



OPEN

DATA DESCRIPTOR

# Computational scanning tunneling microscope image database

Kamal Choudhary<sup>1</sup>✉, Kevin F. Garrity<sup>1</sup>, Charles Camp<sup>1</sup>, Sergei V. Kalinin<sup>2</sup>, Rama Vasudevan<sup>2</sup>, Maxim Ziatdinov<sup>2</sup> & Francesca Tavazza<sup>1</sup>

We introduce the systematic database of scanning tunneling microscope (STM) images obtained using density functional theory (DFT) for two-dimensional (2D) materials, calculated using the Tersoff-Hamann method. It currently contains data for 716 exfoliable 2D materials. Examples of the five possible Bravais lattice types for 2D materials and their Fourier-transforms are discussed. All the computational STM images generated in this work are made available on the JARVIS-STM website (<https://jarvis.nist.gov/jarvisstm>). We find excellent qualitative agreement between the computational and experimental STM images for selected materials. As a first example application of this database, we train a convolution neural network model to identify the Bravais lattice from the STM images. We believe the model can aid high-throughput experimental data analysis. These computational STM images can directly aid the identification of phases, analyzing defects and lattice-distortions in experimental STM images, as well as be incorporated in the autonomous experiment workflows.

## Background & Summary

Since the invention of the scanning tunneling microscope (STM)<sup>1</sup>, this technique has become an essential tool for characterizing material surfaces and adsorbates. In addition to providing atomic insights, STM has been proven useful for characterizing the electronic structure, shapes of molecular orbitals, and vibrational and magnetic excitations<sup>2,3</sup>. It can also be used for manipulating adsorbates and adatoms, and for catalysis and quantum information processing applications<sup>3-8</sup>. Quantum mechanics-based density functional theory (DFT) has often been used to produce virtual STM images for these applications<sup>9,10</sup>. However, a systematic database of such computational STM data is still lacking. As DFT-STM images are constructed using defect-free materials, they provide standard reference images (SRI) that are useful to aid in identifying phases<sup>11,12</sup>, analyzing defects<sup>13,14</sup> and quantifying lattice-distortions<sup>15</sup> in experimental STM images. A DFT-STM database is therefore essential to provide a direct link between atomic positions and images, which can aid experimental analysis. Moreover, the orbital-projected electronic density of states available in our database can help explain which atoms and orbitals contribute to an experimental STM image. Finally, a computational database can provide an accurate training set for developing machine learning (ML) models to rapidly analyze experimental STM images.

STM imaging is particularly well-suited to studying two-dimensional (2D) materials, such as graphene<sup>16</sup>, MoS<sub>2</sub><sup>17</sup>, NbSe<sub>2</sub><sup>18</sup>, WSe<sub>2</sub><sup>19</sup>, WTe<sub>2</sub><sup>20</sup>, FeSe<sup>21</sup>, black-phosphorous<sup>22,23</sup> and SnSe<sup>24</sup>. 2D materials<sup>25,26</sup> has opened diverse areas of application, such as sub-micron level electronics<sup>27</sup>, flexible and tunable electronics<sup>28</sup>, superconductivity<sup>29</sup>, photo-voltaics<sup>30</sup>, water-purification<sup>31</sup>, sensors<sup>32</sup>, thermal-management<sup>33</sup>, energy-storage<sup>34</sup>, medicine<sup>35</sup>, quantum dots<sup>36,37</sup> and composites<sup>38-40</sup>. The surfaces of 2D materials are unique because they lack dangling bonds, allowing them to be exfoliated. This property makes them ideal candidates for building a database of computational STMs images because they don't require thick slabs perpendicular to the surface, which are computationally expensive to simulate accurately, and they do not have surface reconstructions. The generation of STM images for perfect systems is an initial step, and we will extend this project to include defective systems and the effect of thermal noise in the future.

In this work, we use DFT to generate STM images of exfoliable 2D materials. We use the recently developed JARVIS-DFT database (<https://jarvis.nist.gov/jarvisdft>) and select 2D materials with exfoliation energy less than 200 meV/atom. The JARVIS-DFT database contains about 40000 bulk and 1000 two-dimensional materials with their DFT-computed structural, energetic<sup>26</sup>, elastic<sup>41</sup>, optoelectronic<sup>42</sup>, thermoelectric<sup>43</sup>, piezoelectric, dielectric, infrared<sup>44</sup>, solar-efficiency<sup>45</sup>, and topological<sup>46,47</sup> properties. We note that there are several

<sup>1</sup>Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, 20899, USA.

<sup>2</sup>Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA. ✉e-mail: [kamal.choudhary@nist.gov](mailto:kamal.choudhary@nist.gov)

factors that can influence the appearance of experimental or DFT-based STM image predictions, such as the STM-tip material, bias voltage, and the scanning mode, *i.e.* constant-height mode (CHM) vs. constant current mode (CCM). Similarly, there are several methods for simulating STM images using DFT, including Bardeen<sup>48</sup>, Tersoff-Hamman<sup>49</sup> and Chen<sup>3</sup> methods. Here, we present results for constant height and constant current DFT-STM images computed using the Tersoff-Hamann approach<sup>49</sup>, which assumes a non-functionalized (*s*-wave) STM tip. Hence, in the simulation we don't explicitly model the tip and its interactions. The ML model training is based on CHM images. The DFT-STM database currently contains images for 716 exfoliable 2D materials, with additional computations ongoing. All the DFT-STM data will be uploaded into the JARVIS-DFT database.

As a first example application of this database and use artificial intelligence methods<sup>50–53</sup>, we use the computational STM images to train a convolution neural network ML classification model for Bravais-lattices. This model is able to quickly classify STM images into the five lattice classes (square, hexagon, rhombus/centered-rectangle, rectangle and parallelogram/oblique) that are possible for 2D systems. Such classifications are of importance, for example when dealing with phase transitions<sup>54</sup>. They can also be used as an aid to automatic conventional crystallographic image processing of big datasets and to obtain information from noisy images. This work acts as a starting point for identifying the defects in experimental images by providing a collection of ideal STM images for comparison purposes<sup>55</sup>. Ideally one would use an information-theoretic approach, as opposed to deep learning, to enable space group determination with uncertainty quantification, as demonstrated by Moeck, to distinguish the specific subgroups of selected group<sup>56</sup>. However, a pre-screening step can be rapidly accomplished with a suitably trained neural network as shown here, which should then be verified using the approach outlined in ref. <sup>54</sup>. Later, these computational STM trained models can be integrated with experiments for active learning processes<sup>50,57</sup>.

