline or non-gasoline ILs). Accuracy was calculated to evaluate the overall performance of the models. Table 3 provides the formulations of the performance indicators. True positive refers to the output data that were classified as gasoline by the models, and they were indeed gasoline. In contrast, false-positive are the output data classified as gasoline, but they were non-gasoline ILs. On the contrary, true negative refers to the output data that were classified as non-gasoline ILs by the models, and they were truly non-gasoline ILs. In contrast, false negatives are the output data classified as non-gasoline ILs, but they were gasoline.

The second evaluation work compared the differences in mean numbers of prediction probabilities given by the proposed model between different levels of weathered gasoline groups and between other distance groups when employing the point and shoot method under ambient environment by the hand-held Raman spectrometer in the assessment dataset. One-way ANOVA was performed for data analysis. The independent variables were weathered states and distances from the point-and-shoot attachment to the side of the vial, respectively. Mean differences were considered statistically significant when $p < 0.05$.

3. Results and discussion

3.1. Model training

Based on the proposed fine-tuning parameters, the elapsed time for the CNN-GoogLeNet model training was 4 min and 29 s. There were five epochs in the training cycle. Each epoch enabled the CNN algorithm to

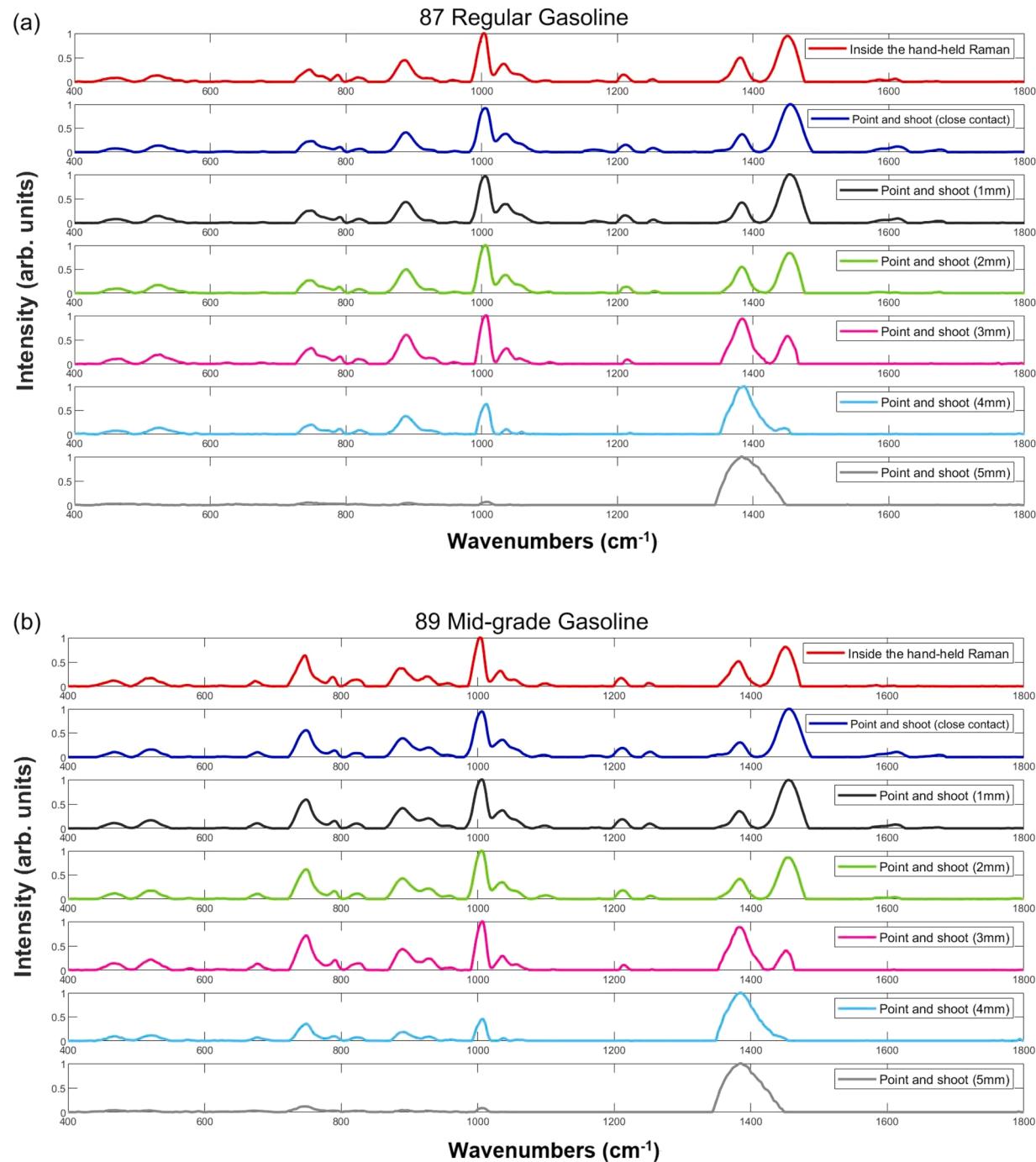


Fig. 7. Comparison of normalized gasoline Raman spectra of standard measuring condition (inside the vial chamber of the hand-held Raman spectrometer) and environmental measuring condition (point and shoot method, the distances from the point-and-shoot attachment to the side of the vial were set as 0, 1, 2, 3, 4, and 5mm) (a) 87 regular grade, (b) 89 mid-grade, and (c) 93 premium grade.

