Machine learning for scattering data: strategies, perspectives and applications to surface scattering

Alexander Hinderhofer,* Alessandro Greco, Vladimir Starostin, Valentin Munteanu, Linus Pithan, Alexander Gerlach and Frank Schreiber

Institute of Applied Physics, University of Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany. *Correspondence e-mail: alexander.hinderhofer@uni-tuebingen.de

Machine learning (ML) has received enormous attention in science and beyond. Discussed here are the status, opportunities, challenges and limitations of ML as applied to X-ray and neutron scattering techniques, with an emphasis on surface scattering. Typical strategies are outlined, as well as possible pitfalls. Applications to reflectometry and grazing-incidence scattering are critically discussed. Comment is also given on the availability of training and test data for ML applications, such as neural networks, and a large reflectivity data set is provided as reference data for the community.

1. Introduction

Machine learning (ML) is receiving enormous attention in essentially all areas of our lives, including in the physical sciences (Erdmann *et al.*, 2021). The application of ML strategies for the analysis of scattering data is particularly attractive (Chen *et al.*, 2021). Here we discuss the status, opportunities, challenges and limitations of ML applied to X-ray and neutron scattering techniques, with specific focus on surface scattering (Feidenhans'l, 1989; Holý *et al.*, 1999; Birkholz, 2006; Als-Nielsen & McMorrow, 2011), which is intended to include interface scattering as well, *i.e.* interfaces between two condensed phases.

One motivation for applying ML in the context of scattering data is simply the hope for faster and more efficient data analysis compared with standard methods. The general theoretical framework for modelling and simulating scattering data is well established. This allows for a simple generation of training data, which is a huge advantage compared with other fields where no direct data generation mechanism exists (*e.g.* computer vision) or where the simulations are very computationally expensive. At the same time, although based essentially on a simple Fourier transform, in addition to optical effects, the conversion of scattering data back into direct information is not straightforward. The acceleration of conventional fitting strategies, which are generally time consuming, with ML methods is also possible if sufficient annotated experimental data are available.

Another motivation is derived from the need to handle huge data volumes and data acquisition rates, which is an almost universal trend in many areas of science. In the scattering world this is due in particular to ever-improving sources with higher flux and to greatly improved detector technology, with area detectors of high resolution and high dynamic range. Real-time experiments (Wang *et al.*, 2021) and high-throughput experiments (Ludwig, 2019; Bai *et al.*, 2018) constitute a particular challenge. In these and many other
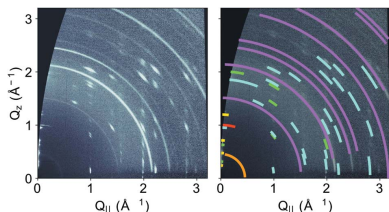
experiments the rate of data production can be overwhelming and simply impossible to handle for traditional screening by humans, triggering a demand for pre-screening and filtering. A suitable ML algorithm to filter and categorize or even analyse the data before a human researcher inspects the data can be extremely valuable.

There are, of course, many data analysis strategies for different applications in physics. Here, we highlight specific applications of ML techniques, but without a detailed technical discussion of the algorithms, for which we refer the reader to the work of Erdmann *et al.* (2021). Before discussing ML strategies specifically applied to surface scattering, we mention some other efforts that apply ML to scattering methods.

For example, work on bulk crystallography started many years ago and has showed impressive progress (Tatlier, 2011; Oviedo *et al.*, 2019; Lee *et al.*, 2020; Bai *et al.*, 2018). Other standard scattering methods, such as small-angle scattering, have also received considerable attention, especially for classification tasks (Song *et al.*, 2020; Ikemoto *et al.*, 2020; Franke *et al.*, 2018; Archibald *et al.*, 2020; Chang *et al.*, 2020). For non-standard coherent scattering methods, such as X-ray photon correlation spectroscopy (XPCS), ML-based analysis using autoencoders has been employed (Konstantinova *et al.*, 2021; Timmermann *et al.*, 2022). For a more general review of ML methods for scattering we refer to a recent review, which also discusses applications in the broader context of scattering experiments, such as spectroscopy methods, theoretical calculations, automated alignment procedures, beam optimization and data filtering (Chen *et al.*, 2021).

The goal of the present paper is to discuss the perspectives of ML applied to the analysis of scattering data in general and surface scattering data in particular. We first explain the key characteristics and challenges for scattering from surfaces. We then discuss three main surface scattering methods [reflectometry, grazing-incidence small-angle scattering (GISAS) and grazing-incidence wide-angle scattering (GIWAS)], each in their own subsection, with specific regard to ML-based data analysis and their specific scattering geometries (Fig. 1). In doing so, we implicitly cover both X-rays and neutrons, but we also comment on the specifics of the two different probes (Section 6). This is followed by a critical discussion of the main challenges, as well as the possible role of a reference database and perspectives for establishing it, for which we offer a starting point (Pithan *et al.*, 2022).

2. Strategies and challenges

ML can be applied to solving many different tasks in the context of surface scattering, each with their own specific challenges:
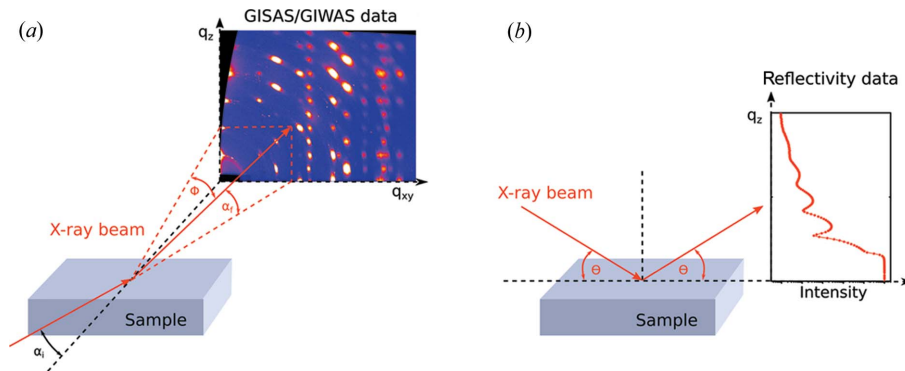

Figure 1
Schematics of typical surface scattering geometries. (*a*) Grazing-incidence small-/wide-angle scattering (GISAS/GIWAS). Data are typically a 2D data set with out-of-plane (q_z) and in-plane (q_{xy}) information. (*b*) Reflectometry recorded under the specular condition, *i.e.* the angle of incidence is always equal to the detector angle. X-ray and neutron reflectometry are 1D data sets.

