

# Machine Learning Determination of the Twist Angle of Bilayer Graphene by Raman Spectroscopy: Implications for van der Waals Heterostructures

Pablo Solís-Fernández\* and Hiroki Ago\*



Cite This: ACS Appl. Nano Mater. 2022, 5, 1356–1366



Read Online

ACCESS |

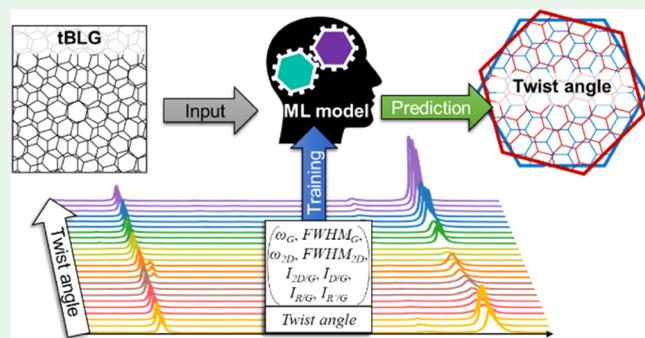
Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** With the increasing interest in twisted bilayer graphene (tBLG) of the past years, fast, reliable, and non-destructive methods to precisely determine the twist angle are required. Raman spectroscopy potentially provides such a method, given the large amount of information about the state of the graphene that is encoded in its Raman spectrum. However, changes in the Raman spectra induced by the stacking order can be very subtle, thus making the angle identification tedious. In this work, we propose the use of machine learning (ML) analysis techniques for the automated classification of the Raman spectrum of tBLG into a selected range of twist angles. The ML classification proposed here is low computationally demanding, providing fast and accurate results with a ~99% agreement with the manual labeling of the spectra. The flexibility and noninvasive nature of the Raman measurements, paired with the predictive accuracy of the ML, is expected to facilitate the exploration of the emerging research field of twisted van der Waals heterostructures. Moreover, the present work showcases how the currently available open-source tools facilitate the study and integration of ML-based techniques.

**KEYWORDS:** bilayer graphene, twisted, Raman spectroscopy, machine learning, artificial intelligence



## INTRODUCTION

Research on twisted bilayer graphene (tBLG) has gained interest in the past few years, sparked in part by the finding of superconductive states at small twist angles.<sup>1</sup> This interest has also expanded to include twisted stacks of other van der Waals (vdW) materials, giving rise to the research field known as twistronics.<sup>2–6</sup> However, determining the twist angle of tBLG or other vdW stacks is not a trivial task. High-resolution microscopy techniques like transmission electron microscopy (TEM)<sup>7–13</sup> or scanning probe microscopies (SPM)<sup>14–17</sup> provide the most accurate angle determinations, with precisions below ~0.01°. The downside is that measurements are time-consuming and require either a free-standing sample or supported on a conductive substrate. Moreover, they provide very local information on sub-micron-sized areas, whereas the twist angle can vary considerably within a few micrometers.<sup>18,19</sup> These techniques are thus not suitable for practical applications, which require large-area characterizations on arbitrary substrates and in relatively short times. Similar accuracies can be obtained from transport measurements of devices under magnetic fields at low temperatures,<sup>1</sup> but the complexity of the measurements and their limitation to small areas make them unsuitable for practical applications. Low-energy electron diffraction (LEED) can provide information about the number of layers and their stacking orientations with sub-degree accuracy for areas larger

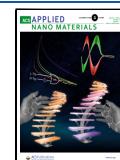
than those obtained from TEM and SPM, although generally requires a conductive substrate and high vacuum conditions.<sup>2,18–20</sup> In the case of isolated tBLG grains with hexagonal shapes, it is also possible to determine the twist angle by optical microscopy.<sup>21</sup> This technique is simple and provides an accuracy of ~1°, but it cannot be applied to the case of tBLG with high coverages or with nonhexagonal shapes, as generally occurs for graphene grown by chemical vapor deposition (CVD).<sup>19</sup>

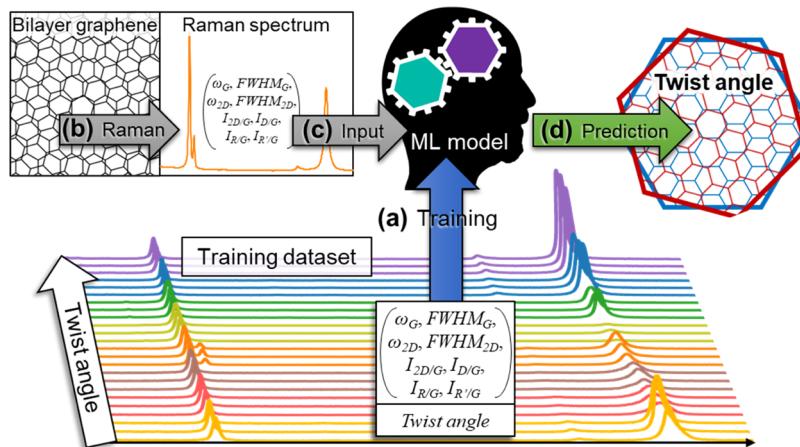
Raman spectroscopy is a noninvasive analysis technique that provides flexibility about the kind of substrate and environment in which the measurement is done, along with the possibility to examine relatively large areas in reasonably short periods of time. It has been widely used in the characterization of graphene, providing extensive information about its characteristics, quality, and electronic state.<sup>22</sup> In particular, the Raman spectrum of graphene depends on the number of layers and on their relative orientation.<sup>7–9,11,21–26</sup> The case of Bernal (ABA) and

Received: November 18, 2021

Accepted: January 6, 2022

Published: January 18, 2022





**Figure 1.** Schematic of the process involved in the ML determination of the twist angle of BLG. (a) ML model is trained using data extracted from a set of Raman spectra for which the twist angle was previously determined (training dataset). (b) Raman features are collected from a different tBLG sample for the determination of its twist angle. (c, d) The trained model can then predict the twist angle of other BLG samples from their Raman signature.

rhombohedral (ABC) few-layer graphene (FLG) has been extensively studied owing to their high occurrence in natural graphite. For such ordered graphene, both the shape of the Raman 2D band and the Raman modes in the spectral region at  $1650\text{--}1800\text{ cm}^{-1}$  are good indicators of the stacking order.<sup>24,25,27</sup> For the case of tBLG, it is also possible to determine the twist angle by Raman spectroscopy, providing even sub-degree accuracies for certain angle ranges.<sup>28</sup> In general, determining the twist angle requires the simultaneous comparison of several features of the Raman spectrum. However, the increased complexity of the spectra can greatly difficult this task.<sup>7,8,10,11,18,26</sup> This complexity is especially evident for low twist angles in which the tBLG undergoes a lattice reconstruction, which has been recently mapped by nano-Raman spectroscopy.<sup>29</sup> Although information of the twist angle of tBLG is encoded in the Raman spectrum, the variations for different angles can be very subtle, often involving small changes in the positions, widths, and/or intensity ratios of the different peaks. These differences are frequently imperceptible at first glance and might easily be overlooked, requiring a careful inspection of the spectrum.<sup>11,21,28</sup> Observing some of them rely on the obtention of high-quality spectra with large signal-to-noise ratios, as it is the case for the subtle changes in the spectral region between the G and 2D bands or at low Raman shifts.<sup>21,26,28,30</sup> Moreover, the Raman spectrum can be significantly altered by the electronic, chemical, and structural states of the graphene, due to the presence of defects, strain, or doping.<sup>22,31–35</sup> This further difficult to determine the twist angle, as each angle is represented by a broad dispersion of spectra instead of by a single ideal spectrum. All of these make the manual determination of the stacking order a time-consuming and nonpractical task. As the twist angle of CVD-grown tBLG can greatly vary point to point,<sup>18,19</sup> it is necessary to introduce automated Raman-based analysis methods for the fast and reliable determination of the stacking order for large areas.

