Document Title:	Deep Learning to Enhance Investigative Lead Information for Automotive Clearcoats
Author(s):	Barry K. Lavine
Document Number:	311114
Date Received:	January 2026
Award Number:	15PNIJ-21-GG-04182-RESS

Final Research Report

Award Number: 15PNIJ-21-GG-04182-RESS

Project Title: Deep Learning to Enhance Investigative Lead Information for Automotive Clearcoats

Project Director/Principal Investigator (PD/PI)
Barry K. Lavine, Regents Professor of Chemistry

Contact Information
bklab@chem.okstate.edu
Department of Chemistry, Oklahoma State University, Stillwater, OK 74078
405-744-5945

Award Recipient Organization:
Oklahoma State University
203 Whitehurst
Stillwater, OK 74078-1016

Project Period
1/01/22 - 12/31/24

Award Amount
$194,000

Major Goals and Objectives

Modern automotive paint consists of a thin e-coat, primer and color coat layer protected by a thick clear coat layer. All too often, the clear coat is the only layer of automotive paint recovered at the crime scene of a "hit-and-run" where damage to a vehicle and/or injury or death to a pedestrian has occurred. In these cases, directly searching an automotive paint database such as the Royal Canadian Mounted Police paint data query (PDQ) database using commercial library search algorithm *will generate a large number (thousands) of potential hits* because of the large number of similar spectra. The inability of Fourier transform infrared (FTIR) spectroscopy and the PDQ database to identify the manufacturer and model of the vehicle from an automotive clear coat sample is a limitation in the use of FTIR spectroscopy and forensic automotive paint databases such as PDQ.

Lavine and coworkers [1-6] have previously demonstrated that pattern recognition when applied directly to mid-infrared absorbance spectra of original equipment manufacturer (OEM) clear coats has the potential to differentiate between similar clear coat IR spectra. The prototype pattern recognition-based infrared (IR) library search system developed by Lavine for OEM clear coats consisted of two separate but interrelated components: search prefilters to reduce the size of the library of a specific manufacturer to an assembly plant or assembly plants corresponding to the unknown paint sample and a cross-correlation library searching algorithm to identify IR spectra most like the unknown in the subset of spectra identified by the search prefilters. This approach has been shown to be successful if information about the automotive manufacturer of the OEM clear coat is provided.

To obtain information about the vehicle manufacturer from an IR spectrum of an automotive clear coat, deep learning as implemented using a four-layer artificial neural network has been investigated. Specifically, we seek to determine the vehicle manufacturer (e.g., General

Motors, Chrysler, Ford, Toyota, Nissan, and Honda) from the IR spectrum of an automotive clear coat. It has been purported that deep learning requires less data preprocessing. Raw data can (in principle) be pipelined directly to the neural network allowing it to learn patterns directly from the data for successful recognition. Selecting wavenumbers of interest or removing wavenumber regions, for example, is (in principle) not necessary as the neural network allows the data to identify key relationships within the data that are crucial for a successful classification. As shown by the results of this investigation, raw data should not be pipelined directly to the neural network and removing wavenumber regions from the spectra can improve the generalization performance of the model.

Project Design and Methods

Clear Coat Infrared Spectra Data Cohort

The 2796 mid-IR spectra of the automotive clear coats provided by the RCMP for our in-house spectral library were obtained from street samples and factor panels. Four different FTIR spectrometers were used to collect the clear coat spectra: Bio-Rad 40A, Bio-Rad 60A, and two Thermo Nicolet 6700 FTIR spectrometers. All four FTIR spectrometers were run at 4 cm^{-1} resolution. A Harrick 4X beam condenser was used in both Bio-Rad instruments whereas a Harrick 6X beam condenser was used in the two Thermo-Nicolet instruments. Each spectrometer was equipped with a DTGS detector. All clear coat samples were between 3 and 4 micrograms and were run using a high-pressure transmission diamond anvil cell [7, 8]. The thickness of the sample (i.e. the clear coat peel) and the pressure applied by the diamond anvil cell in collecting the spectra were such that a transmittance between 7% and 14% was obtained for the carbonyl stretching band (around 1730 cm^{-1}) in all 2796 clear coat FTIR spectra.

