npj | computational materials

Article

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

https://doi.org/10.1038/s41524-026-01970-w

Accelerated discovery of supertetragonal perovskites with giant polarization *via* machine learning

Wenguang Hu, Zebin Wu, Menglu Li, Shan Feng, Hangbo Qi, Xingjian Lu, Xiaotao Zu, Haiyan Xiao ✉ & Liang Qiao ✉

Ferroelectric perovskites with giant spontaneous polarization have extensive applications in electronic devices, energy conversion, sensor and so on. However, the rapid discovery of new perovskites with giant polarization remains an open challenge especially when thousands of candidates are treated. Here, combining machine learning (ML) and first-principles calculations, we successfully predict 8 perovskites with giant polarization from 2021 different possible compounds, among which seven candidates have never been reported before. These perovskites have large c/a ratio and giant polarization compared to the reported ferroelectric perovskites, and room temperature stability. Among them, the polarization of $SnFeO_3$ with G-AFM magnetic ordering is as high as 138.63 μC/cm^2. The non-magnetic $SrPbO_3$ and magnetic $EuSnO_3$ not only exhibit giant polarization, but also possess band gaps close to the ideal value for photovoltaic applications, showing great potential in the field of ferroelectric photovoltaics. Besides, polarity and metallicity coexist in $SnFeO_3$ and $CaTaO_3$, which are suggested to have potential applications in fields such as spintronics and superconductivity. This work thus provides an effective strategy for discovering new functional materials.

In recent years, functional materials have been particularly important in industrial development. Ferroelectric materials have gained extensive attention in the scientific community due to their spontaneous polarization characteristic, which exhibit nonzero electric dipole moment even without an external electric field[1–6]. The most common ferroelectric materials currently include perovskite, layered bismuth oxide, and boracite[7]. Tetragonal perovskite materials with a high c/a ratio often have giant spontaneous polarization, which is referred to as supertetragonal (ST) perovskites[8]. Supertetragonal perovskites, such as $BiFeO_3$, $PbTiO_3$, $BiCoO_3$, and $PbVO_3$, show great potential in the fields of photodetectors, ferroelectric nanocapacitors, and piezoelectric oscillators due to their spontaneous polarization[8–14]. However, these compounds are rare and generally require extreme synthesis conditions. For example, supertetragonal lead titanate ($PbTiO_3$) with an out–of–plane–to–in–plane lattice parameter ratio (c/a) of 1.238 and giant polarization of 236.3 μC/cm^2 is obtained in epitaxial composite thin films of tetragonal $PbTiO_3$ by Zhang and Chen *via* interphase strain[8]. Stress is also a key factor in regulating the c/a ratio and polarization intensity[14–19]. Perovskites such as $BiScO_3$, $BiMnO_3$, $BiFeO_3$ and $BiCoO_3$ exhibit an increase in the c/a ratio with increasing negative pressure[20]. Conversely, under positive pressure, the c/a ratio gradually decreases as the pressure increases. Ion implantation can produce defect-driven true supertetragonal phase with single-domain-state in ferroelectric $BiFeO_3$ thin films with c/a values of ~1.3[16]. Therefore, it is necessary to explore novel supertetragonal perovskites that can exist stably under ambient conditions and exhibit giant polarization for applications in ferroelectric devices.

Due to the complexity and diversity of perovskites structures, Density Functional Theory-based high-throughput calculations are too expensive and time-consuming, not to mention experiments. Fortunately, the launch of the Materials Genome Initiative (MGI) has brought exciting prospects for addressing this dilemma, which aims to accelerate the research and application process of new materials[21–27]. In recent years, with the maturity of material synthesis and characterization, as well as data storage technology, the field of materials has gradually entered the big data era[28]. In the past, the MGI has given rise to a multitude of high-throughput computational platforms and databases such as the Materials Project (MP)[29], Open Quantum Materials Database[30], Harvard Clean Energy Project[31], Electronic Structure Project (ESP)[32], Computational Materials Repository[33], and Novel Materials Discovery[34]. Information like atomic positions, space groups, lattice constants, and symmetries can be found in these databases, providing material engineers with ready access to the properties of known materials.

School of Physics, University of Electronic Science and Technology of China, Chengdu, China. ✉e-mail: hyxiao@uestc.edu.cn; liang.qiao@uestc.edu.cn

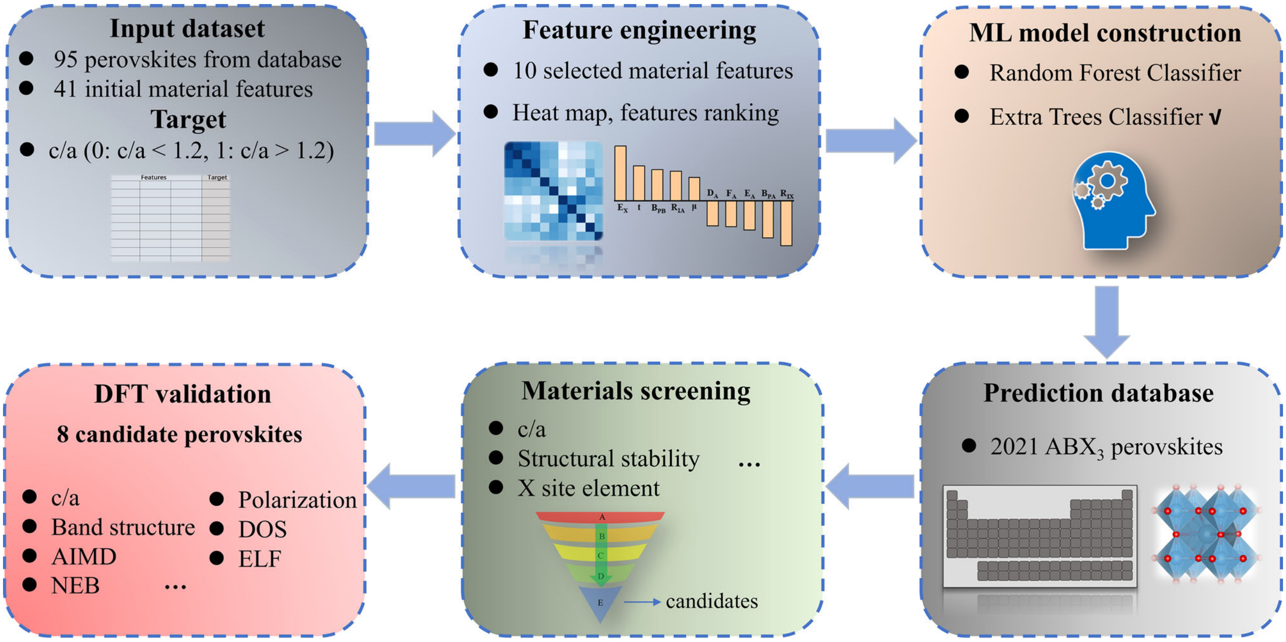


Fig. 1 | Workflow for accelerating discovery of stable supertetragonal perovskites. The entire workflow consists of six parts: input dataset preparation, feature engineering, ML model construction, prediction, materials screening and DFT validation.

With the rapid development of artificial intelligence technology, the machine learning (ML) technique has made significant advancements in materials design. For example, using ML, 15 promising lithium-ion battery solid electrolyte materials were screened out by Jalem et al.[35]. Kaufmann et al. successfully predict the entropy-forming ability of 70 new compositions by ML model, and several of these predictions were validated by density functional theory calculations (DFT) and experimental synthesis[36]. Combining ML and DFT, six orthorhombic lead-free hybrid organic-inorganic perovskites (HOIPs) with proper bandgap for solar cells and room temperature thermal stability were screened out by Lu et al. from 5158 unexplored HOIPs[37]. By bypassing complex quantum mechanics, ML technology can not only greatly accelerate materials design with high accuracy, but also reveal the structure-property relationships within materials from large material datasets.

Here, we develop a target-driven method to discover supertetragonal perovskites ABX_3 (A = monovalent or divalent cation; B = divalent or tetravalent cation; X = oxygen or halide anion) based on ML technique and DFT calculations. We first train our ML model from 95 reported perovskites, and predict the lattice parameter ratio c/a of 2021 different possible perovskites. A close structure-property relationship mapping the c/a of perovskites is concurrently excavated from ML data, in which the ranges of tolerance factor (t), octahedral factor (μ), electronegativity of A-site element (E_A), electronegativity of X-site element (E_x), first ionization of A-site element (F_A), and ionic radius of X-site element (R_{IX}) are defined for supertetragonal perovskites. After further screening, eight stable supertetragonal perovskites are selected as promising ferroelectric materials with giant polarization. These newly screened supertetragonal perovskites provide a valuable theoretical reference for related experimental investigation.

Results

Design framework

Our material screening process consists of 6 parts, i.e., Input dataset, Feature engineering, ML model construction, Prediction dataset, Materials screening, and DFT validation, which are schematically illustrated in Fig. 1. As the foundation of ML, the first step is to prepare a train-test dataset. The train-test dataset consists of input data and output data, which are the descriptor and the target performance, respectively. To remove redundant features and improve the accuracy of the ML model, feature engineering is performed to retain the most important features for the target data. Then, the ML modeling is constructed by using classification algorithms. Next, the trained ML model is applied to a prediction dataset, where the predicted material is screened based on certain constraints. Finally, DFT calculations and Ab initio molecular dynamics (AIMD) simulations are used to study the electronic properties and stability of the candidate perovskites selected from the ML simulation.

Dataset

The input data for this study, containing 95 perovskites, is obtained from Materials Project (MP)[29]. The selected 95 perovskite compounds belong to the family of perovskites of chemical formula ABX_3. Generally, it is considered that if the lattice constant ratio of c/a is greater than 1.2, then the perovskite can be defined as a supertetragonal perovskite material[8]. Accordingly, the value of 1.2 is used as a threshold to divide the c/a values in the dataset. Figure 2a visualizes the division of the dataset, where blue and green dots represent the c/a values less than and greater than 1.2, respectively. For the classification modeling, the c/a values less than 1.2 are considered as 0, and greater or equal to 1.2 as 1, to conduct a binary classification. As shown in Fig. 2b, the numbers 0 and 1 are 76 and 19, respectively. Data analysis and visualization can also reveal the hidden trends and periodicity in the perovskite dataset. The impact on the c/a ratio is evaluated based on several important descriptors, as shown in Fig. 2c–f. It can be observed from the figures that with the increase of E_x, the value of c/a tends to increase; it tends to exhibit a larger c/a when t is between 0.7 and 1.0; the lower the boiling point of the monatomic element at the A-site, the larger the c/a value tends to be; the smaller the ionic radius of the element at the X-site, the larger the c/a value tends to be.

Indeed, there are many choices for the A-, B-, and X-site elements in perovskites. As illustrated in Fig. 3a, when X is F, Cl, Br, or I, the A should be a monovalent cation, and B is a divalent cation; when X is O, the A should be a divalent cation, and B is a tetravalent cation. Consequently, for X = F, Cl, Br, or I, we consider 5 monovalent cations and 43 divalent cations from the periodic table for the A-site and B-site, respectively. In the case of X = O, we consider 43 divalent cations for the A-site and 27 tetravalent cations for the B-site. Thus, we obtain 2021 different possible perovskite compounds, which are the potential candidates we will explore in this study.

