

Review

X-ray Diffraction Data Analysis by Machine Learning Methods—A Review

Vasile-Adrian Surdu ^{1,2}  and Romuald György ^{2,3,*} 

¹ Department of Science and Engineering of Oxide Materials and Nanomaterials, Faculty of Chemical Engineering and Biotechnologies, National University of Science and Technology Politehnica Bucharest, Gheorghe Polizu 1-7, 011061 Bucharest, Romania; adrian.surdu@upb.ro

² Academy of Romanian Scientists, Ilfov 3, 050044 Bucharest, Romania

³ Department of Chemical and Biochemical Engineering, Faculty of Chemical Engineering and Biotechnologies, National University of Science and Technology Politehnica Bucharest, Gheorghe Polizu 1-7, 011061 Bucharest, Romania

* Correspondence: romuald.gyorgy@upb.ro

Abstract: X-ray diffraction (XRD) is a proven, powerful technique for determining the phase composition, structure, and microstructural features of crystalline materials. The use of machine learning (ML) techniques applied to crystalline materials research has increased significantly over the last decade. This review presents a survey of the scientific literature on applications of ML to XRD data analysis. Publications suitable for inclusion in this review were identified using the “machine learning X-ray diffraction” search term, keeping only English-language publications in which ML was employed to analyze XRD data specifically. The selected publications covered a wide range of applications, including XRD classification and phase identification, lattice and quantitative phase analyses, and detection of defects and substituents, as well as microstructural material characterization. Current trends in the field suggest that future efforts pertaining to the application of ML techniques to XRD data analysis will address shortcomings of ML approaches related to data quality and availability, interpretability of the results and model generalizability and robustness. Additionally, future research will likely incorporate more domain knowledge and physical constraints, integrate with quantum physical methods, and apply techniques like real-time data analysis and high-throughput screening to accelerate the discovery of tailored novel materials.

Keywords: X-ray diffraction; phase identification; phase transitions; crystal structure; machine learning; cluster analysis; deep learning; neural networks



Citation: Surdu, V.-A.; György, R. X-ray Diffraction Data Analysis by Machine Learning Methods—A Review. *Appl. Sci.* **2023**, *13*, 9992. <https://doi.org/10.3390/app13179992>

Academic Editor: Alexander J. G. Lunt

Received: 4 August 2023

Revised: 1 September 2023

Accepted: 1 September 2023

Published: 4 September 2023

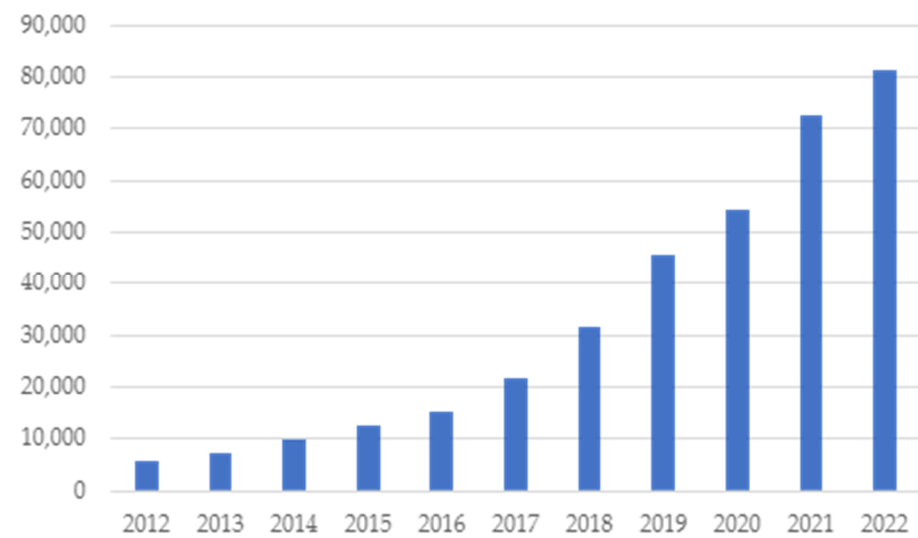


Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

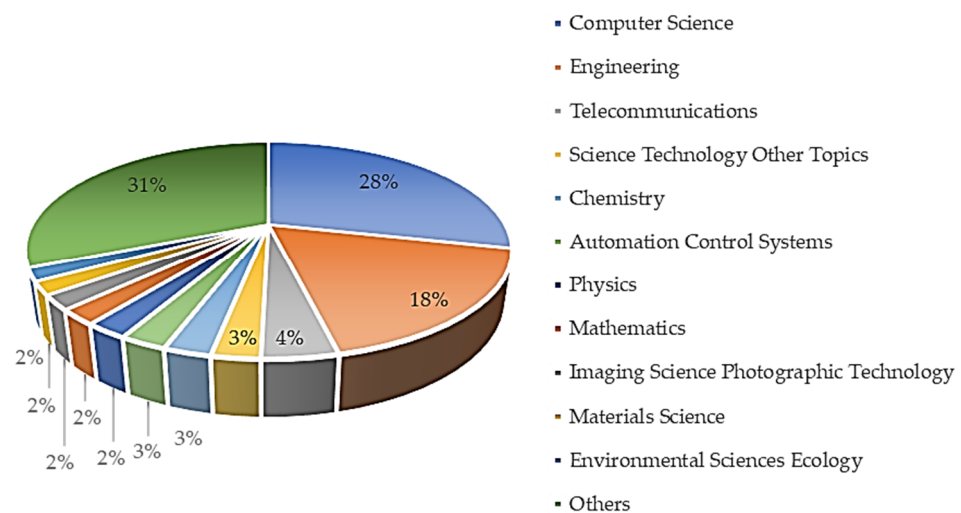
1. Introduction

Machine learning (ML) has recently found numerous applications, and it is being leveraged as a powerful tool in various fields: computer science, engineering, telecommunications, chemistry, physics, mathematics, imaging science, materials science, and environmental sciences. The importance of ML is also confirmed by an increasing trend in the volume of work published over the last decade, according to Web of Science data (Figure 1).

In computer science, ML is used for various tasks, such as natural language processing [1–5], image recognition [6–8], or computer vision [9–14]. In engineering, ML is applied in the optimization and control of complex systems [15–18], the prediction of equipment failures [19–21], or the enhancement of manufacturing processes [22–25]. ML is also extensively used in materials science for materials discovery [26–30], property prediction [31–34], and accelerated materials design [35–37].



(a)



(b)

Figure 1. Number of publications about machine learning, according to Web of Science: (a) yearly publication counts; (b) classification by research area.

Understanding materials' structure, composition, and properties is essential in experimental materials science. Thus, spectroscopy and microscopy are used to characterize the behavior of materials at various scales. The integration of machine learning methods has brought transformative advancements to the analysis of complex data. Several models have been employed to automate the interpretation of intricate spectroscopic data, facilitating the enhancement of signals, feature extraction, compound classification, and property prediction. Similarly, these methods have enabled automated particle detection, crystallographic analysis, and defect recognition in electron microscopy images, surpassing conventional image processing approaches [38–40].

One example is [41], which showed that deep learning has great potential in performing all the steps and emphasized the importance of addressing the estimation of the prediction quality of deep learning models on small datasets with complex covariance structures. Another example is the case of scanning transmission electron microscopy–electron energy loss spectroscopy studies, for which [42] used a principal components

analysis (PCA) algorithm to analyze the momentum-resolved spectra for SiGe quantum dots and Si-SiGe interfaces. In their case, the low acquisition parameters for mapping datasets required the PCA method to improve the signal quality. The authors state that by using this method instead of traditional Fourier filtering or smoothening techniques, the important features are maintained, and their quality is improved.

The focus of this paper was to assess the implications of ML in the field of data analysis of X-ray diffraction (XRD), compare the accuracy of the reported models with that of traditional XRD data analysis procedures, and present future development opportunities.

1.1. Overview of X-ray Diffraction (XRD) Technique

More than 100 years after its discovery, X-ray diffraction is still one of the most powerful and versatile techniques widely employed for understanding crystalline materials' phase composition, structure, and microstructural features. The technique was developed in the early twentieth century when Max von Laue discovered that crystals diffract X-rays, and the obtained pattern reveals the crystal's structure [43]. The findings suggested both the wave-particle duality of the X-ray and the validity of the space lattice hypothesis [44]. When monochromatic X-rays interact with a crystalline material, they undergo constructive and destructive interference caused by the periodic arrangement of atoms within the crystal lattice. The interference generates a diffraction pattern, which is recorded on a detector and, subsequently, used to deduce structural information about the sample. Since its discovery, XRD has been used by engineers, particularly in materials science, chemists, physicists, and biologists, aiding them in the discovery, development, and optimization of novel compounds with tailored properties for numerous practical and research applications [45].

The foundation of XRD lies in Bragg's law, formulated by Sir William Lawrence Bragg and his father Sir William Henry Bragg [46]. Bragg's law states that the conditions for the constructive interference of X-rays in a crystal lattice are determined using the equation:

$$n \cdot \lambda = 2 \cdot d \cdot \sin\theta \quad (1)$$

where:

- n is the order of the diffraction peak (usually 1 for primary peaks);
- λ is the wavelength of the incident X-rays;
- d is the lattice spacing of the crystal planes;
- θ is the angle between the incident X-rays and the crystal plane.

From an experimental setup perspective, XRD typically involves an X-ray source, such as a sealed tube or a synchrotron radiation source, which emits X-rays of a specific wavelength [47]. Next, the X-rays pass either through a system of slits to produce a divergent beam or through a collimator to produce a parallel beam that interacts with the crystalline sample [48,49]. The diffracted X-rays are then collected with a detector, such as a scintillation counter or a semiconductor detector, which records the intensity of diffracted X-rays as a function of the diffraction angle (2θ) [50].

The analysis of an obtained XRD pattern allows researchers to identify the phases in the studied material, determine their relative abundances, and estimate the crystallite size and microstrain of the sample [51–54]. Additionally, the crystallographic orientation can be deduced by analyzing the preferred orientation of the crystallites [55].

X-ray diffraction finds extensive applications in multiple scientific fields. Professionals in chemistry, physics, geology, and materials science use this technique for both qualitative and quantitative analysis [56]. The first application of XRD was in the field of geology for the identification of minerals and rocks, and the technique has decisive contribution in crystal system determination [57]. In the microelectronics industry, qualitative phase analyses, stress measurements, and microstructural features determinations are routinely performed using XRD [58]. In the pharmaceutical industry, XRD is applied to examine formulations by providing polymorph identification, relative abundance, and degree of crystallinity. Moreover, nonambient XRD analysis is useful for the study of moisture

influence on drug properties [59]. Although not as popular, XRD is also used in the forensic sciences for the analysis of soils, explosives, pigments and paints, alloys, metals, or drugs. Compared with other techniques, XRD has several advantages such as the ability to work with small-volume samples; it is a nondestructive method, and it allows for the identification of phases in mixtures [60]. The Fourier analysis of XRD patterns is a clever technique used for the determination of the local arrangement of atoms, which proved the noncrystalline nature of soda–silica glass [61].

