

Supplementary Materials for

Interpretable Deep Learning for Atomicity Classification of Platinum Nanoclusters in STEM Images

Author list:

Keizo Tsukamoto,^{1*} Naoyuki Hirata,¹ Masahide Tona,^{1*} and Atsushi Nakajima^{2*}

Author affiliations:

¹ Ayabo Corporation, 1 Hosogute, Fukukama-cho, Anjo, Aichi 446-0052, Japan.

² Department of Chemistry, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan.

Corresponding author:

*Address correspondence to K. Tsukamoto, M. Tona, and A. Nakajima

E-mail: keizo@ayabo.com; masahide@ayabo.com; nakajima.a7@keio.jp

Contents

Note S1. Comparative evaluation of CNNs and conventional machine learning.	S2
Fig. S1. Comparative performance of conventional machine learning and CNN.	S5
Fig. S2. Confusion matrices for conventional machine learning models and CNN.	S6
Table S1. Class-wise silhouette scores from UMAP embeddings on the outer test set.	S7
Fig. S3. Enlarged UMAP views with misclassification-direction markers.	S8
Fig. S4. Stratified Grad-CAM examples across classes.	S9
Fig. S5. Diameter-binned softmax probability profiles.	S11
Fig. S6. Training speed as a function of network depth and configuration.	S12
Fig. S7. UMAP embeddings with dimensionality sweep and silhouette-scores.	S13
Note S2. Analysis of regression-based atomicity estimation.	S14
Fig. S8. Comparison of prediction behavior between classification and regression models.	S16
Fig. S9. Example of regression-model behavior for an unseen atomicity (Class 41).	S17
Fig. S10. Comparison of class-wise Center Compactness Scores (CCS).	S18
Note S3. Selection of the second input channel and filtering conditions.	S19
Table S2. Classification performance of two-channel (2ch) ResNet-18 models using different filtering conditions for the second input channel.	S21
Fig. S11. Representative example of a Class 70 nanocluster (same particle as shown in Fig. 5e in the main text) after different filtering operations.	S22

Note S1: Comparative evaluation of CNNs and conventional machine learning for nanocluster atomicity classification

To rigorously evaluate the utility of convolutional neural networks (CNNs) for nanocluster (NC) atomicity classification, we constructed benchmark models using conventional machine learning methods—logistic regression (LR) and random forests (RF)—and systematically compared their performance.

Conventional machine learning requires predefined feature design and extraction, which provides the advantage of enabling interpretation through the contributions of individual features. However, feature engineering is both time-consuming and operator-dependent, and variations in experimental conditions often cause contrast fluctuations within the same class, resulting in misclassification. Moreover, while Z-score normalization can mitigate intra-class contrast variability, it simultaneously reduces inter-class differences, which in turn diminishes classification accuracy.

Figure 2 of the main text shows that classification becomes difficult when particle diameter alone is used as a feature, as the five classes exhibit substantial overlap. In contrast, the scatter plot of particle diameter (x-axis) versus mean particle intensity (y-axis) in Fig. S1a1 confirms that, although some overlap remains, separability is improved when intensity is included as an additional dimension. Notably, Class 55 and Class 70 each appear to split into two distinct groups. This separation originates from the deliberate variation in pixel dwell time (exposure conditions): while the majority of images, including those of the other classes, were acquired with a pixel dwell time of 10 μ s, a subset of Class 55 and Class 70 clusters was acquired with 18 μ s. The latter consistently exhibited higher mean pixel intensities (≥ 0.25). By incorporating both acquisition conditions into the analysis, we were able to evaluate the generalization capability of the model against variations in imaging parameters and confirm that the improvement in separability was not limited to a single condition.

Based on these results, we designed ten descriptive features to further improve classification performance, which were grouped into four categories: particle size, shape, pixel intensity statistics, and edge characteristics. The features are as follows:

• Size-related features

1. Particle diameter: calculated as (major axis + minor axis)/2, measured in pixels.

• Shape-related features

2. Aspect ratio: ratio of the major to the minor axis obtained by ellipse approximation.

3. Eccentricity: eccentricity of the fitted ellipse.

4. Circularity: defined as $(4\pi \times \text{area})/(\text{perimeter}^2)$.

• Pixel intensity statistics (calculated within the particle region, where $E[\alpha]$ denotes the expectation of α , and X represents the intensity value of each pixel)

5. Mean pixel intensity: $\mu_I = E[X]$ (1)

6. Standard deviation of pixel intensity: $\sigma_I = \sqrt{E[(X - \mu_I)^2]}$. (2)

7. Skewness: measure of asymmetry of pixel intensity distribution:

$$\text{Skew} = \frac{E[(X - \mu_I)^3]}{\sigma_I^3}. \quad (3)$$

8. Kurtosis: measure of sharpness of pixel intensity distribution peak:

$$\text{Kurt} = \frac{E[(X - \mu_I)^4]}{\sigma_I^4} - 3. \quad (4)$$

• Edge-related features (edge ratio detected using the Canny filter)

9. Edge ratio with parameters $\sigma_C=5$, $\text{th}=60, 90$.

10. Edge ratio with parameters $\sigma_C=10$, $\text{th}=30, 45$.

Using these ten features, we trained logistic regression (LR) and random forest (RF) models and evaluated them by five-fold cross-validation (CV). As shown in Fig. S1b, both LR and RF achieved accuracies of approximately 0.9 when trained on the raw (unnormalized) images. The corresponding confusion matrices (Figs. S2a1 and S2a2) indicate overall good performance across classes; however, misclassification remains non-negligible, particularly for Class 55, with notable confusion between Classes 55 and 70.

Permutation importance analysis (Fig. S1c) revealed that mean pixel intensity is the most influential feature for both LR and RF, followed by particle diameter. This is consistent with the observation that incorporating mean intensity improves separability when size

distributions overlap (Fig. S1a1). Notably, the diameter–intensity scatter plot (Fig. S1a1) shows that Classes 55 and 70 split into two apparent groups along the intensity axis, suggesting that systematic contrast shifts associated with imaging conditions (e.g., exposure parameters) can introduce acquisition-dependent cues that may artificially enhance separability (i.e., spurious separability driven by a shortcut feature). To explicitly test robustness against such acquisition-dependent contrast variability, we applied Z-score normalization to all images and repeated the analysis under identical training and evaluation protocols. In the scatter plot of Z-score–normalized intensity (Fig. S1a2), the distinctive two-group structure in Classes 55 and 70 is largely suppressed, whereas overlap between classes increases. In addition, an intra-class trend becomes apparent, where larger clusters exhibit lower normalized intensity while smaller clusters exhibit higher intensity, which may reflect within-class physical variability (e.g., morphology and/or thickness) rather than acquisition settings. Consistent with the increased inter-class overlap, the accuracies of LR and RF decreased to approximately 0.7 after normalization (Fig. S1b, dotted bars), and the dominant contributor in permutation importance shifted from mean intensity to particle diameter (Fig. S1c, dotted bars). The confusion matrices (Figs. S2b1 and S2b2) confirm increased misclassification across all classes. In contrast, the ResNet-18-based CNN (2ch) maintained high classification accuracy after normalization (Fig. S2b3); corresponding to Fig. 3c in the main text), indicating robustness to systematic contrast variations arising from imaging conditions. Taken together, these results support CNN-based classification as a robust framework for NC atomicity assignment across varying acquisition conditions. Compared with conventional feature-based models that can rely strongly on global intensity statistics, the CNN achieves stable performance for both raw images (which may contain acquisition-dependent contrast shifts) and normalized images (which suppress such shifts and emphasize within-class physical characteristics). We note that the normalization experiment serves as a stress test for sensitivity to acquisition-dependent shortcuts, and the preserved CNN performance is consistent with the use of richer spatial/structural cues.