(i) Classification. The ML algorithm can sort the data of an input data set into categories, such as particle shapes in GISAS data, and the output would be a class for each data set.

(ii) Object detection. The ML procedure can find objects in a data set, for example Bragg reflections in grazing-incidence wide-angle X-ray scattering data, and output object coordinates.

(iii) Parameter extraction. The ML algorithm replaces the conventional fitting process and extracts numerical parameters directly from the data. For example, layer thickness and roughness from reflectometry data could be the output of such an approach (Fig. 2).

(iv) Data processing. For this approach scattering data are typically processed to improve the conventional fitting procedures. For example, the denoising of neutron reflectivity data or XPCS data by an autoencoder has already been demonstrated (Konstantinova *et al.*, 2021; Timmermann *et al.*, 2022).

There are several challenges when trying to apply ML to the analysis of scattering data. The most important one is arguably the well known phase problem (Sivia *et al.*, 1991), which can lead to ambiguous solutions that require additional knowledge for the data to be interpreted correctly.

Furthermore, experimental limitations can reduce the information content in the data. Each setup has specific properties and error sources that need to be taken into account. Differences in the size, shape and divergence of the
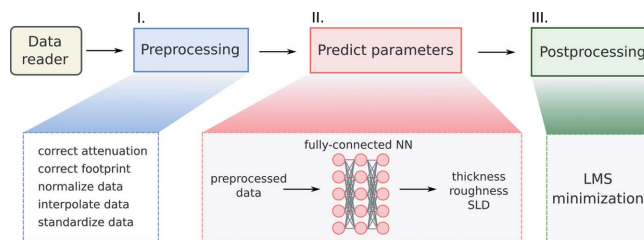

Figure 2
The neural network architecture for extracting thin-film parameters from reflectivity data as demonstrated by Greco *et al.* (2022).

beam, for example, can lead to slightly different measurement results. Also, different measurements might have a different dynamic range in terms of intensity. This is of course affected by the type of source, but also by optical elements in the beam path, such as slits or monochromators, which may be different between setups. Furthermore, the data may look slightly different for different detectors. Similarly, different sample environments may introduce specific noise or background into the measurement. In addition, before any scattering measurement, each sample typically has to be aligned. While this is considered a routine task, the alignment is usually done iteratively and has a finite accuracy, which may have a surprisingly strong impact (Greco *et al.*, 2022).

All these factors are difficult to generalize and usually included in the analysis individually for each experimental setup. Thus, if the results of the measurement are sensitive to these effects, it is difficult to train an ML model that is agnostic to the experimental setup. Therefore, to achieve the highest performance, it is usually necessary to include information about the setup in the model.

The above also implies that the training and test data for neural networks must be of high quality. Training data need to be diverse and accurately labelled to be useful. Currently in most work the training is done with simulated data, since large enough annotated sets of real data are not available. On the other hand, for testing the performance of the neural network a much smaller data set is sufficient. We stress here that real experimental data are often very different from simulated data, which makes it an absolute necessity to judge the performance of a neural network on experimental data.

As with basically all ML applications, the dilemma between high generality with poor performance or high specificity with good performance is also found for surface scattering.

By choosing training data and hyper-parameters the boundary for possible outputs is fixed. For example for X-ray or neutron reflectometry, if we train the neural net only with data from layers without interfacial roughness, we cannot expect that model to perform well on rough layers. On the other hand, increasing the flexibility of the neural network by introducing a wider range of training data usually has a very strong negative impact on the performance of the neural network. Finding the optimum between performance and flexibility is therefore a critical task for ML applications.

3. X-ray and neutron reflectivity

Specular X-ray and neutron reflectometry (XRR and NR), *i.e.* where $\alpha_i = \alpha_f$ and $\phi = 0$, are common techniques for investigating surfaces, thin films and layered structures (Fig. 1). The goal of these measurements is typically to extract different physical parameters for each layer in the sample, such as thickness, roughness and scattering length density (SLD), or in the case of neutrons, even magnetic properties. However, depending on the system studied, the data analysis of reflectivity measurements can be difficult and time consuming. For this reason, several attempts have been made to facilitate data analysis using ML in recent years.

The majority of the publications on this topic focus on the efficient extraction of layer parameters directly from the measured reflectivity curve. The first such published attempt (Greco *et al.*, 2019) demonstrated a fully connected neural network trained to predict the thickness, roughness and electron density of organic thin films based on real-time reflectivity measurements during growth (Fig. 3). The neural network architecture is shown in Fig. 2. The training was done for a fixed substrate using data simulated via a well established theoretical model, such as the Parratt algorithm (Parratt, 1954). The advantage of this method is that, if trained properly, the neural network is well adapted to solving the inverse problem for a given subset of samples and can predict the sample parameters within a fraction of a second with high accuracy. The disadvantage is that a new neural network model must be trained (or at least re-trained) for different sample architectures (*e.g.* different substrates). Other studies have demonstrated that this approach also works in principle for multiple layers (shown for up to three), but the possible parameter range still had to be restricted (Doucet *et al.*, 2021; Mironov *et al.*, 2021). The reason why it is difficult to train a general ML model that is completely agnostic towards the studied system is that, even without considering measurement errors and a finite q_z range, reflectivity problems do not always have a unique solution because of the phase problem (Sivia *et*
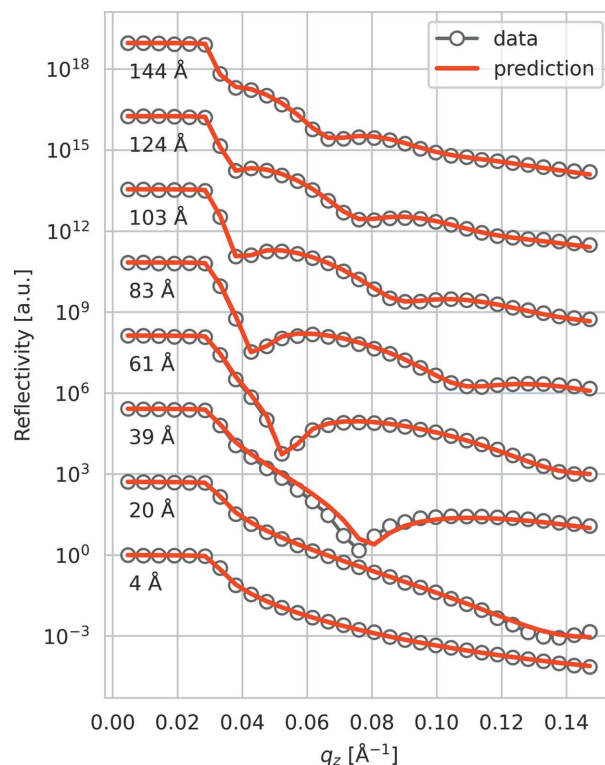

