Prediction of Graphene Oxide Functionalization Using Gradient Boosting: Implications for Material Chemical Composition Identification

Bowen Zheng and Grace X. Gu*



ABSTRACT: Graphene oxides have exhibited alluring potential for state-of-the-art applications such as biomedical devices and functional nanocomposites. The types and concentrations of oxygen-containing functional groups are fingerprints of graphene oxides, dictating the properties and usage of the nanomaterial. Compared to pure graphene, the properties of graphene oxides are more challenging to model from a theoretical perspective, mainly because of the profound but implicit influences of the functional groups within. Machine learning is a potent method to uncover the hidden structure—property relations and to accelerate material discovery. Here, we develop a machine learning-based strategy to determine the functional groups. Trained by mechanical responses upon uniaxial tension computed by reactive molecular dynamics simulations, our proposed gradient boosting machine learning model can accurately identify the chemical composition of graphene oxides in the reserved data set. The difference in prediction accuracies between oxygen coverage and functional group composition is rationalized by graphene oxide molecular mechanisms. The proposed data-driven strategy can contribute to the predictive modeling of functionalized two-dimensional materials of a broad variety.

KEYWORDS: graphene oxide, machine learning, functional groups, molecular dynamics, mechanical property, chemical composition

1. INTRODUCTION

The discovery of graphene¹ has sparked significant research and industrial interest due to its extraordinary electronic,^{2,3} thermal,^{4,5} and mechanical^{6–10} properties. Graphene oxide (GO), one of the best-known graphene derivatives, opens up avenues for a broad spectrum of novel, tunable properties^{11–13} and is widely used in state-of-the-art applications such as biomedical devices,^{14–16} flexible electronics,^{17,18} and functional nanocomposites.^{19,20} GO is composed of a graphene basal plane and oxygen-containing functional groups such as epoxide (-O-), hydroxyl (-OH), carbonyl (C=O), and carboxyl (-COOH) groups. Analysis has shown that GO consists primarily of epoxide and hydroxyl groups and that only a small amount of carbonyl and carboxyl groups is present on the edges of GO sheets.²¹ The absolute and relative concentrations of epoxide and hydroxyl groups can be viewed as the fingerprint of GO, determining multiple important physical properties and therefore the potential usage of the nanomaterial. The control of oxygen content, epoxide-to-hydroxyl group ratio, and spatial distribution of functional groups has enabled rational material design and optimization.^{22,23}

Compared to pure graphene, the mechanical properties of GO are more challenging to model from a theoretical perspective, mainly because of the profound but implicit

Received: February 6, 2021 Accepted: February 9, 2021 Published: February 23, 2021







Figure 1. Schematic of GO sheets and mechanical responses upon tensile loading. (a) Schematic of the GO sheet with a side length of *L*. "a" and "z" refer to the armchair and zigzag directions of the GO basal plane. (b) Shape of the GO sheet upon full relaxation at a temperature of 300 K. (c) Comparison between the stress–strain curves of the present MD simulation and the result in ref.²⁶ The inset shows the GO structure used in this simulation case. (d) Stress–strain curves of GO sheets of various sizes with fixed parameters $\varphi_{\rm O} \approx 0.253$ and $\varphi_{\rm f} \approx 0.561$.

influences of the functional groups within. For example, hydroxyl groups cause GOs to behave in a brittle manner,²⁴ while epoxide groups may potentially increase the ductility via a mechanochemical epoxide-to-ether functional group transformation.^{25,26} In addition, with the same amount of oxygen atoms, different epoxide-to-hydroxyl group ratios yield diverse ultimate strengths.²⁷ Last but not least, the spatial distribution of functional groups triggers specific mechanochemical reactions under certain loading conditions,²⁶ which in turn affect the properties of the GO sheet. The correlation between collective properties of GO and the amount and types of functional groups, however, is not well understood.

Machine learning is a potent method to uncover the hidden structure-property relations and to accelerate new material discovery,^{28–35} which has been applied to study nanomaterials of a broad variety, including GO. Motevalli et al. used classification, regression, and causal inference to understand and predict the causes of defects in GO.³⁶ Motevalli et al. used multiple clustering machine learning models to determine representative structures of GO.³⁷ Amani et al. used regression models to estimate the temperature-dependent moduli of GO-reinforced nanocomposites.³⁸ However, the potential of machine learning has not been sufficiently used in predicting the degree and type of oxidation, which is one of the most defining and fundamental features of GO. The investigation of functionalization can shed light on how GOs can be optimally used in multiple research and technological fields such as flexible electronics, nonlinear optics, gas storage, and

