Journal of Applied Physics

Machine learning 5d-level centroid shift of Ce^{3+} inorganic phosphors

Cite as: J. Appl. Phys. **128**, 013104 (2020); https://doi.org/10.1063/5.0012434
Submitted: 01 May 2020 . Accepted: 16 June 2020 . Published Online: 02 July 2020

Ya Zhuo, Shruti Hariyani, Shihai You, Pieter Dorenbos, and Jakoah Brgoch

Journal of Applied Physics ARTICLE scitation.org/journal/jap

Machine learning 5d-level centroid shift of Ce^{3+} inorganic phosphors

Cite as: J. Appl. Phys. **128**, 013104 (2020); doi: 10.1063/5.0012434
Submitted: 1 May 2020 · Accepted: 16 June 2020 ·
Published Online: 2 July 2020

Ya Zhuo,[1] Shruti Hariyani,[1] Shihai You,[1,2] Pieter Dorenbos,[3] and Jakoah Brgoch[1,a)]

AFFILIATIONS

[1]Department of Chemistry, University of Houston, Houston, Texas 77204, USA
[2]College of Materials, Xiamen University, Xiamen, Fujian 361005, China
[3]Faculty of Applied Sciences, Department of Radiation Science and Technology, Delft University of Technology, Mekelweg 15, 2629 JB Delft, The Netherlands

Note: This paper is part of the special collection on Machine Learning for Materials Design and Discovery
[a)]**Author to whom correspondence should be addressed:** jbrgoch@uh.edu

ABSTRACT

Information on the 5d level centroid shift (ε_c) of rare-earth ions is critical for determining the chemical shift and the Coulomb repulsion parameter as well as predicting the luminescence and thermal response of rare-earth substituted inorganic phosphors. The magnitude of ε_c depends on the binding strength between the rare-earth ion and its coordinating ligands, which is difficult to quantify *a priori* and makes phosphor design particularly challenging. In this work, a tree-based ensemble learning algorithm employing extreme gradient boosting is trained to predict ε_c by analyzing the optical properties of 160 Ce^{3+} substituted inorganic phosphors. The experimentally measured ε_c of these compounds was featurized using the materials' relative permittivity (ε_r), average electronegativity, average polarizability, and local geometry. Because the number of reported ε_r values is limited, it was necessary to utilize a predicted relative permittivity ($\varepsilon_{r,SVR}$) obtained from a support vector regressor trained on data from ~2800 density functional theory calculations. The remaining features were compiled from open-source databases and by analyzing the rare-earth coordination environment from each Crystallographic Information File. The resulting ensemble model could reliably estimate ε_c and provide insight into the optical properties of Ce^{3+}-activated inorganic phosphors.

I. INTRODUCTION

Phosphor-converted white light-emitting diodes (pc-wLEDs) have gained significant attention because they are less toxic than fluorescent lamps and consume less energy than incandescent light bulbs.[1,2] In a typical pc-wLED, a rare-earth substituted inorganic phosphor is excited by blue or near-ultraviolet light from an LED chip. The absorbed photons are then down-converted and re-emitted at longer wavelengths.[3] The resulting combination of the higher energy LED emission and the lower energy phosphor emission covers the entire visible spectrum, thereby appearing as white light. In general, the color quality of this white light is primarily controlled by the phosphor's optical properties, which stem from the rare-earth ion's parity-allowed electric dipole $5d \leftrightarrow 4f$ transition. The magnitude of the d–f separation is particularly significant because it sets the position of the absorption and emission energies. More importantly, it can be manipulated by varying the interaction between the rare-earth 5d-orbitals and the neighboring anion ligands.

