ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Learning Multidimensional Electronegativity of Selected Atom Types in Organic Molecules Using Graph Neural Networks

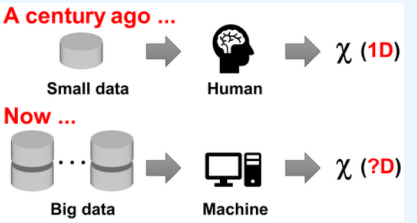
Da Bean Han, Gyoung S. Na, and Hyun Woo Kim*

Cite This: *ACS Omega* 2025, 10, 48740−48749

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Electronegativity can be considered a data-driven concept that has been widely used since Pauling proposed this property. However, updating the electronegativity based on the vast amount of high-quality experimental and computational data has been overlooked. Thus, advances in artificial intelligence (AI), with its ability to manage large data sets and identify underlying patterns, necessitate reconsidering data-driven concepts such as electronegativity. In this work, we present a data-driven method to generate more informative multidimensional electronegativity of atoms in organic molecules using graph neural networks. Although this electronegativity can be extended to any dimension, we focused on 2D electronegativity to do a more detailed classification of the atoms and their covalent bonds. By replacing the conventional electronegativity with the newly proposed one, we observed performance improvement in molecular machine learning tasks. We believe that our findings will be useful in understanding electronegativity and chemical bonds by applying AI-driven methods to chemical studies.



INTRODUCTION

Electronegativity, χ, which explains the ability of atoms to attract electrons in chemical bonds, becomes a key indicator in chemistry since it was proposed by Pauling in 1932.[1,2] Pauling defined electronegativity, χ_{Pauling} by using the energy stabilized after forming a heterodimer of two atoms, A and B as

$$|\chi_A - \chi_B| = \sqrt{D_{AB} - \frac{1}{2}(D_{AA} + D_{BB})} \quad (1)$$

where D_{AB} is the bond energy between atoms A and B and $\frac{1}{2}(D_{AA} + D_{BB})$ is the average bond energy of covalently bonded homodimers. By construction, χ_{Pauling} is a data-driven concept defined by expert intuition in the mid-20th century. For a long time, χ_{Pauling} has played an important role in describing the nature of chemical bonds in molecules and materials.[3−5] Although the concept proposed by Pauling has been the most widely used, electronegativity has been revisited by many researchers. Huggins and Sanderson refined the electronegativity in their separate works.[6−8] Allred completed the modern χ_{Pauling} table by referencing the enthalpy of formation.[9,10] Mulliken defined electronegativity as the mean of ionization energy and electron affinity,[11,12] and Bratsch extended this concept to *sp* hybridizations for elements across the periodic table.[13] Sanderson defined electronegativity as the driving force of bonding, while Iczkowski represented this quantity through the chemical potential.[7,14] Later, Gordy,[15] Allred,[9] and Ghosh[16,17] used atomic radii to explain electronegativity, while Hinze et al. defined it based on orbital characteristics rather than considering the entire atom.[18] Allen proposed an electronegativity using the average energy of valence electrons.[19] These reinterpretations show that electronegativity is a long-established semiempirical concept that can effectively serve as a qualitative guide for predicting electronic density rearrangements in molecules.[20] Moreover, they imply that there is much room for further consideration of the nature of the chemical bonds through electronegativity.

For decades, data sets have become more accurate and extensive, and advances in data science facilitate the processing of such data sets with the help of artificial intelligence (AI), including machine learning (ML). AI enables researchers to discover hidden patterns in the data set, thus driving innovation in many scientific fields.[21−23] Moreover, ML has been applied to various topics in the field of chemistry, including the prediction of molecular properties[24−28] and chemical reactions,[29−33] as well as the modeling of novel molecules and materials.[34−36] Due to its broad impact on various topics, certain aspects remain unaddressed. A more comprehensive review of ML applications in chemistry can be found in recent review and perspective articles.[29,37−42]

Applying ML in chemistry has been accelerated by the accumulation of large molecular data sets, including PubChem,[43] ChEMBL,[44] GBD-17,[45] QM9,[46] MolData,[47] and ZINC.[48] Using these data sets, ML models can predict molecular properties in a data-driven manner.[49] In this regard, molecular representations have been essential for transforming

Received: July 10, 2025
Revised: September 22, 2025
Accepted: September 29, 2025
Published: October 9, 2025

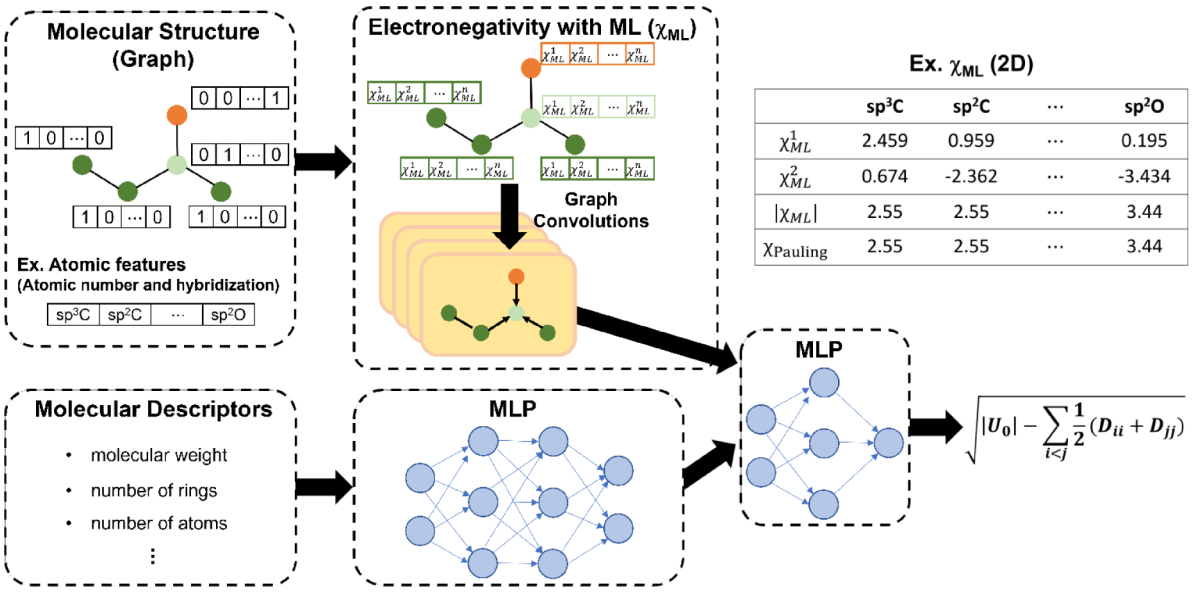


Figure 1. Workflow of χ_{ML} generation. Molecular structures were represented in graphs and further updated through graph convolutions. Multilayer perceptrons (MLPs) were used to predict target values using molecular descriptors and graph features.

structural information in chemical data sets into machine-readable formats.[39,50] As typical ML models are considered blackbox models, explainable AI (XAI) attracted considerable interest in the interpretation of structure−activity relationships,[51,52] predicting reactions,[53,54] and drug design.[55] One might wonder whether large chemical data sets can be used to redefine data-driven concepts such as electronegativity with AI. To the best of our knowledge, the concept of electronegativity has not been extensively explored by employing AI methods on readily accessible chemical databases. In addition, we argue that we can generate more informative electronegativity using AI methods.

