npj computational materials

www.nature.com/npjcompumats

ARTICLE OPEN

Check for updates

XGBoost model for electrocaloric temperature change prediction in ceramics

Jie Gong[1], Sharon Chu[1], Rohan K. Mehta[1] and Alan J. H. McGaughey[1]✉

An eXtreme Gradient Boosting (XGBoost) machine learning model is built to predict the electrocaloric (EC) temperature change of a ceramic based on its composition (encoded by Magpie elemental properties), dielectric constant, Curie temperature, and characterization conditions. A dataset of 97 EC ceramics is assembled from the experimental literature. By sampling data from clusters in the feature space, the model can achieve a coefficient of determination of 0.77 and a root mean square error of 0.38 K for the test data. Feature analysis shows that the model captures known physics for effective EC materials. The Magpie features help the model to distinguish between materials, with the elemental electronegativities and ionic charges identified as key features. The model is applied to 66 ferroelectrics whose EC performance has not been characterized. Lead-free candidates with a predicted EC temperature change above 2 K at room temperature and 100 kV/cm are identified.

npj Computational Materials (2022)8:140 ; https://doi.org/10.1038/s41524-022-00826-3

INTRODUCTION

The electrocaloric (EC) effect is the coupled temperature and entropy change of a dielectric material due to the polarization change that results from the application or removal of an electric field[1]. Electrocaloric cooling offers great potential to build efficient solid-state cooling devices that are quiet, low weight, and compact, making it a promising replacement for noisy, less-efficient vapor compression systems and a candidate for on-chip cooling and wearable cooling devices[2,3]. With the potential to reach 60–70% of the Carnot coefficient of performance[4], EC cooling devices are more efficient than thermoelectric devices[5]. The large electric fields required to operate EC cooling devices are easy to generate compared to the large magnetic fields required in magnetocaloric cooling devices[6].

The EC temperature change (ΔT_{EC}) is a function of material, characterization temperature, and applied electric field. Its magnitude is generally larger around phase transition temperatures and it increases with increasing applied electric field. The key to building an EC cooling device is to find a material with a large ΔT_{EC}, which drives the heat flow, over a wide temperature range near the operating conditions, which are usually around room temperature. The EC effect was first observed in the Rochelle salt in 1930[7] and was intensively studied in the 1960s and 1970s in bulk materials. The resulting ΔT_{EC}s were not practically useful (below 1 K), however, and research interests waned. In 2006, the field was revitalized by the discovery of a large EC temperature change in a $PbZr_{0.95}Ti_{0.05}O_3$ ceramic (12 K at a characterization condition of 499 K and 480 kV/cm)[8]. In 2008, the discovery of large EC temperature changes in PVDF-based polymers[9] (over 12 K at 353 K and 2090 kV/cm for a P(VDF–TrFE) copolymer and at 328 K and 3070 kV/cm for a P(VDF-TrFE-CFE) terpolymer] suggested a path towards economical and environment-friendly fabrication and integration in flexible systems. Significant progress in EC materials development[10–12] and EC device designs[13–16] has been achieved since then. In 2013, Peng et al.[17] reported a giant ΔT_{EC} of 45 K in a $Pb_{0.8}Ba_{0.2}ZrO_3$ thin film at 290 K and 598 kV/cm. In 2017, Ma et al.[18] built a EC cooling device with a flexible P(VDF–TrFE–CFE) polymer and electrostatic actuation that produced a specific cooling power of 2.8 W/g and a coefficient of performance of 13.

Identifying an effective EC material is a non-trivial task. The synthesis of new EC materials relies on the instincts of experts and extensive experimental synthesis of ceramics, polymers, and/or composite materials. Effort has been devoted to theoretical understanding of the EC effect by conducting ab initio simulations[19,20], classical simulations[21,22], and phenomenological theory-based (i.e., Landau–Ginzburg–Devonshire type theory) studies[23,24], but elucidating the physics of the EC effect remains challenging. One major reason is the complex nature of the EC effect, which involves polarization hysteresis, phase transition, sample crystallinity and crystallite size, and interactions between crystalline and amorphous regions[1,3,25].

To aid the search for effective EC materials, we herein apply a data-driven approach to build a machine learning (ML) model to predict the ΔT_{EC} for ceramics based on the material composition, dielectric constant, Curie temperature, and characterization conditions. While a large EC entropy change is equally important, we did not include it as a label because the isothermal entropy change can be directly calculated from the adiabatic temperature change, density, specific heat, and temperature[1].

Machine learning is the field of study where computer programs learn some class of tasks and improve by performance measures from examples or experience[26]. The application of ML methods in materials science dates back to the 1990s[27,28]. Although ML methods were initially used as assistance tools for tasks such as spectral analysis[29], biomolecule binding site prediction[30], and the derivation of quantitative structure–activity relationships[31], they have since become an essential part of the materials research portfolio[32]. Materials databases such as the Materials Project[33], the Inorganic Crystallographic Structure Database (ICSD)[34], Automatic Flow for Materials Discovery (AFLOW)[35], and the Open Quantum Materials Database (OQMD)[36] provide curated and reliable data that can be mined to identify the underlying physics of observed material properties and phenomena. Open access to ML packages and libraries (e.g.,

[1]Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA. ✉email: mcgaughey@cmu.edu

Scikit-learn[37], PyTorch[38], and Tensorflow[39]) facilitates the application of data-driven approaches to materials challenges.

Materials property prediction with data-driven methods typically uses the material compositions and/or basic material properties as input and the property (or properties) of interest as output. The learning problem is to find the best estimate of the property of a material not in the original dataset. The resulting well-trained ML model can reduce human effort that would otherwise be required to synthesize and characterize materials, accelerating the discovery of new materials and uncovering previously unknown correlations between properties[32].

The application of ML methods in the field of materials science is extensive, and we provide four examples that are related to the current study. To support magnetocaloric refrigeration applications, Holleis et al. [40] built neural network, random forest, least absolute shrinkage and selection operator, and support vector regression models to predict the entropy change of single-molecule magnets (SMMs) at a magnetic field of 5 T. They assembled a dataset of more than 60 experimentally synthesized SMMs and designed 16 descriptors that described their structure, dimensionality, and chemical features. They identified four key descriptors that have the largest impact on the entropy change: number of *d*-ions, number of *f*-ions, molecular mass, and ideal spin per ion. Their ML models predicted three hypothetical SMMs for future synthesis.

Stanev et al. used random forest and neural network models to predict the critical temperature of superconducting materials[41]. The features were a combination of Materials Agnostic Platform for Informatics and Exploration (Magpie)[42] features generated from the elemental composition of each material and AFLOW[35] features based on crystallographic and electronic information. The tree-based model reached a coefficient of determination (R^2 score) of 0.88 on the test data for critical temperatures >10 K. Their models offered insight into the mechanisms behind superconductivity in different families of materials and identified over 30 candidate materials from the ICSD. We will use tree-based models because of their interpretability and Magpie for feature generation due to its effectiveness in the absence of structural information[43].

The dielectric constant, a property related to the EC effect, can also be predicted via ML methods. Mannodi-Kanakkithodi et al. used the kernel ridge regression model to predict the dielectric constant of polymers[44]. They represented the polymer structure based on its building blocks and trained the model on density functional (DFT) calculations. Their model showed moderate transferability and can predict the dielectric constant of 6-block and 8-block polymers after training on only 4-block polymer data. They also used a genetic algorithm to optimize the blocks in an evolutionary manner to design polymers with desired dielectric properties.

Su et al.[45] recently developed support vector regression and random forest regression models to predict the ΔT_{EC} of $BaTiO_3$-based ceramics represented by chemical composition (i.e., elemental descriptors of the A-site and B-site elements), temperature, and applied electric field. Two separate regression-based ML models were developed for indirect and direct measurements. A classification model that predicts the expected phase as a function of chemical composition and temperature[46] was introduced to complement the regression models. The combined regression and classification ML models are able to predict a global maximum in ΔT_{EC} near the rhombohedral-to-cubic and tetragonal-to-cubic phase transitions.