Ce^{3+} substituted materials are among the most widely explored rare-earth containing phosphors due to their unique luminescent properties such as broad and highly efficient emission. Typically, the energy gap between the 4f orbitals and the centroid position of the five 5d orbitals for a free Ce^{3+} ion in vacuum occurs deep in the ultraviolet region of the electromagnetic spectrum (~6.35 eV; 51 200 cm^{-1}). The 5d levels can be stabilized (decreased) in energy relative to the free ion ground state by surrounding the cation with anions, called the nephelauxetic effect. This change can be quantified as the 5d centroid shift (ε_c).[4] Additional energy stabilization can also be achieved by varying the different degrees of covalency and geometry of the cation–anion interaction, known as crystal field splitting (ε_{cfs}).[5–7] The optical properties of a phosphor can be reasonably assessed with the knowledge of the centroid shift and crystal field splitting of the 5d energy levels.

Studies have revealed the coordination number, polyhedron size, and degree of polyhedron distortion around the Ce^{3+} ion

control the crystal field splitting of the $4f^{n-1}5d$-levels, whereas the anion type is generally irrelavent.[8–12] This is in contrast to the centroid shift, which is suggested to be closely associated with the anion ligands. The instantaneous position of Ce^{3+} electrons polarize the anion ligands, which results in a self-induced potential that reduces the inter-electron Coulomb repulsion between the Ce^{3+} electrons and causes the centroid energy of the 5d levels to lower.[13] The centroid shift is further influenced by the covalency of the bonding within the crystal structure.[14] Indeed, a significant degree of covalency appears to induce a more substantial nephelauxetic effect, and thus, a large value of ε_c. For example, the bonds in nitrides and oxynitrides generally exhibit stronger covalency than oxides.[15] As a result, nitride and oxynitride compounds cause a more significant downward shift of the 5d energy levels than oxide hosts, correlating to a more substantial centroid shift ε_c.[16–22]

Considering both the ligand polarization and covalency models, ε_c can be estimated following Eq. (1) based on the polarizability of N coordinating anion ligands at distance R_i from the rare-earth, in this case, Ce^{3+},[8,9]

$$\varepsilon_c = 1.79 \times 10^{13} \alpha_{sp} \sum_{i=1}^{N} \frac{1}{(R_i - f\Delta R)^6}, \quad (1)$$

where α_{sp} is the spectroscopic polarizability of the anion, ΔR is the ionic radius difference between Ce^{3+} and the cation for which it substitutes, and f is the correction parameter that accounts for the relaxation of the neighboring anions induced by the rare-earth substitution. α_{sp} is a phenomenological parameter that is not only related to the average polarizability of the N nearest anionic neighbors around Ce^{3+} but also represents the covalency within the structure. Unfortunately, quantitatively assessing α_{sp} remains challenging. Thus, it is not always straightforward to predict centroid shift values using this equation.

Relative permittivity (ε_r), also called the dielectric constant, provides information on polarizability and covalency that can be explicitly measured or calculated,[23,24] making it an obtainable proxy for α_{sp}.[6] In general, ε_r varies significantly depending on the material, temperature, frequency of applied field, and other parameters. Nevertheless, it provides information on the type of bonding in a compound. Materials with a small ε_r tend to have more ionic-like interactions, such as fluorides, whereas materials with a large ε_r tend to be more covalent-like interactions, such as nitrides and sulfides. The value of the relative permittivity can be experimentally obtained from the comparison of the capacity of an air-filled capacitor with that of the same capacitor containing the dielectric substance under specified temperature and frequency. However, these measurements are relatively uncommon, and thus the number of experimentally measured values is small. Fortunately, ε_r is an intrinsic material property that can also be estimated using density functional theory (DFT), allowing for the potential of high-throughput implementation.[25] There are still considerable challenges in computationally determining ε_r; most notably, calculating ε_r using DFT is computationally expensive. DFT can also not easily account for atomic disorder, like site sharing, which is common in inorganic phosphor hosts, and it is currently restricted to smaller unit cells containing typically fewer than one hundred atoms. These limitations have permitted ε_r to be calculated for only a few thousand ordered crystalline compounds, or <10% of the reported inorganic solids. Luckily, the implementation of machine learning provides an avenue to vastly expand the number of ε_r values available with minimal computational cost.