Over the years, XRD instrumentation has undergone significant advancements. Traditional parafocusing instruments were developed during the 1950s and are mostly used in Bragg–Brentano geometry. However, this configuration can introduce significant systematic errors such as specimen displacement. Parallel-beam diffractometers minimize errors arising from sample displacement and transparency but have the disadvantage of poor particle statistics. Modern XRD systems offer high-resolution detectors, faster data acquisition rates, improved sample handling mechanisms, and portability for in situ and operando studies [62].

1.2. Applications of XRD Data Analysis

XRD data analysis is of paramount importance for many scientific and industrial applications. This section highlights the significance of XRD data analysis and its role in advancing materials science, research, and technology.

One of the primary objectives of XRD data analysis, whether acquired on single crystals or polycrystalline samples, is the determination of crystal structure. Thus, the technique is used for the identification of the arrangements of atoms within crystalline materials in terms of lattice parameters, unit-cell dimensions, crystal symmetry and subsequent space group. As emphasized in a paper by Zok [63], the mechanical behavior of solid materials is strongly connected to the crystal structure; consequently, by controlling the processing parameters for obtaining a desired structure, the compressive strength of a material might be controlled.

XRD data analysis enables the identification of different phases in a sample. The set of d lattice spacings and corresponding I (intensity) values are characteristic for a material like a fingerprint is for a human [62]. Moreover, many materials can exist in various crystallization systems (polymorphic forms) and can undergo phase transformations under different temperature or pressure conditions. Identifying the phases accurately is crucial for ensuring material purity, assessing the success of synthesis processes, and characterizing complex multiphase materials.

Quantitative phase analysis allows researchers to determine the relative abundance of different phases in a sample. This determination is useful for various materials, such as cement, ceramics, steel, alloys, electronic materials, and composite materials. In the cement industry, the quantitative phase analysis of clinker provides information for the control of kiln parameters, whereas the analysis of Portland cement provides the quality assessment of the finished product [64]. In the case of traditional ceramics, the abundance of phases and the evaluation of the crystallinity degree is decisive for establishing the thermal processing parameters [65]. Assessing the phase fraction of zirconia polymorphic phases is of great importance in dental ceramics applications for predicting the mechanical behavior of the material [65]. The mechanical behavior of stainless steel is primarily governed by its martensite and austenite content. Even if the material is textured, like in the case of orthodontic wires, assessments can still be made based on X-ray diffraction patterns [66]. In the case of electronic materials, Angus et al. established the crystallization kinetics of $\text{PbZr}_{1-x}\text{Ti}_x\text{O}_3$ using an in situ X-ray diffraction study [67].

XRD data analysis plays a vital role in texture and microstructure characterization. Texture refers to the preferred orientation of crystalline planes in a material, influencing its anisotropic properties. Understanding texture is crucial in fields like metallurgy, where it impacts mechanical properties such as strength and ductility [68]. Additionally, XRD anal-

ysis provides information on crystallite size, microstrain, and defects, which are essential in assessing material stability and mechanical behavior [69].

1.3. Motivation for Machine Learning in XRD Data Analysis

The motivation for incorporating machine learning (ML) techniques into the analysis of XRD patterns stems from the increase in volume of available data and the need for accurate phase identification, as well as the quantification of multiphase mixtures with varying raw data quality. The following points highlight the advantages of ML over traditional methods of XRD data analysis:

- Handling big data: The development of synchrotrons has enabled the fast acquisition of XRD patterns, which results in a significant increase in the amount of data collected during experiments. The fine-tuning of beam-time experiments depends on the analysis of patterns, and, thus, an automatic processing flow would be required to further increase its autonomy. In this regard, machine learning routines using clustering represent a potential solution to the challenges faced by the scientific community [70];
- Automated phase identification: In traditional XRD data analysis, the manual identification of phases in complex samples can be time consuming and error prone, especially when dealing with overlapping peaks or noisy data recorded in cases in which short measurement times are a must. ML algorithms can accurately identify and quantify phases and even predict material features from XRD patterns. Moreover, the successful implementation of the algorithms would save time while also benefitting XRD users who are not experts [71–73];
- Quantitative phase analysis (QPA): Several traditional methods with different complexity and sample preparation requirements are available for the evaluation of phase fractions, including the reference intensity ratio (RIR) method, which requires the introduction of an internal standard calibration [74]; the whole pattern fitting procedure [51–54]; or Rietveld refinement [54]. Each of the traditional methods is time consuming and requires trained personnel to deliver accurate results. ML algorithms, such as regression models and support vector machines, can efficiently estimate phase proportions based on trained patterns, greatly improving the accuracy and speed of QPA [75,76].

The main goal of this study was to assess currently available ML methods for XRD data analysis, their applications, challenges, and limitations, as well as future directions and emerging trends. For this purpose, all results obtained after a search procedure in the Web of Science and Scopus databases using the search term “machine learning X-ray diffraction” were assessed by two reviewers using a blind method. From the total number of 754 entries, 513 were identified as unique, and 11% of these were included in the current review based on several selection criteria:

- The bibliographic source must refer to the use of machine learning methods for the analysis of XRD patterns;
- The bibliographic source must be written in English;
- The bibliographic source represents a peer-reviewed article, conference proceeding, or an edited book.

The findings are presented and compared to traditional XRD data analysis methods.

2. Challenges in Traditional XRD Data Analysis

XRD data processing allows for various applications based on the determined structure parameters and phase composition. However, traditional XRD data analysis methods face several challenges that can hinder the accurate and efficient interpretation of experimental measurements.

2.1. Data Preprocessing and Reduction

Raw XRD data often contain noise, background signals, and artefacts that can obscure diffraction peaks and affect the accuracy of subsequent analysis. Additionally, the sheer volume of data generated by modern XRD instruments can be overwhelming, making data reduction and handling a significant challenge.

Traditional XRD data preprocessing is performed most often by Sonneveld and Visser technique [77], which consists of background and noise level determination followed by the refining of peak positions and intensities. The approximation of the background can be performed by taking into consideration 5% of the data points, each point being expressed as the arithmetic mean of its neighbors. Then, the noise is separated from the peak contribution by assessing the standard deviation and the mean value of noise contribution. After the determinations of background level and noise, peak discovery is performed by searching for negative regions in the second derivative of the scan, which is calculated using a sliding polynomial filter according to Savitsky and Golay [78]. The result of the preprocessing will be a list of peak positions (angles or converted d -spacings) and their corresponding intensities.

2.2. Phase Identification and Crystallographic Analysis

Identifying and distinguishing between multiple phases present in a complex sample can be challenging, especially when diffraction peaks overlap or exhibit broadening due to microstructural effects. Additionally, crystallographic analysis to determine lattice parameters and symmetry requires meticulous peak indexing and fitting.

The successful accomplishment of this task often requires experienced personnel. In terms of procedures, several strategies were reported. Manual search was first implemented by Hanawalt [79] and consisted of comparisons among the three most intense characteristic d -spacings determined from the patterns. Crystallographic analysis of new materials can be performed using a full-pattern fitting algorithm like Rietveld refinement [79], which compares a measured profile with a calculated one from crystal structure data. This method accounts for several contributions to the XRD pattern: scale factor, multiplicity, Lorentz polarization factor, structure factor, absorption, and extinction. Thus, simultaneous phase identification and crystallographic analysis are enabled.

2.3. Quantitative Phase Analysis

Quantitative phase analysis involves determining the relative proportions of different phases in a sample. Traditional methods like those described in Section 1 (RIR, whole pattern fitting procedure, or Rietveld refinement) can be computationally intensive, especially for samples with many phases, leading to long processing times.

2.4. Microstructural Characterization

Extracting microstructural information, such as crystallite size, microstrain, and texture, from XRD data requires specialized techniques and complex mathematical models. Additionally, microstructural effects can lead to peak broadening and distortions, complicating the interpretation of diffraction patterns.

Microstructural characterization can be accomplished through a Scherrer analysis [80], Williamson–Hall plot [81], and Warren–Averbach analysis [82], which relate peak broadening to crystallite size and microstrain. Texture analysis requires specialized mathematical models to deduce preferred crystallographic orientations in polycrystalline materials like March–Dollase model [83].

3. Introduction to Machine Learning

Machine learning (ML) is a type of artificial intelligence where computer algorithms “learn” from example data and can make predictions without being explicitly told what to do or how to achieve their targets. ML is a powerful data analysis tool, used in diverse applications, such as data processing, pattern recognition, and automated decision making.

To build a machine learning model capable of making predictions, training data are first collected and processed, then a (machine learning) model is chosen, which is then trained and evaluated for the intended task [84].

Machine learning encompasses several paradigms, each offering unique approaches to tackle different data analysis challenges. By leveraging these ML techniques, researchers can automate efficient and accurate data interpretation, leading to significant advancements in materials science, chemistry, and other fields. We briefly present five fundamental paradigms of machine learning below.

Supervised learning is a type of ML in which an algorithm learns from sets of labeled data where both the input data and corresponding desired output are provided during training. The goal of supervised learning is to learn (or optimize) the parameters of a mapping function and use it to accurately predict the output for new inputs that are not available to the algorithm during the training phase [84]. Common algorithms and structures used in supervised learning include linear regression (when the output is a continuous variable), support vector machines (SVMs), decision trees (DTs), random forests (RFs), k-nearest neighbors (KNN)s, naïve Bayes (NB), and neural networks (NNs) [85].

- SVMs, which are very suitable for binary classification and linearly separable data, work by transforming (mapping) the input data to a high-dimensional feature space such that different categories become linearly separable [86];
- Decision trees work (as their name implies) by inferring simple if–then–else decision rules from the data features and can be visualized as a piecewise constant approximation of the data [86];
- Random forests (RFs) are ensemble methods that make predictions by aggregating the output of multiple decision trees. Randomness is built into the algorithm to decrease the variance in the predictions of the generated forest. RFs are robust in overfitting and useful for both regression and classification applications. A different ensemble method, called “extremely randomized trees” may be employed to increase the prediction power by reducing the variance [86];
- Nearest neighbor methods predict labels from a predefined number of training samples that are closest to the given input point; in KNNs, this number is a user-defined constant [86];
- Naïve Bayes methods are an application of Bayes’ theorem under the “naïve” assumption that input features are independent from each other [86]. For example, this assumption would be violated when using length, width, and area as input features in the same data analysis workflow;
- Neural networks can identify and encode nonlinear relationships in high-dimensional data; sometimes NNs used in machine learning are referred to as ANNs, where the letter A stands for “artificial”. NNs are composed of layers of “neurons” that mimic their biological counterparts: they have multiple input streams (which work like dendrites) and a single output activation signal (similar in function to an axon). Each layer of neurons has adjustable parameters that are used to compute the output signal. Based on the connectivity between layers, NNs can be categorized as dense (whereby each neuron in a layer is connected to every neuron in the previous layer) or sparse. The term multilayer perceptron (MLP) is sometimes used to refer to modern ANNs; MLPs consist of (at least three) dense layers: input, output, and at least one hidden (other) layer [86].

Unsupervised learning involves finding structure and relationships in data without using explicit (output) data labels. The ML algorithm tries to identify patterns or clusters in the data that are not known a priori, making unsupervised learning useful for tasks such as data exploration, dimensionality reduction, or anomaly detection [84]. Common unsupervised learning algorithms include K-means clustering, Gaussian mixture, fuzzy c-means (FCM), hierarchical clustering, principal component analysis (PCA), and autoencoders [87].