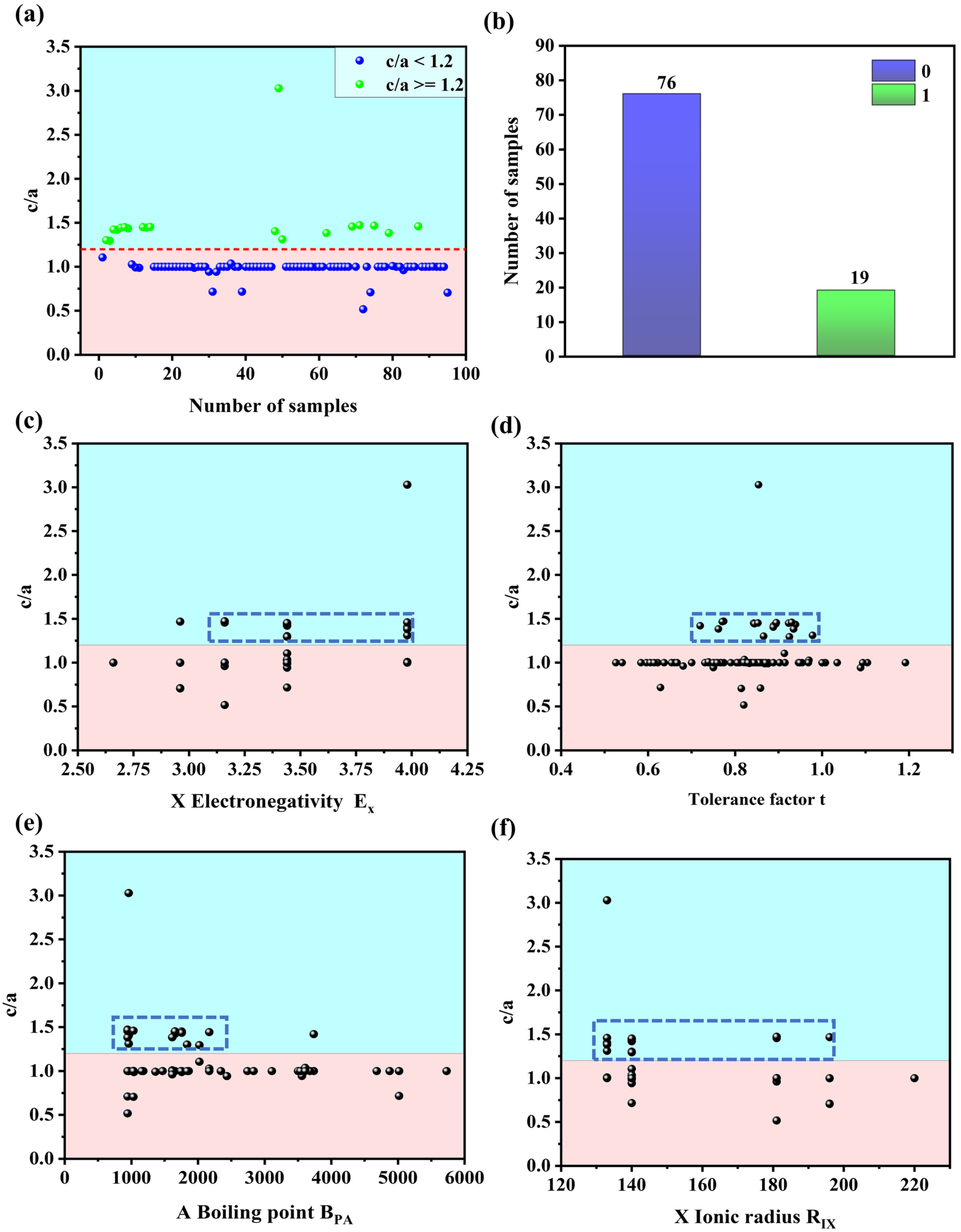


Fig. 2 | Visualization of the perovskite c/a in the train/test dataset and the structure-property relationship between perovskites c/a and features. **a** Visualization of train/test dataset, where green and blue dots represent c/a less and greater than 1.2, respectively; **b** Number of samples in binary classification, where 0 and 1 represent c/a less and greater than 1.2, respectively; **c** The relationship between c/a and electronegativity of X atom; **d** The relationship between and c/a and tolerance factor; **e** The relationship between c/a and boiling point of single substance for B-site element; **f** The relationship between c/a and Ion radius of X-site element. The dotted box represents the most appropriate range for each feature in the train/test dataset.

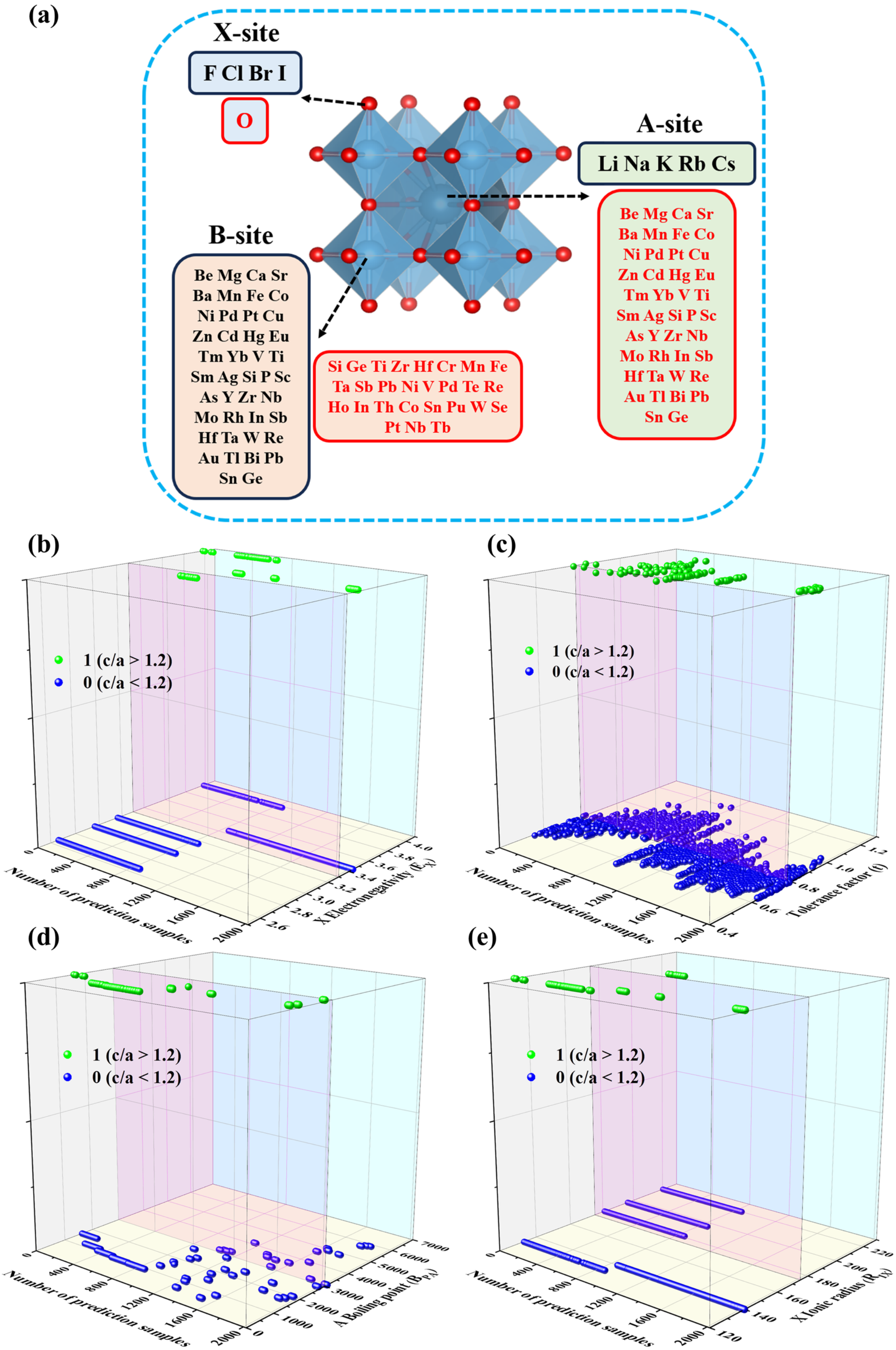


Fig. 3 | Schematic showing the details of the A, B and X-site compositions considered in the virtual space. Both train/test and prediction datasets incorporate the materials modified via the element mixing strategy and the structure-property relationship between c/a values and features in prediction datasets **a** Virtual space for ABX_3 perovskites, **b** Electronegativity of X atom E_x, **c** Tolerance factor t, **d** A Boiling point of single substance B_{PA} and **e** Ion radius of X-site element R_{IX}. The green and blue dots represent c/a greater and less than 1.2.

Feature engineering

For any ML method directed at predefining and specifying material properties, it typically relies on a certain number of features (descriptors)[24,37]. These features not only uniquely define each material in the input dataset, but also are related to the physical and chemical properties that are expected. While there might be many factors influencing the target properties of the material, the number of features must be reasonable. The best strategy is to choose features capable of completely representing the material properties, and the number of such features should be far less than the number of materials in the input dataset, to avoid the curse of dimensionality[38]. In this work, 41 initial features are selected (see Supplementary Table S1 for a complete list of features), including ionic radius, tolerance factor, and electronegativity to collectively describe the chemical space of the perovskites. Pearson correlation coefficients are calculated in the feature engineering step to eliminate the redundant features, which are depicted in the heat map for each feature pair (see Fig. S1). The heat map demonstrates that a significant number of raw features are highly correlated with each other, and some of them should be eliminated. The features with the Pearson correlation coefficients >0.9 or <-0.9 are removed, such that only the appropriate one in the feature pair remains for the next round. Consequently, 10 features are selected from the raw features based on the Pearson correlation coefficients (Fig. 4a). The new feature set contains structural features (tolerance factor t, octahedral factor μ) as well as the elemental properties of A, B and X site ions, including electronegativity of A-site element E_A, first ionization energy of A-site element F_{IA}, density of the simple substance corresponding to A-site element D_A, boiling point of the simple substance corresponding to the A-site element B_{PA}, boiling point of the simple substance corresponding to the B-site element B_{PB}, electronegativity of the X-site element E_X, ionic radius of the A-site element R_{IA} and ionic radius of the X-site element R_{IX}. It turns out that the descriptors most correlated with c/a are E_X, t, B_{PA}, and R_{IX}, as indicated by the feature importance ranking shown in Fig. 4b.

Machine learning models

An appropriate ML algorithm is important for ML modeling. In this work, two ML algorithms, i.e., Random Forest classifier and Extra Trees classifier, are compared, with 70% (66 samples) of the original dataset used as the training set and 30% (29 samples) as the test set. Figures 4c, d represent training and test set models established by using a Random Forest classifier, respectively. As shown in Table S2, this classification algorithm achieves a model *accuracy* of 0.924 on the training set and 0.897 on the test set. However, it is noted that this model still predicts 5 samples originally labeled 1 as type 0 in the training set and predicts 3 samples that should have been type 1 as type 0 in the test set. Due to the limited number of type 1 samples in the dataset, although the predictions of the model for type 0 samples are relatively accurate and the overall accuracy is acceptable, there are still significant errors in predicting type 1 samples, resulting in a *recall* of 0.643 in the training set and only 0.40 in the test set. These results show that the Random Forest classifier model is not suitable for our prediction goals to some extent. Therefore, another ML model is established by using the Extra Trees classifier. The obtained training and test set models are presented in Fig. 4e, f, respectively. It can be observed that this classification algorithm achieves an *accuracy* of 1 in both the training set and test set, where the real type and predicted type are the same, suggesting that all predictions are correct. Moreover, the consistently high accuracy observed in cross-validation (as shown in Fig. S2) further mitigates concerns that the enhanced performance of the model may arise from overfitting. Therefore, the Extra Trees classifier model can be used as a predictive model for further analysis.

Materials screening and first-principles evaluation

We predict the c/a ratio of perovskite materials in the predicted space of 2021 candidates *via* the trained ML model. As shown in Fig. 5, to screen out structurally stable supertetragonal perovskite materials, more screening is required. First, 130 perovskites are screened out from the total 2021 perovskites with ML predicted c/a ratio as type 1. Based on the stability of the structure[37], 109 samples with a tolerance factor between 0.78 and 1 are selected, among which 67 samples with an octahedral factor between 0.4 and 0.7 are further selected. As compared with the 95 input samples from the previous training and test sets included in the 2021 total samples, there are 11 duplicates of perovskite types with c/a ratios greater than 1.2 in the output samples. Consequently, 56 new perovskite materials are selected in total, all of which possess ideal c/a ratios and stable structures. Considering that oxides are relatively rich in earth, as well as cheap and stable[32,39,40], perovskite oxides are retained for further investigations. It should be noted that the reason of including halides in the dataset lies in their contribution to enhancing the diversity of the training and test sets, which in turn facilitates the construction of a more accurate and robust ML model.

