PAPER

Check for updates

Machine learning-assisted design of AlN-based high-performance piezoelectric materials†

Huirong Jing, Chaohong Guan, Yu Yang and Hong Zhu ⓘ *

Dopants play an important role in improving the piezoelectric stress coefficient (e_{33}) of aluminum nitride (AlN)-based piezoelectric materials. However, the existing experimental or computational approaches cannot provide generalized design criteria or fast predictive capabilities for screening high-performance piezoelectric materials over a wide range of composition space. To address this demand, we have designed a general machine learning (ML) strategy to make a comprehensive prediction and exploration of AlN-based piezoelectric materials of various concentrations and compositions. The predicted piezoelectric strain coefficient (d_{33}) was verified to be remarkably consistent with the experimentally available values of Sc-, MgTi-, and MgZr-doped AlN compounds. It is worth noting that an extremely large d_{33} of 202 pC N^{-1} was discovered in $Sc_{0.5}Al_{0.5}N$. Besides, the first ionization energy, the formation energy of decomposition products, and the number of out-of-plane first-nearest-neighbor cation bonds were revealed to be critical physical quantities to facilitate the prediction of the piezoelectric coefficient based on a detailed investigation of the physical mechanism. This study demonstrates the feasibility of the fast prediction and design of high-performance piezoelectric materials with easily accessible features.

1 Introduction

Piezoelectric materials are capable of converting mechanical energy into electric energy and *vice versa* and have great potential for use in many modern applications, such as the health monitoring of nuclear power plants,[1,2] filtering of mobile communication frequency bands,[3,4] and touch-sensitive switches.[5,6] Many of these applications require operation over a wide temperature range. It is known that the piezoelectric coefficient is inversely correlated with the Curie temperature; a high Curie temperature generally corresponds to a low piezoelectric coefficient.[7] Therefore, it is urgent to find or design better piezoelectric materials with high Curie temperatures.

Aluminum nitride (AlN) has attracted considerable attention owing to its high Curie temperature of 1150 °C, high stiffness, high sound speed, high-temperature stability, and good compatibility with complementary metal oxide semiconductors (CMOS), which make it a good candidate to be utilized as a piezoelectric material in film bulk acoustic resonators

(FBAR).[8–12] However, the intrinsic poor piezoelectric stress coefficient (e_{33}) leads to a limited bandwidth, which is detrimental in high-frequency communications. In the past decade, the optimizations of the piezoelectric properties of AlN have been investigated experimentally through alloying with transition metals, such as Sc,[13,14] MgTi,[15,16] and MgZr.[17] Compared to pure AlN, it is reported that transition metal alloying can increase the piezoelectric coefficient by up to 4 times.[13] However, the traditional trial-and-error approach with high experimental costs makes the research and design of high-performance piezoelectric materials difficult. In recent years, several research groups have used first-principles calculations to search for high-performance AlN-based piezoelectric materials and have proposed some new alloying elements, such as Y, Cr, Yb, CaTi, and LiV.[18–22] These results broaden the existing composition regime of AlN-based piezoelectric materials. Another study also attracted our attention, in which Daoust *et al.* found that the alloy configuration has a significant effect on the piezoelectric properties of $ScAlN_2$ alloys,[23] implying the possibility to tune the piezoelectric properties by modulating the atomic configuration. However, these findings are more sporadic and unsystematic. The discovery of high-performance AlN-based piezoelectric materials from huge compositional and configurational spaces is critical but challenging based on traditional experimental or computational approaches.

Machine learning (ML) as a state-of-the-art method has been used to predict many properties of materials,[24–27] and its application towards piezoelectric properties was just initiated in recent years and needs to be accelerated in the near future.[28–31]

University of Michigan – Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China. E-mail: hong.zhu@sjtu.edu.cn

† Electronic supplementary information (ESI) available: Detailed feature descriptions and Pearson correlation matrix among 30 original features. Online GitHub repository with the interactive Jupyter notebook files and data used in developing the ML models are freely available at https://github.com/jinghuirong/Data-and-codes-for-ML-assisted-design-of-high-preformance-materials in the form of spreadsheets. See DOI: https://doi.org/10.1039/d3ta02095a

For instance, a surrogate model was built by Yuan *et al.*[28] to determine the piezoelectric coefficient of $BaTiO_3$-based materials. However, their regression results showed a large deviation around the high piezoelectric values. Based on previous work, He *et al.*[29] enlarged the databases and refined the features, and the predicted piezoelectric coefficient of $BaTiO_3$-based materials showed good consistency with experimental values. Moreover, Choudhary *et al.*[30] calculated the piezoelectric coefficients of non-metallic materials in the JARVIS-DFT database and found that it depends more strongly on the structural descriptors than the chemical descriptors, although they did not obtain a well-developed regression model. Thus, it is highly desirable to establish accurate machine learning models that can predict the piezoelectric properties of a rich set of AlN-based materials with different configurations, while a clear understanding of the underlying design principles from the physical aspects is also necessary.

In our work, high-throughput computations along with the machine learning strategy are used to predict the e_{33} of AlN-based alloys. The whole investigation process can be divided into four steps. First, 260 initial data sets with $2 \times 2 \times 1$ supercells of doped AlN were constructed based on density functional theory calculation. Next, structural features and chemical features were generated for the accurate prediction of piezoelectric properties, where structural features were employed to distinguish distinct configurations for the same chemical formula. Third, six classical machine learning models were built to predict e_{33} and the gradient boosting regression exhibited the best performance as measured by the mean absolute error and coefficient of determination for the test set. Finally, the well-developed model was used to predict the e_{33} of thirteen AlN-based systems in the whole databases within $2 \times 2 \times 2$ supercells, three systems (Sc-, MgTi-, MgZr-) were further calculated by density functional theory (DFT) and the most stable configuration in each of them is identified and verified with experimental results. Our studies open avenues for the fast prediction of e_{33} for AlN-based piezoelectric materials and put forward that governing the structural and chemical properties can effectively guide the design of high-performance piezoelectric materials.