Here, we build an eXtreme Gradient Boosting (XGBoost)[47] model to predict the ΔT_{EC} of a ceramic ferroelectric material given its composition, dielectric constant, Curie temperature, and characterization conditions. The dataset is assembled from available experimental ΔT_{EC} measurements. We include the measurement method (i.e., direct or indirect) and polarization stage (i.e., polarization or depolarization) as categorical variables to better describe the data. The XGBoost model is able to predict ΔT_{EC} with an R^2 score of 0.90 and 0.77 on train and test data, corresponding to root-mean-square errors (RMSEs) of 0.38 K for both. The model correctly identifies the known physics that contribute to a large EC temperature change (i.e., applied electric field and the difference between the characterization and Curie temperatures). We apply the model to search for effective EC materials from 66 ferroelectrics whose EC performance has not been characterized and suggest candidate materials for future experimental verification.

RESULTS

Materials dataset

There are three major categories of EC materials: polymers, ceramics, and polymer–ceramic composites. We built a dataset for EC ceramics due to their wide compositional variety. We extracted information from available literature as most of the material compositions do not appear in well-known materials databases [33–36]. The dataset consists of 97 materials from 45 papers and is available at GitHub[48] in a csv format. Snapshots from the dataset and a flow chart of the data gathering and model construction steps are shown in Fig. 1. More detailed information can be found in the "Methods—Materials dataset preparation" section and Supplementary Note 1.

We extract 7 features for each material: the material composition, temperature (T) and electric field (E) at which the ΔT_{EC} is measured, the phase transition temperature (T_{Curie}), the dielectric constant ϵ at T, and the measurement condition pol (for measurement during polarization or depolarization) and method (for direct or indirect measurements). We encode each material composition with the Magpie package[42]. Compared to direct encoding methods, encoding with Magpie keeps more chemical information from the materials by converting elemental chemical properties of the composition into 145 continuous or discrete numerical features. We then conducted feature selection by dropping features with zero variance, dropping features that have a Pearson correlation coefficient higher than 0.95 with an existing feature, and conducting a backward feature elimination process on these Magpie features. More detailed information on can be found in the "Methods—Feature selection" section and Supplementary Notes 2 and 3.

After these preprocessing steps, we have 4406 data points, each containing the 21 features listed in Table 1 (7 experimental condition/material property features and 14 Magpie features). The label to predict is the ΔT_{EC} at the given conditions (i.e., T and E).

The collected data are plotted as a function of characterization temperature for the full scale in Fig. 2a. Data points with ΔT_{EC} in the range of 0–2 K are plotted as a function of $T-T_{Curie}$ in Fig. 2b. Different colors represent different material compositions and the marker sizes are proportional to the applied electric field. Most of these EC materials have a relatively small temperature change, with a median of 0.36 K and a mean of 1.07 K. Three of the 97 materials have a maximum $\Delta T_{EC} > 30$ K [$Pb_{0.88}La_{0.08}Zr_{0.65}Ti_{0.35}O_3$[49], $Pb_{0.8}Ba_{0.2}ZrO_3$[17], and $Pb_3Mg_{0.65}Nb_{1.3}Ti_{1.05}O_9$[50]], which far exceeds the next largest maximum value of 13 K. These three materials are marked as outliers and are excluded when building the model unless otherwise specified.

Model performance

The XGBoost regression models (more details in the "Methods—XGBoost regression" section) for ΔT_{EC} prediction are built with the best hyperparameter set from a grid search of 6912 combinations (Table 2). Given that XGBoost is unable to extrapolate and can only make reasonable predictions for situations previously encountered in the training history, the materials with the lowest and

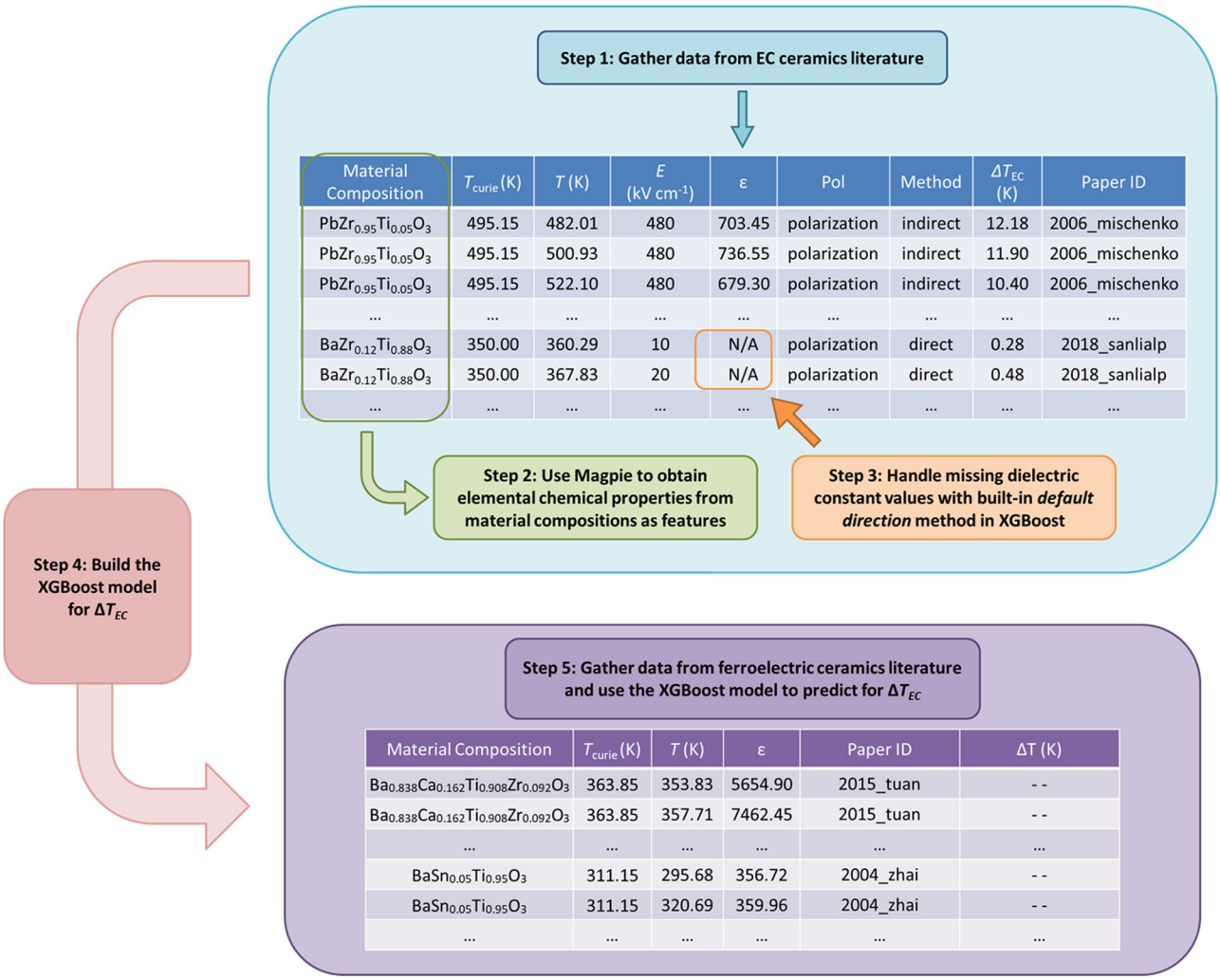

Fig. 1 Flow chart of data gathering and EC temperature change model construction. The five steps are: 1. Data gathering from the EC ceramics literature. 2. Encoding material compositions with Magpie. 3. Handling the missing dielectric constant values with a built-in method in XGBoost. 4. Building the XGBoost model for the EC temperature change prediction. 5. Predicting the EC temperature change for ferroelectric ceramics from the non-EC literature.

highest ΔT_{EC} are forced to be in the training set, unless noted. These materials are $PbZr_{0.95}Ti_{0.05}O_3$[8], whose maximum ΔT_{EC} is above 12 K, and $Bi_{0.5}Na_{0.5}TiO_3$[51], which has the largest negative ΔT_{EC}. To examine the extrapolation capabilities, we forced $PbZr_{0.95}Ti_{0.05}O_3$ to be in the test set and built three models, differentiated by their random seeds, whose results are presented in Supplementary Fig. 3. Although, as expected, the XGBoost models cannot predict a ΔT_{EC} higher than the maximum seen in the training set [8.5 K of $PbZr_{0.97}La_{0.02}(Zr_{0.95}Ti_{0.05})O_3$[52]], they all predict a ΔT_{EC} high in their capability range for $PbZr_{0.95}Ti_{0.05}O_3$. This observation demonstrates that the XGBoost models learned from the underlying physics and can be a useful tool for qualitative prediction and refining the search for new materials.