lubrication. Some examples are as follows. First, the probe of oxygen coverage percentage can help determine if the GO of interest is a semiconductor or an insulator. Carbon atoms comprising of pure graphene are of sp² hybridization. Hydroxyl and epoxide groups are both responsible for the hybridization change from sp² to sp³. A substantial oxygen coverage renders GO insulating, while a low oxygen coverage makes it a semiconductor.³⁹ Second, the degree and type of oxidation can be used to estimate band gaps, which contain information of optical modulation. The combination of sp² and sp³ hybridization can break the symmetry and therefore result in band gaps.⁴⁰ Third, the knowledge of functional groups can be used to evaluate the applicability of GOs in gas storage. The reaction between hydroxyl groups and boronic acids can link GO layers together, forming a layered structure that provides room for gas storage.⁴¹ Fourth, the degree and type of oxidation can shed light on the friction coefficient of GO-based lubricant additives. It is shown that GO containing a high percentage of epoxide groups exhibits a better lubrication ability than GOs with a high content of hydroxyl groups.⁴² The above applications can be much better realized if we can access the functionalization properties of GO in quantitative details.

In this paper, we develop a machine learning-based strategy to determine the functionalization properties of monolayer GO sheets, which are quantified by two nondimensional features: the oxygen-to-carbon atom ratio and the fraction of epoxide groups. Data are prepared based on the mechanical responses upon a uniaxial tensile load, computed by molecular dynamics



Figure 2. Mechanical responses of four GO sheets with $\varphi_0 = 0.253$ and $\varphi_f = 0.561$ and random spatial distributions of functional groups. (a–d) Structures of four exemplar GOs. (e) Stress–strain curves and (f) system potential energy evolutions.

(MD) simulation with a reactive force field. A gradient boosting machine learning model, which is built up by forming an ensemble of weak prediction submodels in a stagewise fashion, is trained and used to predict the functionalization properties of unknown GO configurations. Machine learning features are extracted from both stress-strain relations and potential energy-related metrics, which successfully circumvent the uncertainty from the spatial distribution of functional groups. Physical insights into failure mechanisms associated with different functional groups are provided to rationalize prediction errors. This study demonstrates the power of machine learning models in uncovering complex, hidden structure-property relations in GO, offering possibilities for material discovery of a broader range using data-driven approaches.

2. RESULTS AND DISCUSSION

2.1. GO Configurations and MD Simulation. GOs in the present study are configured by following four sequential steps: (1) construction of a graphene basal plane; (2) assignment of epoxide groups; (3) assignment of hydroxyl groups; and (4) decoration of carbonyl and carboxyl groups on graphene basal plane edges. For the graphene basal plane, square-shaped monolayer graphene sheets with a side length of L = 3 nm are constructed, consisting of 446 carbon atoms in total and 336 nonedge atoms. For the epoxide group assignment, each epoxide group resides on two neighboring nonedge carbon atoms and there are 226 such pairs in total in the graphene basal plane. Additionally, the sp³ hybridization requires that one carbon atom cannot be associated to two epoxide groups. Herein, N_p pairs are randomly picked from all 226 possibilities and pairs that violate the sp³ hybridization requirement are removed from selection. For the hydroxyl group assignment, hydroxyl groups are hosted by nonedge sp²

atoms (not associated to any epoxide group). $N_{\rm a}$ nonedge carbon atoms that are not associated to epoxide groups are randomly picked to host hydroxyl groups. The numbers of epoxide and hydroxyl groups assigned to both sides are roughly equal. Lastly, the edges of the graphene basal plane are decorated with carbonyl and carboxyl groups, where the numbers of both functional groups obey a uniform distribution $\mathcal{U}(0, 20)$ and the locations are random.

To quantify the degree and type of oxidation independent of the absolute GO sheet size, the following two nondimensional features are used to label GO sheets: (1) oxygen-to-carbon atom ratio $\varphi_{\rm O}$, the ratio between total number of oxygen and carbon atoms $\varphi_{\rm O} = N_{\rm O}/N_{\rm C}$, where $N_{\rm O}$ and $N_{\rm C}$ are the total number of oxygen and carbon atoms and (2) fraction of epoxide group $\varphi_{\rm f}$, the ratio of the number of epoxide groups to the total number of hydroxyl and epoxide groups $\varphi_{\rm f} = N_{-\rm O-}/(N_{-\rm O-} + N_{-\rm OH})$, where $N_{-\rm O}$ and $N_{-\rm OH}$ are the total number of existence and hydroxyl functional groups. $\varphi_{\rm O}$ denotes the intensity of oxidation, while $\varphi_{\rm f}$ quantifies the relative concentrations of two distinct functional groups. For example, $\varphi_{\rm f} \rightarrow 0$ and $\varphi_{\rm f} \rightarrow 1$ indicate hydroxyl-rich and epoxide-rich, respectively.

