

ARTICLE OPEN



Towards understanding structure–property relations in materials with interpretable deep learning

Tien-Sinh Vu¹, Minh-Quyet Ha¹, Duong-Nguyen Nguyen¹, Viet-Cuong Nguyen², Yukihiro Abe², Truyen Tran³, Huan Tran⁴, Hiori Kino⁵, Takashi Miyake⁶, Koji Tsuda⁷ and Hieu-Chi Dam^{1,8}✉

Deep learning (DL) models currently employed in materials research exhibit certain limitations in delivering meaningful information for interpreting predictions and comprehending the relationships between structure and material properties. To address these limitations, we propose an interpretable DL architecture that incorporates the attention mechanism to predict material properties and gain insights into their structure–property relationships. The proposed architecture is evaluated using two well-known datasets (the QM9 and the Materials Project datasets), and three in-house-developed computational materials datasets. Train–test–split validations confirm that the models derived using the proposed DL architecture exhibit strong predictive capabilities, which are comparable to those of current state-of-the-art models. Furthermore, comparative validations, based on first-principles calculations, indicate that the degree of attention of the atoms' local structures to the representation of the material structure is critical when interpreting structure–property relationships with respect to physical properties. These properties encompass molecular orbital energies and the formation energies of crystals. The proposed architecture shows great potential in accelerating material design by predicting material properties and explicitly identifying crucial features within the corresponding structures.

npj Computational Materials (2023)9:215; <https://doi.org/10.1038/s41524-023-01163-9>

INTRODUCTION

A central challenge in the field of materials science involves the use of both experience and theory to explore the compositions and structures of materials with specific properties and subsequently validating them through experimentation. Unfortunately, the research and development of materials is a time-consuming endeavor that often relies on serendipity. Materials informatics (MI) has emerged as a rapidly growing interdisciplinary field toward addressing these challenges. This concept employs data-driven methods to extract practical knowledge regarding materials and their related physicochemical phenomena from experimental and computational data, thus ultimately accelerating the discovery of superior materials^{1–4}.

The majority of MI approaches comprise three key components⁵. The first component comprises datasets containing information regarding the structure of the materials, measurement results directly related to these structures, and physical properties relevant to the material development goals. The second component, that is, representation, quantitatively describes the data instances in the first component, collecting a primitive description of materials for identification and analogic inference. The final component is a system that utilizes machine learning or data mining algorithms (either a single approach or a combination of approaches) to extract knowledge from the materials datasets for specific purposes, such as predicting properties or identifying new material compositions and structures.

Traditionally, materials have been characterized based on their elemental compositions and structures. Researchers have primarily relied on their knowledge and experience, often referred to as

tacit knowledge, to predict certain properties of hypothetical materials with specific compositions and structures. Computational chemistry approaches based on quantum mechanics, particularly density functional theory (DFT) simulations, can be used to theoretically verify the compositions and structures of these materials through in-silico computational experimentation. However, despite providing accurate information on the physical properties of hypothetical materials, computational experiments have certain limitations. For example, the vast number of potential hypothetical materials renders the design of materials with desired physical properties time-consuming and expensive due to the exhaustive calculations required. Moreover, researchers require specialized and detailed knowledge to narrow down the potential compositions and material structures.

Unlike traditional approaches, MI approaches initially involve the conversion of primitive data descriptions into appropriate representations that can be used for mathematical reasoning and inference. In particular, MI systems are given the task of estimating qualitative and quantitative relationships between materials based on this transformed representation, allowing them to uncover potential patterns in the material data^{6–8}. The development of material representation (i.e., the design of material descriptors or methods for learning material representation from data) play a crucial role in MI approaches. This is because the effectiveness of an MI algorithm highly depends on the material representation, as it directly affects the algorithm's performance and facilitates the explanation and interpretation of the inference process and prediction results⁹. Recent advancements in automated experiments and high-performance computers have helped acquire

¹Japan Advanced Institute of Science and Technology, Nomi, Ishikawa, Japan. ²HPC SYSTEMS Inc., Minato, Tokyo, Japan. ³Applied Artificial Intelligence Institute, Deakin University, Geelong, VIC, Australia. ⁴School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA. ⁵Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, Tsukuba, Ibaraki, Japan. ⁶Research Center for Computational Design of Advanced Functional Materials, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, Japan. ⁷Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa-shi, Chiba-ken, Japan. ⁸International Center for Synchrotron Radiation Innovation Smart (SRIS), Tohoku University, Aoba-ku, Sendai, Japan. ✉email: dam@jaist.ac.jp

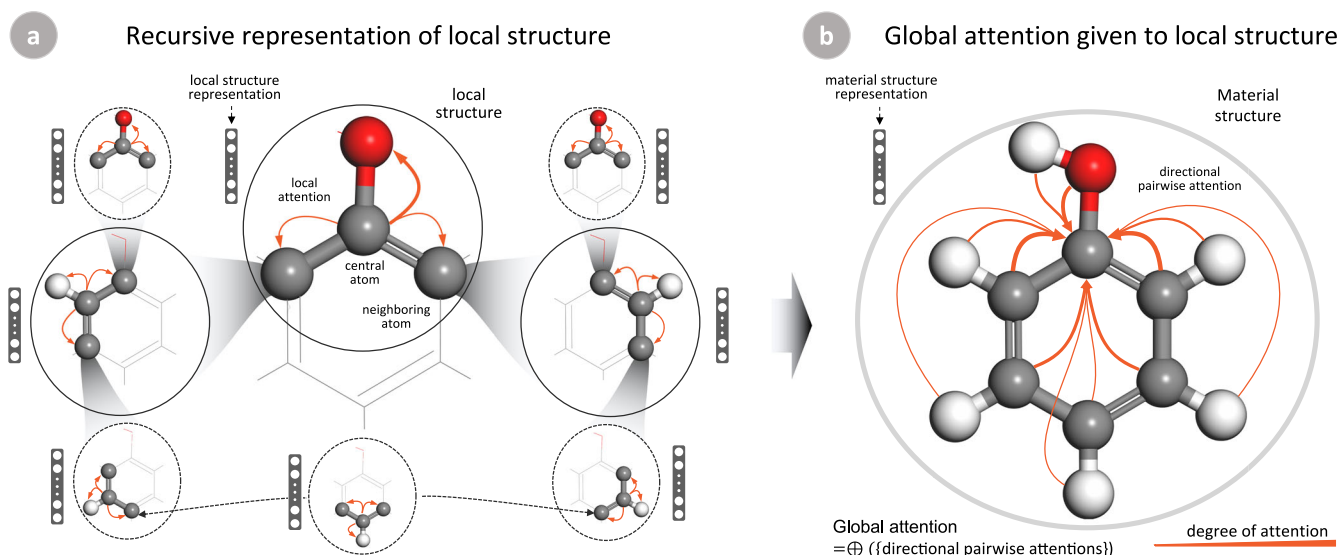


Fig. 1 Illustrations of representations for local structure and material structure. Schematics of (a) the learning recursive representation of a local structure (central atom and its neighboring atoms) within the molecular structure of phenol (C_6H_5OH), and (b) measurement of the global attention given to a local structure when determining representation of the molecular structure. The direction and size of each arrow indicate the degree of attention given to other atoms when establishing the representation of the local structure of a particular atom.

substantial experimental and computational data. Consequently, there is a growing need for the development of explainable and interpretable ML methods to enhance our understanding of physical and chemical phenomena.

Recently, various deep learning (DL)-based ML approaches have been developed to address the challenges associated with material representation and to predict physical properties^{10–13}. A typical example is the DL architecture that uses a continuous-filter convolution layer with filter-generation networks to handle atomistic systems and accurately predict the properties of molecular and crystalline materials¹⁰. Another example is the convolutional neural network based on crystal graphs, which can predict material properties with an accuracy comparable to that of DFT calculations while also providing atomic-level chemical insight¹². In addition to the aforementioned approaches, researchers have developed various other DL architectures to encode the local chemical environments of atoms and improve the prediction accuracy by integrating different types of material descriptors, applying graph neural networks (GNNs), and utilizing many-body tensor representations^{11,13}. Furthermore, several studies have incorporated prior knowledge to construct neural network models that ensure the relationship between structures and properties of materials is learned accurately^{14–16}.

However, a significant challenge confronted by both traditional and DL-based machine learning approaches is the issue of interpretability. Machine learning models often prioritize including all available information rather than selecting an interpretable representation to improve prediction accuracy. The relationship between material representation and its properties is complex and nonlinear, and as a result, machine learning models acting as “black boxes” that do not explicitly reveal correlations. Although statistical evaluations based on existing data often exhibit high prediction accuracies, estimating their predictive capability for new materials is difficult. Gaining a comprehensive understanding machine learning to clarify underlying physicochemical phenomena also remains challenging.

Numerous studies have aimed to enhance model interpretability by incorporating additional information or features. For instance, graph convolutional networks use SMILES strings to represent molecules as inputs, which helps identify crucial fingerprint fragments and facilitating interpretation^{17,18}. Even with

these advancements, these networks still require assistance in accurately predicting the properties of molecular and crystalline materials due to the absence of 3D structural information. Message-passing neural network-based models (MPNNs)^{19–21} employ heuristic bonding information to capture atomic interactions but encounter several challenges with long-range interactions, feature interpretability, global information representation, and scalability when handling large molecule/crystal datasets. With the aim of addressing these limitations, recent works have turned to transformer-based networks^{15,22–28}, which utilize attention mechanisms¹⁶. These networks offer a promising avenue by modeling interatomic reactions between constituent atoms through attention scores, which indicate the significance of each atom in learning the representation of other atoms. Various pooling methods like max or average pooling^{14,15,28–33} are then employed to derive a comprehensive representation of the entire structure. However, extracting meaningful structure–property relationships from these transformer-based networks remains challenging and non-trivial.

