scientific reports

OPEN

Drawing a materials map with an autoencoder for lithium ionic conductors

Yudai Yamaguchi[1], Taruto Atsumi[1], Kenta Kanamori[2], Naoto Tanibata[1], Hayami Takeda[1], Masanobu Nakayama[1✉], Masayuki Karasuyama[2] & Ichiro Takeuchi[2,3,4]

Efforts to optimize known materials and enhance their performance are ongoing, driven by the advancements resulting from the discovery of novel functional materials. Traditionally, the search for and optimization of functional materials has relied on the experience and intuition of specialized researchers. However, materials informatics (MI), which integrates materials data and machine learning, has frequently been used to realize systematic and efficient materials exploration without depending on manual tasks. Nonetheless, the discovery of new materials using MI remains challenging. In this study, we propose a method for the discovery of materials outside the scope of existing databases by combining MI with the experience and intuition of researchers. Specifically, we designed a two-dimensional map that plots known materials data based on their composition and structure, facilitating researchers' intuitive search for new materials. The materials map was implemented using an autoencoder-based neural network. We focused on the conductivity of 708 lithium oxide materials and considered the correlation with migration energy (ME), an index of lithium-ion conductivity. The distribution of existing data reflected in the materials map can contribute to the development of new lithium-ion conductive materials by enhancing the experience and intuition of material researchers.

Research in the field of material informatics (MI) has become increasingly active with recent developments in ceramic material databases and computational performance (CPUs and GPUs). MI combines informatics with materials exploration to accelerate the search for new functional materials. Examples of inorganic materials databases, such as the Materials Project[1] and the Inorganic Crystal Structure Database (ICSD)[2], contain a vast amount of crystal structure data points and material properties estimated using computational simulations. These databases have facilitated many studies in materials science using high-throughput calculations, such as first-principles and force-field calculations[3–5]. By applying MI to these databases or via high-throughput studies, researchers can organize complex relationships between the compositions, structures, and physical properties of materials, facilitating the more efficient search of useful materials. The process typically involves converting material data into descriptors x_i, and learning a prediction function that yields the desired physical properties as the objective variable Y[6]. Descriptors numerically represent crystal structures and serve as an interface between crystal structure and material property data. Researchers (users) can propose candidate new materials, convert them into descriptors, and predict their physical properties instantly[4,7–9]. By identifying promising materials in advance, experimental costs can be significantly reduced. Furthermore, analyzing the constructed prediction functions using methods such as importance variable analysis may provide chemical semantic interpretations of the descriptors[10,11]. However, existing methods face challenges in identifying unknown materials that are not registered in databases. Inorganic crystalline material descriptors contain information derived from composition, crystal structure and other characteristics. Although it is easy to generate descriptors from conventional crystal structure descriptions, such as the lattice parameters and fractional coordinates of ions, reconstruction of crystal structures from those descriptors remains challenging. Consequently, the MI approach, which uses such descriptors, cannot predict the physical properties of undiscovered compounds—i.e., non-registered compounds—in the material database. Attempts have been made to evaluate phase stability and other physical properties by constructing machine learning models using only compositional information, without using structure-derived

[1]Department of Advanced Ceramics, Nagoya Institute of Technology, Gokiso, Showa, Nagoya, Aichi 466-8555, Japan. [2]Department of Computer Science, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi 466-8555, Japan. [3]RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan. [4]Faculty of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan. ✉email: masanobu@nitech.ac.jp

descriptors[12–14]. A composition and its descriptors are reversible, and users can easily obtain evaluation results by simply considering their compositions. However, the accuracy of evaluation results is greatly reduced as material properties are highly dependent on both composition and crystal structure information. Nevertheless, there is significant potential in areas outside the scope of existing material databases. According to a prior report[15], it is estimated that only 16% of inorganic compounds have been discovered in ternary systems and merely 0.6% in quaternary systems. In another paper[16], it was estimated that, in compositional combinations considering charge neutrality, there were 30 million cases in ternary systems and 30 billion cases in quaternary systems. This indicates that novel functional materials are likely hidden in the vast space of undiscovered materials, necessitating an efficient exploration approach. The conventional MI approach alone is inefficient for determining the compositions and structures of undiscovered materials with desired properties. To support specialized researchers and enhance their intuition, this study introduces the concept of a materials map, which visualizes existing datasets in a two-dimensional space. Even today, numerous groundbreaking material discoveries are believed to heavily rely on the intuition and expertise of exceptional specialists. Therefore, the development of MI-derived tools that can effectively and efficiently support these experts is expected to accelerate the discovery process.

One challenge experienced by researchers when using the MI approach arises from the utilization of high-dimensional descriptors as inputs for the MI scheme. To optimize the intuition and perspective of researchers, this study introduces the concept of a *materials map*, which transforms existing datasets into a two-dimensional space. This map evaluates materials based on two intuitive scales: chemical composition and crystal structure. Consequently, it enables the specification of the search range for known materials while simultaneously identifying unexplored materials. Notably, 2D materials maps using fundamental properties, such as electronegativity and ionic radius, have been proposed in several prior studies[17,18]. However, the experience of researchers invaluable when selecting relevant coordinates since conventional materials maps only consider specific indices. Therefore, in this study, the descriptors were dimensionally reduced using an autoencoder, a type of deep learning model, while retaining maximum information related to composition and structure. Furthermore, by applying a neural network structure associated with the target properties, we developed a materials map tailored to user-specified target properties.

Dataset

To validate the materials mapping method developed in this study, we present an application example focusing on candidate solid electrolyte materials in all-solid-state batteries, which are attracting attention as a next-generation energy source. One crucial property of solid electrolyte materials is their high Li-ion conductivity. We used a dataset comprising the data of 708 conductivity simulations of Li–O-based inorganic solid materials, which we reported previously[19]. All compounds were selected from the Materials Project database and matched two conditions: (1) inclusion of Li and O, and (2) incorporation of 25 elements (Mg, Ca, Sr, Ba, Sc, Y, La, Ti, Zr, Hf, Nb, Ta, Zn, B, Al, Ga, In, C, Si, Ge, Sn, P, As, Sb, S). For samples meeting these conditions, the migration energy (ME, denoted as E_{mig} [eV]) of lithium-ion diffusion in the solids was calculated using an automated ion conductivity evaluation algorithm based on the Bond Valence Force Field (BVFF)[20] method, where smaller values of E_{mig} indicate a smaller barrier to Li-ion diffusion and thus better conductivity. The compositional information considered in this study includes the chemical composition ratio and various properties of constituent elements (atomic number; atomic weight; electronegativity; melting point in the metallic state; atomic/ion/covalent bond radii; number of s-, p-, d-, and f-electrons; and Mendeleyev number) expressed in histogram form[21]. Three types of structural information were employed: the Radius Distribution Function (RDF), the Angular Distribution Function (ADF), and the Voronoi diagram. Compositional and structural descriptors were represented by numerical sets of 1782 and 858 dimensions, respectively. All descriptors and E_{mig} values were scaled by the maximum value for each type (compositional descriptors, RDF, ADF, Voronoi diagrams, and E_{mig}) and converted to a range of 0 to 1. All data tables are available in the Supporting Information.

