Computational Materials Science 234 (2024) 112783

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Full length article

Predicting melting temperature of inorganic crystals via crystal graph neural network enhanced by transfer learning

Jaesun Kim [a,1], Jisu Jung [a,1], Sookyung Kim [b], Seungwu Han [a,c,*]

[a] Department of Materials Science and Engineering, Seoul National University, Seoul, 08826, Republic of Korea
[b] Palo Alto Research Center (PARC), Palo Alto, CA, USA
[c] Korea Institute for Advanced Study, Seoul, 02455, Republic of Korea

ARTICLE INFO

Dataset link: https://github.com/MDIL-SNU/TL-GeoCGNN

Keywords:
Graph convolution neural network
Melting temperature
Transfer learning
Material informatics

ABSTRACT

The melting temperature, a crucial material property, is particularly challenging to measure accurately for inorganic crystals. The data-driven approach emerges as a potential solution, although its effectiveness is hindered by the limitation of available data. To counter this challenge, we implement transfer learning, leveraging a vast computational database of atomization energy. We first pre-train a geometric-information-enhanced crystal graph neural network (GeoCGNN) using atomization energies of approximately 36,000 materials that are computed by the density functional theory. Subsequently, the pre-trained model is fine-tuned using melting temperatures measured for 799 crystals, encompassing 83 elements, ranging from unary to quaternary systems. This transfer learning strategy decreases the root mean square error from 407 to 218 K, attesting to a marked improvement in prediction accuracy. Furthermore, transfer learning significantly mitigates error variability across unary, binary, and ternary (or higher-order) systems, thereby enhancing the reliability of predictions across a broader range of crystals. We also show that transfer learning allows effective task adaptation by leveraging representation learned from pre-training. Therefore, it can achieve better prediction performance even with a limited number of data for predicting melting points.

1. Introduction

The melting temperature (T_m) is a vital material property that largely determines operation and processing temperatures [1,2]. However, accurate measurement of T_m is challenging due to the sensitivity of the procedure to experimental conditions, such as sample purity, heating rate, and instrument calibration. Furthermore, complex phase equilibria and incongruent melting complicate the determination of T_m in multicomponent systems.

Consequently, experimental data on T_m are relatively scarce compared to other fundamental properties such as crystal structure or heat of formation. Density functional theory (DFT) based computational methodologies offer an alternative to determine T_m, either through thermodynamic integration or direct simulation of the liquid–solid interface [3,4]. However, these computational approaches can be prohibitively expensive due to the high computational cost of DFT calculations [5]. Recent advancements in machine learning potentials have considerably mitigated these costs, yet the computational time for DFT calculations required in constructing the training set remains significant [6].

Considering the challenges discussed above in both experimental and theoretical aspects, predicting T_m based on simpler parameters would be beneficial. For instance, the Debye model [7] or the equation of state [8] were utilized in predicting T_m within physics-based models. However, these approaches have been validated only for unary metal systems.

More recently, data-driven machine-learning (ML) methods have been attempted in predicting T_m of diverse compounds [9–13]. To be specific, Gu et al. used support vector regression (SVR) with kernels to predict T_m for binary and ternary compound semiconductors [9]. Their database consisted of 25 III-V, II-VI binary compounds and 28 I-III-VI_2, II-IV-V_2 ternary crystals. By encoding atomic features like electronegativity and atomic mass into descriptors, they achieved a mean relative error of 6% in T_m prediction. Saad et al. predicted T_m for 44 AB suboctet solids (e.g., MgAu, LiAl) using ridge regression [10]. They utilized elemental properties of the constituent atoms, such as electronegativity and boiling point, as input features. The median relative error in their predictions was 12.8%.

In another study, Seko et al. used both linear and non-linear regression methods including SVR to predict T_m for 248 A_xB_y binary

* Corresponding author at: Department of Materials Science and Engineering, Seoul National University, Seoul, 08826, Republic of Korea.
E-mail address: hansw@snu.ac.kr (S. Han).
[1] Contributed equally.

https://doi.org/10.1016/j.commatsci.2024.112783
Received 18 August 2023; Received in revised form 24 November 2023; Accepted 3 January 2024
Available online 9 January 2024

crystals that do not contain transition metals [11]. They employed atomic features, similar to the descriptors used in Refs. [9,10], as well as crystal properties like volume and bulk modulus. The incorporation of crystal properties significantly improved the accuracy, resulting in a root mean squared error (RMSE) of 262 K when using SVR. Pilania et al. also utilized SVR with a Gaussian kernel to predict T_m for 46 binary crystals that satisfy the octet rule [12]. In addition to individual atomic features, they introduced crystal properties calculated with DFT, such as Born effective charges and nearest-neighbor distances, achieving an RMSE of 150 K.

However, the above data-driven approaches tend to rely on chemically biased datasets that include only non-transition metals or specific combinations of oxidation states. Such biased datasets may lead to model overfitting in a specific domain and limit the generalizability of the model to broader applications. In this respect, it is notable that Hong et al. developed a graph neural network that can be applied to a more general material class by training the model over experimental and theoretical T_m of 9375 compounds [13]. The model achieved an RMSE of 160 K.

