frontiers | Frontiers in Materials

TYPE Original Research
PUBLISHED 04 December 2024
DOI 10.3389/fmats.2024.1440608

OPEN ACCESS

EDITED BY
John L. Provis,
Paul Scherrer Institut (PSI), Switzerland

REVIEWED BY
Kelvin Xie,
Texas A and M University, United States
Stefanos Papanikolaou,
National Centre for Nuclear Research, Poland

*CORRESPONDENCE
Stefan Sandfeld,
s.sandfeld@fz-juelich.de

RECEIVED 29 May 2024
ACCEPTED 30 October 2024
PUBLISHED 04 December 2024

CITATION
Zhang C, Bos C, Sandfeld S and Schwaiger R (2024) Unsupervised learning of nanoindentation data to infer microstructural details of complex materials.
Front. Mater. 11:1440608.
doi: 10.3389/fmats.2024.1440608

Unsupervised learning of nanoindentation data to infer microstructural details of complex materials

Chen Zhang[1], Clémence Bos[2], Stefan Sandfeld[1,3]* and Ruth Schwaiger[4,5]

[1]Institute for Advanced Simulation – Materials Data Science and Informatics (IAS-9), Forschungszentrum Jülich GmbH, Jülich, Germany, [2]Institute for Applied Materials, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, [3]Chair of Materials Data Science and Informatics, Faculty of Georesources and Materials Engineering, RWTH Aachen University, Aachen, Germany, [4]Institute of Energy and Climate Research – Structure and Function of Materials (IEK-2), Forschungszentrum Jülich GmbH, Jülich, Germany, [5]Chair of Energy Engineering Materials, Faculty of Georesources and Materials Engineering, RWTH Aachen University, Aachen, Germany

In this study, Cu-Cr composites were studied by nanoindentation. Arrays of indents were placed over large areas of the samples resulting in datasets consisting of several hundred measurements of Young's modulus and hardness at varying indentation depths. The unsupervised learning technique, Gaussian mixture model, was employed to analyze the data, which helped to determine the number of "mechanical phases" and the respective mechanical properties. Additionally, a cross-validation approach was introduced to infer whether the data quantity was adequate and to suggest the amount of data required for reliable predictions–one of the often encountered but difficult to resolve issues in machine learning of materials science problems.

KEYWORDS

unsupervised learning, cross-validation, Gaussian mixture model, Cu-Cr composite, mechanical properties, nanoindentation

1 Introduction

Nanoindentation has emerged as a powerful technique for characterizing the mechanical properties of materials at small length scales (Shen, 2019; Golovin, 2021). By applying a force to a sharp indenter tip, while measuring the resulting displacement of the tip into the material, nanoindentation enables the characterization of the mechanical behavior, including hardness and elastic modulus (Oliver and Pharr, 1992), with high spatial resolution. In recent years, there has been growing interest in the application of statistical approaches to nanoindentation data analysis, offering new insights into material properties and behavior.

The traditional approach to nanoindentation testing involves conducting individual tests on small regions of a sample surface and averaging the resulting mechanical properties to obtain a representative value. While this method provides valuable information, it may overlook variations in mechanical properties within the material that may arise from statistical heterogeneity. Statistical nanoindentation techniques, on the other hand, seek to capture and analyze the full distribution of mechanical properties across a sample, allowing for a more comprehensive assessment of its mechanical behavior. Grid

indentation (Constantinides et al., 2006; Nohava et al., 2012; Sanchez-Camargo et al., 2020), for example, is not only used for statistical analysis of the mechanical properties of multiphase materials (Ulm et al., 2007; Haušild et al., 2016; Hintsala et al., 2018), but also for mapping the mechanical properties (Tromas et al., 2012; Hintsala et al., 2018) to study the correlations between microstructural features and corresponding properties. Due to progress in equipment, it has become popular to study vast numbers of indents, which nowadays can be measured in remarkably little time. For example, around 500,000 indentations were reported in a study of an Al-Cu eutectic alloy and a duplex stainless steel (Besharatloo and Wheeler, 2021), 100,000 indentations were performed to analyze the properties of the Taza meteorite (Wheeler, 2021), and 212,500 indents were performed in a study of thermal-barrier coatings (Vignesh et al., 2019).

An important aspect of statistical nanoindentation is the application of advanced data analysis and machine learning techniques (Puchi-Cabrera et al., 2023) to extract meaningful information from indentation data. For example, data deconvolution can be performed by assuming multi-modal Gaussian distributions and fitting the probability distribution functions (Sorelli et al., 2008; Randall et al., 2009) or applying expectation maximization techniques (Veytskin et al., 2017). Recently, the effectiveness of different deconvolution methods was studied (Besharatloo and Wheeler, 2021). A convolutional neural network-based classifier (Kossman and Bigerelle, 2021) was developed to identify whether pop-in events were present in the load-displacement curves from nanoindentation tests to help understand the process that created pop-ins. Graph neural networks have been used for supervised learning of indentation data obtained from a polycrystalline steel in conjunction with electron back-scatter diffraction (EBSD) mappings (Karimi et al., 2023); they are able to also consider non-local information. In another study (Vignesh et al., 2019), the phase level features were extracted from spatial hardness and elastic modulus maps using a deconvolution method based on the KMeans clustering algorithm. The method was also employed for the examination of dual phase and high-strength low-alloy steels (Jentner et al., 2023). Similarly, Bayesian inference methods (Becker et al., 2022; Puchi-Cabrera et al., 2023) offer a probabilistic framework for estimating material properties and uncertainties, providing a more robust and comprehensive analysis of nanoindentation data.

In this study, we utilized Cu-Cr composites with controlled heterogeneity as a model material, characterized by varying fractions of the two material components. Our objective was to assess the effectiveness of the statistical nanoindentation technique in detecting variations in mechanical properties within a material. While the microstructures of the selected materials allow for precise positioning of indents and a conventional analysis, this controlled heterogeneity served as a basis for exploring fundamental methodological aspects, such as data processing, uncertainty quantification, and model selection. Unlike more complex materials with multiple, possibly unknown, phases and heterogeneous microstructures, the controlled heterogeneity in our composites simplifies the interpretation of the nanoindentation data by eliminating the complexities associated with unknown heterogeneity. Another key question that arises in the context of statistical nanoindentation is: How much data is necessary to obtain reliable and meaningful results? We address this question by systematically investigating the effects of data quantity on the statistical analysis of the nanoindentation data.

2 Materials and methodology

Cu-Cr composites as a two-phase model material with different compositions were evaluated. Four materials were studied with different fractions of Cr, i.e., 25 wt% Cr and 60 wt% Cr corresponding to 29.95 at% and 64.40 at% Cr, respectively, as well as Cu and Cr as reference samples. All materials were produced via field-assisted sintering technique (FAST) as described in detail in (von Klinski-Berger, 2015). Briefly, Cu powder with 99.9 at% purity and technically pure Cr powder (99.5 at%) were used and compacted at a temperature of 950°C and a pressure of 40 MPa, except for the Cr sample that was compacted at a temperature of 1,450°C. The composite samples (Figure 1) will be referred to as CuCr25 and CuCr60 according to their nominal compositions. For indentation testing, the sample surfaces were prepared applying standard grinding and polishing techniques using SiC paper and diamond suspensions with decreasing grain size down to 0.1 μm.

A nanoindenter G200 XP (Agilent/Keysight Technologies, Inc., CA, United States) equipped with a diamond Berkovich tip was used to investigate the mechanical properties of the composite samples. The samples were indented to different depths using the so-called Express Test option. Arrays of indents covering areas up to 500 μm × 500 μm were made in different locations on the sample surface. The indentation depths ranged from 200 to 2,000 nm and the distance between individual indents was maintained approximately constant between 20 and 23 μm for all depths. Hardness H and Young's modulus E were determined assuming 1,141 GPa and 0.07 for Young's modulus and Poisson's ratio, respectively, of the diamond tip. The Poisson ratio of the materials was fixed at 0.25. Figure 2 shows an example distribution of hardness and Young's modulus of CuCr60 for an average indentation depth of 1 μm. The marginal hardness distribution is a bimodal distribution, while a trimodal distribution can be seen for the marginal histogram of Young's modulus.

The Cu-Cr composites consisted of two distinct phases with average Cr particle diameters of approximately 30 μm on the indented surface; thus, upon indenting the surface we assume that either one or the other element dominates or the properties of a mixture of both phases is measured. However, the "mixture of elements" might as well be more than "just the sum of its parts" since additional effects, e.g., related to the presence of interfaces, might occur during indentation. Furthermore, we assume that similar local microstructural or chemical properties lead to similar measurement data and that the mechanical properties exhibit gradual changes over the surface.

