DATA PAPER

Evaluating thermal expansion in fluorides and oxides: Machine learning predictions with connectivity descriptors

Yilin Zhang(张轶霖)[1,†], Huimin Mu(穆慧敏)[2,†], Yuxin Cai(蔡雨欣)[1,†], Xiaoyu Wang(王啸宇)[2], Kun Zhou(周琨)[1], Fuyu Tian(田伏钰)[1], Yuhao Fu(付钰豪)[2,3,‡], and Lijun Zhang(张立军)[1,3,§]

[1] *State Key Laboratory of Integrated Optoelectronics, Key Laboratory of Automobile Materials of MOE, College of Materials Science and Engineering, Jilin University, Changchun* 130012, *China*

[2] *State Key Laboratory of Superhard Materials, College of Physics, Jilin University, Changchun* 130012, *China*

[3] *International Center of Computational Method and Software, Jilin University, Changchun* 130012, *China*

(Received 15 April 2023; revised manuscript received 17 April 2023; accepted manuscript online 18 April 2023)

Open framework structures (e.g., ScF_3, $Sc_2W_3O_{12}$, etc.) exhibit significant potential for thermal expansion tailoring owing to their high atomic vibrational degrees of freedom and diverse connectivity between polyhedral units, displaying positive/negative thermal expansion (PTE/NTE) coefficients at a certain temperature. Despite the proposal of several physical mechanisms to explain the origin of NTE, an accurate mapping relationship between the structural–compositional properties and thermal expansion behavior is still lacking. This deficiency impedes the rapid evaluation of thermal expansion properties and hinders the design and development of such materials. We developed an algorithm for identifying and characterizing the connection patterns of structural units in open-framework structures and constructed a descriptor set for the thermal expansion properties of this system, which is composed of connectivity and elemental information. Our developed descriptor, aided by machine learning (ML) algorithms, can effectively learn the thermal expansion behavior in small sample datasets collected from literature-reported experimental data (246 samples). The trained model can accurately distinguish the thermal expansion behavior (PTE/NTE), achieving an accuracy of 92%. Additionally, our model predicted six new thermodynamically stable NTE materials, which were validated through first-principles calculations. Our results demonstrate that developing effective descriptors closely related to thermal expansion properties enables ML models to make accurate predictions even on small sample datasets, providing a new perspective for understanding the relationship between connectivity and thermal expansion properties in the open framework structure. The datasets that were used to support these results are available on Science Data Bank, accessible via the link https://doi.org/10.57760/sciencedb.j00113.00100.

Keywords: first-principles calculations, machine learning, negative thermal expansion, Grüneisen parameter

PACS: 63.20.dk **DOI:** 10.1088/1674-1056/accdca

1. Introduction

Thermal expansion is a ubiquitous phenomenon that materials experience a change in volume in response to changes in temperature, and uncontrolled thermal expansion can impair device performance and accuracy.[1,2] Thermal management issues restrict the practical application of these devices.[3,4] A promising strategy to mitigate this issue is to combine positive and negative expansion materials to achieve low or near-zero thermal expansion.[5–9] Despite ongoing efforts, the discovery of matching negative thermal expansion (NTE) materials has proven challenging due to their rarity and the lack of a unified explanation for NTE, unlike positive thermal expansion (PTE) which can be attributed to anharmonic effects.[10] The current explanations for the intrinsic NTE within a homogenous solid typically involve both structural (e.g., tension effects and rigid unit modes (RUMs)) and electronic mechanisms (e.g., electron transfer, and ferroelectric/magnetic phase transition).[11,12] NTE materials driven by structural factors exhibit a broad temperature range of thermal contraction compared to NTE materials induced by electronic mechanisms.[13,14] This wide operating NTE temperature range makes them promising candidates for various applications, such as high-precision optics on satellites for deep space exploration.[15] Therefore, a deeper understanding of the mechanisms behind structural-driven thermal expansion materials can accelerate the discovery and application of both PTE and NTE materials.

Metal fluorides and oxides in flexible networks are important materials that have received significant attention.[16–22] These compounds consist of polyhedral units linked by bridging atoms, and the polyhedral units are generally rigid. Lattice vibrations occur perpendicular to the axis of the polyhedron, and the collective motion of these rigid units can be analyzed using the concept of RUMs.[23] For example, ZrW_2O_8 is a representative oxide material, where the low-frequency transverse phonon modes of bridging oxygen atoms result in coupled rotations of the essentially corner-sharing rigid polyhedral units,

[†]These authors contributed equally.

