SN Computer Science (2022) 3:465
https://doi.org/10.1007/s42979-022-01370-z

ORIGINAL RESEARCH

Contrastive Metric Learning for Lithium Super-ionic Conductor Screening

Boyu Zhang[1,2] · **Shuo Wang**[3] · **Fuchang Gao**[4]

Received: 6 April 2022 / Accepted: 12 August 2022 / Published online: 3 September 2022

Abstract

High-performance Li-ion battery significantly impacts modern society, and materials with high conductivity play critical roles in battery development. Machine learning (ML) technologies have rapidly changed the field in recent years. However, it is still challenging to predict the high conductors directly due to the lack of validated conductor samples. This paper presents a succinct but effective metric-learning framework for high conductor screening. The material structures are mapped to an optimized feature space using a Siamese network, and an instance-based method is used to classify the input sample. The experiments demonstrate that the proposed method could effectively extract knowledge from imbalanced data and has good performance and generalization ability.

Keywords Metric learning · Contrastive learning · Graph neural network · Material science · Conductor screening

Introduction

Electrode and electrolyte materials with high room-temperature Li-ion conductivity play key roles in developing high-performance batteries. Accurate prediction of material conductivity has significant scientific interest and technological importance. Due to the high computational cost of the first principle (viz., ab initio) methods, it is essential to develop high-efficiency ML technologies to screen candidate materials. The noticeable progress in machine learning has benefited a wide range of areas of chemistry, such as the prediction of new materials and the calculation of material properties. In this paper, we investigate the prediction of materials with high Li-ion conductivity using deep learning.

Several machine learning technologies have been applied to predict materials with high conductivity. For example, Ahmad et al. [1] used a graph convolution neural network [2] to predict a series of mechanical properties of inorganic solids and fed these properties into a theoretical framework [3, 4] to check whether the solids quantify the dendrite initiation with Li metal anode. The screened candidates present an opportunity to obtain both desirable mechanical properties and fast ion conduction. Cubuk et al. [5] embedded atoms to feature vectors and developed an algorithm to screen potential solid lithium-ion conductors using transfer learning. There was one notable advantage that their method is based on formulas, which allowed them to screen billions of candidates in a short period.

However, two major open questions remain in the area. On the one hand, modeling and computational simulation are still challenging for materials due to their inherent complexity. Unlike the fix-sized, two-dimensional image data, the material structures vary in three-dimensional space with an indefinite number of atoms. The varied structures significantly decelerate the applications of advanced machine learning technologies in the area. On the other hand, the percentage of validated materials with high conductivity

This work is financially supported by the National Science Foundation Grant No. NSF 1940270 and NIH Grand P20GM104420.

✉ Boyu Zhang
boyuz@uidaho.edu

1 Institute for Modeling Collaboration and Innovation, University of Idaho, 875 Perimeter Dr MS 1122, Moscow, ID 83844-1122, USA

2 Institute for Interdisciplinary Data Sciences, University of Idaho, 875 Perimeter Dr MS 1122, Moscow, ID 83844-1122, USA

3 Department of Materials Science and Engineering, University of Maryland, College Park, MD 20742, USA

4 Department of Mathematics and Statistical Science, University of Idaho, 875 Perimeter Drive, MS 1103, Moscow, ID 83844-1103, USA

compared to the overall number of candidate materials is minuscule. It is difficult to extract meaningful knowledge from such imbalanced data, and the learned model could be biased due to poor generalization capability. This paper tackles the above problems with a metric-learning framework that transforms the material structures to an optimized latent feature space using k-Nearest Atom Graph Neural Network (k-NAGCN) [6], which is an adaptive Graph Convolution Neural Network (GCNN) for material modeling. The k-NAGCN is trained under the Siamese Network setup to ensure the expected feature space maximizes the distance between the validated fast Li-ion conductors and the rest of the materials and minimizes the distances within both groups. From the feature selection perspective, the training strategy forces the network to focus on the features (local environments) related to conductivity, thus improving the generalization and convergence. A similar strategy has been widely applied to biology identification, such as face recognition, where the data is imbalanced. However, such a method has not been applied to material prediction to our knowledge. After that, a potential candidate could be evaluated using instance-based learning, such as the k-NN algorithm [7], with a group of validated fast Li-ion conductors.

Our contributions are summarized as follows.

- We propose a generic metric-learning framework for fast Li-ion conductor screening. The proposed method could effectively extract knowledge from the imbalanced dataset.
- We introduce the k-NAGCN model to describe the complex material structures and simulate the interatomic interaction.
- We present a comprehensive experiment on a crystal conductivity dataset. The experiment results demonstrate that our method achieves satisfactory performance, even when only a small number of validated fast Li-ion conductors are available.