pass over the entire 350 scalograms in the training dataset. Ninety iterations were used for the stochastic gradient descent algorithm to repeatedly evaluate the gradient and update the descent algorithm weights to minimize the loss function. Fig. 2(a) depicts the training progress. Some variations occurred in both accuracy and loss before the third epoch due to the GoogLeNet model starting to learn the new task in this work. Every ten iterations, the GoogLeNet model validated its performance by using the data it had not seen, which were the prior divided 20% of the scalograms in the training dataset. The validation accuracy reached 100% at the tenth iteration, and the validation loss decreased to

zero at the twentieth iteration. The last epoch and iteration result was a validation accuracy of 100% and a loss of zero. This indicates that the GoogLeNet model has successfully learned to recognize and classify gasoline scalograms from different types of non-gasoline ILs' scalograms.

3.2. Model performance verification

The verification dataset was used to verify the result of the training. The performance of the GoogLeNet model was compared with a logistic

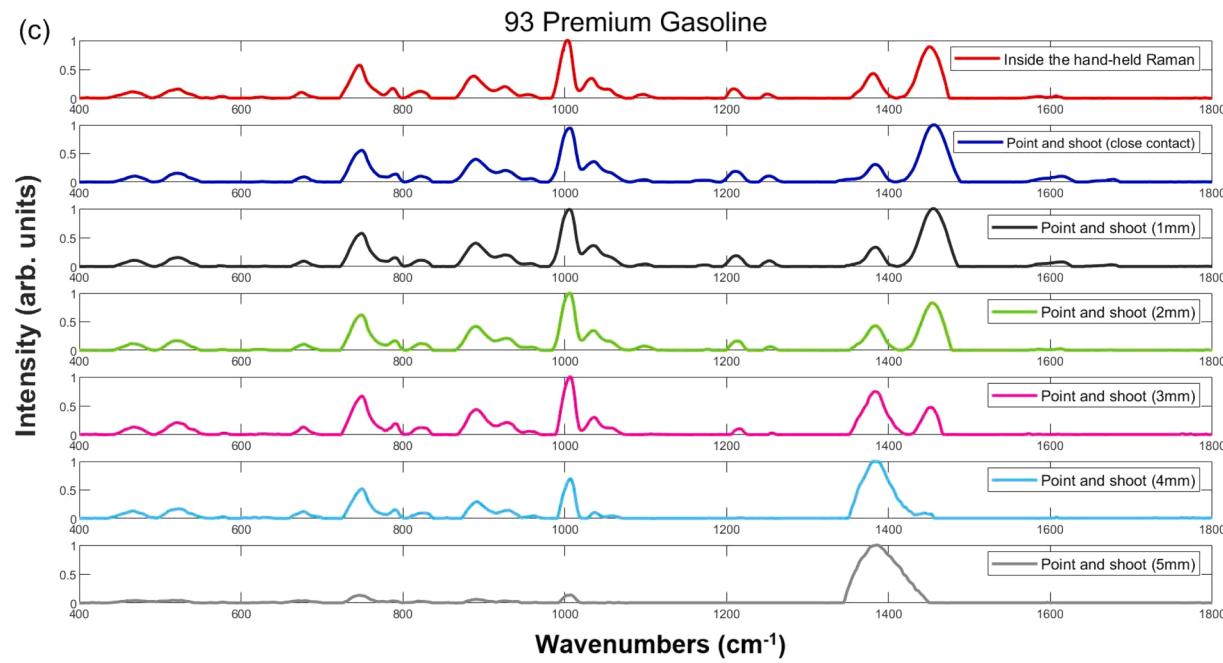


Fig. 7. (continued).

regression model. Fig. 3 provides the classification results of the proposed CNN-GoogLeNet model, the logistic regression model using raw Raman spectra, and the logistic regression model using pre-processed Raman spectra. With transformed scalograms as input data, the re-trained GoogLeNet model identified each scalogram into correct categories (Fig. 3(a)). On the contrary, the logistic regression model misclassified a few raw gasoline Raman spectra into the non-gasoline ILs class (Fig. 3(b)). The logistic regression model made correct classification only when raw Raman spectra were normalized, features extracted, and transformed by PCA (Fig. 3(c)). To quantitatively evaluate the models' performance, several indicators, including precision, sensitivity, specificity, F1 score, and accuracy, were derived from the confusion matrices in Fig. 3. The comparison of the indicators of the models is shown in Table 4. Precision measures the proportion of gasoline data predicted as gasoline class, and all models achieved 100% precision. Sensitivity is used to measure the model's capability to identify all gasoline data in the dataset. In contrast, specificity measures the model's capability to determine all non-gasoline ILs data in the dataset. It is seen that the GoogLeNet model and the logistic regression model using pre-processed Raman spectra obtained the highest scores in both cases. However, when raw Raman spectra were analyzed by the logistic regression model, the score of sensitivity decreased to 0.82. This reveals the unsatisfactory ability of the logistic regression model to detect raw gasoline Raman spectra. The F1 score is a weighted average between precision and sensitivity. Due to the poor sensitivity for raw gasoline spectra, the logistic regression model only obtained 0.9 for the F1 score, whereas the GoogLeNet model still held the best score. The accuracy is viewed as an overall performance indicator. The result shows that the GoogLeNet model had a high performance in a comprehensive aspect, while the logistic regression model required data pre-processing so that improved accuracy could be obtained.