Machine learning has become a valuable tool for the development of optical materials like inorganic scintillators, inorganic phosphors, and inorganic photovoltaics, among other property-specific problems.[26–32] Embedding machine learning into the physical and chemical sciences not only accelerates the discovery of new materials but also provides insightful knowledge on composition–structure–property relationships. For example, a model based on chemical compositions and electronic structure calculations was recently developed to predict the Coulomb repulsion energy and chemical shift for lanthanide ions in a compound. In combination with the key experimentally measured parameters and the physical-based empirical models, the machine learning model can predict the 4f ground state and $5d^1$ excited state electronic configurations of the lanthanide dopant relative to the band edges of the host material, which enables the fast screening for potential candidates in a high-throughput manner.[31]

In this work, we employ a combination of compositional and structural descriptors to build a machine learning regression model that first predicts relative permittivity. This model is trained using ~2800 $\varepsilon_{r,\mathrm{DFT}}$ values extracted from the Materials Project, and allows us to expand our database of $\varepsilon_{r,\mathrm{SVR}}$ values to ~280 000 compositions that are compiled in Pearson's Crystal Database (PCD). We then use a tree-based ensemble learning method implementing extreme gradient boosting (XGB) to predict the centroid shift ($\varepsilon_{c,\mathrm{XGB}}$) for Ce^{3+}-activated inorganic phosphors using a descriptor set including the machine-learning derived relative permittivity as well as average cation electronegativity, average anion polarizability, structure condensation, and Ce^{3+} coordination environment. Our successful development of a machine learning method to approximate the centroid shift is a key step to interpret and predict the luminescence properties of inorganic phosphors. Moreover, the centroid shift data combined with luminescence spectroscopy can provide insights into the chemical shift of 4f-electron binding energy.[33]

II. EXPERIMENTAL

A. Relative permittivity data extraction and model construction

The machine learning model to predict ε_c first requires $\varepsilon_{r,\mathrm{SVR}}$ as a descriptor. $\varepsilon_{r,\mathrm{DFT}}$ of 4867 compounds were therefore extracted from the Materials Project database.[25] The Materials Project calculates $\varepsilon_{r,\mathrm{DFT}}$ using a high-throughput framework based on Density Functional Perturbation Theory (DFPT).[34–36] In some cases, negative values of $\varepsilon_{r,\mathrm{DFT}}$ are present in the database; these were removed from the final training data set. The number of compositions was further reduced to 2991 after cross-referencing with PCD[37] to ensure that all the phases used for machine learning are also experimentally reported, i.e., hypothetical crystal structures were removed. The data distribution plot is provided in the supplementary material (Fig. S1). About 5% of the data possess an $\varepsilon_{r,\mathrm{DFT}} > 20$ and these tend to be transition metal-containing compositions. The multiple oxidation states and narrow (or zero)

bandgap makes these compositions unsuitable as phosphor hosts. Thus, these compounds were also removed by fixing the learning window to $0 < \varepsilon_{r,\mathrm{DFT}} < 20$. Compounds with elements from group 18 (noble gases) and $Z > 83$ (except for U and Th) were also excluded. These criteria reduced the final training set to 2832 compounds.

The descriptors used in this study included 17 distinct compositional variables describing elemental properties such as position on the periodic table, electronic structure, and physical properties as well as their associated math expressions (difference, average, largest value, smallest value, and standard deviation). Additionally, 13 structural descriptors related to variables, including crystal system, space group, and unit cell volume, among others, were incorporated in the machine learning algorithm. The full list of descriptors and mathematical expressions is available in the supplementary material (Table S1). In total, 98 descriptors were used to build the model.