Ground state properties of selected supertetragonal perovskites

After further first-principles calculations validation, there are 11 perovskites that stand out first with ordered structure, including three non-magnetic perovskites, i.e., $CaSnO_3$, $SrPbO_3$, $CaWO_3$, and eight magnetic perovskites, i.e., $EuPbO_3$, $EuSnO_3$, $BiVO_3$, $SnFeO_3$, $SrFeO_3$, $CaTaO_3$, $SrTbO_3$, and $CaTbO_3$. Since magnetic perovskites may exhibit different magnetic ordering states, it is necessary to perform first-principles calculations on different magnetic ordering states for one material to find the lowest energy state (See Table S3 for details). Figure 6b illustrates four different magnetic orderings, i.e., ferromagnetic (FM), A-type antiferromagnetic (A-AFM), C-type antiferromagnetic (C-AFM), and G-type antiferromagnetic (G-AFM), where red and black arrows represent spin up and spin down ordering of magnetic atoms, respectively. Figure 6c illustrates the energy differences between various AFM magnetic states and FM states, as well as the c/a ratios for eight magnetic perovskites. According to Fig. 6c, for magnetic perovskites $SrFeO_3$, $SrTbO_3$, and $CaTbO_3$, the c/a ratios are all less than 1.2. These three perovskites, thus, are excluded from further consideration. Besides, for G-AFM and A-AFM $EuPbO_3$, FM and A-AFM $EuSnO_3$, FM and A-AFM $BiVO_3$, as well as C-AFM and G-AFM $SnFeO_3$, they not only exhibit the lowest energy but also have c/a ratios greater than 1.2. Hence, they are considered as ideal candidates for supertetragonal perovskites. Although the $CaTaO_3$ with FM and A-AFM magnetic orderings do not possess the lowest energy, but their c/a ratios are greater than 1.2, indicative of its potential as supertetragonal perovskites. Obviously, the magnetic orderings of these perovskites have nonnegligible impacts on the lattice parameters and ground state energy of the material, highlighting the importance of magnetism as a factor that researchers cannot ignore. To the end, 8 perovskites are finally screened out, including three non-magnetic perovskites, i.e., $CaSnO_3$, $SrPbO_3$, $CaWO_3$, and five magnetic perovskites, i.e., G-AFM and A-AFM $EuPbO_3$, FM and A-AFM $EuSnO_3$, G-AFM and C-AFM $SnFeO_3$, FM and A-AFM $CaTaO_3$, and FM and A-AFM $BiVO_3$ (see Fig. 6a for detail structures).

Thermal and chemical stabilities of eight selected supertetragonal perovskites

Based on the optimized geometrical structures, AIMD simulations are further carried out to evaluate the thermal stability of selected supertetragonal perovskites. As shown in Fig. 7, although the energies of $SrPbO_3$, $CaWO_3$, and $EuSnO_3$ exhibit relatively large fluctuations, all the screened supertetragonal materials keep an ordered crystal structure within a 5000 fs process at a temperature of 300 K. To clarify the origin of these fluctuations, we present the geometrical structures of all screened materials at their minimum and maximum energy in Fig. S3. For $SrPbO_3$, $CaWO_3$, and $EuSnO_3$, while the temporary structural fluctuations observed at the maxima may suggest local instabilities, such deviations are expected under finite-temperature molecular dynamics simulations and do not necessarily undermine the overall thermodynamic stability of the phase. The recovery to the supertetragonal configuration at the energy minima further indicates that the polar distortion is robust against thermal perturbations, supporting the feasibility of stabilizing these phases experimentally under ambient conditions. The other five materials exhibit only small energy fluctuations throughout the entire AIMD simulation process and retain a well-preserved

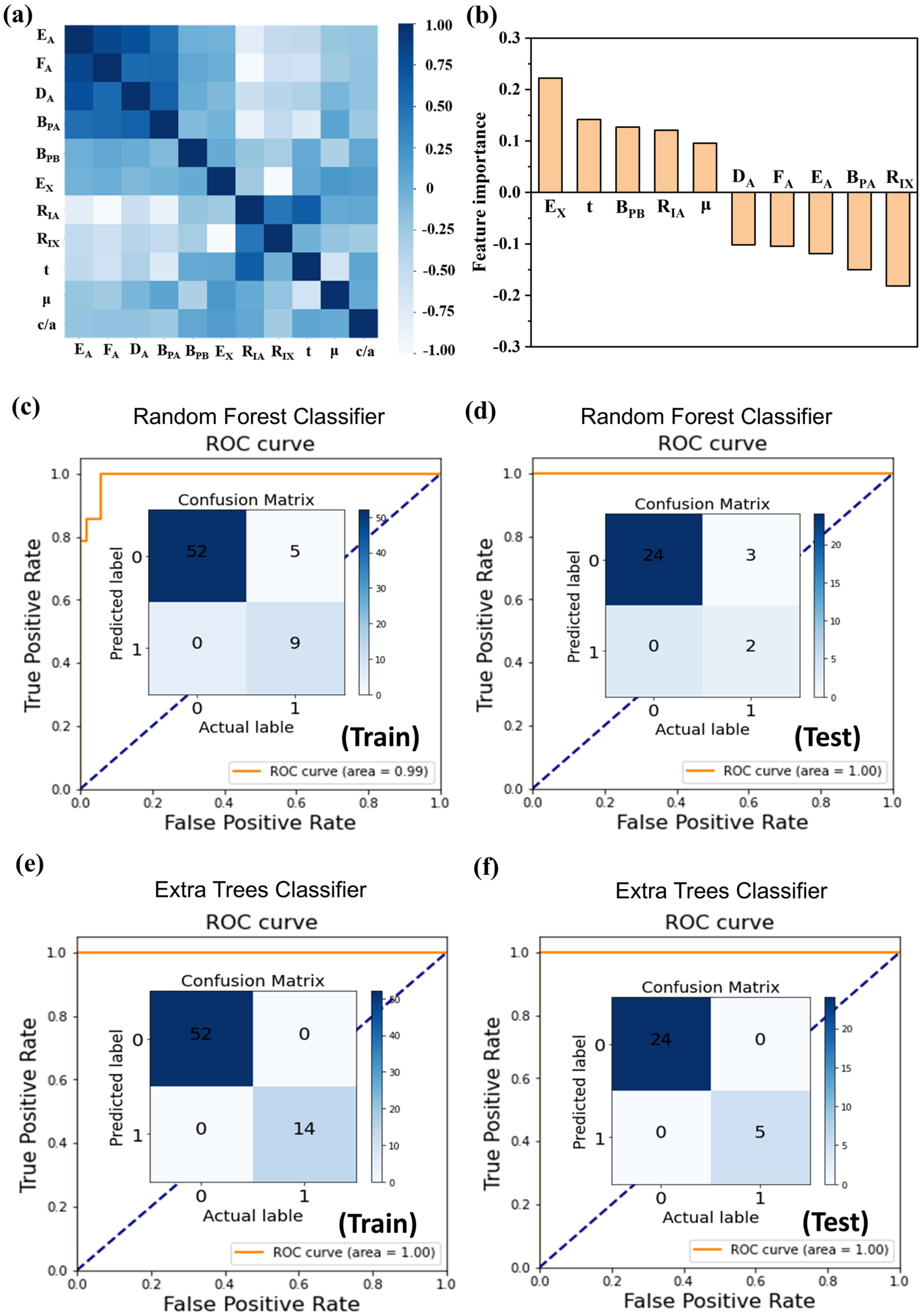


Fig. 4 | Importance and correlation of 10 selected features and confusion matrix between the actual label and predicted label using Random Forest Classifier and Extra Trees Classifier. **a** The heat map of the Pearson correlation coefficient matrix among the 10 selected features and target c/a values. **b** The 10 selected features are ranked according to their Pearson correlations with the target c/a values. **c** Training model for a random forest classifier. **d** Test model for random forest classifier. **e** Training model for Extra Trees classifier and **f** Test model for Extra Trees classifier.

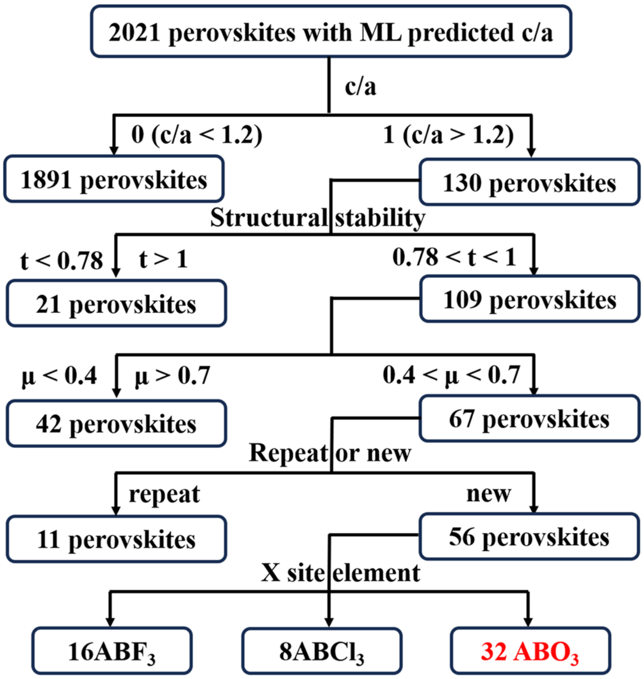


Fig. 5 | Workflow for optimal supertetragonal perovskites screen. The screening is based on four criteria: whether c/a exceeds 1.2, whether the structure is theoretically stable, whether it is new and the species of X-site element.

supertetragonal phase at both the maximum and minimum energy states. Therefore, it can be concluded that the supertetragonal materials screened in this work can remain stable at room temperature.

In addition to the thermal stability analysis, we have also calculated the formation energy to evaluate the chemical stability of screened materials. Table S4 and Table S5 present the chemical potentials of constituent elements and formation energies of the eight screened supertetragonal materials, respectively. It is shown that all formation energies are negative, ranging from −1.2 to −2.7 eV/atom. These results are close to the formation energies of $CaTiO_3$ and $BaTiO_3$ (~ −3.5 eV/atom) that have been reported experimentally[41], suggesting that the screened materials are chemically stable and have the potential to be synthesized experimentally.