The 94 EC ceramics are split into train and test data sets based on their distance in the Magpie feature space[53,54]. The Magpie features of the EC materials are first projected onto a two-dimensional *t*-distributed stochastic neighbor embedding (t-SNE) space. A *k*-means clustering of the projection of the 94 materials was then conducted. An optimal *k* value of 3 was determined from the Elbow Method by plotting the within-cluster sum of squares as a function of *k* and identifying the "Elbow" as *k*. A cluster label is assigned to all data. From each cluster, 75% of the materials are picked as the training data and the remaining 25% are picked as the test data. Different numbers of features are used to build the models. We went from all 21 features in Table 1, to 20 features by removing the dielectric constant, to 7 features by removing all Magpie features, and to 6 features by removing the dielectric constant and all Magpie features. For each feature set, we trained 100 XGBoost models with the same hyperparameters but different random seeds and train/test splits. The R^2 and RMSE results (mean and standard deviation) are summarized in Table 3, where the data in each of the rows corresponds to 100 models. The statistics of the model performance when training with the three outlier materials using different feature sets are provided in Supplementary Table 5. Box plots of the results are provided in Supplementary Fig. 4. The standard deviation is due to randomness in (i) the construction of the XGBoost model and (ii) splitting the materials into the training and testing sets.

In all cases, models trained without the outlier materials have significantly better performance in both the train and test scores. This observation suggests that there are features contributing to a large ΔT_{EC} that are not included in our model. This point is further addressed in the "Discussion".

Training with or without the dielectric constant gives a comparable performance. Models trained with all 21 features have a 0.01 lower R^2 score and 0.01 K higher RMSE on the test set.

Table 1. Features used for ΔT_{EC} prediction.

ID	Feature name	Feature type	Feature description	Source
1	E	Numerical	Applied electric field	Literature
2	T	Numerical	Characterization temperature	Literature
3	T_{Curie}	Numerical	Curie temperature	Literature
4	$T-T_{Curie}$	Numerical	Difference between characterization temperature and Curie temperature	Literature
5	pol	Categorical	Stage where ΔT_{EC} is measured (polarization or depolarization)	Literature
6	method	Categorical	Measurement method (direct or indirect)	Literature
7	ϵ	Numerical	Dielectric constant at T	Lit./XGBoost
8	maxdiff_NfUnfilled	Numerical	Maximum difference of number of unfilled *f*-orbitals among elements in the composition	Magpie
9	maxdiff_NdValence	Numerical	Maximum difference of number of valence *d*-orbitals among elements in the composition	Magpie
10	maxdiff_NdUnfilled	Numerical	Maximum difference of number of unfilled *d*-orbitals among elements in the composition	Magpie
11	NComp	Numerical	Number of elements in the composition	Magpie
12	max_NUnfilled	Numerical	Maximum of unfilled valence orbitals among elements in the composition	Magpie
13	dev_MeltingT	Numerical	Standard deviation of melting temperature among elements in the composition	Magpie
14	maxdiff_Mendeleev Number	Numerical	Maximum difference of Mendeleev number among elements in the composition	Magpie
15	mean_NfValence	Numerical	Mean of number of valence *f*-orbitals	Magpie
16	min_Number	Numerical	Minimum of atomic number among elements in the composition	Magpie
17	mean_NdUnfilled	Numerical	Mean of number of unfilled *d*-orbitals among elements in the composition	Magpie
18	mean_GSbandgap	Numerical	Mean of DFT bandgap of elemental solid among elements in the composition	Magpie
19	maxdiff _GSvolume_pa	Numerical	Maximum difference of DFT-computed volume of elemental solid among elements in the composition	Magpie
20	maxdiff_MeltingT	Numerical	Maximum of difference of melting temperature among elements in the composition	Magpie
21	mean_Row	Numerical	Mean of row number of elements in the composition	Magpie

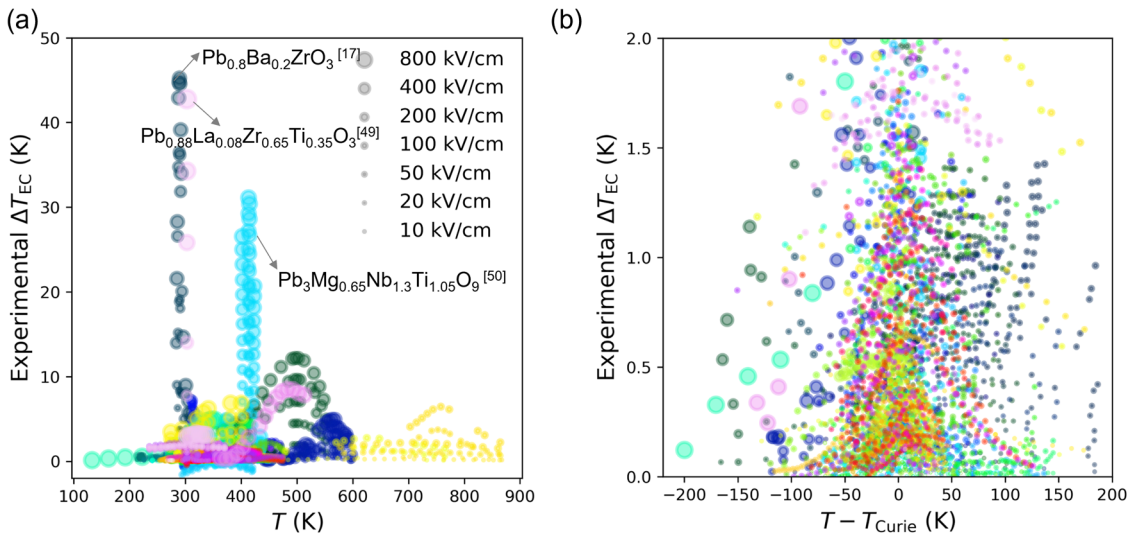

Fig. 2 Experimentally measured EC temperature change (ΔT_{EC}). a Full scale of ΔT_{EC} versus temperature. **b** Data points with ΔT_{EC} between 0 and 2 K versus $T-T_{Curie}$. Different colors are used to distinguish different material compositions. The marker sizes are proportional to the applied electric field.

When training on 7 features (i.e., the Magpie features removed), there is a large decrease in the test performance (−0.2 in R^2 and +0.1 K in RMSE) and an increase in the standard deviations because XGBoost is losing all information about the material composition. Further removal of the dielectric constant (6 features remain) leads to a slight improvement (+0.04 in R^2 and −0.02 K in RMSE) in the test performance compared to using 7 features. This improvement is because the model is introducing noise when imputing the missing dielectric constants without information from the Magpie features. In all cases, the standard deviations of the test sets are an order of magnitude higher than those of the training sets. The main contributor to the large standard deviations in the test sets is poor train/test splits where a large difference exits in the distributions of labels between these two groups. Even though all three outlier materials are removed and $PbZr_{0.95}Ti_{0.05}O_3$ is placed in the training set when building models,

Table 2. XGBoost hyperparameters from grid search.

Parameter name	Meaning	Value
objective	Learning task and corresponding learning objective	reg:pseudohubererror
n_estimators	Number of trees	120
max_depth	Maximum depth of a tree	12
learning_rate	Step size shrinkage to prevent overfitting	0.3
reg_alpha	L1 regularization parameter	5
reg_lambda	L2 regularization parameter	20
colsample_bynode	Subsample ratio of columns for each node (split)	0.3
subsample	Subsample given ratio of the training instances prior to growing trees to prevent overfitting	0.5
gamma	Minimum loss reduction required to make a further partition on a leaf node of the tree	1

Table 3. XGBoost model statistics (mean and standard deviation) using different features.