The 2832×98 data matrix was split into a random train and test subsets with a 9:1 ratio. The training set was then standardized to have a mean of 0 and a variance of 1, and the test set was transformed using the same scalar as the training set. The machine learning model was constructed using a support vector regression (SVR) algorithm with the radial basis function (RBF) kernel.[38] A grid search, which exhaustively evaluates all parameter combinations, was performed on the training set with a 10-fold cross-validation method to choose the best hyper-parameter settings. The searching space was defined as cost (C) values ranging in the log space [0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2], and epsilon values in the real space [0.001, 0.01, 0.1, 1], where cost is the regularization term and epsilon specifies the epsilon-tube within which no penalty is associated in the training loss function. The scikit-learn python implementations of these learning algorithms were used.[39]

B. Centroid shift data extraction and model construction

The development of the machine learning model to predict the centroid shift of Ce^{3+}-activated phosphors involved extracting 219 experimentally reported host compositions and the associated ε_c values from the literature.[5,7–11,40] Before training, data were examined according to two primary criteria. First, phosphors hosts that do not have a reliable crystal structure reported were omitted. The second criterion is that Ce^{3+} must have a single, chemically obvious substitution site or a site that was specifically studied and reported in the paper if the compound contains multiple substitution sites. The final number of training labels was reduced to 160 phosphors. Eight features, including relative permittivity, which was predicted using the first $\varepsilon_{r,\mathrm{SVR}}$ model, weighted average cation electronegativity, weighted average anion polarizability, and local environment related descriptors, were employed to characterize these phosphors. A train-test splitting ratio of 9:1 was also used to randomly split the data into two subsets.

The tree-based extreme gradient boosting (XGB) algorithm was selected to train the centroid shift model owing to its effectiveness in obtaining satisfactory results with small data sets while controlling overfitting.[41] The model was gradient boosted for 50 rounds with a learning rate of 0.15 and a maximal tree depth of 3. These actions were carried out using the scikit-learn API.[39]

III. RESULTS AND DISCUSSION

A. Modeling relative permittivity

The objective of SVR is to find a multi-dimentional function, $f(x)$, that deviates from the training label by a value no greater than epsilon (ε) for each training sample and is simultaneously as flat as possible. The x in $f(x)$ is a set of feature vectors that represent the compounds and the solution of $f(x)$ is the predicted centroid shift value. A regularization parameter, C, also called cost, is introduced to control the compromise between model accuracy and flatness. As C increases, the strength of the regularization increases and then the tolerance for data outside of ε increases. As C approaches 0, the tolerance approaches 0 and the equation collapses into the simplified one. The two parameters, C and ε, must first be optimized to maximize the model performance. Conducting a grid search for the free parameters, provided in the supplementary material (Fig. S2), shows that values of $C = 10^{1.75}$ ($\approx$56.23) and $\varepsilon = 0.1$ give the best model performance based on the model statistics, i.e., coefficient of determination (r^2) and mean absolute error (MAE).

The ε_r was trained using these parameters and 2832 training labels collected from the Materials Project. Figure 1 shows the training data distribution. Over half of the compositions have a DFT-calculated relative permittivity ($\varepsilon_{r,\mathrm{DFT}}$) below 5, and $\approx$95% have an $\varepsilon_{r,\mathrm{DFT}} \leq 12.5$. The SVR model was evaluated with a test set, which was obtained from randomly selecting 10% of the data before training and completed unseen by the model. As shown in Fig. 2, the histograms at the top and right show that both the DFT-calculated relative permittivity ($\varepsilon_{r,\mathrm{DFT}}$) and the machine learning predicted values ($\varepsilon_{r,\mathrm{SVR}}$) in the test set represent the distribution of the entire training labels well. Acceptable agreement is obtained between $\varepsilon_{r,\mathrm{DFT}}$ and $\varepsilon_{r,\mathrm{SVR}}$ with the coefficient of