Methods

In this study, an autoencoder was used to compress each of the compositional and structural descriptors into one dimension each. The autoencoder[22] is a neural network model with specific constraints on the number of neurons and layer shapes. It is an unsupervised machine learning technique that can effectively reduce the dimensionality of input information. Upon receiving an *n*-dimensional input, the number of neurons (dimensions) is reduced (encoded) to the user-specified dimensionality through the application of fully connected layers and activation functions (encoded values), and the encoded values are then expanded (decoded) to ensure that the output's dimensionality matches that of the input (decoded values). When the autoencoder successfully reproduces the input descriptors in its outputs, the encoded values effectively capture all the information from the original descriptors. In this study, the compositional and structural descriptors were compressed to one dimension each and then used to generate a two-dimensional map based on the compositional and structural axes. Furthermore, an intermediate layer was introduced to learn the relationship between ME and the 2D information in the coding layer, allowing for the creation of a materials map associated with ME. A schematic of the neural network model used in this study is presented in Fig. 1. To optimize the model weights, loss functions were defined as Eqs. (1) and (2), corresponding to minimization of the mean squared error (MSE).

$$Loss_{comp} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d} \left(Out\ comp_{i,j} - In\ comp_{i,j} \right)^2, \tag{1}$$
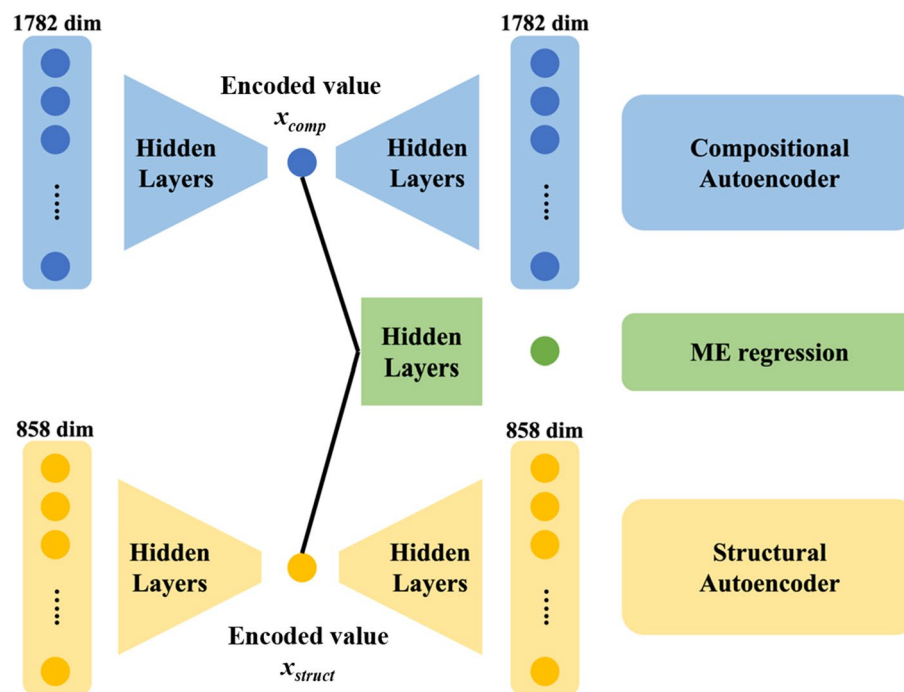
Figure 1. Autoencoder architecture for the materials map. Compositional and structural descriptors are condensed into a one-dimensional format, producing encoded values, denoted as x_{comp} and x_{struct}, which are linked to the migration energy (ME) as the target objective.

$$Loss_{str} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d} \left(Out\,str_{i,j} - In\,str_{i,j} \right)^2, \tag{2}$$

where n and d denote the number of samples and descriptors respectively.

In addition, the two encoded variables derived from the composition and structure descriptors are associated with the lithium ions' ME using a network structure, as illustrated in Fig. 1. The loss function is defined by Eq. (3), which also considers the MSE.

$$Loss_{ME} = \frac{1}{n} \sum_{i=1}^{n} (Pred_i - True_i)^2. \tag{3}$$

The loss function for the entire model is defined as the sum of $Loss_{comp}$, $Loss_{str}$, and $Loss_{ME}$ multiplied by the weight factor W (Eq. 4).

$$Loss_{all} = Loss_{comp} + Loss_{str} + W \times Loss_{ME}. \tag{4}$$

The weight factor, W, controls the priority assigned to ME prediction. A larger W prioritizes the accuracy of ME regression, whereas a smaller W prioritizes the recovery of compositional and structural descriptors. We examined patterns with $W = [0, 0.01, 0.1, 1]$ to balance the accuracy of ME prediction with the reconstruction of input descriptors. The batch size, learning rate, and L2 penalty were set as hyperparameters, and a grid search was conducted to determine the combination that minimized $Loss_{all}$. The number of epochs was determined when $Loss_{all}$ did not improve for 80 consecutive optimizations using the Adam optimizer[23]. Hyperparameter tuning resulted in a batch size of 16, learning rate of 0.001, and L2 penalty of 10^{-6}. To train the materials map autoencoder, 80% of the data were randomly selected and assigned to the training set, and the remaining 20% were used as test data to verify the generalizability of the model.