## Methods

All DFT calculations are carried out with Vienna *ab initio* simulation package (VASP)<sup>58,59</sup> using projected augmented wave (PAW) formalism and using vdW-DF-OptB88 functional<sup>60</sup>. Note that for monolayers, vdW functionals are not strictly necessary. But we include vdW interactions to be consistent with our JARVIS-DFT 3D dataset. Also, we plan to develop multi-layer materials databases, which do require vdW interactions. The vdW functional works for both strongly and weakly bonded systems<sup>41</sup>. All the machine learning trainings are carried using Keras with TensorFlow backend<sup>61</sup>. Note that commercial software is identified to specify procedures, and such identification does not imply recommendation by the National Institute of Standards and Technology. The *k*-point and plane-wave cut-off convergence for each material are obtained using the workflow detailed in ref. <sup>45</sup>. The high-throughput computation and analysis tools will be made available at JARVIS-Tools github page: <https://github.com/usnistgov/jarvis>. The 2D materials are provided with at least 20 Å vacuum in the *z*-direction to avoid self-interactions. The force and energy convergence for DFT self-consistent calculations are 10<sup>−6</sup> eV and 0.001 eV/Å respectively.

The STM images are calculated using the Tersoff-Hamann approach, which is a simple model of an *s*-wave STM tip<sup>49</sup>:

$$n(r, E) = \sum_{\mu} |\psi_{\mu}(r)|^2 \delta(\varepsilon_{\mu} - E) \quad (1)$$

$$I(r, V) \propto \int_{E_F}^{E_F+eV} dE n(r, E) \quad (2)$$

In this approach, the tunneling current *I*, which depends on the tip position *r* and the applied voltage *V*, is proportional to the integrated local density of states (ILDOS). The ILDOS is calculated from the Kohn-sham eigenvectors,  $\psi_{\mu}$ , and eigenvalues,  $\varepsilon_{\mu}$ , where  $\mu$  labels different states.  $E_F$  is the Fermi-energy. Different experiments will choose different applied voltages, but we concentrate on two values, +0.5 eV for positive bias and −0.5 eV for negative bias, which require integrating from  $E_F$  to  $E_F \pm 0.5$  eV. We choose 0.5 eV range for simplicity sake, and other values usually produce qualitatively similar images for metals or small gap semiconductors. However, simulations for other voltages should also be possible with the method and tools discussed in this work.

This method is readily available in DFT software such as VASP<sup>62</sup>. Please note that plane-wave codes like VASP will not accurately describe the exponential decay of the wave functions far away from the atoms, and wave functions may need to be extrapolated in order for STM simulations at large heights such as 7 Å else it can show unphysical effects<sup>63</sup>. Hence, we choose image height relatively close to surface. All the STM images are made at least 20 Å long in the *xy* plane by repeating the primitive unit cell. We choose height 2 Å above the surface (maximum of *z*-coordinate) during the simulations. For constant-current images, we identify iso-surfaces that have a constant ILDOS. The height of these iso-surfaces at each *xy*-coordinates produces the images.

For the machine learning model, we simplify the constant-height STM images using a black/white color-scheme and choose a pixel value of 170 (out of maximum 255) for finding atomic features. We simplify the images because the image produced from the wavefunction is still on a continuous scale (*i.e.* grey image), while for the Bravais lattice classification only requires information on whether an atom is there or not. Based on the lattice-parameters and angles the 2D materials can be classified in five classes: 1) hexagonal, 2) square, 3) rhombus/centered-rectangle, 4) rectangle, 5) parallelograms/oblique. Deep-learning image recognition tasks typically require thousands of training images. To increase the size of our training set, we use several commonly applied image augmentations: random rotations, flipping, zooming in and zooming out. We apply augmentations until all the five classes have at least 10000 images leading 53508 images. Image processing ML models are usually non invariant to the operations mentioned above, which is why the initial dataset is augmented with such

operations. We use a multi-layer network with four convolution layers (with 16, 32, 48 and 64 feature-maps and with kernel-size of 3), four max-pooling layers (with pool-size of (2, 2)) activated by a rectified linear unit (ReLU), one fully-connected 600-nodes layer with ReLU activations, and a fully-connected softmax layer with five outputs. Since the entire dataset is too big to feed to the GPU memory at once, we divide it into multiple smaller batches. The total number of training examples present in a single batch (batch size) is 32 for our NN model. We have 20% dropout before the softmax layer to avoid overfitting. We use ADAM stochastic optimization method for gradient descent with ‘sparse categorical crossentropy’ as loss function. We split the dataset into training, validation, and test sets. We use a 90%-10% train-test split for the entire dataset in such a way that both training and testing data have a proportionate amount of all the five classes. Furthermore, we apply a 90%-10% split on the training data for model-training, validation and generating the learning curve. We apply ‘Early-stopping’ to avoid over-fitting of the model. After the model development, we apply this model on the 10% test-data to evaluate the accuracy of the model. Note that the 10% test-dataset was never used during model development.

During the training, we monitor the train-validation curve (discussed later) to avoid overfitting. We use accuracy, precision, recall, and F1-score to measure the overall and individual class performances. The precision is the ratio  $\frac{TP}{TP + FP}$  where TP is the number of true positives and FP the number of false positives. The recall is the ratio  $\frac{TP}{TP + FN}$  where TP is the number of true positives and FN the number of false negatives. The recall is intuitively the ability of the classifier to find all the positive samples. The F1-score can be interpreted as a weighted harmonic mean of the precision and recall, where an F1-score reaches its best value at 1 and worst score at 0. The overall classification accuracy of the model is given as  $\frac{TP + TN}{TP + TN + FP + FN}$ , where TN represents the number of true-negatives. We also use the confusion matrix to show the percentage correct and incorrect predictions of each class. Both the model and the associated dataset will be made publicly available soon at the JARVIS-DFT website.