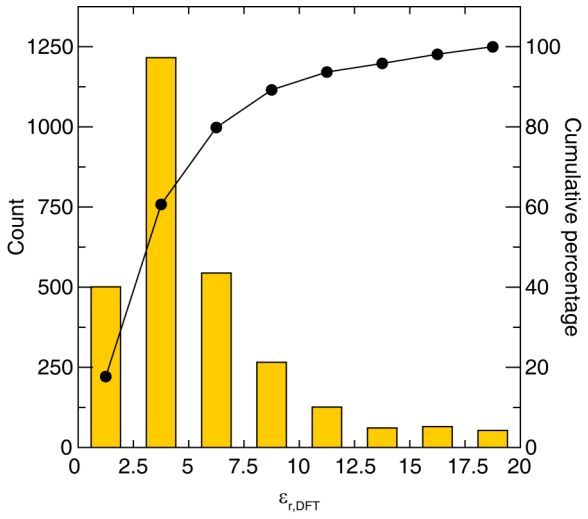

FIG. 1. Data distribution of 2832 training samples that were calculated with DFT. The bars represent the total counts in each bin, and the curve represents the cumulative percentage.

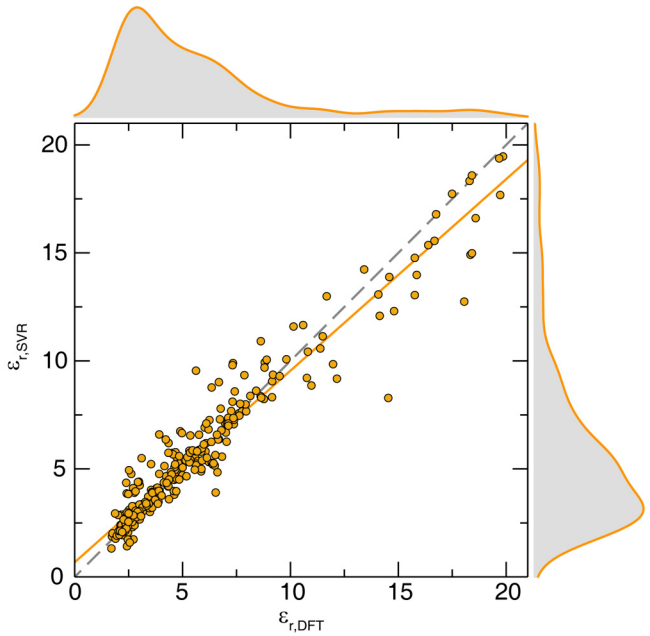

FIG. 2. Predicted relative permittivity ($\varepsilon_{r,SVR}$) vs DFT-calculated relative permittivity ($\varepsilon_{r,DFT}$) is shown for a 10% holdout test set. The ideal line is shown as the dashed gray line and the fit line is shown as the solid yellow line. The curves at the top and right show the histograms of the data.