Related Work

Structure descriptors In material science, an ideal description should be able to capture all the relevant information that is necessary to distinguish materials from others [8]. Generally, the descriptors' quality decides the performance's upper bound. Using the ML perspective, existing features could be categorized into knowledge-guided and data-driven descriptions. The former category usually involves elaborate design by specialists in the field of quantum chemistry. For instance, the Coulomb matrix representation and its extensions [9–11], which are based on Coulomb repulsion between atoms and a polynomial fit of atomic energies to the nuclear charge, have been frequently used. One could find these descriptions reflect higher-level characteristics of the materials. The latter category represents the distribution of each atom and its local structural environment based on the detailed atom distribution. The extracted knowledge is kept in the models and the restrictions. Typical descriptions in this category include atom-centered description [12–16] and graph-based descriptions [2, 11, 17–21]. For these methods, no high-level knowledge is needed when constructing the descriptions, but the knowledge could be learned from the available data samples. The extracted knowledge is implicit in the computational models.

Metric learning Metric learning has been extensively researched in the past decades [22–24]. The core idea of metric learning is to transform the original samples, which are usually more complex, into a new feature space that is optimized for the specific task. Metric learning could be both supervised and unsupervised, based on different target problems. Typical unsupervised metric learning methods involve Principle Component Analysis (PCA) [25], Multiple Dimension Scaling (MDS), Local Linear Embedding (LLE) [26], and so on. The metric learning research involves a series of optimization methods such as Support Vector Machine (SVM) [27], kernel methods [28], and Deep Learning. One successful application of metric learning is face verification [29]. The proposed Siamese network structure that shares parameters that could effectively handle the imbalanced data and its deep learning extension, the 'Deepface' model [30], has been recognized as a major landmark of the development of artificial intelligence (AI) technology. Moreover, the structure has been successfully applied on a series of problems such as signature verification [31], object tracking [32], and one-shot image recognition [33].

In this paper, our framework combines the metric-learning methodology with the graph-based structure description. The former could help us to extract desired knowledge from the extremely imbalanced data, and the latter could effectively describe the complex structures and simulate the interatomic interaction.

Material Screening Framework

The proposed high conductor screening framework has two major stages, i.e., the metric learning and the instance-based learning stage. In this section, we first illustrate the architecture of the Siamese network with the k-NAGCN backbone for metric learning. Then we go into the details of the instance-based algorithm for material classification. A brief discussion is provided at the end of this section.

Siamese Network for Metric-Learning

Metric learning aims to find an optimized mapping function that projects the materials from a high-complexity structure space to a task-related latent feature space. For high conductor screening, the latent features are expected to capture the distinguishing characteristics of the materials. And thus, the high conductivity materials could be easily distinguished from the other lithium crystals and compounds. The task-related metric learning algorithm is effective when the positive sample are rare.

This paper uses Siamese network structure with contrastive loss [34] to find the optimized mapping function.

The architecture of the Siamese network is presented in Fig. 1.

As the name implies, the Siamese network contains two network branches that share the parameters. It needs a pair of input samples, x_i and x_j, to get a pair of latent vectors during the training process. In practice, each sample goes through the network G separately and gets the feature vector $G(x_i)$ and $G(x_j)$. Since the prediction of high conductivity materials is an information retrieval or binary classification problem, the contrastive loss is calculated as follows.

$$L = (1 - Y) \times \frac{1}{2}\left(D_{\mathrm{W}}\right)^2 + \frac{Y}{2} \times \max\left(0, m - D_{\mathrm{w}}\right)^2, \quad (1)$$

where D_{w} is the distance between $G(x_i)$ and $G(x_j)$, and m is an empirical parameter. $Y = 0$ means x_i and x_j belong to the same category. That is, both samples are fast Li-ion conductors or neither. $Y = 1$ means one of the two samples is high conductivity material, and the other is not. Conceptually, when the two input samples belong to the same category, they should be clustered as near as possible; otherwise, they should be separated to a distance m. By minimizing the loss L, the network G will be able to map the sample into an embedded feature space in which fast Li-ion conductors could be easily distinguished from the rest of the materials (see Fig. 2).

Graph-Based Description

The above framework could work in concert with most description methods we mentioned in Section "Related Work". This paper uses an improved graph convolution neural network (k-NAGCN) as the base network in the metric learning framework, considering the characteristics of the crystal structures.

The k-NAGCN model treats the unit cell as a 3D point cloud and describes the atoms using the 3D graph. The prior knowledge extracted from existing materials is summarized in the initial atom features attached to the vertices in the graph. The convolution mechanism in the k-NAGCN model involves the k nearest atoms simultaneously, which is different from the adding up over the pairwise convolution, which could be found in most previous GCNN models. The new convolution mechanism, which involves the 3D atomic positions, could better simulate the atomic interactions and