The data-driven prediction of T_m is fundamentally constrained by the data scarcity that originates from the low-throughput in both experimental and computational procedures. In material science, one way to overcome such data scarcity, in particular experimental data, is to employ transfer learning (TL) that involves pre-training with large theoretical databases and fine-tuning with smaller experimental datasets. For instance, Jha et al. utilized a pre-trained model incorporating DFT formation enthalpy of more than 10,000 materials, which is fine-tuned with the formation enthalpy of 1643 experimental observations [14]. Pratik et al. applied TL to predict the capacitance of 4896 metal–oxide–semiconductor (MOS) capacitors using knowledge transferred from a dataset of 114,000 theoretical capacitance values [15]. Yamada et al. made predictions on the heat capacity and conductivity of polymers [16] by employing the theoretical heat capacities of monomers as a pre-training set.

Motivated by the above discussions, we herein develop a TL approach with an aim to develop a ML model that predicts T_m across a wide range of materials. To the best of our knowledge, this is the first attempt to apply TL in predicting T_m. Given that the database of theoretical T_m is also limited due to the high computational cost of DFT, we opt to utilize a theoretical property that is known for a large number of materials. We also consider that the effectiveness of TL in improving prediction accuracy is enhanced when the physical property of the pre-trained set demonstrates a high correlation with the target property [17].

Thermodynamically, the T_m is closely related to the enthalpy of fusion that quantifies the energy loss of crystals as the interactions between atoms weaken during melting. This quantity in turn can be roughly estimated using the atomization energy, which indicates the energy of interactions between atoms in crystals. For instance, Guinea et al. and Sankaran et al. have found approximately linear relations between T_m and atomization energies for metals and elemental crystals, respectively, particularly at high temperatures [7,18]. Thus, our TL strategy is as follows: We first pre-train a ML model using the atomization energies of approximately 36,000 inorganic crystals collected from the Materials Project [19]. Then, we use TL to fine-tune the model with experimental T_m data for 799 compounds from the CRC Handbook [20]. As a ML model, we utilize the geometric-information-enhanced crystal graph neural network (GeoCGNN) [21], which has been demonstrated to outperform other graph-based models in predicting crystal properties.

2. Method

2.1. Pre-training and target database

First, a pre-trained model is developed by utilizing atomization energy data sourced from the Materials Project [19]. The pre-training database consists of approximately 36,000 crystals provided in GeoCGNN, with the exception of some crystals for which atomization energy data is unavailable. The hyperparameters for the model are set to the same optimized values as those used in the previous study [21]. The dataset is partitioned into training, validation, and test sets with respective ratios of 0.6, 0.2, and 0.2. The mean absolute error of the pre-trained model is 0.038 eV/atom, while the distribution of atomization energy of the database ranges from 0 to 9 eV/atom.

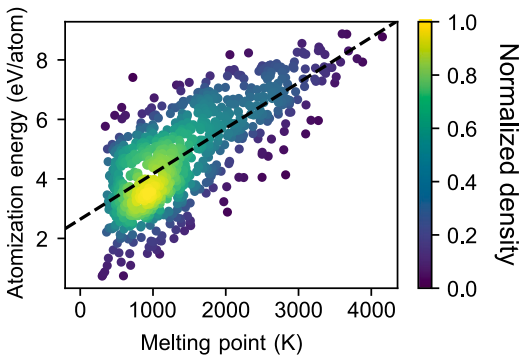

Fig. 1. The correlation between the T_m and atomization energy. The dashed line indicates the linear regression between them. The coefficient of determination (R^2 value) is 0.54.

Next, a database of experimental T_m is compiled using information from the CRC Handbook [20]. Compounds exhibit sublimation, explosion, or thermal decomposition such as peritectic melting were excluded from the database. Since this work primarily focuses on inorganic crystals, we also exclude organic and molecular systems that possess significantly different physical properties from inorganic compounds. The crystal structures were obtained from the Inorganic Crystal Structure Database (ICSD) [22]. For materials that exhibit a phase transition before reaching the melting stage, this procedure may introduce inconsistencies between the crystal structure and its corresponding melting point since GeoCGNN can detect the difference between crystal structures even at identical compositions. Nevertheless, we select the most stable structure in the Materials Project [19] among polymorphs, assuming that this discrepancy would be negligible. In fact, the variation of melting points between polymorphs in unary systems is typically less than 100 K, within a range spanning from 500 K to 2500 K [23]. In total, the database comprises 799 compounds, encompassing 83 different elements with unary to quaternary composition. Among them, the highest T_m of 4153 K is found for TaC and Cs has the lowest T_m of 301.5 K.

