RESEARCH ARTICLE | APRIL 01 2025

Machine learning for predicting physical parameters of atom-vacancy defects from low-frequency noise in few-atom layer MoS_2

Y. Nonaka ; K. Takaki; Y. Kobayashi; J. Haruyama

AIP Advances 15, 045202 (2025)
https://doi.org/10.1063/5.0254351

Machine learning for predicting physical parameters of atom-vacancy defects from low-frequency noise in few-atom layer MoS_2

Cite as: AIP Advances 15, 045202 (2025); doi: 10.1063/5.0254351
Submitted: 14 January 2025 • Accepted: 14 March 2025 •
Published Online: 1 April 2025

Y. Nonaka,[1] K. Takaki,[1] Y. Kobayashi,[1] and J. Haruyama[1,2,a)]

AFFILIATIONS

[1] Faculty of Science and Engineering, Aoyama Gakuin University, 5-10-1 Fuchinobe, Sagamihara, Kanagawa 252-5258, Japan
[2] Institute for Industrial Sciences, The University of Tokyo, 4-6-1 Komaba Meguro-ku, Tokyo 153-8505, Japan

[a)] **Author to whom correspondence should be addressed:** J-haru@ee.aoyama.ac.jp

ABSTRACT

The rapid advancement of deep learning (DL) has significantly expanded its application across various fields, including physics and materials science. DL facilitates the prediction of physical parameters of materials without the need for time-consuming and labor-intensive experiments. Atom-vacancy defects, present in various materials responsible for intriguing physical phenomena, exhibit diverse physical parameters and are a prime example of such applications. These tendencies are even more pronounced in two-dimensional (2D) (atomically) thin van der Waals layers. However, the correlation between different types of atomic defects and their properties remains an area of active exploration. Herein, we propose an innovative DL model and predict the structural parameters of atom-vacancy defects solely from low-frequency noises, arising from stochastic capture and emission of charges through them, in few-atom layer semiconductors molybdenum disulfides (MoS_2). The predicted parameters show good agreement with experimental results. This methodology enables the accurate extraction of structural parameters of various materials, significantly reducing error rates by suppressing excess learning and demonstrating the high potential of DL.

The recent rapid advancements in deep learning (DL) have demonstrated its potential across various domains, including natural language processing, image generation, and the prediction of materials' physical properties.[1–7] In particular, approaches such as Graph Neural Networks (GNNs) for molecular-structure representation learning and diffusion models for novel material design have enabled not only the discovery of innovative physical phenomena and materials but also the estimation of physical parameters of diverse materials, without the need for complicated, time-consuming, and labor-intensive experiments. Furthermore, approaches like Physics-Informed Neural Networks (PINNs), which incorporate physical laws into their loss functions, have garnered significant attention for their ability to ensure physical consistency in predictions.[8]

Atom-vacancy defects in various semiconductors give rise to numerous intriguing physical phenomena and are characterized by various physical parameters,[9] such as activation energy (E_a), capture cross section (σ), trap state density per unit volume (N_T), and electron velocity (v). These parameters are typically determined through time-consuming experimental techniques, such as photoluminescence, high-resolution transmission electron microscopy (HRTEM), deep-level transient spectroscopy (DLTS), and observation of low-frequency (f) (LF) noises. Since these defects can hinder the performance of electronic and photonic devices in certain cases, it would be highly advantageous to predict the physical properties associated with atom-vacancy defects without relying on such labor-intensive measurements.

The influence of atom-vacancy defects becomes more pronounced in two-dimensional (2D) atomically thin van der Waals (vdW) semiconductors [e.g., the transition metal dichalcogenide (TMDC) family such as molybdenum disulfide (MoS_2)], as even few atom-vacancy defects can significantly alter those physical

properties. Various types of atom-vacancy defects have been identified in these materials {e.g., sulfur (S) vacancies in MoS_2 [Fig. 1(a)]}.[9–20] These defects introduce numerous mid-gap states with different E_a values and electric behaviors. Among them, S vacancies have the lowest formation energy and are the most chemically and thermally stable in MoS_2 layers. However, the correlations with various kinds of atom defects are still under exploration.