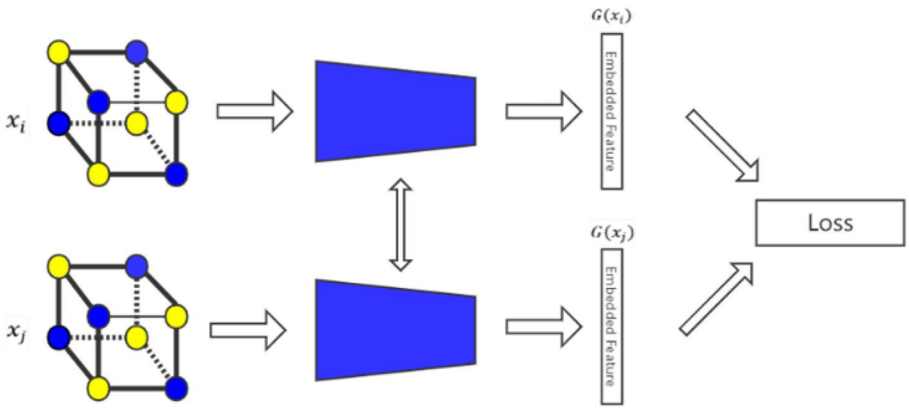

Fig. 1 Siamese network for high conductor prediction

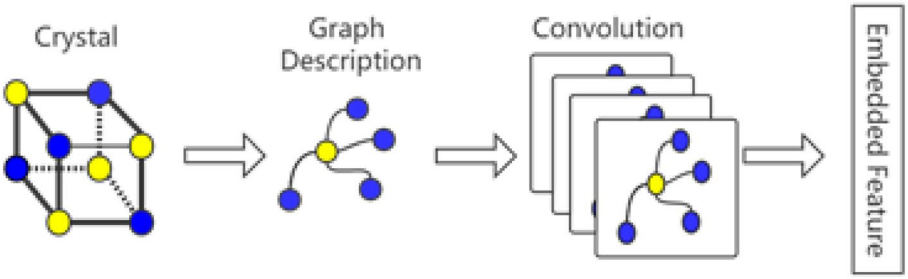

Fig. 2 The structure of k-NAGCN model

achieves superior performance on a series of property prediction tasks. More details about the k-NAGCN model could be found in [6].

Instance-Based Material Screening

High conductors are rare. Only a hand full of validated lithium crystals and compounds are fast conductors. Moreover, only a small portion of all lithium materials have been validated. Therefore, flexibility and generalization capacity are two of the most expected features of the screening methods.

The instance-based method is used in this paper. The effectiveness of the instance-based approach depends on the validity of the previously learned mapping function and the representativeness of the validated sample. The positive sample set (a.k.a., the high conductors) is extremely small, and a complex classification model will overfit easily. The instance-based approach is less prone to overfitting than more complex classification models, such as ANN and SVM. Moreover, the performance of the instance-based approach is expected to improve along with the discovery of new high conductors.

For new material, if it is closer to the verified high conductivity materials in the hidden space and further away from the other materials, the probability that the material has a high conductivity rate is higher. Note the set of verified materials as $X = \{X_h, X_l\}$, where X^h is the set of high conductivity materials, and V^l is the set of other verified materials. Then for an input material x_q, the ratio of the distances is used to evaluate the potential of x_q, as follows.

$$R_{x_q} = \frac{M}{N} \frac{\sum_{i=1,\ldots,N} dist\left(\dot{G}(x_q), G\left(x_i^h\right)\right)}{\sum_{i=1,\ldots,M} dist\left(G(x_q), G\left(x_i^l\right)\right)}, \tag{2}$$

where $dist(a, b)$ is the distance between vector a and b, $G(\cdot)$ is the graph model. And thus, $G(x_q)$ is the corresponding feature vector of material x_q in the latent feature space. The materials with higher R values are more likely to have high conductivities.

Experiments and Results

To validate the proposed framework, we collected experimental data from previous publications. The verified high conductivity materials were reported by He et al. [35]. Besides, the materials that had been proven incapable of being used to develop lithium-ion batteries were used as negative samples.

In detail, 23 materials were identified from [35] according to the common industry standard (room temperature conductivity higher than $(1 \times e^{-4})$. Similarly, 264 negative samples were identified from the previous publications. The negative samples are unqualified for different reasons, including conductivity lower than threshold $(1 \times e^{-4})$, containing specific elements, and so on. The material index in ICSD and the room temperature conductivity of the experimental materials were given in Appendix 1. The method used to verify materials could be found in [35].

As the first step, we verify the effectiveness of the graph-based model (k-NAGCN) in the metric learning process. With the total 287 training samples, we created 82, 369 sample pairs available for training because the Siamese network takes pairwise input. The generated sample pairs were shuffled and split into training (80%), validation (10%), and testing (10%). The model was trained on the training set and evaluated on the validation set for each epoch, and the model was tested on the testing set after the training was finished.

For the hyperparameters in the k-NAGCN, we referred to the conclusion in [6], and set the number of graph convolution layers as three, the number of fully-connected layers as two, feature-length as 96, and the number of atoms involved in the convolution k as 12. The k-NAGCN with the above configuration successfully predicted three conductivity-related properties and was expected to be effective for conductivity prediction.

Figure 3 presents the proposed model's converge curve of the contrastive loss. The k-NAGCN with the above hyperparameters presents good convergence on the experimental data. The training process effectively reduced the contrastive loss, which means the latent feature space maximizes the distances between high conductivity materials and the rest materials and minimizes the distances within high conductivity materials.

