Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.081876

ARTICLE

MSA-ConvNeXt: Predicting Magnetism of Doped Two-Dimensional Nanomaterials via Multi-Scale Convolution and Attention Mechanisms

Yuxuan Feng, Lili Liang*, Guanglu Sun and Yanrui Wei

College of Computer Science and Technology, Harbin University of Science and Technology, Harbin, China

*Corresponding Author: Lili Liang. Email: lianglili@hrbust.edu.cn

Received: 10 March 2026; Accepted: 27 April 2026; Published: 23 July 2026

ABSTRACT: In doped two-dimensional nanomaterials, magnetism is one of the important physical properties. By introducing foreign doping atoms or molecules, the electronic structure of the material can be effectively regulated, leading to changes in magnetic behavior. Currently, magnetic property prediction has achieved considerable results with the help of traditional CNNs, but there are still obvious limitations: (1) The feature extraction of dopant sites is constrained by fixed receptive fields, making it difficult to characterize local structural perturbations in the vicinity of dopant atoms and their spatial influence propagating to surrounding regions; (2) CNNs lack the capability to model long-range dependencies between non-neighboring atoms and their chemical bonds, thereby weakening the representation of long-range interactions within the material. In this study, we propose Multi-Scale and Attention ConvNeXt (MSA-ConvNeXt) based on multi-scale convolution and attention mechanisms, which consists of the following two core modules: (1) The Multi-scale Convolution Attention Block (MCAB), which models local structural perturbations around dopant atoms and their spatial effects via parallel multi-scale convolutions. It uses a serial channel and spatial attention mechanism to adaptively recalibrate multi-scale features, highlighting the response of doping related regions and enhancing the ability to express dopant-site information; (2) The Visual Geometry Group–Swin Transformer (VGG-Swin) architecture extracts structural features of dopant sites using VGG convolutions to prevent the attenuation of structural information during global relationship modeling. Subsequently, the Swin Transformer is introduced, which uses the self-attention mechanism to dynamically weight and globally associate features at different spatial locations, in order to depict the long-range correlations between non-neighboring atoms and their chemical bonds with the dopant-site. Experiments conducted on a doped two-dimensional nanomaterial dataset constructed from the CMR database demonstrate that the proposed model outperforms existing methods in terms of accuracy and F1-score. Specifically, MSA-ConvNeXt achieves an accuracy of 91.66%, representing an improvement of 1.65% over the next best model. In addition, all experimental results are averaged over multiple independent runs (with five different random seeds), demonstrating the stability and reliability of the model's performance. Ablation studies further validate the effectiveness of each module design.

KEYWORDS: Doped two-dimensional nanomaterials; magnetic property prediction; multi-scale convolution; attention mechanism; ConvNeXt

1 Introduction

With the continuous in-depth research on two-dimensional nanomaterials such as graphene, the quantum confinement effect in low-dimensional systems has triggered the reconstruction of band structure and exchange interactions, causing two-dimensional nanomaterials to exhibit unique properties in electronic, optical, and magnetic aspects, thus becoming an important research direction in the field of materials

science [1,2]. However, intrinsic two-dimensional nanomaterials typically exhibit low magnetic ordering temperatures and limited magnetic anisotropy, which restricts their practical applications in magnetic devices [3]. Therefore, effective magnetic control strategies are urgently needed.

Among various modulation strategies, doping is regarded as an efficient and highly designable modification approach [4,5]. By introducing foreign atoms to substitute lattice sites or occupy interstitial positions, doping can alter the local electronic state distribution and exchange interactions, thereby achieving precise tuning of magnetism [6]. Furthermore, distinct combinations of dopant species, concentrations, and lattice sites significantly influence the magnetic properties of the material [7]. Traditional experimental methods entail complex conditions and high costs during preparation and characterization, rendering them inadequate to support large-scale trial-and-error screening [8,9].

Therefore, the efficient prediction of the magnetic properties of doped 2D nanomaterials has become a critical issue for improving material design efficiency. Magnetism fundamentally relies on the interactions between the local atomic environment and the overall structure, exhibiting complex non-linear correlations [10]. In recent years, deep learning methods have been introduced into the field of material property prediction [11,12]. Among these approaches, graph neural networks have made significant progress in material property prediction by effectively modeling crystal structures based on atomic nodes and their connectivity relations. Meanwhile, with the emergence of structural image datasets for doped two-dimensional nanomaterials, data-driven prediction methods based on visual representations have also attracted increasing attention. It offers a new perspective for magnetic property prediction through structural feature modeling. However, most existing studies directly migrate general-purpose visual network architectures, lacking an inductive bias design specifically tailored to the structural characteristics of doped 2D nanomaterials. This makes it difficult to simultaneously characterize the local structural perturbations induced by dopant sites and the overall long-range correlations of the material, thereby limiting the accuracy of magnetic property prediction [13–18]. Specifically, on the one hand, traditional convolutional neural networks employ kernels with fixed receptive fields, restricting their capacity for multi-scale information modeling. In the structural image representation of doped 2D nanomaterials, it is challenging for them to simultaneously capture the local structural perturbations in the vicinity of dopant atoms and their propagating effects toward larger spatial scales [15,17]. On the other hand, although certain methods embed attention mechanisms within convolutional modules to strengthen feature representation [18,19], they still lack the capability to explicitly model long-range dependencies across different spatial regions, making it difficult to adequately capture the structural correlations among non-neighboring atoms, their chemical bonds, and the dopant sites.

To address the aforementioned issues, this paper constructs a four-stage progressive feature modeling framework, achieving multi-level feature representation from local to global through a shallow-to-deep structural design. The first three stages employ multi-scale convolutional structures to model the local structural perturbations in the vicinity of dopant atoms and their effects across different spatial scales, thereby enhancing the network's perceptual capacity for multi-scale structural information. Subsequently, in the high-level feature space of the fourth stage, a global dependency modeling module based on a self-attention mechanism is introduced to explicitly characterize the long-range structural correlations among different spatial regions, thereby achieving unified modeling of the coupling relationship between local perturbations and the overall structure. Through this collaborative design of "local structure enhancement—global relationship modeling," the prediction of magnetic properties in doped 2D nanomaterials is realized. The main contributions of this paper are summarized as follows:

- We propose a multi-scale convolutional attention block (MCAB) to extract doping-related structural information. It utilizes multiple depthwise separable convolution branches to capture local patterns at

different scales, followed by adaptive fusion via learnable weights. Channel and spatial attention are then applied to recalibrate key responses, enhancing the representation of structural features around dopant sites.

