Microscopy and Microanalysis, 2024, **00**, 1–10
https://doi.org/10.1093/mam/ozae068

Original Article

Microscopy AND Microanalysis

Identifying Sample Provenance From SEM/EDS Automated Particle Analysis via Few-Shot Learning Coupled With Similarity Graph Clustering

Jasmine Eshun, Natalie C. Lamar, Sinan G. Aksoy, Sarah Akers, Benjamin Garcia, Heather Cunningham, George Chin Jr, and Jenna A. Bilbrey*

National Security Directorate, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA 99352, USA

***Corresponding author:** Jenna A. Bilbrey, E-mail: jenna.pope@pnnl.gov

Abstract

Automated particle analysis (APA) provides a vast amount of compositional data via energy-dispersive X-ray spectroscopy along with size and shape data via scanning electron microscopy for individual particles in a sample. In many instances, APA data are leveraged to support identification of the source of a sample based on the detection of particles of a specific composition. Often, the particles that provide context make up a minuscule portion of the sample. Additionally, the interpretation of complex samples can be difficult due to the diversity of compositions both in the mixture and within a particle. In this work, we demonstrate a method to compute and cluster similarity graphs that describe inter-particle relationships within a sample using a multi-modal few-shot learning neural network. As a proof-of-concept, we show that samples known to have been exposed to gunshot residue can be distinguished from samples occasionally mistaken for gunshot residue. Our workflow builds upon standard APA techniques and data processing methods to unveil additional information in a readily interpretable and quantitatively comparable format.

Key words: automated particle analysis, deep learning, few-shot neural network, graph theory, scanning electron microscopy–energy-dispersive X-ray spectroscopy

Introduction

Automated particle analysis (APA) has widespread use across a variety of scientific domains and industries, including analyses of inhaled particulate matter (Lowers et al., 2018; Hayden et al., 2023), soil mineralogy for environmental and forensic investigations (Pirrie et al., 2009; Pirrie & Rollinson, 2011; Schulz et al., 2020), and criminal forensics of gunshot residue (ASTM International, 2017; Charles et al., 2023). APA provides compositional data via energy-dispersive X-ray spectroscopy (EDS) along with size and shape data via scanning electron microscopy (SEM) of particles in a sample. A typical APA survey scan captures images and spectra for thousands to tens of thousands of individual particles to provide a snapshot of the overall composition of the sample.

Real-world samples can be difficult to confidently characterize due to the diversity of compositions both in the mixture and within individual particles. The National Institute of Standards and Technology (NIST) has developed an open-source software, DTSA-II (Goldstein et al., 2018), that quantifies elemental compositions obtained from automated fitting of the EDS spectra. Using the set of normalized k-ratios extracted from each spectrum (Ritchie, 2023), one can apply rule-based algorithms to assign particles to user-defined material classes. Such hierarchical classification scheme has proven useful in a variety of applications, including the detection of laboratory contamination (Lindstrom et al., 2013) and atmospheric transport of Mn-enriched rock varnish (Ortiz-Montalvo et al., 2018).

Often, hierarchical sorting algorithms assign particles to a single material class in a "greedy" manner, meaning that the order of the ruleset has a large influence on the ultimate class assignment, with rules encountered early in the hierarchy receiving elevated priority. The ruleset and hierarchy are determined manually based on the intended application, which can lead to disagreement in the make-up of a sample depending on the assigned order. In addition, a single class assignment does not allow particles with complex compositions to fall into multiple classes, nor does it express relationships between material classes that are compositionally similar.

One way of understanding relationships in complex samples is through the calculation of pairwise similarities. For instance, Huber et al. developed a deep-learning similarity measure for mass spectra (Huber et al., 2020; Huber et al., 2021*a*), which was then applied to predict structural similarity in complex mixtures based on tandem mass spectra (Huber et al., 2021*b*). Wei et al. (2022) used predicted similarity scores of NMR spectra to identify compounds in mixtures. Routh et al. (2023) used the similarity of embedded X-ray absorption fine structure spectra for the speciation of heterogeneous nanocatalysts. Here, we define a multi-modal similarity measure between particles examined by APA using a few-shot learning (FSL) approach.

Traditionally, neural networks used for classification require large, balanced training sets to learn accurate representations

Received: December 5, 2023. **Revised:** May 6, 2024. **Accepted:** July 4, 2024

of all classes. However, many real-world problems call for decisions on samples whose corresponding class has few or no examples in the original training set. This is especially true for APA of environmental samples (here, we define as mixtures of naturally occurring geologic minerals and man-made compounds including but not limited to metal alloys, plastics, industrial effluents, construction materials, and glass), which can contain particles with rarely observed or previously unobserved compositions. FSL provides a solution by transforming a multiclass classification problem into a binary classification problem, producing a similarity score between two samples. FSL networks are more dynamic than traditional supervised networks in that they quickly respond to new and unseen samples.

When making predictions on new samples, many FSL approaches employ a support set, or a pre-collected library of known samples. Choosing the appropriate support set requires a predetermined knowledge of the sample composition, which is often unknown for environmental samples. To avoid the use of a support set, we propose a workflow that compares all particles in a sample against one another. The set of similarity scores output by the FSL network represents the complex relationships between particles in a sample. Because of the combinatorial nature of such pairwise comparison, the resulting large set of similarity scores requires an abstract mathematical representation for interpretation.

The set of similarity scores output by our FSL framework can be mathematically represented as a similarity graph. Graphs have been used to analyze complex relationships in diverse domains such as omics data (Berger et al., 2013), chemical reaction networks (Zeigarnik et al., 1996; Craciun & Feinberg, 2006), and descriptions of chemical space (Coley, 2021; Dunn et al., 2022; Scalfani et al., 2022). Similarity graph representations are advantageous and ubiquitous across data science because they enable graph methodologies that perform fundamental analyses, such as clustering or partitioning, entity importance ranking, and generating node embeddings (von Luxburg, 2007). Crucially, graph analyses can utilize not only the local similarity information (i.e., those related to a single data point) but also relations between data points through longer, complex chains. This richer information yields additional insights into micro- and macroscale relationships in the data. For instance, clusters of overlapping material classes within a sample, as well as potentially outlying or misclassified particles, are readily observed. The similarity graph formulation also allows quantitative analysis of the graph structure, providing a basis for cross-sample comparison.

In forensic analysis, APA is used to locate gunshot residue in a sample, which provides important trace evidence to support reconstruction of a crime scene (Shrivastava et al., 2021). In this work, we demonstrate the utility of similarity graphs in APA through examination of a dataset containing samples from firearm discharge, firework residue, and automobile brake dust—the latter two of which are often confused with firearm discharge. A multi-modal FSL network that ingests both EDS spectra and SEM images from APA data is used to produce similarity scores between all particles in a sample, which are formulated into a similarity graph. We then perform a quantitative analysis on the similarity graphs and show that clustering of the computed graph metrics affords complete separation of samples obtained from firearm discharge from samples obtained from firework residue or brake dust.

