

Transfer Learning for Deep Learning-based Prediction of Lattice Thermal Conductivity

L. Klochko and M. d’Aquin
Université de Lorraine, LORIA, Nancy F-54000, France

A. Togo
*Center for Basic Research on Materials (CBRM),
National Institute for Materials Science (NIMS), Tsukuba, Ibaraki 305-0047, Japan*

L. Chaput
*Université de Lorraine, LEMTA, Nancy F-54000, France
Institut Universitaire de France, 1 rue Descartes, 75231 Paris, France*

Machine learning promises to accelerate the material discovery by enabling high-throughput prediction of desirable macro-properties from atomic-level descriptors or structures. However, the limited data available about precise values of these properties have been a barrier, leading to predictive models with limited precision or the ability to generalize. This is particularly true of lattice thermal conductivity (LTC): existing datasets of precise (ab initio, DFT-based) computed values are limited to a few dozen materials with little variability. Based on such datasets, we study the impact of transfer learning on both the precision and generalizability of a deep learning model (ParAI site). We start from an existing model (MEGNet [1]) and show that improvements are obtained by fine-tuning a pre-trained version on different tasks. Interestingly, we also show that a much greater improvement is obtained when first fine-tuning it on a large datasets of low-quality approximations of LTC (based on the AGL model) and then applying a second phase of fine-tuning with our high-quality, smaller-scale datasets. The promising results obtained pave the way not only towards a greater ability to explore large databases in search of low thermal conductivity materials but also to methods enabling increasingly precise predictions in areas where quality data are rare.

I. INTRODUCTION

Machine learning models have advanced material research in several fields, including various domains of physics [2–5], quantum chemistry [6, 7], drug discovery [8], and cancer studies [9–11]. In particular, recent studies have proposed various models to predict the physical properties of materials [1, 12, 13]. These models utilize diverse datasets, input data, and tuned neural network designs for specific purposes. However, the applicability of those models, i.e. their efficacy when applied on large databases, remains limited by their precision and generalizability, which are in turn dependent on the quality and size of the available training data. In other words, unsurprisingly, machine learning models trained on small databases of similar materials do not perform well on large databases of diverse materials [14].

Predicting the low lattice thermal conductivity (LTC) of crystal compounds on the basis of their structure and physical properties is one of those tasks that are made challenging by the lack of quality data. However, it is critical, as the ability to identify low-LTC materials on a scale could have profound implications for the design and optimization of materials in various applications, from electronics to energy storage [15, 16]. The difficulty of this problem lies in the complexity of the relationship between the structure of a material and its thermal properties. Although large data sets on material properties are available through databases such as AFLOW [17], OQMD [18], Materials Project [19], and JARVIS [20] available data on LTC is either based on approximate models (such as AGL [21, 22]) and therefore not usable to build precise machine learning-based predictive models, or rely on ab initio, DFT-based computation which are too expensive to run at large scales and are therefore of very small in sizes.

In this study, we apply a two-stage transfer learning methodology as a way to take advantage of both types of dataset to achieve greater levels of precision and generalizability in deep learning models for predicting LTC. Transfer learning as applied in this article is a process in which a model trained on a first task is reused as a starting point for training on a second similar task. The idea is that initial patterns can be learned in larger, relevant data, which will bootstrap the learning in smaller, more targeted data. The aim is to show the benefits of relying on an existing model having demonstrated good performance in predicting a different property than LTC in transfer learning, but also how this can be pushed further in a second stage of transfer learning, by using the larger, low quality datasets for LTC to pre-train models for predicting the smaller, high quality datasets.

II. METHODOLOGY

Transfer learning [23] involves the use of a machine learning model that has been pre-trained on a large dataset and subsequently fine-tuning it on a smaller domain-specific dataset. This strategy is particularly effective when working with limited data, as it allows us to capitalize on the knowledge gained from broader datasets, as demonstrated in multiple applications for material properties [24, 25]. Here, we describe the datasets, model, and the process used to carry out our two-stage transfer learning process.

A. Data

As mentioned above, the accuracy and reliability of any machine learning model is highly dependent on the quality of the data used for training and validation. In this work, we rely on four datasets, each contributing to our analysis. The first two are derived from calculations of the first principle of anharmonic lattice dynamics [26, 27], looking at a small set of materials with specific structures. The third is a combination of the first two, used to obtain a slightly larger and slightly more diverse dataset. Finally, the fourth can be seen to represent a large dataset that contains low-precision values for LTC.

