Computational Materials Science 199 (2021) 110730

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

# Prediction of the Curie temperatures of ferroelectric solid solutions using machine learning methods

Evan M. Askanazi [N], Suhas Yadav [N], Ilya Grinberg [N]

*Department of Chemistry, Bar Ilan University, Ramat Gan 5211401, Israel*

ARTICLE INFO

*Keywords:*
Machine learning
Computational materials
Ferroelectrics
Solid Solutions

ABSTRACT

In this work, we studied the application of machine learning for the prediction of the Curie temperature ($T_c$) of ferroelectric $BiMe'Me''O3$-$PbTiO_3$ systems (where Me′ and Me″ are metal cations). We found that among the studied K-nearest neighbor (KNN), support vector regression (SVR) and random forest (RF) methods, RF obtains the best performance and is insensitive to the choice of hyperparameters. SVR results show a strong sensitivity to the choice of hyperparameters and obtained $T_c$ predictions with significantly lower accuracy than RF even after hyperparameter optimization. KNN results show poor accuracy and are essentially unusable with an incomplete feature set and are only qualitatively accurate with a complete feature set. With regard to the choice of features for accurate prediction of the Ferroelectric (FE) systems, we find that Bi content and B-cation valence, ionic radius and ionic displacements form the irreducible set of features such that these features or their equivalents must be used to obtain quantitatively accurate $T_c$ predictions. We also find that homovalent and heterovalent $BiMe'Me''O_3$-$PbTiO_3$ solid solutions form distinct classes of compounds with different behaviors so that both types must be included in the input data set to obtain high predictive accuracy. Our work confirms that for the small data sets typically available in materials science, careful selection of the input system, features and ML methods is required to enable accurate model construction and discovery of previously unknown relationships, but can be achieved in a systematic manner.

## 1. Introduction

Inspired by the success of ML methods in other fields and the availability of several large datasets such as the Materials Project and AFLOW, the application of machine learning (ML) to materials science has seen rapid growth in recent decade. In particular, several ML studies have been carried out on ferroelectric oxide compounds predicting several new ferroelectrics with higher ferroelectric-to-paraelectric Curie temperatures ($T_c$) [1-6]. The subsequent experimental synthesis and characterization of the predicted compositions mostly confirmed the accuracy of ML predictions but also revealed large deviations from the ML-predicted $T_c$ values in several cases. It is known that the successful application of ML to materials science faces several challenges due to the generally small size of the data set available from experimental sources that makes the predictive ability of ML models sensitive to the selection of the input data and to the choice of the ML method and the features used in the ML model.[1] This means that an understanding of the physical and chemical relationships for the materials of interest is necessary for the construction of accurate ML models.

In this work, we seek to understand how ML can be used for predicting Curie temperatures of perovskite oxide ferroelectrics and for understanding and revealing their previously unknown important features. This is relevant because perovskite oxide ferroelectrics are an important class of materials with a wide variety of fascinating physical properties and applications that are controlled by the composition of the perovskite oxide solid solutions. In particular, we use the example of solid solutions of $PbTiO_3$ with Bi-based perovskite oxides (Bi-$PbTiO_3$). These solid solutions are currently some of the most promising candidates for obtaining high piezoelectric performance and high temperature operation stability due to their high Curie temperatures. The Curie temperature $T_c$ is one of the most important and widely reported properties of ferroelectric materials and because it is derived from the dynamics of the system, it is difficult to evaluate it directly using first-principles methods. Therefore, it would be useful to develop ML methods for accurate estimation of $T_c$ that could enable the rapid identification of new FE materials with high $T_c$ values suitable for high-temperature applications.

To obtain an accurate description of Bi-$PbTiO_3$ systems, we need a

N Corresponding authors.
*E-mail addresses:* emaskanazi@gmail.com (E.M. Askanazi), suhasyadav22@gmail.com (S. Yadav), ilya.grinberg@biu.ac.il (I. Grinberg).

https://doi.org/10.1016/j.commatsci.2021.110730
Received 29 March 2021; Received in revised form 14 June 2021; Accepted 13 July 2021
Available online 21 July 2021

flexible and powerful ML method, a fully descriptive set of features and a set of input data systems that fully characterizes the phase space of possible chemical and physical interactions in Bi-$PbTiO_3$ materials. Additionally, an effective measure of the accuracy of the predictive ML model is necessary. Failure to meet any one of these requirements will lead to poor prediction accuracy. Therefore, we seek to identify (i) the best machine learning method for use on this problem, (ii) the set of features that provide a full (or reasonably close to full) description of $T_c$ for this class of materials, (iii) the systems that fully explore the phase space of possible materials, and (iv) the accuracy metrics that should be used to obtain the most physically relevant results for our problem.

The question of what features and systems constitute a fully descriptive set for the problem of interest is related to the physical and chemical interactions in the target class of materials. Clearly, physically similar systems and similar features will not provide any valuable new information. Therefore, identification of different features that provide increased accuracy for ML predictions indicates that these features have different physical meaning. Identification of different classes of systems that provide increased predictive accuracy when they are included in the input data set implies that these classes display different chemical and physical interactions and properties.

For example, we do not expect $BaTiO_3$-based solid solutions to behave like $PbTiO_3$-based solid solutions whereas based on experimental data and general crystal chemistry arguments we know that Ba and Sr are similar to each other and Bi and Pb are similar to each other. This should be manifested by a weak prediction ability of a model trained on Pb/Bi-based system for solid solutions including $BaTiO_3$ and $SrTiO_3$. Conversely, an identification of Pb-based and Ba-based materials as different based on ML results would point to a physical difference between these systems that could then be related to the differences in their atomic properties such as electronegativity and filling of valence electron orbitals. Thus, in our work, in addition to obtaining a predictive model for the design of new high-T FE oxides, we also seek to use ML to confirm existing heuristics derived from experiment and first-principles calculations and to demonstrate additional, perhaps more subtle relationships and classifications for Bi-$PbTiO_3$ solid solutions.

Here, we first examine the $T_c$ prediction performance of different ML learning methods on the set of materials and features used in a previous ML study. Then, we expand the set of features to obtain improved predictive accuracy and identify the key features that determine the $T_c$ of this class of ferroelectric materials. The final model shows good predictive performance for a wide range of Bi-$PbTiO_3$ materials and identifies important features beyond those commonly used in previous work on the design of new ferroelectric Bi-$PbTiO_3$ materials.

## 2. Methodology

### 2.1. Machine learning methods and prediction accuracy measures

Here, we use three prominent ML methods, namely Support Vector Regression (SVR), K Nearest Neighbor (KNN) and Random Forest (RF), and analyze their predictive abilities and their sensitivity to the characteristics of the materials data used to optimize them.

SVR is a vector clustering learning technique that performs regression and classification of unlabeled nonlinear input data through kernel methods. SVR seeks to divide the input data into different classes by splitting the data with hyperplanes. The support vectors refer to the data points closest to the hyperplane that determine the position of the hyperplane. SVR finds global as opposed to local minima in modeling; additionally, it is less prone to overfitting than other ML methods and Deep Learning methods. Due to its advantages, SVR has traditionally been used for data sets with large numbers of input features relative to the number of data points. We utilize the radial basis function as a kernel function to map input data into a nonlinear space so that SVR can perform nonlinear regression. Previous studies implementing SVR obtained promising accuracy in predicting band gaps and melting points in binary, tertiary and quaternary compounds compared to the results obtained using neural networks and conventional regression models [7,8]. The drastic superiority of SVR to linear regression for the prediction of crystal structures of binary compounds has also been established [9].

KNN categorizes the input data points by finding which sets of neighboring data points have the closest proximity. When used for regression modeling, the KNN computes the average of the clusters designated as comprising the nearest neighbors of the given data point. For KNN, the only necessary input parameters are the number of neighboring points used to compute the averages of clusters for a data point and the distance function used to compute the nearest neighbors, However, this leads to high sensitivity to outliers and noise compared to SVR and RF.

RF models are an ensemble learning method. They split the data into classification through a creation of a network of decision trees to determine how to partition the data into component categories. Independently sampled random input data vectors are fitted to each of the decision trees formed in the procedure. The trees are formed from a root node, layers of split or internal nodes and terminal or leaf nodes containing the final regression values. Each node in the decision tree is obtained from the split among a random set of input vector features that allows for the most accurate regression and/or classification. When using RF for regression models, the predictions of all decision trees are averaged for the final fit. RF has been demonstrated to be a useful technique for the application of ML in materials science in the prediction of the metal alloy crystal structures [10] and in the prediction of the energies of absorption on metal alloys [11]. RF models are a popular method because the way in which they classify and generate regression models is particularly suitable for preventing the overfitting of data with large numbers of input features. This is due to their use of large numbers of decision trees. They are also known for being able to treat moderately small data sets better than other Machine and Deep Learning models.