Our recent work has revealed that the stochastic capture and emission of charge carrier through such defects lead to large hysteresis loops in the statical electric properties of field-effect transistors (FETs) using few-atom-layer MoS_2 (Fig. 1).[9] The temperature dependence and HRTEM observations indicated that these effects originate from pairs of neighboring in-plane S vacancy defects [Fig. 1(a)], which are predominantly located near the interface between the MoS_2 flake and the substrate, with $E_a \sim 0.35$ eV. LF noise measurements clarified a significant f shift in two $1/f^2$-dependent regimes, implying stochastic behaviors of electric charges through the S vacancy pairs, with high-speed charge (or spin) transitions across low kinetic energy barriers between narrow discrete states.

However, such measurements for individual 2D layer materials are time-consuming and labor-intensive. It is therefore essential to accurately predict the structure parameters of materials using DL without relying on these measurements. Herein, we develop an innovative DL model and predict the structural parameters (E_a) of S atom-vacancy defects solely from simple LF noise [i.e., power spectral density (PSD)] measurements in few-atom-layer MoS_2.[9–11] The predicted parameters indicate good agreement with the experimental results, confirming the high effectiveness of the developed model.

In the present research, vdW-fabricated few-layer MoS_2 flakes (of a thickness of ~6 nm) are employed with a current flow area of ~5 × 10 μm^2 between two Au/Ti electrodes [Figs. 1(b) and 1(c)]. Few-layer MoS_2 flakes are affixed onto pre-fabricated two Au/Ti electrodes on the SiO_2/Si substrate through mechanical exfoliation of bulk MoS_2 (purchased from HQ Graphene Co.) using Scotch tape and poly-dimethyl siloxane (PDMS) sheets, utilizing a 2D heterostructure transfer system (HQ graphene Co). Prior to exfoliation, the SiO_2/Si substrate surface is plasma-cleaned for 30 minutes under high vacuum. The thickness of ~6 nm is confirmed by atomic force microscopy [Fig. 1(d)] and Raman spectroscopy. After flake transfer, each sample is covered by a high-quality hBN monolayer. All processes are conducted inside a glovebox with an Ar-gas atmosphere. Dynamic behaviors (LF noises and PSD) associated with the S-vacancy pairs are measured at room temperature (T) using the set shown in Fig. 1(e), following the methodology described in Ref. 9, for the investigation of the present DL model.

The dataset for constructing the DL model (i.e., PSD values and structure parameters) is generated employing the structural parameters of MoS_2 reported in Ref. 10. For building the DL model, a total of 40 000 data points, which consist of 28 000 training samples, 4000 validation samples, and 8000 test samples, are employed. The PSD is calculated based on these parameters using the following two equations, covering the frequency range of 1–3000 Hz:

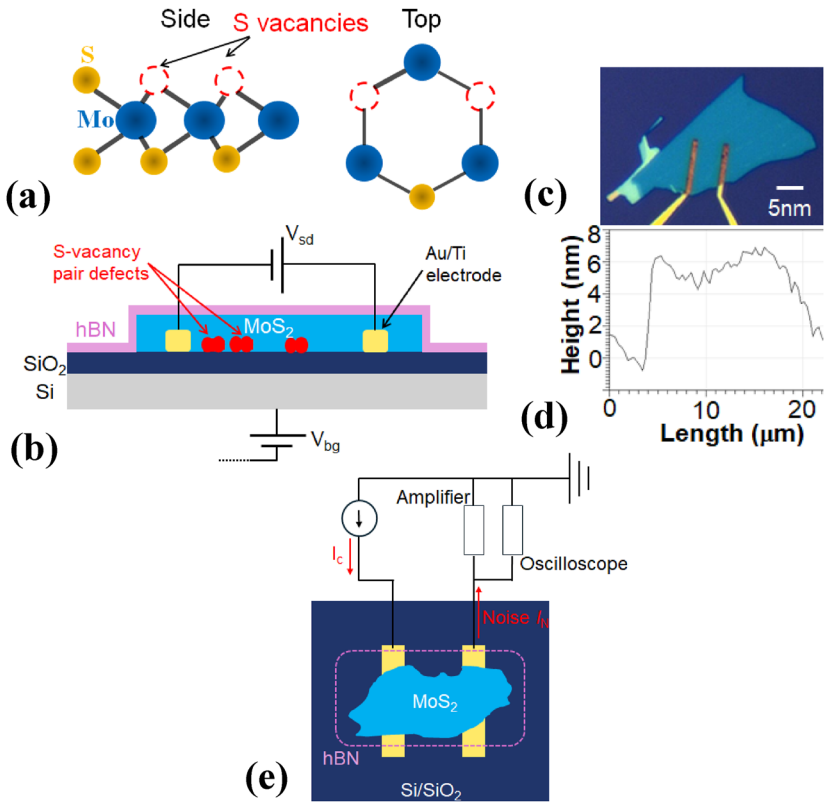

FIG. 1. (a) Schematic cross-sectional and top views of an S-vacancy pair in a mono-quantum layer of MoS_2. (b) Schematic views of the sample cross section used for LF noise measurements and DL modeling, highlighting the MoS_2/SiO_2 junction interface, including the structure shown in (a). (c) Optical microscope image of the sample described in (b), showing a few-atom-layer MoS_2 flake attached to two pre-fabricated Au/Ti electrodes on a SiO_2/Si substrate. The flake is fabricated using vdW mechanical exfoliation of bulk MoS_2 with Scotch tape and a PDMS sheet transfer technique. (d) Cross-sectional AFM image of (c). (e) Schematic setup for LF noise measurements using the sample shown in (b). Noise current (I_N) and PSD are measured with an oscilloscope at room temperature, while varying the constant I_{sd} current (I_c) and V_{bg} in (b).

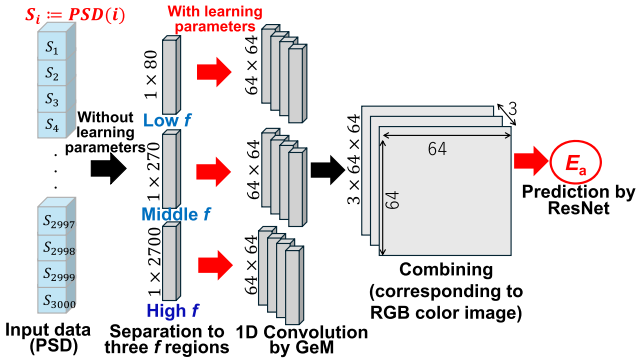

FIG. 2. Schematic illustrations depicting the construction of the proposed DL model.

$$S_I(f) = \frac{A_\alpha(I_{bias})I_{bias}^2}{1+\left(\frac{f}{f_\alpha}\right)^2} + = \frac{A_\beta(I_{bias})I_{bias}^2}{1+\left(\frac{f}{f_\beta}\right)^2} + C, \quad (1)$$

$$\tau(T) = \frac{1}{2\pi f(T)} = \frac{1}{\sigma^{(0)} \times \exp\left(\frac{-E}{k_B T}\right) \times N_T \times v}. \quad (2)$$

Equation (1) describes the PSD spectrum modeled using two Lorentzian functions, as provided in Ref. 10. In this equation, I_{bias} represents the constant current used for LF noise measurements, A_α and A_β are the amplitudes of the Lorentzian functions corresponding to widths f_α and f_β, respectively, and C denotes a flat background accounting for the measured system noise spectrum. Equation (2) is the Shockley–Read–Hall (SRH) statistics equation, which incorporates the structural parameters mentioned above. It describes the capture and emission of charge carriers through energy levels introduced by defects.

The proposed DL model is constructed as follows (Fig. 2):

(1) Input Data: PSD data, generated as described above (a 1D vector with 3000 points), are used as input.
(2) Segmentation: The input PSD data are divided into three f segments. This introduces an inductive bias based on the hypothesis that different f regions carry independent features.
(3) Feature Extraction: A 1D Convolutional Neural Network (CNN) with Generalized Mean Pooling (GeM) and the Swish activation function is applied to extract features from the segmented f-region data.
(4) Data Combination: The segmented data are combined into a 3D tensor, structurally analogous to the RGB components in image processing. This transformation facilitates the use of image-processing techniques that are widely adopted in DL applications.
(5) Final Prediction: The combined 3D tensor is fed into a ResNet-based model to predict E_a.