- The K-means method is used for partitioning the data into a predetermined number of K disjoint clusters, which are chosen with the aim to evenly distribute the variance between different clusters [86];
- Gaussian mixture models are probabilistic in nature and try to represent the input data as a mixture of a finite number of Gaussian distributions with unknown parameters to be learned during training [86];
- In fuzzy clustering, points are not assigned (only) to specific clusters; instead, each point has an association (weight) with each cluster. Since each point can belong to more than one cluster, fuzzy c-means is sometimes referred to as soft K-means [86,88];
- Hierarchical clustering works by successively merging or splitting clusters to create a tree-like (nested) representation of the data. In agglomerative clustering, a hierarchy is built using a bottom-up approach (each observation starts as a single-item cluster, and clusters are successively merged until a single, all-encompassing cluster is formed) [86];
- PCA is a linear decomposition technique used for reducing the dimensionality of the data by projecting it onto a lower dimensional space while preserving the most amount of variance; in kernel PCA, the algorithm is applied to a transformed version of the data [86,88];
- Autoencoders use ANNs to learn an encoder–decoder pair that can efficiently represent unlabeled data: the encoder compresses the input data, while the decoder reconstructs an output from the compressed version of the input. Autoencoders are suitable for unsupervised feature learning and data compression [86].

Deep learning utilizes artificial neural networks with multiple layers (deep architectures) to learn hierarchical representations from data [84]. Common algorithms include convolutional neural networks (CNN), and recurrent neural networks (RNNs, which are more suitable for sequential data such as speech in natural language processing applications) [87].

- CNNs, belonging to the artificial neural network group, are commonly used in image data analysis. Their name stems from the mathematical operation convolution, which is used in at least one of the neuron layers, instead of the simpler matrix multiplication used by regular ANNs [86];
- The architecture of RNNs makes them suitable for identifying patterns in sequences of data and are used for applications such as speech and natural language processing. In contrast to regular ANNs, in which calculations are performed layer-by-layer from input to output, in recursive NNs information can also flow backward, allowing the output from some nodes to affect their inputs in the future (in subsequent evaluations of the neural network), thus introducing an internal state useful for inferring meaning in text processing based on words previously read by the algorithm [86,89];
- Long short-term memory (LSTM) units were introduced within the RNN framework to enable RNNs to learn over thousands of steps, which would have not been possible otherwise because of the problem of vanishing or exploding gradients (that accumulate and compound over multiple iterations of the NN) [86,89].

In reinforcement learning (RL) an agent learns to make decisions by repeatedly interacting with an environment. The agent receives feedback (rewards or penalties) based on its actions and uses this information to tune its parameters and improve its decision-making process over multiple iterations. It is commonly used in robotics, computer games, and control systems [84].

Transfer learning can be used when the required knowledge for one task or domain can be leveraged by using insight gained in a different but related task or domain. Instead of training a model from scratch for a specific task, transfer learning allows pretrained models to be reused and fine-tuned, often with limited labeled data [90].

4. Applications of Machine Learning in XRD Data Analysis

Over the last decade, ML has found various applications in XRD data analysis, revolutionizing the way researchers extract information from XRD patterns. In this section, we explore how ML techniques have been applied to different aspects of XRD data analysis. An overview of machine learning algorithms and their use cases in X-ray diffraction data analysis is presented in Figure 2.

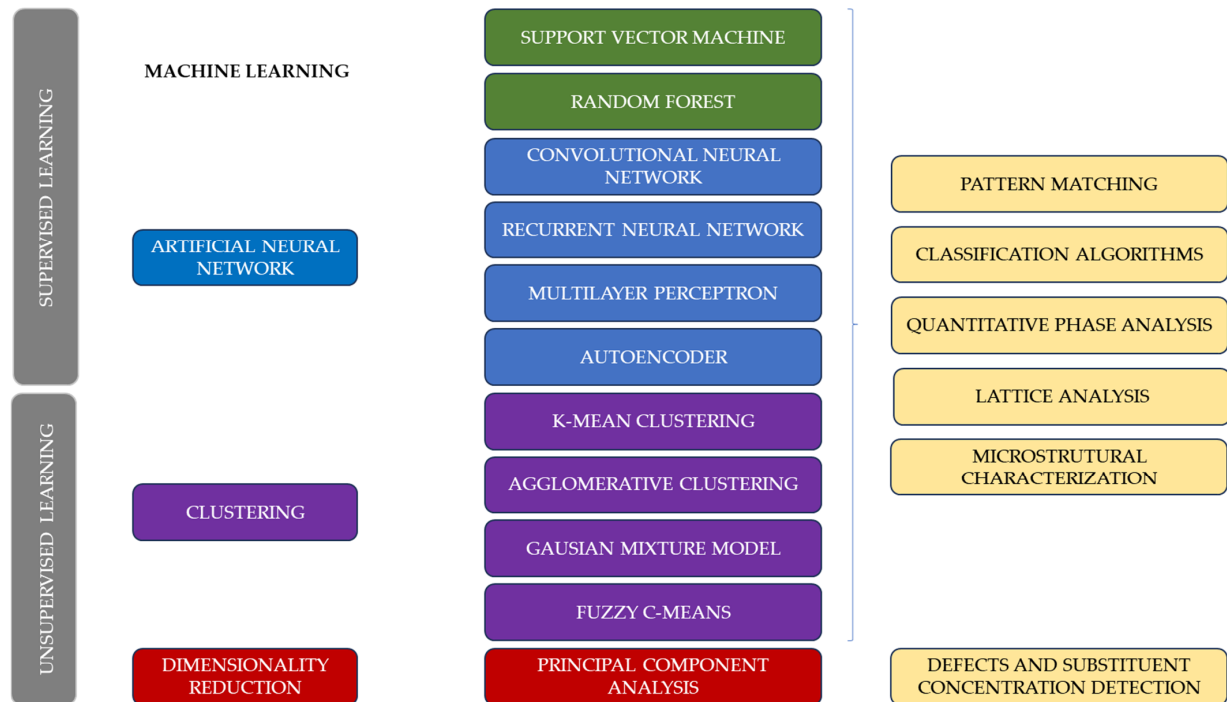


Figure 2. Machine learning algorithms used in X-ray diffraction data analysis.

The remainder of this section presents the main results from publications that have used ML as part of their XRD data analysis, as well as some challenges that are inherent to ML. Many of the works cited herein have expressed the performance of their classification algorithms in terms of accuracy, which is a measure of how often ML models correctly predict the desired outcome. Accuracy is calculated by dividing the number of correct predictions by the total number of predictions made by the model [88]. Precision is another measure that is often used in binary classification tasks in which the algorithm predicts whether an item belongs (or not) to a target category. Precision is calculated as the proportion of correctly classified items out of all items that were predicted to be part of the target category [88]. Both accuracy and precision must have values between 0 and 1, and they were reported for predictions made on the test set, unless explicitly stated otherwise.

4.1. Pattern Matching and Classification Algorithms

Wang et al. [91] applied support vector machine and deep learning methods (convolutional neural network) to extract image features from synchrotron data streams and compared the accuracy of their algorithm against synthetic and real datasets. Image features to be identified and classified belong to several groups (experiments, instrumentation, imaging, scattering features, samples, materials, and specific substances), as depicted in Table 1. The experimental dataset consisted of 2832 grayscale X-ray images and the dataset used to train the CNN consisted of 100,000 synthetic (simulated) images. In what concerns the deep learning algorithm, the basic units were a convolutional layer, a subsampling/pooling layer, and an activation layer based on ReLU (rectified linear unit) activation function. The mean average precision determined for the SVM algorithm on the synthetic dataset was

0.6705, compared with 0.771 for the CNN algorithm, which showed superior performance in this instance.

Table 1. Attributes groups to be identified from synchrotron data streams images. Data from reference [91].

Group Number	Group Attributes	Labels
G1	Experiments	GIWAXS, GISAXS, TSAXS, TWAXS, GTSAXS, Theta sweep, and phi sweep.
G2	Instrumentation	Beam off image, photonics CCD, MarCCD, Linear beamstop, saturation, asymmetric (left/right), and circular beamstop.
G3	Imaging	Specular rod, weak scattering, 2D detector obstruction, strong scattering, saturation artifacts, misaligned, beam streaking, blocked, bad beam shape, direct, object obstruction, empty cell, parasitic slit scattering, and point detector obstruction.
G4	Scattering Features	Horizon, peaks: isolated, ring: oriented z, halo: isotropic, ring: isotropic, ring: textured, higher orders: 2 to 3, ring: oriented xy, vertical streaks, peaks: many / field, diffuse high-q: isotropic, higher orders: 4 to 6, higher orders: 7 to 10, Bragg rods, ring: anisotropic, peaks: along ring, diffuse low-q: isotropic, Yoneda, halo: oriented z, high background, ring: spotted, peak: line Z, peaks: line xy, diffuse low-q: anisotropic, many rings, diffuse low-q: oriented z, diffuse low-q: oriented xy, diffuse specular rod, smeared horizon, symmetry ring: 4-fold, higher orders: 10 to 20, ring doubling, halo: anisotropic, specular rod peaks, ring: oriented other, peaks: line, diffuse high-q: oriented z, peak doubling, halo: oriented xy, diffuse high-q: oriented xy, peaks: line other, waveguide streaks, higher orders: 20 or more, substrate streaks/Kikuchi, diffuse low-q: oriented other, halo: spotted, diffuse low-q: spotted, and diffuse high-q: spotted.
G5	Samples	Thin film, ordered, single crystal, grating, amorphous, composite, nanoporous, powder, and polycrystalline.
G6	Materials	Polymer, block-copolymer, and superlattice.
G7	Specific Substances	P3HT, SiO ₂ , PCBM, rubrene, PS-PMMA, silicon, MWCNT, PDMS, AgBH, and LaB ₆ .

The detection of synchrotron image features was also studied by Czyzewski et al. [92], which aimed to identify seven types of flaws: ice rings, diffuse scattering, background rings, nonuniform detector responses, loop scattering, strong background, and digital artefacts. The group compared several algorithms (SVM; naïve Bayes—NB; k-nearest neighbors—KNNs; random forest—RF) with CNN, in which they used different inputs (cartesian coordinates and polar coordinates with different interpolation methods: *min* and *max*). The dataset used in the study comprised 6311 diffraction images from the Integrated Resource for Reproducibility in Macromolecular Crystallography, from which 5048 were used as a training set, 631 were used as a validation set, and, subsequently, 632 were used in the testing set. The accuracy of class-specific predictive performance for the different classifier algorithms is shown in Table 2. Their results clearly show that CNN performance was consistently better than any other of the tested classifiers. Moreover, it is worth mentioning that the differences among the CNNs were generally approximately 0.01–0.02.

Table 2. Accuracy of the class-specific predictive performance for the different classifier algorithms. Data from reference [92].