Figure 3
Real-time XRR data sets of a growing organic thin film fitted using the *mlreflect* pipeline (Greco *et al.*, 2022). The grey circles represent XRR measurements at different times during growth. The numbers in ångströms refer to the corresponding film thickness determined via a manual least-mean-squares fit. The red curves are simulations based on the respective neural network predictions. Figure adapted with permission from Greco *et al.* (2019) under a Creative Commons Attribution 4.0 International License, https://creativecommons.org/licenses/by/4.0/.

al., 1991). Including background, noise and roughness can further increase the level of ambiguity, even for a simple system, such as a single layer on a substrate (Greco *et al.*, 2021). Since the above-mentioned neural network models try to approximate an inverse function that maps a given reflectivity curve to a unique set of sample parameters, the solution space necessarily needs to be restricted in such a way as to achieve a mostly unique mapping.

In another approach (Loaiza & Raza, 2021), this problem is tackled by identifying different symmetry-based families of SLD profiles that can be uniquely distinguished. The main idea is that, if the SLD family of the studied system is known, the neural network can predict the complete SLD profile of any sample within that family. While promising, this approach has, however, not yet been tested with experimental data where the above-mentioned experimental conditions apply.

Kim & Lee (2021) demonstrated a different neural network architecture employing a mixture density model that predicts a probability density for the sample parameters in the form of several superimposed multi-modal Gaussians in the solution space. This has the advantage of the network yielding several possible solutions at once, with the height of the Gaussians representing the likelihood of a given solution and the widths of the Gaussians yielding implicit error estimates.

Other groups have tried to employ autoencoder architectures for the analysis of reflectivity data. For example, Andrejevic *et al.* (2021) trained a variational autoencoder to compress reflectivity curves from polarized neutron reflectometry into an information-dense latent space. They deliberately designed the architecture in such a way that the sample parameters can be retrieved from the latent-space variables. Furthermore, the idea is that further fitting in the latent space

is easier than fitting the reflectivity curve in q space, because there are fewer local minima in the objective function. A different application of autoencoders was shown by Aoki *et al.* (2021) where an autoencoder was trained to denoise neutron reflectivity measurements which can then be analysed more easily through conventional means. This can help to reduce the integration time that is necessary to achieve a suitable signal-to-noise ratio during the measurement. This is particularly useful for neutron reflectometry where the flux is typically several orders of magnitude lower than for synchrotron radiation.

While already quite varied, all of the approaches published so far still suffer from the problem of being specific to only a subset of samples. In some cases, the neural networks are even specialized to only a certain combination of materials. This shows the necessity of prior physical knowledge to narrow down the task for the neural network. In all of these examples, this physical knowledge is inserted into the model via the selection of the training data. This means that models must be trained for every subset of problems, which can be non-trivial. Furthermore, after a given model is trained, there is no way to use additional knowledge, *e.g.* from other measurements, to exclude certain solutions. Therefore, in the future it would be interesting to explore neural network architectures that allow the input of knowledge about the studied system during inference time, *i.e.* after the model has already been trained. However, successfully training such a model might be challenging and would arguably require a substantially different neural network architecture from what has been published so far.

Most of the training and testing of neural networks in this context is done with simulated data, since large quantities of
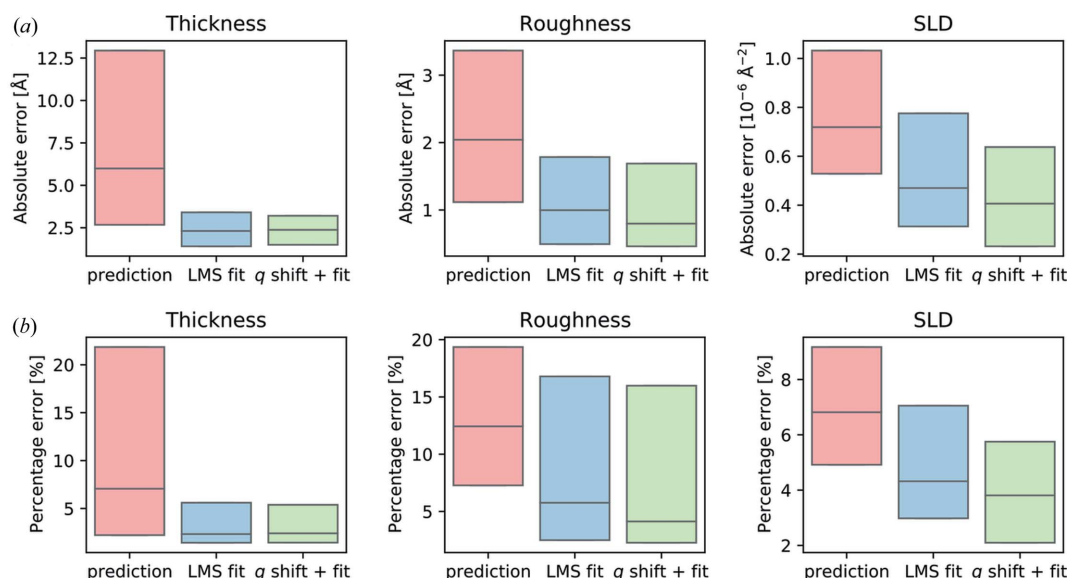

Figure 4
(*a*) Box plots of the absolute errors for 242 measured reflectivity curves for each of the three predicted parameters. The upper and lower edges of the boxes represent the first and third quartiles, respectively, with the horizontal lines inside the boxes denoting the medians. The pink boxes represent the error compared with the pure neural network predictions. The blue boxes represent the error after applying a simple least-mean-squares minimization using the neural network predictions as starting parameters. The green boxes show the error when a q_z shift optimization was performed before this fit. (*b*) The same box plots of the median error but as a percentage of the ground truth. Figure adapted with permission from Greco *et al.* (2022) under a Creative Commons Attribution 4.0 International License, https://creativecommons.org/licenses/by/4.0/.

varied and labelled experimental data are difficult to obtain. Recent work (Greco *et al.*, 2022) has shown, however, that the performance of a model on simulated data is not a good estimate for the performance on real data. While most of the published work demonstrates results on at least some experimental data sets, it frequently does not represent a large and varied set of reflectivity curves. Without such a representative data set, however, it is difficult to judge the general applicability of a given method. As a result, future work should strive to test their performance on larger data sets. In addition, it might be useful to collect a large body of data that can be shared among research groups for standardized testing, as is common in other ML communities.