In this study, inspired by Pauling's concept in eq 1, we proposed a ML-based electronegativity, χ_{ML}, based on the QM9 data set with stable small organic molecules. We can suggest χ_{ML} as a multidimensional vector for various types of atoms by employing linear regression (LR) and graph convolutional networks (GCNs).[56,57] LR was employed to transform one-hot encoding vectors into multidimensional electronegativites. We selected LR in this work due to its simplicity, interpretability, and lower risk of overfitting.[58,59] Among various graph neural networks (GNNs),[60,61] we used GCNs in this study as they are a representative example and widely used in the ML community to ensure that multidimensional electronegativity is effectively generated using well-established methods. As shown in Figure 1, the initial atomic feature vector of each atom was transformed to multidimensional χ_{ML}, and then χ_{ML} was used as the input variable for the corresponding atom to train GCN. Thus, GCN utilized the connections between atoms in a molecular structure to update the atomic features. GCN allowed us to generate atomic properties by incorporating bond information within the molecule. All GNNs in this work employed molecular descriptors, as illustrated in Figure 1. With these, we were able to generate χ_{ML} that effectively explains eq 1 from the perspective of AI. For the five elements (H, C, N, O, and F), which are commonly found in organic compounds, we classified bond types in a detailed manner compared to the traditional approach using $\Delta\chi_{Pauling}$. Our analysis revealed categories of carbon−carbon bonds that were not noticed by the traditional approach. Our AI-aided approach broadened our understanding of covalent bonds and could be expanded to all chemical bonds. Furthermore, the newly generated electronegativity, χ_{ML}, has significant potential in molecular ML. The prediction accuracy of the ML models was improved when χ_{ML} was used as input variables of molecular ML. Additionally, we found that applying the data-driven categories of covalent bonds to molecular ML provided a more accurate prediction of molecular properties with relational graph convolutional networks (RGCNs).[62] We believe that our data-driven multidimensional electronegativity shows the potential to enhance our understanding of chemical bonds. Furthermore, our findings also suggest the possibility of applying AI to refine other data-driven concepts in chemistry.[63,64]

METHODS

Graph Convolutional Neural Networks. In this work, we used graph convolutional networks (GCNs)[56] and their heterogeneous graph version, relational graph convolutional networks (RGCNs)[62] using the PyTorch framework[57] and the PyTorch Geometric library.[65] GCN extracts information from graph-structured data by iteratively updating node information based on node features and their connectivity. In molecular ML, molecules are modeled as graphs, where atoms correspond to nodes and chemical bonds correspond to edges. Through iterative update and message passing, molecular representations are extracted, and each update step can be represented as

$$H^{(l+1)} = \sigma(\tilde{A}H^{(l)}W^{(l)}) \quad (2)$$

Here, $H^{(l)}$ represents the feature matrix of the atoms at the lth layer, where each row corresponds to the feature vector of a single atom. The feature matrix of the $(l + 1)$th layer, $H^{(l+1)}$, was then obtained by using the feature matrix of the previous layer, $H^{(l)}$. $\tilde{A}$ is the normalized adjacency matrix containing the connectivity information of the atoms, and σ is the activation function. We selected the sigmoid linear unit (SiLU) as the

https://doi.org/10.1021/acsomega.5c06731
ACS Omega 2025, 10, 48740−48749

activation function in this work based on its improved performance over the rectified linear unit (ReLU).[66] During the training phase of GCN, the weight parameter, $W^{(l)}$, is optimized to predict the data in the training set. As an extended version of GCN, RGCN treats heterogeneous relationships between nodes in the graph by assigning different weights to each edge.[62] The graph convolution in RGCN is given by

$$H^{(l+1)} = \sigma\left(\hat{A}_0 H^{(l)} W_0^{(l)} + \sum_{r\in\mathcal{R}} \tilde{A}_r H^{(l)} W_r^{(l)}\right) \tag{3}$$

where R denotes the set of bond types and $\tilde{A}_r$ and $\hat{A}_0$ are adjacency matrices for the rth bond type and the atom itself, respectively. In practice, trainable weight parameters are $W_r^{(l)}$ and $W_0^{(l)}$. The relationship-specific information is learnable by optimizing $W_r^{(l)}$ while the other information is by optimizing $W_0^{(l)}$. The SiLU activation function was chosen to ensure consistency between the GCNs and RGCNs. In this work, GCNs were used to generate χ_{ML} and to predict molecular properties. RGCNs were used to predict molecular properties through a more detailed classification with χ_{ML}.

ML Approach for Generating Electronegativity. We redefined electronegativity using a data-driven approach, as ML effectively revealed hidden patterns in data. To achieve this, we chose a sufficiently large data set with appropriate molecular properties. We selected the QM9 data set that contains properties of about 134k organic molecules.[46] Modifying the atomization energy at 0 K, U_0, we intuitively modeled the target value to generate χ_{ML} as

$$f(\{\chi_{ML}\}) \approx \sqrt{|U_0| - \sum_{i<j} \frac{1}{2}(D_{ii} + D_{jj})} \tag{4}$$

where f denotes an ML model function with a set of atomic electronegativity as inputs. Here, the absolute value of U_0 in the QM9 data set can be considered as the accumulated value of D_{AB} in eq 1 for all atoms in the given molecule. Here, i and j denote the indices of atoms in the molecule. The bond energies of homodimers, D_{ii} and D_{jj}, were taken from Pauling's work:[1] 4.44 eV for H, 3.60 eV for C, 1.44 eV for N, 1.49 eV for O, and 2.80 eV for F. Using eq 4, we can generate χ_{ML} that closely matched those of eq 1. For the proof of concept, we generated χ_{ML} using linear functions, since they are simple and more readily interpretable than complex nonlinear functions, and also showed remarkable performance.[58,59] We also checked the nonlinear transformation of one-hot encoding vectors. However, we did not observe significant advantages of nonlinear functions over the linear function in the electronegativity generation layer. Therefore, we used the linear function to transform one-hot encoding vectors into electronegativity. We first distinguished atoms using one-hot encoding. For example, when five elements in the data set were categorized by atomic number, hydrogen and carbon were represented as [1, 0, 0, 0, 0] and [0, 1, 0, 0, 0], respectively. Then, the one-hot encoding vectors were converted to χ_{ML} through eq 5.

$$\chi_{ML} = Wx + b \tag{5}$$

where $W \in \mathbb{R}^{n\times m}$ is the weight matrix of the linear function that transforms a m-dimensional one-hot encoding vector into a n-dimensional vector, χ_{ML}, $b \in \mathbb{R}^n$ is the bias vector that shifts the transformed result along each axis, and x is one-hot encoding vector for each atom. Then, we used χ_{ML} as input variables for GCNs. Our goal was to identify new atomic properties that most accurately predict the target property, y defined on the right-hand side of eq 4,

$$\arg\min_{\chi_{ML}} \mathcal{L}(\chi_{ML}) = \arg\min_{\chi_{ML}} \frac{1}{N}\sum_{i=1}^{N} (y_i - f(\{\chi_{ML}\}))^2 \tag{6}$$

where N is the number of data points and f is the GCN model for the prediction. The loss function, denoted as $\mathcal{L}$, is the mean squared error (MSE) in this work. Although χ_{ML} could be generated in any dimension, two-dimensional vectors were used for simplicity in this work. For each case, the generation of χ_{ML} was repeated ten times. We examined three versions of one-hot encoding vectors containing different types of atomic information, as detailed in Table 1. Atomic information such as the atomic number, hybridization, and aromaticity was used to generate a one-hot vector for each atom. These vectors were then used to classify atom types within the data set. Of course, other atomic characteristics could be used to provide a more detailed description of the atom types. Additional atomic information can be found in Supporting Information (Figure S1).