Ferroelectric properties of eight selected supertetragonal perovskites

Subsequently, ferroelectric properties of these perovskites are investigated. Considering that the ability to switch polarization under an electric field is the key to define a ferroelectric material[42,43], here we perform nudged elastic band (NEB) calculations to evaluate the polarization reversal pathways and energy barriers of the screened materials (see Fig. S4). The calculated energy barriers for the eight compounds lie in the range of 0.175–0.720 eV/f.u., comparable with the values of ferroelectric $CuCrS_2$ (0.23 eV/f.u.)[44] and $BiFeO_3$ (0.43 eV/f.u.)[45]. Although the supertetragonal phases of $CaWO_3$, $EuSnO_3$, and $CaTaO_3$ are predicted to be metastable, they still exhibit inversion-symmetry breaking and maintain distinct polar distortions. Obviously, the screened materials possess reversible polarization, indicative of their potential to be ferroelectrics. In addition, the exact values of polarization are calculated and summarized in Table 1. For non-magnetic

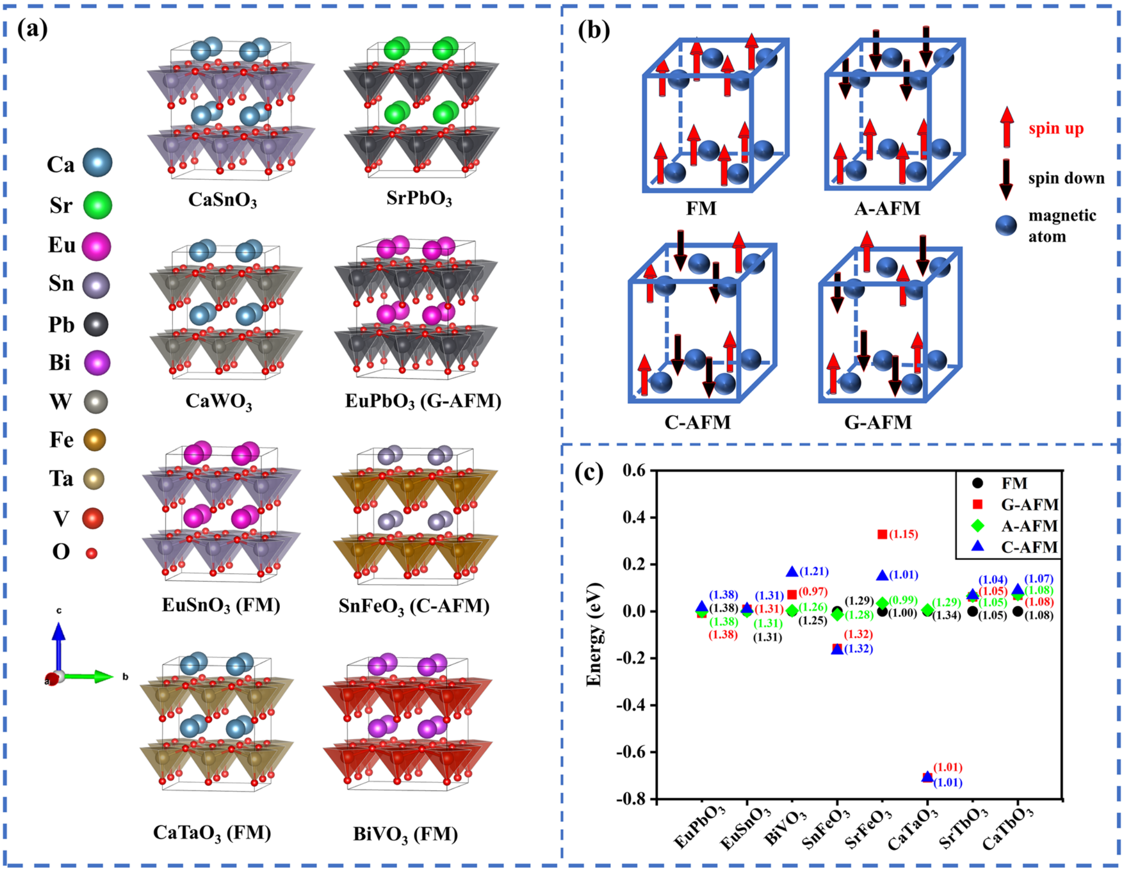


Fig. 6 | The supertetragonal perovskites screened by ML and DFT. a optimized geometrical structures of 8 selected supertetragonal perovskites, **b** illustration of different magnetic orderings where red and black arrows represent spin up and spin down, **c** energy differences between various magnetic orderings and the FM ordering, and c/a values for magnetic perovskites with different magnetic orderings. The value inside the parentheses is the c/a ratio.

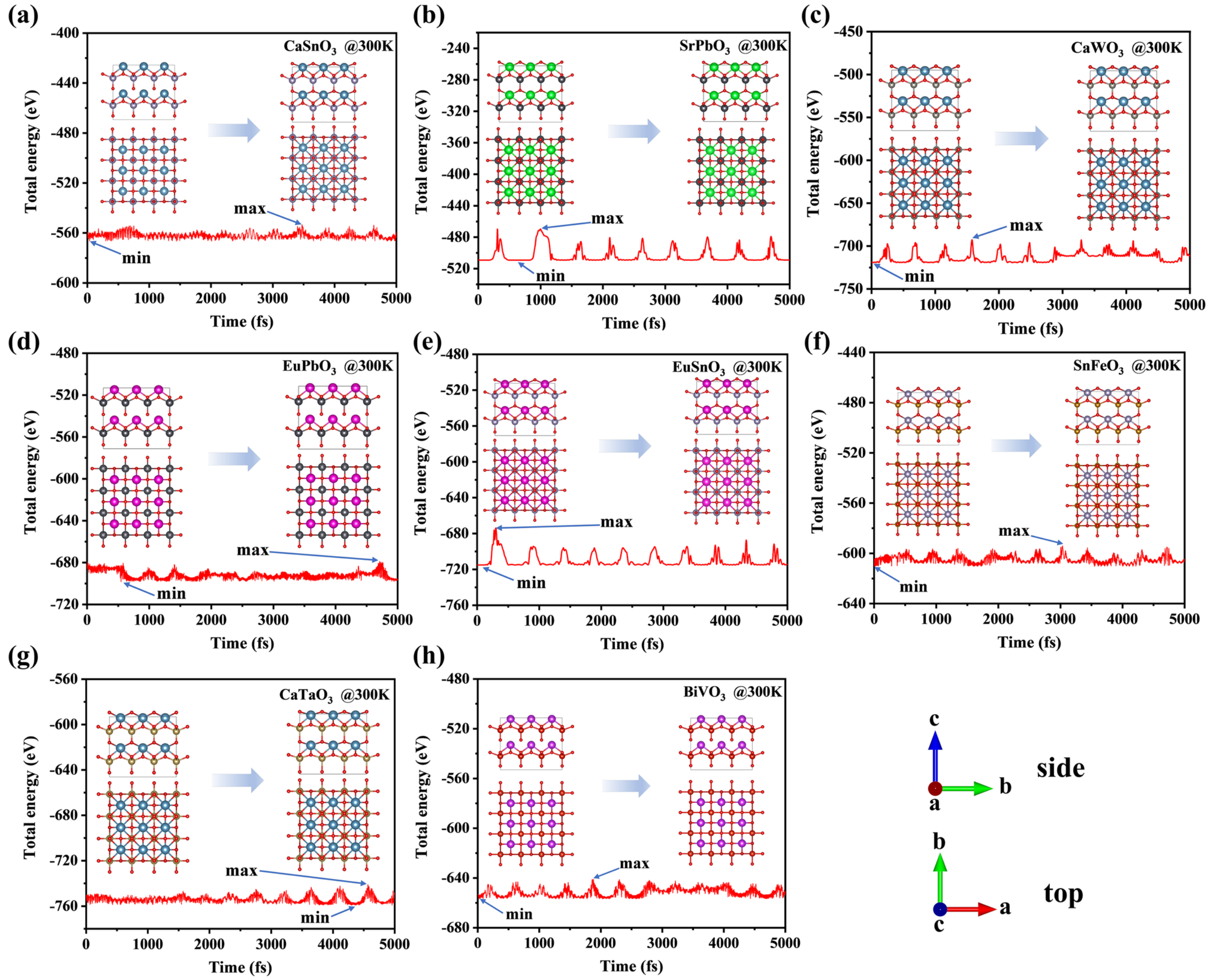


Fig. 7 | Variation of total energy with time for 8 selected supertetragonal perovskites. **a** CaSnO$_3$, **b** SrPbO$_3$, **c** CaWO$_3$, **d** EuPbO$_3$, **e** EuSnO$_3$ **f** SnFeO$_3$ **g** CaTaO$_3$ and **h** BiVO$_3$. The AIMD simulations are performed at room temperature (300 K) by using the Nosé-Hoover method. Geometrical structures of each compound before and after AIMD simulations are presented. Min and max represent the minimum and maximum energy points, respectively.

Table 1 | Lattice constants, polarization, and bandgap of eight selected supertetragonal perovskites

Formula	Magnetic ordering	a (Å)	c (Å)	c/a	MM (μB)	Polarization (μC/cm^2)	Eg (eV) (PBE)	Eg (eV) (HSE)
CaSnO$_3$	—	3.823	5.077	1.328	—	103.248	1.21	2.99
SrPbO$_3$	—	4.004	5.401	1.349	—	99.860	\	1.14
CaWO$_3$	—	3.668	4.996	1.362	—	117.438	\	0.44
EuPbO$_3$	G-AFM	3.986	5.484	1.376	6.557	98.095	\	0.78
	A-AFM	3.988	5.490	1.377	6.556	105.186	\	0.58
EuSnO$_3$	FM	3.911	5.120	1.309	6.701	90.996	\	1.20
	A-AFM	3.910	5.120	1.310	6.702	95.255	\	1.23
SnFeO$_3$	G-AFM	3.656	4.835	1.322	3.468	126.013	\	\
	C-AFM	3.660	4.823	1.318	3.403	138.627	\	\
CaTaO$_3$	FM	3.740	5.022	1.343	0.458	126.249	\	\
	A-AFM	3.790	4.908	1.295	0.371	127.779	\	\
BiVO$_3$	FM	3.787	4.748	1.254	1.683	124.633	\	\
	A-AFM	3.777	4.760	1.260	1.662	124.166	\	0.68

MM magnetic moment of Eu, Fe, Ta, and V. E$_g$: bandgap.

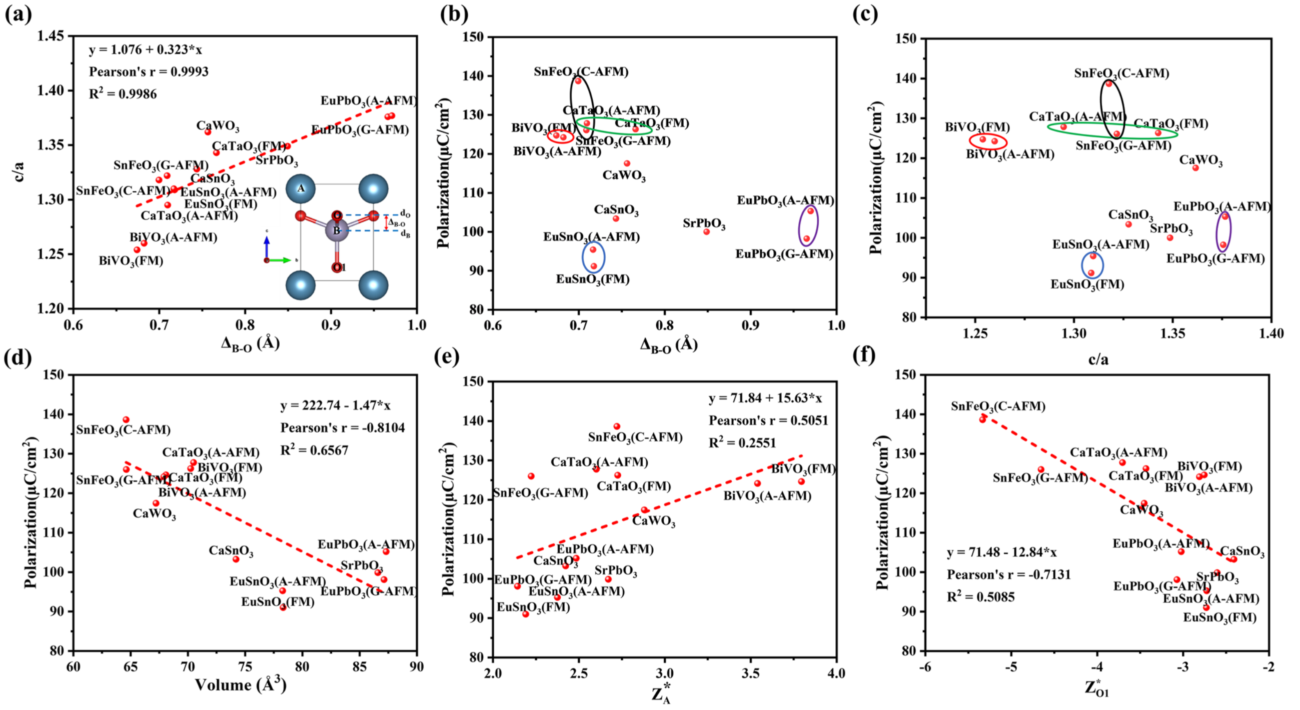


Fig. 8 | Illustration of ion displacement and the relationships between structural parameters and polarization in supertetragonal perovskites. The relationships between **a** $\Delta_{B\text{-}O}$ and c/a ratio; **b** $\Delta_{B\text{-}O}$ and polarization; **c** c/a ratio and polarization; **d** volume of unit cell and polarization; **e** born effective charge of A-site atom (Z^*_A) and polarization; **f** born effective charge of O_1 (Z^*_{O1}) and polarization of 8 screened supertetragonal perovskites. In (**a**), d_O and d_B represent the displacement of the B-site atom and O atom along the *c* axis and $\Delta_{B\text{-}O} = |d_B - d_O|$, respectively. Since O has two different sites, the O at the top of the octahedra is defined as O_1.