Methods

Pairwise Similarity Prediction

Our FSL network aims to predict a similarity score between two samples, rather than assigning a discrete class to each individual sample, through the use of "twin" networks (Figs. 1a, 1b). To perform pair classification, the data for each particle are individually passed through feature extraction layers which learn the ideal features to distinguish pairs of different classes and minimize difference between pairs of the same class. In this work, a 1D AlexNet (Krizhevsky et al., 2017) architecture was used for feature extraction from EDS spectra, and a 2D AlexNet architecture was used for feature extraction from SEM images. The feature vectors are then passed through a series of fully connected layers with L_2 regularization and dropout. Dropout is activated during both training and inference to reduce overfitting and allow uncertainty quantification (Gal & Ghahramani, 2016). The incorporation of dropout uncertainty in twin networks has precedent in studies of change detection in images (Wang et al., 2020) and comparisons of tandem mass spectra (Huber et al., 2021*b*). During training, the weights of the shared feature extraction layers, as well as the weights of the shared UQNet layers, are updated simultaneously, as indicated by double sided arrows in Figure 1a. The resultant vectors from each modality are then concatenated in the fusion step to produce a single vector per particle. The fused vectors are passed into an entropy function, in this case, the squared difference. In the single modality case, this fusion step is omitted and the UQNet output is fed directly into the entropy function. Finally, the entropy scalar is pass through the sigmoid activation function to produce a similarity score ranging from 0 to 1.

The FSL network is trained to learn an embedding function that maps input examples into a meaningful vector space that captures relevant information about the input examples, enabling similarity comparisons between them. The goal of training is to minimize the binary cross-entropy loss, for which we employed the ADAM (Adaptive Moment Estimation) optimization algorithm. The binary cross-entropy loss function is commonly used in binary classification tasks, as it provides a suitable loss metric for evaluating the dissimilarity between predicted probabilities and true binary labels. The gradients of the loss with respect to the network's learnable parameters are calculated through backpropagation, and the parameters are updated in the opposite direction of the gradients to minimize the loss. Model training was parallelized across two NVIDIA V100 GPUs, an adaptive learning rate beginning at 0.001 was applied, the batch size was set to 256, and training was stopped once the validation loss did not decrease for ten consecutive epochs. Dropout was set to 0.3 and applied during both training and inference. During inference, ten repeat predictions were performed for each particle pair to obtain a mean similarity score (Fig. 1c).

The final form of a neural network is highly dependent on the data used for training, requiring careful curation of the training set. Due to the large number of particles measured during APA, we apply a set of unsupervised methods to automatically curate the data. First, the isolation forest anomaly detection method (Liu et al., 2008), as implemented in the scikit-learn library, was used to detect outliers among EDS spectra within a material class. Outlier spectra were observed to have additional or missing peaks compared with other spectra in the corresponding material class, indicating a difference

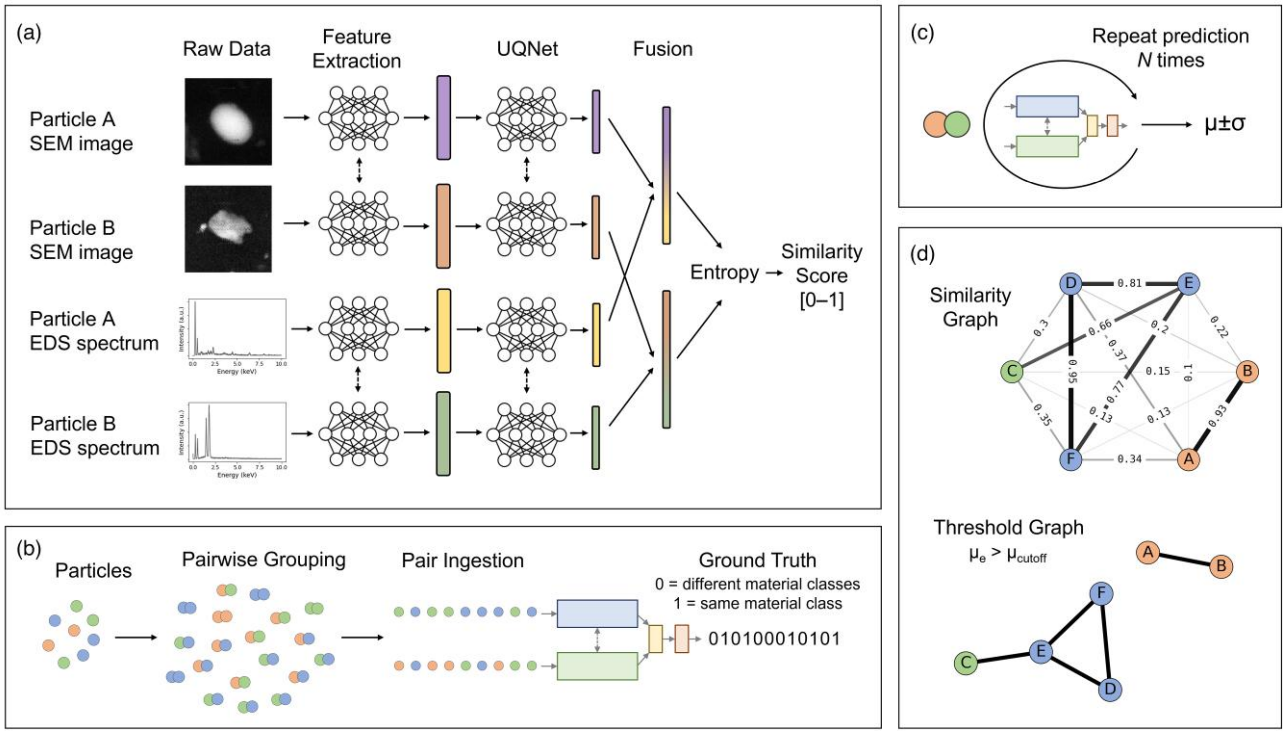

Fig. 1. (**a**) Key components of the analysis workflow: (**a**) multi-modal few-shot learning (FSL) network to predict similarity scores between particle pairs, (**b**) training of the FSL network, (**c**) similarity predictions incorporating model uncertainty, and (**d**) formulation of a similarity graph from pairwise similarity scores.

in elemental composition. Particles with outlier spectra were dropped from the training set.

The training set for the FSL network is composed of particle pairs. Particle pairs of the same material class are labeled 1, while particle pairs of different material classes are labeled 0. However, some of the material classes are broadly defined and may have several subgroups. Therefore, we further refined same-class pairs by computing a spectral similarity score between pairs of EDS spectra using Spec2Vec (Huber et al., 2021*a*). Spec2Vec was inspired by Word2Vec, a natural language processing algorithm that provides similarity scores between words in sentences. Spec2Vec transforms peaks in a spectrum into a textual representation to compute spectral similarity scores based on a learned embedding. Spec2Vec was originally developed for metabolomics analyses using spectra from tandem mass spectrometry. In this work, we re-tool their algorithm for use with EDS spectra and trained a Spec2Vec network over each material class. Once the Spec2Vec scores were obtained for pairs within a material class, a threshold was applied based on the q-th percentile, and pairs whose Spec2Vec score did not meet this threshold were not included in the training set. We found a percentile of 2.2 to provide adequate filtering of dissimilar spectra within the same material class.