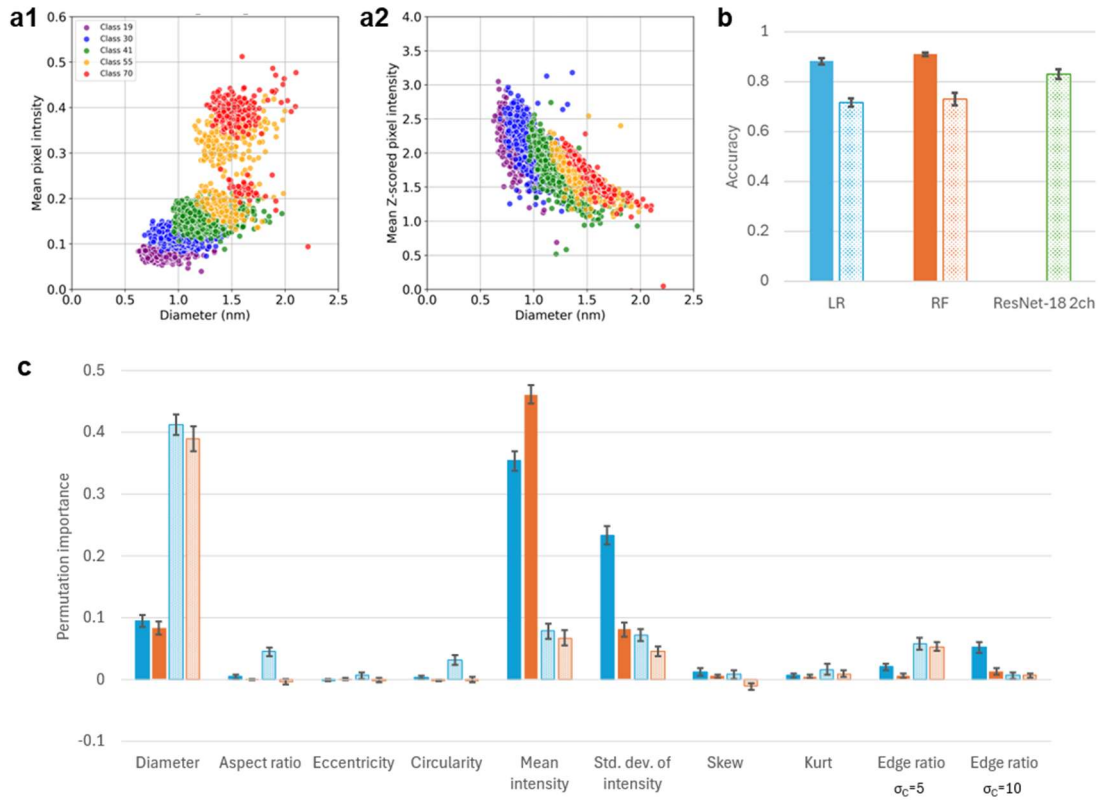


Fig. S1 | Comparative performance of conventional machine learning and CNN-based classification for nanocluster atomicity.

a1 Scatter plot of particle diameter versus mean pixel intensity obtained from raw STEM images. **a2** Scatter plot of particle diameter versus mean pixel intensity after Z-score normalization, highlighting increased overlap among classes. **b** Classification accuracy of logistic regression (LR), random forest (RF), and ResNet-18 2ch (raw+LCN) across five atomicity classes. **c** Permutation importance of ten features used in LR and RF models, showing strong dependence on mean pixel intensity, standard deviation of pixel intensity and diameter, whereas other features contribute less to classification accuracy. In panels **b** and **c**, solid bars indicate results obtained from raw images, whereas dotted bars indicate results from Z-score-normalized images. Error bars represent standard deviations over five-fold cross-validation.

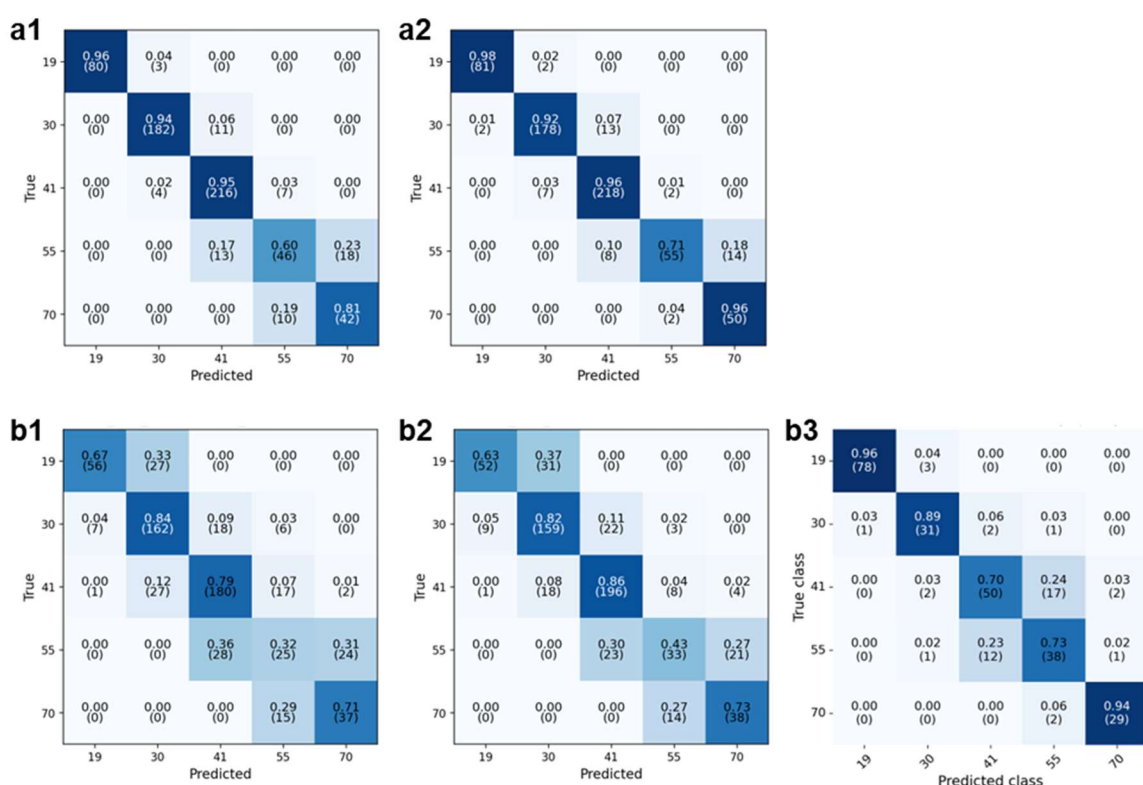


Fig. S2 | Confusion matrices comparing classification performance of conventional machine learning models and CNN.

a1 and **a2** Confusion matrices obtained using raw STEM images. **b1–b3** Confusion matrices obtained using Z-score-normalized images. Columns represent predicted classes and rows represent true classes across five atomicity categories (19, 30, 41, 55, and 70 Pt atoms). **b1** Logistic regression (LR), **b2** random forest (RF), and **b3** ResNet-18 2ch (raw+LCN) models.