As mentioned in the introduction, we choose the atomization energy as the target property for the pre-trained model because it can be collected for a large number of materials and is expected to have a strong correlation with T_m. In Fig. 1, atomization energies and T_m are scatter-plotted for crystals included in this study. It is seen that the two physical quantities are well correlated, showing an R^2 value of 0.54. Since the atomization energy varies between 0 and 9 eV/atom, while T_m ranges from 300 to 4000 K, the learnable parameters in the optimized model could differ significantly, even though there is a high correlation between these properties. To address this issue, a normalization process is employed to align the scales between the pre-trained and target properties [17,24]. Herein, a linear regression function $y = 1.529(x/1000) + 2.645$ (see a dashed line in Fig. 1) is used to adjust the range of the pre-training set to the target values of T_m. Additionally, to investigate the effect of pre-training feature selection on the accuracy of the resulting model, we also collect the formation energy and band gap energy from the Materials Project [19] as potential pre-training features, despite their low R^2 values of 0.011 and 0.072, respectively (see Fig. S1).

2.2. Transfer learning

Like pre-training, the TL is carried out by partitioning the T_m database into training, validation, and test sets with proportions of 0.6, 0.2, and 0.2, respectively. To preserve the knowledge from the pre-trained model, TL adopts two strategies: reducing the learning rate and freezing layers [17]. Both of them aim to optimize the parameters of the model from the pre-trained space, taking advantage of convergence compared to random initialization. In the reduced learning rate approach, after loading all pre-trained parameters, the model updates the parameters slightly to fit target properties with a reduced learning rate. We tried a reduced learning rate of 1, 0.75, 0.5, and 0.25 times the original learning rate of $1e^{-3}$. In similar manner, by fixing the parameters in the frozen layers, the model can retain knowledge from pre-training. The GeoCGNN model consists of an embedding block, a series of gated convolution blocks, which includes a convolution layer and a single perceptron layer, and an output block [21]. We select the frozen layers in an embedding block and gated convolution blocks. For the full details of the architecture of the GeoCGNN model, we refer to Ref. [21].

Since there are a huge number of possible combinations of frozen layers and learning rates, we employ the Taguchi method to identify the optimal choice [25]. To be specific, a mixed orthogonal L16 ($2^{12}4^1$) array in the Taguchi method is utilized, in which the first 12 columns indicate whether a certain layer is frozen or not, and the final column chooses a reduced learning rate from four values described above. For the case of atomization energy as a pre-training dataset, the three frozen layers and the same learning rate with pre-training are chosen.

3. Results and discussions

3.1. Model performance

As detailed in the previous section, the model (called Model-TL hereafter) was initially pre-trained with atomization energies of approximately 36,000 crystals, and then trained on the T_ms of 799 crystals. To investigate the impact of TL, we develop another model, referred to as Model-FS. The Model-FS is trained 'from scratch' using T_ms of the same crystals as in Model-TL, but without utilizing any pre-training process, unlike Model-TL. In Fig. 2, we present a scatter plot comparing the experimental T_m with the predicted T_m values for the test set (160 materials). For the Model-FS in Fig. 2(a), the RMSE is 407 K (or the mean relative error of 25.5%). The accuracy significantly improves with the Model-TL with an RMSE of 218 K (or the mean relative error of 12.6%). In addition, outlier crystals in Model-FS such as Re, SiC, Ce_2S_3, and $KAsO_3$ [see dashed circles in Fig. 2(a)] become more accurately described by Model-TL. This indicates that the inclusion of atomization energy information in Model-TL contributes to its ability to learn the relationship between the crystal structure and T_m. The error of Model-TL is comparable to those in previous studies [9–13]. However, we note that the previous studies focused on specific material domains [9–12] or ten times larger database than the present work [13].

As a test, when formation energy and band gap energy were used for pre-training, the resulting RMSEs in predicting melting points were 269 K and 450 K, respectively. These performances are inferior compared to the Model-TL pre-trained with atomization energy, which achieved an RMSE of 218 K. Our findings underscore the significance of the resolution in the pre-training feature and the relationship between the pre-training feature and target property in TL. In particular, atomization energy aids in distinguishing unary crystals with various energy values, whereas formation energy fails in this aspect, providing a constant value of zero for unary crystals. Furthermore, employing band gap energy as a pre-training feature resulted in decreased model accuracy (450 K) compared to Model-FS which is trained solely on melting points without pre-training (407 K). This can be attributed to the indirect influence of electronic properties, like band gap energy, on the melting mechanisms, unlike atomization energy and formation energy. Additionally, the absence of band gaps in metallic compounds results in a null value for band gap energy, further complicating the performance of the model. Therefore, TL can be more advantageous when a pre-training feature is directly related to the target property with a distinguishable value. Conversely, if the physical meaning of the pre-training feature is not related to that of the target property, it is more beneficial to exclude such features from the TL.