For the SVR and RF methods, the hyperparameters governing how these methods break down the data can be fully optimized. This is done through K-fold cross-validation. A given data set is split into 10 subsets; one subset is taken for a test set within this k-fold input set and the other 9 are chosen as a training set. The hyperparameters are optimized for the training set, and evaluated on the test set and the process is repeated for all 10 subsets until the optimal hyperparameters are found. These subsets are formed from randomly generated samples of data taken without replacement; therefore, bootstrap aggregation is not used. The search for the optimal hyperparameters for each iteration can be done through Grid Search or Random Search. For SVR, the hyperparameters are cost, gamma and epsilon. The cost refers to the penalty for training errors in the SVR regression and effectively controls the tradeoff between the training error and model complexity. Gamma is the kernel function coefficient and epsilon determines the accuracy of the fitting method. These hyperparameters are optimized through Grid Search. For RF, the hyperparameters are the number of trees in the forest, maximum depth of the decision trees in the forest, the number of features to consider when looking for splitting, the minimum samples required to split an internal node and to be in a leaf node and whether or not bootstrap sampling is used. Due to a substantially larger number of possible hyperparameter combinations for RF, Randomized Search was used to optimize the RF hyperparameters. The hyperparameters for the various methods are presented in Tables 1 and 2

To measure the accuracy of the prediction of the different ML methods, we use the standard mean absolute error (MAE) measure that has been widely used in various ML studies. However, MAE is an aggregate measure of accuracy that may not reflect the properties of interest for research in ferroelectric materials. Furthermore, for the large range of $T_c$ values examined here, an MAE of 50 K may appear to indicate a fairly accurate prediction. In fact, such an MAE often results in a large scatter in the range of $T_c$ of interest to FE researchers that would make the ML prediction useless for providing guidance for the

**Table 1**
Hyper parameters for various Machine Learning Methods trained on the regression data set with data points taken from Balachandran et al. [14].

| Machine Learning Method | Hyper Parameters |
|---|---|
| SVR Default | γ = 0.008, ε = 0.1 and C = 1 |
| SVR Hyperparameters Optimized Without B-site element Valence | γ = 5, ε = 10 and C = 1000 |
| SVR Hyperparameters Optimized With B-site element Valence | γ = 0.005, ε = 5 and C = 1000 |
| RF Default | # of Trees = 100, Maximum # of Features= # of Input Data Points, Maximum Tree Depth = None, Minimum # of Points to Split Node = 2, Minimum # of Points Allowed in Node = 1, Bootstrap Sampling = True |
| RF Hyperparameters Optimized Without B-site Element Valence | # of Trees = 125, Maximum # of Features = Square Root of # of Input Data Points, Maximum Tree Depth = 60 Node Splits, Minimum # of Points to Split Node = 2, Minimum # of Points Allowed in Node = 1, Bootstrap Sampling = False |
| RF Hyperparameters Optimized With B-site Element Valence | # of Trees = 38, Maximum # of Features= # of Input Data Points, Maximum Tree Depth = 50 Node Splits, Minimum # of Points to Split Node = 2, Minimum # of Points Allowed in Node = 1, Bootstrap Sampling = False |

**Table 2**
Hyperparameters for various Machine Learning Methods trained on the expanded feature regression data set with the input data taken from [23-29].

| Machine Learning Method | Hyper Parameters |
|---|---|
| SVR Default | γ = 0.008, ε = 0.1 and C = 1 |
| SVR Hyperparameters Optimized Without B-site element Valence | γ = 5, ε = 10 and C = 1000 |
| SVR Hyperparameters Optimized With B-site element Valence | γ = 0.005, ε = 5 and C = 1000 |
| RF Default | # of Trees = 100, Maximum # of Features= # of Input Data Points, Maximum Tree Depth = None, Minimum # of Points to Split Node = 2, Minimum # of Points Allowed in Node = 1, Bootstrap Sampling = True |
| RF Hyperparameters Optimized Without B-site Element Valence | # of Trees = 137, Maximum # of Features = Square Root of # of Input Data Points, Maximum Tree Depth = 30 Node Splits, Minimum # of Points to Split Node = 2, Minimum # of Points Allowed in Node = 1, Bootstrap Sampling = False |
| RF Hyperparameters Optimized With B-site Element Valence | # of Trees = 82, Maximum # of Features= # of Input Data Points, Maximum Tree Depth = 90 Node Splits, Minimum # of Points to Split Node = 2, Minimum # of Points Allowed in Node = 1, Bootstrap Sampling = False |

identification of new solid solutions for synthesis.

Thus, since investigations and design of new FE materials have been focused on the design of new high-temperature solid solutions, a measure that evaluates the accuracy of the ML method in the high-temperature range (700–850 K) is necessary. Therefore, to examine the suitability of our ML models for the design of high-$T_c$ ferroelectric materials, we fit a linear regression to the predicted $T_c$ values plotted versus experimental $T_c$ values in the 700–850 K range and use the slope, the intercept and the $R$ correlation coefficients as measures of prediction accuracy. For a perfectly accurate predictor, the slope, intercept and $R$ will be equal to 1, 0 and 1, respectively. Deviation of the slope and the intercept from these values indicates a possibly systematic error of the model predictions compared to experimental results.

### 2.2. Input data sets

The regression-based predictions for the $T_c$ values of desired materials may depend strongly on the selection of the materials used in the input data set and the set of features used to characterize the materials in the input data set. In this analysis, we focus on the $x$BiMe′$_{1-y}$Me″$_y$O$_3$-(1 − $x$)PbTiO$_3$ solid perovskite solutions [12]. In perovskite solutions the Bi-based end member is often itself a solid solution with two B-site cations (e.g. $BiMg_{2/3}Nb_{1/3}O_3$) that we designate as Me′ and Me″. For these solutions, 1-$y$ and $y$ are the stoichiometric coefficients for the Me′ and Me″ cations. Previous data analysis of such solid solutions used classification to identify solutions that can form in the perovskite structures and used regression to identify the candidate high $T_c$ solutions [13-15]. Composition, tolerance factor, ionic displacement ($D$), the valence (oxidation state or formal charge) of the constituent B-site metal ions, the Born effective charge (Z*), the ratios of the Me′ and Me″ in the Bi end-member and polarization ($P$) values (defined as $P = \Sigma_i (x_i Z_i D_i)$ over ion index $i$ where $Z_i$ is the $i$th ion's Z* and $D_i$ is its ionic displacement) are all fundamentally important attributes of the input data that will contribute to the ability of the ML model to make accurate regression-based $T_c$ predictions. Tolerance factor and ionic displacement have been established as being able to predict $T_c$ to some degree in simple linear regression modeling [16-19]. The relationship between the Curie temperature and the atomic displacement from the high-symmetry positions structure has been well established [19-21] and it was suggested that for some BiMe′Me″O$_3$-PT compositions, $T_c$ depends on the tolerance factor [12].

Here, we first use the set of compositions and features used in the previous ML learning of high-$T_c$ Bi-PbTiO$_3$ ferroelectrics performed by Balachandran et al. [14] This data set contains 135 ferroelectric compounds. For this data set, we also add the valences of Me′ and Me″ as the features for model training. This will distinguish between the BiMe′$_{1-y}$Me″$_y$O$_3$ compounds where Me′ and Me″ have the same valence of + 3 (for example, $BiScO_3$ and $BiSc_{0.5}Fe_{0.5}O_3$) and the compounds where Me′ and Me″ have different valences (e.g. + 2 and + 4 such as $BiMg_{0.5}Zr_{0.5}O_3$). As was found by density functional theory calculations, [22] these two classes of Bi-PbTiO$_3$ ferroelectric solid solutions show different behaviors, suggesting that valence of the substituent Me′ and Me″ may be an important feature.

Additionally, we constructed a second data set with an expanded set that contains 120 Bi-PbTiO$_3$ ferroelectric systems [23-29]. Unlike the dataset of Balachandran et al. that uses a product of A-site and B-site ionic displacements, we characterize each ferroelectric system using the displacement ($D$) parameters for the different metal elements derived from density functional theory calculations in a previous work [15] for the A-site and B-site separately and also use the Born effective charge Z* and valence (oxidation state) of the B-site cations. In addition to containing a tolerance factor that is based on the ratio of the A-site and B-site ionic sizes of the Bi-based end-member (specifically, using a tolerance factor defined as $t_{f,BiMe'1-yMe''yO3} = (r_A + r_O)/(\sqrt{2}(r_B + r_O))$ where $r_A$ , $r_B$ and $r_O$ are the average ionic radii of the A-Site, B-Site and O atom in BiMe′$_{1-y}$Me″$_y$O$_3$), this input set also uses individual ionic sizes of the A-site and B-sites as features. This allows us to examine whether the individual ionic sizes of the A-site and B-site are important for $T_c$ or whether they merely serve to set the tolerance factor of the system. Thus, a richer set of features is available for the second data set.