Machine learning (ML) comprises a series of techniques that rely on statistics to categorize new data based on the training of a model (supervised ML), or to find patterns in uncategorized data (unsupervised ML).<sup>36</sup> ML-based methods are being actively introduced in different aspects of the research and handling of 2D materials.<sup>37–40</sup> Recently, ML has been proved to be effective in determining the twist angle of simulated portions

of the Raman spectrum mainly around the G band.<sup>41</sup> The structural relaxation at low twist angles to minimize the AA-stacking area leads to phononic reconstructions, expected to produce evident changes in the Raman G band that include splitting and shift.<sup>41,42</sup> However, real Raman spectra can differ significantly from the simulated ones, with the positions, widths, and relative intensities of the most characteristic Raman peaks being heavily affected by things like strain and doping.<sup>22</sup> ML has also been used to identify selected twist angles of BLG made by artificial stacks of single-layer graphene (SLG).<sup>43</sup> Clustering, an unsupervised ML paradigm, has also been used to assemble uncategorized Raman data into groups roughly corresponding to different kinds of BLG, but without sensitivity toward the twist angle.<sup>44</sup>

In this work, we propose an easy, fast, and low computationally demanding ML-based analytical method to determine the stacking order of tBLG from its Raman spectrum. As shown in Figure 1, this method involves extracting selected features of the Raman spectrum of tBLG with enough information to train an ML model able to infer the twist angle within some predefined ranges. The accuracy of the ML predictions exceed a 99% agreement with the manual labeling of the spectra. The combination of this predictive accuracy with the flexibility and noninvasive nature of the Raman measurements is expected to facilitate and accelerate the emerging research field of twisted vdW heterostructures. Given the prevalence of Raman signals in vdW materials and their reported sensitivity to the twist angles, the proposed method can be applied to virtually any heterostructure. This has practical implications for the advance and understanding of the recent findings in the field of twistronics. The present work also introduces a practical example of how currently existing available open-source tools allow easy and effective integration of ML-based techniques with conventional research methods, potentially increasing the efficiency of the research.<sup>45</sup>

## EXPERIMENTAL SECTION

The BLG was grown by CVD on Cu–Ni thin films supported by c-plane sapphire substrates, at a temperature of  $1085\text{ }^\circ\text{C}$ .<sup>19</sup> To test the validity of the ML predictions for graphene grown on a more commonly used catalyst, isolated graphene grains were also grown at  $1050\text{ }^\circ\text{C}$  on an  $80\text{ }\mu\text{m}$  thick Cu foil (Nilaco Co.). However, the data obtained for

graphene grown on Cu foils was not used in the training dataset. After the CVD, graphene was transferred to a  $\text{SiO}_2$  substrate using a PMMA support film. Confocal Raman spectroscopy was performed on the transferred graphene using a Nanofinder 30 spectrometer (Tokyo Instruments, Inc.), with a laser excitation of 532 nm ( $\sim 1.6 \text{ mW}$ , 0.5 s exposure), a 100 $\times$  objective (0.9 NA), and a 600 gr/mm grating.

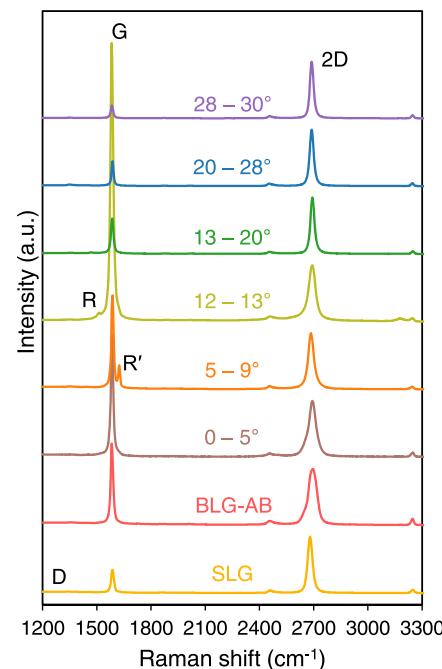
The complete training dataset was generated from  $\sim 6000$  individual Raman spectra collected from graphene samples obtained from eight different CVD batches, all of them using Cu-Ni catalyst thin films. The open-source Python library scikit-learn was used to perform the ML modeling.<sup>45</sup> A simplified version of the code and dataset are available under an open-source license and can be found at [https://github.com/2DMatSci/ML\\_Raman\\_tBLG](https://github.com/2DMatSci/ML_Raman_tBLG). The dataset was used without any processing except for the support vector machine (SVM), multilayer perceptron (neural network), and k-nearest neighbors models, for which each feature was centered around the mean and then scaled to unit variance. To optimize the hyperparameters of each of the models, the original dataset was split into training and test subsets. The new training subset was then used to find the best hyperparameters via stratified k-fold cross-validation,<sup>46</sup> while the held-out test subset was only used to evaluate the performance of the different models.<sup>47</sup> The accuracy of the trained models was finally determined by comparing the predictions of the model and the manual labeling of a different set of Raman mappings not used during the training process.

Prior to the unsupervised clustering, the data for SLG and BLG-AB was removed from the training set, and each feature was scaled by its maximum absolute value. The dataset was then projected onto a two dimensional space by a neighborhood components analysis transformation, to differentiate the manually determined classes.<sup>48</sup> The clustering algorithm employed was a spectral clustering, which allows us to decide the number of clusters beforehand.<sup>49</sup>

## RESULTS AND DISCUSSION

**Determining the Twist Angle of tBLG by Raman Spectroscopy.** Several characteristics of a given graphene sample can be obtained by the inspection of its Raman spectrum, including the strain, doping, presence of defects, or thickness.<sup>22,35</sup> The twist angle of tBLG can also be roughly estimated from the Raman characteristics.<sup>11,21</sup> Figure 2 shows the average spectra for CVD-grown SLG, AB-stacked BLG (BLG-AB), and for tBLG with different twist angle ranges, measured after transfer to  $\text{SiO}_2$  substrates. The averages were obtained from the  $\sim 6000$  Raman spectra shown in Figure S1, which were used without any further preprocessing, such as noise reduction or background subtraction. These spectra were collected from graphene grown in different CVD batches, to obtain the wider possible range of angles, and to counteract the effect of small sample to sample differences and unintentional variations of strain or doping levels.<sup>19</sup> Careful inspection of each individual Raman spectrum allowed us to assign them to one of the eight different classes shown in Figure 2. This classification was done by comparing the specific characteristics of each spectrum with spectra reported in the literature and precisely labeled by alternative experimental techniques.<sup>11,21,50</sup> The whole range of twist angles from  $0^\circ$  to  $30^\circ$  is covered by the classification, except for twist angles between  $9^\circ$  and  $12^\circ$  for which we could not find an unambiguous way to label the Raman spectra. As will be discussed below, it should be possible to properly label such angles either by including low-frequency Raman modes in the analysis<sup>21</sup> or by directly measuring the twist angles and comparing them with the current Raman data. This manual process to classify the Raman spectra is laborious and prone to errors, and as the twist angle can greatly vary over the BLG surface,<sup>19</sup> it is difficult to scale it to large CVD-grown BLG areas.

Once the spectra are classified, a few measurable features were extracted from each spectrum (Figure 1).<sup>43,44</sup> The selected



**Figure 2.** Average spectra of SLG, BLG-AB, and tBLG with different twist angles, measured using a 532 nm excitation. Each spectrum corresponds to one of the classes included in the training dataset for the ML models. The twist angles were determined by comparison with the existing literature. The spectra are normalized to the 2D band intensity and vertically shifted for clarity, with the most characteristic Raman bands being indicated with labels.

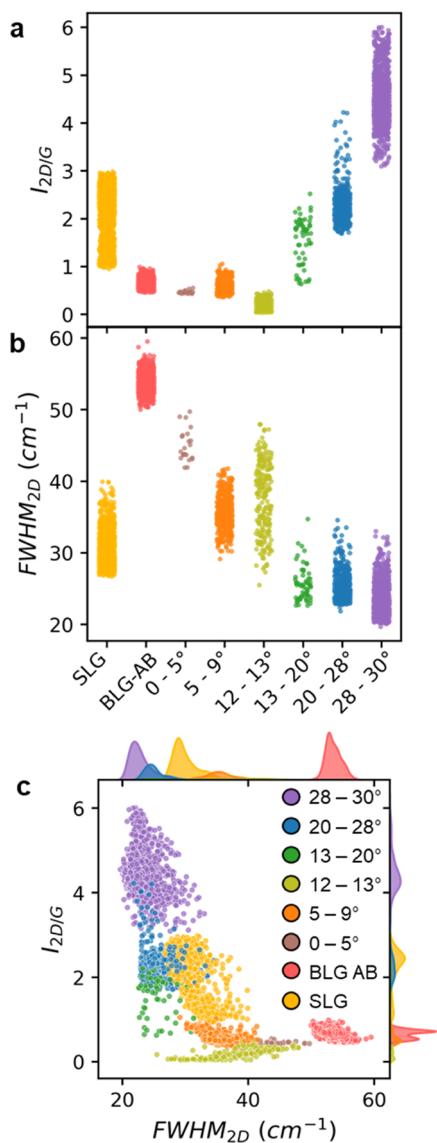
features include the shift ( $\omega$ ) and full width at half-maximum (FWHM) of each of the main graphene Raman bands (G and 2D) and the relative intensities of the 2D, D, R, and R' bands with respect to the G band ( $I_{2D/G}$ ,  $I_{D/G}$ ,  $I_{R/G}$ , and  $I_{R'/G}$ ). The relative band intensities were used instead of the actual intensities, as they are less prone to variations from different measurements. As shown in Figure 2, the R and R' bands can be found around the G band for certain twist angles. The ratios  $I_{R/G}$  and  $I_{R'/G}$  were only collected for spectra with noticeable R and R' bands, and considered to be 0 for the rest. All of the features are represented in Figure S2, with their average values shown in Table 1 for the different twist angle ranges. The bottom row of Table 1 corresponds to the values for the mutual information of each feature and the assigned classes for the twist angle. This is a measure of the shared information between the corresponding feature and the twist angle.<sup>51</sup> Hence, the features with the largest values are expected to be the most relevant to determine the twist angle. Taken separately, none of these features allow to unambiguously determine the twist angle. This even happens for the two features with the largest mutual information, for which Figure 3a,b shows the corresponding values of all of the spectra separated by the mentioned classes. As seen in Figure 3a, the intensity ratio  $I_{2D/G}$  is not sensitive to angles between  $0^\circ$  (BLG-AB) and  $9^\circ$ , with data points for these classes overlapping in the range  $0.5 \lesssim I_{2D/G} \lesssim 1$ . On the other hand, Figure 3b shows that FWHM<sub>2D</sub> fails in discriminating twist angles between  $13^\circ$  and  $30^\circ$ , for which most of the data points lie in the range  $20 \text{ cm}^{-1} \lesssim \text{FWHM}_{2D} \lesssim 30 \text{ cm}^{-1}$ .

