Toward Practical Augmentation of Raman Spectra for Deep Learning Classification of Contamination in HDD

Seksan Laitrakun[1], Somrudee Deepaisarn[1]*, Sarun Gulyanon[2], Chayud Srisumarnk[1], Nattapol Chiewnawintawat[1], Angkoon Angkoonsawaengsuk[1], Pakorn Opaprakasit[1], Jirawan Jindakaew[1], and Narisara Jaikaew[1]

[1]Sirindhorn International Institute of Technology, Thammasat University, Pathum Thani 12120, Thailand
[2]College of Interdisciplinary Studies, Thammasat University, Pathum Thani 12120, Thailand

Abstract

Deep learning techniques provide powerful solutions to several pattern-recognition problems, including Raman spectral classification. However, these networks require large amounts of labeled data to perform well. Labeled data, which are typically obtained in a laboratory, can potentially be alleviated by data augmentation. This study investigated various data augmentation techniques and applied multiple deep learning methods to Raman spectral classification. Raman spectra yield fingerprint-like information about chemical compositions, but are prone to noise when the particles of the material are small. Five augmentation models were investigated to build robust deep learning classifiers: weighted sums of spectral signals, imitated chemical backgrounds, extended multiplicative signal augmentation, and generated Gaussian and Poisson-distributed noise. We compared the performance of nine state-of-the-art convolutional neural networks with all the augmentation techniques. The LeNet5 models with background noise augmentation yielded the highest accuracy when tested on real-world Raman spectral classification at 88.33% accuracy. A class activation map of the model was generated to provide a qualitative observation of the results.

Index Terms: Raman Spectral Classification, Data Augmentation, Convolutional Neural Networks (CNN), Hard Disk Drive, Small Particle

I. INTRODUCTION

Hard disk drives (HDD) are essential for data storage in computer systems. To achieve a high areal density, the flying height at the head-disk interface (HDI) is minimized to the order of nanometers [1]. At this scale, even small contaminants can enter the HDI and potentially cause HDD failure. There are two main sources of contamination [2]: inside the HDD itself, because of cracked particles in the disk, and in the HDD assembly production line. In this work, we are interested in identifying the contaminating particles from the HDD assembly production line so that we can specify the contamination stage and prevent it from recurring.

Spectroscopic techniques are commonly used to study and investigate the characteristics of contaminating particles in HDD. Examples of these techniques are time-of-flight-secondary ion mass spectrometry (TOF-SIMS), Fourier transform infrared (FT-IR) spectroscopy, X-ray photoelectron spectroscopy (XPS), energy-dispersive X-ray spectroscopy (EDS), and Raman spectroscopy [3, 4].

Raman spectroscopy can be used to identify the material types. This principle is based on measuring the molecular vibrational energy stages, which reveal the unique characteristics of the testing materials. A contaminated particle can be

represented by a Raman spectrum, which is a plot of signal intensity against wavenumber. Because of the unique spectral pattern of each sample (*i.e.* locations of a set of peaks), a human expert can identify the corresponding material based on the observed Raman spectrum pattern. However, the capability to analyze Raman spectra is limited by the signal-to-noise ratio, which is influenced by various parameters such as fluorescence noise, chemical noise, electrical noise, detector limitation, thermal noise, and other environmental causes. Noise appears in the spectra as an unwanted background superimposed on actual signals, which sometimes causes difficulties in determining the Raman spectral patterns using conventional commercial software.

Several studies have focused on applying deep-learning (DL) models to identify substances of interest using Raman spectra. The studies in [5-8] showed that DL models can automatically extract useful features and outperform conventional methods based on machine learning, such as K-nearest neighbors. However, DL methods require a large amount of labeled data to train models. Obtaining numerous Raman spectra is time-consuming and tedious. Data augmentation provides reasonable solutions to boost the size of the training data by simulating samples with some added variations to the original samples, which helps improve the robustness and generalization of the models.

In this study, we examine the impact of noise augmentation in terms of the performance and robustness of Raman spectral classification, such that the material identification of contaminants present in HDD is improved. The following five types of augmentation were examined: (1) weighted sums of the spectral signals; (2) imitated chemical backgrounds, *i.e.*, influenced by substrates and air; (3) extended multiplicative signal augmentation (EMSA), mimicking noise from physical variations related to scattering and instrumental effects; (4) generated Gaussian-distributed noise; and (5) generated Poisson-distributed noise. Nine state-of-the-art convolutional neural networks (CNN) were trained on the augmented data, and their performances were evaluated by classifying a test dataset of the measured noisy spectra. Our contributions include (a) an empirical finding that measures the effectiveness of increasing the performance and robustness of DL models using data augmentation based on computed noise, and (b) an analysis of noise characteristics for Raman spectra of small particles through computed noise, where the models are validated on real-world noisy spectra to suggest a practical application of spectral augmentation. Part of this work was presented in our previous conference paper [9], in which a preliminary comparative study of noise augmentation and DL techniques was performed on a small number of ideally clean spectra.