Features	Train R^2	Test R^2	Train RMSE (K)	Test RMSE (K)
21	0.90 ± 0.02	0.58 ± 0.19	0.37 ± 0.03	0.64 ± 0.16
20	0.89 ± 0.02	0.58 ± 0.19	0.38 ± 0.02	0.64 ± 0.16
7	0.87 ± 0.02	0.40 ± 0.42	0.41 ± 0.03	0.75 ± 0.26
6	0.79 ± 0.03	0.44 ± 0.33	0.52 ± 0.04	0.73 ± 0.22

there are other materials with larger ΔT_{EC}s than the majority of the data (e.g., 8.5 K in $Pb_{0.97}La_{0.02}Zr_{0.95}Ti_{0.05}O_3$[52], 7.2 K in $BaZr_{0.2}Ti_{0.8}O_3$[55], and 6.9 K in $PbSc_{0.5}Ta_{0.5}O_3$[56]) that may not be placed in the training set. In contrast, training and testing on the same split with 100 random seeds only leads to a standard deviation of 0.04 (0.03 K) in the R^2 (RMSE) of the test set.

The parity plot for the predicted ΔT_{EC} versus the experimental values for one of the high-performing models trained with the full 21 features is shown in Fig. 3. This model will also be used for feature analysis and predictions. For this model, the R^2 (RMSE) scores for the train and test data are 0.90 (0.38 K) and 0.77 (0.38 K). In the high ΔT_{EC} range (above 7 K), the model underpredicts ΔT_{EC} due to the scarcity of materials that show a giant EC effect (3 out of 94 materials and 27 out of 4227 data points have $\Delta T_{EC} > 7$ K).

A test R^2 value of 0.77 for a model without material microstructural information, as we have built, is comparable to results from previous studies where tree-based models were used for dielectric constant prediction. For example, Qin et al. [57] trained five commonly used ML models with 32 intrinsic chemical, structural, and thermodynamic features to predict the dielectric constants of ceramics. They gathered 254 single-phase materials from the experimental literature and their random forest model achieved an R^2 score of 0.76 on the test set. Takahashi et al. [58] built random forest models to predict the electronic and ionic contributions to the static dielectric constant. Their dataset consisted of approximately 1200 metal oxides from density functional perturbation theory calculations. The introduction of structural descriptors to the original compositional descriptors improved the model performance. The test R^2 score increased from 0.87 to 0.89 for the electronic contribution and from 0.65 to 0.73 for the ionic contribution.

Feature analysis

We conducted feature analysis on the XGBoost ΔT_{EC} model whose parity plot is shown in Fig. 3. The impurity-based feature importance is calculated by XGBoost by measuring the total gain (i.e., improvement in accuracy) across all splits where the feature is used. The higher the feature importance value, the more important that feature is. Impurity-based feature analysis tends to favor features with high cardinality (i.e., more unique values). As such, permutation-based feature importance, which measures the decrease of the model score when the values in a single feature are randomly shuffled, is also reported in Supplementary Fig. 5. In both methods, the six most important features are the same.

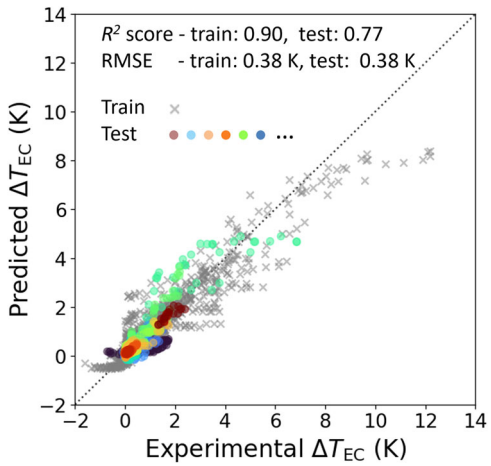

Fig. 3 Parity plot of the predicted EC temperature change from one of the high-performing XGBoost models versus the experimental values. The gray crosses denote the training materials and the colored dots denote the test materials (different colors are used to distinguish different material compositions).

The impurity-based feature importance is presented in Fig. 4. The applied electric field E ranks first, followed by $T - T_{Curie}$. These observations are in agreement with the known physics: ΔT_{EC} of an EC material increases with the applied electric field (before electrical breakdown) and is usually the largest around its Curie temperature as the material goes through a phase transition. We further analyzed the features by setting $E = 100$ kV/cm and $T = T_{Curie}$ for ΔT_{EC} predictions, thus ruling out their influence, and then calculating the Pearson correlation coefficient of the remaining features with the predicted ΔT_{EC}. As shown in the color map in Fig. 4, the predicted ΔT_{EC} shows mildly positive correlations to T (Pearson correlation coefficient of 0.28) and ϵ (Pearson correlation coefficient of 0.42). A Pearson correlation coefficient heat map among all features is reported in Supplementary Fig. 6.

The 16 Magpie features synergistically help the model distinguish between different materials. As such, the direct interpretation of an individual feature can be difficult. The predicted ΔT_{EC} shows moderately positive correlations to

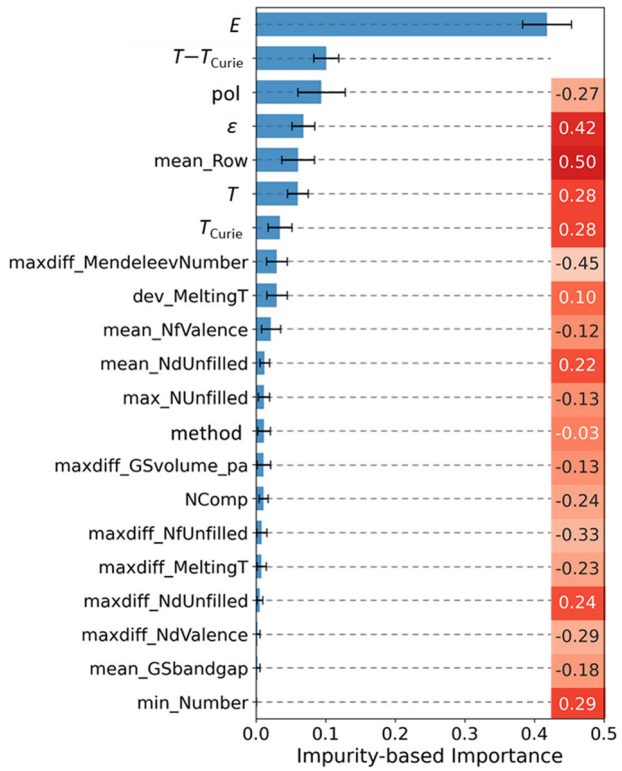

Fig. 4 Impurity-based feature importance analysis. The error bar comes from the standard deviation of 100 random seeds. The color map shows the Pearson correlation coefficient of each feature with the predicted ΔT_{EC} in the model shown in Fig. 3 after ruling out the influence of E and $T - T_{Curie}$ by setting $E = 100$ kV/cm and $T = T_{Curie}$.

mean_Row (Pearson correlation coefficient of 0.50) after ruling out the influence of E and $T-T_{Curie}$. This result is consistent with the observation that the ceramics in our dataset with a negative EC effect usually contain sodium, while 9 of the top 10 materials with largest experimental ΔT_{EC} contain lead.

maxdiff_MendeleevNumber is ranked as the most important Magpie feature in Fig. 4. Recall that in the feature selection, we dropped Magpie features with a Pearson correlation coefficient >0.95 with existing features. In this case, MaxIonicChar, maxdiff_Electronegativity, and min_Electronegativity show a Pearson correlation coefficient of 0.96, 0.96, and −0.96 with maxdiff_MendeleevNumber. It is interesting that this set of four features is picked up by the XGBoost model among all 145 Magpie features (or 110 features of non-zero variance). As shown in Fig. 4, the predicted ΔT_{EC} has a moderately negative correlation to maxdiff_MendeleevNumber (−0.45). This feature set could be related to the electronic and ionic polarization in EC ceramics, but microstructural information (e.g., phases and grain size) for these materials is required before making firm conclusions.

Although the measurement condition pol is ranked as the third highest in the impurity-based feature importance, it is ranked fifth in the permutation-based feature importance. The model tends to predict a 0.3 K higher ΔT_{EC} if pol indicates polarization instead of depolarization. The measurement method does not stand out, as might have been expected. We also note that the differences in the measured ΔT_{EC} from pol and method in literature are generally below 0.2 K, smaller than the model RMSE (0.38 K).

