Nanoscale

PAPER

View Article Online
View Journal | View Issue

Cite this: *Nanoscale*, 2024, **16**, 13663

Determining the orderliness of carbon materials with nanoparticle imaging and explainable machine learning†

Mikhail Yu. Kurbakov,[a] Valentina V. Sulimova,[a] Andrei V. Kopylov,[a] Oleg S. Seredin,[a] Daniil A. Boiko,[b] Alexey S. Galushko,[b] Vera A. Cherepanova[b] and Valentine P. Ananikov *[b]

Received 5th March 2024,
Accepted 20th June 2024

DOI: 10.1039/d4nr00952e

rsc.li/nanoscale

Carbon materials have paramount importance in various fields of materials science, from electronic devices to industrial catalysts. The properties of these materials are strongly related to the distribution of defects—irregularities in electron density on their surfaces. Different materials have various distributions and quantities of these defects, which can be imaged using a procedure that involves depositing palladium nanoparticles. The resulting scanning electron microscopy (SEM) images can be characterized by a key descriptor—the ordering of nanoparticle positions. This work presents a highly interpretable machine learning approach for distinguishing between materials with ordered and disordered arrangements of defects marked by nanoparticle attachment. The influence of the degree of ordering was experimentally evaluated on the example of catalysis *via* chemical reactions involving carbon–carbon bond formation. This represents an important step toward automated analysis of SEM images in materials science.

1. Introduction

Carbon materials play a crucial role in modern materials science,[1] with applications in electronics,[2,3] sensors,[4,5] adsorption[6,7] and purification,[8,9] and catalysis.[10,11] The applicability of specific carbon materials strongly depends on their structure,[12,13] making its determination an important task.

Scanning electron microscopy (SEM) is one of the major methods for studying these materials.[14–16] This method involves scanning the material's surface with an electron beam, providing high resolution and simplifying surface morphology analysis.[17] However, several material surface features, such as defects (irregularities in the atom arrangement and, consequently, electron density), are difficult to study. Advanced methods such as atomic force microscopy can be used, but they only cover a limited sample surface area.

Previously, a new approach was developed based on the significant difference in the binding energies of palladium nanoparticles to material surfaces.[18] By depositing nanoparticles on the surface, they can be used as markers for defects. In some cases, analysis is further simplified by objects with higher atomic numbers appearing brighter in the SEM images (*Z*-contrast).

In recent years, computational modeling in general and particularly machine learning algorithms were actively employed in nanotechnology.[19–21] A significant contribution was made in optimizing synthesis of nanomaterials,[22–25] analyzing nano-scale properties,[26,27] developing datasets,[28,29] new algorithms,[30,31] and revealing correlations between structure and properties,[32] as well as to evolve methodology applied to micro- and nanoscale dynamics[33,34] and spectroscopy.[35]

However, electron microscopy image analysis connected with particles arrangement insights remains a significant problem. Much work has focused on detecting individual particles,[36,37] but research on more global material features is limited, especially considering materials ordering to be involved in human analysis.[38,39] One major target for analysis is distinguishing between ordered and disordered materials and determining their overall order. Notably, the intuitive understanding of the ordered arrangement that each researcher may possess does not allow for reliable scientific research without its formalization.

Despite the fact that deep learning approaches currently occupy leading positions in solving machine learning problems, especially computer vision problems, the interpretability of the entire set of several million internal parameters of models has not yet been solved. As such, heatmaps of various types are often used to explain the features of the neural network model (visually investigating model behavior) rather than dependencies in the source data.[58] The main purpose of this study is to formalize the concept of the ordered arrange-

[a]*Tula State University, Lenina Ave. 92, 300012 Tula, Russia*
[b]*Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninsky Prospekt 47, Moscow, 119991, Russia. E-mail: val@ioc.ac.ru; https://AnanikovLab.ru*
†Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4nr00952e

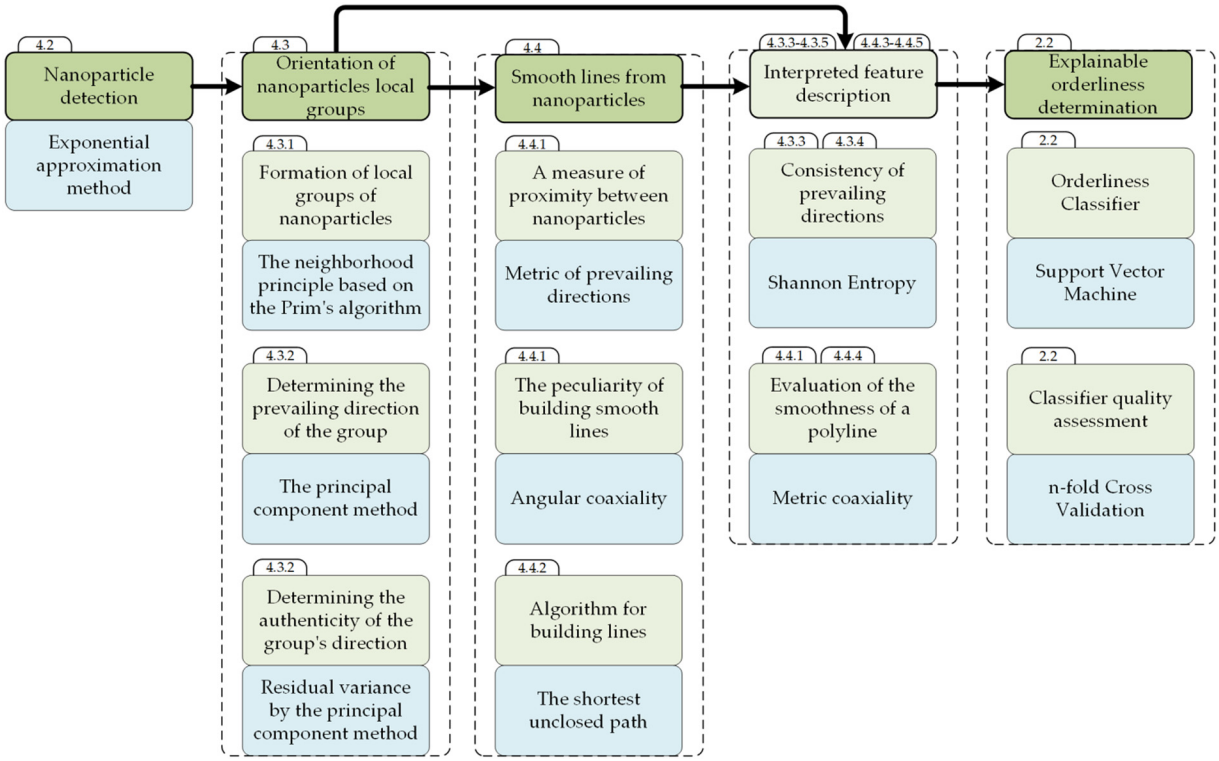

Fig. 1 The main stages of the ordering analysis of nanoparticles arrangement in a SEM image. The numerical labels correspond to the section numbers of the present article.

ment of particles based solely on their coordinates in the image.

The project described here aims to provide researchers with intelligible and controllable attributes of orderliness that can lead to a reliable interpretation of underlying chemical processes. This work presents a novel approach for determining the order of carbon materials as a measure of their underlying structure. The method uses data from the visualization of metal nanoparticles and formalizes the concept of ordering the relative position of these nanoparticles, taking into account the specificity of human perception. As a result, we propose a set of characteristics that, on the one hand, is in good agreement with intuitive understanding and, on the other hand, allows us to quantitatively describe data in terms of understandable and easily interpreted physical parameters. In this case, the classifiers training and statistical quality assessment of their work allowed us to check the adequacy of the constructed features. A plausible connection between the ordering of defects on carbon materials and catalytic activity is discussed based on experimental electron microscopy analysis of catalysts before and after the reaction.

2. Results and discussion

Experimental data consisted of electron microscopy images of the surface of carbon materials with attached metal nanoparticles. A previously developed procedure was involved using Pd as a contrast agent.[40] The idea is to highlight the positions of defect sites with easily detectable markers. Pd nanoparticles are preferentially and selectively attached to defect sites on the carbon surface. Thus, the defect sites, which are typically indistinguishable on the surface, become clearly visible, and geometrical patterns and spatial arrangements of defects can be recorded with electron microscopy images (see ESI section 2† for feature description).

2.1 General modeling approach

The proposed approach to determining the ordering of nanoparticle positions in SEM images involves forming an interpretable characteristic description of the concept of ordering, which is consistent with unique human perception.

Fig. 1 illustrates the primary steps in analyzing the ordering, presented in the form of a flowchart diagram. First, nanoparticles are detected in the SEM image. Next, based on their positions, an interpretable feature description is generated, which is necessary for comprehensive explainable ordering analysis.

A preliminary step in forming a feature description is detecting nanoparticles in the SEM image. After that, information about the positions of the nanoparticles (coordinates of their centers) in the image becomes available. A previously proposed nanoparticle detection method based on the exponential approximation of image fragments[41] was successfully applied to real SEM images of nanoparticles. Fig. 2 provides an example of the detection results.