[‡]Corresponding author. E-mail: fuyuhaoy@gmail.com

[§]Corresponding author. E-mail: lijun_zhang@jlu.edu.cn

© 2023 Chinese Physical Society and IOP Publishing Ltd

leading to NTE.[24,25] Besides, there exists a close correlation between the connectivity of structural units and the thermal expansion behavior. In $Cu_2V_2O_7$[26] and ZrW_2O_8,[25,27] the dominant connection pattern between polyhedral units is shared vertices, and the corresponding coefficient of volumetric thermal expansion can reach to −60.6 ppm/K and −28.2 ppm/K, respectively. Conversely, for materials that exhibit large PTE, such as Fe_2SiO_4 (37.8 ppm/K)[28] and $NaMgF_3$ (88 ppm/K),[29] their structures are primarily established through coplanar and collinear connections among units.

In fact, framework structures can be viewed as being connected by structural units, and their connectivity can theoretically be extracted. Recent studies have begun to investigate the relationship between thermal expansion and structural characteristics.[30–32] The concept of average atomic volume (AAV) was proposed to study cyanide NTE systems, where CN ligands bridge polyhedral units.[31] The study suggests a positive correlation between NTE intensity and AAV. Specifically, a larger AAV, which indicates greater vibrational degrees of freedom of the transverse group (CN ligands), is associated with a more significant NTE. Additionally, an investigation was conducted to examine the impact of twin boundaries on thermal expansion.[32] The results showed that the thermal expansion of SnO_2 twinning decreased significantly with an increase in the number of corner-sharing oxygen anions between polyhedrons. Conversely, as the number of edge-sharing connections between polyhedrons increased, the thermal expansion behavior intensified. These studies suggest that the connectivity in flexible network materials has a significant influence on the thermal expansion behavior. Specific connection patterns between polyhedral units can lead to a decrease in thermal expansion or even NTE. Fortunately, with the advancement of machine learning (ML) techniques,[33–39] it is now possible to leverage ML models to uncover the relationship between connectivity and thermal expansion behavior.

In this study, a combination of first-principles calculations and ML algorithms was used to predict the thermal expansion behavior of binary and ternary oxides and fluorides with a framework structure. The selection of materials for this study was based on previous literature reports, including 246 samples. Initially, an efficient algorithm was developed to extract the connectivity in framework structure materials. Then, ML algorithms were employed to predict the thermal expansion behavior of oxides and fluorides based on structural features and elemental features. Despite the challenges posed by a limited dataset and complex symmetry, the accuracy of 92% is achieved by the trained model by leveraging connectivity features based on the validation set. Moreover, the trained models effectively predict eight negative NTE candidate materials in new binary oxides and fluorides, obtained by replacing metal atoms with main-group elements. Two of them have been experimentally and theoretically confirmed, while the remaining six materials, reported for the first time in this study, were validated by negative averaged Grüneisen parameters through first-principles calculations.

2. Methods

We employed first-principles density functional theory (DFT) calculations implemented in the Vienna *Ab initio* Simulation Package code (VASP) to relax the structure and determine the total energy.[40–42] The ion–electron interactions and exchange–correlation functional were described by the projector augmented wave (PAW) method[43] and the generalized gradient approximation[44] of Perdew, Burke, and Ernzerhof (PBE),[45] respectively. To ensure accurate structural optimizations, we used an energy cutoff of 520 eV for the plane-wave basis set with a *k*-points mesh of $2\pi \times 0.16$ Å^{-1}. Using our in-house developed high-throughput computing codes (JAMIP, Jilin Artificial-intelligence aided Materials-design Integrated Package),[46,47] we optimized all structures until the total energy and residual forces of each atom converged to 10^{-4} eV and were smaller than 10^{-3} eV·Å^{-1}, respectively. To classify the thermal expansion behavior (PTE/NTE), we trained six binary classification models, including logistic regression (LR),[38] *k*-nearest neighbor classifier (*k*-NN),[48] support vector machine (SVM),[36] decision tree (DT),[39] random forest (RF),[34] and gradient boosting decision tree (GBDT),[35] based on the 5-fold cross-validation approach. 80% of the total dataset was selected as the training set and 20% as the validation set. The hyperparameters of each model were determined through a grid search method to identify the optimal parameters, and accuracy was used to evaluate the performance of these models.

Phonon dispersions and the phonon density of states (PDOS) of the NTE material predicted by the ML models were calculated through the real-space supercell method and the finite displacement approach implemented in the PHONOPY package.[49] The volumetric coefficient of thermal expansion can be written as

$$\alpha_{\mathrm{V}}(T) = \frac{C_{\mathrm{V}}(T)\,\bar{\gamma}(T)}{BV}, \tag{1}$$

where C_{V}(T), $\bar{\gamma}(T)$, B and V represent the specific heat, averaged Grüneisen parameter, bulk modulus, and volume, respectively. At any temperature, C_{V}, B, and V are positive quantities. Therefore, the sign of $\bar{\gamma}(T)$ plays a crucial role in determining the thermal expansion behavior of a material. The averaged Grüneisen parameter is expressed as

$$\bar{\gamma}(T) = \frac{1}{C_{\mathrm{V}}(T)} \sum_{\boldsymbol{q},\nu} \gamma_{\boldsymbol{q},\nu} C_{\boldsymbol{q},\nu}, \tag{2}$$

where $C_{\boldsymbol{q},\nu}$ and $\gamma_{\boldsymbol{q},\nu}$ represent the mode heat capacity and mode Grüneisen parameters at wave vector $\boldsymbol{q}$ and band in-

dex ν, respectively. The mode Grüneisen parameters were obtained by computing the logarithmic derivatives of phonon frequencies with respect to volume change, achieved by expanding/compressing the lattice with strain $\varepsilon = \pm 1\%$. A phonon momenta $\boldsymbol{q}$ mesh of $20 \times 20 \times 20$ was used in the calculations of the thermal properties.