The observed low impurity- and/or permutation-based importance of some Magpie features (e.g., maxdiff_NdUnfilled, mean_GSbandgap, and min_number) suggest that they can be removed. Given that our XGBoost model is rigorously regularized (i.e., with tree pruning and restrictions in the splitting of nodes), the model is not sensitive to the inclusion of less relevant features.

Prediction on ferroelectric materials

We now examine the predicted ΔT_{EC} for 66 ferroelectric ceramics whose EC performance has not been reported. We use the XGBoost ΔT_{EC} model whose parity plot is shown in Fig. 3 for the predictions. The electric field is set to 100 kV/cm. The features for the characterization conditions are set as direct measurement (method) and polarization (pol). The predicted ΔT_{EC} as a function of temperature (for the temperatures at which the dielectric constant is reported) are plotted in Fig. 5a for the 66 ferroelectric ceramics and in Fig. 5b for the original EC dataset. To simplify the plots, we grouped materials into families based on compositional similarities. For example, $(1-x)PbMg_{1/3}Nb_{2/3}O_{3-x}PbTiO_3$ is denoted as PMN-PT and $BaZr_xTi_{1-x}O_3$ is denoted as BZT. A full list of material compositions with their family group is provided in a separate csv file[48], as described in Supplementary Note 1.

For the ferroelectric ceramics from the non-EC literature, $PbZr_{0.57}Ti_{0.43}O_3$ has the highest predicted ΔT_{EC} of 3.6 K at 100 kV/cm and 600 K. The EC material $PbZr_{0.3}Ti_{0.7}O_3$[59] exhibits the largest ΔT_{EC} of 4.1 K at 100 kV/cm and 753 K. We identify good candidates in Table 4: $Pb_{0.89}La_{0.11}Zr_{0.68075}Ti_{0.29175}O_3$[60] is predicted to have a ΔT_{EC} of 2.9 K at 328 K and the lead-free $Ba_{0.91}Ca_{0.09}Ti_{0.86}Zr_{0.14}O_3$[61] is predicted to have a ΔT_{EC} of 2.6 K at 318 K.

Although no material from the non-EC literature exceeds the performance of the state-of-the art EC materials, the top lead-free candidates in Table 4 are worth exploring. By adding new measurements of promising materials to the existing dataset, active learning techniques can be used to iteratively train new models and search for promising materials[53]. Active learning approaches, such as Bayes' theorem on materials discovery[62] and uncertainty-driven active learning for neural network potential development[63], have demonstrated success in reducing experimental trials and/or computational cost.

DISCUSSION

The predictive ability of our XGBoost model demonstrates that a physics-informed data-driven approach is a promising avenue to studying EC materials. Moving forward, an organized database dedicated for ferroelectrics and/or EC materials would be useful to the community. To further improve the model's predictive capability and ability to elucidate underlying physical meaning, more detailed information on the EC materials is required. Many of the factors that influence ΔT_{EC}, such as sample size, morphology, and crystallinity, are not reported in the literature and are therefore not accessible to our model. Representative examples are provided below.

Going from a thick film to a thin film leads to an increased ΔT_{EC} at a moderate voltage[3,25] and at the same time changes the dielectric properties of a material. For example, the maximum dielectric constant in $Pb_{0.8}Ba_{0.2}ZrO_3$ is around 12,000 in a sintered bulk sample[64] while it has a value of 1200 in a 320 nm thin film[17]. Gao et al.[65] demonstrated the effect of film thickness with antiferroelectric $Pb_{0.82}Ba_{0.08}La_{0.10}(Zr_{0.90}Ti_{0.10})O_3$ (PBLZT) thick films. ΔT_{EC} values of 25.1, 19.8, and 13.9 K were measured for film thicknesses of 1.0, 1.5, and 2.0 mm at 700 kV/cm and room temperature. They attributed the decrease of ΔT_{EC} as film thickness increases to the reduction of the preferred orientation. The PBLZT thick films changed from a ⟨100⟩-preferred orientation to random orientation with increased film thickness.

Phase coexistence can contribute to a large ΔT_{EC} and many EC ceramics are synthesized with compositions from the morphotropic phase boundary[17,50,66]. For example, thin film $Pb_{0.8}Ba_{0.2}ZrO_3$ has a giant ΔT_{EC} of 45 K at room temperature as it goes through an electric field-induced transition from an orthorhombic antiferroelectric

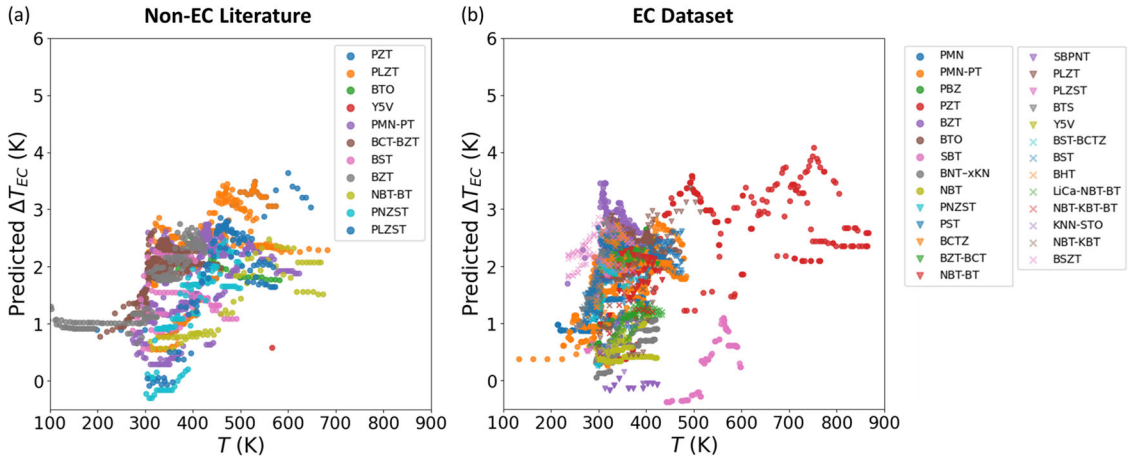

Fig. 5 Predicted ΔT_{EC} at 100 kV/cm and varying temperature for ferroelectric ceramics. a Predictions on the non-EC literature. **b** Predictions on the EC dataset.

Table 4. Potential EC candidates from the ferroelectric dataset predicted at 100 kV/cm.

Material composition	T (K)	Predicted ΔT_{EC}
$PbZr_{0.57}Ti_{0.43}O_3$[80]	600	3.6
$PbZr_{0.9}Ti_{0.1}O_3$[81]	530	3.5
$Pb_{0.89}La_{0.11}Zr_{0.68075}Ti_{0.29175}O_3$[60]	328	2.9
$BaZr_{0.2}Ti_{0.8}O_3$[82]	373	2.7
$Ba_{0.91}Ca_{0.09}Ti_{0.86}Zr_{0.14}O_3$[61]	318	2.6
$BaSn_{0.05}Ti_{0.95}O_3$[83]	312	2.5

phase to a rhombohedral ferroelectric phase[17]. The large entropy change from random nanoregions to ordered nanoregions in thin film $Pb_{0.8}Ba_{0.2}ZrO_3$ under an electric field may also contribute to the giant EC effect[17].

Crystal orientation and phase can lead to a different polarization response and hence a different ΔT_{EC}[50,67–70]. Luo et al.[68] found the maximum ΔT_{EC} for ⟨111⟩- and ⟨001⟩-oriented $0.71PbMg_{1/3}Nb_{2/3}O_3$-$0.29PbTiO_3$ single crystals to be 2.0 and 2.3 K at 50 kV/cm. Bai et al.[66] demonstrated with $(1-x)PbMg_{1/3}Nb_{2/3}O_{3-x}PbTiO_3$ ($x = 0.3$–0.35) that the tetragonal-cubic phase transition induced a much larger ΔT_{EC} (0.69 K) than the rhombohedral-cubic and rhombohedral-tetragonal transitions. The tetragonal phase has the largest polarization vector and the complete disappearance of dipoles in the cubic phase induces a much larger polarization change than their rearrangement in the rhombohedral phase. A difference of 0.4 K in ΔT_{EC} is also observed among their ⟨001⟩, ⟨110⟩, and ⟨111⟩ crystals.