- We propose the VGG-Swin architecture, which achieves local structural pre-enhancement of multi-scale features by introducing VGG convolutions prior to window attention. It utilizes the hierarchical window self-attention mechanism of the Swin Transformer to model long-range structural dependencies across different spatial regions. This effectively characterizes the correlations among non-neighboring atoms, their chemical bonds, and dopant sites, elevating the discriminative capability of the high-level feature space.

Based on the aforementioned modules, we construct the MSA-ConvNeXt model and systematically validate it on the magnetic property prediction task for doped 2D nanomaterials. Experimental results showed that the model achieved a prediction accuracy of 91.66%.

2 Related Works

Early research on material property prediction primarily relied on Graph Neural Networks (GNNs) to explicitly model crystal structures. SchNet [20] encoded interatomic distances through continuous filter convolutions, achieving effective prediction of molecular potential energy surfaces; CGCNN [11] extended graph convolutions to the property prediction of crystalline materials; GATGNN [21] introduces a global attention mechanism to enhance structural relationship modeling. ALIGNN [22] incorporates edge-level information by jointly learning atom and bond graphs, improving the representation of complex structural dependencies. These methods have made significant progress in atomic-level relationship modeling. With the development of visual modeling technologies, some studies began mapping material structures into continuous image representations to leverage the advantages of Convolutional Neural Networks (CNNs) in spatial feature extraction. For instance, McNutt et al. [23] transformed the molecular docking problem into an image recognition task; Takiguchi et al. [24] directly predicted the properties of single-molecule magnets from molecular structures. Such methods can better preserve the continuous spatial distribution characteristics of material structures and enhance the representational capacity for complex, non-linear structure-property relationships. In the field of magnetic property prediction, Kovacs et al. [25] utilized CNNs to model magnetic domain images, achieving the prediction of hysteresis performance and validating the effectiveness of convolutional networks in magnetic structural image analysis. Talapatra et al. [26] further directly regressed magnetic parameters from magnetic force microscopy images, exploring the mapping relationship between visual features and magnetic responses. Zhang et al. [17] combined convolutional and Transformer architectures to predict the magnetism of doped 2D materials, attempting to introduce global relationship modeling on the basis of local feature extraction. These methods predominantly focus on local feature extraction, and the synergistic modeling of dopant-induced local structural perturbations and overall long-range structural correlations remains insufficient.

As the application of convolutional networks in material structural image modeling continues to deepen, how to enhance multi-scale modeling capabilities while maintaining the advantages of local continuous structural representation has become an important research direction. Compared to pure Transformer architectures, convolutional networks possess a more stable inductive bias when processing material images with continuous spatial distribution characteristics. Therefore, structural improvements based on modern convolutional architectures have emerged as a viable pathway. ConvNeXt [27] enhances multi-scale feature representation through designs such as large-kernel convolutions, inverted bottleneck structures, and layer normalization; E-ConvNeXt [19] introduces cross-stage partial connections and channel attention mechanisms to maintain high accuracy while reducing computational complexity, thereby further improving

the efficiency and representational capacity of the model. However, these variants are mainly designed for natural images and are not specifically optimized for doped two-dimensional nanomaterials, especially lacking a mechanism for jointly modeling local structural perturbations and long-range correlations.

To address the aforementioned deficiencies, this paper constructs the MSA-ConvNeXt model based on a multi-scale convolution and attention fusion mechanism. This achieves fine-grained modeling of structural details in the vicinity of dopant sites and their cross-scale diffusion effects, while simultaneously strengthening the correlational representation among non-neighboring regions.

3 Materials and Methods

3.1 Data

The dataset used in this study is derived from a publicly available image dataset of doped two-dimensional nanomaterials, which was constructed based on structural data from the Computational Materials Repository (CMR) [28]. Zhang et al. [17] employed the Jmol-16.1 molecular visualization tool to convert the structural data into images by rendering three-dimensional atomic coordinates into standardized structural images, thereby establishing a two-dimensional material image database. Some example images are shown in Fig. 1.

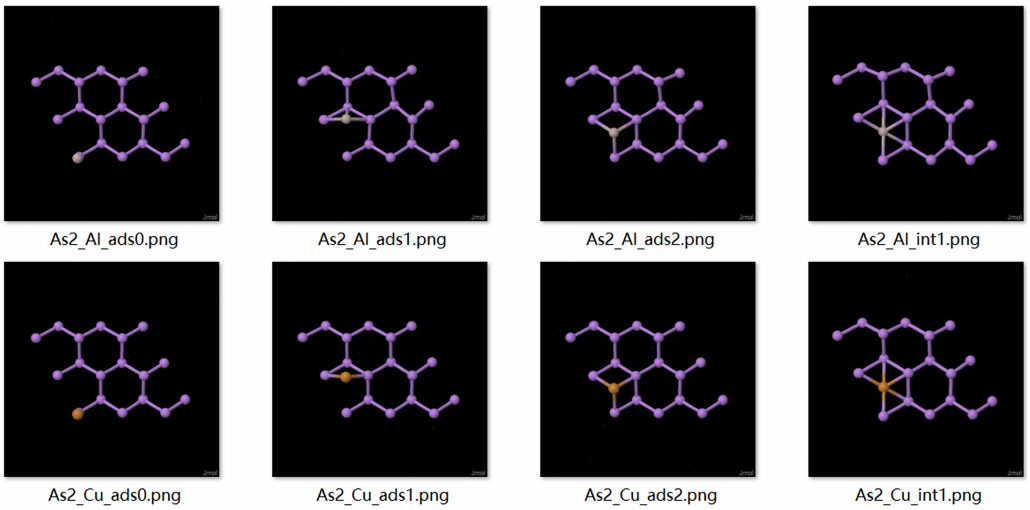


Figure 1: Images of two-dimensional nanomaterials with doping. Each image is named in three parts: the first part represents the main element, the second part represents the doping element, the third part represents the doping method, and additional numerical identifiers are used to distinguish repeated names.