Thus, the data-science task is to analyze a number of data records consisting of the given feature variables E (Young's modulus) and H (hardness), measured for four different materials and at different depths, i.e., 300, 400 and 1,000 nm. The whole dataset $\mathcal{D}$ consists of N data records and can be written as the set of pairs

$$\mathcal{D} = \{(E_i, H_i)\}_{i=1 \quad \ldots N}.$$

The goal of this work is to perform a clustering analysis, during which data points with similar elastic properties are grouped

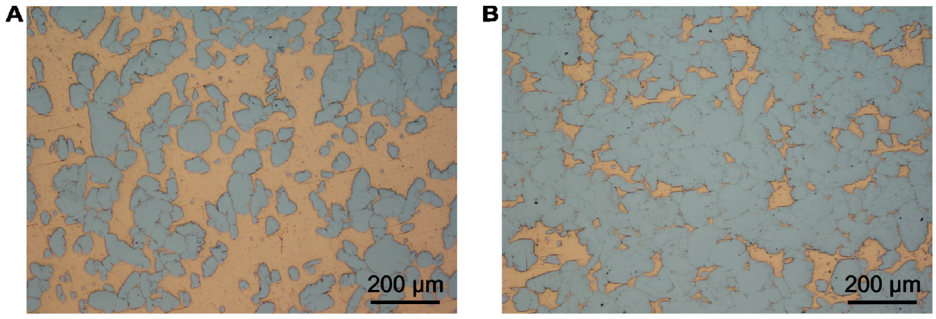

FIGURE 1
Optical micrographs of Cu-Cr composites produced by field-assisted sintering technique with **(A)** 25 wt% Cr, (referred to as CuCr25), and **(B)** 60 wt% Cr (referred to as CuCr60) were investigated by indentation.

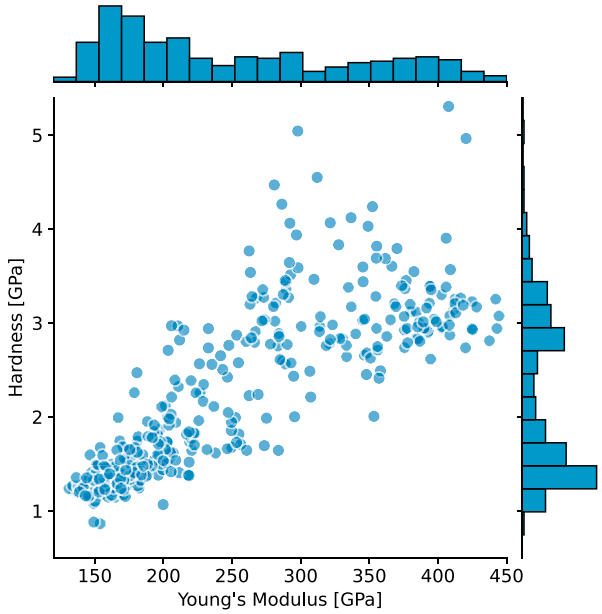

FIGURE 2
Scatter plot with marginal histograms of all obtained values for Young's modulus and hardness measured for CuCr60 at an average indentation depth of 1 µm.

together, i.e., for each pair (E_i, H_i) in the feature space, we determine the categorical variable y_i that gives the number of the respective cluster. Additionally, the total number of different clusters needs to be determined. Since annotated training data does not exist, this has do be done in an unsupervised manner.

For the clustering, we use the Gaussian mixture model (GMM), which is a robust and well-established probabilistic clustering model in the statistics literature [see, e.g. (Bishop and Nasrabadi, 2006; Reynolds, 2009; Reynolds and Rose, 1995)]. The GMM technique has been used successfully in a wide range of materials science applications. It has, for example, been applied to the automated analysis and visualization of continuum fields in atomistic simulations to extract distributions of total strain, elastic strain, and rotation for individual grains (Prakash and Sandfeld, 2022), or to determine the so-called *grain orientation spread* based on electron back-scatter diffraction (Yeo et al., 2023). In the analysis of high resolution high-angle annular dark-field scanning transmission electron microscopy data, the GMM was used to estimate the number of atoms of crystalline nanostructures assuming that the total scattered intensity is proportional to the number of atoms per atom column (De Backer et al., 2013). Furthermore, X-ray diffraction investigations using GMM allowed the automatic extraction of charge density wave order parameters and the detection of intraunit cell ordering and its fluctuations from a series of high-volume X-ray diffraction measurements (Venderley et al., 2022).

Using nanoindentation data to investigate the distribution of heterogeneous materials has great potential, provided the GMM technique is applied properly and a sufficient amount of data is available. Here, each "phase" is assumed to correspond to an individual Gaussian distribution of Young's modulus and hardness. These materials "phases" are commonly referred to as *components* in the context of machine learning. We assume that the distribution of experimental data was generated by a combination of Gaussian processes, which are represented by the probability density functions (PDFs) $\mathcal{N}$ for each component j. The resulting superposition is then given by

$$p(\mathcal{D};\boldsymbol{\Phi}) = \sum_{j=1}^{k} \alpha_j \quad \mathcal{N}\left(\mathcal{D} \mid \boldsymbol{\theta}_j\right),$$

where k is the number of components of the model, $\alpha_j > 0$ are the weights of each component j, the $\boldsymbol{\theta}_j$ are the vectors of parameters for the Gaussian, and $\boldsymbol{\Phi} = \{\alpha_1, \ldots, \alpha_k, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k\}$ is a short notation for the whole set of parameters governing the Gaussian mixture model. For a multivariate Gaussian, the component j of the superimposed function is given by the parameters $\boldsymbol{\theta}_j = \{\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\}$, where $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ are the mean value and the covariance matrix, respectively. $\boldsymbol{\mu}_j$ describes the location of the component j in the feature space, while the covariance matrix $\boldsymbol{\Sigma}_j$ characterizes the j-th component data distributed around $\boldsymbol{\mu}_j$. The objective of the training process is to estimate the values for the model parameters of the Gaussian distribution that best align with the training data $\mathcal{D}^{\text{train}} \subset \mathcal{D}$. The model parameters $\boldsymbol{\Phi}_k$ are iteratively determined while assuming the number of clusters k to be predefined. Typically, the superposition of Gaussians is computed using the maximum

likelihood method based on a set of candidate models that differ in the number of clusters generated using the expectation maximization algorithm. For further details of the algorithm used, please refer to the appendix.

The Bayesian Information Criterion (*BIC*) is employed as a selection criterion to identify the optimal model. It is well established and commonly applied in model selection (Gideon, 1978), providing a measure for assessing the accuracy of the unsupervised GMM:

$$BIC = -2 \ln \mathcal{L} + d \ln N, \tag{1}$$

where d denotes the number of parameters of the model, and $\mathcal{L}$ is the maximum likelihood achieved by the model used. The first term represents the maximized likelihood of a model, and the second term introduces a penalty for the number of parameters to mitigate the risk of overfitting. The model with the lowest *BIC* value indicates the highest likelihood which has the best predictive capability for the observed data. In this study, the GMM analysis has been implemented using the open-source Python package "scikit-learn" (Pedregosa et al., 2011). A comparison in Section 3.4 additionally justifies the choice of the GMM model.

3 Results and discussion

In the following, we will start with the description of the data cleaning process and the analysis of the one-component metallic composites using a 1D Gaussian mixture model. Then, the CuCr25 and CuCr60 composites are investigated using both 1D and 2D GMM for an average indentation depth of 1 μm. Finally, a comprehensive analysis is conducted to examine the influence of sample size on the robustness of the model as well as the influence of indentation depth.

3.1 Preparation of the datasets

Figure 3A illustrates the original datasets comprising the measured E and H values of the four materials tested. Due to variations in the height of the materials' regions on the polished surface, the actual recorded maximum indentation depths ranged from 200 to 2,000 nm, deviating from the nominal values set in the experiments. Here, we focused our analysis on depths ranging from 800 to 1,200 nm. Subsequently, the data underwent cleaning and filtering processes to remove measurement errors, particularly outliers with unrealistically high values as well as other invalid data. For CuCr25, data within the ranges of $100 \leq E \leq 400$ GPa and $0.8 \leq H \leq 4.5$ GPa were retained, while for CuCr60 the data range was $100 \leq E \leq 500$ GPa and $1.0 \leq H \leq 5.0$ GPa. The resulting cleaned data, comprising approximately 98% of the original data, is presented in Figure 3B. Notably, the distributions of the two pure metals exhibit lower variances compared to the Cu-Cr composites. Further preprocessing of the data, such as standardization, was found to have no significant impact on the training results and was therefore not included in this study.

3.2 Mechanical properties of the pure Cu and pure Cr specimens

The mechanical properties of pure Cu and pure Cr were analyzed for the indentation depth of 800–1,200 nm. The probability density function (PDF) plots of E and H are depicted in Figure 4A-E showing the mean values of $E = 118.80$ GPa and $H = 0.91$ GPa for pure Cu, and $E = 371.24$ GPa and $H = 3.21$ GPa for pure Cr. The bin size was selected to ensure an approximately equal number of bins covering the range of all PDFs. To utilize the GMM, it is important to demonstrate that the distributions are roughly normally distributed, despite the GMM's inherent robustness. Thus, the Shapiro-Wilk test was employed to assess the normality of the distributions (Öztuna et al., 2006), resulting in p-values of 0.95 for E and 0.14 for H in pure Cu, and 0.95 for E and 0.14 for H in pure Cr. The test accepts the normality hypothesis, when the p-value exceeds 0.05, confirming that the data are sufficiently normally distributed. The optimal number of components was 1 for both the Cu and Cr specimens (as shown in Figure 4C and Figure 4F), as determined by the BIC analysis. Regarding determination of the component numbers, the result of 1D GMM for E with high p-values is therefore more reliable than the fit of H. Nonetheless, because H is strongly correlated with local microstructure details, such as phase and grain boundaries or dislocations, the fitting result of H can be used to infer local microstructural characteristics. This can also be seen from the small variation at the beginning of the BIC plot for hardness in Figure 4F, where the BIC values for $k = 2$ differ only marginally from the one at $k = 1$ that we had identified as the optimal one.