In the present research, GeM pooling methods are employed as the feature extraction layer, instead of standard pooling methods.

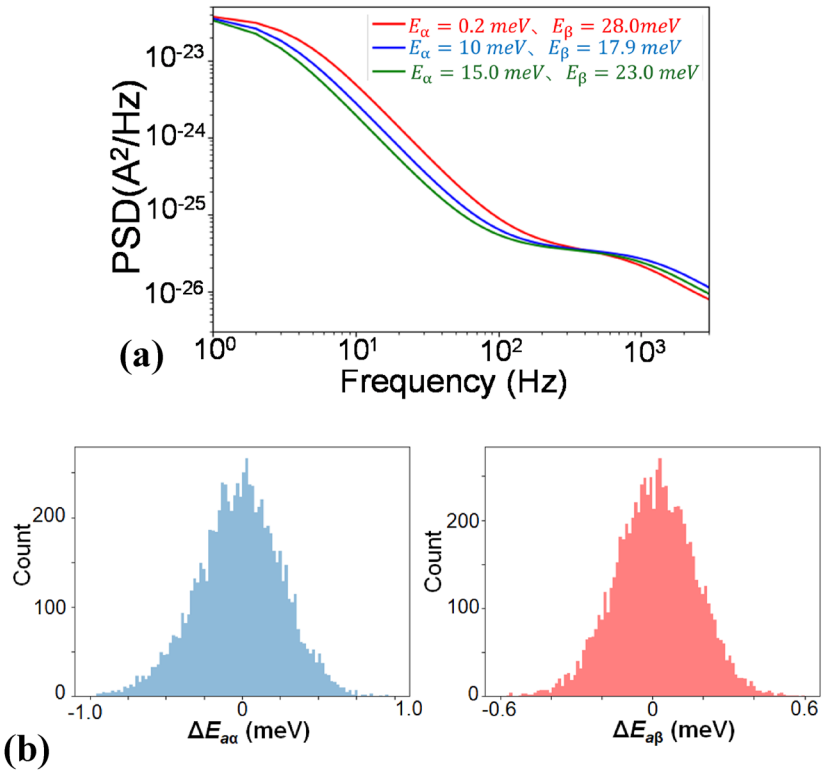

FIG. 3. (a) Predicted examples of PSD vs f relationships for three sets of $E_{a\alpha}$ and $E_{a\beta}$, obtained using the present model with the parameters listed in Table I. (b) Distribution of error values Δ for $E_{a\alpha}$ (left) and $E_{a\beta}$ (right), predicted from 8000 PSD datasets.

Standard pooling methods include average and maximum poolings, which need to be appropriately selected depending on the task.[21] Average pooling is advantageous for tasks that reflect global trends of the entire data (e.g., image classification), as it uniformly aggregates feature distributions across the entire input. On the other hand, maximum pooling extracts the most prominent features in local regions and is suitable for tasks where local information is vital (e.g., object detection). However, when predicting E_a from the PSD spectrum generated in the present MoS_2, there is no clear insight into which method would contribute more to improving prediction accuracy, except for Ref. 22. Therefore, we adopt the GeM pooling method, which is learnable and automatically allows us to optimize feature representations for E_a prediction.

Figure 3(a) illustrates the predicted examples of the PSD vs f relationships for three sets of $E_{a\alpha}$ and $E_{a\beta}$, which correspond to shallower and deeper E_a levels, respectively. These results are obtained using this DL model, with the parameters listed in Table I. $E_{a\alpha}$ and $E_{a\beta}$ values predicted from the PSD show good agreement with the correct solutions, validating the appropriateness of the constructed DL model. The error values for $E_{a\alpha}$ and $E_{a\beta}$ predicted from 8000 PSD datasets are shown in Fig. 3(b). They imply the mean absolute errors of $|\Delta E_{a\alpha}|$ = 0.2 meV and $|\Delta E_{a\beta}|$ = 0.1 meV, compared to the experimental values of $E_{a\alpha}$ ~10 meV and $E_{a\beta}$ ~23 meV, reaffirming the strong predictive relevance and accuracy of the developed model.