In this study, we propose an interpretable DL architecture that incorporates the attention mechanism to predict material structure properties and provide meaningful information about the structure–property relationships. The proposed architecture initiates by learning the representation of local structures of atoms within a material structure through the recursive application of attention mechanisms to the local structures of the neighboring atoms (Fig. 1a). The local structure of an atom includes the atom itself (central atom), its neighboring atoms, and the arrangement of the neighboring atoms around the central atom. Finally, the material structure representation is derived from the representations of these local structures of the atoms. This architecture utilizes the attention mechanism to incorporate information about the geometrical arrangement of neighboring atoms into the representations of local structures. Moreover, it quantitatively measures the degree of attention given to each local structure from a global perspective when determining the representation of the material structure (Fig. 1b). Consequently, through the training of the model with a specific target property, our approach aids in the interpretation of the structure–property relationships in materials.

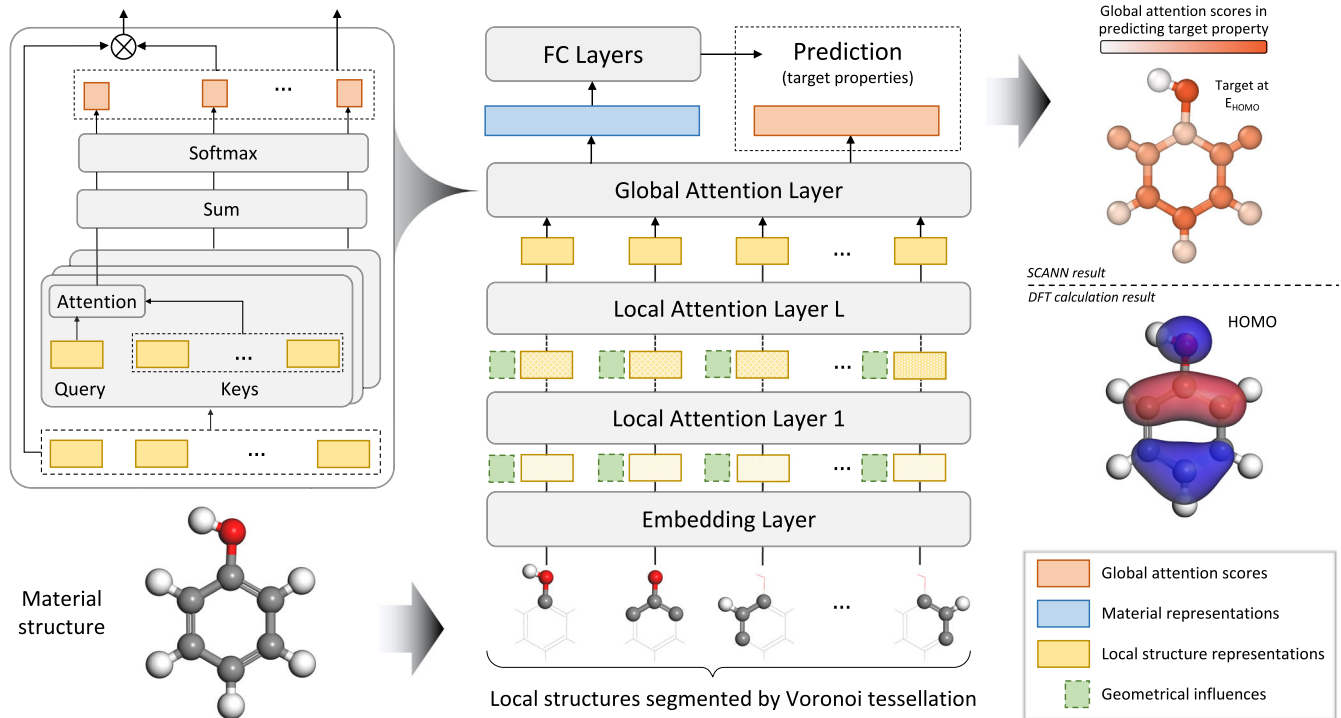


Fig. 2 Overview of proposed SCANN architecture. The SCANN is formed by stacking an embedding layer and local attention layers to learn the representations of various local structures in a material. In the readout stage, a global attention layer is used to assess the attention scores of these local structures. The attention score indicates the degree of attention that should be paid to a local structure to accurately represent the material and predict its physical property. The material representation is linearly combined based on the representations of its local structures with their corresponding attention scores. Fully connected (FC) layers are applied to the material representation to estimate the property of the material.

RESULTS

SCANN framework

We introduce a DL architecture called the self-consistent attention neural network (SCANN). The SCANN focuses on representing material structures from local structures of atoms with learned weights, thus facilitating the prediction and interpretation of material properties. The key objective of SCANN is to recursively learn consistent representations of these local structures within the material (as shown in Fig. 1a), which are then appropriately combined to obtain an overall representation of the material structure.

In this study, each material structure S in a dataset \mathcal{D} is represented using the atomic numbers and the corresponding coordinates of its M atoms. By employing Voronoi tessellation, a set of neighboring atoms \mathcal{N}_i is identified for each atom a_i in the structure S . Then, a vector \mathbf{g}_{ij}^0 is defined as the geometrical influence of a neighboring atom a_j on atom a_i ($1 \leq j \neq i \leq M$) based on the Euclidean distance and Voronoi solid angle between them (Supplementary Section IVA). The use of the Voronoi method here clearly determines the neighboring atoms in a local structure based on material domain knowledge and in alignment with the logical design of our method.

SCANN employs an embedding layer to express the atomic information of each atom a_i in S by an h -dimensional vector \mathbf{c}_i^0 . Hereinafter, we denote the matrix $\mathbf{C}^0 = [\mathbf{c}_i^0]_{1 \leq i \leq M}$ as $[\mathbf{c}_1^0, \mathbf{c}_2^0, \dots, \mathbf{c}_M^0]$. The SCANN architecture comprises of a series of L local attention layers and a global attention layer, each utilizing attention mechanisms¹⁶ to represent the local structures within a material structure and the material structure itself, respectively. The layer-wise design of the local attention layers enables SCANN to iteratively learn and enhance the consistency of local structure representations, thereby providing insights

regarding long-range interactions between these local structures (Supplementary Section IVB). For instance, the representation vector \mathbf{c}_i^{l+1} of the local structure $\{a_i, \mathcal{N}_i\}$ at the $(l+1)$ th local attention layer can be derived as follows:

$$\begin{aligned} \mathbf{c}_i^{l+1} &= \text{LocalAttention}^{l+1}(\mathbf{c}_i^l, \mathbf{C}_{\mathcal{N}_i}^l \times \mathbf{G}_{\mathcal{N}_i}^l) \\ &= \text{Attention}(\mathbf{q}_i^l, \mathbf{K}_{\mathcal{N}_i}^l) + \mathbf{q}_i^l \\ &= \text{softmax}(\mathbf{q}_i^{lT} \mathbf{K}_{\mathcal{N}_i}^l) \mathbf{K}_{\mathcal{N}_i}^l + \mathbf{q}_i^l, \end{aligned} \quad (1)$$

where \mathbf{c}_i^l is the central atom at layer l th, $\mathbf{C}_{\mathcal{N}_i}^l = [\mathbf{c}_j^l]_{a_j \in \mathcal{N}_i}$ denotes its neighboring local structures, and the geometrical influence of the neighboring atoms $\mathbf{G}_{\mathcal{N}_i}^l = [\mathbf{g}_{ij}^l]_{a_j \in \mathcal{N}_i}$. Herein, the local attention layer employed key-query attention¹⁶ $\text{Attention}(\mathbf{q}_i^l, \mathbf{K}_{\mathcal{N}_i}^l)$, in which the query vector \mathbf{q}_i^l and key matrix $\mathbf{K}_{\mathcal{N}_i}^l$ are defined as $\mathbf{c}_i^l \mathbf{W}_q^l$ and $(\mathbf{C}_{\mathcal{N}_i}^l \times \mathbf{G}_{\mathcal{N}_i}^l) \mathbf{W}_k^l$, respectively.

Previous studies have commonly represented a material structure by a combination of its local structures, typically using operators such as summation or pooling. However, these operators either assume equal contribution from all local structures (sum and average-pooling)^{14,15,28–30} or focus on a single specific local structure (max- and min-pooling)^{31–33}, which can hinder the elucidation of structure-property relationships. To address this issue, SCANN represents a material structure as a linear combination of the representation vectors of its local structures, with the global attention (GA) scores of each local structure serving as the coefficients (Supplementary Section IVC). We preserve the structural information of S from all representations of its local structures obtained at the final local attention layer to produce \mathbf{C}^L , where $\mathbf{C}^L = [\mathbf{c}_i^L]_{1 \leq i \leq M}$. The global attention layer subsequently learns a suitable representation of the material structure by considering representations of its constituent local structures. This enables the accurate prediction of material

properties:

$$\begin{aligned} \mathbf{x}_S &= \text{GlobalAttention}(\mathbf{C}^L) \\ &= \text{SAttention}(\mathbf{Q}^g, \mathbf{K}^g) \\ &= \rho(\mathbf{Q}^{gT} \mathbf{K}^g) \mathbf{K}^g \\ &= \boldsymbol{\alpha}^g \mathbf{K}^g, \end{aligned} \quad (2)$$

where $\mathbf{Q}^g = \mathbf{C}^L \mathbf{W}_q^g$ and $\mathbf{K}^g = \mathbf{C}^L \mathbf{W}_k^g$ are the query and key matrices, $\boldsymbol{\alpha}^g = [\alpha_1^g, \alpha_2^g, \dots, \alpha_M^g]$ are the corresponding GA scores of each local structure \mathbf{c}_i . Herein, the global attention layer employed key-query attention SAttention; however, instead of the softmax function, a weighting function $\rho(\cdot)$ is applied to the attention matrix $\mathbf{A} = \mathbf{Q}^{gT} \mathbf{K}^g$ to evaluate the GA scores. In detail, we obtain $s_j = \sum_{i=1}^M [\mathbf{A}(\mathbf{1} - \mathbf{I})]_{ij}$ as the sum of each column j within the attention matrix \mathbf{A} (the identity matrix is denoted as \mathbf{I}). Then, we define $\rho(\mathbf{A}) = \text{softmax}([s_1, s_2, \dots, s_M])$. Subsequently, this approach can measure the amount of attention (GA scores) that should be given to a local structure by summing all corresponding directional pairwise attention scores from other local structures (Fig. 1b).