To further investigate the classification problem with the logistic regression model, it is found that the most easily confused class was the gasoline class in raw Raman spectra. The raw Raman spectral data in this dataset contained 87 regular, 89 mid-grade, and 93 premium grades. As shown in Fig. 4(a–c), peaks at around 520, 788, 1004, 1033, 1211, 1381, and 1607 cm^{-1} indicate the occurrence of aromatic ring chain vibrations, which resulted from the presence of the main contents in gasoline, such as toluene [47]. Another peak at 1451 cm^{-1} shows that methylene

scissoring and methyl bending were found [47]. The predictive power of each gasoline grade by the logistic regression model was 0.47(7/15), 1 (15/15), and 1(15/15), respectively. The reason for 87 regular gasoline to have a relatively weak predictive power could be attributed to different additives in lower and higher performance gasoline. In Fig. 4(a), the 87 regular grade exhibited exclusive variations at peaks located at 674, 745, 925, and 957 cm^{-1} . After taking several additional pre-processed measures, 87 regular grades finally obtained an improved predictive power of 53% (1–15/15). In contrast, the proposed method employed a simple step to transform all Raman spectra into scalograms. As a result, the high predictive power of three gasoline grades (15/15, 15/15, and 15/15) by the re-trained GoogLeNet model was achieved. This reveals that the proposed method held advantages over the conventional machine learning method by showing no dependency on skilled feature extraction procedures. In summary, the primary advantage of the CNN model trained on CWT transformed images was that no manual data preprocessing was applied, such as peak normalization and PCA. In addition, the superior capability of the CNN model to solve the issue of intra-variations in input data by efficiently learning and generalizing the features in scalograms is also demonstrated.

To verify the benefit of the CWT process of Raman spectra in transfer learning, the same training data without CWT transformation (i.e., images of plotted raw Raman spectra without axis) were used to train the GoogLeNet. As shown in Fig. 2(b), although the model achieved a 100% validation accuracy at the end of the training, the learning curves showed a more unstable learning accuracy during the entire training process. In contrast, the validation accuracy of the scalogram-based CNN model achieved 100% with one training epoch, whereas more training epochs (4 epochs) were needed for using the images of raw Raman spectra for transfer learning. Moreover, the prediction probabilities of the gasoline verification data using CWT transformed images as training inputs achieved 99.96% ($n = 45$, SD = 0.005%) compared to that of 99.39% ($n = 45$, SD = 0.0064%) using images of raw Raman spectra as training inputs. Similarly, the prediction probabilities of the non-gasoline verification data using CWT transformed images achieved 100% ($n = 50$) compared to that of 96% ($n = 50$, SD = 0.58%) when images of raw Raman spectra were used as the training inputs. These data suggest that CWT transformed images worked better as inputs for CNN models in transfer learning.