The number of points collected in the wavelength range using the Thermo-Nicolet instrument (4000 to 400 cm^{-1}) varied from 1878 to 1958 points, whereas the FTIR spectra collected on the two Bio-Rad instruments for the same wavelength range and resolution were represented by 1944 points. To address the problem of spectral alignment, each IR spectrum was normalized to the helium neon laser frequency of 15798.0 cm^{-1}. The laser frequency value was set to that measured at the aperture setting, which makes the sample peak positions independent of aperture setting. However, this results in a change in the data point spacing and the location of the data points. Although the default laser frequency of the spectrometer is 15798.3 cm^{-1}, 15798.0 cm^{-1} was used because it solved problems that occurred when importing the spectra from GRAMS. This also ensured proper spectral alignment along the wavelength axis for the imported Bio-Rad spectra to the Thermo-Nicolet instrument. After performing this normalization, the Bio-Rad instruments were comparable to FTIR spectra collected on the two Thermo-Nicolet instruments. To authenticate proper spectral alignment, FTIR spectra of known samples measured on both the Bio-Rad and Thermo-Nicolet instruments were compared using vector subtraction before and after performing the frequency normalization procedure. Subtraction yielded a nonzero response at each wavelength before frequency normalization but zero at each point after normalization, indicating that spectral alignment had been achieved. After frequency normalization, each FTIR spectrum (4000 to 400 cm^{-1}) was represented by 1869 points.

Many clear coat samples evaluated in this study were from the same manufacturer within a limited production year range (2000 – 2010). This makes the comparisons among the FTIR spectra in the database more challenging and tests the limits of the proposed neural network methodology to discern subtle but significant differences in the FTIR spectrum of automotive clear coats. Although most of the clear coats in our in-house spectral library were acrylic melamine styrene or acrylic melamine styrene polyurethane, there were other paint chemistries represented in the data cohort, e.g., acrylic melamine, and acrylic styrene polyurethane. The clear coat samples analyzed were obtained from metallic automotive substrates as automotive paint samples for plastic substrates are often applied at the plant that manufactures the component, not at the plant where the vehicle is assembled.

Data Analysis

The 2796 mid-IR clear coat spectra were divided into a training set of 2237 spectra and a prediction set of 559 spectra. Clear coat IR spectra comprising the prediction set were selected by random lot. To evaluate the performance of the neural network models and set the meta parameters of the neural network, ten-fold cross validation was performed. The 2237 spectra were partitioned into ten equal sized sets. The model would be trained on eight of the partitions with the nineth partition serving as a validation set (to select meta parameters for the model and to decide when to stop the training). The tenth partition was used to evaluate the model (test partition). This process was repeated ten times, each time using a different fold as a validation set or test set. Thus, ten models are created and trained independently, each using a different partition for evaluation. Finally, the results from each test set are averaged to produce a robust estimate of the model's performance.

For this problem, the convolutional layer was simplified as a stack of 1D filters. Individual network models used weight regularization, activity regularization and dropout layers to avoid overfitting and improve the generalization of the network model without causing a reduction in training. L1 regularization of activities in the earlier layers was found to be beneficial as well as using rectified linear activation functions for the output of each layer except for the last layer where soft max was used.

Each clear coat IR spectrum in the training set was baseline corrected using the rubber-band method [9] and normalized to unit length. For outlier analysis, the generalized distance test [10] as implemented by SCOUT [11] was applied to each class (vehicle manufacturer) in the training set to identify and delete anomalous samples.