The simultaneous use of different features is expected to improve the sensitivity, e.g., by pairing features as in Figure S3. The specific case of  $I_{2D/G}$  and FWHM<sub>2D</sub> (Figure 3c) shows how each of the classes forms a cluster. The presence of these clusters

**Table 1.** Average Values of the Raman Features for Each of the Twist Angle Ranges Used to Train the ML Models<sup>a</sup>

twist angle	$\bar{\omega}_G$ (cm <sup>-1</sup> )	$\overline{\text{FWHM}}_G$ (cm <sup>-1</sup> )	$\bar{\omega}_{2D}$ (cm <sup>-1</sup> )	$\overline{\text{FWHM}}_{2D}$ (cm <sup>-1</sup> )	$\overline{I_{2D/G}}$	$\overline{I_{D/G}}$	$\overline{I_{R/G}}$	$\overline{I_{R'/G}}$
SLG	1587.02 ± 3.17	14.82 ± 1.45	2680.27 ± 3.31	30.15 ± 2.2	2.18 ± 0.5	0.07 ± 0.04		
BLG-AB	1583.35 ± 2.08	15.17 ± 1.15	2694.60 ± 2.5	53.55 ± 1.25	0.66 ± 0.1	0.04 ± 0.01		
0–5°	1584.24 ± 0.65	15.34 ± 0.79	2693.05 ± 1.21	45.03 ± 2.21	0.46 ± 0.03	0.03 ± 0.003		
5–9°	1586.77 ± 2.05	13.31 ± 1.69	2685.43 ± 2.39	35.55 ± 2.3	0.57 ± 0.14	0.04 ± 0.01		0.25 ± 0.12
12–13°	1582.91 ± 3.23	15.89 ± 2.29	2691.08 ± 5.66	37.65 ± 4.92	0.19 ± 0.12	0.01 ± 0.004	0.04 ± 0.01	
13–20°	1584.98 ± 1.83	15.41 ± 1.73	2692.69 ± 2.01	25.66 ± 2.42	1.55 ± 0.48	0.04 ± 0.01	0.05 ± 0.01	
20–28°	1587.04 ± 1.2	12.03 ± 1.45	2688.75 ± 2.43	25.20 ± 1.88	2.26 ± 0.36	0.07 ± 0.02		
28–30°	1583.44 ± 1.34	15.39 ± 1.28	2688.84 ± 2.61	23.03 ± 2.01	4.47 ± 0.49	0.07 ± 0.03		
MI	0.390	0.295	0.769	1.220	1.266	0.613	0.182	0.284

<sup>a</sup>The bottom row shows the mutual information between each of the features and the classes of the training dataset.



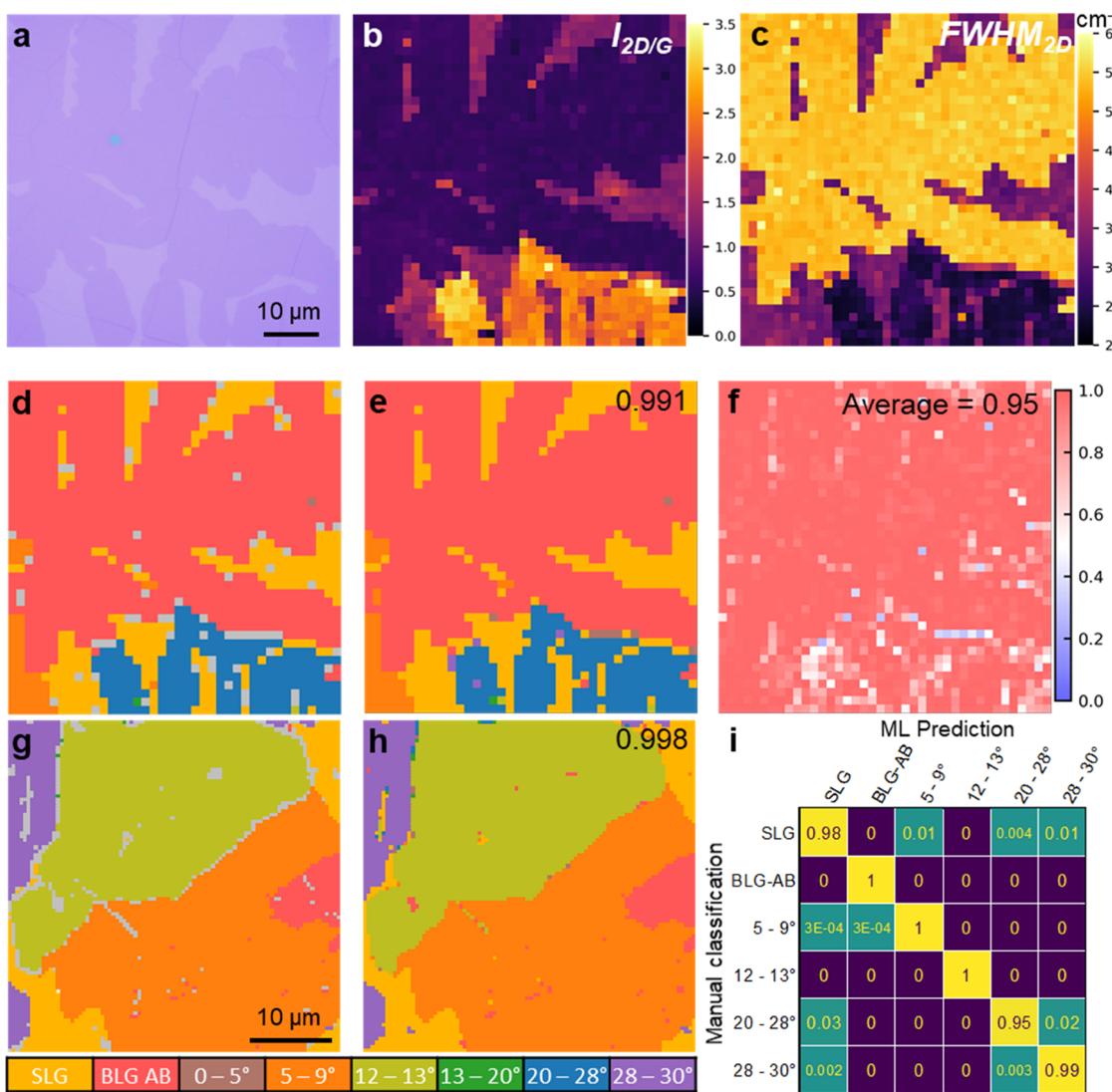
**Figure 3.** (a, b) Distribution of  $I_{2D/G}$  (a) and  $\text{FWHM}_{2D}$  (b) for the different classes. (c) Scatter plot of  $\text{FWHM}_{2D}$  vs  $I_{2D/G}$  for each of the classes. The probability distributions are included at the top ( $\text{FWHM}_{2D}$ ) and right ( $I_{2D/G}$ ) sides of (c).

provides a faster way to label new experimental data compared with the manual inspection of the whole spectrum. Although the differentiation of the classes is now more evident than in Figure 3a,b, some of the clusters still overlap (Figure S4), preventing the unequivocal determination of the twist angle by simply using

two features. The overlap of the clusters is reduced by further increasing the number of features ( $n$ ), as they will now be embedded in an  $n$ -dimensional hyperspace. By combining all of the available features shown in Figure S2 ( $n = 8$ ), it is thus possible to increase the accuracy in the determination of the twist angle. To confirm this, we tried to determine the classes of the graphene region shown in Figure 4a. The Raman maps for  $I_{2D/G}$  and  $\text{FWHM}_{2D}$  are shown in Figure 4b,c, with the maps for the whole set of features in Figure S5. The contrast in the Raman mapping images indicates the existence of different graphene regions, comprising different numbers of layers and twist angles. Combining the data from all of these maps allows us to manually assign a class to each of the spectra, as shown in Figure 4d. Most of the spectra of the mapping could be classified this way, except for the few points marked in gray that are usually found in limits between two different regions. The manual assignment was also performed at other areas, which together include all of the classes (Figure S6). Following this method, it was thus possible to distinguish not only between SLG and BLG but also different twist angles for the latter. But this approach still requires a considerable manual input and a careful inspection of the Raman data in an  $n$ -dimensional hyperspace. Moreover, it can also hinder the identification of finer details that might potentially go unnoticed, thus preventing to reach a higher precision in the determination of the twist angle. An alternative that can improve the manual labeling is to simplify the problem by decreasing the dimensionality of the feature hyperspace.<sup>48</sup> When performed carefully, the projection of the original features into a lower-dimensional space can retain most of the original information while making the analysis simpler (Figure S7).