## Data Records

After the calculations, the metadata is stored in the Javascript Object Notation Files (JSON) format which can be easily integrated with databases such as MongoDB. The dataset is made publicly available through the JARVIS-STM (<https://jarvis.nist.gov/jarvisstm>) web-app. The web-app provides both constant-height and constant-current simulation features and allows the user to change the chosen height or current value. We have made the dataset publicly available through Figshare repository<sup>64</sup> as well. The dataset consists of positive and negative bias constant-height images in Joint Photographic Experts Group (JPEG) format for the 2D materials under investigation. In addition to the images, we provide the raw input/output files for the calculations (including PARCHG files) at to enhance reproducibility of the work that could be used for generating both constant height and constant current images and for a given size of the xy-dimension.

## Technical Validation

**Validation of DFT simulated images.** We simulate computational STM images of 716 exfoliable materials ( $E_{\text{f}} < 200$  meV/atom) using the Tersoff-Hamann approach. We compare computational STM images with those from experiments for graphene<sup>16</sup>, 2H-MoS<sub>2</sub><sup>17</sup>, 2H-NbSe<sub>2</sub><sup>18</sup>, 2H-WSe<sub>2</sub><sup>19</sup>, 1T'-WTe<sub>2</sub><sup>20</sup>, FeSe<sup>21</sup>, black-P<sub>22,23</sub>, SnSe<sup>24</sup>, Bismuth<sup>64,65</sup>. We chose these systems because we could find well-characterized experimental images in the literature. Qualitatively, we observe that the patterns in the computational and experimental STMs are very similar (see Fig. 1).

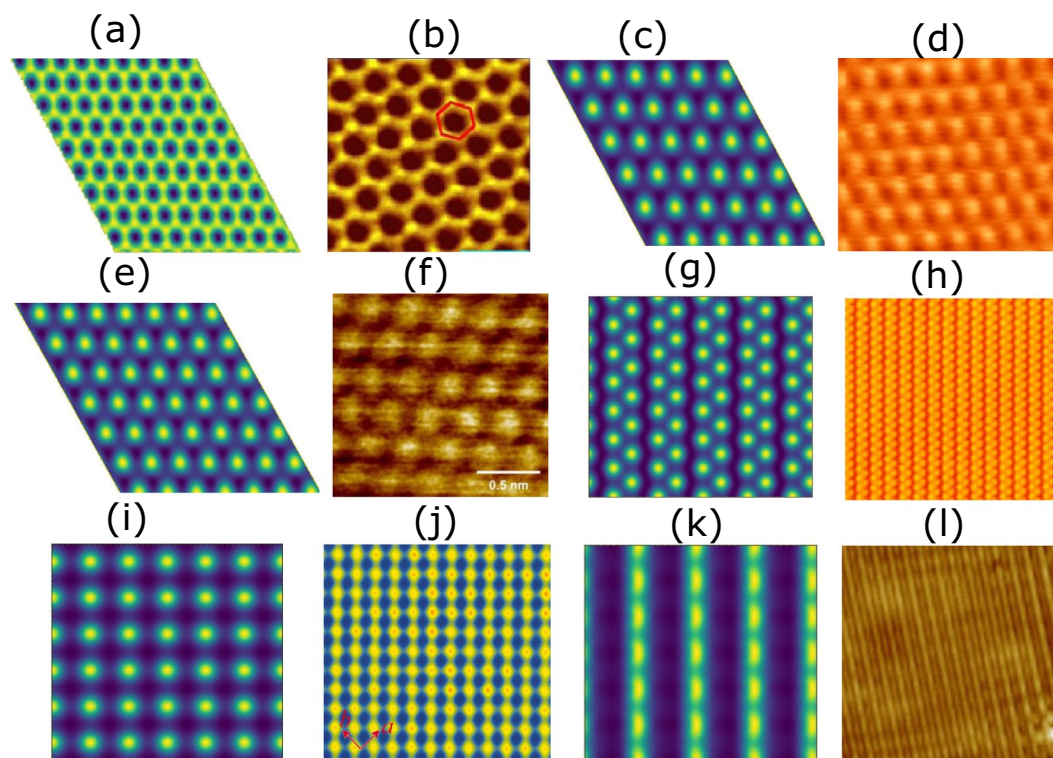
Experimental STM images for each system can be found in appropriate reference. Note that we are able to predict the STM for 2D vdW materials very well because they lack dangling bonds. Such images with non-vdW systems such as Si (111)<sup>66</sup> would require bigger simulation cells in the xy direction to accommodate reconstructions, as well as many additional layers to converge the calculations.

The DFT-STM can be used for distinguishing phases such as the 2D-monolayer 2H-MoTe<sub>2</sub> (JVASP-670) and 1T'-MoTe<sub>2</sub> (JVASP-673) phases, as shown in constant height positive bias conditions in Fig. 2. Such phase-identifications can be helpful in providing insight into phase-transformation mechanisms during experiments.

The 2H-phase is semiconducting material with hexagonal symmetry, as is evident from the crystal structure in Fig. 2a. The positive + 0.5 eV bias constant height image of this structure is shown in Fig. 2b. The electronic states in this range are dominated by Mo (*d*-orbital) states, hence the brighter spots in the STM are dominated by Mo *d*-orbitals, which can be understood by analyzing the projected density of states (Fig. 3). As shown in Fig. 2c, the fast Fourier transform (FT) of the simulated STM image in Fig. 2b shows hexagonal symmetry. Similarly, the crystal structure, STM image and FT of rectangular 1T'-MoTe<sub>2</sub> is shown in Fig. 2d–f respectively. We note that the FT of the STM image of a rectangular system with a multi-atom cell is not a simple rectangle. We show examples of variation of height in Å and current in arbitrary units in Fig. 2g,h and i for 2H-MoTe<sub>2</sub>. The constant height for 2H-MoTe<sub>2</sub> in Fig. 2b is for 3 Å while that in Fig. 2g is for 5 Å with respect to the highest atom in the cell. Clearly, the hexagonal patterns remain the same, but the structure around the atoms slightly changes due to the change in height. This is because as we move the hypothetical STM tip, we probe different layers of charge density. Similarly, we show the current variation based STM images for 0.01 and 0.05 a.u.<sup>-3</sup> eV<sup>-1</sup> in Fig. 2h,i. Note that it is difficult to quantitatively compare the computational and experimental STM images because the tunneling-current is critically dependent on the specific experimental setup.