Table S1| Class-wise silhouette scores from UMAP embeddings on the outer test set before and after FT. Table S1 compares, on a per-class basis, the silhouette scores (SS) calculated from UMAP embeddings of the outer test set obtained with the LOIO-R3 model and those calculated from UMAP embeddings of the same outer test set after fine-tuning (FT) using representative labels. After FT, SS increased for many classes, suggesting improved intra-class compactness and enhanced inter-class separability in the embedding space. These trends are consistent with the classification accuracy improvement observed after FT and support the notion that FT updated the feature representations in a direction more suitable for class discrimination.

	class					SS (mean $\pm$ std)
	19	30	41	55	70	
Outer test: Baseline (LOIO-R3)	0.76	0.13	0.36	0.26	0.73	0.48 $\pm$ 0.41
Outer test: +FT	0.83	0.37	0.31	0.27	0.80	0.53 $\pm$ 0.39

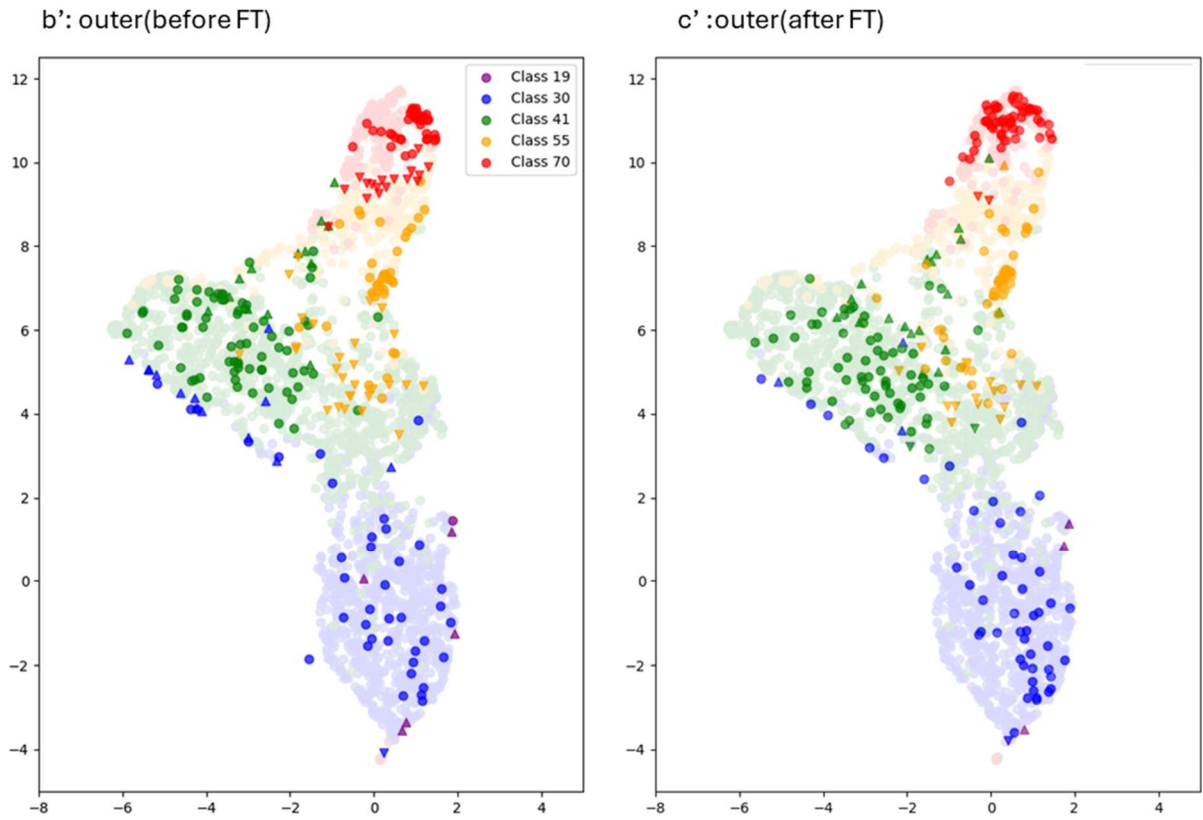


Fig. S3 | Enlarged UMAP views with misclassification-direction markers (single-deposition LOIO outer-test). Magnified views of the LOIO outer-test projections onto the UMAP reference space (trained on inner-CV features; see Fig. 4 in the main text) are shown for before fine-tuning (**b'**) and after fine-tuning (**c'**). The outer-test points are overlaid on the UMAP embedding obtained from the inner test. Point colors indicate the true class (19, 30, 41, 55, 70). Marker shapes encode the relation between the predicted and true classes: ● correct prediction, ▲ predicted larger class than the true class, and ▼ predicted smaller class than the true class. These enlarged views facilitate inspection of how fine-tuning affects class mixing near cluster boundaries (e.g., near the Class 55–70 region), complementing the inset views in Figs. 4**b** and 4**c** (main text).