Consequently, the physical property y_S of the material structure S can be predicted from the learned representation \mathbf{x}_S with fully connected layers F_S , as follows:

$$\hat{y}_S = F_S(\mathbf{x}_S) \quad (3)$$

This design of the SCANN architecture, particularly the inclusion of a fully connected layer, is tailored to capture the complex and nonlinear relationships between their representations and properties. Furthermore, the GA scores $\boldsymbol{\alpha}^g$ of the local structures, obtained from the global attention layer, help identify key factors that contribute to understanding the structure–property relationships of the material. A comprehensive depiction of the proposed SCANN architecture is presented in Fig. 2.

Experimental design

In this study, we implement two versions of deep learning models using the proposed architecture, each trained independently on different datasets with distinct target properties, with the aim of evaluating the architecture's performance in predicting target properties and its ability to provide information regarding the structure–property relationships (interpretability) across five molecular and crystal structure datasets (Table 1). The properties of these datasets are determined through quantum mechanical calculations using DFT. The predictive capability is evaluated by partitioning the data into train-validation-test sets. The models are then trained on the training set and optimized to minimize the mean absolute error (MAE) on the validation set. The MAEs of the predictions for the

Dataset	#Size	#Atoms	Properties
QM9 ⁶	130,831	4 to 29	E_{HOMO} , E_{LUMO} , E_{gap} , α , C_v
Fullerene-MD ³⁷	3000	60, 70, 72	E_{HOMO} , E_{LUMO}
Pt/Graphene-MD ³⁷	21,666	103	ΔU
SmFe ₁₂ -CD ³⁸	3307	13	ΔE
MP 2018.6.1 ^{34,35}	69,239	1 to 296	ΔE , E_g

The table shows information of five datasets regarding eight properties analyzed with the SCANN models, including dataset size (number of molecules/crystals - #Size), number of atoms present in structures (#Atoms), and the specific physical properties examined.

E_{HOMO} (meV) Energy of the highest occupied molecular orbital, E_{LUMO} (meV) Energy of the lowest unoccupied molecular orbital, E_{gap} (meV) Energy HOMO-LUMO gap, α (Bohr³) Isotropic polarizability, C_v (cal mol⁻¹ K⁻¹) Heat capacity at 298 K, ΔU (eV) Deformation energy, ΔE (meV atom⁻¹) Formation energy per atom, E_g (meV) Band gap.

target properties on the test sets are reported for comparison with other models in the literature. Hereafter, the implemented models based on the SCANN architecture, trained with the respective datasets, will be referred to as SCANN models. A detailed explanation regarding the used datasets is given in Supplementary Section IVE.

We evaluated the predictiveness of the SCANN models based on a comparison with seven DL models using the QM9⁶ and Materials Projects^{34,35} (version MP 2018.6.1) datasets. Among the compared models, SchNet³⁰, MEGNet³⁵, and CGCNN¹² utilize graph neural networks to represent molecules or crystals as atomistic graphs. Cormorant¹⁴ and SE(3)-Trans¹⁵ are GNN variants that incorporate physical constraints, such as covariant or equivalence principles, on the 3D coordinates of atoms. Conversely, ALIGNN³⁶, the current leading network, uses an additional line graph where bonds serve as nodes, and edges convey angular information between bonds in addition to the atomistic graph. This enables ALIGNN to represent the geometrical arrangement of triplets of atoms in a molecule or crystal.

Furthermore, the interpretability of the SCANN models is assessed by examining the relationship between the learned GA scores of the local structures and the corresponding results from first-principles calculations. The results demonstrate the ability of the SCANN models to provide valuable information regarding the structure–property relationships of materials in four scenarios: the local structures and HOMO/LUMO molecular orbitals (QM9⁶ and Fullerene-MD³⁷), the deformation energy ΔU and the deformation of the Pt/graphene structures (Pt/graphene-MD³⁷), and the derived crystal formation energy and the substitution atom species and sites of SmFe₁₂-based compounds (SmFe₁₂-CD³⁸).

Evaluation of the predictive performance

Train-validation-test splits are performed in an 80:10:10 ratio to evaluate the predictive capability of SCANN in predicting five physical material properties (E_{HOMO} , E_{LUMO} , E_{gap} , α , and C_v) in the QM9 dataset. Six DL methods with the MAE of the predictions derived from the models are also employed for comparison. The evaluation process is repeated five times to obtain an average MAE for the test set, thereby providing a robust assessment of the predictive capabilities of the models^{14,15}. Similar to previous practices^{30,35,36}, our study employs train-validation-test configurations of 60000–5000–4239 for the MP 2018.6.1 dataset.

Table 2 presents the average MAE scores obtained from five training runs of the SCANN models, alongside the scores of competing models on the QM9 dataset. The ALIGNN outperforms

Table 2. Comparative evaluation of SCANN, SCANN⁺, and six other DL models predicting five physical properties using the QM9 dataset.

	E_{HOMO} (meV)	E_{LUMO} (meV)	E_{gap} (meV)	α (Bohr ³)	C_v (cal mol ⁻¹ K ⁻¹)
WaveScatt ³⁹	85	76	118	0.160	0.049
SchNet ³⁰	41	34	63	0.235	0.033
MEGNet ³⁵	38	31	61	0.081	0.030
Cormorant ¹⁴	34	38	61	0.085	0.026
SE(3)-Trans ¹⁵	35	33	53	0.142	0.054
ALIGNN ³⁶	21	19	38	0.056	–
SCANN	41	37	61	0.141	0.05
SCANN ⁺	32	31	52	0.115	0.041

The bold numbers denote the lowest mean absolute errors (MAEs) among the eight models.

The dash symbol (–) indicates the result has not been reported yet.

E_{HOMO} Energy of the highest occupied molecular orbital, E_{LUMO} Energy of the lowest unoccupied molecular orbital, E_{gap} Energy gap, α Isotropic polarizability, C_v Heat capacity at 298 K.

Table 3. Comparative evaluation of SCANN and four other DL models in predicting two physical properties on the MP 2018.6.1 dataset.

Target	Unit	CGCNN ¹²	SchNet ³⁰	MEGNet ³⁵	ALIGNN ³⁶	SCANN	SCANN ⁺
ΔE	meV atom ⁻¹	39	35	28	22	29	28
E_g	meV	388	–	330	210	260	225

The bold numbers denote the lowest mean absolute errors (MAEs) among the six models.

The dash symbol (–) indicates the result has not been reported yet.

ΔE Formation energy per atom, E_g Band gap.

all competing models across the four properties under consideration. In comparison, the MAEs of the SCANN models are 2 to 2.5 times higher than those of ALIGNN. Despite this gap, SCANN exhibits competitive performance relative to other remaining models, particularly in predicting E_{HOMO} , E_{LUMO} , and E_{gap} . For a comprehensive analysis of the predictive capabilities of these models, please see Supplementary Section IIIA. The predictiveness of the SCANN models on the MP 2018.6.1 dataset is detailed in Table 3. Similar to the result on the QM9 dataset, the ALIGNN model yields the highest prediction accuracy among all competitors. Meanwhile, the MAEs of the SCANN models for predicting the formation energy (ΔE) and band gap (E_g) achieve 29 meV atom⁻¹ and 260 meV, which are 24% and 19% higher than those of ALIGNN, respectively. Notably, the SCANN model has a comparable result with the MEGNet model and exhibits enhancements over the CGCNN and SchNet models regarding ΔE predictions. For E_g , the SCANN model outperforms the CGCNN and the MEGNet models by 32% and 21%, respectively.

Incorporating conventional prior knowledge (e.g., several dozen atomic features and bonding information between atoms^{12,35,36}) or adding physical constraints (e.g., equivalencies, covariates, and equations^{14,15,39}) to the learning process for structure representations can enhance the prediction accuracies. For instance, the ALIGNN model outperformed all competitor models by introducing additional angular information between triplets of atoms, while others only considered two-body interactions (distances and/or bond valences). To improve the descriptiveness for the geometrical structure of molecules or crystals, we develop another version of the SCANN, the SCANN⁺, with minor adjustments to the original one by employing the Voronoi solid angle embedding layer and updating the geometrical information through multiple LocalAttention layers (Supplementary Section IVD). The additional updates significantly improve the predictive power of the method; the SCANN⁺ models outperform all other competitors, except for the ALIGNN model, in predicting the electronic properties (E_{HOMO} , E_{LUMO} , and E_{gap}) on the QM9 dataset; ΔE and E_g on the MP 2018.6.1 dataset) that are sensitive to the geometrical structure of molecules or crystals (Tables 2 and 3).

However, these strategies introduce higher dimensionality and may bring consequent biases into the model by favoring certain materials, overlooking others, or oversimplifying complex phenomena due to constraints or potential inaccuracies in the heuristic information assigned during the training phase. Consequently, such issues could hamper the clear understanding of structure-property relationships, which is the primary objective of this study. Additionally, for the QM9 dataset, the widely accepted “chemical accuracy” thresholds are 43 meV for the three energy-related properties, E_{HOMO} , E_{LUMO} , and E_{gap} ; 0.1 Bohr³ for the isotropic polarizability α ; and 0.05 cal mol⁻¹ K⁻¹ for the heat capacity C_v at 298 K⁴⁰. Notably, the SCANN models demonstrated a prediction error of 41 meV for E_{HOMO} , 34 meV for E_{LUMO} and 0.05 cal mol⁻¹ K⁻¹ for C_v , indicating that chemical accuracy thresholds were achieved for these properties.