2   Methods

2.1   First-principles calculations

First-principles calculations based on DFT were performed with the Vienna *Ab initio* Simulation Package (VASP), using the projector augmented-wave (PAW) method.[32,33] The wave functions were expanded in a plane-wave basis set with a kinetic energy cutoff of 520 eV. The exchange–correlation interaction was treated within the generalized gradient approximation (GGA) in the Perdew–Burke–Ernzerhof (PBE) format.[34] In the DFT calculations, the Γ-centered k-point meshes with a grid density of 5000 were adopted. All structures were fully relaxed until the total energy was less than 10^{-6} eV and the force on each atom was no more than 0.01 eV Å^{-1}. All computations and data analyses were performed using the Materials Project high-throughput software: Pymatgen,[35] FireWorks,[36] and MPWorks

(available at **http://www.github.com/materialsproject**). The piezoelectric stress coefficients were calculated by using the preset workflow based on density functional perturbation theory (DFPT), while the elastic tensors were computed by the workflow that fits the strain–stress relationship by using six independent components of Green-Lagrange strain tensor with four magnitudes ($\pm 0.5\%$, $\pm 1\%$) for the original structure.[37,38]

2.2   Data set for AlN-based materials

In a $2 \times 2 \times 1$ supercell, 260 AlN-based materials with the formula $A_xB_yAl_{1-x-y}N$ were considered, where A and B include Li^+, Na^+, Be^{2+}, Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}, Sc^{3+}, Y^{3+}, B^{3+}, Ga^{3+}, Ti^{4+}, Zr^{4+}, Hf^{4+}, Si^{4+}, V^{5+}, Nb^{5+}, Ta^{5+} and $x + y$ is in the range of $[0, 0.5]$ and the average valence state of A_xB_y is 3+. To reduce the randomness of the doping element designation in the compounds and make the subsequent regression more reliable, for the quaternary compound, A and B correspond to doping elements with smaller and larger ionic radii, respectively, while for the ternary compound, there is only one doping element denoted by A, and the B is assumed to be empty.

Structural and chemical features have been extracted for ML. The chemical features utilized in this work were generated by the Matminer python package,[39] including ionic radius, electronegativity difference, formation energy, and some others. Concurrently, some structural features were defined based on the wurtzite AlN crystal structure (Fig. 2b). To distinguish different crystal structures, configuration-related information, such as the uniformity of the distribution of doping elements and the number of nearest cation pairs, was extracted. In general, the predicted performance is more accurate after structural optimization.[40] However, structural features in this work involve the number of bonds and atomic distribution, rather than the exact bond lengths. To build a fast prediction model of the piezoelectric coefficient, we extracted all features from the established doping model without DFT geometry optimizations. Detailed feature descriptions are given in the ESI material.†

2.3   Machine learning

The ML part of this work was done using the scikit-learn Python library.[41] Least absolute shrinkage and selection operator (Lasso), AdaBoost regressor (AdaBoost), support vector regressor (SVR), random forest regressor (RFR), eXtreme gradient boosting regressor (XGBoost) and gradient boosting regressor (GBR) algorithms were selected to predict e_{33} of AlN-based materials. The best-performing model was selected by considering its performance on the test set. Before applying various supervised learning models, standardization of the input data is an important function of data preprocessing. In our case, we standardized our data to conform to the criterion of standard normal distribution. Our original data was divided into a training set and a test set with a ratio of $(8:2)$. Hyperparameter searches were carried out using the GridSearchCV algorithm and evaluated using 10-fold cross-validation error rate estimators. The evaluation of the prediction accuracy of the

regression model was characterized using the coefficient of determination (R^2) and mean absolute error (MAE). R^2 indicates how well the ML model predicts the target data, whereas MAE designates the average magnitude of the error. Each value displayed in this work is the average of R^2/MAE, by dividing the data set 50 times.

3 Results and discussion

Fig. 1 illustrates the flowchart for predicting the piezoelectric properties of AlN-based compounds through active learning. First, piezoelectric data for 260 ternary and quaternary AlN-based alloys were obtained from high-throughput DFT calculations. Considering the computation costs and the diversity of structure configuration, only the structures with the $2 \times 2 \times 1$ supercell were included in the initial data set. Then, the chemical and structural features were extracted to capture the underlying mechanism for high piezoelectric properties. After establishing several regression models, the best-performing one was selected for further optimization. Finally, using the well-established model, the e_{33} of thirteen AlN-based systems were predicted in the whole databases within $2 \times 2 \times 2$ supercells. Unlike the typical screening processes that first consider the stability and then other properties, we have initially focused on the piezoelectric properties for a wide variety of structures and then take stability into consideration. The first reason is that stability screening at the beginning will lead to a lack of configuration diversity, which is detrimental to the accuracy of the model prediction and the availability of design criteria. The second reason is that the experimental radio frequency magnetron reactive co-sputtering can break the stability limit and generate metastable phases.[41,42] Therefore, we decided to determine the most likely structure based on stability criteria after the development of the ML model for e_{33}, for which the predicted piezoelectric coefficients have been validated with the calculated and experimental values.

The distribution of the initial data set is shown in Fig. 2. The cation elemental symbol and the number of occurrences of each element in the initial data set are shown in Fig. 2a. An average valence of +3 for the cation dopants, A and B, was maintained for the charge neutral and semiconducting behavior of the doped AlN. Fig. 2c shows the distribution of e_{33}, of which a large number lies between 1.0 C m^{-2} to 2.5 C m^{-2}. As shown in Fig. 2d, the initial data set with doping concentrations ($x + y$) of 25% and 50% dominates.

To capture the underlying information on piezoelectric properties, the material features were divided into two categories. The first type of features are chemical properties, such as ionic radius, first ionization energy, and decomposition band gap, and were calculated by the sum of element properties or decomposition product properties times their corresponding molar ratios. Some of these features were also used in previous piezoelectric material research.[29] The second type is structural features (*e.g.* distribution of dopant ions, number of nearest-neighbor cation pairs), which were first proposed to distinguish different configurations with the same formula. For the features with a Pearson correlation coefficient larger than 0.9 or less

than −0.9, we selected the one that is more related to the e_{33} as the representative of the physical or chemical intuitions. The number of features thus decreased from 30 to 24, and the remaining features included five chemical features and nineteen structural features, as summarized in Table 1. The Pearson correlation matrix and a complete definition of features can be found in the ESI materials.†

Next, six classical machine learning models were built to train the piezoelectric data: (i) least absolute shrinkage and selection operator (Lasso), (ii) AdaBoost regressor (AdaBoost), (iii) support vector regressor (SVR), (iv) random forest regressor (RFR), (v) eXtreme gradient boosting regressor (XGBoost), (vi) gradient boosting regressor (GBR), as shown in Fig. 3a–f. The training set and test set of the scatter plot for individual models are shown with orange hollow circles and green-filled circles, respectively. The performance of each model was evaluated by MAE and R^2, shown in Fig. 3g and h. Among them, the GBR model performed the best, with MAE and R^2 values of 0.22/0.21, and 0.89/0.79 for the training/test set, respectively. Therefore, GBR was used as the preferred ML model in the subsequent exploration.