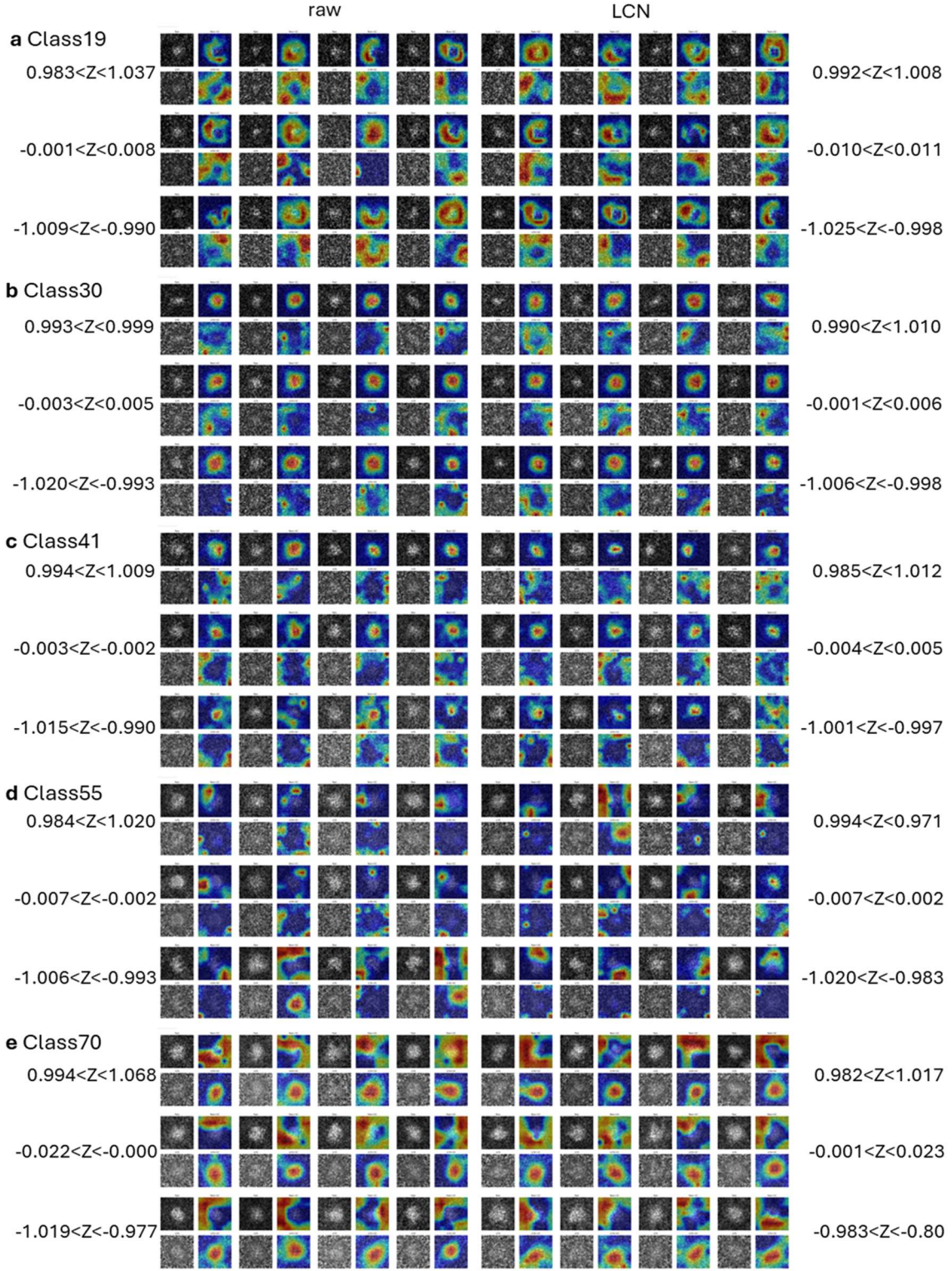


Fig. S4 | Stratified Grad-CAM examples across classes. Grad-CAM visualizations for the 2-channel ResNet-18 classifier (raw + LCN, $\sigma_L=24$) using the LOIO representative model (Round 3). For each class (19, 30, 41, 55, and 70), multiple particle examples are shown.

CONFIDENTIAL

1 For each particle, the top row shows the raw channel (raw image and raw Grad-CAM), and
2 the bottom row shows the LCN channel (LCN image and LCN Grad-CAM). Particles were
3 selected by stratifying the Center Compactness Score (CCS), computed from each channel's
4 Grad-CAM map, into class-wise z-scores and sampling from three groups: near the class
5 mean ($z \approx 0$), near +1 standard deviation ($z \approx +1$), and near -1 standard deviation ($z \approx -1$).
6 From each group, four particles were selected; with two channels displayed per particle, this
7 results in eight panels per class ($4 \text{ particles} \times 2 \text{ channels}$). The z-score ranges of the
8 selected particles are indicated next to each row. If fewer particles satisfied the target z-score
9 window, the closest available particles to the target value (0, +1, or -1) were selected. This
10 montage complements Fig. 5 in the main text by demonstrating that the class-dependent
11 spatial attention trends are consistently observed across multiple particles rather than being
12 driven by a small number of examples.

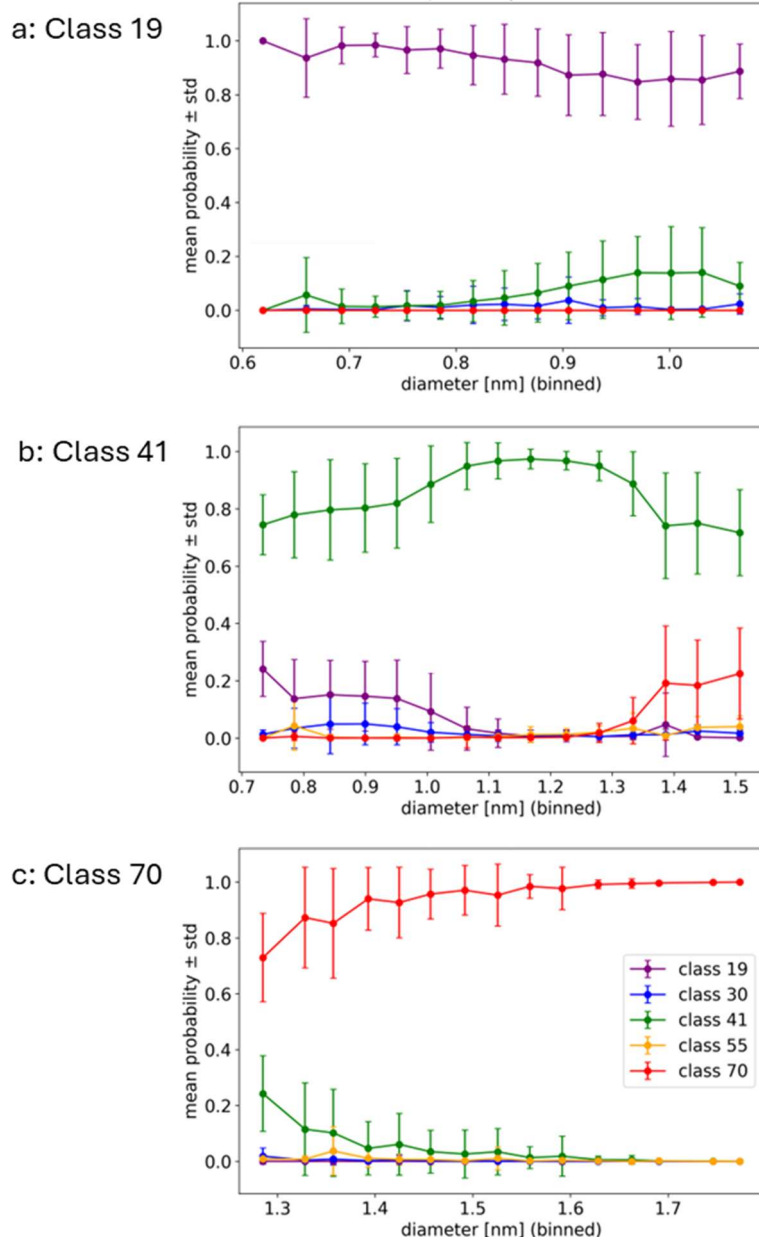


Fig. S5 | Diameter-binned softmax probability profiles indicating continuous transitions between adjacent classes. softmax probabilities (mean $\pm$ variability) are plotted as a function of diameter bins for each class. Near class boundaries, probabilities shift smoothly toward neighboring classes, consistent with structural continuity. (a: Class 19, b: Class 41, and c: Class 70).

To examine class-wise classification behavior, we analyzed the mean transition of the softmax outputs for groups of samples binned by particle diameter. The resulting changes in the predicted probability distributions were consistent with structural continuity, showing smooth transitions primarily between neighboring classes.

Specifically, for Class 19 (Fig. S5a), the predicted probability shifted continuously from Class 19 to Class 41 as particle diameter increased. For Class 41 (Fig. 5b), the probabilities of Classes 19 and 30 increased on the small-diameter side, whereas those of Classes 55 and 70 increased on the large-diameter side, indicating that Class 41 behaves as an intermediate class that continuously connects to adjacent classes on both sides. For Class 70 (Fig. 5c), the probabilities of Classes 41 and 55 increased as particle diameter decreased, confirming that features of neighboring classes tend to emerge in the boundary region.