II. LITERATURE REVIEW

DL approaches have shown potential in computer vision and natural language processing [10]. One of the most popular models is a convolutional neural network (CNN), which comprises a series of convolutional layers for feature extraction, followed by layers for classification. Notable DL models include LeNet-5 [11], AlexNet [12], VGG16 [13], GoogLeNet [14], ResNet [15], SqueezeNet [16], Xception [17], DenseNet [18], and MobileNet [19]. Several state-of-the-art CNN models have been applied to Raman spectral classification in many application areas. Ho et al. [20] proposed a CNN based on ResNet to classify 30 bacterial pathogens using Raman spectra. The proposed CNN model yielded an accuracy of 82%, outperforming the baseline models (*i.e.,* logistic regression and support vector machine). Chen et al. [21] analyzed serum Raman spectra to classify patients as having no cancer, lung cancer, or glioma. The authors investigated and compared four neural networks: a multilayer perceptron, a simple recursive neural network, a simple CNN, and AlexNet. Zhang et al. [22] proposed DL-based methods to identify patients with membranous nephropathy using the Raman spectra of serum, urine, and DL models. Among the investigated models (AlexNet, GoogLNet, and ResNet), AlexNet yields the best accuracy. Chang et al. [23] collected the Raman spectra of oral cancer tissues and normal oral tissues and applied five DL models (AlexNet, VGGNet, ResNet50, MobileNetV2, Transformer) to classify them. The results showed that ResNet50 outperformed the other networks.

DL is a technique that requires a large amount of data to build models of the relationships between input and output pairs. Data augmentation was necessary to increase the number of samples in the training dataset by adding useful variations to the collected samples. Several methods have been introduced and applied, such as adding Gaussian noise [21], using an autoencoder [24], and using a generative adversarial network [25]. In this study, we investigated five types of data augmentation and compared them based on the classification performances of nine CNN models.

III. MATERIALS AND METHODS

In this study, we investigated five augmentation methods to boost the number of training spectra by comparing the classification performance of various CNN models. The framework is illustrated in Fig. 1. Considering 10 substances (*i.e.,* sample classes), we collected 50 spectra for each substance. Because these spectra were subject to fluorescence
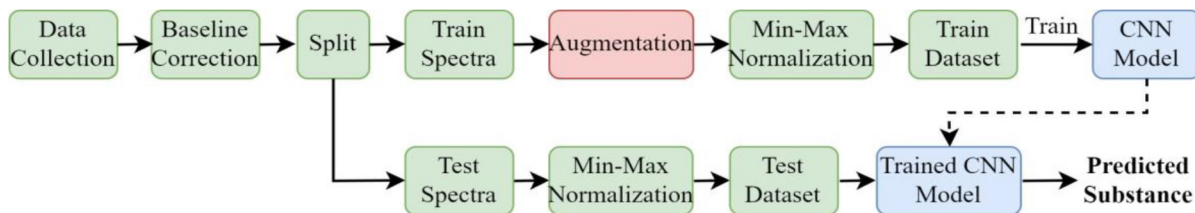
Fig. 1. Framework of this study. Training and test spectra go through the same preprocessing steps (green) except for the augmentation (red) which is performed only on the training data before they are fed to the deep learning models (blue).

noise, a baseline correction was performed. The set of 50 spectra for each substance was divided into training and test sets. In this study, 20 spectra per substance were assigned to the training set, and the other 30 spectra per substance were assigned to the test set. As the size of the spectra in the training set was small (200 spectra in total), it was not sufficient to train a model. Therefore, augmentation methods were used to create 200 synthetic spectra for each class. Thereafter, the entire synthetic dataset consisting of 2,000 spectra was used to train the model. The trained model was used to classify an independent test dataset consisting of 300 spectra (30 spectra per class). More details are provided below.

A. Data Collection

All computational experiments in this study were performed using a dataset containing 500 noisy spectra of substances that presumably contaminated the HDD during fabrication. Fig. 2 shows the process of Raman spectral acquisition from a small contaminating particle in the HDD with a schematic diagram. Unfortunately, the smaller the particle, the more challenging the Raman spectrum. In this study, the samples were prepared in the form of micro- to nano-sized particles suspended in a suitable solvent to obtain noisy spectral signals that imitate the contamination of HDD. The dispersion was then dropped onto the HDD. After solvent evaporation, the particles were deposited on the surface of the HDD to acquire the Raman spectra. To generate noisy Raman spectra, a laser was fired at the rims of the particles. This helped achieve small-sized contamination-like spectra, which generally varied from micron to submicron levels. Clean spectra were acquired for larger particles.