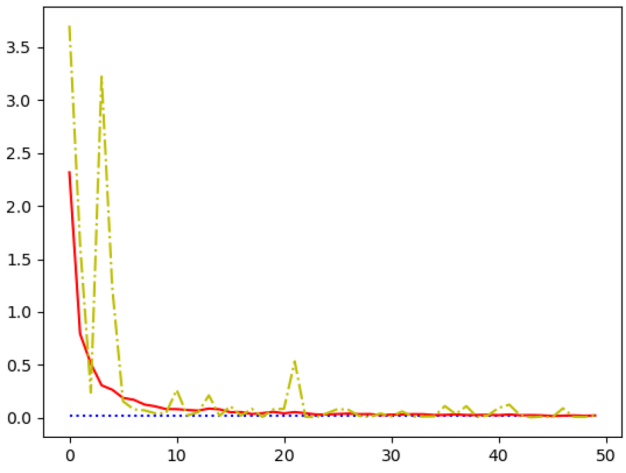

Fig. 3 The training curve using the contrastive loss. The solid red line is the loss of training data, the yellow dash-dot line is the loss of validation data, and the blue dot line is the value of contrastive loss for testing data

The proposed framework was tested for the classification task with the latent feature space learned. 'leave-one-out' cross-validation was used in the experiment. In detail, the experiment was repeated $N + M$ times, where N and M are the number of positive and negative samples, respectively. For each time, one material was 'leave-out' as the testing sample, and the rest of the samples were used to train a k-NAGCN model. The training process used 90% pairs as training and the rest 10% pairs as validation. The training process stopped when the contrastive loss did not reduce for 15 epochs, or the total training epochs reached 50. Then the instance-based method described in Section "Instance-Based Material Screening" was used to predict the label of the testing sample.

The performance was evaluated with three metrics, accuracy, sensitivity, and specificity.

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{3}$$

$$\text{Sens} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{4}$$

$$\text{Spec} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{5}$$

The proposed method could work with most distance measurements. During our experiments, we tested Cosine distance and Euclidean distance. We also tested the impact of hyperparameter k, which represents the number of neighbor atoms involved in calculating graph convolution. It could be found in Table 1 that the proposed method has satisfactory performance for the fast Li-ion conductor screening, with both high specificity and sensitivity. For the instance-based method, the Euclidean distance performed better than the Cosine distance on sensitivity and overall accuracy. Moreover, the impact of k is significant. In general, increasing the number of involved atoms benefited the performance, where the model with $k = 12$ outperformed the model with $k = 4$ and $k = 8$. However, the larger k will significantly increase the computational cost and over smooth the atom features, reducing the performance.

The results in Table 1 prove that the learned network can map the structures to an optimized latent feature space where high conductors have a smaller distance from each other. Considering the graph-based description and convolution mechanism, we presume the conductivity of the materials is highly correlated with specific local structures in the unit cell. However, due to the graph-based model's pooling process, the model lacks explainability, and we could not identify the local structure that contributed to the conductivity. Increasing the explainability of the model will be one of the most critical parts of our future work. To a certain extent, the above results support the idea in [35], which claims the conductivity is related to some local topology features, including the Li sites' size and percolation radius, and minimum bond length.

Conclusion and Future Work

In this paper, we propose a succinct but effective metric-learning framework for screening lithium materials with high conductivity. The proposed Siamese network framework, which contains two k-NAGCN that share parameters and perform in parallel, could effectively map the high-complexity material structures into an optimized latent feature space. Then, the instance-based method is used to distinguish the high conductors from the rest of the materials. The experiments demonstrate the effectiveness of the graph-based description and metric learning algorithm. It also proves that the conductivity of the materials is related to specific local structures. In future research, we will continue the study from two directions. First, we will screen materials in the larger dataset using the proposed method and validate the candidates. Second, we will enhance the explainability of the model by introducing an attention mechanism into the graph-based model and investigating the relationship between specific local structures and material conductivity.

Appendix A: Experimental materials

See Table 2.

Table 1 The performance of high conductor prediction using different measurements

	Euclidean			Cosine		
	$k = 4$	$k = 8$	$k = 12$	$k = 4$	$k = 8$	$k = 12$
Accuracy (%)	84.17	86.54	87.11	82.19	85.76	86.76
Sensitivity (%)	86.34	89.42	91.30	83.26	86.96	86.96
Specificity (%)	82.06	85.50	86.74	81.04	83.58	86.74