Class	Classifier						
	SVM	NB	KNN	RF	CNN: Cartesian	CNN: Polar-Min	CNN: Polar-Max
Artifact	0.85	0.78	0.87	0.91	0.94	0.93	0.92
Background Ring	0.72	0.61	0.72	0.86	0.92	0.91	0.90
Diffuse Scattering	0.93	0.45	0.93	0.93	0.96	0.95	0.97
Ice Ring	0.14	0.80	0.93	0.95	0.99	0.99	0.98
Loop Scattering	0.70	0.62	0.71	0.83	0.94	0.95	0.96
Nonuniform Detector Response	0.45	0.68	0.75	0.81	0.87	0.89	0.89
Strong Background	0.90	0.87	0.89	0.93	0.94	0.91	0.93

Chakraborty and Sharma [93,94] compared several algorithms (RF, KNN, decision tree, SVM, and gradient boosting) with the CNN for the purpose of the classification of crystal systems into seven categories: triclinic, monoclinic, orthorhombic, tetragonal, hexagonal, rhombohedral, and cubic. The training dataset consisted of 164 compounds extracted from the Inorganic Crystal Structure Database with a similar composition, expected crystal symmetry, and space group. Their work showed that the CNN performed better than the other studied algorithm achieving a cross-validation accuracy for crystal system classification of 95.6% as compared to 55% for naïve Bayes, 64.3% for KNN, 68.5% for logistic regression, 56.5% for RF, 45.6% for decision trees, 67.1% for SVM, 62.3% for decision trees and 65.4% for deep neural network.

Massuyeay et al. [95] explored RF and CNN to distinguish between perovskite and non-perovskite-type materials in a series of hybrid lead halides. The synthetic (simulated) dataset was based on 998 crystal structures from the Cambridge Structural Database: 375 perovskite-type compounds (50 chlorides, 105 bromides, and 220 iodides) and 623 non-perovskite-type compounds (50 chlorides, 139 bromides, and 426 iodides). The study also used experimentally measured X-ray powder diffraction data on 23 freshly prepared lead halides: 9 previously published (and reported in Cambridge Structural Database) and 14 new compounds. The categories used for the classification were perovskite and nonperovskite. On the one hand, in the RF algorithm, the number of trees was set to 100, with a maximum of 10 levels in tree, a minimum number of 2 samples on a leaf, a minimum number of samples to split a node of 10, and a step size for the XRD patterns of 2.18° . On the other hand, the CNN was designed with 23 layers and simulated patterns acted as 1D input. The mean values of the accuracy obtained after the classification were 0.92 in the case of CNN and 0.89 in the case of RF. In what concerns the 23 experimentally synthesized samples, the mean values of accuracy were 0.73 for CNN and 0.78 for RF. The lower accuracy obtained for the experimentally raw patterns was explained by the authors in terms of the different effects, such as the preferential orientation and different signal/noise ratio [92–95].

In geothermal fields, the classification of rock cuttings is important for understanding the geothermal system and for selecting a promising site [96]. Rock cuttings containing 24 minerals (Table 3) were obtained from two wells in the Hachimantai geothermal field, which may have formed during hydrothermal alteration according to Ishitsuka et al. For the assessment of three ML algorithms, namely, K-mean clustering, Gaussian mixture model, and agglomerative clustering [96], the authors prepared a dataset of 88 simulated samples with four mineral distributions along a well down to 1000 m with a depth spacing of 10 m. The classification of the samples was performed using three labels: quartz index, temperature, and depth.

Table 3. List of minerals identified from XRD data for the rock cuttings. Data from reference [96].

Mineral Group	Mineral
Clay Minerals	Smectite, chlorite, sericite, and kaolinite
Zeolite Minerals	Laumontite and wairakite
Silica Minerals	Tridymite and cristobalite
Silicate Minerals	Clinopyroxene, epidote, prehnite, antrophyllite, and biotite, cordierite, and talc
Oxide Minerals	Magnetite, ilmenite, hematite, anatase, and rutile
Sulfide Minerals	Marcasite
Sulfate Minerals	Anhydrite and alunite
Carbonate Minerals	Calcite

The K-means clustering and Gaussian mixture algorithms provided similar results, whereas the agglomerative clustering showed unique classification outcomes. The methodology proposed by the authors is applicable to other boreholes in geothermal fields.

In materials science, the composition–structure–properties (CSP) paradigm is often used for predicting materials behavior under certain conditions. Yuan et al. [97] developed a supervised machine-learning algorithm to classify materials encountered in aviation security determinations based on CSP and XRD patterns without material identification. For this purpose, a dataset of 206 relevant materials in stream of commerce baggage (explosives, prohibited flammables, acids, plastic, metals, food, etc.) was prepared. It is worth mentioning that the dataset included both crystalline and amorphous compounds, which can easily be discerned from XRD data. The dataset was classified by crystalline/noncrystalline, solid/liquid, explosive/nonexplosive, prohibited/allowed classes with satisfactory results, as the authors state.

In what concerns pattern matching and classification, ML methods offer rapid automation and complex pattern recognition in XRD data analysis, improving accuracy and adaptability. However, their effectiveness relies on the availability of labeled data, and complex models might overfit noise. Further improvement of the models might arise from using larger databases compared to those reported by the authors in our review, such as Crystallography Open Database (505,398 entries) or Powder Diffraction File (1,186,076 entries). Conventional methods lack automation and struggle with intricate patterns, but they are more interpretable and require less data.

4.2. Quantitative Phase Analysis

Phase identification and phase-fraction determination of multiphase inorganic compounds were performed by Lee et al. [98] for the Li-La-Zr-O compositional system using a data-driven approach. The authors prepared a training dataset starting from a total of 218 known inorganic compounds from the Li-La-Zr-O quaternary compositional system, which comprised 21 independent structures. In the simulation process, lattice parameters variation and randomly chosen peak profile parameters, as well as mixing parameters, were considered. Two training datasets containing 89,943 (D1) and 180,056 (D2) synthetic patterns were generated for the phase identification algorithms. For phase-fraction prediction a total of 13,930,000 (D3) XRD patterns were prepared. Moreover, a real-world dataset was obtained by acquiring XRD patterns of conventionally prepared inorganic powders. The prepared samples were synthesized from Li_2O , La_2O_3 , and ZrO_2 powders by mixing and subsequent firing at 1000 °C.

Phase identification was performed using the CNN, KNN, RF, and SVM algorithms. A comparison of the highest test accuracy values obtained in each case are presented in Table 4, which shows the superior performance of the CNN algorithm over other ML methods included in the study.

Table 4. Phase identification test accuracy values for the CNN, KNN, RF, and SVM. Data are from Reference [98].

	Dataset	CNN	KNN	RF	SVM
Synthetic dataset	D1-trained	94.36%	12.15%	56.82%	33.60%
	D2-trained	96.47%	13.08%	63.62%	42.74%
Real-world dataset	D1-trained	88.88%	24.44%	17.78%	13.33%
	D2-trained	91.11%	22.22%	15.56%	13.33%

In what concerns the phase-fraction determination, the authors used an artificial neural network with a fixed architecture. This algorithm was also compared in terms of test errors by assessing the mean square error and R-square values of the phase-fraction regressions (Table 5). From the presented values, one can assess the KNN and SVM algorithms as best performing at the phase-fraction regression tasks.

Table 5. Phase-fraction regression of the MSE and R² values for the ANN, KNN, RF, and SVM. Data are from Reference [98].

	Dataset	ANN	KNN	RF	SVM
Synthetic dataset	MSE	0.004612	0.002507	0.003987	0.001809
	R ²	0.923253	0.956168	0.930789	0.968471
Real-world dataset	MSE	0.008260	0.008035	0.006453	0.002423
	R ²	0.821816	0.860250	0.894196	0.958704

Quantification of the mineral composition of gas hydrate-bearing sediments was performed by Park et al. [99] using various algorithms including CNN, recurrent neural network (RNN), multilayer perceptron (MLP), RF, and long-short term memory (LSTM). A total of 488 materials with complex compositions, including 12 minerals (quartz, albite, opal-A, calcite, muscovite, dolomite, chlorite, kaolinite, illite, pyrite, NaCl, and K-feldspar) were quantified using the mentioned algorithms. The algorithms showed promising results for predicting mineral composition even for those which showed an amorphous broad peak. On the other hand, for samples with low opal-A content, compared to the others in the dataset, all algorithms had high errors compared to the traditional indexing method. This might have occurred because the training dataset contained only hundreds of patterns, and, because of this fact, the RF algorithm showed the highest possibility among the studied ones to predict mineral compositions.

The ML methods used for the quantitative phase analysis offer automation and adaptability advantages, while the conventional methods are established but time consuming. The ML algorithms showed good results for the identification of composition and limited accuracy in terms of phase fraction determination. To the best of our knowledge, to date there is no available database containing experimental patterns with complex compositions and different phase fractions, which may contribute to the limited success of these ML models. For the quantitative phase analysis task, a hybrid model using ML for identification and classification coupled with conventional phase fraction determination based on Rietveld refinement or whole powder pattern fitting procedures would be the most suitable approach from our perspective.

4.3. Lattice Analysis

Pasha et al. [100] proposed a specialized learning engine for identifying the cubic structure of materials regardless of their composition. Their approach borrows from human gene regulation theory to conduct the training of a group of distributed neural networks, where each neural network is managed by the engine in a similar manner to how genes are regulated inside the human body. The application of this approach to the classification of cubic lattices showed an accuracy rate over 99% over a representative range of materials.

Since the proposed method is also computationally efficient, it can potentially be used for partial automation of the complex XRD analysis task.

The space group determination problem from powder XRD patterns was studied by Vecsei et al. [101] employing a dense neural network method for the classification of crystal symmetry. The model was trained on theoretically computed XRD patterns and tested on both theoretical and experimental data. The authors report a space group classification accuracy on real experimental data of approximately 54%, with incorrectly classified structures often exhibiting close symmetry to the correct space group. The certainty of the predicted space group was ascertained using a softmax activation function for the output layer, which was chosen because it produces results that can be interpreted as a probability distribution. Notably, when uncertain predictions were dropped (i.e., those for which the highest softmax probability was less than 0.45), the classification accuracy for the remaining data (approximately half of the initial dataset) improved to about 82%. Thus, the method may be used to automatically characterize a subset of the data (for which the algorithm has arbitrarily high certainty), leaving the remaining diagrams to be processed manually.

Suzuki et al. [102] also approached crystal system and space group classification using ML models. Their most successful model, which exceeded 90% accuracy for crystal system classification, is built on an extremely randomized trees (ensemble) algorithm. The positions of the ten left-most peaks (low values of 2θ) in the XRD were included as part of the input to the model to mimic the process that human experts perform, while the decision tree model was chosen to obtain an interpretable model that would provide insight into the classification process. The model also performed space group classification, with an accuracy of 80% for the most likely candidate; when considering a list of the five top candidates, the accuracy increased to above 92% (probability that the list contains the correct space group). Despite the generally high accuracy, the model significantly underperformed on the triclinic crystal system, with an accuracy just below 50%; this reduced performance was attributed to a shortage of triclinic training data, suggesting that any ML-based approach would be affected by this issue.

A report by Oviedo et al. [103] introduces a CNN-based ML model for space group and crystal dimensionality. The ML algorithms they tested used both experimental and computed XRD patterns, which were generated using a data augmentation process based on domain knowledge. As stated by the authors, physics-informed data augmentation is more robust (avoids overfitting), model independent, and offers higher interpretability compared with explicit regularization, while also being more robust than noise-based data augmentation. The reported accuracy for the dimensionality and space group classification was 93% and 89%, respectively, for a set of 115 thin film metal halides, when using data augmentation. Interestingly, the classification accuracy was reduced to approximately 84% and 80% (for dimensionality and space group, respectively) when using only simulated XRD patterns for training and keeping all of the experimental data for the testing phase, highlighting both the importance of data augmentation and some of its limitations for the data analysis applications of real XRD data.