Ten common contaminants were observed in HDD product lines. These contaminations can cause disk failure if present in sensitive parts such as the magnetic head. The ten classes of selected samples were cellulose, polycarbonate (PC), low-density polyethylene (LDPE), high-density polyethylene (HDPE), polyethylene terephthalate (PET), polyvinyl chloride (PVC), polytetrafluoroethylene (PTFE), polyoxymethylene (POM), polyether ether ketone (PEEK), and polypropylene (PP). Each class contained 50 spectra.
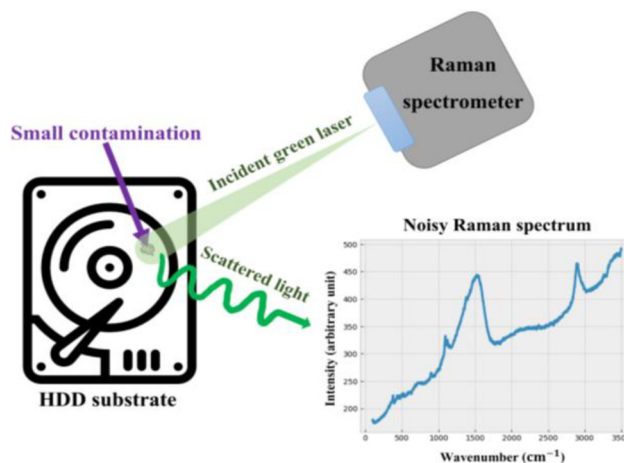

Fig. 2. Schematic diagram indicating the process of Raman spectral acquisition from a small contamination particle in HDD. Raman spectra can be used to characterize and identify the particle types. Due to the relatively small-sized particles compared to the laser beam, the acquired Raman spectrum is noisy and difficult to identify.

The modality used for the spectral acquisition was the Raman spectrometer (Model: DXR, Thermo Fisher Scientific Inc., USA), which is equipped with the 532 nm laser (visible green light). Appropriate calibration was performed prior to acquisition. The laser power was set at a constant value of 10 mW. The acquisition range was 99-3500 cm^{-1}. A spectrum was generated from accumulated signals of 100 cycles using a 50 μm pinhole. Photobleaching was performed for up to 3 min.

B. Baseline Correction and Data Splitting

We applied an improved modified polynomial (IModPoly) baseline algorithm [26] to remedy fluorescence noise. The baseline-corrected spectra of each substance are shown in Fig. 3. Thereafter, for each substance, 20 spectra were selected to create a synthetic training dataset, and the remaining 30 spectra were retained in the test set. Specifically, the test dataset consisted of 300 measured noisy spectra, with each class containing 30 spectra.
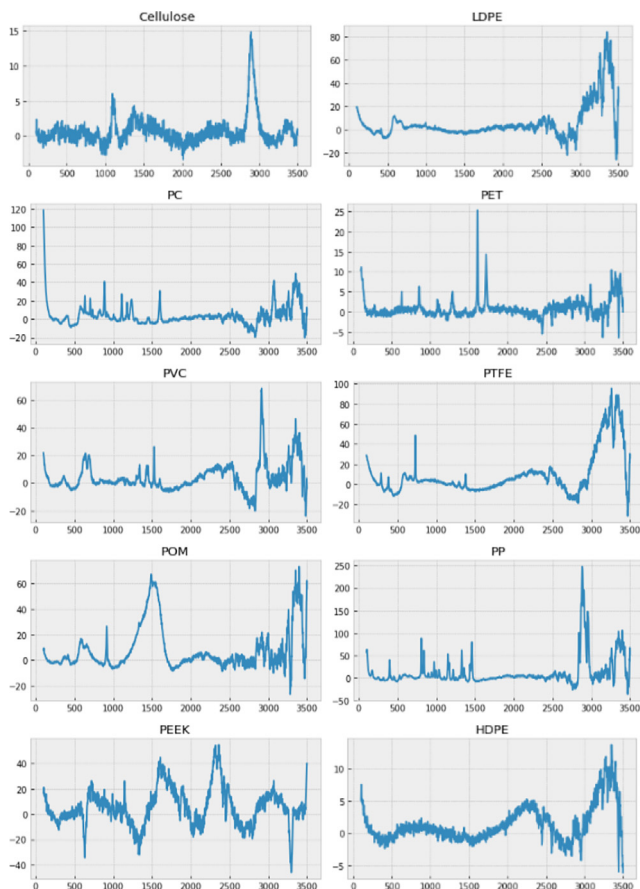
Fig. 3. Spectra of 10 substances after baseline correction. The horizontal and vertical axes represent wavenumber in cm^{-1} and signal intensity, respectively.