The DL model is applied to the PSD measurement results of our multiple MoS_2 flake samples obtained from Fig. 1(e) [Fig. 4(a)] to predict their E_a and N_T values. Distribution of the predicted results of $E_{a\alpha}$ and $E_{a\beta}$ pairs for three different N_T values is presented in Fig. 4(b). When the N_T value from Ref. 10 (~5 × 10^{11}/cm^2) is employed, the predicted $E_{a\alpha}$ values become negative. In contrast, as the N_T ratio increases slightly (by a factor of 1.1–1.3 times), the $E_{a\alpha}$ values shift to positive and reach the most relevant range ($E_{a\alpha}$ ~1.0–1.5 meV and $E_{a\beta}$ ~5–6 meV) at N_T ~ 1.1 ×, the value from Ref. 10. These results are in approximately good agreement with the previously reported experimental values of $E_{a\alpha}$ ~1.5 meV and $E_{a\beta}$ ~3 meV for N_T ~10^{12}/cm^2 in our earlier studies on MoS_2 flakes, which were confirmed through HRTEM, temperature-dependent resistance, and LF noise measurements.[9] This model enables accurate predictions of structure parameter values without the need for time-consuming and labor-intensive experimental measurements. Although the predicted $E_{a\beta}$ ~5–6 meV values are larger, this result implies the strong effectiveness of the proposed DL model.

One of the causes for this discrepancy between the predicted $E_{a\beta}$ and the experimental values (~3 meV) is attributable to an imbalance in the output scales of the problem.[23] Since the input data have more data points in the high-frequency region than in the low-frequency region, the DL model tends to underestimate $E_{a\alpha}$. To address this imbalance during training, a weight factor of 3.16 is applied to encourage more careful learning of $E_{a\alpha}$. This weight potentially causes larger values of $E_{a\beta}$. Adjusting the weight factor to 2.25 to equalize the scales of $E_{a\alpha}$ and $E_{a\beta}$, or implementing a relative error evaluation that weights each target by the inverse of its variance, may be beneficial for further improvement of the prediction accuracy.

We discuss additional advantages of the present DL model, which outperforms the conventional 1D CNN model in several important aspects. The best loss values are significantly improved, decreasing from 0.021 in the conventional 1D CNN model to 0.003 in the present model. For the conventional 1D CNN model [red curve in Fig. 5(a)], the test loss initially decreases slightly around epoch 0 (i.e., at early stages of learning) but then increases as the number of epochs grows, indicating typical excess-learning behavior. Moreover, the test loss diverges from the training loss (blue curve) and exhibits oscillatory behavior, highlighting unstable learning. In contrast, these issues are not observed in the present DL model. The test loss smoothly converges along with the training loss, achieving stable learning [Fig. 5(b)]. Furthermore, the present model uses ~5.5 times more parameters than the conventional model (i.e., 11 × 10^6 parameters [Fig. 5(b)] compared to 2 × 10^6 [Fig. 5(a)],