#### Training an ML Model to Determine the Stacking Order.

To increase the processing speed and precision of the labeling of the Raman spectra, we applied ML analysis methods. The  $n$ -dimensional vector of data conformed by the chosen features of a Raman spectrum can be easily interpreted and processed by relatively simple ML models. As we are using supervised learning, the first step is training the chosen model using Raman data that has already been labeled (Figure 1a). The models used here are multiclass classifiers for which the training data is categorized into a finite set of classes, corresponding to the previously shown SLG, BLG-AB, and tBLG with several twist angle ranges (Figure 2). The dataset used to train the model was the same used in the manual classifications, by assigning a class to the spectra in Figure S1 and extracting the relevant Raman features (Figures S2 and S3). The training ultimately consists of feeding the ML model with both the values of the Raman features and the manual classification of each spectrum. Each model has a different set of internal variables, known as hyperparameters, that have a large impact on the

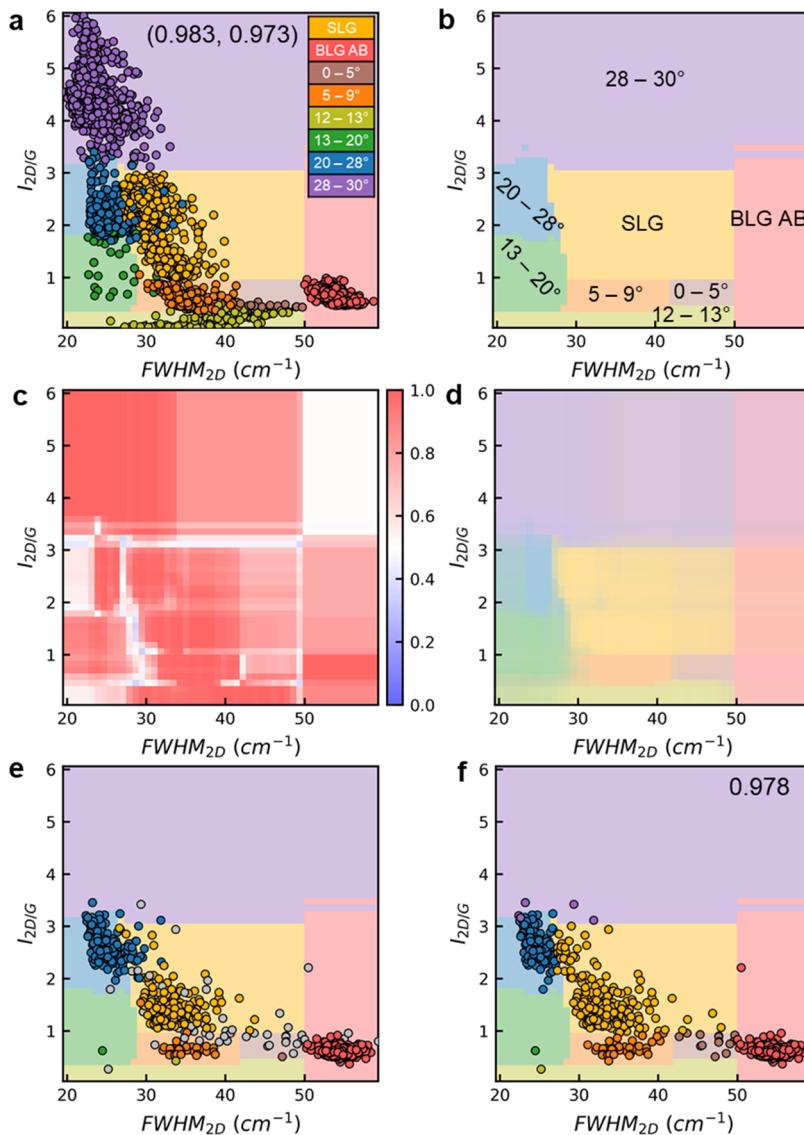


**Figure 4.** (a) Optical image of CVD-grown graphene consisting of SLG (light) and BLG (dark) areas. (b, c) Raman mappings of  $I_{2D/G}$  (b) and  $FWHM_{2D}$  (c) for the area in (a). (d, e) Mapping of the twist angle determined by manually inspecting the characteristics of the Raman spectra (d) and by a trained random forest ML model (e). Predictions in (e) match  $\sim 99.1\%$  (value in the top right corner) with the manual determination in (d). This value excludes the points in (d) that could not be manually labeled (gray). (f) Confidence score of the random forest model to each of the predictions shown in (e). The value in the top right corner of (f) shows the averaged value of the confidence. (g, h) Manual classification (g) and ML prediction (h) of a different area (accuracy  $\sim 99.8\%$ ). (i) Confusion matrix for the ML predictions in (e) and (h), normalized by rows.

performance of the model but cannot be derived during the training process. As specified in the [Experimental section](#), these hyperparameters were optimized for each of the different models by cross-validation.

The trained model is then ready to predict the twist angle of tBLG from its Raman spectrum ([Figure 1b–d](#)). Owing to the inclusion of a class for SLG in the training dataset, the model is also able to discriminate between BLG and SLG. It is important to note that all of the ML predictions done in this work are for data that was not present in the training set, being collected from different CVD batches. This ensures that the model can generalize to new unseen data and that the high predicting accuracies are not due to problems like overfitting of the training data.<sup>46</sup> To carry out the prediction, the  $n$ -dimensional vector containing the Raman features of data unseen by the trained model is used as input. In the case of the area shown in [Figure 4a](#), the trained model was supplied with the maps containing the Raman features for each point ([Figures 4b,c and S5](#)). The

prediction using a random forest model is presented in [Figure 4e](#), matching  $\sim 99.1\%$  with the values obtained by manual labeling ([Figure 4d](#)). Several areas were studied, as it was not possible to find a single area of graphene containing all of the different classes that can be predicted. [Figure 4g,h](#) shows the manual classification and the model prediction of another area. Both areas together include all of the classes but  $0–5^\circ$  and  $13–20^\circ$ , which are the least frequent twist angles. All of the classes are represented when including the areas shown in [Figure S6](#), each collected from a different sample. The average accuracy of the model predictions for the areas in [Figures 4e,h and S6](#) is  $\sim 99.0\%$ , obtained as a weighted average to account for differences in the size of the areas. [Figure 4i](#) shows the confusion matrix for the predictions in [Figure 4e,h](#), while the confusion matrix in [Figure S8](#) summarizes the predictions from all of the measured areas. The confusion matrix evidences that all of the classes achieve good accuracy ratios, with most of the mistakes made by the ML model being due to confusion between the SLG



**Figure 5.** (a) Scatter plot of  $FWHM_{2D}$  vs  $I_{2D}/G$  used for training a random forest classifier, with the numbers in parenthesis representing the accuracy of the model on the train and test sets, respectively. The colored background shows the corresponding decision boundary of the model. (b) Decision boundary for the random forest classifier trained with the data shown in (a). (c) Confidence score of the model for each point of the decision boundary. (d) Decision boundary with the colors determined according to the weighted average of the probabilities assigned to each of the classes. (e, f) Decision boundary including scatter plots for the maps in Figure 4b,c with the manually assigned classes (e) and those predicted by the model (f). The number in the top right corner of (f) represents the accuracy of the model with respect to the manual labeling, excluding the points in (e) that could not be manually labeled (gray colored).

and tBLG with twist angles of 20–30°. The model used in these calculations is a random forest classifier, being a commonly employed model that performs well on a wide range of problems, and it is relatively easy to understand.<sup>52</sup> The performance of other ML algorithms was also studied, with the results of the predictions being shown in Figure S9. After adjusting their corresponding hyperparameters, all of the selected models clearly outperform the results obtained from the dummy models that were also included in Figure S9.

With some exceptions, during the prediction, the ML models assign a probability for the input data to belong to each of the classes. The class predicted by the model is that with the highest probability, known as the confidence score and which can be interpreted as a confidence of the model on the prediction. Figure 4f shows the confidence score of the random forest model to the predictions in Figure 4e. The high value of the average

confidence score, ~0.95, is consistent with the high accuracy of the prediction. The areas with the lower probabilities mostly coincide with areas in which the manual labeling was not possible, such as in the boundaries between different stacking regions, and with areas in which the predictions are more likely to disagree with the manual labeling.

To better understand the training and prediction processes, a random forest model was trained using only the pair of features with the largest values of mutual information, *i.e.*,  $I_{2D}/G$  and  $FWHM_{2D}$ , as shown in Figure 5a. This decreases the accuracy of the predictions but allows to visualize the decision boundary of the model, which divides the feature space ( $FWHM_{2D}$ ,  $I_{2D}/G$ ) into regions for which different classes are predicted. The resulting decision boundary is displayed as the colored background of Figure 5a,b, obtained by plotting the class predicted by the model on each point of the feature space. The

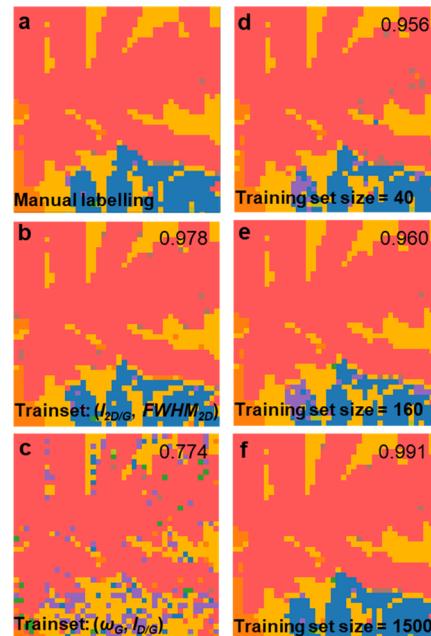
fact that the decision boundary appears simple indicates that the model is not overfitting the training set.<sup>46</sup> This is in accordance with the high accuracy values obtained on both the train and test sets (values shown in parenthesis in Figure 5a), and with the high accuracy of the predictions of Figures 4e,h and S6. For comparison, in Figure S10, there are examples of decision boundaries for overfitted and underfitted SVM models. The overfitted model shows a complex decision boundary that closely follows the data from the training set. Consequently, it has a perfect accuracy on the training set, but the much lower value obtained in the test set indicates that it does not generalize well to predict new results. Conversely, the underfitted model shows extremely simple boundaries, with similarly low accuracy values for both the training and test sets. These results stress the importance of choosing suitable hyperparameters to obtain reliable predictions from the models.