Similarity Graphs

After training, inference is performed over all pairs of particles particle pairs in the sample, producing $\frac{n(n-1)}{2}$ similarity scores, where n is the number of particles. Therefore, a 500-particle sample would have 124,750 associated similarity scores. A mathematically convenient way to represent these data is in the form of a similarity graph $G = (V, E)$, where the set of vertices $V = \{v_1, \ldots, v_n\}$ corresponds to the particles in the sample and the set of edges E corresponds to the set of pairwise similarity scores. When all pairwise similarities are included in G, the graph is a "complete graph" and can be denoted K_n. Thresholding the graph by removing edges representing similarity scores below a certain threshold value improves visualizations of G, as shown in Figure 1d, and supports the computation of quantitative metrics describing the graph structure.

To calculate structural metrics without employing an arbitrary threshold, we employ a scheme that averages over all possible similarity score filterings. Filtering G is necessary because many graph metrics have no natural weighted analog and unweighted graph metrics are not meaningful when applied to K_n, which always links every pair of particles with a score. Accordingly, we filter G by removing edges whose associated similarity score value does not meet a specified threshold value. Rather than limit our analysis to graphs derived from a single threshold, our approach captures how the graph structure evolves over all thresholds: we compute the graph metric under consideration for each threshold choice, and weight the value based on the range of scores for which thresholding yields the same graph. More precisely, if the observed similarity scores are $0<s_1<s_2<\ldots<s_k<1$, then for each threshold s_i, we

1. Construct a filtered similarity graph consisting only of edges with scores greater than s_i,
2. Compute the graph metric on this filtered graph, and
3. Scale the output value by $(s_{i+1} - s_i)$,

repeating and summing this quantity over all observed s_i. McAssey & Bijma (2015) took a similar "weighted average"

approach for computing the graph clustering coefficient on similarity data. In practice, however, the number of distinct, pairwise similarity scores grows rapidly with the number of particles, and it is neither computationally tractable nor necessary to compute each graph metric at each similarity score. Instead, we find that choosing 50 equally spaced thresholding values—that is, thresholding in increments of 0.02—yields sufficiently accurate approximations across the graph metrics we consider.

For quantitative comparison of the different graphs, we compute a set of graph-level metrics that that provide structural information about each graph. Macindoe & Richards (2010) developed a method for quantitatively comparing graphs based on structural features. The metrics described in their work—leadership, bonding, and diversity—were targeted toward properties of social networks. Leadership quantifies the level at which a particular node overwhelmingly influences the edge connectivity; in our case, a high leadership value indicates that a small number of particles have higher-than-average similarities to a large proportion of particles in the sample, while a low leadership value indicates roughly equal similarity distributions between particle pairs. Bonding is a measure graph transitivity that described the overall probability for a network to have adjacent nodes interconnected. This metric describes the presence of communities within the graph, affording a description of how particles cluster within a sample. Diversity is a normalized measure of the number of disjoint edge pairs whose endpoints are not linked. This metric explores the separation between communities in the graph, in other words, the separation of groupings of similar particles in the sample. We extend their set of metrics to include class assortativity, degree assortativity, algebraic connectivity, and p-smoothness. Assortativity describes the preference for vertices to be linked to other vertices that are similar in a defined way. Class assortativity examines this preference based on material class of the particle, while degree assortativity examines this preference based on number of edges connected to the vertex. Algebraic connectivity reflects how well connected the overall graph is (i.e., are some populations of particles completely dissimilar from others), while p-smoothness measures the entropy in the size of connected components in the graph (i.e., are these dissimilar populations of disproportionate size). We find this set of seven quantitative metrics to sufficiently describe various aspects of the particles in the sample and their relationship to one another. For the analyzed samples, pairwise correlation coefficients were computed between all metrics to check for redundancy.

Clustering

The seven metrics provide a set of descriptors that can be used to group the similarity graphs. In this work, we apply Density-Based Spatial Clustering of Applications with Noise (DBSCAN), as implemented in the scikit-learn package, as a post-analysis tool to confirm our hypothesis that our coupled few-shot learning-similarity graph method can be used to characterize samples. Density-based techniques (such as DBSCAN) tend to be more robust toward arbitrary-shaped clusters and have the benefit of denoting outliers that do not belong to any detected cluster, compared with partition-based clustering techniques (such as k-means) and hierarchical clustering techniques (such as agglomerative clustering). Note that, except for that of the noise class, the numerical label of the cluster is arbitrary; the important factor is samples contained within each cluster.

Dataset

APA data consisting of SEM images with paired EDS spectra were obtained from NIST (Ritchie & Renolds, 2021). Along with the raw data, each particle is assigned a material class based on a user-defined ruleset that assigns particles according to their normalized k-ratios and some numeric morphology metrics. The dataset contains 30 APA samples from three broad categories: automobile brake dust, firework residue, and firearm discharge. The hierarchy of samples is shown in Figure 2. Twelve brake dust samples were obtained from all four tires of three different vehicles—two Ford Explorers (denoted A and B) and one Chevy Caprise. Ten fireworks samples were obtained from sparklers, spinners, and Roman candles at various collection periods both before and after ignition. Eight firearm discharge samples were taken by sampling the hands of five volunteer shooters after firing a firearm at a firing range.

The samples were separated into training and analysis sets. Samples reserved for the training set included the four tires of Ford Explorer A (20,000 particles), all samples collected from spinners (20,000 particles), and samples from both hands of shooter #3 (51,604 particles). The analysis set contained samples of brake dust from all four tires of Ford Explorer B and the Chevy Caprise (39,999 particles), sparklers and Roman candles (25,247 particles), and firearm discharge from the remaining shooters (111,468 particles). Particles assigned to the "Other" material class were removed from the training set, due to lack of specificity in the class description, but retained in the analysis sets.

Results and Discussion

SEM/EDS is the technique of choice for the detection of inorganic gunshot residue (GSR) in particle samples (ASTM International, 2017; Ritchie et al., 2020; Charles et al., 2023). Both SEM images and EDS spectra are used to support a positive indication of GSR particulate through assessment of the shape (spheroidal), size (0.5–5.0 μm in diameter), and elemental composition (presence of lead, barium, and antimony). Detection is typically performed in two phases: 1) discovery of "GSR-like" particles using APA and automated classification and 2) confirmation of GSR via manual re-examination of the identified GSR-like particles, often through newly acquired SEM images and EDS spectra. The second step is key to the correct identification of GSR, but requires careful examination by an experienced analyst and can be labor-intensive if many GSR-like particles are found.

The changing composition of ammunition, which affects both elemental composition and particle morphology, can muddle determinations of GSR (Feeney et al., 2020; Tahirukaj et al., 2021). Moreover, a recent study of firearm gunpowders showed some to be coated with additional elements, such as chrome, manganese, and iron (Burnett et al., 2021), further complicating elemental-based determinations of GSR. To further confound analyses of real-world samples, automobile brake pads can produce particles that are both compositionally and morphologically similar to gunshot residue. While modern automotive brake pads produce a small number of particles compositionally similar to GSR, these "GSR-like" particles are commonly found alongside iron-containing particles (Torre et al., 2001; Tucker et al., 2017).