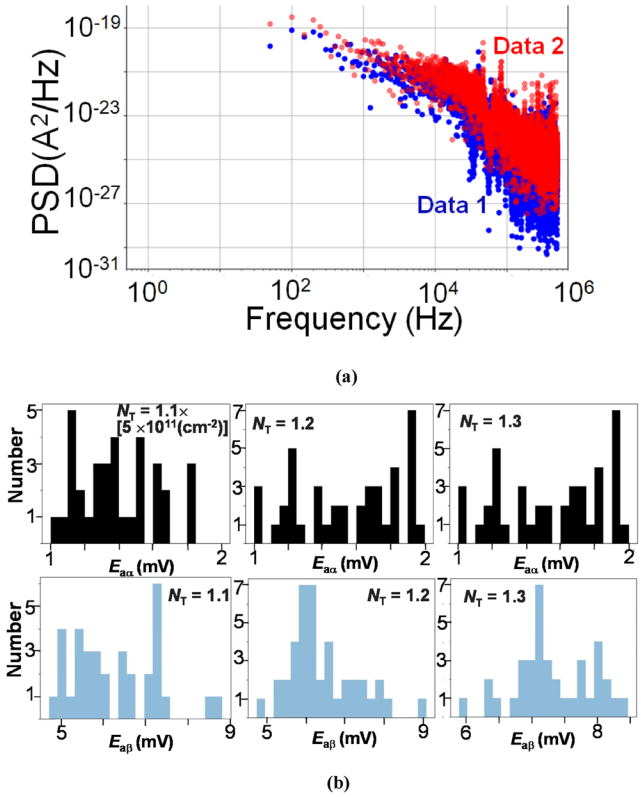

FIG. 4. (a) Examples of PSD measurement results for the two MoS_2 flake samples obtained from Fig. 1(e). (b) Distribution of predicted pairs of $E_{a\alpha}$ and $E_{a\beta}$ values for three different N_T values, including those shown in Fig. 4(a).

TABLE I. Structure parameters of defects used for the prediction shown in Fig. 3.

N_T	4×10^{11} (cm^2)
V	10^7 cm/s
k_B	8.6×10^{-5} eV/k
T	300 K
σ_α	10^{-23} cm^2
σ_β	$\sigma_\alpha \times 830$ cm^2

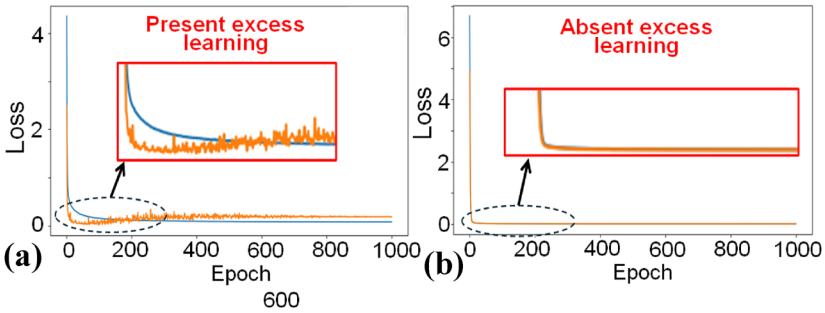

FIG. 5. Loss curves for the conventional 1DCNN model (a) and the present DL model (b) using the number of parameters of 2×10^6 and 11×10^6, respectively. The insets show expanded views of the regions enclosed by dotted circles.

demonstrating superior learning performance. These results suggest that the inductive bias introduced by the data segmentation method [see Eq. (2)] effectively suppresses excess-learning behavior and enhances the stability of the learning process. This improvement underscores the robustness and reliability of the present DL model.

Beyond the advantages in accuracy and computational efficiency, we also examined the model's robustness to quantization. In the case of the DL model using a conventional 1D CNN, performance significantly deteriorates when the model parameters are quantized (i.e., rounded off to integer values, a common practice in DL fields). In contrast, the present DL model exhibits minimal performance loss after quantization. Specifically, when all parameter values are rounded off to one significant digit, the loss increases by only 0.001. This robustness against quantization can be attributed to the convergence of the model's solution to a broadly flat optimal region within the parameter space. Such flat regions are well known for their high generalization performance, enabling the model to provide appropriate solutions even for various unknown data. This characteristic underscores the superior adaptability and stability of the present DL model.

We have conducted a detailed comparative analysis between our proposed DL model and the 1D CNN. Here, in addition to the 1D CNN, other multiple existing architectures, such as AvgNet (GNN-based model), ResNet 1D, Long Short-Term Memory (LSTM), nonstationary transformers, and TimesNet, are compared.[24–26] For evaluation metrics, we use the mean squared error (loss) and inference speed. As the results shown in Fig. 6 and Table II clearly indicate, our model achieves the lowest loss value of 0.003, outperforming all other architectures in prediction accuracy.