To prevent data leakage, the dataset was first split at the level of original samples into training, validation, and test sets with a ratio of 8:1:1. Data augmentation was applied only to the training set to improve the model's robustness to structural perturbations. The augmentation strategies included horizontal and vertical flipping, rotation, affine transformation, mirror symmetry, Gaussian noise perturbation, brightness and contrast adjustment, scaling, and translation. After preprocessing, the dataset contained a total of 12,760 structural images, all of which were uniformly resized to a resolution of 500 × 500. The magnetic moments obtained from the CMR database were used as labels, and the samples were divided into different magnetic categories according to the range of magnetic moment values, thereby formulating a magnetic classification task.

3.2 Training Settings

The MSA-ConvNeXt model constructed in this study adopts the main hyperparameter settings shown in Table 1.

Table 1: Hyperparameter settings of the model.

Hyperparameter	Value
Input channels	3 (RGB)
Stage depth configuration	[3, 3, 9, 3]
Kernel sizes of multi-branch convolutions	$3 \times 3/5 \times 5/7 \times 7$
Drop rate of stochastic depth	0.3
Initial value for LayerScale	1×10^{-6}
Window size	7
Number of attention heads	24
Label smoothing coefficient (ε)	0.1
Batch size	32
Number of training epochs	300

To ensure the stability and reproducibility of the experimental results, the hardware and software environments were uniformly configured. The specific configurations are listed in Table 2.

Table 2: Hardware and software environment settings.

Environment	Configuration
Python	3.9
PyTorch	1.10.0
GPU	NVIDIA Tesla V100
GPU memory	32 GB
CPU	4 cores

To comprehensively and fairly evaluate model performance in multi-class tasks, Accuracy, Macro-Precision, Macro-Recall, and Macro-F1 are adopted. All metrics are computed via macro-averaging to ensure equal class weighting and mitigate class imbalance.

3.3 MSA-ConvNeXt Network

The proposed MSA-ConvNeXt model adopts a four-stage progressive feature extraction framework, as shown in Fig. 2. The first three stages are equipped with the Multi-scale Convolution Attention Block (MCAB), which consists of a Multi-Scale Local Feature Fusion (MSLFF) module and a Convolutional Block Attention Module (CBAM). This design is intended to address the characteristic pattern in structural images of doped two-dimensional nanomaterials, where local perturbations around dopant sites are pronounced and their spatial influence gradually extends outward. Specifically, convolutional branches with different scales are employed to characterize local structural information over different spatial ranges, while the attention mechanism is introduced to enhance the representation of key responses in the fused multi-scale features. In the fourth stage, a VGG-Swin (Visual Geometry Group-Swin Transformer) architecture is introduced for global feature modeling. Local features are first strengthened by lightweight VGG-style

convolutions, and cross-region information interaction is then achieved through the Transformer, thereby capturing long-range structural dependencies and global structural correlations.

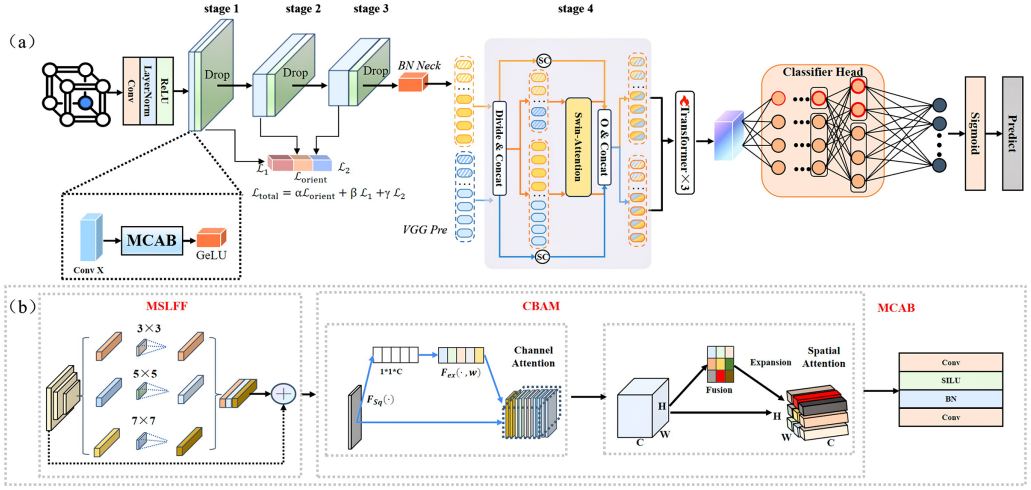


Figure 2: Overall architecture of the MSA-ConvNeXt model.

3.4 Multi-Scale Convolution and Attention Mechanism Module

To address the problem that structural images of doped two-dimensional nanomaterials exhibit local perturbations at multiple scales and that a single receptive field cannot effectively capture structural features over different spatial ranges, this study designs the MCAB. The block consists of the MSLFF and the CBAM. By combining multi-scale feature extraction with attention guidance, MCAB emphasizes the characterization of local perturbations around dopant sites and their spatial diffusion patterns, thereby enhancing the model's ability to represent local structural information.

Specifically, MSLFF employs three parallel depthwise separable convolution branches with kernel sizes of 3×3, 5×5, and 7×7 to extract structural features over different spatial receptive ranges, enabling the network to capture both local perturbations and broader structural information within the same layer.

$$\boldsymbol{F}_k = \boldsymbol{W}_k * \boldsymbol{X}, \quad k \in \{3,5,7\} \tag{1}$$

To enable the network to adaptively select receptive fields, learnable parameters $\alpha = [\alpha_3, \alpha_5, \alpha_7]$ are introduced, and the fusion weights for each branch are obtained via the softmax function:

$$w_k = \frac{e^{\alpha_k}}{\sum_{j\in\{3,5,7\}} e^{\alpha_j}} \tag{2}$$

The final multi-scale convolutional output is obtained by weighted fusion of the three branches:

$$\boldsymbol{F}_{\mathrm{MSLFF}} = w_3\boldsymbol{F}_3 + w_5\boldsymbol{F}_5 + w_7\boldsymbol{F}_7 \tag{3}$$

After multi-scale feature extraction, channel attention and spatial attention mechanisms are introduced to adaptively enhance the fused features. The channel attention module extracts channel-wise response information through global average pooling and max pooling, and employs a multilayer perceptron to model nonlinear dependencies among channels, thereby generating channel-level weights and strengthening feature responses related to magnetic variations. The spatial attention module, in turn, applies average

pooling and max pooling along the channel dimension, followed by a 7 × 7 convolution to generate the spatial attention weight $\boldsymbol{M}_s$, thereby highlighting critical structural regions.