XRD patterns collected at several temperatures can explain structural transitions. In this regard, fluctuations in XRD patterns are an effect of the charge density waves that show the change in the size of the unit cell and of those that involve intra-unit cell distortions. Venderley et al. [104] developed an unsupervised and interpretable machine learning algorithm for the study of phase transitions in $(\text{Ca}_x\text{Sr}_{1-x})_3\text{Rh}_4\text{Sn}_{13}$ and $\text{Cd}_2\text{Re}_2\text{O}_7$ materials and plotted the phase diagram of the former; by applying the introduced model to the analysis of thousands of Brillouin zones, the authors demonstrate the potential application of their approach to the real-time analysis of temperature dependencies and automation of the inverse scattering problem [105]. The same model [104] was applied by Kautzsch et al. [106] to reveal the structural evolution of the kagome superconductors AV_3Sb_5 ($A = \text{K}, \text{Rb}, \text{and Cs}$) through the charge density wave order parameter.

X-ray Laue microdiffraction scan analyses were indexed using a machine learning method based on clustering and labeling algorithms by Song et al. [107]. Their model was tested on four materials (CuAlMn, AuCuZn, and CuAlNi alloys and BaTiO₃ ceramics). To increase the computational efficiency of the approach and allow the model to be harnessed as part of real-time processing in a synchrotron pipeline, the original Laue patterns were processed with a CNN autoencoder for dimensionality reduction. Dropout layers were used as part of the CNN architecture to mitigate overfitting, and PCA was applied to the output of the CNN encoder to further reduce feature space dimensionality.

The analysis of the phase transformations in Ni-Ti-Co thin films was performed by Al Hasan et al. [108] aided by unsupervised hierarchical clustering machine learning. The ML model describes phase mixtures belonging to multiple cubic structures (*Pm3m*, *Fm3m*, and *Im3m*), as well as orthorhombic and hexagonal structures. A total of 177 XRD patterns were analyzed and, ultimately, clustered into six groups based on composition. Together with the crystal structure, phase, and thermal hysteresis behavior, this study maps the material properties of Ni-Ti-Co alloys onto the chemical composition space.

In another approach of unsupervised machine learning, a fuzzy c-means (FCM) clustering algorithm was used by Narayanachari et al. [109] to classify tantalum oxynitride thin film structures obtained by pulsed laser deposition. The unsupervised ML analysis grouped XRD patterns into four clusters, which corresponded to mixtures having similar chemical and phase composition. Their overall results showed that the proposed procedure (including experimental methods and ML data analysis) is efficient and could enable the identification of deposition parameters for obtaining a desired phase.

ML techniques used for lattice analysis tasks demand substantial training data and might lack transparency in decision making, limiting their interpretability. The accuracy of the models restricts their use, especially for complex structure analysis (triclinic and monoclinic) or for space group determinations in cases where features showing the difference among several space groups are not apparent. Conventional methods, while often slower and manual, provide a well-established framework for crystallographers to validate results.

4.4. Defects and Substituent Concentration Detection

Determination of substituent concentrations in [Sm_{1-y}Zr_y]Fe_{12-x}Ti_x crystal structures was performed by Utimula et al. [110] using a dynamic time-warping (DTW) analysis of simulated XRD patterns coupled with the Ward linkage method for clustering based on Euclidian distances between pairs of time series. The method had an accuracy of approximately 96% distinguishing different Sm/Zr substitution concentrations. The method is less suitable for distinguishing between XRD patterns for Fe/Ti substitutions, with a success rate of only 33%. While this issue can be mitigated by performing the clustering on a different dissimilarity measure, such as DTW weighted by the magnetization per unit volume of the sample, these data (i.e., magnetization) will not always be available for experimental XRD patterns. The authors of the work state that their algorithm is applicable to other systems where atomic substitutions within a phase must be tuned.

In a different publication by Utimula et al. [111] an autoencoder was used to compress XRD patterns to two dimensions. The hidden layers used ReLU activation functions, while the final layers used tanh (hyperbolic tangent) for the encoder and a linear function for the decoder. Although the features learned by the autoencoder algorithm do not have physical significance in general, in this case, they appear to be related to the composition of the samples, which is justifiable through the connection among atomic substitutions, lattice constants, and XRD peak shifts. Clustering of the feature space was performed, using local information (linear interpolation) instead of unsupervised ML algorithms, such as k-means, since the different groups did not form simply connected regions within the feature space. The authors applied the autoencoder to assess the significance of XRD peaks by removing peaks and measuring the shift observed on the feature space. Additionally, the feature space could also be used to simulate the XRD patterns of small concentration

changes by conducting an interpolation on the feature space instead of performing more costly computer simulations or experiments.

For the determination of defects and substituent concentration detection, ML techniques have the advantage of providing the results in fewer steps compared with the conventional, time-consuming approaches. However, clustering models show a good accuracy in distinguishing substituent concentrations only in cases where there is a substantial difference between the element regularly placed on a certain crystallographic site and the substituent.

4.5. Microstructural Characterization

Strain profiles from XRD data in irradiated or ion-implanted materials were determined by Boulle et al. [112] using a CNN model built on usual convolutional, max pooling, and batch normalization layers, together with dense neuron layers. While the accuracy of the results was above 90% for individual parameters, the accuracy for the complete strain profile (the key output of the modeling effort) ranged between 50% and 82%. The highest accuracies for the strain profile were achieved when training the CNN for separate strain ranges: 82% when considering only strains above 0.5%, and 76% for the lower strain region (in the 0.5–2% range) using a purposefully trained CNN. The lowest reported accuracy (50%) for the strain profile was applicable to strains below 2% when the CNN was trained using the complete strain range.

Residual stress in rails was also determined from X-ray measurements by applying a dimensionality reduction technique to X-ray data characteristic of normal rail regions, followed by multivariate statistical analysis based on the Gaussian distribution; finally, the presence of stress was identified using anomaly detection [113]. After performing dimensionality reduction on the original data using either PCA, kernel PCA, or an autoencoder, each datapoint was given an anomaly score corresponding to the local amount of damage. The accuracy of the model was assessed by comparing the location of the cracks in the rail with the anomalous zones, with appropriate thresholds for the anomaly scores, adjusted for each dimensionality reduction method. Out of the three dimensionality reduction algorithms, the autoencoder resulted in the highest accuracy in residual stress measurement. This is not a surprising result, given that autoencoders are based on neural networks that can represent nonlinear relationships using neuron activation functions, while PCA is a linear analysis technique.

ML methods such as neural networks and autoencoder excel at nonlinear relationships, capturing nuances conventional methods might miss. However, their success in microstructural characterization relies on high-quality training data and model interpretability remains a challenge. Despite good results for determining individual parameters, because of the limited availability of experimental data, the accuracy of the models employed for several parameters has unsatisfactory accuracy to date. Until further improvement of the databases or the models, conventional methods are preferred in complex situations.

4.6. Challenges and Limitations of Machine Learning

Overall, machine learning is an effective tool for XRD data analysis, but it also faces several challenges and limitations that must be addressed to make the technique more reliable and ubiquitous.

Data quality: ML models generally require large, diverse, and high-quality datasets for effective training. Obtaining enough labeled XRD data for specific materials or conditions can be time consuming and expensive. Additionally, the noise and artifacts present in experimental XRD data can decrease the performance of ML models. To mitigate data scarcity and diversity, data augmentation [93], transfer learning [114], or domain adaptation might be employed [115]. Additionally, preprocessing steps, such as noise filtering and background subtraction, help mitigate the impact of noise in the data.

Interpretability: ML algorithms can lack transparency in their decision-making process, making it difficult to interpret and explain the underlying reasons behind predictions

made by ML models, potentially hindering their use in scientific research. Efforts are being made to develop more transparent ML models through techniques such as feature visualization and saliency maps. Furthermore, hybrid approaches that combine ML with more traditional methods can leverage the advantages of ML while providing more interpretable results [116].

Generalizability: ML models may fail to generalize to new materials not seen by the model during training; accurate predictions across different crystal structures, lattice parameters and phase compositions require robust ML models. While generalization can be improved by the selection of appropriate features, data representation, and model architecture, the acquisition of datasets spanning the full range of possible variations is the best option for tackling generalization issues [103].

Model robustness refers to the ability of ML models to perform well under different conditions and input data quality. Overfitting negatively impacts robustness and occurs when ML models adjust to the training data instead of uncovering the meaningful patterns. To prevent overfitting and improve the model's robustness, researchers can employ regularization techniques to prevent extreme parameter values, cross validation to assess the model's performance on multiple subsets of the data, and early stopping of the model's training when the validation set's performance starts to degrade [117].

5. Conclusions and Future Development

The use of machine learning (ML) for processing X-ray diffraction (XRD) measurements has been increasing at an accelerating rate over the last decade, as computers have become more powerful, and both ML and XRD have been streamlined and enhanced. Based on current trends, it seems that ML will continue to be harnessed for XRD data analysis, and the technique will continue to be expanded and improved.

Future research is likely to focus on developing ML models that incorporate domain knowledge and physical constraints into the learning process. By integrating the fundamental principles of crystallography and materials science, the predictions of ML models can be more meaningful and in line with known physical property values. Alternatively, ML can be used in combination with physics-based models to obtain more interpretable, accurate, and physically meaningful predictions [110].

Quantum mechanical methods (such as density functional theory, DFT) combined with ML is a promising frontier in materials science, which can provide highly accurate and detailed insights into the properties and electronic structure of materials. Hybrid models integrating DTF calculations with ML can accelerate material discovery by increasing the efficiency and accuracy of material property predictions [111].

In operando and in situ studies, XRD patterns are acquired continuously during reaction or phase transformations. Real-time data analysis will facilitate and accelerate decision making and feedback, allowing researchers to monitor and control processes as they unfold, ultimately leading to deeper insights into material behavior under a wide range of conditions [107].

In our opinion, currently, machine learning is most helpful when used in conjunction with established mathematical models and domain knowledge. Hybrid approaches benefit both from the speed and flexibility that ML techniques can offer and from the rigors of physics informed analysis that guarantee the validity of the results up to the limitations of current scientific knowledge. XRD data analysis needs new, faster methods, but they should only be adopted if they can ensure the accuracy of their results.

Linking ML with combinatorial material analysis and high-throughput screening techniques can accelerate the discovery of novel materials with tailored properties: automated ML models can analyze vast libraries of XRD patterns to identify promising materials and suggest targeted experimental designs for further investigation, with huge potential for applications in fields like catalysis, energy materials, and drug development.