To further test the generalization of GBR models trained in AlN-based compounds with the $2 \times 2 \times 1$ supercell, the piezoelectric coefficients of 30 new data sets with the $2 \times 2 \times 2$ supercell, which is usually considered to be the smallest supercell that can simulate the actual disordered systems, were calculated based on the DFT calculations and compared with the GBR model predicted piezoelectric coefficient (see Fig. 4a). Although the e_{33} of the test set is distributed on both sides of $y = x$, it has a large dispersion as compared to the result in the $2 \times 2 \times 1$ supercell. The corresponding MAE is 0.59 C m^{-2} (Fig. 4c), which indicates that it is not appropriate for predicting the $2 \times 2 \times 2$ supercell directly with the model only trained on the $2 \times 2 \times 1$ supercell. Such a deviation might mainly come from the change in structural features. For example, in the $2 \times 2 \times 1$ supercell, the distance between cations is only up to the out-of-plane second-nearest-neighbor, the higher-order nearest-neighbor cation pairs exist in larger supercells. To alleviate the limitations caused by the low-dimensional structural features, we calculated 100 new structure data sets with the $2 \times 2 \times 2$ supercell, which include a variety of doping systems and have a wide range of piezoelectric values. We successively added i ($i = 0$–100 with an interval of 10) data from the new data set into the initial $2 \times 2 \times 1$ data set as a training set. To avoid contingency, we extracted i data 20 times and averaged the obtained MAE (see Fig. 4b and c). After adding 100 new data sets, the average MAE reached a stable value of about 0.27 C m^{-2}, which is close to the value in small supercells and the accuracy met our requirements. The parity plot in Fig. 4b shows good consistency between the DFT-calculated e_{33} and ML-predicted e_{33}. The application of this model in larger supercells is shown in Fig. S3.†

To gain physical insights from our model, the SHapley Additive exPlanation (SHAP) values derived from the optimized GBR model are shown in Fig. 5a, which were used to quantify the feature contributions. The features are ranked based on overall importance, and importance is defined as the mean
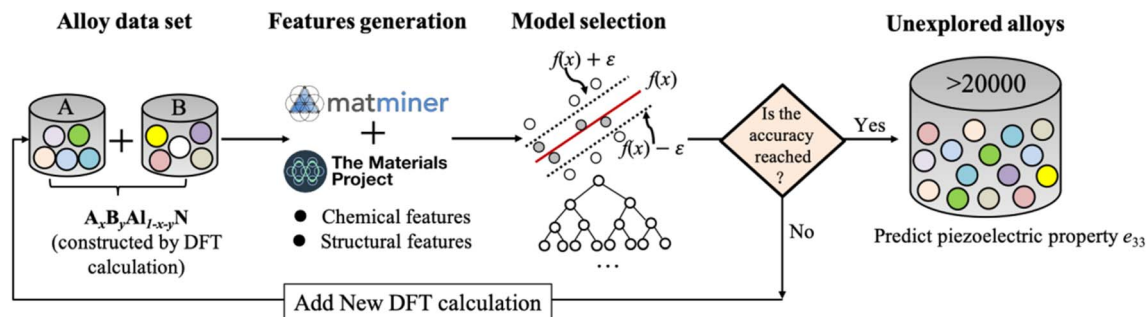
Fig. 1 The flowchart for predicting e_{33} for the target materials by combining DFT simulations and machine learning.

absolute SHAP value of all the points in the data set. In particular, a positive SHAP value indicates a forward contribution to the target, while a negative value represents an adverse contribution. A higher absolute value of SHAP indicates a stronger correlation. The colors in Fig. 5a denote the value of the feature variable, where red means a relatively large feature value and blue means a small value.[44] Here, we only displayed 15 features according to the feature importance. Notably, FIE is the most important feature among them. This is reasonable because the larger the FIE, the harder it is for atoms to lose electrons,
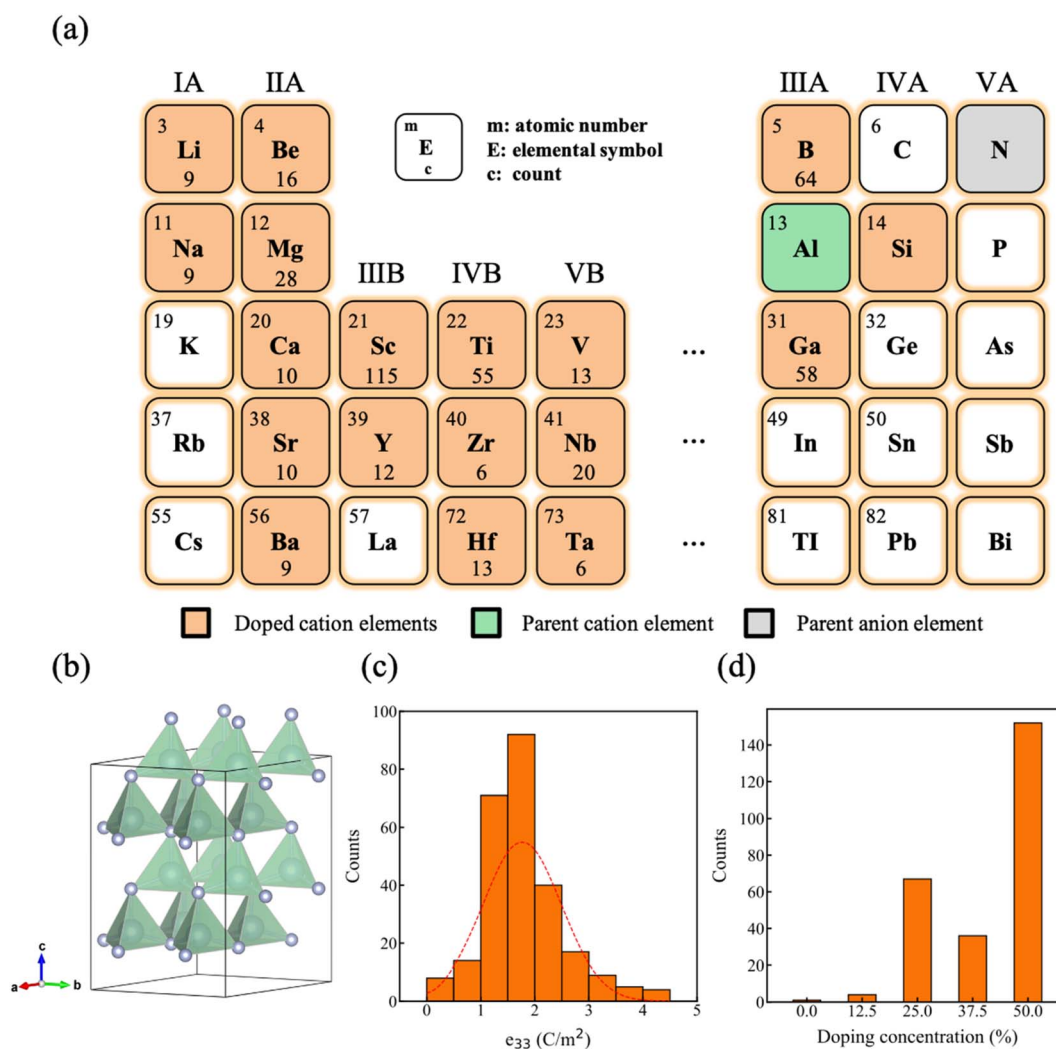