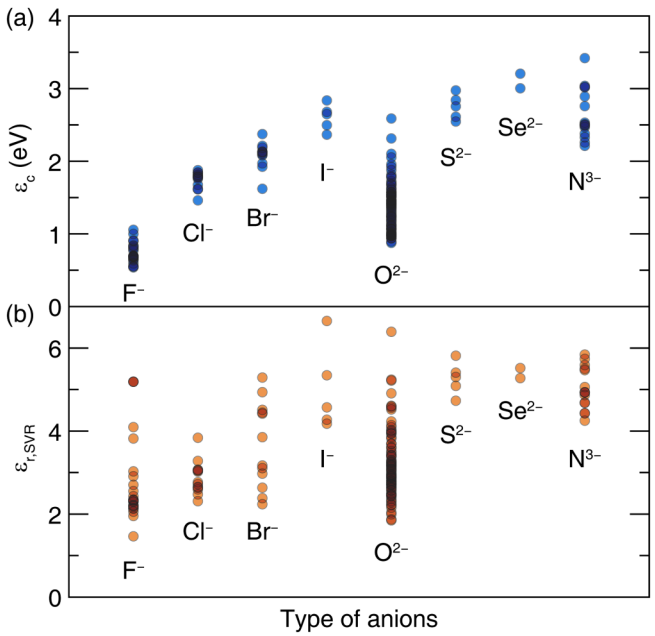

FIG. 3. (a) The centroid shift (ε_c) and (b) the machine learning obtained relative permittivity ($\varepsilon_{r,SVR}$) for Ce^{3+} inorganic compounds displayed against the type of anions. The darker regions represent a higher density of data points.

determination (r^2), and mean absolute error (MAE) being 0.93 and 0.65, respectively. 246 out of 284 (87%) compounds in the test set were predicted with an error of less than 25%. There is a slight but noticeable underestimation for $12.5 < \varepsilon_r \leq 20$, which is most likely due to the limited number of training samples in this region. Nevertheless, this model is useful to estimate $\varepsilon_{r,SVR}$ for over 270 000 compounds compiled in PCD.

B. Connecting relative permittivity to the centroid shift prediction

To build the centroid shift model, 219 experimentally measured centroid shift (ε_c) data were collected from the literature for Ce^{3+}-activated inorganic phosphors. Data reduction was then performed. For example, although ε_c data are available for $BaCaBO_3F$ and Sr_2SiO_4, they were excluded because there is no clear information on the Ce^{3+} substitution site or the associated optical properties stemming from substitution on each site. Indeed, the obtained value of ε_c for these systems could stem from Ce^{3+} occupying several crystallographic sites. This ambiguity can cause issues in machine learning; thus, these data were removed. Data sanitization reduced the final training set to 160 data, including 53 (oxy) halides, 85 oxides, 5 sulfides, 2 selenides, and 15 nitrides. All of these phosphors contain a single substitution site, or if multiple crystallographic sites are present, the rare-earth site occupancy was specifically investigated and reported in the literature. The data are provided in the supplementary material (Table S2).

Analyzing these data by plotting ε_c against the type of anions present in the composition [Fig. 3(a)] demonstrates that the centroid shift tends to follow some trends. Fluorides have the smallest ε_c as dictated by the ionic bonding of these phases, whereas sulfides, selenides, and nitrides possess a relatively larger ε_c values resulting from their more covalent bonding. Although a crude estimation of ε_c can be made based on the anions present in the host compounds, ε_c may still vary greatly for compounds that are composed of the same type of anion. This variation is particularly noticeable for oxides and nitrides. One possible explanation is that the susceptibility of O^{2-} and N^{3-} to the electronegativity of the surrounding cations is larger than anions such as F^-. In other words, the anion polarizability and covalency are affected more strongly by the cations in oxides and nitrides, leading to greater variability in the properties. This observation is a bit unfortunate because these compound types are among the most common classes of inorganic phosphor hosts, yet, they have the biggest range of centroid shifts.

The same general trend was also observed in the plot of $\varepsilon_{r,SVR}$ against the type of anions. Because the experimentally measured or DFT calculated relative permittivity for these 160 compositions is unknown, the $\varepsilon_{r,SVR}$ model was used for this comparison. As shown in Fig. 3(b), $\varepsilon_{r,SVR}$ in general shows a trend as $I^- > Br^- > Cl^- > F^-$ in halides, and sulfides, selenides, and nitrides possess a larger $\varepsilon_{r,SVR}$ than oxides. These trends agree well with those in ε_c, as expected. There is a slightly larger spread in this plot, however, meaning that using $\varepsilon_{r,SVR}$ solely to predict ε_c is subject to error. For example, although $LaPO_4:Ce^{3+}$ and $GdAlO_3:Ce^{3+}$ have a very close $\varepsilon_{r,SVR}$ (4.23 and 4.25, respectively), they