Results and discussion

Figure 2 shows values of $Loss_{comp}$, $Loss_{str}$, and $Loss_{ME}$ obtained for different values of W, demonstrating a trade-off relationship between the reconstruction of composition and structure descriptors ($Loss_{comp}$ and $Loss_{str}$) and the regression of migration energy ($Loss_{ME}$) as a function of W. We adopted a weight factor of $W = 0.1$ to balance the performance of both processes. Figure 3a–c depict diagnostic plots of the test data for composition, structure, and ME regression predictions at W = 0.1. Flattened values for the composition (1782 dimensions) and structural (858 dimensions) descriptors are displayed in Fig. 3a,b. The coefficient of determination, R^2 score, exceeded 0.8, indicating the accurate reconstruction of both descriptor types. Figure 3d–f compare the input and output descriptors for the three selected samples that exhibit the largest loss values ($Loss_{all}$), confirming that the rough shapes of the descriptors were reconstructed. On the other hand, the ME regression (Fig. 3c) exhibits a very low
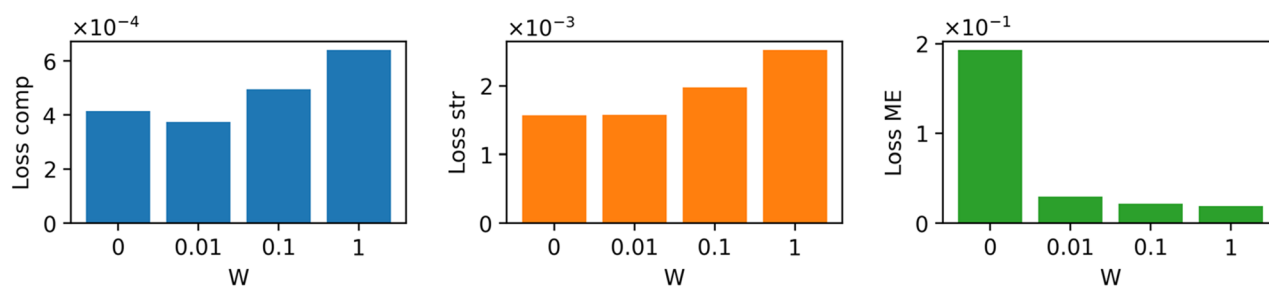
Figure 2. Loss functions of (**a**) compositional descriptors, (**b**) structural descriptors, and (**c**) ME regression, represented as functions of the weight parameter W. This parameter governs the trade-off in performance between composition/structure reconstruction and ME prediction.

R^2 score of 0.29, which stems from the insufficient prediction power of the original compositional and structural descriptors as reported in our paper[19].

Figure 4 shows a map of the constructed materials, with the horizontal axis obtained by compressing the compositional descriptor (1782 dimensions) to one dimension, and a vertical axis likewise obtained by compressing the structural descriptor (858 dimensions) to one dimension. Because each axis was produced by an autoencoder, no units or clear chemical interpretation are present. The colors of the plots in the figure correspond to the MEs of lithium ions, with larger and smaller values represented in dark red and dark blue, respectively. Note that the research and development of battery materials requires high lithium-ion conductivities, which correspond to materials with a low ME. Encoded values for compositional and structural descriptors, and calculated and predicted ME for all 708 materials are listed in Supporting Information. In the figure, the crowed plots indicate areas where many materials with similar compositions and structures have been reported, and where the material search has progressed relatively. However, such studies have never been conducted in empty plot areas, which correspond to unknown materials. These areas are considered risks in the context of material exploration owing to the high possibility of materials not being synthesized. Hence, the sparse plot areas might be promising in terms of both synthesizability and discovery of unknown materials. In Fig. 4, the plotted colors shift from red to blue (corresponding to a decrease in ME) as the value of structural coding increases, confirming that ME is primarily linked to structural information. Appropriate control of the crystal structure is crucial in obtaining highly conductive lithium-ion materials. As the compositional coding increased, the plot distribution became narrower, concentrating in a range from −2 to 2 on the structural axis. Two possible reasons are suggested: (1) the range of possible crystal structures is limited with respect to chemical composition, and such a combination of composition and structure is not possible; or (2) vacancies in the map correspond to materials that researchers have not yet worked with. In any case, the materials map presented in Fig. 4 may be useful for summarizing reported research, visualizing unknown material areas, and providing a bird's-eye view for materials researchers. In particular, the inclusion of an ME regression component in the autoencoder shown in Fig. 1 is largely beneficial in the search of high-ionic-conductivity materials, as the user can roughly evaluate the ME even in the non-plotted coordinates of the material map. Further details are discussed herein.

Although the significance analysis of variables encoded by an autoencoder is generally a challenging task, we investigated factors affecting the composition and structure axes using a trial-and-error approach. The relationship between composition, structure, and high Li-ion conductivity was also considered. As a result, we infer that the lithium-to-oxygen ratio is thought to be one of the factors affecting the compositional axis. In Fig. 5, the molar ratio of (a) Li to (b) O in the composition is plotted as a function of the encoded value of the compositional descriptor x_{comp}. At $x_{comp} > \sim 1.0$, the molar ratio of Li tends to be $< \sim 0.3$, whereas at $x_{comp} < \sim 1.0$, it ranges from 0 to 0.6 (Fig. 5a). Similarly, the molar ratio of oxygen tends to be > 0.55 at $x_{comp} < \sim 1.0$ (Fig. 5b), except for compounds whose molar ratio of oxygen was < 0.3. Thus, the compositional feature at $x_{comp} > \sim 1.0$ appears to reflect relatively low Li and high O concentrations, which indicates that the proportions of lithium and oxygen are factors associated with ME. This conclusion is supported by the connection between x_{comp} and ME in the neural network model illustrated in Fig. 1. However, the correlation between x_{comp} and the ratio of Li and/or O is rather poor even at $x_{comp} > 1$; therefore, numerous other composition-related parameters must be intricately involved in the encoded value of x_{comp}.

The relationships between the molar ratio of lithium to oxygen in the composition and the structural axes is displayed in Fig. 5c,d, respectively. A more distinct dependence on the molar ratios of Li and O is evident on x_{struct} as compared to x_{comp}. Specifically, the molar ratios of Li and O are scattered at less than 0.3 and greater than 0.5, respectively, at $-5 < x_{struct} < 0$. However, an increase in the lithium proportion and decrease in the oxygen proportion are clearly visible at $x_{struct} > 0$. This suggests that the structural axis correlates with the lithium and oxygen concentrations in the lattice at $x_{struct} > 0$. We inferred that the compositional information of the molar ratios of Li or O was indirectly extracted from the radius distribution function (RDF) of Li–O, Li–Li, etc., although these descriptors were classified as structural when $x_{struct} > 0$. In addition, regions exceeding zero on the structural axis represent clusters of materials with relatively small ME (blue distribution). Therefore, we suggest that the dense distribution of lithium and low oxygen concentration in the lattice are advantageous for achieving high ionic conduction.