Dataset1: This dataset contains 96 materials from [26] in the rocksalt, zincblende, and wurtzite structures that could be unambiguously identified in the Material Project Database. The LTC values are obtained using the phono3py software package [28, 29] using the YAML files available through the PhononDB (<https://github.com/atztogo/phonondb>) repository. Obtaining prediction with low deviation from those values is the central motivation for this work.

Dataset2: This dataset includes thermal conductivity data for 143 half-Heusler compounds, as reported in [30]. This dataset adds another set of materials, while being itself significantly more specific than the previous one: it focuses only on one specific structure, for which the range of LTC values is significantly narrower ([2.17, 34.51] W/mK, compared to [0.51, 1769.00] W/mK in Dataset1).

MIX dataset: This dataset is a combination of the two datasets mentioned above, integrating the properties of Dataset1 and Dataset2 to enhance the diversity and scope of our model training.

AFLOW AGL dataset: This dataset contains 5578 materials obtained from the AFLOW-LIB repository [17] together with their corresponding thermal conductivity obtained with the use of a quasi-harmonic Debye-Grüneisen model [21, 22].

In addition, for all cases, logarithmic scaling is applied to all values of LTC, followed by standardization using the parameters of the corresponding dataset. For validation purposes, each data set is split 9 times, keeping 80% for training. In other words, each dataset is associated with 9 different randomly selected validation sets that represent each 20% of the total data set that is not used for training. Any result shown later in this article is measured as an average over those 9 validation sets and the corresponding models for a given dataset.

B. Model

Our model for predicting LTC from the properties and structure of materials (called ParAI site) is based, as shown in Figure 1, on the addition of a multilayer perceptron (MLP, a dense, feed-forward, fully connected layer) on top of an existing, pre-trained model. In line with our transfer learning approach, the idea here is to use an existing model (the pre-trained model) that has already shown its ability to predict properties of materials as a foundation to be adapted for the task of LTC prediction. More concretely, ParAI site is based on connecting the last hidden layer of the pre-trained model to the input of the added MLP.

A first step, therefore, for establishing this model is the selection of the most appropriate pre-trained model from the top-performing models cataloged on MatBench [31]. We only considered models based on an unambiguous identification of the materials as input and on a representation of their structures (in the form of crystallographic information files, CIFs). An initial set of tests were carried out with Dataset1 using the model of Figure 1 with each candidate pre-trained model to validate the model’s performance and ascertain its suitability for the specific challenges associated with predicting the thermal conductivity in crystal compounds.

Following validation (see Fig. 2), we chose the model that combined the best performance (measured by the mean absolute percentage error, MAPE) and was most consistent with the features of our dataset. Despite the better

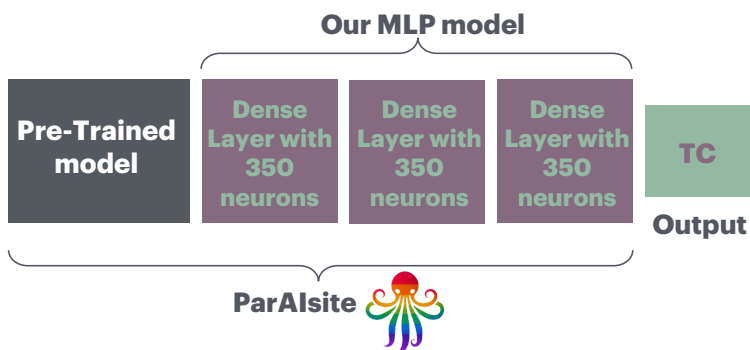


FIG. 1. The ParAIsite model architecture. Each dense layer contains 350 neurons. The output property is the thermal conductivity (TC).