Fig. 2 Distribution of initial 260 AlN-based compounds. (a) The doped cations in the periodic table (colored in light orange). (b) The crystal structure of the wurtzite phase. Distribution of (c) the calculated piezoelectric coefficient e_{33} and (d) doping concentration ($x + y$) in the initial data set.

resulting in weaker bonding with N atoms. According to our previous studies,[45] weaker bonding will help to improve the piezoelectric coefficient of AlN. The next important feature is FED, which shows a negative correlation with the piezoelectric coefficient. We define FED as the average formation energy of the most stable decomposed products of AlN-based compounds. In previous work, Tholander *et al.*[21] investigated the piezoelectric properties of ternary and quaternary wurtzite nitride alloys and concluded that the larger energy difference between the tetrahedral phase and the layered hexagonal phases of the parent $A_{0.5}B_{0.5}N$ alloy will lead to a strong piezoelectric response. In our work, to decrease the computational cost, FED was used instead. Looking over the decomposed products, the majority of their structures are quite different from that of AlN (space group: $P6_3/mmc$). Thus, the lower the FED, the greater energy difference of $A_xB_yAl_{1-x-y}N$ will be between their parent stable phase and wurtzite phase, which will benefit the piezoelectric coefficient. To directly show the correlation between features and the piezoelectric coefficient, another symbolic regression algorithm, gplearn, was adopted to establish a basic mathematical expression that can accurately describe the relationship between the genetic programming (GP) processed features and the target values. During processing, the symbolic transformer can be used to reduce the feature dimension by automatically generating new combined features in chemical features and structural features, respectively.[46] Fig. 5b illustrates the two-dimensional projection of the piezoelectric coefficient as a function of the chemical features and structural features. The expressions obtained from chemical and structural features are as follows:

$$\text{Chemical factor}: \; -0.28(\text{FED} - \text{FIE}) + \frac{2.02 - \text{BGD}}{\text{FED} + 3.17} + 1.22, \quad (1)$$

$$\text{Structural factor}: \; \frac{N_o^{1st}(A-A)}{(N_o^{4th}(A-A))A_z + 3.48} + \frac{N_o^{1st}(B-B)}{B_{xy} + 3.48}$$
$$+ 0.23(N_o^{1st}(B-B)) + 1.86, \quad (2)$$

Consistent with the trend of feature importance in the SHAP plot, FIE, and FED are positively correlated and inversely correlated with the piezoelectric coefficient, respectively. $N_o^{1st}(A-A)$, and $N_o^{1st}(B-B)$, play a key role as structural features in enhancing the piezoelectric coefficient.

Fig. 5c and d display two typically different dopant distributions, namely out-of-plane first-nearest-neighbor bond ($1NN_o$) and in-plane first-nearest-neighbor bond ($1NN_i$). The $1NN_o$ bond shown in Fig. 5c is in the beveled edge of a tetrahedron and has a smaller angle to the *c*-axis as compared to the horizontal direction. In the tetrahedron containing $1NN_o$, two doped cations are on the upper and lower sides of the central nitrogen atom, respectively. In contrast, the $1NN_i$ bond is almost in the same plane and perpendicular to the *c*-axis. In the tetrahedron of Fig. 5d, the two doping cations that comprise the $1NN_i$ bond are above the central nitrogen atom. $N_o^{1st}(A-A)$, ($N_o^{1st}(B-B)$) measures the number of $1NN_o$, while $N_i^{1st}(A-A)$ ($N_i^{1st}(B-B)$) measures the number of $1NN_i$. In the crystal structure of Fig. 5c, $N_o^{1st}(A-A)$ ($N_o^{1st}(B-B)$) is equal to 2. A tetrahedron was used as an example for illustration. In this case, the upward force generated by three transversal bonds (two Al–N bonds and a dopant-N bond) will compensate for the downward force generated by the longitudinal bond (a dopant-N bond) on the nitrogen atom to a certain extent, leading to an almost equilibrium force ($F_1 \approx F_2$) on the nitrogen atom along the *c*-direction. If the bonding strength between the doping element and nitrogen is weak, as the quantity of $N_o^{1st}(A-A)$ ($N_o^{1st}(B-B)$) increases, the bonding energy around the nitrogen decays continuously and positively affects the increase in the piezoelectric coefficient, which has been demonstrated in our former studies.[45] However, as in Fig. 5d, if doped cation pairs are in the horizontal direction (here, it is 1), there will be a great discrepancy in the forces on both sides of the nitrogen along the *c*-direction, resulting in $F_1 \ll F_2$ or $F_1 \gg F_2$. At this point, a further increase of $N_i^{1st}(A-A)$ ($N_i^{1st}(B-B)$) will have little effect on the piezoelectric response. TF, obtained by the division of the average radius of the cation by the radius of the anion, is negatively correlated with the piezoelectric coefficient. We used the graphical diagram in Fig. 5e to explain this