Mechanical responses of various GO sheets upon uniaxial tensile loads are computed by reactive MD simulations, of which the implementation details are provided in the Supporting Information. The structure of GO in Figure 1(a) immediately after full relaxation at 300 K is shown in Figure 1(b), indicating that GO upon loading is not flat and shows a zigzag sheet-like structure. First, to validate the simulation setup, the stress-strain curve of a GO sheet with a side length of L = 3 nm is recorded and compared against a stress-strain curve in ref.²⁶ with L = 2 nm. To draw a fair comparison with the reported results, following ref.,²⁶ the fraction of oxidized carbon atoms of all four GO sheets is set as 0.36 and the

epoxide-to-hydroxyl functional group ratio is set as 4:1. Also, the same effective thickness $t_e = 0.75$ nm is used to calculate the normal stress component in the zigzag direction σ_{z}^{26} The comparison result is shown in Figure 1(c), which exhibits a good agreement despite not knowing the exact locations of functional groups and the difference in size. The von Mises stress σ is used for the rest of stress-strain curves in this paper, of which the calculation method is provided in the Supporting Information. Next, the effect of GO size is investigated. We perform MD simulations of GOs with side lengths of 3, 5, 7, and 10 nm, with fixed generic parameters $\varphi_0 \approx 0.253$ and $\varphi_f \approx$ 0.561, of which the results are shown in Figure 1(d). As can be seen, despite the difference in size and the uncertainty in functional group distribution, the results agree well in the general trend and the ultimate strength, indicating the reliability and reasonable dimensional scalability of the current simulation settings. Despite larger GO sheets (L = 7 nm and L= 10 nm) render smoother stress-strain curves, L = 3 nm enjoys the advantage of much reduced computational expenses, which makes this choice of GO size more realistic in the preparation of a large amount of data. This also poses greater challenges to machine learning implementation to overcome the fluctuation issue, which is no less uncommon in an experimental setting. A successful prediction with L = 3 nmcan indicate good performance with noisy data. In addition, L = 3 nm creates more variations of functional group distribution compared to L = 2 nm in the reported results,²⁶ which facilitates the generation of more diverse GO configurations that benefit the training and extrapolation ability of the machine learning model. The temperature effect toward the mechanical responses and the stability of GO is also investigated. Stress-strain curves of GO with parameters φ_{O} = 0.253 and $\varphi_{\rm f}$ = 0.561 under various temperatures (100, 200, 300, 400, and 1000 K) are recorded and are shown in Figure S1(a). The results show that as the temperature increases, GO is slightly softened and the mechanical properties are not subject to major changes. Figure S1(b) shows the GO structure immediately after relaxation at T = 1000 K, which suggests that no thermal instability has occurred. To approximate a realistic setting, the rest of the MD simulations in this paper are performed at 300 K.

It is critical that a specific combination of the oxygen-tocarbon atom ratio and the epoxide-to-hydroxyl group ratio is not sufficient to reflect a single GO configuration because the parameter pair does not possess information of locations of functional groups. In fact, there exists an effectively infinite number of different functional group spatial distributions, whose effect is nontrivial. The effect of distribution can exist when investigating multiple randomly generated GO sheets with the same φ_{0} and φ_{f} . To this end, four GO sheets with φ_{0} ≈ 0.253 and $\varphi_{\rm f} \approx 0.561$ but randomly diverse functional group distributions are simulated, of which the structures after relaxation are shown in Figure 2(a-d). The stress-strain curves of these GO sheets are shown in Figure 2(e), which indicates that the functional group distribution has a profound effect on the mechanical properties of GO. Although the ultimate strengths are roughly on the same level, the ductility and curve shape differ dramatically. This finding implies that one unique combination of $\varphi_{\rm O}$ and $\varphi_{\rm f}$ does not yield one unique mechanical response. Therefore, to reliably predict $\varphi_{
m O}$ and $\varphi_{\rm f}$ using mechanical responses of GO, more useful features need to be used besides the ones extracted from stress-strain relations.

To this end, the energy-related phenomena are visited. The system potential energy evolutions of the four GO sheets mentioned above are computed, as shown in Figure 2(f). It is shown that there are substantial differences in these curves. GO sheets with the same $\varphi_{\rm O}$ and $\varphi_{\rm f}$ are characterized by distinct starting energy (energy at thermal vibration before subject to loading), maximum energy increase, and the general curve shape. It can be assumed that these differences in strainstrain relations and system potential energy evolutions stem from different functional group distributions. It is envisioned that by extracting features from both stress-strain curves and potential energy evolutions, a unique (or approximately unique) combination of φ_0 and φ_f may be obtained, which makes machine learning-based prediction possible. If so, the issue of random functional group spatial distribution can be circumvented. It is worth noting that energy-related quantities can be more challenging to measure in experiments compared to mechanical stress and strain. To this end, when extracting features from potential energy curves, we will take the experimental viability into account.