C. Data Augmentation

In this study, we consider a situation in which the number of measured noisy spectra is limited, which is insufficient for effective model training. As previously mentioned, 20 noisy spectra measured per class were selected to create a synthetic training dataset. Our goal was to obtain 200 synthetic spectra per class (2,000 synthetic spectra in total). Therefore, we synthesized 10 spectra from every spectrum in the training set using different augmentation methods and compared the results of each method.

We investigated five spectral augmentation techniques: weighted sums of spectral signals imitated chemical backgrounds from substrates and air, extended multiplicative signal augmentation (EMSA) [27], generated Gaussian-distributed noise, and generated Poisson-distributed noise. We determined a data augmentation approach that performed well in classifying noisy Raman spectra against a baseline approach of generating randomly weighted sums of signal intensity.

1) Weighted-Sum Augmentation

Using this method, a synthetic spectrum X was generated by the weighted sum of 20 measured spectra: $X = \sum_{i=1}^{20} w_i S_i$ where S_i is a measured spectrum and w_i is a weight value between 0 and 1.

2) Background Noise Augmentation

In this study, we were interested in the contaminations found on the following substrates in the hard disk head: aluminum oxide (Al_2O_3), aluminum oxide coated with diamond-like carbon (Al_2O_3 + DLC), aluminum oxide-titanium carbide (Al_2O_3 + TiC), and aluminum oxide-titanium carbide coated with diamond-like carbon (Al_2O_3 + TiC + DLC). As a result, the measured Raman spectrum of a substance can be affected by substrate and air noise. We recorded 10 spectra for each substrate and air noise. Based on the background noise augmentation, each synthetic spectrum was generated by superimposing the measured, substrate, and air noise spectra.

3) Extended Multiplicative Signal Augmentation

Extended multiplicative signal augmentation (EMSA) is based on the theoretical concepts underlying extended multiplicative signal correction (EMSC), which is a baseline correction algorithm that is applied to Raman and infrared spectral data. The EMSC eliminates unwanted backgrounds, where modes of variation due to chemical, instrumental, and physical background signals can be modeled with a mathematical expression. EMSA uses knowledge from the extracted background noise to suggest spectral augmentation by adding known captured noise variations [27]. To vary the distortion, Gaussian-distributed random numbers with zero mean and varying standard deviations were applied to alter the parameter coefficients in the EMSA model, resulting in various augmented spectra [26].

4) Statistical Noise Augmentation

There are two popular statistical noise models for simulating the behaviors of natural noise that occur in instrumental data acquisition processes [28]: Gaussian and Poisson noise. The Gaussian noise model represents environmental and electrical noises. Based on a Gaussian noise model, random variations drawn from the same statistical distribution were added to the measured spectrum. By contrast, the Poisson noise model denotes signal-dependent shot noise. Random Poisson noise can be simulated by calculating the square root of signal intensity multiplied by a Gaussian random number [29,30].

D. Min-Max Normalization

The range of the signal intensities (either synthetic or measured) varied from one spectrum to another. Min-max nor-

malization was applied to scale the signals into a common range. In this study, we scaled all spectra using min-max normalization over the training and test datasets. Let X be a spectrum and $X = [x_1, x_2, \cdots, x_k]$ The normalized spectrum $\overline{X} = [\overline{x}_1, \overline{x}_2, \cdots, \overline{x}_k]$ can be computed from $\overline{x}_k = (x_k - x_{\min})/(x_{\max} - x_{\min})$, where $x_{\max}$ and $x_{\min}$ are the maximum and minimum of spectral intensity values, $x_1, x_2, \cdots, x_k$, respectively.

E. Convolutional Neural Networks

We investigated and compared the augmentation methods based on the performance of nine state-of-the-art CNN models: LeNet5 [11], AlexNet [12], VGG16 [13], GoogLeNet [14], ResNet [15], SqueezeNet [16], Xception [17], DenseNet [18], and MobileNet [19]. As they were proposed for image processing, their structures originally consisted of two-dimensional (2D) convolutional and 2D pooling layers. In spectral classification applications, we implemented one-dimensional (1D) versions of these CNN models by replacing any 2D layers with the corresponding 1D layers. Hyper-parameters, such as the number of filters and filter sizes, were kept the same as in the original models.