$$\boldsymbol{z}_{\text{avg},c} = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{X}_c(i,j) \tag{4}$$

$$\boldsymbol{z}_{\max,c} = \max_{i,j}\boldsymbol{X}_c(i,j) \tag{5}$$

$$\boldsymbol{F} = \left[\boldsymbol{F}_{\text{avg}};\boldsymbol{F}_{\max}\right] \in \mathbb{R}^{2\times H\times W} \tag{6}$$

$$\boldsymbol{M}_s = \sigma\left(\text{Conv}_{7\times 7}(\boldsymbol{F})\right) \in \mathbb{R}^{1\times H\times W} \tag{7}$$

The cascaded design of the multi-scale depthwise convolution module and the CBAM attention mechanism establishes a collaborative modeling scheme across structural hierarchies and spatial dimensions, thereby enhancing the model's ability to represent dopant-site information.

3.5 VGG–Swin Architecture

During the global structural modeling stage, the network adopts a VGG-Swin architecture to achieve collaborative optimization of local detail enhancement and high-level relational modeling. This module first applies lightweight convolutions to enhance local information in the input features, and then employs a self-attention mechanism to model cross-region dependencies, thereby improving the model's ability to jointly represent local perturbations and global structural correlations in doped two-dimensional nanomaterials. Specifically, a 3 × 3 convolution is first used to extract local structural information from the input features. Subsequently, to ensure spatial consistency between the enhanced features and the original window features, adaptive average pooling is applied to the output of the convolutional branch.

$$\boldsymbol{X}_e = \text{AdaptiveAvgPool}\left(\boldsymbol{X}_{\text{VGG}}\right) \tag{8}$$

On this basis, the aligned locally enhanced features are residually fused with the original input features. The fused features are then fed into a self-attention module to model dependencies across spatial regions. Finally, the features are further updated through residual connections and a feed-forward network.

$$\boldsymbol{A} = \text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\top}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{9}$$

$$\boldsymbol{Y} = \left(\tilde{\boldsymbol{X}} + \boldsymbol{A}\right) + \text{MLP}\left(\tilde{\boldsymbol{X}} + \boldsymbol{A}\right) \tag{10}$$

This architecture enables the model to learn more stable and discriminative high-level feature representations by effectively integrating locally enhanced information with global relational information.

3.6 Definition of the Loss Function

During model training, cross-entropy loss with label smoothing was adopted as the optimization objective. For the magnetic property prediction task of doped two-dimensional nanomaterials, the magnetic categories of samples were obtained by discretizing the magnetic moments of materials into different value ranges. Since the magnetic response of a material is jointly affected by factors such as dopant species, local coordination environment, and structural distortion, different magnetic categories may exhibit a certain

degree of similarity in the feature space. Therefore, a label smoothing strategy was introduced to improve training stability and enhance the model's robust discrimination among different magnetic categories.

$$L = -(1-\varepsilon)\log(p_y) - \frac{\varepsilon}{C}\sum_{i=1}^{C}\log(p_i) \tag{11}$$

where ε denotes the smoothing factor, C is the total number of classes, p_i is the predicted probability for the ith class, and y is the ground-truth class.

4 Results

4.1 Analysis of Experimental Results

4.1.1 Comparative Results and Discussion

This study selects the baseline model and several deep learning models for comparison, and the results are presented in Table 3.

Table 3: Training results of magnetic property prediction for doped two-dimensional nanomaterials. Best results are in **bold**, second best are underlined.

Model	Optimization	Loss	Precision	Recall	F1-Score	ACC
ConvNeXt	Momentum	1.0482	0.6323	0.5680	0.5919	0.8147
	SGD	1.0369	0.6626	0.5746	0.6067	0.8115
	RMSProp	1.1412	0.6495	0.5549	0.5818	0.7519
	AdamW	1.0482	0.6683	0.6123	0.6362	0.8321
DenseNet	Momentum	0.6733	0.7273	0.7273	0.7273	0.8973
	SGD	0.7192	0.6909	0.6515	0.6706	0.8678
	RMSProp	0.6238	0.6458	0.6212	0.6333	0.8337
	AdamW	0.6772	0.6713	0.6704	0.6619	0.8581
ResNet50	Momentum	0.8168	0.6818	0.6212	0.6501	0.8683
	SGD	0.5821	0.5802	0.5606	0.5702	0.8236
	RMSProp	1.0696	0.6989	0.6515	0.6744	0.8727
	AdamW	0.8487	0.7612	0.7799	0.7675	0.8635
Res2Net	Momentum	0.5775	0.6909	0.6515	0.6706	0.8616
	SGD	0.7089	0.2802	0.3333	0.3045	0.7622
	RMSProp	0.5476	0.6818	0.6818	0.6818	0.8794
	AdamW	0.8215	0.7069	0.7048	0.7055	0.8862
ResNeXt	Momentum	1.0132	0.7045	0.6818	0.6930	0.8738
	SGD	1.0387	0.6515	0.6515	0.6515	0.8515
	RMSProp	0.8795	0.7045	0.6818	0.6930	0.8895
	AdamW	0.8183	0.6978	0.7188	0.7078	0.8917
Swin-Transformer	Momentum	1.6550	0.0505	0.0909	0.0649	0.4095
	SGD	1.6481	0.0505	0.0909	0.0649	0.4051
	RMSProp	1.6611	0.0505	0.0909	0.0649	0.4051
	AdamW	1.6192	0.0505	0.0909	0.0649	0.4051
Swin-ResNet	Momentum	1.1695	0.7907	0.7972	0.7893	0.8945
	SGD	1.0875	0.8750	0.8848	0.8552	0.9001
	RMSProp	1.1447	0.8741	0.8708	0.8578	0.8793
	AdamW	1.1065	0.7116	0.7114	0.7093	0.8982

(Continued)

Table 3 (continued)