2.2. Machine Learning Results. Our machine learning model implementation using computational tensile test results can be broken down into the following three phases: (1) sample preparation, (2) feature extraction, and (3) model training and testing. For sample preparation, 1570 sample instances of mechanical responses of GO are computed for training and validation, while 100 sample instances are reserved for testing. The distributions of all 1670 sample points on the $N_{-O-}-N_{-OH}$ plane and the $\varphi_O-\varphi_f$ plane are shown in Figure S2. While $\varphi_{\rm f}$ naturally ranges from 0 to 1, $\varphi_{\rm O}$ ranges from 0.1 to 0.8. In an experimental setting, φ_0 varies from 0.25 to 0.75,40 which has been covered by the range of our sample space. Data for training and validation are randomly shuffled together and then separated into $N_{\text{training}} =$ 1470 and $N_{\text{validation}} = 100$ before use, where N_{training} and $N_{
m validation}$ are the number of training and validation data. During training, the shuffle-and-split process is repeated 20 times where each time is based on a different random seed. The aim of this practice is to avoid good results by chance and to ensure better generalized results. Test data totaling $N_{\text{test}} = 100$ are not involved in the shuffling with training and validation data, implying that the test data represent future unseen configurations. A good test accuracy can indicate promising extrapolation ability of the machine learning model.

For feature extraction, the features to extract from both stress-strain relations and potential energy properties comprise of (1) ultimate strength σ_{u} , (2) strain at ultimate strength $\varepsilon_{\rm w}$ and (3) total potential energy at room-temperature-free vibration $U_{\rm V}$ (the starting energy of potential energy evolution). In an experimental setting, vibrational energy distributions can be measured using a scanning force microscope⁴³ or by interferometry,⁴⁴ after which the total potential energy can be computed. From a physics point of view, the above features can be related to the target properties, namely, the oxygen-to-carbon atom ratio and the relative concentrations of functional groups, in the following ways. First, ultimate strength σ_{u} is dependent on the sheer amount and types of bond interactions due to functionalization. For a pure graphene, the ultimate strength is dictated only by carbon-carbon (C-C) bonds formed by sp^2 carbon atoms. Oxidation not only adds new carbon-oxygen (C-O) covalent bond interactions but also disturbs the sp² carbon atom lattice by introducing sp³ hybridization, thus affecting the ultimate



Figure 3. Extracted features as a function of φ_0 and φ_f (a) σ_w (b) ε_w and (c) U_V as a function of φ_0 and φ_f Data are from the training and the validation sets.

strength. Second, strain at ultimate strength ε_u is impacted by both the amount of hydroxyl and epoxide groups, where the former tends to result in brittleness and the latter can trigger mechanochemical epoxide-to-ether functional group transformation to enhance ductility. The coexistence of two functional groups gives rise to uncertainty. Third, the total potential energy at room-temperature-free vibration U_V is highly related to molecular weight and therefore to oxygen percentage. The general trend is that the higher the oxygen percentage, the higher the molecular weight, the lower the potential energy when GO is at rest. In addition, the relative concentrations of functional groups also play an implicit role in affecting the potential energy. $\sigma_{uv} \varepsilon_{uv}$ and U_V as a function of φ_O and φ_f are shown in Figure 3, suggesting that U_V is the strongest indicator of GO functionalization, followed by σ_u and ε_u .

The procedure of machine learning model training and testing is described as follows. For the i^{th} sample instance, the sample vector can be formulated as $\mathbf{x}_i = [\sigma_{u,i} \ \varepsilon_{u,i} \ U_{V,i}]^T \in \mathbb{R}^3$ and the label vector can be formulated as $\mathbf{y}_{\!_i} = [\varphi_{\mathrm{O},i} \; \varphi_{\mathrm{f},i}]^{\mathrm{T}} \in \mathbb{R}^2$. A gradient boosting machine learning model, which is built by forming an ensemble of weak prediction submodels in a stagewise fashion, is used to map the above features to φ_{Ω} and $\varphi_{\rm f}$. Tunable hyperparameters include learning rate (shrinkage factor) α and the number of boosting stage to perform M. Algorithmic details of the gradient boosting model are provided in Table S1. The framework of machine learning implementation is summarized in the flow chart shown in Figure 4. The coefficient of determination R^2 as well as the mean squared error (MSE) are used as metrics to quantify the model predictions. Particularly for the machine learning training session, the averages of R^2 and MSE over 20 random shuffle-and-splits, denoted as R_{avg}^2 and MSE_{avg}, are used as metrics. The hyperparameter tuning results of $\varphi_{\rm O}$ and $\varphi_{\rm f}$ in the training session are shown in Figure S3. The results provide an optimal hyperparameter combination $\alpha^* = 0.11$ and $M^* = 70$, scoring R_{avg}^2 = 0.980 and MSE_{avg} = 0.000488 for φ_0 and R_{avg}^2 = 0.906 and $MSE_{avg} = 0.00661$ for φ_{f} Convergence with respect to the size of the training data set is shown by plotting R_{avg}^2 and MSE_{avg} of φ_{O} and φ_{f} on the validation data as a function of the number of training data, as shown in Figures S5,S6.