Table 1 Chemical and structural features adopted in this work

| Types	Feature abbreviation	Description
Chemical features	FED	The average formation energy of decomposition products (*e.g.* AlN, ScN, Be_3N_2…)
	BGD	The average band gap of decomposition products
	FIE	The average first ionization energy of cations
	Nd	The average number of d orbital electron
	TF	Tetrahedral factor, the average value of the cation radius divided by the anion radius
Structural features	$N_{i/o}^{jth}(A-A)$, $N_{i/o}^{jth}(B-B)$, $N_{i/o}^{jth}(AB-AB)$	The number of in-plane/out-of-plane *j*th (*j*th=1st, 2nd, 3rd, 4th) nearest-neighbor A–A, B–B, A–B bonds
	A_{xy}, A_z, B_{xy}, B_z	Uniformity of the distribution of doped cations along *xyz* and *z* directions (A, B represent different cation elements)
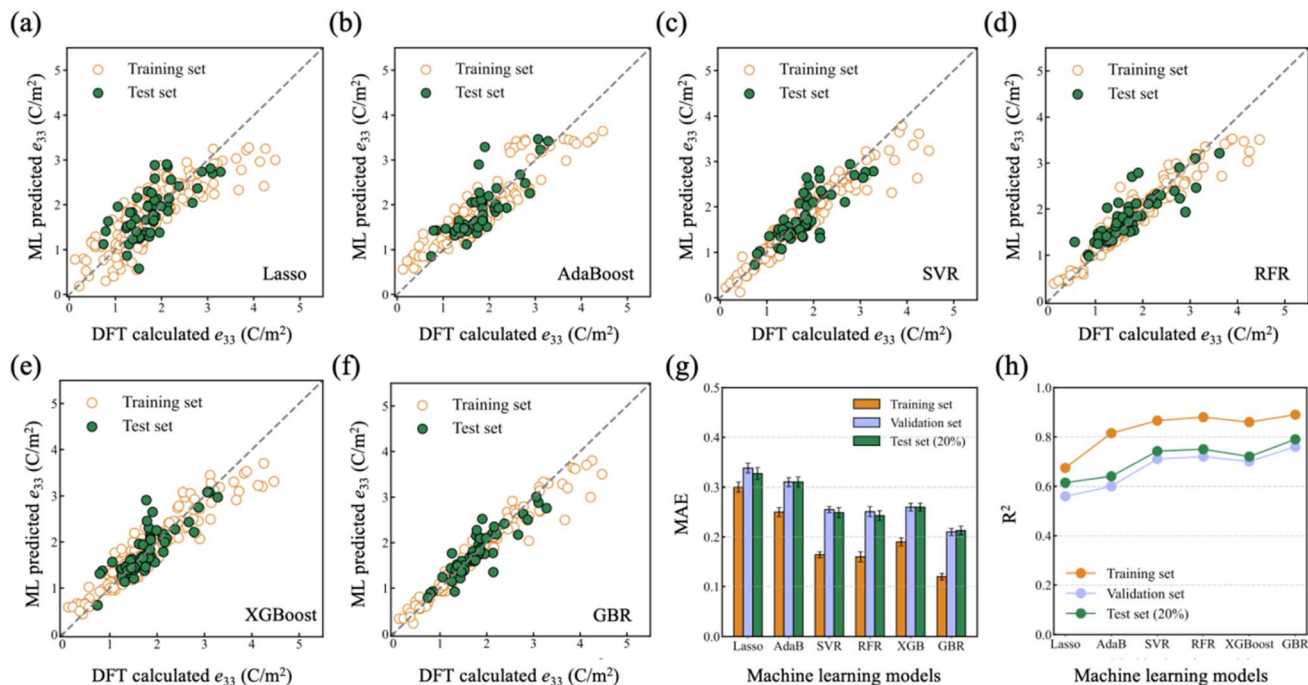
Fig. 3 Regression results of the machine learning algorithms in predicting the piezoelectric coefficient. Parity plots comparing the predicted e_{33} using (a) Lasso, (b) AdaBoost (c) SVR, (d) RFR, (e) XGBoost, and (f) GBR ML models, and that obtained from the $2 \times 2 \times 1$ supercell database (from DFT calculations). The gray dashed lines indicate the $y = x$ function. Comparison of (g) average MAE and (h) average R^2 for 50 random divisions of training, validation and test datasets among six models.

mechanism. According to the definition of TF, the larger the radius of the doped cation, the larger the TF. When doped with a cation with the same valence state as Al but a larger radius (larger TF), the nitrogen ion on the top of the dopant will be pushed upward, which will lead to an upward shift of the negative charge center. At the same time, three Al ions adjacent to the doping cation will also be pushed up due to the size effect of the doping cation, causing the positive charge center to move up. However, this upward shift compared to N is weak in the c-direction since Al is obliquely above and further away from the

doping cation. As a consequence, the distance between the positive and negative charge centers is shortened, leading to a decrease in polarization ($P_2 < P_1$) and thus a decrease in the piezoelectric coefficient.

We used the well-established GBR model for the quick prediction of e_{33} in the database of AlN-based materials and the results are displayed in Fig. 6a. The database was constructed using thirteen AlN-based systems with $2 \times 2 \times 2$ supercells, each system containing two doping concentrations (25%, 50%). After the removal of equivalent structures and existing
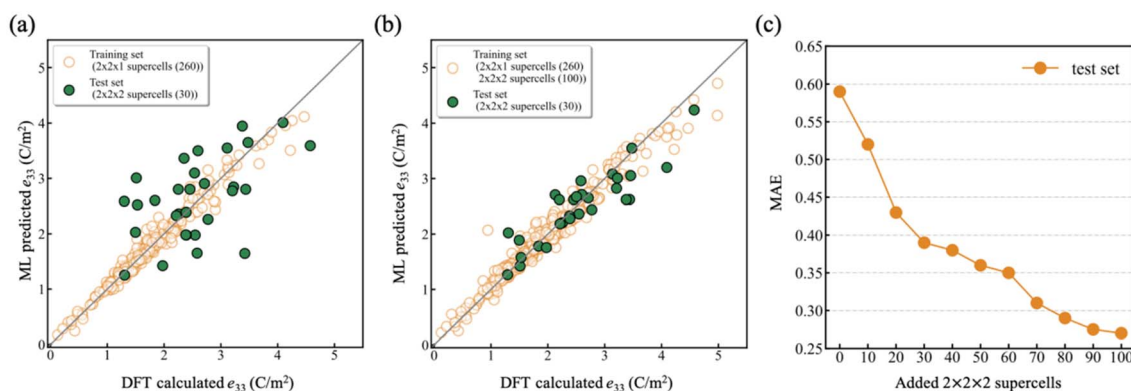