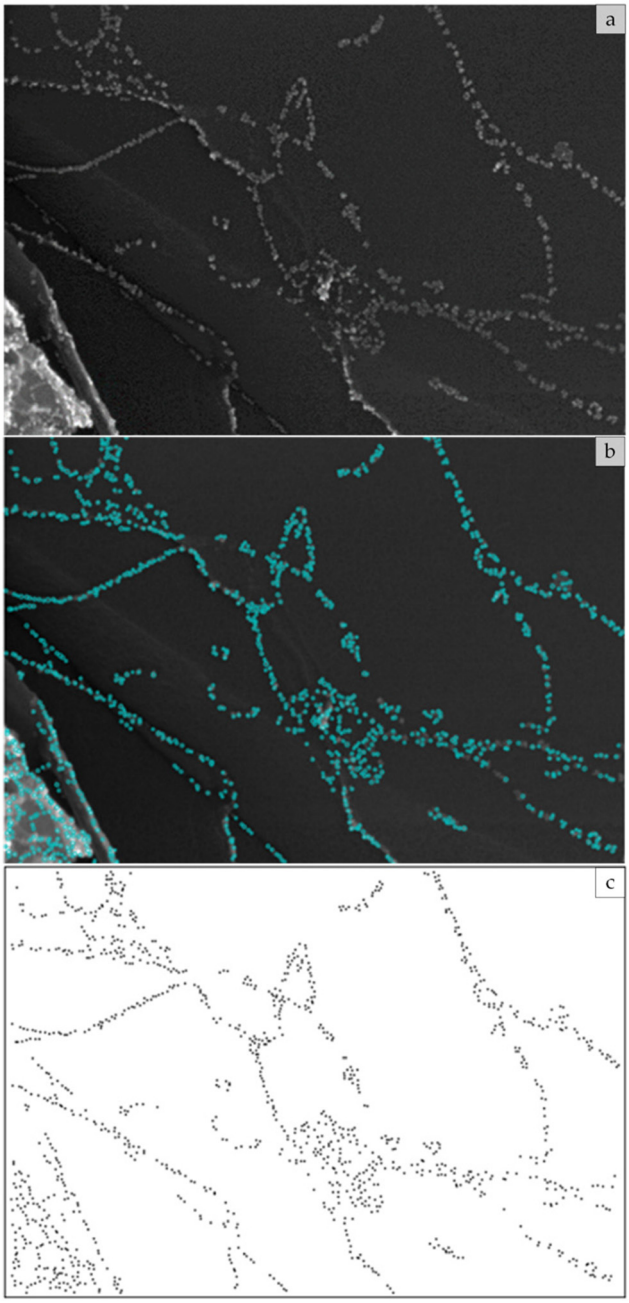

Fig. 2 Image analysis: (a) the original SEM image; (b) the nanoparticles arrangement in the background of the SEM image; and (c) the map of nanoparticles arrangement used for order analysis.

To analyze the ordering, we used only the nanoparticle arrangement information, eliminating factors such as background, average brightness, and SEM image resolution (Fig. 2c).

An attempt to formalize the concept of orderliness within the proposed approach is based on two interrelated assumptions about the nanoparticle arrangement.

2.1.1 First assumption. In the case of an ordered nanoparticles arrangement, local groups of nanoparticles will have similar (consistent) orientations. Moreover, it is possible to measure not only the orientations but also their determination reliability, which indicates how close the arrangement of nanoparticles in a local group is to a straight line. For example, Fig. 3a shows that in the areas of orderliness, there are reliably determined orientations that are consistent for neighboring groups of nanoparticles. In contrast, disordered areas are characterized by the presence of chaotic orientations with low reliability.

This assumption makes it possible to form an orderliness characteristic group based on orientations (O-group):

• O_1 feature. General consistency of orientations. This characteristic is a general estimate of the consistency of nanoparticle local group orientations, regardless of their reliability. This reflects the idea that the more consistent the orientations are, the more ordered the nanoparticles arrangement is.

• O_2 feature. Partial consistency of orientations. As mentioned above, a SEM image may contain disordered regions with unreliable orientations of local nanoparticles groups (Fig. 3a). Considering such orientations can distort the estimate of the order of nanoparticle arrangement. Therefore, this

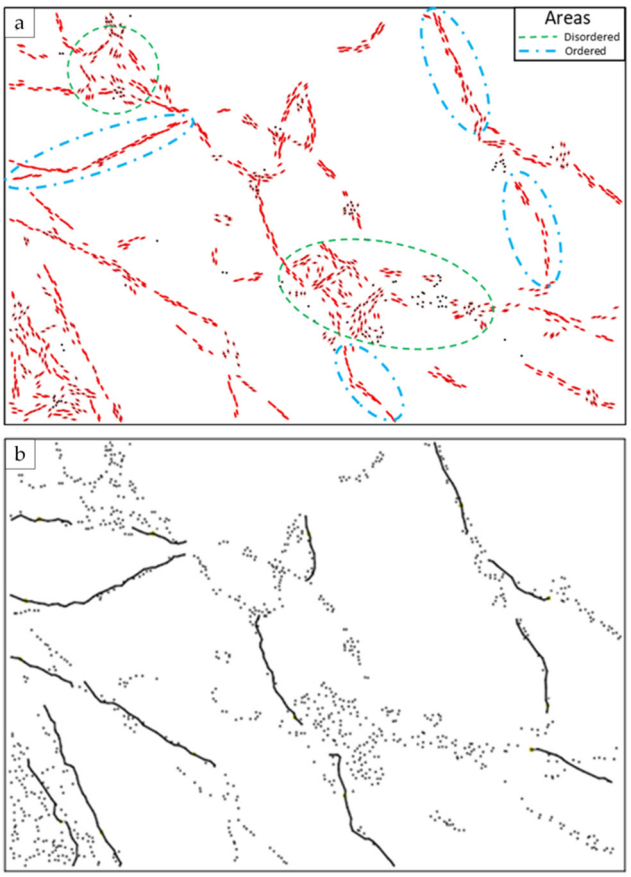

Fig. 3 Illustration of the orderliness assumptions: (a) orientations of local groups are shown in red; the line brightness corresponds to the reliability of determining the orientation; (b) smooth lines in ordered areas based on the connection of nearby nanoparticles.

characteristic estimates the consistency of local directions (such as O_1) but takes into account only those that have high reliability.

• O_3 feature. The fraction of reliable orientations. This characteristic is based on estimating the number of local nanoparticles groups with highly reliable orientation, and reflects the idea that the more groups have highly reliable orientations, the more ordered the nanoparticle arrangement is.

Determining the orientations of local groups of nanoparticles and calculating features based on them are described in more detail in subsection 4.3 "Features based on orientations of nanoparticles local groups".

It should be noted that in the ordered areas, the neighboring orientations of local nanoparticle groups (shown in red in Fig. 3a) already form some appearance of smooth lines visible by the human eye. This observation is the basis of the following assumption about orderliness.

2.1.2 Second assumption. In the case of an ordered arrangement of nanoparticles, nearby nanoparticles can be connected to obtain smooth polylines. Fig. 3b illustrates this assumption. It is easy to see that in the ordered areas (for which examples are indicated in Fig. 3a by blue dash-dot ellipses), there are smooth, long enough lines in Fig. 3b. In contrast, the connection of nearby nanoparticles in disordered areas generally prevents long smooth lines from being obtained.

This assumption makes it possible to form a group of orderliness characteristics based on smooth lines (L-group):

• L_1 feature. Number of constructed lines. This characteristic reflects the idea that the more lines can be constructed, the more ordered the nanoparticle arrangement is.

• L_2 feature. Smoothness of the constructed lines. This characteristic is a generalized estimate of all constructed lines in terms of smoothness (the integrated index of local similarity of small polyline fragments with a straight line). This reflects the idea that the smoother lines can be formed, the more ordered the nanoparticle arrangement is.

• L_3 feature. Rectilinearity of the constructed lines. Like the L_2 characteristic, this feature is a generalized estimate of all constructed lines but from the point of view of similarity to a straight line throughout the polyline.

• L_4 feature. The fraction of connected nanoparticles. This characteristic suggests that the more nanoparticles that are connected into lines, the more ordered the nanoparticle arrangement is.

The formation of smooth polylines and the calculation of features based on them are fully described in subsection 4.4 "Features based on smooth lines".

Combining features based on the orientations of local groups of nanoparticles and features based on smooth lines makes it possible to form a well-interpreted feature description for each of the SEM images. This description forms an important basis for further explainable analysis of nanoparticles arrangement orderliness using explainable machine learning methods,[42,43] as opposed to the use of unexplainable deep neural networks.[44]

2.2. Determination of explainable orderliness

As a result of the application of the proposed approach, an interpretable feature description was obtained, containing seven ordering features for each SEM image of the dataset[48] (see subsection 4.1 "Data collection").