In practice, exceeding the threshold for chemical accuracy by increasing the model’s complexity is unnecessary, particularly when the data employed are derived via DFT calculations. Such an

approach can lead to overfitting or biases, thereby potentially sacrificing model interpretability and the elucidation of underlying chemical principles. Therefore, we investigate the relationship between the structures of the molecules in the QM9 dataset and their properties (E_{HOMO} and E_{LUMO}) by using the GA scores obtained from the SCANN models instead of those from the more complex SCANN⁺ model, which possesses higher dimensionality and more parameters (Supplementary Section IID).

Supplementary Section IIIA presents the evaluation of SCANN’s predictive capabilities on three in-house-developed material datasets, demonstrating its broad adaptability and high accuracy in diverse prediction scenarios. The remarkable prediction accuracy of SCANN confirms its practical applicability and suggests that the interpretation derived from the attention scores provides valuable insights into key structure–property relationships for the investigated material properties. In the following sections, we examine the correspondence between the obtained GA scores of the local structures and the corresponding results from first-principles calculations to assess the interpretability of the SCANN models (Supplementary Section IIE–G).

Correspondence between the learned attentions of local structures and the molecular orbitals of small molecules

For the small molecules in the QM9 dataset, the SCANN models demonstrate a remarkable correspondence between the obtained GA scores of the local structures and molecular orbitals results obtained via DFT calculations. As a representative example, Fig. 3 shows the comparison between the GA scores of the local structures and the HOMO/LUMO orbitals obtained from DFT calculations for four molecules. Notably, an apparent correspondence between the relative GA scores of the local structures and the HOMO orbitals of the dimethyl butadiene molecule (*cis*-2,3-dimethyl-1,3-butadiene) is evident (Fig. 3a). Furthermore, the GA scores of the local structures can be easily linked to the interpretation that dimethyl butadiene readily undergoes the Diels–Alder reaction. Similarly, the correspondence between the HOMO orbital and the GA scores of the local structures is apparent for the thymine molecule (5-methyl pyrimidine-2,4 (1*H*,3*H*)-dione), which is a nucleobase in DNA (Fig. 3b).

Moreover, similar correspondences are confirmed between the GA scores of the local structures and the LUMO orbitals obtained from the DFT calculations for methyl acrylate (methyl prop-2-enoate) and dimethyl fumarate (dimethyl(2*E*)-but-2-enedioate). Methyl acrylate is a reagent that is commonly used in the synthesis of various pharmaceutical intermediates⁴¹, whereas dimethyl fumarate has been proposed to exhibit immunomodulatory properties without causing significant immunosuppression⁴²; thus, it has been evaluated as a potential treatment for COVID-19⁴³. The apparent correspondence between the LUMO orbitals and the GA scores of the local structures of these two molecules (Fig. 3c, d) further highlight that the attention scores of the SCANN model provide valuable insights to interpret the structure–property relationships of molecules. Further investigations show that the obtained GA scores from the SCANN⁺ models are almost consistent with those of the SCANN models for these molecules (Supplementary Fig. 1).

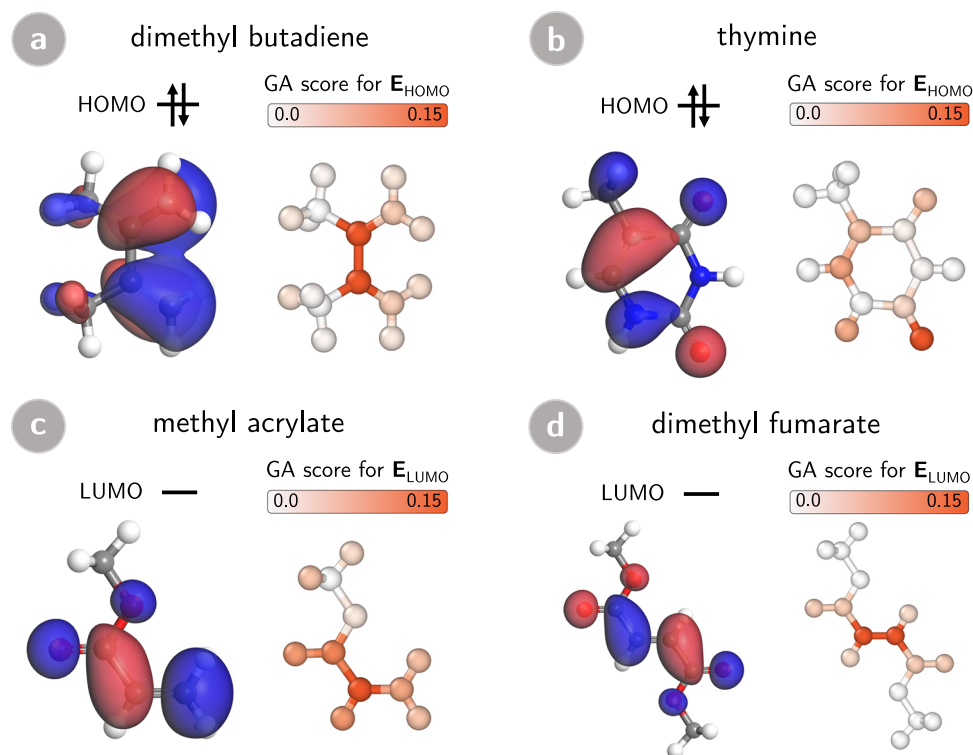


Fig. 3 Visualizations of structure–property relationships for molecules in QM9 dataset. Correspondence between obtained GA scores and molecular orbitals of four molecules: (a) dimethyl butadiene, (b) thymine, (c) methyl acrylate, and (d) dimethyl fumarate. For each molecule, the left side of the figure illustrates the wave function of the HOMO (a, b), or the LUMO (c, d), as calculated using the DFT approach. The isosurfaces with positive and negative values of the wave functions are represented by blue and red lobes, respectively. The right-side figures display the GA scores of the local structures derived from the SCANN models for interpreting the corresponding molecular orbitals. The coloration of atoms indicates their estimated GA scores, while the coloration of links between them does not signify the sign or nodes of the molecular orbital wave functions.

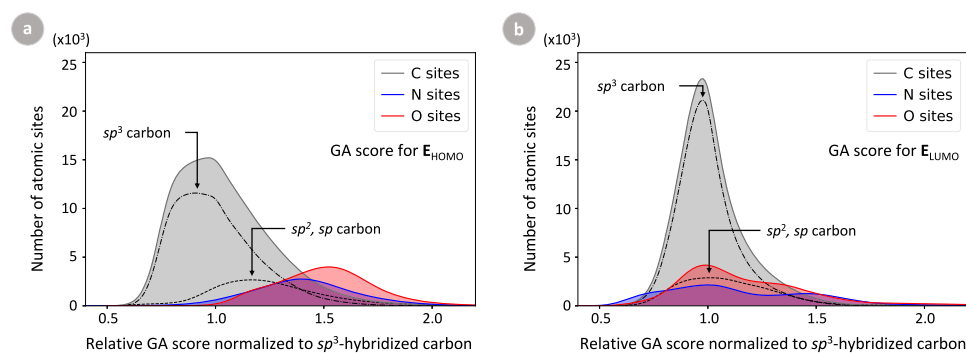


Fig. 4 Correspondence between obtained GA scores of carbon, nitrogen, and oxygen atomic sites and molecular orbitals of molecular structures in QM9 dataset. Statistics of the relative GA scores for E_{HOMO} (a) and E_{LUMO} (b) for all carbon, nitrogen, and oxygen atomic sites in the molecular structures of the QM9 dataset, calculated based on the average GA score of sp^3 -hybridized carbon atoms in each molecule. Gray, blue, and red lines and filled regions represent the statistics for carbon, nitrogen, and oxygen sites, respectively.

All carbon, nitrogen, and oxygen atomic sites in the QM9 dataset were statistically analyzed for a systematic evaluation of the GA scores obtained by the SCANN models. Since the GA scores of atomic sites were normalized to 1, the relative GA scores were calculated based on the average GA score of the sp^3 -hybridized carbon atoms in each molecule. Molecules without any sp^3 -hybridized carbon atoms were excluded (Fig. 4). The analysis of the GA scores for the HOMO energy reveals that the influence on HOMO follows the order of oxygen, nitrogen, and carbon. Specifically, sp^3 -hybridized carbon sites have a lower influence compared to sp^2 -hybridized or sp -hybridized carbon sites (Fig. 4a). These findings align with

the electronegativity and bonding characteristics of the elements. Oxygen and nitrogen exhibit strong electronegativity and electron-rich regions in π -bonds, leading to a more significant electron density shift and higher HOMO energy localized around oxygen, nitrogen, and carbon sites with double or triple bonds.

In contrast, the GA scores for the LUMO energy show no significant difference among the three elements. This observation is consistent with the understanding that unoccupied orbitals primarily influence the LUMO energy, resulting in a less pronounced difference in electronegativity compared to its impact on the HOMO energy (Fig. 4b).

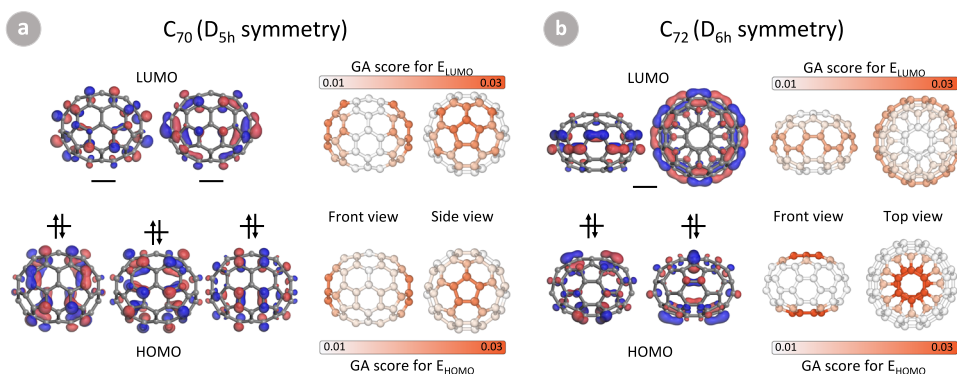


Fig. 5 Visualizations of structure–property relationships for fullerene molecules. Correspondence between obtained GA scores and the molecular orbitals of (a) C_{70} and (b) C_{72} . For each molecule, the left side of the figure illustrates the wave functions of the degenerate HOMO (bottom) and LUMO (top) orbitals, as calculated by the DFT approach. The isosurfaces with positive and negative values of the wave functions are represented by the blue and red lobes, respectively. The figure on the right displays the GA scores of local structures obtained using the SCANN model for the corresponding property. The coloration of atoms indicates their estimated GA scores, while the coloration of links between them does not signify the sign or nodes of the molecular orbital wave functions.