For all systems, the $T_c$ values are taken from experimental studies as evaluated by dielectric constant measurements. When evaluating the capabilities of the Machine Learning models, a given regression data set is divided such that 90% of the input points are used for training and 10% are used for the test sets. Additionally, we create a set of 23 compounds [30-39] that are not included in the input data for validation testing of the obtained ML models. The regression data sets are available as supplementary files and are designated BalachandranRegression for the Balachandran dataset and ExpandedRegression for the expanded set

of features regression data set.

## 3. Results and discussion

We first examine ML models trained on the Balachandran data set using the three features of Bi content, tolerance factor and ionic displacement and then examine how the addition of the valences of the Me′ and Me″ atoms in $BiMe'_{1-y}Me''_yO_3$ as features affects the accuracy of ML prediction. The average + 3 valence of the B-site of the Bi-based end-member can be obtained with + 3 cations as in $BiScO_3$ and $BiFeO_3$ or $BiSc_{1-y}Fe_yO_3$, or as a combination of + 2 and + 4 cations as in $BiMg_{1/2}Ti_{1/2}O_3$ and $BiMg_{1/2}Zr_{1/2}O_3$, or as a combination of + 2 and + 5 cations as in $BiMg_{2/3}Nb_{1/3}O_3$ or as a combination of + 2 and + 6 cations as in $BiMg_{3/4}W_{1/3}O_3$. In principle, the different combinations of valences (formal charges) on the B-site of the Bi-based end-member may have different effects on $T_c$ of the solid solution and therefore the valence of the individual Me′ and Me″ cations should be included as features. For the SVR and RF models, we use both models with hyperparameter tuning and models with the default values of hyperparameters.

Fig. 1 shows the MAE values for the training and test set of the KNN, SVR and RF ML models. It is observed that without hyperparameter tuning the MAE is in the order of SVR, KNN, RF. Parameter tuning strongly increases SVR accuracy so that the order changes to KNN, SVR, RF. The superiority of RF is quite strong, with the MAE for RF equal to half of the MAE for the hyperparameter-tuned SVR and one third of the MAE of KNN. Based on the results of our analysis of structure–property trends in $PbTiO_3$-based ferroelectric solid solutions [40], we then include the valence of the Me′ and Me″ atoms in $BiMe'_{1-y}Me''_yO_3$ as a feature in addition to Bi content, tolerance factor and displacement. This will distinguish between the $BiMe'_{1-y}Me''_yO_3$ compounds where Me′ and Me″ have the same valence of + 3 (for example, $BiScO_3$ and $BiSc_{0.5}Fe_{0.5}O_3$) and the compounds where Me′ and Me″ have different valences (e.g. + 2 and + 4 such as $BiMg_{0.5}Zr_{0.5}O_3$). As found by DFT calculations, these two classes of Bi-$PbTiO_3$ ferroelectric solid solutions show different behaviors, suggesting that valence of the substituent Me′ and Me″ may be an important feature. With the inclusion of Me′ and Me″ valence feature, we obtain a reduction in the MAE values by a factor of ≈2 for all three methods. The order of accuracy is still unchanged with RF showing the best results and reaching MAE as low as 13 K. Such superiority of RF for small data sets has been noted in previous work. [41-43]. The strong improvement in the MAE due to the inclusion of the Me′ and Me″ valences for all three methods indicates that Me′ and Me″ valences play an important role in determining $T_c$; this has not been appreciated in previous ML studies. Examination of the slope, intercept and $R$ values for the region of interest with $T_c$ values between 700 and 850 K clearly shows the superiority of the RF model, even when the Me′ and Me″ valence information is included in the fitting, because only RF obtains the slope, intercept and $R^2$ values close to 1, 0 and 1, respectively. By contrast, SVR shows large intercept and low slope value even when Me′ and Me″ valences are included as features, and all other methods show even worse performance. The substantial dependence of SVR on parameter optimization shows a relative lack of robustness of the SVR model when trained on this particular set of input data features. Thus, our results show that RF is the best method for the prediction of $T_c$ of these materials. Our results also demonstrate the importance of going beyond the MAE for the evaluations of the quality of the ML model predictions. As shown in Fig. 2, the MAE of 40 or higher implies a poor ability for overall $T_c$ prediction and essentially zero predictive ability for

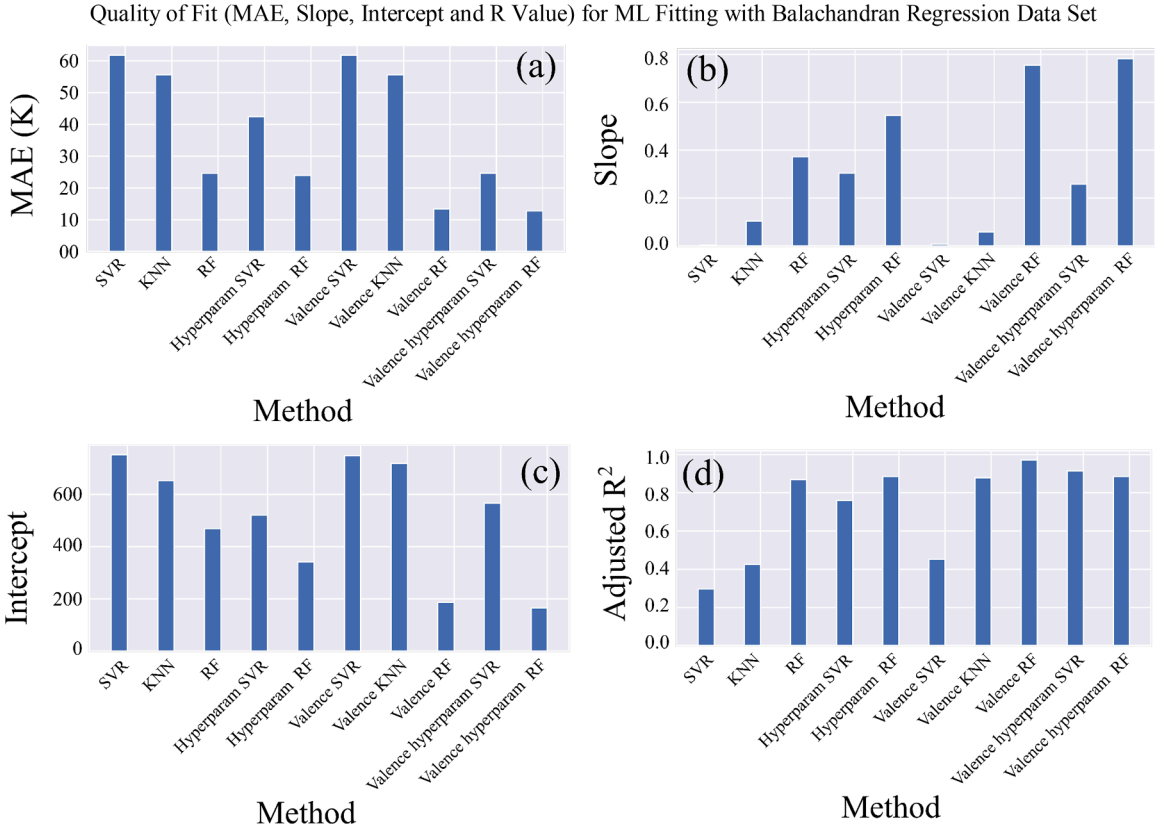


**Fig. 1.** 1a) MAE for the Balachandran et al. test data set [13] using initial features, for the five left bar values, and initial features and valence of Me′ and Me″ for the five right bar values. Initial features are defined as composition ($x$) along with solid solution $t_f$ and ionic displacement (D), both of which are defined [13]. Hyperparam refers to fitting of the test data using fine-tuned hyperparameters. 1b) Slope for the 700–850 K range for the different ML methods using the same input features 1c) y-intercept values for the 700–850 K range for the different ML methods using the same input features 1d) $R^2$ values of the fit in the 700–850 K range for the different ML methods using the same input features.