- (O_1) General consistency of orientations;
- (O_2) Partial consistency of orientations;
- (O_3) The fraction of reliable orientations;
- (L_1) Number of lines constructed
- (L_2) Smoothness of the constructed lines;
- (L_3) Rectilinearity of the constructed lines;
- (L_4) The fraction of connected nanoparticles.

Fig. 4 shows a graphical representation of each of the SEM images in the space of three features (O_2, O_3 and L_4), the most informative for determining whether ordering occurred.

It also shows that the classes of ordered and disordered images are locally concentrated even when only a portion of the proposed features is used. This suggests that the feature description reflects the real relationship between the ordering and the nanoparticle arrangement.

The issue of determining the order of particles is a novel one, with an example of using convolutional neural networks described previously.[38]

The classifier quality was estimated using a 5-fold cross-validation[45,46] procedure. It should be noted that the experimental datasets are typically unbalanced (for example, 750 ordered and 250 disordered images in the dataset used in the present study). Therefore, when forming cross-validation folds, stratification must be carried out.[47] For more information about the quality indicators used, see subsection 4.6.

The results of applying the proposed approach were compared to the results of the previous work[38] that solved the considered problem using deep neural networks. It is important to note that the literature study was focused on detecting nanoparticle ordering and therefore provides information only about the "ordered" target class (Table 1). Note that the whole image was used as the initial data for training the neural network, and not the coordinates of the particles, as in the proposed approach.

Table 1 shows the main quality indicators of 5-fold cross-validation for three convolutional networks, where the target class is ordered images. Table 2 shows the main quality indicators by 5-fold cross-validation for the Support Vector Machines (SVM) linear classifier based on the proposed interpreted features, where the target class is both ordered and disordered images.

Tables 1 and 2 demonstrate that the SVM linear classifier, which utilizes only 7 interpretable features, performs marginally better than the approach based on convolutional neural networks that use abstract parameters in the tens and hundreds of millions.

The results obtained confirm our assumption that the proposed characteristics reflect the real relationship between the ordering and nanoparticles arrangement. From this point of view, the type of classifier used does not play a large role since

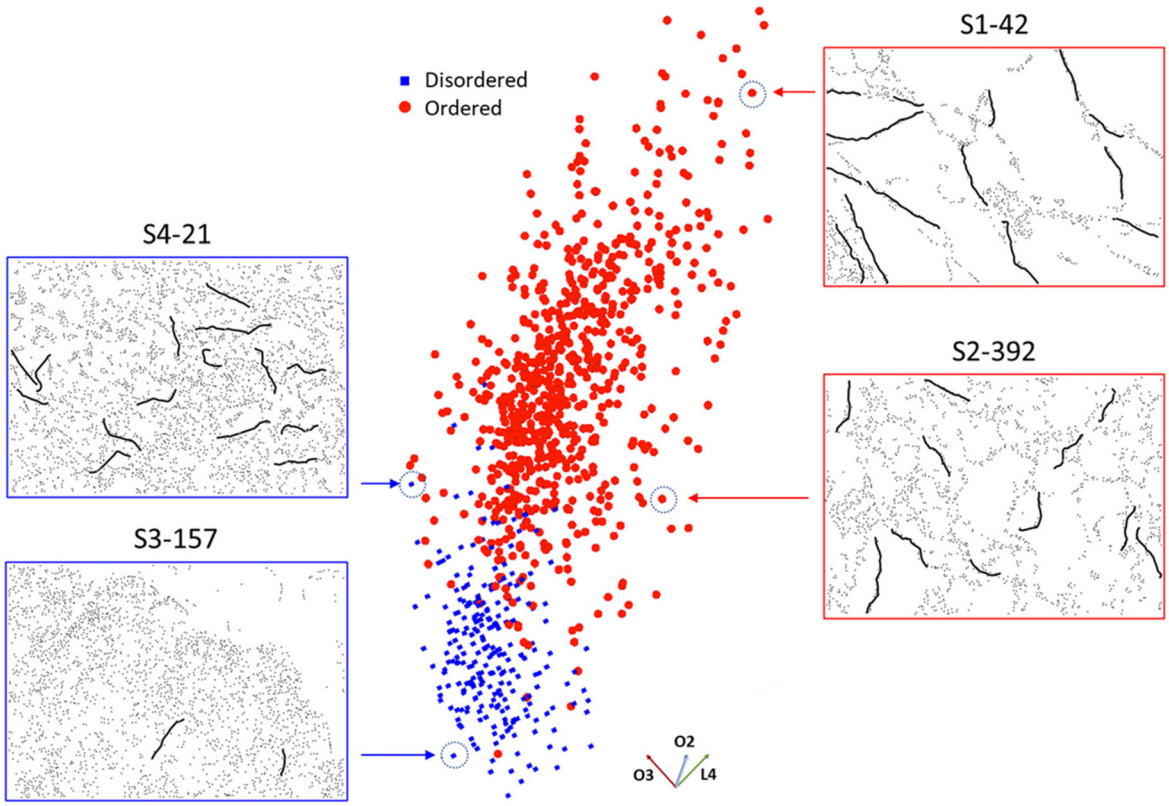

Fig. 4 Representation of SEM images in the space of the three most informative features (O_2, O_3, L_4).

Table 1 The quality indicators for 5-fold cross-validation of three neural networks[38]

Neural network	Number of parameters	Accuracy	Precision	Recall	*F*	AUC
AlexNet	57 M	0.80	0.71	1.00	0.83	0.92
ResNet34	21 M	0.95	0.91	1.00	0.95	0.98
VGG-13	129 M	0.95	0.91	1.0	0.95	1.0

Table 2 The quality indicators for 5 stratified folds cross-validation of a linear SVM

Target class	Accuracy	Precision	Recall	*F*	AUC
Ordered (750)	0.957	0.97	0.98	0.97	0.989
Disordered (250)		0.93	0.90	0.91	

interpretability is achieved at the expense of the feature space. Linear SVM was chosen because it is theoretically justified and allows visualization of the separability of objects in a system of explicable features (see ESI, section 5†).

2.3. Nanoparticle ordering under catalytic conditions

To study the behavior of catalysts with different degrees of nanoparticle ordering under catalytic conditions, we selected two reactions—the Suzuki–Miyaura reaction with Pd on nanoglobular carbon (disordered, see Fig. 5a and b showing before and after the reaction cases, respectively) and the Mizoroki–Heck reaction with Pd on graphite (ordered, see Fig. 5c and d showing before and after the reaction cases, respectively).

The Suzuki–Miyaura reaction was carried out under relatively mild conditions (70 °C; see Methods section, subsection 4.7), and palladium deposited on nanoglobular carbon was chosen as the catalyst (Fig. 5a). This type of support has a random distribution of surface defects, so that the deposited nanoparticles are also distributed chaotically. Examination of the sample by electron microscopy after the catalytic reaction showed that the nanoparticles were almost completely dissolved or detached from the support (Fig. 5b). In addition, accumulation of agglomerates may indicate particle movement on the surface followed by agglomeration.

The harsher conditions of the Mizoroki–Heck reaction (140 °C; see Methods section, subsection 4.7) were chosen to demonstrate the behavior of palladium on a graphite support. As previously shown, this type of support exhibits an ordered arrangement of defects, as shown in Fig. 5c and d. However, the experiment showed that more stringent reaction con-

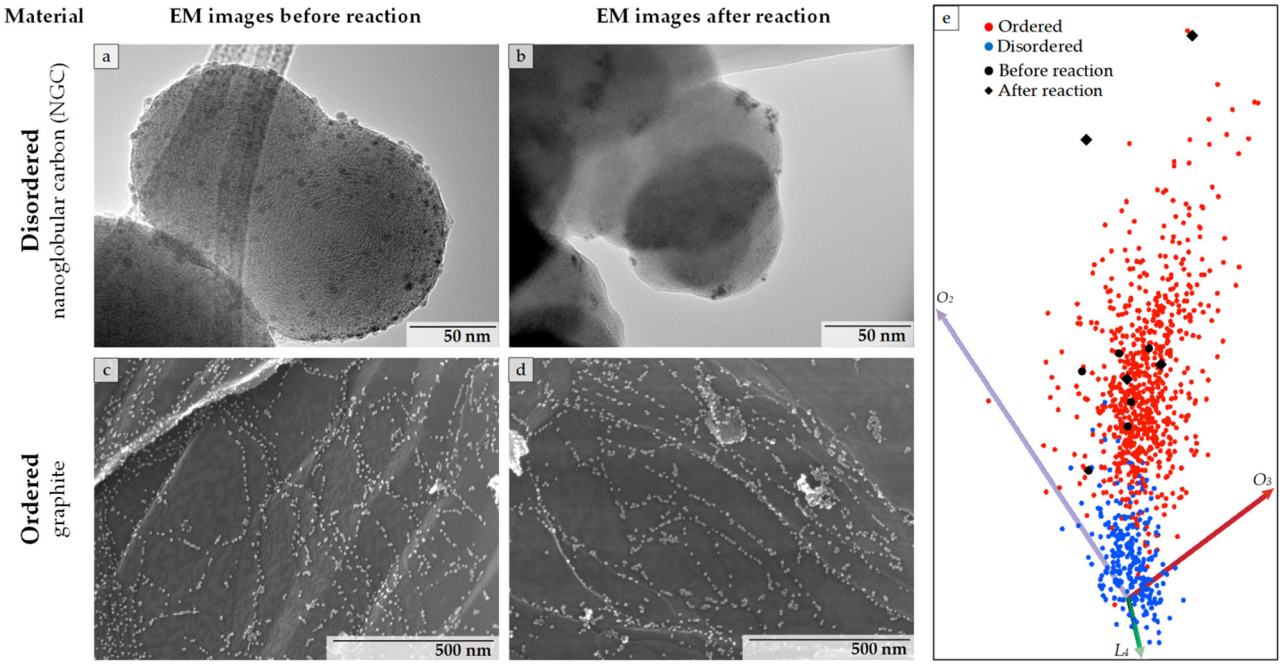

Fig. 5 Analysis of images before and after reaction: (a–d) examples of catalyst behavior with ordered and disordered arrangements under catalytic conditions; (e) representation of SEM images before and after reaction in the space of the three most informative features (O_2, O_3, L_4).

ditions did not result in the pattern observed in the Suzuki–Miyaura reaction. It was found that support with an ordered defect array is more resistant to the metal leaching phenomenon.