The structure of the CNN model used in the spectral classification is shown in Fig. 4. It consists of a feature extractor and classifier. A CNN model was applied to extract a set of features (*i.e.*, the output of the flattened layer). These features were subsequently entered into the classifier to predict the corresponding sample classes. Each classifier used in this study consists of three dropout layers and three fully connected (FC) layers. The first two FC layers had 512 and 256 neurons, respectively, and used a rectified linear unit (ReLU) as the activation function. The last FC layer, which functions as the output layer, has 10 neurons (equal to the number of

Fig. 4. The structure of the CNN model consists of two parts: feature extraction and classification.

classes) and uses softmax as the activation function. Because each CNN model has a different network structure, number of layers, types of layers, and hyperparameters, a different set of features was extracted. Consequently, the CNN models exhibit different performances.

IV. RESULTS

A. Experimental Setup

We generated five synthetic training datasets from 200 measured noisy spectra using the weighted sum, background noise, EMSA, Gaussian noise, and Poisson noise methods. The last two methods are statistical noise-augmentation methods. Each training dataset consisted of 2,000 synthetic spectra (10 classes, 200 spectra per class). All the trained models were evaluated using the same test dataset containing 300 measured noisy spectra.

The models were trained by minimizing the categorical cross-entropy. The Adam optimizer was used with the following default values: training rate = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-7}$. The batch size and the number of epochs were 32 and 30, respectively.

The classification performances were measured using the accuracy score (%), which was obtained from Equation (1):

$$Acc = \left[\frac{1}{N}\sum_{n=1}^{N}\frac{TP_n + TN_n}{TP_n + FP_n + TN_n + FN_n}\right] \times 100 \quad (1)$$

where N is the number of considered substances (classes); TP_n is the true-positive number; TN_n is the true-negative number; FP_n is the false-negative number; and FN_n is the false-negative number. Subscript n refers to the n-th class.

B. Performance Comparison

This section compares the five augmentation methods. Table 1 lists the accuracy scores of the nine CNN models trained using five synthetic training datasets from the weighted-sum, background noise, EMSA, Gaussian, and Poisson methods, as explained in Section III-C. The signal-to-noise ratios for the Gaussian and Poisson methods were
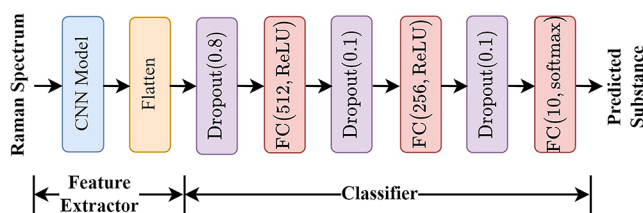
Table 1. Accuracy scores (%) of CNN models trained using synthesis spectra from different augmentation methods. Bold indicates the highest accuracy score for each CNN model

Augmentation	LeNet5	AlexNet	VGG16	GoogLeNet	ResNet	SqueezeNet	Xception	DenseNet	MobileNet
Weighted Sum	48.67	44.67	46.67	52.67	44.00	44.00	59.33	63.33	47.00
Background Noise	**88.33**	**82.33**	**69.33**	**77.33**	**79.00**	**73.67**	**88.00**	**82.33**	**77.00**
EMSA	51.33	45.00	44.33	33.33	21.33	19.33	48.00	52.00	16.33
Gaussian	68.67	54.33	64.33	67.67	50.00	46.33	57.33	67.67	52.00
Poisson	82.67	65.00	63.67	73.67	43.00	63.67	58.67	64.67	59.00

set to 20 dB. Accuracy scores were obtained by classifying the test dataset, which consisted of 300 measured noisy spectra.

In addition, we compared the accuracy scores with those of CNN models built using the weighted-sum training dataset. In this method, each synthetic spectrum was computed using a random weighted sum of 20 measured noisy spectra of the same class. Similarly, 200 synthetic spectra were generated for each class. The weighted-sum method represents the baseline performance because it eliminates the effects of noise and produces more data with fewer variations.

The following findings were obtained. First, unlike the weighted-sum method, the other augmentation methods created training datasets by adding more variations to the original sample spectra. Consequently, they achieved higher accuracy scores than the weighted-sum method in many cases, except for the EMSA. Second, background noise augmentation resulted in the highest accuracy score among the methods of interest. Third, LeNet5 achieved the highest accuracy scores in many cases, even though it was a simple and shallow CNN model. Finally, LeNet5, which was trained using the dataset created using the background noise method, achieved the highest accuracy score of 88.33%. An EMSA provides augmented spectra with known functional variations. Similarly, the Gaussian and Poisson noise methods add known statistical distributions to augmented spectra. In contrast, the background noise method adds practical complexity to the training data. This allowed the classification models to seek and learn from the actual variations within the labeled spectra rather than the underlying backgrounds, making the generated spectra robust for training DL models.