$CaSnO_3$, $SrPbO_3$, and $CaWO_3$, the polarizations are predicted to be 103.248, 99.86, and 117.438 μC/cm^2, respectively. As for magnetic perovskites, it is found that magnetic ordering has different influences in different compounds. For $EuPbO_3$ and $EuSnO_3$, small differences exist in the polarization between different magnetic states, since the polarizations of 105.186 and 95.255 μC/cm^2 for their A-AFM state are only 7.091 and 4.259 μC/cm^2 larger than those of G-AFM and FM states, respectively. As for $SnFeO_3$, the influence of magnetic ordering on the polarization is larger than the cases of $EuPbO_3$ and $EuSnO_3$, as indicated by the difference of 12.614 μC/cm^2 between G-AFM and C-AFM states. In the case of $CaTaO_3$ and $BiVO_3$, the magnetic ordering has almost no effect on polarization, which only shows the difference of 1.53 and 0.467 μC/cm^2 between FM and A-AFM states, respectively. Obviously, all the perovskites exhibit giant polarization, varying from 90.996 to 138.627 μC/cm^2. As compared with the traditional ferroelectric material $BiFeO_3$ with spontaneous polarization of 150 μC/cm^2 driven by epitaxial strain[46] and defect engineering[47], the $PbTiO_3$ with polarization of 75 μC/cm^2 with high Curie temperature and high thermoelectric coefficient[48,49], the $BiCoO_3$ with polarization of 126 μC/cm^2 predicted by DFT within the local spin-density approximation[50] and the $PbVO_3$ with polarization above 100 μC/cm^2 under ambient and high pressure[14], the supertetragonal perovskites that are screened out in this work are stable at room temperature and can be obtained without any external conditions, suggesting that they are all potential excellent ferroelectric materials. Notably, the polarization of $BiVO_3$ has been predicted to be 128 μC/cm^2 by Abbassi et al. employing DFT calculation with the generalized gradient approximation (GGA) and modified Becke Johnson approximation[51], which agrees well with our calculated value of ~125 μC/cm^2, further validating the robust predictive performance of our ML model and the reproducibility of our polarization results.

The intrinsic spontaneous polarization, which is widely recognized as the most important feature of a ferroelectric perovskite, depends on its structural information, such as axial ratio c/a or off-center ion displacements[52–54]. To explore the origin of the different polarization of eight perovskite oxides, we further analyze the relationship between polarization and structural parameters of the predicted materials (see Table S6 for details). Figure 8a presents the relationship between $\Delta_{B\text{-}O}$ and the c/a ratio. It can be seen that the $\Delta_{B\text{-}O}$ is linearly correlated with c/a to a certain degree, since the A-site atom has more space for displacement than the B-site atom due to the fact that A is located between oxygen octahedra, while B is inside the octahedra[54]. Consequently, as the $\Delta_{B\text{-}O}$ increases, the A-site atom will displace much more than the B-site atom to keep the structure stable, resulting in an increase in c/a. Figures 8b, c illustrate the relationships between the $\Delta_{B\text{-}O}$ and polarization as well as between the c/a and polarization, respectively. Obviously, there is no strong correlation between $\Delta_{B\text{-}O}$ and polarization, as well as between c/a ratio and polarization in different supertetragonal materials. This is different from the case of $PbTiO_3$[54], for which polarization shows a positive correlation with c/a ratio. The presented results suggest that this mechanism apply only to the same material, but not to different kinds of materials. Furthermore, we conduct a detailed analysis of the born effective charges and volumes of eight supertetragonal perovskites. As shown in Fig. 8d, the polarization shows a strong linear negative correlation with the volume, i.e., the smaller the volume, the larger the polarization. This suggests that the polarization of the system is affected by its size, which is mainly because a reduced unit cell volume enhances the dipole density per unit area, thus giving rise to the relatively large polarization. Theoretically, Zhang et al. tuned the structure of supertetragonal $PbTiO_3$ via strain to modulate the polarization strength; Tariq et al. changed the unit cell volume by varying the film thickness of $ATiO_3$ (A = Ba and Pb) and applying strain engineering, and found that the polarization strength is affected[54,55]. These investigations demonstrate that the change in volume has a significant impact on the polarization strength, agreeing well with our calculations. Besides, it is found that there is no relationship between polarization and born effective charge of the B-site atom (Z^*_B) as well as the O atom (Z^*_O), as shown in Fig. S10. However, Fig. 8e, f shows that the polarization is almost linearly correlated with the Born effective charge of the A-site atom (Z^*_A) and O_1 atom (Z^*_{O1}). This finding agrees well with the results of feature engineering shown in Fig. 4b, suggesting that the type of A-site element has a significant impact on the polarization of supertetragonal perovskites.

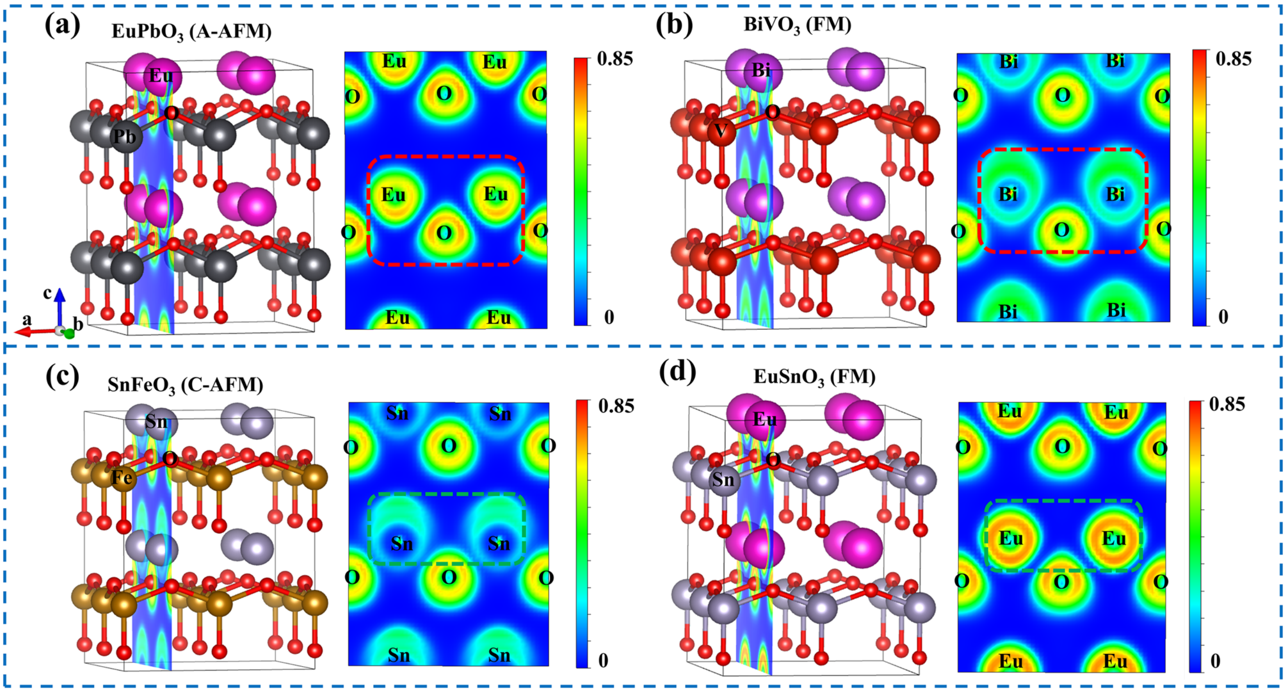


Fig. 9 | The electron localization function (ELF) of representative supertetragonal perovskites in (1 0 0) plane. **a**, **b** ELF and 2D image for A-AFM EuPbO$_3$ and FM BiVO$_3$ with the biggest and smallest c/a ratio, and **c**, **d** for C-AFM SnFeO$_3$ and FM EuSnO$_3$ with the biggest and smallest polarization in 8 screened supertetragonal perovskites, respectively.

To gain insight into the mechanisms underlying the giant c/a and polarization of these supertetragonal perovskites, the electron localization function (ELF)[56] on the (100) plane of representative perovskites is further calculated and shown in Fig. 9. The ELF values can be ranged from 0 to 1. If it is close to 0, it means that the cation-anion bond is of ideal ionic character; if it is close to 1, it represents that the cation-anion bond is close to ideal covalent nature[57]. The ELF for A-AFM EuPbO$_3$ and FM BiVO$_3$, which are characterized by the largest and smallest c/a ratios among eight screened supertetragonal perovskites, is illustrated in Fig. 9a, b, respectively. It is evident that the charge density of the A-site atom Eu is significantly higher than that of Bi, which results in a stronger interaction between Eu and O compared to the interaction between Bi and O[58]. Thus, the displacement of O towards Eu is larger than that of Bi, leading to the fact that $\Delta_{\text{Pb-O}}$ is larger than $\Delta_{\text{V-O}}$. Since the c/a ratio has a strong positive correlation with $\Delta_{\text{B-O}}$, as shown in Fig. 8b, the c/a ratio of A-AFM EuPbO$_3$ is larger. For the C-AFM SnFeO$_3$ and FM EuSnO$_3$ with the largest and smallest polarization in eight screened supertetragonal perovskites, the ELF is illustrated in Fig. 9c, d, respectively. One can see from the figures that the density of electron gas surrounding the A-site atom Sn is significantly lower than that of Eu. For cations, fewer localized electrons typically exhibit greater positive oxidation states, which corresponds to a larger Born effective charge[59]. This analysis, in conjunction with the insights from Fig. 8d–e, accounts for the observed larger polarization in C-AFM SnFeO$_3$.

Electronic structures of eight selected supertetragonal perovskites

Furthermore, the electronic structures of eight selected supertetragonal perovskites are studied (see Figs. S6 and S7 for more details). The PBE functional under GGA approximation is employed first. As shown in Figs. S8 and S9, only CaSnO$_3$ has a bandgap value of 1.21 eV and the other seven perovskites show metallic characteristics. Considering that the GGA/PBE functional generally underestimates the band gap of materials, the Heyd–Scuseria–Ernzerhof (HSE) screened hybrid functional is further employed, which is known to be more successful than standard DFT in predicting the bandgaps of a wide range of materials[60–62]. The specific values of the bandgaps are presented in Table 1. As illustrated in Fig. 10a–c, three non-magnetic supertetragonal perovskites exhibit insulating or semiconducting behavior. The band structures of CaSnO$_3$ and SrPbO$_3$ show indirect character with bandgap values of 2.99 eV and 1.14 eV, respectively, while CaWO$_3$ is of a direct bandgap about 0.44 eV. Among five magnetic supertetragonal perovskites, the SnFeO$_3$, CaTaO$_3$, and FM BiVO$_3$ exhibit no bandgap, indicative of metallic characteristics, while the EuPbO$_3$, EuSnO$_3$, and A-AFM BiVO$_3$ are of semiconducting character (Fig. 10d–h). It is noted that the non-magnetic SrPbO$_3$ and magnetic EuSnO$_3$ not only exhibit giant polarization, but also possess band gaps close to the ideal value of 1.2 eV for photovoltaic applications, showing great potential in the field of ferroelectric photovoltaics[63–65]. Besides, in SnFeO$_3$ and CaTaO$_3$, polarity and metallicity coexist in the same phase, i.e., SnFeO$_3$ and CaTaO$_3$ are polar metals in nature. Therefore, these two compounds may have wide implications in the fields of topology, ferroelectricity, magnetoelectricity, spintronics, and superconductivity[66–69].