Table 1. Atomic Information for the Generation of One-Hot Encoding Vectors and the Corresponding χ_{ML}

Symbol	Atomic information
χ_{ML}^{AN}	Atomic number
χ_{ML}^{AN+HY}	Atomic number, Hybridization
$\chi_{ML}^{AN+HY+AR}$	Atomic number, Hybridization, Aromaticity

Hyperparameter Optimization and Model Evaluation. A total of 133,885 molecules in QM9 data set were used in this work.[46] For computational efficiency, hyperparameter optimization of the GCN[56] models was carried out using 10% of the data set and 5-fold cross-validation as implemented in the scikit-learn package.[67] The list of hyperparameters and their optimal values used in this work is summarized in Tables S1 and S2. Then, we generated multidimensional electronegativity using the entire QM9 data set. The Adam optimizer was employed to minimize the mean square error (MSE). The coefficient of determination (R^2 score) was used as the evaluation metric during the hyperparameter optimization process. MSE was used as the loss function, and root mean squared error (RMSE) was used as model performance metrics. They are defined as

$$R^2 \text{ score} = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{7}$$

$$\text{MSE} = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n} \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \tag{9}$$

where $\hat{y}$, y_i, and $\bar{y}$ are predicted, true, and average values, respectively, and n is the number of samples.

https://doi.org/10.1021/acsomega.5c06731
ACS Omega 2025, 10, 48740−48749

RESULTS AND DISCUSSION

Analysis of χ_{ML}. First, we analyzed two-dimensional χ_{ML} as shown in Figures 2 and S2−S4. We found that χ_{ML} offered a

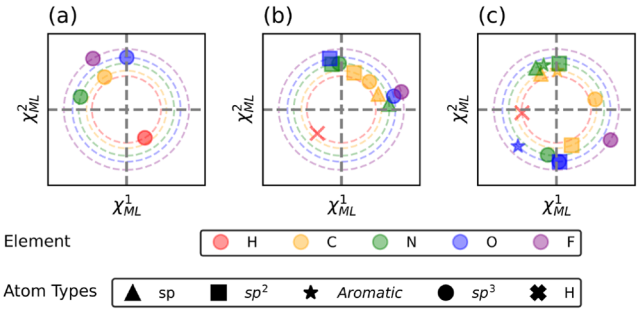


Figure 2. Scattering patterns of two-dimensional (a)χ_{ML}^{AN}, (b)χ_{ML}^{AN+HY}, and (c)$\chi_{ML}^{AN+HY+AR}$, where $\chi_{ML} = (\chi^1_{ML}, \chi^2_{ML})$. Thus, χ^1_{ML} and χ^2_{ML} represent each component of the two-dimensional electronegativity. The $\chi_{Pauling}$ values were represented by concentric circles centered at the origin. Atom types were distinguished by symbols and color, and hydrogen was marked with an "X" symbol in parts (b) and (c).

more diverse perspective on atomic properties than one-dimensional $\chi_{Pauling}$. In Figure 2, $\chi_{Pauling}$ was represented by dotted lines centered at the origin, while χ_{ML} was shown with the symbols and colors as explained in the legend below. Upon analyzing these results, we observed that these patterns are different from conventional perceptions. In Figure 2b, for example, sp^3 oxygen (blue circle) appeared to be more similar to sp nitrogen (green triangle) than to sp^2 oxygen (blue square). In Figure 2c, we observed a range of distributions for χ_{ML} of carbons (yellow). For example, the sp^2 and aromatic carbon atoms (indicated by the yellow square and star, respectively) were positioned at opposite ends. In contrast, the sp^2 carbon and sp^2 oxygen were located in close proximity to each other. These observations revealed that ML distinguished atom types in the 2D space, often in contrast with our intuition. In all cases in Figure 2, hydrogen was largely different from other heavy atoms. Through these, we confirmed that ML can distinguish atoms with concepts beyond the conventional understanding from one-dimensional $\chi_{Pauling}$. Generation of χ_{ML} involved multidimensional optimizations through the generation of random initial weights of ML models. Thus, generalization of our findings should be discussed with caution. To this end, we statistically analyzed χ_{ML} using a clustering algorithm.

By using the K-means algorithm,[67,68] we set the number of clusters based on the distribution pattern of χ_{ML} shown in Figures 2 and S2−S4. In Figure 3, hydrogen is clearly isolated from all other atom types. We also found that strong correlations among sp atoms and among sp^2 atoms in χ_{ML}^{AN+HY} (Figure 3b), as well as between aromatic nitrogen and oxygen, and between sp^2 nitrogen and oxygen in $\chi_{ML}^{AN+HY+AR}$ (Figure 3c). These correlations were not readily rationalized by classical chemical intuition. The atom type classifications inferred from χ_{ML} differ from those defined in $\chi_{Pauling}$, suggesting a reorganization of the chemical space that may not align with conventional classification schemes. However, the ML-based approach comes with several limitations. In particular, the optimization could become trapped in local minima or saddle points, which led to

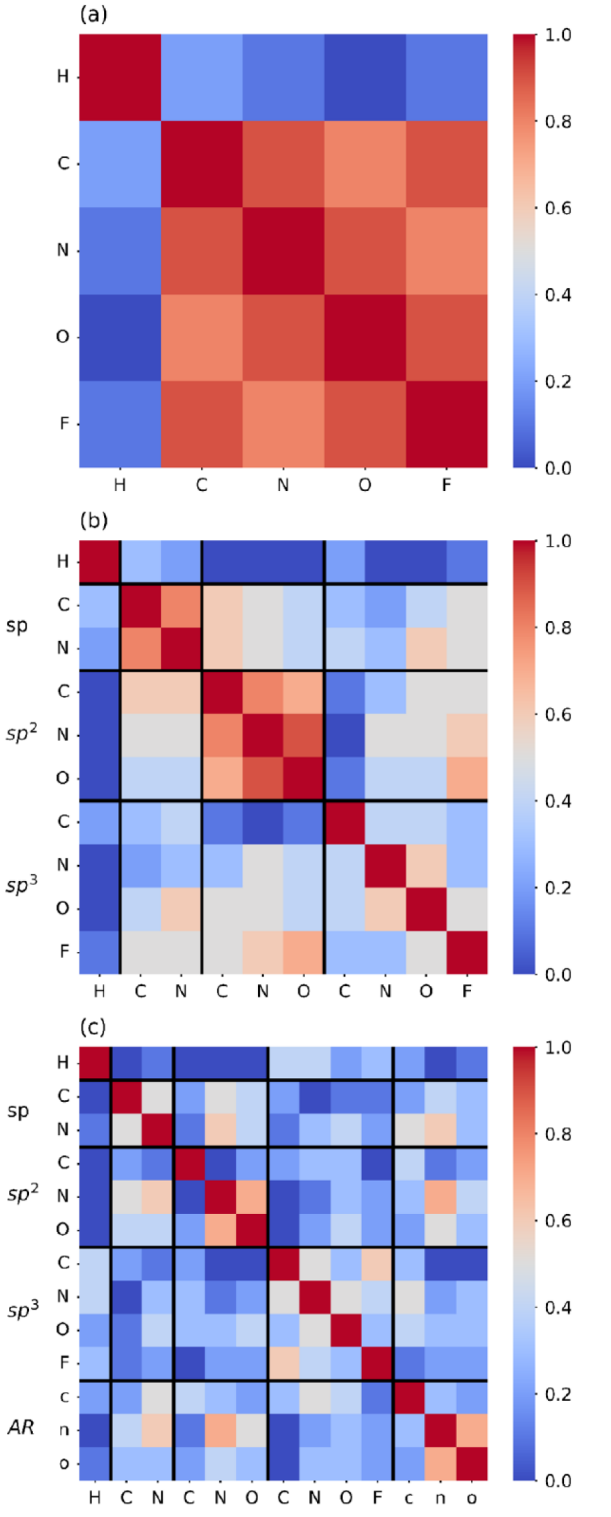


Figure 3. Probability of each atom type in (a)χ_{ML}^{AN}, (b)χ_{ML}^{AN+HY}, and (c)$\chi_{ML}^{AN+HY+AR}$ belonging to the same cluster. For each case, χ_{ML} was generated ten times. In (c), the upper (lower) case atom symbols were for aliphatic (aromatic) atoms.

premature convergence, an intrinsic issue that remains unresolved in the present study.