The Li–O materials used in this study include representative host structures for fast Li-ion conductors, such as the Perovskite-type[24–26], Garnet-type[27–31], and NASICON-type structures[32–38]. Specific ID numbers from the
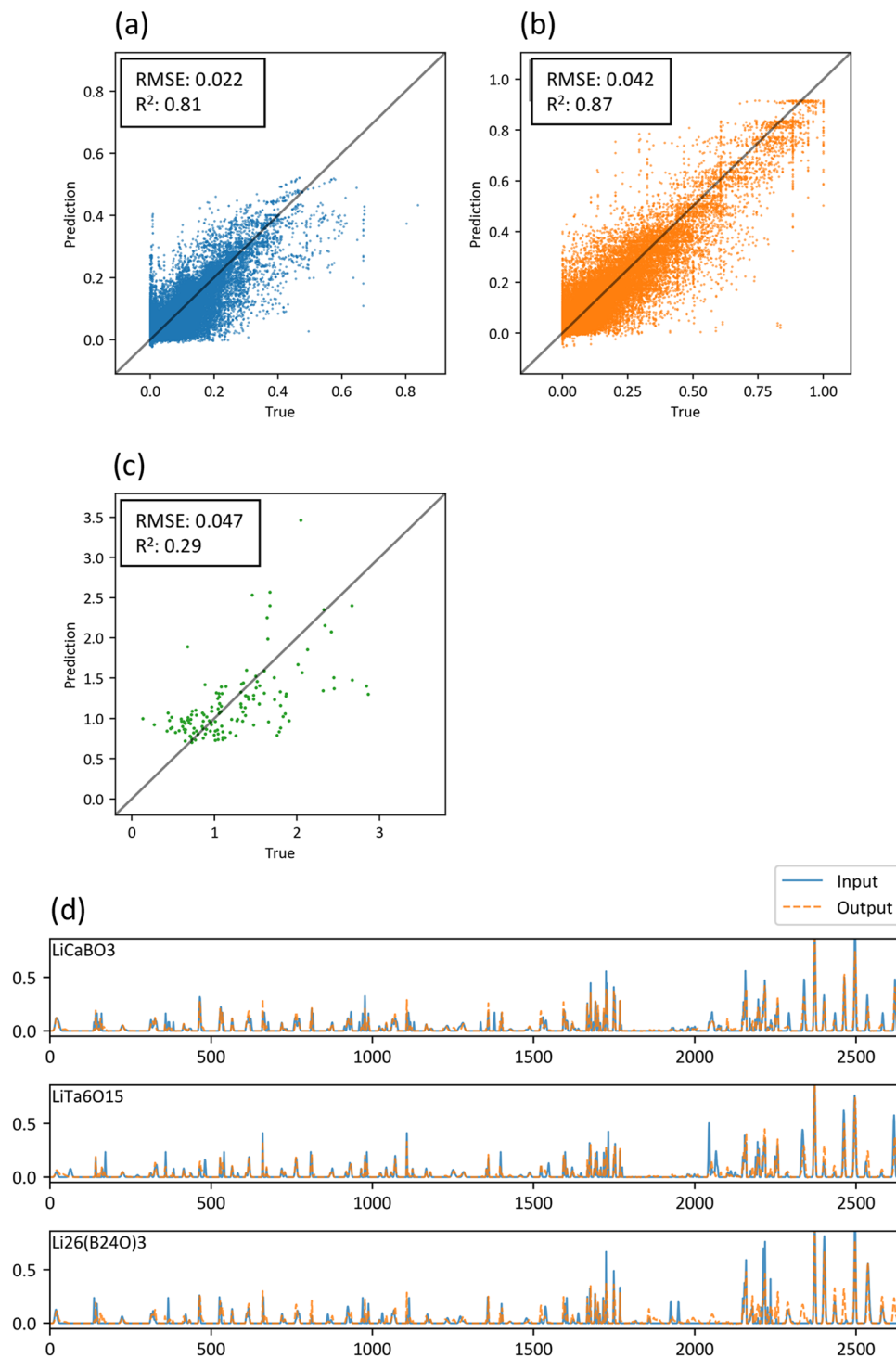
Figure 3. Reconstruction evaluation of compositional/structural descriptors and ME prediction performance of the materials map autoencoder. Diagnosis plots of (**a**) compositional and (**b**) structural descriptors, as well as (**c**) ME predictions, along with (**d–f**) three examples of input and output histogram-descriptors that correspond to the three poorest reconstruction loss functions. Compositional descriptors are represented by numbers 0–1782, while structural descriptors are represented by numbers 1783–2640.