Correspondence between the learned attentions of local structures and molecular orbitals of fullerene molecules

To further evaluate the interpretability of the proposed method, the correspondence between the obtained GA scores of the local structures and the molecular orbitals obtained from DFT calculations for fullerene molecules is examined. Supplementary Fig. 2 shows the GA scores of the local structures for the HOMO and LUMO energies of the C_{60} molecule (I_h symmetry). In this case, the target molecule has a truncated icosahedral structure composed of 20 hexagons and 12 pentagons, with all carbon atoms exhibiting equivalent local structures. The SCANN model estimates identical GA scores for all local structures of the C_{60} molecule, thus indicating its ability to handle large and symmetric molecules.

As the number of carbon atoms in the fullerene molecule increases, the symmetry of the C_{70} (D_{5h} symmetry) and C_{72} (D_{6h} symmetry) molecules becomes slightly broken, and the local structures of the carbon atoms in these molecules are no longer equivalent. Figure 5 demonstrates the significant correspondence between the GA scores of the local structures and the HOMO and LUMO results obtained from DFT calculations for the C_{70} and C_{72} molecules. The GA scores of the local structures in the C_{70} and C_{72} molecules exhibit a five-fold (top view) and six-fold (top view) symmetry upon the prediction of the HOMO energy, respectively. These results align with the structural symmetry and degenerate HOMO orbitals of the two fullerene molecules. Notably, the C_{70} molecule possesses an additional 10-carbon ring, forming a plane symmetry, resulting in a planar symmetry of its HOMO with the node situated on that ring's plane. The SCANN model reveals a clear correspondence between the HOMO of the C_{70} molecule and the GA scores of the local structures (Fig. 5a), along with the LUMO and their corresponding GA scores. Furthermore, the shapes of LUMO and HOMO of the C_{72} molecule exhibit a perfect correspondence with the GA scores of the local structures obtained using the SCANN models (Fig. 5b). Compared to C_{60} , the C_{72} molecule has an additional ring of 24 carbon atoms with six-fold symmetry, consisting of 12 pairs of carbon–carbon bonds in five-membered carbon rings. The high GA scores of the local structures in the ring indicate the localization of the LUMO of the C_{72} molecule on the ring. In contrast, the HOMO orbitals are located on two opposite sides of the ring and are also captured by the local structures with high GA scores. This evaluation experiment provides further confirmation that SCANN-derived GA scores offer valuable insights for understanding the structure–property relationship, even for large molecules.

Correspondence between the learned attentions of local structures and structural deformation in Pt/graphene

Figure 6a presents the GA scores of the local structures obtained by the SCANN model for predicting the deformation energy of a system comprising a platinum atom adsorbed on a graphene flake. The deformation energy is defined as the difference between the total energy of the deformed and optimized structures. A detailed examination of the obtained GA scores reveals that local structures with high GA scores possess relatively elongated carbon–carbon bonds (Fig. 6b). Additionally, the carbon atoms that form high local curvatures upon the formation of a convex from the planar structure of the sp^2 hybridization bonding network received high GA scores (Fig. 6c).

The results obtained from the experiment on the system where a platinum atom was adsorbed on a graphene flake reveal that the GA scores obtained by the SCANN model exhibit a high correspondence with the observed structural deformations. In particular, the high GA scores for the increased carbon–carbon bond lengths and the convexed carbon atoms align well with the contribution to the deformation energy, as determined by DFT calculations. This finding indicates that the GA scores generated by SCANN are reliable indicators of structural deformations in such systems, demonstrating the model's capability to capture and interpret the underlying material instability. These results validate the usefulness of SCANN in understanding and predicting structural deformations in materials, particularly in cases involving the interaction of different elements or adsorption onto surfaces.

Correspondence between the learned attentions of local structures and stability of $SmFe_{12}$ -based crystal structures

The SCANN model's ability to predict the formation energy of $SmFe_{12}$ -based crystal structures is evaluated by analyzing the GA scores of atomic sites. The focus is on understanding the effects of substituting Fe sites with other elements on the derived formation energy and the stabilization of the crystal structure, as well as the influence of the elemental substitution on the formation energies of other Sm and Fe sites. It should be noted that the GA scores of the local structures are normalized to ensure that the sum of the attention scores of all local structures in the crystal structure is equal to one.

For instance, Fig. 7a shows the GA scores of the local structures obtained using the SCANN model for predicting the formation energies of the $SmFe_{12}$, $SmFe_{11}Mo$, $SmFe_{11}Co$, and $SmFe_{11}Al$ crystal structures. For the optimized $SmFe_{12}$ crystal structure, all Fe sites receive identical GA scores, indicating a symmetric cage of Fe atoms surrounding the Sm atoms. Additionally, the negligible

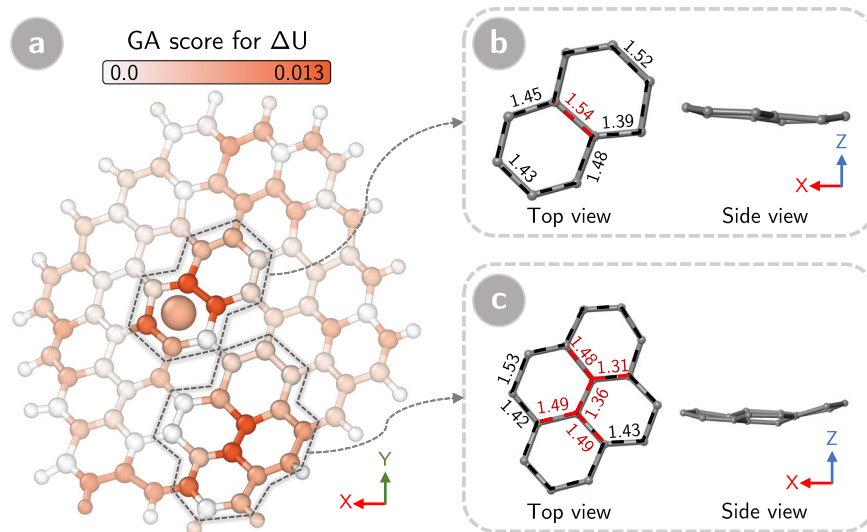


Fig. 6 Visualization of relationship between the adsorption energy and the deformation of a graphene flake with a platinum atom adsorbed on a graphene flake. **a** Visualization of the GA scores obtained from the SCANN model for the Pt/Graphene system with a deformation. The coloration of atoms indicates their estimated GA scores, while the coloration of links between them does not signify the sign or nodes of the molecular orbital wave functions. Structural visualizations of the high attention local structures during the deformation: **(b)** elongated carbon–carbon bond, and **(c)** convexed carbon–carbon configuration. The distance from the carbon atom to the adjacent carbon atom (in Å) is highlighted to show the distortion caused by the deformation.

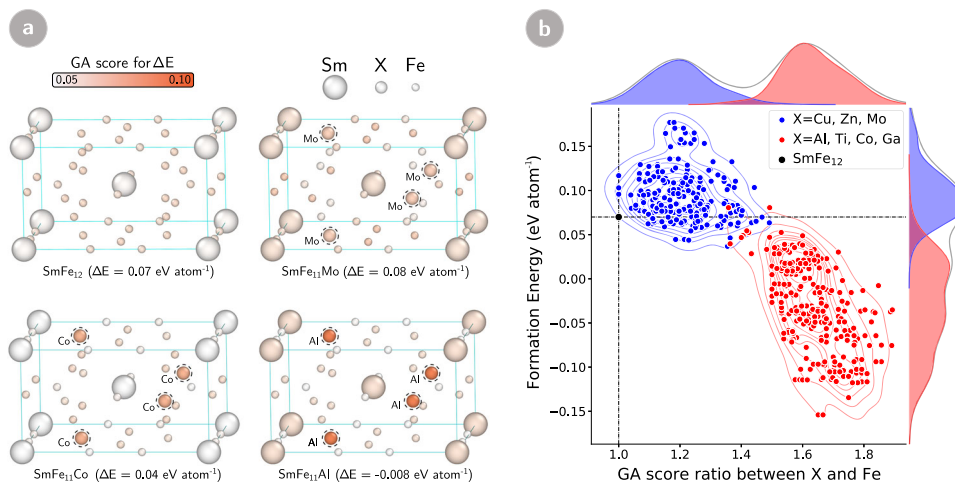


Fig. 7 Visualization of relationship between structure and formation energy obtained from SCANN model for crystalline magnetic materials in SmFe_{12} -CD. **a** Visualization of the GA scores estimated by the model for atomic sites in the SmFe_{12} , $\text{SmFe}_{11}\text{Mo}$, $\text{SmFe}_{11}\text{Co}$, and $\text{SmFe}_{11}\text{Al}$ crystal structures. **b** Correlation between the ratio of GA scores of the substitution sites to the minimum GA scores among the Fe sites and the formation energy, as calculated via the DFT approach, in crystal structures substituted by a single type of element.

difference in GA scores between the Sm and Fe sites suggests that when analyzing the formation energy of the SmFe_{12} crystal structure, greater attention should be given to the Fe sites rather than the Sm sites. This implies that Sm atoms are comfortably placed within the cage of Fe atoms in the SmFe_{12} crystal structure.