Furthermore, we examined the correlation between χ_{ML} and traditional atomic properties to study the physical meaning of χ_{ML} as shown in Figure S5. We found that the two components

https://doi.org/10.1021/acsomega.5c06731
ACS Omega 2025, 10, 48740−48749

of χ_{ML} showed distinct correlations with atomic properties. For example, when atom types were defined using only atomic number, one component showed a positive correlation, while the other exhibited a negative correlation with specific atomic properties. When hybridization was also considered, the components showed a similar correlation pattern only with electron affinity. Additionally, when aromaticity was included, this trend was observed for both $\chi_{Pauling}$ and electron affinity. These results suggest that if one component is proportional to a physically meaningful atomic property, then the other shows the opposite behavior for the same property. Interestingly, this emerged without any modification to the objective function during the electronegativity generation process. Thus, by refining the objective functions to provide a more physical interpretation, one can improve the physical meaning of multidimensional electronegativity. In addition, it is important to note that, unlike the electronegativities considered previously,[11−13] our framework lacks theoretical transparency due to the blackbox nature of ML models. However, we believe that our framework produces electronegativities that are at least consistent across several atom types. The strength of our approach lies in its flexibility to generate electronegativities as more molecular data become available in the future.

Analysis of the Chemical Bond Using χ_{ML}. The nature of chemical bonds is typically characterized by the difference in $\chi_{Pauling}$ between the two atoms that form the bond. Using $\chi_{Pauling}$, we can classify chemical bonds as ionic and covalent bonds, and the covalent bonds are further classified as polar and nonpolar. Most organic molecules are composed of hydrogen and carbon atoms with additional main-group elements, such as nitrogen, oxygen, fluorine, sulfur, and phosphorus. For these elements, the values of $\chi_{Pauling}$ are in almost half of the possible range, roughly from 0.8 to 4.0. Therefore, when chemical bonds in organic molecules are examined, almost all bonds are treated as covalent bonds or as polar and nonpolar covalent bonds. As we generated multidimensional electronegativity, χ_{ML} from the perspective of AI, we postulated that new chemical insights about chemical bonds can be revealed. To explore this, we examined how AI interpreted chemical bonds using the multidimensional electronegativity, χ_{ML}, along with the K-means clustering algorithm (Figures 4 and S6). The multidimensional χ_{ML} generated in this study could provide a more comprehensive description of electronegativity differences and not just the difference between two values. By considering two-dimensional χ_{ML} as a vector in a two-dimensional space, we could describe the difference between χ_{ML} values using not only the distance but also the angle between the χ_{ML} vectors. In this study, with χ_{ML}, we described the difference between the two electronegativity values using the Euclidean distance and the angular difference as shown in Figure 4. We analyzed $\chi_{ML}^{AN+HY+AR}$ for carbon−carbon bonds (Figure 4b,c). Using conventional $\chi_{Pauling}$, all carbon−carbon bonds were considered nonpolar covalent bonds. However, with $\chi_{ML}^{AN+HY+AR}$, the different types of carbon−carbon bonds could be distinguished by a more detailed classification. The $\chi_{ML}^{AN+HY+AR}$ was generated based on four types of carbon atoms: *sp*, sp^2, sp^3, and aromatic carbons (*c*). Therefore, there are a total of six possible carbon−carbon bond types after bonds between identical carbon types. These bonds would be expected to have similar $\Delta\chi_{ML}^{AN+HY+AR}$ values, with small differences in both angle and distance.

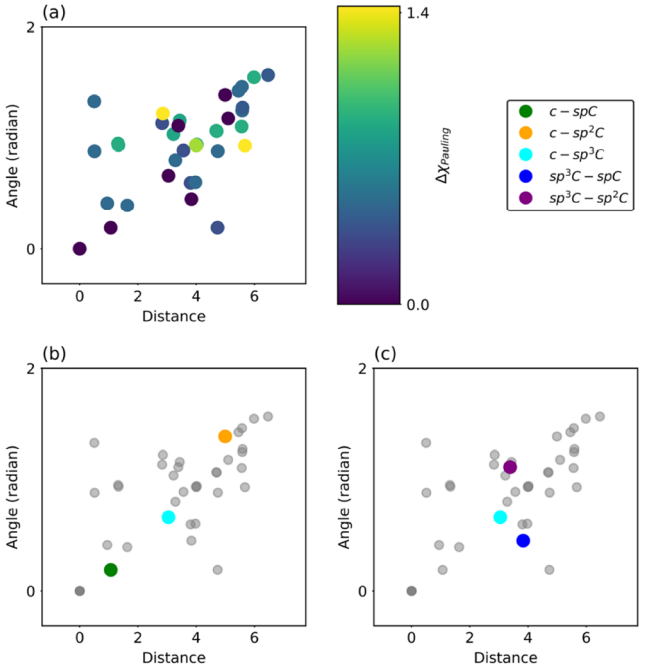


Figure 4. Clustering analysis of $\Delta\chi_{ML}$ using two-dimensional $\chi_{ML}^{AN+HY+AR}$. Each point corresponds to a chemical bond where *x*-axis and *y*-axis represent the Euclidean distance and angle between χ_{ML} values of two atoms, respectively. (a) $\Delta\chi_{Pauling}$ between atoms was shown with different colors while (b) three aromatic carbon-containing bonds and (c) three sp^3 carbon-including bonds excluding bonds with identical atom types were highlighted.

However, the results were completely different. For example, the aromatic carbon-containing bonds and the sp^3 carbon-containing bonds showed noticeable patterns as shown in Figure 4b,c. These results, which were beyond our intuition, should be considered with caution due to the intrinsic stochastic nature of ML. Therefore, to verify the existence of a common pattern, we generated $\chi_{ML}^{AN+HY+AR}$ ten times with different initial weights of the neural networks. We then applied the K-means algorithm to identify hidden clustering patterns, as shown in Figure 5.

With the two-dimensional χ_{ML}, we analyzed the clustering patterns of chemical bonds and explored how AI recognized these bonds as shown in Figures 5 and S7,S8. Among the notable patterns, we analyzed five distinct groups, highlighted with black boxes and red labels in Figure 5. In the first group, carbon−carbon bonds rarely appeared in the same cluster. This suggests that our approach has the potential to distinguish subtle variations even among the carbon−carbon bonds involving diverse carbon hybridizations. In the second group, bonds involving sp^2-hybridized carbon and either sp^2-hybridized nitrogen, oxygen, or aromatic nitrogen formed a distinct and consistently observed cluster. The third and fourth groups exhibited a similar clustering pattern involving sp^3-hybridized and aromatic carbon. This is consistent with Figure 3c as sp^2 nitrogen, oxygen, and aromatic nitrogen were similarly located. The fifth group, bonds between identical atom types ($\Delta\chi_{ML} = 0$) were assigned to a distinct cluster. Using χ_{ML}, we could generate a more detailed chemical bond classification that was not observed with conventional electronegativity. To evaluate its effect on molecular ML, we