show a large discrepancy in ε_c ($\varepsilon_c = 1.07$ eV for $LaPO_4:Ce^{3+}$ and $\varepsilon_c = 1.70$ eV for $GdAlO_3:Ce^{3+}$).[7,42,43] The failure stems from the fact that a simple descriptor like $\varepsilon_{r,SVR}$ is not sufficient to account for all of the impacting factors of the centroid shift, such as the size effect. For instance, permittivity can be expressed as a sum of contributions from the different atoms in a compound, and a large cation, such as Cs^+, takes a large volume of the lattice, but it may not contribute much to the permittivity. However, its large size most likely leads to a smaller ε_c, according to Eq. (1).

C. Machine learning the centroid shift

The ability to predict the centroid shift is clearly multidimensional. An ensemble learning method was therefore constructed to predict ε_c accurately with features chosen based on numerical equations for centroid shift and the crystal structures. The features and the corresponding notations used in this work are provided in Table I, and includes $\varepsilon_{r,SVR}$, average cation electronegativity defined in Eq. (2),

$$\chi_{av} = \frac{\sum n_i z_i \chi_i}{\sum n_i z_i}, \tag{2}$$

where n_i is the stoichiometric index of cation i with charge z_i and χ_i is the Pauling electronegativity.[44] The average anion polarizability is calculated from Eq. (3),

$$\alpha_{av} = \frac{\sum m_e \alpha_e}{\sum m_e}, \tag{3}$$

where m_e is the stoichiometric index of anion e, and α_e is the polarizability. In addition, ΔR is the difference between the ionic radius compiled by Shannon of Ce^{3+} and the cation (R_M) for which Ce^{3+} substitutes.[45] The final descriptor is condensation (cond.), which is the ratio between the number of anions and the number of cations contained in the chemical formula.

The model was subsequently trained using a tree-based XGB algorithm with 10% of the data held out for model evaluation. To account for randomness in the test data selection, ten different test sets were examined, which yielded an average r^2 of 0.90. Figure 4(a) shows the results from one of these test sets. The detailed statistics are provided in the supplementary material (Table S3). The statistics of the test set prediction show compelling agreement between

TABLE I. The feature set used to predict the centroid shift modeled using XGB.

Variable	Notation
Relative permittivity	$\varepsilon_{r,SVR}$
Average cation electronegativity	χ_{av}
Average anion polarizability	α_{av}
Ionic radius	R_M
Difference in radius	ΔR
Average bond length	R_{av}
Coordination number	N
Condensation	cond.

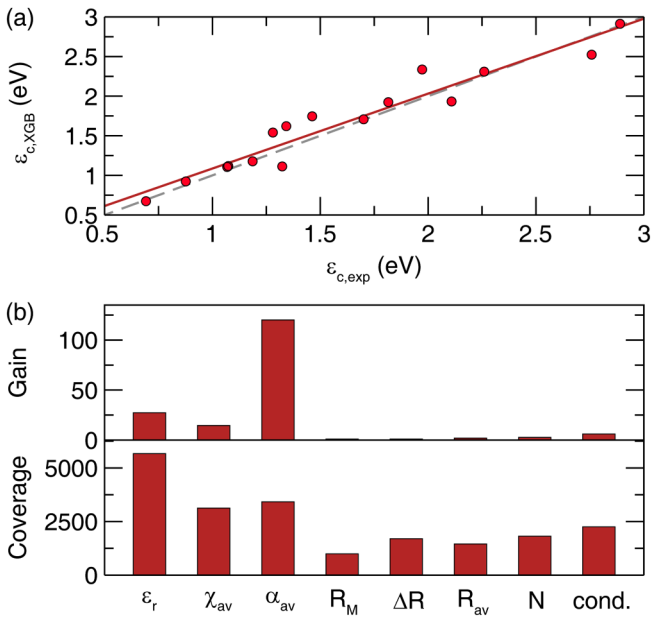

FIG. 4. (a) Machine learning predicted centroid shift ($\varepsilon_{c,XGB}$) against the experimentally measured centroid shift ($\varepsilon_{c,exp}$). The test set plotted here represents 10% of the training set. The ideal line is shown as the dashed gray line and the fit line is shown as the solid red line. (b) Feature importance of the XGB model in terms of gain and coverage.