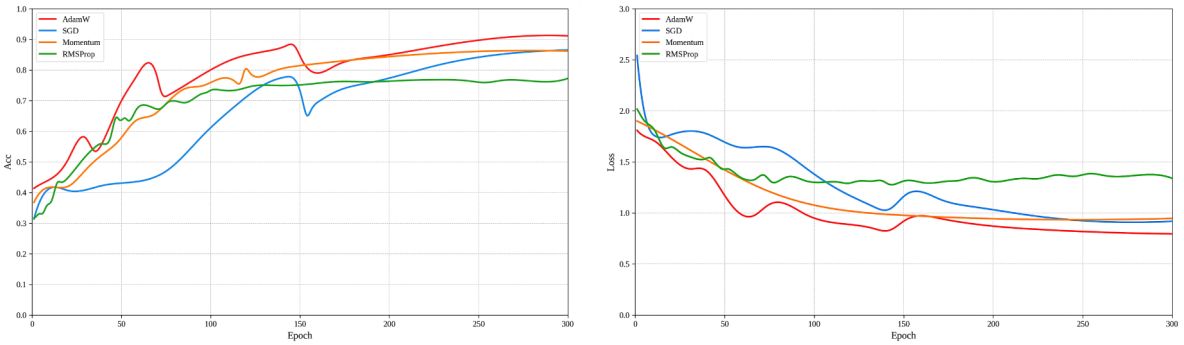
Model	Optimization	Loss	Precision	Recall	F1-Score	ACC
MSA-ConvNeXt (ours)	Momentum	0.9340	0.7450	0.7137	0.7276	0.8689
	SGD	0.9081	0.7352	0.8027	0.7544	0.8678
	RMSProp	1.1939	0.6635	0.5798	0.6129	0.7811
	AdamW	0.8008	0.8636	0.9202	0.8800	**0.9166**

Experimental results demonstrate that conventional convolutional neural networks achieve reasonable performance under different optimizer settings. Compared with Swin–ResNet50, the proposed model constructs a task-specific local–global collaborative modeling paradigm to capture the coexistence of local structural perturbations and long-range structural correlations in doped two-dimensional nanomaterial magnetic property prediction. Swin–ResNet50 enhances local feature extraction through a convolutional backbone and models cross-region dependencies via the Swin Transformer, which alleviates the limitation of pure Transformer models in capturing fine-grained local structures. However, its local representations still rely on fixed-scale convolutional kernels, lacking explicit modeling of structural perturbations around doping sites and their cross-scale propagation, and it does not perform targeted feature recalibration on critical channels and spatial regions.

In contrast, MSA-ConvNeXt introduces a multi-scale convolutional attention module to adaptively capture local structural features at different scales. In the global modeling stage, a VGG-Swin architecture is incorporated to enhance cross-region dependency modeling, enabling global relationships to be established upon more discriminative local representations. Under different optimizer settings, MSA-ConvNeXt consistently achieves strong performance, reaching the highest accuracy of 0.9166 with the AdamW optimizer and outperforming all comparative models. These results demonstrate the superior overall performance of the proposed method for magnetic property prediction in doped two-dimensional nanomaterials.

4.1.2 Analysis of Accuracy and Loss Curves under Different Optimizers

To systematically evaluate the influence of different optimization algorithms on the training dynamics of MSA-ConvNeXt, this section presents the validation accuracy and loss curves of four optimizers, namely AdamW, SGD, Momentum, and RMSProp, over the training epochs, as shown in Fig. 3. The vertical axis represents the validation accuracy and loss, while the horizontal axis denotes the training epochs.



(**a**) Validation ACC Curves under Four Optimizers (**b**) Validation Loss Curves under Four Optimizers

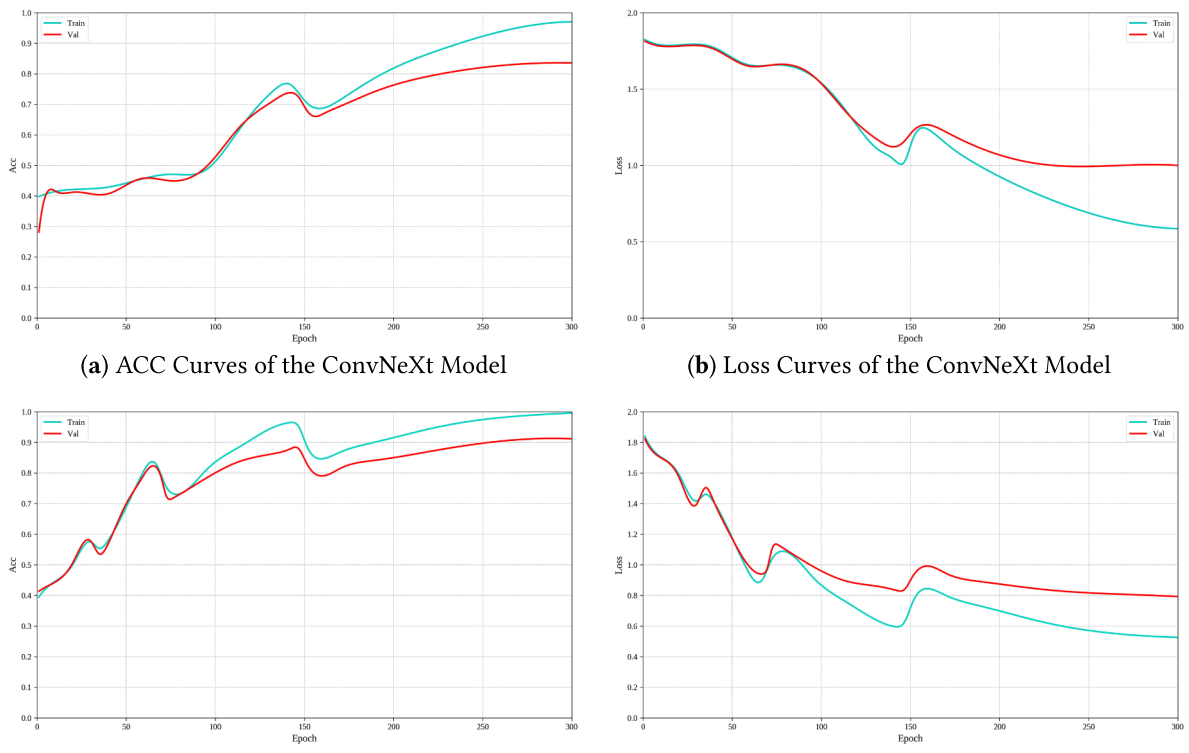
Figure 3: Validation ACC and loss curves of MSA-ConvNeXt.