https://doi.org/10.1021/acsomega.5c06731
ACS Omega 2025, 10, 48740−48749

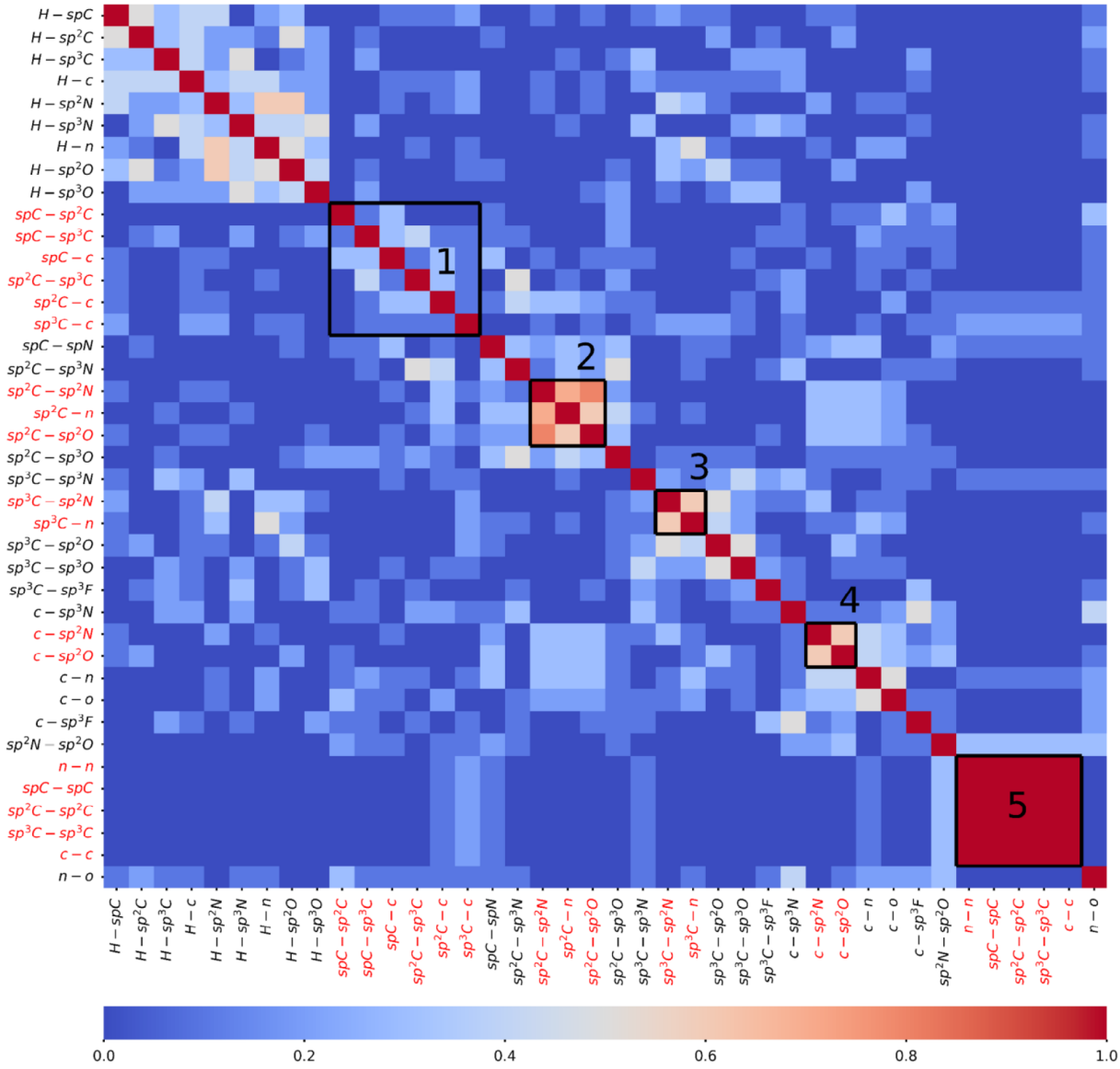


Figure 5. Probability of each bond ($\Delta\chi^{AN+HY+AR}$) belonging to the same cluster. Probabilities were represented by a color scale with red and blue indicating higher and lower probabilities, respectively. Five groups of remarkable patterns were identified, distinguished by black boxes in the figure and red text on the axes.

used a variant of the GCN model, RGCN,[62] which is capable of representing such detailed bond classification as edge types.

Prediction of Molecular Properties. All edges or chemical bonds in a molecular graph are treated equally in the GCN model. For example, in the QM9 data set, the bonds between five elements (C, H, O, N, and F) are regarded as the same type of bond. This is conceptually similar to using conventional electronegativity $\chi_{Pauling}$ for bond classification. All bonds are treated as a single bond type, a covalent bond. With the conventional GCN,[56] atomic feature vectors are updated via eq 2 in each convolution layer, where all bonds are treated the same. In contrast, RGCN treats heterogeneous relationships between nodes in the graph by assigning different weights to each edge.[62] In the RGCN model used in this work, the chemical bonds within a molecule were categorized into several types, and each type of bond had different weights. Thus, a more flexible representation of RGCN can handle the complex relationships in chemical systems[69,70] and can be written as eq 3. In this study, we showed that χ_{ML} provided a more detailed and novel classification of both atoms and bonds within molecules from the perspective of AI. To assess the effectiveness of χ_{ML} in molecular ML, we compared several models: the RGCN model using two-dimensional χ_{ML} alone and the RGCN model using χ_{ML} and conventional atomic properties, X, and the GCN model using X and $\chi_{Pauling}$. These models were evaluated on four properties from the QM9 data set: (a) polarizability (α), (b) electronic spatial extent (r^2), (c) HOMO−LUMO gap, and (d) dipole moment (μ) as shown in Figure 6. As shown in Figure 1, molecular descriptors were also used in all cases.

The two-dimensional χ_{ML} used to predict molecular properties is tabulated in Table 2, and a total of 19 conventional atomic properties from Mendeleev and RDKit package,[71,72] X, are tabulated in Table S3. The complete list of χ_{ML} is available in our GitHub repository (https://github.com/hwk-grp/en_multi_mol).

We chose the GCN model with $\chi_{Pauling}$ and X as the baseline, which is the most commonly used feature setting in typical

https://doi.org/10.1021/acsomega.5c06731
ACS Omega 2025, 10, 48740−48749

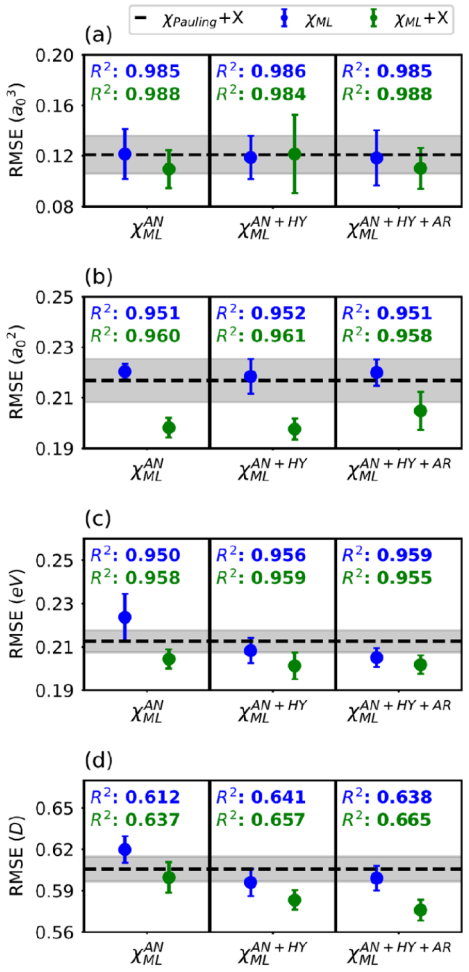


Figure 6. Comparison of prediction performance of RGCN with χ_{ML} (blue), RGCN with χ_{ML} + X (green) and GCN with $\chi_{Pauling}$ + X (black). R^2 score is shown at the top. From the baseline model, R^2 scores were (a) 0.985, (b) 0.953, (c) 0.955, and (d) 0.630.