As an example for a performance test, Greco *et al.* (2022) compared a neural network prediction with conventional fitting results from 242 experimental XRR sets (Fig. 4). The setup there was the fit of three parameters (thickness, roughness, SLD) of a layer on an Si substrate. The median of the initial prediction results (pink boxes in Fig. 4) is in the range of 7–12% of the ground truth. If the initially predicted parameters are used as starting parameters for a least-mean-squares fit, the median of the error decreases to around 5% compared with the ground truth. This result is acceptable for many applications and can be refined further with post-processing. For instance the screening for experimental errors in q (q shift) in combination with a least-squares fit leads to a further decrease in error (green boxes in Fig. 4).

Other issues for XRR and NR which may be critical for analysis with an ML approach include the limited dynamic range of the measurement, since it essentially defines the maximum q_z range that can be measured. This of course limits the amount of information that can be extracted from the data. In addition, for *in situ* reflectivity measurements, the time it takes to perform a scan can be important. If the observed real-time change in the sample is on the same time scale as the time it takes to measure one reflectivity curve, it may happen that different parts of the curve are measured under different sample conditions. For example, during *in situ* annealing of a sample, changes in roughness or thickness may be continuously ongoing during a reflectivity scan. While a human researcher may notice this effect and apply necessary corrections when analysing the data, it is difficult to include this in an ML model.

4. GISAS (GISAXS and GISANS)

Grazing-incidence small-angle X-ray and neutron scattering (GISAXS and GISANS) [Fig. 1(*a*)] are surface-sensitive techniques used to probe the morphology of surfaces with statistically relevant averaging (Levine *et al.*, 1989; Sinha *et al.*, 1988). GISAS has been employed for numerous applications, such as investigating the deposition of metallic nanoparticles on surfaces (Schwartzkopf *et al.*, 2013) and elucidating the morphology of nanostructured polymer thin films (Müller-Buschbaum, 2003).

GISAS experiments employ a scattering geometry where the surface sensitivity is achieved by the grazing incidence of

the incoming beam and the grazing exit of the outgoing beam. If α_i is below the angle of total reflection α_c the transmission of the beam is strongly limited and the amplitude from the reflected beam is increased (Tolan, 1999).

In contrast to bulk methods like powder diffraction, where scattering data in many geometries can often be reduced to a one-dimensional problem, for surface scattering in grazing-incidence geometry such a projection onto one dimension is not possible due to the substrate surface, which breaks the radial symmetry. In addition, the scattering background from the surface is usually anisotropic and can include complex diffuse scattering from the substrate (Sinha *et al.*, 1988). Therefore, GIWAXS and GISAXS data are 2D, which is associated with specific challenges for ML-based analysis

GISAS data can provide information about the morphological parameters of the surfaces studied, such as the number of layers with different thicknesses and densities, the shape and size distributions of nanoparticles on top of or embedded in the layers, or the densities and spatial ordering of nanoparticles. The conventional analysis obtains the corresponding parameters by solving the inverse problem via iterative adjustments of the parameters and minimizing the difference between the measured and simulated data. This fitting routine is typically slow and would greatly benefit from automated ML-based tools.

In general, assumptions about the studied structures might be necessary to reduce complexity and avoid ambiguity in the analysis. Therefore, some of the existing ML solutions for automated GISAS analysis focus on particular morphological models. In this way, convolutional neural networks (CNNs) that extract nanoparticle orientations have been developed (Van Herck *et al.*, 2021; Liu *et al.*, 2019). Fig. 5 shows the typical CNN training workflow with augmented GISAS data (Van Herck *et al.*, 2021). A possible extension to this approach would involve training an ML model to extract nanoparticle size, interparticle distance and surface layer roughness from GISAS data given the underlying assumptions about the morphological model. Due to unavoidable discrepancies between the simulated and experimental data, further improvement in this direction might require building a database with manually analysed GISAS images, which is a challenging task. An alternative approach involves modern data augmentation techniques that can reproduce experimental
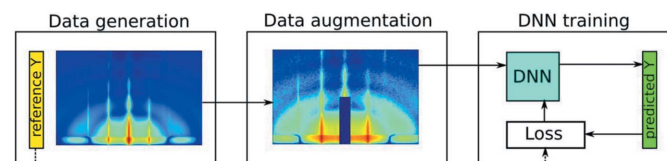
Figure 5
A sketch of the workflow for training a CNN to extract parameters from augmented GISAS data. The GISAS data shown are a simulation of ordered nanoparticles on a surface. Data augmentation is done with several noise sources (including Poisson noise) and cropping of data in the centre to account for the beam stop that is typically present in experimental data. Image reproduced with permission from Van Herck *et al.* (2021) under a Creative Commons Attribution 4.0 International License, https://creativecommons.org/licenses/by/4.0/.

artefacts via the generative adversarial network (GAN) technique and its variants (Goodfellow *et al.*, 2014).

A significantly easier approach is the category classification of GISAS images on the basis of specific characteristics. Ikemoto *et al.* (2020) used a simple CNN to classify GISAXS data according to the shape of the nanoparticles (capsule, spheroid, ellipsoid, truncated spheroid, hemispheroid, prism, hexagonal prism or cylinder). This approach could be used to select the initial model for an iterative fitting of the GISAXS pattern. Also, a CNN was trained to predict 17 different attributes of X-ray scattering images (including GISAXS measurements) from a predefined list (Wang *et al.*, 2017). An interactive visualization system for X-ray scattering images with multiple attributes was introduced by Huang *et al.* (2021). The performance of the multilabel annotation task was improved by Guan *et al.* (2018, 2020). The annotation process for the classification tasks is substantially simpler and faster than a comprehensive analysis, and future development in this direction would benefit from aggregating the corresponding data sets into a standard database available for the community.