As shown in Fig. 11, for non-magnetic supertetragonal perovskites (Fig. 11a–c), the valence band maxima (VBM) are partly contributed by the O 2*p* orbitals and some orbitals from the metal atoms at the B site, while the conduction band minima (CBM) are mainly contributed by the orbitals of the B site atom. For EuPbO$_3$ with G-AFM magnetic ordering and EuSnO$_3$ with FM magnetic ordering (Fig. 11d, e), the VBM is partly contributed by the Eu 4*f* orbital, and the CBM is mainly contributed by O 2*p* orbitals. For SnFeO$_3$ with C-AFM magnetic ordering (Fig. 11f), CaTaO$_3$ with FM magnetic ordering (Fig. 11g), and BiVO$_3$ with FM ordering (Fig. 11h), they have no bandgap, and the electrons distributed on the Fermi level are primarily contributed by the *d* orbitals of metal atoms at the B site.

Physical insights between descriptors and structures

To further understand the structure-property relationships in the eight predicted materials, we analyze descriptors with physical significance. It is easy to find that the descriptor of electronegativity is important for the c/a ratio from Fig. 4b. Electronegativity describes the ability of an atom or ion to attract electrons in a chemical bond[70]. Here, the electronegativity of the A-site element is denoted as E_A, which ranges from 0.79 to 2.54 (see source

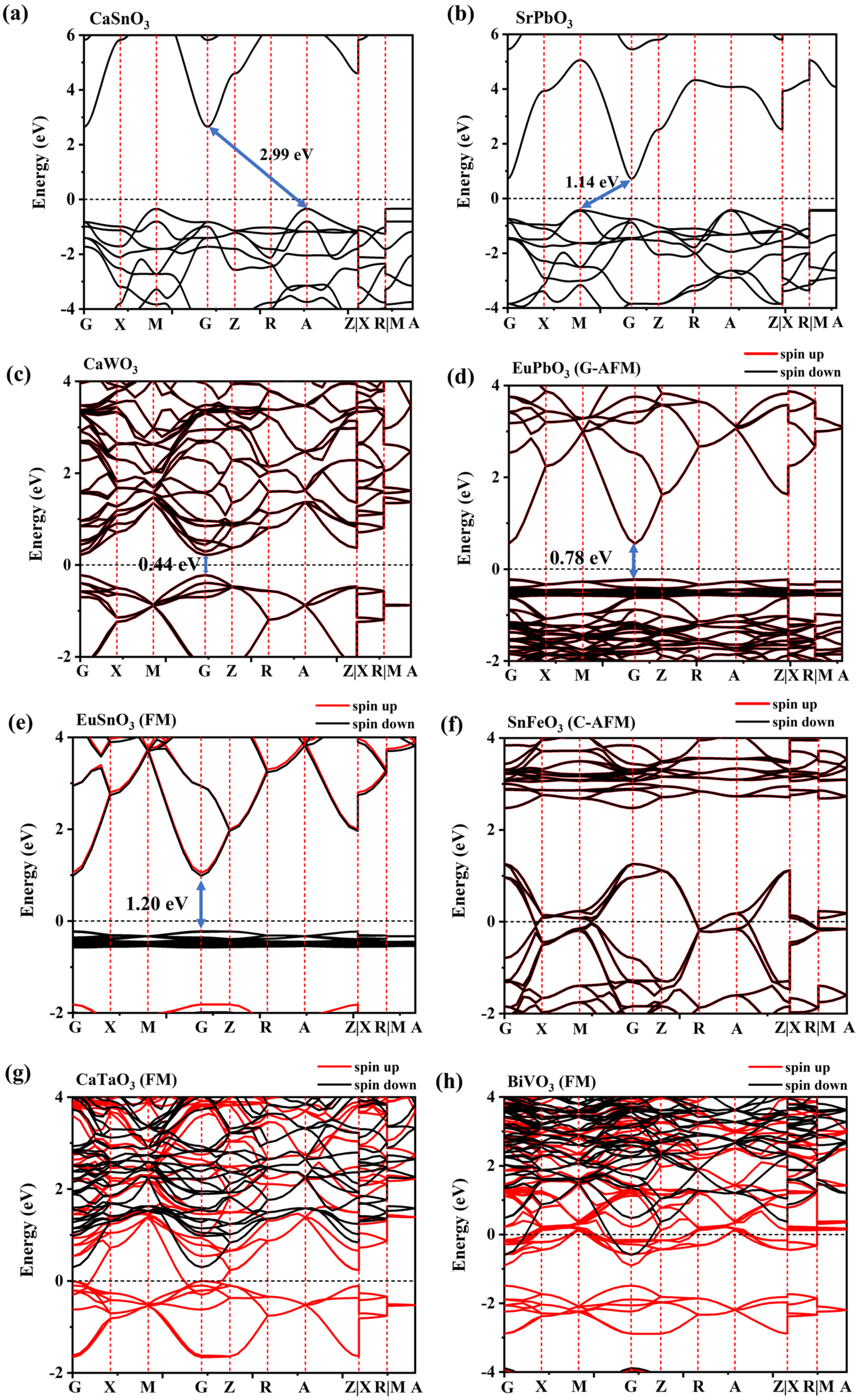
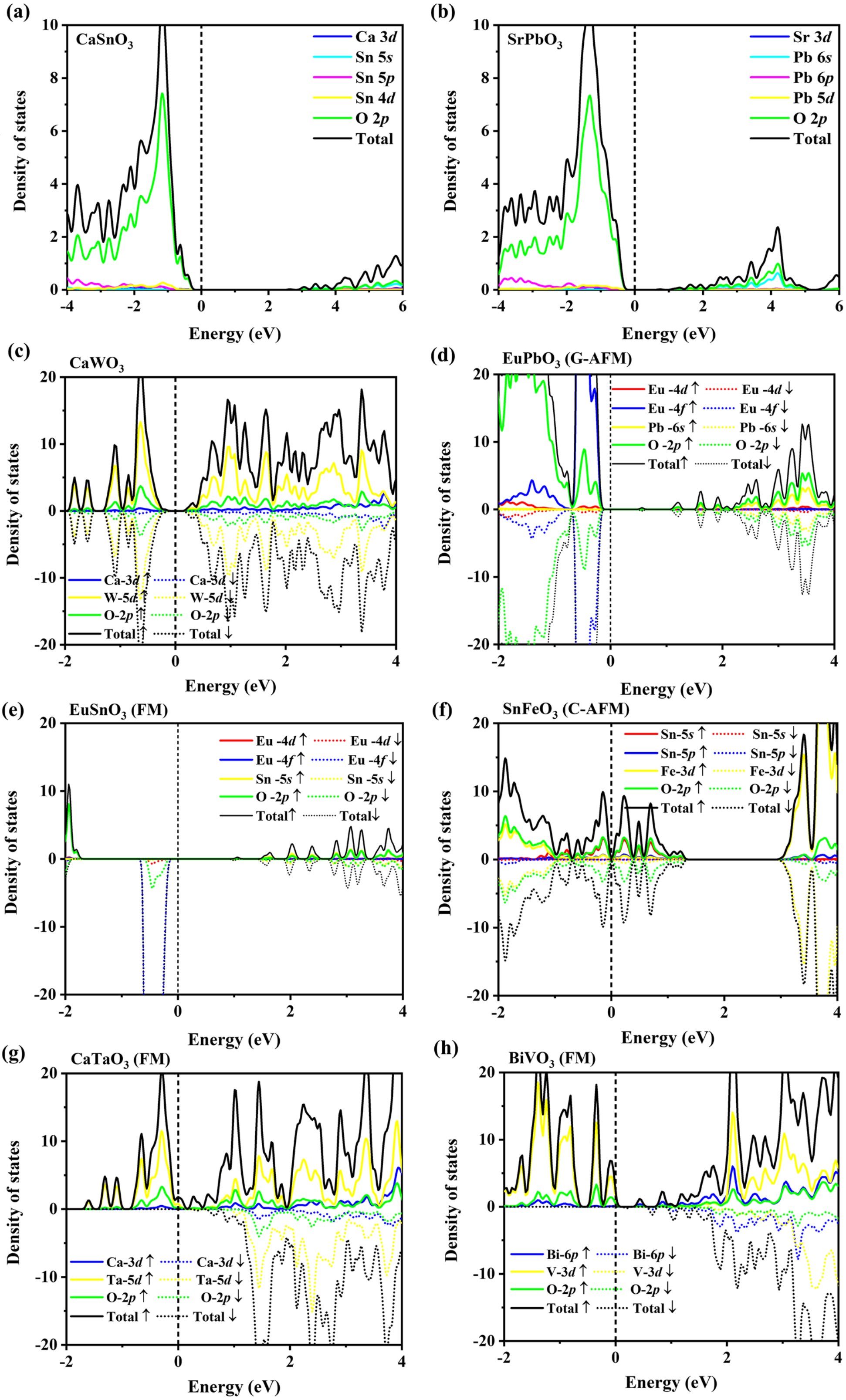


Fig. 10 | The band structures for 8 selected supertetragonal perovskites obtained by HSE functional. **a** CaSnO$_3$, **b** SrPbO$_3$, **c** CaWO$_3$, **d** EuPbO$_3$ with G-AFM magnetic ordering, **e** EuSnO$_3$ with FM magnetic ordering, **f** SnFeO$_3$ with C-AFM magnetic ordering, **g** CaTaO$_3$ with FM magnetic ordering and **h** BiVO$_3$ with FM magnetic ordering. The red and black lines represent spin-up and spin-down states, respectively.

data). Figure 4b shows that E_A is negatively correlated with c/a, i.e., a smaller E_A corresponds to a larger c/a. As shown in Table 2, the screened eight perovskites have E_A values close to 1, with the highest being 2.02, which are within the lower end of the range. The electronegativity of element X (E_X) is positively correlated with c/a and has a value range of 2.66–3.98 (see source data). The E_X of oxygen is 3.44, which lies towards the higher end of the range. One can see from Fig. 4b that the value of c/a is positively correlated with the electronegativity of the X-site element and negatively correlated

Fig. 11 | Density of states for 8 selected super-tetragonal perovskites obtained by the HSE functional. **a** $CaSnO_3$, **b** $SrPbO_3$, **c** $CaWO_3$, **d** $EuPbO_3$ with G-AFM magnetic ordering, **e** $EuSnO_3$ with FM magnetic ordering, **f** $SnFeO_3$ with C-AFM magnetic ordering, **g** $CaTaO_3$ with FM magnetic ordering and **h** $BiVO_3$ with FM magnetic ordering. The solid and dash lines represent spin-up and spin-down states, respectively.



with the electronegativity of the A-site element. This is because X-site ions tend to gain electrons, while A-site ions tend to lose electrons. The stronger the electronegativity of the X-site element, the easier it is to attract the A-site elements. Conversely, the lower the electronegativity of the A-site element, the easier it is to be attracted. This leads to a structural offset, resulting in an increasing trend in the c/a ratio.

F_A represents the first ionization energy of the A-site element, and it shows a negative correlation with the c/a ratio (see Fig. 4b). The predicted F_A ranges from 3.89 to 10.49 KJ/mol (see source data). Most of the selected eight perovskites have F_A values around 5–6, which are towards the lower end of the range. The first ionization energy is the amount of energy required to remove the outermost electron from an atom[71]. An atom with a lower first

Table 2 | Eight selected perovskites with relevant descriptor values

Formula	E_A	F_A (KJ/mol)	D_A (g/ml)	B_{PA} (K)	B_{PB} (K)	E_X	R_{IA} (pm)	R_{IX} (pm)	t	M
$CaSnO_3$	1	6.1132	1.54	1757	2875	3.44	99	140	0.8086	0.4928
$EuSnO_3$	1.2	5.6704	5.24	1802	2875	3.44	117	140	0.8695	0.4929
$EuPbO_3$	1.2	5.6704	5.24	1802	2022	3.44	117	140	0.8355	0.5536
$SrPbO_3$	0.95	5.6949	2.64	1655	2022	3.44	112	140	0.8193	0.5536
$SnFeO_3$	1.96	7.3439	7.29	2875	3134	3.44	112	140	0.8977	0.4179
$CaTaO_3$	1	6.1132	1.54	1757	5731	3.44	99	140	0.8125	0.4857
$CaWO_3$	1	6.1132	1.54	1757	5828	3.44	99	140	0.8204	0.4714
$BiVO_3$	2.02	7.2856	9.81	1837	3680	3.44	103	140	0.8678	0.4143

ionization energy is more easily attracted by an ion with a negative charge. Therefore, the smaller the first ionization energy of the A element, the more easily it is attracted by the X-site element, leading to structural deviation and an increasing trend in c/a ratio.