The value of the confidence score for each point of the decision boundary is shown in Figure 5c. Regions with low probability values can be found in areas void of training data, as in the top right region. This does not affect the prediction accuracy of the model and only reflects the uncertainty in areas of the feature space for which experimental data is not physically allowed (e.g., graphene does not concurrently show broad and intense 2D bands). Especially problematic for the model predictions are the low probability values in boundaries in which training data for different classes are close to each other. This highlights the need for properly trained models that can maintain high accuracies in such regions. The uncertainty of the model is also represented in Figure 5d, for which the color of each point is a weighted linear combination obtained from the probabilities assigned to each class. In contrast to the sharp boundaries of the decision boundary (Figure 5b), the boundaries between different classes are now blurred, reflecting the low confidence of the model for the predictions in these regions.

Figure 5e,f shows how the model trained only with  $I_{2D/G}$  and  $FWHM_{2D}$  classifies the Raman mapping from Figure 4a. In Figure 5e, the classes are manually assigned (Figure 4d), while Figure 5f shows the result of the ML prediction (additional details are shown in Figure S11). The class of some of the dots in Figure 5e do not correspond with the region of the decision boundary on which they stay. This occurs for some of the dots assigned a 20–28° twist (blue dots) but located at the lower part of the 28–30° region (purple area) and at the top left part of the SLG region (yellow area). Considering how the decision boundary plot is constructed, the prediction from the model trivially follows the regions delimited by the boundaries, and hence those dots are being mislabeled by the model and decrease its accuracy. This mislabeling mainly occurs near the boundaries delimiting the regions, which have low confidence scores (Figure 5c). A minor thing to note is that the model can assign a class for the spectra that could not be manually classified (gray dots in Figure 5e). Although it is not possible to ascertain the accuracy of the model in these cases, the outcome seems reasonable for this case when inspecting the mapping in Figure S11. Ultimately, the confidence score of each prediction can be used as a threshold value to discard ML predictions with low scores.

**Limitations of the ML Models.** It is important to understand the limitations of the proposed ML approach and the factors that influence the accuracy of the predictions. One of the most important factors in the ML analysis is the quality of the training data. Here, we analyze the impact of the chosen Raman

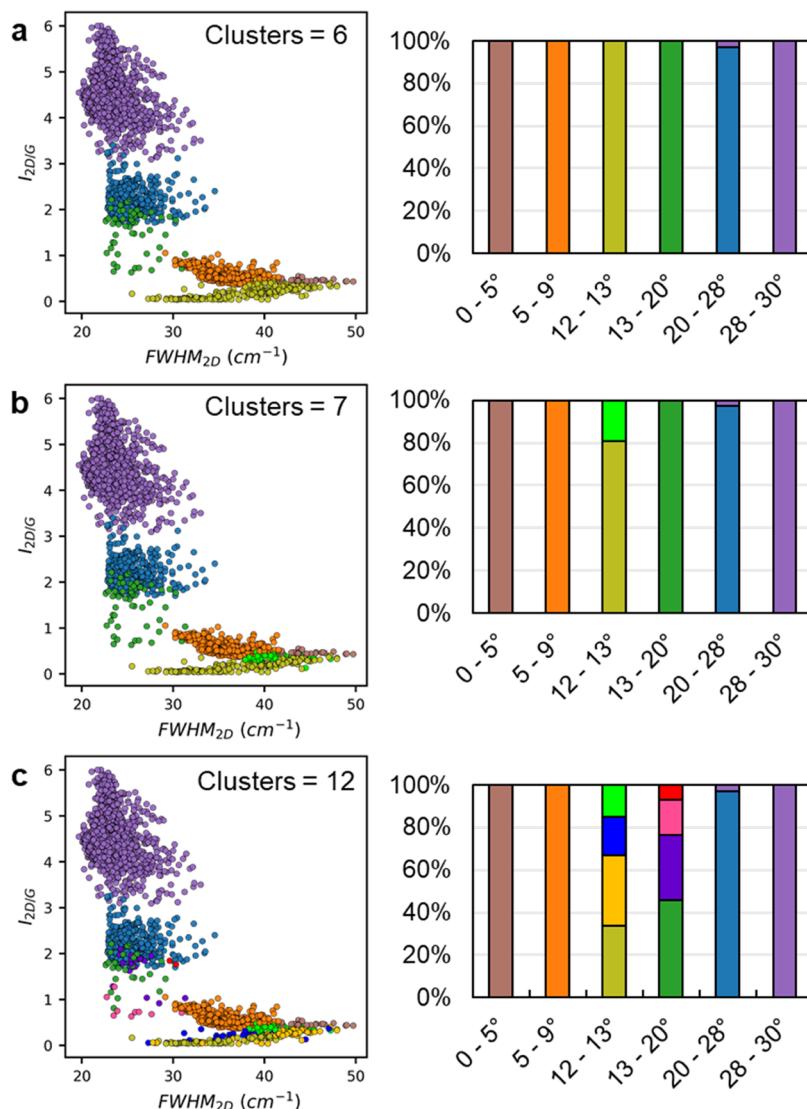
features and of the size of the training set. As previously mentioned in Table 1, not all of the features have the same relevancy on the ML accuracy. To see this in a practical case, we compare the results of the manual labeling (Figure 6a) with the



**Figure 6.** Effect of the features and the size of the training dataset on the predictions of a random forest model. (a) Manual labeling. (b, c) Predictions for a model trained using only the features ( $FWHM_{2D}$ ,  $I_{2D/G}$ ) (b), and ( $\omega_G$ ,  $I_{D/G}$ ) (c). (d–f) Predictions for models trained with dataset sizes of 40 (d), 160 (e), and 1500 (f) instances. The numbers in the top right corner represent the accuracy of the model in each case for this specific area.

predictions of models trained with only two features. When choosing two features with large values of mutual information, such as  $I_{2D/G}$  and  $FWHM_{2D}$ , the model retains a relatively good accuracy of ~97.8% (Figures 6b and S11), only slightly lower than the value obtained when using the complete set of features (~99.1%). In contrast, selecting less relevant features, such as  $\omega_G$  and  $I_{D/G}$ , decreases the accuracy to ~77.9% (Figures 6c and S11). This highlights the importance of selecting features that are adequate to the problem to obtain the best possible outcome.

The ideal size of the training dataset depends on the specific problem and of the model used, with the most complex cases requiring datasets with several thousands of instances. Our training set was obtained from ~6000 spectra, with the accuracy being expected to decrease for smaller sets. This is evidenced in the evolution of the predictions in Figure 6d–f for increased sizes of the training set. Figure S12a shows the accuracy of a random forest model for different sizes of the training set. The accuracy was obtained for the same mappings shown in Figures 4e and S6. Even for the smallest sets tested, with 40–160 instances, relatively high accuracies of ~91% were attained. This shows that modest set sizes can still provide reasonably good predictions of the stacking order. As shown by the error mappings in Figure S12, most of the errors of these smaller sets arise from the confusion between classes with similar spectra, e.g., between BLG-AB and 0–5°, or between 20–28 and 28–30°. Thus, for small sets, the accuracy can be potentially increased by focusing on obtaining additional training instances of such classes. A steep rise in the accuracy is observed when the



**Figure 7.** Clustering results of the tBLG classes of the dataset for (a) 6, (b) 7, and (c) 12 clusters. The left column shows the  $\text{FWHM}_{2D}$  vs  $I_{2D/G}$  scatter plot of the dataset, with each color corresponding to one of the found clusters. The right column shows how the original classes relate to the found clusters.

training set is increased, reaching a steady value of  $\sim 91\%$  for sizes over  $\sim 1500$ . This indicates that the best strategy to further increase the accuracy for these set sizes is to improve the quality of the dataset rather than continue increasing its size.