3. Results and discussion

Our dataset is based on the collection of PTE and NTE materials reported in the literature (see Table S1 in supplementary materials), with structures sourced from ICSD and the Materials Project. As shown in Fig. 1, there are a total of 246 samples, including 206 oxides (60 binaries and 146 ternaries) and 40 fluorides (25 binaries and 15 ternaries) with 46 symmetries. All structures were optimized using first-principles DFT calculations. Then, we extracted the corresponding structural features, including connection patterns, AAV, and porosity, using our developed algorithm. In addition, we incorporated four elemental features. With the aid of ML algorithms, we performed binary classification training for thermal expansion behavior (PTE/NTE). Subsequently, high-accuracy trained models were used to predict NTE candidates in a new dataset generated by substituting main group elements into the binary oxides and fluorides in the initial dataset. Among the screened NTE candidates, we further identified thermodynamically stable materials and verified their NTE properties by negative Grüneisen parameter according to Eq. (2).

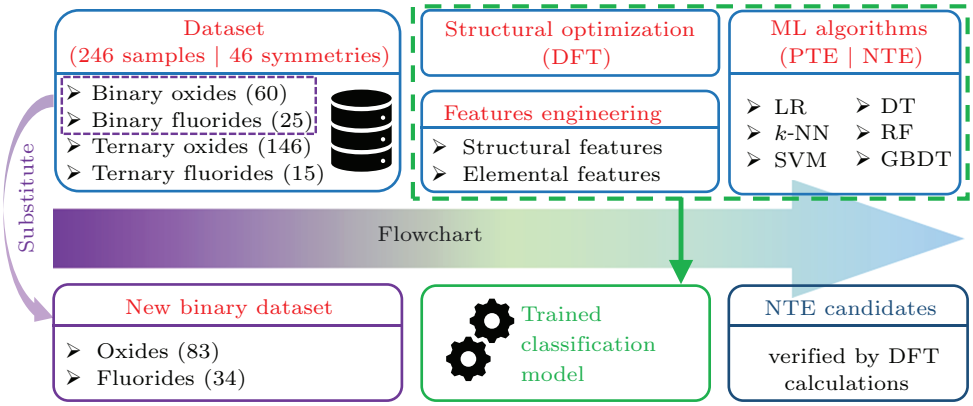


Fig. 1. Research flowchart including the composition of the database, six ML algorithms employed, prediction and validation of NTE candidates in the new unexplored binary dataset.

Recent studies have begun to explore the relationship between structural degrees of freedom and material thermal expansion, but they are still limited to simple geometric descriptors or finite symmetric systems.[31–33] For instance, the AAV concept is defined as the average volume occupied by each atom in the structural unit, but most of the structural features are lost due to its relatively simple averaging statistical method.[31] Furthermore, research on polyhedral configurations is restricted to a few symmetric structures or fixed chemical formulas.[33] The main challenge is the lack of a reasonable method for determining polyhedral units and no feasible statistical method for describing the connectivity between these units. Herein, we propose an efficient algorithm to address these challenges and further deepen our understanding of the thermal expansion behavior of materials by exploring the connectivity between structural units.

Taking the NTE oxide $Dy_2Mo_4O_{15}$ with complex geometric configurations as an example, its structure consists of DyO_7 pentagonal bipyramids, MoO_4 tetrahedra, and MoO_5 trigonal bipyramids,[50] as shown in Fig. 2(a). By selecting the n-th cation (central atom) in such polyhedral units and calculating the bond lengths l_i (i corresponds to the index of bond lengths from short to long) between neighboring anions (terminal atoms), the bond length distribution presents a step-like distribution (Fig. 2(b)). The bond lengths within the same polyhedron are close, while the bond lengths between anions of different polyhedrons are distributed in another value step. Therefore, for the coordination environment of each central atom, we can use the bond length change rate ρ_i^{n} to obtain the number of terminal atoms:

$$\rho_i^{\mathrm{n}} = \frac{\Delta l_i^{\mathrm{n}}}{l_i^{\mathrm{n}}} = \frac{l_{i+1}^{\mathrm{n}} - l_i^{\mathrm{n}}}{l_i^{\mathrm{n}}}, \tag{3}$$

where the coordination number corresponds to the maximum change rate index i. As shown in Fig. 2(b), the peaks of ρ_i^{n} indicate that Dy is coordinated with 7 oxygen atoms (the grey peak), while Mo exhibits 4- and 5-coordination environments, corresponding to the blue and red peaks, respectively.