Based on lattice parameter information in 2D plane, the 2D materials lattices can be classified in 5 types: hexagon, square, rectangle, rhombus/centered-rectangle, and parallelograms/oblique. We classify all the 2D materials in our database, with the distribution shown in Fig. 4a. Most of the 2D materials in our database are hexagonal, followed by rectangular and square lattices. In Fig. 4, we give examples of materials in each lattice type, in each case showing the atomic positions, a constant height STM image, and the fast Fourier transform (FT) of the STM



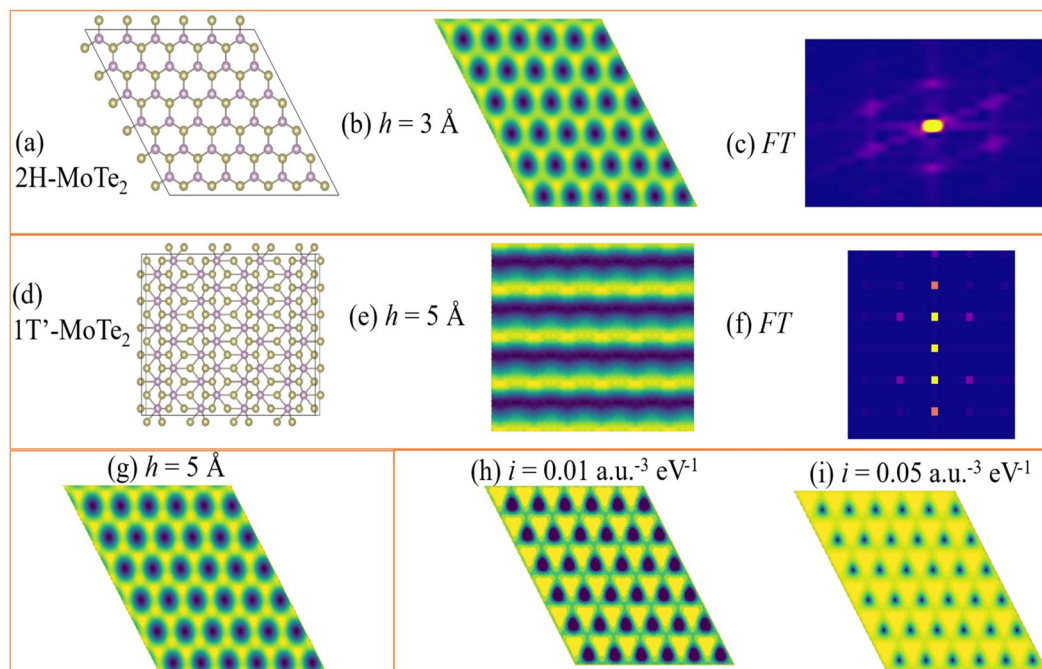


**Fig. 1** Comparison of computational and experimental STM images. Columns 1 and 3 are computational images while 2 and 4 are corresponding experimental images. (a) DFT-STM image of graphene, (b) experimental STM image of graphene [Reprinted with permission from ref. <sup>16</sup>. Copyright 2019 American Physical Society], (c) DFT-STM image of 2H-NbSe<sub>2</sub>, (d) experimental STM image of 2H-NbSe<sub>2</sub> [Reprint from ref. <sup>18</sup> with the permission of AIP Publishing.], (e) DFT-STM image of 2H-MoS<sub>2</sub>, (f) experimental STM image of 2H-MoS<sub>2</sub> [Reprint from ref. <sup>17</sup> with the permission of AIP Publishing.], (g) DFT-STM image of black phosphorous, (h) experimental STM image of black-P [Adapted with permission from ref. <sup>22</sup>, Copyright 2019 American Physical Society], (i) DFT-STM image of FeSe, (j) experimental STM image of FeSe [Reprint from ref. <sup>21</sup> with the permission of AIP Publishing.], (k) DFT-STM image of 1T'-WTe<sub>2</sub>, (l) experimental STM image of 1T'-WTe<sub>2</sub> [Reprinted with permission from ref. <sup>20</sup> Copyright 2017 by the American Physical Society].

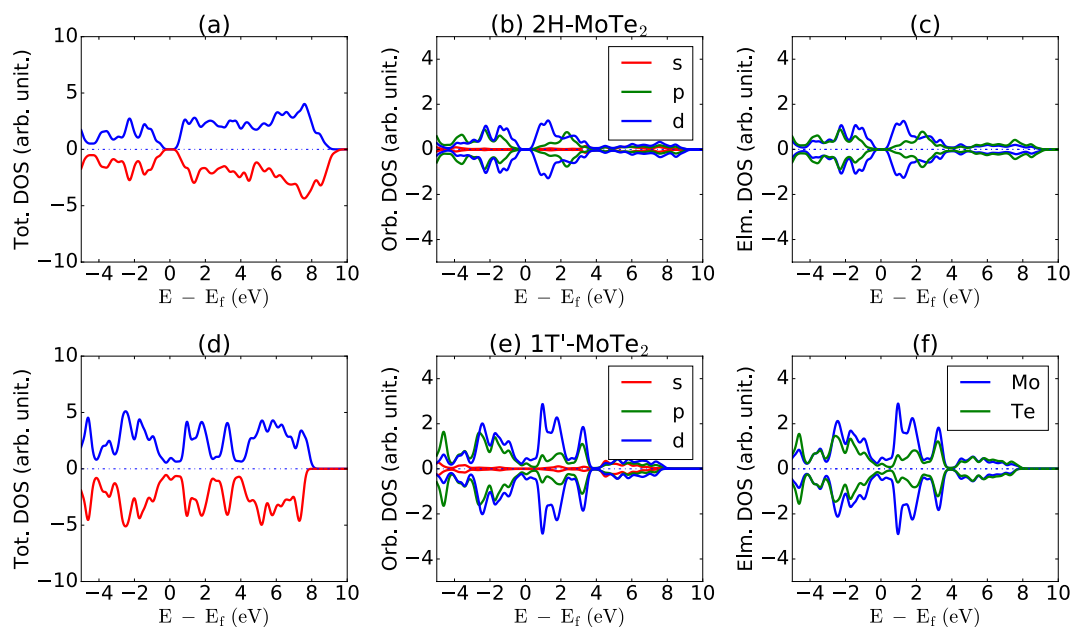
image. An example of hexagonal lattice is shown in Fig. 4b graphene (JVASP-667). It is one of the most widely investigated 2D materials. The STM positive bias image for graphene is shown in Fig. 4c. An FT of the image Fig. 4c is shown in Fig. 4d. It is clear from Fig. 4d that there is a hexagonal pattern due to hexagonal symmetry in graphene. Similarly, for the square lattice example, FeTe (JVASP-6667), the crystal structure, STM, and FT are shown in Fig. 4e–g. Fe *d*-states mainly contributes to the STM image in Fig. 4f. The FT of this image shows square-like patterns in Fig. 4g. Similarly, Fig. 4h gives the crystal structure of VCIO (JVASP-8933), and its STM and FT show a rectangular pattern (Fig. 4i,j). AuI (JVASP-6187) has a centered-rectangle structure, as shown in Fig. 4k. The lattice constants are 4.274 Å and the angle between them is 93.2 degrees. The Au *d*-orbitals contribute most to the STM image. The atomic and orbital projected density of systems for all the systems here is given in the supplementary information (Supplementary Fig. S1) and the respective webpages for each material. The FT in the Fig. 4m,p shows a noticeable blur, which can be caused by the truncation of the infinite slab to a finite image. Note that the mathematical FT of a perfectly periodic system would have ideal/sharp peaks. However, we purposefully truncate the images and include white spaces to mimic experimental images. Hence, they won't be perfectly sharp. Figure 4n shows As<sub>2</sub>Se<sub>3</sub> (JVASP-13544), an example material with an oblique unit cell with lattice constants of 4.4 and 12.9 Å and an angle of 109.9 degrees. The FT of the STM in Fig. 4p is difficult to interpret.

