11-2023

Materials Characterization: Can Artificial Intelligence Be Used to Address Reproducibility Challenges?

Miu Lun Lau
Boise State University

Abraham Burleigh
Illinois Institute of Technology

Jeff Terry
Illinois Institute of Technology

Min Long
Boise State University

JVST A
Journal of Vacuum Science & Technology A

REVIEW

pubs.aip.org/avs/jva

Materials characterization: Can artificial intelligence be used to address reproducibility challenges?

Cite as: J. Vac. Sci. Technol. A 41, 060801 (2023); doi: 10.1116/6.0002809
Submitted: 3 May 2023 · Accepted: 20 September 2023 ·
Published Online: 3 November 2023

Miu Lun Lau,[1] Abraham Burleigh,[2] Jeff Terry,[2,3,4,5,a)] and Min Long[1,b)]

AFFILIATIONS

[1]Department of Computer Science, Boise State University, Boise, Idaho 83725
[2]Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
[3]Department of Mechanical, Materials, and Aerospace Engineering, Illinois Institute of Technology, Chicago, Illinois 60616
[4]Department of Social Sciences, Illinois Institute of Technology, Chicago, Illinois 60616
[5]Department of Biology, Illinois Institute of Technology, Chicago, Illinois 60616

Note: This paper is part of the Special Topic Collection: Reproducibility Challenges and Solutions II with a Focus on Surface and Interface Analysis.

[a)]**Electronic mail:** terryj@iit.edu
[b)]**Electronic mail:** minlong@boisestate.edu

ABSTRACT

Material characterization techniques are widely used to characterize the physical and chemical properties of materials at the nanoscale and, thus, play central roles in material scientific discoveries. However, the large and complex datasets generated by these techniques often require significant human effort to interpret and extract meaningful physicochemical insights. Artificial intelligence (AI) techniques such as machine learning (ML) have the potential to improve the efficiency and accuracy of surface analysis by automating data analysis and interpretation. In this perspective paper, we review the current role of AI in surface analysis and discuss its future potential to accelerate discoveries in surface science, materials science, and interface science. We highlight several applications where AI has already been used to analyze surface analysis data, including the identification of crystal structures from XRD data, analysis of XPS spectra for surface composition, and the interpretation of TEM and SEM images for particle morphology and size. We also discuss the challenges and opportunities associated with the integration of AI into surface analysis workflows. These include the need for large and diverse datasets for training ML models, the importance of feature selection and representation, and the potential for ML to enable new insights and discoveries by identifying patterns and relationships in complex datasets. Most importantly, AI analyzed data must not just find the best mathematical description of the data, but it must find the most physical and chemically meaningful results. In addition, the need for reproducibility in scientific research has become increasingly important in recent years. The advancement of AI, including both conventional and the increasing popular deep learning, is showing promise in addressing those challenges by enabling the execution and verification of scientific progress. By training models on large experimental datasets and providing automated analysis and data interpretation, AI can help to ensure that scientific results are reproducible and reliable. Although integration of knowledge and AI models must be considered for the transparency and interpretability of models, the incorporation of AI into the data collection and processing workflow will significantly enhance the efficiency and accuracy of various surface analysis techniques and deepen our understanding at an accelerated pace.

I. INTRODUCTION

Following the remarkable success of artificial intelligence (AI) within the past few decades, there has been rapid advancement of AI being applied to different fields, such as in natural language processing,[1–4] robotics,[5,6] and many others. In materials science, we are just beginning to see the use of AI in the analysis of material characterization data from techniques such as x-ray photoelectron spectroscopy (XPS), extended x-ray absorption fine structure

(EXAFS), and other material characterization techniques. The AI methods in material characterization can help in designing future experiments and in analyzing collected data, thereby allowing experimenters to better understand material properties and behavior under different conditions or to discover the optimal conditions for material synthesis. AI methods enable material scientists to analyze the large and diverse datasets generated by modern characterization techniques and/or models. The purpose of this paper specifically is to review the role and the future potential of various AI techniques in materials characterization. We will start by reviewing common AI methods and the current challenges of AI implementation (including assuring that the results are both physically and chemically relevant). We will finish with a discussion on the application of AI to materials characterization.

AI is the broad concept of computers and/or instruments containing computers being able to carry out tasks in ways that could be considered "intelligent." It is the science of making machines perform tasks that historically would have required the intelligence of humans to have been completed in the past. AI includes a range of technologies and methods, such as traditional rule-based systems, expert systems, machine learning, and more. Typical methods of AI include nearest neighbors,[7] Bayesian method,[8] symbolic programming,[9] and others. Alternatively, machine learning (ML) is defined as a subset of AI, which utilizes mathematical models to transform data into knowledge[10] without sending a computer direct instruction to learn. The input for a learning algorithm consists of training data, while the output comprises knowledge like guidelines for identifying cats in pictures. Typically, this knowledge is represented as a computer program capable of executing specific tasks. Deep learning (DL) is a subset of ML. It is a technique for implementing ML that is based on the structure and function of the brain (specifically, it mimics the workings of the human neural network using multiple layers of neurons). DL has attracted the most attention in the past decade due to applications in image recognition and language processing.[11,12] It utilizes artificial neural networks (NNs) of multiple layers, forming deep neural networks (DNNs), as a crucial manifestation[13] to discover disentangled representations in high-dimensional data,[14] enabling the software to learn new tasks autonomously instead of relying solely on programmers to create hand-crafted rules. AI, ML, and DL can be differentiated as follows: AI encompasses a board range of intelligence systems, whereas ML is a subset of AI that uses algorithms to learn from data and make intelligence predictions and decisions, while DL is a subset of ML that uses DNNs to mimic the human brain and achieve advanced pattern recognition tasks. These relationships are shown in Fig. 1, which highlights the major distinctions among them.

Implementation of the process of ML typically requires several key steps: (1) data selection and retrieval, (2) data cleaning and feature engineering, (3) model training and evaluation, and (4) model validation and deployment. Figure 2 shows those steps which will be discussed below.

Data collection. A suitable collected dataset representing a problem of interest needs to be curated. ML relies heavily on data to learn and make predictions, so the quality and relevance of the data can have a significant impact on the accuracy and effectiveness of the system. In order to identify appropriate and reliable data sources, it is paramount for the user to understand and define the problem that the ML system is intended to solve and the types of data that will be needed to address the problem. The retrieved raw datasets may not be able to be directly applied without eliminating possible data bias or errors. This step involves operations like cleaning and transforming the data, dealing with missing or inconsistent (e.g., unphysical) values, and handling outliers. The cleaning and identifying of each dataset will ensure that the data are suitable for use in ML models and that the data are actually a proper representative of the model. Additionally, the curation of the dataset might need to be performed multiple times in order to have a complete enough dataset to reach the desired model performance and accuracy.

Model training and evaluation. A mathematical model can be trained using the previously constructed dataset. One must carefully select a model representative of their problem, such as a classification model for images or a regression model for predictions. During training, the user must actively monitor the training process by tracking the model's metrics, such as loss, accuracy, and performance; and then analyze any trends that may emerge to make necessary adjustments to improve the model. It is also required to tune a large amount of hyperparameters that exist in many models in order to obtain accurate results. One must perform experiments on/with the optimal hyperparameter values to test them. These steps can be repeated multiple times until a satisfactory result is obtained.

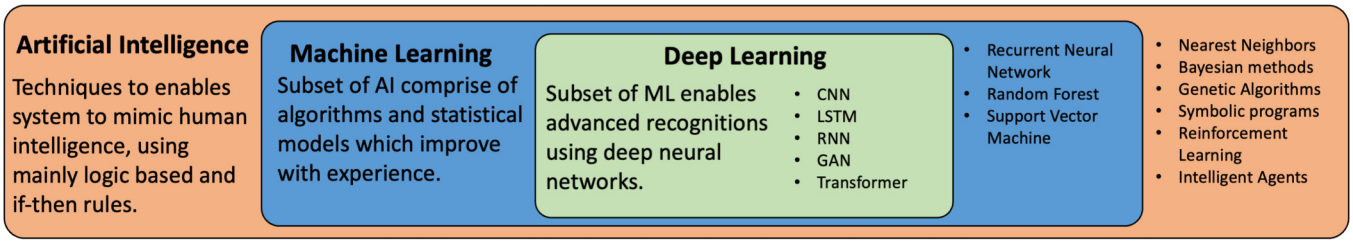

FIG. 1. Distinctions between AI, ML, and DL: AI encompasses all intelligent systems, ML is a subset of AI that uses algorithms to learn from data and make predictions, while DL is a subset of ML that uses DNNs to mimic the human brain and achieve advanced pattern recognition tasks. Examples include support vector machine and regression as AI; recurrent neural network as ML; and CNN, long short-term memory (LSTM), recurrent neural networks, and generative adversarial networks (GNNs) as DL.

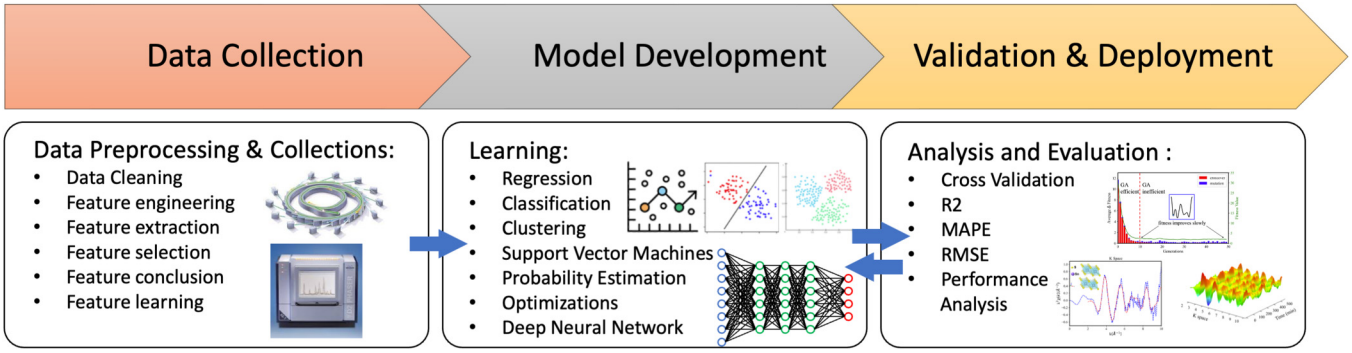

FIG. 2. Flow chart of ML in scientific discovery. In *data collection*, scientific data from experiments are processed, cleaned, and selected for the next step. In *model development*, the data are trained depending on the data and desired output (i.e., prediction of material properties will be regression, crystal structure identification will be classification). After sufficient model development, the model will need to be validated and deployed for prediction. See the text for description of acronyms.

Model deployment. Finally, the trained model needs to be deployed into standard use. To obtain meaningful results and retain the model's high quality of performance, accuracy, and efficiency, the model must be continually trained, tested, and validated. The trained model needs to be continually refined during deployment.

An example of this ML development process is shown in Fig. 2. As the model would likely be put into practice by users with limited knowledge of the model, multiple metrics need to be monitored to ensure there is no drop in accuracy or performance while in operation. Finally, if new data become available or if the performance of the model deteriorates, the developer must retrain using the new data and deploy the new model. The model may be evaluated[15] using the standard computer science techniques of cross-validation, R-squared (R2), mean absolute percentage error, root mean square error (RMSE), and performance analysis.

A. Challenges to implementation

There are several challenges in the AI and ML-facilitated scientific discovery process. First, the quality and quantity of the available datasets influence the understanding of physicochemical processes that can be explained by the characterization data.

One of the key challenges that must be overcome to enable more use of AI tools is data quality and bias. While AI systems rely heavily on data to learn and make predictions, if the data are incomplete, inaccurate, or biased, it can cause system predictions to be inaccurate.[16] Additionally, most modern AI models can be complex and difficult to interpret, making it extremely challenging to understand how the system is making decisions. This effectively turning the system into a blackbox.[17] Other challenges include overfitting and generalization, which can lead to poor performance on new data, as well as the need for significant computing resources to train and run some AI models. Overall, these challenges highlight the need for careful planning, evaluation, and monitoring when using AI methods. In addition, AI does have the added complication of there being ongoing efforts to address ethical and societal concerns related to these technologies.

A number of techniques have been developed to address the issues of data quality and bias. These techniques do take different approaches to address the need for high quality, training data. Data augmentation[18] creates new training data by applying transformations to existing data. Imputation[19] can fill in missing values with estimated values based on the available data. Transfer learning uses a pretrained model on a related task[20] to improve the model performance. Ensemble methods combine multiple models to reduce the impact of missing data. Active learning selects the most informative data points for labeling by a human expert to improve model performance.[21] Finally, Bayesian methods[22] incorporate prior knowledge into the model to reduce the impact of missing data.

Second, data imbalance is a problem in classification tasks where there is an unequal distribution of instances across classes of interests. Several techniques are employed to address this challenge. For example, the resampling method aims to obtain a balanced dataset via either oversampling the minority class, undersampling the majority class, or a combination of both. Oversampling involves creating duplicates of the minority class instances, while undersampling involves removing instances from the majority class. However, both of these techniques can lead to cases of overfitting or underfitting, respectively. Synthetic data generation methods[23] can also be used to create additional instances of the minority class, such as Synthetic Minority Over-Sampling Technique,[24] which creates synthetic instances by interpolating between existing instances in the minority class. Cost-sensitive learning[25] assigns different costs to different classes to adjust the bias of the model with respect to the minority class. This can be done by increasing the cost of misclassifying the minority class or decreasing the cost of misclassifying the majority class. Ensemble methods[26] combine multiple models to improve the overall performance of the model. The models can be trained on different subsets of the data or with different algorithms. Anomaly detection techniques[27] can be used to identify the minority class as an anomaly and classify the data accordingly and can be done using techniques such as one-class support vector machine (SVM) or isolation forests.

B. Challenges to the scientific community: Reproducibility

Two of the major issues affecting the scientific community include the lack of reproducibility[28] and the large amount of experimental data being produced by modern characterization tools.[29] In materials characterization, both of these issues limit the utility of the data collected during the experiments such as during the operation of user facilities such as synchrotron radiation facilities. This hurts the scientific community as there is a possibility that important science is lost. Whether data are lost due to inactivity or were improperly analyzed, the knowledge that is lost cannot be used to improve our understanding of the physical world. At worst, time is wasted trying to replicate incorrect conclusions.[30]

Recent reviews of the scientific literature have attempted to quantify the size of the reproducibility problem.[31–33] These reviews have examined the quality of the XPS analyses in three upper tier journals that emphasize next-generation materials synthesis and characterization. They reviewed published data for accuracy and completeness. They found that ca. 30% of the XPS analyses in these journals were incorrect or interpreted wrongly. Based on other reports, it is not uncommon to similarly find that 20%–30% of the analyses of data from other materials characterization techniques are also incorrect.[34,35] However, the actual problem is probably even more severe than these numbers suggest. Given the proliferation of incorrect analyses appearing in the literature,[30,32,34–36] it appears that experts in the synthesis and/or development of a particular type of material may not always possess a detailed understanding of all the analytical methods that may have been used to characterize them. Thus, the structure, traditions, constraints, and pressures of the current scientific endeavor often lead to the publication of faulty or misleading data analysis.[36,37]

Increasing the use of AI can make significant contributions toward addressing the reproducibility problem in a number of ways. First, AI-based automated experiment replication can facilitate analyzing datasets to identify the necessary steps, parameters, and conditions required to reproduce an experiment accurately. By automating this process, researchers can verify and validate previous findings, enhancing reproducibility. Second, AI can provide insights into the decision-making process of very complex models. By combining theory and characterization, AI can improve our understanding of materials synthesis. For instance, in synchrotron studies, it is crucial to employ autonomous data processing and data fitting techniques. By combining data fitting with predictive simulations, researchers can gain deeper insights into the three-dimensional structure of materials.[38] This integrated approach accelerates the development of predictive simulations by incorporating new theories, algorithms, and information. The ultimate goal is to establish a close connection between predictive theory and experiment, reducing the need for trial-and-error measurements and expediting the pace of discovery and understanding.

C. Challenges to the scientific community: Data quantity

Further complicating these issues is the fact that advances in instrumentation have made it possible to acquire data orders of magnitude more rapidly than it can be analyzed.[29,39] It is now possible to collect data in real time during a chemical reaction on the surface of a catalyst or in a battery cell, or to monitor the deposition of a thin film or catalyst with multiple techniques. In many cases, thousands of spectra characterizing the chemical reactions or thin film depositions must be analyzed. The increasingly sophisticated imaging modes of many instruments, e.g., time-of-flight secondary ion mass spectrometry (ToF-SIMS), have similarly led to an enormous increase in the amount of data collected on samples. A standard ToF-SIMS image of a surface can easily contain 10 000 mass spectra. This number then increases substantially when this technique is used for depth profiling. These vast quantities of data greatly increase the likelihood that novice users of the characterization tools are employed to analyze the data. This increases the likelihood of making a mistake that enters into and then propagates through the literature.

D. Role of artificial intelligence

There are three key AI research pathways that must be addressed for AI techniques to have significant impact in addressing the aforementioned challenges. AI techniques need significant improvement in the areas of (i) content and reliability analysis of the scientific literature; (ii) automated analysis and interpretation of results of characterization experiments; and preferably, (iii) automated experimental design for reliable materials characterization. The remainder of this section outlines the components of these pathways and their interrelationships.