Fig. 4 Generalization of the GBR model. Parity plots comparing DFT-calculated e_{33} against ML-predicted e_{33} via GBR models with (a) 260 and (b) 360 training data, respectively, 30 new structures with the $2 \times 2 \times 2$ supercell as the test set are not included in the training dataset. (c) The MAE of the test set when 0–100 data sets with the $2 \times 2 \times 2$ supercell was added to the training dataset.
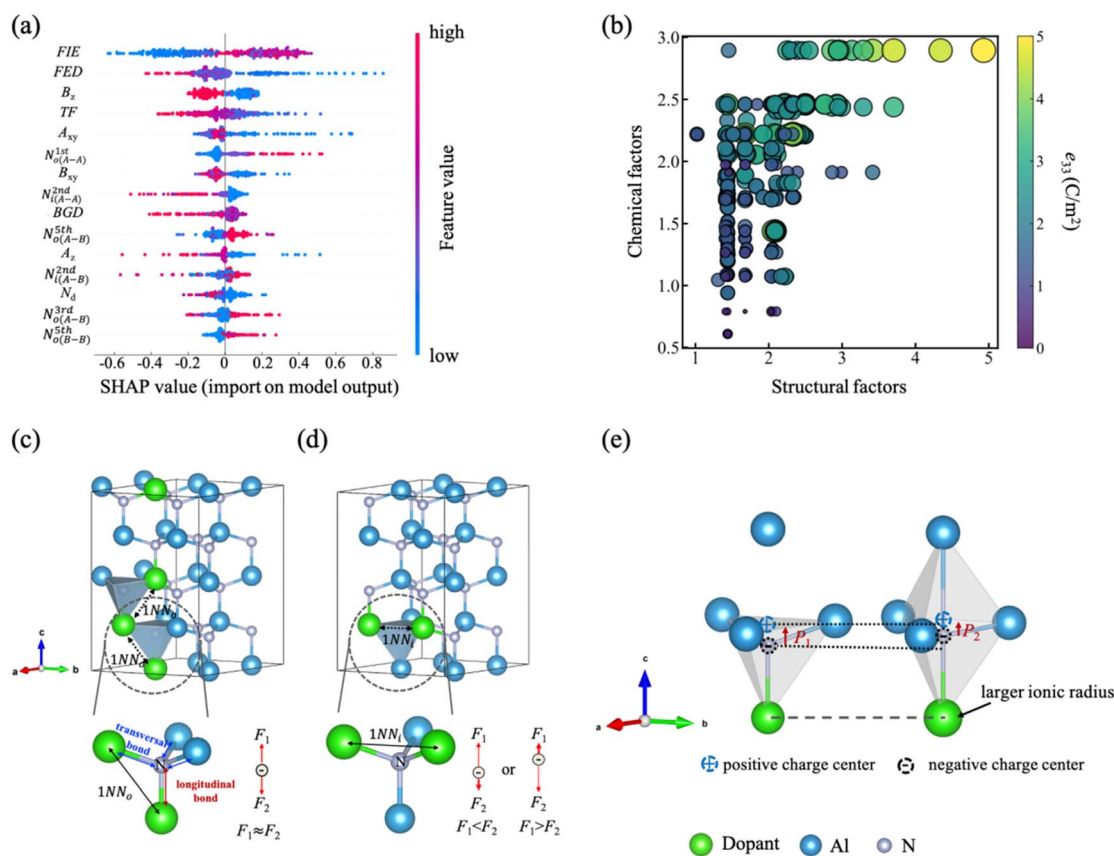
Fig. 5 Analysis of features. (a) SHAP plot summarizing the first 15 features for 390 points in the data set in order of increasing importance. (b) The two-dimensional scatter plot used to describe the relationship between the structural and chemical contributions and e_{33}. The color and size of the circle represent the value of the piezoelectric coefficient. Schematic diagram of the effect of the structural features (c) $N_o^{1st}(A-A)$ $(N_o^{1st}(B-B))$, (d) $N_i^{1st}(A-A)$ $(N_i^{1st}(B-B))$, and (e) chemical feature TF on the piezoelectric coefficient. $1NN_o$ and $1NN_i$ in (a) and (b) represent the out-of-plane first-nearest-neighbor bond and in-plane first-nearest-neighbor bond, respectively.
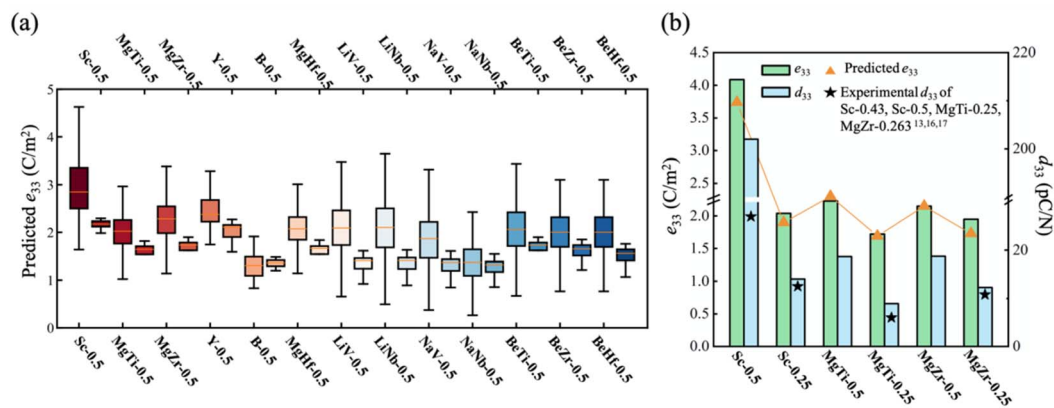

Fig. 6 Prediction and verification of the piezoelectric coefficient in the $2 \times 2 \times 2$ supercell. (a) Distribution of the piezoelectric values of the 13 systems in the $2 \times 2 \times 2$ supercell. (b) A comparison of piezoelectric properties in experiments, calculations, and predictions values.

structures in the training set, there were still about three thousand structures in each quaternary system and one hundred in each ternary system, resulting in more than 20 000 structures in the whole database. Fig. 6a shows the distribution of the predicted e_{33} for the specific chemical formula

considered. In line with our expectations, the distribution of e_{33} is wider for high doping concentrations, while relatively narrow for low doping concentrations. The main reason is that the configuration corresponding to high concentration is far greater than that of low concentration, leading to a large

possibility of an e_{33} value. Furthermore, we found that the e_{33} fluctuated slightly for the B-doped AlN system, implying that the influence of configuration on the $B_xAl_{1-x}N$ system is weak, which might arise from the strong B–N bond. In contrast, for $Li_{0.25}V_{0.25}Al_{0.5}N$ (LiV-0.5) and $Li_{0.25}Nb_{0.25}Al_{0.5}N$ (LiNb-0.5) systems, the e_{33} distribution was pretty wide, which indicates a strong configurational dependence.

Among the thirteen systems, Sc-, MgTi-, and MgZr- doped AlN alloys have been synthesized experimentally. To compare our predicted piezoelectric coefficient with the available experimental values, we first predicted and ranked the formation energy of structures in the system using the MegNet neural network algorithm,[47] then selected the most stable one to make further DFT calculations. The piezoelectric strain coefficient (d_{33}) was calculated by the formula

$$d_{ij} = \sum_{k=1}^{6} e_{ik} \left(C^{-1} \right)_{kj}, \quad (3)$$

where d_{ij}, e_{ij}, and C_{ij} are piezoelectric strain tensors, piezoelectric stress tensors, and elastic tensors, respectively. As shown in Fig. 6b, our predicted e_{33} of six structures are similar to the DFT calculated results. Furthermore, the calculated d_{33} for $Sc_{0.25}Al_{0.75}N$ and $Mg_{0.125}Zr_{0.125}Al_{0.75}N$ are almost the same as the experimental values, while the calculated $Mg_{0.125}Ti_{0.125}Al_{0.75}N$ is slightly larger than that of the experiment. More surprisingly, the calculated d_{33} of $Sc_{0.5}Al_{0.5}N$ is 202 (pC/N), which is nearly seven times the results of Akiyama *et al.*[13] ($Sc_{0.43}Al_{0.67}N$); this gives experimenters more confidence to achieve the high piezoelectric coefficient, though it should be noted that high concentration substitution may be challenging due to the limited solubility and may require special synthesis approaches.[43,48]

The success of the ML model in predicting the piezoelectric coefficient provides an opportunity to reduce time-consuming DFPT calculations by utilizing easily available chemical features and structural features. The strong correlation between chemical features (such as FIE and FED) and the piezoelectric coefficient suggests the great influence of doping elements. In addition, the fluctuation of the piezoelectric coefficient under the determined chemical formula implies configurational dependence. Hence, structural and chemical properties are both important for enhancing the piezoelectric performance, such as strong first ionization energy and a large number of out-of-plane first-nearest-neighbor cation pairs, which can be considered as criteria for the design of high-performance AlN-based piezoelectric materials. Furthermore, we believe that this method of training machine learning models is also applicable to other piezoelectric material systems. There are two reasons for this. First, the chemical features are only related to the chemical formula, which is generic for other systems. Second, the structural features are obtained without lattice optimization and are derived from the distribution of doping elements under the lattice parameters of the parent material, which make it easy to generalize. However, building a model that is applicable to a mixture of different systems needs further study in the future.