After the training process, predictions are performed on the reserved test set using optimal hyperparameters α^* and M^* , which yields $R^2 = 0.978$ and MSE = 0.000499 for φ_0 and $R^2 = 0.870$ and MSE = 0.0103 for $\varphi_{\dot{p}}$ slightly inferior to the performance on the validation set. Prediction results of φ_0 and



Figure 4. Flow-chart presentation of the machine learning procedure.

 $\varphi_{\rm f}$ are shown in Figure 5. The differences between the predicted label and the true label (denoted as $d_{\rm O} = \varphi_{\rm O} - \hat{\varphi}_{\rm O}$ and $d_{\rm f} = \varphi_{\rm f} - \hat{\varphi}_{\rm f}$ where $\hat{\varphi}_{\rm O}$ and $\hat{\varphi}_{\rm f}$ are the predicted labels) of all 100 sample points in the test set are plotted against the true labels, as shown in Figure 5(a,b). For φ_{0} , the difference between the predicted label and the true label for most sample points ranges between +0.06 and -0.06, and for $\varphi_{\rm fr}$ the difference for most sample points ranges between +0.02 and -0.02 (excluding one sample point with a true label $\varphi_f = 1.0$). For both labels, no dependency of prediction error on the absolute value of true label is detected. Based on the results of Figure 5(a,b), histograms of $|d_0|$ and $|d_f|$ are constructed, as shown in Figure 5(c,d). For φ_0 , 63% of GO sheets achieved an error of <1%, while 99% of GO sheets achieved an error of <5%. For $\varphi_{\rm fr}$ 58% of GO sheets achieved an error of <5%, while 95% of GO sheets achieved an error of <20%. Additionally, a kernel ridge regression (KRR) model, which allows to progressively build up model complexity without adding to the computational cost (the kernel trick), is used to make predictions. Tunable hyperparameters include polynomial kernel degree p and regularization parameter λ . Using the optimal hyperparameters after tuning, KRR achieved an inferior prediction performance compared to gradient boosting. The algorithmic details and the hyperparameter tuning process of KRR are provided in the Supporting Information. When setting $\lambda = 0$ and p = 1, KRR equates a



Figure 5. Evaluation of machine learning prediction accuracies of φ_0 and φ_f on the test data. Scatter plots of (a) d_0 against φ_0 and (b) d_f against φ_{f^*} . Histograms of (c) $|d_0|$ and (d) $|d_f|$. The data point with a d_f of ~-0.7 is not represented in (d).

multivariate least squares regression, of which the results are also provided in the Supporting Information.

One observation regarding the prediction of φ_0 and φ_f is that the former can be more accurately predicted with models varying from gradient boosting to multivariate least squares regression, while the latter is far more demanding. In fact, using only $U_{\rm V}$, our gradient boosting model can yield a R^2 of ~0.9, as shown in Figure S7. A R^2 close to 1 can be achieved if all features are used for both gradient boosting and KRR models. Our rationale is that φ_0 is directly related to the molecular weight of GO, which poses a major effect on the potential energy especially when GO is at rest (when U_V is extracted). Assisted by features related to mechanical stress and strain, $\varphi_{
m O}$ can be accurately pinned down. Meanwhile, $\varphi_{
m f}$ contains little molecular weight information. The composition of functional groups does not have a significant impact on the potential energy as molecular weight, reflected by the failure to predict $\varphi_{\rm f}$ with only $U_{\rm V}$. Instead, $\varphi_{\rm f}$ dictates the types of bonding interaction, making the difference between hydroxyl and epoxide group properties to play a major role under different $\varphi_{\rm f}$ values. For example, the difference in failure mechanisms of hydroxyl and epoxide groups begins to take effect. Hydroxyl groups lead to a brittle failure mechanism where only C-C bonds are broken. Each hydroxyl group attaches to only one carbon atom on the basal plane, and the failure associated with a hydroxyl group happens only to one of the C-C bonds in its immediate vicinity. The C-O bonds in hydroxyl groups remain intact at all times. Conversely, epoxide groups can result in a progressive, ductile failure. Each epoxide group attaches to two bonded carbon atoms on the basal plane. If stress causes the bond between these two carbon atoms to break, the two initially bonded carbon atoms can still be

connected by two C-O bonds, transforming the epoxide group to a new ether group. These two C-O bonds can continue to stretch and bear loads. After the stress reaches the C-O bond strength, one of the two C-O bonds will break and result in the failure of the local area. It is possible that the C-O bond strength is never reached. This happens when a catastrophic failure of the GO sheet has already taken place caused by crack propagation elsewhere. In addition, the mechanical properties of epoxide groups are sensitive to orientation. If the two bonded carbon atoms are aligned in the stretching direction, the epoxide group is more likely to fail; if the two bonded carbon atoms are perpendicular to the stretch, the epoxide group is less likely to fail. The above mechanisms together with random functional group spatial distributions increase the problem complexity drastically. One possibility to tackle the elevated complexity is to train neural networks with entire stress-strain curves as the input vectors, as inspired by ref.⁴⁵ We will consider neural networks in our future work that involves more complex systems and physics.