The first pathway involves extracting experimental parameters and results from the peer-reviewed literature and assessing the validity and self-consistency of the information in each individual paper. Because of the amount of incorrect data within the literature, it will be necessary to continually assess the database and reevaluate the interpretation of the individual papers used to populate a database of accurate characterization results. Once the database is populated, information can be pulled from the database to improve and automate the collection and analysis of characterization of these new materials.

The second pathway for automated characterization analysis of experimental results requires one to develop strong combinatorial search methods among all possible combinations of experimental parameters to find the one(s) that most likely produced the observed signal. This often requires two stages: (a) a prefilter to determine a relatively small set of potential components to narrow the phase or parameter space that must be considered, and then (b) intelligent combinatorial search among them for the set of components giving rise to the observed signal.

For the third pathway of automated experimental design, we expect that artificial intelligence tools will need to try multiple complementary approaches. A knowledge-based approach can extract experimental design concepts and rules from subject matter experts that can be used to choose appropriate characterization techniques and parameters for a given case. A case-based approach will find similar cases in Materials Characterization Databases and suggest similar experimental parameters.

Unfortunately, at this time, our review indicates that none of the available tools fully reach the level of meeting the goals of all three pathways that we see being necessary for full AI automated

analysis. Hence, the goal of this review is to highlight the current applications of ML in various materials characterization methods and to assess whether the current AI methods are suitable for addressing the reproducibility issues currently plaguing scientific progress. The use of AI techniques needs to prevent the unwanted result of conclusions based upon unsatisfactory analysis being propagate throughout the literature and to enhance the peer-review process that helps advance science. This review paper is organized as follows: Section II introduces several materials characterization techniques. Section III reviews ML applications to those techniques. Finally, Sec. IV presents our discussion and conclusion.

II. MATERIAL CHARACTERIZATION TECHNIQUES

We selected the following set of materials characterization techniques for this review of AI analysis applications: XPS, x-ray absorption near edge structure (XANES), EXAFS, x-ray diffraction (XRD), x-ray emission spectroscopy (XES), nanoindentation, scanning electron microscopy (SEM), energy dispersive x-ray spectroscopy (EDS), transmission electron microscopy (TEM), grazing-incidence small-angle x-ray scattering (GISAXS), grazing-incidence wide-angle x-ray scattering (GIWAXS), and neutron scattering. These techniques are widely utilized in the characterization of solid, liquid, and gaseous reaction products and share the base analysis process of fitting measured spectra with theoretically informed line shapes. This process makes them amenable to analysis with a combinatorial search in physically allowed parameter space. Remember that the goal of materials characterization analysis is to determine the best chemically relevant fit and not necessarily the best mathematical fit.

A. XPS

In conventional photoemission (XPS), x rays of known energy are directed onto a sample and the ejected electrons are detected. One of the most fundamental equations of XPS is a restatement of the principle of conservation of energy,

$$h\nu = E_B + E_k + \phi_{\text{spectrometer}}, \tag{1}$$

where $h\nu$ is the energy of incident x ray, E_B is the binding energy of the electron, E_k is the kinetic energy of the electron, and $\phi_{\text{spectrometer}}$ is the work function of the spectrometer. In core level photoemission, the last term is usually relatively small (about 4.5 eV) compared to the E_B and E_k. One of two types of XPS scans is usually acquired: a survey (broad) scan over ca. 1000 eV or a narrow (detail) scan over ca. 20 eV. The survey scan is useful for identifying the elements in the sample. Narrow scans are important for obtaining chemical information from the material. That is, while the peak positions (E_B) of XPS peaks identify the elements in a sample, finer energy shifts (usually about 1–4 eV), known as chemical shifts, encode physicochemical information within the measurement. In many problems, these chemical shifts are initial state effects that result from the oxidation state of the element in question.[40] In general, peak fitting is required in XPS data analysis because the fundamental widths of XPS peaks are comparable to their chemical shifts and first principles theory has not been sufficient to calculate the spectra.[41] Multiple software packages have been developed to facilitate the XPS fitting process, such as CASAXPS,[42] AVANTAGE,[43] FITXPS,[44] AANALYZER,[45] XPS MULTIQUANT,[46] and XPS VIEWER.[47]

B. XAS: XANES and EXAFS

The x-ray absorption spectra (XAS) of a material typically consist of two distinct regions: (1) the XANES region and (2) the EXAFS region. The XANES region measures the absorption just below and just above the edge in a range of 50 eV. As a result, XANES is used primarily for analysis of the formal oxidation state and coordination chemistry.[48–50] The EXAFS region is used to understand the local atomic structure of materials. It has been used to study materials from a myriad of applications.[49] These include materials synthesis and design of batteries,[51] assessment of material performance for energy materials,[52] and evaluation of radiation damage for nuclear energy applications.[53] An XAS experiment measures the photon absorption as a function of photon energy in the x-ray region of the electromagnetic spectrum. At the energies of typical XAS experiments, the measured spectrum is dominated by the photoabsorption component.[53–55] However, when the absorbing atom has other atoms located around it, the x-ray absorption spectrum has additional oscillations beyond the edge.[53,55] These additional oscillations are called the fine structure. The fine structure in the EXAFS region encodes the local atomic structure around the absorbing atom of chemical compounds being measured.

EXAFS data is an excellent candidate for structural refinement using AI methods because small changes in the path parameters result in large changes in the theoretical path model. Unfortunately, there is some art in this analysis and it can result in a complicated fitting process. Manual fitting of XANES/EXAFS is challenging and requires a good understanding of the theory and practice of XANES/EXAFS analysis. The three most commonly used theoretical programs for calculating XAS are FEFF (Full Multiple Scattering),[56] GNXAS (N-body Green's function),[57] and EXCURVE (Exact Curved Wave).[58,59] First, the analyst must decide which of the relevant x-ray physics must be included. Second, they need to initialize basic parameters and apply material dependency constraints. Third, an optimization process is required, such as, least-squared; cubic splines least squared; splines smoothing;[60,61] or nonlinear optimization such as Levenberg-Marquardt.[62] There are multiple software suites exist and can be used in the fitting process, such as ATHENA,[63] WINXAS,[64] IFEFFIT,[56] and LARCH.[65] Lastly, the fitted result still needs to be validated against other methods and the relevant statistics need to be analyzed to ensure the spectra are well fitted and are consistent with the proposed model.

C. XES

XES is a materials characterization technique used to study the electronic structure and chemical states of elements in materials. It is similar to XAS but instead it focuses on analyzing the emission of characteristic x-ray photons by the elements in a sample. XES is a relaxation process after excitation.[50,66]

The x-ray photons that are detected are generated from the refilling of a core hole by one of the outer-shell electrons of the atoms in the sample. From the emitted x ray, information about the chemical states and electronic structure of the sample can be obtained. The energy of the emitted x-ray photons directly

corresponds to the difference in energy levels of the electrons involved in the transitions, which depend on the specific elements present in the sample and the surrounding chemical environment.[67]

XES can provide valuable insights into the oxidation states, coordination environments, and electronic configurations of atoms in a material. By analyzing the energy and intensity of the emitted x rays, researchers can determine the materials' chemical composition, bonding, and electronic properties. XES spectra are often fit using more generic programs designed for fitting other data types.

D. XRD

XRD is used primary for phase identification of crystalline materials. It is based on the principles of crystallography, which relate the diffraction of x rays to the arrangement of atoms in a crystal lattice. The mechanism of XRD is a beam of x rays impinges upon a sample, and one measures the outgoing pattern of x rays scattered by the atoms in the sample. By analyzing the positions and intensities of the peaks, information about the crystal structure, lattice parameters, and atomic positions of the material can be obtained.

XRD is widely used in many fields, including materials science, chemistry, geology, and biology to analyze a wide range of materials, including metals, ceramics, polymers, and pharmaceuticals.[68–72] In addition to determining the crystal structure of materials, XRD can also be used to quantify the amount or concentration of different components in a sample. This is achieved by comparing the intensity of the diffraction peaks to a calibration curve or reference standard.[73]

Typical analysis software for XRD involves shape fitting of each excitation peak, which corresponds to an element or compound. There exist a large variety of software analysis packages for XRD, such as FULLPROF,[74] X'PERT HIGHSCORE PLUS,[75] TOPAS,[76] and GSAS II.[77] Peak libraries are typically included in each software, which contains diffraction patterns for common materials. The patterns are generated using theoretical calculations or experimental measurements and represent the diffraction behavior of the materials under specific XRD conditions.

E. Nanoindentation

Nanoindentation is a materials characterization technique commonly used to probe mechanical properties by extracting information from very shallow regions of a sample on the order of micrometers.[78] It uses an indenter with a known material hardness and performs many indents on one sample to retrieve a statistical mean of the hardness and elastic modulus values from that region of the sample. These values can be calculated using the material stiffness extracted from fitting the unloading curve using the Oliver–Pharr method,[79] or by analyzing oscillations of the tip during loading by continuous stiffness measurement.[80] When a material undergoes nanoindentation, two types of material deformation occurs: plastic and elastic. During elastic deformation, the material is able to fully recover its original shape (for sufficiently small indents) resulting in an unloading curve that closely traces the loading curve. At the other extreme, a completely plastic deformation results in an unloading curve that is nearly vertical as the material fails to recover leaving a large residual indent.

Nanoindentation is useful for characterizing the mechanical properties of materials such as thin films, irradiated materials, coatings, and composites, where traditional mechanical testing methods may not be suitable.[81–83] It is also used to study the effect of different processing parameters on a material's mechanical properties.[84]

F. SEM

SEM is used primarily to provide information on the surface topology and thereby uncover its crystalline structure, chemical composition, and electrical properties.[85,86] The relatively low energy of the electron (2–40 keV) only allows observations of the top 1 μm. A common technique for microstructural characterization in the SEM is electron backscatter diffraction (EBSD). It generates a diffraction pattern (DP) from a sample surface tilted at $\approx$20°–30° away from the electron beam to reduce the path length and increase the number of detected electrons undergoing diffraction.[87,88] This restricts the sampled volume to around 20 nm in depth from the surface. In addition, EBSD measures the microtexture of the material with a spatial resolution of 100 (Ref. 89) to 500 nm[87] depending on factors such as accelerating voltage and probe current as well as material properties.[88] The most commonly extracted information from these EBSD measurements is phase identification and orientation mapping, which help to understand grain sizes, crystal structures, and grain orientation. Another common technique in SEM analysis is EDS, which is used extensively for analyzing the atomic composition of samples for concentrations as low as 250 PPM for many elements.[90]

The obtained SEM images can be analyzed through software[91] to extract features about the material. For example, DIGITALMICROGRAPH[92] is commonly used in analysis if EDS is involved; AVIZO[93] or ZEISS ZEN[94] is used for 3D SEM results. ASTROEBSD[95] is a good package for analysis of EBSD data.

G. TEM

TEM is a powerful technique used to study the structure and composition of materials at the nanoscale level. TEM operates by firing an electron beam through a sufficiently thin sample (<100 nm).[85] The electrons are focused by a series of electromagnetic lenses, and the resulting image is formed by the electrons that pass through the sample and focused onto a fluorescent screen or a digital detector.[96] Due to the interactions of the electrons with the atoms in the crystal, a visual image of the internal structure of the crystal is obtained.

TEM provides much greater magnification than does SEM, allowing examination of the sample's crystal structure, defects, and chemical composition of materials with much greater spatial precision. TEM is widely used in materials science, biology, and other fields to investigate the structure and properties of a wide range of materials, including metals, semiconductors, ceramics, polymers, and biological materials.[97–102] TEM is also used to study the behavior of materials under different conditions, such as high temperatures or high pressures,[103,104] and to understand the mechanisms of various physical and chemical processes. Techniques such as selected area electron diffraction (SAED) and high resolution electron microscopy (HREM) enable site specific examination of a material's microstructure. SAED allows for the determination of

lattice spacing and crystal orientation from regions smaller than 500 nm across.[105] The crystal lattice planes of the same region of interest can be directly imaged using HREM for identification of defect structures and determination of their concentrations.[106,107] There are many software packages[108] used to facilitate the analysis of TEM, such as IMAGE PLUS PRO,[109] AVIZO,[93] DIGITALMICROGRAPH,[110] EMAN2,[111] and IMOD.[112]

H. GISAXS/GIWAXS

GISAXS (Refs. 113–115) and GIWAXS (Refs. 114 and 116) are noninvasive surface-sensitive techniques that use x-ray scattering to probe nanoscale structures at surfaces and interfaces. The incident angle is what separates GISAS and GIWAS apart and dictates what kind of material information can be obtained. In GISAXS, a monochromatic x-ray beam is incident on the sample surface at a very shallow angle ($<0.5°$). This ensures that the penetration depth remains small, probing only the top surface layer of the sample. In GIWAS, the incident angle is set at much wider (typically 0.5°–5°), and it is typically used to study a more bulk-representative crystal structure. GISAXS is highly sensitive to lateral density fluctuations and roughness at surfaces and interfaces, and therefore special care must be taken during sample preparation. The result is that GISAXS can characterize nanoscale features like roughness, domains, particles, and lamellae. GIWAXS usually requires much less preparation and care when preparing the sample.

The GISAXS diffraction pattern contains a wealth of structural information that can be extracted through analysis and modeling. The analysis can use the measured peak positions, intensities, and widths to extract important structural parameters such as domain sizes, roughness, and crystalline phases. GISAXS can probe samples in different environments (air, liquid) and under external stimuli (heating, shear flow). The use of GISAXS has been used to characterize the structure of thin films,[117] semiconductors,[118] solar cells,[119] polymers,[120] biological materials,[121] and nanoparticle growth.[122,123]

GIWAXS is used for studying the crystal structure of thin films, fibers, and other materials that have a preferred orientation or alignment. It can provide information about the degree of crystallinity, the size and shape of individual crystals or domains, the orientation and texture of the crystal lattice, and the presence of defects or disorder in the crystal structure. Like GISAXS, GIWAXS is also sensitive to changes in the crystal structure that occur as a result of external stimuli, such as temperature, pressure, or chemical reactions.

GISAXS provides quantitative insights into roughness, patterning, and particle distributions. It provides insight at surfaces, interfaces, and buried interfaces for studying surface chemistry. GIWAXS provide insights into crystalline and crystal orientation. In addition, these method are typically use in conjunction to other surface techniques to further understand and verify the material. Software packages used for this diffraction pattern analysis include BORNAGAIN,[124] GISAGUI,[125] FIT2D,[126] MCSAS,[127] PYFAI,[128] and CRYSTFEL.[129]

I. Neutron scattering

Neutron scattering is a powerful technique used in the field of materials science to study the structure and properties of materials at the atomic and molecular level.[130–132] It involves directing a beam of neutrons at a sample and analyzing the energy, intensity, and direction from which neutrons are scattered by the material.

Neutron scattering has a wide range of applications in materials research, including the study of metals,[133] polymers,[134] ceramics,[135] and biological materials.[136] Neutron scattering is a valuable tool for advancing our understanding of the fundamental properties of matter and for developing new technologies that benefit society. Neutron scattering has numerous advantages that make it a valuable tool for studying the structure and properties of materials. For example, the wavelength of thermal neutrons is comparable to the interatomic spacing in materials, which allows for a detailed analysis of the atomic structure. Additionally, neutrons have a kinetic energy comparable to that of atoms in a solid, which makes them sensitive to the motions and vibrations of atoms within a material.

Another advantage of neutron scattering is that neutrons are highly penetrating, allowing experiments to provide insight into the bulk materials' properties. This also allows the sample to be contained, making it useful for studying hazardous or radioactive materials. Furthermore, the weak interaction between neutrons and matter aids in the interpretation of scattering data. Neutron scattering also has isotopic sensitivity, which allows for contrast variation in materials, making it easier to distinguish between different elements. Additionally, the neutron magnetic moment couples to the magnetic field, allowing neutrons to "see" unpaired electron spins.

However, there are also some disadvantages to using neutron scattering. One major disadvantage is that neutron interactions are weak, which results in low signals and the need for large samples. Additionally, some elements, such as cadmium, boron, and gadolinium, absorb neutrons strongly, which limits the materials that can be studied.[132] Finally, there are kinematic restrictions that prevent access to all energy and momentum transfers, which can limit the range of materials and phenomena that can be studied using neutron scattering. Popular software for the analysis of neutron scattering data include FULLPROF[74] and GSAS II.[77]

III. AI APPLICATIONS IN MATERIALS CHARACTERIZATION

The adoption of AI methods by the materials characterization community has enabled us to rapidly identify patterns and make predictions based on the input data.[137] Specifically, users can analyze large and increasing amounts of data and extract relevant features or properties that may be too difficult or time-consuming to identify manually. Given the increase in productivity, ML has the potential to revolutionize materials characterization by enabling faster and more accurate analysis of complex datasets. This is particularly important in the big data era when data are produced orders of magnitude more rapidly than it can be analyzed manually.