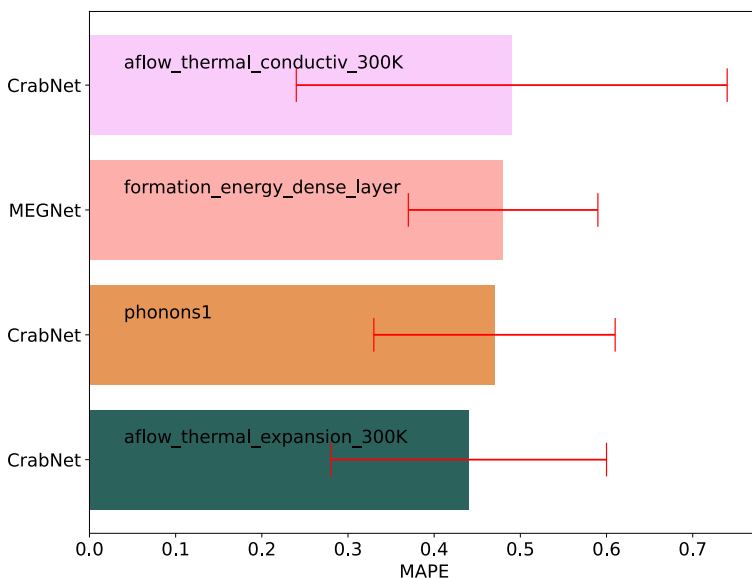


FIG. 2. Results of validation tests of ParAIsite when using top-performing models cataloged on MatBench [31] as pre-trained models. As one can see, the MEGNet model [1] shows better stability compared to the CrabNet model [32] on Dataset1. Here, the name of the datasets on which model was pre-trained are indicated inside the box, and the error variability are shown in red lines.

average performance of the CrabNet model [32] on Dataset1, results over multiple runs showed a lot of variability, demonstrating that this model was too unstable to be used. That result led us to use the graph-based neural network MEGNet model [1], which was pre-trained on the formation energy data of 62,315 compounds from the Materials Project Database as of the 2018.6.1 version. By fine-tuning MEGNet on our thermal conductivity data, we aim to improve the prediction accuracy and better understand the thermal properties of the crystal compounds in our data set.

C. Model training

In order to solve the complexity of predicting LTC, we adopted a three-step method to systematically improve model performance and evaluate the impact of transfer learning (TL) between the steps. In summary, those steps move from no transfer learning at all, to two levels of pre-training/transfer learning being applied, allowing us to evaluate how pre-trained models and fine-tuning affect the model’s ability to generalize across datasets.

Step 1 (no pretraining): In this step, we train and test ParAIsite on our four datasets using an uninitialized MEGNet (and MLP) model. The training is therefore done from scratch (with random initial weights), without any form of transfer learning.

Step 2 (using a pre-trained MEGNet): In this step, we train and test ParAIsite on our four datasets using a pre-trained MEGNet, that is, where MEGNet is initialized with the weights obtained through the training carried out by its authors on the task of predicting formation energy.

Step 3 (transfer-learning over fine-tuned AFLOW model): In this step, we train and test ParAIsite on Dataset1, Dataset2 and MIX, taking as a starting point the best model after fine-tuning using the AFLOW AGL dataset. In other words, we apply a second round of fine-tuning, over the one already done for MEGNet, and our own on AFLOW AGL. By pretraining the whole model further on a larger, but lower quality dataset, we anticipate that training on the three other, more precise datasets will converge to better performances.

At each step, for each model, training is performed for 300 epochs with a fixed random seed of 42, ensuring reproducible results. As already mentioned, each of these steps is repeated 9 times to ensure statistical robustness, and the averaged results for validation loss across all steps are presented in Figs. 3–4. For all training steps, we use MAPE as the loss function, having applied normalisation and scaling.

III. RESULTS AND DISCUSSION

A first straightforward conclusion that can be drawn from training and cross-validating in each step with the considered datasets is that the double steps of transfer learning is having a significant effect on the performance of the model on Dataset1, and generally improve its capacity to generalize if the considered dataset is varied enough. Indeed, in Table I we show the average and standard deviation (over the 9 runs) of MAPE of models trained on the training subsets of each dataset (rows) and tested on the validation subsets of each dataset (columns).