4 Conclusion

We have employed a high-throughput calculation combined with machine learning techniques to predict and design AlN-based piezoelectric properties. To capture the materials' information, structural features and chemical features were applied for the accurate prediction of piezoelectric properties. After the feature-importance analysis, we found that structural and chemical features play a synergistic role in the prediction of e_{33}, especially FID, FED, and $N_o^{1st}(A-A)$ ($N_o^{1st}(B-B)$). Based on the best-performing model, we predicted the piezoelectric distribution of thirteen AlN-based systems within the $2 \times 2 \times 2$ supercell, and a wide distribution of e_{33} can be observed in various systems. Additionally, we selected three systems for detailed validation, and the predicted e_{33} of six structures were similar to the DFT calculated results. Furthermore, the calculated d_{33} for $Sc_{0.25}Al_{0.75}N$ and $Mg_{0.125}Zr_{0.125}Al_{0.75}N$ were almost the same as the experimental values, while the calculated d_{33} of $Sc_{0.5}Al_{0.5}N$ was about seven times larger than the largest value in the experiment, which implies more space for experimentalists to achieve the higher piezoelectric coefficient by optimizing their fabrication process. The prediction effect will be further optimized by including more data in the future. Our studies not only establish a fast prediction tool for AlN-based piezoelectric materials with easily accessible features, they also provide insight into the underlying design concept of high-performance piezoelectric materials, which could be instructive for future experiments.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 S. Zhang and F. Yu, Piezoelectric materials for high temperature sensors, *J. Am. Ceram. Soc.*, 2011, **94**, 3153–3170.

2 M. J. Schulz, M. J. Sundaresan, J. Mcmichael, D. Clayton, R. Sadler and B. Nagel, Piezoelectric materials at elevated temperature, *J. Intell. Mater. Syst. Struct.*, 2003, **14**, 693–705.

3 P. Muralt, J. Antifakos, M. Cantoni, R. Lanz and F. Martin, Is there a better material for thin film BAW applications than A1N?, *IEEE Ultrason. Symp.*, 2005, **1**, 315–320.

4 C. Xiong, X. Sun, K. Y. Fong and H. X. Tang, Integrated high frequency aluminum nitride optomechanical resonators, *Appl. Phys. Lett.*, 2012, **100**, 171111.

5 D. B. Deutz, N. T. Mascarenhas, J. B. J. Schelen, D. M. de Leeuw, S. van der Zwaag and P. Groen, Flexible piezoelectric touch sensor by alignment of lead-free

alkaline niobate microcubes in PDMS, *Adv. Funct. Mater.*, 2017, **27**, 1700728.

6 H. Khanbareh, K. de Boom, B. Schelen, R. B. N. Scharff, C. C. L. Wang, S. van der Zwaag and P. Groen, Large area and flexible micro-porous piezoelectric materials for soft robotic skin, *Sens. Actuators, A*, 2017, **263**, 554–562.

7 S. Zhang, R. Xia, L. Lebrun, D. Anderson and T. R. Shrout, Piezoelectric materials for high power, high temperature applications, *Mater. Lett.*, 2005, **59**, 3471–3475.

8 N. Sinha, G. E. Wabiszewski, R. Mahameed, V. V. Felmetsger, S. M. Tanner, R. W. Carpick and G. Piazza, Piezoelectric aluminum nitride nanoelectromechanical actuators, *Appl. Phys. Lett.*, 2009, **95**, 053106.

9 M. A. Moram and S. Zhang, ScGaN and ScAlN: emerging nitride materials, *J. Mater. Chem. A*, 2014, **2**, 6042–6050.

10 R. C. Turner, P. A. Fuierer, R. E. Newnham and T. R. Shrout, Materials for high temperature acoustic and vibration sensors: A review, *Appl. Acoust.*, 1994, **41**, 299–324.

11 C. Fei, X. Liu, B. Zhu, D. Li, X. Yang, Y. Yang and Q. Zhou, AlN piezoelectric thin films for energy harvesting and acoustic devices, *Nano Energy*, 2018, **51**, 146–161.

12 H. Tian, J. Liu, B. Dong, J. C. Skehan, M. Zervas, T. J. Kippenberg and S. A. Bhave, Hybrid integrated photonics using bulk acoustic resonators, *Nat. Commun.*, 2020, **11**, 1–8.

13 M. Akiyama, T. Kamohara, K. Kano, A. Teshigahara, Y. Takeuchi and N. Kawahara, Enhancement of piezoelectric response in scandium aluminum nitride alloy thin films prepared by dual reactive cosputtering, *Adv. Mater.*, 2009, **21**, 593–596.

14 K. R. Talley, S. L. Millican, J. Mangum, S. Siol, C. B. Musgrave, B. Gorman, M. Holder, A. Zakutayev and G. L. Brennecka, Implications of heterostructural alloying for enhanced piezoelectric performance of (Al,Sc)N, *Phys. Rev. Mater.*, 2018, **2**, 063802.

15 S. A. Anggraini, M. Uehara, K. Hirata, H. Yamada and M. Akiyama, Effects of different divalent cations in mTi-based codopants (m= Mg or Zn) on the piezoelectric properties of AlN thin films, *Ceram. Int.*, 2020, **46**, 4015–4019.

16 B. Wang, K. Aryana, J. T. Gaskins, P. E. Hopkins, S. V. Khare and D. Gall, Structural Stabilization and Piezoelectric Enhancement in Epitaxial $(Ti_{1-x}Mg_x)_{0.25}Al_{0.75}N$ (0001) Layers, *Adv. Funct. Mater.*, 2020, **30**, 2001915.

17 T. Yokoyama, Y. Iwazaki, Y. Onda, T. Nishihara, Y. Sasajima and M. Ueda, Effect of Mg and Zr co-doping on piezoelectric AlN thin films for bulk acoustic wave resonators, *IEEE Trans. Ultrason. Ferroelectrics Freq. Control*, 2014, **61**, 1322–1328.

18 A. Assali, F. Laidoudi, R. Serhane, F. Kanouni and O. Mezilet, Highly enhanced electro-acoustic properties of YAlN/sapphire based surface acoustic wave devices for next generation of microelectromechanical systems, *Mater. Today Commun.*, 2021, **26**, 102067.