3. CONCLUSIONS

In this paper, we develop a machine learning-based strategy to determine the functional group contents of monolayer GO sheets. By constructing a feature space with stress-based and energy-based mechanical responses computed by reactive MD simulations, the oxygen-to-carbon atom ratio and the relative concentrations of epoxide and hydroxyl groups in GO are predicted by gradient boosting. Despite the uncertainty brought by the random, uncontrolled functional group spatial distributions, our feature selection from both stress–strain relation and system potential energy enables the machine learning model to overcome this issue. The best prediction results of the reserved test data show that for the oxygen-tocarbon atom ratio, all test samples have achieved a prediction error of <5%, while for the fraction of epoxide groups, 95% of the test samples have achieved a prediction error of <20%. The difference in prediction accuracies between the oxygen coverage and the functional group composition is rationalized by GO molecular mechanisms. The proposed data-driven strategy may also shed light on the predictive modeling and identification of functionalized two-dimensional materials of a broad variety. In the current research, the targeted features mainly focus on the chemical composition of GOs and do not reflect structural properties such as locations of specific functional groups, which, together with predicting properties of multilayer and defect-containing GOs, will be addressed in a future work.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsanm.1c00384.

MD simulation details, stress calculation method, correlation between machine learning features, KRR results, hyperparameter tuning process results, and details of machine learning algorithms (PDF)

AUTHOR INFORMATION

Corresponding Author

Grace X. Gu – Department of Mechanical Engineering, University of California, Berkeley, California 94720, United States; o orcid.org/0000-0001-7118-3228; Email: ggu@ berkeley.edu

Author

Bowen Zheng – Department of Mechanical Engineering, University of California, Berkeley, California 94720, United States

Complete contact information is available at: https://pubs.acs.org/10.1021/acsanm.1c00384

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work used the Extreme Science and Engineering Discovery Environment (XSEDE) Bridges system, which is supported by the National Science Foundation grant number ACI-1548562. The authors also acknowledge support from an NVIDIA GPU seed grant.

REFERENCES

(1) Novoselov, K. S.; Geim, A. K.; Morozov, S. V.; Jiang, D.; Zhang, Y.; Dubonos, S. V.; Grigorieva, I. V.; Firsov, A. A. Electric Field Effect in Atomically Thin Carbon Films. *Science* **2004**, *306*, 666–669.

(2) Berger, C.; Song, Z.; Li, T.; Li, X.; Ogbazghi, A. Y.; Feng, R.; Dai, Z.; Marchenkov, A. N.; Conrad, E. H.; First, P. N.; de Heer, W. A. Ultrathin Epitaxial Graphite: 2D Electron Gas Properties and a Route toward Graphene-based Nanoelectronics. *J. Phys. Chem. B* **2004**, *108*, 19912–19916.

(3) Withers, F.; Dubois, M.; Savchenko, A. K. Electron properties of fluorinated single-layer graphene transistors. *Phys. Rev. B* 2010, *82*, No. 073403.

(4) Balandin, A. A. Thermal properties of graphene and nanostructured carbon materials. *Nat. Mater.* **2011**, *10*, 569–581.

(5) Balandin, A. A.; Ghosh, S.; Bao, W.; Calizo, I.; Teweldebrhan, D.; Miao, F.; Lau, C. N. Superior Thermal Conductivity of Single-Layer Graphene. *Nano Lett.* **2008**, *8*, 902–907.

(6) Zheng, B.; Gu, G. X. Tuning the graphene mechanical anisotropy via defect engineering. *Carbon* **2019**, *155*, 697–705.

(7) Lee, C.; Wei, \bar{X} .; Kysar, J. W.; Hone, J. Measurement of the Elastic Properties and Intrinsic Strength of Monolayer Graphene. *Science* **2008**, 321, 385–388.

(8) Zheng, B.; Gu, G. X. Recovery from mechanical degradation of graphene by defect enlargement. *Nanotechnol.* **2019**, *31*, 085707.

(9) Zhang, P.; Ma, L.; Fan, F.; Zeng, Z.; Peng, C.; Loya, P. E.; Liu, Z.; Gong, Y.; Zhang, J.; Zhang, X.; Ajayan, P. M.; Zhu, T.; Lou, J. Fracture toughness of graphene. *Nat. Commun.* **2014**, *5*, 3782.