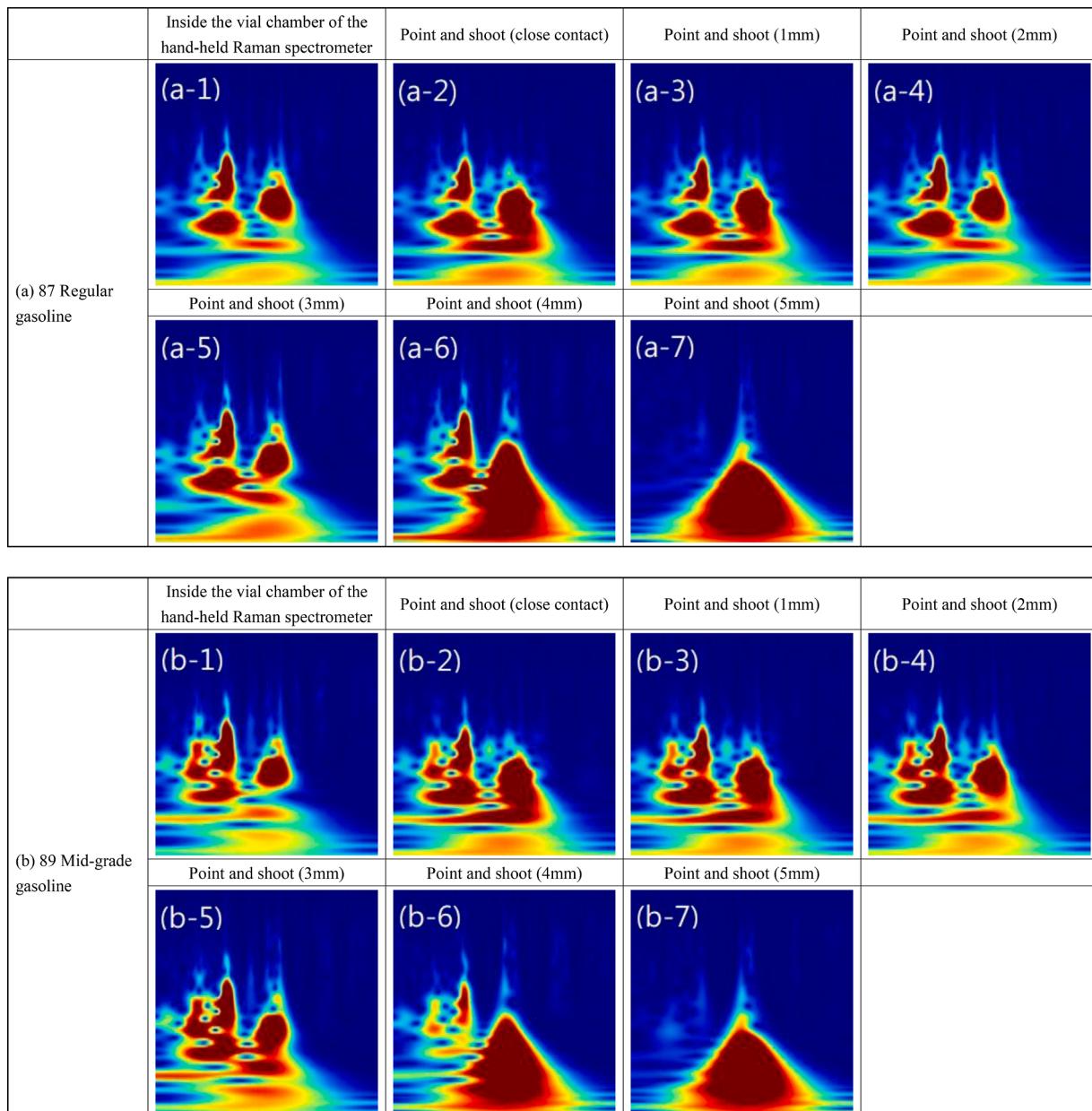


Fig. 8. Comparison of scalograms of standard measuring condition and environmental measuring condition (a) 87 regular grade, (b) 89 mid-grade, and (c) 93 premium grade: (a-1/b-1/c-1) inside the vial chamber of the hand-held Raman spectrometer, (a-2/b-2/c-2) point-and-shoot attachment at close contact of the side of the vial, (a-3/b-3/c-3) a distance of 1mm from the point-and-shoot attachment to the side of the vial, (a-4/b-4/c-4) a distance of 2mm from the point-and-shoot attachment to the side of the vial, (a-5/b-5/c-5) a distance of 3mm from the point-and-shoot attachment to the side of the vial, (a-6/b-6/c-6) a distance of 4mm from the point-and-shoot attachment to the side of the vial, (a-7/b-7/c-7) a distance of 5 mm from the point-and-shoot attachment to the side of the vial.

3.3. Model performance assessment

The assessment dataset was used to evaluate the effect of sample nature and environmental testing conditions on the performance of the proposed method. Sample nature was evaluated by three levels of weathered gasoline (70, 50, and 25%). Compared to the neat gasoline samples, weathered gasoline samples typically contain less amount of light components and a higher amount of heavy components. For example, as shown in Fig. 5, the increased Raman peaks at 1248, 1586, and 1611 cm^{-1} represented heavy components in gasoline samples because they were less volatile. In contrast, the peak intensity decreased at 897, and 1205 cm^{-1} could represent the loss of light components due to evaporation. Note that this comparison was controlled by normalizing

the Raman spectrum using the peak at 1004 cm^{-1} .

The predictive power was 1(30/30), 0(0/30), and 0.17(5/30) for 70%, 50%, and 25% weathered gasoline, respectively (Table 5(a)). It was found that the proposed method had a decreased performance on the weathered samples. The decreased predictive power in 50 and 25% weathered gasoline was due to the evaporation of the volatile compounds in gasoline, which altered the spectral data in peak intensity, peak ratios with a few new peaks appeared due to weathering effect, as shown in Fig. 5. The Raman spectra alteration also affected the pattern of scalograms in different weathered states when employing CWT for image transformation. As shown in Fig. 6, the patterns of scalograms changed a lot due to weathering. This suggested that the CWT process is sensitive to the peak intensity and pattern of the Raman spectrum. As a