molecular property predictions with GCN models. With these atomic and molecular features, we varied the number of clusters of $\Delta\chi_{ML}$ and checked their effects on the performance of the RGCN models, as shown in Figure S9. Considering the convergence of the RGCN model errors, we selected the optimal number of clusters, which is about one-third of the total number of bond types found in the data set. In the case of χ_{ML}^{AN}, where atoms and bonds were classified only by atomic number, performance improvement over the baseline (black dotted line) was observed only in limited cases. However, using additional information such as hybridization (χ_{ML}^{AN+HY}) and aromaticity ($\chi_{ML}^{AN+HY+AR}$) noticeably improved prediction performance. With conventional electronegativity such as $\chi_{Pauling}$, it is difficult to capture the detailed classification of chemical bonds within organic molecules, limiting application of molecular ML using RGCN. By contrast, χ_{ML} enables a more detailed classification of chemical bonds within organic molecules and can enhance molecular representations in ML models such as RGCN. These results show the practical utility of the χ_{ML} in molecular ML. While ML models with χ_{ML} performed comparably to those with $\chi_{Pauling}$ in predicting polarizability (α, Figure 6a), they showed enhanced performance for three other molecular properties: electronic spatial extent (r^2, Figure 6b), highest occupied molecular orbital-lowest unoccupied molecular orbital gap (HOMO−LUMO gap, Figure 6c), and dipole moment (μ, Figure 6d). This trend is consistent with the fundamental concept of electronegativity, which describes how electron distributions change due to the formation of chemical bonds. Three properties, r^2, HOMO−LUMO gap, and μ that showed improved performance, can be regarded as molecular properties, that respond to variations in electron distribution within a molecule. In contrast, α is more strongly correlated to molecular size. In this case, we hypothesized that molecular descriptors could play an important role. Overall, the observed improvements in prediction performance suggest that further refinement and systematic construction of χ_{ML} could lead to deeper insight into chemical bonds from the perspective of AI. In our view, this could be a promising direction for future research. We believe that this approach may offer chemists a novel framework for encoding atomic-, bond-, and molecular-level features in molecular ML.

Table 2. Two-dimensional χ_{ML} used in molecular property prediction.

Symbol	Element	χ_{ML}^{1}	χ_{ML}^{2}	Norm
χ_{ML}^{AN}	H	1.171	-1.862	2.20
	C	-1.390	2.138	2.55
	N	-2.922	0.839	3.04
	O	0.014	3.440	3.44
	F	-2.131	3.361	3.98
χ_{ML}^{AN+HY}	H	-1.539	-1.572	2.20
	C(sp)	2.338	1.018	2.55
	C(sp^2)	0.803	2.420	2.55
	C(sp^3)	1.774	1.832	2.55
	N(sp)	3.017	0.371	3.04
	N(sp^2)	-0.600	2.980	3.04
	N(sp^3)	-0.151	3.036	3.04
	O(sp^2)	-0.753	3.356	3.44
	O(sp^3)	3.325	0.882	3.44
	F(sp^3)	3.807	1.162	3.98
$\chi_{ML}^{AN+HY+AR}$	H	-2.189	-0.218	2.20
	C(sp)	-0.993	2.349	2.55
	C(sp^2)	0.960	-2.363	2.55
	C(sp^3)	2.459	0.675	2.55
	c	0.054	2.549	2.55
	N(sp)	-1.315	2.741	3.04
	N(sp^2)	0.173	3.035	3.04
	N(sp^3)	-0.550	-2.990	3.04
	n	-0.812	2.929	3.04
	O(sp^2)	0.195	-3.434	3.44
	O(sp^3)	0.172	-3.436	3.44
	o	-2.446	-2.419	3.44
	F(sp^3)	3.436	-2.008	3.98

CONCLUSION

In summary, we revisited the Pauling electronegativity that was proposed about a century ago by using a data set of organic molecules and graph-based machine learning (ML) models. The newly proposed electronegativity, χ_{ML} enabled detailed analysis of chemical bonds through the classification of atoms and their chemical bonds. We also found that χ_{ML} showed

https://doi.org/10.1021/acsomega.5c06731
ACS Omega 2025, 10, 48740−48749

better performance when it was used as an input variable in molecular ML. These findings suggest that ML will be effectively employed to reconsider other data-driven chemical concepts in chemistry,[63] and these ML-guided concepts can be utilized in chemical applications such as the design of new organic molecules for drug discovery and retrosynthesis.[73−75] Of course, the current perspective would benefit from further elaboration with larger chemical data sets and more advanced ML models in the future. While we focused on two-dimensional electronegativity for its simplicity and interpretability, it is important to note that the proposed approach can be used to generate electronegativity or other atomic properties in a higher-dimensional space.[76,77] Furthermore, with the available data set of inorganic materials,[78,79] we can generate multidimensional electronegativity for most of the elements in the periodic table, and this is in progress. We believe that this research will provide new insights into the understanding of chemical bonds.

■ ASSOCIATED CONTENT

Data Availability Statement

The code and data used in this work are deposited on GitHub (https://github.com/hwk-grp/en_multi_mol). All the parameter, input, and output files necessary to reproduce the key results are included.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.5c06731.

List of hyperparameters, optimal hyperparameter values, atomic properties used as atomic feature vectors, scattering patterns of two-dimensional electronegativity, Pearson correlation analysis, probability analysis of chemical bonds, and effect of the number of chemical bond clusters on prediction error (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Hyun Woo Kim − *Department of Chemistry, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, Republic of Korea; Center for Quantum Technology, Korea Institute of Science and Technology (KIST), Seoul 02792, Republic of Korea*; Email: hwk@gist.ac.kr

Authors

Da Bean Han − *Department of Chemistry, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, Republic of Korea*

Gyoung S. Na − *Digital Chemical Research Center, Korea Research Institute of Chemical Technology (KRICT), Daejeon 34114, Republic of Korea*; orcid.org/0000-0001-9803-0782

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.5c06731

Author Contributions

All authors contributed to the conceptualization of the study, methodology, and software development. Data curation was performed by H.W.K. and D.B.H. Formal analysis was carried out by D.B.H. Funding acquisition was managed by H.W.K. Project administration and supervision were undertaken by H.W.K. Validation was performed by H.W.K. and D.B.H. Visualization was performed by D.B.H. The original draft was written by D.B.H. with support from H.W.K., and all authors contributed to the review and editing of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (NRF-2022R1F1A107620912 and RS-2024-00411137). This work was also supported by Korea Institute of Science and Technology (KIST) under grand numbers of 2E32971-24-P025 and 2E33571-25-P003.

■ REFERENCES

(1) Pauling, L. The nature of the chemical bond. IV. The energy of single bonds and the relative electronegativity of atoms. *J. Am. Chem. Soc.* **1932**, *54*, 3570−3582.

(2) Pauling, L. *The nature of the chemical bond and the structure of molecules and crystals: an introduction to modern structural chemistry*; Cornell university press, 1960; Vol. *18*.

(3) Aldossary, A.; Head-Gordon, M. Non-iterative method for constructing valence antibonding molecular orbitals and a molecule-adapted minimum basis. *J. Chem. Phys.* **2022**, *157* (9), 0095443.

(4) Shaik, S.; Danovich, D.; Hiberty, P. C. On the nature of the chemical bond in valence bond theory. *J. Chem. Phys.* **2022**, *157* (9), 0095953.

(5) Schön, C.-F.; Bergerem, S. V.; Mattes, C.; Yadav, A.; Grohe, M.; Kobbelt, L.; Wuttig, M. Classification of properties and their relation to chemical bonding: Essential steps toward the inverse design of functional materials. *Sci. Adv.* **2022**, *8* (47), No. eade0828.