Materials Project database (mp-id), chemical compositions, ME values, and encoded values for composition and structure descriptors are listed in Table 1. These data points are plotted on the material map in Fig. 6a, where each structure is relatively clustered. However, the garnet- and NASICON-type structures are located
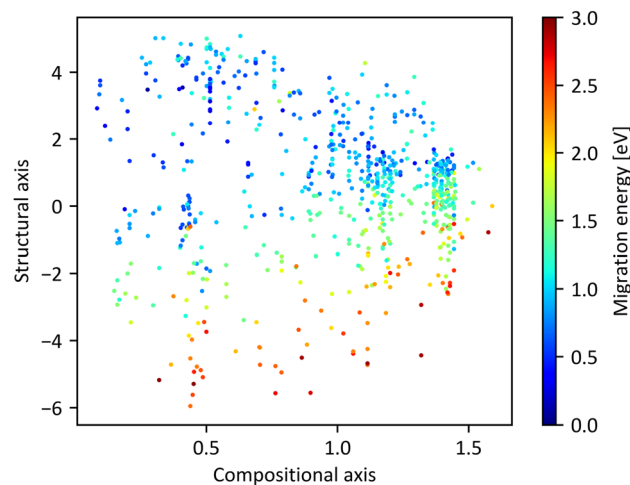
Figure 4. Map for Li-ion conductive materials in the Li–O system. The horizontal and vertical axes correspond to the compressed compositional and structural descriptors, respectively, achieved through the autoencoder approach.
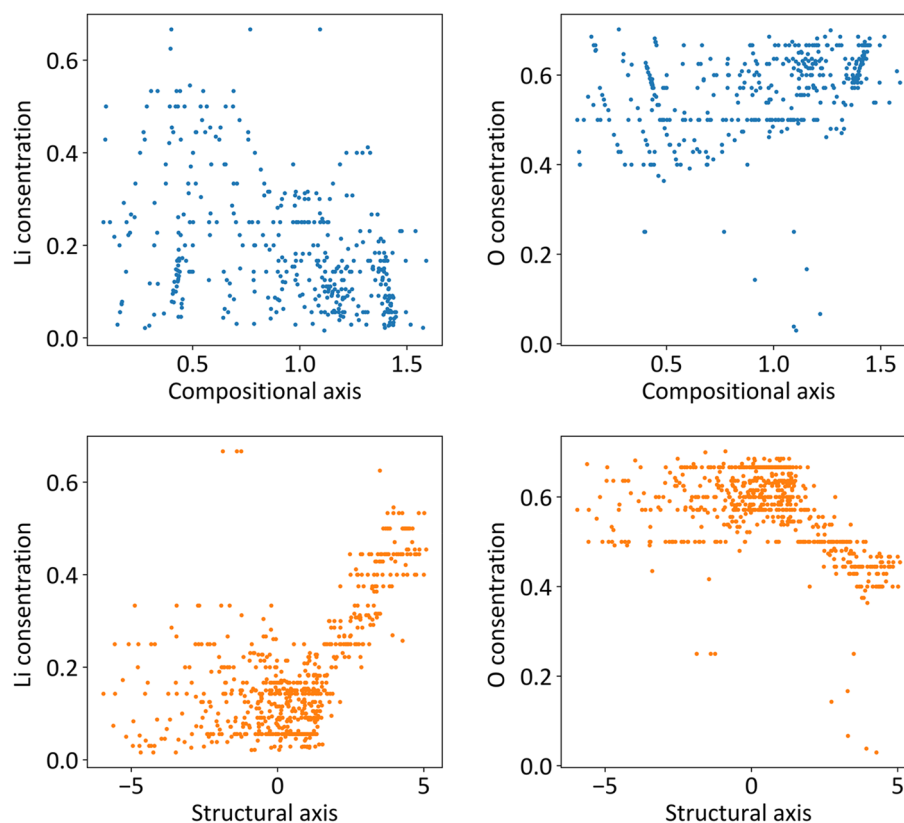

Figure 5. Correlation between encoded compositional variables and (**a**) Li concentration and/or (**b**) O concentration. Relationship between encoded structural encoded variables and (**c**) Li concentration and/or (**d**) O concentration.

far from each other. Although differences in the materials map with respect to crystal structure (type) depend on the structural axis (vertical axis), the same structure exhibits a larger variation on the structural axis than on the compositional axis. The distributions of the Garnet and NASICON structures exhibit similar values on the structural axis, which may be simply interpreted as them being distinguished solely by composition rather than structure. One possible reason for such a counterintuitive conclusion may be the strong dependence of the structural axis on the local structure of Li and O ions, as suggested previously. In other words, the structural axis

				Encoded values	
mp-id	Chemical formula	Structure type	ME/eV	Compositional descriptor (x_{comp})	Structural descriptor (x_{struct})
mp-761986	$Li_5La_3Nb_2O_{12}$	Garnet	0.983	0.207	−0.536
mp-775628	$Li_5La_3(SbO_6)_2$	Garnet	0.827	0.434	−0.517
mp-1200057	$Li_7La_3(SnO_6)_2$	Garnet	0.923	0.322	3.272
mp-554747	$Li_5La_3Nb_2O_{12}$	Garnet	0.543	0.207	1.342
mp-774721	$Li_5La_3Nb_2O_{12}$	Garnet	1.018	0.207	−0.269
mp-559776	$Li_5La_3Ta_2O_{12}$	Garnet	0.529	0.334	1.300
mp-774437	$Li_5La_3Nb_2O_{12}$	Garnet	0.842	0.207	−0.657
mp-942733	$Li_7a_3Zr_2O_{12}$	Garnet	0.861	0.178	3.172
mp-779434	$Li_5La_3Nb_2O_{12}$	Garnet	0.837	0.207	−0.530
mp-10499	$LiZr_2(PO_4)_3$	NASICON	0.850	0.890	0.501
mp-759280	$LiZr_2(PO_4)_3$	NASICON	1.016	0.890	0.982
mp-773074	$LiZr_2(PO_4)_3$	NASICON	1.261	0.890	−0.160
mp-681439	$LiZr_2(PO_4)_3$	NASICON	0.385	0.890	0.876
mp-541661	$LiZr_2(PO_4)_3$	NASICON	0.865	0.890	0.918
mp-1222522	$LiLa_3Ti_4O_{12}$	Perovskite	2.964	0.320	−5.177
mp-768320	$LiLa_5Ti_8O_{24}$	Perovskite	0.760	0.297	−0.221

Table 1. Details of Perovskite, Garget, and NASICON materials in this dataset. Materials Project ID numbers (mp-id), chemical formulae, calculated migration energies (ME), and encoded values for compositional and structural descriptors, x_{comp} and x_{struct}, are listed. Note that both 1782 compositional and 858 structural descriptors are transformed into one-dimensional encoded values, referred to as x_{comp} and x_{struct}, as illustrated in Fig. 1.
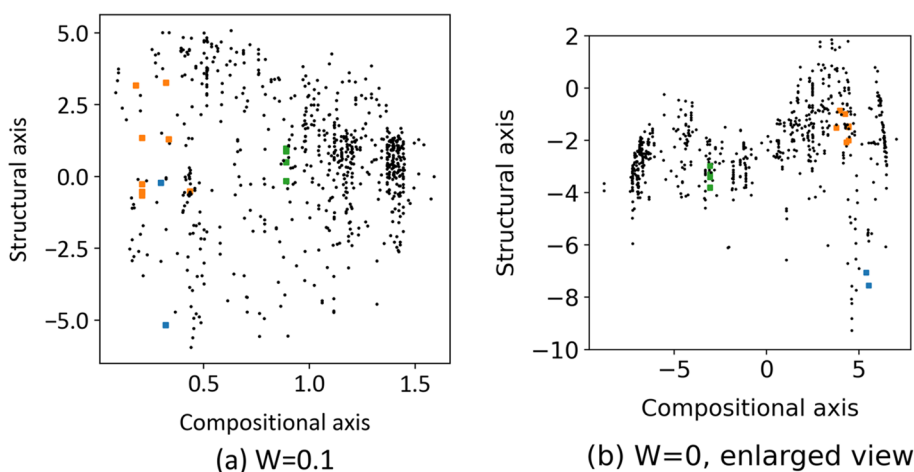
Figure 6. Visualization of typical Li-ion conductive materials (Perovskite, Garnet, NASICON-type materials) visualized on a materials map. Distribution of representative Li-ion conductive materials in the map at (**a**) $W = 0.1$ and (**b**) $W = 0$. Note that there are no discernible relationships between compositional/structural descriptors and ME values when W = 0.