Author Contributions: Conceptualization, V.-A.S.; methodology, V.-A.S.; formal analysis, V.-A.S. and R.G.; visualization, V.-A.S., writing—original draft preparation, V.-A.S. and R.G.; writing—review and editing, V.-A.S. and R.G.; supervision, V.-A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Raj, C.; Agarwal, A.; Bharathy, G.; Narayan, B.; Prasad, M. Cyberbullying Detection: Hybrid Models Based on Machine Learning and Natural Language Processing Techniques. *Electronics* **2021**, *10*, 2810. [\[CrossRef\]](#)
2. Olthof, A.W.; Shouche, P.; Fennema, E.M.; Ijpma, F.F.A.; Koolstra, R.H.C.; Stirler, V.M.A.; van Ooijen, P.M.A.; Cornelissen, L.J. Machine Learning Based Natural Language Processing of Radiology Reports in Orthopaedic Trauma. *Comput. Methods Programs Biomed.* **2021**, *208*, 106304. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Bashir, M.F.; Arshad, H.; Javed, A.R.; Kryvinska, N.; Band, S.S. Subjective Answers Evaluation Using Machine Learning and Natural Language Processing. *IEEE Access* **2021**, *9*, 158972–158983. [\[CrossRef\]](#)
4. Mollaei, N.; Cepeda, C.; Rodrigues, J.; Gamboa, H. Biomedical Text Mining: Applicability of Machine Learning-Based Natural Language Processing in Medical Database. In Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies—Volume 4: BIOSTEC, Online, 9–11 February 2022; pp. 159–166. [\[CrossRef\]](#)
5. Houssein, E.H.; Mohamed, R.E.; Ali, A.A. Machine Learning Techniques for Biomedical Natural Language Processing: A Comprehensive Review. *IEEE Access* **2021**, *9*, 140628–140653. [\[CrossRef\]](#)
6. Zhang, Z.H. Image Recognition Methods Based on Deep Learning. In *3D Imaging—Multidimensional Signal Processing and Deep Learning, Volume 1*; Smart Innovation, Systems and Technologies Series; Springer: Singapore, 2022; Volume 297, pp. 23–34.
7. Wang, Y.S.; Hu, X. Machine Learning-Based Image Recognition for Rural Architectural Planning and Design. *Neural Comput. Appl.* **2022**, 1–10. [\[CrossRef\]](#)
8. Jabnoui, H.; Arfaoui, I.; Cherni, M.A.; Bouchouicha, M.; Sayadi, M. Machine Learning Based Classification for Fire and Smoke Images Recognition. In Proceedings of the 2022 8th International Conference on Control, Decision and Information Technologies (CODIT'22), Istanbul, Turkey, 17–20 May 2022; pp. 425–430. [\[CrossRef\]](#)
9. Shah, S.S.H.; Ahmad, A.; Jamil, N.; Khan, A.U.R. Memory Forensics-Based Malware Detection Using Computer Vision and Machine Learning. *Electronics* **2022**, *11*, 2579. [\[CrossRef\]](#)
10. Medeiros, E.C.; Almeida, L.M.; Teixeira, J.G.D. Computer Vision and Machine Learning for Tuna and Salmon Meat Classification. *Informatics* **2021**, *8*, 70. [\[CrossRef\]](#)
11. Yin, H.; Yi, W.L.; Hu, D.M. Computer Vision and Machine Learning Applied in the Mushroom Industry: A Critical Review. *Comput. Electron. Agric.* **2022**, *198*, 107015. [\[CrossRef\]](#)
12. Shah, N.; Bhagat, N.; Shah, M. Crime Forecasting: A Machine Learning and Computer Vision Approach to Crime Prediction and Prevention. *Vis. Comput. Ind. Biomed. Art* **2021**, *4*, 9. [\[CrossRef\]](#)
13. Mahadevkar, S.V.; Khemani, B.; Patil, S.; Kotecha, K.; Vora, D.R.; Abraham, A.; Gabralla, L.A. A Review on Machine Learning Styles in Computer Vision-Techniques and Future Directions. *IEEE Access* **2022**, *10*, 107293–107329. [\[CrossRef\]](#)
14. Khan, A.A.; Laghari, A.A.; Awan, S.A. Machine Learning in Computer Vision: A Review. *EAI Endorsed Trans. Scalable Inf. Syst.* **2021**, *8*, e4. [\[CrossRef\]](#)
15. Mun, C.H.; Rezvani, S.; Lee, J.; Park, S.S.; Park, H.W.; Lee, J. Indirect Measurement of Cutting Forces during Robotic Milling Using Multiple Sensors and a Machine Learning-Based System Identifier. *J. Manuf. Processes* **2023**, *85*, 963–976. [\[CrossRef\]](#)
16. Kim, N.; Barde, S.; Bae, K.; Shin, H. Learning Per-Machine Linear Dispatching Rule for Heterogeneous Multi-Machines Control. *Int. J. Prod. Res.* **2023**, *61*, 162–182. [\[CrossRef\]](#)
17. Piat, J.R.; Dafflon, B.; Bentaha, M.L.; Gerphagnon, Y.; Moalla, N. A Framework to Optimize Laser Welding Process by Machine Learning in a SME Environment. In *Product Lifecycle Management. PLM in Transition Times: The Place of Humans and Transformative Technologies, PLM 2022*; Springer: Cham, Switzerland, 2023; Volume 667, pp. 431–439.
18. Carpanzano, E.; Knüttel, D. Advances in Artificial Intelligence Methods Applications in Industrial Control Systems: Towards Cognitive Self-Optimizing Manufacturing Systems. *Appl. Sci.* **2022**, *12*, 10962. [\[CrossRef\]](#)
19. Hashemnia, N.; Fan, Y.Y.; Rocha, N. Using Machine Learning to Predict and Avoid Malfunctions: A Revolutionary Concept for Condition-Based Asset Performance Management (APM). In Proceedings of the 2021 IEEE PES Innovative Smart Grid Technologies—ASIA (ISGT ASIA), Brisbane, Australia, 5–8 December 2021.
20. Xu, D.; Chen, L.Q.; Yu, C.; Zhang, S.; Zhao, X.; Lai, X. Failure Analysis and Control of Natural Gas Pipelines under Excavation Impact Based on Machine Learning Scheme. *Int. J. Press. Vessels Pip.* **2023**, *201*, 104870. [\[CrossRef\]](#)

21. Shcherbatov, I.; Lisin, E.; Rogalev, A.; Tsurikov, G.; Dvorak, M.; Strielkowski, W. Power Equipment Defects Prediction Based on the Joint Solution of Classification and Regression Problems Using Machine Learning Methods. *Electronics* **2021**, *10*, 3145. [\[CrossRef\]](#)
22. Nuhu, A.A.; Zeeshan, Q.; Safaei, B.; Shahzad, M.A. Machine Learning-Based Techniques for Fault Diagnosis in the Semiconductor Manufacturing Process: A Comparative Study. *J. Supercomput.* **2023**, *79*, 2031–2081. [\[CrossRef\]](#)
23. Ko, H.; Lu, Y.; Yang, Z.; Ndiaye, N.Y.; Witherell, P. A Framework Driven by Physics-Guided Machine Learning for Process-Structure-Property Causal Analytics in Additive Manufacturing. *J. Manuf. Syst.* **2023**, *67*, 213–228. [\[CrossRef\]](#)
24. Dogan, A.; Birant, D. Machine Learning and Data Mining in Manufacturing. *Expert Syst. Appl.* **2021**, 166. [\[CrossRef\]](#)
25. Acosta, S.M.; Oliveira, R.M.A.; Sant’Anna, A.M.O. Machine Learning Algorithms Applied to Intelligent Tyre Manufacturing. *Int. J. Comput. Integr. Manuf.* **2023**, 1–11. [\[CrossRef\]](#)
26. Gao, C.C.; Min, X.; Fang, M.H.; Tao, T.Y.; Zheng, X.H.; Liu, Y.G.; Wu, X.W.; Huang, Z.H. Innovative Materials Science via Machine Learning. *Adv. Funct. Mater.* **2022**, *32*, 2108044. [\[CrossRef\]](#)
27. Peterson, G.G.C.; Brgoch, J. Materials Discovery through Machine Learning Formation Energy. *J. Phys.-Energy* **2021**, *3*, 022002. [\[CrossRef\]](#)
28. Fuhr, A.S.; Sumpter, B.G. Deep Generative Models for Materials Discovery and Machine Learning-Accelerated Innovation. *Front. Mater.* **2022**, *9*, 865270. [\[CrossRef\]](#)
29. Fang, J.H.; Xie, M.; He, X.Q.; Zhang, J.M.; Hu, J.Q.; Chen, Y.T.; Yang, Y.C.; Jin, Q.L. Machine Learning Accelerates the Materials Discovery. *Mater Today Commun.* **2022**, *33*, 104900. [\[CrossRef\]](#)
30. Juan, Y.F.; Dai, Y.B.; Yang, Y.; Zhang, J. Accelerating Materials Discovery Using Machine Learning. *J. Mater. Sci. Technol.* **2021**, *79*, 178–190. [\[CrossRef\]](#)
31. Hou, H.B.; Wang, J.F.; Ye, L.; Zhu, S.J.; Wang, L.G.; Guan, S.K. Prediction of Mechanical Properties of Biomedical Magnesium Alloys Based on Ensemble Machine Learning. *Mater. Lett.* **2023**, *348*, 134605. [\[CrossRef\]](#)
32. Magar, R.; Farimani, A.B. Learning from Mistakes: Sampling Strategies to Efficiently Train Machine Learning Models for Material Property Prediction. *Comput. Mater. Sci.* **2023**, *224*, 112167. [\[CrossRef\]](#)
33. Rong, C.; Zhou, L.; Zhang, B.W.; Xuan, F.Z. Machine Learning for Mechanics Prediction of 2D MXene-Based Aerogels. *Compos. Commun.* **2023**, *38*, 101474. [\[CrossRef\]](#)
34. Chan, C.H.; Sun, M.Z.; Huang, B.L. Application of Machine Learning for Advanced Material Prediction and Design. *EcoMat* **2022**, *4*, e12194. [\[CrossRef\]](#)
35. Sendek, A.D.; Ransom, B.; Cubuk, E.D.; Pellouchoud, L.A.; Nanda, J.; Reed, E.J. Machine Learning Modeling for Accelerated Battery Materials Design in the Small Data Regime. *Adv. Energy Mater.* **2022**, *12*, 2200553. [\[CrossRef\]](#)
36. Pei, Z.R.; Rozman, K.A.; Dogan, O.N.; Wen, Y.H.; Gao, N.; Holm, E.A.; Hawk, J.A.; Alman, D.E.; Gao, M.C. Machine-Learning Microstructure for Inverse Material Design. *Adv. Sci.* **2021**, *8*, 2101207. [\[CrossRef\]](#) [\[PubMed\]](#)
37. He, J.J.; Li, J.J.; Liu, C.B.; Wang, C.X.; Zhang, Y.; Wen, C.; Xue, D.Z.; Cao, J.L.; Su, Y.J.; Qiao, L.J.; et al. Machine Learning Identified Materials Descriptors for Ferroelectricity. *Acta Mater.* **2021**, *209*, 116815. [\[CrossRef\]](#)
38. McSweeney, D.M.; McSweeney, S.M.; Liu, Q. A Self-Supervised Workflow for Particle Picking in Cryo-EM. *IUCr* **2020**, *7*, 719–727. [\[CrossRef\]](#)
39. Ramakrishnan, R.; Dral, P.O.; Rupp, M.; Von Lilienfeld, O.A. Quantum Chemistry Structures and Properties of 134 Kilo Molecules. *Sci. Data* **2014**, *1*, 140022. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Xie, T.; Grossman, J.C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018**, *120*, 145301. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Luo, R.; Popp, J.; Bocklitz, T. Deep Learning for Raman Spectroscopy: A Review. *Analytica* **2022**, *3*, 287–301. [\[CrossRef\]](#)
42. Gadre, C.A.; Yan, X.; Song, Q.; Li, J.; Gu, L.; Huyan, H.; Aoki, T.; Lee, S.W.; Chen, G.; Wu, R.; et al. Nanoscale Imaging of Phonon Dynamics by Electron Microscopy. *Nature* **2022**, *606*, 292–297. [\[CrossRef\]](#)
43. Friedrich, W.; Knipping, P.; Laue, M. Interferenzerscheinungen Bei Röntgenstrahlen. *Ann. Phys.* **1913**, *346*, 971–988. [\[CrossRef\]](#)
44. Authier, A. *Early Days of X-ray Crystallography*; Oxford University Press: Oxford, UK, 2013.
45. Singh, A.K. *Advanced X-ray Techniques in Research and Industry*; IOS Press: Amsterdam, The Netherlands, 2005; ISBN 1586035371.
46. Bragg, W.L.; Thomson, J.J. The Diffraction of Short Electromagnetic Waves by a Crystal. *Proc. Camb. Philos. Soc. Math. Phys. Sci.* **1914**, *17*, 43–57.
47. Withers, P.J. Synchrotron X-ray Diffraction. In *Practical Residual Stress Measurement Methods*; Wiley: Hoboken, NJ, USA, 2013; pp. 163–194. ISBN 9781118402832.
48. Li, Y.; Beck, R.; Huang, T.; Choi, M.C.; Divinagracia, M. Scatterless Hybrid Metal-Single-Crystal Slit for Small-Angle X-ray Scattering and High-Resolution X-ray Diffraction. *J. Appl. Crystallogr.* **2008**, *41*, 1134–1139. [\[CrossRef\]](#)
49. Wohlschlögel, M.; Schüllli, T.U.; Lantz, B.; Welzel, U. Application of a Single-Reflection Collimating Multilayer Optic for X-ray Diffraction Experiments Employing Parallel-Beam Geometry. *J. Appl. Crystallogr.* **2008**, *41*, 124–133. [\[CrossRef\]](#)
50. Saha, G.B. Scintillation and Semiconductor Detectors. In *Physics and Radiobiology of Nuclear Medicine*; Saha, G.B., Ed.; Springer: New York, NY, USA, 2006; pp. 81–107. ISBN 978-0-387-36281-6.
51. Maniammal, K.; Madhu, G.; Biju, V. X-Ray Diffraction Line Profile Analysis of Nanostructured Nickel Oxide: Shape Factor and Convolution of Crystallite Size and Microstrain Contributions. *Phys. E Low Dimens. Syst. Nanostruct.* **2017**, *85*, 214–222. [\[CrossRef\]](#)