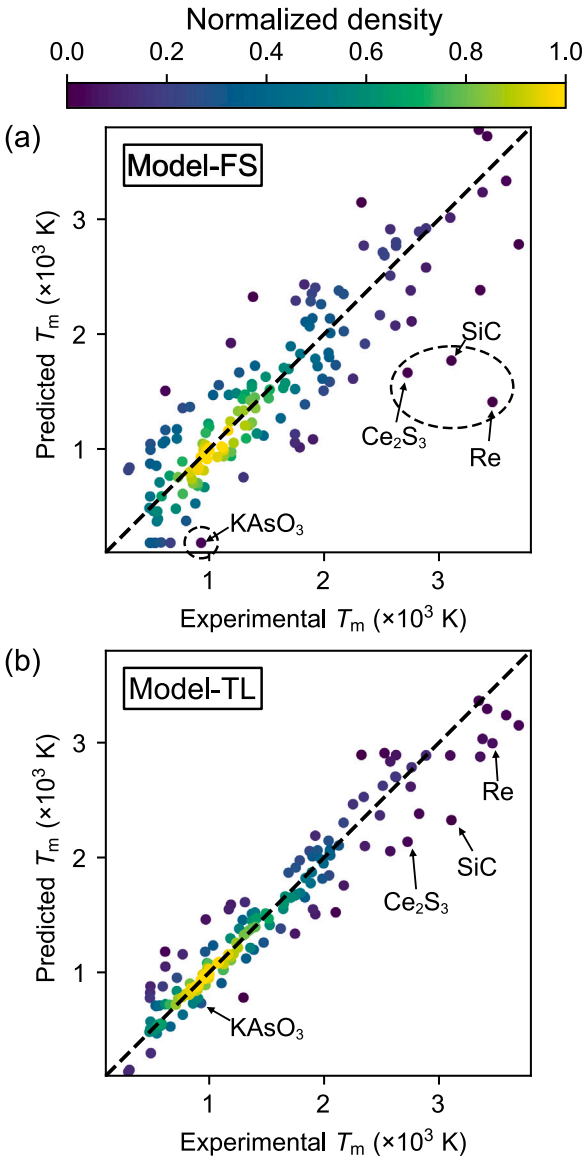

Fig. 2. Parity plots of reference T_m and T_m predicted by (a) Model-FS and (b) Model-TL for the test set (160 materials). Several crystals with high errors at Model-FS are highlighted in dashed circles.

3.2. Model accuracy depending on material types

In this subsection, we examine the model accuracy and the impact of TL depending on the material types. We first classify the database (799 materials) with respect to the number of elements, i.e., unary, binary, and ternary (or higher-order) materials. The distributions of population and T_m are shown at the top and bottom of Fig. 3(a), respectively. It is seen that the binary crystals dominate the population. Such an imbalance in the dataset can lead to significant errors in the model, especially for relatively sparse datasets like unary crystals.

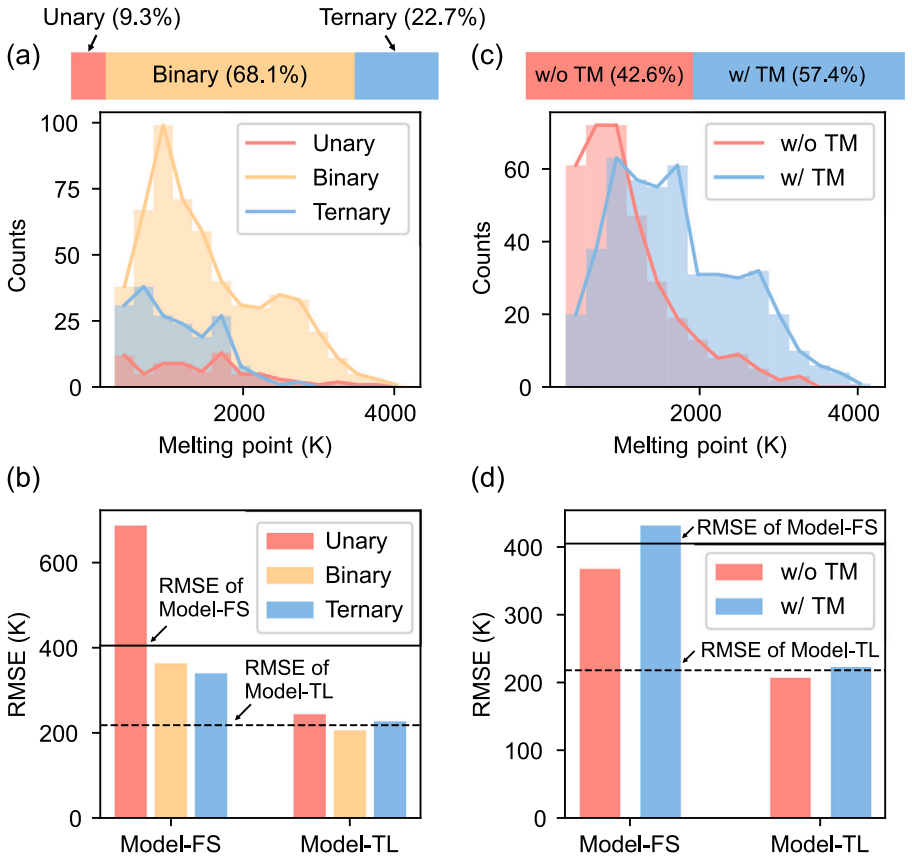

Fig. 3. Distribution and test set RMSE of T_m by each crystal group. (a) Distribution in the database and (b) test set RMSE of T_m by the number of elements in the crystals. (c) Distribution in the database and (d) test set RMSE of T_m with and without transition metals. The solid and dashed lines in (b) and (d) indicate the RMSE of Model-FS and Model-TL, respectively.