is expected to be strongly influenced by the Li coordination state and arrangement information, rather than the similarity of the host structure. For example, a significant difference in structural values was observed between $Li_7La_3Zr_2O_{12}$ and $Li_7La_3Sn_2O_{12}$, which belong to the same garnet-type material. Specifically, the lithium ions in $Li_7La_3Zr_2O_{12}$ are preferentially distributed in octahedral sites[39], whereas those in $Li_7La_3Sn_2O_{12}$ as well as other compositions are distributed in tetrahedral sites[31]. This suggests that the structural axis is not dependent on the host structure, as classified by the Perovskite, Garnet, and NASICON types, but on the local structure around Li. This is likely because both the structural and compositional axes are associated with the target-variable ME in the neural network model, as shown in Fig. 1.

To confirm this, we reconstructed the material map by setting the weight of ME loss ($Loss_{ME}$) to zero; i.e., $W = 0$. Consequently, the encoded variables became independent of ME. Figure 6b presents the resulting material map. Apparently, materials with Garnet, NASICON, and Perovskite structures are plotted as clusters and exhibit different structural encoded values in descending order of the structure axis (Fig. 6b). This suggests that structural descriptors (RDF, ADF, and Voronoi polyhedron) can capture the grouping of crystal host structures, even without considering Li-ion conductivity.

Figure 7 presents a heat map of the ME predicted for arbitrary coordinates on the material map using a neural network from two encoded variables to the ME (see Fig. 1). Although predictive performance for ME is low (see Fig. 3c), and, in some areas, the predicted ME values do not match the measured values shown in the plots, the map generally matches the distribution trend of the measured values. The region around $x_{comp} \sim 0.0$ and $0 < x_{struct} < 4$ on the materials map corresponds to a lower predicted ME values, indicating higher Li-ion conductivity. Furthermore, in this region, materials registered in the database are sparsely distributed, and the possibility of synthesizing real materials is high. As shown in Fig. 5, the area is characterized by high proportions of lithium in the compounds. For example, this area prominently features garnet-type solid electrolytes, which have a higher molar ratio of Li ions than NASICON- or perovskite-type solid electrolytes. Therefore, the concentration of Li ions may be an efficient guideline for the optimization of garnet compounds by metal substitution[40]. Furthermore, several undiscovered compounds may be present in this area. Unfortunately, in principle, specifying the composition and structure for given coordinates (x_{comp}, x_{struct}) on the material map is difficult, as an accurate conversion from descriptors to crystal structure remains infeasible. However, a material map may inspire experienced researchers to examine the distribution of materials around the focused coordinates. For example, Fig. 8 shows the compositional distribution of materials in proximity of a relevant area, which may support the researchers' intuitions. In addition, descriptors can be restored by decoding the input coordinates on the materials map. Although the crystal structure cannot be directly restored from descriptor information, obtaining information that can be used as material design guidelines is possible. In summary, these findings provide valuable insights for material researchers, enabling them to explore the distribution of materials, predict properties, and recover descriptors using the material map.

Conclusion

In the field of materials informatics for inorganic materials, there are existing databases containing the data of hundreds of thousands of crystal structures. Vigorous research is underway to predict material properties using machine learning, as well as develop fast and efficient methods for optimizing functional materials. However, the vast space of registered materials in these databases suggests the presence of numerous undiscovered materials with desirable properties. Furthermore, it is difficult to systematically search a database using conventional MI methods.

In this study, we designed a materials map to visualize the distribution of known materials by using compositional and structural axes as the basis for organizing the materials. The materials map offers a quick and comprehensive overview of the current state of material exploration. For instance, areas densely populated on the materials map indicate extensive research efforts, reducing the likelihood of discovering new materials. Conversely, sparsely populated areas suggest uncertainty in material synthesis. Additionally, we successfully linked structurally and compositionally encoded values to ME values, allowing us to visualize regions on the materials map where high ionic conductivity is expected. By analyzing both the density of data points and predicted ionic conductivity values on the materials map, material researchers can make informed decisions about their next exploration area, considering results and associated risks. This approach has the potential to facilitate the discovery of materials with exceptional ionic conductivity. While other machine learning techniques such as t-Distributed Stochastic Neighbor Embedding (t-SNE)[41] and Principal Component Analysis (PCA)[42], can reduce descriptor dimensionality, they face challenges in decoding descriptors from arbitrarily chosen encoded values and establishing connections between encoded values and objective variables, such as migration energy in this context. In this regard, the autoencoder-based materials map we present here offers a distinct advantage. Unfortunately, although the specification of encoded variables for optimized ionic conductivity by the
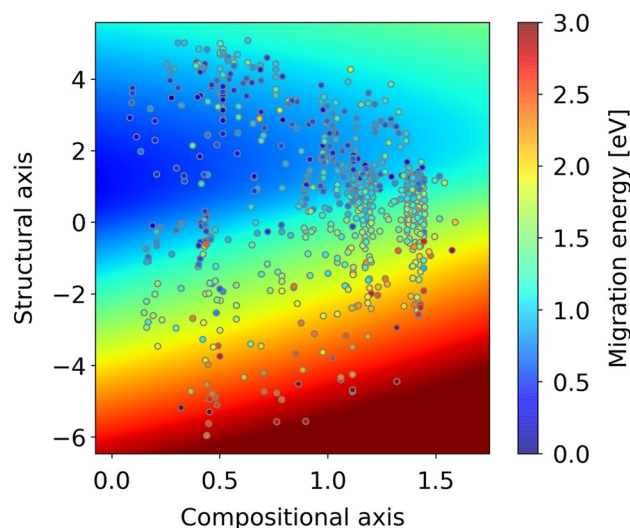
Figure 7. Materials map with predicted ME distribution (Background color represents predicted ME values; data points are obtained from materials simulations).
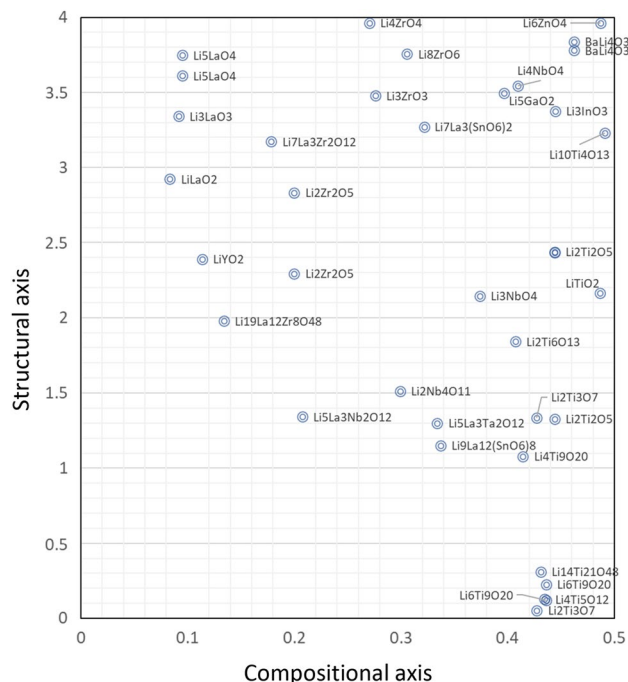
Figure 8. Enlarged materials map highlighting areas with anticipated high lithium conductivity and limited research. Compositions are labeled in the vicinity of their respective data points.