High crystallinity, a smaller grain size, and isometric grain shape are beneficial to the EC performance[25,71]. By introducing Pr to $SrBi_2(Nb_{0.2}Ta_{0.2})_2O_9$, Axelsson et al.[71] synthesized dense samples with no secondary phase. The highly isometric grain shape of the Aurivillius phases in combination with a small grain size increased the grain boundary density, leading to an increase in the dielectric strength. The dopant Pr^{3+} without lone pair electrons also leads to a diffuse relaxor-ferroelectric phase transition and shifted the phase transition temperature to a lower value. Other factors such as the substrate material and orientation, as well as the sample growth temperature can also influence ΔT_{EC} by introducing clamping, misfit strain, and/or thermal stresses[72]. The order of the phase transition in an EC material impacts the shape of ΔT_{EC} versus T and is also worth considering[3].

Enforcing physical laws in ML models (e.g., tailoring neural network architectures, designing kernel-based regression models, encoding simple symmetries, or choosing appropriate loss functions) provides another avenue for fast and accurate training and improved generalization ability[73]. In the field of EC cooling, ML architectures that associate ΔT_{EC} with the polarization at a given E and T, or that incorporate mathematical constraints from the Maxwell relations are potential approaches. Parallel efforts for EC polymers would also be of great interest.

METHODS

Materials dataset preparation

We collected 97 materials from 45 papers. For each of these materials, we extracted ΔT_{EC} at different characterization temperatures T and applied electric fields E and then included these quantities as features for the predictive model. As another feature, we extracted the dielectric constant ϵ at different temperatures (and at the lowest reported frequency if measured at multiple frequencies) and conducted linear interpolation to obtain values at the temperatures where the ΔT_{EC} is measured. 34 of the 97 materials do not have a dielectric constant reported with their ΔT_{EC}. We used the built-in algorithm from the XGBoost[47] model to handle the missing values. The phase transition (e.g., ferroelectric to paraelectric) temperature is recorded as the feature T_{Curie}. When the Curie temperature or phase transition temperature is not explicitly reported, we extracted the temperature where the peak of ΔT_{EC} occurs as T_{Curie}. A precalculated feature $T-T_{Curie}$ is also included. The categorical feature *method* is used to denote how ΔT_{EC} is measured. The direct measurement (where ΔT_{EC} is measured directly using thermometers or by recording the heat flow using calorimeters[3]) and the indirect measurement (where ΔT_{EC} is obtained from polarization change via the Maxwell relations[1,74]) may provide different values and/or trends for ΔT_{EC} for the same material and conditions[55,71,75–77]. The categorical feature pol records whether ΔT_{EC} is measured in the polarization or depolarization stage. Due to small losses associated with the polarization hysteresis and Joule heating, the measured EC temperature increase in polarization can be ~0.2 K larger than the EC temperature decrease in depolarization[76,77]. Although the phenomenon of the EC effect should be nominally reversible and independent of measurement conditions[78], we use pol and method to help distinguish the source of the experimentally measured ΔT_{EC}.

Each of the 97 materials has a different number of entries in the dataset based on how many electric fields and/or temperatures they were measured at. We trimmed data points when a material had over 100 entries, as these ΔT_{EC}s are recorded at small electric field and/or temperature intervals.

Three of the 97 materials are marked as outliers: $Pb_{0.88}La_{0.08}Zr_{0.65}Ti_{0.35}O_3$[49], $Pb_{0.8}Ba_{0.2}ZrO_3$[17], and $Pb_3Mg_{0.65}Nb_{1.3}Ti_{1.05}O_9$[50]. These materials have a giant EC effect, with a maximum ΔT_{EC} above 30 K, while the maximum ΔT_{EC} of the rest of the materials does not exceed 13 K. There may be microstructural effects contributing to the giant EC effect in these three outlier materials that cannot be captured by the descriptors

generated from material compositions and commonly reported properties. $Bi_{0.5}Na_{0.5}TiO_3$ (BNT)[51] has the smallest ΔT_{EC} of −1.6 K. A negative ΔT_{EC} is known as the negative EC effect, where the isofield polarization of a material increases with increasing temperature (e.g., an antiferroelectric–ferroelectric transition with increasing temperature)[78]. In total, 16 materials have one or more data points where ΔT_{EC} is smaller than zero. Of these, the magnitude of ΔT_{EC} is <0.5 K except for three BNT-based materials.

Feature selection

Each material composition was converted into 145 features with the Magpie package[42]. The Magpie features are obtained using the minimum, maximum, and mean of elemental chemical properties and their positions in the periodic table. Given the large quantity of generated features, not all are relevant in our study. We conducted feature selection to remove redundant features, which could add noise to the model. First, we removed Magpie features that have zero variance in our dataset, which leads to 111 remaining Magpie features. For example, as all the collected EC ceramics contain oxygen, the minimum atomic weight and the minimum melting temperature of the individual elements have the same value. We then reduced the feature collinearity by dropping features that have a Pearson correlation coefficient higher than 0.95 with an existing feature, leading to 38 Magpie features (listed in Supplementary Table 1). A list of features dropped due to a large Pearson correlation coefficient with existing features is listed in Supplementary Table 2. The choice of the feature to keep from a set of highly correlated features has a negligible influence on the model performance (see Supplementary Fig. 1 and Supplementary Table 3).

Next, we conducted a backward feature elimination process on these Magpie features. The three outlier points with giant ΔT_{EC} are temporarily removed in this step to avoid high variance in the cross-validation R^2 score. We started with 45 features (38 Magpie features plus 7 features extracted from literature), computed the performance of the XGBoost model after eliminating each Magpie feature, and then removed the least-significant feature until no improvement was observed. 14 Magpie features (feature IDs 8–21 in Table 1) remained after this process. The number of features kept versus the cross-validation R^2 score is plotted in Supplementary Fig. 2. The order of features being dropped in this process is listed in Supplementary Table 4.

XGBoost regression

XGBoost[47] is an open-source library with an efficient and scalable implementation of the gradient boosting framework[79]. Gradient boosting minimizes the prediction error with a gradient descent algorithm and produces a model in the form of a set of weak prediction models (decision trees in this case). During the training, gradient boosting adds new regression trees one at a time to reduce the residual (i.e., the difference between the model predictions and the label values). Existing trees in the model remain untouched, which slows down the rate of overfitting. The output of the new tree is combined with the output of existing trees until the loss is minimized below a threshold or the specified limit of trees (e.g., maximum depth) is reached. After the trees are built, XGBoost can apply tree pruning, where the size of the trees is reduced by pruning nodes from bottom (leaves) to top (root) if the loss reduction of having that node is smaller than the regularization parameter *gamma*.

We applied XGBoost with the Scikit-learn package[37]. The hyperparameters were tuned with a grid search using training data. A five-fold cross-validation resampling technique was used. When examining each hyperparameter set, the training data was randomly split into five groups of approximately equal size. Four groups of the data were used for the training set and the remaining group was used as the validation set. This procedure was repeated five times, leading to five validation R^2 scores for each hyperparameter set. The set with the highest averaged validation R^2 score was selected.

The dielectric constant is a relevant property for the EC performance of a material. A linear interpolation of the dielectric constant at the temperatures where ΔT_{EC} was measured provided data for 63 of the materials in the dataset. The missing dielectric constant values in the other 34 materials were handled by the default built-in method in the XGBoost algorithm[47]. Essentially, each of the decision nodes in XGBoost has a default direction, such that when a missing value is encountered during splitting in the tree branch, the instance is classified in that default direction.

DATA AVAILABILITY

The data used to train and test the machine learning models are available at https://github.com/Jie0705/MachieneLearning-EC-Dataset.git.

CODE AVAILABILITY

All codes from this study are available from the corresponding author upon reasonable request.