52. Uvarov, V.; Popov, I. Metrological Characterization of X-Ray Diffraction Methods for Determination of Crystallite Size in Nano-Scale Materials. *Mater. Charact.* **2007**, *58*, 883–891. [\[CrossRef\]](#)
53. Epp, J. 4—X-Ray Diffraction (XRD) Techniques for Materials Characterization. In *Materials Characterization Using Nondestructive Evaluation (NDE) Methods*; Hübschen, G., Altpeter, I., Tschuncky, R., Herrmann, H.-G., Eds.; Woodhead Publishing: Sawston, UK, 2016; pp. 81–124, ISBN 978-0-08-100040-3.
54. Chipera, S.J.; Bish, D.L. Fitting Full X-ray Diffraction Patterns for Quantitative Analysis: A Method for Readily Quantifying Crystalline and Disordered Phases. *Adv. Mater. Phys. Chem.* **2013**, *3*, 47–53. [\[CrossRef\]](#)
55. Sitepu, H.; O'Connor, B.H.; Li, D. Comparative Evaluation of the March and Generalized Spherical Harmonic Preferred Orientation Models Using X-ray Diffraction Data for Molybdate and Calcite Powders. *J. Appl. Crystallogr.* **2005**, *38*, 158–167. [\[CrossRef\]](#)
56. Jenkins, R.; Snyder, R.L. *Introduction to X-ray Powder Diffractometry*; Wiley: New York, NY, USA, 1996; Volume 138.
57. Reventos, M.M.; Descarrega, J.M.A. Mineralogy and Geology: The Role of Crystallography since the Discovery of X-ray Diffraction in 1912. *Mineralogía y Geología: El papel de la Cristalografía desde el descubrimiento de la difracción de Rayos X en 1912. Rev. Soc. Geol. España* **2012**, *25*, 133–143.
58. Okoro, C.; Levine, L.E.; Xu, R.; Hummler, K.; Obeng, Y.S. Nondestructive Measurement of the Residual Stresses in Copper Through-Silicon Vias Using Synchrotron-Based Microbeam X-Ray Diffraction. *IEEE Trans. Electron. Devices* **2014**, *61*, 2473–2479. [\[CrossRef\]](#)
59. Bunaciu, A.A.; Udriștioiu, E.G.; Aboul-Enein, H.Y. X-ray Diffraction: Instrumentation and Applications. *Crit. Rev. Anal. Chem.* **2015**, *45*, 289–299. [\[CrossRef\]](#)
60. Kotrly, M. Application of X-ray Diffraction in Forensic Science. *Z. Kristallogr. Suppl.* **2006**, *23*, 35–40. [\[CrossRef\]](#)
61. Warren, B.E.; Biscoe, J. Fourier Analysis of X-ray Patterns of Soda-Silica Glass. *J. Am. Ceram. Soc.* **1938**, *21*, 259–265. [\[CrossRef\]](#)
62. Misture, S.T. X-ray Powder Diffraction. In *Encyclopedia of Materials: Technical Ceramics and Glasses*; Pomeroy, M., Ed.; Elsevier: Oxford, UK, 2021; pp. 549–559. ISBN 978-0-12-822233-1.
63. Zok, F.W. Integrating Lattice Materials Science into the Traditional Processing–Structure–Properties Paradigm. *MRS Commun.* **2019**, *9*, 1284–1291. [\[CrossRef\]](#)
64. Scarlett, N.V.Y.; Madsen, I.C.; Manias, C.; Retallack, D. On-Line X-ray Diffraction for Quantitative Phase Analysis: Application in the Portland Cement Industry. *Powder Diffr.* **2001**, *16*, 71–80. [\[CrossRef\]](#)
65. Conconi, M.S.; Gauna, M.R.; Serra, M.F.; Suarez, G.; Aglietti, E.F.; Rendtorff, N.M.; Gonnet, M.B.; Aires, B.; Aires, B. Quantitative Firing Transformations of Triaxial Ceramic by X-Ray Diffraction Methods. *Ceramica* **2014**, *60*, 524–531. [\[CrossRef\]](#)
66. Cheary, R.W.; Ma-Sorrell, Y. Quantitative Phase Analysis by X-ray Diffraction of Martensite and Austenite in Strongly Oriented Orthodontic Stainless Steel Wires. *J. Mater. Sci.* **2000**, *35*, 1105–1113. [\[CrossRef\]](#)
67. Wilkinson, A.P.; Speck, J.S.; Cheetham, A.K.; Natarajan, S.; Thomas, J.M. In Situ X-ray Diffraction Study of Crystallization Kinetics in PbZr_{1-x}Ti_xO₃ (PZT, x = 0.0, 0.55, 1.0). *Chem. Mater.* **1994**, *6*, 750–754. [\[CrossRef\]](#)
68. Purushottam raj purohit, R.R.P.; Arya, A.; Bojjawar, G.; Pelerin, M.; Van Petegem, S.; Proudhon, H.; Mukherjee, S.; Gerard, C.; Signor, L.; Mocuta, C.; et al. Revealing the Role of Microstructure Architecture on Strength and Ductility of Ni Microwires by In-Situ Synchrotron X-ray Diffraction. *Sci. Rep.* **2019**, *9*, 79. [\[CrossRef\]](#)
69. Prasetya, A.D.; Rifai, M.; Mujamilah; Miyamoto, H. X-ray Diffraction (XRD) Profile Analysis of Pure ECAP-Annealing Nickel Samples. *J. Phys. Conf. Ser.* **2020**, *1436*, 012113. [\[CrossRef\]](#)
70. Wang, C.; Steiner, U.; Sepe, A. Synchrotron Big Data Science. *Small* **2018**, *14*, 1802291. [\[CrossRef\]](#) [\[PubMed\]](#)
71. Suzuki, Y. Automated Data Analysis for Powder X-ray Diffraction Using Machine Learning. *Synchrotron. Radiat. News* **2022**, *35*, 9–15. [\[CrossRef\]](#)
72. Laalam, A.; Boualam, A.; Ouadi, H.; Djezzar, S.; Tomomewo, O.; Mellal, I.; Bakelli, O.; Merzoug, A.; Chemmakh, A.; Latreche, A.; et al. Application of Machine Learning for Mineralogy Prediction from Well Logs in the Bakken Petroleum System. In Proceedings of the SPE Annual Technical Conference and Exhibition, Houston, TX, USA, 3–5 October 2022; p. D012S063R002. [\[CrossRef\]](#)
73. Zhao, B.; Greenberg, J.A.; Wolter, S. Application of Machine Learning to X-ray Diffraction-Based Classification. In *Anomaly Detection and Imaging with X-Rays (ADIX) III*; SPIE: Bellingham, WA, USA, 2018; p. 1063205. [\[CrossRef\]](#)
74. Hillier, S. Accurate Quantitative Analysis of Clay and Other Minerals in Sandstones by XRD: Comparison of a Rietveld and a Reference Intensity Ratio (RIR) Method and the Importance of Sample Preparation. *Clay Miner.* **2000**, *35*, 291–302. [\[CrossRef\]](#)
75. Lee, D.; Lee, H.; Jun, C.-H.; Chang, C.H. A Variable Selection Procedure for X-Ray Diffraction Phase Analysis. *Appl. Spectrosc.* **2007**, *61*, 1398–1403. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Greasley, J.; Hosein, P. Exploring Supervised Machine Learning for Multi-Phase Identification and Quantification from Powder X-ray Diffraction Spectra. *J. Mater. Sci.* **2023**, *58*, 5334–5348. [\[CrossRef\]](#)
77. Visser, J.W.; Sonneveld, E.J. Automatic Collection of Powder Data from Photographs. *J. Appl. Crystallogr.* **1975**, *8*, 1–7.
78. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [\[CrossRef\]](#)
79. Hanawalt, J.D. Phase Identification by X-ray Powder Diffraction Evaluation of Various Techniques. *Adv. X-ray Anal.* **1976**, *20*, 63–73. [\[CrossRef\]](#)
80. Scherrer, P. Estimation of the Size and Internal Structure of Colloidal Particles by Means of Röntgen. *Nachr. Ges. Wiss. Göttingen* **1918**, *2*, 96–100.