It can be observed that the composition of a polyhedron consists of one central atom and several terminal atoms (vertices). Therefore, by analyzing the vertex sharing between polyhedral units, we can determine the connection patterns between them. The number of shared vertices between each polyhedron and its neighboring polyhedrons is shown in Fig. 2(c). Taking the MoO_5 trigonal bipyramid in $Dy_2Mo_4O_{15}$ as an example (with five vertices, i.e., oxygen atoms labeled as

O1, O2, O3, O4, and O5), it is connected to three DyO_7 pentagonal bipyramids (labeled as A, B, C) and one MoO_5 trigonal bipyramid (labeled as D) and MoO_4 tetrahedron (labeled as E). The shared vertices are listed in Fig. 2(c) as a connection matrix. Based on the connection matrix, we find that the O1 vertex is shared with polyhedral units A and E, and O2 is also shared with polyhedron A. O3, O4, and O5 are shared with polyhedral units B, C, and D, respectively. In addition, the vertex sharing among the neighboring polyhedrons (A, B, C, D, E) can also be tracked, which is designated by the black dashed box. It is noticeable that among the polyhedral units, only A and E share the O1 (marked in red with coordinate $\boldsymbol{r}_1$), while the others are mutually independent, meaning that no other oxygen atom shares the same coordinates except for O1. Therefore, we can conclude that the MoO_5 trigonal bipyramid shares only one point (vertex) with B, C, and D units, and this type of connection can be named as "co-point" connection, as shown in Fig. 3. O1 and O2 form an edge that is shared between MoO_5 and unit A (named "co-edge" connection), while unit E is connected through an "edge-connected" connection.

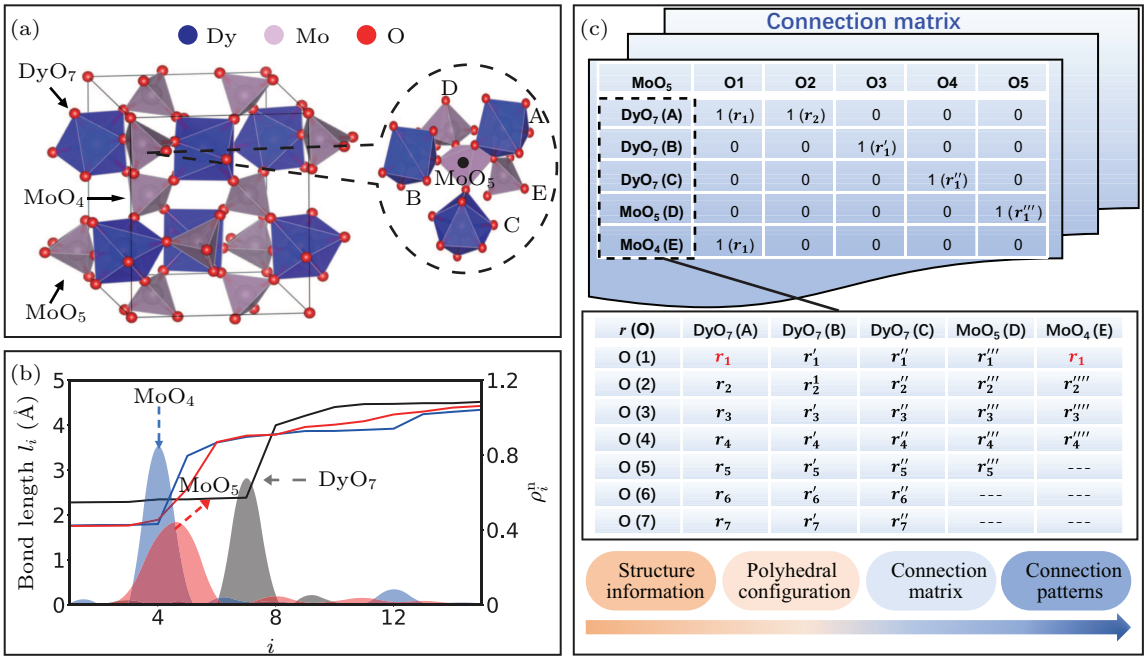


Fig. 2. (a) Schematic representation of the $Dy_2Mo_4O_{15}$ structure, which comprises three distinct types of polyhedral units (DyO_7 pentagonal bipyramids, MoO_4 tetrahedra, and MoO_5 trigonal bipyramids). The inset shows the connectivity of MoO_5 trigonal bipyramids with surrounding polyhedrons. (b) Determination of polyhedral configurations based on bond length variation rates. (c) Connection matrix of the MoO_5 trigonal bipyramids with surrounding polyhedrons and the shared vertices of surrounding polyhedral units.