For the crystal structures with Mo substitutions, the GA scores of the Mo and Sm sites are estimated to be the same as those of the Fe sites. However, for crystal structures with Co or Al substitutions, the GA scores of the Co and Al sites are significantly higher than those of Fe sites. The GA score results for the three crystal structures indicate that Mo substitution has little effect on the cage of Fe atoms, whereas the Sm sites become nonnegligible in interpreting the formation energy of the $\text{SmFe}_{11}\text{Mo}$ crystal structure. This suggests that Sm atoms are less comfortably placed within the Fe and Mo atom cages in the substituted crystal structure. By contrast, for crystal structures substituted with Co or

Al, the GA scores of the Co and Al sites are significantly higher than those of Fe sites, indicating that the Co and Al sites should be the central focus of attention when interpreting the formation energy of the $\text{SmFe}_{11}\text{Co}$ and $\text{SmFe}_{11}\text{Al}$ crystal structures, respectively. Moreover, the GA scores of Fe sites exhibit a slight decrease, indicating that the Fe atoms become more comfortably placed in the substituted crystal structures.

To validate the aforementioned interpretation, the ratio of the GA scores of the substitution sites to the minimum GA scores among the Fe sites was calculated for each crystal structure. Subsequently, the relationship between this ratio and the estimated formation energies of the structures was investigated using DFT calculations. Figure 7b shows that the crystal structures substituted with a single type of element can be divided into two groups: one with Cu, Zn, and Mo substitutions, and the other with Al, Ti, Co, and Ga substitutions. Interestingly, it was observed that

crystal structures with higher local structure GA scores for the substitution sites possess lower formation energies, whereas those with lower local structure GA scores for the substitution sites possess higher formation energies. These results highlight the potential of the SCANN model in estimating the local structure GA scores for a rational discussion of SmFe₁₂-substituted crystal structures and their formation energies. While additional first-principles calculations are necessary for each specific crystal structure to fully understand the relationship between the substitution elements, the substitution sites, and the crystal structure stability, these results indicate the potential usefulness of SCANN. The local structure GA score provides valuable information and indicates key focus points for understanding the stability of crystalline material structures. Thus, this study offers valuable insights that can contribute to the development of more efficient and effective methods for designing crystal material structures.

DISCUSSION

This study proposes SCANN, an attention-based DL architecture designed for material dataset analysis. SCANN leverages attention mechanisms to learn from material datasets, predict material properties, and interpret the underlying characteristics of material structures. By applying attention recursively to neighboring local structures, SCANN learns representations of atomic local structures in a self-consistent manner. The architecture then combines these local structure representations to create a comprehensive representation of the entire material structure, enabling precise property predictions. During the learning process, global attention scores are estimated, indicating the importance of each local structure in representing the overall material structure. Experimental results based on five molecular and crystalline material structure datasets demonstrated the excellent predictive capability of SCANN for different material properties. Furthermore, an in-depth qualitative analysis of the global attention scores of local structures revealed that the trained models can extract essential information from material datasets, facilitating a deeper understanding of the structure–property relationships in both molecular and crystalline materials. The ability of the proposed architecture to interpret the attention scores can aid in identifying critical features and accelerating the material design process.

However, it is important to acknowledge that the interpretability of attention scores in DL models is still a subject of debate and lacks clear guidelines^{44–46}. Several factors need to be considered, such as the correlation analysis of attention scores, alternative interpretability metrics, and counterfactual analysis, to validate meaningful explanations of the relationships. Additionally, the quantification and assessment of uncertainty in attention score estimation are essential. Despite these challenges, the findings of this study demonstrate the potential of attention mechanisms in uncovering valuable information that can provide a better understanding of structure–property relationships in materials.

METHODS

Characterization of material structure

Given a material structure S with the property of interest $y_S \in \mathbb{R}$ containing M atoms ($\mathcal{A}_S = \{a_1, a_2, \dots, a_M\}$), we consider the structure S as a geometrical arrangement of M local structures. Each local structure consists of a central atom, its neighboring atoms, and their arrangement around the central atom. To determine the neighboring atoms and segment each material structure into local structures, we employ the definition of O’Keeffe^{47,48} instead of the assumption about chemical bonds between the atoms in the structure. According to O’Keeffe’s

definition, all atoms at these atomic sites share Voronoi polyhedron faces with the atomic site of an atom under consideration (the central atom of the local structure) and are regarded as neighboring atoms. Subsequently, the local structures of the neighboring atoms are referred to as the neighboring local structures. By incorporating the information from the Voronoi polyhedron faces, we assess the geometrical influences of neighboring atoms on the central atoms for conveying the structural information of structure S to SCANN for learning the appropriate representation of S .

For each atom a_i in the structure S , by using the Voronoi tessellation, we can determine $\mathcal{N}_i \subset \mathcal{A}_S$, which contains N atoms whose atomic sites share Voronoi polyhedron faces with an atomic site of a_i . Subsequently, the geometrical influence of a neighboring atom $a_j \in \mathcal{N}_i$ on atom a_i is represented by a vector $\mathbf{g}_{ij} \in \mathbb{R}^h$, which is defined by the element-wise multiplication of the Euclidean distance d_{ij} (Å) and Voronoi solid angle⁴⁸ $\theta_{ij} \in [0, 4\pi]$ information between the atoms, as follows:

$$\mathbf{g}_{ij}^0 = \mathbf{DE}(d_{ij}) \times \frac{\theta_{ij}}{\max(\theta_{ik})}, \quad (4)$$

where $\mathbf{DE}(d_{ij})$ is a distance embedding layer representing the distance d_{ij} as an h -dimensional vector (Supplementary Section IIB). As a result, for each atom a_i , we obtain a matrix $\mathbf{G}_{\mathcal{N}_i}^0 = [\mathbf{g}_{ij}^0]_{a_j \in \mathcal{N}_i}$ representing the geometrical influences of the neighboring atoms of atom a_i . Each row of the matrix consists of a vector \mathbf{g}_{ij}^0 that represents the geometrical influence of atom a_j on atom a_i .

Local structure representation

Similar to other DL architectures, the SCANN employs an embedding layer (Supplementary Section IIA) to express the atomic information of each atom a_i in S as an h -dimensional vector $\mathbf{c}_i^0 \in \mathbb{R}^h$. Through training, the vector representation \mathbf{c}_i^0 is updated and refined to represent the atom more appropriately for accurately predicting property y_S of material structure S .

To learn representations for local structures in material structure S , a local attention layer that utilizes the atomic and geometrical arrangement of atomic sites is proposed. The design of the local attention layer is based on the dot-product key-query attention¹⁶, $\text{Attention}(\mathbf{q}, \mathbf{K}) = \text{softmax}(\mathbf{q}^T \mathbf{K}) \mathbf{K}$, where $\mathbf{q} \in \mathbb{R}^h$ and $\mathbf{K} \in \mathbb{R}^{h \times h}$ denote the query vector and key matrix, respectively. In addition, SCANN consists multiple local attention layers to iteratively update the representation of local structures in a layer-wise manner; the $(l+1)$ th local attention layer uses the representations of local structures constructed from the l th layer as inputs. As a result, this design enables SCANN to efficiently capture information on long-range interaction between local structures in the material structure.

For instance, the representation $\mathbf{c}_i^{l+1} \in \mathbb{R}^h$ of local structure $\{a_i, \mathcal{N}_i\}$ at the $(l+1)$ th local attention layer is derived from the representation vectors in the preceding layer of itself (\mathbf{c}_i^l), its neighboring local structures ($\mathbf{C}_{\mathcal{N}_i}^l = [\mathbf{c}_j^l]_{a_j \in \mathcal{N}_i}$), and the geometrical influence of the neighboring atoms \mathcal{N}_i on atom a_i ($\mathbf{G}_{\mathcal{N}_i}^l$) as follows:

$$\begin{aligned} \mathbf{c}_i^{l+1} &= \text{LocalAttention}^{l+1}(\mathbf{c}_i^l, \mathbf{C}_{\mathcal{N}_i}^l \times \mathbf{G}_{\mathcal{N}_i}^l) \\ &= \text{Attention}(\mathbf{q}_i^l, \mathbf{K}_{\mathcal{N}_i}^l) + \mathbf{q}_i^l, \end{aligned} \quad (5)$$

where $\mathbf{q}_i^l = \mathbf{c}_i^l \mathbf{W}_q^l$ and $\mathbf{K}_{\mathcal{N}_i}^l = (\mathbf{C}_{\mathcal{N}_i}^l \times \mathbf{G}_{\mathcal{N}_i}^l) \mathbf{W}_k^l$; $\mathbf{W}_k^l, \mathbf{W}_q^l \in \mathbb{R}^{h \times h}$ are learnable parameters of the local attention layer and are shared between local structures. For the SCANN model, the geometry influences at each layer is kept the same as initial layer $\mathbf{g}_{ij}^l = \mathbf{g}_{ij}^0$. The detailed implementation of the local attention layer is described in Supplementary Section IIC.

Owing to the application of multiple local attention layers, the attention information regarding a target property between local

structures in a material structure S can be passed through the attention relationships between neighboring local structures. In our evaluation experiments, SCANN models employ L local attention layers, with the value of L tailored to optimize performance for each dataset, as explained in Supplementary Section IIIB. Consequently, we preserve the structural information of S from the representations of all its local structures obtained from the final local attention layer, to produce \mathbf{C}^L , where $\mathbf{C}^L = [\mathbf{c}_i^L]_{a_i \in A_S}$.

Material structure representation

To represent a material structure S , simple operators, such as the sum or pooling operator, are typically applied to integrate the representations of all local structures in S . However, such operators consider that either the contribution of each local structure to the final structure representation is equal (sum and average-pooling operator)^{14,15,28–30} or that the property depends on only the specific local structures in the material structure whereas the others have zero impact (max- and min-pooling operators)^{31–33}. Therefore, designing appropriate combination operators for specific target properties is challenging and requires prior hypotheses regarding the structure–property relationships. To overcome this problem, SCANN again utilizes the dot-product key-query attention¹⁶ to coherently learn the representation of local structures and integrate them into the representation of material structure in a target-dependent manner.