The high sensitivity of the Raman spectra to the state of graphene can also limit the accuracy of the ML determination of the twist angle. This is because the spectra included in the training dataset are acquired under certain specific conditions, and hence the model might be unable to properly classify spectra acquired under different conditions. It is thus important to establish the limitations of the training and determine when the model can start failing. One of the possible constraints is using a different substrate for the graphene, as the substrate is known to affect the parameters on which the training of the model is based.<sup>31</sup> In our case the training dataset was obtained for graphene supported on  $\text{SiO}_2$ , and hence the reliability of the model is expected to decrease for other substrates. As an example, Figure S13a,b shows the predictions for graphene supported on c-plane sapphire and on quartz, respectively. While for sapphire the results are still good, the performance for the quartz substrate decreased significantly, thus requiring an

additional training set for such substrate. The different processes to which graphene can be subjected, such as during device fabrication, can also potentially alter its Raman spectrum. Figure S13c shows that the performance of the ML model on a BLG field-effect transistor (FET) maintains a reasonable accuracy. Another important factor can be the growth conditions of graphene. While our model was trained using graphene grown on thin Cu–Ni films supported on sapphire,<sup>19</sup> the graphene of Figure S13d was grown on a commercial Cu foil under a different CVD condition. This results in differences in the graphene as evidenced in the optical images, with the BLG of Figure S13d being present only at the center of isolated SLG grains, instead of the high-coverage BLG used for the training set. The accuracy of the ML prediction decreases below 95% for this different graphene, with most of the errors occurring at the boundary between areas with different stacking. It should be noted that the specific classes present in this area (SLG, 20–28 and 28–30°) are those that the model is more prone to mix up (Figure S8), making this a borderline case.

Finally, it is worth mentioning that this sensitivity of the Raman spectra toward the specific condition of the graphene can

also be used as an opportunity to expand the capacities of the ML model. Thus, with appropriate training sets, it seems conceivable to train models that can also determine the amount of strain or doping of the graphene,<sup>32</sup> or even the stacking order of graphene with respect to other 2D layered materials used as substrates.<sup>53</sup>

### Improvement of the Precision of the Twist Angle Determination.

Although the method proposed here provides high accuracies for the determination of the twist angle of tBLG, the angle ranges of the different classes are still broad for most real applications. It is thus important to understand if smaller ranges for the predicted angles can be potentially achieved. Using clustering analysis, we first tried to look in our training set for hidden information that can arise from different twist angles within each of the classes. Clustering comprises a range of unsupervised learning techniques that allow finding groups (clusters) of instances in the dataset without prior knowledge of the classes.<sup>43,44,49</sup> Instances pertaining to the same cluster are expected to be more similar to each other than to those in other clusters. We chose to use spectral clustering, an algorithm that allows us to decide the number of clusters, hence allowing us to start from the same number of classes that we have determined. For simplicity, the data of SLG and BLG-AB were removed from the dataset. Hence, the dataset now includes only the six classes of tBLG. Figure 7a shows the results after clustering the dataset using six clusters. These six clusters mainly correspond to the manual classification, with the only difference being a small mixture of the manual classes 20–28 and 28–30°. Except for this, each of the original classes is represented by a single cluster, as can be appreciated in Figure S14. Increasing the number of clusters resulted in some of the classes being subdivided into several clusters, as shown in Figure 7b,c for 7 and 12 clusters, respectively. Although it is neither possible to determine the nature of these new clusters nor to assert that they effectively correspond to well-defined subranges of twist angles, the fact that the original classes do not share clusters seems to indicate that they might originate in differences in the twist angle within the given class. The different metrics shown in Figure S15 indicate that the ideal number of clusters might range between 6 and 11. This opens an opportunity to increase the precision of the angle ranges of the supervised ML models by increasing the number of classes of the training dataset.

One way to attain this is to use a more precise determination of the twist angles of the training dataset, for example, by TEM measurements.<sup>9</sup> This would allow us to narrow down some of the angle ranges of the training classes, hence supplying more information for the ML models. In this sense, the R and R' bands seem promising candidates to increase the accuracy in the range of 5–20°, as their shift and intensities are known to be affected by the twist angle.<sup>9,54</sup> Our data also point to this possibility, with the shift of the R band changing for different spectra as shown in Figure S1. Correspondingly, some other overlooked trends in the features might be used by the models to increase the precision while keeping high accuracies. Another unexplored possibility to narrow down the predicted angle ranges is to use multiple excitation wavelengths for the Raman measurement. The trend of the characteristics of some Raman bands, such as  $\omega$ , FWHM, and relative intensity, largely depends on the wavelength of the laser excitation.<sup>7,28</sup> This provides additional information that can be used to obtain narrower angles for the classes of the training spectra, with accuracies that can be of the order of 0.1° for certain angle ranges.<sup>28</sup> Similarly, additional Raman bands and the inclusion of different regions of the spectra

can also be used to improve the angle accuracy, especially for certain angle ranges.<sup>26</sup> For example, the low-frequency part of the spectrum ( $\lesssim 200 \text{ cm}^{-1}$ ) shows a variety of modes that are sensitive to both the twist angle and the number of layers.<sup>21,26,55,56</sup> This makes this area of the spectrum highly promising for achieving a more precise determination of the stacking angle and for covering the gap in the range 9–12°.<sup>21</sup> Moreover, the presence of similar low-frequency bands in other 2D materials allows us to expand this method to different vdW heterostructures.<sup>57</sup> Finally, another possibility to enhance the precision can be the use of whole (or parts) of the Raman spectra. This has already been proved with spectra simulated by first principles,<sup>41</sup> and more recently with trainsets limited to a few selected angles of artificially stacked tBLG.<sup>43</sup> However, the method employed here using some selected features of the spectra results in much faster and simpler calculations.

## CONCLUSIONS

We have demonstrated the feasibility to integrate ML-based methods for the automated determination of the twist angle of CVD-grown tBLG. The proposed method provides accuracies that exceed 99% compared with the manual labeling of the twist angles. Moreover, the method is not computationally demanding, providing predictions for whole Raman mappings comprising hundreds of spectra in a matter of seconds even on average desktop computers. The precision of the predicted angle ranges can be possibly increased by improving the quality of the data used for training the model. Finally, the flexibility of the proposed method makes it reasonable to expand it to determine the amount of strain and doping of graphene,<sup>32</sup> and the twist angle of other 2D materials<sup>57</sup> and even of heterostacks.<sup>53</sup> Paired with the flexibility and noninvasive nature of Raman measurements, we expect that this method facilitates the research of tBLG and of its intriguing properties and potential applications. The proposed method has also the potential to be adapted to study the twist angles of different vdW heterostructures, providing a useful and simple analysis tool with practical applications for the advancement and understanding of the nascent field of twistronics.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsanm.1c03928>.

Training dataset, prediction using alternative ML models, effects on the model prediction of the size of the training set and of the features used, and predictions on alternative graphene areas (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

Pablo Solís-Fernández – Global Innovation Center (GIC), Kyushu University, Fukuoka 816-8580, Japan;  [orcid.org/0000-0003-1001-5874](https://orcid.org/0000-0003-1001-5874); Email: [pablosolisfernandez@gmail.com](mailto:pablosolisfernandez@gmail.com)

Hiroki Ago – Global Innovation Center (GIC), Kyushu University, Fukuoka 816-8580, Japan; Interdisciplinary Graduate School of Engineering Science, Kyushu University, Fukuoka 816-8580, Japan;  [orcid.org/0000-0003-0908-5883](https://orcid.org/0000-0003-0908-5883); Email: [h-ago@gic.kyushu-u.ac.jp](mailto:h-ago@gic.kyushu-u.ac.jp)

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acsanm.1c03928>

## Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by the JSPS KAKENHI grant numbers JP18H03864, JP19K22113, 21K18878, and JSPS Transformative Research Areas (A) "Science of 2.5 Dimensional Materials" program (21H05232, 21H05233), JST CREST grant numbers JPMJCR18I1, JPMJCR20B1, and the JSPS A3 Foresight Program.