(6) Huggins, M. L. Bond energies and polarities1. *J. Am. Chem. Soc.* **1953**, *75*, 4123−4126.

(7) Sanderson, R. An interpretation of bond lengths and a classification of bonds. *Science* **1951**, *114*, 670−672.

(8) Sanderson, R. An explanation of chemical variations within periodic major groups. *J. Am. Chem. Soc.* **1952**, *74*, 4792−4794.

(9) Allred, A.; Rochow, E. Electronegativities of carbon, silicon, germanium, tin and lead. *J. Inorg. Nucl. Chem.* **1958**, *5*, 269−288.

(10) Allred, A. L. Electronegativity values from thermochemical data. *J. Inorg. Nucl. Chem.* **1961**, *17*, 215−221.

(11) Mulliken, R. S. A new electroaffinity scale; together with data on valence states and on valence ionization potentials and electron affinities. *J. Chem. Phys.* **1934**, *2*, 782−793.

(12) Mulliken, R. S. Electronic structures of molecules XI. Electroaffinity, molecular orbitals and dipole moments. *J. Chem. Phys.* **1935**, *3*, 573−585.

(13) Bratsch, S. G. Revised Mulliken electronegativities: I. Calculation and conversion to Pauling units. *J. Chem. Educ.* **1988**, *65*, 34.

(14) Iczkowski, R. P.; Margrave, J. L. Electronegativity. *J. Am. Chem. Soc.* **1961**, *83*, 3547−3551.

(15) Gordy, W. A new method of determining electronegativity from other atomic properties. *Phys. Rev.* **1946**, *69*, 604.

(16) Ghosh, D. C.; Chakraborty, T. Gordy's electrostatic scale of electronegativity revisited. *J. Mol. Struct.: THEOCHEM* **2009**, *906*, 87−93.

(17) Ghosh, D. C.; Chakraborty, T.; Mandal, B. The electronegativity scale of Allred and Rochow: revisited. *Theor. Chem. Acc.* **2009**, *124*, 295−301.

(18) Hinze, J.; Whitehead, M.; Jaffe, H. H. Electronegativity II. Bond and orbital electronegativities. *J. Am. Chem. Soc.* **1963**, *85*, 148−154.

(19) Allen, L. C. Electronegativity is the average one-electron energy of the valence-shell electrons in ground-state free atoms. *J. Am. Chem. Soc.* **1989**, *111*, 9003−9014.

(20) Politzer, P.; Murray, J. S. Electronegativity: A continuing enigma. *J. Phys. Org. Chem.* **2023**, *36*, No. e4406.

(21) Yoon, D.; Jang, J.-H.; Choi, B. J.; Kim, T. Y.; Han, C. H. Discovering hidden information in biosignals from patients using artificial intelligence. *Korean J. Anesthesiol.* **2020**, *73*, 275−284.
(22) Roostaee, M.; Meidanshahi, R. Hidden Pattern Discovery on Clinical Data: an Approach based on Data Mining Techniques. *J. Artif. Intell. Data Min.* **2023**, *11*, 343−355.
(23) Abiodun, O. I.; Jantan, A.; Omolara, A. E.; Dada, K. V.; Umar, A. M.; Linus, O. U.; Arshad, H.; Kazaure, A. A.; Gana, U.; Kiru, M. U. Comprehensive review of artificial neural network applications to pattern recognition. *IEEE Access* **2019**, *7*, 158820−158846.
(24) Wu, Y.; Wang, G. Machine learning based toxicity prediction: from chemical structural description to transcriptome analysis. *Int. J. Mol. Sci.* **2018**, *19*, 2358.
(25) Panteleev, J.; Gao, H.; Jia, L. Recent applications of machine learning in medicinal chemistry. *Bioorg. Med. Chem. Lett.* **2018**, *28*, 2807−2815.
(26) Lodaya, K.; Ricke, N. D.; Chen, K.; Van Voorhis, T. Machine Learning Identification of Active Sites in Graphite-Conjugated Catalysts. *J. Phys. Chem. C* **2023**, *127*, 2303−2313.
(27) Sigmund, L. M.; Shree, S. S. V.; Albers, A.; Erdmann, P.; Paton, R. S.; Greb, L. Predicting Lewis Acidity: Machine Learning the Fluoride Ion Affinity of p-Block-Atom-Based Molecules. *Angew. Chem., Int. Ed.* **2024**, *63*, No. e202401084.
(28) Liu, Y.; Li, Z. Predict ionization energy of molecules using conventional and graph-based machine learning models. *J. Chem. Inf. Model.* **2023**, *63*, 806−814.
(29) Meuwly, M. Machine learning for chemical reactions. *Chem. Rev.* **2021**, *121*, 10218−10239.
(30) Joung, J. F.; Fong, M. H.; Roh, J.; Tu, Z.; Bradshaw, J.; Coley, C. W. Reproducing Reaction Mechanisms with Machine-Learning Models Trained on a Large-Scale Mechanistic Dataset. *Angew. Chem., Int. Ed.* **2024**, *63*, No. e202411296.
(31) Zheng, Z.; Florit, F.; Jin, B.; Wu, H.; Li, S.-C.; Nandiwale, K. Y.; Salazar, C. A.; Mustakis, J. G.; Green, W. H.; Jensen, K. F. Integrating machine learning and large language models to advance exploration of electrochemical reactions. *Angew. Chem., Int. Ed.* **2024**, *64*, No. e202418074.
(32) Adebar, N.; Keupp, J.; Emenike, V. N.; Kühlborn, J.; Vom Dahl, L.; Möckel, R.; Smiatek, J. Scientific deep machine learning concepts for the prediction of concentration profiles and chemical reaction kinetics: Consideration of reaction conditions. *J. Phys. Chem. A* **2024**, *128*, 929−944.
(33) Guan, Y.; Lee, T.; Wang, K.; Yu, S.; McWilliams, J. C. SNAr regioselectivity predictions: machine learning triggering DFT reaction modeling through statistical threshold. *J. Chem. Inf. Model.* **2023**, *63*, 3751−3760.
(34) Terrones, G. G.; Huang, S.-P.; Rivera, M. P.; Yue, S.; Hernandez, A.; Kulik, H. J. Metal−organic framework stability in water and harsh environments from data-driven models trained on the diverse ws24 data set. *J. Am. Chem. Soc.* **2024**, *146*, 20333−20348.
(35) Zhang, L.; Zhang, X.; Chen, C.; Zhang, J.; Tan, W.; Xu, Z.; Zhong, Z.; Du, L.; Song, H.; Liao, S.; et al. Machine Learning-Aided Discovery of Low-Pt High Entropy Intermetallic Compounds for Electrochemical Oxygen Reduction Reaction. *Angew. Chem.* **2024**, *136*, No. e202411123.
(36) Gebauer, N. W. A.; Gastegger, M.; Hessmann, S. S. P.; Müller, K.-R.; Schütt, K. T. Inverse design of 3d molecular structures with conditional generative neural networks. *Nat. Commun.* **2022**, *13*, 973.
(37) Keith, J. A.; Vassilev-Galindo, V.; Cheng, B.; Chmiela, S.; Gastegger, M.; Müller, K.-R.; Tkatchenko, A. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chem. Rev.* **2021**, *121*, 9816−9872.
(38) Schwaller, P.; Vaucher, A. C.; Laplaza, R.; Bunne, C.; Krause, A.; Corminboeuf, C.; Laino, T. Machine intelligence for chemical reaction space. *Wiley Interdiscip. Rev.: comput. Mol. Sci.* **2022**, *12*, No. e1604.
(39) Damewood, J.; Karaguesian, J.; Lunger, J. R.; Tan, A. R.; Xie, M.; Peng, J.; Gómez-Bombarelli, R. Representations of materials for machine learning. *Annu. Rev. Mater. Res.* **2023**, *53*, 399−426.
(40) Ding, Y.; Qiang, B.; Chen, Q.; Liu, Y.; Zhang, L.; Liu, Z. Exploring chemical reaction space with machine learning models: Representation and feature perspective. *J. Chem. Inf. Model.* **2024**, *64*, 2955−2970.
(41) Back, S.; Aspuru-Guzik, A.; Ceriotti, M.; Gryn'ova, G.; Grzybowski, B. A.; Gu, G. H.; Hein, J.; Hippalgaonkar, K.; Hormázabal, R.; Jung, Y.; et al. Accelerated chemical science with AI. *Digit. Discovery* **2024**, *3*, 23−33.
(42) Racki, A.; Paduszynski, K. Recent advances in the modeling of ionic liquids using artificial neural networks. *J. Chem. Inf. Model.* **2025**, *65*, 3161−3175.
(43) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; et al. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* **2019**, *47*, D1102−D1109.
(44) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47*, D930−D940.
(45) Ruddigkeit, L.; Van Deursen, R.; Blum, L. C.; Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864−2875.
(46) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; Von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **2014**, *1* (1), 140022.
(47) Arshadi, A. K.; Salem, M.; Firouzbakht, A.; Yuan, J. S. MolData, a molecular benchmark for disease and target based machine learning. *J. Cheminf.* **2022**, *14*, 10.
(48) Sterling, T.; Irwin, J. J. ZINC 15−ligand discovery for everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324−2337.
(49) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.* **2018**, *9*, 513−530.
(50) Wigh, D. S.; Goodman, J. M.; Lapkin, A. A. A review of molecular representation in the age of machine learning. *Wiley Interdiscip. Rev.: comput. Mol. Sci.* **2022**, *12*, No. e1603.
(51) Harren, T.; Matter, H.; Hessler, G.; Rarey, M.; Grebner, C. Interpretation of structure−activity relationships in real-world drug design data sets using explainable artificial intelligence. *J. Chem. Inf. Model.* **2022**, *62*, 447−462.
(52) Mastrolorito, F.; Gambacorta, N.; Ciriaco, F.; Cutropia, F.; Togo, M. V.; Belgiovine, V.; Tondo, A.; Trisciuzzi, D.; Monaco, A.; Bellotti, R.; et al. Chemical Space Networks Enhance Toxicity Recognition via Graph Embedding. *J. Chem. Inf. Model.* **2025**, *65*, 1850−1861.
(53) Kim, J.; Gu, G. H.; Noh, J.; Kim, S.; Gim, S.; Choi, J.; Jung, Y. Predicting potentially hazardous chemical reactions using an explainable neural network. *Chem. Sci.* **2021**, *12*, 11028−11037.
(54) Probst, D. An explainability framework for deep learning on chemical reactions exemplified by enzyme-catalysed reaction classification. *J. Cheminf.* **2023**, *15*, 113.
(55) Jiménez-Luna, J.; Grisoni, F.; Schneider, G. Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* **2020**, *2*, 573−584.
(56) Kipf, T. N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv*, **2016**, .
(57) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. et al.Pytorch: An imperative style, high-performance deep learning library. *arXiv*, **2019**, .
(58) Song, L.; Langfelder, P.; Horvath, S. Random generalized linear model: a highly accurate and interpretable ensemble predictor. *BMC Bioinf.* **2013**, *14*, 5.
(59) Shimakawa, H.; Kumada, A.; Sato, M. Extrapolative prediction of small-data molecular property using quantum mechanics-assisted machine learning. *Npj Comput. Mater.* **2024**, *10* (1), 11.
(60) Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P. S. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 4−24.