Although the nanoparticles are still present in the images before and after the reaction, their location may change, which will help shed light on the dynamic processes in the solution. Fig. 5e shows the results before and after the reaction for a larger image number of the ordered material. For these images, which are not included in the dataset,[48] the proposed ordering parameters were calculated (see the ESI, section 6†), which allows them to be displayed in the appropriate space. Fig. 5e shows that these images are usually ordered. On average, the order after the reaction is greater than the order before the reaction.

These results confirm the importance of the effects underlying nanoparticle ordering in dynamic processes occurring during chemical reactions. The development of automated methods for nanoparticle ordering analysis will contribute to the development of new, more efficient catalytic systems in the future.

3. Conclusions

This work presents a new approach for identifying ordered and disordered arrangements of palladium nanoparticles to determine the distribution of defects on carbon material surfaces. The approach includes the calculation of various features from two main groups: nanoparticle orientations and polyline construction. High performance was achieved, which is on par with that of much more complex neural network models previously reported.

Notably, the proposed approach based on explicable data analysis allows us to explicitly interpret the classification result based on formalized ordering features, which is impossible for a neural network represented as a "black box" model. This is important because the proposed approach can form the basis of a more general indicator of orderliness—the degree of orderliness.

We also showed that nanoparticle ordering is strongly related to the dynamic processes occurring in chemical reaction mixtures. Undoubtedly, the application of these models will have a significant impact on automating SEM image analysis in carbon material research and material science in general.

4. Methods

4.1 Data collection

A previously developed dataset containing 750 images (with a particle ordering effect) and additional 250 images (without an ordering effect) was utilized.[48] The SEM images were acquired in TIFF format with a 1280 × 890 resolution. The images were separated into two groups: those with predominant ordering and those with predominant disordering effects.

These SEM images were obtained using a field-emission scanning electron microscope (FE-SEM) Hitachi SU8000. The operation conditions involved secondary electron mode at an accelerating voltage of 10–30 kV and an operating distance of 6–12 mm.

4.2 Nanoparticle detection

Nanoparticles detection was made by early proposed the exponential approximation method,[41] which includes five stages: (1) pre-processing, (2) selecting small image fragments, (3) making exponential approximations for each image fragment, (4) detecting fragments containing a nanoparticle, and (5) determining the radius of each particle whose center lies in the fragment's center. See the original paper[41] for a more detailed description.

4.3 Features based on the orientations of nanoparticle local groups (O-features)

Each detected particle forms a local group of nanoparticles located close to it.

In accordance with the proposed approach, the orientation of a local group of nanoparticles is understood as the prevailing direction along which the nanoparticles of this group line up.

Therefore, this section describes the following:

- The proposed method for forming local groups of nanoparticles prevailing directions (subsection 4.3.1),
- The proposed method for computing the prevailing direction for a local group of nanoparticles (subsection 4.3.2) and
- Three orientation-based features (O-features) (subsections 4.3.3–4.3.5).

4.3.1 Forming local groups of nanoparticles. Let N be the number of nanoparticles detected in an SEM image and $\boldsymbol{E} = [e_{ij};\ i, j = 1,..., N]$ be a matrix of the Euclidean distances between their centers.

Each of the detected nanoparticles is the starting point for the formation of a local group (so initially, the local group consists of only one nanoparticle).

A new nanoparticle for adding to a group is selected as the nanoparticle with the minimum Euclidean distance to the nearest nanoparticle of the group. The proper nanoparticle can be easily found on the basis of the neighborhood graph constructed by Prim's algorithm.[49]

In the simplest case, the adding process is continued until forming a group of the given size has finished. However, it should be noted that in a number of cases, the minimum distance nanoparticle can be situated far enough from the group. This is especially true for regions in a SEM image with low local nanoparticle density. It is evident that adding far nanoparticle is undesirable because it can lead to a distortion of the group properties.

To solve this problem, we propose the use of an early stopping criterion based on the special threshold, which represents the average local density of nanoparticles in areas with their most intense accumulation and can be estimated on the basis of $k{\cdot}N$ minimal distances between nanoparticles $e_j^* \in \boldsymbol{E}$, $j = 1, ..., k{\cdot}N$:

$$d = w_d \cdot \frac{1}{k \cdot N} \sum_{i=1}^{k \cdot N} e_i^*, \quad (1)$$

where k is the proportionality coefficient, which defines the number of used minimal distances and is the parameter of the proposed method, and w_d is a weight coefficient.

Thus, the modified Prim's algorithm for forming a local group of nanoparticles can be represented as follows.

Algorithm 1. Modified Prim's algorithm

```
G = {i}                  # Indices of nanoparticles of the group
E = {e_11, ..., e_NN}    # Euclidean distance matrix
s                        # Maximum number of nanoparticles in one group
1 while (|G| < s):
2     j = argmin e_GG     # index of nanoparticle closest to the group
3     if (∃g∈G: e_gj ≤ d) then:
4         G = G∪j
5     else: break
```

As a result of the proposed procedure, the located groups may be completely different, partially overlap, or be exactly the same. The number of nanoparticles in each local group is upper bounded by some predefined value s, which is the parameter of the proposed method. At the same time, small groups of nanoparticles (with a size less than four) were excluded from further analysis.

4.3.2. Computing the prevailing direction for a group of nanoparticles. Since each nanoparticle is represented by the coordinates of its center, the determination of the prevailing direction is performed for a group of points in two-dimensional space. Formally, such a task can be set as the search for such a vector (line), the sum of the distances from the objects of the group to which is minimal. Such a problem is equivalent to finding the main component, which can be solved analytically for the two-dimensional case. In computer vision, this task can be solved using principal component analysis (PCA).[50,51]

In this case, the direction of the local group of nanoparticles corresponds to the maximum eigenvector of the covariance matrix. The tilt angle of this vector Θ relative to the horizontal can be calculated by the following formula:

$$\Theta = \arctan\left(\frac{2 \cdot \mu'_{11}}{\mu'_{20} - \mu'_{02}}\right), \quad (2)$$

where μ'_{11}, μ'_{20}, and μ'_{02} are elements of the covariance matrix of a pair of vectors composed of the coordinates of the nanoparticles centers. To estimate the reliability of the prevailing direction, the image eccentricity[51] is used:

$$q = \left(1 - \frac{\lambda_{\min}}{\lambda_{\max}}\right)^2, \quad (3)$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and minimum eigenvalues of the covariance matrix, respectively.

This estimate takes values in the range [0, 1] and shows how much the arrangement of nanoparticles is "elongated" in the prevailing direction. The best possible value $q = 1$ is reached when all the nanoparticles in the local group are located on the same straight line.

Fig. 6 shows a visual representation of the main characteristics for computing O-features for disordered and ordered arrangements of nanoparticles in SEM images.

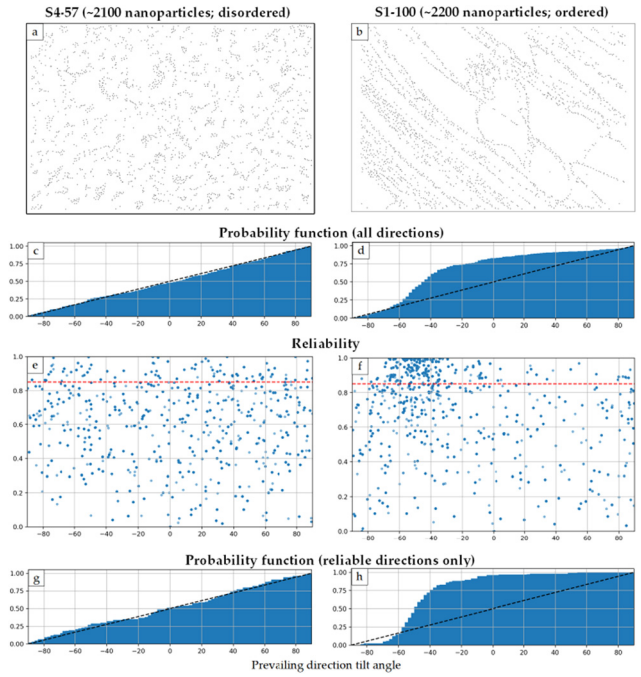

Fig. 6 Visual representation of the main characteristics for computing O-features for disordered (left) and ordered (right) arrangements of nanoparticles.