3.3 Mechanical properties of the CuCr25 and CuCr60 specimens

The histograms of E and H of the CuCr composites, together with their best-fit models are shown in Figure 5A and Figure 5D (left panel) and Figure 5B and Figure 5E (middle panel). The right panel (Figure 5C and Figure 5F) shows the BIC as a function of the number of components.

The BIC values for both the modulus and the hardness cover a range of around 500 and ≈200, respectively, if $k = 1$ is excluded. Though not obvious, this is an important difference as compared to the previous investigation of pure Cu and Cr, as detailed below.

How many free parameter need to be determined by how many data points? Taking CuCr60, for example, there are $N \approx 300$ valid measurements for E determined at an average depth of 1 μm. Assuming an optimal number of clusters of $k = 3$, the number of parameters to be measured in a 1-dimensional analysis are three mean values (μ_1, μ_2, μ_3), three variance values $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$, and two coefficients that determine the relative weights of the three Gaussians (α_1, α_2). Given that the sum of all weights should be 1, which means $\sum_{j=1}^{k} \alpha_j = 1$ and $\alpha_3 = 1 - (\alpha_1 + \alpha_2)$, the number of parameters to be determined is $d = 8$. Thus, as a first estimate we conclude that the amount of data should be more than sufficient in this case (assuming that the variance of the data is small).

Are the differences of the BIC values as a function of k large or small? To answer this question, we need to understand how much variation results from a small change in the dataset. The calculated

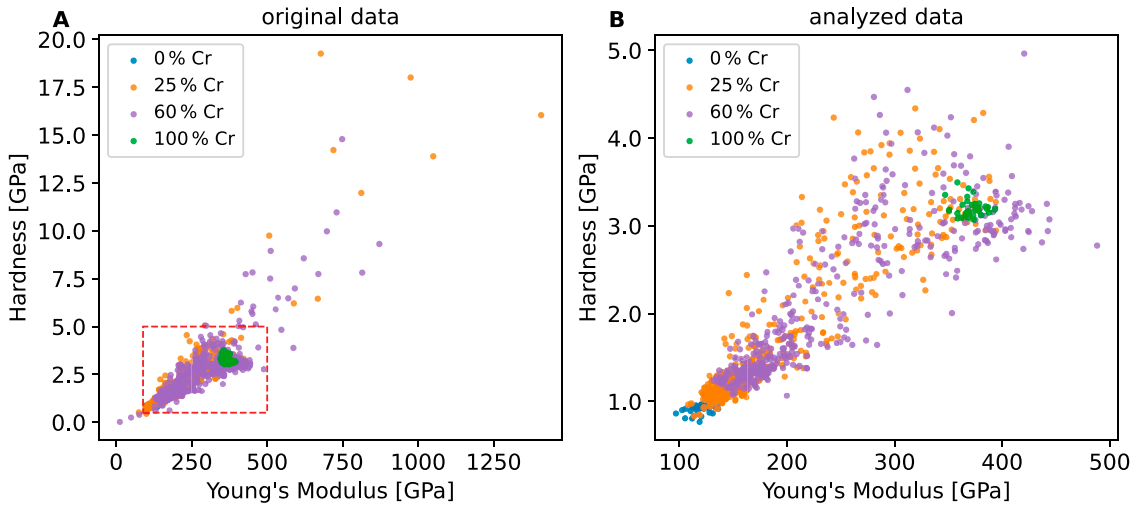

FIGURE 3
The experimentally obtained distributions of Young's modulus and hardness for 0 wt%, 25 wt%, 60 wt% and 100 wt% Cr content. **(A)** The experimental data in its original form, **(B)** The cleaned and preprocessed data set used in our analysis. The rectangle in **(A)** indicates the region illustrated in **(B)**.

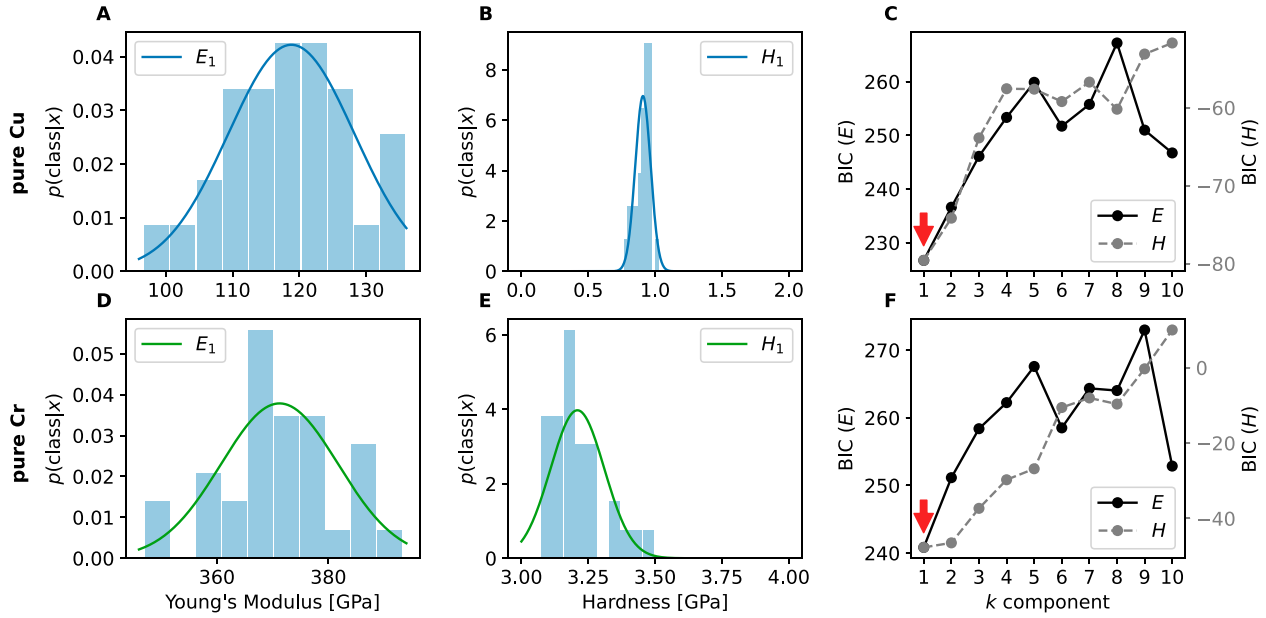

FIGURE 4
Probability density functions (left and middle column) and plot of BIC (right column) of pure Cu and Cr. The BIC values are shown for both, Young's modulus (left "*y*" axis) and hardness (right "*y*" axis). The top row shows the data for Cu **(A–C)**, the bottom row for Cr **(D–F)**.

values of the logarithm of $\mathcal{L}$ (cf. Equation 1) of this dataset are between −10 and −4. According to Equation 1, the first term varies between $-2\times-4=8$ and $-2\times-10=20$ and the second term is $8\times\ln 300\approx 45$. The sum of the two terms is in between 53 and 65. Assuming a minor variation in 1 out of 300 data points, the BIC variation should fall between 53 and 65. In other words, the BIC very effectively captures changes in the data. If the BIC difference between the two models for this dataset is greater than ≈50, it already suggests that the corresponding number of components k is indeed the more likely one. In this example, the difference between a BIC value for k and $k+1$ is only slightly smaller which suggests that the differences are relevant and not just "noise". Additionally, taking a look at larger values of k, there is a very clear trend that points to the minimum. In conclusion, our 1D GMM analysis of Young's modulus reveals that both composites have three mechanical phases at 1 µm depth. Table 1 summarizes the results of the Gaussian mixture models evaluated for the four different materials.