Train on \ Test on	Dataset1	Dataset2	MIX	AFLOW
Step 1:				
Dataset1	0.82 (0.21)	2.53 (2.17)	1.74 (1.26)	2.16 (1.26)
Dataset2	0.51 (0.09)	0.42 (0.07)	0.43 (0.05)	0.48 (0.05)
MIX	0.69 (0.15)	0.77 (0.15)	0.83 (0.07)	0.97 (0.27)
AFLOW	0.55 (0.34)	1.18 (0.27)	0.93 (0.27)	0.62 (0.33)
Step 2:				
Dataset1	0.76 (0.29)	3.09 (2.40)	2.07 (1.40)	2.32 (1.94)
Dataset2	0.55 (0.14)	0.42 (0.06)	0.44 (0.04)	0.52 (0.10)
MIX	0.66 (0.14)	0.75 (0.16)	0.81 (0.11)	1.10 (0.47)
AFLOW	0.44 (0.14)	1.27 (0.48)	0.94 (0.30)	0.50 (0.03)
Step 3:				
Dataset1	0.34 (0.15)	1.26 (0.44)	0.87 (0.27)	0.57 (0.10)
Dataset2	0.55 (0.10)	0.78 (0.20)	0.63 (0.18)	0.62 (0.07)
MIX	0.37 (0.14)	0.78 (0.31)	0.69 (0.19)	0.59 (0.06)

TABLE I. Validation error (MAPE) across training steps and datasets.

Taking as example Dataset1, which is considered particularly challenging, we can see that **with** no transfer learning, models reach on average a 82% error on average when tested on Dataset1’s validation subsets. The error, training and testing again on the relevant subsets of Dataset1 falls to 76% when using the original pre-trained MEGNet model (Step 1) falls to an average of 76%, and falls significantly further, to 34% at Step 2. In other words, transfer learning

has had a significant effect, especially taking as starting point a model that has already been fine-tuned for a similar task (i.e. predicting approximated LTC through the AFLOW dataset).

This effect of the two steps of transfer learning is particularly visible in Figure 3, which shows the evolution of the average MAPE (over 9 runs) in the validation subsets of Dataset1 during the training iterations (epochs). Comparing this evolution between Step 1 and Step 2 shows that starting from a relevant pre-trained model, even if made for a different task, enabled the model to converge faster to slightly lower values of MAPE. We can also see that the MAPE on the validation subsets does not rise up in Step 2 as much as it did during the Step 1 training process, showing that the model was less prone to overfitting in this case. Looking at the chart for Step 3, we can see a significantly different behavior, with the MAPE value converging very quickly to much lower values.

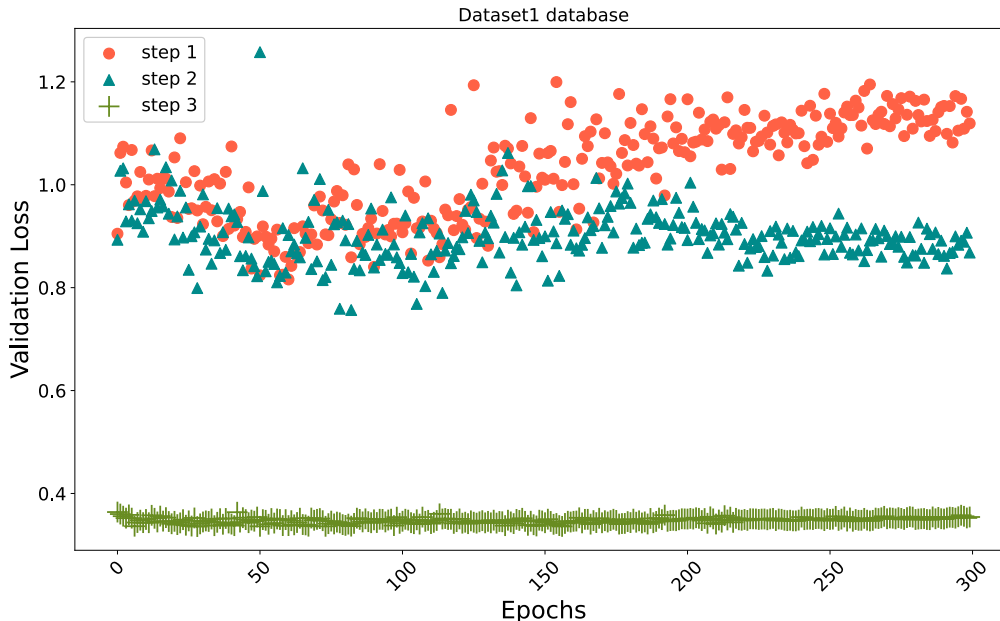


FIG. 3. Validation loss (MAPE) for models trained and tested on Dataset1 across training epochs.