Project Results and Findings

Initial Studies. Our initial studies focused on developing models to discriminate clear coat IR spectra by manufacturer using 1-NN, decision trees, three-layer artificial neural network, artificial neural network with four deep layers and convolutional artificial neural network with four deep layers. For this comparison, both the convolutional neural network and the artificial neural network with four deep layers were at the same stage of their parameter search. For the artificial neural network with four deep layers and the convolutional artificial neural network, the effects of removing sample outliers and performing baseline correction on the spectra were evaluated. The presence of the convolutional layer in the neural network did not improve its performance. Outliers, however, impacted the performance of the neural network and eliminating them prior to training improved model performance. Baseline correction also improved the classification performance of the network. Deep learning can achieve 100% on the training data so the problem is obtaining good generalization using information/regularization beyond fitting training data for class prediction.

Full and Reduced Spectral Range. Artificial neural network models with 4 deep layers were developed for the full spectral range and for the following segments: 1500 to 600 cm^{-1}, 1844 to 667 cm^{-1}, 1641 to 667 cm^{-1}, and 1641 to 860 cm^{-1}. The neural network model developed using the spectral range 1641 to 667 cm^{-1} yielded the best results (90% for cross validated accuracy and 88.3% for the accuracy of the prediction of the 559 spectra comprising the external test set). For this comparison, which involved 80,000 iterations, all artificial neural network models were at the same stage of their parameter search.

Two-Way Classifications Using Ten-Fold Cross Validation. Classification studies were performed comparing the six vehicle manufactures pairwise to better understand the challenges inherent in this pattern recognition problem. The mean value for 10-fold cross validation was used to assess accuracy for each pairwise comparison. From an examination of the results for these pairwise comparisons, it is evident that differentiating General Motors from Ford and differentiating Chrysler from Ford are the most difficult of the classifications.

Infrared Spectra of Clear Coats. The clear coat layer, like the color coat and primer-surfacer and e-coat layers, has features in its IR spectrum unique to the vehicle manufacturer. Using neural networks, fingerprint patterns in the IR spectra of clear coats characteristic of the vehicle manufacturer could be identified. Such information can serve to quantify the discrimination power of OEM clear coat paint samples encountered in actual case work.

Applicability of Research

The research project described in this final summary overview is directly targeted to the development of new approaches for the forensic examination of automotive clear coat, both at the investigative lead stage and at the court room testimony stage. Direct impact on over 57 local, state, and federal forensic laboratories in the United States that are currently using the PDQ database is anticipated. There will also be impact on international forensic laboratories using the PDQ database including the Forensic Laboratory Services Division of the RCMP, the Centre of Forensic Science in Toronto, Canada, the ENFSI network of European forensic science institutes, the Australian Police Services, and the New Zealand Public Services.

Participants and Other Collaborating Organizations

Members of the Oklahoma State University (OSU) team that participated in this study include Dr. Douglas Heisterkamp (Associate Professor of Computer Science at OSU), Dr. Collin G. White (postdoctoral researcher), Elizabeth Donkor (Graduate Research Assistant) and Chamika Eranga Liyanarachchi (Graduate Research Assistant). Our outside partner has been The Oklahoma State Bureau of Investigation.

Publications

1. B. K. Lavine, K. S. Booksh, S. L. Neal, C. G. White, and D. R. Heisterkamp, "A Perspective on Chemical Data Science," Sensors, In Preparation.
2. B. K. Lavine, C. G. White, and D. R. Heisterkamp, "Deep Leaning to Enhance Investigative Information from Automotive Clear Coats," in Chemometric Analysis and Machine Learning, Edited by Harvey Hou and Peter He, Springer. In Preparation
3. B. K. Lavine, C. G. White, and D. R. Heisterkamp, Artificial Neural Networks for Discrimination of Automotive Clearcoats by Vehicle Manufacturer," Forensic Chemistry, in preparation.