As stated above, our GMM analysis of CuCr25 shows the presence of three mechanical phases as well as differences in the properties of nominally identical phases. In CuCr25, for example,

10.3389/fmats.2024.1440608

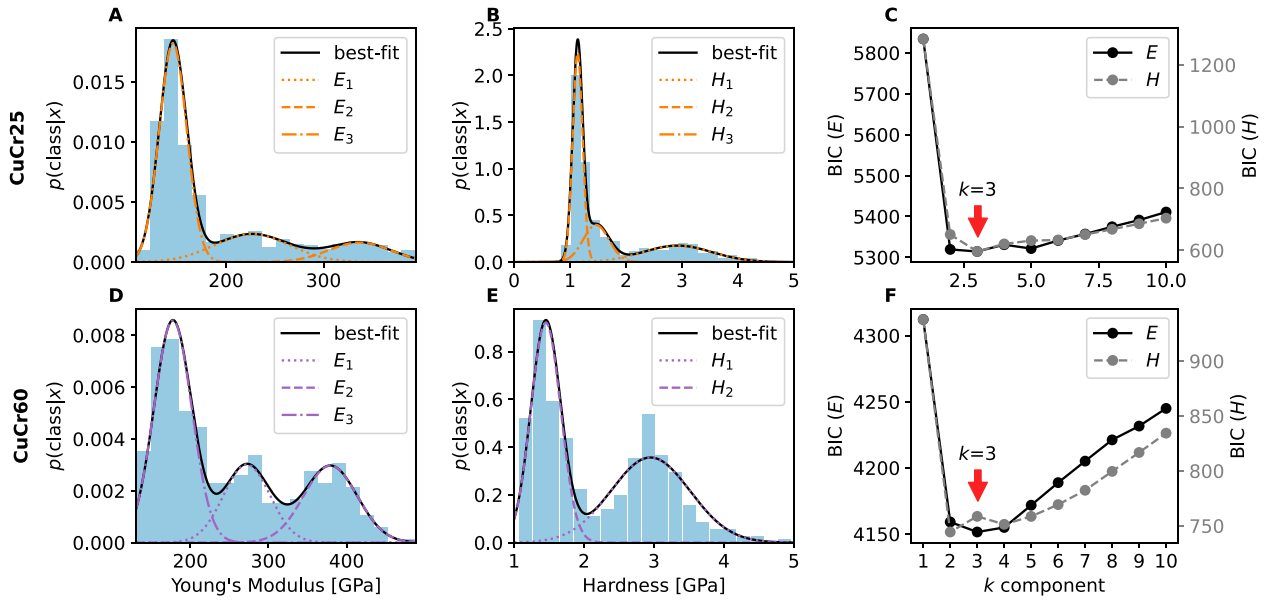

FIGURE 5
1D GMM results of CuCr25 and CuCr60 at 800–1,200 nm indentation depth. CuCr25: **(A)** Histogram of E_i and the best fit (solid line), **(B)** Histogram of H_i and the best fit (solid line), **(C)** BIC of H and E; CuCr60: **(D)** Histogram of E_i and the best fit model, **(E)** Histogram of H_i and the best fit model, **(F)** BIC of H and E.

TABLE 1 1D mechanical property fitting based on the optimal BIC results at 800–1,200 nm depth.

Cr (wt%)	Number of data	Mechanical property	Average (GPa)	Standard deviation	Percentage (%)
0 (pure Cu)	30	E_1	118.80	9.45	100
		H_1	0.91	0.06	100
25 (CuCr25)	513	E_1	145.55	14.12	64.5
		E_2	226.50	38.42	22.6
		E_3	337.02	31.69	12.9
		H_1	1.14	0.09	50.1
		H_2	1.47	0.23	22.6
		H_3	2.92	0.62	27.3
60 (CuCr60)	364	E_1	177.35	24.17	52.2
		E_2	272.17	29.75	22.4
		E_3	379.32	34.34	25.4
		H_1	1.45	0.22	50.0
		H_2	2.96	0.56	50.0
100 (pure Cr)	31	E_1	371.24	10.54	100
		H_1	3.21	0.10	100

the Cu-rich phase accounts for 64.5 vol% (defined by E_1 as the lowest value, closest to pure Cu), while the combined volumes of E_2 and E_3 amount to 35.5 vol%. These results indeed coincide well with the 35.5 vol% Cr estimated by optical microscopy (Bos, 2019). The lowest modulus value for the Cu phase was found in the compacted Cu sample (i.e., E_1), followed by E_1 of CuCr25 and CuCr60. The highest modulus of the Cr phase was found for CuCr60 (i.e., E_3) and the compacted pure Cr sample (i.e., E_1 in Table 1) almost reaching the modulus value of pure Cr.

These differences in the mechanical properties fitted by GMM are likely related to the diffusion of one phase into the other, the presence of foreign particles or pores, or the influence of the surrounding material. Assuming that a small amount of Cr, i.e., 0.4 at% - 3 at%, can be dissolved in the Cu matrix (Jacob et al., 2000; Chakrabarti and Laughlin, 1984), the Cr solid solution likely contributes to the difference of the modulus of the Cu phase in the composite samples. In addition, the presence of Cr nanoparticles of 100–200 nm in size was reported (von Klinski-Berger, 2015), which as well results in a higher Young's modulus value of the Cu phase in CuCr25 and CuCr60. Finally, considering the relative densities of 99.4% and 98.3% for CuCr25 and CuCr60, respectively, compared to 97.7% for the pure Cu sample, a Young's modulus value with an estimated reduction up to 9% (Lebedev et al., 1995) can be expected. The reduction of the modulus of the Cr phase in CuCr25 (i.e., E_3) can be explained by the surrounding softer Cu phase. While in general also the hardness fitting of CuCr25 supports the presence of three mechanical phases, the phase fractions are different. Again, this is not surprising, since hardness is a local property and strongly related to microstructural details.

We now take a look at the outcomes of the 2D GMM with independent feature variables E and H shown in Figure 6. The distribution of three and four components or mechanical phases of CuCr25 and CuCr60, respectively, are shown in Figure 6A-Band Figure 6D-E. The red points in the graphs represent the average value of each component, while the concentric ellipses with different orientations (covariance) represent the different components. The model selection criteria BIC are shown as a function of the number of components in the right panel (Figure 6C and Figure 5F).

Based on Figure 6C, one could say that four mechanical phases are the ideal match for the CuCr25 composite. Given the anomaly in the upper left corner of Figure 6B, though, the best assumption remains at three, which will be further discussed in Section 3.4. As shown in Table 2, in the 2D GMM, which combines both E and H, the estimated amount of Cr in CuCr25 was 38.9 vol%, which is close to the actual experimental findings (Bos, 2019).

The 2D GMM analysis (in Figure 5F) also reveals that CuCr60 contains three mechanical phases in the depth range of 800–1,200 nm. The fitted result for the volume of Cr in 1D is 47.8 vol% based on the modulus values, which is less than the nominal value (i.e., 65 vol% Cr). By contrast, the 2D GMM result indicates a Cr volume fraction of 56.6 vol% (Table 2). The difference between the 1D GMM (E) and 2D GMM results is related to H, while the difference between the fitting results and the experimental data is due to the amount of data and variation of microstructures over the samples. Note that the number of datapoints for CuCr60 is by 40% less than that for CuCr25; an insufficient amount of data can be the source of inaccuracies, which we are addressing in the following.

In the above analysis of CuCr60, the size of the dataset was 364, collected over an area of 500 × 500 μm^2. To increase the size of the dataset, we merged it with two more nanoindentation areas (100 × 100 μm^2 and 300 × 300 μm^2) and analyzed the data using the procedures described above. The results are summarized in Tables 3, 4. Merging the three datasets now includes indentation depths ranging from 500 nm to almost 2,000 nm with 97.8% of the data lying between 800–1,200 nm indentation depth. As shown in Figure 7A, the distribution ranges of the data are congruent indicating that the microstructure of the material was comparable over the different areas indented.

As shown in Figure 7B, the most likely number of phases determined by analyzing Young's Modulus is three, while two phases are most probable when analyzing the hardness. The 2D fit, though, also indicates $k = 3$, as can be seen in Figure 7C. The individual 2D fits for the different areas of 300 × 300 μm^2 and 500 × 500 μm^2, each reflect three mechanical phases with 67.1 vol% Cr and 59.5 vol% Cr, respectively. Not unexpectedly, the merged data then yielded a Cr fraction of 61.5 vol%. The somewhat different results reflect the variation of the microstructure over the sample surface and underscores the importance of identifying the characteristic microstructures or increasing the size of the dataset. How should one decide if this is a two or three phase system? From the point of materials science, one would probably choose only 2 phases, however, the analysis showed, that the properties of the third "phase" are rather distinct and can be recognized as a separate one.

3.4 Cross-validation of GMM

The amount of required training data has been an often encountered concern in clustering algorithms for machine learning. In contrast to the model selection criteria used in the GMM with BIC, we will now focus on the robustness and validity of the clustering results as a function of the size of the dataset. Clustering together with cross validation is used to evaluate the effect of data size on the results and to identify the amount of experimental data required to achieve the same level of performance. We applied the following procedure:

1. Given the whole dataset $\mathcal{D}$, the categorical variable y_i is generated by a clustering algorithm A_k. It constructs a solution $\mathbf{Y} \coloneqq A_k(\mathcal{D})$, where $\mathbf{Y} = \{y_1, \ldots, y_i\}$, and $y_i = \{1, \ldots, k\}$ is assigned to the cluster. In our case, labels y_i were generated using the KMeans and GMM clustering algorithms, with the prediction of the optimal GMM algorithm serving as the ground truth.
2. k-fold cross-validation is conducted on the dataset $\mathcal{D}$: First, divide the data set into k equal parts (the "folds"), then choose $(k-1)$ folds for training and the remaining fold for testing. Then, validate the model using the testing data set $\mathcal{D}^{\text{test}}$ after training the model with the training data set $\mathcal{D}^{\text{train}}$. Perform k rounds of cross-validation using multiple training datasets. Note, that k indicates the number of folds and *not* the number of clusters as above!
3. The adjusted Rand index (ARI) is used to represent the performance level of the clustering algorithms. In general, the ARI approaches 0.0 for random clustering, reaches 1.0