Similar conclusions can be drawn on the MIX dataset, even if those are less strong: The models trained on their training subsets and tested on their validation subsets reach 83%, 81% and 69% in steps 1, 2 and 3 respectively. As seen in Figure 4, MAPE also reaches lower values in Step 2 compared to Step 2 and significantly lower ones yet in Step 3. However, in this case, overfitting appears early in the training process (around epoch 50) in all steps. This is probably explained by the fact that the MIX dataset combines Dataset1, which is well predicted through the transfer learning process, and Dataset2 which, as we discuss below, did not work as well.

Regarding Dataset2, the results shown in Table I, the results obtained are against the ones obtained for the other two datasets. In fact, when training and testing on subsets of Dataset2, the best average MAPE obtained where 42%, 42% and 78% in steps 1, 2 and 3 respectively. In other words, the third step, the double pre-training, significantly worsens the results obtained. This can be explained by the fact that this dataset contains a very restricted range of LTCs compared to the others, in particular AFLOW. In other words, having learned in Step 2 to predict a wide range of LTC values through the AFLOW dataset, the model did not adapt well to the very specific set of materials in Dataset2. We can observe this issue in Figure 5 where, in Step 3, the training starts with a high error rate and is overfit almost immediately.

The issue mentioned above, of the lack of variety in Dataset2, is part of the motivation for integrating the MIX dataset as well. One final conclusion that can be drawn from the results shown in Table I is that, in accordance with our assumption, a more varied dataset (MIX) tends to generalize better, but also that the double pre-training process applied here helps support this ability to generalize. This is visible in the last line of the table, which shows that models trained on the MIX dataset are better on average across all datasets (including AFLOW) than those trained

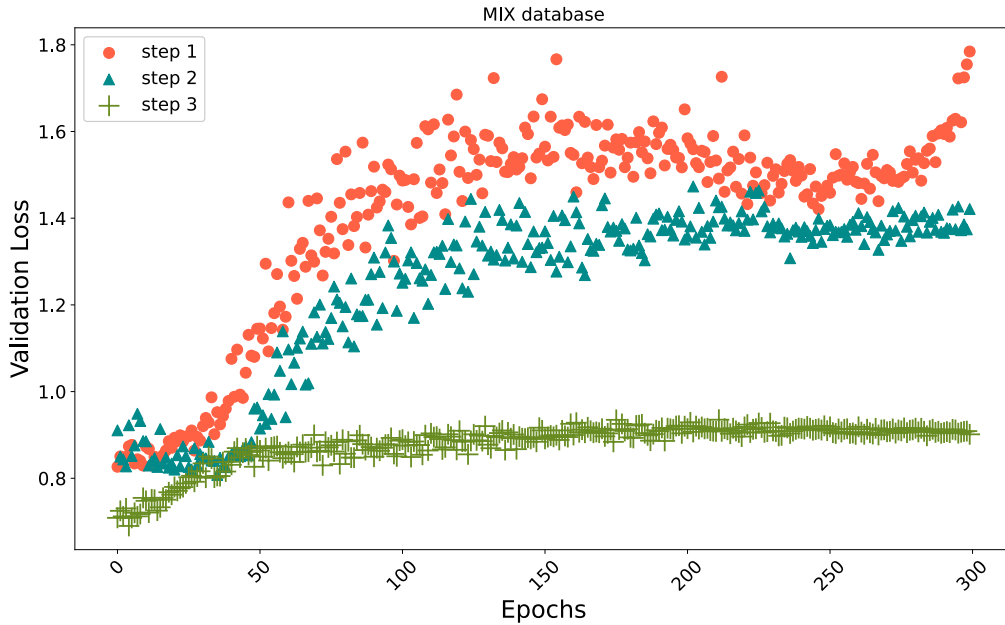


FIG. 4. Validation loss (MAPE) for models trained and tested on MIX across training epochs.