In Fig. 3(b), we evaluate the accuracy of the model and the influence of TL for each subset. With regards to Model-FS, it demonstrates satisfactory predictions for binary and ternary systems but encounters increased errors when predicting unary systems. The observed outcome can be ascribed to two key factors: Firstly, unary systems represent only a small portion of the database. Secondly, their unique data characteristics complicate accurate prediction. To elaborate, unary systems predominantly consist of elemental metal crystals, and these metal elements often form similar crystal structures such as face-centered cubic (FCC) and body-centered cubic (BCC). For example, both Na and Fe adopt BCC structures with comparable bond lengths (4.50 and 4.52 Å, respectively) [26]. As a result, the edge features of the crystal graph in GeoCGNN are almost indistinguishable for these elements, and the primary distinguishing factor is the node feature that one-hot encodes the atomic number. However, despite the similarities in input data, output T_m differs significantly, with Na having a T_m of 371 K and Fe having a T_m of 1811 K. The output range for unary systems extends from 300 to 3500 K [see Fig. 3(a)], making accurate prediction of T_m challenging for the Model-FS given the limited available information.

In Fig. 3(b), the application of transfer learning, Model-TL, results in improved prediction accuracy, along with reduced error deviations between unary, binary, and ternary systems. In particular, the unary systems exhibit the most significant improvement in accuracy when transfer learning is applied, likely a consequence of the features internalized during pre-training: Among the material types in the present database, elemental metals display the highest correlation coefficient ($R^2 = 0.79$) between atomization energy and T_m, thereby indicating a particularly effective knowledge transfer.

Conversely, the enhancement in accuracy for ternary systems is relatively modest. This can be attributed to the weak correlation between atomization energy and the T_m of ternary crystals. For example, inorganic crystals containing polyatomic ions, such as sulfate and carbonate, are predominantly found within the ternary crystal group. During the melting process, these polyatomic ions remain undissociated [27], but atomization energy presumes the complete dissociation of all bonds for each atom within the crystal. Consequently, ternary crystals, with T_m in the range of 500 to 1500 K, exhibit a relatively high average atomization energy of 5.01 eV/atom, whereas unary and binary systems present average atomization energies of 3.10 and 3.78 eV/atom, respectively. This implies limited improvements in prediction accuracy for ternary or higher-order crystals when using atomization energy of the entire system as the pre-training set.

One might have an interest in particular systems for TL, where the correlation between a pre-training feature and target property is concerned. For instance, ternary systems show the upshifted linear correlation due to their unique melting mechanism compared to unary and binary systems as discussed before. Because the atomization energy and melting point data for ternary compounds alone had an R^2 value of 0.57 [see Fig. S2(a)], we reduced the RMSE of the ternary system from 230 K to 207 K, underscoring the benefits of targeted data collection. Based on the fact that atomization energy correlates well with T_m by representing the thermal stability of crystals, other thermodynamic properties might be beneficial. Herein, we investigate thermal properties that are related to hull energy as a pre-training set for ternary system. To be more specific, we utilize decomposition energy and formation energy. These properties might offer insights to the dissociation into a stable phase, rather than just individual atoms, addressing the distinctive melting processes observed in ternary compounds, as previously discussed. However, decomposition energy, representing the energy difference between a compound and its competing compounds in a given chemical space [28], showed negligible correlation with melting points in ternary systems with R^2 of less

than 0.01. This can be attributed to the presence of other stable solid compounds within the same chemical space. Conversely, the formation energies exhibited a moderate correlation with an R^2 value of 0.51 with melting points in ternary systems [see Fig. S2(b)] compared to entire systems with R^2 value of 0.011. For ternary compounds alone, using formation energy and melting point data as a pre-training feature and target property, respectively, resulted in an RMSE of 236 K, lower than the RMSE of 269 K observed when using formation energy for the entire dataset. This suggests that formation energy can be an effective alternative to atomization energy for enhancing model accuracy, particularly when atomization energy data is challenging to collect. These results highlight the importance of focusing on the target material system with relevant pre-training features. Concentrating on a specific class of materials enables the model to capture relationships pertinent to that class, leading to improved accuracy.

According to Saad et al. the inadequate representation of d electrons resulted in significant prediction errors in crystals containing d-block elements [10]. To investigate if this issue persists in our models, we classify the crystals into two categories: those containing transition metals (TMs) and those without. As illustrated in Fig. 3(c), the number of crystals both with (w/ TM) and without TMs (w/o TM) is similar.

In Fig. 3(d), we assess the prediction accuracy for crystals with and without TMs. The error level remains practically identical in both cases, at variance with Ref. [10], where higher errors were noted for crystals containing TMs. This discrepancy can be traced back to the difference in the prediction model. In Ref. [10], TM prediction was dependent on predefined descriptors, which might fail to capture the complex behavior of d electrons adequately. In contrast, GeoCGNN utilizes learnable parameters for atomic feature embedding, facilitating the representation of elements regardless of their shell structures. Moreover, the enhancement in prediction performance achieved through TL is similarly significant for both crystal subsets, irrespective of the presence or absence of TMs.