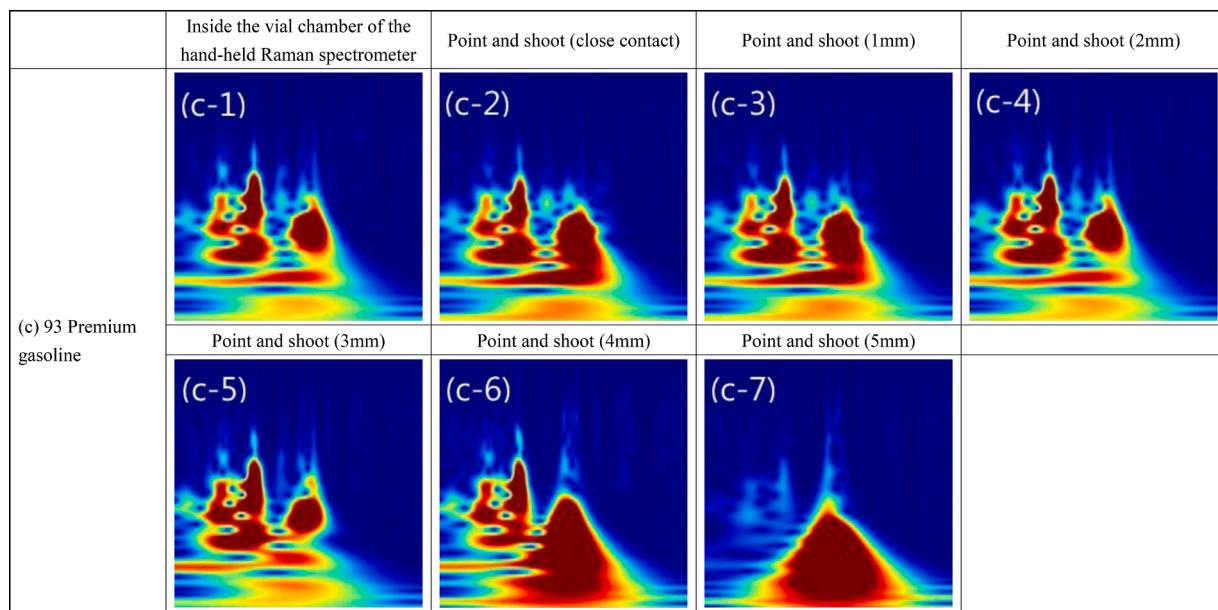


Fig. 8. (continued).

result, the pattern of the scalograms transformed from the Raman spectrum could be a source of features for subsequent deep CNN model training. Moreover, the means of prediction probabilities given by the GoogLeNet model between un-weathered gasoline and three levels of weathered gasoline have statistically significant differences, as reported in Table 6. This provides evidence that the CNN model could detect the change of the pattern in scalograms or, to be specific, the change of the information in the original spectral data. This proves our hypothesis that regarding Raman spectra as an analog to time-series-based data and applying CWT for image transformation could facilitate the deep CNN transfer learning.

The environmental testing condition was evaluated by setting different measuring distances between the side of the vial and the point-and-shoot attachment of the hand-held Raman spectrometer. The predictive power was found to be 0.67(2/3), 1(3/3), and 0.67(2/3) for measuring distances at 1, 2, and 3 mm, respectively, while the rest of the distances were not feasible to classify gasoline scalograms (Table 5(b)). The decreased predictive power could associate with the focal length of the laser. The spots without laser being focused resulted in altered Raman spectra. Fig. 7 demonstrates three grades of gasoline Raman spectra under various measuring distances. It can be seen that a distance of 2 mm from the point-and-shoot attachment to the side of the vial acquired the most similar Raman spectra to the spectra that were measured inside of the hand-held Raman spectrometer. The remaining distances caused the issue of peak pattern alteration. Once again, scalograms transformed from those altered Raman spectra also exhibited obvious variations in the pattern. As shown in Fig. 8, scalograms transformed from those Raman spectra collected at a distance of 2 mm maintained the most resembling pattern compared to the standard scalograms. The ANOVA result also shows significant differences in predicting probabilities given by the GoogLeNet model among all measuring distances (Table 6). All those findings corroborate the sensitivity of CWT to variations in original spectral data and the superior ability of the CNN model to learn features and detect changes in scalograms mentioned above.

The assessment of the GoogLeNet model performance also reveals some limitations for the proposed method. The model's performance may decrease for a volatile substance because the spectral profiles alter due to changing its physical properties. In our study, the high performance of the weathered gasoline classification could maintain at least to 70% weathered state with an average prediction probability close to

90% (Table 5(a)). In addition, the ambient environment involving the issue of laser focusing can also impact the prediction result. The optimal spot for gasoline samples being measured under an ambient environment in our work was a distance of 2 mm from the point-and-shoot attachment to the side of the vial (Table 5(b)). However, it should be noted that those input data used in the assessment dataset were not included in the model training process. Future work on extending the training database can improve the effect of sample nature and environmental testing conditions on the performance of the proposed model.