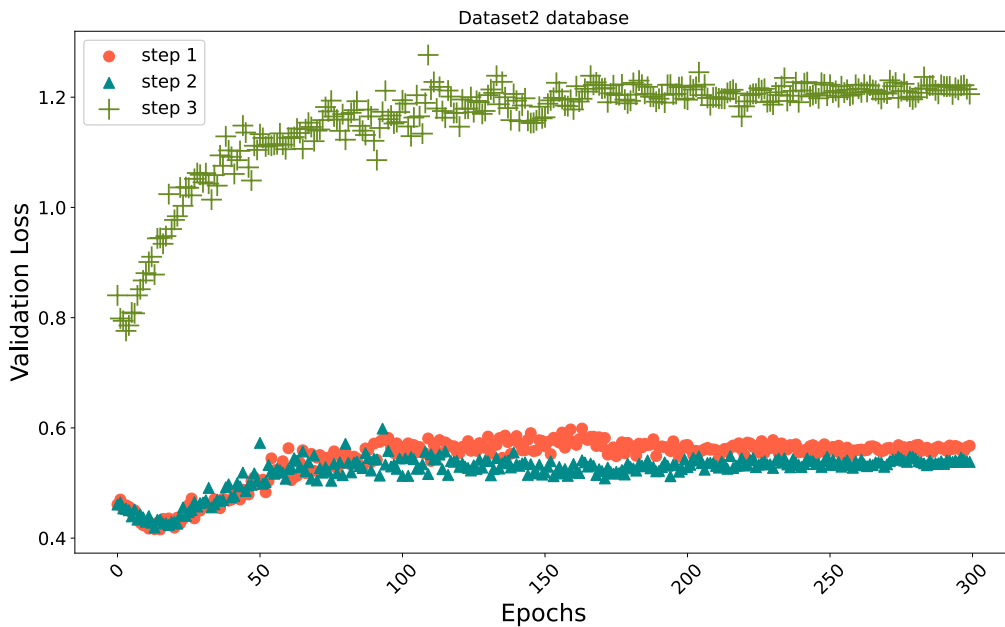


FIG. 5. Validation loss (MAPE) for models trained and tested on Dataset2 across training epochs.

on the other two datasets in the same step. In most cases, models trained on MIX or Dataset1 achieved better results when tested with other datasets at Step 3 than at other steps.

IV. CONCLUSION

In this work, we tested using multiple steps of transfer learning (pre-training and fine-tuning) a machine learning model that we developed to predict the thermal conductivity of crystal compounds. Despite the availability of large datasets for general material properties, databases specifically focused on thermal conductivity are limited, which introduced a challenge to our task. For this reason, the model was trained and validated on three different high-precision datasets, one being a mixture of the other two. We also used a less precise, but larger volume dataset as part of one of the transfer learning steps.

Through a series of experiments involving different training configurations, we found that double transfer learning, which includes an additional phase of training on external data, proved to be effective in reaching not only better precision but also better generalizability and reduced overfitting. This is true for the datasets that show a broader range of values for LTC, and more diversity in the type of material included. For those, the error rate (MAPE) decreased consistently as we progressed through the steps, particularly in Step 3. This indicates that transfer learning, when applied judiciously, can enhance model performance on small datasets.

However, for our less varied dataset, transfer learning had the opposite effect. Double TL caused the model to rapidly overfit, as the specific nature of the dataset was incompatible with the broader generalization achieved through additional external training. In other words, transfer learning was most effective when the dataset was representative of a wide range of materials. In contrast, when dealing with highly specialized datasets, additional training phases may introduce confusion rather than improvement.

To provide concrete validation of the best performing models, we applied them to obtain predictions for stable materials in the Material Project Database. LTC for $(BaSbO_3)_2$ (mp-9127) was then calculated using a robust ab initio calculation, as our models consistently found that it has a relatively low thermal conductivity. The result of the computation ($7.1 \text{ W / m} \cdot \text{K}$) was on the same order of magnitude as the predictions of our models ($1.23 \text{ W/m} \cdot \text{K}$). This agreement underscores the ability of the model to capture critical trends in LTC prediction, even for datasets it was not trained directly on.

In summary, while double transfer learning shows great promise in improving model accuracy and generalization, its success heavily depends on the dataset’s diversity and scope. The results obtained are very promising but also demonstrate that the choice of dataset and training approach is crucial when predicting thermal conductivity. A greater availability of a broader range of datasets of LTC, whether of high precision for training or of approximate precision for pre-training, is therefore expected to enable us to reach better results in the future.