An attention mechanism-based layer, called the global attention layer, is proposed to quantitatively model the attention distribution required across each constituent local structure. This layer aims to obtain a more appropriate representation for the entire structure S . The global attention layer is designed to learn an optimal representation of structure S from data, which subsequently facilitates the construction of a highly accurate predictive model for the target property y_S . The representation vector \mathbf{x}_S of the structure S is formulated by aggregating the representations of all the constituent local structures according to the obtained global attention (GA) scores, as shown below:

$$\begin{aligned} \mathbf{x}_S &= \text{GlobalAttention}(\mathbf{C}^L) = \rho(\mathbf{A}) \mathbf{K}^g \\ &= \mathbf{a}^g \mathbf{K}^g = \sum_{i=1}^M a_i^g \mathbf{k}_i^g, \end{aligned} \quad (6)$$

where $\mathbf{A} = \mathbf{Q}^g \mathbf{T} \mathbf{K}^g \in \mathbb{R}^{M \times M}$, which $\mathbf{Q}^g = \mathbf{C}^L \mathbf{W}_q^g$ and $\mathbf{K}^g = \mathbf{C}^L \mathbf{W}_k^g$ are the query and key matrices, respectively; further, $\mathbf{W}_q^g, \mathbf{W}_k^g \in \mathbb{R}^{h \times h}$ are the learnable parameters of the global attention layer. A weighting function $\rho(\cdot)$ is applied to the attention matrix \mathbf{A} to evaluate the GA scores paid to the local structures. As a result, we obtain $\rho(\mathbf{A}) = \text{softmax}([s_1, s_2, \dots, s_M])$, in which $s_j = \sum_{i=1}^M [\mathbf{A}(1 - \mathbf{I})]_{ij}$ is the sum of each column j within the attention matrix \mathbf{A} (the identity matrix is denoted as \mathbf{I}).

The function $\rho(\cdot)$ is designed based on the assumption that heightened attention should be allocated to local structures whose representations are crucial for accurately representing the other local structures in S . This attention allocation allows a the precise prediction of target property y_S . In essence, a local structure that garners higher cumulative attention scores from all the other local structures should be prioritized when representing material structure S . As a result, the degree of attention to a local structure $\{a_i, \mathcal{N}_i\}$ in S is quantitatively modeled by summing all the attention received from other local structures. For a detailed implementation of the global attention layer, please refer to Supplementary Section IID.

Consequently, the physical property y_S of the material structure S can be predicted from the learned representation \mathbf{x}_S , as shown below:

$$\hat{y}_S = F_S(\mathbf{x}_S), \quad (7)$$

where $F_S : \mathbb{R}^h \rightarrow \mathbb{R}$ is represented by two fully connected (FC) layers. The weight matrices and bias vectors of the network are learned by training the prediction model.

Furthermore, the GA scores $\mathbf{a}^g = [a_1^g, a_2^g, \dots, a_M^g]$, which describe the degree of attention given to the corresponding local structures for representing S , are used to reveal critical aspects that help interpret the structure–property relationship of interest. Notably, the attention to local structures discussed here signifies the amount of information these local structures contribute to appropriately represent S for the accurate prediction of y_S .

SCANN⁺

The SCANN⁺ introduced the embedding vector for the Voronoi solid angle θ_{ij} as follows:

$$\mathbf{g}_{ij}^0 = \mathbf{DE}(d_{ij}) \times \mathbf{AE}(\theta_{ij}), \quad (8)$$

where $\mathbf{DE}(d_{ij})$ and $\mathbf{AE}(\theta_{ij})$ are the distance and angle embedding layers corresponding to the distance d_{ij} and angle θ_{ij} of an h -dimensional vector (Supplementary Section IIB). In addition, the geometry influences between the neighbor \mathbf{c}_j^l and the center \mathbf{c}_i^l are updated based on the following formulation:

$$\mathbf{g}_{ij}^{l+1} = F_g^l([\mathbf{c}_i^l \oplus \mathbf{g}_{ij}^l \oplus \mathbf{c}_j^l]) + \mathbf{g}_{ij}^l, \quad (9)$$

where \oplus denotes the concatenating vectors and $F_g : \mathbb{R}^h \rightarrow \mathbb{R}^h$ is a fully connected (FC) layer.

Model training

The training of the DL model using the proposed architecture begins with the initialization of all learnable parameters. All weighting matrices such as $\mathbf{W}_q^l, \mathbf{W}_k^l, \mathbf{W}_q^g$ and \mathbf{W}_k^g are initialized to random matrices using Glorot Uniform⁴⁹, while the entries of all bias vectors are initialized to zero. The dropout layer and attention dropout¹⁶ are applied in the local attention layers with a rate of 0.1 for better regularization.

In the training process, all parameters of the proposed DL model are updated by minimizing a loss function using Adam optimization⁵⁰ with a scheduled learning rate decay ranging from 5×10^{-4} to 10^{-4} . To predict the physical property y_S of a material structure S in training dataset \mathcal{D} , the loss function is defined as follows:

$$\mathcal{L} = \frac{1}{|\mathcal{D}|} \sum_{S \in \mathcal{D}} (y_S - \hat{y}_S)^2 \quad (10)$$

Each model is trained on an individual Tesla A100-PCIe graphics processing unit (GPU) with a memory capacity of 40 gigabytes. Remarkably, there are fewer than one million parameters in SCANN, primarily affected by the configuration settings of the number of LocalAttention layers (L). Supplementary Table II contains the epoch-wise time cost with a batch size 128 for the QM9 dataset. SCANN excels in term of training times per epoch and is distinguished by its commendable memory efficiency, and it is thus highly suitable for practical deployment.

Dataset information

QM9⁶. This computational dataset comprises of data of 133,885 drug-like organic molecules composed of C, H, O, N, and F. However, 3054 files were removed due to the questionable geometric stability¹⁴ that 130,831 molecules remained were used for the experiments. Five physical properties from the QM9 dataset are used as targets for evaluating the predictive capability of the SCANN models. These properties include the energy of the highest occupied molecular orbital (E_{HOMO}), the energy of the lowest unoccupied molecular orbital (E_{LUMO}), the gap between the energies ($E_{\text{gap}} = E_{\text{LUMO}} - E_{\text{HOMO}}$), the isotropic polarizability (α), and the heat capacity

at 298 K (C_v). In the experiment, the predictive capability of the SCANN models is compared with that of recent state-of-the-art DL models^{14,30,35,39}.

Fullerene-MD³⁷. This is an in-house-developed computational material dataset comprising the data of three well-known fullerene molecules, viz. C_{60} (I_h), C_{70} (D_{5h}), and C_{72} (D_{6h}). It includes optimized structures and 3000 deformed structures obtained from molecular dynamics simulations (1000 structures for each molecule). The HOMO (E_{HOMO}) and LUMO (E_{LUMO}) energies of these structures are determined using DFT calculations, similar to the approach used in the QM9 dataset. Experiments are performed on this dataset to evaluate the predictive capability of the SCANN models for HOMO and LUMO energies and to assess the interpretability of the models' predictions for these properties. A distinctive feature of all structures in this dataset is that they only contain carbon atoms. Furthermore, due to the symmetric nature of fullerene molecules, the local structures within each molecule are highly similar with only minor differences. Therefore, this dataset allows for a precise evaluation of the interpretability of the SCANN model. In the evaluation experiment using this dataset, SCANN models pre-trained on the QM9 dataset are applied to train the prediction models for the HOMO and LUMO energies of the fullerene molecules.

Pt/graphene-MD³⁷. This dataset is also an in-house-developed computational material dataset representing a system composed of a platinum atom adsorbed on a graphene flake terminated by hydrogen atoms^{51,52}. It consists of data from approximately 21,000 optimized and deformed structures generated through molecular dynamics simulations. The adsorption energies of these structures are determined using DFT calculations, similar to the approach used in the QM9 dataset. There are two main purposes of the experiments conducted on this dataset is twofold—to evaluate the predictive performance of the SCANN models for deformation energies of the structures (ΔU) and to assess the interpretability of the models' predictions for these deformation energies. The structural characteristic of this dataset is the presence of a two-dimensional honeycomb network of carbon atoms forming the graphene flake. Although the local structures of each carbon atom in the system exhibit slight distortions from the ideal sp^2 hybridization structure⁵², this dataset enables a quantitative evaluation of the interpretability of the SCANN models in terms of the distortion of the honeycomb network on the graphene surface.

SmFe₁₂-CD³⁸. This dataset is an in-house-developed computational material dataset containing the data of crystalline magnetic materials. It comprises the data from 3307 optimized structures of SmFe₁₂-based compounds, along with their corresponding formation energies (ΔE) as the target properties. The dataset was generated by introducing partial substitutions of Mo, Zn, Co, Cu, Ti, Al, and Ga into the iron sites of the original SmFe₁₂ structure, which exhibits notable magnetic properties. Subsequently, the structures were optimized, and their formation energies were assessed using DFT calculations. A detailed explanation regarding the DFT calculation method used to create this dataset can be found in a previous work³⁸. By using this dataset, the predictive capability of the SCANN models for the formation energies of the structures (ΔE) and the interpretability of the models are quantitatively evaluated to investigate the structural stability of the SmFe₁₂-based structures.

MP-2018.6.1^{34,35}. This dataset is a time-versioned snapshot of the Materials Project dataset. Since solid-state datasets are continuously updated, we specifically chose the MP version used by

previous works to facilitate a direct comparison of model performance with the literature. We specifically utilized the Materials Project version that was updated until June 1, 2018, which encompasses 69,239 crystal structures. Two physical properties from the material dataset are used as targets for evaluating the predictive capability of the SCANN models. These properties include the DFT-computed formation energy per atom (ΔE) and the band gaps (E_g). In the experiment, the predictive capability of the SCANN models is compared with that of recent state-of-the-art DL models.