## REFERENCES

- (1) Cao, Y.; Fatemi, V.; Fang, S.; Watanabe, K.; Taniguchi, T.; Kaxiras, E.; Jarillo-Herrero, P. Unconventional Superconductivity in Magic-Angle Graphene Superlattices. *Nature* **2018**, *556*, 43–50.
- (2) de Jong, T. A.; Jobst, J.; Yoo, H.; Krasovskii, E. E.; Kim, P.; Molen, S. J. van. der. Measuring the Local Twist Angle and Layer Arrangement in van der Waals Heterostructures. *Phys. Status Solidi B* **2018**, *255*, No. 1800191.
- (3) Alexeev, E. M.; Ruiz-Tijerina, D. A.; Danovich, M.; Hamer, M. J.; Terry, D. J.; Nayak, P. K.; Ahn, S.; Pak, S.; Lee, J.; Sohn, J. I.; Molas, M. R.; Koperski, M.; Watanabe, K.; Taniguchi, T.; Novoselov, K. S.; Gorbachev, R. V.; Shin, H. S.; Fal'ko, V. I.; Tartakovskii, A. I. Resonantly Hybridized Excitons in Moiré Superlattices in van der Waals Heterostructures. *Nature* **2019**, *567*, 81.
- (4) Yang, Y.; Li, J.; Yin, J.; Xu, S.; Mullan, C.; Taniguchi, T.; Watanabe, K.; Geim, A. K.; Novoselov, K. S.; Mishchenko, A. In Situ Manipulation of van der Waals Heterostructures for Twistrionics. *Sci. Adv.* **2020**, *6*, No. eabd3655.
- (5) Lee, H. Y.; Al Ezzi, M. M.; Raghuvanshi, N.; Chung, J. Y.; Watanabe, K.; Taniguchi, T.; Garaj, S.; Adam, S.; Gradečák, S. Tunable Optical Properties of Thin Films Controlled by the Interface Twist Angle. *Nano Lett.* **2021**, *21*, 2832–2839.
- (6) Ji, H. G.; Solís-Fernández, P.; Erkiliç, U.; Ago, H. Stacking Orientation-Dependent Photoluminescence Pathways in Artificially Stacked Bilayer WS<sub>2</sub> Nanosheets Grown by Chemical Vapor Deposition: Implications for Spintronics and Valleytronics. *ACS Appl. Nano Mater.* **2021**, *4*, 3717–3724.
- (7) Kim, K.; Coh, S.; Tan, L. Z.; Regan, W.; Yuk, J. M.; Chatterjee, E.; Crommie, M. F.; Cohen, M. L.; Louie, S. G.; Zettl, A. Raman Spectroscopy Study of Rotated Double-Layer Graphene: Misorientation-Angle Dependence of Electronic Structure. *Phys. Rev. Lett.* **2012**, *108*, No. 246103.
- (8) Havener, R. W.; Zhuang, H.; Brown, L.; Hennig, R. G.; Park, J. Angle-Resolved Raman Imaging of Interlayer Rotations and Interactions in Twisted Bilayer Graphene. *Nano Lett.* **2012**, *12*, 3162–3167.
- (9) Lu, C.-C.; Lin, Y.-C.; Liu, Z.; Yeh, C.-H.; Suenaga, K.; Chiu, P.-W. Twisting Bilayer Graphene Superlattices. *ACS Nano* **2013**, *7*, 2587–2594.
- (10) Yeh, C.-H.; Lin, Y.-C.; Chen, Y.-C.; Lu, C.-C.; Liu, Z.; Suenaga, K.; Chiu, P.-W. Gating Electron–Hole Asymmetry in Twisted Bilayer Graphene. *ACS Nano* **2014**, *8*, 6962–6969.
- (11) Chen, X.-D.; Xin, W.; Jiang, W.-S.; Liu, Z.-B.; Chen, Y.; Tian, J.-G. High-Precision Twist-Controlled Bilayer and Trilayer Graphene. *Adv. Mater.* **2016**, *28*, 2563–2570.
- (12) Yoo, H.; Engelke, R.; Carr, S.; Fang, S.; Zhang, K.; Cazeaux, P.; Sung, S. H.; Hovden, R.; Tsien, A. W.; Taniguchi, T.; Watanabe, K.; Yi, G.-C.; Kim, M.; Luskin, M.; Tadmor, E. B.; Kaxiras, E.; Kim, P. Atomic and Electronic Reconstruction at the van der Waals Interface in Twisted Bilayer Graphene. *Nat. Mater.* **2019**, *18*, 448.
- (13) Lin, Y.-C.; Motoyama, A.; Solís-Fernández, P.; Matsumoto, R.; Ago, H.; Suenaga, K. Coupling and Decoupling of Bilayer Graphene Monitored by Electron Energy Loss Spectroscopy. *Nano Lett.* **2021**, *21*, 10386–10391.
- (14) Li, G.; Luican, A.; Lopes dos Santos, J. M. B.; Castro Neto, A. H.; Reina, A.; Kong, J.; Andrei, E. Y. Observation of van Hove Singularities in Twisted Graphene Layers. *Nat. Phys.* **2010**, *6*, 109–113.
- (15) Kerelsky, A.; McGilly, L. J.; Kennes, D. M.; Xian, L.; Yankowitz, M.; Chen, S.; Watanabe, K.; Taniguchi, T.; Hone, J.; Dean, C.; Rubio, A.; Pasupathy, A. N. Maximized Electron Interactions at the Magic Angle in Twisted Bilayer Graphene. *Nature* **2019**, *572*, 95–100.
- (16) Luo, Y.; Engelke, R.; Mattheakis, M.; Tamagnone, M.; Carr, S.; Watanabe, K.; Taniguchi, T.; Kaxiras, E.; Kim, P.; Wilson, W. L. In Situ Nanoscale Imaging of Moiré Superlattices in Twisted van der Waals Heterostructures. *Nat. Commun.* **2020**, *11*, No. 4209.
- (17) Uri, A.; Grover, S.; Cao, Y.; Crosse, J. A.; Bagani, K.; Rodan-Legrain, D.; Myasoedov, Y.; Watanabe, K.; Taniguchi, T.; Moon, P.; Koshino, M.; Jarillo-Herrero, P.; Zeldov, E. Mapping the Twist-Angle Disorder and Landau Levels in Magic-Angle Graphene. *Nature* **2020**, *581*, 47–52.
- (18) Beechem, T. E.; Ohta, T.; Diaconescu, B.; Robinson, J. T. Rotational Disorder in Twisted Bilayer Graphene. *ACS Nano* **2014**, *8*, 1655–1663.
- (19) Solís-Fernández, P.; Terao, Y.; Kawahara, K.; Nishiyama, W.; Uwanno, T.; Lin, Y.-C.; Yamamoto, K.; Nakashima, H.; Nagashio, K.; Hibino, H.; Suenaga, K.; Ago, H. Isothermal Growth and Stacking Evolution in Highly Uniform Bernal-Stacked Bilayer Graphene. *ACS Nano* **2020**, *14*, 6834–6844.
- (20) Takesaki, Y.; Kawahara, K.; Hibino, H.; Okada, S.; Tsuji, M.; Ago, H. Highly Uniform Bilayer Graphene on Epitaxial Cu-Ni(111) Alloy. *Chem. Mater.* **2016**, *28*, 4583–4592.
- (21) He, R.; Chung, T.-F.; Delaney, C.; Keiser, C.; Jauregui, L. A.; Shand, P. M.; Chancey, C. C.; Wang, Y.; Bao, J.; Chen, Y. P. Observation of Low Energy Raman Modes in Twisted Bilayer Graphene. *Nano Lett.* **2013**, *13*, 3594–3601.
- (22) Ferrari, A. C.; Basko, D. M. Raman Spectroscopy as a Versatile Tool for Studying the Properties of Graphene. *Nat. Nanotechnol.* **2013**, *8*, 235–246.
- (23) Gupta, A.; Chen, G.; Joshi, P.; Tadigadapa, S.; Eklund, P. C. Raman Scattering from High-Frequency Phonons in Supported n-Graphene Layer Films. *Nano Lett.* **2006**, *6*, 2667–2673.
- (24) Cong, C.; Yu, T.; Sato, K.; Shang, J.; Saito, R.; Dresselhaus, G. F.; Dresselhaus, M. S. Raman Characterization of ABA- and ABC-Stacked Trilayer Graphene. *ACS Nano* **2011**, *5*, 8760–8768.
- (25) Lui, C. H.; Malard, L. M.; Kim, S.; Lantz, G.; Laverge, F. E.; Saito, R.; Heinz, T. F. Observation of Layer-Breathing Mode Vibrations in Few-Layer Graphene through Combination Raman Scattering. *Nano Lett.* **2012**, *12*, 5539–5544.
- (26) Schäpers, A.; Sonntag, J.; Valerius, L.; Pestka, B.; Strasdas, J.; Watanabe, K.; Taniguchi, T.; Morgenstern, M.; Beschoten, B.; Dolleman, R. J.; Stampfer, C. Raman Imaging of Twist Angle Variations in Twisted Bilayer Graphene at Intermediate Angles. *2021*, arXiv:2104.06370 arXiv.org e-Print archive. <https://arxiv.org/abs/2104.06370>.
- (27) Henni, Y.; Ojeda Collado, H. P.; Nogajewski, K.; Molas, M. R.; Usaj, G.; Balseiro, C. A.; Orlita, M.; Potemski, M.; Faugeras, C. Rhombohedral Multilayer Graphene: A Magneto-Raman Scattering Study. *Nano Lett.* **2016**, *16*, 3710–3716.
- (28) Moutinho, M. V. O.; Eliel, G. S. N.; Righi, A.; Gontijo, R. N.; Paillet, M.; Michel, T.; Chiu, P.-W.; Venezuela, P.; Pimenta, M. A. Resonance Raman Enhancement by the Intralayer and Interlayer Electron–Phonon Processes in Twisted Bilayer Graphene. *Sci. Rep.* **2021**, *11*, No. 17206.
- (29) Gadelha, A. C.; Ohlberg, D. A. A.; Rabelo, C.; Neto, E. G. S.; Vasconcelos, T. L.; Campos, J. L.; Lemos, J. S.; Ornelas, V.; Miranda, D.; Nadas, R.; Santana, F. C.; Watanabe, K.; Taniguchi, T.; van Troeye, B.; Lamparski, M.; Meunier, V.; Nguyen, V.-H.; Paszko, D.; Charlier, J.-C.; Campos, L. C.; Cançado, L. G.; Medeiros-Ribeiro, G.; Jorio, A. Localization of Lattice Dynamics in Low-Angle Twisted Bilayer Graphene. *Nature* **2021**, *590*, 405–409.