Received: 17 December 2021; Accepted: 10 June 2022;
Published online: 01 July 2022

REFERENCES

1. Lines, M. E. & Glass, A. M. *Principles and Applications of Ferroelectrics and Related Materials* (Oxford University Press, 1977).
2. Scott, J. F. Applications of modern ferroelectrics. *Science* **315**, 954–959 (2007).
3. Scott, J. F. Electrocaloric materials. *Annu. Rev. Mater. Res.* **41**, 229–240 (2011).
4. Correia, T. & Zhang, Q. *Electrocaloric Materials* (Springer, 2014).
5. Zhao, D. & Tan, G. A review of thermoelectric cooling: materials, modeling and applications. *Appl. Therm. Eng.* **66**, 15–24 (2014).
6. Gómez, J. R., Garcia, R. F., Catoira, A. D. M. & Gómez, M. R. Magnetocaloric effect: a review of the thermodynamic cycles in magnetic refrigeration. *Renew. Sust. Energ. Rev.* **17**, 74–82 (2013).
7. Kobeko, P. & Kurtschatov, J. Dielektrische eigenschaften der seignettesalzkristalle. *Z. Phys.* **66**, 192–205 (1930).
8. Mischenko, A. S., Zhang, Q., Scott, J. F., Whatmore, R. W. & Mathur, N. D. Giant electrocaloric effect in thin-film $PbZr_{0.9}Ti_{0.05}O_3$. *Science* **311**, 1270–1271 (2006).
9. Neese, B. et al. Large electrocaloric effect in ferroelectric polymers near room temperature. *Science* **321**, 821–823 (2008).
10. Li, Q. et al. Relaxor ferroelectric-based electrocaloric polymer nanocomposites with a broad operating temperature range and high cooling energy. *Adv. Mater.* **27**, 2236–2241 (2015).
11. Zhang, G. et al. Nanoconfinement-induced giant electrocaloric effect in ferroelectric polymer nanowire array integrated with aluminum oxide membrane to exhibit record cooling power density. *Adv. Mater.* **31**, 1806642 (2019).
12. Chen, Y. et al. An all-scale hierarchical architecture induces colossal room-temperature electrocaloric effect at ultralow electric field in polymer nanocomposites. *Adv. Mater.* **32**, 1907927 (2020).
13. Epstein, R. I. & Malloy, K. J. Electrocaloric devices based on thin-film heat switches. *J. Appl. Phys.* **106**, 064509 (2009).
14. Guo, D. et al. Design and modeling of a fluid-based micro-scale electrocaloric refrigeration system. *Int. J. Heat Mass Transf.* **72**, 559–564 (2014).
15. Defay, E. et al. Enhanced electrocaloric efficiency via energy recovery. *Nat. Commun.* **9**, 1827 (2018).
16. Gong, J. & McGaughey, A. J. H. Device-level thermodynamic model for an electrocaloric cooler. *Int. J. Energy Res.* **44**, 5343–5359 (2020).
17. Peng, B., Fan, H. & Zhang, Q. A giant electrocaloric effect in nanoscale antiferroelectric and ferroelectric phases coexisting in a relaxor $Pb_{0.8}Ba_{0.2}ZrO_3$ thin film at room temperature. *Adv. Funct. Mater.* **23**, 2987–2992 (2013).
18. Ma, R. et al. Highly efficient electrocaloric cooling with electrostatic actuation. *Science* **357**, 1130–1134 (2017).
19. Ponomareva, I. & Lisenkov, S. Bridging the macroscopic and atomistic descriptions of the electrocaloric effect. *Phys. Rev. Lett.* **108**, 167604 (2012).
20. Prosandeev, S., Ponomareva, I. & Bellaiche, L. Electrocaloric effect in bulk and low-dimensional ferroelectrics from first principles. *Phys. Rev. B* **78**, 052103 (2008).
21. Kühn, M. & Kliem, H. Monte carlo simulations of ferroelectric properties based on a microscopic model for pvdf. *Phys. Status Solidi B* **245**, 213–223 (2008).
22. Yu, Y.-J. & McGaughey, A. J. H. Energy barriers for dipole moment flipping in pvdf-related ferroelectric polymers. *Chem. Phys.* **144**, 014901 (2016).
23. Ma, Y.-B., Albe, K. & Xu, B.-X. Lattice-based Monte Carlo simulations of the electrocaloric effect in ferroelectrics and relaxor ferroelectrics. *Phys. Rev. B* **91**, 184108 (2015).
24. Ma, Y.-B. et al. State transition and electrocaloric effect of $BaZr_x Ti_{1-x}O_3$: simulation and experiment. *J. Appl. Phys.* **121**, 024103 (2017).
25. Valant, M. Electrocaloric materials for future solid-state refrigeration technologies. *Prog. Mater. Sci.* **57**, 980–1009 (2012).
26. Mitchell, T. M. et al. *Machine Learning* Vol. 45, 870–877 (McGraw Hill, Burr Ridge, IL, 1997) . .
27. Gasteiger, J. & Zupan, J. Neural networks in chemistry. *Angew. Chem. Int. Ed. Engl.* **32**, 503–527 (1993).
28. Sumpter, B. G., Getino, C. & Noid, D. W. Theory and applications of neural computing in chemical science. *Annu. Rev. Phys. Chem.* **45**, 439–481 (1994).