**Machine learning model development.** Having prepared our database, we now train a ML model (JARVIS-STMnet) following the flow-chart in Fig. 5.

In Fig. 6 we show the convolution neural network training and the learning curves for the deep learning model. We monitor the learning curve as in Fig. 6a. We see that after the 5<sup>th</sup> epoch the training and validation accuracy curves begin to diverge, so we stop further training. We obtain 90.1% accuracy on the validation set and 90.0% accuracy on the 10% test-set, which was never used during the training process. The difference between the training and the validation curve is small, implying low overfitting. We apply the trained model on the 10% test-set data and the confusion matrix is shown in Fig. 6b. We also provide precision, recall and F1 scores in Table 1. The baseline accuracy of the model is  $1/5 = 20\%$ . Clearly, the overall accuracy is more than 4 times higher than the random-guessing baseline model. Also, all the scores in Table 1 are more than 0.85, indicating that the model performs much better than a random guessing model. Note that although the accuracy is a measure of

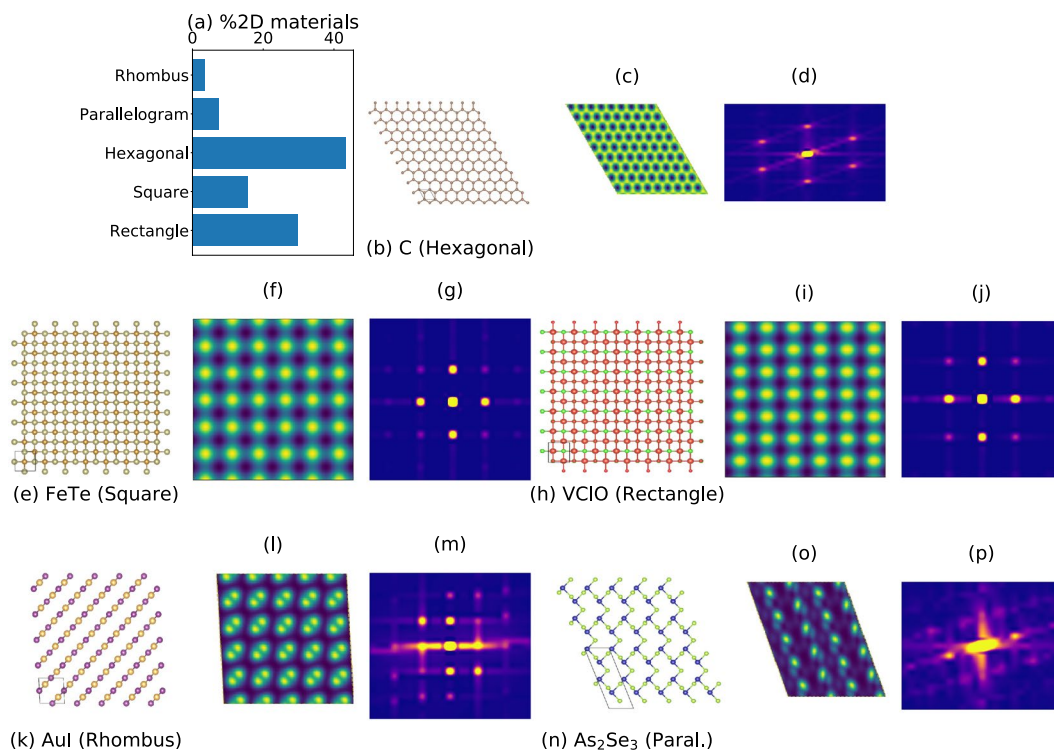


**Fig. 2** STM images of MoTe<sub>2</sub> in 2H and 1T' phases, their fast Fourier transform (FT), height and current dependence. (a) crystal structure of 2H-MoTe<sub>2</sub> (JVASP-670) (b) its constant-height STM image at the height of 3 Å, (c) FT of the STM image, (d) crystal structure of 1T'-MoTe<sub>2</sub> (JVASP-673) (e) its STM image at the height of 5 Å, (f) FT of the STM image, (g) constant-height STM image of 2H-MoTe<sub>2</sub> at the height of 5 Å to compare with Fig. b, (h) constant current image for 2H-MoTe<sub>2</sub> at constant current 0.01 a.u.<sup>-3</sup> eV<sup>-1</sup>, (i) constant current image for 2H-MoTe<sub>2</sub> at constant current 0.05 a.u.<sup>-3</sup> eV<sup>-1</sup>.