Taken together, these results suggest that even in diameter-overlap regimes, the model bases its decisions not on particle diameter itself but on particle-specific structural characteristics, yielding physically consistent probability transitions.

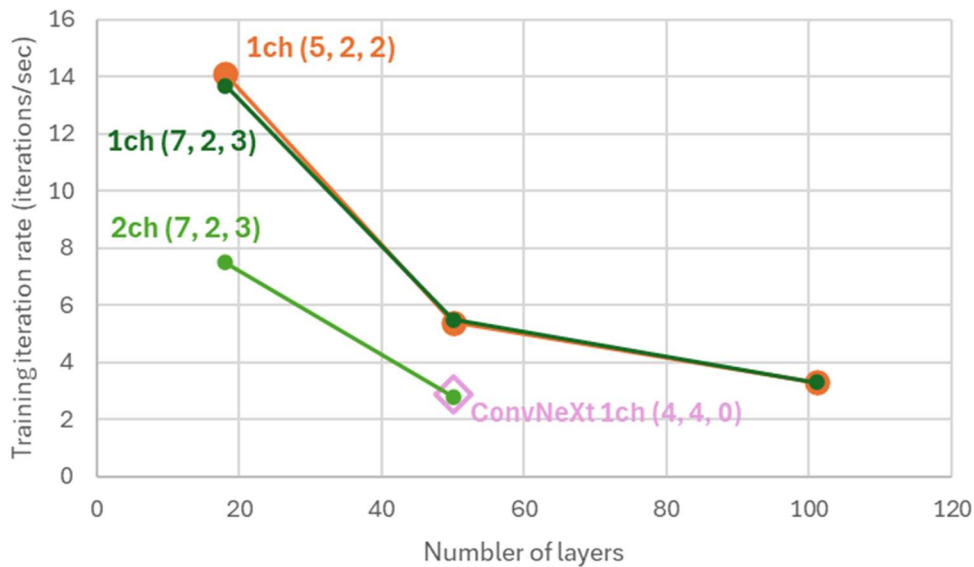


Fig. S6 | Training iteration speed as a function of network depth and configuration. Training throughput (iterations/s) was measured on an RTX 3060 GPU using CUDA event timing (torch.cuda.Event). The label (k, s, p) denotes the convolution kernel size (k), stride (s), and padding (p). Overall, iteration speed decreases with increasing network depth, and models with larger strides tend to run faster. ResNet-18 consists of 18 layers, as indicated by its name.

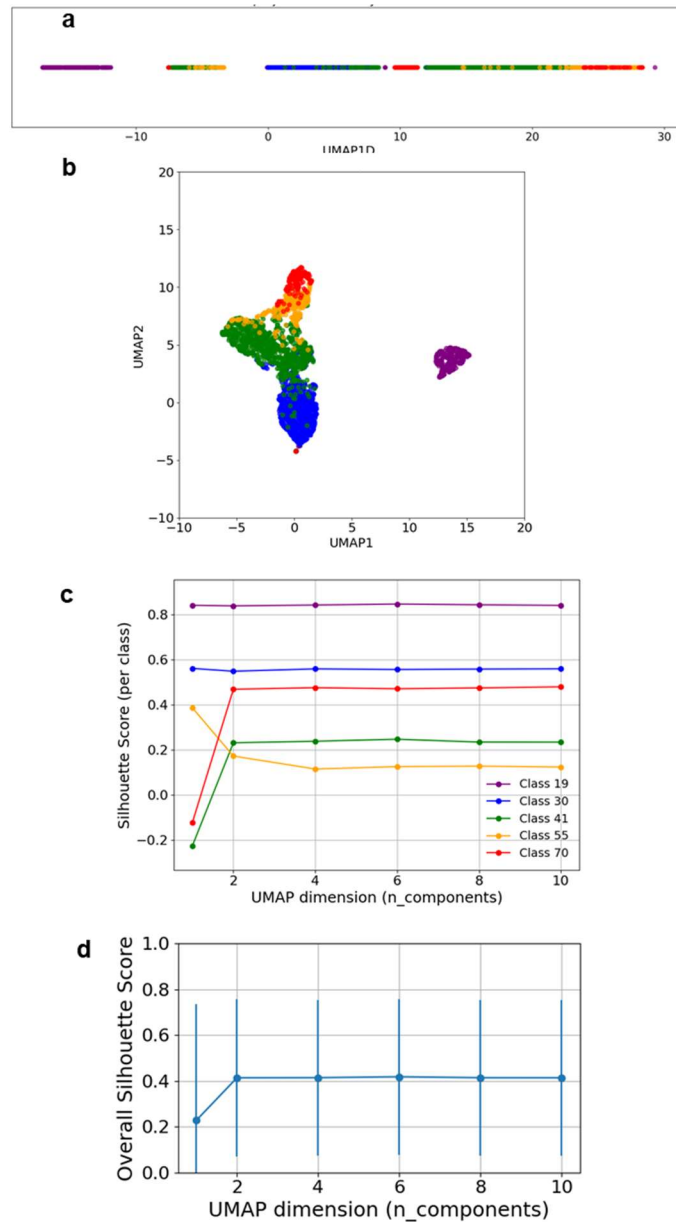


Fig. S7 | UMAP embeddings with dimensionality sweep and silhouette-score evaluation (LOIO-R3, inner test). **a** One-dimensional UMAP projection (n_components = 1, where n_components denotes the dimensionality of the UMAP embedding space). **b** Two-dimensional UMAP projection (n_components = 2), identical to Fig. 4a in the main text. Points are colored by the true class (19, 30, 41, 55, 70). **c** Per-class silhouette scores as a function of the UMAP embedding dimension (n_components). **d** Overall (class-averaged) silhouette score as a function of n_components. The largely stable silhouette scores from 2D to 10D indicate that the cluster separation is not an artifact specific to the 2D visualization, whereas the reduced scores in 1D reflect insufficient embedding capacity to unfold the cluster structure.

Note S2. Analysis of regression-based atomicity estimation.

In the main text, we primarily analyzed Pt nanoclusters with experimentally well-defined atomicities by formulating the task as a classification problem over known atomicity classes. At the same time, atomicity is intrinsically an integer-valued quantity, and a regression-based approach that estimates atomicity as a continuous variable from STEM images represents a potentially viable alternative.

Accordingly, in Note S2 we construct a regression model using the same network architecture and evaluation protocol as the classification model, with the aim of performing a direct comparison and examining its behavior in a supplementary manner. In particular, we focus on (i) how the regression results correspond to the classification outcomes and (ii) how the regression model behaves when applied to an unseen atomicity, based on experimental observations.

The regression model employs the same ResNet-18 architecture as the classification model, with the output of the final fully connected layer replaced by a one-dimensional continuous output that directly predicts atomicity. The input consists of either raw STEM images (1 channel) or combined raw + LCN images (2 channels), and the mean squared error (MSE) loss function is used for training.