29. Thomsen, J. & Meyer, B. Pattern recognition of the 1h nmr spectra of sugar alditols using a neural network. *J. Magn. Reson. (1969)* **84**, 212–217 (1989).
30. Keil, M., Exner, T. E. & Brickmann, J. Pattern recognition strategies for molecular surfaces: Iii. Binding site prediction with a neural network. *J. Comput. Chem.* **25**, 779–789 (2004).
31. Agrafiotis, D. K., Cedeno, W. & Lobanov, V. S. On the use of neural network ensembles in qsar and qspr. *J. Chem. Inf. Comput. Sci.* **42**, 903–911 (2002).
32. Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A. & Kim, C. Machine learning in materials informatics: recent applications and prospects. *NPJ Comput. Mater.* **3**, 1–13 (2017).
33. Jain, A. et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
34. Bergerhoff, G., Brown, I. & Allen, F. et al. Crystallographic databases. *IUCr* **360**, 77–95 (1987).
35. Curtarolo, S. et al. Aflow: an automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218–226 (2012).
36. 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).
37. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
38. Paszke, A. et al. Pytorch: an imperative style, high-performance deep learning library. In *Advances in Neural Information Processing* (eds. Wallach, H. et al.) Vol. 32, 8024–8035 (Curran Associates, Inc., 2019).
39. Abadi, M. et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467 (2016).
40. Holleis, L., Shivaram, B. & Balachandran, P. V. Machine learning guided design of single-molecule magnets for magnetocaloric applications. *Appl. Phys. Lett.* **114**, 222404 (2019).
41. Stanev, V. et al. Machine learning modeling of superconducting critical temperature. *NPJ Comput. Mater.* **4**, 1–14 (2018).
42. Ward, L., Agrawal, A., Choudhary, A. & Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *NPJ Comput. Mater.* **2**, 1–7 (2016).
43. Cao, Z. et al. Convolutional neural networks for crystal material property prediction using hybrid orbital-field matrix and magpie descriptors. *Crystals* **9**, 191 (2019).
44. Mannodi-Kanakkithodi, A., Pilania, G., Huan, T. D., Lookman, T. & Ramprasad, R. Machine learning strategy for accelerated design of polymer dielectrics. *Sci. Rep.* **6**, 20952 (2016).
45. Su, M., Grimes, R., Garg, S., Xue, D. & Balachandran, P. V. Machine-learning-enabled prediction of adiabatic temperature change in lead-free $BaTiO_3$-based electrocaloric ceramics. *ACS Appl. Mater. Interfaces* **13**, 53475–53484 (2021).
46. Tian, Y. et al. Determining multi-component phase diagrams with desired characteristics using active learning. *Adv. Sci.* **8**, 2003165 (2021).
47. Chen, T. & Guestrin, C. Xgboost: a scalable tree boosting system. *In Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794 (ACM, New York, NY, USA, 2016). [Online]. Available: https://doi.org/10.1145/2939672.2939785
48. Gong, J., Chu, S., Mehta, K. M., & McGaughey, A. J. H. Link to GitHub Data set: https://github.com/Jie0705/MachieneLearning-EC-Dataset.git (2022).
49. Lu, S. et al. Organic and inorganic relaxor ferroelectrics with giant electrocaloric effect. *Appl. Phys. Lett.* **97**, 162904 (2010).
50. Saranya, D., Chaudhuri, A. R., Parui, J. & Krupanidhi, S. Electrocaloric effect of PMN-PT thin films near morphotropic phase boundary. *Bull. Mater. Sci.* **32**, 259–262 (2009).
51. Jiang, X., Luo, L., Wang, B., Li, W. & Chen, H. Electrocaloric effect based on the depolarization transition in $(1-x)Bi_{0.5}Na_{0.5}TiO_{3-x}KNbO_3$ lead-free ceramics. *Ceram. Int.* **40**, 2627–2634 (2014).
52. Hao, X., Yue, Z., Xu, J., An, S. & Nan, C.-W. Energy-storage performance and electrocaloric effect in (100)-oriented $Pb_{0.97}La_{0.02}$ $(Zr_{0.95}Ti_{0.05})$ O_3 antiferroelectric thick films. *J. Appl. Phys.* **110**, 064109 (2011).
53. Meredig, B. et al. Can machine learning identify the next high-temperature superconductor? Examining extrapolation performance for materials discovery. *Mol. Syst. Des. Eng.* **3**, 819–825 (2018).
54. Huan, T. D. et al. A universal strategy for the creation of machine learning-based atomistic force fields. *NPJ Comput. Mater.* **3**, 1–8 (2017).
55. Qian, X.-S. et al. Giant electrocaloric response over a broad temperature range in modified $BaTiO_3$ ceramics. *Adv. Funct. Mater.* **24**, 1300–1305 (2014).
56. Correia, T. et al. PST thin films for electrocaloric coolers. *J. Phys. D: Appl. Phys.* **44**, 165407 (2011).
57. Qin, J., Liu, Z., Ma, M. & Li, Y. Machine learning approaches for permittivity prediction and rational design of microwave dielectric ceramics. *J. Materiomics* **7**, 1284–1293 (2021).
58. Takahashi, A., Kumagai, Y., Miyamoto, J., Mochizuki, Y. & Oba, F. Machine learning models for predicting the dielectric constants of oxides based on high-throughput first-principles calculations. *Phys. Rev. Mater.* **4**, 103801 (2020).
59. Qiu, J. & Jiang, Q. Effect of electric field on electrocaloric effect in $Pb(Zr_{1-x}Ti_x)O_3$ solid solution. *Phys. Lett. A* **372**, 7191–7195 (2008).
60. Kumar, A. et al. High energy storage properties and electrical field stability of energy efficiency of $(Pb_{0.89}La_{0.11})(Zr_{0.70}Ti_{0.30})_{0.9725}$ O_3 relaxor ferroelectric ceramics. *Electron. Mater. Lett.* **15**, 323–330 (2019).
61. Liu, W. & Ren, X. Large piezoelectric effect in Pb-free ceramics. *Phys. Rev. Lett.* **103**, 257602 (2009).
62. Kusne, A. G. et al. On-the-fly closed-loop materials discovery via bayesian active learning. *Nat. Commun.* **11**, 1–11 (2020).
63. Wang, W., Yang, T., Harris, W. H. & Gómez-Bombarelli, R. Active learning and neural network potentials accelerate molecular screening of ether-based solvate ionic liquids. *Chem. Commun.* **56**, 8920–8923 (2020).
64. Pokharel, B. P. & Pandey, D. Dielectric studies of phase transitions in $(Pb_{1-x}Ba_x)$ ZrO_3. *J. Appl. Phys.* **88**, 5364–5373 (2000).
65. Gao, H., Hao, X., Zhang, Q., An, S. & Kong, L. B. Thickness-dependent electrocaloric effect of $Pb_{0.82}Ba_{0.08}La_{0.10}(Zr_{0.90}Ti_{0.10})O_3$ antiferroelectric thick films. *J. Alloys Compd.* **690**, 131–138 (2017).
66. Bai, Y., Wei, D. & Qiao, L.-J. Control multiple electrocaloric effect peak in Pb $(Mg_{1/3}Nb_{2/3})$ O_3-$PbTiO_3$ by phase composition and crystal orientation. *Appl. Phys. Lett.* **107**, 192904 (2015).
67. Sebald, G. et al. Electrocaloric and pyroelectric properties of 0.75Pb $(Mg_{1/3}Nb_{2/3})$ O_3–$0.25PbTiO_3$ single crystals. *J. Appl. Phys.* **100**, 124112 (2006).
68. Luo, L., Dietze, M., Solterbeck, C.-H., Es-Souni, M. & Luo, H. Orientation and phase transition dependence of the electrocaloric effect in 0.71 $PbMg_{1/3}Nb_{2/3}O_3$–$0.29PbTiO_3$ single crystal. *Appl. Phys. Lett.* **101**, 062907 (2012).
69. Peräntie, J., Hagberg, J., Uusimäki, A. & Jantunen, H. Electric-field-induced dielectric and temperature changes in a ⟨011⟩-oriented Pb $(Mg_{1/3}Nb_{2/3})$ O_3–$PbTiO_3$ single crystal. *Phys. Rev. B* **82**, 134119 (2010).
70. Goupil, F. L., Berenov, A., Axelsson, A.-K., Valant, M. & Alford, N. M. Direct and indirect electrocaloric measurements on ⟨001⟩ - $PbMg_{1/3}Nb_{2/3}O_3$–$30PbTiO_3$ single crystals. *J. Appl. Phys.* **111**, 124109 (2012).
71. Axelsson, A.-K., Le Goupil, F., Valant, M. & Alford, N. M. Electrocaloric effect in lead-free aurivillius relaxor ferroelectric ceramics. *Acta Mater.* **124**, 120–126 (2017).
72. Zhang, J., Heitmann, A., Alpay, S. & Rossetti Jr, G. Aspects of the electrocaloric behavior of ferroelectric thin films: a review of the predictions of the Landau–Ginzburg theory. *Integr. Ferroelectr.* **125**, 168–175 (2011).
73. Karniadakis, G. E. et al. Physics-informed machine learning. *Nat. Rev. Phys.* **3**, 422–440 (2021).
74. Tuttle, B. A. & Payne, D. A. The effects of microstructure on the electrocaloric properties of $Pb(Zr,Sn,Ti)O_3$ ceramics. *Ferroelectrics* **37**, 603–606 (1981).
75. Lu, S. G. et al. Comparison of directly and indirectly measured electrocaloric effect in relaxor ferroelectric polymers. *Appl. Phys. Lett.* **97**, 202901 (2010).
76. Kar-Narayan, S. & Mathur, N. D. Direct and indirect electrocaloric measurements using multilayer capacitors. *J. Phys. D: Appl. Phys.* **43**, 032002 (2010).
77. Moya, X. et al. Giant electrocaloric strength in single-crystal $BaTiO_3$. *Adv. Mater.* **25**, 1360–1365 (2013).
78. Moya, X. & Mathur, N. Caloric materials for cooling and heating. *Science* **370**, 797–803 (2020).
79. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat* **29**, 1189–1232 (2001).
80. Ortega, N., Kumar, A. & Katiyar, R. Enhanced polarization and dielectric properties of $Pb(Zr_{1-x}Ti_x)O_3$ thin films. *J. Appl. Phys.* **104**, 074114 (2008).
81. Buixaderas, E. et al. Ultrabroadband dielectric spectroscopy and phonons in $(Pb_{1-x/2}La_x)(Zr_{0.9}Ti_{0.1})O_3$. *J. Appl. Phys.* **108**, 104101 (2010).
82. Chen, H. et al. Microstructure and dielectric properties of $BaZr_xTi_{1-x}O_3$ ceramics. *J. Mater. Sci.: Mater. Electron.* **19**, 379–382 (2008).
83. Zhai, J., Shen, B., Yao, X., Zhang, L. & Chen, H. Dielectric properties of $Ba(Sn_xTi_{1-x})$ O_3 thin films grown by a sol–gel process. *J. Am. Ceram. Soc.* **87**, 2223–2227 (2004).

ACKNOWLEDGEMENTS

We thank Prof. Levent Burak Kara for helpful discussions on machine learning model construction. This work was supported by the U.S. National Science Foundation under Grant No. CBET-1605000.

AUTHOR CONTRIBUTIONS

J.G. designed and carried out the project. J.G., S.C., and R.K.M. conducted data extraction from the literature. The project was supervised by A.J.H.M. All authors discussed the results and contributed to writing 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/s41524-022-00826-3.

Correspondence and requests for materials should be addressed to Alan J. H. McGaughey.

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

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