autoencoder-derived materials map is feasible, the reconstruction of crystal structures is technically difficult owing to the irreversibility between the structural data and their descriptors. We, therefore, believe that integrating experience and knowledge of materials researchers may be necessary for overcoming this difficulty, and we anticipate that the 2D visualization of materials information will support researchers' in intuitively understanding material distributions.

Data availability

All the input and output data are available in the Supporting Information.

References

1. Jain, A. *et al.* Formation enthalpies by mixing GGA and GGA + U calculations. *Phys. Rev. B* **84**, 45115. https://doi.org/10.1103/PhysRevB.84.045115 (2011).
2. Bergerhoff, G., Hundt, R., Sievers, R. & Brown, I. D. The inorganic crystal structure data base. *J. Chem. Inf. Comput. Sci.* **23**, 66–69. https://doi.org/10.1021/ci00038a003 (1983).
3. Aizu, S. *et al.* Screening chloride Li-ion conductors using high-throughput force-field molecular dynamics. *J. Am. Ceram. Soc.* **12**, 1–10 (2023).
4. Nakayama, M. *et al.* Data-driven materials exploration for Li-ion conductive ceramics by exhaustive and informatics-aided computations. *Chem. Rec.* **19**, 771–778. https://doi.org/10.1002/tcr.201800129 (2019).
5. Karplus, M. & Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **347**, 631–639 (1990).
6. Chen, C. *et al.* A critical review of machine learning of energy materials. *Adv. Energy Mater.* **10**, 1903242. https://doi.org/10.1002/aenm.201903242 (2020).
7. Nyshadham, C. *et al.* Machine-learned multi-system surrogate models for materials prediction. *NPJ Comput. Mater.* **5**, 51. https://doi.org/10.1038/s41524-019-0189-9 (2019).
8. Himanen, L., Geurts, A., Foster, A. S. & Rinke, P. Data-driven materials science: Status, challenges, and perspectives. *Adv. Sci.* **6**, 1900808. https://doi.org/10.1002/advs.201900808 (2019).
9. Pyzer-Knapp, E. O., Li, K. & Aspuru-Guzik, A. Learning from the harvard clean energy project: The use of neural networks to accelerate materials discovery. *Adv. Funct. Mater.* **25**, 6495–6502. https://doi.org/10.1002/adfm.201501919 (2015).
10. Wang, G., Fearn, T., Wang, T. & Choy, K. L. Machine-learning approach for predicting the discharging capacities of doped lithium nickel–cobalt–manganese cathode materials in Li-ion batteries. *ACS Cent. Sci.* **7**, 1551–1560. https://doi.org/10.1021/acscentsci.1c00611 (2021).
11. Kondo, R., Yamakawa, S., Masuoka, Y., Tajima, S. & Asahi, R. Microstructure recognition using convolutional neural networks for prediction of ionic conductivity in ceramics. *Acta Mater.* **141**, 29–38. https://doi.org/10.1016/j.actamat.2017.09.004 (2017).
12. Meredig, B. *et al.* Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* **89**, 94104. https://doi.org/10.1103/PhysRevB.89.094104 (2014).
13. Seko, A., Hayashi, H. & Tanaka, I. Compositional descriptor-based recommender system for the materials discovery. *J. Chem. Phys.* **148**, 241719. https://doi.org/10.1063/1.5016210 (2018).
14. Atsumi, T. *et al.* Chemical composition data-driven machine-learning prediction for phase stability and materials properties of inorganic crystalline solids. *Phys. Status Solidi B Basic Res.* **259**, 2100525 (2022).