Fig. 6a and b show the nanoparticles detected in the disordered (left) and ordered (right) SEM images that would be used for further analysis.

Fig. 6c and d show graphs of the probability distribution functions for all prevailing direction tilt angles, where the dashed black line corresponds to the case of a uniform distribution of tilt angles.

Fig. 6e and f show the prevailing direction tilt angles jointly with their reliabilities, where the dashed red line shows the threshold reliability value. It should be noted that some local groups can be characterized by the same angle and reliability values; thus, they fall into the same point on the graph. The point brightness indicates the number of local groups with the same characteristics (the more local groups there are, the brighter the color is).

Fig. 6g and h show graphs of the probability distribution functions for only high-reliability prevailing direction tilt angles (for which the reliability is higher than the threshold – only points above the red dotted line in Fig. 6e and f).

As shown in Fig. 6, the representation of main characteristics of the ordered and disordered nanoparticles arrangements differ significantly from each other. Therefore, these features are expected to be quite informative for further analysis to distinguish between ordered and disordered nanoparticle arrangements.

4.3.3 O_1 feature. General consistency of orientations. The features O_1 and O_2 are based on the assumption that the ordered arrangement of nanoparticles is characterized by the presence of a significant number of local groups with similar (or even identical) orientations (these orientations are referred to as consistent orientations here), which occurs because the nanoparticles in the corresponding groups are arranged along long smooth lines. At the same time, for the disordered arrangement of nanoparticles, a significantly more uniform distribution of the orientations is typical. Therefore, in the case of a disordered arrangement, a low consistency of orientations of local groups of nanoparticles occurs.

A quantitative measure of the general consistency of orientations (prevailing directions) can be computed on the basis of the Shannon entropy.[52]

The tilt angle in the prevailing direction always takes values in the range [−90°, +90°]. To calculate the Shannon entropy, this range is divided into m intervals of equal length, and the empirical probabilities p_i of the angle falling into each interval $i = 1, \ldots, m$ are computed.

Then, the value of the Shannon entropy H can be calculated by the following formula:

$$H = -\sum_{i=1}^{m} p_i \cdot \log_2 p_i. \quad (4)$$

Note that the maximum possible entropy value is limited and can be reached in the case of a uniform distribution:[59]

$$H* = \log_2 m. \quad (5)$$

The final value of the O_1 feature can be calculated as the ratio (4) on (5) and reflects the effective value.[59] The negative sign in the ratio is required to normalize the values of the features – the greater, the better.

$$O_1 = -\frac{H}{H^*}. \quad (6)$$

In this case, $O_1 \in [-1, 0]$, and the highest value of 0 can be achieved when all the nanoparticles are arranged in a single line. The value of *bins* in this work was taken as 90.

4.3.4 O_2 feature. Partial consistency of the orientations. This feature is a special case of the general consistency of orientations. The difference is that only the tilt angles of those prevailing directions are taken into account, and their reliability exceeds the specified threshold $q_{\min}$.

4.3.5 O_3 feature. The proportion of reliable orientations. Each prevailing direction is characterized by its reliability (3), which is calculated from the set of reliability values $\boldsymbol{Q} = \{q_i, \ldots, q_N\}$, where N is the number of detected nanoparticles. Additionally, we consider a subset of reliability values that exceed the specified threshold:

$$Q^* = \{q_1^*, \ldots, q_n^*\} \subseteq Q,\ q_i^* > q_{\min} \quad (7)$$

The proposed O_3 defines the proportion of reliable orientations taken into account at that reliability value:

$$O_3 = \left(\sum_{i=1}^{n} q_i^*\right) / \sum_{i=1}^{N} q_i. \quad (8)$$

In this case, $O_3 \in [0, 1]$, where the larger its value is, the more ordered the nanoparticle arrangement is.

4.4 Features based on smooth lines (L-features)

Since, in this work, nanoparticles are represented by the coordinates of their centers, the problem of constructing lines from nanoparticles is, in fact, a problem of connecting points on a plane.

The natural approach is to connect dots that are close to each other. Generally, this problem is solved by constructing the shortest unenclosed path (SUP).[53] However, in this case, there are two problems due to the specifics of the applied task being solved.

The first problem is related to the building smooth lines. By a smooth line here, we mean a line, each small section of which is similar to a straight line. The presence of these lines is typical for images with an ordered arrangement of nanoparticles. However, the use of SUP together with the traditional Euclidean distance in most cases leads to the construction of strongly curved lines, the presence of which cannot serve as an indicator of nanoparticle orderliness.

To solve this problem, we propose a new adaptive metric that, in addition to the Euclidean distances between points, considers the prevailing directions of nanoparticle local groups (subsection 4.3.1) and their reliability, as well as the consistency of a new point (which is a candidate for adding to the line) with the already constructed part of the polyline to possess the smoothness property. The description of the proposed metric, named the metric of prevailing directions (MPD), is given in subsection 4.4.1 of this section.

The second problem is related to the fact that the SUP method is focused on connecting all points into a single line, while within the framework of the applied problem being solved, it is necessary to build separate long smooth lines, not necessarily using all the available points. As a result, some nanoparticles may remain not belonging to any line at all.

In this regard, subsection 4.4.2 of this section proposes a modification of the SUP method, which allows us to take into account the indicated specifics.

Subsections 4.4.3–4.4.5 contain, respectively, a description of the interpreted features L_1–L_4 based on the lines constructed in accordance with the proposed approach.

4.4.1. Metric of prevailing directions (MPD). On the one hand, the proposed metric of prevailing directions is constructed as the weighted combination of the Euclidean distance and the prevailing directions with their corresponding reliability:

$$m_{ij} = C \cdot e_{ij} + (1 - C) \cdot 2 \cdot \max\left[\sin\left(\left|\theta_i - \theta_j\right|\right), \frac{\left(\varepsilon_i + \varepsilon_j\right)}{2}\right], \quad (9)$$

where e_{ij} is the Euclidean distance between the centers of nanoparticles i and j; θ_i is the tilt angle of the prevailing direction for the local group corresponding to the i-th nanoparticle (2); ε_i is the unreliability of the corresponding i-th prevailing direction, which is the reciprocal of the reliability q_i (3): $\varepsilon_i = 1 - q_i \in [0;1]$; and C is the proportionality coefficient, which allows us to adjust the degree of influence of individual parts of the metric on its resulting value.

Note that the difference in the tilt angles of the prevailing directions can be characterized by one of the adjacent angles at the intersection of these directions. Since the sine of the adjacent angles is not important for estimating the difference, we use the sine of the difference in tilt angles of the prevailing directions. Additionally, the use of a sine allows one to normalize the magnitude of the angle difference so that the larger its value is, the further away the nanoparticles are located.

According to the metric, the greater the average unreliability of determining prevailing directions is, the greater the distance of the metric is (the respective nanoparticles are more distant from each other). At the same time, if the difference in tilt angle is large, then the corresponding nanoparticles will be considered distant even if the average unreliability is small due to the occurrence of maximum operation.

On the other hand, in the process of constructing a specific line, the value of the MPD can be corrected to ensure the smoothness of the constructed lines *via* so-called angular coaxiality coax_{ijk}, which is similar to the cosine similarity measure;[54] however, in contrast, it is scaled to the limits of [0,1]:

$$\mathrm{coax}_{ijk} = \frac{1 - \cos\alpha}{2}, \quad (10)$$

where α is the angle between segments (i, j) and (j, k).

Fig. 7 illustrates the concept of angular coaxiality.

The resulting corrected MPD value is defined as follows:

$$m_{ij}^{*} = m_{ij} + w_{\mathrm{coax}} \times (1 - \mathrm{coax}_{ijk}^{2}), \quad (11)$$

where w_{coax} is a weight coefficient of the angular coaxiality.

It should be noted that the corrected MPD values (11), in contrast to the basic MPD values (9), are dynamically changed in the process of constructing each line and cannot be computed in advance.

Fig. 8 shows the contributions of the main parts of the corrected metric of the prevailing directions (11) for three consecutive steps of choosing the nearest nanoparticle.