C. Confusion Matrix

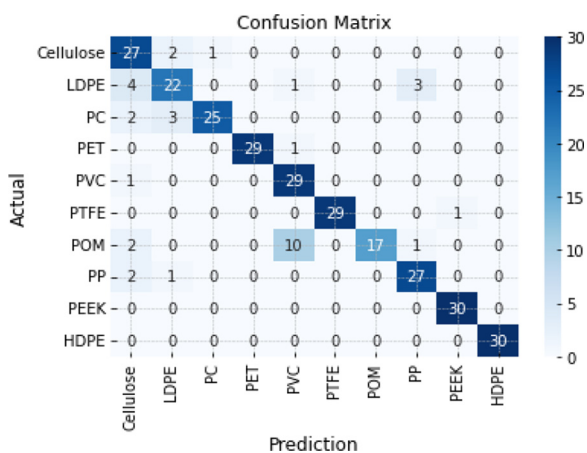We further analyzed the classification performance using a confusion matrix. Fig. 5 shows the confusion matrix according to LeNet5 trained using the dataset created using the background noise method (which offers the best accuracy score, as previously mentioned). The results were based on the classification of the test dataset, which consisted of 30 spectra per substance. The rows represent the actual substances, and the columns denote the predicted substances. The model performed very well in classifying all the substances except POM. Only 17 POM spectra were correctly identified, whereas 10 were misclassified as PVC.

D. Class Activation Map of the Best Model

A class activation map (CAM) [31] was used to understand how a model predicted the output. In spectral classification, CAMs can be applied for building heat maps to show the values and their wavenumbers on which the model focuses when computing the prediction. In Fig. 6, we used HiResCAM [32] to show the CAMs of 10 substances according to LeNet5 trained on the dataset created from the
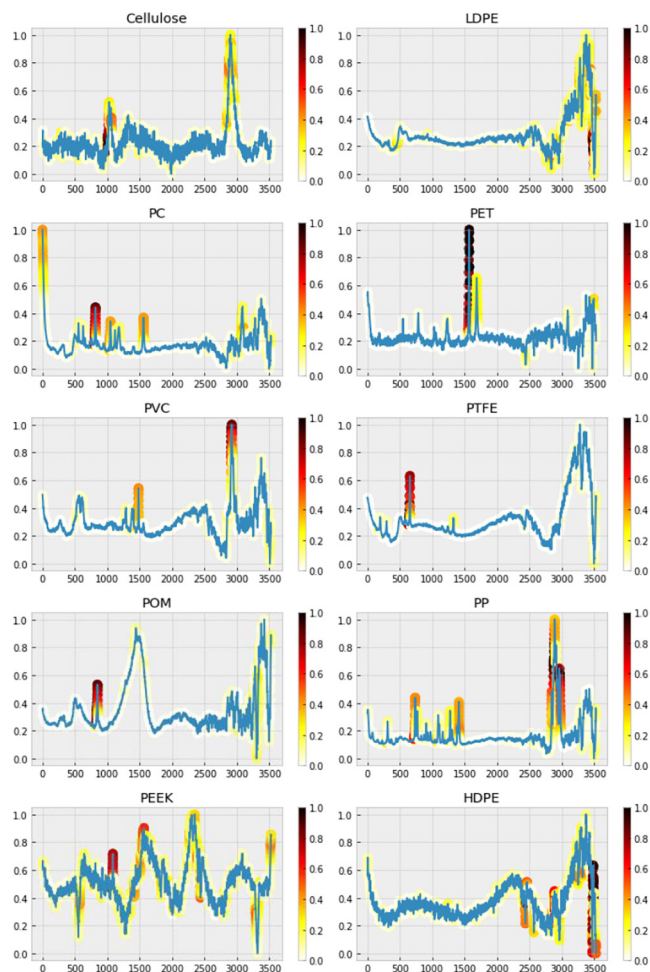

Fig. 6. Class activation maps of LeNet-5 computed by HiResCAM [26] on 10 substances. The horizontal and vertical axes represent wavenumber in cm^{-1} and signal intensity, respectively.

Fig. 5. Confusion matrix according to LeNet5 trained by the augmented dataset created from the background-noise method.

background noise method. Compared with the manner in which human experts recognize spectra, the model clearly focuses on peak patterns to differentiate substances, except for LDPE and HDPE. For LDPE and HDPE, we observed common peak locations with the other substances. Consequently, the model seeks other locations to indicate LDPE and HDPE.