MSA-ConvNeXt introduces parallel multi-scale convolution branches and attention-based feature recalibration, which creates strong collaborative optimization demands across different layers and branches. AdamW effectively coordinates the learning of these modules through its adaptive update mechanism, enabling early integration of multi-scale features and faster attention to magnetism-related structural regions. Its faster accuracy improvement and loss reduction in the early stage indicate both strong convergence capability and better compatibility with the proposed multi-scale convolution–attention design.

In comparison, SGD shows slower ACC improvement and loss reduction in the early stage, suggesting limited update efficiency and a longer convergence process. Momentum exhibits relatively smooth curves, indicating that the momentum term helps alleviate local oscillations, although its final performance remains slightly lower than that of AdamW. RMSProp achieves rapid initial loss reduction but enters a plateau earlier in the middle and late stages, indicating weaker sustained optimization ability. Considering convergence speed, fluctuation characteristics, and final performance, AdamW was selected for subsequent experiments.

4.1.3 Comparison of the Convergence Characteristics of ConvNeXt and MSA-ConvNeXt

To further analyze the convergence behavior and generalization ability of MSA-ConvNeXt during training, this section presents the ACC and loss curves of the training and validation sets for both the baseline ConvNeXt model and MSA-ConvNeXt under the AdamW optimizer, as shown in Fig. 4.



(**a**) ACC Curves of the ConvNeXt Model (**b**) Loss Curves of the ConvNeXt Model

(**c**) ACC Curves of the MSA-ConvNeXt Model (**d**) Loss Curves of the MSA-ConvNeXt Model

Figure 4: Comparison of the ACC and loss curves of ConvNeXt and MSA-ConvNeXt.

The ACC curves show that MSA-ConvNeXt achieves faster convergence and higher final classification accuracy than the original ConvNeXt throughout training. In the early stage, the validation ACC of MSA-ConvNeXt increases more rapidly, indicating that the multi-scale convolution design helps capture local perturbations around dopant sites and their spatial diffusion patterns more effectively. Although both models

exhibit fluctuations during the learning-rate adjustment stage, MSA-ConvNeXt recovers more quickly. In the later stage, the gap between its training and validation curves remains within a reasonable range, with no obvious overfitting, suggesting good generalization ability.

The loss curves further demonstrate that MSA-ConvNeXt exhibits a smoother and faster downward trend, with both training and validation losses consistently lower than those of ConvNeXt. Overall, the multi-scale feature extraction, attention-based feature enhancement, and global dependency modeling jointly improve the model's suitability for magnetic property prediction of doped two-dimensional nanomaterials.

4.2 Ablation Study of the MCAB Module

To systematically evaluate the functional position of the MCAB module within the network and its impact on performance, this section introduces the MCAB module at different stages of the network while maintaining the fourth-stage VGG-Swin global modeling structure unchanged. An ablation analysis is conducted through stage-wise and cross-stage combinations, with the results presented in Table 4.

Table 4: Ablation study results of the MCAB module. Best results are in **bold**, second best are underlined.

Methods	Accuracy
ConvNeXt	0.8321
Stage 1	0.8537
Stage 2	0.8635
Stage 3	0.8743
Stage 1 + 2	0.8917
Stage 2 + 3	0.9047
MSA-ConvNeXt	**0.9166**

As can be observed from the experimental results, after introducing the MCAB module in any single stage, the model accuracy improves compared to the baseline ConvNeXt (0.8321). When the MCAB is introduced solely in the first stage, the accuracy increases to 0.8537; when introduced in the second and third stages separately, the accuracy further improves to 0.8635 and 0.8743, respectively, demonstrating a trend of gradual performance enhancement as the network hierarchy deepens. This phenomenon indicates that with the elevation of the semantic level of features, the modeling capability of multi-scale convolution and attention mechanisms for structural information progressively strengthens. Under cross-stage combination configurations, the performance improvement is more pronounced. The Stage 1 + 2 combination achieves 0.8917, while the Stage 2 + 3 combination further elevates to 0.9047, which is approaching the performance of the complete model. This suggests that middle- and high-level features play a more critical role in semantic representation for the magnetic property prediction task of doped two-dimensional nanomaterials. Multi-scale structural modeling and attention recalibration in the middle and high-level stages can more effectively enhance the structural representations relevant to magnetic discrimination. When the MCAB acts simultaneously on the shallow, middle, and deep stages, the model achieves optimal performance (0.9166). This result demonstrates that although a single-stage introduction can yield stable gains, a multi-stage synergistic configuration enables more comprehensive multi-scale structural modeling. This allows the characterization of local structural perturbations and feature recalibration to form complementarities across different levels, thereby further elevating overall prediction performance and model stability.

4.3 Ablation Study of the VGG-Swin Architecture

To verify the role of the VGG-Swin architecture in global structural modeling, this section conducts an ablation analysis on the innovations in the fourth stage of the model under the condition of keeping the convolutional backbone structure unchanged. It compares the impact of different module configurations on model performance, and the results are shown in Table 5. Here, Model1 indicates that only the fourth stage of the baseline model is replaced with the VGG-Swin architecture, while Model2 indicates that the MCAB module is enabled in the first three stages and the ConvNeXt block in the fourth stage is retained.

Table 5: Ablation results of the VGG-Swin architecture. Best results are in **bold**, second best are underlined.

Methods	Accuracy
ConvNeXt	0.8321
Model1	0.8657
Model2	0.9026
MSA-ConvNeXt	**0.9166**

As can be observed from the results in the table, the accuracy of the baseline ConvNeXt model is 0.8321. When the VGG-Swin architecture is introduced only in the fourth stage (Model1), the model accuracy increases to 0.8657, indicating that the global relationship modeling mechanism based on window self-attention can to some extent enhance the model's representational capacity for long-range structural correlations. When the MCAB module is introduced only in the first three stages (Model2), the accuracy improves to 0.9026, demonstrating that multi-scale convolution and the CBAM attention mechanism play a more significant role in modeling local structural perturbations and can effectively enhance fine-grained structural representations related to dopant sites. In the complete model (MSA-ConvNeXt), when both the MCAB and VGG-Swin architecture are combined simultaneously, the model accuracy further improves to 0.9166, achieving optimal performance.