V. CODE AND DATA AVAILABILITY

The code and data that are required to reproduce the results of paper or re-utilizing for their own purposes are shown on <https://github.com/liudak1/ParAIsite.git>

ACKNOWLEDGMENTS

The authors acknowledge Lorraine Université d’Excellence (<https://www.univ-lorraine.fr/lue/>) for the financial support of Liudmyla Klochko to work on this project.

VI. AUTHOR CONTRIBUTIONS

All authors discussed the results and contributed equally to the manuscript. L. Klochko developed and carried out the implementation of the model and wrote the first versions of the manuscript. M. d’Aquin provided the domain expertise in deep learning and transfer learning necessary for this work and revised the manuscript. L. Chaput provided the necessary domain expertise in crystallography and thermophysical properties of material and revised the manuscript. A. Togo developed the computer codes used to compute the thermal conductivity from the first principle.

VII. COMPETING INTERESTS

There is no competing interests to declare.

-
- [1] C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong, Graph networks as a universal machine learning framework for molecules and crystals, *Chemistry of Materials* **31**, 3564–3572 (2019).
 - [2] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, Physics-informed machine learning, *Nature Reviews Physics* **3**, 422–440 (2021).
 - [3] C. K. H. Borg, E. S. Muckley, C. Nyby, J. E. Saal, L. Ward, A. Mehta, and B. Meredig, Quantifying the performance of machine learning models in materials discovery, *Digital Discovery* **2**, 327–338 (2023).
 - [4] A. Asensio Ramos, M. C. M. Cheung, I. Chifu, and R. Gafeira, Machine learning in solar physics, *Living Reviews in Solar Physics* **20**, 10.1007/s41116-023-00038-x (2023).
 - [5] W.-B. He, Y.-G. Ma, L.-G. Pang, H.-C. Song, and K. Zhou, High-energy nuclear physics meets machine learning, *Nuclear Science and Techniques* **34**, 10.1007/s41365-023-01233-z (2023).
 - [6] A. C. Mater and M. L. Coote, Deep learning in chemistry, *Journal of Chemical Information and Modeling* **59**, 2545–2559 (2019).
 - [7] V. Vinod, U. Kleinekathöfer, and P. Zaspel, Optimized multifidelity machine learning for quantum chemistry, *Machine Learning: Science and Technology* **5**, 015054 (2024).
 - [8] C. Sarkar, B. Das, V. S. Rawat, J. B. Wahlang, A. Nongpiur, I. Tiewsoh, N. M. Lyngdoh, D. Das, M. Bidarolli, and H. T. Sony, Artificial intelligence and machine learning technology driven modern drug discovery and development, *International Journal of Molecular Sciences* **24**, 2026 (2023).
 - [9] B. Zhang, H. Shi, and H. Wang, Machine learning and ai in cancer prognosis, prediction, and treatment selection: A critical approach, *Journal of Multidisciplinary Healthcare* **Volume 16**, 1779–1791 (2023).
 - [10] K. Swanson, E. Wu, A. Zhang, A. A. Alizadeh, and J. Zou, From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment, *Cell* **186**, 1772–1791 (2023).
 - [11] A. Yaqoob, R. Musheer Aziz, and N. K. verma, Applications and techniques of machine learning in cancer classification: A systematic review, *Human-Centric Intelligent Systems* **3**, 588–615 (2023).
 - [12] A. Ihalage and Y. Hao, Formula graph self-attention network for representation-domain independent materials discovery, *Advanced Science* **9**, 10.1002/adv.202200164 (2022).
 - [13] Y. Suzuki, T. Taniai, K. Saito, Y. Ushiku, and K. Ono, Self-supervised learning of materials concepts from crystal structures via deep neural networks, *Machine Learning: Science and Technology* **3**, 045034 (2022).
 - [14] Y. Zhang and C. Ling, A strategy to apply machine learning to small datasets in materials science, *npj Computational Materials* **4**, 10.1038/s41524-018-0081-z (2018).
 - [15] Y. Qian, Y. Luo, Y. Li, T. Xiong, L. Wang, W. Zhang, S. Gang, X. Li, Q. Jiang, and J. Yang, Enhanced electromagnetic wave absorption, thermal conductivity and flame retardancy of BCN@LDH/EP for advanced electronic packing materials, *Chem. Eng. J.* **467**, 143433 (2023).
 - [16] T. Huang, T. Wang, J. Jin, M. Chen, and L. Wu, Design of silicon rubber/bn film with high through-plane thermal conductivity and ultra-low contact resistance, *Chemical Engineering Journal* **469**, 143874 (2023).
 - [17] C. E. Calderon, J. J. Plata, C. Toher, C. Oses, O. Levy, M. Fornari, A. Natan, M. J. Mehl, G. Hart, M. Buongiorno Nardelli, and S. Curtarolo, The aflow standard for high-throughput materials science calculations, *Computational Materials Science* **108**, 233–238 (2015).
 - [18] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, and C. Wolverton, The open quantum materials database (oqmd): assessing the accuracy of dft formation energies, *npj Computational Materials* **1**, 10.1038/npj-compumats.2015.10 (2015).
 - [19] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, Commentary: The materials project: A materials genome approach to accelerating materials innovation, *APL Materials* **1**, 10.1063/1.4812323 (2013).
 - [20] K. Choudhary, K. F. Garrity, A. C. E. Reid, B. DeCost, A. J. Bicchi, A. R. Hight Walker, Z. Trautt, J. Hattrick-Simpers, A. G. Kusne, A. Centrone, A. Davydov, J. Jiang, R. Pachter, G. Cheon, E. Reed, A. Agrawal, X. Qian, V. Sharma, H. Zhuang, S. V. Kalinin, B. G. Sumpter, G. Pilania, P. Acar, S. Mandal, K. Haule, D. Vanderbilt, K. Rabe, and F. Tavazza, The joint automated repository for various integrated simulations (jarvis) for data-driven materials design, *npj Computational Materials* **6**, 10.1038/s41524-020-00440-1 (2020).
 - [21] M. Blanco, E. Francisco, and V. Luaña, Gibbs: isothermal-isobaric thermodynamics of solids from energy curves using a quasi-harmonic debye model, *Computer Physics Communications* **158**, 57–72 (2004).
 - [22] C. Toher, J. J. Plata, O. Levy, M. de Jong, M. Asta, M. B. Nardelli, and S. Curtarolo, High-throughput computational screening of thermal conductivity, debye temperature, and grüneisen parameter using a quasiharmonic debye model, *Physical Review B* **90**, 10.1103/physrevb.90.174107 (2014).
 - [23] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, A comprehensive survey on transfer learning (2019).