Table 2 Materials used in experiments

ICSD Index	Comp	SigmaRT
182963		Containing: Ti, Fe
83501	$Li_{5.2}Hf_{12.2}P_{18}O_{72}$	Containing: Hf
85714	$LiPO_3$	Li–P–O
411410	$Li_2B_2Se_5$	Selenide
78030		Li sharing sites with: Ca, Na
241135		Containing: Mo, Fe
248315		Containing: Mo, Fe
169477		Containing: Fe
186520	$LiFeSiO_4$	Containing: Fe
188009		Li–Se–O, Li–S–O, Li sharing sites with: Zn
290033		Containing: Fe
81857	$LiLaTiO_4$	small radius
18313	Li_2TiF_6	9.20E–01
100324		Li–C–O
68463	$Li_4Ge_5O_{12}$	Melting
39604	$LiKBaZnF_6$	Small radius
186519	Li_2FeSiO_4	Containing: Fe
157430		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
39610		Containing: V
16175		Containing: Mo, Fe
184019		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
51754	$Li_3Al(BO_3)_2$	3.60E–01
92395	Li_4SeO_5	Low D
202500		Containing: Mn
55717		Containing: V
84602	$Li_4Mo_3O_8$	Containing: Mo
246132	Li_2FeSiO_4	Containing: Fe
201959	$Li(MoO_3)_3$	Containing: Mo
75900		Containing: Mo, Fe
16176		Containing: Mo, Fe
58		Li–S–O
59742	$Li_{0.1}Mg_{1.6}Cu_{4.3}O6$	Li sharing sites with: Mg
84943	$Li_3Mo_3P_3O_{17}$	Containing: Mo
169799		Li–Se–O, Li–S–O, Li sharing sites with: Zn
51033	$Li_{1.1}Ti_{5.7}O_{12}$	Containing: Ti
162953	$Li_2MgSi_5O_{12}$	Mg blocking
262643	$Li_2In_2GeS_6$	4.50E–02
189825	$LiTiS_2$	2D network
34256	$LiBO_2$	4.50E–01
155721		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
16713		Li–C–O
188761	$LiMg_4Al(MoO_4)_6$	Containing: Mo
9987	$Li_3Ga(BO_3)_2$	3.00E–02
39669	$Li_{1.39}K_2Y_{0.87}F_6$	Small radius
66941		Li–C–O
247557		Containing: Mo, Fe
93650	$Li_2Mn_2(SO_4)_3$	Containing: S, Mn
184018		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
187706		Containing: Mo, Lu, Ho. Li sharing sites with: Y, La, Gd
200222	$LiEu(SO_4)_2$	Containing: Eu
89903	$LiSb_3O_8$	1D network

Table 2 (continued)

ICSD Index	Comp	SigmaRT
62359	$Li_2MgTi_9O_{20}$	Containing: Ti
167775		Containing: Mo, V
409384	Li_3GaF_6	2.40E+00
191532	$Li_7Mn_{4.2}O_{12}$	Containing: Mn
65176		Li–Si–O, Li–Ge–O, Containing: Co
21026		Li–Si–O, Li–Ge–O, Containing: Co
245516	$Li_2Si_3O_7$	Li–Si–O
191593		Containing: Mo, Lu, Ho. Li sharing sites with: Y, La, Gd
261830		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
402111	$LiNa_2GaAs_2$	Pnictide
186498		Containing: Fe
200984	$K_{1.2}Li_{10.8}Sb_{12}O_{36}$	Li sharing sites with: K
180318	$Li_6P_2S_{7.8}$	Li–P–S
20403		Containing: Eu, Mo, Ho
151759		Li sharing sites with: Na
164080	$Li_8Ti_{21}Zn_4O_{32}$	Containing: Ti
262048		Containing: Fe
161794	$LiFe_3SiPO_8$	Containing: Fe
9430	$LiBe_2Na_2F_7$	Melting
9129		Containing: V
202439	$LiAlLa_4O_8$	Small *r*
416278		Containing: Mo, Fe
186451		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
280108	$LiGaSi_2O_6$	Bad network
69300		Li–Si–O
263140	$Na_5Li(Mo_3O_{10})3$	Containing: Mo
184023		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
87414	LiSrN	8.10E–07
188749	$Li_4Zr_8V_{3.2}P_{8.8}O_{48}$	Containing: V
159316		Containing: Ti
23723		Li–S–O
163706		Containing: Mo, Lu, Ho. Li sharing sites with: Y, La, Gd
236294	$Li_2MgBa(PO_4)_2$	Mg blocking
400552	$LiIO_4$	Li–I–O
290325	$Na_{23.3}Li_{0.7}Mo_{44}O_{144}$	Containing: Mo
2512		Li–S–O
248310		Li–Ge–O
26297	Li_6TeO_6	Melting
241136		Containing: Mo, Fe
169864		Containing: Ti, Fe
16689		Containing: V
20558		Containing: V
35728	$LiFeSnO_4$	Containing: Fe
37118	$LiReO_4$	Containing: Re
51188		Containing: Ti, Fe
161187		Containing: Mo, Lu, Ho. Li sharing sites with: Y, La, Gd
156058		Containing: Mo, Fe
247985	$Li_2MnP_2O_7$	Containing: Mn
245699		Containing: V
66942		Li–C–O
85439	$Li_4Mo_5O_{17}$	Containing: Mo

Table 2 (continued)