3.3. Effect of TL by dataset size and composition

Given the challenges associated with measuring T_m, it would be important to predict T_m with a small number of experimental observations. To assess the effect of TL for limited size of dataset, we randomly sample 20 to 480 training data points from the original database. For all the sizes of the training set, we evaluate the RMSE on a common test set with five independent models. In Fig. 4, we present the RMSE of both Model-FS and Model-TL for each data size. We found significant variations in model performance based on the size of the training set [see Fig. 4(a)]. The accuracy of Model-TL trained with only 50 crystals with RMSE of 422 K is comparable to that of Model-FS trained with the entire database with RMSE of 407 K. It is also seen that as the data size shrinks, the accuracy of Model-FS degrades more rapidly than Model-TL. This indicates that the role of TL becomes increasingly important as the dataset size decreases, offering high accuracy for T_m prediction in limited data size.

We have conducted further analyses to assess the impact of different subsets of training data on model accuracy. We categorized the training set into unary, binary, and ternary groups based on the number of elements in the compounds. This division was meant to enable the model to specialize in particular material classes. We labeled the model as 'Model-TL-unary', 'Model-TL-binary', and 'Model-TL-ternary' for the model exclusively trained by melting points of unary, binary, and ternary systems, respectively. Their maximum training set sizes were 50, 313, and 125. For the Model-TL-unary, the RMSE decreased steadily from 732 K to 559 K with an increasing size of training set size from 20 to 50 data points. However, for binary and ternary systems, the RMSE remains relatively constant across varying the size of the training set. This distinct discrepancy likely stems from the inherent similarity in the crystal graph of unary systems. Every single data point in unary systems would be beneficial in increasing the accuracy by

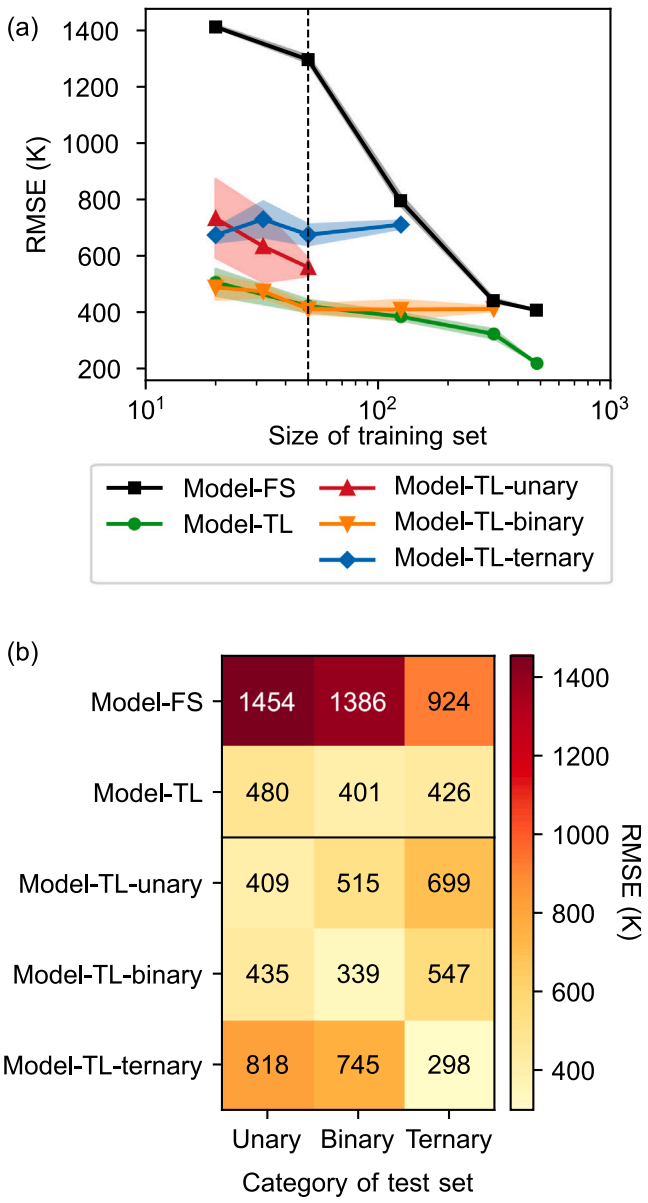

Fig. 4. Test set RMSEs of Model-FS, Model-TL, and its variations. (a) RMSEs for varying the size of the training set. (b) RMSEs for subsets in the test set at training points of 50. The shades in (a) stand for the standard deviation of 5 models.

distinguishing the melting points between crystals that have similar crystal graphs. At an equal training set size of 50 data points [see the dashed line in Fig. 4(a)], the Model-TL-binary achieves the lowest RMSE of 411 K, compared to 559 K and 674 K for the Model-TL-unary and Model-TL-ternary, respectively. This outcome can be attributed to the binary system comprising about 70% of the test set. However, as the training set size increased, the RMSE for the model trained on binary system data deviated from Model-TL, indicating that knowledge from binary systems meets challenges in predicting the melting points of other material classes. In addressing the performance of the model across specific material classes, we have analyzed the impact of varying the subsets in the training set [see Fig. 4(b)]. The results indicate that models tailored to specific material classes demonstrate superior performance in predicting the melting points of their respective target classes compared to the general Model-TL. The improvements in RMSE for Model-TL-unary, Model-TL-binary, and Model-TL-ternary are 71 K, 62 K, and 128 K, respectively. However, it is noteworthy that while

these specialized models excel in their respective domains, they exhibit increased RMSEs when predicting melting points outside their trained material class. The results from the impact of the size of the training set and subset division regarding transition metals can also be explained in the same manner (see Fig. S3). Therefore, for practitioners focused on a particular material class, prioritizing data collection and training within that specific subset offers a more effective approach than compiling a more generalized dataset.