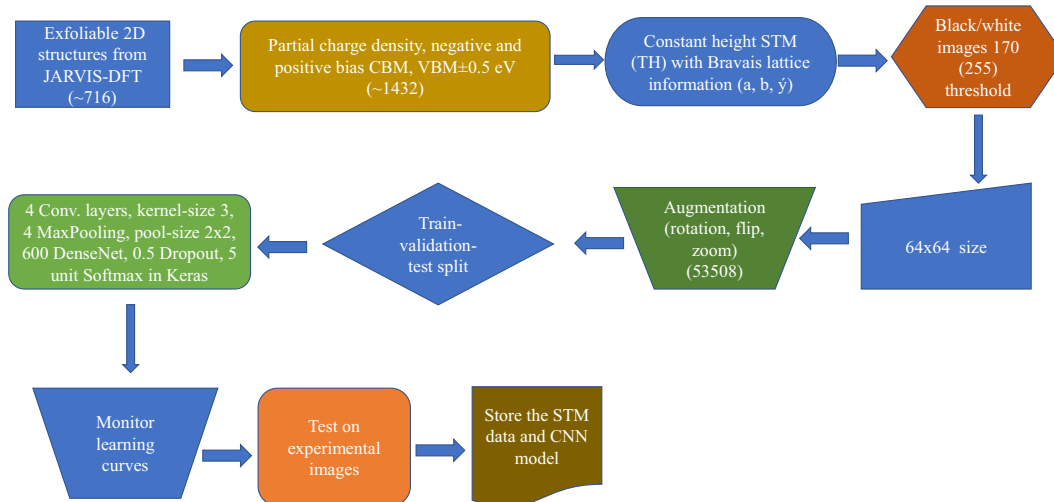


**Fig. 3** Total (a,d), element (b,e) and orbital(c,f) projected density of states of materials discussed in Fig. 1. The Fermi-energies are set to zero. Total up and down spins are shown as blue and red lines respectively in (a, d).

the overall model, it is important to investigate the prediction accuracy for each class of the model. A confusion matrix with high diagonal element values signifies high accuracy. It is clear from the Fig. 6b that the model performs excellently for hexagonal, centered rectangle and square lattices, and less well for the rectangle and oblique lattice types. Moving beyond simulated STM images, as an initial validation, we apply the model to nine experimental images discussed above for an initial more realistic test step for graphene<sup>16</sup>, 2H-MoS<sub>2</sub><sup>17</sup>, 2H-NbSe<sub>2</sub><sup>18</sup>, 2H-WSe<sub>2</sub><sup>19</sup>, 1T'-WTe<sub>2</sub><sup>20</sup>, FeSe<sup>21</sup>, black-P<sup>22,23</sup>, SnSe<sup>24</sup>, Bismuth<sup>64,65</sup>. We find that the model predicts the correct class



**Fig. 4** Types of 2D lattices with examples. **(a)** 2D lattice-type distribution in the database. **(b–d)** crystal structure, constant height STM (CHS) and FT for graphene. Similar images are also provided for **(e–g)** FeTe, **(h–j)** VCIO, **(k–m)** Aul, **(n–p)** As<sub>2</sub>Se<sub>3</sub>.

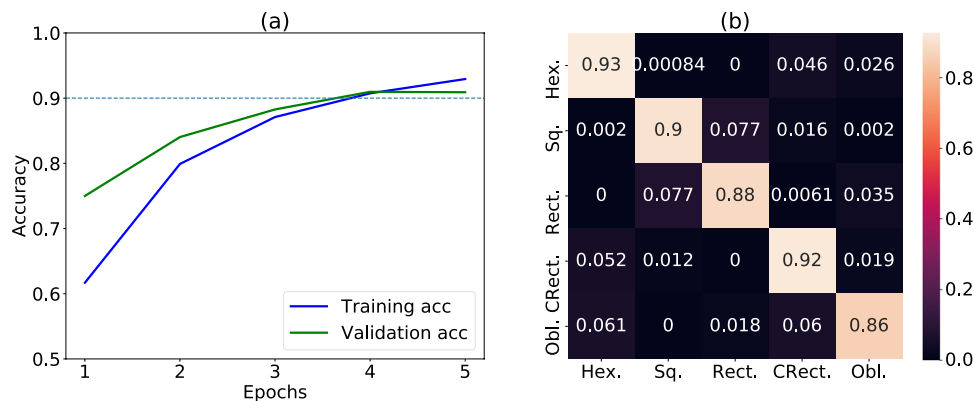


**Fig. 5** Flow-chart showing the steps involved in the generation of STM-dataset and the machine learning process.

for seven of them. Performing a more systemic analysis of our model's accuracy on experimental images would require a database of hundreds of experimental images, and such a database is currently not available. We hope this work will spur the development of such a database. Also, as we make the entire dataset publicly available, and we hope that other researchers could apply their machine-learning models on this dataset.

### Usage Notes

We introduce the first systematic database of scanning tunneling microscope (STM) images obtained using density functional theory (DFT) for two-dimensional (2D) materials. Specifically, the database is constructed using the Tersoff-Hamann method for constant-height images. Although only defect free materials are considered in this work, STM image dataset with defects will be developed soon. We anticipate that this dataset and methods



**Fig. 6** Performance of the machine learning model. **(a)** learning curve, **(b)** confusion matrix.

Lattice	Precision	Recall	F1-score
Hexagonal	0.91	0.93	0.92
Square	0.90	0.90	0.90
Rectangle	0.91	0.88	0.90
Centered Rectangle	0.86	0.92	0.89
Oblique	0.91	0.86	0.88
Accuracy			0.90

**Table 1.** Classification report of classifying 2D constant-height STM images into lattice-types.

used will provide a useful tool in fundamental and application-related studies of materials. Experimental verification provides insight into understanding the applicability and limitation of our DFT data. Based on the list of data, the user will be able to choose particular materials for specific applications. Data mining, data analytics, and artificial-intelligence tools then can be added to guide screening of materials.

### Code availability

Python-language-based codes with examples are given at the JARVIS-Tools page <https://github.com/usnistgov/jarvis>.