4. Conclusion

This study proposed a new AI system that integrated the techniques of hand-held Raman spectroscopy, image transformation, and deep CNN transfer learning. Experimental results suggested several findings for our new platform. First, CWT was feasible to process Raman scattering signals to facilitate AI. The pattern in the transformed imaginary data provided the possibility of feature extraction during a deep CNN training. Second, the proposed CNN model could identify Raman patterns in the transformed images by CWT. This means that the CNN model coupling with CWT could detect features in Raman spectra for sample classification. The model training and verification both resulted in ceiling-level performance on gasoline detection in terms of precision, sensitivity, specificity, F1 score, and accuracy. It is found that the performance could be affected by the evaporation of gasoline and environmental measuring conditions. However, high performance could be maintained at a level down to 70% weathered samples and an optimal distance of 2 mm from the point-and-shoot attachment of the hand-held Raman spectrometer to the side of the vial. Overall, the proposed method had no dependency on handcrafted engineering. Its prominent learning and generalizing ability allowed for accurate and automated gasoline detection without analysts' intervention to interpret data. Furthermore, the system's output included prediction probability so that the reliability and characterizations of the experimental results could be evaluated. More recent AI models, such as efficientnet-v2, could be tested to update the weights and the loss function because these parameters may significantly impact the AI performance. The proposed method could be a promising platform to assist crime scene investigation.

Declaration of Competing 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.

Acknowledgments

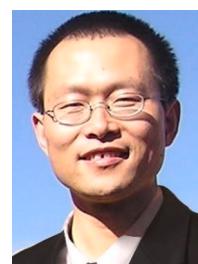
Ting-Yu Huang appreciated the funding support from the Ministry of Education, Taiwan.