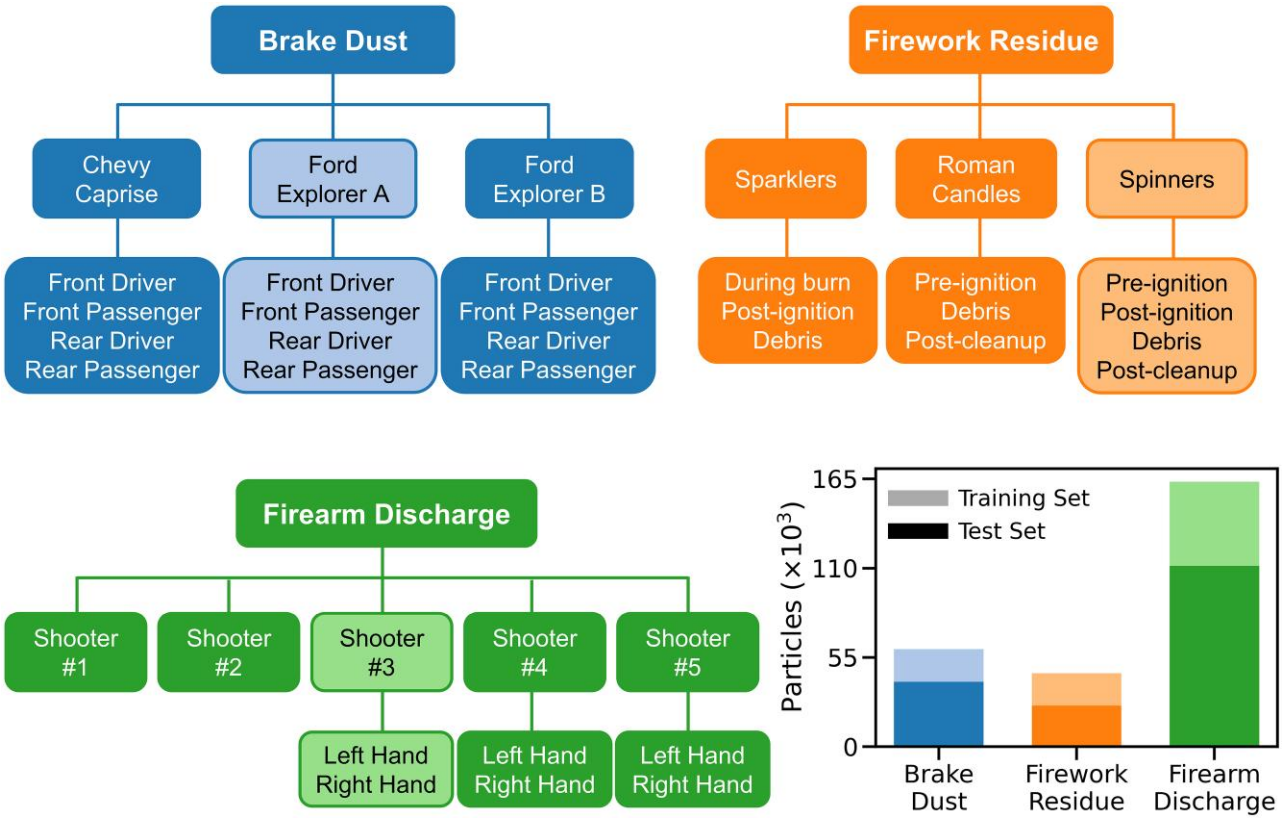

Fig. 2. Hierarchy of sample categories and number of particles contained in the NIST APA dataset. Samples in shaded boxes were used to train the FSL network, while samples in the solid boxes were reserved for analysis.

Table 1. Number of Particles Assigned to a GSR Material Class for Each Sample Category.

	Brake Dust	Firework Residue	Firearm Discharge
GSR.0	0	0	2,310
GSR.1	23	2	3,381
GSR.2	5,548	544	4,276
GSR.3	0	1	6
GSR.6	4	39	99
GSR.Ba-Sb	77	14	520
GSR.Pb-Ba	235	362	703
GSR.Pb-Sb	1	1	533
GSR.Sr-bearing	154	605	1,471
Total	6,042 (10.0%)	1,568 (3.5%)	13,299 (8.2%)

Also given is the total number of particles labeled GSR in the sample class, with the percent of the total sample in parenthesis.

The presence of copious iron-containing particles in a sample can be used to rule out GSR, but a false negative result may occur, for example, when sample contains a mixture of brake pad dust and low amounts of GSR (Tucker et al., 2017).

Table 1 gives a count of particles labeled GSR for each sample class (brake dust, firework residue, and firearm discharge) for the NIST APA dataset of samples known to have been exposed to gunshot residue and from samples occasionally mistaken for gunshot residue (Ritchie & Renolds, 2021) applied in this work. Supplementary Table S1 gives the same count broken out by specific sample. Notably, each sample contains particles labeled as GSR. Surprisingly, the brake dust class contains a higher percentage of particles labeled GSR than the firearm discharge class (10.0% versus 8.2%, respectively). The high count of GSR-like particles in the brake dust class are predominately from the two rear tires of the Chevy Caprise, which accounts for 36.6% of all GSR-like particles in the brake dust class.

In this work, we present an analysis method to discriminate firearm discharge from confounding samples that considers both GSR-like and non-GSR-like particles. We apply a multimodal neural network that ingests both EDS spectra and SEM images and determines a similarity score between pairs of particles. The set of similarity scores between all particle pairs in the sample is formatted into a similarity graph, which wholistically describes the inter-particle relationships in the sample. Quantitative graph metrics are computed for each sample, and clustering of the metrics allows discrimination of samples collected from firearm discharge from those collected from brake dust or firework residue.

The SEM images obtained by APA surveys are low resolution and do not provide enough information to categorize particles based on morphology. However, the general particle shape and brightness, which is related to the atomic number, may provide ancillary information to enhance particle similarity predictions. We hypothesize that the addition of SEM images along with EDS spectra to the FSL network will improve particle characterization by enhancing the learned embedding with additional information. To confirm this, we train two networks: one using only EDS spectra (single modality network) and one using both EDS spectra and SEM images (multi-modal network).