Received: 23 June 2020; Accepted: 6 January 2021;

Published online: 11 February 2021

### References

- Binnig, G., Rohrer, H., Gerber, C. & Weibel, E. Surface studies by scanning tunneling microscopy. *Phys. Rev. Lett.* **49**, 57 (1982).
- Mugarza, A. *et al.* Spin coupling and relaxation inside molecule–metal contacts. *Nat. Commun.* **2**, 490 (2011).
- Chen, C. J. *Introduction to scanning tunneling microscopy*. Vol. 4 (Oxford University Press on Demand, 1993).
- Gross, L. *et al.* High-resolution molecular orbital imaging using a p-wave STM tip. *Phys. Rev. Lett.* **107**, 086101 (2011).
- Eigler, D. M. & Schweizer, E. K. Positioning single atoms with a scanning tunnelling microscope. *Nature* **344**, 524 (1990).
- Stipe, B., Rezaei, M. & Ho, W. Single-molecule vibrational spectroscopy and microscopy. *Science* **280**, 1732–1735 (1998).
- Hirjibehedin, C. F. *et al.* Large magnetic anisotropy of a single atomic spin embedded in a surface molecular network. *Science* **317**, 1199–1203 (2007).
- Yang, K. *et al.* Coherent spin manipulation of individual atoms on a surface. *Science* **366**, 509–512 (2019).
- Barth, J., Brune, H., Ertl, G. & Behm, R. Scanning tunneling microscopy observations on the reconstructed Au (111) surface: Atomic structure, long-range superstructure, rotational domains, and surface defects. *Phys. Rev. B* **42**, 9307 (1990).
- Magonov, S. N. & Whangbo, M.-H. *Surface analysis with STM and AFM: experimental and theoretical aspects of image analysis*. (John Wiley & Sons, 2008).
- Poirier, G. *et al.* Identification of the facet planes of phase I TiO<sub>2</sub> (001) rutile by scanning tunneling microscopy and low energy electron diffraction. *J. Vac. Sci. Tech. B* **10**, 6–15 (1992).
- Burk, B., Thomson, R., Zettl, A. & Clarke, J. Charge-density-wave domains in 1T-TaS<sub>2</sub> observed by satellite structure in scanning-tunneling-microscopy images. *Phys. Rev. Lett.* **66**, 3040 (1991).
- Vancsó, P. *et al.* The intrinsic defect structure of exfoliated MoS<sub>2</sub> single layers revealed by Scanning Tunneling Microscopy. *Sci. Rep.* **6**, 29726 (2016).
- Liu, H. *et al.* Line and point defects in MoSe<sub>2</sub> bilayer studied by scanning tunneling microscopy and spectroscopy. *ACS Nano* **9**, 6619–6625 (2015).
- Dubout, Q. *et al.* Giant apparent lattice distortions in STM images of corrugated sp<sup>2</sup>-hybridised monolayers. *New J. Phys.* **18**, 103027 (2016).
- Li, G., Luican, A. & Andrei, E. Y. Scanning tunneling spectroscopy of graphene on graphite. *Phys. Rev. Lett.* **102**, 176804 (2009).
- Mills, A. *et al.* Ripples near edge terminals in MoS<sub>2</sub> few layers and pyramid nanostructures. *App. Phys. Lett.* **108**, 081601 (2016).
- Wang, J. *et al.* A variable-temperature scanning tunneling microscope operated in a continuous flow cryostat. *Rev. Sci. Instr.* **90**, 093702 (2019).
- Liu, H. *et al.* Molecular-beam epitaxy of monolayer and bilayer WSe<sub>2</sub>: a scanning tunneling microscopy/spectroscopy study and deduction of exciton binding energy. *2D Mater.* **2**, 034004 (2015).