Overall, each polyhedron in the structure was assigned a corresponding connection matrix, and the atoms (cations and anions) in each polyhedron were labeled. Based on these matrices and the vertices sharing information about neighboring polyhedral units, we could analyze more complex polyhedral connection patterns. As shown in Fig. 3, there are eight connection patterns for framework oxides and fluorides, including co-plane (C_1), co-edge (C_2), co-point (C_3), tri-point (C_4), quadri-point (C_5), edge-connected (C_6), atom-bridged (C_7), and unbridged connections (C_8). These connectivity features were averaged over each polyhedral unit to mitigate errors arising from disparate structural unit sizes. Together with the AAV and porosity (the percentage of non-polyhedral space in the structure), these features constitute a total of ten structural features. In addition to these structural features, we also incorporated elemental features based on atomic mass, including cation density (CD), anion mass density (AD), total atomic mass density (TAD = CD + AD), cation-to-anion mass ratio (CMR). All features and labels (i.e., PTE and NTE) are listed in Table S2.

As shown in Fig. 4(a), based on the features described above, the ML models achieved accuracies of 84% (SVM), 86% (LR, *k*-NN and DT), and 92% (RF, GBDT), respectively. The ensemble models, RF and GBDT, outperformed the others with higher accuracy, and their ROC curves reached 90% and 91% (Fig. 4(b)), indicating a high probability of correctly classifying positive and negative expansion behaviors. Additionally, facing the challenges of the limited dataset and complex symmetry, the results of the confusion matrix (the inset of Fig. 4(b)) demonstrated that descriptors based on structural connectivity can effectively identify the thermal expansion behavior of materials. Then, the new dataset was obtained by substituting metal elements in binary fluorides and oxides with main group elements, yielding a new compound named "*x*_initial structure", where *x* denotes the substituting element. For instance, Cr_MoO_3 signifies the substitution of Cr for Mo in MoO_3. And the new feature set is listed in Table S3. The trained RF model and GBDT model were used to predict the

thermal expansion behavior of these materials. To illustrate the influence of connectivity on thermal expansion in a more intuitive way, we defined a simple connectivity factor χ based on connection patterns as follows:

$$\chi = \left(\sum_{i=3,4,5,8} C_i - \sum_{j=1,2,6,7} C_j\right) \cdot \mathrm{AAV} \cdot \mathrm{porosity}, \quad (4)$$

where $\sum_{i=3,4,5,8} C_i$ represents the sum of connection patterns that promote higher degrees of freedom between units (e.g., co-point (C_3)), while $\sum_{j=1,2,6,7} C_j$ represents the sum of connection patterns that decrease degrees of freedom (e.g., co-plane (C_1)). The product of AAV and polyhedral porosity is used to evaluate the overall freedom of structure. As shown in Figs. 3(c) and 3(d), materials with predicted probabilities below 50% were classified as PTE (colored red), while those above 50% were classified as NTE (colored blue). We predicted 10 fluorides and 25 oxides with NTE (see Table S4). It is evident that the connectivity factor effectively distinguishes between PTE and NTE materials, with a general trend of the larger factor indicating the higher likelihood of NTE. This trend demonstrates that there is a close correlation between the connectivity of structural units in framework materials and their thermal expansion behavior.

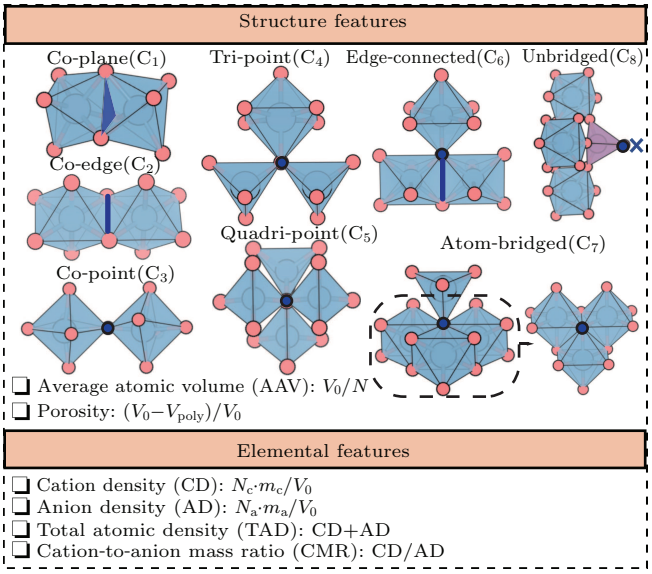


Fig. 3. Structural and elemental features used for ML model training, including eight connection patterns, AAV, and porosity as structural features, and four elemental features based on atomic mass. V_0 and V_{poly} represent the unit cell volume and the volume of polyhedral structures in the unit cell, respectively. N, N_c, and N_a represent the total number of atoms, cations, and anions in the unit cell, respectively, while m_c and m_a correspond to the masses of cations and anions.