The training set consists of 260,164 pairs of particles, half composed of different classes and half of the same class. The validation set consists of 32,520 pairs, and the test set consists

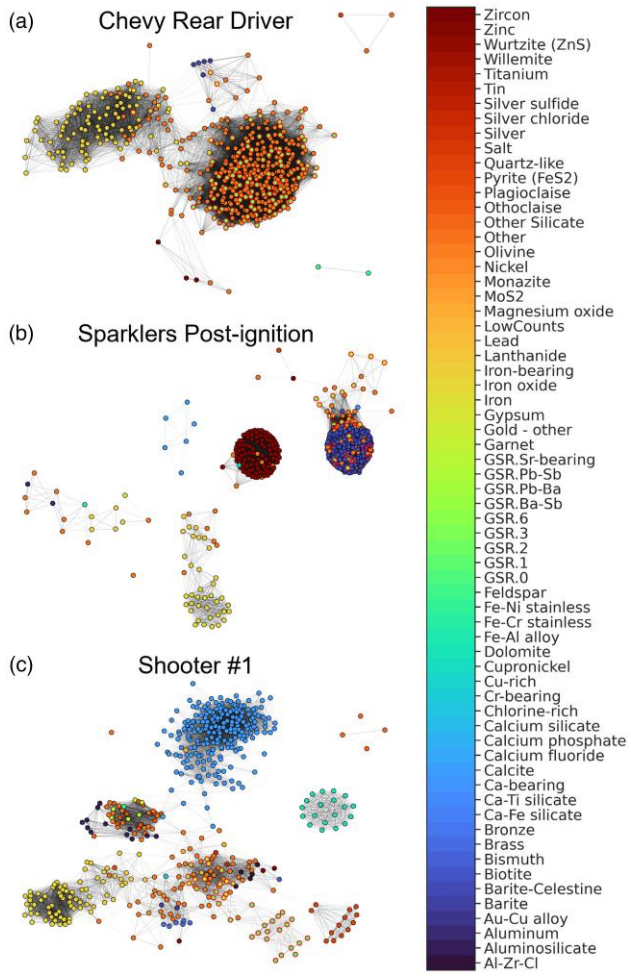

Fig. 3. Thresholded multi-modal similarity graphs (similarity score ≥ 0.8) from the three sample classes: (**a**) break dust from the Chevy's rear driver tire, (**b**) firework residue from sparklers post-ignition, and (**c**) firearm discharge from shooter #1. Particles assigned to a GSR class are circled in red. Nodes are colored according to the material class denoted in the color bar. Note that the color map is consistent across samples, and not all particle types are present in each graph.

of 32,528 pairs, both with 1:1 splits. The same training, validation, and test sets were used in both the single modality and multi-modal networks. Training proceeded successfully for both networks (Supplementary Fig. S1); however, the multi-modal network required more training epochs to converge. The networks were considered to have converged when the validation loss did not decrease over ten consecutive epochs. Initially, the multi-modal network was stuck in a local minimum until around ten epochs, which did not occur with the single modality network. Training converged at 25 epochs for the single modality network, with an optimal validation loss of 0.188 and validation accuracy of 94.0%, and at 40 epochs for the multi-modal network, with an optimal validation loss 0.200 and validation accuracy of 93.8%. Receiver operating characteristic (ROC) curves and the corresponding area under the curve (AUC) were used to analyze model performance through comparison of true positive and false positive rates. AUC values range from 0 (all model predictions are wrong) to 1 (all model predictions are correct). ROC curves were calculated for the validation and test sets for the trained networks, with both networks giving AUCs of 0.98 for both sets. This demonstrates that both networks were well-trained.

Inference was performed and similarity graphs were produced for samples not included in the training set: eight automobile brake dust, six firework residue, and six firearm discharge samples were analyzed. To measure the consistency and reliability of the similarity graphs, ten similarity graphs were created from 500-particle subsamples of each APA scan. The Roman candle debris sample consisted of 247 particles, and so only a single graph was generated.

Individual similarity graphs provide information on groupings of material classes in a sample. Figure 3 shows an example thresholded multi-modal similarity graph for each sample category. In general, at the 0.8 threshold, the brake dust samples show high connectivity among the majority of particles, mainly due to the large number of particles containing iron in these samples. In contrast, the samples from firearm discharge show greater separation between clusters of highly similar particles. In the firework residue classes, Roman candles at all collection periods and sparklers during burn show separation similar to the firearm discharge samples, while sparkler samples taken post-ignition and from debris show connectivity intermediate to the firearm discharge and brake dust samples. These readily observable differences in the graph structure, even at an arbitrary similarity score threshold, lead us to suppose that samples could be grouped according to quantitative metrics describing their graph structure.

For cross-sample comparison, we computed a set of graph metrics that quantitatively describe the inter-particle relationships within a sample, which are described in the Methods section. We hypothesize that these graph metrics can be used to minimize the false positive identification of GSR in brake dust and firework samples, by associating the inter-particle relationships with the source of a given sample.

The addition of SEM images to the particle description will further improve the ability to distinguish GSR-containing samples. The similarities derived from the multi-modal model capture additional structural information about the particles, which increases the reliability of the multi-modal model in comparison to the single modality model in distinguishing samples that have a true positive identification of GSR. To demonstrate the needs for multi-modal similarity graphs, we computed graph metrics and perform clustering for on two separate groups: 1) graph metrics derived from single modality similarity graphs and 2) graph metrics derived from multi-modal similarity graphs.

Before applying a clustering technique to the graph metrics, we use pair plots to assess the potential of clusters forming. Figure 4 shows pair plots of metrics from the multi-modal analysis, while Supplementary Figure S2 shows the corresponding plots for the single modality analysis. In both cases, emergent clusters are observed even in pairwise comparisons of the metrics. Notably, samples from firearm discharge do not break out into separate groups, as is observed for brake dust and firework residue samples. Some metrics distinguish sample classes more than others; for example, algebraic connectivity and degree assortativity help to distinguish brake dust from firework residue and firearm discharge.

To compare the distributions of single modality and multi-modal graph metrics resulting from the single modality and multi-modal networks, we computed the Wasserstein distance between the distributions for each graph metric across the three sample classes (Table 2). The addition of SEM images