DATA AVAILABILITY

All the computational material datasets related to this article have been deposited to a Zenodo repository³⁷.

CODE AVAILABILITY

The Python implementations for training the SCANN models and predicting physical properties have been deposited to a GitHub repository⁵³.

Received: 3 June 2023; Accepted: 1 November 2023;

Published online: 07 December 2023

REFERENCES

- Agrawal, A. & Choudhary, A. Perspective: materials informatics and big data: realization of the “fourth paradigm” of science in materials science. *APL Mater.* **4**, 053208 (2016).
- Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A. & Kim, C. Machine learning in materials informatics: recent applications and prospects. *Npj Comput. Mater.* **3**, 54 (2017).
- Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
- Siriwardane, E. M. D., Zhao, Y., Perera, I. & Hu, J. Generative design of stable semiconductor materials using deep learning and density functional theory. *Npj Comput. Mater.* **8**, 164 (2022).
- Ward, L. & Wolverton, C. Atomistic calculations and materials informatics: a review. *Curr. Opin. Solid State Mater. Sci.* **21**, 167–176 (2017).
- Ramakrishnan, R., Dral, P. O., Rupp, M. & Von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. data* **1**, 1–7 (2014).
- Himanan, L., Geurts, A., Foster, A. S. & Rinke, P. Data-driven materials science: status, challenges, and perspectives. *Adv. Sci.* **6**, 1900808 (2019).
- Zhao, Y. et al. Physics guided deep learning for generative design of crystal materials with symmetry constraints. *Npj Comput. Mater.* **9**, 38 (2023).
- Rupp, M., Tkatchenko, A., Müller, K.-R. & Von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301 (2012).
- Schütt, K. T. et al. How to represent crystal structures for machine learning: towards fast prediction of electronic properties. *Phys. Rev. B* **89**, 205118 (2014).
- Karamad, M. et al. Orbital graph convolutional neural network for material property prediction. *Phys. Rev. Mater.* **4**, 093801 (2020).
- Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
- Rahaman, O. & Gagliardi, A. Deep learning total energies and orbital energies of large organic molecules using hybridization of molecular fingerprints. *J. Chem. Inf. Model.* **60**, 5971–5983 (2020).
- Anderson, B., Hy, T. S. & Kondor, R. Cormorant: covariant molecular neural networks. *NeurIPS* **32**, 14537–14546 (2019).
- Fuchs, F. B., Worrall, D. E., Fischer, V. & Welling, M. Se(3)-transformers: 3d rotation equivariant attention networks. *NeurIPS* **33**, 1970–1981 (2020).
- Vaswani, A. et al. Attention is all you need. *NeurIPS* **30**, 6000–6010 (2017).
- Duvenaud, D. K. et al. Convolutional networks on graphs for learning molecular fingerprints. *NeurIPS* **28**, 2224–2232 (2015).
- Wu, Z. et al. Moleculenet: a benchmark for molecular machine learning. *Chem. Sci.* **9**, 513–530 (2018).
- Fung, V., Zhang, J., Juarez, E. & Sumpter, B. G. Benchmarking graph neural networks for materials chemistry. *Npj Comput. Mater.* **7**, 84 (2021).
- Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. *ICML* **70**, 1263–1272 (2017).

21. Yang, K. et al. Analyzing learned molecular representations for property prediction. *J. Chem. Inf. Model.* **59**, 3370–3388 (2019).
22. Chen, P., Jiao, R., Liu, J., Liu, Y. & Lu, Y. Interpretable graph transformer network for predicting adsorption isotherms of metal-organic frameworks. *J. Chem. Inf. Model.* **62**, 5446–5456 (2022).
23. Cao, Z., Magar, R., Wang, Y. & Barati Farimani, A. MOFormer: self-supervised transformer model for metal-organic framework property prediction. *J. Am. Chem. Soc.* **145**, 2958–2967 (2023).
24. Kang, Y., Park, H., Smit, B. & Kim, J. A multi-modal pre-training transformer for universal transfer learning in metal-organic frameworks. *Nat. Mach. Intell.* **5**, 309–318 (2023).
25. Korolev, V. & Protsenko, P. Accurate, interpretable predictions of materials properties within transformer language models. *Patterns* **4**, 100803 (2023).
26. Das, K., Goyal, P., Lee, S.-C., Bhattacharjee, S. & Ganguly, N. Crysmmnet: multi-modal representation for crystal property prediction. *PMLR* **216**, 507–517 (2023).
27. Gunning, D. et al. XAI-Explainable artificial intelligence. *Sci. Robot.* **4**, eaay7120 (2019).
28. Moran, M., Gaultois, M. W., Gusev, V. V. & Rosseinsky, M. J. Site-net: using global self-attention and real-space supercells to capture long-range interactions in crystal structures. *Digit. Discov.* **2**, 1297–1310 (2023).
29. Pham, T.-L. et al. Learning materials properties from orbital interactions. *J. Phys. Conf. Ser.* **1290**, 012012 (2019).
30. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet—a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
31. Wu, Z. et al. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **32**, 4–24 (2021).
32. Schweidtmann, A. M. et al. Physical pooling functions in graph neural networks for molecular property prediction. *Comput. Chem. Eng.* **172**, 108202 (2023).
33. Keyulu, X., Weihua H., Leskovec, J. & Jegelka, S. How Powerful are Graph Neural Networks? In Proceedings of the 7th International Conference on Learning Representations (ICLR 2019), New Orleans, LA, USA, 2019.
34. Jain, A. et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
35. Chen, C., Ye, W., Zuo, Y., Zheng, C. & Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* **31**, 3564–3572 (2019).
36. Choudhary, K. & DeCost, B. Atomistic line graph neural network for improved materials property predictions. *Npj Comput. Mater.* **7**, 185 (2021).
37. Vu, T.-S. & Chi, D. H. Fullerene structures and Pt adsorbed on Graphene structures with HOMO, LUMO and Total energy properties. *Zenodo*, <https://zenodo.org/record/7792716> (2023).
38. Nguyen, D.-N., Kino, H., Miyake, T. & Dam, H.-C. Explainable active learning in investigating structure-stability of $\text{SmFe}_{12-a-\beta}\text{X}_a\text{Y}_\beta$ structures X, Y {Mo, Zn, Co, Cu, Ti, Al, Ga}. *MRS Bull.* **48**, 31–44 (2022).
39. Hirn, M., Mallat, S. & Poilvert, N. Wavelet scattering regression of quantum chemical energies. *Multiscale Model. Simul.* **15**, 827–863 (2017).
40. Faber, F. A. et al. Prediction errors of molecular machine learning models lower than hybrid dft error. *J. Chem. Theory Comput.* **13**, 5255–5264 (2017).
41. Ohara, T. et al. Acrylic acid and derivatives. *Ullmann's Encyclopedia of Industrial Chemistry* 1–21 (Wiley Online Library, 2020).
42. Schulze-Topphoff, U. et al. Dimethyl fumarate treatment induces adaptive and innate immune modulation independent of Nrf2. *PNAS* **113**, 4777–4782 (2016).
43. Mantero, V. et al. Covid 19 in dimethyl fumarate treated patients with multiple sclerosis. *J. Neurol.* **268**, 2023–2025 (2021).
44. Jain, S. & Wallace, B. C. Wallace. Attention is not Explanation. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume1 (Long and Short Papers), 3543–3556, Association for Computational Linguistics, Minneapolis, Minnesota.
45. Wiegrefe, S. & Pinter, Y. Attention is not not Explanation. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 11–20, Association for Computational Linguistics, Hong Kong, China.
46. Grimsley, C., Mayfield, E. & R.S. Bursten, J. Why attention is not explanation: Surgical intervention and causal reasoning about neural models. *LREC* **12**, 1780–1790 (2020).
47. O'Keeffe, M. A proposed rigorous definition of coordination number. *Acta Crystallogr. A: Found. Adv.* **35**, 772–775 (1979).
48. Pham, T. L. et al. Machine learning reveals orbital interaction in materials. *Sci. Technol. Adv. Mater.* **18**, 756 (2017).
49. Glorot, X. & Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. *PMLR* **9**, 249–256 (2010).
50. Kingma, D. & Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA, 2015.
51. Chi, D. H. et al. Electronic structures of Pt clusters adsorbed on (5,5) single wall carbon nanotube. *Chem. Phys. Lett.* **432**, 213–217 (2006).
52. Dam, H. C. et al. Substrate-mediated interactions of Pt atoms adsorbed on single-wall carbon nanotubes: Density functional calculations. *Phys. Rev. B* **79**, 115426 (2009).
53. Vu, T. S. Python implementation of Self-Consistent Attention-based Neural Network - SCANN. *GitHub*, <https://github.com/sinhvt3421/scann-material> (2023).

ACKNOWLEDGEMENTS

This work was supported by the JSPS KAKENHI Grants 20K05301, JP19H05815, 20K05068, and JP23H05403; and the JST-CREST Program (Innovative Measurement and Analysis) JPMJCR2235, Japan.

AUTHOR CONTRIBUTIONS

T.-S.V., M.-Q.H., D.-N.N., T.T., and H.-C.D. conceived and designed the experiments. T.-S.V., M.-Q.H., V.-C.N., and H.-C.D. performed the experiments. T.-S.V., D.-N.N., H.K., and H.-C.D. analyzed the data. T.-S.V., D.-N.N., and H.-C.D. contributed the materials and analysis tools. T.-S.V., M.-Q.H., and H.-C.D. wrote the paper. T.-S.V., M.-Q.H., D.-N.N., Y.A., T.T., H.T., H.K., T.M., K.T. and H.-C.D. reviewed and edited the paper.

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/s41524-023-01163-9>.

Correspondence and requests for materials should be addressed to Hieu-Chi Dam.

Reprints and permission information is available at <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 Author(s) 2023