A. XPS

AI methods can accelerate the identification of the potential components that exist in the XPS spectrum with the least amount of human intervention. Many early attempts use the fitting of the raw spectrum combined with various basic line shape functions,

i.e., Gaussian and Lorentzian, to automate the fitting process, or by comparing data against quantitative results. However, recent attempts have applied DNNs to fit XPS spectra.

Drera *et al.*[139] demonstrated the use of a DNN in identifying and quantifying XPS survey spectra. They employed a convolutional neural network (CNN) to analyze the chemical composition of multicomponent XPS by surveying data from a spectral library and performing model training on synthetic spectra. The training data consisted of approximately 100 000 XPS spectra, and model training used electron scattering theory in the transport approximation, which generalized well to about 500 well-characterized experimental examples. Although their work was based on synthetic spectra, the technique demonstrated the potential future where it is possible to directly fit intricate experiential spectra.

Aarva *et al.*[140,141] used open source data available from XAS/XPS database to construct an interatomic potential mirroring density functional theory (DFT). They then applied AI data clustering algorithms to atomic motifs for quantifying spectral analysis for fingerprinting the spectrum. Clustering algorithms are a type of unsupervised learning for pattern identification.[142] The database from which they obtained the training data contained a total of 240 samples containing differing amounts of oxygen in the range from 10 to 20 at. %. They ran multiple AI trials using the clustering algorithm, which results in the multiple predicted spectra shown for each sample in Fig. 3.

Similarly, Park *et al.*[143] also proposed the use of CNN to perform human-free analyses of one-dimensional spectroscopic data, such as data in the frequency domain. In this study, a modified novel network model was applied to experimental XPS of graphene, MoS_2, and WS_2. Six different network architectures were trained and compared, with the proposed "squeeze-and-excitation" network structure (SENet) having demonstrating the best performance in material property predictions. The study also examined the ideal relationship between training performance and the choice of loss function.

Golze *et al.*[138] applied ML algorithms to XPS data and developed a predictive model for carbon-based material. Two CHO-(carbon, hydrogen, oxygen) based structure databases were created for XPS training purposes: one for large CHO materials and one for small CHO-containing molecules. Using the libraries, multiple ML methods such as principal component analysis (PCA) and partial least squares regression (PLSR) were used to analyze the XPS data and develop a predictive model. PCA (see Chatterjee *et al.*[144] for more on PCA) was first used to reduce the dimensionality of the XPS data into their basic components, and afterward PLSR was applied to develop a model that could predict the thickness and composition of the carbon-based material. To validate the model's accuracy, cross-validation was applied to the model, and the model was able to successfully predict binding energy. It was also capable of generating synthetic XPS spectra for CHO from the model, enabling rapid analysis.

Figure 3 shows the comparison between the computational and ML prediction models for XPS predictions of C 1S spectra of rGO. The top panel represents predictions based on a starting precursor with COOH-rich graphene oxide (GO), and the bottom panel shows predictions based on a starting precursor with OH-rich GO. Series 1(a) and Series 2(b) demonstrate, respectively, the variations in predictions based on the different precursor compositions. Currently, it is not clear what the best AI methodology for fitting XPS spectra will be. XPS may be the most difficult of the material characterization tools for AI analysis to handle due to all the potential complications of the physics of the measurement. AI analysis of XPS will need to be able to address spin–orbit splitting, Coster–Kronig broadening,[145] and multiple types of inelastic scattering. We are not surprised that majority of the AI work done on XPS to date has focused on simpler 1s based analysis.

B. EXAFS

The objective of EXAFS analysis is to account for the changes in the absorption spectrum from the scattering of electrons from near-neighbor atoms, this in turn provides details about the surrounding crystal structure, such as coordination numbers and Debye–Waller factors.[146] The main challenge in analyzing EXAFS spectra comes about from trying to understand all of the potential physical differences in the structure that can effect these scattering events, which can make it difficult to extract accurate structural information.[147] Another challenge is the availability of data compared to other material characterization techniques. EXAFS data can be difficult and time-consuming to acquire. Historically, it required the use of a synchrotron facility, and this has limited the amount of data that is available for ML training. However, recent advances in data sharing and open science have led to the availability of large and diverse datasets that can be used for ML applications.[148] Manual modeling of EXAFS spectrum is typically performed with small numbers of short-distance scattering paths. However, there have been many attempts in accelerating the analysis process via the use of AI techniques to enable rapid analysis of EXAFS spectra using high-throughput pipeline.

Terry *et al.*[149] demonstrated the use of a genetic algorithm (GA) based software, EXAFS NEO, to quantify EXAFS data. The GA-based technique has also been applied to large scale sets of EXAFS spectra to examine *in situ* structural changes. This has included applying the analysis pipeline to Li-ion batteries where it was used to observe the process of Li insertion and extraction during the cycling process. This analysis pipeline was also able to identify patterns and extract meaningful material structural data from extreme noisy data. A surface plot of SnS_2 of Li-Sn scattering path batteries as a k-space and time is shown in Fig. 4 to demonstrate the changes in phase amplitude as the battery cycles.

Timoshenko *et al.*[150] have applied DL techniques to construct an NN to extract EXAFS parameters for metals and oxide materials. The model was applied to tackle the challenging problem of monitoring the catalyst structure and composition under *operando* conditions. The usage of an NN can allow for the processing of a large number of experimental spectra and make possible time sensitive analysis such as the time-dependent evolution of the local structure around the catalytic active species.

Additionally, Martini *et al.*[151] examined the use of an inverse AI method to study the photoelectron backscattering phases and extract the amplitudes of single and multiple scattering paths. The AI model was developed based on the method of Timoshenko *et al.*[150] and has been extended for use in the EXAFS case and generalized for different configurations. The result has been applied to

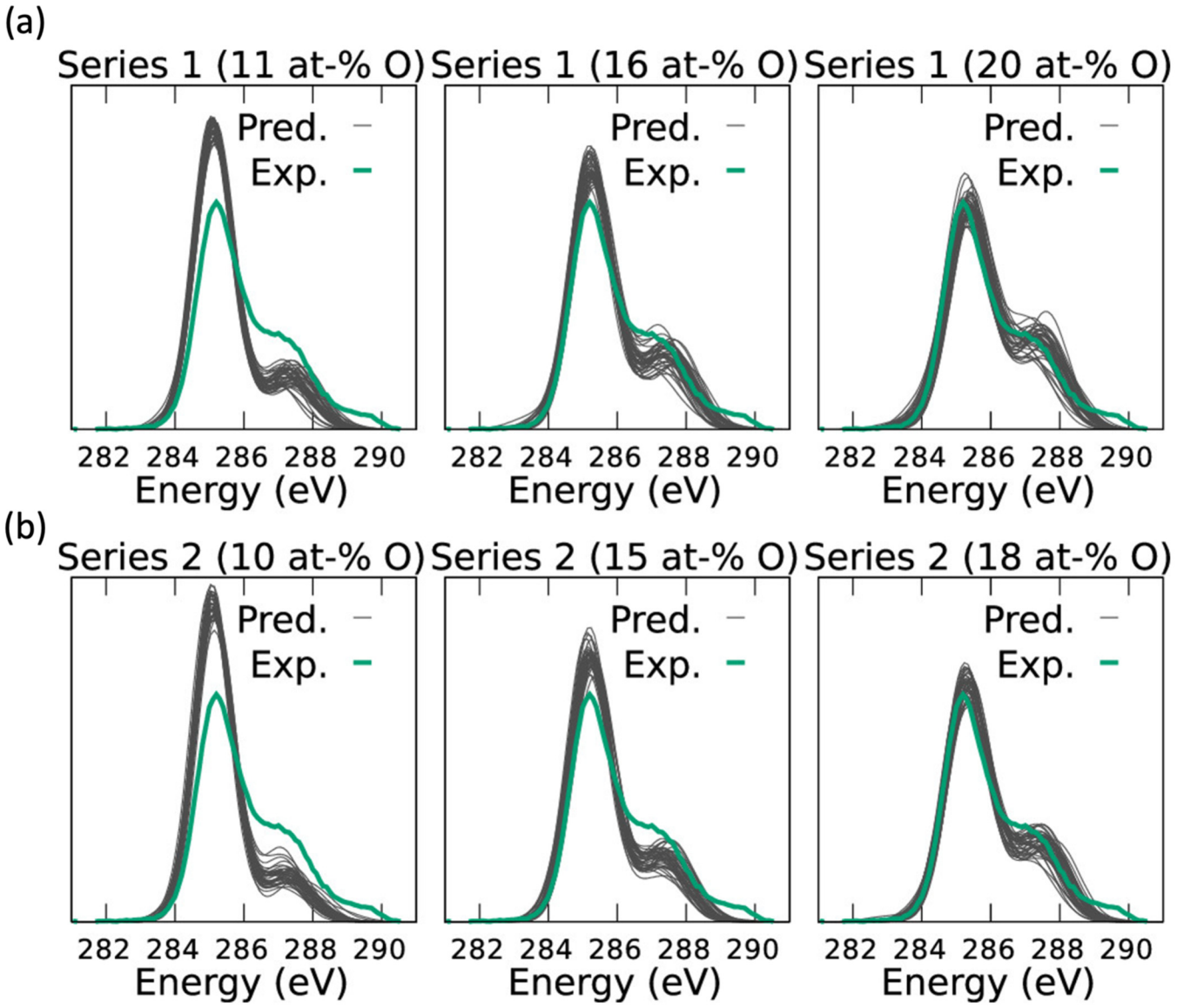

FIG. 3. XPS predictions of C 1s spectra of rGO (reduced graphene oxide) between computational and ML prediction model (Ref. 138). The top and bottom panels represent different starting precursors. Series 1(a) was generated from COOH-rich GO and series 2(b) was from OH-rich GO. Reprinted from Aarva *et al.*, Chem. Mater. **31**, 9243 (2019). Copyright 2019 Chemistry of Materials, licensed under a Creative Commons license.

$KAu(Cn_2)$ and $[RuCl_2(CO)_3]_2$ and demonstrated the ability to extract theoretical phases, amplitude, and electron free paths based on the input geometry alone. The inverse method presented in this paper was separated into two sections: (1) identifying the set of paths with the highest impact to the total EXAFS signal and (2) generating a set of nonlinear functions to approximate the contribution of each path. Once the most significant paths are selected, the fitting with the signal parameters is performed via minimizing the sum of squares between theory and signal data. Figure 5 shows an example using the inverse method for the difference between the experimental fits and the best fit. This shows (a) the modulus without phase correction and (b) the imaginary component of both the experimental and best fit EXAFS spectra for $KAu(CN)_2$. These results demonstrate the efficacy of AI in accurately fitting experimental data. Additionally, the lower part displays the specific contributions from each scattering path.

Finally, Kido *et al.*[152] have developed an automatic collection and visualization pipeline of EXAFS analysis and applied the

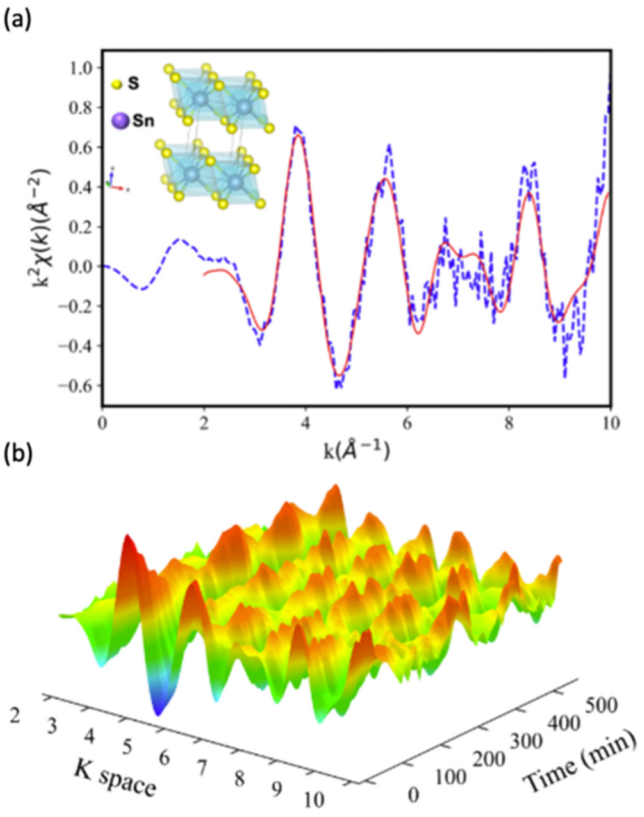

FIG. 4. (a) Single snapshot of the GA automatically fitted model (solid line) and raw spectra of SnS_2 (dashed line) at approximately 125 min of a battery cycle. (b) Surface plot of SnS_2 battery during *in situ* measurement from Terry *et al.* (Ref. 149). The x axis is K space and y axis is cycling time. A battery cycling can be observed as the Li ion is being depleted (highest regions) and inserted (minimum regions).

pipeline to MoO_3 analysis. This analysis utilized "Thorough Search," which examines all possible rational structure bounded by the limits of the experiments. The specific AI method utilizes both K-mean and PCA to perform unsupervised learning. The pipeline has the potential to be applied on other materials.

C. XANES

Compared to EXAFS, XANES only focuses on the region of the XAS spectra just below and above the rising edge when the atom is excited during the x-ray photoabsorption process. There have been many methods with potential to utilize AI algorithms to accelerate the analysis process.

Rankine *et al.*[153] used DNNs to estimate XANES edge position and peak intensities. It has been applied primarily to iron by scraping datasets from the Materials Project.[154] The resulting fitted DNN generates predictions in the order of seconds and the accuracy of peak positions is within sub-eV resolution.

Furthermore, Liu *et al.*[155] utilized NNs to uncover the relationship between the XANES spectra and the structural parameters. The NN was fed using synthetic data, and the resulting model has been applied to copper oxide system. The fitted NN has demonstrated the capability of obtaining material parameters, such as average particle size and oxidation state.

Mizoguchi and Kiyohara[156] applied multiple ML techniques to characterize and extract material structure and properties from XANES spectra. Hierarchical clustering was first used to categorize the spectra, and then a decision tree was applied to create labeled training data. A feed-forward NN was then constructed from the training data for material prediction. It has demonstrated accurate prediction of upward of six material properties such as bond length, angle, and Voroni volume.

To address the lack of XANES data, Torrisi *et al.*[157] applied random forest techniques to create XANES-based descriptors. It obtained XANES datasets from the Materials Project[154] and Open Quantum Materials Database.[158] Their work has been successful at determining two material related properties (Bader charge and Nearest-neighbor distance).

Kiyohara and Mizoguchi[159] mapped the XANES spectrum to a radial distribution function (RDF) before applying it in an NN. The conversion to RDF can help reduce the significant experimental noise inherent in XANES spectra. It has been applied to silicon oxides and used to estimate parameters such as bond length and coordination number.

Similarly, Khan *et al.*[160] utilized PYFITIT[161] to generate synthetic spectra of transition metals. Martini *et al.*[161] developed PYFITIT[161] for quantitative analysis of XANES spectra. PYFITIT is an upgraded version of FITIT.[162] The new version contains many improvements in analysis such as PCA and spectral unmixing. It also incorporates ML algorithms such as ridge regression, Extra Trees, and LightGBM. The new version is written in PYTHON and contains a built-in graphical user interface and additional accessibility features.

Guda *et al.*[163] utilized PCA on XANES spectra, and their algorithm attempted to establish various relationships between edge position, intensity, and minima and maxima of the curvature. Using PCA, they concluded that good fits were obtained using linear basis functions and second order polynomials.

Finally, Trejo *et al.*[164] developed a pipeline for rapid analysis of XANES spectra using experiments and *ab initio* simulations/modeling. The data from experiment feed into an ML model of random forest and NN to gain insights into the evolution of ZnS under atomic layer deposition (ALD).

Figure 6 shows an example of such a framework. It highlights the XANES fitting analysis pipeline described by Trejo *et al.*[164] This pipeline involved collecting data from both experimental measurements and an existing database, which were utilized to create *ab initio* simulations for training purposes. The results were then used to construct the dataset. This workflow or pipeline enables swift analysis of XANES spectra while requiring minimal user input. This combination of high-throughput experiments, modeling, and ML can be adapted to other systems for rapid analysis.

D. XES

The fitting of XES spectra involves fitting measured data with conventional line shapes such as Gaussian, Lorentzian, and Voigt.