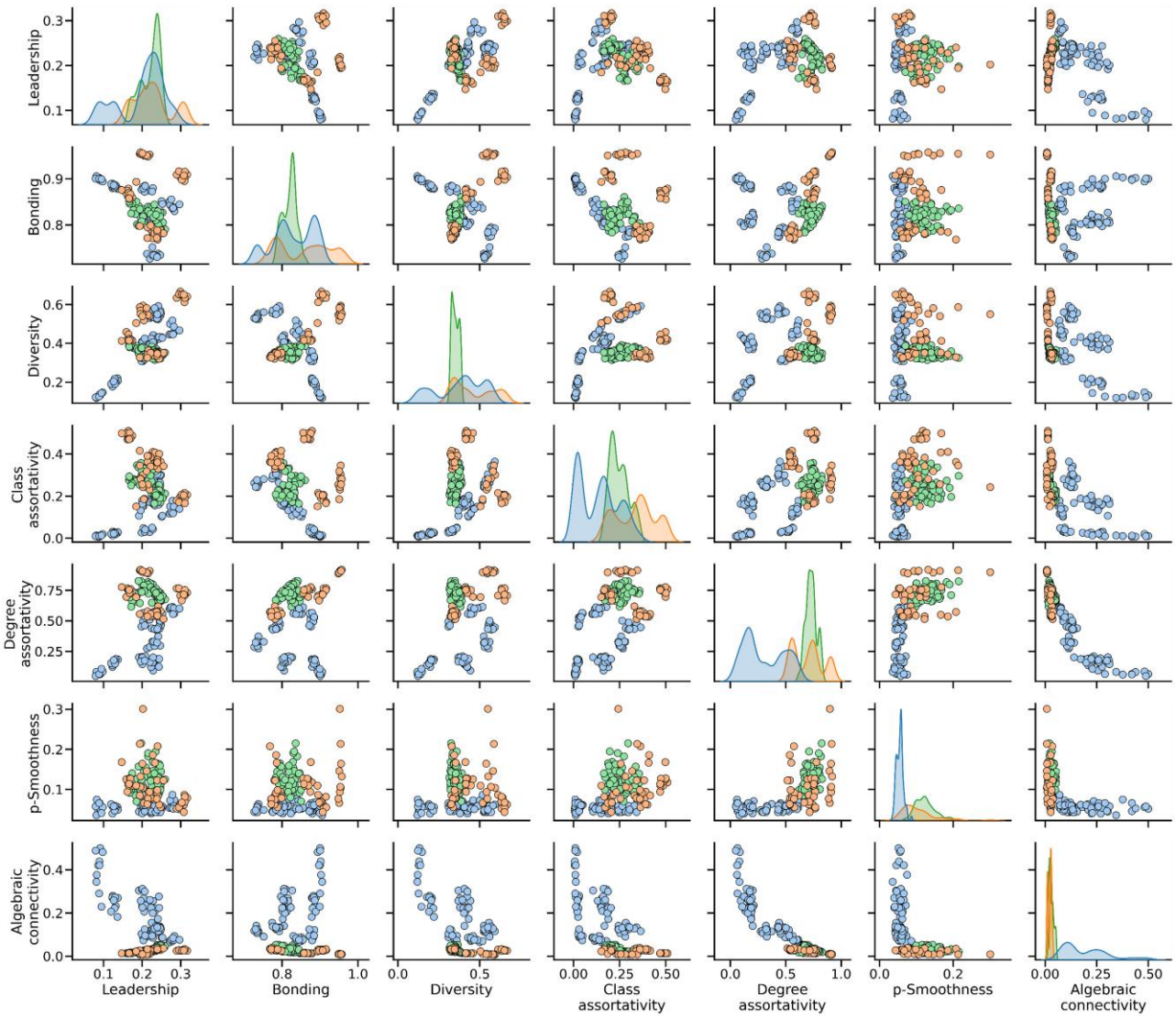

Fig. 4. Pair plots of the multi-modal graph metrics for the brake dust (blue), firework residue (orange), and firearm discharge (green) sample classes. The plots along the diagonal show the distribution of the corresponding metric.

Table 2. Wasserstein Distances Between the Single Modality and Multi-Modal Graph Metric Distributions for the Brake Dust, Firework Residue, and Firearm Discharge Sample Classes.

	Leadership	Bonding	Diversity	Class Assortativity	Degree Assortativity	p-Smoothness	Algebraic Connectivity
Brake dust	0.017	0.041	0.043	0.026	0.058	0.004	0.036
Firework residue	0.018	0.014	0.019	0.021	0.037	0.016	0.004
Firearm discharge	0.012	0.006	0.031	0.009	0.039	0.006	0.002

had the largest effect on the brake dust class, which showed the largest Wasserstein distances for five of the seven metrics (bonding, diversity, class assortativity, degree assortativity, and algebraic connectivity). The firework residue class showed the largest Wasserstein distances for the leadership and p-smoothness metrics. Notably, samples from the firearm discharge class showed comparatively smaller changes in graph metric distributions.

We then applied the DBSCAN clustering algorithm to both sets of single modality and multi-modal graph metrics. DBSCAN is an unsupervised technique that clusters data into high-density regions clearly separated by regions of low density. In this technique, data in the low-density regions are categorized as noise. DBSCAN has two tunable parameters: epsilon and minimum number of samples. Epsilon is the maximum distance between two data points for them to be considered in the same neighborhood. Minimum number of samples is the number of samples in a neighborhood required for a data point to be considered a core point (including itself). In this work, we used an epsilon value of 0.9 and a minimum samples value of 5.

Figure 5 shows the results of clustering the metrics derived from the single modality and multi-modal networks; specific samples in each cluster are given in Table 3 for the multi-

modal model and Supplementary Table S2 for the single modality model. From clustering of the single modality metrics, three subgraphs from firearm discharge (all from shooter #2), two from brake dust (both from Chevy front passenger), and two from firework residue (Roman candle pre-ignition and Roman candle debris) were classified as noise, while for the multi-modal model, only the firework residue class had subgraphs classified as noise (three from Roman candle post-cleanup, one from Roman candle debris, two from sparkler post-ignition, and one from sparkler during burn). The 247-particle Roman candle debris sample fell into noise for both models. Subgraphs that fall into noise are dissimilar from any of the identified clusters, indicating that these subgraphs are not characteristic of the samples and do not support the ability to associate the sample class with the graph metrics. With the multi-modal model, all subgraphs that fall into the noise category are from the firework residue class, and all firearm discharge subgraphs are able to be clustered.

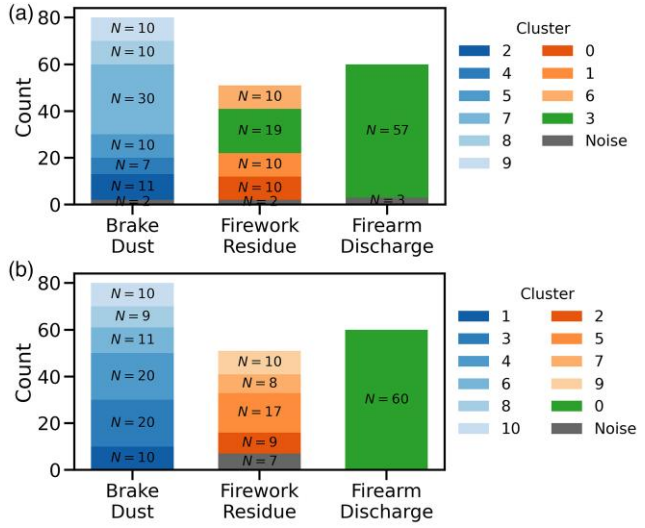

Fig. 5. Subsample classifications based on DBSCAN clustering of the graph metrics for subgraphs of samples using the (**a**) single modality and (**b**) multi-modal FSL networks.

Although clustering with the single modality group produced a cluster containing the remaining firearm discharge samples ($N = 57$), the cluster was not unique to firearm discharge and included several firework residue samples, specifically Roman candle pre-ignition ($N = 9$) and Roman candle post-cleanup ($N = 10$). The addition of SEM data in the multi-modal group resolved this overlap. Clustering with the multi-modal group formed a cluster composed of all shooter observations ($N = 50$). Moreover, this cluster did not include samples from any of the other classes, and no firearm discharge samples categorized as noise.

In addition, both modalities captured meaningful information in the clusters outside of the firearm discharge class. The multi-modal group provided additional context with information on vehicle brand, location of the tire on the vehicle, and collection period for the fireworks. In the single modality group, Chevy rear driver samples were clustered with the Ford rear driver and passenger samples ($N = 30$). The multi-modal model provided better segregation on tire location and vehicle brand: the Chevy and Ford samples always fell in separate clusters.

In the single modality group, sparkler samples fell into separate clusters based on collection period, while Roman candles did not. Roman candle samples not categorized as noise ($N = 2$) were clustered with the firearm discharge samples. For the multi-modal group, the sparkler samples were also clustered based on collection period, although some samples that previously fell into clusters were categorized as noise ($N = 3$). In the multi-modal group, Roman candle samples did not segregate by collection period but were no longer clustered with firearm discharge samples.