Training and evaluation follow the same protocol as for the classification model, consisting of a Leave-One-Image-Out (LOIO) outer-test framework with representative model selection based on five-fold cross-validation within the inner set. In addition, to examine adaptation behavior under domain shifts arising from differences in imaging conditions, lightweight fine-tuning (FT) was applied to the outer-test data in the same manner as for the classification model, and model behavior before and after FT was compared.

Figure S8 compares the confusion matrices obtained from the classification model (2-channel: raw + LCN) with the prediction results of the regression model using the same network configuration. The classification results shown are identical to those in Figs. 3a–c of the main text. For the regression model, the results are summarized using the mean and standard deviation of the predicted atomicity values for each atomicity class.

The width of the prediction distributions produced by the regression model tends to

1 increase for class pairs that are more prone to confusion in the classification confusion
2 matrix (e.g., between atomicities 30–41 and 41–55). That is, although the regression model
3 outputs continuous values, the manner in which the prediction distributions broaden is
4 consistent with the inter-class ambiguity observed in the classification task, indicating that
5 the regression formulation does not eliminate the structural uncertainty reflected in the
6 confusion matrix.

7 Next, to explicitly demonstrate how the regression model behaves for an unseen atomicity,
8 we performed an additional analysis as an illustrative application example (Fig. S9).

9 **S2.4.1 Construction of an inner regression model excluding Class 41 (Fig. S9a)**

10 A regression model was trained after completely excluding all data corresponding to
11 atomicity = 41, yielding an inner model that contains no information about Class 41 during
12 training.

13 **S2.4.2 Regression prediction for unseen Class 41 before fine-tuning (Fig. S9b)**

14 The constructed inner model was applied to Class 41 data that were not used during
15 training, and regression predictions were obtained. The predicted values exhibited a broad
16 distribution centered near 41, with a mean of 43.16 and a standard deviation of 8.49. Thus,
17 for an unseen atomicity, the regression model produced predictions distributed over a finite
18 range rather than converging to a single estimated value.

19 **S2.4.3 Regression prediction after fine-tuning (Fig. S9c)**

20 Next, lightweight fine-tuning was performed using a subset of the Class 41 data. The model
21 was then evaluated on the remaining Class 41 samples that were not used for fine-tuning.
22 As a result, the prediction distribution became clearly narrower and appeared as a finite-
23 width distribution with a mean of 42.12 and a standard deviation of 2.74.

24 These results indicate that fine-tuning does not force the regression predictions to collapse
25 to a single value, but instead acts to narrow the prediction range for an unseen atomicity.

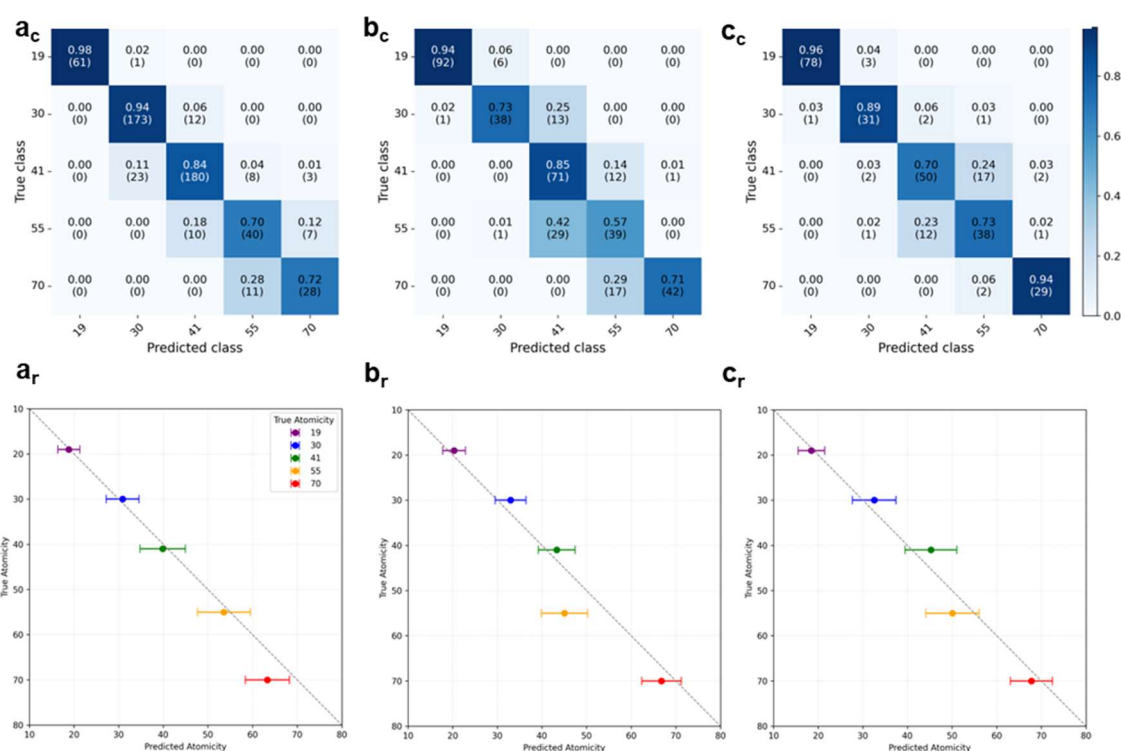


Fig. S8 | Comparison of prediction behavior between classification and regression models. Results obtained using the 2-channel (raw + LCN) model under the Leave-One-Image-Out (LOIO) evaluation framework are shown. The top row (**a_c–c_c**) presents the confusion matrices of the classification model, while the bottom row (**a_r–c_r**) shows the results of the regression model in which only the final layer of the same network architecture was replaced with a continuous-valued output. Regression results are summarized using the mean and standard deviation of the predicted atomicity values for each atomicity class. Panels **a**, **b**, and **c** correspond to the inner-test, outer-test, and post-fine-tuning results, respectively.