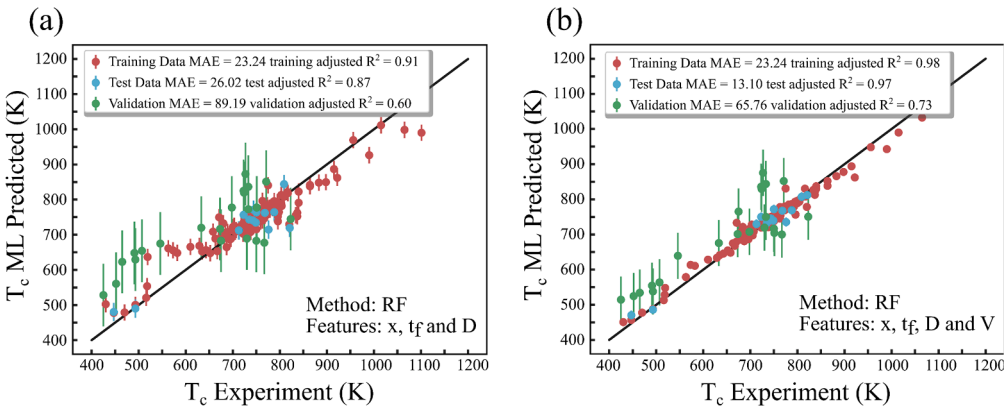


**Fig. 2.** Quality of Machine Learning Prediction for Training and Test Set for the Balachandran et al. dataset [13]. The RF method was used for ML fitting. Initial Features are defined as composition ($x$) along with solid solution $t_f$ and ionic displacement (D) defined in [13]. The RF method was used for ML fitting. Hyperparameters were optimized using 10-fold cross validation. Training data (red) and test data (blue) refer to the training and test data sets randomly generated from splitting of the input regression data. Test compounds (green) refers to the validation test set. The MAE is shown in the error bars. a) Uses initial Features 2b) Uses initial features and the valence (V) of Me′ and Me″.

the 700–850 K region of interest, while the MAE of 10–20 implies excellent overall $T_c$ prediction ability and moderate accuracy for the 700–850 K region of interest.

We now examine the performance of the different ML methods in more detail using the plots of Experimental vs Predicted $T_c$ (Fig. 2). For the RF method, Fig. 2a show that using the Bi content, tolerance factor and displacement, the training set $T_c$ values are reproduced well in the entire $T_c$ range. For the test set, the overall $T_c$ values are also reproduced well, but the predictive ability of the ML model in the 700–850 K range is relatively poor, with the linear fit through the test points in this range showing a slope and intercept that significantly deviate from 1 and 0, respectively. By contrast, inclusion of the Me′ and Me″ valences leads to an essentially exact prediction of $T_c$ even for the test set data in the 700–850 K as shown in Fig. 2b.

Examination of the prediction quality for the validation set shows much worse results for some $BiSc_yFe_{1-y}O3$-$PbTiO_3$ compounds and for the $BiNi_{3/4}W_{1/4}O_3$-$PbTiO_3$ compounds. For predicting the $T_c$ values for the validation test set, there is an inherent difficulty in obtaining accurate values for compounds with sufficiently different input features relative to the regression data set. The validation test contains $BiSc_yFe_{1-y}O_3$-$PbTiO_3$ compounds with x = 0.2 and 0.3 (Me′ and Me″ coefficient values of 0.2 or 0.3 for Sc and of 0.7 or 0.8 for Fe). By contrast, the training set only contains $BiSc_yFe_{1-y}O_3$-$PbTiO_3$ with x = 0.5. Thus, the training set does not have sufficient information regarding the influence of the change in the Sc/Fe ratio for accurate prediction. Therefore, our ML model cannot properly evaluate the $T_c$ values for $BiSc_yFe_{1-y}O_3$-$PbTiO_3$ compounds with y = 0.2 and 0.3 and makes relatively inaccurate predictions.

To study the relative importance and the contributions of the Bi-content, tolerance factor ($t_f$), displacement and the Me′ and Me″ valence features, we first examine the feature importance of the final RF model and then evaluated the MAE of the RF models limited to different

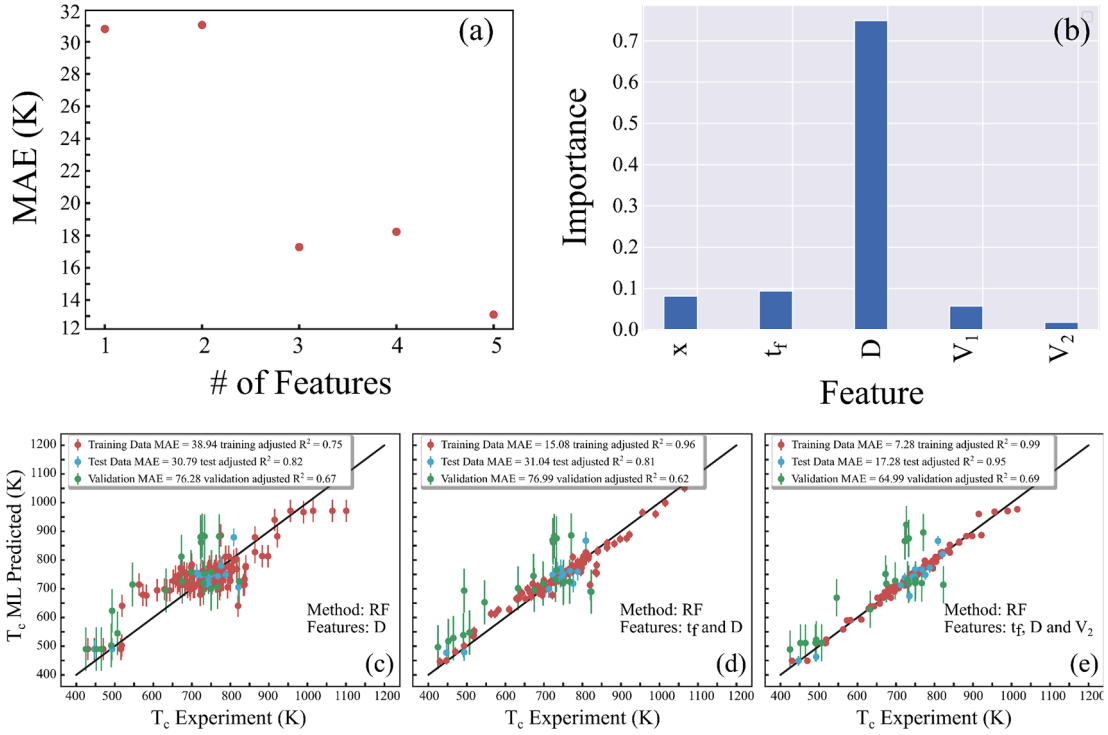


**Fig. 3.** Feature Plots the Balachandran et al. dataset [13]. The RF method was used for ML fitting. 3a. Error vs number of features when optimal feature set is chosen for each # of features 3b Feature Importance for composition ($x$) along with solid solution $t_f$ and ionic displacement (D), both of which are defined in [13] and valence of Me′ and Me″. 3c Quality of machine learning prediction for one optimal feature 3d Quality of machine learning for two optimal features 3e quality of machine learning for three optimal features.

numbers of features (1–5) with the results shown in Fig. 3. As shown in Fig. 3a, the test set MAE values for RF models with 1 and 2 features only are 31 K and the MAE then drops to 17 K for RF models with 3 features and to 13 K for the full model (Fig. 2). It is observed from Fig. 3b that displacement is the dominant feature with × and $t_f$ as secondary features and the valences of the Me′ and Me″ atoms ($V_1$ and $V_2$) as the least important features. The plots of experimental versus predicted Tc for these RF models with limited numbers of features are shown in Fig. 3c-e. For the varying feature numbers, backward sequential feature selection, a greedy feature algorithm, was utilized. The displacement and tolerance factor were shown to be the two most important features whereas $V_2$ was shown to be the third most important. This, along with the results of fitting with and without $V_1$ and $V_2$ (Fig. 2) shows that features which are relatively less important when included does not always reveal predicting ability when said features are excluded. This is likely because in this feature set the Me′ and Me″ valence information acts as a sort of numerical label for the data. Therefore, it is important to include them for a high quality fit because they allow for separation of data based on a distinct ionic feature of Me′ and Me″. However, when they are included in the fitting they are not quantified as important features due to their lack of numerical variance relative to the other features. Examination of these figures also shows that the MAE of 31 corresponds to only qualitative accuracy for $T_c$ prediction while MAE of 10–20 implies excellent overall $T_c$ prediction ability and moderate accuracy for the 700–850 K region of interest. Therefore, our results show that inclusion of displacement is crucial for achieving even moderate predictive ability and the use of other features including the Me′ and Me″ valences is necessary for quantitatively accurate prediction in the 700–850 K range.

To summarize, our results for the Balachandran data set show that the combination of RF method and the four features of Bi content, displacement, $t_f$ and valence are necessary to obtain quantitative accuracy. The use of SVR or KNN method leads to only qualitatively accurate predictions even with all the necessary features. The omission of the displacement features makes the model incapable of achieving even qualitative accuracy regardless of the ML method utilized.