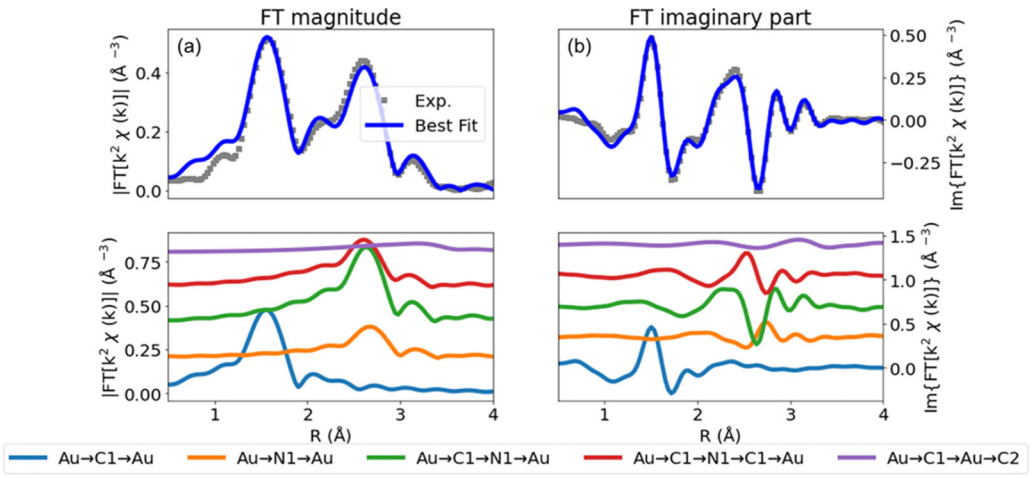

FIG. 5. (a) Phase uncorrected modulus and (b) imaginary part of the experimental and best fit EXAFS spectra for $KAu(CN)_2$. The result demonstrates the power of the AI in fitting the experimental data. The bottom is the corresponding individual contributions of single and multiple scattering paths. Reprinted with permission from Martini *et al.*, J. Phys. Chem. A **125**, 7080 (2021). Copyright 2023 American Chemical Society.

The resulting fitting can be used to deduce the physical oxidation state information.

Tetef *et al.*[165] presented a comprehensive computational study on the application of unsupervised machine learning techniques for extracting chemically relevant information from XANES and valence-to-core x-ray emission spectroscopy (VtC-XES) data. This study focused primarily on the classification of a diverse ensemble of organic sulfur molecules. By progressively reducing the

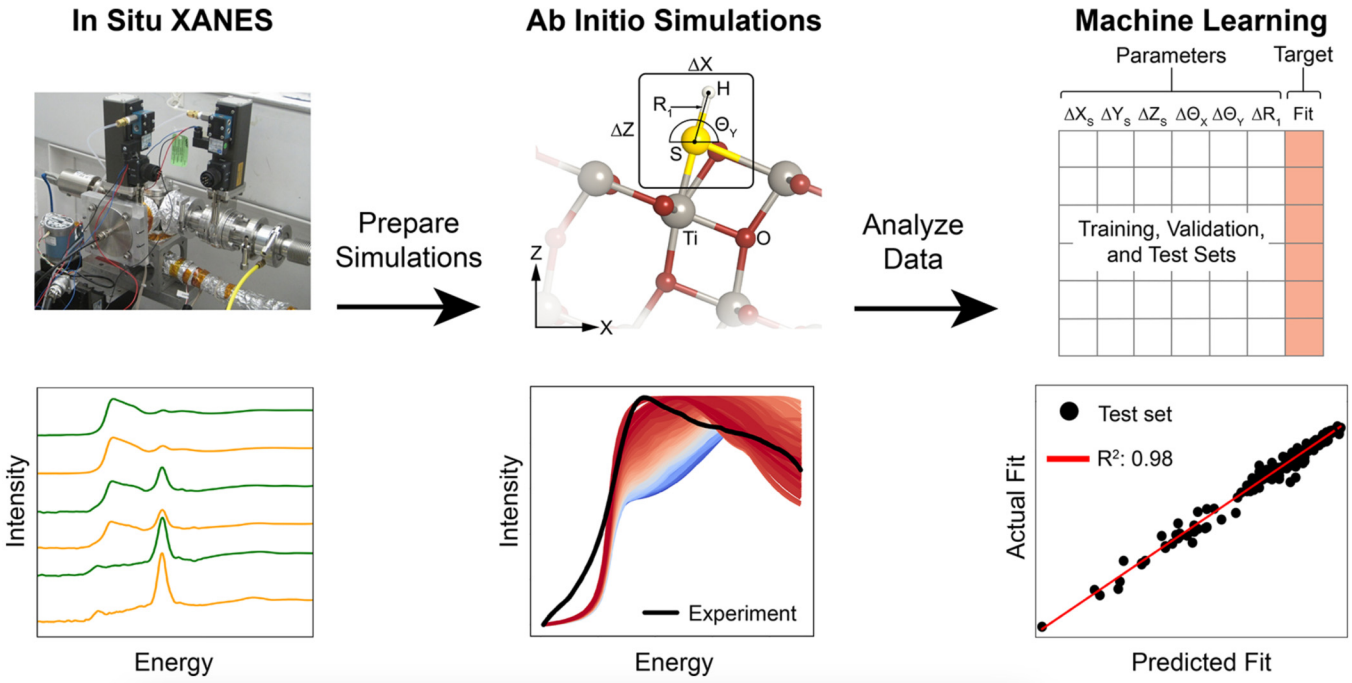

FIG. 6. Analysis pipeline of XANES fits from Trejo *et al.* (Ref. 164). The data were collected from both experiments and existing databases, which were used to construct the *ab initio* simulations used for training purposes. This workflow/pipeline allows for rapid analysis of XANES spectra with minimal input from the user. Reprinted (adapted) with permission from Trejo *et al.*, Chem. Mater. **31**, 8937 (2019). Copyright 2023 American Chemical Society.

constraining assumptions of the unsupervised ML algorithm from PCA to variational autoencoder (VAE) and t-distributed stochastic neighbor embedding (t-SNE), they were able to improve sensitivity while refining chemical state information. Even when reducing the embedding dimensionality to 2, t-SNE demonstrates the ability to distinguish not only oxidation state and general sulfur bonding environment but also the aromaticity of the bonding radical group. The overall accuracy of the t-SNE was 87%. In addition, this method was able to identify finer details in the electronic structure within aromatic or aliphatic subclasses. Overall, t-SNE demonstrated excellent potential in superior classification and also able to find new chemically relevant clusters not observed by other methods. This allow for the differentiation of different sulfur subclasses.

Penfold and Rankine[166] extended XANESNET[153,167] DNN to predict the line shape of VtC-XES spectra for first-row transition metal using the K-edge. It demonstrated that despite the strong sensitivity of VtC-XES to the electronic structure of the system under investigation, DNN is capable of replicating the principal spectral features using only the local coordination geometry of the transition metal complexes. This was achieved by representing the coordination geometry as a feature vector encoded with weighted atom-centered symmetry functions.

Furthermore, three methods were implemented and evaluated for assessing uncertainty in the predictions made by the VtC-DNN: (1) deep ensembles, (2) Monte-Carlo (MC) dropout, and (3) bootstrap resampling. The results indicated that bootstrap resampling yielded the best performance when evaluated on separate testing data. The developed models achieved an average accuracy of around 20% in predicting K-edge VtC-XES spectral intensities within the selected windows, which was higher than previous reports for K-edge XANES spectra.[167] The higher percentage error was partly due to regions in the VtC-XES spectrum where near-zero target intensities can result in significant errors. Unlike XANES spectra, VtC-XES is dominated by electronic transitions, making the direct link between local atomic structure and the measured spectrum less straightforward. The Bootstrap approach effectively estimated uncertainty between measured and predicted spectra, showing a stronger association compared to other methods. This suggested that uncertainty arose from the composition of the reference dataset. The uncertainty metric also enabled active learning, where structures can be selectively added to the training set based on prediction uncertainty. Additionally, a strong correlation was observed between the uncertainty predicted by bootstrap resampling and the error between the target and predicted VtC-XES spectra.

Hwang *et al.*[168] presented a methodology for analyzing 3d transition metal XES using a user-friendly program called *AXEAP2*. Their research aimed to address the challenges associated with data analysis in XES and to improve the efficiency of parameter optimization. The methodology involved three main steps. First, the XES calculation was performed using the CTM4XAS[169] code, which is widely used for calculating x-ray absorption and emission spectra, and used here to provide the initial theoretical spectra. Next, a convolution process was applied to the calculated XES spectra, which modified the spectra by applying a broadening function to account for instrumental effects and other experimental factors. This helped to better match the calculated spectra with the experimental data. Lastly, to optimize the parameters in the spectral simulations, a GA was employed which efficiently explored multiple parameter spaces to find optimized values that best fit the experimental data.

To aid in result interpretation, a viewer function was developed, enabling users to visualize and analyze the obtained spectra and parameter values. This feature enhanced the user-friendliness of the program and facilitated a better understanding of the analysis results. The performance of the methodology was evaluated and applied to manganese (Mn), cobalt (Co), and nickel (Ni) oxides. In total, six spectra were used to evaluate the performance and capabilities of *AXEAP2*, which does not provide uncertainty on the parameters used as a gene. A comparison method based on the relationship between RMSE and material parameter was used, and theoretical spectra were generated within physical/chemical boundaries. Acceptable values for reduction factor of $3d-3d$ electrons Coulomb force (F_{dd}), crystal field energy ($10Dq$), and reduction factor of $3p3d$ electrons exchange force (G_{pd}) were found, whereas spin–orbital coupling reduction factor (SO), reduction factor of 3p3d electrons Coulomb force (Fpd), and splitting point to apply different broadening (SP) did not have a significant effect on the data and were not determined. The spin state was determined by comparing the spin pairing energy with the optimized F_{dd} and $10Dq$ values based on atomic multiple theory and crystal field theory. While some samples had low RMSE $10Dq$ values that were similar to values calculated or used in the previous literature, *AXEAP2* reported a range of acceptable values from 0.5 to 2.5 eV and the lowest RMSE value of 2.2–2.5 eV. The G_{pd} value was found to be the most sensitive compared to other parameters, suggesting that the $3p3d$ exchange force played a key role in contributing to the energy difference between $K\beta_{1,3}$ and $K\beta''$ even in MnO and $MnCO_3$ spectra with the same electron number and configuration. *AXEAP2* demonstrates the capabilities to numerically estimate the spin state, $3d3p$ electron exchange force, and the broadening effect that arises from the core-hole lifetime.

E. XRD

XRD has similar peak fitting and parameterization requirements as XPS. The positions of the peaks in a powder diffraction pattern are determined by the lattice constants (unit cell parameters), and the peak intensities are determined by the atomic positions in the unit cell.[170] The powder pattern is, thus, a characteristic fingerprint of the material and used for phase identification. Many algorithms for such phase identification are included in commercial software and in the Powder Diffraction File database.[171]

Suzuki *et al.*[172] have applied similar work to μ-XRD and nano-XRD using random forest techniques for classifying space group. The dataset contained over 180 K XRD patterns from the Inorganic Crystal Structure Database (ICSD), and the resulting model yielded a prediction accuracy of 83.62% on space group prediction (230 class) and 93.07% on crystal system prediction (7 class) using only ten 2θ positions. The analysis pipeline demonstrated the possibility of accelerated materials characterization by combining high-throughput experiments and real-time analysis.

In another study, Suzuki *et al.*[173] demonstrated the capability of data-driven results for classification of space-group and crystal system in ML using XRD data. They included over 170 000 XRD spectra from the Materials Project.[154] They examined various ML

models such as logistic regression, k-nearest neighbors (KNN), decision tree, random forest, and extremely randomized tree. The tree-ensemble based model resulted in around 92% accuracy for system classification.

Similarly, Iwasaki *et al.*[174] proposed a high-throughput analysis pipeline using *ab initio* calculation and nonsupervised ML method. The method has been applied to a set of FeCoNi alloys with data from scanning microbeam XRD. The model utilized nonnegative matrix factorization to decompose the XRD data into singular structural XRD spectra, from which structural rate can then be uncovered. The result demonstrated mappings of material parameters such as Kerr rotation and magnetic moment as a function of FeCoNi composition.

Sivaraman *et al.*[175] incorporated simulation and ML to amorphous transition metal oxides for use-cases in fuel cell, catalytic, ceramic, and thermal exchange applications. Their methodology combined XRD and simulation methods, using pair distribution function measurements and ML based Gaussian approximation potential (GAP) method, to extract atomic structures and property information on bonding, density, diffusion, and conductivity. The main factor was the use of ML interatomic potentials to bridge the gap between classical interatomic potentials and *ab initio* molecular dynamics simulations by fitting directly to DFT reference datasets. The interatomic GAP is constructed through an active learning and fitting process that was initialized by configuration sampling of experimental data, with total energy written as a sum of local energies given by two-body and many-body terms. Finally, it utilized *ab initio* DFT to fit the XRD spectra.

Figure 7 shows an example of the pipeline, which involves the construction of interatomic fits. In this pipeline, two main components were identified. First, XRD data were utilized to gather information about the arrangement of atoms in a material, then ML techniques are employed to develop models that accurately describe the interatomic interactions. This combination of experimental and computational methods enables the creation of reliable interatomic fits. Second, the modeling of ML data workflow plays a crucial role in developing GAP from diffraction data. ML algorithms are employed to analyze the collected diffraction data and extract meaningful features that capture the interatomic interactions. These features are then used to train the underlying ML models. This ML-based workflow provides a powerful tool for understanding and predicting the behavior of materials based on their interatomic interactions.

Lee *et al.*[176] used a different approach to develop a novel NN architecture using a CNN with transformer encoder and variational autoencoder to identify symmetry (crystal systems and space-group) and property prediction (such as band gap and formation energy). The resulting fully convolutional network (FCN) contains over 1.3 million parameters and requires training for both weight, bias, and hyperparameters. The predictions with the ICSD dataset using the VAE-FCN (92.12%) outperform the baseline contextually guided CNN (82.17%) in crystal system identification.

Finally, Venderley *et al.*[177] developed a unsupervised ML pipeline, x-ray diffraction temperature clustering (X-TEC) for XRD analysis. It was used to extract charge density wave ordering and intraunit cell ordering from temperature fluctuation. The pipeline allowed for automatic analysis for XRD by performing preprocessing, X-TEC, and visualization.

F. Nanoindentation

The nanoindentation parameter can be optimized using a combination of algorithmic method (least square) and functional fitting. However, the fitting process is largely material dependent.

Konstantopoulos *et al.*[178] demonstrated the viability of using NN for analysis of nanoindentation data for fiber-matrix material such as carbon fiber based materials. The development process involved a multistep pipeline. The first step involved normalization of the data and extracting similar data using k-means clustering.[179] The next step involved training with some type of NN, such as multilayer perceptron (MLP), Stuttgart Neural Network, average neural network (avNNet), or SVM. The resulting models were compared using ML metrics to determine the best fitted models and hyperparameters. Afterward, transfer learning was applied by extrapolation of the prediction of the best trained models to a validation dataset of various indentation depths. It was observed that SVM performed the best with 67% on the test dataset and 72.7% on the validation dataset.

Kossman and Bigerelle[180] have employed the use of CNN model to identify pop-in and pop-out events in nanoindentation. The training dataset contained a total of 755 load and unloaded displacement curve, where there were 342 curves with pop-ins and 402 curves without pop-in. These curves were converted into square matrices (50×50) and used as inputs for the CNN model. The model successfully differentiated between pop-in and non-pop-in curves with an accuracy of approximately 93% in the training and validation datasets, indicating minimal risk of overfitting. These results demonstrated the use of direct spectrum data as inputs for usefulness in artificial intelligence and computer vision models for analyzing nanoindentation data.