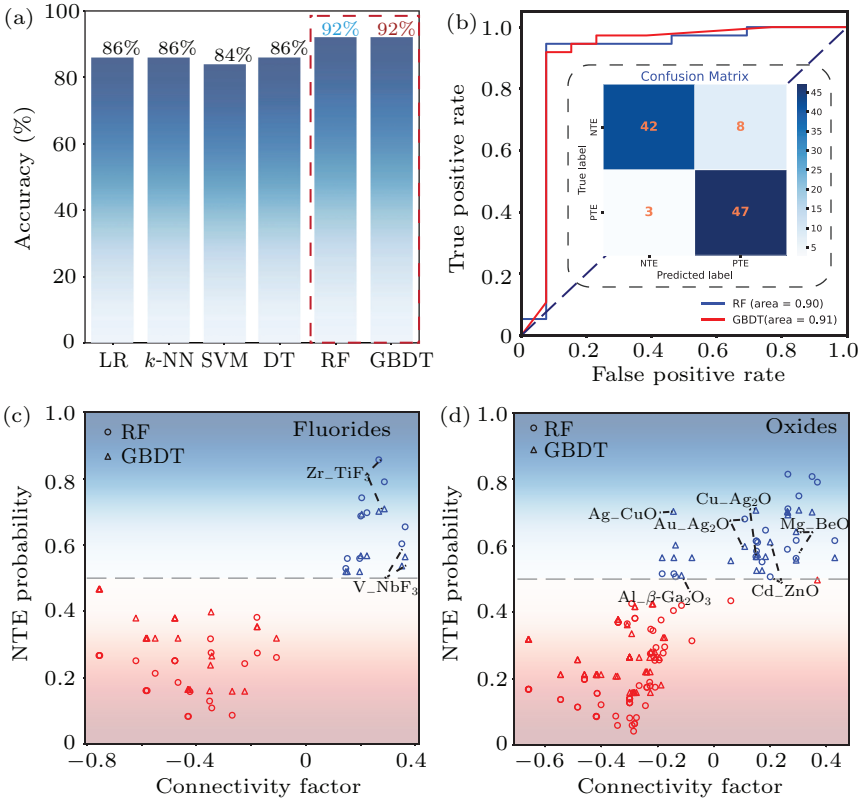


Fig. 4. (a) Accuracy of six trained binary classification models. (b) ROC curve and confusion matrix of the RF and GBDT models with high accuracy. Predicted NTE candidates in new binary (c) fluorides and (d) oxides ranked by connectivity factor in order using the trained RF and GBDT models.

To identify potential candidates for NTE, we conducted first-principles calculations to optimize the structures of the predicted materials and calculated their phonon spectra. The absence of imaginary frequencies in the phonon spectra is an important indicator of thermodynamic stability in solids. Eight compounds exist stable phonon spectra without imagination frequency, including two fluorides and six oxides labeled in Figs. 4(c) and 4(d), respectively. For binary fluorides, V_NbF_3 and Zr_TiF_3 have the same cubic structure formed by the corner-sharing octahedra. Their phonon spectra dot-

ted the low-frequency negative Grüneisen parameters (marked by blue circles), as shown in Figs. S1a and S1b. From the analysis of PDOS, it is evident that NTE primarily arises from the low-frequency phonon vibrations of the F atoms. It is noteworthy in binary oxides that the trained ML modes accurately predicted the reported NTE of Au_2O (i.e., Au_Ag_2O) and Cu_2O (i.e., Cu_Ag_2O), which have been experimentally and theoretically confirmed.[51,52] Their NTE can be attributed to the dominant low-frequency liberational modes of the metal atoms, which is consistent with the results of PDOS shown in Figs. S1(c) and S1(e).

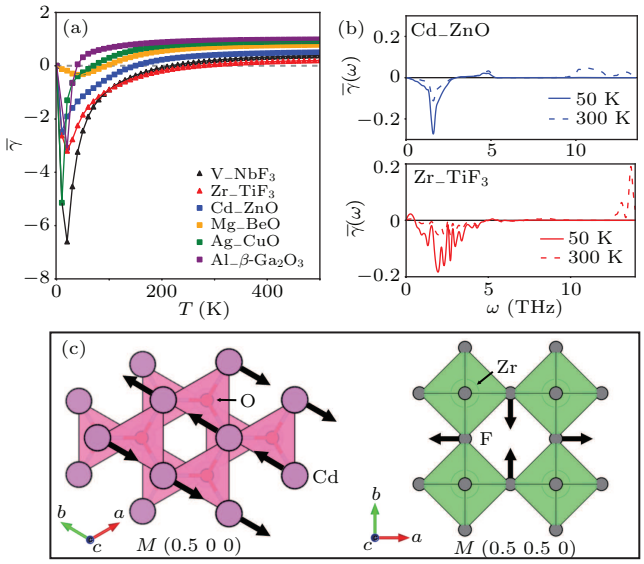


Fig. 5. (a) Temperature-dependent Grüneisen parameters for six predicted thermodynamically stable NTE materials. (b) Reduction of low-frequency phonon modes with increasing temperature in the oxide Cd_ZnO and the fluoride Zr_TiF_3. (c) Low-frequency phonon vibration modes of bridged atoms in the oxide Cd_ZnO and the fluoride Zr_TiF_3 at the high-symmetry points featured with large negative Grüneisen parameters.