(10) Zheng, B.; Gu, G. X. Stress Field Characteristics and Collective Mechanical Properties of Defective Graphene. J. Phys. Chem. C 2020, 124, 7421–7431.

(11) Zhu, Y.; Murali, S.; Cai, W.; Li, X.; Suk, J. W.; Potts, J. R.; Ruoff, R. S. Graphene and Graphene Oxide: Synthesis, Properties, and Applications. *Adv. Mater.* **2010**, *22*, 3906–3924.

(12) Chen, D.; Feng, H.; Li, J. Graphene Oxide: Preparation, Functionalization, and Electrochemical Applications. *Chem. Rev.* **2012**, *112*, 6027–6053.

(13) Chen, C.-T.; Martin-Martinez, F. J.; Ling, S.; Qin, Z.; Buehler, M. J. Nacre-inspired design of graphene oxide-polydopamine nanocomposites for enhanced mechanical properties and multi-functionalities. *Nano Futures* **2017**, *1*, No. 011003.

(14) Chung, C.; Kim, Y.-K.; Shin, D.; Ryoo, S.-R.; Hong, B. H.; Min, D.-H. Biomedical Applications of Graphene and Graphene Oxide. *Acc. Chem. Res.* **2012**, *46*, 2211–2224.

(15) Georgakilas, V.; Tiwari, J. N.; Kemp, K. C.; Perman, J. A.; Bourlinos, A. B.; Kim, K. S.; Zboril, R. Noncovalent Functionalization of Graphene and Graphene Oxide for Energy Materials, Biosensing, Catalytic, and Biomedical Applications. *Chem. Rev.* **2016**, *116*, 5464– 5519.

(16) Jian, W.; Hui, D.; Lau, D. Nanoengineering in biomedicine: Current development and future perspectives. *Nanotechnol. Rev.* **2020**, *9*, 700–715.

(17) Jeong, H. Y.; Kim, J. Y.; Kim, J. W.; Hwang, J. O.; Kim, J.-E.; Lee, J. Y.; Yoon, T. H.; Cho, B. J.; Kim, S. O.; Ruoff, R. S.; Choi, S.-Y. Graphene Oxide Thin Films for Flexible Nonvolatile Memory Applications. *Nano Lett.* **2010**, *10*, 4381–4386.

(18) Li, X.; Tang, Y.; Song, J.; Yang, W.; Wang, M.; Zhu, C.; Zhao, W.; Zheng, J.; Lin, Y. Self-supporting activated carbon/carbon nanotube/reduced graphene oxide flexible electrode for high performance supercapacitor. *Carbon* **2018**, *129*, 236–244.

(19) Hu, S.-H.; Chen, Y.-W.; Hung, W.-T.; Chen, I.-W.; Chen, S.-Y. Quantum-Dot-Tagged Reduced Graphene Oxide Nanocomposites for Bright Fluorescence Bioimaging and Photothermal Therapy Monitored In Situ. *Adv. Mater.* **2012**, *24*, 1748–1754.

(20) Yang, X.; Tu, Y.; Li, L.; Shang, S.; Tao, X.-M. Well-Dispersed Chitosan/Graphene Oxide Nanocomposites. *ACS Appl. Mater. Interfaces* **2010**, *2*, 1707–1713.

(21) Sun, Y.; Tang, X.; Bao, H.; Yang, Z.; Ma, F. The effects of hydroxide and epoxide functional groups on the mechanical properties of graphene oxide and its failure mechanism by molecular dynamics simulations. *RSC Adv.* **2020**, *10*, 29610–29617.

(22) Wei, Z.; Wang, D.; Kim, S.; Kim, S.-Y.; Hu, Y.; Yakes, M. K.; Laracuente, A. R.; Dai, Z.; Marder, S. R.; Berger, C.; King, W. P.; de Heer, W. A.; Sheehan, P. E.; Riedo, E. Nanoscale Tunable Reduction of Graphene Oxide for Graphene Electronics. *Science* **2010**, *328*, 1373–1376.

(23) Loh, K. P.; Bao, Q.; Eda, G.; Chhowalla, M. Graphene oxide as a chemically tunable platform for optical applications. *Nat. Chem.* **2010**, *2*, 1015.

(24) Tavakol, M.; Montazeri, A.; Aboutalebi, S. H.; Asgari, R. Mechanical properties of graphene oxide: The impact of functional groups. *Appl. Surf. Sci.* **2020**, 525, 146554.

(25) Verma, A.; Parashar, A. Molecular dynamics based simulations to study failure morphology of hydroxyl and epoxide functionalised graphene. *Comput. Mater. Sci.* **2018**, *143*, 15–26.

(26) Wei, X.; Mao, L.; Soler-Crespo, R. A.; Paci, J. T.; Huang, J.; Nguyen, S. T.; Espinosa, H. D. Plasticity and ductility in graphene oxide through a mechanochemically induced damage tolerance mechanism. *Nat. Commun.* **2015**, *6*, 8029.