Finally, Burleigh *et al.*[181] have extended materials characterization framework of Terry *et al.*,[149] *Neo*, for the analysis of nanoindentation load-displacement curves. They utilized a GA method in conjunction with the Oliver and Pharr[79] to overcome the limitations of least squares fitting (LSF) in fitting materials with sharply peaked unloading curves, such as polycrystalline isotropic graphites. The existing GA based materials characterization tool, Neo, was extended to fit nanoindentation data, with the new nanoindentation software requiring minimal user input to produce meaningful fit parameters. Fused silica (FS) and Al reference standards were used to benchmark the quality of the fits using this method. After validation, the method was applied to graphite and a high entropy alloy (HEA) consisting of body-centered cubic (BCC) and face-centered cubic (FCC) phases. For FS, the calculated material characteristics were slightly larger than those found in the literature, but consistent with LSF results for fits on the same tests. In the case of the aluminum sample, the GA results were consistent with those found in the literature and were 10.5% and 5.7% smaller than the HYSITRON software hardness and modulus results, respectively. When compared against LSF, the GA produced fits where the contact geometry fits well within the range predicted by theory, while the LSF is much above this range for the graphite specimen. For these tests, the elastic modulus computed using the GA-based tool was approximately 5% less than those from the LSF but follow similar trends between individual indentations, while the hardness results were consistent between the two methods. When applied to

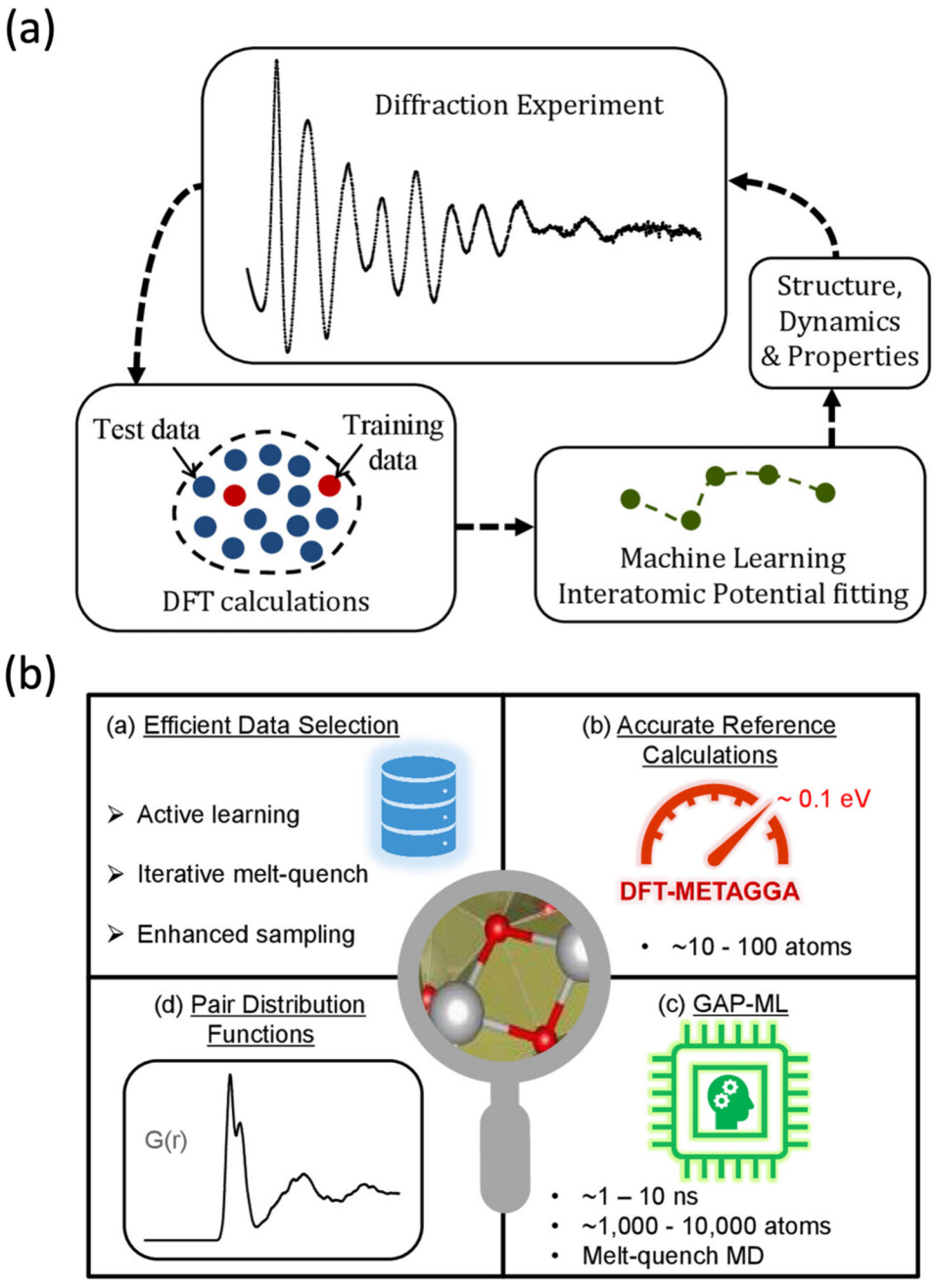

FIG. 7. (a) Analysis pipeline of the construction of interatomic fits using XRD and ML. (b) The modeling of ML data workflow for developing GAP from diffraction data. Reprinted from Sivaraman *et al.*, J. Phys. Soc. Jpn. **91**, 091009 (2022). Copyright 2022 JPS, licensed under a Creative Commons License.

HEA, our GA-based tool results were consistent with the HYSITRON software as well as the LSF used for analysis of the fused silica and graphite samples. All three fitting algorithms showed the same trends across all tests and produced consistent estimates for the hardness and modulus of the FCC and BCC phases. The HEA fitting result can be seen in Fig. 8, which shows the (a) inverse pole figure, (b) higher hardness, (c) reduced modulus, (d) HYSITRON hardness value, (e) hardness, and (f) reduced modulus using GA. The row of indentation corresponds to the result seen in earlier figure, which examine a specific row of data. The result

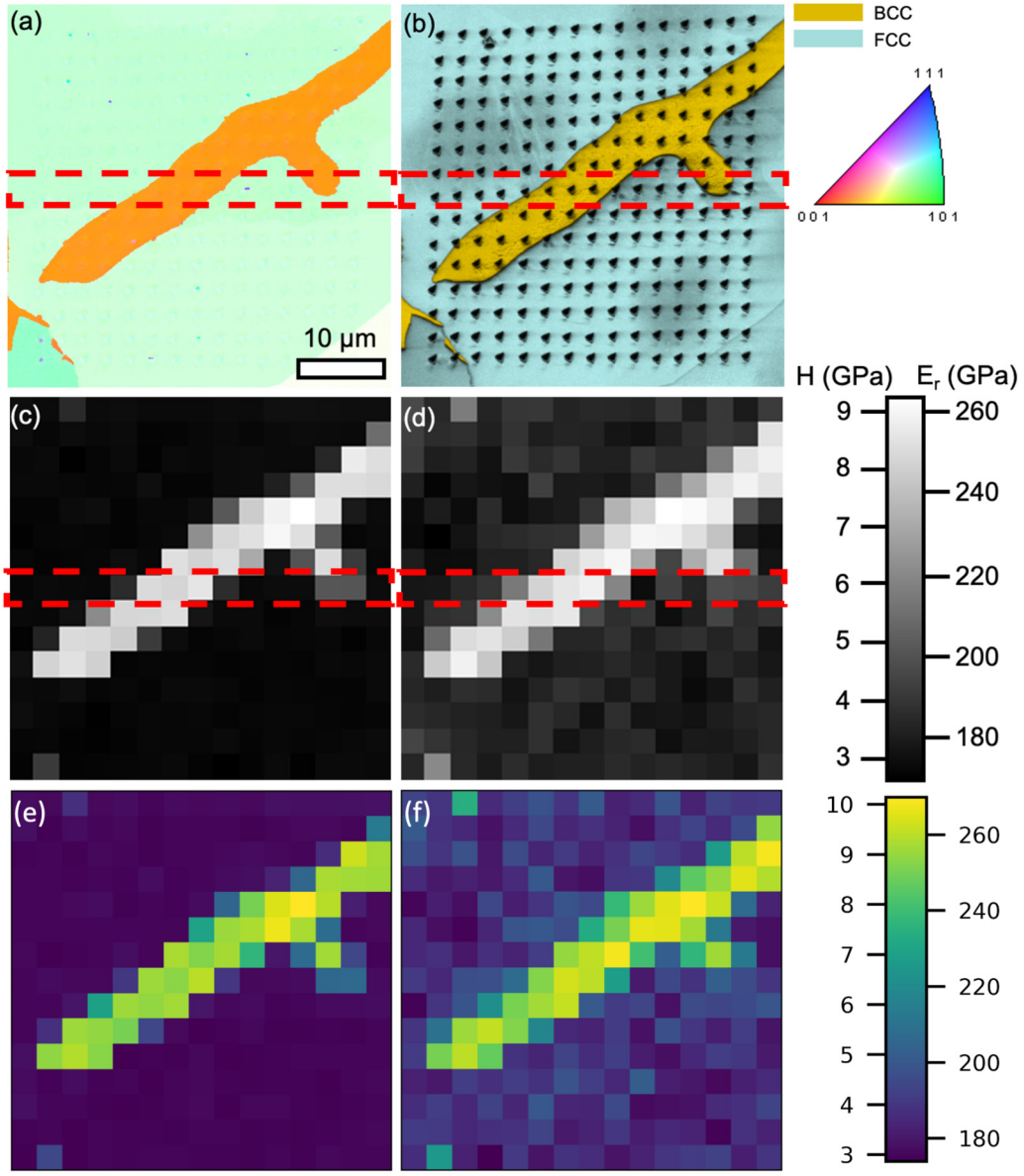

FIG. 8. Various analysis techniques were applied to generate maps of an HEA sample. These include an inverse pole figure obtained from EBSD (a), a phase map (b), and measurements of hardness (c) and reduced modulus (d) determined via HYSITRON software. Additionally, accelerated GA-based Neo, as described in Ref. 181, was utilized to measure hardness (e) and reduced modulus (f). Reprinted with permission from Burleigh *et al.*, Appl. Surf. Sci. **612**, 155634 (2023). Copyright 2023 Elsevier.

demonstrates the accuracy of the high-throughput workflow to obtain the hardness and reduced modulus mapping, allowing determination of the material properties.

G. SEM

SEM commonly uses image analysis techniques in the data analysis; therefore, most ML applications to SEM resolve around classification of chemical species present in the image or reconstruction technique to improve image quality.

de Haan *et al.*[182] proposed a DL-based method for improving the lateral resolution of SEM images through the use of a neural network. The approach involves training a CNN with coregistered high- and low-resolution SEM images of the same set of samples, which is then utilized for blind super-resolution of individual SEM images. This technique helps mitigate issues such as sample

charging and beam damage without compromising image quality or requiring additional sample preparation steps. Unlike traditional image enhancement methods, this approach is versatile and can be applied to a wide range of sample types using only a single SEM image as input. This approach has been demonstrated on gold-on-carbon specimens, with a result of increasing the resolution by a factor of 2x. This increase allowed for a 4X reduction in the number of electrons required from the conventional SEM, further decreasing the time required for image acquisition. DL based super-resolution techniques have been shown to generalize other types of samples or magnification factors, particularly when larger datasets or similar samples were used, mainly through the use of transfer learning once a well generalized model has been trained.

Figure 9 shows an example of the image with the resolution enhancement. It demonstrates the ability in upsampling the input image given the original input image when compared with a Au nanorod. Additionally, cross-sectional slices displaying various spatial features are included to highlight the significant improvement in resolution. By employing CNN algorithms, this approach achieves a remarkable increase in resolution, up to 2X, without compromising the image quality. Furthermore, it effectively reduces the number of electrons required to achieve a desired 4X resolution enhancement.

Ge *et al.*[183] proposed a workflow of using DL on micoscopic imaging, specifically for materials science. It involves the generalization of tasks such as specific task analysis, preparation of work, model design, feature analysis, and validation. Current applications of DL primarily focus on extracting structural information, such as morphology recognition and tracking, phase identification, and defect analysis. They also suggest the use of generative models for super-resolution reconstruction and linkage construction between experimental conditions and microstructures. In the future, the increasing use of DL in microscopic imaging analysis presents both challenges and opportunities. While current work mainly employ similar models and architectures, other powerful algorithms like LSTM and graph neural networks (GNNs) can be used to handle complex data structures in materials research data.

Similarly, a CNN model has been developed by Kaufmann, *et al.*[184] using the Xception architecture[185] for the phase mapping of EBSD data and tested against traditional Hough transformation based techniques for six different materials. The CNN-based model was then compared to Hough-based EBSD results that have used EDS measurements to manually determine the phase of certain parts of the image in order to improve mapping, referred to as optimized EBSD. Training was done using Adam optimization[186] on images that had not been prefiltered and the datasets may have contained partial or low quality DPs. The CNN performed well, identifying the correct Bravais lattice automatically and reproducing phase maps consistent with the traditional analysis even in the case of Rutile quartz when the model was not trained with quartz DPs. When analyzing a sample from the Campo del Cielo meteorite, the CNN was able to correctly identify a higher percentage of the schreibersite phase that was recessed below the surface than the traditional method. For samples of Fe–Al and a cycled thermal barrier coating (five FCC phases, one rhombohedral), the CNN produced phase maps that were superior to the traditional EBSD maps and consistent with optimized EBSD using EDS classification. Finally, in the case of 430 stainless steel, the Hough-based EBSD map incorrectly characterized most of the sample as martensite while the CNN produced a phase map consistent with that produced by a pattern quality thresholding technique often used to discern these two phases.

Shiratori *et al.*[187] developed a decision tree model to predict the dimensions of gold nanorods over a wide range of sizes, using approximately 450 nanorod geometries and corresponding scattering spectra obtained from finite-difference time-domain simulations. The resulting model deduced relationship between two spectral features (linewidth and resonance energy). Validations of the model were performed using experimental spectra and sizes obtained from correlated scanning electron microscopy images, and the model generated dimension predictions of gold nanorods within approximately 10% of their true values over a wide range of sizes. Analysis of the decision tree structure shows that a relationship with resonance energy and linewidth of the localized surface plasmon resonance was sufficient to predict nanorod dimensions, outperforming more complicated models.

H. TEM

Compared to SEM, TEM provides high-resolution images of the internal structure, composition, and crystallographic information of samples. However, the usage of ML in both SEM and TEM is very similar, which involves the analysis of TEM images to identify and quantify features of interest in the images.

Wen *et al.*[188] proposed the use of unsupervised ML to characterize the size of nanoparticles in TEM images. The resulting model allowed for accurate and rapid analysis of the TEM sample images, but was also robust against factors such as image quality, imaging modalities, and participle dispersions from experimental noise. The analysis pipeline contained multiple steps. It started with background removal. A separate background from that of the foreground was removed using a dynamic brightness threshold filter. Then, the edges of the particles were identified using a Canny Edge detection algorithm.[189] This algorithm was shown to correctly identify single particles but encountered difficulty when multiple particles were aggregated or overlapping. To overcome this, a convexity filtering method was instead applied, which showed substantial improvement in edge identification. Afterward, the processing data were used to train a unsupervised ML classification algorithm, which calculated the approximate particle size. The shape of each particle contour was parameterized using Hu moments, and the hierarchical agglomerative clustering method[190] was used with the average linkage to classify the parameterized particle shapes. The optimum numbers of clusters were determined automatically by applying internal cluster validity indexes.[191] This method was highly automated. It did not require initial steps from the user. It is generally applicable to any nanoparticle samples without the need of new raw data for training.

Computer vision algorithms can be utilized in the automation of TEM micrograph analysis for the purpose of feature identification and classification. One such tool is the Crystallographic Tool Box (CRYSTBOX) software[192,193] for the analysis of electron diffraction patterns. CRYSTBOX utilizes blob detection techniques such as Gaussian detection, Hessian response, and the difference of Gaussian[194] to

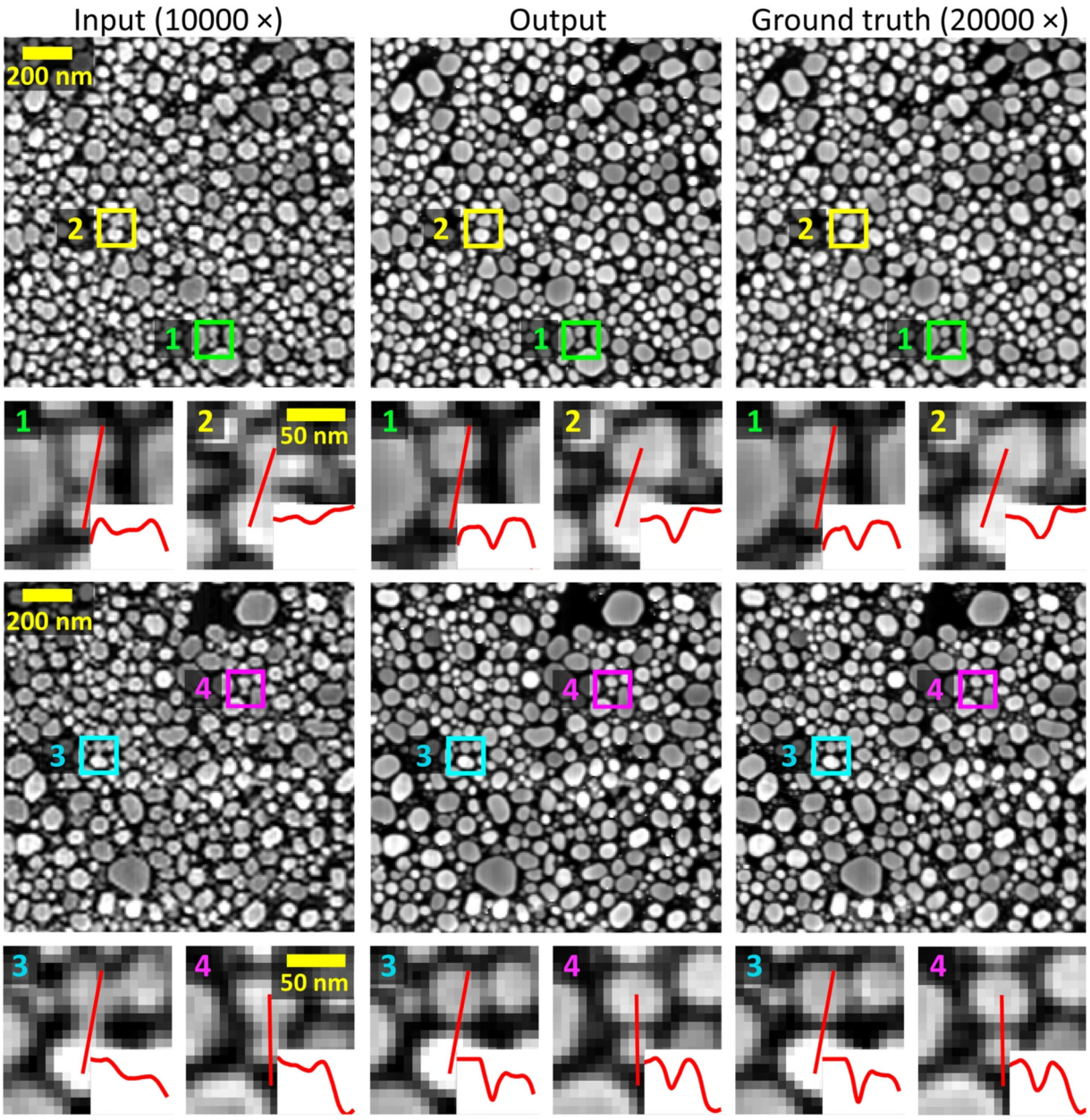

FIG. 9. Example of image upsampling using DNN between original input image, output, and the ground truth by de Haan *et al.* (Ref. 183). The image also includes cross-sectional slices of different spatial features to demonstrate the noticeable resolution enhancement. The use of CNN algorithms enables upward of 2X resolution increase without any compromise to the image quality, while also reducing the number of electrons requires to achieve desire resolution by 4X. Reprinted from de Haan *et al.*, Sci. Rep. **9**, 12050 (2019). Copyright 2019 Springer Nature, licensed under a creative commons license.

locate diffraction spots before applying it to the RANSAC algorithm,[195] which matches the diffraction pattern to a lattice. For ring diffraction patterns, the central beam spot can be automatically detected and the beam stopper removed from the image through iterative thresholding before the mean radial profile is extracted. Morenko and Ostaeva[196] compared the results of the CRYSTBOX automated tools and those of another automated analysis program, ProcessDiffraction, with manual counting and two "semiautomatic" tools, IMAGEJ[91] and DIGITALMICROGRAPH.[92,110] For comparisons, two simulated (ZnO and TiO_2) and one measured diffraction pattern

(copper-containing nanoparticles) were analyzed, and it was found that all of the methods were in agreement with the literature and of comparable precision. However, both of the automated programs failed to identify some reflections, although they were more time efficient. The Watershed method[197] is another computer vision algorithm that is available in the form of open source libraries. It is an image segmentation algorithm that has been applied to diffraction spot identification[198] as well.