We then examined the ML performance set of the Bi-PTO systems with a broader set of features. As shown in Fig. 4, here as well, RF shows the best performance while KNN shows the worst performance with SVR in between. Hyperparameter optimization has a huge effect on SVR accuracy, converting SVR from essentially useless to quantitatively accurate and also somewhat increases the accuracy of RF. The same lack of robustness seen with SVR for the previous set of input features is therefore apparent with this feature set. Unlike for the Balachandran data set, the addition of Me′ and Me″ valences to the initial feature set does not improve the performance of SVR and RF, indicating that the information provided by the Me′ and Me″ valences is already contained in the initial feature set. It is likely that Me′ and Me″ valence information is already contained in the Z* and in the stoichiometric coefficients of the Me′Me″ in our data set and therefore, addition of explicit valence data does not improve the quality of the prediction results. This is supported by the correlation matrix of the different features presented in Fig. 5 that shows a strong correlation between the Me′ and Me″ coefficients and valence. The similar MAE values obtained for our data set and the Balachandran data set when a complete set of input features characterizing the compounds is incorporated indicate that both sets of systems and both sets of features provide a close to full description of the Bi-PTO solid solutions

Next, we examine the predictive ability of different features for the experimental $T_c$ values (Fig. 6). We find that Bi content × , average displacement $D$ and polarization $P$ are the dominant features for the final RF model. For determining prediction accuracy with a select number of features (6c-6e) backward sequential feature selection was again used. For the RF models with limited number of features, the test set $T_c$ prediction accuracy as measured by the MAE rapidly improves as the

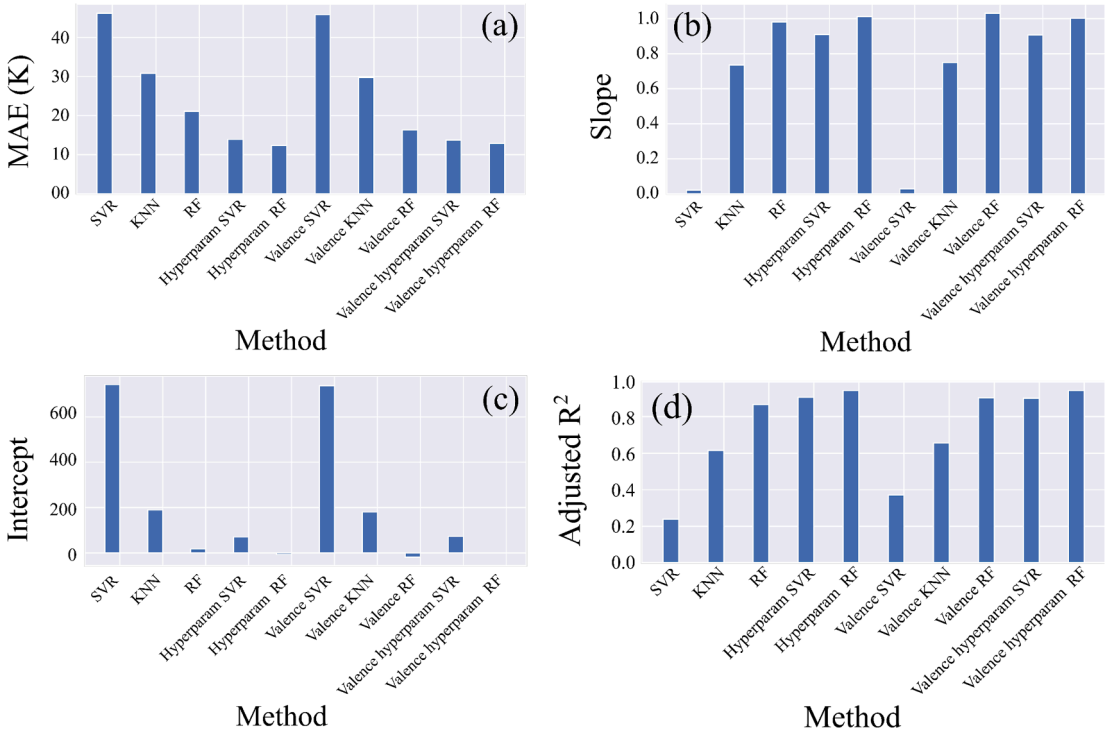


**Fig. 4.** 4a) MAE for the expanded set of features regression data set [23-29] using an input feature set comprising × , $t_f$, BiMe′1-yMe″yO3 ($t_{f,BiMe'1-yMe''yO3} = (r_A + r_O)/(\sqrt{2}(r_B + r_O))$), ionic radii, ionic displacement, Born effective charges $Z^*_i$ and coefficients of Me′ and Me″ content, average ionic radii ($R_{avg} = R_1C_1 + R_2C_2$), P where $P = \Sigma_i (x_i Z_i D_i)$ where $D_i$ are ionic displacements) and average ionic displacement ($D_{avg} = D_1C_1 + D_2C_2$) for the five left bar values, and the input feature set plus valence of Me′ and Me″ for the five right bar values. Hyperparam refers to fitting of the test data using fine-tuned hyperparameters. 4b) Slope in the 700–850 K range for the different ML methods using the same input features 4c) y-intercept values in the 700–850 K range for the different ML methods using the same input features 4d) R values for the fit in the 700–850 K range for the different ML methods using the same input features.

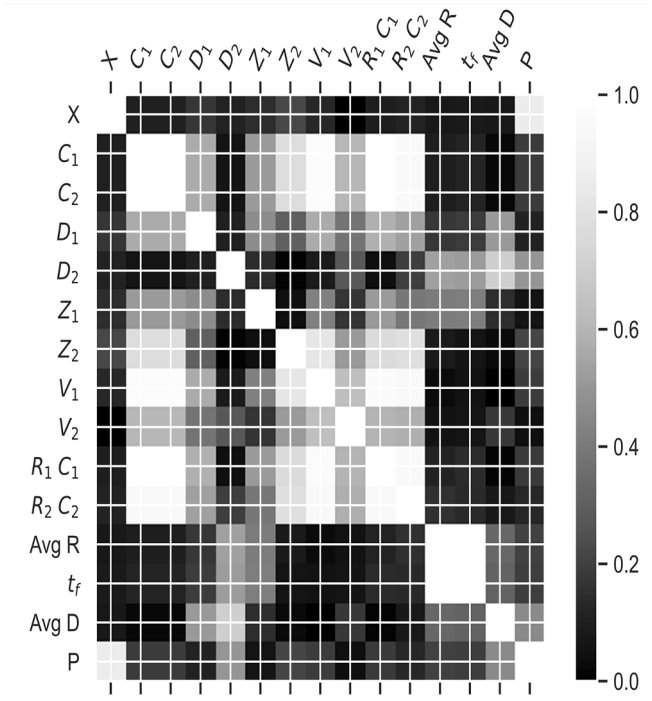


**Fig 5.** Correlation matrix of input features for the expanded feature test data set [23-29]. The RF method was used for ML fitting. XPTO (The fraction of $PbTiO_3$ present in the $BiMe'Me''O_3$-PT compounds, designated X in the matrix), Me′ and Me″ stoichiometric coefficients ($C_1$ and $C_2$), Ionic Displacements ($D_1$ and $D_2$), Born effective ion charges Z and Radii of Me′ and Me″ ($R_1$ and $R_2$), Average Ionic Radii (Avg R), Average Ionic Displacement (Avg D), valence of ions ($V_1$ and $V_2$), tolerance factor ($t_f$) and P where $P = \Sigma_i (x_i Z_i D_i)$.

number of features increases from 1 (45 K) to 2 (20 K). The use of additional features leads to further improvement of only about 5 K in the test MAE until the full set of features is used, obtaining MAE of 12 K. We attribute the high predictive power of the two-feature RF model here to the inclusion of the *P* that contains information about the ionic contents and Me′ and Me″ valences because *P* is defined as the $\Sigma_i x_i Z_i D_i$ where *i* runs over all cation species in the system. The importance of *P* confirms the previously found empirical relationship between *P* and $T_c$ of ferroelectric solid solutions. [20,21]. In the 11 feature set (6e) $D_{avg}$, *x*, $V_1$ and $V_2$ also suffice for accurate predictions since information about ionic content and valence is also contained. This is shown in the strong correlation of $D_{avg}$ and *x* with P (Fig. 5).

While essentially exact accuracy is obtained for the test set $T_c$ values by our RF model, the validation set still shows some errors for the $BiSc_{1-y}Fe_yO_3$-$PbTiO_3$ systems and particularly large errors are obtained for the $BiNi_{3/4}W_{1/4}O_3$-$PbTiO_3$ systems. Due to the similarity of the input features (*D*, *R*, *V*, *Z**) for Ni and Mg, the $T_c$ values for the $BiNi_{3/4}W_{1/4}O_3$-$PbTiO_3$ systems are predicted to be similar to those of the $BiMg_{3/4}W_{1/4}O_3$-$PbTiO_3$ compounds. However, experimentally $BiNi_{3/4}W_{1/4}O_3$-$PbTiO_3$ systems show much lower $T_c$ values. We note that our expanded feature set contains several $BiNi_{1/2}Ti_{1/2}O_3$-$PbTiO_3$ and $BiNi_{1/2}Zr_{1/2}O_3$-$PbTiO_3$ systems for which $T_c$ values in good agreement with experiment are obtained by the ML model. This suggests that the large Ni content ($Ni_{3/4}$) gives rise to a novel behavior in the $BiNi_{3/4}W_{1/4}O_3$-$PbTiO_3$ systems that is absent in other solid solutions. The origin of this effect is unclear and points out the need for a further investigation of the $BiNi_{3/4}W_{1/4}O_3$-$PbTiO_3$ systems and the incorporation of additional input features into ML learning for accurate prediction of Curie temperatures.