Fig. 8a, d and g show the traditional Euclidean distance between points. Fig. 8b, e and h show the basic MPD distance (9), which is based on the Euclidean distance and the direction

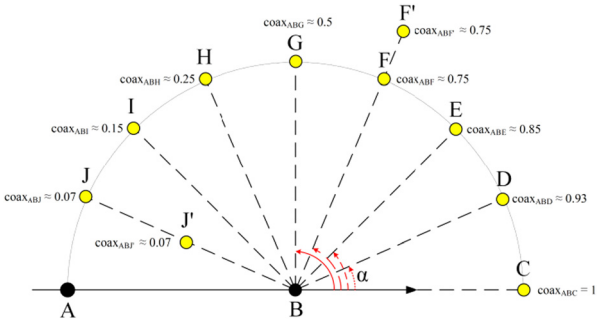

Fig. 7 Illustration of the idea of angular coaxiality, where the last point and last but one point of the line are marked in black and the yellow points are candidates for adding to the line.

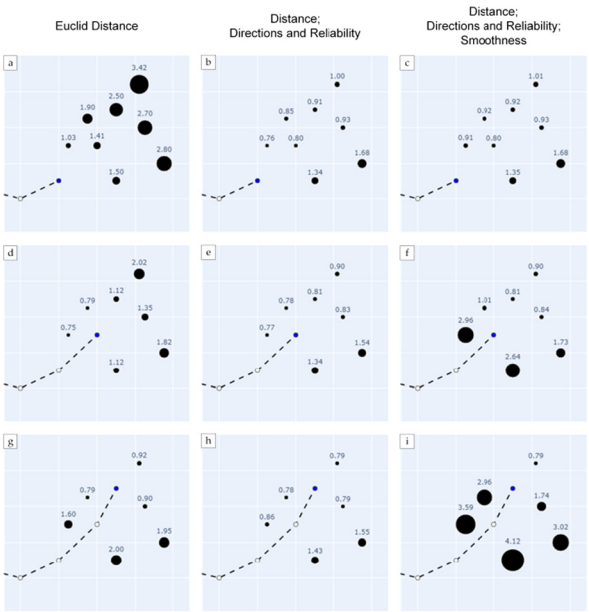

Fig. 8 Contributions of the MPD parts during the construction of the line. The black color shows the points that can potentially be added to the line. Their size corresponds to the distance to the last point of the line already constructed (indicated in blue). The white color shows the points that are already connected in the line.

information jointly related to the reliability. Fig. 8c, f and i show the corrected MPD (11), which allows greater smoothness of the line under construction to be reached.

4.4.2. Constructing lines by the modified shortest unclosed path (SUP) method. In accordance with the proposed approach, line formation is performed point by point, first, it is necessary to select the initial point of the line. It is evident that the best initial point is a nanoparticle with the maximum reliability (3) of the local nanoparticles group associated with it since such a nanoparticle corresponds to the maximum "elongation" of the local group of nanoparticles along the prevailing direction.

Each next point to add to the line can be found on the basis of the shortest unclosed path (SUP) method,[55] which we modify to incorporate dynamically computed distances (11), elongate the line on two sides (by adding new points before the first and after the end point) and stop line formation if the corrected MPD distance (11) exceeds the adopted threshold; this approach is computed as a special case of (9):

$$\mathrm{thr} = C \cdot d + (1 - C) \times w_{\mathrm{thr}}, \tag{12}$$

where d is the nanoparticles density, estimated for a SEM image in accordance with (1); $w_{\mathrm{thr}} = 1$ is the value obtained for limit angle θ and unreliability ε values in (9); and C is the same proportionality coefficient of MPD as in (9).

The description of the proposed algorithm is given in a general form, where the term "point" implies the center of the nanoparticle and the term "index" implies the ordinal number of a nanoparticle:

Algorithm 2. Construction of lines *via* the modified shortest unclosed path method

$\boldsymbol{L} = \langle i \rangle$	# Indexes of points forming a line (i – index of the starting point)
$\boldsymbol{M} = \{m_{11}, \ldots, m_{NN}\}$	# Matrix of the basic metric of prevailing distances (4)
N	# Number of detected nanoparticles
1 while ($\lvert\boldsymbol{L}\rvert \le N$):	
2	$l = \boldsymbol{L}_0$ # indexes of the leftmost point in the line
3	$r = \boldsymbol{L}_{\lvert\boldsymbol{L}\rvert}$ # indexes of the rightmost point in the line
4.1	# indexes of the closest points to the line on the left $\boldsymbol{J} = \{j_1, \ldots, j_x\}$, $\forall j \in \boldsymbol{J}$: $m_{lj} \le \mathrm{thr}$; $\boldsymbol{J} \cap \boldsymbol{L} = \varnothing$
4.2	# corrected MPD metric values (6) for the nearest points to the left of the line $\boldsymbol{ML} = \{ml_1, \ldots, ml_x\}$
4.3	# index of the closest point to the line on the left in the corrected MPD metric $j = \mathrm{argmin}\ \boldsymbol{ML}$
5.1	# indexes of the closest points to the line on the right $\boldsymbol{K} = \{k_1, \ldots, k_y\}$, $\forall k \in \boldsymbol{K}$: $m_{rk} \le \mathrm{thr}$; $\boldsymbol{K} \cap \boldsymbol{L} = \varnothing$
5.2	# corrected MPD metric values (6) for the nearest points to the right of the line $\boldsymbol{MR} = \{mr_1, \ldots, mr_y\}$
5.3	# index of the closest point to the line on the right in the corrected metric $k = \mathrm{argmin}\ \boldsymbol{MR}$
6	if (($\boldsymbol{J} \neq \varnothing$) and ($ml_j \le mr_k$)) then:
7	$\boldsymbol{L} = \langle j, \boldsymbol{L} \rangle$
8	else if (($\boldsymbol{K} \neq \varnothing$) and ($mr_k \le ml_j$)) then:
9	$\boldsymbol{L} = \langle \boldsymbol{L}, k \rangle$
10	else: break.

One pass of this algorithm allows one line to be contracted. To construct a new line, a new starting point should be chosen, and the algorithm should be reapplied.

To exclude the starting line from unreliable points, we set the minimum reliability value $q_{\min}$ that acts as the threshold while choosing the starting point.

4.4.3. L_1 feature. Number of lines constructed. Note that the lines are not informative if they combine a small number of nanoparticles, so it is necessary to take into account only those lines whose length in the nanoparticles will be greater than or equal to a certain threshold $L_{\min}$. In the future, we will discuss only such lines, and their number corresponds to the value of this feature.

4.4.4. L_2 and L_3 features. Smoothness and rectilinearity of the constructed lines. The smoothness and the rectilinearity of the constructed lines directly correspond to human perceptions of orderliness. To formalize these notions, we introduce a special characteristic named the metric coaxiality of a polyline fragment and a straight line.

Let $P = (p_1, \ldots, p_n)$ be a constructed polyline consisting of n ordered nanoparticles and $P^* = (p_a, \ldots, p_b)$ be a fragment. Then, the metric coaxiality for this fragment can be defined as follows:

$$\mathrm{coax}_{P^*} = e(p_a, p_b) / \sum_{i=a}^{b} e(p_i, p_{i+1}), \tag{13}$$

where $e(x, y)$ is the Euclidean distance between the centers of nanoparticles x and y.

The rectilinearity of a single polyline expresses its similarity to a straight line throughout it and can be calculated as the metric coaxiality (13) of the full polyline. The L_3 characterizes the rectilinearity of all constructed lines at once and is computed by averaging individual rectilinearity values. The smoothness of a one polyline expresses its local similarity to a straight line and can be computed as the average metric coaxiality of all polyline fragments of some size f_{size}.

The L_2, like the L_3, characterizes the smoothness of all constructed lines at once and is computed by averaging individual smoothness values.

4.4.5. L_4 features. The proportion of connected nanoparticles. This feature is calculated as the ratio of the number of nanoparticles that are connected by all the constructed lines to the total number of detected nanoparticles.

4.5. Model implementation and training

When calculating the proposed features of the SEM images, the following parameter values were set:

- The proportionality coefficient for early stopping in local groups formation (subsection 4.3.1): $k = 3$;
- The weight coefficient for estimating the local nanoparticles density in a SEM image (subsection 4.3.1): $w_d = 1.5$;
- The maximum number of nanoparticles in a local group (subsection 4.3.1): $s = 8$;
- The reliability threshold for computing the partial consistency of orientations (subsection 4.3.4, subsection 4.4.2): $q_{\min} = 0.85$;
- Proportionality coefficient to adjust the degree of influence of individual parts of the proposed metric of prevailing directions (subsection 4.4.1): $C = 0.025$;
- Weight coefficient of the angular coaxiality in the metric of prevailing directions to ensure line smoothness (subsection 4.4.1): $w_{\text{coax}} = 1.75$;
- Minimum line length in nanoparticles (subsection 4.4.3): $L_{\min} = 12$;
- The size of a polyline local fragment is used to estimate the smoothness of the constructed lines (subsection 4.4.4): $f_{\text{size}} = 6$.

The number of lines found depends not only on the nature of the image but also on the parameters of the search algorithm, which vary from all possible lines (Fig. 9c and f) to no-lines (Fig. 9d). The optimal values of the parameters given above were chosen to guarantee a significant number of lines on the ordered images (Fig. 9b) and a small number of short lines on the disordered images (Fig. 9e). For more information about the effect of the algorithm parameters on the construction of lines, see the ESI section 1.†

The proposed algorithms were implemented in Python.