4. Conclusions

In summary, we suggested a TL procedure for predicting the T_m for inorganic crystals by pre-training the ML model with a large DFT database of atomization energies. Owing to the underlying relation between T_m and atomization energy, TL provides a significant enhancement of about 46% in accuracy compared to training T_m from scratch. Other pre-training features such as formation energy and band gap energy resulted in RMSEs of 269 and 450 K, respectively, which are higher than 218 K from atomization energy due to their low resolution in specific crystals and weaker correlation with target properties.

Moreover, it effectively reduced the deviation in error levels between different crystal groups, enabling more reliable predictions for a broader range of crystals. Although we observed the enhancement of TL is diminished on ternary crystals where the correlation between the pre-training feature and target property is concerned, by using only the atomization energy of ternary crystals, we can reduce RMSE from 230 K to 207 K with respect to ternary crystals. The capability of TL in effective task adaptation for a limited dataset was demonstrated, as it achieves a similar level of accuracy while utilizing only about 10% of the original dataset. Additionally, we confirmed that accuracy enhancement respective to the size of the dataset also varies by the composition of the training set. The present study illustrates how the computational database can effectively mitigate the scarcity of experimental data, a common challenge inherent in material science.

CRediT authorship contribution statement

Jaesun Kim: Formal analysis, Software, Writing – original draft. **Jisu Jung:** Conceptualization, Formal analysis, Writing – review & editing. **Sookyung Kim:** Writing – review & editing. **Seungwu Han:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The modified GeoCGNN code integrated with TL is available at https://github.com/MDIL-SNU/TL-GeoCGNN. At this repository, the atomization energies and melting points retrieved from the Materials Project (Jain et al., 2013) and CRC Handbook (Lide, 2004), respectively, are also uploaded with their MP-ids and CAS registration numbers as CSV format.

Acknowledgments

The authors acknowledge the financial support from the Virtual Engineering Platform Project of the Ministry of Trade, Industry and Energy (MOTIE) of Korea (grant number: P0022336). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (00247245). The computations were conducted in Korea Institute of Science and Technology Information (KISTI) National Supercomputing Center (KSC-2022-CRE-0283).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.commatsci.2024.112783.