https://doi.org/10.1021/acsomega.5c06731
ACS Omega 2025, 10, 48740−48749

(61) Corso, G.; Stark, H.; Jegelka, S.; Jaakkola, T.; Barzilay, R. Graph neural networks. *Nat. Rev. Methods Primers* **2024**, *4* (1), 17.
(62) Schlichtkrull, M.; Kipf, T. N.; Bloem, P.; van den Berg, R.; Titov, I.; Welling, M. Modeling Relational Data with Graph Convolutional Networks. In *The Sematic Web ESWC 2018*; Springer: Cham, 2018.
(63) Artrith, N.; Butler, K. T.; Coudert, F.-X.; Han, S.; Isayev, O.; Jain, A.; Walsh, A. Best practices in machine learning for chemistry. *Nat. Chem.* **2021**, *13*, 505−508.
(64) Tawfik, S. A.; Russo, S. P. Naturally-meaningful and efficient descriptors: machine learning of material properties based on robust one-shot ab initio descriptors. *J. Cheminf.* **2022**, *14*, 78.
(65) Fey, M.; Lenssen, J. E. Fast graph representation learning with PyTorch Geometric. *arXiv* **2019**.
(66) Elfwing, S.; Uchibe, E.; Doya, K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Netw.* **2018**, *107*, 3−11.
(67) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(68) Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129−137.
(69) Gurniak, E. J.; Yuan, S.; Ren, X.; Branicio, P. S. Harnessing graph convolutional neural networks for identification of glassy states in metallic glasses. *Comput. Mater. Sci.* **2024**, *244*, 113257.
(70) Wu, Z.; Jiang, D.; Hsieh, C.-Y.; Chen, G.; Liao, B.; Cao, D.; Hou, T. Hyperbolic relational graph convolution networks plus: a simple but highly efficient QSAR-modeling method. *Briefings Bioinf.* **2021**, *22* (5), bbab112.
(71) Mentel, L. *mendeleev − A Python resource for properties of chemical elements, ions and isotopes*, https://github.com/lmmentel/mendeleev.
(72) Landrum, G.; Tosco, P.; Kelley, B.; Cosgrove, D.; Vianello, R.; *rdkit/rdkit: 2023_03_2 (Q1 2023) Release*; zenodo, 2023
(73) Zhong, Z.; Song, J.; Feng, Z.; Liu, T.; Jia, L.; Yao, S.; Hou, T.; Song, M. Recent advances in deep learning for retrosynthesis. *Wiley Interdiscip. Rev.: comput. Mol. Sci.* **2024**, *14*, No. e1694.
(74) Ali, R. S. A. E.; Meng, J.; Khan, M. E. I.; Jiang, X. Machine learning advancements in organic synthesis: A focused exploration of artificial intelligence applications in chemistry. *Artif. Intell. Chem.* **2024**, *2*, 100049.
(75) Strieth-Kalthoff, F.; Szymkuc, S.; Molga, K.; Aspuru-Guzik, A.; Glorius, F.; Grzybowski, B. A. Artificial Intelligence for Retrosynthetic Planning Needs Both Data and Expert Knowledge. *J. Am. Chem. Soc.* **2024**, *146*, 11005−11017.
(76) Onwuli, A.; Hegde, A. V.; Nguyen, K. V.; Butler, K. T.; Walsh, A. Element similarity in high-dimensional materials representations. *Digit. Disc.* **2023**, *2*, 1558−1564.
(77) Park, H.; Onwuli, A.; Butler, K. T.; Walsh, A. Mapping inorganic crystal chemical space. *Faraday Discuss* **2025**, *256*, 601−613.
(78) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002.
(79) Barroso-Luque, L.; Shuaibi, M.; Fu, X.; Wood, B. M.; Dzamba, M.; Gao, M.; Rizvi, A.; Zitnick, C. L.; Ulissi, Z. W. Open materials 2024 (omat24) inorganic materials dataset and models. *arXiv*, **2024**, .

https://doi.org/10.1021/acsomega.5c06731
ACS Omega 2025, 10, 48740−48749