I. GISAXS/GIWAXS

ML can also be used to solved many challenges faced in GISAXS/GIWAXS, such as particle classification, object detection for Bragg reflection, parameter extraction for layer thickness and roughness, and data processing to improve conventional fitting.[199] However, application of scattering data to ML is challenging due to the phase problem and experimental limitations such as differences in setup properties, measurement results, detectors, and sample environments, as well as finite accuracy of sample alignment.

Loaiza and Raza[200] discussed using artificial NNs to infer a material's scattering length density (SLD) profile from experimental reflectivity curves in the context of specular neutron and x-ray reflectometry, especially in GISAS. Their goal was to explore a new approach to the ill-posed, noninvertible problem of SLD profile inference, using ML to analyze large datasets of simulated reflectivity curves and SLD profiles. The approach involved replacing detailed layer-by-layer quantitative descriptions of samples (SLDs, thicknesses, roughnesses) with parameter-free curves $\rho(z)$, allowing *a priori* assumptions to be used in terms of the sample family to which a given sample belongs. They demonstrated that properly trained DNNs can recover plausible SLD profiles when presented with previously unseen simulated reflectivity curves. However, additional information about the instrument, such as its resolution and Q range, would be needed to apply the approach to real experiments. The proposed approach offers the advantages of faster batch analysis for large datasets and a new paradigm for describing sample physical models.

Similarity, Van Herck *et al.*[201] have proposed the use of DNN to rapidly analyze GISAXS data. As an inital case study, the authors extracted rotational distributions of hexagonal nanoparticle arrangements, with the training set sourced from BORNAGAIN.[124] The dataset was generated from synthetic labeled data, which underwent data augmentation to expand the size of the dataset. The resulting input dataset comprised of 50 000 training samples and 5000 validation samples. The two specific neural networks that were selected were DenseNet121 and DenseNet169.[202] Every layer was connected in a feed-forward manner, allowing feature maps to propagate to subsequent layers and alleviate the vanishing gradient problem while promoting feature reuse. This neural network architecture also has a relatively small number of parameters, which is essential for neutron and x-ray scattering data due to the limited amount of training data. For the training process, Kullback–Leibler divergence[203] was used as the loss metric during training, using a stochastic gradient descent algorithm as the optimizer and adaptive learning rate. The learning rate was actively adjusted during training, with a warm-up phase where the learning rate linearly increased up to its base value, followed by a decrease at each step. The resulting algorithms were applied to three different experiments with experiment 3 shown in Fig. 10. The expert scientists suggest that rotational distribution for experiment 3 contained only two rotations, while the predictions using neural network were significantly different. To indirectly check the accuracy of the prediction, a simulation was performed with predicted and uniform orientational distributions, and the results were compared to the measured pattern. The uniform distribution provided a better match to the experimental data than the expert scientist's guess, particularly in reproducing the diffuse scattering. This work shows great promise that it can be extended to other data analysis tasks and scattering techniques that require predicting distributions.

Archibald *et al.*[204] used weighted k-nearest neighbors (wKNN) as an ML algorithm for classifying unknown datasets in small angle scattering data. Unlike using neural networks, wKNN is a lazy learning algorithm that compares the unknown data against the entire set of training data each time, which can be very computationally intensive. This approach employed the Euclidean distance

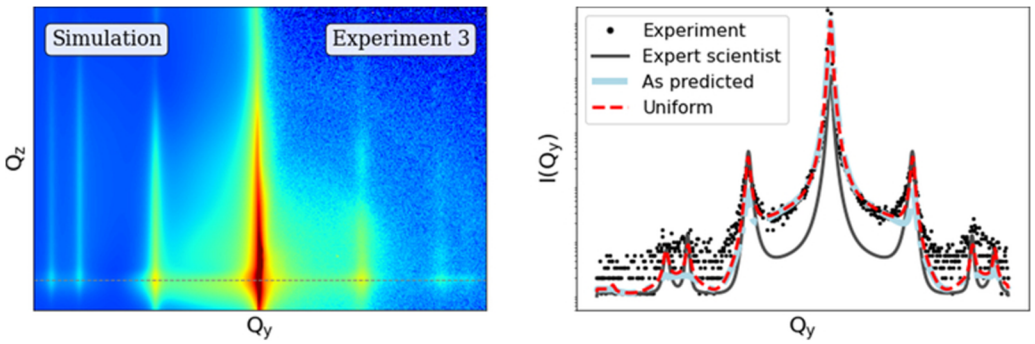

FIG. 10. Distribution predicted by DenseNet169 for experiment 3 (iron nanoparticles on the surface of ferroelectric $BaTiO_3$) was compared to a simulation by Van Herck *et al.* (Ref. 201) with axis labels omitted intentionally. The range of Q_y values varied approximately between −0.7 and 0.7 nm^{-1}, while Q_z ranged between 0.3 and 1.4 nm^{-1}. Reprinted from Van Herck *et al.*, Mater. Res. Express **8**, 045015 (2021). Copyright 2021 IOP, licensed under a Creative Commons license.

as the criterion for classification and was able to perform actual data fitting during classification to inform the user of the model parameters that best fit the data for the selected model. This method was tested against 61 models implemented in commonly used analysis software SASVIEW[205] and showed good performance against many of the models, especially when a Gaussian process with or without integrated fitting was employed. However, closely related models based on spheres or cylinders presented challenges for classification.

J. Neutron scattering

For neutron scattering, application of ML involved model training to uncover potential structural parameters. Conventional methods of processing neutron scattering data involve the usage of physics-based forward models and comparing against profiles generated experimentally, which can be very computationally intensive.

Garcia-Cardona *et al.*[206] proposed the use of powder diffraction data generated from GSAS II,[77] as input into both shallow and DL models to accelerate materials discovery by aiding experts in the determination of structures. Multiple ML models were applied such as unconstrained least squares, multilabel regressor with gradient boosting, and SVM, to material structure prediction. It was uncovered that these "shallow" models can achieve up to 90% prediction accuracy for certain classes of materials, but deep learning models such as CNN are better for more general predictions with accuracy up to 92%. Future model improvement requires exploration of deep learning to alleviate limitations in shallow ML models.

Kanazawa *et al.*[207] proposed a new approach to implementing an adaptive policy for multistep decision making, using a small-angle neutron scattering (SANS) experiment in materials science as

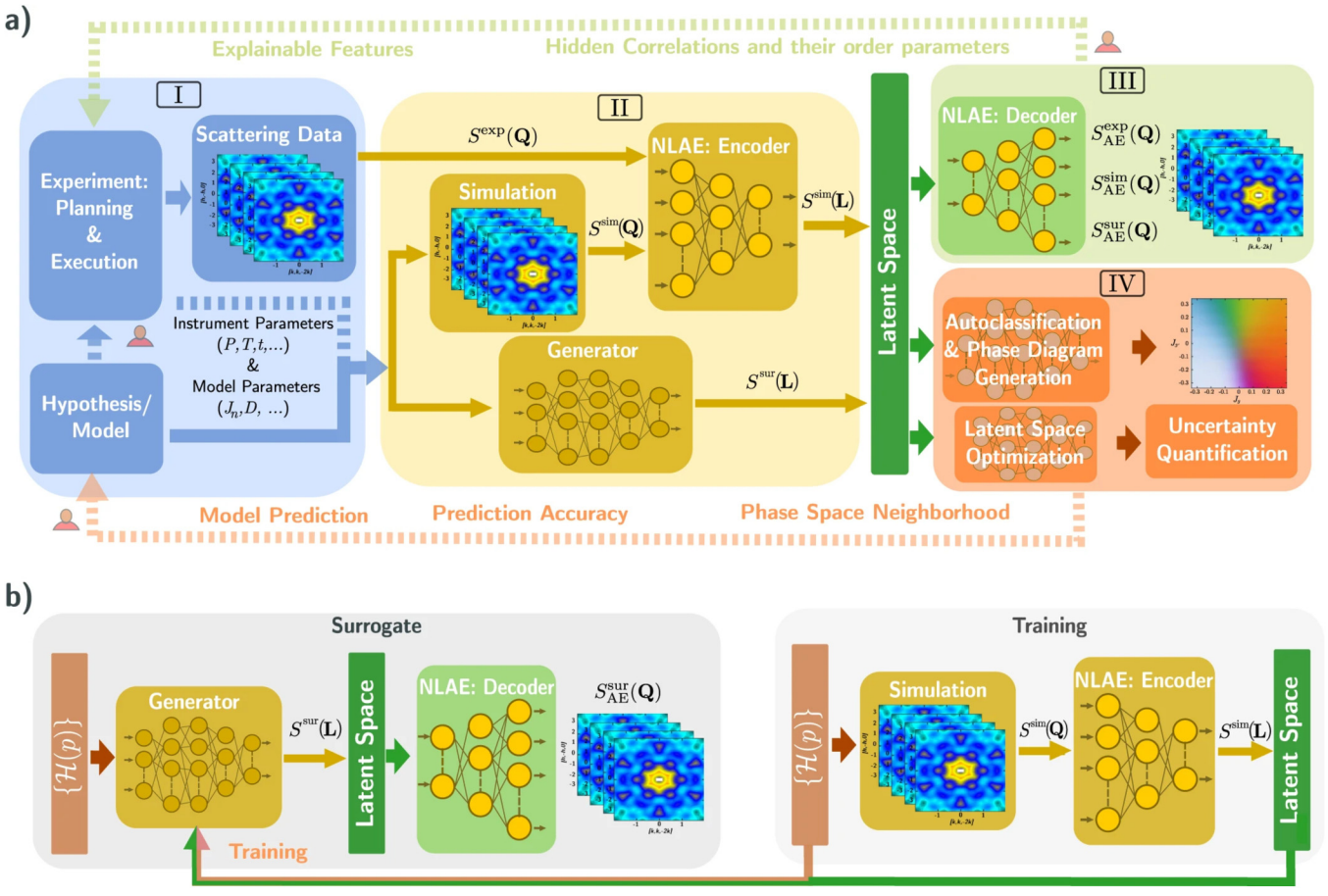

FIG. 11. (a) The proposed ML workflow for solving direct and inverse scattering problem. The main workflow was separated into four different block: block I involved scattering experiment design and optimization, block II was designating parameter space and information compression. Block III was structure and properties' predictions using NLAE. Finally, block IV was parameter space predictions from the latent space. (b) Design of a surrogate model used to predict the latent space S(L) and S(Q) for a given set of parameters. The training process involved using MC simulations and NLAE encoding. Simulations were done in areas of interest, and the surrogate was trained using the new MC result to improve the prediction accuracy. Reprinted from Samarakoon *et al.*, Commun. Mater. **3**, 84 (2022). Copyright 2022 Springer Nature, licensed under a Creative Commons license.

a proving ground. Even though this is not ML, machine learning can instead be easily incorporated into the analysis process. Two methods were proposed, one where decision is based on real-time similarity search, and the second where an adaptive database search is used to achieve future measurement confirmation. Though these two methods used numerical experiments, it was demonstrated to accelerate an SANS experiment by a factor of 2–3 as compared to random sampling without prior planning. It is expected that similar methods can also be applied to speed up a small-angle x-ray scattering (SAXS) experiment. Future directions related to this research involve extending the scheme to SANS experiments for nonspherical scatters and also the requirement to address potential issues surrounding sequential optimization problems beyond SANS experiments. In addition, the effect of noisy data must be understood.

Samarakoon *et al.*[208] proposed a workflow that applied ML to enable high-level, real-time feedback for neutron scattering. It involved using nonlinear autoencoders (NLAEs) to compress data from computationally expensive simulations of neutron scattering into a latent space. Using the compressed data, a generative model was then created, which allowed for identification of areas of interest and experiment planning. Description of the pipeline is shown in Fig. 11. Additionally, hierarchical clustering was used to categorize theoretical phase diagrams and identify experimental phases from measurements. The NLAE also provided accurate parameter determination and data handling. The authors demonstrated the capabilities of this approach on the highly frustrated magnet $Dy_2Ti_2O_7$ under pressure, demonstrating that hydrostatic pressures of up to 1.3 GPa can modify the magnetic interactions of the material, leading to prediction of a magnetic phase transition at higher pressures.

IV. CONCLUSION AND DISCUSSION

Integration of AI methods into the analysis of materials characterization data has been a rapidly expanding research area in recent years. AI algorithms have enabled more efficient and accurate analysis of complex material systems that may not be feasible using convectional methods, leading to advancements in the understanding of their material properties. Here, we reviewed the application of various AI algorithms in popular surface characterization techniques such as XPS, XANES, EXAFS, XRD, XES, nanoindentation, SEM, TEM, and GISAS/GIWAXS. A summary of the characterization methods and the AI methods that were reviewed are presented in Table I.

Such applications have also enabled rapid interpretation of spectra or images data, allowing for reduction in analysis time, which leads to an overall increase in the scientific productivity. Similarly, AI algorithms such as DNN and GA have been used to analyze XANES and EXAFS data, providing more comprehensive and accurate information about the local atomic and electronic structure of materials. AI has also been applied to image characterization technique such as SEM and TEM to enable rapid characterization of physical structure. The application of AI in other characterization techniques, such as XRD, nanoindentation, SEM, TEM, and GISAS, has also been successful in automating data processing and analysis, leading to more efficient and accurate material analysis.

The application of AI into various material characterization pipelines has allowed scientists to not only increase the rate of analysis, but also to study materials to a greater depth than is possible conventionally. The process of training an AI model requires a large amount of data from either experiment or theoretical modeling. Fortunately, the next generation of material characterization experiments excels at creating large datasets. By leveraging AI abilities in data ingestion both with and without human input, a more accurate and data driven conclusion will be made compared to convectional experiments.

In addition, modeling and simulations can complement experimental methods. In fact, many of the approaches examined here involved a combination of experiments and modeling. Together, modeling and simulation can efficiently explore the scope of parameter space and is much more practical than conducting a large number of experiments separately. Furthermore, modeling can assist in filling in gaps in data and meeting the AI requirements of a vast and varied dataset.

All in all, incorporating AI into material characterization analysis pathways offers promising methods to address reproducibility issues in material characterization. These methods include the use of AI algorithms to analyze large datasets generated by various characterization techniques, where they can be used to identify patterns and extract meaningful information. AI can help to standardize the data analysis process and reduce the potential for human error. AI can be used to verify that all necessary parameters and metadata are reported,[149] which is a major issue in the literature.[33] Additionally, AI results can be used to inform the data collection pipeline. AI can be used to optimize experimental conditions, such as identifying the optimal parameter settings for XRD or SEM imaging. AI can also be used to develop predictive models, which can aid in interpreting large data results. By leveraging AI in material characterization, we can improve reproducibility, reduce variability in analysis, and accelerate the pace of scientific discovery. To date, none of the tools reviewed here are capable of full automation of the materials characterization data analysis process for data currently being collected. There is still much work to be done to extract parameters from either literature or material databases. Most of the current AI based tools are focused on determination of experimental parameters, which leaves very few avenues available for pre-experiment planning. As a result, we recommend that further research should be directed toward alleviating these deficiencies.

Although AI's ability to inform the scientific process is well detailed above, there are several challenges and potential problems. There is always the potential for misuse of AI models. There is a lack of expertise among computer scientists in materials science along with the corresponding lack of expertise among material scientists in computer science. The complexity of materials and data availability can hamper the ability to further expand the use of AI. These may require more training of experts and the expansion of open databases holding raw data. The interpretability and operation of AI models is always open to discussion. It is obvious to us that AI models need to be continually monitored by experts in order to detect any shifts in accuracy and be retrained as soon as new data become available. Experts in both materials characterization and AI

TABLE I. Summary of surface characterization methods and their corresponding AI usage references in this review.