The results DBSCAN clustering of the graph metrics confirmed the hypothesis that similarity information can be leveraged to identify samples acquired from firearm discharge.

Table 3. Subsample Classifications Based on DBSCAN Clustering of the Graph Metrics for Subgraphs of Samples Using the Multi-Modal Few-Shot Network.

	DBSCAN Cluster											
Sample	Noise	0	1	2	3	4	5	6	7	8	9	10
Chevy front driver	0	0	0	0	0	0	0	1	0	9	0	0
Chevy front passenger	0	0	0	0	0	0	0	10	0	0	0	0
Chevy rear driver	0	0	0	0	0	0	0	0	0	0	0	10
Chevy rear passenger	0	0	10	0	0	0	0	0	0	0	0	0
Ford front driver	0	0	0	0	10	0	0	0	0	0	0	0
Ford front passenger	0	0	0	0	10	0	0	0	0	0	0	0
Ford rear driver	0	0	0	0	0	10	0	0	0	0	0	0
Ford rear passenger	0	0	0	0	0	10	0	0	0	0	0	0
Sparkler during burn	1	0	0	9	0	0	0	0	0	0	0	0
Sparkler post-ignition	2	0	0	0	0	0	0	0	8	0	0	0
Sparkler debris	0	0	0	0	0	0	0	0	0	0	10	0
Roman candle pre-ignition	0	0	0	0	0	0	10	0	0	0	0	0
Roman candle post-cleanup	3	0	0	0	0	0	7	0	0	0	0	0
Roman candle debris	1	0	0	0	0	0	0	0	0	0	0	0
Shooter #1	0	10	0	0	0	0	0	0	0	0	0	0
Shooter #2	0	10	0	0	0	0	0	0	0	0	0	0
Shooter #4 left hand	0	10	0	0	0	0	0	0	0	0	0	0
Shooter #4 right hand	0	10	0	0	0	0	0	0	0	0	0	0
Shooter #5 left hand	0	10	0	0	0	0	0	0	0	0	0	0
Shooter #5 right hand	0	10	0	0	0	0	0	0	0	0	0	0

Notably, the single modality model failed to isolate the firearm discharge samples from Roman candle samples, while the multi-modal model succeeded in separation, indicating that the inclusion of both EDS spectra and SEM images is necessary to accurately describe inter-particle relationships within a sample.

Conclusions

APA is used to locate particles that provide support for identification of the source of a sample. Typically, such particles comprise a small proportion of the sample, making this search akin to the search for a needle in a haystack. Furthermore, changing formulations and compositions complicate the identification of specific predefined "needles".

In this work, we demonstrate a method to quantitatively describe and cluster samples based on characteristics of the full particle set measured by APA. Although we provide proof-of-concept in the context of identifying firearm discharge, our method can be applied broadly to APA data in general. Our combined FSL and similarity graph approach was shown to successfully distinguish samples obtained from firearm discharge from samples obtained from automobile brake dust or firework residue. A multi-modal model that ingests both EDS spectra and SEM images was required for complete separation of all firearm discharge samples.

We note that our method is intended to support, rather than replace, the analyst, especially in situations with potentially severe real-world consequences. The methods presented should only be used to provide further assurance that a correct determination was made and should not replace or outrank standard practices.

Availability of Data and Materials

The automated particle analysis (SEM/EDS) dataset (Ritchie & Renolds, 2021) used in this study is openly available on the NIST website (https://data.nist.gov/od/id/mds2-2476). Code necessary to apply our method and support the findings of this study is available at https://github.com/pnnl/particlefsl.

Supplementary Material

The supplementary information contains number of GSR-like particles for each sample, training results for both FSL networks, pair plot of graph metrics from the single modality network, and full results of clustering of metrics from the single modality network.

To view supplementary material for this article, please visit https://doi.org/10.1093/mam/ozae068.

Acknowledgments

The authors thank Dr. John Henry Scott of NIST and Dr. Nicholas Ritchie of NIST for insightful conversations on automated particle analysis. Pacific Northwest National Laboratory (PNNL) is operated for the United States Department of Energy by Battelle Memorial Institute under contract DE-AC05-76RL01830. This research was performed using resources available through Research Computing at PNNL.

Author Contributions Statement

J.A.B. and G.C. conceived of the idea. H.C. and B.G. provided interpretation of the experimental data. S.A. developed the multi-modal network. N.C.L. and S.G.A. implemented the graph metrics. J.E. trained the FSL network, computed the graph metrics, and performed clustering analysis. J.E. and J.A.B. developed the software. All authors contributed to the writing of the manuscript.

Financial Support

The authors gratefully acknowledge the Department of Defense for funding this exploration of ML-assisted particle analysis.

Conflict of Interest

The authors declare that they have no competing interest.

References

ASTM International (2017). *ASTM E1588-17: Standard Practice for Gunshot Residue Analysis by Scanning Electron Microscopy/Energy Dispersive X-Ray Spectrometry*. Conshohocken, PA, USA: ASTM International West. (n.d.)

Berger B, Peng J & Singh M (2013). Computational solutions for omics data. *Nat Rev Genet* **14**, 333–346. https://doi.org/10.1038/nrg3433

Burnett BR, Nunziata F & Gentile C (2021). Examination of firearm gunpowders by scanning electron microscopy/energy dispersive X-ray analysis. *J Forensic Sci* **66**, 709–718. https://doi.org/10.1111/1556-4029.14621

Charles S, Geusens N & Nys B (2023). Interpol review of gunshot residue 2019 to 2021. *Forensic Sci Int Synerg* **6**, 100302. https://doi.org/10.1016/j.fsisyn.2022.100302

Coley CW (2021). Defining and exploring chemical spaces. *Trends Chem* **3**, 133–145. https://doi.org/10.1016/j.trechm.2020.11.004

Craciun G & Feinberg M (2006). Multiple equilibria in complex chemical reaction networks: II. The species-reaction graph. *SIAM J Appl Math* **66**, 1321–1338. https://doi.org/10.1137/050634177

Dunn TB, Seabra GM, Kim TD, Juárez-Mercado KE, Li C, Medina-Franco JL & Miranda-Quintana RA (2022). Diversity and chemical library networks of large data sets. *J Chem Inf Model* **62**, 2186–2201. https://doi.org/10.1021/acs.jcim.1c01013

Feeney W, Vander Pyl C, Bell S & Trejos T (2020). Trends in composition, collection, persistence, and analysis of IGSR and OGSR: A review. *Forensic Chem* **19**, 100250. https://doi.org/10.1016/j.forc.2020.100250

Gal Y & Ghahramani Z (2016). Dropout as Bayesian approximation: representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning, PMLR, 48*, pp. 1050–1059. New York, NY.