Software

The Keras (neural network) models developed in this project and the data used for training as Tensor-Flow data-streams are housed in a repository at www.box.com. For an individual to access box.com, it is necessary for the user to create an account and then contact Barry K. Lavine (barry.lavine@okstate.edu) who then will grant access to the user. It will be necessary for the user to provide Lavine with the email address associated with the user's box account. A fully functioning Windows executable version of Keras software is available to the public. The executable includes a Python implementation of the rubber-band baseline correction algorithm that will allow the user to stack the baseline with the original unknown spectrum and send it to the desired Keras model.

Dissemination Activities

1. B. K. Lavine, C. G. White, and D. R. Heisterkamp, "Convolutional Neural Networks to Enhance Investigative Lead Information from Automotive Clear Coats," SCIX 2024, Raleigh, NC, 10/22/2024.
2. B. K. Lavine, C. G. White, and D. R. Heisterkamp, "Deep Learning to Enhance Investigative Lead Information from Automotive Clear Coats," Eastern Analytical Symposium, Plainsboro, NJ, 11/20/2024.
3. B. K. Lavine, C. G. White, and D. R. Heisterkamp, "Artificial Neural Networks to Enhance Investigative Lead Information from Automotive Clear Coats," Research Frontier Symposium, Alabama State University, 3/13/2025.

References

1. Barry K. Lavine, Nikhil Mirjankar, Scott Ryland, and Mark Sandercock, "Wavelets and Genetic Algorithms Applied to Search Prefilters for Spectral Library Matching in Forensics," Talanta**,** 2011, 87, 46-52.
2. B. K. Lavine, A. Fasasi, N. Mirjankar, and M. Sandercock, "Development of Search Prefilters for Infrared Library Searching of Clear Coat Paint Smears," Talanta, 2014, 119, 331–340.
3. B. K. Lavine, A. Fasasi, N. Mirjankar, and M. Sandercock, "Development of Search Prefilters for Infrared Library Searching of Clear Coat Paint Smears," Talanta, 2014, 119,

331–340.

4. B. K. Lavine, A. Fasasi, N. Mirjankar, and C. White, "Search Prefilters for Library Matching of Infrared Spectra in the PDQ Database using the Autocorrelation Transformation," Microchem. J., 2014, 113, 30–35.
5. B. K. Lavine, A. Fasasi, N. Mirjankar, K. Nishikida, and J. Campbell, "Simulation of Attenuated Total Reflection Infrared Absorbance Spectra – Applications to Forensic Analysis of Automotive Clearcoats," Appl. Spectros., 2014, 68(5), 608-615.
6. M. Maric, W. van Bronswijk, K. Pitts, and S. W. Lewis, "Characterization and Classification of Automotive Clear Coats for Forensic Purposes," J. Raman Spec., 47(8), 948-955.
7. F. T. Tweed, R. Cameron, J. S. Deak, P. D. Rodgers, "The Forensic Microanalysis of Paints, Plastics, and Other Materials by an Infrared Diamond Cell Technique," Forensic Sci., 1974, 4, 211-218.
8. P. G. Rodgers, R. Cameron, N. S. Cartwright, W. H. Clark, J. S. Deak, and E. W. Norman, "The Classification of Automotive Paint by Diamond Window Infrared Spectrophotometry, Part III: Case Histories," Can. Soc. Forens. Sci. J. 1976, 9(4), 103-111.
9. X. Shen, S. Ye, L. Xu, R. Hu, L. Jin, H. Xu, J. Liu, and W. Liu, "Study on Baseline Correction Methods for the Fourier Transform Infrared Spectra with Different Signal-to-Noise Ratios," Appl. Optics, 2018, 57(20), 5794-5799.
10. S. J. Schwager, and B. H. Margolin, "Detection of Multivariate Normal Outliers," Annals Stat. 1982, 10, 943-953.
11. Martin A. Stapanian, Forest C. Garner, Kirk E. Fitzgerald, George T. Flatman, John M. Nocerino, J. Chemom., 1993, 7, 165-176