5. GIWAS (GIWAXS and GIWANS)

Grazing-incidence wide-angle X-ray and neutron scattering (GIWAXS and GIWANS) are key methods for investigating crystalline structures on surfaces (Feidenhans'l, 1989). The scattering geometry is essentially the same as GISAS [Fig. 1(*a*)] with the only difference being the detected range and resolution in *q*. The wide-angle geometry allows the resolution of Bragg reflections and therefore the analysis of crystal structures and domain orientations. The technique is particularly suitable for *in situ* measurements that enable investigation of crystallization processes or phase transitions in real time, which typically result in hundreds of thousands of images obtained per experimental day.

In contrast to GISAS images with rather complex continuous diffraction features, GIWAS data mostly contain distinct Bragg peaks superimposed on a scattering background and other experimental artefacts such as detector gaps. In general, the characteristics of the Bragg peaks such as their positions, angular and radial sizes, and intensities allow us to obtain information about unit-cell parameters, crystal size distribution or relative fractions of coexisting phases. ML algorithms are ideally suited for processing GIWAXS images by identifying diffraction peaks. However, there are substantially fewer publications on machine learning for GIWAS data than for GISAS. Most of the current approaches focus on preliminary filtering of huge amounts of data. For instance, Wang *et al.* (2017) used a neural network to classify both GISAXS and GIWAXS images. Other approaches are required for quantitative analysis of GIWAXS data.

The way GIWAXS images are processed depends on the application. In some measurements, the expected structures are known and the task of phase determination is simplified to a comparison of the obtained diffraction peak positions with a predefined list of crystal structures. In other cases, the structures are unknown and a correct structure determination

might require indexing algorithms and certain adjustments to the experimental setup (such as larger *q* ranges, absence of coexisting phases *etc.*). Moreover, the diffraction peak characteristics are required for further quantitative analysis. Thus, GIWAS image processing can be split into the peak detection task and further steps using algorithms determined by the specific application. Such an approach allows these complex tasks to be separated into sub-tasks which are easier to improve and test. Separating peak detection and further analysis also allows the use of the same peak detection model for a wide variety of different samples and experimental setups with a wide range of different applications.

Sullivan *et al.* (2019) and Liu *et al.* (2020) employed deep learning methods to accelerate and improve a Bragg peak fitting routine. The first fully automated peak detection approach for GIWAXS images was demonstrated by Starostin *et al.* (2022). A neural network trained on synthetic GIWAXS images allowed them to obtain a list of detected features (areas with coordinates in a 2D image) which could be passed on to other algorithms for peak indexing, structure matching, unit-cell refinement or texture analysis (Fig. 6). Similarly to GISAS, the quality of the peak detection analysis can be improved if annotated experimental data are used for the
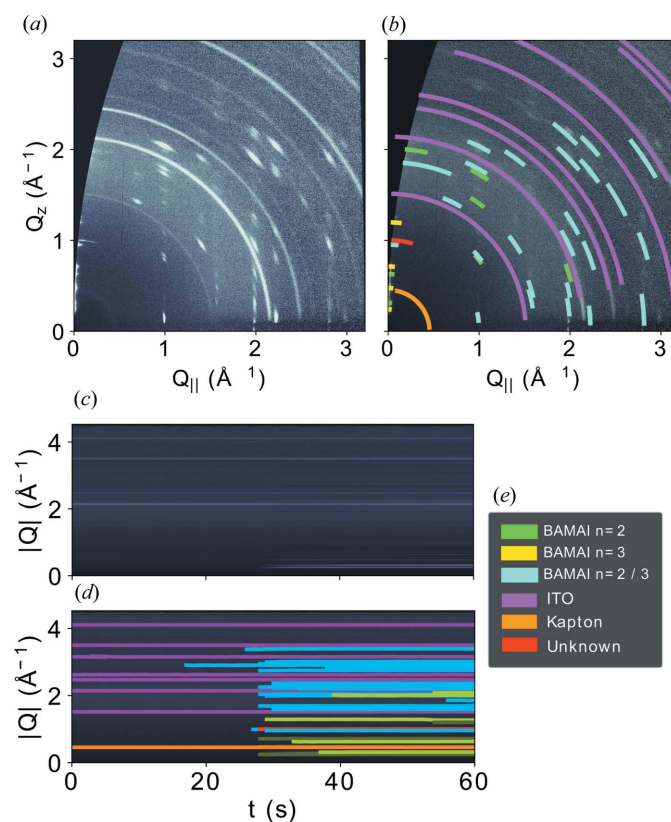

Figure 6
Schematic of GIWAXS peak finding demonstrated for a 2D perovskite thin-film structure during crystallization. (*a*) GIWAXS data, (*b*) annotated with Bragg reflections detected by the ML algorithm at *t* = 60 s. (*c*) The radial positions of the detected peaks over time, with (*d*) the found peak positions. (*e*) Colour code for peak origins. Image reproduced with permission from Starostin *et al.* (2022) under a Creative Commons Attribution 4.0 International License, https://creativecommons.org/licenses/by/4.0/.

training or at least for formal testing. The use of GANs for data augmentation in this case can be particularly complicated since it is challenging to control the appearance of each diffraction peak on the generated images.

In general, the peak finding procedure is followed by the indexing step, which constitutes a highly challenging task. The existing indexing tools either are based on slow iterative routines (Savikhin *et al.*, 2020) or require complementary data from the specular geometry (Kainz *et al.*, 2021), which are unavailable for real-time GIWAXS measurements. Thus, ML techniques might be suitable for accelerating the indexing procedure. Related efforts in this direction are ML-based crystal structure identification methods for 1D X-ray diffraction (XRD) bulk measurements (Tatlier, 2011; Oviedo *et al.*, 2019; Lee *et al.*, 2020). However, to the best of our knowledge, there have been no published attempts to employ ML to accelerate indexing of GIWAS data so far.

The grazing-incidence geometry leads to a highly asymmetric footprint of the beam on the sample surface, and the associated Bragg reflection shape in GIWAXS experiments depends on the resulting resolution, refractive index and penetration depth. The varying shape leads to challenges for the reliable identification of diffraction features in GIWAXS data.

Also, a non-trivial texture, such as a partially preferred domain orientation, may be challenging, in particular in combination with the grazing-incidence geometry and the associated distortion of the scattering signal.