Based on the corresponding methods from the scikit-learn[56] package, the following steps were implemented: calculations of the prevailing directions of nanoparticles local groups and of the corresponding reliabilities (principal component analysis – decomposition.PCA); training of a linear SVM classifier (svm.SVC: the core is linear, the regularization parameter is 10); evaluation of the classifier quality (cross-validation – model_selection. StratifiedKFold: the number of folds is 5).

Detection of nanoparticles was performed based on a parallel algorithm proposed in our previous work.[57]

The authors implemented algorithms were used for the formation of local groups of nanoparticles and the construction of smooth lines based on the proposed modification of the shortest unclosed path method.

Depending on the number of nanoparticles in the original SEM image, the operating time (excluding the detection stage) of the proposed implementation varies from a couple of seconds for ~1000 nanoparticles to several dozens of minutes for ~20 000 nanoparticles. In the dataset under study, the most common number of nanoparticles in the SEM image corresponds to ~5000, which is processed in a few minutes. The indicated time costs correspond to calculations on a personal computer with the following characteristics: processor – Intel® Core™ i7-9700k (3.6/4.9 GHz); RAM – 16 Gb (DDR4, 3866 MHz); SSD: 256 Gb, operating system – Windows 10 ×64. Parallel computing technologies were not used in this experiment.

4.6. Quality indicators

The main classification quality indicators are calculated based on the confusion matrix of classification, which contains the following values: TP – correctly classified ordered images, TN – correctly classified disordered images, FP – ordered images classified as disordered, and FN – disordered images classified as ordered. Then the classification Accuracy[60] is the ratio of correctly recognized images to the total number of images:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}}. \tag{14}$$

At the same time, most classifiers can balance the decision rule either toward increasing the number of correctly recognized positive class objects (ordered) or toward reducing the number of incorrectly classified negative class objects (disordered) using some hyperparameters. In this regard, such characteristics as Precision,

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{15}$$

which is understood as the proportion of correctly recognized objects among all objects recognized as positive, and Recall,

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{16}$$

which represents the proportion of correctly recognized objects among all positive objects, are often used.

The F-measure (F) is a widely known measure that attempts to combine these two indicators and characterize the quality of the classifier with a single number.[60] It is defined as the harmonic mean between Precision and Recall:

$$F = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = 2\frac{\text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}}. \tag{17}$$

The AUC[61] is estimated as the area bounded by the ROC curve and the axis of false positive classifications (FPR). The

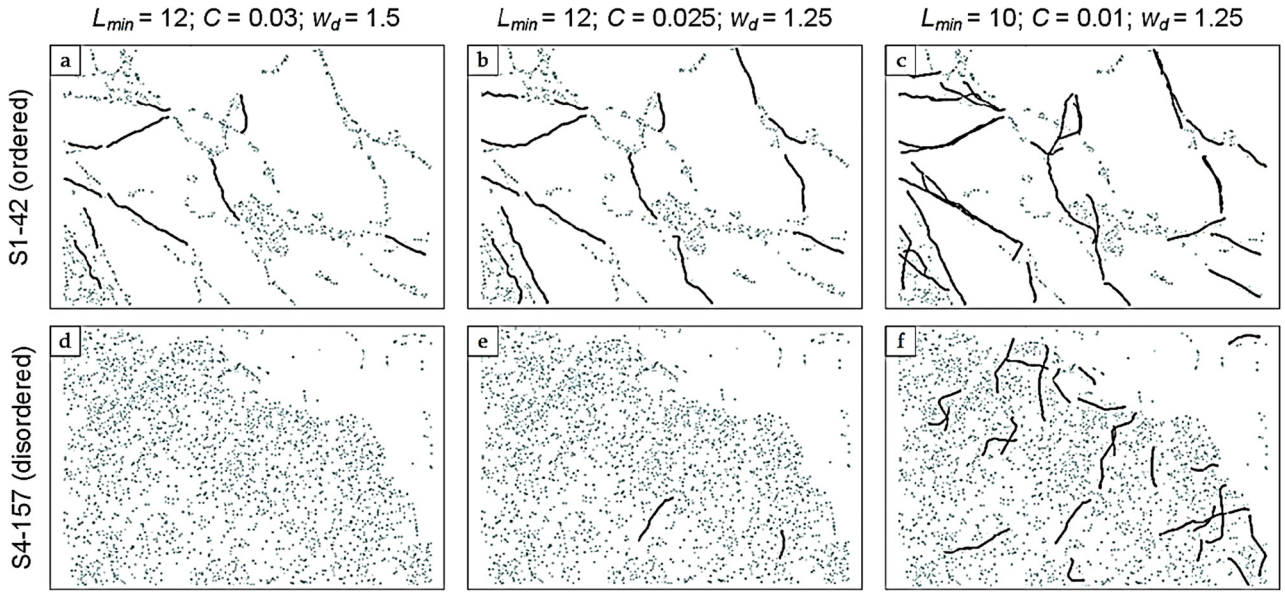

Fig. 9 Illustration of the dependence of the construction of smooth lines on the parameters of the proposed approach for ordered and disordered images.

ROC curve reflects the ratio of the sensitivity of the algorithm (TPR) and its specificity (FPR):

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \quad (18)$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}. \quad (19)$$

4.7. Catalysts with different ordering under catalytic conditions

4.7.1 Preparation of Pd on carbon. 4 ml of chloroform was poured into each one of 2 vials, and 0.4 mg of the $Pd_2dba_3{\cdot}CHCl_3$ complex was added. Carbon material was placed in the resulting solution, and the mixture was heated at 50 °C until bleaching of solution. After the precipitation of palladium, the resulting catalyst was washed with acetone and dried under vacuum. Then, graphite and on NGC (nanoglobular carbon) with deposited palladium were used as catalysts.

4.7.2 Suzuki–Miyaura reaction on Pd/NGC. A microdroplet of the Pd/NGC catalyst suspension in propanol-2 containing approximately 100 000 particles was placed on a sodium chloride crystal and used as a catalyst. 4-Iodo-nitrobenzene (18.7 mg, 7.5×10^2 mmol), 4-OMe-phenylboronic acid (14.8 mg, 10×10^{-2} mmol) and K_2CO_3 (13.8 mg, 10×10^{-2} mmol) were dissolved in 1.6 ml of ethanol and 0.4 ml of water. The mixture was heated at 70 °C for 8 h. At the end of the reaction, a carbon-coated copper grid for TEM was placed in the system so that some amount of catalyst particles was deposited on it. The sample morphology was studied using a transmission electron microscope (Hitachi HT7700). The operation conditions involved a bright-field TEM mode at a 100 kV accelerating voltage.

4.7.3 Mizoroki–Heck reaction on Pd/graphite. 4-Bromo-nitrobenzene (15.2 mg, 7.5×10^{-2} mmol), styrene (7.8 mg, 7.5×10^{-2} mmol) and triethylamine (7.6 mg, 7.5×10^{-2} mmol) were dissolved in 2 ml of DMF. The previously obtained Pd/graphite was used as a catalyst. The mixture was heated at 140 °C for 1 h. Before SEM analysis, graphite was washed with acetone and dried under vacuum at 150 °C for 30 minutes. The sample morphology was studied using a Regulus 8230 (Hitachi) scanning electron microscope at a 20–30 kV accelerating voltage.

Author contributions

Conceptualization, V.P.A.; formal analysis, V.V.S., M.Y.K., A.V.K. and O.S.S.; investigation, D.A.B., V.V.S., M.Y.K., A.V.K., O.S.S. and V.A.C.; methodology, D.A.B., V.V.S., M.Y.K., A.V.K., O.S.S. and V.A.C.; project administration, V.P.A.; software, V.V.S., M.Y.K., A.V.K. and O.S.S.; supervision, V.P.A.; validation, V.V.S., M.Y.K., A.V.K. and O.S.S.; visualization, M.Y.K.; writing—original draft, D.A.B., V.V.S., M.Y.K., A.V.K., O.S.S. and V.A.C.; writing—review & editing, all authors. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare no competing interests.

Acknowledgements

This research is funded by the Ministry of Science and Higher Education of the Russian Federation (project FEWG-2024-0001). A comparative study of various methods is carried out

partly using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University.