V. DISCUSSION, CONCLUSIONS AND FUTURE WORK

Practical approaches for Raman Spectral data augmentation, specifically to improve the DL classification of small contaminants in HDD, were investigated. Nine deep learning classification models were applied to the data from five different augmentation techniques. The results were examined by comparing the classification performance and robustness of the classifiers, assuming that they resulted from the variations generated during augmentation. Augmented data are required to mimic the real-world behavior of noise in Raman spectra to achieve good classification performance. Consequently, an appropriate noise augmentation model can enhance the performance of classifiers when the amount of original spectral data is insufficient for training the DL models. The background noise augmentation resulted in the highest accuracy scores, and the LeNet5 model revealed the highest accuracy scores in many cases, even with the application of a simple CNN model. LeNet5 combined with the background noise method achieved the highest accuracy score of 88.33%. The CAM of LeNet5 provides evidence of the robustness of the model because the highlighted wavenumbers and peaks match those used by chemists to characterize these substances.

Moreover, the gain in accuracy of DL models, regardless of the model, indicates that background noise augmentation has the potential to characterize the variations of noisy spectra acquired from small particles in real-world practice. This allows classification models to learn actual variations overlaid on noisy backgrounds. Hence, the results demonstrate the application of computed noise and data augmentation in a data-centric approach to artificial intelligence, which allows the training of high-performance DL models, even if the available spectral dataset is small.

This study paves the way for future research. For example, a detailed analysis of the statistical noise behavior underlying Raman spectra can be performed to create better approximations of real spectra and further improve the performance of DL classification. The other is the development of classification pipelines that apply deep learning models for denoising purposes to improve the quality of the spectra before attempting to classify them. The latter may also assist classification models in achieving greater accuracy, whereas clean spectra are informative by-products that can be investigated using various techniques.

REFERENCES

[1] G. Guo, C. Bi, and A. A. Mamun, *Hard Disk Drive: Mechatronics and Control*, FL, Boca Raton: CRC, 2006.

[2] R. Nagarajan, "Survey of cleaning and cleanliness measurement in disk drive manufacture," *Precision Cleaning*, pp. 13-21, Feb. 1997.

[3] A. Rosenkranz, L. Freeman, B. Suen, Y. Fainman, and F. E. Talke, "Tip-enhanced Raman spectroscopy studies on amorphous carbon films and carbon overcoats in commercial hard disk drives," *Tribology Letters*, vol. 66, no. 2, pp. 1-6, Mar. 2018. DOI: 10.1007/s11249-018-1005-2.

[4] M. Kansiz, C. Prater, E. Dillon, M. Lo, J. Anderson, C. Marcott, A. Demissie, Y. Chen, and G. Kunkel, "Optical photothermal infrared microspectroscopy with simultaneous Raman – A new non-contact failure analysis technique for identification of <10 μm organic contamination in the hard drive and other electronics industries," *Microscopy Today*, vol. 28, no. 3, pp. 26-36, May 2020. DOI: 10.1017/S1551929520000917.

[5] X. Fan, W. Ming, H. Zeng, Z. Zhang, and H. Lu, "Deep learning-based component identification for the Raman spectra of mixtures," *Analyst*, vol. 144, no. 5, pp. 1789-1798, Jan. 2019. DOI: 10.1039/C8AN02212G.

[6] X. Zhang, T. Lin, J. Xu, X. Luo, and Y. Ying, "DeepSpectra: An end-to-end deep learning approach for quantitative spectral analysis," *Analytica Chimica Acta*, vol. 1058, pp. 48-57, Jun. 2019. DOI: 10.1016/j.aca.2019.01.002.

[7] W. Zhang, W. Feng, Z. Cai, H. Wang, Q. Yan, and Q. Wang, "A deep one-dimensional convolutional neural network for microplastics classification using Raman spectroscopy," *Vibrational Spectroscopy*, vol. 124, p. 103487, Jan. 2023. DOI: 10.1016/j.vibspec.2022.103487.

[8] X. Qiu, X. Wu, X. Fang, Q. Fu, P. Wang, X. Wang, S. Li, and Y. Li, "Raman spectroscopy combined with deep learning for rapid detection of melanoma at the single cell level," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 286, p. 122029, Feb. 2023. DOI: 10.1016/j.saa.2022.122029.

[9] S. Gulyanon, S. Deepaisarn, C. Srisumarnk, N. Chiewnawintawat, A. Angkoonsawaengsuk, S. Laitrakun, P. Opaprakasit, P. Rakpongsiri, T. Meechamnan, and D. Sompongse, "A comparative study of noise augmentation and deep learning methods on Raman spectral classification of contamination in hard disk drive," in *2022 17th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, Chiang Mai, Thailand, pp. 1-6, 2022. DOI: 10.1109/iSAI-NLP56921.2022.9960277.

[10] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 3rd ed., Sebastopol, CA: O'Reilly Media, 2022.

[11] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998. DOI: 10.1109/5.726791.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, Jun. 2017. DOI: 10.1145/3065386.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv* 2014. [Online] Available: https://arxiv.org/abs/1409.1556.