6. X-rays versus neutrons

The above appears to be written mostly from the perspective of X-rays but in fact applies largely to both X-ray and neutron scattering, in particular when considering diffraction applications. For inelastic or quasi-elastic scattering (Grimaldo *et al.*, 2019), or other forms of addressing the dynamics such as XPCS (Sinha *et al.*, 2014; Timmermann *et al.*, 2022), there are more significant differences, but these are not the focus of the present paper. Nevertheless, also for diffraction, we should note some specific features of neutron scattering compared with X-ray scattering (Greco *et al.*, 2021).

In most cases, for both X-rays and neutrons, the scattering follows kinematic theory, except for *e.g.* perfect crystals and optical effects at surfaces (total reflection *etc.*). The key difference concerns the elementary scattering processes and the resulting scattering length and cross sections. These differences lead to the following consequences, all of which can impact the quality of the ML analysis:

(i) For X-rays, the scattering length depends on the number of electrons and can only be positive; for neutrons, the interaction with the nucleus can be both attractive and repulsive. Thus, positive and negative scattering lengths are possible, which can lead to the absence of total external reflection.

(ii) Furthermore, different isotopes of the same element can have very different scattering lengths for neutrons, which allows for contrast tuning through isotopic substitution (Fragneto-Cusani, 2001).

(iii) For neutrons, absorption is usually smaller than for X-rays.

(iv) Neutron scattering is dependent on the nuclear spin. This can introduce an incoherent part of the nuclear scattering. For diffraction applications this can lead to an enhanced background. We note that for quasi-elastic scattering (energy resolved) this can be exploited to study the dynamics (Grimaldo *et al.*, 2019).

(v) Because of the magnetic moment of neutrons, the magnetic structure of the sample can be studied (Ankner & Felcher, 1999). For non-magnetic samples, a magnetic reference layer can be employed (Treece *et al.*, 2019; Skoda *et al.*, 2022) which, for a given sample, produces different scattering patterns depending on the polarization of the beam. These patterns can then be co-refined in a common analysis procedure to reduce ambiguities.

(vi) Many neutron sources use pulsed/polychromatic beams with subsequent energy resolution to measure different q simultaneously. Since the intensity of each wavelength in the spectrum is not constant, counting statistics can be different for different wavelengths or q values (and generally there tends to be a lower incident flux than with X-rays). This affects how the noise is modelled in ML applications (Aoki *et al.*, 2021) and, interestingly, these 'sparse sampling' concepts can also be applied to time-resolved (low-counting) X-ray data (Mareček *et al.*, 2022).

7. Availability of reference data

A major challenge regarding the successful implementation of ML strategies to analyse data arises from the limited availability of suitable test, validation and training data sets consisting of raw experimental data and the corresponding already-performed data analysis. While the training may be aided by simulated (*i.e.* synthetic) data, the true performance test for a neural network to be successful in the analysis of scattering data can only be for real experimental data with their intricacies.

There has been some progress regarding the availability of experimental raw data through data portals and the policies of large-scale facilities (Dimper *et al.*, 2019) and federated data catalogues like PaNOSC and DataFed (Götz *et al.*, 2020; Stansberry *et al.*, 2019) following the FAIR principles (Wilkinson *et al.*, 2016) and endorsing open science (Bezjak *et al.*, 2018). In the current state, however, this can be seen rather as a source of data sets from individual experiments (*e.g.* Scoppola *et al.*, 2020) that provide a significant portion of the metadata generated at large-scale facilities by a specific instrument, but currently do not map the end-to-end process of a scientific experiment (Doucet, 2020). Even more important is the lack of experiment-specific annotated data to enable ML-based data analysis in surface scattering. Today, if available, individual data sets from a limited number of samples may be found attached to individual scientific publications, *e.g.* Doucet *et al.* (2021), as there is no common domain-specific data repository, such as the PDB for protein crystallography (Berman *et al.*, 2000), which allows the

retrieval of analysed data and raw data in a systematic fashion (Helliwell *et al.*, 2019). This lack is also due to the more individual and non-standardized character of the respective experiments in surface scattering compared with *e.g.* protein crystallography.

Nevertheless, larger-scale and comprehensive collections of technique-specific data sets are feasible. For example, *MLExchange* is a platform for easing the use of AI/ML tools by scientific communities by bringing together models, data sets and processing workflows on a single platform (Hexemer *et al.*, 2021). Other recent initiatives such as DAPHNE4NFDI (DAPHNE4NFDI Consortium, 2023) or the Tübingen Cluster of Excellence for Machine Learning in Science (Universität Tübingen – Cluster of Excellence, 2023) are also working towards this goal. Scientific community-driven groups such as ORSO (Open Reflectometry Standards Organization; Arnold *et al.*, 2022) can also help in this context.

As a step towards better availability of reference data, in conjunction with the present publication a collection of ML-ready data sets of X-ray reflectometry are published by Pithan *et al.* (2022).

8. Outlook and summary

We have discussed the status, opportunities, challenges and limitations of ML as applied to X-ray and neutron scattering techniques. In general, ML-based methods are much faster and can be more easily automated compared with conventional fitting and modelling approaches. However the latter are typically more precise, since an expert user controls each step of the analysis. This direct control is, on the other hand, typically time consuming, which makes ML methods preferable for large data volumes and time-critical applications.

In the past few years significant progress has been made in applying ML methods to surface scattering, but some critical milestones are yet to be reached before the scientific community can use ML methods for routine data analysis tasks in surface scattering. Certainly the most relevant missing part is access to large and diverse annotated data sets for testing and comparing the performance of different ML analysis approaches. We are confident that the recent formation of data science consortia will be an important ingredient for the success of ML in scattering.

9. Supporting information – details of XRR data set

We have compiled, and published on Zenodo, a collection of experimental XRR curves, together with corresponding box-model parameters that fit the measured data, and these can be used to test, train or validate ML models (Pithan *et al.*, 2022). From the authors' point of view the provided data set is intended as a nucleation site for a corresponding reference database. We plan to extend this data set to include a larger variety of models, materials, substrate materials and NR data, and we explicitly welcome external contributions to further versions of the data collection.

References

Als-Nielsen, J. & McMorrow, D. (2011). *Elements of Modern X-ray Physics*, 2nd ed. Chichester: John Wiley & Sons.

Andrejevic, N., Chen, Z., Nguyen, T., Fan, L., Heiberger, H., Lauter, V., Zhou, L.-J., Zhao, Y.-F., Chang, C.-Z., Grutter, A. & Li, M. (2021). *arXiv*:2109.08005.

Ankner, J. & Felcher, G. (1999). *J. Magn. Magn. Mater.* **200**, 741–754.