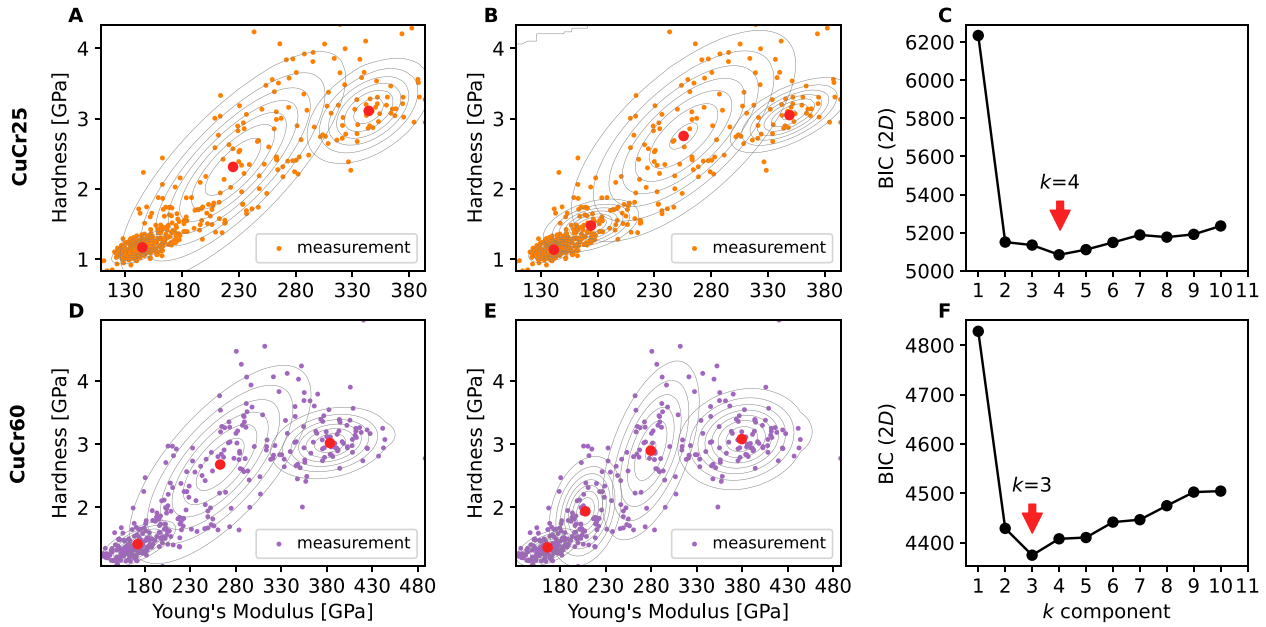

FIGURE 6
2D Gaussian mixture model clustering of CuCr25 and CuCr60. CuCr25: **(A)** Three components. **(B)** Four components and **(C)** 2D BIC; CuCr60: **(D)** Three components **(E)** Four components and **(F)** 2D BIC. The ellipses in **A**, **B**, **D**, and **E** are isolines of the Gaussian distributions and the red points represent the average values of the different components.

TABLE 2 2D mechanical property fitting based on the optimal BIC results at 800–1,200 nm depth.

Cr (wt%)	Number of data	Average *E* (GPa)	Average *H* (GPa)	Percentage (%)
25 (CuCr25)	513	144.91	1.17	61.1
		220.92	2.25	27.1
		340.52	3.18	11.8
60 (CuCr60)	364	172.32	1.42	43.4
		262.63	2.67	34.6
		383.35	3.02	22.0

for identical clusterings, but can go as low as −0.5 for highly discordant clusterings (Chacón and Rastrojo, 2023).

Figure 8A illustrates the outcomes of *k*-fold cross-validation conducted on 364 sets of CuCr60 data (500 × 500 μm^2) using two distinct clustering techniques, i.e., KMeans and GMM. The GMM algorithm consistently outperformed KMeans, as evidenced by the results. Figure 8B presents the results of *k*-fold cross-validation using the GMM algorithm across varying data sizes for CuCr60.

For the CuCr60 composite, a total of 571 datapoints were collected by merging data from three different indentation areas. We developed an approach to assess varying sizes of datasets, starting with taking 50 data records as the dataset of interest, and increasing by increments of 50 until reaching a dataset containing 550 data points. During the process of training a model, a fraction of $(k-1)/k$ of the entire datadataset was used as training data $\mathcal{D}^{\text{train}}$. For example, for 5-fold cross validation ($k = 5$), this entailed using 40, 80, 120, … 440 datapoints for training, accompanied by 10, 20, 30, … 60 datapoints for testing the model. During *k*-fold cross-validation, unique selections of training data were made, resulting in a total of five iterations for each fixed amount of datapoints. This iterative approach was adopted to gain insights into model performance concerning various data quantities, while minimizing any inaccuracies that may arise due to sampling bias.

Overall, our analysis indicates that the GMM algorithm exhibits superior accuracy when the average model performance approaches 0.9. Nevertheless, in the pursuit of enhanced generalizability, we aim to minimize errors further, ideally pushing the lower boundary of performance beyond 0.9. To achieve this, it is recommended to initiate the training process using a dataset comprising more

10.3389/fmats.2024.1440608

TABLE 3 1D mechanical property fitting of CuCr60 based on the optimal BIC results at 500 nm–2000 nm depth encompassing data of three different arrays.

Indentation area (μm^2)	Number of data	Mechanical property	Average (GPa)	Standard deviation (GPa)	Percentage (%)
100 × 100	23	E_9	—	—	—
		H_1	1.83	0.21	39.1
		H_2	3.15	0.29	60.9
300 × 300	154	E_1	157.79	12.02	29.2
		E_2	280.30	70.47	70.8
		H_1	1.42	0.18	37.3
		H_2	3.02	0.74	62.7
500 × 500	394	E_1	177.86	24.92	49.5
		E_2	275.26	31.88	24.9
		E_3	378.94	35.93	24.6
		H_1	1.44	0.22	45.5
		H_2	2.95	0.62	54.5
merged data	571	E_1	175.12	24.10	46.0
		E_2	269.55	34.10	28.5
		E_3	373.37	37.46	25.5
		H_1	1.45	0.22	42.8
		H_2	2.98	0.65	57.2

TABLE 4 2D mechanical property fitting of CuCr60 based on the optimal BIC results at 500 nm–2000 nm depth encompassing data of three different arrays.

Indentation area (μm^2)	Average E (GPa)	Average H (GPa)	Percentage (%)
100 × 100	252.19	2.46	78.4
	393.06	3.27	21.6
300 × 300	159.47	1.39	32.8
	229.10	2.29	31.9
	337.91	3.51	35.2
500 × 500	172.25	1.41	40.5
	267.99	2.74	38.1
	383.12	3.01	21.4
merged data	170.49	1.42	38.5
	264.73	2.78	40.4
	379.28	3.70	21.1

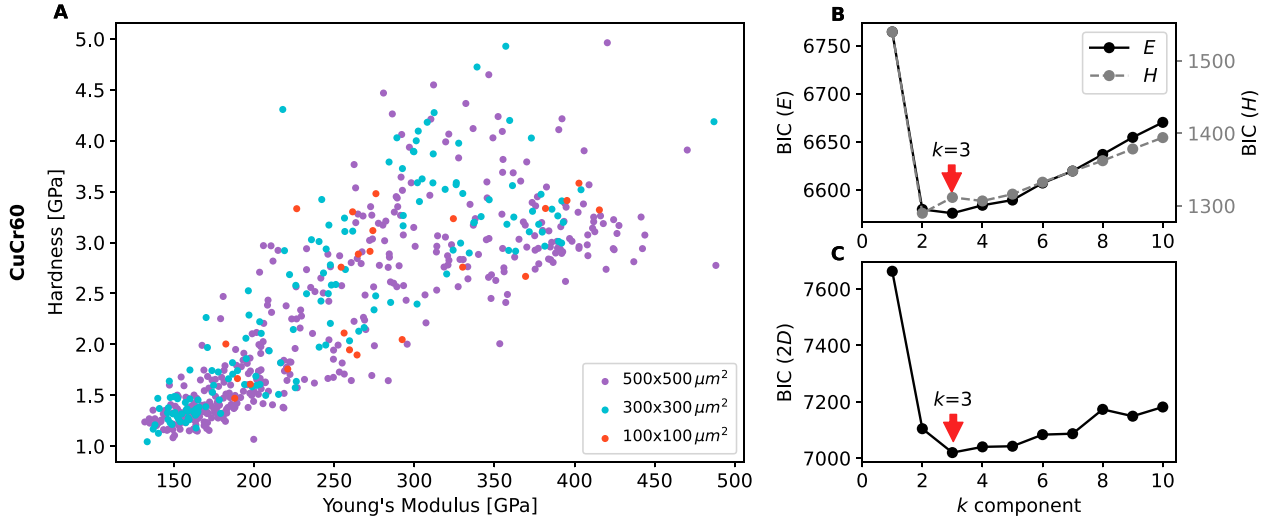

FIGURE 7
(A) Distribution of the CuCr60 data over the depth range of 500–2,000 nm (combined datasets for indentation arrays of 100 × 100 μm^2, 300 × 300 μm^2, and 500 × 500 μm^2) **(B)** 1D BIC for *E* and *H*, and **(C)** 2D BIC for *E* and *H*.