ICSD Index	Comp	SigmaRT
419061	$LiAsS_2$	1.10E–01
152282		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
30982	$LiAlSiO_4$	2D network
41199		Li–I–O
55718		Containing: V
421130		Li–P–S
67535	$LiGeBO_4$	5.10E–08
42701		Containing: Mo, V
81249		Li–Si–O
71471	$Li_4VAs_2O_9$	Containing: V
401208	LiK_2GaAs_2	Pnictide
51332	$Li_8Ti_{7.4}Fe_{8.6}P_{24}O_{96}$	Containing: Fe, Ti
410714		Li–O–Se
71058	$Li_5P_2N_5$	8.24E+01
100516	$LiBa_3Ti_5Sb_3O_{21}$	Small *r*
247556		Containing: Mo, Fe
188866		Li sharing sites with: Ca, Na
67845		Containing: V
27672	LiK_2AlF_6	Small radius
27008	$Li_3Na_3In_2F_{12}$	Small radius
92312	$Li_3Na_3N_2$	8.10E+00
245933		Containing: Mo, Fe
248309		Li–Ge–O
162961		Containing: Mo, Fe
69133		Li–C–O
86167	$Li_4Co_2Ge_9O_{16}$	Containing: Co
51443		Containing: V
416695	$Li_2Mo_4O_{13}$	Containing: Mo
424700		Containing: Mo, Lu, Ho. Li sharing sites with: Y, La, Gd
416279	$Li_{26.9}Ta_{4.3}Mo_24O_{96}$	Containing: Mo
84775	$Li_3Cu_2SbO_6$	Li sharing sites with: Cu
174132		Containing: Mo, V
202897		Containing: Ti, Fe
93540	$Li_2Ti_3Bi_4O_12$	Bad network
608360	$LiAlS_2$	5.00E–02
200926	$LiIn(MoO_4)_2$	Containing: Mo
241134		Containing: Mo, Fe
88458	$Li_2V_2(SO_4)_3$	Containing: S, V
30276		Li–S–O
249456		Containing: Eu, Mo
2899		Containing: V
8262		Containing: Ti, Fe
21012		Containing: V
165579		Li–Cl–O
183763		Containing: Fe
78819	$Li_{10}BrN_3$	3.60E–01
59243		Li–P–O
245389		Li–C–O
401169		Containing: Mo, Fe
27318	$LiBa_4Sb_3O_{12}$	Small radius
4185		Containing: Mo

Table 2 (continued)

ICSD Index	Comp	SigmaRT
246757		Containing: Mo, Fe
424835	Li_3AsS_3	Low D
184016		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
85171	Li_3AlF_6	2.00E–02
94530		Containing: Fe
20559		Containing: V
245847		Containing: V
99503	$Li_4CaB_2O_6$	2D network
249455		Containing: Eu, Mo
40457	$LiSbS_2$	1.40E–01
39849		Containing: Mo, Fe
402147	LiK_2InAs_2	Pnictide
186040		Containing: Mo, Lu, Ho. Li sharing sites with: Y, La, Gd
32713	$LiPN_2$	Low D
246302	Li_2SeO_4	Low D
50993	$Li_{4.5}Y_{3.3}O_8$	Li sharing sites with: Y
186041		Containing: Mo, Lu, Ho. Li sharing sites with: Y, La, Gd
48106	Li_2SeO_4	Li–O–Se
424698		Containing: Fe
87774	$LiMn_2O_4$	Containing: Mn
281292		Containing: Mo, Fe
30249		Li–Al–O
262642	$Li_2In_2SiS_6$	1.40E+00
15642	Li_2BaSi	Silicide
172581	$LiTiOAsO_4$	Low D
16215		Containing: Mo, Fe
201816		Li–I–O
261829		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
170956		Containing: Mo, Fe
638		Li–P–O
67264	$Li_7Br_3O_2$	1.40E–03
248311		Li–Ge–O
63519	$LiMgIn(MoO_4)_3$	Containing: Mo
25105	$Li_3Ba_2Ti_{9.25}O_{22}$	Ti blocking
34361	$Li_4Ge_9O_{20}$	Li–Ge–O
67536	$LiSiBO_4$	3.10E–01
65126	$LiAlGeO_4$	2D network
186450		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
35676	$LiGeTe_2$	Telenide
184022		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
28388	$LiGaO_2$	2D network
39814		Li–P–O
171375	$LiBF_4$	Melting
84763	Li_5Cl_2N	7.00E–02
187059		Containing: Mo, Lu, Ho. Li sharing sites with: Y, La, Gd
27007	$Li_3Na_3Sc_2F_{12}$	Small radius
169478		Containing: Fe
182966	$Li_2Ti_6O_{13}$	1D network
59640	$Li_4ZnP_2O_8$	Low D
67110		Li–Si–O
201817		Li–I–O

Table 2 (continued)