Aoki, H., Liu, Y. & Yamashita, T. (2021). *Sci. Rep.* **11**, 22711.

Archibald, R. K., Doucet, M., Johnston, T., Young, S. R., Yang, E. & Heller, W. T. (2020). *J. Appl. Cryst.* **53**, 326–334.

Arnold, T., Murphy, B., Stahn, J., Skoda, M., Maranville, B., Nelson, A., Kinane, C. & McCluskey, A. (2022). Open Reflectometry Standards Organisation (ORSO), https://www.reflectometry.org/.

Bai, J., Xue, Y., Bjorck, J., Le Bras, R., Rappazzo, B., Bernstein, R., Suram, S. K., Van Dover, R. B., Gregoire, J. M. & Gomes, C. P. (2018). *AI Mag.* **39**, 15–26.

Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. & Bourne, P. (2000). *Nucleic Acids Res.* **28**, 235–242.

Bezjak, S., Clyburne-Sherin, A., Conzett, P., Fernandes, P., Görögh, E., Helbig, K., Kramer, B., Labastida, I., Niemeyer, K., Psomopoulos, F., Ross-Hellauer, T., Schneider, R., Tennant, J., Verbakel, E., Brinken, H. & Heller, L. (2018). *Open Science Training Handbook*, https://www.fosteropenscience.eu/content/open-science-training-handbook.

Birkholz, M. (2006). *Thin Film Analysis by X-ray Scattering*. Weinheim: Wiley-VCH.

Chang, M.-C., Wei, Y., Chen, W.-R. & Do, C. (2020). *MRS Commun.* **10**, 11–17.

Chen, Z., Andrejevic, N., Drucker, N. C., Nguyen, T., Xian, R. P., Smidt, T., Wang, Y., Ernstorfer, R., Tennant, D. A., Chan, M. & Li, M. (2021). *Chem. Phys. Rev.* **2**, 031301.

DAPHNE4NFDI Consortium (2023). *DAPHNE4NFDI*, https://www.daphne4nfdi.de.

Dimper, R., Götz, A., de Maria, A., Solé, V., Chaillet, M. & Lebayle, B. (2019). *Synchrotron Rad. News*, **32**(3), 7–12.

Doucet, M. (2020). *Driving Scientific and Engineering Discoveries Through the Convergence of HPC, Big Data and AI*, edited by J. Nichols, B. Verastegui, A. B. Maccabe, O. Hernandez, S. Parete-Koon & T. Aheran, pp. 257–268. Cham: Springer International Publishing.

Doucet, M., Archibald, R. K. & Heller, W. T. (2021). *Mach. Learn. Sci. Technol.* **2**, 035001.

Erdmann, M., Glombitza, J., Kasieczka, G. & Klemradt, U. (2021). *Deep Learning for Physics Research*. Singapore: World Scientific.

Feidenhans'l, R. (1989). *Surf. Sci. Rep.* **10**, 105–188.

Fragneto-Cusani, G. (2001). *J. Phys. Condens. Matter*, **13**, 4973–4989.

Franke, D., Jeffries, C. M. & Svergun, D. I. (2018). *Biophys. J.* **114**, 2485–2492.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. (2014). *Advances in Neural*

Information Processing Systems, edited by Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence & K. Weinberger. Red Hook: Curran Associates.

Götz, A., Bertelsen, M., Bodera Sempere, J., Campbell, A., Carboni, N., Caunt, S., De Maria Antolinos, A., Dimper, R. E. J., Fangohr, H., Fortmann-Grote, C., Gliksohn, F., Hall, J., Holm Rod, T., Kieffer, J., Kluyver, T., Perrin, J.-F., Pugliese, R., Richter, T., Rosca, R., Schrettner, L., Solé, V. A., Taylor, J. & Vincet, T. (2020). *Proceedings of the 17th International Conference on Accelerator and Large Experimental Physics Control Systems, ICALEPCS2019*, 5–11 October 2019, New York, USA, pp. 694–701. Geneva: CERN.

Greco, A., Starostin, V., Edel, E., Munteanu, V., Rußegger, N., Dax, I., Shen, C., Bertram, F., Hinderhofer, A., Gerlach, A. & Schreiber, F. (2022). *J. Appl. Cryst.* **55**, 362–369.

Greco, A., Starostin, V., Hinderhofer, A., Gerlach, A., Skoda, M. W. A., Kowarik, S. & Schreiber, F. (2021). *Mach. Learn. Sci. Technol.* **2**, 045003.

Greco, A., Starostin, V., Karapanagiotis, C., Hinderhofer, A., Gerlach, A., Pithan, L., Liehr, S., Schreiber, F. & Kowarik, S. (2019). *J. Appl. Cryst.* **52**, 1342–1347.

Grimaldo, M., Roosen-Runge, F., Zhang, F., Schreiber, F. & Seydel, T. (2019). *Q. Rev. Biophys.* **52**, e7.

Guan, Z., Qin, H., Yager, K. G., Choo, Y. & Yu, D. (2018). 29th British Machine Vision Conference (BMVC), 3–6 September 2018, Newcastle upon Tyne, UK, Abstract No. 245.

Guan, Z., Yager, K. G., Yu, D. & Qin, H. (2020). *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1–5 March 2020, Snowmass, Colorado, USA, pp. 2190–2198. New York: IEEE.

Helliwell, J. R., Minor, W., Weiss, M. S., Garman, E. F., Read, R. J., Newman, J., van Raaij, M. J., Hajdu, J. & Baker, E. N. (2019). *IUCrJ*, **6**, 341–343.

Hexemer, A., Zwart, P., McReynolds, D., Green, A. & Chavez Esparza, T. (2021). *MLExchange.* Version 1. https://www.osti.gov/doecode/biblio/61623.

Holý, V., Pietsch, U. & Baumbach, T. (1999). *High-Resolution X-ray Scattering from Thin Films and Multilayers.* Berlin: Springer.

Huang, X., Jamonnak, S., Zhao, Y., Wang, B., Hoai, M., Yager, K. G. & Xu, W. (2021). *IEEE Trans. Vis. Comput. Graph.* **27**, 1312–1321.

Ikemoto, H., Yamamoto, K., Touyama, H., Yamashita, D., Nakamura, M. & Okuda, H. (2020). *J. Synchrotron Rad.* **27**, 1069–1073.

Kainz, M. P., Legenstein, L., Holzer, V., Hofer, S., Kaltenegger, M., Resel, R. & Simbrunner, J. (2021). *J. Appl. Cryst.* **54**, 1256–1267.