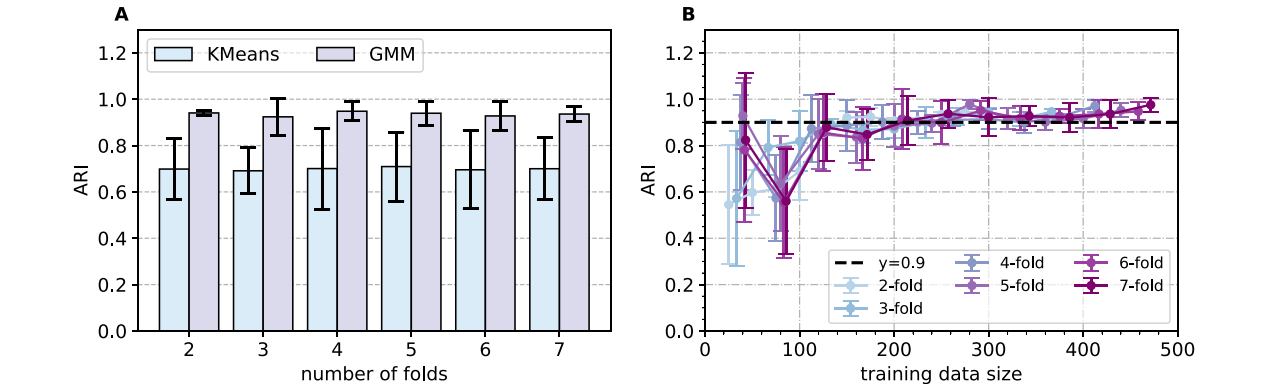

FIGURE 8
k-fold cross validation results of CuCr60. **(A)** the performance of two unsupervised models, KMeans and GMM, measured by the ARI value and as a function of the number of folds of the *k*-fold cross-validation. **(B)** ARI values of *k*-fold cross-validation using GMM with varying the amount of training data.

than 400 datapoints as a starting point, as it appears to yield more reliable results.

The identical approach for sampling was employed in the examination of CuCr25. Figure 9 shows the outcomes of *k*-fold cross validation of CuCr25 with different numbers of clusters. Based on the 2D GMM analysis described above, the optimal model contained three clusters. However, we also notice that the BIC values determined for four and five components are comparable to those determined for three (Figure 6C). When the BIC values do not indicate a significant difference, the *k*-fold cross-validation can certainly provide a hint as to which model is superior. In the case of three clusters, the ARI values increased with the amount of data, whereas with other numbers of clusters the values did not exhibit an upward trend in conjunction with predicted values lower than 0.8. As a supplement to BIC, the *k*-fold cross-validation method can be used to determine how robust a certain number of clusters is and, more importantly, whether the amount of data is sufficient for training.

3.5 Effect of indentation depth

Figure 10A shows the 2D mapping of *E* for the CuCr25 dataset containing indents in the depth range between 200 and 600 nm. White color indicates positions for which no nanoindentation data was collected, blue represents the Cu-rich zones, and red represents the Cr-rich zones with color gradients in the vicinity of phase boundaries. Figure 10B shows the 1D Gaussian mixture model fitting results for Young's modulus indicating that the microstructure primarily consists of three mechanical phases, with the third

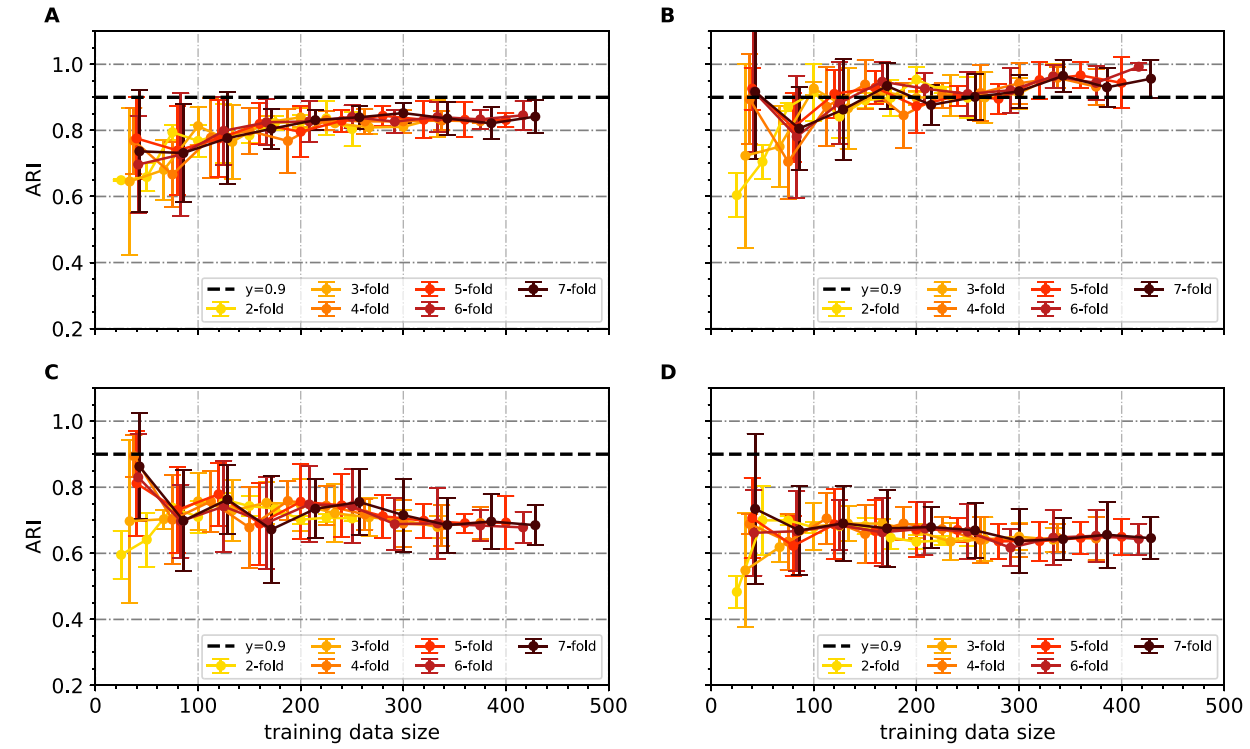

FIGURE 9
k-fold cross validation results of CuCr25 with different numbers of clusters for the GMM model and data size. **(A)** two component **(B)** three component **(C)** four component **(D)** five component. A total of 513 datapoints were collected for the CuCr25 at 1 μm depth. The proportion of datapoints used as training data in a *k*-fold cross-validation is determined by the value of *k*.

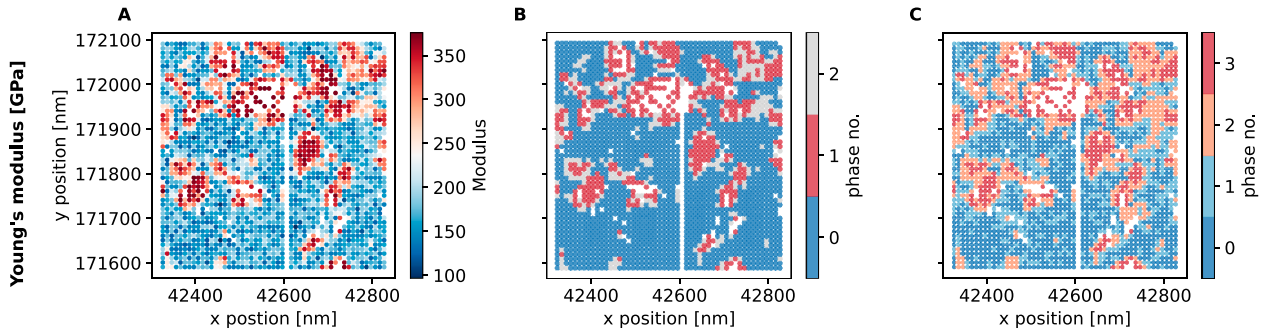

FIGURE 10
Elastic modulus of CuCr25 based on idents in the depth range of 200–600 nm. **(A)** Young's modulus distribution over the testing area. **(B)** Young's modulus distribution of 1D GMM clustering. **(C)** 2D GMM clustering results ($k = 4$).

phase mainly located at the intersection of the Cu-rich and Cr-rich regions. Figure 10C displays the 2D Gaussian mixture model's fitted results, with an additional fourth term.

The clustering results assign fractions of the Cu-phase at different depths to the fourth mechanical phase. Therefore, datasets containing data at different depths, need to be evaluated in subsets when determining the number of mechanical phases. Accordingly, we further examined the phase composition using a 1D Gaussian mixing model for different depth ranges covering 100 nm each. The results are shown in Table 5. Based on this analysis, the occurrence of additional phases can be observed at higher depths, i.e., in the 400 nm and 500 nm regimes, along with a significant reduction in Cu content.