Subsequently, we calculated the temperature-dependent averaged Grüneisen parameters for the remaining materials that have not been experimentally investigated, including V_NbF_3, Zr_TiF_3, Cd_ZnO, Mg_BeO, Ag_CuO, and Al_β-Ga_2O_3. As shown in Fig. 5(a), these materials exhibit negative $\bar{\gamma}(T)$ regions, meaning the negative volumetric thermal expansion behavior according to Eq. (2). Meanwhile, it also validates the feasibility of predicting NTE materials based on connectivity between polyhedral units. We chose Cd_ZnO and Zr_TiF_3 as representative materials to investigate the mechanisms underlying their NTE behavior, as they exhibit relatively wide NTE temperature ranges. The NTE of Cd_ZnO was induced by the low-frequency vibrational modes of the Cd metal atoms (the upper panel in Fig. 5(b)), while the NTE of Zr_TiF_3, as mentioned earlier, was attributed to the low-frequency vibrational modes of the F atoms (the lower panel in Fig. 5(b)). By examining the low-frequency phonon modes at the high-symmetry point *M* featured with large negative Grüneisen parameters obviously in the phonon spectra (Figs. S1(b) and S1(d)), we observed that the low-frequency transverse vibration of the bridge atoms (F and Cd) drives the rotation of polyhedral units, as shown in Fig. 5(c), leading to the thermal contraction. Thereby, it is reasonable that with increasing temperature, the low-frequency phonon modes weaken while the high-frequency phonon modes strengthen, facilitating the transition towards positive expansion, as depicted in Fig. 5(b).

4. Conclusion

We propose an algorithm to characterize the connectivity in an open framework system and construct a descriptor set containing connection patterns between polyhedral units and elemental information for the ML classification model of material thermal expansion properties. Despite the small size and significant structural and chemical composition differences in our training/test dataset, which is based on experimental data and contains slightly over 200 samples, our trained model achieves a high prediction accuracy of 92%. This result demonstrates the effectiveness of our descriptor set in accurately predicting the thermal expansion behavior in this system. Furthermore, by substituting elements in experimental structures, we construct a new unexplored dataset of binary compounds (83 oxides and 34 fluorides). Based on the trained ML model, we predict eight thermodynamically stable NTE materials, including two materials that have been experimentally and theoretically confirmed, and the remaining six materials that have not been reported before. We confirm the NTE properties of these six materials through first-principles calculations, further demonstrating the reliability of our prediction model. Our results deepen the understanding of the relationship between the connectivity of framework structures and thermal expansion behavior. The connectivity-based descriptors provide a novel insight into predicting thermal expansion behavior, which could accelerate the design of materials with tailored thermal expansion properties.

Data availability statement

The data that support the findings of this study are openly available in Science Data Bank via the link https://doi.org/10.57760/sciencedb.j00113.00100.

Acknowledgements

Project supported by the National Natural Science Foundation of China (Grant Nos. 12004131, 22090044, 62125402, and 92061113). Calculations were performed in part at the high-performance computing center of Jilin University.

References

[1] Moore A L and Shi L 2014 *Mater. Today* **17** 163
[2] Takeuchi Y, Sakamoto M and Sata T 1982 *Precis. Eng.* **4** 19