Goldstein JI, Newbury DE, Michael JR, Ritchie NWM, Scott JHJ & Joy DC (2018). DTSA-II EDS software. In *Scanning Electron Microscopy and X-Ray Microanalysis*, Goldstein JI, Newbury DE, Michael JR, Ritchie NWM, Scott JHJ & Joy DC (Eds.), pp. 235–264. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-6676-9_17

Hayden L, Strausborger S & Lewin-Smith M (2023). Automated particle analysis using field-emission scanning electron microscopy (FE-SEM) and energy dispersive X-ray spectroscopy (EDS) to characterize inhaled particulate matter (PM) in biopsied lung tissue. *Microsc Microanal* **29**, 235–243. https://doi.org/10.1093/micmic/ozac015

Huber F, Ridder L, Verhoeven S, Spaaks JH, Diblen F, Rogers S & Van Der Hooft JJJ (2021*a*). Spec2Vec: Improved mass spectral similarity scoring through learning of structural relationships. *PLoS Comput Biol* **17**, e1008724. https://doi.org/10.1371/journal.pcbi.1008724

Huber F, van der Burg S, van der Hooft JJJ & Ridder L (2021*b*). MS2DeepScore: A novel deep learning similarity measure to compare tandem mass spectra. *J Cheminform* **13**, 84. https://doi.org/10.1186/s13321-021-00558-4

Huber F, Verhoeven S, Meijer C, Spreeuw H, Castilla EMV, Geng C, van der Hooft JJJ, Rogers S, Belloum A, Diblen F, *et al.* (2020). matchms-processing and similarity evaluation of mass spectrometry data. *J Open Source Softw* **5**(52), 2411. https://doi.org/10.21105/joss.02411

Krizhevsky A, Sutskever I & Hinton GE (2017). ImageNet classification with deep convolutional neural networks. *Commun ACM* **60**, 84–90. https://doi.org/10.1145/3065386

Lindstrom AP, Ritchie NW & Newbury DE (2013). Use of an automated SEM to detect laboratory contamination. *Microsc Microanal* **19**, 1066–1067. https://doi.org/10.1017/S1431927613007320

Liu FT, Ting KM & Zhou Z-H (2008). Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pp. 413–422. Pisa, Italy. https://doi.org/10.1109/ICDM.2008.17

Lowers HA, Breit GN, Strand M, Pillers RM, Meeker GP, Todorov TI, Plumlee GS, Wolf RE, Robinson M, Parr J, Miller R, Groshong S, Green F & Rose C (2018). Method to characterize inorganic particulates in lung tissue biopsies using field emission scanning electron microscopy. *Toxicol Mech Methods* **28**, 475–487. https://doi.org/10.1080/15376516.2018.1449042

Macindoe O & Richards W (2010). Graph comparison using fine structure analysis. In *2010 IEEE Second International Conference on Social Computing*. Minneapolis, MN: IEEE. https://doi.org/10.1109/SocialCom.2010.35

McAssey MP & Bijma F (2015). A clustering coefficient for complete weighted networks. *Netw Sci* **3**, 183–195. https://doi.org/10.1017/nws.2014.26

Ortiz-Montalvo DL, Vicenzi EP, Ritchie NW, Grissom CA, Livingston RA, Weldon-Yochim Z, Conny JM & Wight SA (2018). Chemical compound classification by elemental signatures in Castle dust using SEM automated X-ray particle analysis. *Microsc Microanal* **24**, 718–719. https://doi.org/10.1017/S1431927618004087

Pirrie D, Power MR, Rollinson GK, Wiltshire PEJ, Newberry J & Campbell HE (2009). Automated SEM-EDS (QEMSCAN®) mineral analysis in forensic soil investigations: Testing instrumental reproducibility. In *Criminal and Environmental Soil Forensics*, pp. 411–430. Dordrecht, NL: Springer. https://doi.org/10.1007/978-1-4020-9204-6_26

Pirrie D & Rollinson GK (2011). Unlocking the applications of automated mineral analysis. *Geology Today* **27**, 226–235. https://doi.org/10.1111/j.1365-2451.2011.00818.x

Ritchie NWM (2023). What is the best way to extract a k-ratio from an EDS spectrum? *Microsc Microanal* **29**(Supplement_1), 229–-230https://doi.org/10.1093/micmic/ozad067.102

Ritchie NWM, DeGaetano D, Edwards D, Niewoehner L, Platek F & Wyatt JM (2020). Proposed practices for validating the performance of instruments used for automated inorganic gunshot residue analysis. *Forensic Chem* **20**, 100252. https://doi.org/10.1016/j.forc.2020.100252

Ritchie NWM & Renolds A (2021). *Automated Particle Analysis (SEM/EDS) Data from Samples Known to Have Been Exposed to Gunshot Residue and from Samples Occasionally Mistaken for Gunshot Residue—Like Brake Dust and Fireworks*. National Institute of Standards and Technology. https://data.nist.gov/od/id/mds2-2476. Accessed March 30, 2023.

Routh PK, Marcella N & Frenkel AI (2023). Speciation of nanocatalysts using X-ray absorption spectroscopy assisted by machine learning. *J Phys Chem C* **127**, 5653–5662. https://doi.org/10.1021/acs.jpcc.3c00571

Scalfani VF, Patel VD & Fernandez AM (2022). Visualizing chemical space networks with RDKit and NetworkX. *J Cheminform* **14**, 87. https://doi.org/10.1186/s13321-022-00664-x

Schulz B, Sandmann D & Gilbricht S (2020). SEM-based automated mineralogy and its application in geo- and material sciences. *Minerals* **10**, 1004. https://doi.org/10.3390/min10111004

Shrivastava P, Jain VK & Nagpal S (2021). Gunshot residue detection technologies—A review. *Egypt J Forensic Sci* **11**, 11. https://doi.org/10.1186/s41935-021-00223-9

Tahirukaj M, Olluri B & Surleva A (2021). A study of the effect of working parameters and validation of SEM/EDS method for determination of elemental composition of commonly encountered GSR samples in shooting events in Kosovo. *J Forensic Sci* **66**, 2393–2404. https://doi.org/10.1111/1556-4029.14803

Torre C, Mattutino G, Vasino V & Robino C (2001). Brake linings: A source of non-GSR particles containing lead, barium, and antimony. *J Forensic Sci* **47**, 494–504. https://doi.org/10.1520/JFS2001093

Tucker W, Lucas N, Seyfang KE, Kirkbride KP & Popelka-Filcoff RS (2017). Gunshot residue and brakepads: Compositional and morphological considerations for forensic casework. *Forensic Sci Int* **270**, 76–82. https://doi.org/10.1016/j.forsciint.2016.11.024

von Luxburg U (2007). A tutorial on spectral clustering. *Stat Comput* **17**, 395–416. https://doi.org/10.1007/s11222-007-9033-z

Wang M, Tan K, Jia X, Wang X & Chen Y (2020). A deep siamese network with hybrid convolutional feature extraction module for change detection based on multi-sensor remote sensing images. *Remote Sens* **12**, 205. https://doi.org/10.3390/rs12020205

Wei W, Liao Y, Wang Y, Wang S, Du W, Lu H, Kong B, Yang H & Zhang Z (2022). Deep learning-based method for compound identification in NMR spectra of mixtures. *Molecules* **27**, 3653. https://doi.org/10.3390/molecules27123653

Zeigarnik AV, Temkin ON & Bonchev D (1996). Application of graph theory to chemical kinetics. 3. Topological specificity of multiroute reaction mechanisms. *J Chem Inf Comput Sci* **36**, 973–981. https://doi.org/10.1021/ci950268n