These results demonstrate that local multi-scale structural modeling and the global self-attention mechanism are complementary in function. Their synergistic interaction constructs a more hierarchical and discriminative structural representation, thereby elevating the overall magnetic property prediction performance.

4.4 Ablation Study of Model Depth

To systematically verify the impact of depth allocation strategies at different stages on model performance, this study adopts four representative depth variants based on the ConvNeXt backbone, including an overall shallow structure (V1:Uniform-2), a uniformly deepened structure (V2:Uniform-3), a mid-stage enhanced structure (V3:Stage-3-Enhanced), and the finally adopted MSA-ConvNeXt. By comparatively analyzing the performance differences corresponding to various depth allocation methods, the importance of middle- and high-level semantic modeling for predicting the magnetic properties of doped two-dimensional nanomaterials can be further revealed, providing a theoretical basis for the design of the final backbone structure. The results are presented in Table 6.

As can be seen from the results, with the increase in the overall depth of the network, the model performance exhibits a steady upward trend. Compared to 0.8819 for the shallow structure V1, the accuracy of V2, which uniformly deepens all stages, improves to 0.9014, indicating that increasing network capacity can enhance feature representation capability. When more depth is further concentrated in the third stage to form

V3, the accuracy improves to 0.9084, indicating that compared to uniform deepening, reinforcing middle- and high-level features is more conducive to elevating semantic modeling capabilities in magnetic property prediction. Middle-level features aid in aggregating local structural patterns and meso-scale configuration information, while high-level features are more directly associated with the final magnetic discrimination decision. Finally, building upon the reinforcement of middle-level and high-level depth and integrating the multi-scale convolution attention module with the VGG-Swin global modeling mechanism, MSA-ConvNeXt achieves the optimal performance of 0.9166. This demonstrates that combining staged depth allocation with local-global synergistic modeling can further enhance the model's representational capacity for the magnetic structural features of doped two-dimensional nanomaterials.

Table 6: Impact of different depth allocations across stages on model performance. Best results are in **bold**, second best are underlined.

Model	Depths Configuration	Accuracy
V1:Uniform-2	[2, 2, 2, 2]	0.8819
V2:Uniform-3	[3, 3, 3, 3]	0.9014
V3:Stage-3-Enhanced	[2, 2, 6, 2]	0.9084
MSA-ConvNeXt	**[3, 3, 9, 3]**	**0.9166**

4.5 Visualization Analysis of Model Attention Mechanism

To further analyze the discriminative basis and interpretability of the proposed model, this study conducts a visualization analysis of the attention responses of MSA-ConvNeXt. As illustrated in Fig. 5, the input structural images of doped two-dimensional (2D) nanomaterials, their corresponding attention heatmaps, and the overlay results are displayed. It can be observed that the high-response regions of the model are primarily concentrated in the vicinity of dopant sites and their surrounding key structural units rather than the background regions. This indicates that the model can adaptively focus on local structural areas that potentially contribute to magnetic discrimination.

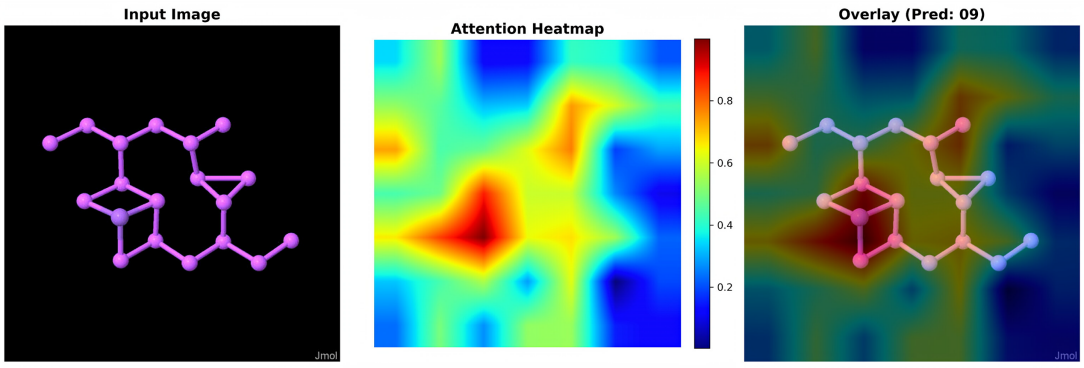


Figure 5: Visualization of the model's attention mechanism.

From a structural perspective, these visualization results reflect the reinforcement effect of the MCAB module on local structural perturbations in the shallow and middle stages, while also demonstrating the integration capability of the VGG-Swin architecture for cross-regional structural relationships in the deep stages. Overall, the distribution of attention responses is consistent with the aforementioned ablation experiments and quantitative comparison results, further supporting the rationality of multi-scale feature

modeling and the global self-attention mechanism in the magnetic property prediction task of doped 2D nanomaterials from the perspective of interpretability.

5 Conclusion

Magnetic property prediction of doped two-dimensional nanomaterials is important for the screening and design of novel magnetic materials. To address the limitation of conventional convolutional neural networks in jointly modeling local structural perturbations around dopant sites and long-range structural correlations, this study proposes the MSA-ConvNeXt model. Experimental results show that the proposed model achieves strong classification performance, reaching an accuracy of 91.66% and outperforming the selected comparison models overall. Ablation studies further confirm the effectiveness of multi-scale feature extraction, attention-based recalibration, and global relational modeling.

Nevertheless, several limitations remain. The current experiments are mainly conducted on a single public dataset, and the generalization ability of the model still needs to be validated on more diverse data sources and in more complex scenarios. In addition, model size, computational complexity, and inference efficiency have not been systematically evaluated. Future work will focus on multi-source validation, efficiency assessment, and physics-informed interpretability analysis to further improve the reliability and applicability of the proposed model.

Acknowledgement: None.

Funding Statement: This work was supported by the Heilongjiang Provincial Discipline Innovation Project (No. LJGXCG2024-F10).