ICSD Index	Comp	SigmaRT
416101	$Li_{7.23}B_7Se_{15}$	Selenide
71035	Li_6KBiO_6	1.10E+01
30253	$Li_3Na_3Al_2F_{12}$	Small radius
402083	$LiNa_2AlP_2$	Pnictide
246756	$Li_2Zn_2(MoO_4)_3$	Containing: Mo
425095	$Li_7Mn(BO_3)_3$	Containing: Mn
186497		Containing: Fe
98615	Li_4SiO_4	Li–Si–O
182259		Containing: Mo, V
262370		Containing: Eu, Mo, Ho
65127	$LiGaGeO_4$	2D network
54021	$Li_8Bi_2(MoO_4)_7$	Containing: Mo
81074	$LiMo(PO_4)_2$	Containing: Mo
186042		Containing: Mo, Lu, Ho. Li sharing sites with: Y, La, Gd
20032		Li–I–O
51630		Li–P–O
246250	$Li_{1.5}Ta_{10.5}P_{18}O_{72}$	Li sharing sites with: Ta
247558		Containing: Mo, Fe
6134		Containing: Mo
94490		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
94491		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
31050	$Li_6Ge_2O_7$	Li–Ge–O
8222		Li–Si–O
200947	$Li_2Zr(MoO_4)_3$	Containing: Mo
180011	$Li_{0.5}TiO_2$	High Ea
4155		Containing: Mo
424697		Containing: Fe
20346		Containing: Lu
246758		Containing: Mo, Fe
185582		Containing: Fe
14360	Li_2BeF_4	2D network
18004	Li_6ZrBeF_{12}	1.00E–02
261831		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
246859	$Li_4P_2O_7$	Li–P–O
23406	Li_2GeF_6	4.00E–04
32028	LiYSi	silicide
246133	$Li_2V(PO_4)_2$	Containing: V
201390		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
247255	Li_8N_2Se	1.09E+01
84703	$NaLi_2Mo_2P_3O_{14}$	Containing: Mo
32029	LiYGe	Ge anion
418488		Li–P–S
245392		Li–C–O
39580		Containing: Mo, Lu, Ho. Li sharing sites with: Y, La, Gd
20557		Containing: V
415120	Li_2TeS_3	Bad network
245393	Li_4CO_4	Li–C–O
156006	LiMoIO6	Containing: Mo
413238		Li–Cl–O
261832		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
65175		Li–Si–O, Li–Ge–O, Containing: Co

Table 2 (continued)

ICSD Index	Comp	SigmaRT
181478		Containing: Mo, Lu, Ho. Li sharing sites with: Y, La, Gd
79018	$LiMo_3P_3O_{16}$	Containing: Mo
188763		Containing: Mo, V
247257	Li_8N_2Te	1.20E–01
262791		Containing: Mo, Fe
401554	$Li_3Er(NO_3)_6$	Containing: N
300254	$Li_2Pr(NO_3)_5$	Containing: N
422680	LiCNO	CN anion
402341	LiBeN	1.20E–07
246759		Containing: Mo, Fe
200357	$Li_{0.6}Si_{1.3}Zn_{1.1}O_4$	melting
39581		Containing: Mo, Lu, Ho. Li sharing sites with: Y, La, Gd
39608		Containing: V
261834		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
2642		Li–I–O
23815		Li–Al–O
184015		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
48011		Containing: Ti, Fe
248414		Li–P–O
44957	$LiYS_2$	2D network
84649	Li_4NCl	2.80E+00
184021		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
153522	$LiTiOPO_4$	Low D
248415	$Li_4P_2O_7$	Li–P–O
261833		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
35727		Containing: Ti, Fe
39609		Containing: V
426263		Containing: Ti
184020		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
182961		Containing: Ti, Fe
184017		Containing: Mo, Eu, Li sharing sites with: Tb, Nd, La, Gd, Sm
50573	$Li_3Mo_2P_3O_{14}$	Containing: Mo

For positive samples, the conductivities were given. For negative samples, the reasons were given

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Ahmad Z, Xie T, Maheshwari C, Grossman JC, Viswanathan V. Machine learning enabled computational screening of inorganic solid electrolytes for suppression of dendrite formation in lithium metal anodes. ACS Cent Sci. 2018;4(8):996–1006.
2. Xie T, Grossman JC. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. Phys Rev Lett. 2018;120(14):145301.
3. Ahmad Z, Viswanathan V. Stability of electrodeposition at solid-solid interfaces and implications for metal anodes. Phys Rev Lett. 2017;119(5):056003.
4. Sendek AD, Yang Q, Cubuk ED, Duerloo K-AN, Cui Y, Reed EJ. Holistic computational structure screening of more than 12000 candidates for solid lithium-ion conductor materials. Energy Environ Sci. 2017;10(1):306–20.
5. Cubuk ED, Sendek AD, Reed EJ. Screening billions of candidates for solid lithium-ion conductors: a transfer learning approach for small data. J Chem Phys. 2019;150(21):214701.
6. Zhang B, Zhou M, Jianzhong W, Gao F. Predicting the materials properties using a 3D graph neural network with invariant representation. IEEE Access. 2022;10(10):62440–9.
7. Gou J, Lan D, Zhang Y, Xiong T. A new distance-weighted k-nearest neighbor classifier. J Inf Comput Sci. 2012;9(6):1429–36.
8. Schmidt J, Marques MRG, Botti S, Marques MAL. Recent advances and applications of machine learning in solid-state materials science. npj Comput Mater. 2019;5(1):1–36.
9. Rupp M, Tkatchenko A, Müller K-R, Von Lilienfeld OA. Fast and accurate modeling of molecular atomization energies with machine learning. Phys Rev Lett. 2012;108(5):058301.