(27) Soler-Crespo, R. A.; Gao, W.; Xiao, P.; Wei, X.; Paci, J. T.; Henkelman, G.; Espinosa, H. D. Engineering the Mechanical Properties of Monolayer Graphene Oxide at the Atomic Level. *J. Phys. Chem. Lett.* **2016**, *7*, 2702–2707.

(28) Chen, C.-T.; Gu, G. X. Machine learning for composite materials. *MRS Commun.* **2019**, *9*, 556–566.

(29) Schmidt, J.; Marques, M. R. G.; Botti, S.; Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* **2019**, *5*, 83.

(30) Zheng, B.; Gu, G. X. Machine Learning-Based Detection of Graphene Defects with Atomic Precision. *Nano-Micro Lett.* **2020**, *12*, 181.

(31) Hanakata, P. Z.; Cubuk, E. D.; Campbell, D. K.; Park, H. S. Accelerated search and design of stretchable graphene kirigami using machine learning. *Phys. Rev. Lett.* **2018**, *121*, 255304.

(32) Chen, C. T.; Gu, G. X. Generative Deep Neural Networks for Inverse Materials Design Using Backpropagation and Active Learning. *Adv. Sci.* **2020**, 1902607.

(33) Jacobs, E. W.; Yang, C.; Demir, K. G.; Gu, G. X. Vibrational detection of delamination in composites using a combined finite element analysis and machine learning approach. *J. Appl. Phys.* **2020**, *128*, 125104.

(34) Pilania, G.; Wang, C.; Jiang, X.; Rajasekaran, S.; Ramprasad, R. Accelerating materials property predictions using machine learning. *Sci. Rep.* **2013**, *3*, 2810.

(35) Zhang, Z.; Hong, Y.; Hou, B.; Zhang, Z.; Negahban, M.; Zhang, J. Accelerated discoveries of mechanical properties of graphene using machine learning and high-throughput computation. *Carbon* **2019**, *148*, 115–123.

(36) Motevalli, B.; Sun, B.; Barnard, A. S. Understanding and Predicting the Cause of Defects in Graphene Oxide Nanostructures Using Machine Learning. *J. Phys. Chem. C* **2020**, *124*, 7404–7413.

(37) Motevalli, B.; Parker, A. J.; Sun, B.; Barnard, A. S. The representative structure of graphene oxide nanoflakes from machine learning. *Nano Futures* **2019**, *3*, No. 045001.

(38) Amani, M. A.; Ebrahimi, F.; Dabbagh, A.; Rastgoo, A.; Nasiri, M. M. A machine learning-based model for the estimation of the temperature-dependent moduli of graphene oxide reinforced nano-composites and its application in a thermally affected buckling analysis. *Eng. Computation* **2020**, 1–11.

(39) Eda, G.; Mattevi, C.; Yamaguchi, H.; Kim, H.; Chhowalla, M. Insulator to Semimetal Transition in Graphene Oxide. *J. Phys. Chem. C* 2009, *113*, 15768–15771.

(40) Johari, P.; Shenoy, V. B. Modulating Optical Properties of Graphene Oxide: Role of Prominent Functional Groups. *ACS Nano* **2011**, *5*, 7640–7647.

(41) Burress, J. W.; Gadipelli, S.; Ford, J.; Simmons, J. M.; Zhou, W.; Yildirim, T. Graphene Oxide Framework Materials: Theoretical Predictions and Experimental Results. *Angew. Chem Int. Ed.* **2010**, *49*, 8902–8904.

(42) Gupta, B.; Kumar, N.; Panda, K.; Kanan, V.; Joshi, S.; Visoly-Fisher, I. Role of oxygen functional groups in reduced graphene oxide for lubrication. *Sci. Rep.* **2017**, *7*, 45030.

(43) Garcia-Sanchez, D.; van der Zande, A. M.; Paulo, A. S.; Lassagne, B.; McEuen, P. L.; Bachtold, A. Imaging Mechanical Vibrations in Suspended Graphene Sheets. *Nano Lett.* **2008**, *8*, 1399– 1403.

(44) Bunch, J. S.; van der Zande, A. M.; Verbridge, S. S.; Frank, I. W.; Tanenbaum, D. M.; Parpia, J. M.; Craighead, H. G.; McEuen, P. L. Electromechanical Resonators from Graphene Sheets. *Science* **2007**, *315*, 490–493.

(45) Lewis, C.; Erikson, J. W.; Sanchez, D. A.; McClure, C. E.; Nordin, G. P.; Munro, T. R.; Colton, J. S. Use of Machine Learning with Temporal Photoluminescence Signals from CdTe Quantum Dots for Temperature Measurement in Microfluidic Devices. *ACS Appl. Nano Mater.* **2020**, *3*, 4045–4053.