- (30) Cong, C.; Yu, T.; Saito, R.; Dresselhaus, G. F.; Dresselhaus, M. S. Second-Order Overtone and Combination Raman Modes of Graphene Layers in the Range of 1690–2150 cm<sup>-1</sup>. *ACS Nano* **2011**, *5*, 1600–1605.
- (31) Wang, Q. H.; Jin, Z.; Kim, K. K.; Hilmer, A. J.; Paulus, G. L. C.; Shih, C.-J.; Ham, M.-H.; Sanchez-Yamagishi, J. D.; Watanabe, K.; Taniguchi, T.; Kong, J.; Jarillo-Herrero, P.; Strano, M. S. Understanding and Controlling the Substrate Effect on Graphene Electron-Transfer Chemistry via Reactivity Imprint Lithography. *Nat. Chem.* **2012**, *4*, 724–732.
- (32) Lee, J. E.; Ahn, G.; Shim, J.; Lee, Y. S.; Ryu, S. Optical Separation of Mechanical Strain from Charge Doping in Graphene. *Nat. Commun.* **2012**, *3*, No. 1024.
- (33) Rozada, R.; Solís-Fernández, P.; Paredes, J. I.; Martínez-Alonso, A.; Ago, H.; Tascón, J. M. D. Controlled Generation of Atomic Vacancies in Chemical Vapor Deposited Graphene by Microwave Oxygen Plasma. *Carbon* **2014**, *79*, 664–669.
- (34) Solís-Fernández, P.; Okada, S.; Sato, T.; Tsuji, M.; Ago, H. Gate-Tunable Dirac Point of Molecular Doped Graphene. *ACS Nano* **2016**, *10*, 2930–2939.
- (35) Kinoshita, H.; Jeon, I.; Maruyama, M.; Kawahara, K.; Terao, Y.; Ding, D.; Matsumoto, R.; Matsuo, Y.; Okada, S.; Ago, H. Highly Conductive and Transparent Large-Area Bilayer Graphene Realized by MoCl<sub>5</sub> Intercalation. *Adv. Mater.* **2017**, *29*, No. 1702141.
- (36) Kotsiantis, S. B.; Zaharakis, I. D.; Pintelas, P. E. Machine Learning: A Review of Classification and Combining Techniques. *Artif. Intell. Rev.* **2006**, *26*, 159–190.
- (37) Masubuchi, S.; Morimoto, M.; Morikawa, S.; Onodera, M.; Asakawa, Y.; Watanabe, K.; Taniguchi, T.; Machida, T. Autonomous Robotic Searching and Assembly of Two-Dimensional Crystals to Build van der Waals Superlattices. *Nat. Commun.* **2018**, *9*, No. 1413.
- (38) Tanaka, K.; Hachiya, K.; Zhang, W.; Matsuda, K.; Miyauchi, Y. Machine-Learning Analysis to Predict the Exciton Valley Polarization Landscape of 2D Semiconductors. *ACS Nano* **2019**, *13*, 12687–12693.
- (39) Frey, N. C.; Akinwande, D.; Jariwala, D.; Shenoy, V. B. Machine Learning-Enabled Design of Point Defects in 2D Materials for Quantum and Neuromorphic Information Processing. *ACS Nano* **2020**, *14*, 13406–13417.
- (40) Han, B.; Lin, Y.; Yang, Y.; Mao, N.; Li, W.; Wang, H.; Yasuda, K.; Wang, X.; Fatemi, V.; Zhou, L.; Wang, J. I.-J.; Ma, Q.; Cao, Y.; Rodan-Legrain, D.; Bie, Y.-Q.; Navarro-Moratalla, E.; Klein, D.; MacNeill, D.; Wu, S.; Kitadai, H.; Ling, X.; Jarillo-Herrero, P.; Kong, J.; Yin, J.; Palacios, T. Deep-Learning-Enabled Fast Optical Identification and Characterization of 2D Materials. *Adv. Mater.* **2020**, *32*, No. 2000953.
- (41) Sheremeteva, N.; Lamparski, M.; Daniels, C.; Van Troeye, B.; Meunier, V. Machine-Learning Models for Raman Spectra Analysis of Twisted Bilayer Graphene. *Carbon* **2020**, *169*, 455–464.
- (42) Lamparski, M.; Troeye, B. V.; Meunier, V. Soliton Signature in the Phonon Spectrum of Twisted Bilayer Graphene. *2D Mater.* **2020**, *7*, No. 025050.
- (43) Vincent, T.; Kawahara, K.; Antonov, V.; Ago, H.; Kazakova, O. Data Cluster Analysis and Machine Learning for Classification of Twisted Bilayer Graphene. 2021, arXiv:2107.08919. arXiv.org e-Print archive. <https://arxiv.org/abs/2107.08919>.
- (44) Wang, H.; Yao, Z.; Jung, G. S.; Song, Q.; Hempel, M.; Palacios, T.; Chen, G.; Buehler, M. J.; Aspuru-Guzik, A.; Kong, J. Frank-van der Merwe Growth in Bilayer Graphene. *Matter* **2021**, *4*, 3339–3353.
- (45) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (46) Cawley, G. C.; Talbot, N. L. C. On Over-Fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079–2107.
- (47) Wang, A. Y.-T.; Murdock, R. J.; Kauwe, S. K.; Oliynyk, A. O.; Gurlo, A.; Brzozka, J.; Persson, K. A.; Sparks, T. D. Machine Learning for Materials Scientists: An Introductory Guide toward Best Practices. *Chem. Mater.* **2020**, *32*, 4954–4965.
- (48) Goldberger, J.; Roweis, S.; Hinton, G.; Salakhutdinov, R. Neighbourhood Components Analysis. *Adv. Neural Inf. Process. syst.* **2004**, *17*, 513–520.
- (49) Xu, D.; Tian, Y. A Comprehensive Survey of Clustering Algorithms. *Ann. Data. Sci.* **2015**, *2*, 165–193.
- (50) Sun, L.; Wang, Z.; Wang, Y.; Zhao, L.; Li, Y.; Chen, B.; Huang, S.; Zhang, S.; Wang, W.; Pei, D.; Fang, H.; Zhong, S.; Liu, H.; Zhang, J.; Tong, L.; Chen, Y.; Li, Z.; Rummeli, M. H.; Novoselov, K. S.; Peng, H.; Lin, L.; Liu, Z. Hetero-Site Nucleation for Growing Twisted Bilayer Graphene with a Wide Range of Twist Angles. *Nat. Commun.* **2021**, *12*, No. 2391.
- (51) Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating Mutual Information. *Phys. Rev. E* **2004**, *69*, No. 066138.
- (52) Biau, G.; Scornet, E. A Random Forest Guided Tour. *TEST* **2016**, *25*, 197–227.
- (53) Eckmann, A.; Park, J.; Yang, H.; Elias, D.; Mayorov, A. S.; Yu, G.; Jalil, R.; Novoselov, K. S.; Gorbachev, R. V.; Lazzeri, M.; Geim, A. K.; Casiraghi, C. Raman Fingerprint of Aligned Graphene/h-BN Superlattices. *Nano Lett.* **2013**, *13*, 5242–5246.
- (54) Carozo, V.; Almeida, C. M.; Ferreira, E. H. M.; Cançado, L. G.; Achete, C. A.; Jorio, A. Raman Signature of Graphene Superlattices. *Nano Lett.* **2011**, *11*, 4527–4534.
- (55) Campos-Delgado, J.; Cançado, L. G.; Achete, C. A.; Jorio, A.; Raskin, J.-P. Raman Scattering Study of the Phonon Dispersion in Twisted Bilayer Graphene. *Nano Res.* **2013**, *6*, 269–274.
- (56) Liang, L.; Zhang, J.; Sumpter, B. G.; Tan, Q.-H.; Tan, P.-H.; Meunier, V. Low-Frequency Shear and Layer-Breathing Modes in Raman Scattering of Two-Dimensional Materials. *ACS Nano* **2017**, *11*, 11777–11802.
- (57) Lin, K.-Q.; Holler, J.; Bauer, J. M.; Parzefall, P.; Scheuck, M.; Peng, B.; Korn, T.; Bange, S.; Lupton, J. M.; Schüller, C. Large-Scale Mapping of Moiré Superlattices by Hyperspectral Raman Imaging. *Adv. Mater.* **2021**, *33*, No. 2008333.

## □ Recommended by ACS

### Polarized Raman Spectroscopy for Determining Crystallographic Orientation of Low-Dimensional Materials

Bo Xu, Jin Zhang, et al.

AUGUST 02, 2021

THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS

READ ▶

### Raman Tensor of WSe<sub>2</sub> via Angle-Resolved Polarized Raman Spectroscopy

Minge Jin, Feng Huang, et al.

NOVEMBER 22, 2019

THE JOURNAL OF PHYSICAL CHEMISTRY C

READ ▶

### Study of the Molecular Bending in Azobenzene Self-Assembled Monolayers Observed by Tip-Enhanced Raman Spectroscopy in Scanning Tunneling Mode

A. Foti, R. Ossikovski, et al.

OCTOBER 09, 2019

THE JOURNAL OF PHYSICAL CHEMISTRY C

READ ▶

### Molecular Perturbation Effects in AFM-Based Tip-Enhanced Raman Spectroscopy: Contact versus Tapping Mode

Giovanni Luca Bartolomeo, Renato Zenobi, et al.

NOVEMBER 12, 2021

ANALYTICAL CHEMISTRY

READ ▶

Get More Suggestions >