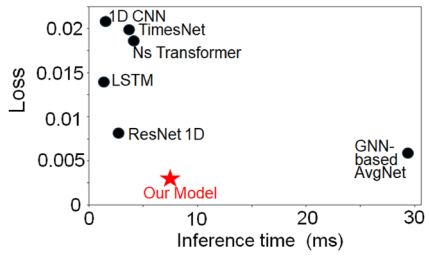

FIG. 6. Comparison of test loss vs inference time for various neural network architectures as plotted from Table II.

In particular, GNN-based AvgNet demonstrates the second-best performance with an accuracy of 0.006, while it suffers from a significant decrease in inference speed, down to 29.187 ms. This is because the computational complexity increases proportionally to the square of the sequence length. The data used in the present study have a sequence length of 3000, which is quite large for graph data. When applying GNNs to such data, practical challenges arise not only in terms of learning stability and decreased inference speed but also in the requirement for expensive computational environments.

Our model achieves the highest prediction accuracy (loss: 0.003) while maintaining an acceptable balance in inference speed (7.487 ms). Particularly for this task, which deals with large-scale sequential data as training data, our proposed DL model is considered the optimal selection as it does not require expensive computational environments while maintaining accuracy.

In conclusion, we have predicted the structural parameters (i.e., $E_{a\alpha}$ ~1.0–1.5 meV and $E_{a\beta}$ ~5–6 meV at N_T ~5.5 $\times 10^{11}/cm^2$) of atom-vacancy defects solely from LF noise (PSD) data, originating from stochastic capture and emission of charges through them, in few-atom layer MoS_2 by developing an innovative DL model. The predicted E_a and N_T values approximately agreed with the experimental results.[9] This result was attributed to the separation of PSDs to the three f segments introducing an inductive bias, the extraction of parameters from each segment using a 1D CNN with GeM, and those combining into 3D tensors, analogous to RGB components in image processing, to leverage advanced DL techniques. This DL model allows precise prediction of structural parameters for various materials without the need for complex and hassle-intensive measurements, highlighting the significant potential of DL in advancing material science.

TABLE II. Comparison of test loss vs inference time among various neural network architectures.

Model	Loss	Inference speed (ms)
1D CNN	0.021	1.512
ResNet 1D	0.008	2.602
LSTM	0.014	1.384
Nonstationary transformer	0.019	4.136
TimesNet	0.020	3.655
AvgNet (GNN)	0.006	29.187
Our model	0.003	7.487

We thank M. Kosugi, S. Furuichi, T. Kohno, T. Kikkawa, T. Taniguchi, K. Watanabe, K. Suenaga, E. Saitoh, and S. Maruyama for their technical contributions, fruitful discussions, and encouragement. The work at the Aoyama Gakuin University was partly supported by the Aoyama Gakuin University Research Institute Grant Program for the creation of innovative research.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Y. Nonaka: Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (lead). **K. Takaki**: Data curation (supporting). **Y. Kobayashi**: Data curation (supporting). **J. Haruyama**: Conceptualization (lead); Supervision (lead); Writing – original draft (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article.