CONFIDENTIAL

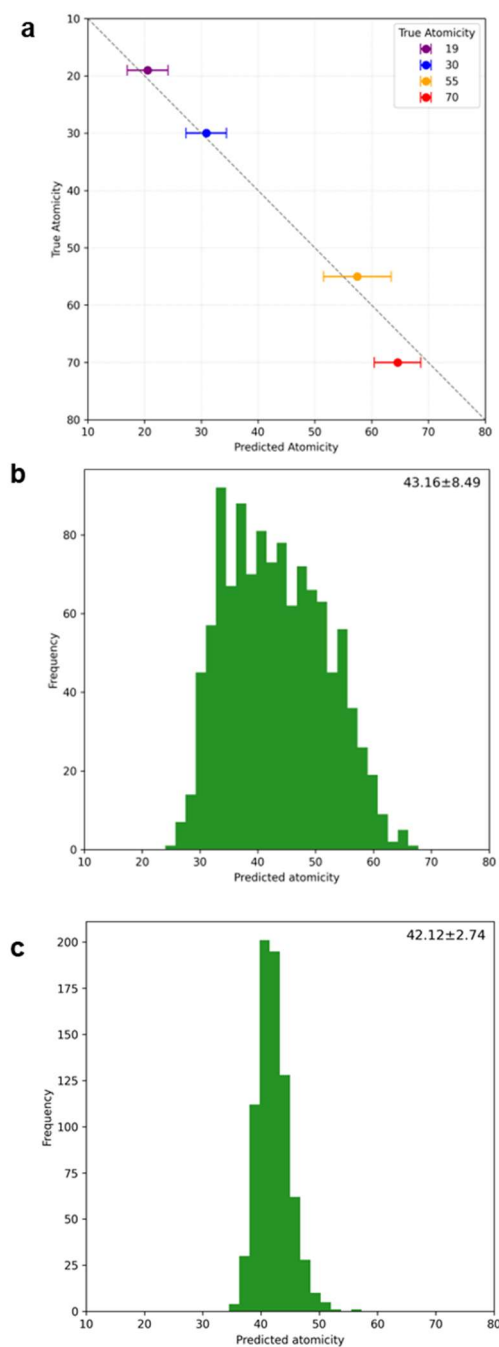


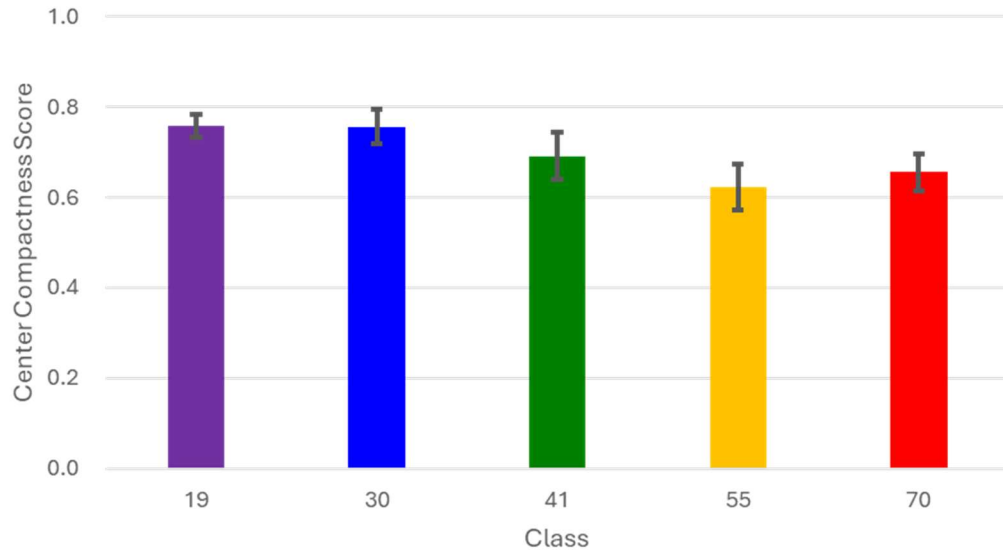
Fig. S9 | Example of regression-model behavior for an unseen atomicity (Class 41). **a** Prediction results of an inner regression model trained with all data corresponding to atomicity = 41 excluded.

b Regression predictions for Class 41 data that were not used during training (before fine-tuning).

c Regression predictions for Class 41 data not used for fine-tuning, after lightweight fine-tuning using a subset of the Class 41 data.

In each panel, the distributions represent histograms of the predicted atomicity values, with the mean and standard deviation indicated.

F



f

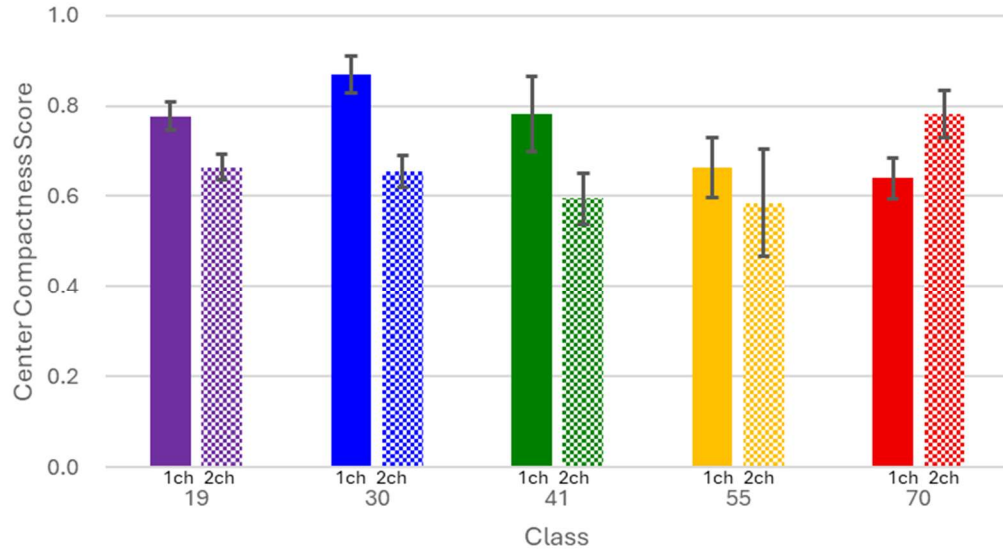


Fig. S10 | Comparison of class-wise Center Compactness Scores (CCS). The top panel (F) shows the class-wise CCS obtained from the 1-channel (1ch) model. The bottom panel (f) compares the CCS values of the 1ch and 2-channel (2ch) models for each class (identical to Fig. 5f in the main text). In the 2ch model, the CCS trends vary with class size, while preserving the spatial tendency observed in the 1ch model corresponding to the raw channel, and the LCN channel provides a distinct spatial distribution. This class-size-dependent behavior suggests a differentiation of the spatial contribution patterns between input channels in the 2ch model. Error bars represent standard deviations.

Note S3. Selection of the second input channel and filtering conditions.

In the main text, we adopted a two-channel (2ch) model consisting of raw STEM images and a Local Contrast Normalization (LCN)-processed channel. When applying this framework to practical applications, the choice of the second input channel is not unique and may depend on design considerations such as robustness, stability, and reproducibility. In Note S3, we therefore present a comparative evaluation of different filtering conditions for the second channel, in order to clarify the motivation for selecting the LCN filter with $\sigma_L = 24$ used in the main text.

During model development, we examined multiple two-channel configurations based on raw images combined with either LCN or Difference-of-Gaussians (DoG) filtering, using several parameter settings for each method. For LCN, the standard deviation of the Gaussian kernel (σ_L) was varied, while for DoG, pairs of Gaussian scales (σ_{D1} , σ_{D2}) were tested. All models shared the same ResNet-18 architecture and were trained and evaluated using identical protocols. A representative visual comparison of the filtered images for the same nanocluster particle is shown in Fig. S11, illustrating how the apparent spatial structure and contrast distribution depend on the filtering method and parameter choice.

Performance was evaluated using the single-deposition dataset by assessing particle-level classification accuracy. In addition to standard 5-fold cross-validation performed on all images, robustness against domain shift was examined using the Leave-One-Image-Out (LOIO) protocol, including inner-test, outer-test performance and performance after lightweight fine-tuning (FT). The results of these comparisons are summarized in Table S2.

Overall, both LCN- and DoG-based two-channel models achieved comparable mean classification accuracies under several parameter settings. However, notable differences were observed in terms of stability across conditions. In particular, the LCN-based models exhibited relatively small performance fluctuations across different σ_L values, whereas the DoG-based models showed a stronger dependence on the chosen parameter pairs, especially in LOIO outer-test evaluations. Among the LCN conditions tested, $\sigma_L = 24$ did not necessarily yield the highest peak accuracy in every metric, but consistently provided balanced performance with comparatively small standard deviations across cross-validation,

1 outer-test, and fine-tuned evaluations. In contrast, while some DoG configurations achieved
2 similar mean accuracies, their performance varied more strongly with parameter selection,
3 making it less straightforward to identify a single robust operating point.