Characterization	AI methods	References
XPS	CNN to analyze the chemical composition of multicomponent XPS.	Drera *et al.*[139]
	Data clustering algorithms for quantifying spectral analysis.	Aarva *et al.*[140,141]
	CNN is used to perform human-free analyses of 1D spectroscopic data.	Park *et al.*[143]
	PCA and PLSR are used to create CHO predictive model.	Golze *et al.*[138]
EXAFS	GA applied to examine *in situ* Li-Sn battery structural changes.	Terry *et al.*[149]
	NN to extract EXAFS parameters for metals and oxide materials.	Timoshenko *et al.*[150]
	Inverse method to examine the backscattering phases and amplitudes.	Martini *et al.*[151]
	K-mean and PCA are used to examine all possible rational structures.	Kido *et al.*[152]
XANES	DNNs to estimate XANES edge position and peaks.	Rankine *et al.*[153]
	NN used to uncover the structural parameters.	Liu *et al.*[155]
	Hierarchical clustering, Decision Tree and NN used to predict bond length, angle, and Voroni volume.	Mizoguchi and Kiyohara[156]
	Random forest is used to create XANES based descriptors from databases.	Torrisi *et al.*[157]
	Training data are mapped to RBF before used for NN training and applied to SiO.	Kiyohara and Mizoguchi[159]
	PCA and spectral un-mixing for analysis of transition metals.	Khan *et al.*[160]
	PCA used to uncover spectra edge position, intensity, and minima/maxima of the curvature.	Guda *et al.*[163]
	Ab initio modeling used to construct dataset, which feeds into random forest and NN to analyze ZnS.	Trejo *et al.*[164]
XRD	Random forest techniques used for classifying crystal space group.	Suzuki *et al.*[172]
	Logistic regression, KNN, Decision Tree, Random Forest, and Extremely Randomized Tree were compared for space-group and crystal system classification.	Suzuki *et al.*[173]
	High-throughput analysis of FeCoNi using *ab initio* calculation and nonsupervised learning.	Iwasaki *et al.*[174]
	Experimental, modeling, and GAP method are used to extract material properties.	Sivaraman *et al.*[175]
	CNN with Transformer encoder and FCN to identify symmetry and property prediction.	Lee *et al.*[176]
	XTEC pipeline for XRD and XRD temperature clustering.	Venderley *et al.*[177]
XES	PCA, VAE, and t-SNE are used to classify VtC-XES data of organic sulfur.	Tetef *et al.*[165]
	DNN is used to predict line shape of VtC-XES spectra.	Penfold and Rankine[166]
	Convolution and GA is used to optimizes line-shape parameters of Mn, Co, and Ni oxides.	Huwang *et al.*[168]
Nanoindentation	MLP, SNN, avNN, and SVM are used for analysis of fiber-matrix material.	Konstantopoulos *et al.*[178]
	CNN used to identify pop-in and pop-out events.	Kossman and Bigerelle[180]
	GA used to analyze unloading and loading curves of graphite and HEA.	Burleigh *et al.*[181]
SEM	CNN used for image enhancement of Au and C samples.	de Haan *et al.*[182]
	DL is used to extract structural information, such as morphology recognition and tracking, phase identification, and defect analysis.	Ge *et al.*[183]
	Xception CNN is used on EBSD for phase mapping on Fe–Al.	Kaufmann *et al.*[184]
	Decision tree model used to predict the dimensions of Au nanorods.	Shiratori *et al.*[187]
TEM	Unsupervised ML used to characterize the nanoparticles size.	Wen *et al.*[188]
	Gaussian detection and Hessian response are used to match diffraction patterns of ZnO and TiO_2.	Klinger *et al.*[192,193]
GISAXS/ GIWAXS	NN is used to infer SLD profile from experimental reflectivity curves size.	Loaiza and Raza[200]
	NN is used to extract rotational distributions of hexagonal nanoparticle arrangements.	Van Herck *et al.*[201]
	wKNN is used to classify material properties.	Archibald *et al.*[204]
Neutron Scattering	DL models to accelerate analysis of powder diffraction.	Garcia-Cardona *et al.*[206]
	Adaptive policy with multistep decision making is used to accelerate analysis.	Kanazawa *et al.*[207]
	NLAE is used to create auto-descriptors to categorize theoretical phase diagrams and identify experimental phases.	Samarakoon *et al.*[208]

must be vigilant in checking that the output of the AI models does not drift over time to unphysical values. While much of this monitoring will need to be driven by these subject matter experts because AI algorithms will continue to become more sophisticated. The users of the AI analysis tools must also pay close attention to the output. The analyst using the AI tools is responsible for their interpretation of the data. The end user must ask themselves if the obtained results make sense in terms of the sample being analyzed,

the application for which it is being used, and the chemical and physical properties expected to be observed. Ultimately, it is analysts job to carefully select and ensure their data are accurate and free of bias and ensure that AI models are appropriately used and interpreted such that the result from the AI model accurately represents both the physics and data.

The future of AI in material characterization techniques is promising, with the potential to revolutionize the field of materials science. With continued development of AI algorithms and advancements in computing power, AI models could become even more accurate and reliable in predicting and characterizing materials in the era of big data. Potential applications include high-throughput screening, where AI models could rapidly screen large numbers of potential materials to identify promising candidates, or in the design of novel energy materials to optimize our use of energy, such as batteries and solar cells, by improving their efficiency and performance. Additionally, AI could be used to improve the accuracy of predicting material properties, analyze large amounts of data from experiments, and discover new materials with novel properties. Overall, the future of AI in material characterization techniques is exciting, and it has the potential to enable faster and more accurate characterization of materials, leading to the discovery of new materials with novel properties and applications.

ACKNOWLEDGMENTS

M.L. thanks the support from the National Science Foundation (NSF) under Grant No. 2213494. The work was partially supported by the Fermi National Accelerator Laboratory, managed and operated by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. DOE.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Miu Lun Lau: Formal analysis (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Abraham Burleigh:** Writing – original draft (equal). **Jeff Terry:** Conceptualization (equal); Funding acquisition (equal); Project administration (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Min Long:** Conceptualization (equal); Funding acquisition (equal); Project administration (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

Data sharing is not applicable as no new data were created or analyzed in this study.

REFERENCES

[1]W. Khan, A. Daud, J. A. Nasir, and T. Amjad, Kuwait J. Sci. **43**, 95 (2016).
[2]R. Socher, Y. Bengio, and C. D. Manning, "Deep learning for NLP (without magic)," in *Tutorial Abstracts of ACL 2012, ACL '12* (Association for Computational Linguistics, Kerrrville, TX, 2012), p. 5.
[3]A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, Madison, WI, 24-26 July 1998 (Association of Computing Machinery, New York, NY, 1998), pp. 92–100.
[4]J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," arXiv:1810.04805 (2018).
[5]H. A. Pierson and M. S. Gashler, Adv. Rob. **31**, 821 (2017).
[6]N. Sünderhauf *et al.*, Int. J. Rob. Res. **37**, 405 (2018).
[7]L. E. Peterson, Scholarpedia **4**, 1883 (2009).
[8]G. F. Cooper and E. Herskovits, Mach. Learn. **9**, 309 (1992).
[9]N. Rochester, Transactions of the IRE Professional Group on Electronic Computers **EC-2**, 10 (1953).
[10]S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms* (Cambridge University, Cambridge, 2014).
[11]Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).
[12]I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT, Cambridge, MA, 2016).
[13]J. Schmidhuber, Neural Networks **61**, 85 (2015).
[14]Y. Bengio *et al.*, Found. Trends Mach. Learn. **2**, 1 (2009).
[15]N. Seliya, T. M. Khoshgoftaar, and J. Van Hulse, "A study on the relationships of classifier performance metrics," in *2009 21st IEEE International Conference on Tools with Artificial Intelligence*, Newark, NJ, 2-4 November 2009 (IEEE Computer Society, Washington, DC, 2009), pp. 59–66.
[16]J. G. Carbonell, R. S. Michalski, and T. M. Mitchell, "An overview of machine learning," in *Machine Learning: An Artificial Intelligence Approach*, edited by R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (Springer, Berlin/Heidelberg, 1983), pp. 3-23.
[17]C. Rudin, Nat. Mach. Intell. **1**, 206 (2019).
[18]C. Shorten and T. M. Khoshgoftaar, J. Big Data **6**, 1 (2019).
[19]A. R. T. Donders, G. J. Van Der Heijden, T. Stijnen, and K. G. Moons, J. Clin. Epidemiol. **59**, 1087 (2006).
[20]K. Weiss, T. M. Khoshgoftaar, and D. Wang, J. Big Data **3**, 1 (2016).
[21]O. Sener and S. Savarese, "Active learning for convolutional neural networks: A core-set approach," arXiv:1708.00489 (2017).
[22]J. L. Schafer and J. W. Graham, Psychol. Methods **7**, 147 (2002).
[23]S. I. Nikolenko, *Synthetic Data for Deep Learning* (Springer, New York, 2021), Vol. 174.
[24]N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, J. Artif. Intell. Res. **16**, 321 (2002).
[25]C. Elkan, "The foundations of cost-sensitive learning," in *International Joint Conference on Artificial Intelligence* (Lawrence Erlbaum Associates Ltd, California City, CA, 2001), Vol. 17, pp. 973–978.
[26]T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems: First International Workshop, MCS 2000 Proceedings* Cagliari, Italy, 21–23 June (Springer, New York, 2000), pp. 1–15.
[27]V. Chandola, A. Banerjee, and V. Kumar, ACM Comput. Surv. (CSUR) **41**, 1 (2009).
[28]National Academies of Sciences, Engineering, and Medicine, *Reproducibility and Replicability in Science* (The National Academies, Washington, DC, 2019).
[29]C. Wang, U. Steiner, and A. Sepe, Small **14**, 1802291 (2018).
[30]M. Linford, V. Jain, and G. Major, "Gross errors in XPS peak fitting," in *AVS66 Abstract Book* (AVS, New York, NY, 2019), Vol. 72, see https://avs.org/AVS/files/e1/e1a0333e-4ffc-4a68-8f19-3e3b1dd9733d.pdf.
[31]M. R. Linford *et al.*, Microsc. Microanal. **26**, 1 (2020).
[32]G. H. Major *et al.*, Appl. Surf. Sci. **38**, 061203 (2020).
[33]G. H. Major, B. M. Clark, K. Cayabyab, N. Engel, C. D. Easton, J. Čechal, D. R. Baer, J. Terry, and M. R. Linford, J. Vac. Sci. Technol. A **41**, 043201 (2023).
[34]R. D. Chirico *et al.*, J. Chem. Eng. Data **58**, 2699 (2013).

[35]J. Park, J. D. Howe, and D. S. Sholl, Chem. Mater. **29**, 10487 (2017).
[36]D. R. Baer and I. S. Gilmore, J. Vac. Sci. Technol. A **36**, 068502 (2018).
[37]G. H. Major *et al.*, J. Vac. Sci. Technol. A **41**, 038501 (2023).
[38]F. Oviedo *et al.*, npj Comput. Mater. **5**, 60 (2019).
[39]G. T. Whiting, F. Meirer, and B. M. Weckhuysen, "Operando EXAFS and XANES of catalytic solids and related materials," in *XAFS Techniques for Catalysts, Nanomaterials, and Surfaces* (Springer International, New York, 2016), pp. 167–191.
[40]V. Gupta, H. Ganegoda, M. H. Engelhard, J. Terry, and M. R. Linford, J. Chem. Educ. **91**, 232 (2014).
[41]P. M. Sherwood, J. Vac. Sci. Technol. A **14**, 1424 (1996).
[42]N. Fairley, CasaXPS **2**, 1999 (2013).
[43]See https://www.thermofisher.com/order/catalog/product/IQLAADGACKFAKRMAVI for "Avantage" (2023) (accessed 30 January 2023).
[44]D. Adams and J. Andersen, see http://www.sljus.lu.se/download.html for "FitXPS version 2.12" (2010).
[45]RDATAA, "Aanalyzer: A peak fitting program for photoemission data," https://www.rdataa.com/aanalyzer (2023).
[46]M. Mohai, Surf. Interface Anal. **36**, 828 (2004).
[47]NIST X-ray Photoelectron Spectroscopy Database, *NIST Standard Reference Database Number 20* (National Institute of Standards and Technology, Gaithersburg, MD, 2000).
[48]J. Rehr and A. Ankudinov, Coord. Chem. Rev. **249**, 131 (2005).
[49]M. Newville, Rev. Mineral. Geochem. **78**, 33 (2014).
[50]P. Zimmermann, S. Peredkov, P. M. Abdala, S. DeBeer, M. Tromp, C. Müller, and J. A. van Bokhoven, Coord. Chem. Rev. **423**, 213466 (2020).
[51]G. Aquilanti, M. Giorgetti, R. Dominko, L. Stievano, I. Arčon, N. Novello, and L. Olivi, J. Phys. D: Appl. Phys. **50**, 074001 (2017).
[52]W.-S. Yoon *et al.*, Sci. Rep. **4**, 6827 (2014).
[53]M. Li, D. Olive, Y. Trenikhina, H. Ganegoda, J. Terry, and S. A. Maloy, J. Nucl. Mater. **441**, 674 (2013).
[54]R. Prins and D. Koningsberger, *X-ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS, and XANES* (Wiley, New York, 1988).
[55]V. Biebesheimer, E. Marques, D. Sandstrom, F. Lytle, and R. Greegor, J. Chem. Phys. **81**, 2599 (1984).
[56]M. Newville, J. Synchrotron Radiat. **8**, 322 (2001).
[57]A. Filipponi and A. Di Cicco, TASK Q. **4**, 575 (2000).
[58]R. Joyner, K. Martin, and P. Meehan, J. Phys. C Solid State **20**, 4005 (1987).
[59]M. Feiters, R. Strange, and N. Binsted, *International Tables for Crystallography* (Wiley, Hoboken, NJ, 2020), Vol. 1, pp. 1-8.
[60]M. Alain, M. Jacques, M.-B. Diane, and P. Karine, J. Phys. Conf. Ser. **190**, 012034 (2009).
[61]B. Ravel, "Quantitative EXAFS analysis," in *X-Ray Absorption and X-Ray Emission Spectroscopy: Theory and Applications* (Wiley, New York, 2016), Vol. 281.
[62]S. Roweis, Levenberg-Marquardt Optimization (University of Toronto, Toronto, 1996), Vol. 52.
[63]B. Ravel and M. Newville, J. Synchrotron Radiat. **12**, 537 (2005).
[64]T. Ressler, J. Phys. IV **7**, C2 (1997).
[65]M. Newville, J. Phys. Conf. Ser. **430**, 012007 (2013).
[66]D. Urch, Q. Rev., Chem. Soc. **25**, 343 (1971).
[67]K. O. Kvashnina and A. C. Scheinost, J. Synchrotron Radiat. **23**, 836 (2016).
[68]M. Neelakantan, S. Marriappan, J. Dharmaraja, T. Jeyakumar, and K. Muthukumaran, Spectrochim. Acta, Part A **71**, 628 (2008).
[69]K. Thamaphat, P. Limsuwan, and B. Ngotawornchai, Agric. Nat. Resour. **42**, 357 (2008).
[70]P. Bonneau, P. Garnier, E. Husson, and A. Morell, Mater. Res. Bull. **24**, 201 (1989).
[71]D. Hummer, P. J. Heaney, and J. Post, Powder Diffr. **22**, 352 (2007).
[72]G. Poralan, J. Gambe, E. Alcantara, and R. Vequizo, IOP Conf. Ser.: Mater. Sci. Eng. **79**, 012028 (2015).
[73]A. Chauhan and P. Chauhan, J. Anal. Bioanal. Tech. **5**, 1 (2014).
[74]J. Rodríguez-Carvajal, *An Introduction to the Program FullProf* (Laboratoire Leon Brillouin (CEA-CNRS), Paris, France, 2001).
[75]T. Degen and J. van den Oever, Powder Diffr. **24**, 163 (2009).
[76]A. A. Coelho, J. Appl. Crystallogr. **51**, 210 (2018).
[77]B. H. Toby and R. B. Von Dreele, J. Appl. Crystallogr. **46**, 544 (2013).
[78]C. A. Schuh, Mater. Today **9**, 32 (2006).
[79]W. C. Oliver and G. M. Pharr, J. Mater. Res. **7**, 1564 (1992).
[80]X. Li and B. Bhushan, Mater. Charact. **48**, 11 (2002).
[81]A. Gouldstone, H.-J. Koh, K.-Y. Zeng, A. Giannakopoulos, and S. Suresh, Acta Mater. **48**, 2277 (2000).
[82]S. Bull, J. Phys. D: Appl. Phys. **38**, R393 (2005).
[83]S.-H. Lee, S. Wang, G. M. Pharr, and H. Xu, Composites, Part A **38**, 1517 (2007).
[84]G. Pharr and W. Oliver, MRS Bull. **17**, 28 (1992).
[85]K. D. Vernon-Parry, III-Vs Rev. **13**, 40 (2000).
[86]D. Joy and D. Newbury, AIP Conf. Proc. **449**, 653 (1998).
[87]V. Randle, "Texture," in *Encyclopedia of Materials: Science and Technology* (Elsevier, New York, 2001), pp. 9119–9129.
[88]D. E. Laughlin and K. Hono, *Physical Metallurgy* (Newnes, New York, NY, 2014).
[89]S. Zaefferer, Ultramicroscopy **107**, 254 (2007).
[90]D. E. Newbury and N. W. Ritchie, Microsc. Microanal. **25**, 1075 (2019).
[91]M. D. Abràmoff, P. J. Magalhães, and S. J. Ram, Biophotonics Int. **11**, 36 (2004).
[92]D. R. Mitchell, Microsc. Res. Tech. **71**, 588 (2008).
[93]See https://www.thermofisher.com/us/en/home/electron-microscopy/products/software-em-3d-vis/avizo-software.html for "Thermofisher Avizo" (2023) (accessed 30 January 2023).
[94]See https://www.zeiss.com/microscopy/en/products/software/zeiss-zen.html for "Zeiss Zen" (2023) (accessed 30 January 2023).
[95]T. B. Britton, V. S. Tong, J. Hickey, A. Foden, and A. J. Wilkinson, J. Appl. Crystallogr. **51**, 1525 (2018).
[96]J. Li, T. Malis, and S. Dionne, Mater. Charact. **57**, 64 (2006).
[97]J. Yu, J. Liu, J. Zhang, and J. Wu, Mater. Lett. **60**, 206 (2006).
[98]C. Kübel, A. Voigt, R. Schoenmakers, M. Otten, D. Su, T.-C. Lee, A. Carlsson, and J. Bradley, Microsc. Microanal. **11**, 378 (2005).
[99]J. Lian, L. Wang, K. Sun, and R. C. Ewing, Microsc. Res. Tech. **72**, 165 (2009).
[100]D. Kim, J. S. Lee, C. M. Barry, and J. L. Mead, Microsc. Res. Tech. **70**, 539 (2007).
[101]A. Cossins, *Temperature Biology of Animals* (Springer Science & Business Media, New York, 2012).
[102]R. Egerton, P. Li, and M. Malac, Micron **35**, 399 (2004).
[103]M. Dehmas, J. Lacaze, A. Niang, and B. Viguier, Adv. Mater. Sci. Eng. **2011**, 940634 (2011).
[104]R. Neumann, "Materials research with energetic heavy ions at GSI," in *Physics and Engineering of New Materials* (Springer, New York, 2009), pp. 311–319.
[105]R. Schwarzer, Textures Microstruct. **20**, 696572 (1993).
[106]C. Karthik, J. Kane, D. P. Butt, W. Windes, and R. Ubic, J. Nucl. Mater. **412**, 321 (2011).
[107]J. Eapen, R. Krishna, T. D. Burchell, and K. Murty, Mater. Res. Lett. **2**, 43 (2014).
[108]M. Schorb, I. Haberbosch, W. J. Hagen, Y. Schwab, and D. N. Mastronarde, Nat. Methods **16**, 471 (2019).
[109]I.-P. Plus, "Image processing software," https://imagej.net/ (1994).
[110]B. Schaffer "Digital micrograph," in *Transmission Electron Microscopy: Diffraction, Imaging, and Spectrometry*, edited by C. B. Carter and D. B. Williams (Springer, Cham, 2016), pp. 167-196.
[111]G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, and S. J. Ludtke, J. Struct. Biol. **157**, 38 (2007).
[112]J. R. Kremer, D. N. Mastronarde, and J. R. McIntosh, J. Struct. Biol. **116**, 71 (1996).
[113]A. Hexemer *et al.*, J. Phys. Conf. Ser. **247**, 012007 (2010).
[114]A. Mahmood and J.-L. Wang, Sol. RRL **4**, 2000337 (2020).
[115]P. Müller-Buschbaum, "Structure determination in thin film geometry using grazing incidence small-angle scattering," in *Polymer Surfaces and Interfaces:*