Kim, K. T. & Lee, D. R. (2021). *J. Appl. Cryst.* **54**, 1572–1579.

Konstantinova, T., Wiegart, L., Rakitin, M., DeGennaro, A. M. & Barbour, A. M. (2021). *Sci. Rep.* **11**, 14756.

Lee, J.-W., Park, W. B., Lee, J. H., Singh, S. P. & Sohn, K.-S. (2020). *Nat. Commun.* **11**, 86.

Levine, J. R., Cohen, J. B., Chung, Y. W. & Georgopoulos, P. (1989). *J. Appl. Cryst.* **22**, 528–532.

Liu, S., Melton, C. N., Venkatakrishnan, S., Pandolfi, R. J., Freychet, G., Kumar, D., Tang, H., Hexemer, A. & Ushizima, D. M. (2019). *MRS Commun.* **9**, 586–592.

Liu, Z., Sharma, H., Park, J.-S., Kenesei, P., Almer, J., Kettimuthu, R. & Foster, I. (2020). *arXiv*:2008.08198.

Loaiza, J. M. C. & Raza, Z. (2021). *Mach. Learn. Sci. Technol.* **2**, 025034.

Ludwig, A. (2019). *NPJ Comput. Mater.* **5**, 70.

Mareček, D., Oberreiter, J., Nelson, A. & Kowarik, S. (2022). *J. Appl. Cryst.* **55**, 1305–1313.

Mironov, D., Durant, J. H., Mackenzie, R. & Cooper, J. F. K. (2021). *Mach. Learn. Sci. Technol.* **2**, 035006.

Müller-Buschbaum, P. (2003). *Anal. Bioanal. Chem.* **376**, 3–10.

Oviedo, F., Ren, Z., Sun, S., Settens, C., Liu, Z., Hartono, N. T. P., Ramasamy, S., DeCost, B. L., Tian, S. I., Romano, G., Kusne, A. G. & Buonassisi, T. (2019). *NPJ Comput. Mater.* **5**, 60.

Parratt, L. G. (1954). *Phys. Rev.* **95**, 359–369.

Pithan, L., Greco, A., Hinderhofer, A., Gerlach, A., Kowarik, S., Rußegger, N., Dax, I. & Schreiber, F. (2022). *Reflectometry Curves (XRR and NR) and Corresponding Fits for Machine Learning*, https://doi.org/10.5281/zenodo.6497437.

Savikhin, V., Steinrück, H.-G., Liang, R.-Z., Collins, B. A., Oosterhout, S. D., Beaujuge, P. M. & Toney, M. F. (2020). *J. Appl. Cryst.* **53**, 1108–1129.

Schwartzkopf, M., Buffet, A., Körstgens, V., Metwalli, E., Schlage, K., Benecke, G., Perlich, J., Rawolle, M., Rothkirch, A., Heidmann, B., Herzog, G., Müller-Buschbaum, P., Röhlsberger, R., Gehrke, R., Stribeck, N. & Roth, S. V. (2013). *Nanoscale*, **5**, 5053.

Scoppola, E., Fragneto, G., Kuhrts, L. & Micciulla, S. (2020). *Lipid Bilayers at Soft Liquid/Liquid Interfaces. Data Set*, https://doi.esrf.fr/10.15151/ESRF-ES-187132524.

Sinha, S. K., Jiang, Z. & Lurio, L. B. (2014). *Adv. Mater.* **26**, 7764–7785.

Sinha, S. K., Sirota, E. B., Garoff, S. & Stanley, H. B. (1988). *Phys. Rev. B*, **38**, 2297–2311.

Sivia, D. S., Hamilton, W. A., Smith, G. S., Rieker, T. P. & Pynn, R. (1991). *J. Appl. Phys.* **70**, 732–738.

Skoda, M. W., Conzelmann, N. F., Fries, M. R., Reichart, L. F., Jacobs, R. M., Zhang, F. & Schreiber, F. (2022). *J. Colloid Interface Sci.* **606**, 1673–1683.

Song, G., Porcar, L., Boehm, M., Cecillon, F., Dewhurst, C., Goc, Y. L., Locatelli, J., Mutti, P. & Weber, T. (2020). *EPJ Web Conf.* **225**, 01004.

Stansberry, D., Somnath, S., Breet, J., Shutt, G. & Shankar, M. (2019). *DataFed: Towards Reproducible Research via Federated Data Management.* Las Vegas: IEEE.

Starostin, V., Munteanu, V., Greco, A., Kneschaurek, E., Pleli, A., Bertram, F., Gerlach, A., Hinderhofer, A. & Schreiber, F. (2022). *NPJ Comput. Mater.* **8**, 101.

Sullivan, B., Archibald, R., Azadmanesh, J., Vandavasi, V. G., Langan, P. S., Coates, L., Lynch, V. & Langan, P. (2019). *J. Appl. Cryst.* **52**, 854–863.

Tatlier, M. (2011). *Neural Comput. Appl.* **20**, 365–371.

Timmermann, S., Starostin, V., Girelli, A., Ragulskaya, A., Rahmann, H., Reiser, M., Begam, N., Randolph, L., Sprung, M., Westermeier, F., Zhang, F., Schreiber, F. & Gutt, C. (2022). *J. Appl. Cryst.* **55**, 751–757.

Tolan, M. (1999). *X-ray Scattering from Soft-Matter Thin Films: Materials Science and Basic Research.* Berlin: Springer.

Treece, B. W., Kienzle, P. A., Hoogerheide, D. P., Majkrzak, C. F., Lösche, M. & Heinrich, F. (2019). *J. Appl. Cryst.* **52**, 47–59.

Universität Tübingen – Cluster of Excellence (2023). *Machine Learning: New Perspectives for Science*, https://uni-tuebingen.de/en/research/core-research/cluster-of-excellence-machine-learning/home/.

Van Herck, W., Fisher, J. & Ganeva, M. (2021). *Mater. Res. Expr.* **8**, 045015.

Wang, B., Yager, K., Yu, D. & Hoai, M. (2017). *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 24–31 March 2017, Santa Rosa, California, USA, pp. 697–704. New York: IEEE.

Wang, J., Wang, W., Chen, Y., Song, L. & Huang, W. (2021). *Small Methods*, **5**, 2100829.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. & Mons, B. (2016). *Sci. Data*, **3**, 160018.