R_{IA} refers to the radius of the A-site ion, with a predicted spatial value range of 43–167 pm (see source data). As shown in Fig. 4b, there is a positive correlation between R_{IA} and the c/a ratio. The majority of the eight selected perovskites have R_{IA} around 110 pm, which is on the larger side of the range. On the other hand, the Ionic radius of element X (R_{IX}) is negatively correlated with the c/a value (see Fig. 4b), with a predicted spatial range of 133–220 pm (see source data). The R_{IX} of oxygen is 140 pm, which is on the smaller side of the range. The ionic radius of an element is determined by the number of charges contained in the ion and the changes in the electron configuration[72]. The smaller the nuclear charge of an A-site ion, the larger its radius will be, and the smaller the nuclear charge, the more easily the ion can be attracted by X-site ions. The same analysis applies to X-site ions as well, leading to a structural shift and an increasing trend in c/a ratio.

The structure-property relationships between c/a values and features in prediction datasets are shown in Fig. 3b–e. When E_X and t are larger, c/a tends to be greater than 1.2; whereas when B_{PA} and R_{IX} are smaller, c/a tends to be less than 1.2. The descriptor values and their importance rankings corresponding to the selected perovskites are aligned, from which we can get several guiding suggestions for subsequent predictions of supertetragonal perovskites with large c/a ratios, i.e., the elements on the A-site should have a smaller electronegativity, a lower first ionization energy, a lower density, a lower boiling point, and a larger ionic radius. Besides, the boiling point of the B-site atom should be high, the ionic radius associated with the X-site element should be small, and its electronegativity should be large. This provides guidance for accelerating the discovery of supertetragonal materials with giant polarization, as well as for other functional materials.

Discussions

Combining ML technology and DFT calculations, we have developed a highly efficient method for discovering supertetragonal perovskites with giant polarization. Through screening 2021 perovskite materials, we successfully identify eight stable supertetragonal perovskites, i.e., $CaSnO_3$, $SrPbO_3$, $CaWO_3$, $EuPbO_3$, $EuSnO_3$, $SnFeO_3$, $CaTaO_3$, and $BiVO_3$, among which seven candidates have not been reported before. The eight supertetragonal perovskites selected by ML exhibit significantly giant polarization, varying from 90.996 to 138.627 $\mu C/cm^2$. The further NEB calculations show that these materials possess switchable polarization under an external field and are promising ferroelectrics. Unlike other normal ferroelectric perovskites, which need extreme synthesis conditions, they are thermally stable under ambient condition and no external conditions are needed, thereby leading to a substantial reduction in material development costs. Furthermore, these selected supertetragonal perovskites demonstrate superior theoretical stability and hold promise for practical production applications. Among them, $SrPbO_3$ and $EuSnO_3$ possess an ideal bandgap of around 1.2 eV and giant polarization, which are thus proposed to be potential ferroelectric photovoltaic candidates. Meanwhile, $SnFeO_3$ and $CaTaO_3$ are predicted to be polar metals, which have potential applications in the fields of topology, spintronics, and superconductivity. These results underscore the potential of our research approach in facilitating the design of novel functional materials for practical application. More importantly, we have established a clear structure-property relationship that offers valuable insight for the discovery of supertetragonal materials with giant polarization. Our finding highlights the crucial impact of various factors on the c/a ratio of tetragonal perovskites, such as the electronegativity of the X and A-site elements, tolerance factor, ionic radius of the X-site element, and boiling point of the single substance for the B-site element, providing guidance for the design of perovskites with enhanced polarization.

Methods

Density functional theory calculation

All the DFT calculations are performed using the Vienna Ab Initio Simulation Package[73–75]. The interaction between ions and electrons is described using the projector augmented wave[76] method, and the exchange-correlation effects between electrons are treated using the functional of Perdew–Burke–Ernzerhof (PBE)[77] under the generalized gradient approximation (GGA)[78]. The kinetic-energy cutoff is set to be 600 eV and a $9 \times 9 \times 7$ Gamma k-point mesh is employed. The convergence criterion for total energies is 1×10^{-5} eV. The Gaussian smearing (ISMEAR = 0) and SIGMA = 0.05 (corresponding to a smearing temperature of approximately 580 K) are applied. For magnetic perovskites, it is necessary to consider different magnetic orderings, and the calculations are performed with a $2 \times 2 \times 2$ supercell consisting of 40 atoms and $4 \times 4 \times 3$ Gamma k-point mesh. The electronic structures are calculated by hybrid DFT in the framework of HSE[60] based on the relaxed structures. AIMD simulations are performed at room temperature by using the Nosé-Hoover method[79,80]. The polarization switching pathway is determined using the NEB method[81]. The Born effective charges are calculated using density functional perturbation theory[82]. It should be noted that the Born effective charges are calculated under the adiabatic approximation. In particular, the polar metals investigated in our work are fundamentally different from traditional metals (see Fig. S11 for details). In polar metals, the electrons around atoms are localized, which leads to a large electron linear response under perturbation, resulting in finite Born effective charges. For the investigated systems in this work, there are several atoms with large atomic numbers, such as Eu or Pb. To investigate if the spin-orbit coupling (SOC) effect is necessary, we select $EuPbO_3$ as a representative candidate and calculate its geometrical structure and band structure, considering the SOC effect. Table S7 compares the lattice parameters, and Fig. S5 compares the band structure of $EuPbO_3$ obtained by the PBE and PBE + SOC methods. It can be seen that consideration of the SOC effect has a slight effect on the geometrical structure and electronic structure. Therefore, the SOC effect is not considered in this work.

Machine learning model construction

The Random Forest classifier algorithm[83] is adopted to build the ML models, and the hyperparameters are: max depth = 4, random state = 0, n jobs = 2. Besides, the Extra Trees classifier algorithm[84] is also adopted to build the ML models, and the hyperparameters are: n estimators = 10, max depth = None, min samples split = 2, random state = 0.

In this work, we handle a binary classification task, so the confusion matrix is a 2×2 matrix that reports the number of *true positives* (*TP*), *true negatives* (*TN*), *false positives* (*FP*), and *false negatives* (*FN*) as follows[85,86]:

$$\begin{bmatrix} TN & FN \\ FP & TP \end{bmatrix} \tag{1}$$

The performance indicator can be used to compare different classifiers. We usually take the model with the best performance and the common indicators are as follows:

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{2}$$

$$precision = \frac{TP}{TP + FP} \tag{3}$$

$$recall = \frac{TP}{TP + FN} \tag{4}$$

Formation energy

The formation energy of ABO_3 perovskites, ($E_f^{ABO_3}$) is calculated according to Equation:

$$E_f^{ABO_3} = E_t^{ABO_3} - \mu_A - \mu_B - 3\mu_O \tag{5}$$

where $E_t^{ABO_3}$ is the total energy of the perovskites, $\mu_{A/B}$ is the chemical potential of element A/B in its bulk phase, and μ_O is 1/2 chemical potential in oxygen gas phase[87,88].

Spontaneous polarization

An excellent ferroelectric material should have a large spontaneous polarization. The polarization can be estimated by:

$$\Delta P_\alpha \cong \sum\nolimits_{j\beta}\left(\partial P_\alpha/\partial\mu_{j\beta}\right)\left(\mu_{j\beta} - \mu_{0j\beta}\right) = \frac{e}{\Omega}\sum\nolimits_{j\beta} Z^*_{j\beta}\Delta\mu_{j\beta} \tag{6}$$

where $\Delta\mu_{j\beta}$ is the displacement of ions j in the Cartesian direction β, $Z^*_{j\alpha\beta}$ is the Born effective charge tensor, e is the charge of an electron, and Ω is the cell volume[89,90].

Data availability

The source data and codes used in this study are available at the https://github.com/Winger0203/ML_ABX3_ST_perovskites.

Received: 11 December 2024; Accepted: 13 January 2026;
Published online: 28 January 2026

References

1. Xu, Y. *Ferroelectric materials and their applications*. (Elsevier, 2013).
2. Martin, L. W. & Rappe, A. M. Thin-film ferroelectric materials and their applications. *Nat. Rev. Mater.* **2**, 1–14 (2016).
3. Mikolajick, T. et al. Next generation ferroelectric materials for semiconductor process integration and their applications. *J. Appl. Phys.* **129**, 100901 (2021).
4. Yang, Y. et al. An ultrathin ferroelectric perovskite oxide layer for high-performance hole transport material free carbon based halide perovskite solar cells. *Adv. Funct. Mater.* **29**, 1806506 (2019).
5. Zhang, C. C. et al. Polarized ferroelectric polymers for high-performance perovskite solar cells. *Adv. Mater.* **31**, 1902222 (2019).
6. Han, X., Ji, Y. & Yang, Y. Ferroelectric photovoltaic materials and devices. *Adv. Funct. Mater.* **32**, 2109625 (2022).
7. Rabe, K. M., Ahn, C. H. & Triscone, J.-M. *Physics of ferroelectrics: a modern perspective*. Vol. 105 (Springer Science & Business Media, 2007).
8. Zhang, L. et al. Giant polarization in super-tetragonal thin films through interphase strain. *Science* **361**, 494–497 (2018).
9. Zhang, L. et al. Controllable ferromagnetism in super-tetragonal $PbTiO_3$ through strain engineering. *Nano Lett.* **20**, 881–886 (2019).
10. Li, M. et al. Microstructure-dependent thermal stability of super-tetragonal nanocomposite films through in situ TEM/EELS study. *ACS Appl. Mater. Interfaces* **14**, 52316–52323 (2022).
11. Lachheb, M. et al. Surface and bulk ferroelectric phase transition in super-tetragonal $BiFeO_3$ thin films. *Phys. Rev. Mater.* **5**, 024410 (2021).
12. Fan, Z. et al. Stable ferroelectric perovskite structure with giant axial ratio and polarization in epitaxial $BiFe_{0.6}Ga_{0.4}O_3$ thin films. *ACS Appl. Mater. Interfaces* **7**, 2648–2653 (2015).
13. Shpanchenko, R. V. et al. Synthesis, structure, and properties of new perovskite $PbVO_3$. *Chem. Mat.* **16**, 3267–3273 (2004).
14. Belik, A. A., Azuma, M., Saito, T., Shimakawa, Y. & Takano, M. Crystallographic features and tetragonal phase stability of $PbVO_3$, a new member of $PbTiO_3$ family. *Chem. Mat.* **17**, 269–273 (2005).
15. Belik, A. A. et al. Neutron powder diffraction study on the crystal and magnetic structures of $BiCoO_3$. *Chem. Mat.* **18**, 798–803 (2006).
16. Tinte, S., Rabe, K. M. & Vanderbilt, D. Anomalous enhancement of tetragonality in $PbTiO_3$ induced by negative pressure. *Phys. Rev. B* **68**, 144105 (2003).
17. Damodaran, A. R., Breckenfeld, E., Chen, Z., Lee, S. & Martin, L. W. Enhancement of ferroelectric Curie temperature in $BaTiO_3$ films via strain-induced defect dipole alignment. *Adv. Mater.* **26**, 6341–6347 (2014).
18. Wang, J. et al. Negative-pressure-induced enhancement in a freestanding ferroelectric. *Nat. Mater.* **14**, 985–990 (2015).
19. Kvasov, A. et al. Piezoelectric enhancement under negative pressure. *Nat. Commun.* **7**, 12136 (2016).
20. Singh, A., Singh, V. N., Canadell, E., Íñiguez, J. & Diéguez, O. Polymorphism in Bi-based perovskite oxides: a first-principles study. *Phys. Rev. Mater.* **2**, 104417 (2018).
21. Le, T., Epa, V. C., Burden, F. R. & Winkler, D. A. Quantitative structure–property relationship modeling of diverse materials properties. *Chem. Rev.* **112**, 2889–2919 (2012).
22. Tabor, D. P. et al. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.* **3**, 5–20 (2018).
23. Balachandran, P. V., Kowalski, B., Sehirlioglu, A. & Lookman, T. Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning. *Nat. Commun.* **9**, 1668 (2018).
24. Hu, W., Zhang, L. & Pan, Z. Designing two-dimensional halide perovskites based on high-throughput calculations and machine learning. *ACS Appl. Mater. Interfaces* **14**, 21596–21604 (2022).
25. Xu, G. et al. Machine learning-accelerated discovery and design of electrode materials and electrolytes for lithium ion batteries. *Energy Storage Mater.* **72**, 103710 (2024).
26. Chen, J. et al. Accelerating discovery of next-generation power electronics materials via high-throughput ab initio screening. *npj Comput. Mater.* **11**, 249 (2025).
27. Yin, P. et al. Machine-learning-accelerated design of high-performance platinum intermetallic nanoparticle fuel cell catalysts. *Nat. Commun.* **15**, 415 (2024).
28. Korolev, V., Mitrofanov, A., Eliseev, A. & Tkachenko, V. Machine-learning-assisted search for functional materials over extended chemical space. *Mater. Horiz.* **7**, 2710–2718 (2020).
29. Jain, A. et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
30. Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *JOM* **65**, 1501–1509 (2013).
31. Curtarolo, S. et al. AFLOWLIB. ORG: a distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235 (2012).