4 Based on these observations, the LCN channel with $\sigma_L = 24$ was selected for the main
5 analyses in this study, prioritizing stability and reproducibility over marginal gains in peak
6 performance. Importantly, this choice should not be interpreted as implying that LCN is
7 universally superior to other filters. Rather, the present comparison illustrates a typical
8 design scenario in which multiple reasonable preprocessing options exist, and the final
9 selection is guided by robustness considerations relevant to experimental deployment. We
10 expect that, depending on imaging conditions, noise characteristics, and specific application
11 goals, alternative filtering strategies or parameter choices may be appropriate. The results
12 provided here are intended to serve as a practical reference for readers who wish to adapt
13 the proposed two-channel framework to their own experimental settings.

14

Table S2 | Classification performance of two-channel (2ch) ResNet-18 models using different filtering conditions for the second input channel. Results are shown for raw+LCN models with varying Gaussian kernel widths (σ_L) and raw+DoG models with different scale pairs (σ_{D1} , σ_{D2}). Performance was evaluated using 5-fold cross-validation, Leave-One-Image-Out (LOIO) inner/outer test evaluation, and fine-tuned test accuracy. Values are reported as mean $\pm$ standard deviation.

Model ResNet-18	All-Images 5-Fold Cross-Validation	Leave-One-Image-Out (LOIO) Evaluation		
	Validation Accuracy (mean $\pm$ std)	LOIO Inner-CV Validation Accuracy (mean $\pm$ std)	Test Accuracy (outer-test, mean $\pm$ std)	Fine-Tuned Test Accuracy (after FT, mean $\pm$ std)
2ch model (raw+LCN)				
$\sigma_L=3.0$	0.82 ± 0.02	0.84 ± 0.01	0.76 ± 0.03	0.78 ± 0.03
$\sigma_L=9.0$	0.85 ± 0.01	0.85 ± 0.00	0.81 ± 0.04	0.82 ± 0.04
$\sigma_L=24$	0.87 ± 0.02	0.87 ± 0.00	0.80 ± 0.04	0.83 ± 0.02
$\sigma_L=36$	0.87 ± 0.02	0.87 ± 0.00	0.78 ± 0.06	0.83 ± 0.05
2ch model (raw+DoG)				
$\sigma_D=0.8, 1.6$	0.84 ± 0.01	0.85 ± 0.01	0.78 ± 0.03	0.78 ± 0.04
$\sigma_D=2.0, 3.0$	0.84 ± 0.03	0.85 ± 0.01	0.80 ± 0.05	0.83 ± 0.03
$\sigma_D=5.0, 8.0$	0.85 ± 0.02	0.87 ± 0.00	0.77 ± 0.04	0.83 ± 0.01
$\sigma_D=10, 15$	0.85 ± 0.02	0.86 ± 0.01	0.78 ± 0.02	0.83 ± 0.01

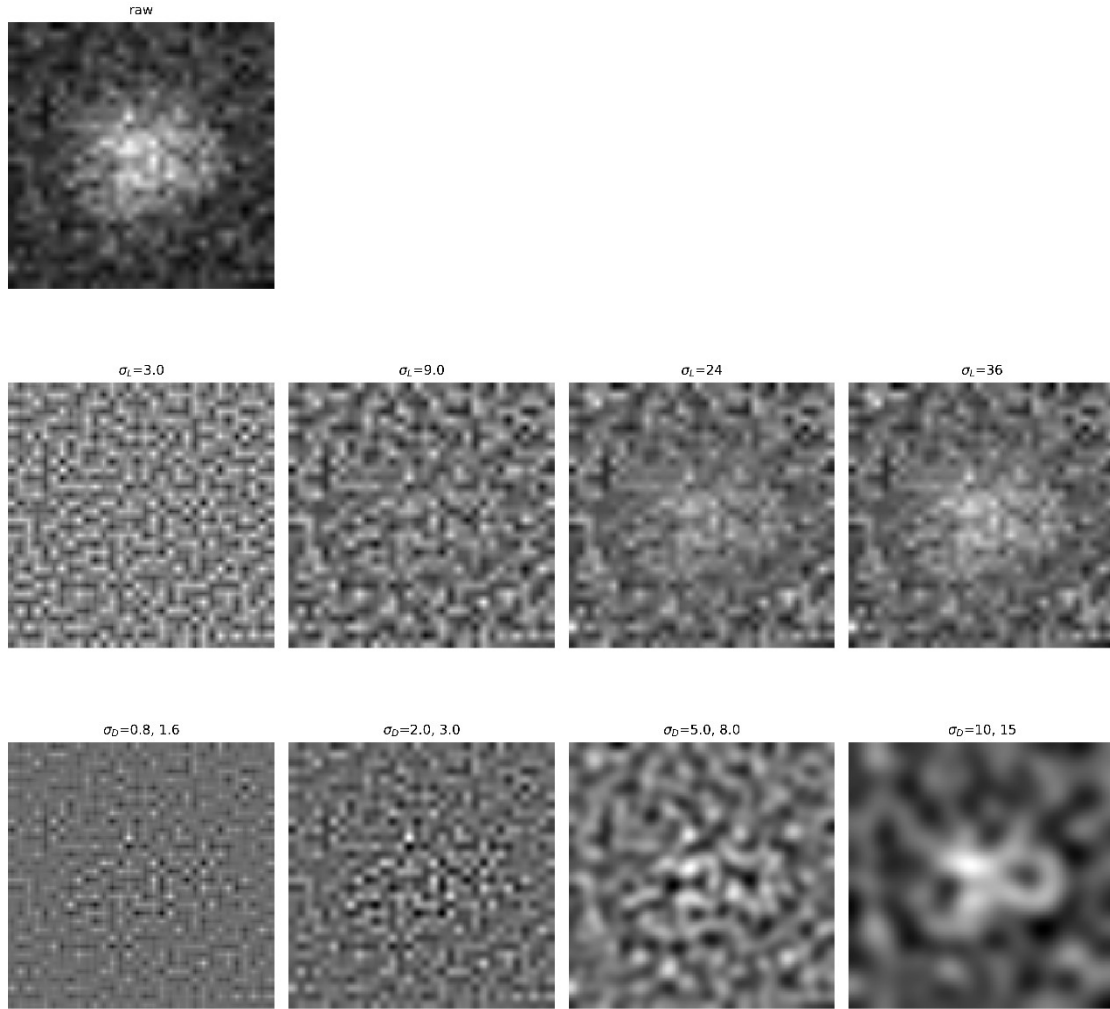


Fig. S11 | Representative example of a Class 70 nanocluster (same particle as shown in Fig. 5e in the main text) after different filtering operations. The top panel shows the raw STEM image. The middle panels show LCN-filtered images with different Gaussian kernel widths (σ_L), and the bottom panels show DoG-filtered images with different scale pairs (σ_{D1} , σ_{D2}). The comparison illustrates how the apparent spatial structure and contrast distribution depend on the filtering method and parameter choice.