We then investigated the transferability of the ML models between the different parts of the phase space of Bi-PTO solutions. Since we found that Me′ and Me″ valences play an important role in predicting $T_c$, we divided the Bi $Me'Me''O_3$-PTO systems into those with homovalent Me′

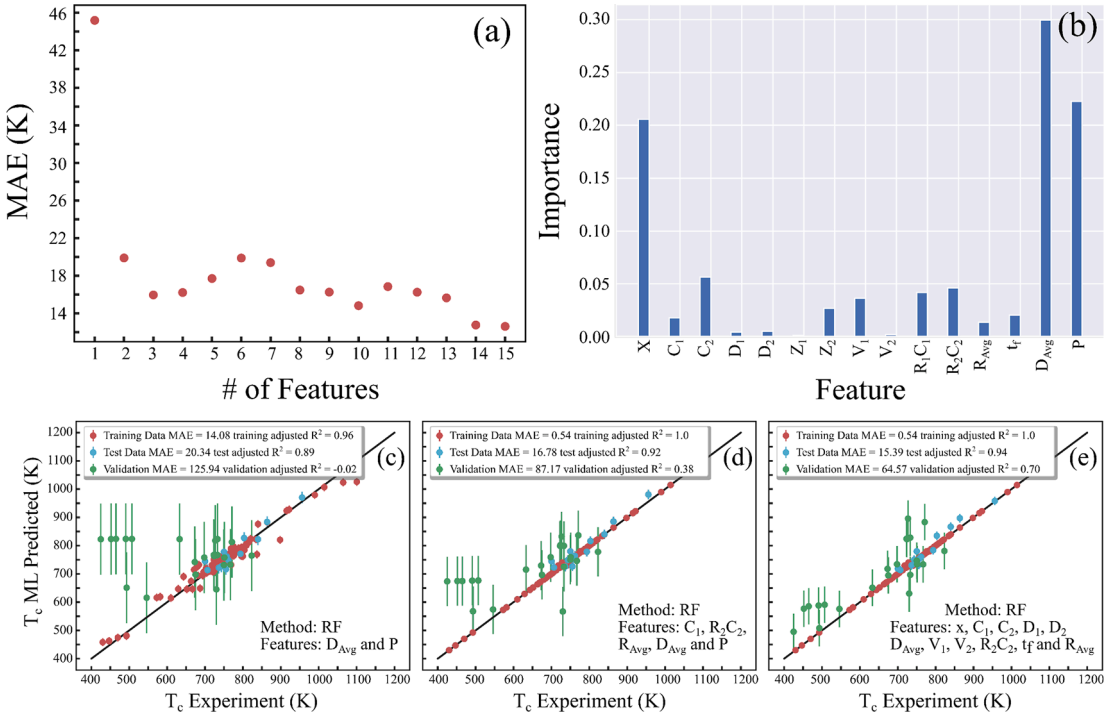


**Fig. 6.** Feature Plots for the expanded feature dataset [23-29]. The RF method was used for ML fitting. The RF method was used for ML fitting. 6a. Error vs number of features when optimal feature set is chosen for each # of features 6b Feature Importance for the input set comprising × , $t_{BiMe'1-yMe''yO3}$ ($t_{f\,BiMe'1-yMe''yO3} = (r_A + r_O)/(\sqrt{2}(r_B + r_O))$) ionic displacement and radii of Me′ and Me″ ($D_1$, $D_2$, $R_1$ and $R_2$), Born effective ion charges $Z_i$, and Me′ and Me″ stoichiometric coefficients ($C_1$ and $C_2$), average ionic displacement ($Davg = D_1C_1 + D_2C_2$) and average ionic radii ($Ravg = R_1C_1 + R_2C_2$) and P where $P = \Sigma_i (x_i Z_i D_i)$ over ion index i and valence of Me′ and Me″. 6c Quality of machine learning prediction for two optimal features 6d Quality of machine learning for five optimal features 6e quality of machine learning for eight optimal features.

in the Bi-based end member (e.g. $BiScO_3$ and $Bi(Sc_{1-y}In_y)O_3$) and heterovalent Me′Me″ Bi-based end member (e.g. $BiMg_{1/2}Zr_{1/2}O_3$ and $BiZn_{3/4}$ $W_{1/4}O_3$). We then created two test sets, one containing both homovalent and heterovalent Bi endmember solutions (the broad set) and another that only contains the homovalent Bi-based endmembers (narrow set). Then, we divided the remaining solid solution systems into different groups and used them for training the KNN, SVR and RF ML models. Fig. 7a-7c show the MAE of the $T_c$ values predicted for the broad set by the ML models trained on heterovalent solid solutions, mixed homovalent and heterovalent solutions, and homovalent solutions, respectively. Fig. 7d-7f show the MAE of the predicted $T_c$'s for the narrow ($Bi(Sc,In)O_3$) set for the ML models trained on heterovalent solid solutions, mixed homovalent and heterovalent solutions, and homovalent solutions, respectively. Examination of Fig. 7a-c shows that SVR gets the best results for 7b and 7c and the second best, behind RF, for case 7a. This is most likely because of the relative absence of overfitting obtained for the SVR and RF. Due to the small training data sets, overfitting is likely and will have a strong effect on the test set prediction quality. Comparison of the SVR results in Fig. 7a-c shows that the MAE obtained using the homovalent input only is significantly worse than the MAE obtained using heterovalent and mixed input data. When using a homovalent test set, the MAE was much smaller for three training sets and surprisingly was the largest for the homovalent training set. These results reveal that homovalent and heterovalent $BiMe'Me''O_3$ end member show different behaviors so that both kinds of systems must be included for predicting the full range of Bi-PTO solid solutions.

## 4. Conclusion

Our study of the application of machine learning for the prediction of $T_c$ of Bi-PTO systems revealed that application of ML techniques to ferroelectric oxides is not trivial and obtaining accurate results requires the correct choice of the ML method, features and input systems. We found that among the studied KNN, SVR and RF methods, RF obtains the best performance and is insensitive to the choice of hyperparameters. SVR results show a strong sensitivity to the choice of hyperparameters and SVR obtained $T_c$ predictions with significantly lower accuracy than RF even after hyperparameter optimization. KNN results show poor accuracy and are essentially unusable with an incomplete feature set and are only qualitatively accurate with a complete feature set. Regarding the choice of features for characterization of the FE systems, we find that Bi content and B-cation valences, ionic radius and ionic displacements form the essential set of features needed for quantitatively accurate $T_c$ predictions for most Bi-$PbTiO_3$ ferroelectric solid solutions. Nevertheless, the prediction of $T_c$ values for Ni-containing materials is still challenging. This is most likely due to limitations of the feature sets used to predict $T_c$ values. We find that homovalent and heterovalent $BiMe'Me''O_3$-$PbTiO_3$ solid solutions form distinct classes of materials with different behaviors. Therefore, homovalent and heterovalent $BiMe'Me''O_3$-$PbTiO_3$ solid solutions must be both included in the input data set to obtain high predictive accuracy. Overall, our work confirms that for the small data sets typically available in materials science, proper identification of input systems, features and ML methods is necessary to obtain qualitatively and quantitatively accurate regression models and reveal connections between Curie temperature and key characteristics of materials. However, in most cases, even with limited available materials data, this can be achieved in a systematic manner. We believe that these conclusions are generally applicable to other properties of ferroelectric oxides and to other classes of functional materials.

The raw data required to reproduce these findings are available to download from [13-1523-39].

The processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

## CRediT authorship contribution statement

**Evan M. Askanazi:** Conceptualization, Methodology, Formal analysis, Investigation. **Suhas Yadav:** . **Ilya Grinberg:** Conceptualization, Methodology, Funding acquisition, Investigation.