10. Faber F, Lindmaa A, von Lilienfeld OA, Armiento R. Crystal structure representations for machine learning models of formation energies. Int J Quantum Chem. 2015;115(16):1094–101.
11. Schütt KT, Glawe H, Brockherde F, Sanna A, Müller K-R, Gross EKU. How to represent crystal structures for machine learning: towards fast prediction of electronic properties. Phys Rev B. 2014;89(20): 205118.
12. Bartók AP, Kondor R, Csányi G. On representing chemical environments. Phys Rev B. 2013;87(18): 184115.
13. Behler J, Parrinello M. Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys Rev Lett. 2007;98(14): 146401.
14. Artrith N, Urban A. An implementation of artificial neural-network potentials for atomistic materials simulations: performance for tio2. Comput Mater Sci. 2016;114:135–50.
15. Behler J. Perspective: Machine learning potentials for atomistic simulations. J Chem Phys. 2016;145(17): 170901.
16. Seko A, Takahashi A, Tanaka I. Sparse representation for a potential energy surface. Phys Rev B. 2014;90(2): 024101.
17. Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph networks as a universal machine learning framework for molecules and crystals. Chem Mater. 2019;31(9):3564–72.
18. Schütt KT, Arbabzadah F, Chmiela S, Müller KR, Tkatchenko A. Quantum-chemical insights from deep tensor neural networks. Nat Commun. 2017;8(1):1–8.
19. Kearnes S, McCloskey K, Berndl M, Pande V, Riley P. Molecular graph convolutions: moving beyond fingerprints. J Comput Aided Mol Des. 2016;30(8):595–608.
20. Duvenaud D, Maclaurin D, Aguilera-Iparraguirre J, Gómez-Bombarelli R, Hirzel T, Aspuru-Guzik A, Adams RP. Convolutional networks on graphs for learning molecular fingerprints. 2015. arXiv preprint arXiv:1509.09292.
21. Schütt KT, Sauceda HE, Kindermans P-J, Tkatchenko A, Müller K-R. Schnet—a deep learning architecture for molecules and materials. J Chem Phys. 2018;148(24): 241722.
22. Kulis B. Metric learning: a survey. Found Trends Mach Learn. 2012;5(4):287–364.
23. Yang L, Jin R. Distance metric learning: a comprehensive survey. Mich State Univ. 2006;2(2):4.
24. Bellet A, Habrard A, Sebban M. A survey on metric learning for feature vectors and structured data. 2013. arXiv preprint arXiv:1306.6709.
25. Yang J, Zhang D, Frangi AF, Yang J. Two-dimensional pca: a new approach to appearance-based face representation and recognition. IEEE Trans Pattern Anal Mach Intell. 2004;26(1):131–7.
26. Saul LK, Roweis ST. Think globally, fit locally: unsupervised learning of low dimensional manifolds. In: Departmental Papers (CIS), 2003; p. 12.
27. Wang L. Support vector machines: theory and applications, vol. 177. Berlin: Springer Science & Business Media; 2005.
28. Yin X, Chen S, Enliang H, Zhang D. Semi-supervised clustering with metric learning: an adaptive kernel method. Pattern Recogn. 2010;43(4):1320–33.
29. Chopra S, Hadsell R, LeCun T. Learning a similarity metric discriminatively, with application to face verification. In: 2005 IEEE Computer Society Conference on computer vision and pattern recognition (CVPR'05), 2015; volume 1, pages 539–546.
30. Taigman Y, Yang M, Ranzato M, Wolf L. Deepface: closing the gap to human-level performance in face verification. In: 2014 Proceedings of the IEEE conference on computer vision and pattern recognition, vol 1, 2014. pp 1701–8.
31. Dey S, Dutta A, Ignacio Toledo J, Ghosh SK, Lladós J, Pal U. Signet: convolutional Siamese network for writer independent offline signature verification. 2017. arXiv preprint arXiv:1707.02131.
32. Bertinetto L, Valmadre J, Henriques JF, Vedaldi A, Torr PHS. Fully-convolutional Siamese networks for object tracking. In: Hua G, Jégou H, editors. 2016 European conference on computer vision. Cham: Springer; 2016. pp. 850–65.
33. Koch G. Siamese neural networks for one-shot image recognition. Dissertation, University of Toronto. 2015.
34. Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping. In: 2006 IEEE Computer Society Conference on computer vision and pattern recognition (CVPR'06), 2006; volume 2, pages 1735–1742. IEEE.
35. He X, Bai Q, Liu Y, Nolan AM, Ling C, Mo Y. Crystal structural framework of lithium super-ionic conductors. Adv Energy Mater. 2019;9(43):1902078.