[14] C. Szegedy, W. Liu, Y. Jia, P. Semanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with

convolutions," *arXiv* 2014. [Online] Available: https://arxiv.org/abs/1409.4842.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas: NV, pp. 770-778, 2016. DOI: 10.1109/CVPR.2016.90.

[16] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," *arXiv* 2016. [Online] Available: https://arxiv.org/abs/1602.07360.

[17] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *arXiv* 2016. [Online] Available: https://arxiv.org/abs/1610.02357.

[18] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu: HI, pp. 4700-4708, 2017. DOI: 10.1109/cvpr.2017.243.

[19] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv* 2017. [Online] Available: https://arxiv.org/abs/1704.04861.

[20] C.-S. Ho, N. Jean, C. A. Hogan, L. Blackmon, S. S. Jeffrey, M. Holodniy, N. Banaei, A. A. E. Saleh, S. Ermon, and J. Dionne, "Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning," *Nature Communications*, vol. 10, p. 4927, Oct. 2019. DOI: 10.1038/s41467-019-12898-9.

[21] C. Chen, W. Wu, C. Chen, F. Chen, X. Dong, M. Ma, Z. Yan, X. Lv, Y. Ma, and M. Zhu, "Rapid diagnosis of lung cancer and glioma based on serum Raman spectroscopy combined with deep learning," *Journal of Raman Spectroscopy*, vol. 52, no. 11, pp. 1798-1809, Aug. 2021. DOI: 10.1002/jrs.6224.

[22] X. Zhang, X. Song, W. Li, C. Chen, M. Wusiman, L. Zhang, J. Zhang, J. Lu, C. Lu, and X. Lv, "Rapid diagnosis of membranous nephropathy based on serum and urine Raman spectroscopy combined with deep learning methods," *Scientific Reports*, vol. 13, p. 3418, Feb. 2023. DOI: 10.1038/s41598-022-22204-1.

[23] X. Chang, M. Yu, R. Liu, R. Jing, J. Ding, J. Xia, Z. Zhu, X. Li, Q. Yao, L. Zhu, and T. Zhang, "Deep learning methods for oral cancer detection using Raman spectroscopy," *Vibrational Spectroscopy*, vol. 126, p. 103522, May 2023. DOI: 10.1016/j.vibspec.2023.103522.

[24] J. Houston, F. G. Glavin, and M. G. Madden, "Robust classification of high-dimensional spectroscopy data using deep learning and data synthesis," *Journal of Chemical Information and Modeling*, vol. 60, no. 4, pp. 1936-1954, Mar. 2020. DOI: 10.1021/acs.jcim.9b01037.

[25] M. Wu, S. Wang, S. Pan, A. C. Terentis, J. Strasswimmer, and X. Zhu, "Deep learning data augmentation for Raman spectroscopy cancer tissue classification," *Scientific Reports*, vol. 11, p. 23842, Dec. 2021. DOI: 10.1038/s41598-021-02687-0.

[26] J. Zhao, H. Lui, D. I. McLean, and H. Zeng, "Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy," *Applied Spectroscopy*, vol. 61, no. 11, pp. 1225-1232, Nov. 2007. DOI: 10.1366/000370207782597003.

[27] U. Blazhko, V. Shapaval, V. Kovalev, and A. Kohler, "Comparison of augmentation and pre-processing for deep learning and chemometric classification of infrared spectra," *Chemometrics and Intelligent Laboratory Systems*, vol. 215, p. 104367, Aug. 2021. DOI: 10.1016/j.chemolab.2021.104367.

[28] N. K. Afseth and A. Kohler, "Extended multiplicative signal correction in vibrational spectroscopy, a tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 117, pp. 92-99, Aug. 2012. DOI: 10.1016/j.chemolab.2012.03.004.

[29] J. Salmon, Z. Harmany, C.-A. Deledalle, and R. Willett, "Poisson noise reduction with non-local PCA," *Journal of Mathematical Imaging and Vision*, vol. 48, no. 2, pp. 279-294, Feb. 2014. DOI: 10.1007/s10851-013-0435-6.

[30] H. Paik, N. Sastry, and I. SantiPrabha, "Effectiveness of noise jamming with white gaussian noise and phase noise in amplitude comparison monopulse radar receivers," in *2014 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, Bangalore, India, pp. 1-5, 2014. DOI: 10.1109/conecct.2014.6740286.

[31] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas: NV, pp. 2921-2929, 2016. DOI: 10.1109/cvpr.2016.319.

[32] R. L. Draelos and L. Carin, "Use HiResCAM instead of Grad-CAM for faithful explanations of convolutional neural networks," *arXiv* 2020. [Online] Available: https://arxiv.org/abs/2011.08891.