## Declaration of Competing Interest

The authors declare that they have no known competing financial

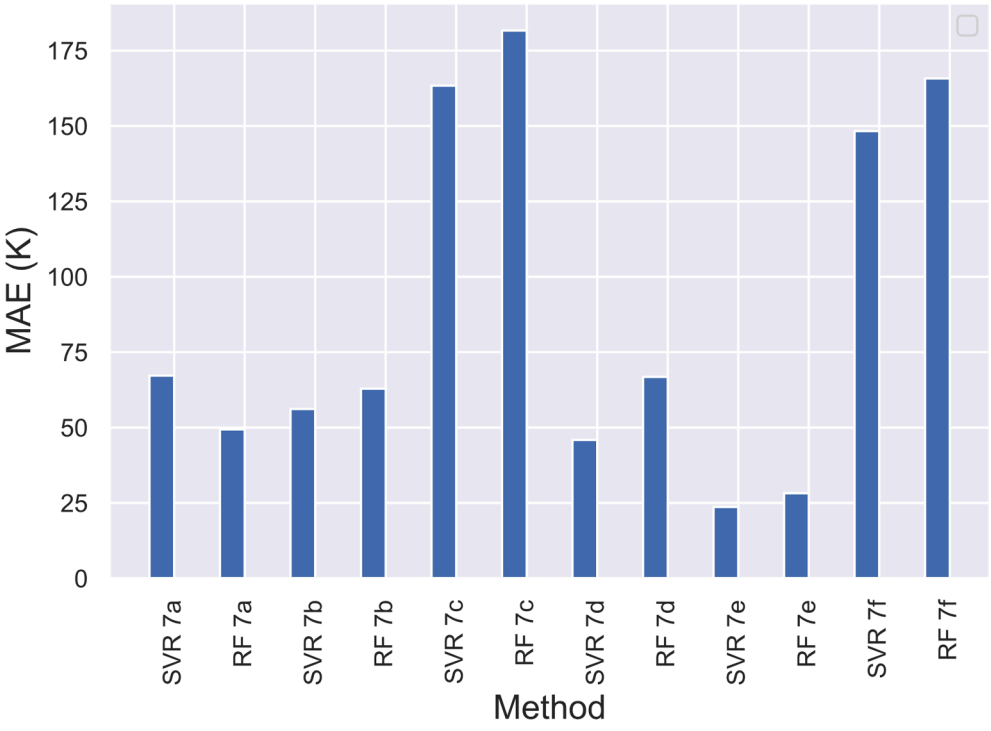


**Fig. 7.** MAE for SVR, RF, KNN, SVR and RF for expanded test data set [23-29] using optimized hyperparameters is displayed for various subsets of regression data used as training and test sets. The RF method was used for ML fitting. 7 a-c shows MAE for predictions using $BiScFeO_3$-$PbTiO_3$, $BiZnWO_3$-$PbTiO_3$, $BiZnZrO_3$-$PbTiO_3$ and $BiMgZrO_3$-$PbTiO_3$ as test set. Fig. 7a shows MAE using $BiZnTiO_3$-$PbTiO_3$, $BiMgWO_3$-$PbTiO_3$, BiMgNbO3-$PbTiO_3$ and $BiZnNbO_3$-$PbTiO_3$ as input data. Fig. 7b displays MAE using $BiMgTiO_3$-$PbTiO_3$, $BiNiTiO_3$-$PbTiO_3$, $BiGaScO_3$-$PbTiO_3$ and $BiNiNbO_3$-$PbTiO_3$. Fig. 7c provides MAE using $BiCoFeO_3$-$PbTiO_3$, $BiFeO_3$-$PbTiO_3$, $BiMnO_3$-$PbTiO_3$, $BiAlO_3$-$PbTiO_3$, $BiLuO_3$-$PbTiO_3$, $BiGaO_3$-$PbTiO_3$, $BiInO_3$-$PbTiO_3$ and $BiScO_3$-$PbTiO_3$. Fig. 7 d-f gives MAE values for predictions using $BiScInO_3$-$PbTiO_3$ as test set. Fig. 7d is using $BiZnTiO_3$-$PbTiO_3$, $BiMgWO_3$-$PbTiO_3$, $BiMgNbO_3$-$PbTiO_3$ and $BiZnNbO_3$-$PbTiO_3$ as input data. Fig. 7e is using $BiMgTiO_3$-$PbTiO_3$, $BiNiTiO_3$-$PbTiO_3$, $BiGaScO_3$-$PbTiO_3$ and $BiNiNbO_3$-$PbTiO_3$. Fig. 7f is using $BiCoFeO_3$-$PbTiO_3$, $BiFeO_3$-$PbTiO_3$, $BiMnO_3$-$PbTiO_3$, $BiAlO_3$-$PbTiO_3$, $BiLuO_3$-$PbTiO_3$, BiGaO3-$PbTiO_3$, $BiInO_3$-$PbTiO_3$ and $BiScO_3$-$PbTiO_3$.

interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by the Israel Science Foundation.