References

- [1] Intentionally Set Fires in Residential Buildings (2008–2010), Topical Fire Report Series, U.S. Fire Administration, 13 (2012). <https://www.usfa.fema.gov/downloads/pdf/statistics/v13i10.pdf> (accessed March 19, 2021).
- [2] H.L. Birks, A.R. Cochran, T.J. Williams, G.P. Jackson, The surprising effect of temperature on the weathering of gasoline, *Forensic Chem.* 4 (2017) 32–40, <https://doi.org/10.1016/j.jforc.2017.02.011>.
- [3] ASTM. E1412 –19, Standard practice for separation of ignitable liquid residues from fire debris samples by passive headspace concentration with activated charcoal, ASTM International, West Conshohocken, PA, 2019, <https://doi.org/10.1520/E1413-19>.
- [4] ASTM E1618 - 19, Standard Test method for ignitable liquid residues in extracts from fire debris samples by gas chromatography-mass spectrometry, ASTM International, West Conshohocken, PA, 2019, doi: 10.1520/E1618-19.
- [5] L.C. Cocco, C.I. Yamamoto, O.F. Von Meien, Study of correlations for physicochemical properties of Brazilian gasoline, *Chemom. Intell. Lab. Syst.* 76 (2005) 55–63, <https://doi.org/10.1016/j.chemolab.2004.09.004>.
- [6] P. Doble, M. Sandercock, E. Du Pasquier, P. Petocz, C. Roux, M. Dawson, Classification of premium and regular gasoline by gas chromatography/mass spectrometry, principal component analysis and artificial neural networks, *Forensic Sci. Int.* 132 (2003) 26–39, [https://doi.org/10.1016/s0379-0738\(03\)0002-1](https://doi.org/10.1016/s0379-0738(03)0002-1).
- [7] G. McLaughlin, K.C. Doty, I.K. Lednev, Raman spectroscopy of blood for species identification, *Anal. Chem.* 86 (2014) 11628–11633, <https://doi.org/10.1021/ac5026368>.
- [8] H. Lu, S. Tian, L. Yu, X. Lv, S. Chen, Diagnosis of hepatitis B based on Raman spectroscopy combined with a multiscale convolutional neural network, *Vib. Spectrosc.* 107 (2020), 103038, <https://doi.org/10.1016/j.vibspec.2020.103038>.
- [9] H. Chen, C. Chen, H. Wang, C. Chen, Z. Guo, D. Tong, et al., Serum Raman spectroscopy combined with a multi-feature fusion convolutional neural network diagnosing thyroid dysfunction, *Optik* 216 (2020), 164961, <https://doi.org/10.1016/j.ijleo.2020.164961>.
- [10] J. Dong, M. Hong, Y. Xu, X. Zheng, A practical convolutional neural network model for discriminating Raman spectra of human and animal blood, *J. Chemom.* (2019) 33, <https://doi.org/10.1002/cem.3184>.
- [11] H. Dies, J. Raveendran, C. Escobedo, A. Docoslis, Rapid identification and quantification of illicit drugs on nanodendritic surface-enhanced Raman scattering substrates, *Sens. Actuators B* 257 (2018) 382–388, <https://doi.org/10.1016/j.snb.2017.10.181>.
- [12] K.C. Doty, I.K. Lednev, Differentiation of human blood from animal blood using Raman spectroscopy: a survey of forensically relevant species, *Forensic Sci. Int.* 282 (2018) 204–210, <https://doi.org/10.1016/j.forsciint.2017.11.033>.
- [13] E. Widjaja, G.H. Lim, A novel method for human gender classification using Raman spectroscopy of fingernail clippings, *Analyst* 133 (2008) 493, <https://doi.org/10.1039/b712389b>.
- [14] D. Wiktelius, L. Ahlinder, A. Larsson, K. Höjer Holmgren, R. Norlin, P. O. Andersson, On the use of spectra from portable Raman and ATR-IR instruments in synthesis route attribution of a chemical warfare agent by multivariate modeling, *Talanta* 186 (2018) 622–627, <https://doi.org/10.1016/j.talanta.2018.02.108>.
- [15] C. Gasser, M. Göschl, J. Ofner, B. Lendl, Stand-off hyperspectral Raman imaging and random decision forest classification: a potent duo for the fast, remote identification of explosives, *Anal. Chem.* 91 (2019) 7712–7718, <https://doi.org/10.1021/acs.analchem.9b00890>.
- [16] F. Lussier, V. Thibault, B. Charron, G.Q. Wallace, J.F. Masson, Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering, *TrAC Trends Anal. Chem.* 124 (2020), 115796, <https://doi.org/10.1016/j.trac.2019.115796>.
- [17] R. Zhang, H. Xie, S. Cai, Y. Hu, G. Liu, W. Hong, Z. Tian, Transfer-learning-based Raman spectra identification, *J. Raman Spectrosc.* 51 (2019) 176–186, <https://doi.org/10.1002/jrs.5750>.
- [18] G.W. Lindsay, Convolutional neural networks as a model of the visual system: past, present, and future, *J. Cogn. Neurosci.* (2020) 1–15, https://doi.org/10.1162/jocn_a.01544.
- [19] Z. Qiu, J. Chen, Y. Zhao, S. Zhu, Y. He, C. Zhang, Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network, *Appl. Sci.* 8 (2018) 212, <https://doi.org/10.3390/app802012>.
- [20] P.M. Cheng, H.S. Malhi, Transfer learning with convolutional neural networks for classification of abdominal ultrasound images, *J. Digit. Imaging* 30 (2016) 234–243, <https://doi.org/10.1007/s10278-016-9929-2>.
- [21] J. Liu, M. Osadchy, L. Ashton, M. Foster, C.J. Solomon, S.J. Gibson, Deep convolutional neural networks for Raman spectrum recognition: a unified solution, *Analyst* 142 (2017) 4067–4074, <https://doi.org/10.1039/c7an01371>.
- [22] K. Thenmozhi, U.Srinivasulu Reddy, Crop pest classification based on deep convolutional neural network and transfer learning, *Comput. Electron. Agric.* 164 (2019), 104906, <https://doi.org/10.1016/j.compag.2019.104906>.
- [23] A. Elouali, Z. Elberrichi, N. Elouali, Hate speech detection on multilingual twitter using convolutional neural networks, *Rev. D'Intell. Artif.* 34 (2020) 81–88, <https://doi.org/10.18280/ria.340111>.
- [24] S Parthasarathy, I Tashev, Convolutional neural network techniques for speech emotion recognition, in: *Proceedings of the 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2018, pp. 121–125, <https://doi.org/10.1109/IWAENC.2018.8521333>.
- [25] M. Coşkun, A. Uçar, Ö. Yıldırım, Y. Demir, Face recognition based on convolutional neural network, *2017 International Conference on Modern Electrical and Energy Systems* (2017) 376–379, <https://doi.org/10.1109/MEES.2017.8248937>.
- [26] L. Blanger, A.R. Panisson, A face recognition library using convolutional neural networks, *Int. J. Eng. Res. Sci. Technol.* 3 (2017) 84–92, <https://doi.org/10.25125/engineering-journal-ijsr-aug-2017-25>.
- [27] S.S. Yadav, S.M. Jadhav, Deep convolutional neural network based medical image classification for disease diagnosis, *J. Big Data* 6 (2019), <https://doi.org/10.1186/s40537-019-0276-2>.
- [28] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, M. Adam, A. Gertych, et al., A deep convolutional neural network model to classify heartbeats, *Comput. Biol. Med.* 89 (2017) 389–396, <https://doi.org/10.1016/j.compbiomed.2017.08.022>.
- [29] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, S. Mougiakakou, Lung pattern classification for interstitial lung diseases using a deep convolutional neural network, *IEEE Trans. Med. Imaging* 35 (2016) 1207–1216, <https://doi.org/10.1109/tmi.2016.2535865>.
- [30] C. Wachinger, M. Reuter, T. Klein, DeepNAT: deep convolutional neural network for segmenting neuroanatomy, *Neuroimage* 170 (2018) 434–445, <https://doi.org/10.1016/j.neuroimage.2017.02.035>.
- [31] T. Guo, J. Dong, H. Li, Y. Gao, Simple convolutional neural network on image classification, in: *IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, 2017, pp. 721–724, <https://doi.org/10.1109/ICBDA.2017.8078730>.
- [32] M.E. Paoletti, J.M. Haut, J. Plaza, A. Plaza, A new deep convolutional neural network for fast hyperspectral image classification, *ISPRS J. Photogramm. Remote Sens.* 145 (2018) 120–147, <https://doi.org/10.1016/j.isprsjprs.2017.11.021>.
- [33] F. Sultana, A. Sufian, P. Dutta, Advancements in image classification using convolutional neural network, in: *Proceedings of the 4th International Conference on Research in Computational Intelligence and Communication Networks (ICRCIN)*, 2018, pp. 122–129, <https://doi.org/10.1109/ICRCIN.2018.8718718>.
- [34] M.J. Afridi, A. Ross, E.M. Shapiro, On automated source selection for transfer learning in convolutional neural networks, *Pattern Recognit.* 73 (2018) 65–75, <https://doi.org/10.1016/j.patcog.2017.07.019>.
- [35] L. Shuo, J.O. Nyagilo, D.P. Dave, J.X. Gao, Continuous wavelet transform based partial least squares regression for quantitative analysis of Raman spectrum, *IEEE Trans. Nanobiosci.* 12 (2013) 214–221, <https://doi.org/10.1109/tnb.2013.2272888>.
- [36] P.M. Ramos, I. Ruisánchez, Noise and background removal in Raman spectra of ancient pigments using wavelet transform, *J. Raman Spectrosc.* 36 (2005) 848–856, <https://doi.org/10.1002/jrs.1370>.
- [37] G. Cooper, M. Kubik, K. Kubik, Wavelet based Raman spectra comparison, *Chemom. Intell. Lab. Syst.* 107 (2011) 65–68, <https://doi.org/10.1016/j.chemolab.2011.01.010>.
- [38] T.Y. Huang, J.C.C. Yu, Development of crime scene intelligence using a hand-held Raman spectrometer and transfer learning, *Anal. Chem.* 93 (2021) 8889–8896, <https://doi.org/10.1021/acs.analchem.1c01099>.
- [39] I. Daubechies, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, 1992.
- [40] J.M. Lilly, S.C. Olhede, Generalized morse wavelets as a superfamily of analytic wavelets, *IEEE Trans. Signal Process.* 60 (2012) 6036–6041, <https://doi.org/10.1109/tsp.2012.2210890>.
- [41] J.M. Lilly, S.C. Olhede, Higher-order properties of analytic wavelets, *IEEE Trans. Signal Process.* 57 (2009) 146–160, <https://doi.org/10.1109/tsp.2008.2007607>.
- [42] Lilly, J. M. (2019), jLab: A data analysis package for Matlab, v. 1.6.6, <http://www.jmlilly.net/jmlsoft.html> (accessed March 19, 2021).
- [43] J.M. Lilly, S.C. Olhede, On the analytic wavelet transform, *IEEE Trans. Inf. Theory* 56 (2010) 4135–4156, <https://doi.org/10.1109/TIT.2010.2050935>.
- [44] K. Thenmozhi, U. Reddy, Srinivasulu, Crop pest classification based on deep convolutional neural network and transfer learning, *Comput. Electron. Agric.* 164 (2019), 104906, <https://doi.org/10.1016/j.compag.2019.104906>.
- [45] L.A. Reisner, A. Cao, A.K. Pandya, An integrated software system for processing, analyzing, and classifying Raman spectra, *Chemom. Intell. Lab. Syst.* 105 (2011) 83–90, <https://doi.org/10.1016/j.chemolab.2010.09.011>.
- [46] C. Militello, L. Rundo, S. Vitabile, V. Conti, Fingerprint classification based on deep learning approaches: experimental findings and comparisons, *Symmetry* 13 (2021) 750, <https://doi.org/10.3390/sym13050750>.
- [47] K.M. Tan, I. Barman, N.C. Dingari, G.P. Singh, T.F. Chia, W.L. Tok, Toward the development of Raman spectroscopy as a nonperturbative online monitoring tool for gasoline adulteration, *Anal. Chem.* 85 (2013) 1846–1851, <https://doi.org/10.1021/ac3032349>.