32. Ortiz, C., Eriksson, O. & Klintenberg, M. Data mining and accelerated electronic structure theory as a tool in the search for new functional materials. *Comput. Mater. Sci.* **44**, 1042–1049 (2009).
33. Landis, D. D. et al. The computational materials repository. *Comput. Sci. Eng.* **14**, 51–57 (2012).
34. Draxl, C. & Scheffler, M. NOMAD: the FAIR concept for big data-driven materials science. *MRS Bull.* **43**, 676–682 (2018).
35. Jalem, R., Nakayama, M. & Kasuga, T. An efficient rule-based screening approach for discovering fast lithium ion conductors using density functional theory and artificial neural networks. *J. Mater. Chem. A* **2**, 720–734 (2014).
36. Kaufmann, K. et al. Discovery of high-entropy ceramics via machine learning. *npj Comput. Mater.* **6**, 42 (2020).
37. Lu, S. et al. Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nat. Commun.* **9**, 3405 (2018).
38. Nasrabadi, N. M. Pattern recognition and machine learning. *J. Electron. Imaging* **16**, 049901 (2007).
39. Moure, C. & Peña, O. Recent advances in perovskites: processing and properties. *Prog. Solid State Chem.* **43**, 123–148 (2015).
40. Hwang, J. et al. Perovskites in catalysis and electrocatalysis. *Science* **358**, 751–756 (2017).
41. Horton, M. K. et al. Accelerated data-driven materials science with the Materials Project. *Nat. Mater*. **24**, 1–11 (2025).
42. Mu, X. et al. Polarization switching of HfO_2 ferroelectric in bulk and electrode/ferroelectric/electrode heterostructure. *npj Comput. Mater.* **11**, 126 (2025).
43. Xu, R. et al. Ferroelectric polarization reversal via successive ferroelastic transitions. *Nat. Mater.* **14**, 79–86 (2015).
44. Zhong, T., Li, X., Wu, M. & Liu, J.-M. Room-temperature multiferroicity and diversified magnetoelectric couplings in 2D materials. *Natl. Sci. Rev.* **7**, 373–380 (2020).
45. Ravindran, P., Vidya, R., Kjekshus, A., Fjellvåg, H. & Eriksson, O. Theoretical investigation of magnetoelectric behavior in $BiFeO_3$. *Phys. Rev. B* **74**, 224412 (2006).
46. Zhang, J. et al. Microscopic origin of the giant ferroelectric polarization in tetragonal-like $BiFeO_3$. *Phys. Rev. Lett.* **107**, 147602 (2011).
47. Chen, C. et al. Controllable defect driven symmetry change and domain structure evolution in $BiFeO_3$ with enhanced tetragonality. *Nanoscale* **11**, 8110–8118 (2019).
48. Tuttle, B., Payne, D. & Mukherjee, J. Ferroelectric materials for dielectric power conversion. *Ferroelectrics* **27**, 219–222 (1980).
49. Ali, H. E. H., Ricote, J., Calzada, M., Bretos, I. & Jiménez, R. The influence of the crystallization temperature on the reliability of $PbTiO_3$ thin films prepared by chemical solution deposition. *J. Eur. Ceram. Soc.* **37**, 1449–1458 (2017).
50. Uratani, Y., Shishidou, T., Ishii, F. & Oguchi, T. First-principles predictions of giant electric polarization. *Jpn. J. Appl. Phys.* **44**, 7130 (2005).
51. Abbassi, A., Zaari, H., Azahaf, C., Ez-Zahraouy, H. & Benyoussef, A. Spontaneous polarization and magnetic investigation of $BiXO_3$ (X= Co, Mn, Fe, V, Zn): first-principle study. *J. Supercond. Nov. Magn.* **29**, 487–491 (2016).
52. Ederer, C. & Spaldin, N. A. Effect of epitaxial strain on the spontaneous polarization of thin film ferroelectrics. *Phys. Rev. Lett.* **95**, 257601 (2005).
53. Qi, T., Grinberg, I. & Rappe, A. M. Correlations between tetragonality, polarization, and ionic displacement in $PbTiO_3$-derived ferroelectric perovskite solid solutions. *Phys. Rev. B* **82**, 134113 (2010).
54. Zhang, F. et al. Correlations between polarization and structural information of supertetragonal $PbTiO_3$. *Phys. Rev. B* **105**, 024106 (2022).
55. Tariq, A. & Nazir, S. Tunable ferroelectric polarization of the bulk and free standing $ATiO_3$ (A = Ba and Pb) thin films via unit-cell thicknesses and strain engineering. *AIP Adv*. **7**, https://doi.org/10.1063/1.5009038 (2017).
56. Savin, A., Nesper, R., Wengert, S. & Fässler, T. F. ELF: the electron localization function. *Angew. Chem. Int. Ed.* **36**, 1808–1832 (1997).
57. Liu, J. et al. Electronic structure and anion engineering for perovskite oxysulfide BaTi (O, S)$_3$. *J. Vac. Sci. Technol. A* **40**, 012801 (2022).
58. Li, M. et al. Supertetragonal $BaZrS_3$: a promising perovskite sulphide with giant ferroelectricity and low band gap. *Sci. China-Phys. Mech. Astron.* **67**, 1–8 (2024).
59. Xie, L. & Zhu, J. The electronic structures, born effective charges, and interatomic force constants in $BaMO_3$ (M= Ti, Zr, Hf, Sn): a comparative first-principles study. *J. Am. Ceram. Soc.* **95**, 3597–3604 (2012).
60. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **118**, 8207–8215 (2003).
61. Deák, P., Aradi, B., Frauenheim, T., Janzén, E. & Gali, A. Accurate defect levels obtained from the HSE06 range-separated hybrid functional. *Phys. Rev. B* **81**, 153203 (2010).
62. Deák, P., Aradi, B. & Frauenheim, T. Polaronic effects in TiO_2 calculated by the HSE06 hybrid functional: dopant passivation by carrier self-trapping. *Phys. Rev. B* **83**, 155207 (2011).
63. Matsuo, H., Noguchi, Y. & Miyayama, M. Gap-state engineering of visible-light-active ferroelectrics for photovoltaic applications. *Nat. Commun.* **8**, 207 (2017).
64. Yang, S. et al. Above-bandgap voltages from ferroelectric photovoltaic devices. *Nat. Nanotechnol.* **5**, 143–147 (2010).
65. Paillard, C. et al. Photovoltaics with ferroelectrics: current status and beyond. *Adv. Mater.* **28**, 5153–5168 (2016).
66. Bhowal, S. & Spaldin, N. A. Polar metals: principles and prospects. *Ann. Rev. Mater. Res.* **53**, 53–79 (2023).
67. Cao, Y. et al. Artificial two-dimensional polar metal at room temperature. *Nat. Commun.* **9**, 1547 (2018).
68. Zhang, J. et al. A correlated ferromagnetic polar metal by design. *Nat. Mater.* **23**, 1–8 (2024).
69. Oh, Y. S., Wang, L., Lee, H., Choi, W. S. & Kim, T. H. Polar perturbations in functional oxide heterostructures. *Adv. Funct. Mater.* **33**, 2302261 (2023).
70. Iczkowski, R. P. & Margrave, J. L. Electronegativity. *J. Am. Chem. Soc.* **83**, 3547–3551 (1961).
71. Lang, P. F. & Smith, B. C. Ionization energies of atoms and atomic ions. *J. Chem. Educ.* **80**, 938 (2003).
72. Rahm, M., Hoffmann, R. & Ashcroft, N. Atomic and ionic radii of elements 1–96. *Chem. Eur. J.* **22**, 14625–14632 (2016).
73. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758 (1999).
74. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
75. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169 (1996).
76. Blochl, P. E. Projector augmented-wave method. *Phys. Rev. B Condens. Matter* **50**, 17953–17979 (1994).
77. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
78. White, J. & Bird, D. Implementation of gradient-corrected exchange-correlation potentials in Car-Parrinello total-energy calculations. *Phys. Rev. B* **50**, 4954 (1994).
79. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
80. Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695 (1985).
81. Sheppard, D., Xiao, P., Chemelewski, W., Johnson, D. D. & Henkelman, G. A generalized solid-state nudged elastic band method. *J. Chem. Phys*. **136**, 074103 (2012).
82. Baroni, S., De Gironcoli, S., Dal Corso, A. & Giannozzi, P. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.* **73**, 515 (2001).
83. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **26**, 217–222 (2005).

84. Sharaff, A. & Gupta, H. in *Advances in Computer Communication and Computational Sciences: Proc. IC4S*. 189–197 (Springer, 2017).
85. Caelen, O. A Bayesian interpretation of the confusion matrix. *Ann. Math. Artif. Intell.* **81**, 429–450 (2017).
86. Deng, X., Liu, Q., Deng, Y. & Mahadevan, S. An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Inf. Sci.* **340**, 250–261 (2016).
87. Emery, A. A. & Wolverton, C. High-throughput DFT calculations of formation energy, stability and oxygen vacancy formation energy of ABO_3 perovskites. *Sci. Data* **4**, 1–10 (2017).
88. Guo, X. et al. Tackling the activity and selectivity challenges of electrocatalysts toward the nitrogen reduction reaction via atomically dispersed biatom catalysts. *J. Am. Chem. Soc.* **142**, 5709–5721 (2020).
89. Neaton, J., Ederer, C., Waghmare, U., Spaldin, N. & Rabe, K. First-principles study of spontaneous polarization in multiferroic $BiFeO_3$. *Phys. Rev. B* **71**, 014113 (2005).
90. Zhang, S. et al. Band-gap reduction in $(BiCrO_3)_m/(BiFeO_3)_n$ superlattices: designing low-band-gap ferroelectrics. *Phys. Rev. Appl.* **10**, 044004 (2018).

Acknowledgements

H.Y. Xiao was supported by the Joint Funds of the National Natural Science Foundation of China (Grant No. U1930120). L. Qiao acknowledges the support by National Natural Science Foundation of China (Grant No. 52072059). The theoretical calculations are performed using the supercomputer resources at TianHe-1 located at National Supercomputer Center in Tianjin.

Author contributions

H.X. conceptualized the project. H.X. and L.Q. supervised the project. W.H. conducted the DFT calculations, data set preparation, and training of the machine learning model. W.H., Z.W., M.L., S.F., H.Q., X.L., and X.Z. analyzed the data and interpreted the results. W.H. drafted the manuscript. All authors reviewed and edited the manuscript and have given approval to 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/s41524-026-01970-w.

Correspondence and requests for materials should be addressed to Haiyan Xiao or Liang Qiao.

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