References

1 M.-M. Titirici, R. J. White, N. Brun, V. L. Budarin, D. S. Su, F. del Monte, J. H. Clark and M. J. MacLachlan, *Chem. Soc. Rev.*, 2015, **44**, 250–290.
2 D. D. L. Chung, *J. Mater. Sci.*, 2004, **39**, 2645–2661.
3 C. Wang, K. Xia, H. Wang, X. Liang, Z. Yin and Y. Zhang, *Adv. Mater.*, 2019, **31**, 1801072.
4 S. W. Lee, W. Lee, Y. Hong, G. Lee and D. S. Yoon, *Sens. Actuators, B*, 2018, **255**, 1788–1804.
5 M. A. Beluomini, J. L. da Silva, A. C. de Sa, E. Buffon, T. C. Pereira and N. R. Stradiotto, *J. Electroanal. Chem.*, 2019, **840**, 343–366.
6 C. Moreno-Castilla, *Carbon*, 2004, **42**, 83–94.
7 F. Yu, Y. Li, S. Han and J. Ma, *Chemosphere*, 2016, **153**, 365–385.
8 A. Durairaj, J. Liu, X. Lv, S. Vasanthkumar and T. Sakthivel, *Biomass Convers. Biorefin.*, 2023, **13**, 3247–3258.
9 N. An, Y. Jiang, Z. Wang, Q. Sun, B. Guo, B. Gao, W. Zhou and Q. Li, *Sep. Purif. Technol.*, 2022, **301**, 122003.
10 P. Serp and J. L. Figueiredo, *Carbon Materials for Catalysis*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2008.
11 E. Lam and J. H. T. Luong, *ACS Catal.*, 2014, **4**, 3393–3410.
12 A. Takakura, K. Beppu, T. Nishihara, A. Fukui, T. Kozeki, T. Namazu, Y. Miyauchi and K. Itami, *Nat. Commun.*, 2019, **10**, 3040.
13 J.-P. Salvetat, J.-M. Bonard, N. H. Thomson, A. J. Kulik, L. Forro, W. Benoit and L. Zuppiroli, *Appl. Phys. A: Mater. Sci. Process.*, 1999, **69**, 255–260.
14 S. Che, K. Lund, T. Tatsumi, S. Iijima, S. H. Joo, R. Ryoo and O. Terasaki, *Angew. Chem.*, 2003, **115**, 2232–2235.
15 K. Morishita and T. Takarada, *J. Mater. Sci.*, 1999, **34**, 1169–1174.
16 O.-W. Achaw, *Scan Electron Microsc*, InTech, 2012.
17 J. I. Goldstein and H. Yakowitz, *Practical Scanning Electron Microscopy*, Springer US, Boston, MA, 1975.
18 E. O. Pentsak, A. S. Kashin, M. V. Polynski, K. O. Kvashnina, P. Glatzel and V. P. Ananikov, *Chem. Sci.*, 2015, **6**, 3302–3313.
19 L. Pokrajac, A. Abbas, W. Chrzanowski, G. M. Dias, B. J. Eggleton, S. Maguire, E. Maine, T. Malloy, J. Nathwani, L. Nazar, A. Sips, J. Sone, A. van den Berg, P. S. Weiss and S. Mitra, *ACS Nano*, 2021, **15**, 18608–18623.
20 P. Zhang, Z. Guo, S. Ullah, G. Melagraki, A. Afantitis and I. Lynch, *Nat. Plants*, 2021, **7**, 864–876.
21 K. J. Jenewein, L. Torresi, N. Haghmoradi, A. Kormanyos, P. Friederich and S. Cherevko, *J. Mater. Chem. A*, 2024, 3072–3083.
22 F. Mekki-Berrada, Z. Ren, T. Huang, W. K. Wong, F. Zheng, J. Xie, I. P. Siyu Tian, S. Jayavelu, Z. Mahfoud, D. Bash, K. Hippalgaonkar, S. Khan, T. Buonassisi, Q. Li and X. Wang, *npj Comput. Mater.*, 2021, **7**, 55.
23 A. A. Volk, R. W. Epps, D. T. Yonemoto, B. S. Masters, F. N. Castellano, K. G. Reyes and M. Abolhasani, *Nat. Commun.*, 2023, **14**, 1403.
24 H. Lv and X. Chen, *Nanoscale*, 2022, **14**, 6688–6708.
25 Q. Gromoff, P. Benzo, W. A. Saidi, Ch. M. Andolina, M.-J. Casanove, T. Hungria, S. Barre, M. Benoit and J. Lam, *Nanoscale*, 2024, **16**, 384–393.
26 K. Faraz, T. Grenier, Ch. Ducottet and T. Epicier, *Sci. Rep.*, 2022, **12**, 2484.
27 A. S. Galushko, D. A. Boiko, E. O. Pentsak, D. B. Eremin and V. P. Ananikov, *J. Am. Chem. Soc.*, 2023, **145**(16), 9092–9103.
28 T. Liu, Z. Y. Tho and A. S. Barnard, *Digital Discovery*, 2024, **3**, 422–435.
29 A. Colliard-Granero, J. Jitsev, M. H. Eikerling, K. Malek and M. J. Eslamibidgoli, *ACS Nanosci. Au*, 2023, **3**(5), 398–407.
30 Q. Lu, *RSC Adv.*, 2021, **11**, 35879–35886.
31 C. Wang, G. Zhang, Z. Cheng and W. Zhou, *Computer Vision – ACCV* 2022, 13842.
32 S. Luan, E. Chen, J. John and S. Gaitanaros, *Sci. Adv.*, 2023, **9**, eadi1453.
33 L. Yao, Z. Ou, B. Luo, C. Xu and Q. Chen, *ACS Cent. Sci.*, 2020, **6**, 1421–1430.
34 Y. Sun, X. Zhang, R. Huang, D. Yang, J. Kim, J. Chen, E. H. Ang, M. Li, L. Li and X. Song, *Nanoscale*, 2024, **16**, 2945–2954.
35 N. M. Sabanes, M. D. Eaton, S. Moreno-Da Silva, A. Naranjo and E. M. Perez, *Nanoscale*, 2024, **16**, 2048–2059.
36 G. Guven and A. B. Oktay, *26th Signal Processing and Communications Applications Conference (SIU)*, IEEE, 2018, 1–4.
37 A. Yu. Kharin, *Ultramicroscopy*, 2020, **219**, 113125.
38 D. A. Boiko, E. O. Pentsak, V. A. Cherepanova, E. G. Gordeev and V. P. Ananikov, *Chem. Sci.*, 2021, **12**, 7428–7441.
39 G. Mather, *Foundations of Perception*, Psychology Press, 2006.
40 E. O. Pentsak, A. S. Kashin, M. V. Polynski, K. O. Kvashnina, P. Glatzelc and V. P. Ananikov, *Chem. Sci.*, 2015, **6**, 3302–3313.
41 D. A. Boiko, V. V. Sulimova, M. Yu. Kurbakov, A. V. Kopylov, O. S. Seredin, V. A. Cherepanova, E. O. Pentsak and V. P. Ananikov, *Nanomaterials*, 2022, **12**, 3914.
42 F. Charmet, H. C. Tanuwidjaja, S. Ayoubi, P.-F. Gimenez, Y. Han, H. Jmila, G. Blanc, T. Takahashi and Z. Zhang, *Ann. Telecommun.*, 2022, **77**, 789–812.
43 N. Burkart and M. F. Huber, *J. Artif. Intell. Res.*, 2021, **70**, 245–317.
44 G. Ras, N. Xie, M. Van Gerven and D. Doran, *J. Artif. Intell. Res.*, 2022, **73**, 329–397.
45 V. N. Vapnik, *Statistical Learning Theory*, Wiley, 1998.
46 R. Kohavi, in *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, Morgan Kaufmann Publishers Inc., 1995, 1137–1143.

47 M. Shahrokh Esfahani and E. R. Dougherty, *Bioinformatics*, 2014, **30**, 242–250.
48 D. A. Boiko, E. O. Pentsak, V. A. Cherepanova and V. P. Ananikov, *Sci. Data*, 2020, **7**, 101.
49 R. C. Prim, *Bell Syst. Tech. J.*, 1957, **36**, 1389–1401.
50 A. N. Gorban, B. Kégl, D. C. Wunsch and A. Y. Zinovyev, *Principal Manifolds for Data Visualization and Dimension Reduction*, Springer, Berlin, 2008.
51 M.-K. Hu, *IEEE Trans. Inf. Theory*, 1962, **8**, 179–187.
52 C. E. Shannon, *Bell Syst. Tech. J.*, 1948, **27**, 623–656.
53 O. Seredin, E. Surkov, A. Kopylov and S. Dvoenko, *ICABDE 2021. Lecture Notes on Data Engineering and Communications Technologies*, Springer, Cham, 2022, pp. 279–299.
54 A. H. Cheetham and J. E. Hazel, *J. Paleontol.*, 1969, **43**, 1130–1136.
55 E. E. Surkov, O. S. Seredin and A. V. Kopylov, in *2023 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology*, IEEE, 2023, 221–224.
56 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.
57 M. Y. Kurbakov and V. V. Sulimova, *Int. Arch. Photogramm., Remote Sens. Spatial Inform. Sci.*, 2023, **XLVIII-2/W3-2023**, 145–150.
58 R. Ibrahim and M. Omair Shafiq, *ACM Comput. Surv.*, 2023, **55**, 1–37.
59 T. M. Cover; and J. A. Thomas, *Elements of Information Theory*, Wiley, 1991.
60 C. J. van Rijsbergen, *Proceedings of the Joint IBM/University of Newcastle upon Tyne Seminar on Data Base Systems*, 79, 1979, 1–14 P.
61 D. M. Powers, *Int. J. Mach. Learn. Technol.*, 2011, 37–63.