The hardness distribution for the CuCr25 and CuCr60 composites at depths ranging from 200 nm to 1200 nm is shown in Figure 11. The data was grouped in 50 bins of equal size with the shaded area indicating the standard deviation. For metals typically an indentation size effect is observed, i.e., the hardness increases with decreasing indentation depth (Wang et al., 2021; Ma et al., 2021; Pharr et al., 2010). This indentation size effect is

TABLE 5 Indentation results of CuCr25 and CuCr60 composites in the depth range 200 nm–600 nm determined by 2D GMM.

Displacement into Surface (nm)	CuCr25				CuCr60			
	Young's Modulus	Average (GPa)	Standard Deviation	Percentage (%)	Young's Modulus	Average (GPa)	Standard Deviation	Percentage (%)
200–300 nm	E_1	163.72	23.81	95.2	E_1	162.10	27.37	76.2
	E_2	272.26	46.67	4.8	E_2	274.61	38.95	23.8
300–400 nm	E_1	163.55	24.29	85.8	E_1	161.90	29.28	34.1
	E_2	290.88	38.88	14.2	E_2	296.06	36.07	65.9
400–500 nm	E_1	158.35	19.65	13.4	E_1	182.94	36.95	14.7
	E_2	249.79	25.65	22.4	E_2	275.25	24.17	47.8
	E_3	303.13	18.81	32.9	E_3	326.85	21.85	37.5
	E_4	351.83	15.34	31.3				
500–600 nm	E_1	176.90	21.06	50.7	E_1	198.21	31.98	69.7
	E_2	242.53	30.22	25.2	E_2	307.28	36.63	30.3
	E_3	341.30	24.67	24.1				

10.3389/fmats.2024.1440608

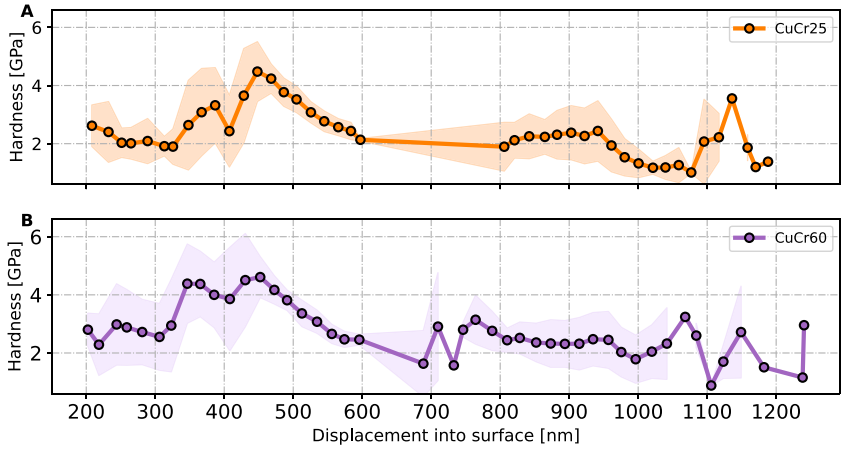

FIGURE 11
Hardness of CuCr25 and CuCr60 composites at different indentation depths. **(A)** CuCr25, with 3,812 data points in the range 200–600 nm, and **(B)** CuCr60, with 3,254 data points in the range 200–600 nm depth. The mean of the hardness with the respective standard deviation as shaded area is shown. The mean hardness assumes an approximately constant value for depths greater than 600 nm.

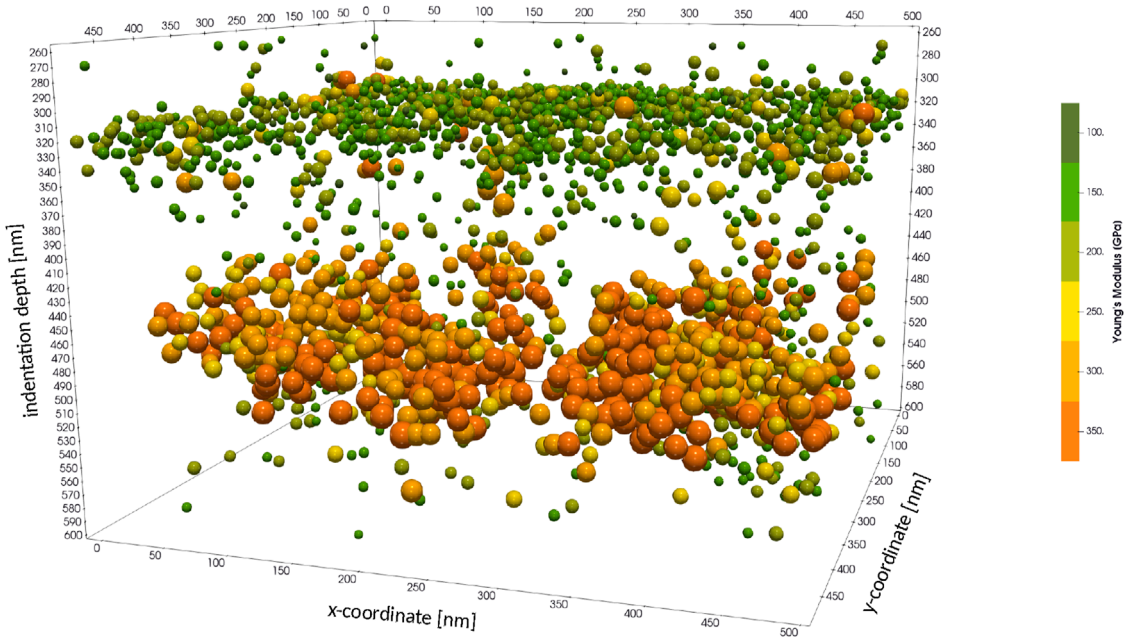

FIGURE 12
Young's modulus at different locations *x* and *y* and for different indentation depths (vertical axis). The colors (as well as the diameter) of the spherical marker denote the values of Young's modulus.

observed here for indentation depths between 450 and 600 nm, while for lower depths it is obscured by the combined effects of microstructural changes over the depth as well as the presence of at least two phases. This apparent nanoscale softening may also be influenced by the distance between indents relative to the size of the different phases possibly favoring a particular property value.

Figure 12 depicts the distribution of Young's modulus in CuCr25 in the range between 250–600 nm, whose 2D top view is depicted in Figure 10A). The distribution of Young's modulus exhibits a certain degree of continuity, and the majority of the changes are gradual. The green hue represents the Cu-phase, while the orange color represents the Cr-phase. Observation reveals that the yellow and light-green portions (the third mechanical phase) can be considered the Cu and Cr intersection region. Figure 12 depicts the marker size adjusted according to the indent size on the surface, which increases as the indentation depth increases. Based on this observation and also on the analysis results listed in Table 5, we observe an increase of the Cr-phase in both the CuCr25 and the CuCr60 composites in the range

of 200–500 nm, which can explain the increase in hardness in the low-depth regimes (cf. Figure 11). Thus, it is critical to identify the characteristic microstructures not only on the surface tested but also in the sub-surface region of the material.

4 Conclusion

Cu-Cr composites were studied by indentation as a model material to evaluate the ability to determine the properties of individual phases as well as the number of phases present. 1D and 2D Gaussian mixture models were trained and the most likely number of components identified based on the BIC. Using cross-validation we showed that the GMM gave more accurate results than KMeans clustering. Investigating the dependence of the cross-validation results on the size of the datasets helped to understand what a reasonable amount of data for the training of such models might be. Our analysis revealed that 450 data-pairs were sufficient for accurate phase volume prediction.

Clearly, the presented GMM-based method has limitations, e.g., it implicitly assumes that indents into different phases also result in different properties. As a consequence, it must be expected that the predictions are getting worse the more similar two (or more) phases become. GMM is a simple yet robust method, which is not able to utilize all available microstructural information. For example, locality effects due to the fact that nearby indents may be correlated cannot be captured. Extending GMM, e.g., in terms of engineered features, could be a possible solution.

While large datasets produced in a very short time are quite impressive, also conventional nanoindentation approaches with typically fewer data can be used. Variations of the phase composition over the depth as well as the distance between indents relative to the size of the microstructural features may favor a particular phase. To avoid such bias indentation depth and characteristic microstructures need to be critically evaluated.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://zenodo.org/doi/10.5281/zenodo.8336071

Author contributions

CZ: Formal Analysis, Investigation, Software, Visualization, Writing–original draft, Data curation, Methodology. CB: Data curation, Investigation, Writing–review and editing. SS: Conceptualization, Supervision, Validation, Writing–original draft, Writing–review and editing, Formal Analysis, Resources. RS: Conceptualization, Supervision, Validation, Writing–original draft, Formal Analysis, Resources, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. StS gratefully acknowledges funding through the Helmholtz Foundation Model Initiative of the project “SOL-AI”.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmats.2024.1440608/full#supplementary-material

References

Becker, B. R., Hintsala, E. D., Stadnick, B., Hangen, U. D., and Stauffer, D. D. (2022). Automated analysis method for high throughput nanoindentation data with quantitative uncertainty. *J. Appl. Phys.* 132, 185101. doi:10.1063/5.0098493

Besharatloo, H., and Wheeler, J. M. (2021). Influence of indentation size and spacing on statistical phase analysis via high-speed nanoindentation mapping of metal alloys. *J. Mater. Res.* 36, 2198–2212. doi:10.1557/s43578-021-00214-5

Bishop, C. M., and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, 4. Springer.