References

[1] W.G. Fahrenholtz, G.E. Hilmas, Ultra-high temperature ceramics: Materials for extreme environments, Scr. Mater. 129 (2017) 94–99, http://dx.doi.org/10.1016/j.scriptamat.2016.10.018.
[2] M. Valant, D. Suvorov, R.C. Pullar, K. Sarma, N.M. Alford, A mechanism for low-temperature sintering, J. Eur. Ceram. Soc. 26 (2006) 2777–2783, http://dx.doi.org/10.1016/j.jeurceramsoc.2005.06.026.
[3] O. Sugino, R. Car, Ab initio molecular dynamics study of first-order phase transitions: Melting of silicon, Phys. Rev. Lett. 74 (1995) 1823–1826, http://dx.doi.org/10.1103/PhysRevLett.74.1823.
[4] D. Alfè, First-principles simulations of direct coexistence of solid and liquid aluminum, Phys. Rev. B 68 (2003) 064423, http://dx.doi.org/10.1103/PhysRevB.68.064423.
[5] L.-F. Zhu, J. Janssen, S. Ishibashi, F. Körmann, B. Grabowski, J. Neugebauer, A fully automated approach to calculate the melting temperature of elemental crystals, Comput. Mater. Sci. 187 (2021) 110065, http://dx.doi.org/10.1016/j.commatsci.2020.110065.
[6] K. Lee, Y. Park, S. Han, Ab initio construction of full phase diagram of MgO–CaO eutectic system using neural network interatomic potentials, Phys. Rev. Mater. 6 (2022) 113802, http://dx.doi.org/10.1103/PhysRevMaterials.6.113802.
[7] F. Guinea, J.H. Rose, J.R. Smith, J. Ferrante, Scaling relations in the equation of state, thermal expansion, and melting of metals, Appl. Phys. Lett. 44 (1984) 53–55, http://dx.doi.org/10.1063/1.94549.
[8] J.H. Li, S.H. Liang, H.B. Guo, B.X. Liu, Four-parameter equation of state of solids, Appl. Phys. Lett. 87 (2005) 194111, http://dx.doi.org/10.1063/1.2128071.
[9] T. Gu, W. Lu, X. Bao, N. Chen, Using support vector regression for the prediction of the band gap and melting point of binary and ternary compound semiconductors, Solid State Sci. 8 (2006) 129–136, http://dx.doi.org/10.1016/j.solidstatesciences.2005.10.011.
[10] Y. Saad, D. Gao, T. Ngo, S. Bobbitt, J.R. Chelikowsky, W. Andreoni, Data mining for materials: Computational experiments with AB compounds, Phys. Rev. B 85 (2012) 104104, http://dx.doi.org/10.1103/PhysRevB.85.104104.
[11] A. Seko, T. Maekawa, K. Tsuda, I. Tanaka, Machine learning with systematic density-functional theory calculations: Application to melting temperatures of single- and binary-component solids, Phys. Rev. B 89 (2014) 054303, http://dx.doi.org/10.1103/PhysRevB.89.054303.
[12] G. Pilania, J.E. Gubernatis, T. Lookman, Structure classification and melting temperature prediction in octet AB solids via machine learning, Phys. Rev. B 91 (2015) 214302, http://dx.doi.org/10.1103/PhysRevB.91.214302.
[13] Q.-J. Hong, S.V. Ushakov, A.v.d. Walle, A. Navrotsky, Melting temperature prediction using a graph neural network model: From ancient minerals to new materials, Proc. Natl. Acad. Sci. USA 119 (2022) e2209630119, http://dx.doi.org/10.1073/pnas.2209630119.
[14] D. Jha, K. Choudhary, F. Tavazza, W.-K. Liao, A. Choudhary, C. Campbell, A. Agrawal, Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning, Nat. Commun. 10 (2019) 5316, http://dx.doi.org/10.1038/s41467-019-13297-w.
[15] S. Pratik, P.-N. Liu, J. Ota, Y.-L. Tu, G.-W. Lai, Y.-W. Ho, Z.-K. Yang, T.S. Rawat, A.S. Lin, Mapping oxidation and wafer cleaning to device characteristics using physics-assisted machine learning, ACS Omega 7 (2022) 933–946, http://dx.doi.org/10.1021/acsomega.1c05552.
[16] H. Yamada, C. Liu, S. Wu, Y. Koyama, S. Ju, J. Shiomi, J. Morikawa, R. Yoshida, Predicting materials properties with little data using shotgun transfer learning, ACS Cent. Sci. 5 (2019) 1717–1730, http://dx.doi.org/10.1021/acscentsci.9b00804.
[17] J. Lee, R. Asahi, Transfer learning for materials informatics using crystal graph convolutional neural network, Comput. Mater. Sci. 190 (2021) 110314, http://dx.doi.org/10.1016/j.commatsci.2021.110314.
[18] K. Sankaran, S. Clima, M. Mees, G. Pourtois, Exploring alternative metals to Cu and W for interconnects applications using automated first-principles simulations, ECS J. Solid State Sci. Technol. 4 (2015) N3127–N3133, http://dx.doi.org/10.1149/2.0181501jss.
[19] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: The materials project: A materials genome approach to accelerating materials innovation, APL Mater. 1 (2013) 011002, http://dx.doi.org/10.1063/1.4812323.
[20] D.R. Lide, CRC Handbook of Chemistry and Physics Vol. 85, CRC Press, 2004.
[21] J. Cheng, C. Zhang, L. Dong, A geometric-information-enhanced crystal graph network for predicting properties of materials, Commun. Mater. 2 (2021) 1–11, http://dx.doi.org/10.1038/s43246-021-00194-3.
[22] D. Zagorac, H. Müller, S. Ruehl, J. Zagorac, S. Rehme, Recent developments in the inorganic crystal structure database: theoretical crystal structure data and related features, J. Appl. Crystallogr. 52 (2019) 918–925, http://dx.doi.org/10.1107/s160057671900997x.

[23] H. Okamoto, T.B. Massalski, Correct and incorrect phase diagram features, in: Methods for Phase Diagram Determination, Elsevier, 2007, pp. 51–107.
[24] A. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, Pattern Recognit. 38 (2005) 2270–2285, http://dx.doi.org/10.1016/j.patcog.2005.01.012.
[25] K.-L. Tsui, An overview of taguchi method and newly developed statistical methods for robust design, IIE Trans. 24 (1992) 44–57, http://dx.doi.org/10.1080/07408179208964244.
[26] M. Rahm, R. Hoffmann, N.W. Ashcroft, Atomic and ionic radii of elements 1–96, Chem. Eur. J. 22 (2016) 14625–14632, http://dx.doi.org/10.1002/chem.201602949.
[27] J. Ding, L. Du, G. Pan, J. Lu, X. Wei, J. Li, W. Wang, J. Yan, Molecular dynamics simulations of the local structures and thermodynamic properties on molten alkali carbonate K_2CO_3, Appl. Energy 220 (2018) 536–544, http://dx.doi.org/10.1016/j.apenergy.2018.03.116.
[28] C.J. Bartel, A. Trewartha, Q. Wang, A. Dunn, A. Jain, G. Ceder, A critical examination of compound stability predictions from machine-learned formation energies, npj Comput. Mater. 6 (2020) 1–11, http://dx.doi.org/10.1038/s41524-020-00362-y.