20. Jia, Z.-Y. *et al.* Direct visualization of a two-dimensional topological insulator in the single-layer 1T'-WTe<sub>2</sub>. *Phys. Rev. B* **96**, 041108 (2017).
21. Song, C.-L. *et al.* Direct observation of nodes and twofold symmetry in FeSe superconductor. *Science* **332**, 1410–1413 (2011).
22. Kumar, A. *et al.* STM study of exfoliated few layer black phosphorus annealed in ultrahigh vacuum. *2D Mater.* **6**, 015005 (2018).
23. Kiraly, B., Hauptmann, N., Rudenko, A. N., Katsnelson, M. I. & Khajetoorians, A. A. Probing single vacancies in black phosphorus at the atomic level. *Nano Lett.* **17**, 3607–3612 (2017).
24. Duvjir, G. *et al.* Origin of p-type characteristics in a SnSe single crystal. *App. Phys. Lett.* **110**, 262106 (2017).
25. Choudhary, K., Kalish, I., Beams, R. & Tavazza, F. High-throughput Identification and Characterization of Two-dimensional Materials using Density functional theory. *Sci. Rep.* **7**, 5179 (2017).
26. Choudhary, K. *et al.* The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *npj Comput Mater* **6**, 173 (2020).
27. Fiori, G. *et al.* Electronics based on two-dimensional materials. *Nat. Nanotechnol.* **9**, 768–779 (2014).
28. Akinwande, D., Petrone, N. & Hone, J. Two-dimensional flexible nanoelectronics. *Nat. Comm.* **5** (2014).
29. Navarro-Moratalla, E. & Jarillo-Herrero, P. Two-dimensional superconductivity: The Ising on the monolayer. *Nat. Phys.* **12**, 112–113 (2016).
30. Bubnova, O. 2D materials: Hybrid interfaces. *Nat. Nanotechnol.* **16**, 497 (2016).
31. Dervin, S., Dionysiou, D. D. & Pillai, S. C. 2D nanostructures for water purification: graphene and beyond. *Nanoscale* (2016).
32. Cui, S. *et al.* Ultrahigh sensitivity and layer-dependent sensing performance of phosphorene-based gas sensors. *Nat. Commun.* **6**, 8632 (2015).
33. Lee, M.-J. *et al.* Thermoelectric materials by using two-dimensional materials with negative correlation between electrical and thermal conductivity. *Nat. Commun.* **7**, 12011 (2016).
34. Zhang, X., Hou, L., Ciesielski, A. & Samori, P. 2D Materials Beyond Graphene for High-Performance Energy Storage Applications. *Adv. Energy Mater.* **6**, 1600671 (2016).
35. Boland, C. S. *et al.* Sensitive, high-strain, high-rate bodily motion sensors based on graphene–rubber composites. *ACS nano* **8**, 8819–8830 (2014).
36. Wang, X., Sun, G., Li, N. & Chen, P. Quantum dots derived from two-dimensional materials and their applications for catalysis and energy. *Chem. Soc. Rev.* **45**, 2239–2262 (2016).
37. Chakraborty, C., Kinnischtzke, L., Goodfellow, K. M., Beams, R. & Vamivakas, A. N. Voltage-controlled quantum light from an atomically thin semiconductor. *Nat. nanotechnol.* **10**, 507–511 (2015).
38. Castellanos-Gomez, A. Why all the fuss about 2D semiconductors? *Nat. Photon.* **10**, 202 (2016).
39. Flat talk. *Nat. Photon.* **10**, 205 (2016).
40. Rodenas, T. *et al.* Metal–organic framework nanosheets in polymer composite materials for gas separation. *Nat. Mater.* **14**, 48–55 (2015).
41. Choudhary, K., Cheon, G., Reed, E. & Tavazza, F. Elastic properties of bulk and low-dimensional materials using van der Waals density functional. *Phys. Rev. B* **98**, 014107 (2018).
42. Choudhary, K. *et al.* Computational screening of high-performance optoelectronic materials using OptB88vdW and TB-mBJ formalisms. *Sci. Data* **5**, 180082 (2018).
43. Choudhary, K., Garrity, K. & Tavazza, F. Data-driven Discovery of 3D and 2D Thermoelectric Materials. *J. Phys.: Condens. Matter* **32**, 475501 (2020).
44. Choudhary, K. *et al.* High-throughput Density Functional Perturbation Theory and Machine Learning Predictions of Infrared, Piezoelectric and Dielectric Responses. *npj Comput. Mat.* **6**, 64 (2020).
45. Choudhary, K. *et al.* Accelerated Discovery of Efficient Solar Cell Materials Using Quantum and Machine-Learning Methods. *Chem. Mater.* **31**(15), 5900 (2019).
46. Choudhary, K., Garrity, K. F., Jiang, J., Pachter, R. & Tavazza, F. Computational Search for Magnetic and Non-magnetic 2D Topological Materials using Unified Spin-orbit Spillage, Screening. *npj Comput. Mat.* **6**, 49 (2020).
47. Choudhary, K., Garrity, K. F. & Tavazza, F. High-throughput Discovery of topologically Non-trivial Materials using spin-orbit spillage. *Sci. Rep.* **9**, 1–8 (2019).
48. Bardeen, R. Tunneling from a many-particle point of view. *Phys. Rev. Lett.* **6**, 57 (1961).
49. Tersoff, J. & Hamann, D. R. Theory of the scanning tunneling microscope. *Phys. Rev. B* **31**, 805 (1985).
50. Vasudevan, R. K. *et al.* Materials science in the artificial intelligence age: high-throughput library generation, machine learning, and a pathway from correlations to the underpinning physics. *MRS Comm.* **9**(3), 821 (2019).
51. Rickman, J. M., Lookman, T. & Kalinin, S. V. Materials informatics: From the atomic-level to the continuum. *Acta Mater.* **168**, 473 (2019).
52. Hill, J., Mannodi-Kanakkithodi, A., Ramprasad, R. & Meredig, B. *Computational Materials System Design* 193–225 (Springer, 2018).
53. Choudhary, K., DeCost, B. & Tavazza, F. Machine learning with force-field-inspired descriptors for materials: Fast screening and mapping energy landscape. *Phys. Rev. Mat.* **2**, 083801 (2018).
54. Vasudevan, R. K. *et al.* Mapping mesoscopic phase evolution during E-beam induced transformations via deep learning of atomically resolved images. *npj Comput. Mat.* **4**, 30 (2018).
55. Moeck, Peter. On classification approaches for crystallographic symmetries of noisy 2D periodic patterns. *IEEE Transactions on Nanotechnology* **18**, 1166–1173 (2019).
56. Ziatdinov, M. *et al.* Deep learning of atomically resolved scanning transmission electron microscopy images: chemical identification and tracking local transformations. *ACS Nano* **11**, 12742 (2017).
57. Sk, R., Deshpande, A. & Engineering. Unveiling the emergence of functional materials with STM: metal phthalocyanine on surface architectures. *Mol. Syst. Design & Engineerin* **4**, 471 (2019).
58. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169 (1996).
59. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mat. Sci.* **6**, 15–50 (1996).
60. Klimeš, J., Bowler, D. R. & Michaelides, A. Chemical accuracy for the van der Waals density functional. *J. Phys. Cond. Mat.* **22**, 022201 (2009).
61. Abadi, M. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv*: **1603**, 04467 (2016).
62. Lounis, S. Theory of scanning tunneling microscopy. *arXiv* **1404**, 0961 (2014).
63. Tersoff, J. Method for the calculation of scanning tunneling microscope images and spectra. *Phys. Rev. B* **40**, 11990 (1989).
64. Choudhary, K. *et al.* Computational Scanning Tunneling Microscope Image Database. figshare <https://doi.org/10.6084/m9.figshare.c.3883270> (2020).
65. Song, F. *et al.* Low-temperature growth of bismuth thin films with (111) facet on highly oriented pyrolytic graphite. *ACS Appl. Mater. Interfaces* **7**, 8525 (2015).
66. Smeu, M., Guo, H., Ji, W. & Wolkow, R. A. Electronic properties of Si (111)-7×7 and related reconstructions: Density functional theory calculations. *Phys. Rev. B* **85**, 195315 (2012).



## Acknowledgements

K.C., K.F.G., C.C. and F.T. thank National Institute of Standards and Technology for funding, computational and data-management resources. S.V.K acknowledges support by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division. This research was in part supported by and conducted at the Center for Nanophase Materials Sciences (R.V.K., M.Z.), which is a DOE Office of Science User Facility. We also thank the computational support from XSEDE computational resources.

## Author contributions

K.C. developed the workflow, carried out the DFT calculations and trained the machine learning model. K.C. and K.G. analyzed the DFT data. C.C., S.K., R.V., M.Z. helped in the machine learning training. R.V., S.K. and M.Z. helped in the experimental validation. All contributed in writing the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-021-00824-y>.

**Correspondence** and requests for materials should be addressed to K.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2021