19 S. Manna, K. R. Talley, P. Gorai, J. Mangum, A. Zakutayev, G. L. Brennecka, V. Stevanović and C. V. Ciobanu, Enhanced piezoelectric response of AlN via CrN alloying, *Phys. Rev. Appl.*, 2018, **9**, 034026.

20 J. Jia and T. Yanagitani, Origin of enhanced electromechanical coupling in (Yb, Al)N nitride alloys, *Phys. Rev. Appl.*, 2021, **16**, 044009.

21 C. Tholander, F. Tasnádi, I. A. Abrikosov, L. Hultman, J. Birch and B. Alling, Large piezoelectric response of quaternary wurtzite nitride alloys and its physical origin from first principles, *Phys. Rev. B*, 2015, **92**, 174119.

22 M. Noor-A-Alam, O. Z. Olszewski, H. Campanella and M. Nolan, Large Piezoelectric Response and Ferroelectricity in Li and V/Nb/Ta Co-Doped w-AlN, *ACS Appl. Mater. Interfaces*, 2021, **13**, 944–954.

23 P. Daoust, P. Desjardins, R. A. Masut, V. Gosselin and M. Côté, Ab initio piezoelectric properties of $Al_{0.5}Sc_{0.5}N$: Impact of alloy configuration on the $d_{33,f}$ piezoelectric strain coefficient, *Phys. Rev. Mater.*, 2017, **1**, 055402.

24 G. S. Na, S. Jang and H. Chang, Predicting thermoelectric properties from chemical formula with explicitly identifying dopant effects, *npj Comput. Mater.*, 2021, **7**, 1–11.

25 R. Kumar and A. K. Singh, Chemical hardness-driven interpretable machine learning approach for rapid search of photocatalysts, *npj Comput. Mater.*, 2021, **7**, 1–13.

26 T. Wang, C. Zhang, H. Snoussi and G. Zhang, Machine learning approaches for thermoelectric materials research, *Adv. Funct. Mater.*, 2020, **30**, 1906041.

27 A. C. Rajan, A. Mishra, S. Satsangi, R. Vaish, H. Mizuseki, K.-R. Lee and A. K. Singh, Machine-learning-assisted accurate band gap predictions of functionalized MXene, *Chem. Mater.*, 2018, **30**, 4031–4038.

28 R. Yuan, Y. Tian, D. Xue, D. Xue, Y. Zhou, X. Ding, J. Sun and T. Lookman, Accelerated search for $BaTiO_3$-based ceramics with large energy storage at low fields using machine learning and experimental design, *Adv. Sci.*, 2019, **6**, 1901395.

29 J. He, *et al.*, Accelerated discovery of high-performance piezocatalyst in $BaTiO_3$-based ceramics via machine learning, *Nano Energy*, 2022, **97**, 107218.

30 K. Choudhary, K. F. Garrity, V. Sharma, A. J. Biacchi, A. R. Hight Walker and F. Tavazza, High-throughput density functional perturbation theory and machine learning predictions of infrared, piezoelectric, and dielectric responses, *npj Comput. Mater.*, 2020, **6**, 1–13.

31 D. Xue, P. V. Balachandran, R. Yuan, T. Hu, X. Qian, E. R. Dougherty and T. Lookman, Accelerated search for $BaTiO_3$-based piezoelectrics with vertical morphotropic phase boundary using Bayesian learning, *Proc. Natl. Acad. Sci.*, 2016, **113**, 13301–13306.

32 G. Kresse and J. Furthmiiller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, *Comput. Mater. Sci.*, 1996, **6**, 15–50.

33 G. Kresse and J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, *Phys. Rev. B*, 1996, **54**, 11169.

34 J. P. Perdew, K. Burke and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.*, 1997, **77**, 3865.

35 S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson and

G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, *Comput. Mater. Sci.*, 2013, **68**, 314–319.

36 A. Jain, S. P. Ong, W. Chen, B. Medasani, X. Qu, M. Kocher, M. Brafman, G. Petretto, G.-M. Rignanese, G. Hautier, D. Gunter and K. A. Persson, FireWorks: a dynamic workflow system designed for high-throughput applications, *Concurrency Comput. Pract. Ex.*, 2015, **27**, 5037–5059.

37 M. de Jong, *et al.*, Charting the complete elastic properties of inorganic crystalline compounds, *Sci. Data*, 2015, **2**, 150009.

38 X. Wu, D. Vanderbilt and D. R. Hamann, Systematic treatment of displacements, strains, and electric fields in density-functional perturbation theory, *Phys. Rev. B*, 2005, **72**, 035105.

39 L. Ward, *et al.*, Matminer: An open source toolkit for materials data mining, *Comput. Mater. Sci.*, 2018, **152**, 60–69.

40 L. Chen, Y. Tian, X. Hu, S. Yao, Z. Lu, S. Chen, X. Zhang and Z. Zhou, A universal machine learning framework for electrocatalyst innovation: a case study of discovering alloys for hydrogen evolution reaction, *Adv. Funct. Mater.*, 2022, **32**, 2208418.

41 F. Pedregosa, *et al.*, Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.

42 C. Höglund, J. Birch, B. Alling, J. Bareño, Z. Czigány, P. O. Å. Persson, G. Wingqvist, A. Zukauskaite and L. Hultman, *J. Appl. Phys.*, 2010, **107**, 123515.

43 A. M. Holder, *et al.*, Novel phase diagram behavior and materials design in heterostructural semiconductor alloys, *Sci. Adv.*, 2017, **3**, e1700270.

44 M. I. M. Kusdhany and S. M. Lyth, New insights into hydrogen uptake on porous carbon materials via explainable machine learning, *Carbon*, 2021, **179**, 190–201.

45 H. Jing, Y. Wang, Q. Wen, X. Cai, K. Liu, W. Li, L. Zhu, X. Li and H. Zhu, Large piezoelectric and elastic properties in B and Sc codoped wurtzite AlN, *J. Appl. Phys.*, 2022, **131**, 245108.

46 J. Zheng, X. Sun, J. Hu, S. B. Wang, Z. Yao, S. Deng, X. Pan, Z. Pan and J. Wang, Symbolic transformer accelerating machine learning screening of hydrogen and deuterium evolution reaction catalysts in MA_2Z_4 Materials, *ACS Appl. Mater. Interfaces*, 2021, **13**, 50878–50891.

47 C. Chen, W. Ye, Y. Zuo, C. Zheng and S. P. Ong, Graph networks as a universal machine learning framework for molecules and crystals, *Chem. Mater.*, 2019, **31**, 3564–3572.

48 C. Tholander, J. Birch, F. Tasnádi, L. Hultman, J. Palisaitis, P. O. Å. Persson, J. Jensen, P. Sandström, B. Alling and A. Žukauskaitė, Ab initio calculations and experimental study of piezoelectric $Y_xIn_{1-x}N$ thin films deposited using reactive magnetron sputter epitaxy, *Acta Mater.*, 2016, **105**, 199–206.