Characterization, Modification and Applications (Springer, Berlin/Heidelberg, 2008), pp. 17-46.
[116]J. Perlich *et al.*, Rev. Sci. Instrum. **81**, 105105 (2010).
[117]P. Müller-Buschbaum, Eur. Polym. J. **81**, 470 (2016).
[118]F. Pietra, F. T. Rabouw, W. H. Evers, D. V. Byelov, A. V. Petukhov, C. de Mello Donegá, and D. Vanmaekelbergh, Nano Lett. **12**, 5515 (2012).
[119]W. Wang *et al.*, J. Mater. Chem. A **3**, 8324 (2015).
[120]D. Smilgies, P. Busch, C. M. Papadakis, and D. Posselt, Synchrotron Radiat. News **15**, 35 (2002).
[121]A. Hexemer and P. Müller-Buschbaum, IUCrJ **2**, 106 (2015).
[122]S. Vajda, R. E. Winans, J. W. Elam, B. Lee, M. J. Pellin, S. Seifert, G. Y. Tikhonov, and N. A. Tomczyk, Top. Catal. **39**, 161 (2006).
[123]E. Metwalli *et al.*, Langmuir **29**, 6331 (2013).
[124]G. Pospelov, W. Van Herck, J. Burle, J. M. Carmona Loaiza, C. Durniak, J. M. Fisher, M. Ganeva, D. Yurov, and J. Wuttke, J. Appl. Crystallogr. **53**, 262 (2020).
[125]Z. Jiang, J. Appl. Crystallogr. **48**, 917 (2015).
[126]A. Hammersley, "FIT2D: An introduction and overview," Internal Report ESRF97HA02T (European Synchrotron Radiation Facility, Grenoble, France, 1997), Vol. 68, p. 58.
[127]I. Breßler, B. R. Pauw, and A. F. Thünemann, J. Appl. Crystallogr. **48**, 962 (2015).
[128]G. Ashiotis, A. Deschildre, Z. Nawaz, J. P. Wright, D. Karkoulis, F. E. Picca, and J. Kieffer, J. Appl. Crystallogr. **48**, 510 (2015).
[129]T. A. White, R. A. Kirian, A. V. Martin, A. Aquila, K. Nass, A. Barty, and H. N. Chapman, J. Appl. Crystallogr. **45**, 335 (2012).
[130]S. W. Lovesey, *Theory of Neutron Scattering from Condensed Matter. Vol. 1. Nuclear Scattering* (Oxford University, New York, 1986), p. 270.
[131]S. W. Lovesey, *Theory of Neutron Scattering from Condensed Matter* (Clarendon, Oxford, UK, 1986), Vol. 2, p. 310.
[132]D. Price and K. Skold, *Neutron Scattering* (Academic, New York, 1987).
[133]J. L. Rowsell, J. Eckert, and O. M. Yaghi, J. Am. Chem. Soc. **127**, 14904 (2005).
[134]J. Higgins, H. Benoit, and H. Benô, *Polymers and Neutron Scattering*, Oxford Science Publications (Clarendon, Oxford, UK, 1996), p. 436.
[135]A. J. Allen, J. Am. Ceram. Soc. **88**, 1367 (2005).
[136]B. Jacrot, Rep. Prog. Phys. **39**, 911 (1976).
[137]K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, Nature **559**, 547 (2018).
[138]D. Golze, M. Hirvensalo, P. Hernández-León, A. Aarva, J. Etula, T. Susi, P. Rinke, T. Laurila, and M. A. Caro, Chem. Mater. **34**, 6240 (2022).
[139]G. Drera, C. M. Kropf, and L. Sangaletti, Mach. Learn.: Sci. Technol. **1**, 015008 (2020).
[140]A. Aarva, V. L. Deringer, S. Sainio, T. Laurila, and M. A. Caro, Chem. Mater. **31**, 9243 (2019).
[141]A. Aarva, V. L. Deringer, S. Sainio, T. Laurila, and M. A. Caro, Chem. Mater. **31**, 9256 (2019).
[142]R. Xu and D. Wunsch, IEEE Trans. Neural Networks **16**, 645 (2005).
[143]S.-H. Park, H. Park, H. Lee, and H.-S. Kim, J. Korean Phys. Soc. **79**, 1199 (2021).
[144]S. Chatterjee, B. Singh, A. Diwan, Z. R. Lee, M. H. Engelhard, J. Terry, H. D. Tolley, N. B. Gallagher, and M. R. Linford, Appl. Surf. Sci. **433**, 994 (2018).
[145]D. Coster and R. D. L. Kronig, Physica **2**, 13 (1935).
[146]B. K. Teo, *EXAFS: Basic Principles and Data Analysis* (Springer Science & Business Media, New York, 2012), Vol. 9.
[147]M. Newville, J. Synchrotron. Radiat. **8**, 96 (2001).
[148]K. Asakura, H. Abe, and M. Kimura, J. Synchrotron. Radiat. **25**, 967 (2018).
[149]J. Terry *et al.*, Appl. Surf. Sci. **547**, 149059 (2021).
[150]J. Timoshenko, H. S. Jeon, I. Sinev, F. T. Haase, A. Herzog, and B. R. Cuenya, Chem. Sci. **11**, 3727 (2020).
[151]A. Martini *et al.*, J. Phys. Chem. A **125**, 7080 (2021).
[152]D. Kido, T. Wada, and K. Asakura, e-J. Surf. Sci. Nanotechnol. **21**, 231 (2023).
[153]C. D. Rankine, M. M. Madkhali, and T. J. Penfold, J. Phys. Chem. A **124**, 4263 (2020).
[154]A. Jain *et al.*, APL Mater. **1**, 011002 (2013).
[155]Y. Liu *et al.*, J. Chem. Phys. **151**, 164201 (2019).
[156]T. Mizoguchi and S. Kiyohara, Microscopy **69**, 92 (2020).
[157]S. B. Torrisi, M. R. Carbone, B. A. Rohr, J. H. Montoya, Y. Ha, J. Yano, S. K. Suram, and L. Hung, npj Comput. Mater. **6**, 109 (2020).
[158]S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, and C. Wolverton, npj. Comput. Mater. **1**, 1 (2015).
[159]S. Kiyohara and T. Mizoguchi, J. Phys. Soc. Jpn. **89**, 103001 (2020).
[160]L. U. Khan, Z. U. Khan, L. Blois, L. Tabassam, H. F. Brito, and S. J. Figueroa, Inorg. Chem. **62**, 2738 (2023).
[161]A. Martini *et al.*, Comput. Phys. Commun. **250**, 107064 (2020).
[162]G. Smolentsev and A. V. Soldatov, Comput. Mater. Sci. **39**, 569 (2007).
[163]A. A. Guda *et al.*, npj Comput. Mater. **7**, 203 (2021).
[164]O. Trejo *et al.*, Chem. Mater. **31**, 8937 (2019).
[165]S. Tetef, N. Govind, and G. T. Seidler, Phys. Chem. Chem. Phys. **23**, 23586 (2021).
[166]T. Penfold and C. Rankine, Mol. Phys. **121**, e2123406 (2023).
[167]C. D. Rankine and T. Penfold, J. Chem. Phys. **156**, 164102 (2022).
[168]I.-H. Hwang, S. D. Kelly, M. K. Y. Chan, E. Stavitski, S. M. Heald, S.-W. Han, N. Schwarz, and C.-J. Sun, J. Synchrotron Radiat. **30**, 923 (2023).
[169]E. Stavitski and F. M. De Groot, Micron **41**, 687 (2010).
[170]B. B. He, *Two-Dimensional X-Ray Diffraction* (Wiley, New York, 2018).
[171]I. C. Madsen, N. V. Scarlett, and N. A. Webster, "Quantitative phase analysis," in *Uniting Electron Crystallography and Powder Diffraction* (Springer, New York, 2012), pp. 207–218.
[172]Y. Suzuki, H. Hino, Y. Takeichi, T. Hawai, M. Kotsugi, and K. Ono, Microsc. Microanal. **24**, 142 (2018).
[173]Y. Suzuki, H. Hino, T. Hawai, K. Saito, M. Kotsugi, and K. Ono, Sci. Rep. **10**, 21790 (2020).
[174]Y. Iwasaki, M. Ishida, and M. Shirane, Sci. Technol. Adv. Mater. **21**, 25 (2020).
[175]G. Sivaraman, G. Csanyi, A. Vazquez-Mayagoitia, I. T. Foster, S. K. Wilke, R. Weber, and C. J. Benmore, J. Phys. Soc. Jpn. **91**, 091009 (2022).
[176]B. D. Lee, J.-W. Lee, W. B. Park, J. Park, M.-Y. Cho, S. Pal Singh, M. Pyo, and K.-S. Sohn, Adv. Intell. Syst. **4**, 2200042 (2022).
[177]J. Venderley *et al.*, Proc. Natl. Acad. Sci. **119**, e2109665119 (2022).
[178]G. Konstantopoulos, E. P. Koumoulos, and C. A. Charitidis, Mater. Des. **192**, 108705 (2020).
[179]A. Ahmad and L. Dey, Data Knowl. Eng. **63**, 503 (2007).
[180]S. Kossman and M. Bigerelle, Materials **14**, 7027 (2021).
[181]A. Burleigh *et al.*, Appl. Surf. Sci. **612**, 155734 (2023).
[182]K. de Haan, Z. S. Ballard, Y. Rivenson, Y. Wu, and A. Ozcan, Sci. Rep. **9**, 12050 (2019).
[183]M. Ge, F. Su, Z. Zhao, and D. Su, Mater. Today Nano **11**, 100087 (2020).
[184]K. Kaufmann, C. Zhu, A. S. Rosengarten, D. Maryanovsky, H. Wang, and K. S. Vecchio, Microsc. Microanal. **26**, 458 (2020).
[185]F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, 21-26 July 2017 (IEEE Computer Society, Los Alamitos, CA, 2017), pp. 1251–1258.
[186]D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980 (2014).
[187]K. Shiratori, L. D. C. Bishop, B. Ostovar, R. Baiyasi, Y.-Y. Cai, P. J. Rossky, C. F. Landes, and S. Link, J. Mater. Chem. C **125**, 19353 (2021).
[188]H. Wen, J. M. Luna-Romera, J. C. Riquelme, C. Dwyer, and S. L. Chang, Nanomaterials **11**, 2706 (2021).
[189]W. Rong, Z. Li, W. Zhang, and L. Sun, "An improved canny edge detection algorithm," in *2014 IEEE International Conference on Mechatronics and Automation* (IEEE, New York, 2014), pp. 577–582.
[190]F. Murtagh and P. Legendre, J. Classif. **31**, 274 (2014).
[191]O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, Pattern Recognit. **46**, 243 (2013).
[192]M. Klinger and A. Jäger, J. Appl. Crystallogr. **48**, 2012 (2015).

[193]M. Klinger, *Crystbox—Crystallographic Toolbox* (Institute of Physics of the Czech Academy of Sciences, Prague, 2015).
[194]D. G. Lowe, Int. J. Comput. Vision **60**, 91 (2004).
[195]M. A. Fischler and R. C. Bolles, Commun. ACM **24**, 381 (1981).
[196]I. Morenko and G. Ostaeva, AIP Conf. Proc. **2467**, 20040 (2022).
[197]A. S. Kornilov and I. V. Safonov, J. Imaging **4**, 123 (2018).
[198]S. Nebaba, D. Zavyalov, and A. Pak, "Patterns detection in SAED images of transmission electron microscopy," in *CEUR Workshop Proceedings*, Moscow Region, Russia, 9-13 November 2020 (CEUR-WS, Aachen, Germany, 2020), Vol. 2763, pp. 319–322.
[199]A. Hinderhofer, A. Greco, V. Starostin, V. Munteanu, L. Pithan, A. Gerlach, and F. Schreiber, J. Appl. Crystallogr. **56**, 3 (2023).
[200]J. M. C. Loaiza and Z. Raza, Mach. Learn.: Sci. Technol. **2**, 025034 (2021).
[201]W. Van Herck, J. Fisher, and M. Ganeva, Mater. Res. Express **8**, 045015 (2021).
[202]G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, 21-26 July 2017 (IEEE, Los Alamitos, CA, 2017), pp. 4700–4708.
[203]J. M. Joyce, "Kullback-Leibler divergence," in *International Encyclopedia of Statistical Science* (Springer, New York, 2011), pp. 720–722.
[204]R. K. Archibald, M. Doucet, T. Johnston, S. R. Young, E. Yang, and W. T. Heller, J. Appl. Crystallogr. **53**, 326 (2020).
[205]M. Doucet *et al.* (2019). "SasView version 5.0.4," Zenodo. https://doi.org/10.5281/zenodo.4467702.
[206]C. Garcia-Cardona, R. Kannan, T. Johnston, T. Proffen, K. Page, and S. K. Seal, "Learning to predict material structure from neutron scattering data," in *2019 IEEE International Conference on Big Data (Big Data)* (IEEE, New York, 2019), pp. 4490–4497.
[207]T. Kanazawa, A. Asahara, and H. Morita, J. Phys.: Mater. **3**, 015001 (2019).
[208]A. Samarakoon, D. A. Tennant, F. Ye, Q. Zhang, and S. A. Grigera, Commun. Mater. **3**, 84 (2022).