15. Villars, P. & Iwata, S. Pauling file verifies/reveals 12 principles in materials science supporting four cornerstones given by Nature. *Chem. Met. Alloys* **6**, 81–108 (2013).
16. Davies, D. W. *et al.* Computational screening of all stoichiometric inorganic materials. *Chemistry* **1**, 617–627. https://doi.org/10.1016/j.chempr.2016.09.010,Pubmed:27790643 (2016).
17. Mooser, E. & Pearson, W. B. On the crystal chemistry of normal valence compounds. *Acta Cryst.* **12**, 1015–1022. https://doi.org/10.1107/S0365110X59002857 (1959).
18. Evans, R. C., Muller, O. & Roy, R. The major ternary structural families. Berlin, Heidelberg, and New York (Springer Verlag), 1974. ix + 487 pp., 46 figs. Price DM 76, US $29.30. *Mineral. Mag.* **40**, 210–211. https://doi.org/10.1180/minmag.1975.040.310.16 (1975).
19. Nakayama, M. Materials informatics for discovery of ion conductive ceramics for batteries. *J. Ceram. Soc. Jpn.* **129**, 286–291. https://doi.org/10.2109/jcersj2.21030 (2021).
20. Nakayama, M., Kimura, M., Jalem, R. & Kasuga, T. Efficient automatic screening for Li ion conductive inorganic oxides with bond valence pathway models and percolation algorithm. *Jpn. J. Appl. Phys.* **55**, 01AH05. https://doi.org/10.7567/JJAP.55.01AH05 (2016).
21. Jalem, R. *et al.* A general representation scheme for crystalline solids based on Voronoi-tessellation real feature values and atomic property data. *Sci. Technol. Adv. Mater.* **19**, 231–242. https://doi.org/10.1080/14686996.2018.1439253 (2018).
22. Hinton, G. E. & Salakhutdinov, R. R. Reducing the Dimensionality of Data with Neural Networks. *Science.* **313**, 504–507. https://doi.org/10.1126/science.1127647 (2006).
23. Kingma, D. P. & Ba, J. L. *Conf. Track Proc. Int. Conf. Learn. Represent. ICLR* 1–15 (2015).
24. Itoh, M., Inaguma, Y., Jung, W. H., Chen, L. & Nakamura, T. High lithium ion conductivity in the perovskite-type compounds. *Solid State Ionics* **70–71**, 203–207 (1994).
25. Stramare, S., Thangadurai, V. & Weppner, W. Lithium lanthanum titanates: A review. *Chem. Mater.* **15**, 3974–3990. https://doi.org/10.1021/cm0300516 (2003).
26. Yang, Z. & Nakayama, M. Materials simulation for the modulated arrangement in A-site deficient perovskite-type lithium doped lanthanum niobate as solid electrolytes for all solid-state Li ion batteries. *J. Phys. Soc. Jpn.* **91**, 91010. https://doi.org/10.7566/JPSJ.91.091010 (2022).
27. Thangadurai, V. & Weppner, W. Recent progress in solid oxide and lithium ion conducting electrolytes research. *Ionics* **12**, 81–92. https://doi.org/10.1007/s11581-006-0013-7 (2006).
28. Murugan, R., Thangadurai, V. & Weppner, W. Fast lithium ion conduction in garnet-type $Li_7La_3Zr_2O_{12}$. *Angew. Chem. Int. Ed.* **46**, 7778–7781. https://doi.org/10.1002/anie.200701144 (2007).
29. Yang, L. *et al.* Interrelated interfacial issues between a Li7 La3 Zr2 O12-based garnet electrolyte and Li anode in the solid-state lithium battery: A review. *J. Mater. Chem. A* **9**, 5952–5979. https://doi.org/10.1039/D0TA08179E (2021).
30. Wang, C. *et al.* Garnet-type solid-state electrolytes: Materials, interfaces, and batteries. *Chem. Rev.* **120**, 4257–4300. https://doi.org/10.1021/acs.chemrev.9b00427 (2020).
31. Jalem, R. *et al.* Insights into the lithium-ion conduction mechanism of garnet-type cubic Li5La3Ta2O12 by ab-initio calculations. *J. Phys. Chem. C* **119**, 20783–20791. https://doi.org/10.1021/acs.jpcc.5b05068 (2015).
32. Xu, X., Wen, Z., Wu, X., Yang, X. & Gu, Z. Lithium ion-conducting glass ceramics of Li1.5 Al0.5 Ge1.5 (PO4)3–xLi2 O (x=0.0–0.20) with good electrical and electrochemical properties. *J. Am. Ceram. Soc.* **90**, 2802–2806. https://doi.org/10.1111/j.1551-2916.2007.01827.x (2007).
33. Safanama, D. & Adams, S. High efficiency aqueous and hybrid lithium-air batteries enabled by Li1.5 Al0.5 Ge1.5 (PO4)3 ceramic anode-protecting membranes. *J. Power Sources* **340**, 294–301. https://doi.org/10.1016/j.jpowsour.2016.11.076 (2017).
34. Xie, H., Li, Y. & Goodenough, J. B. NASICON-type Li1+2xZr2−xCax(PO4)3 with high ionic conductivity at room temperature. *RSC Adv.* **1**, 1728–1731. https://doi.org/10.1039/c1ra00383f (2011).
35. Hou, M., Liang, F., Chen, K., Dai, Y. & Xue, D. Challenges and perspectives of NASICON-type solid electrolytes for all-solid-state lithium batteries. *Nanotechnology* **31**, 132003. https://doi.org/10.1088/1361-6528/ab5be7 (2020).
36. Nakayama, M. *et al.* Na superionic conductor-type LiZr2(PO4)3 as a promising solid electrolyte for use in all-solid-state Li metal batteries. *Chem. Commun.* **58**, 9328–9340. https://doi.org/10.1039/d2cc01526a (2022).
37. Takeda, H. *et al.* Process optimisation for NASICON-type solid electrolyte synthesis using a combination of experiments and Bayesian optimisation. *Mater. Adv.* **3**, 8141–8148. https://doi.org/10.1039/D2MA00731B (2022).
38. Fukuda, H. *et al.* Bayesian optimisation with transfer learning for NASICON-type solid electrolytes for all-solid-state Li-metal batteries. *RSC Adv.* **12**, 30696–30703. https://doi.org/10.1039/d2ra04539g (2022).
39. Jalem, R. *et al.* Concerted migration mechanism in the Li ion dynamics of garnet-type $Li_7La_3Zr_2O_{12}$. *Chem. Mater.* **25**, 425–430. https://doi.org/10.1021/cm303542x (2013).
40. Kataoka, K. *et al.* Development of a compact all-solid-state lithium secondary battery using single-crystal electrolyte—Towards realizing oxide-type all-solid-state lithium secondary batteries. *Synthesiology* **12**, 29–40. https://doi.org/10.5571/syntheng.12.1_29 (2021).
41. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
42. Jolliffe, I. T. & Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. A Math. Phys. Eng. Sci.* **374**, 20150202. https://doi.org/10.1098/rsta.2015.0202 (2016).

Acknowledgements

Author contributions

M.N., M.K and I.T. conceived and directed the project. Y.Y., T.A., and M.N. performed descriptor building and computational evaluations of migration energies. Y.Y. and K.K. coded the software used in this study. Y.Y. analyzed output data. The manuscript was mainly written by Y.Y. and M.N. through the contributions of all authors. All the authors have approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-43921-1.

Correspondence and requests for materials should be addressed to M.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.