## References

[1] J. Schmidt, M.R.G. Marques, S. Botti, et al., Recent advances and applications of machine learning in solid-state materials science, npj Comput. Mater. 5 (2019) 83, https://doi.org/10.1038/s41524-019-0221-0.
[2] P. Raccuglia, K.C. Elbert, P.D.F. Adler, C. Falk, M.B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, A.J. Norquist, Machine-learning-assisted materials discovery using failed experiments, Nature 533 (7601) (2016) 73–76, https://doi.org/10.1038/nature17439.
[3] D. Xue, P.V. Balachandran, J. Hogden, J. Theiler, D. Xue, T. Lookman, Accelerated search for materials with targeted properties by adaptive design, Nat. Commun. 7 (1) (2016), https://doi.org/10.1038/ncomms11241.
[4] D. Xue, P.V. Balachandran, R. Yuan, T. Hu, X. Qian, E.R. Dougherty, T. Lookman, Accelerated search for BaTiO3-based piezoelectrics with vertical morphotropic phase boundary using Bayesian learning, Proc. Natl. Acad. Sci. USA 113 (47) (2016) 13301–13306, https://doi.org/10.1073/pnas.1607412113.
[5] V. Duros, J. Grizou, W. Xuan, Z. Hosni, D.-L. Long, H.N. Miras, L. Cronin, Human versus robots in the discovery and crystallization of gigantic polyoxometalates, Angew. Chem. Int. Ed. 56 (36) (2017) 10815–10820, https://doi.org/10.1002/anie.201705721.
[6] S. Lu, Q. Zhou, L. Ma, Y. Guo, J. Wang, Rapid Discovery of Ferroelectric Photovoltaic Perovskites and Material Descriptors via Machine Learning, Small Methods 3 (11) (2019) 1900360, https://doi.org/10.1002/smtd.v3.1110.1002/smtd.201900360.
[7] T. Gu, W. Lu, X. Bao, N. Chen, Using support vector regression for the prediction of the band gap and melting point of binary and ternary compound semiconductors, Solid State Sci. 8 (2) (2006) 129–136, https://doi.org/10.1016/j.solidstatesciences.2005.10.011.
[8] L. Weston, C. Stampfl, Machine learning the band gap properties of kesterite I2−II−IV−V4 quaternary compounds for photovoltaics applications, Phys. Rev. Mater. 2 (2018), 085407. https://link.aps.org/doi/10.1103/PhysRevMaterials.2.085407.
[9] Oliynik, A., Adutwum, L., Harynuk, J., Mar, A. Classifying Crystal Structures of Binary Compounds AB through Cluster Resolution Feature Selection and Support Vector Machine Analysis Chem. Mater. 2016, 28, 18, 6672–6681 (2016). Doi: 10.1021/acs.chemmater.6b02905.
[10] M.A. Shandiz, R. Gauvin, Application of machine learning methods for the prediction of crystal system of cathode materials in lithium-ion batteries, Comput. Mater. Sci. 117 (2016) 270–278.
[11] T. Toyao, et al., Toward effective utilization of methane: machine learning prediction of adsorption energies on metal alloys, J. Phys. Chem. C 122 (2018) 8315–8326, https://doi.org/10.1016/j.commatsci.2016.02.021.
[12] R. Eitel, et al., New high temperature morphotropic phase boundary piezoelectrics based on Bi(Me)O3PbTi3 ceramics, Jpn. J. Appl. Phys. 40 (2001) 5999, https://doi.org/10.1143/jjap.40.5999.
[13] P.V. Balachandran, B. Kowalski, A. Sehirlioglu, T. Lookman, Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning, Nat. Commun. 9 (1) (2018), https://doi.org/10.1038/s41467-018-03821-9.
[14] P.V. Balachandran, B. Kowalski, A. Sehirlioglu, T. Lookman, Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning (2018) Doi: .
[15] P.V. Balachandran, et al., Predictions of new ABO3 perovskite compounds by combining machine learning and density functional theory, Phys. Rev. Mater. 2 (2018), 043802, https://doi.org/10.1103/PhysRevMaterials.2.043802.
[16] I. Grinberg, M.R. Suchomel, P.K. Davies, A.M. Rappe, Predicting morphotropic phase boundary locations and transition temperatures in Pb- and Bi-based perovskite solid solutions from crystal chemical data and first-principles calculations, J. Appl. Phys. 98 (9) (2005) 094111, https://doi.org/10.1063/1.2128049.
[17] T. Qi, I. Grinberg, A.M. Rappe, Correlations between tetragonality, polarization, and ionic displacement in PbTiO3derived ferroelectric perovskite solid solutions, Phys. Rev. B 82 (2010), 134113, https://doi.org/10.1103/PhysRevB.82.134113.
[18] P.V. Balachandran, S.R. Broderick, K. Rajan, Identifying the “inorganic gene” for high-temperature piezoelectric perovskites through statistical learning, Proc. R. Soc. A Math. Phys. Eng. Sci. 467 (2011) 2271–2290.
[19] S.C. Abrahams, S.K. Kurtz, P.B. Jamieson, Atomic displacement relationship to curie temperature and spontaneous polarization in displacive ferroelectrics, Phys. Rev. 172 (1968) 551–553, https://doi.org/10.1098/rspa.2010.0543.
[20] I. Grinberg, A.M. Rappe, Nonmonotonic TC trends in Bi-based ferroelectric perovskite solid solutions, Phys. Rev. Lett. 98 (2007), 037603, https://doi.org/10.1103/PhysRevLett.98.037603.
[21] I. Grinberg, A.M. Rappe, First principles calculations, crystal chemistry and properties of ferroelectric perovskites, Phase Transit. 80 (4-5) (2007) 351–368, https://doi.org/10.1080/01411590701228505.
[22] Samanta, Atanu & Yadab, Suhas & Shafir, Or & Gu, Zongquan & Meyers, Cedric & Wu, Liyan & Chen, Dongfang & Pandya, Shishir & York, Robert & Martin, Lane & Spanier, Jonathan & Grinberg, Ilya. Advancing from phenomenological to predictive theory of ferroelectric oxide solution properties through consideration of domain walls. (2021).
[23] L. Feng, Z.-G. Ye, Phase Diagram and Phase Transitions in the Relaxor Ferroelectric Pb(Fe2/3W1/3)O3–PbTiO3 System, J. Solid State Chem. 163, (2), (2002). Doi: 10.1006/jssc.2001.9433.
[24] M. Dambekalne, K. Bormanis, A. Sternberg, I. Brante, Relaxor ferroelectric PbSc1/2Nb1/2O3 —PbZn1/3Nb2/3O3 — PbMg1/3Nb2/3O3 ceramics, Ferroelectrics 240 (1) (2000) 1487–1494, https://doi.org/10.1080/00150190008227974.
[25] V. Tennery, K. Hang, Ferroelectric and Structural Properties of the Pb(Sc1/2Nb1/2)1-∞-Ti∞O3 System, J. Am. Ceram. Soc.. 51 (2006) 671–674, https://doi.org/10.1111/j.1151-2916.1968.tb15925.x.
[26] Singh, Satendra & Yusuf, S. & Yoon, Songhak & Baik, Sunggi & Shin, Namsoo & Pandey, Dhananjai. Ferroic transitions in the multiferroic (1 - X)Pb(Fe 1/2Nb 1/2)O 3-xPbTiO 3 system and its phase diagram. Acta Materialia. 58. 5381-5392. (2010).
[27] K. Kusumoto, T. Sekiya, Processing and Properties of Relaxor Ferroelectric PNN–PT Perovskite Ceramics, Ferroelectrics. 240 (2000) 1593–1600, https://doi.org/10.1016/j.actamat.2010.06.014.
[28] Gehring, Peter & Phelan, Daniel & Rodriguez, Efrain & Ye, Zuo-Guang. Phase diagram and skin effect of the relaxor ferroelectric (1-x)Pb(Mg1/3Nb2/3)O3+xPbTiO3. (2012). Doi: 10.1080/01411594.2014.989226.
[29] D. La-Orauttapong, Beatriz Noheda, Z.-G Ye, Gehring, Peter & Toulouse, Jean & Cox, D. & Shirane, G.. Phase diagram of the relaxor ferroelectric (1-x)Pb(Zn1/3Nb2/3)O3-xPbTiO3. Phys. Rev. B. 65. (2002). Doi: 10.1016/j.ssc.2004.04.036.
[30] J. Chen, J. Li, L. Fan, N. Zou, P. Ji, L. Liu, L. Fang, H. Kang, X. Xing, Enhanced piezoelectric and antiferroelectric properties of high-TC perovskite of Zr-substituted Bi(Mg1/2Ti1/2)O3-PbTiO3, J. Appl. Phys. 112 (7) (2012) 074101, https://doi.org/10.1063/1.4755790.
[31] I. Sterianou, D.C. Sinclair, I.M. Reaney, T.P. Comyn, A.J. Bell, Investigation of High Curie Temperature (1–x)BiSc1−yFeyO3–xPbTiO3 Piezoelectric Ceramics, J. Appl. Phys. 106 (8) (2009) 084107, https://doi.org/10.1063/1.3253585.
[32] Sterianou, Iasmi & Reaney, I. & Sinclair, Derek & Woodward, David & Hall, David & Bell, Andrew & Comyn, T.. High-temperature (1−x)BiSc1/2Fe1/2O3-xPbTiO3 piezoelectric ceramics. Appl. Phys. Lett. 87. 242901. (2005). https://doi.org/10.1063/1.2140067.
[33] S. Zhang, C.A. Randall, T.R. Shrout, Characterization of Perovskite Piezoelectric Single Crystals of 0.43BiScO3–0.57PbTiO3 With High Curie Temperature, J. Appl. Phys. 95 (8) (2004) 4291–4295, https://doi.org/10.1063/1.1682694.
[34] Chen, Jun & Tan, Xiaoli & Jo, Wook & Rodel, Jurgen. Temperature Dependence of Piezoelectric Properties of High-Tc (1−x)Bi(Mg1/2Ti1/2)O3–xPbTiO3. J. Appl. Phys. 106. 034109 - 034109. (2009). https://doi.org/10.1063/1.3191666.
[35] Hu, Zhongqiang & Chen, Jianguo & Li, Meiya & Li, Xiaotian & Liu, Guoxi & Dong, Shuxiang. Morphotropic Phase Boundary and High Temperature Dielectric, Piezoelectric, and Ferroelectric Properties of (1−x)Bi(Sc3/4In1/4)O3-XPbTiO3 Ceramics. J. Appl. Phys. 110. 064102-064102. (2011). Doi: 10.1063/1.3638123.
[36] Zhang, Linxing & Chen, Jun & Zhao, Hanqing & Fan, Longlong & Rong, Yangchun & Deng, Jinxia & Yu, Ranbo & Xing, Xianran. Temperature-independent ferroelectric property and characterization of high-TC 0.2Bi(Mg1/2Ti1/2)O3-0.8PbTiO3 thin films. Appl. Phys. Lett. 103. (2013). Doi: 10.1063/1.4819205.
[37] D.M. Stein, I. Grinberg, A.M. Rappe, P.K. Davies, Multiple dielectric transitions in the PbTiO3-Bi(Zn1/2Ti1/2)O3-Bi(Mg1/2Ti1/2)O3 system, J. Appl. Phys - J APPL PHYS. 110 (7) (2011) 074110, https://doi.org/10.1063/1.3646559.
[38] Jiang, Yihang & Qin, Baoquan & Yue, Xi & Zhao, Yi & Jiang, Yuzhi & Xiao, Dingquan & Zhu, Jianguo. Feature of Morphotropic Phase Boundary of yBiGaO3–(1−x−y)BiScO3–xPbTiO3 High-Temperature Piezoelectric Ceramics. J. Appl. Phys. 103. 074116-074116. (2008). https://doi.org/10.1063/1.2871683.
[39] D. Pang, Z. Yi, Ferroelectric, piezoelectric properties and thermal expansion of new Bi(Ni 3/4 W 1/4)O 3 –PbTiO 3 solid solutions, RSC Adv. 7 (32) (2017) 19448–19456, https://doi.org/10.1039/C7RA01638G.
[40] A Samanta, Atanu, Yadav, Suhas; Shafir, Or, Gu, Zongquan, Meyers,Cedric J. G., Wu, Liyan, Chen, Dongfang; Pandya, Shishir, York, Robert A., Martin, Lane W., Spanier, Jonathan E., Grinberg, Ilya. Advancing from phenomenological to predictive theory of ferroelectric oxide solution properties through consideration of domain walls (2021).
[41] W. Wu, C. Zucca, A.S. Muhaimeed, W.M. Al-Shafie, A.M. Fadhil Al-Quraishi, V. Nangia, M. Zhu, G. Liu, Soil salinity prediction and mapping by machine learning regression in Central Mesopotamia, Iraq. Land Degradation and Development. 29 (11) (2018) 4005–4014, https://doi.org/10.1002/ldr.3148.
[42] D. Hussain, A.A. Khan, Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan, Earth Sci. Inform 13 (3) (2020) 939–949, https://doi.org/10.1007/s12145-020-00450-z.
[43] H. Pang, A. Lin, M. Holford, B.E. Enerson, L.u. Bin, M.P. Lawton, E. Floyd, H. Zhao, Pathway analysis using random forests classification and regression, Bioinformatics 22 (16) (2006) 2028–2036, https://doi.org/10.1093/bioinformatics/btl344.