81. Williamson, G.K.; Hall, W.H. X-Ray Line Broadening from Filled Aluminium and Wolfram. *Acta Metall.* **1953**, *1*, 22–31. [\[CrossRef\]](#)
82. Bourniquel, B.; Sprauel, J.M.; Feron, J.; Lebrun, J.L. Warren-Averbach Analysis of X-Ray Line Profile (Even Truncated) Assuming a Voigt-like Profile. In *International Conference on Residual Stresses: ICRS2*; Beck, G., Denis, S., Simon, A., Eds.; Springer: Dordrecht, The Netherlands, 1989; pp. 184–189. ISBN 978-94-009-1143-7.
83. Dollase, W.A. Correction of Intensities of Preferred Orientation in Powder Diffractometry: Application of the March Model. *J. Appl. Crystallogr.* **1986**, *19*, 267–272. [\[CrossRef\]](#)
84. Alzubi, J.; Nayyar, A.; Kumar, A. Machine Learning from Theory to Algorithms: An Overview. *J. Phys. Conf. Ser.* **2018**, *1142*, 012012. [\[CrossRef\]](#)
85. Pane, S.A.; Sihombing, F.M.H. Classification of Rock Mineral in Field X Based on Spectral Data (SWIR & TIR) Using Supervised Machine Learning Methods. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *830*, 012042. [\[CrossRef\]](#)
86. Colliot, O. (Ed.). *Machine Learning for Brain Disorders*; Neuromethods; Springer: New York, NY, USA, 2023; Volume 197, ISBN 978-1-0716-3194-2.
87. Ige, A.O.; Mohd Noor, M.H. A Survey on Unsupervised Learning for Wearable Sensor-Based Activity Recognition. *Appl. Soft Comput.* **2022**, *127*, 109363. [\[CrossRef\]](#)
88. Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2019; ISBN 9781492032649.
89. Schmidt, R.M. Recurrent Neural Networks (RNNs): A Gentle Introduction and Overview. *arXiv* **2019**. [\[CrossRef\]](#)
90. Oh, S.; Ashiquzzaman, A.; Lee, D.; Kim, Y.; Kim, J. Study on Human Activity Recognition Using Semi-Supervised Active Transfer Learning. *Sensors* **2021**, *21*, 2760. [\[CrossRef\]](#) [\[PubMed\]](#)
91. Wang, B.; Guan, Z.; Yao, S.; Qin, H.; Nguyen, M.H.; Yager, K.; Yu, D. Deep Learning for Analysing Synchrotron Data Streams. In Proceedings of the 2016 New York Scientific Data Summit (NYSDS), New York, NY, USA, 14–17 August 2016. [\[CrossRef\]](#)
92. Czyzewski, A.; Krawiec, F.; Brzezinski, D.; Porebski, P.J.; Minor, W. Detecting Anomalies in X-Ray Diffraction Images Using Convolutional Neural Networks. *Expert Syst. Appl.* **2021**, *174*, 114740. [\[CrossRef\]](#) [\[PubMed\]](#)
93. Chakraborty, A.; Sharma, R. See Deeper: Identifying Crystal Structure from X-Ray Diffraction Patterns. In Proceedings of the 2020 International Conference on Cyberworlds (CW), Caen, France, 29 September–1 October 2020; pp. 49–54. [\[CrossRef\]](#)
94. Chakraborty, A.; Sharma, R. A Deep Crystal Structure Identification System for X-Ray Diffraction Patterns. *Vis. Comput.* **2022**, *38*, 1275–1282. [\[CrossRef\]](#)
95. Massuyeau, F.; Broux, T.; Coulet, F.; Demessence, A.; Mesbah, A.; Gautier, R. Perovskite or Not Perovskite? A Deep-Learning Approach to Automatically Identify New Hybrid Perovskites from X-ray Diffraction Patterns. *Adv. Mater.* **2022**, *34*, 2203879. [\[CrossRef\]](#)
96. Ishitsuka, K.; Ojima, H.; Mogi, T.; Kajiwar, T.; Sugimoto, T.; Asanuma, H. Characterization of Hydrothermal Alteration along Geothermal Wells Using Unsupervised Machine-Learning Analysis of X-ray Powder Diffraction Data. *Earth Sci. Inform.* **2022**, *15*, 73–87. [\[CrossRef\]](#)
97. Yuan, S.; Wolter, S.D.; Greenberg, J.A. Classification-Free Threat Detection Based on Material-Science-Informed Clustering. In *Anomaly Detection and Imaging with X-rays (ADIX) II*; SPIE: Bellingham, WA, USA, 2017; Volume 10187, p. 101870K. [\[CrossRef\]](#)
98. Lee, J.W.; Park, W.B.; Kim, M.; Pal Singh, S.; Pyo, M.; Sohn, K.S. A Data-Driven XRD Analysis Protocol for Phase Identification and Phase-Fraction Prediction of Multiphase Inorganic Compounds. *Inorg. Chem. Front.* **2021**, *8*, 2492–2504. [\[CrossRef\]](#)
99. Park, S.Y.; Son, B.K.; Choi, J.; Jin, H.; Lee, K. Application of Machine Learning to Quantification of Mineral Composition on Gas Hydrate-Bearing Sediments, Ulleung Basin, Korea. *J. Pet. Sci. Eng.* **2022**, *209*, 109840. [\[CrossRef\]](#)
100. Pasha, M.F.; Rahmat, R.F.; Budiarto, R.; Syukur, M. A Distributed Autonomous Neuro-Gen Learning Engine and Its Application to the Lattice Analysis of Cubic Structure Identification Problem. *Int. J. Innov. Comput. Inf. Control* **2010**, *6*, 1005–1022.
101. Vecsei, P.M.; Choo, K.; Chang, J.; Neupert, T. Neural Network Based Classification of Crystal Symmetries from X-Ray Diffraction Patterns. *Phys. Rev. B* **2019**, *99*, 245120. [\[CrossRef\]](#)
102. Suzuki, Y.; Hino, H.; Hawai, T.; Saito, K.; Kotsugi, M.; Ono, K. Symmetry Prediction and Knowledge Discovery from X-Ray Diffraction Patterns Using an Interpretable Machine Learning Approach. *Sci. Rep.* **2020**, *10*, 21790. [\[CrossRef\]](#) [\[PubMed\]](#)
103. Oviedo, F.; Ren, Z.; Sun, S.; Settens, C.; Liu, Z.; Hartono, N.T.P.; Ramasamy, S.; DeCost, B.L.; Tian, S.I.P.; Romano, G.; et al. Fast and Interpretable Classification of Small X-Ray Diffraction Datasets Using Data Augmentation and Deep Neural Networks. *NPJ Comput. Mater.* **2019**, *5*, 60. [\[CrossRef\]](#)
104. Venderley, J.; Mallayya, K.; Matty, M.; Krogstad, M.; Ruff, J.; Pleiss, G.; Kishore, V.; Mandrus, D.; Phelan, D.; Poudel, L.; et al. Harnessing Interpretable and Unsupervised Machine Learning to Address Big Data from Modern X-Ray Diffraction. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2109665119. [\[CrossRef\]](#)
105. Samarakoon, A.M.; Alan Tennant, D. Machine Learning for Magnetic Phase Diagrams and Inverse Scattering Problems. *J. Phys. Condens. Matter* **2022**, *34*, 044002. [\[CrossRef\]](#)
106. Kautzsch, L.; Ortiz, B.R.; Mallayya, K.; Plumb, J.; Pokharel, G.; Ruff, J.P.C.; Islam, Z.; Kim, E.A.; Seshadri, R.; Wilson, S.D. Structural Evolution of the Kagome Superconductors A V3Sb5 (A = K, Rb, and Cs) through Charge Density Wave Order. *Phys. Rev. Mater.* **2023**, *7*, 024806. [\[CrossRef\]](#)
107. Song, Y.; Tamura, N.; Zhang, C.; Karami, M.; Chen, X. Data-Driven Approach for Synchrotron X-Ray Laue Microdiffraction Scan Analysis. *Acta Crystallogr. A Found. Adv.* **2019**, *75*, 876–888. [\[CrossRef\]](#)

108. Al Hasan, N.M.; Hou, H.; Gao, T.; Counsell, J.; Sarker, S.; Thienhaus, S.; Walton, E.; Decker, P.; Mehta, A.; Ludwig, A.; et al. Combinatorial Exploration and Mapping of Phase Transformation in a Ni-Ti-Co Thin Film Library. *ACS Comb. Sci.* **2020**, *22*, 641–648. [\[CrossRef\]](#)
109. Narayanachari, K.V.L.V.; Bruce Buchholz, D.; Goldfine, E.A.; Wenderott, J.K.; Haile, S.M.; Bedzyk, M.J. Combinatorial Approach for Single-Crystalline Taon Growth: Epitaxial β -Taon (100)/ α -Al₂O₃ (012). *ACS Appl. Electron. Mater.* **2020**, *2*, 3571–3576. [\[CrossRef\]](#)
110. Utimula, K.; Hunkao, R.; Yano, M.; Kimoto, H.; Hongo, K.; Kawaguchi, S.; Suwanna, S.; Maezono, R. Machine-Learning Clustering Technique Applied to Powder X-Ray Diffraction Patterns to Distinguish Compositions of ThMn12-Type Alloys. *Adv. Theory Simul.* **2020**, *3*, 2000039. [\[CrossRef\]](#)
111. Utimula, K.; Yano, M.; Kimoto, H.; Hongo, K.; Nakano, K.; Maezono, R. Feature Space of XRD Patterns Constructed by an Autoencoder. *Adv. Theory Simul.* **2023**, *6*, 2200613. [\[CrossRef\]](#)
112. Boulle, A.; Debelle, A. Convolutional Neural Network Analysis of X-Ray Diffraction Data: Strain Profile Retrieval in Ion Beam Modified Materials. *Mach. Learn. Sci. Technol.* **2023**, *4*, 015002. [\[CrossRef\]](#)
113. Mitsui, S.; Sasaki, T.; Shinya, M.; Arai, Y.; Nishimura, R. Anomaly Detection in Rails Using Dimensionality Reduction. *ISIJ Int.* **2023**, *63*, 170–178. [\[CrossRef\]](#)
114. Wu, L.; Yoo, S.; Suzana, A.F.; Assefa, T.A.; Diao, J.; Harder, R.J.; Cha, W.; Robinson, I.K. Three-Dimensional Coherent X-Ray Diffraction Imaging via Deep Convolutional Neural Networks. *NPJ Comput. Mater.* **2021**, *7*, 175. [\[CrossRef\]](#)
115. Chang, M.-C.; Tung, C.-H.; Chang, S.-Y.; Carrillo, J.M.; Wang, Y.; Sumpter, B.G.; Huang, G.-R.; Do, C.; Chen, W.-R. A Machine Learning Inversion Scheme for Determining Interaction from Scattering. *Commun. Phys.* **2022**, *5*, 46. [\[CrossRef\]](#)
116. Kløve, M.; Sommer, S.; Iversen, B.B.; Hammer, B.; Dononelli, W. A Machine-Learning-Based Approach for Solving Atomic Structures of Nanomaterials Combining Pair Distribution Functions with Density Functional Theory. *Adv. Mater.* **2023**, *35*, 2208220. [\[CrossRef\]](#)
117. Lee, B.D.; Lee, J.-W.; Park, W.B.; Park, J.; Cho, M.-Y.; Pal Singh, S.; Pyo, M.; Sohn, K.-S. Powder X-Ray Diffraction Pattern Is All You Need for Machine-Learning-Based Symmetry Identification and Property Prediction. *Adv. Intell. Syst.* **2022**, *4*, 2200042. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.