Ting-Yu Huang is a doctoral student with the Department of Forensic Science in the College of Criminal Justice at Sam Houston State University, TX, USA. Ting-Yu received her BA and MA, both in Fire Science at Central Police University, Taiwan, in 2009 and 2015. She was awarded Taiwan Government Fellowship for Studying Abroad in 2019. Ting-Yu is working on a novel AI system for ignitable liquid and fire debris analysis.



Qingzhong Liu is an associate professor of the Department of Computer Science at Sam Houston State University (SHSU), TX, USA, and he has been approved for professor promotion by SHSU. Dr. Liu's research interests include information assurance, digital forensics, multimedia forensics, bioinformatics, machine intelligence, and computational applications. His study has been funded by the US Army, NSF, NIJ, and NIH. In 2020, he was selected among the World's top 2% scientists in Artificial Intelligence and Image Processing. <https://orcid.org/0000-0002-2006-5413>



Jianzhong Wang is a professor with the Department of Mathematics and Statistics at Sam Houston State University, TX, USA. Dr. Wang's research interests and expertise are in (1) wavelet analysis and their application; (2) mathematical methods in image science; (3) high-dimensional data analysis and dimensionality reduction.



Jorn Yu is a professor of Forensic Science with the College of Criminal Justice at Sam Houston State University, TX, USA. Dr. Yu is certified by the American Board of Criminalistics in comprehensive criminalistics (ABC-CC). He is also a Fellow with the American Academy of Forensic Sciences. Dr. Yu is interested in the chemical analysis of trace evidence and crime scene investigation. The goal in his research laboratory is to develop investigative intelligence by chemical forensics. <https://orcid.org/0000-0002-4899-7909>