Author Contributions: The authors confirm their contribution to the paper as follows: study conception and design: Yuxuan Feng and Lili Liang; analysis and interpretation of results: Yuxuan Feng and Guanglu Sun; draft manuscript preparation: Yuxuan Feng and Yanrui Wei. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The source code supporting this study is available at https://github.com/qingke-369/MSA-ConvNeXt.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chang C, Chen W, Chen Y, Chen Y, Chen Y, Ding F, et al. Recent progress on two-dimensional materials. Wuli Huaxue Xuebao/Acta-Phys-Chim Sin. 2021;37(12):2108017.
2. Khan K, Tareen AK, Aslam M, Wang R, Zhang Y, Mahmood A, et al. Recent developments in emerging two-dimensional materials and their applications. J Mater Chem C. 2020;8(2):387–440. doi:10.1039/c9tc04187g.
3. Zhang S, Xu R, Luo N, Zou X. Two-dimensional magnetic materials: structures, properties and external controls. Nanoscale. 2021;13(3):1398–424.
4. Lin Z, Carvalho BR, Kahn E, Lv R, Rao R, Terrones H, et al. Defect engineering of two-dimensional transition metal dichalcogenides. 2D Mater. 2016;3(2):022002.
5. Chen W, Chen Q, Zhang J, Zhou L, Tang W, Wang Z, et al. Electronic and magnetic properties of transition-metal-doped monolayer B_2S_2 within GGA + U framework. RSC Adv. 2024;14(5):3390–9. doi:10.1039/d3ra08472h.
6. Chen X, Shao YT, Chen R, Susarla S, Hogan T, He Y, et al. Pervasive beyond room-temperature ferromagnetism in a doped van der Waals magnet. Phys Rev Lett. 2022;128(21):217203. doi:10.1103/physrevlett.128.217203.

7. An Z, Su Y, Ni S, Guan Z. Carrier doping modulates 2D intrinsic ferromagnetic $Mn_2Ge_2Te_6$ monolayer, high Curie temperature, large magnetic crystal anisotropy. J Phys Chem C. 2022;126(27):11330–40. doi:10.1021/acs.jpcc.2c02087.
8. Schleder GR, Padilha AC, Acosta CM, Costa M, Fazzio A. From DFT to machine learning: recent approaches to materials science—a review. J Phys Mater. 2019;2(3):032001.
9. Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED. Scaling deep learning for materials discovery. Nature. 2023;624(7990):80–5. doi:10.1038/s41586-023-06735-9.
10. Zhong D, Seyler KL, Linpeng X, Cheng R, Sivadas N, Huang B, et al. Van der Waals engineering of ferromagnetic semiconductor heterostructures for spin and valleytronics. Sci Adv. 2017;3(5):e1603113. doi:10.1126/sciadv.1603113.
11. Xie T, Grossman JC. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. Phys Rev Lett. 2018;120(14):145301. doi:10.1103/physrevlett.120.145301.
12. Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph networks as a universal machine learning framework for molecules and crystals. Chem Mater. 2019;31(9):3564–72. doi:10.1021/acs.chemmater.9b01294.
13. Khan A, Ghorbanian V, Lowther D. Deep learning for magnetic field estimation. IEEE Trans Magn. 2019;55(6):1–4. doi:10.1109/tmag.2019.2899304.
14. Kwon HY, Yoon H, Lee C, Chen G, Liu K, Schmid A, et al. Magnetic Hamiltonian parameter estimation using deep learning techniques. Sci Adv. 2020;6(39):eabb0872. doi:10.1126/sciadv.abb0872.
15. Pollok S, Bjørk R, Jørgensen PS. Inverse design of magnetic fields using deep learning. IEEE Trans Magn. 2021;57(7):1–4. doi:10.1109/tmag.2021.3082431.
16. Li H, Tang Z, Gong X, Zou N, Duan W, Xu Y. Deep-learning electronic-structure calculation of magnetic superstructures. Nat Comput Sci. 2023;3(4):321–7. doi:10.1038/s43588-023-00424-3.
17. Zhang Y, Zhou C, Liang F, Liu G, Zhu J. Magnetic prediction of doped two-dimensional nanomaterials based on Swin-ResNet. Coatings. 2024;14(10):1271. doi:10.3390/coatings14101271.
18. Ozdemir B, Sermet F, Pacal I. Attention-enhanced ConvNeXt for accurate, efficient, and interpretable crack detection. Expert Syst Appl. 2025;129165. doi:10.1016/j.eswa.2025.129165.
19. Wang F, Li H, Chao W, Zhuo Z, Ji Y, Peng C, et al. E-convNeXt: a lightweight and efficient convNeXt variant with cross-stage partial connections. arXiv:250820955. 2025.
20. Schütt KT, Sauceda HE, Kindermans PJ, Tkatchenko A, Müller KR. SchNet—a deep learning architecture for molecules and materials. J Chem Phys. 2018;148(24):058301. doi:10.1063/1.5019779.
21. Louis SY, Zhao Y, Nasiri A, Wang X, Song Y, Liu F, et al. Graph convolutional neural networks with global attention for improved materials property prediction. Phys Chem Chem Phys. 2020;22(32):18141–8. doi:10.1039/d0cp01474e.
22. Choudhary K, DeCost B. Atomistic line graph neural network for improved materials property predictions. npj Comput Mater. 2021;7(1):185. doi:10.1038/s41524-021-00650-1.
23. McNutt AT, Francoeur P, Aggarwal R, Masuda T, Meli R, Ragoza M, et al. GNINA 1.0: molecular docking with deep learning. J Cheminform. 2021;13(1):43.
24. Takiguchi Y, Nakane D, Akitsu T. The prediction of single-molecule magnet properties via deep learning. IUCrJ. 2024;11(2):182–9. doi:10.1107/s2052252524000770.
25. Kovacs A, Exl L, Kornell A, Fischbacher J, Hovorka M, Gusenbauer M, et al. Image-based prediction and optimization of hysteresis properties of nanocrystalline permanent magnets using deep learning. J Magn Magn Mater. 2024;596:171937. doi:10.2139/ssrn.4541646.
26. Talapatra A, Gajera U, P SP, Arout Chelvane J, Mohanty JR. Understanding the magnetic microstructure through experiments and machine learning algorithms. ACS Appl Mater Interfaces. 2022;14(44):50318–30. doi:10.1021/acsami.2c12848.
27. Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A convnet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022 Jun 18–24; New Orleans, LA, USA. p. 11976–86.
28. Davidsson J, Bertoldo F, Thygesen KS, Armiento R. Absorption versus adsorption: high-throughput computation of impurities in 2D materials. npj 2D Mater Appl. 2023;7(1):26.