[3] Moore A L and Shi L 2014 *Mater. Today* **17** 163
[4] Cai Y, Faizan M, Mu H, Zhang Y, Zou H, Zhao H J, Fu Y and Zhang L 2023 *Front. Phys.* **18** 43303
[5] Cen D, Wang B, Chu R, Gong Y, Xu G, Chen F and Xu F 2020 *Scripta Mater* **186** 331
[6] Ding L, Wang C, Na Y, Chu L and Yan J 2011 *Scripta Mater.* **65** 687
[7] Takenaka K, Hamada T, Kasugai D and Sugimoto N 2012 *J. Appl. Phys.* **112** 083517
[8] Takenaka K and Ichigo M 2014 *Compos. Sci. Technol.* **104** 47
[9] Yan X, Miao J, Liu J, Wu X, Zou H, Sha D, Ren J, Dai Y, Wang J and Cheng X 2016 *J. Alloys Compd.* **677** 52
[10] Takenaka K 2012 *Sci. Technol. Adv. Mat.* **13** 013001
[11] Dove M T and Fang H 2016 *Rep. Prog. Phys.* **79** 066503
[12] Attfield J P 2018 *Front. Chem.* **6** 371
[13] Coates C S and Goodwin A L 2019 *Mater. Horizons* **6** 211
[14] Takenaka K 2018 *Front. Chem.* **6** 267
[15] Lalpoor M, Eskin D G and Katgerman L 2011 *Int. J. Mater. Res.* **102** 1286
[16] Mittal R, Gupta M K and Chaplot S L 2018 *Prog. Mater. Sci.* **92** 360
[17] Ji T, He S, Ai F, Wu J, Yan L, Hu J and Liao M 2021 *Nano Res.* **14** 3423
[18] Li C W, Tang X, Muñoz J A, Keith J B, Tracy S J, Abernathy D L and Fultz B 2011 *Phys. Rev. Lett.* **107** 195504
[19] Liu Y, Wang Z, Wu M, Sun Q, Chao M and Jia Y 2015 *Comp. Mater. Sci.* **107** 157
[20] Huu H T, Viswanath N S M, Vu N H, Lee J W and Im W B 2021 *Nano Res.* **14** 3977
[21] Wang L, Yuan P F, Wang F, Sun Q, Liang E J, Jia Y and Guo Z X 2014 *Phys. Lett. A* **378** 2906
[22] Wu J, Li Q, Shuck C E, Maleski K, Alshareef H N, Zhou J, Gogotsi Y and Huang L 2022 *Nano Res.* **15** 535
[23] Grima J N, Bajada M, Scerri S, Attard D, Dudek K K and Gatt R 2015 *Proc. Royal Soc. Math. Phys. Eng. Sci.* **471** 20150188
[24] Vila F D, Hayashi S T and Rehr J J 2018 *Front. Chem.* **6** 296
[25] Mary T A, Evans J S O, Vogt T and Sleight A W 1996 *Science* **272** 90
[26] Wang H, Yang M, Chao M, Guo J, Gao Q, Jiao Y, Tang X and Liang E 2019 *Solid State Ionics* **343** 115086
[27] Ernst G, Broholm C, Kowach G R and Ramirez A P 1998 *Nature* **396** 147
[28] Suzuki I, Seya K, Takei H and Sumino Y 1981 *Phys. Chem. Minerals* **7** 60
[29] Zhao Y, Weidner D J, Parise J B and Cox D E 1993 *Phys. Earth Planet. In.* **76** 1
[30] Mu H, Zhang Y, Zou H, Tian F, Fu Y and Zhang L 2023 *J. Phys. Chem. Lett.* **14** 190
[31] Gao Q, Wang J, Sanson A, Sun Q, Liang E, Xing X and Chen J 2020 *J. Am. Chem. Soc.* **142** 6935
[32] Zhang Y, Mu H, Xing B, Zou H, Fu Y and Zhang L 2022 *Phys. Rev. Mater.* **6** 083603
[33] Peng J, Gunda N S H, Bridges C A, Lee S, Haynes J A and Shin D 2022 *Comp. Mater. Sci.* **210** 111034
[34] Liaw A and Wiener M 2002 *R news* **2** 18
[35] Friedman J H 2001 *Ann. Statistics* **29** 1189
[36] Cortes C and Vapnik V 1995 *Mach. Learn.* **20** 273
[37] Zhao R, Xing B, Mu H, Fu Y and Zhang L 2022 *Chin. Phys. B* **31** 056302
[38] Freedman D A 2009 *Statistical models: theory and practice* (Cambridge University)
[39] Quinlan J R 1986 *Mach. Learn.* **1** 81
[40] Kresse G 1995 *J. Non-cryst. Solids* **192** 222
[41] Kresse G and Furthmüller J 1996 *Comp. Mater. Sci.* **6** 15
[42] Kresse G and Joubert D 1999 *Phys. Rev. B* **59** 1758
[43] Blöchl P E 1994 *Phys. Rev. B* **50** 17953
[44] Perdew J P, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865
[45] Perdew J P, Chevary J A, Vosko S H, Jackson K A, Pederson M R, Singh D J and Fiolhais C 1992 *Phys. Rev. B* **46** 6671
[46] Zhao X G, Zhou K, Xing B, Zhao R, Luo S, Li T, Sun Y, Na G, Xie J, Yang X, Wang X, Wang X, He X, Lv J, Fu Y and Zhang L 2021 *Sci. Bull.* **66** 1973
[47] Zhao G, Xie J, Zhou K, Xing B, Wang X, Tian F, He X and Zhang L 2022 *Chin. Phys. B* **31** 037104
[48] Altman N S 1992 *The American Statistician* **46** 175
[49] Togo A and Tanaka I 2015 *Scripta Mater.* **108** 1
[50] Sebastian L, Sumithra S, Manjanna J, Umarji A M and Gopalakrishnan J 2003 *Mater. Sci. Eng. B* **103** 289
[51] Sanson A, Rocca F, Dalba G, Fornasini P, Grisenti R, Dapiaggi M and Artioli G 2006 *Phys. Rev. B* **73** 214305
[52] Gupta M K, Mittal R, Chaplot S L and Rols S 2014 *J. Appl. Phys.* **115** 093507