- [24] S. Kong, D. Guevarra, C. P. Gomes, and J. M. Gregoire, Materials representation and transfer learning for multi-property prediction, *Applied Physics Reviews* **8**, 10.1063/5.0047066 (2021).
- [25] N. Hoffmann, J. Schmidt, S. Botti, and M. A. L. Marques, Transfer learning on large datasets for the accurate prediction of material properties (2023).
- [26] A. Seko, A. Togo, H. Hayashi, K. Tsuda, L. Chaput, and I. Tanaka, Prediction of low-thermal-conductivity compounds with first-principles anharmonic lattice-dynamics calculations and bayesian optimization, *Physical Review Letters* **115**, 10.1103/physrevlett.115.205901 (2015).
- [27] A. Togo, L. Chaput, and I. Tanaka, Distributions of phonon lifetimes in brillouin zones, *Physical Review B* **91**, 10.1103/physrevb.91.094306 (2015).
- [28] A. Togo, L. Chaput, T. Tadano, and I. Tanaka, Implementation strategies in phonopy and phono3py, *J. Phys. Condens. Matter* **35**, 353001 (2023).
- [29] L. Chaput, Direct solution to the linearized phonon boltzmann equation, *Phys. Rev. Lett.* **110**, 265506 (2013).
- [30] H. Miyazaki, T. Tamura, M. Mikami, K. Watanabe, N. Ide, O. M. Ozkendir, and Y. Nishino, Machine learning based prediction of lattice thermal conductivity for half-heusler compounds using atomic information, *Scientific Reports* **11**, 10.1038/s41598-021-92030-4 (2021).
- [31] A. Dunn, Q. Wang, A. Ganose, D. Dopp, and A. Jain, Benchmarking materials property prediction methods: the matbench test set and automatminer reference algorithm, *npj Computational Materials* **6**, 10.1038/s41524-020-00406-3 (2020).
- [32] A. Y.-T. Wang, S. K. Kauwe, R. J. Murdock, and T. D. Sparks, Compositionally restricted attention-based network for materials property predictions, *npj Computational Materials* **7**, 77 (2021).