REFERENCES

[1]J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, "Recent advances and applications of machine learning in solid state materials science," npj Comput. Mater. **5**, 83 (2019).
[2]S. Daimon, K. Tsunekawa, S. Kawakami, T. Kikkawa, R. Ramos, K. Oyanagi, T. Ohtsuki, and E. Saitoh, "Deciphering quantum fingerprints in electric conductance," Nat. Commun. **13**, 3160 (2022).
[3]P. V. Balachandran, B. Kowalski, A. Sehirlioglu, and T. Lookman, "Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning," Nat. Commun. **9**, 1668 (2018).
[4]D. Morgan and R. Jacobs, "Opportunities and challenges for machine learning in materials science," Annu. Rev. Mater. Res. **50**, 71–103 (2020).
[5]A. K. Nair, C. M. Da Silva, and C. H. Amon, "Machine-learning-derived thermal conductivity of two-dimensional TiS_2/MoS_2 van der Waals heterostructures," APL Mach. Learn. **2**, 036115 (2024).
[6]C. E. Ekuma, "Computational synthesis of a new generation of 2D-based perovskite quantum materials," APL Mach. Learn. **2**, 026102 (2024).
[7]Á. Díaz Carral, S. Gravelle, and M. Fyta, "*In silico* design and prediction of metastable quaternary phases in Cu–Ni–Si–Cr alloys," APL Mach. Learn. **2**, 046109 (2024).
[8]G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," Nat. Rev. Phys. **3**, 422 (2021).
[9]M. Kosugi *et al.*, "Atom-vacancy-defect-derived electric hysteresis loops and stochastic low-frequency noises in few-atom layer MoS_2," ACS Appl. Mater. Interfaces **16**(46), 64190 (2024).
[10]S. H. Song *et al.*, "Probing defect dynamics in monolayer MoS_2 via noise nanospectroscopy," Nat. Commun. **8**, 2121 (2017).
[11]V. K. Sangwan, H. N. Arnold, D. Jariwala, T. J. Marks, L. J. Lauhon, and M. C. Hersam, "Low-frequency electronic noise in single-layer MoS_2 transistors," Nano Lett. **13**, 4351 (2013).
[12]Y. Zhao *et al.*, "Electrical spectroscopy of defect states and their hybridization in monolayer MoS_2," Nat. Commun. **14**, 44 (2023).
[13]J. Y. Kim, Ł. Gelczuk, M. P. Polak, D. Hlushchenko, D. Morgan, R. Kudrawiec, and I. Szlufarska, "Experimental and theoretical studies of native deep-level defects in transition metal dichalcogenides," npj 2D Mater. Appl. **6**, 75 (2022).
[14]K. C. Santosh, R. C. Longo, R. Addou, R. M. Wallace, and K. Cho, "Impact of intrinsic atomic defects on the electronic structure of MoS_2 monolayers," Nanotechnology **25**, 375703 (2014).
[15]C.-P. Lu, G. Li, J. Mao, L.-M. Wang, and E. Y. Andrei, "Bandgap, mid-gap states, and gating effects in MoS_2," Nano Lett. **14**, 4628 (2014).
[16]H.-P. Komsa and A. V. Krasheninnikov, "Native defects in bulk and monolayer MoS_2 from first principles," Phys. Rev. B **91**, 125304 (2015).
[17]W. Zhou *et al.*, "Intrinsic structural defects in monolayer molybdenum disulfide," Nano Lett. **13**, 2615 (2013).
[18]Y. Katagiri *et al.*, "Gate-tunable atomically thin lateral MoS_2 Schottky junction patterned by electron beam," Nano Lett. **16**, 3788 (2016).
[19]Y. Nagamine *et al.*, "Optoelectronic properties of laser-beam-patterned few-layer lateral MoS_2 Schottky junctions," Appl. Phys. Lett. **117**, 043101 (2020).
[20]M. Kosugi *et al.*, "Electrical hysteresis characteristics in photogenerated currents on laser-beam-derived in-plane lateral 1D MoS_2-Schottky junctions," AIP Adv. **12**, 105210 (2022).
[21]Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proceedings of 27th International Conference on Machine Learning (ICML'10)* (Omnipress, 2010), p. 111.
[22]S. Lee, J. Jeon, and H. Lee, "Probing oxygen vacancy distribution in oxide heterostructures by deep learning-based spectral analysis of current noise," Appl. Surf. Sci. **604**, 154599 (2022).
[23]S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput. **9**(8), 1735 (1997).
[24]Y. Liu, H. Wu, J. Wang, and M. Long, "Non-stationary transformers: Exploring the stationarity in time series forecasting," in Advances in Neural Information Processing Systems (NeurIPS, 2022), Vol. 35, Main Conference Track.
[25]H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "TimesNet: Temporal 2D-variation modeling for general time series analysis," ICLR 2023 Poster (2023).
[26]Q. Xuan, J. Zhou, K. Qiu, Z. Chen, D. Xu, S. Zheng, and X. Yang, "AvgNet: Adaptive visibility graph neural network and its application in modulation classification," IEEE Trans. Network Sci. Eng. **9**(3), 1516 (2022).