Bos, C. (2019). *Micromechanical characterization of heterogeneous materials, statistical analysis of nanoindentation data*. Karlsruhe: Karlsruhe Institute of Technology (KIT). doi:10.5445/IR/1000098226

Chacón, J. E., and Rastrojo, A. I. (2023). Minimum adjusted rand index for two clusterings of a given size. *Adv. Data Analysis Classif.* 17, 125–133. doi:10.1007/s11634-022-00491-w

Chakrabarti, D. J., and Laughlin, D. E. (1984). The cr-cu (chromium-copper) system. *Bull. Alloy Phase Diagrams* 5, 245–268. doi:10.1007/bf02868543

Constantinides, G., Chandran, K. R., Ulm, F.-J., and Van Vliet, K. (2006). Grid indentation analysis of composite microstructure and mechanics: principles and validation. *Mater. Sci. Eng. A* 430, 189–202. doi:10.1016/j.msea.2006.05.125

De Backer, A., Martinez, G. T., Rosenauer, A., and Van Aert, S. (2013). Atom counting in haadf stem using a statistical model-based approach: methodology, possibilities, and inherent limitations. *Ultramicroscopy* 134, 23–33. doi:10.1016/j.ultramic.2013.05.003

Gideon, S. (1978). Estimating the dimension of a model. *Ann. Statistics* 6, 461. doi:10.1214/aos/1176344136

Golovin, Y. (2021). Nanoindentation and mechanical properties of materials at submicro- and nanoscale levels: recent results and achievements. *Phys. Solid State* 63, 1–41. doi:10.1134/s1063783421010108

Haušild, P., Materna, A., Kocmanová, L., and Matějíček, J. (2016). Determination of the individual phase properties from the measured grid indentation data. *J. Mater. Res.* 31, 3538–3548. doi:10.1557/jmr.2016.375

Hintsala, E., Hangen, U., and Stauffer, D. (2018). High-throughput nanoindentation for statistical and spatial property determination. *JOM* 70, 494–503. doi:10.1007/s11837-018-2752-0

Jacob, K., Priya, S., and Waseda, Y. (2000). A thermodynamic study of liquid Cu-Cr alloys and metastable liquid immiscibility. *Zeitschrift für Metallkunde/Materials Res. Adv. Tech.* 91, 594–600. doi:10.1515/ijmr-2000-910710

Jentner, R. M., Srivastava, K., Scholl, S., Gallardo-Basile, F.-J., Best, J. P., Kirchlechner, C., et al. (2023). Unsupervised clustering of nanoindentation data for microstructural reconstruction: challenges in phase discrimination. *Materialia* 28, 101750. doi:10.1016/j.mtla.2023.101750

Karimi, K., Salmenjoki, H., Mulewska, K., Kurpaska, L., Kosińska, A., Alava, M. J., et al. (2023). Prediction of steel nanohardness by using graph neural networks on surface polycrystallinity maps. *Scr. Mater.* 234, 115559. doi:10.1016/j.scriptamat.2023.115559

Kossman, S., and Bigerelle, M. (2021). Pop-in identification in nanoindentation curves with deep learning algorithms. *Materials* 14, 7027. doi:10.3390/ma14227027

Lebedev, A., Berunkov, Y., Romanov, A., Kopylov, V., Filonenko, V., and Gryaznov, V. (1995). Softening of the elastic modulus in submicrocrystalline copper. *Mater. Sci. Eng. A* 203, 165–170. doi:10.1016/0921-5093(95)09868-2

Ma, X., Higgins, W., Liang, Z., Zhao, D., Pharr, G. M., and Xie, K. Y. (2021). Exploring the origins of the indentation size effect at submicron scales. *Proc. Natl. Acad. Sci.* 118, e2025657118. doi:10.1073/pnas.2025657118

Nohava, J., Hausild, P., Houdková, S., and Enzl, R. (2012). Comparison of isolated indentation and grid indentation methods for hvof sprayed cermets. *J. Therm. Spray. Tech.* 21, 651–658. doi:10.1007/s11666-012-9733-6

Oliver, W. C., and Pharr, G. M. (1992). An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. *J. Mater. Res.* 7, 1564–1583. doi:10.1557/jmr.1992.1564

Öztuna, D., Elhan, A. H., and Tüccar, E. (2006). Investigation of four different normality tests in terms of type 1 error rate and power under different distributions. *Turkish J. Med. Sci.* 36, 171–176.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.

Pharr, G. M., Herbert, E. G., and Gao, Y. (2010). The indentation size effect: a critical examination of experimental observations and mechanistic interpretations. *Annu. Rev. Mater. Res.* 40, 271–292. doi:10.1146/annurev-matsci-070909-104456

Prakash, A., and Sandfeld, S. (2022). Automated analysis of continuum fields from atomistic simulations using statistical machine learning. *Adv. Eng. Mater.* 24, 2200574. doi:10.1002/adem.202200574

Puchi-Cabrera, E. S., Rossi, E., Sansonetti, G., Sebastiani, M., and Bemporad, E. (2023). Machine learning aided nanoindentation: a review of the current state and future perspectives. *Curr. Opin. Solid State Mater. Sci.* 27, 101091. doi:10.1016/j.cossms.2023.101091

Randall, N. X., Vandamme, M., and Ulm, F.-J. (2009). Nanoindentation analysis as a two-dimensional tool for mapping the mechanical properties of complex surfaces. *J. Mater. Res.* 24, 679–690. doi:10.1557/jmr.2009.0149

Reynolds, D. (2009). "Gaussian mixture models," in *Encyclopedia of Biometrics*. Editors S. Z. Li, and A. Jain(Boston, MA: Springer). doi:10.1007/978-0-387-73003-5_196

Reynolds, D. A., and Rose, R. C. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Trans. Speech Audio Process.* 3, 72–83. doi:10.1109/89.365379

Sanchez-Camargo, C.-M., Hor, A., Salem, M., and Mabru, C. (2020). A robust method for mechanical characterization of heterogeneous materials by nanoindentation grid analysis. *Mater. and Des.* 194, 108908. doi:10.1016/j.matdes.2020.108908

Shen, Y. (2019). "Nanoindentation for testing material properties," in *Handbook of mechanics of materials*. Editors S. Schmauder, C. Chen, K. Chawla, N. Chawla, W. Chen, and Y. Kagawa (Singapore: Springer), 1981–2012.

Sorelli, L., Constantinides, G., Ulm, F.-J., and Toutlemonde, F. (2008). The nano-mechanical signature of ultra high performance concrete by statistical nanoindentation techniques. *Cem. Concr. Res.* 38, 1447–1456. doi:10.1016/j.cemconres.2008.09.002

Tromas, C., Arnoux, M., and Milhet, X. (2012). Hardness cartography to increase the nanoindentation resolution in heterogeneous materials: application to a ni-based single-crystal superalloy. *Scr. Mater.* 66, 77–80. doi:10.1016/j.scriptamat.2011.09.042

Ulm, F.-J., Vandamme, M., Bobko, C., Alberto Ortega, J., Tai, K., and Ortiz, C. (2007). Statistical indentation techniques for hydrated nanocomposites: concrete, bone, and shale. *J. Am. Ceram. Soc.* 90, 2677–2692. doi:10.1111/j.1551-2916.2007.02012.x

Venderley, J., Mallayya, K., Matty, M., Krogstad, M., Ruff, J., Pleiss, G., et al. (2022). Harnessing interpretable and unsupervised machine learning to address big data from modern x-ray diffraction. *Proc. Natl. Acad. Sci.* 119, e2109665119. doi:10.1073/pnas.2109665119

Veytskin, Y., Tammina, V., Bobko, C., Hartley, P., Clennell, M., Dewhurst, D., et al. (2017). Micromechanical characterization of shales through nanoindentation and energy dispersive x-ray spectrometry. *Geomech. Energy Environ.* 9, 21–35. doi:10.1016/j.gete.2016.10.004

Vignesh, B., Oliver, W., Siva Kumar, G., and Sudharshan Phani, P. (2019). Critical assessment of high speed nanoindentation mapping technique and data deconvolution on thermal barrier coatings. *Mater. Des.* 181, 108084. doi:10.1016/j.matdes.2019.108084

von Klinski-Berger, K. (2015). *Charakterisierung von Kupfer-Chrom-Verbundwerkstoffen für die Schalttechnik*. Technische Universität Darmstadt.

Wang, J., Volz, T., Weygand, S., and Schwaiger, R. (2021). The indentation size effect of single-crystalline tungsten revisited. *J. Mater. Res.* 36, 2166–2175. doi:10.1557/s43578-021-00221-6

Wheeler, J. (2021). Mechanical phase mapping of the taza meteorite using correlated high-speed nanoindentation and edx. *J. Mater. Res.* 36, 94–104. doi:10.1557/s43578-020-00056-7

Yeo, T., Shigematsu, N., and Katori, T. (2023). Dynamically recrystallized grains identified via the application of Gaussian mixture model to ebsd data. *J. Struct. Geol.* 104800. doi:10.1016/j.jsg.2023.104800