$\varepsilon_{c,exp}$ and $\varepsilon_{c,XGB}$ with a root-mean-squared error (RMSE) of 0.18 eV and a mean absolute error (MAE) of 0.13 eV. These results indicate that the feature set does a reasonable job determining a Ce^{3+} phosphor's centroid shift.

Analyzing the feature importance in this model can provide insight into the crystal-chemical properties that control the centroid shift. Different feature importance matrices are accessible from the XGB model. The principle of developing a tree is that a new split on a node will be added only if the split results in a more accurate prediction. The gain can quantify the improvement in accuracy due to the split of a specific feature at a node. The gain in feature importance is the total gain across all splits in which the feature is used, implying the relative contribution of the corresponding feature to the model. As shown in the top panel of Fig. 4(b), the average anion polarizability (α_{av}) has a significantly higher absolute gain than all other features demonstrating it is the most important feature for predicting ε_c. Moreover, relative permittivity and average cation electronegativity (χ_{av}) also show a non-negligible influence on the model performance, while the remaining five features have a minimal total gain. Although the remaining five features have low gain, removing them from the model and re-training leads to considerably worse performance. All of the descriptors are essential to obtain $\varepsilon_{c,XGB}$.

Feature importance can also be evaluated based on the coverage matrix, which describes the relative quantity of samples related to a feature. Here, the trend is slightly different. As shown in the bottom panel of Fig. 4(b), the coverage matrix suggests that $\varepsilon_{r,SVR}$

is the most crucial feature in the coverage matrix. More than 5000 observations were associated with the split of $\varepsilon_{r,\mathrm{SVR}}$ in the boosting process. α_{av} and χ_{av} relate to ~2500 observations, while R_M is the least important although it still influences 989 observations. Overall, $\varepsilon_{r,\mathrm{SVR}}$, χ_{av}, and α_{av} are the three most essential features related to the centroid shift regardless of the feature importance matrix. Not surprisingly, the average anion polarizability is more correlated to the centroid shift than the average cation electronegativity. The local geometry of the luminescent center and condensation contribute less to the model yet, they are still necessary to predict ε_c.

It is vital to quantitatively estimate the position of the 5d-excited states for rare-earth ions in host crystal structures for designing inorganic phosphors. A phosphor with a large centroid shift tends to have a longer wavelength emission and using this prediction of centroid shift makes it possible to roughly estimate the emission color. The ability to predict ε_c even with moderate quantitative accuracy will allow researchers to accelerate phosphor discovery by performing a top-level screening of new phosphor hosts with desired optical properties. Moreover, the ε_c model can be combined with the knowledge of crystal field splitting and the phosphor host's bandgap to predict the location of 5d levels with respect to the bottom of the conduction band.[40] This combination would provide insight into the thermal quenching behavior of the phosphor that has thus far remained largely empirical.

IV. CONCLUSIONS

Combining high-throughput DFT calculations and machine learning techniques provides a unique framework to predict the relative permittivity of over 270 000 compounds compiled in inorganic structure databases. The predicted relative permittivity was then combined with chemical properties and local geometry of 160 Ce^{3+}-activated phosphors to predict the centroid shift. Ensemble learning methods were successfully employed to predict the values of the centroid shift, which is beneficial to interpret and predict the luminescence properties and thermal quenching behaviors for Ce^{3+}-doped phosphors. The results combined with the information of crystal field splitting can also be extended into the estimation of properties of other rare-earth doped materials, for example, the Eu^{2+} inter 4f-electron Coulomb repulsion energy in compound. In addition, the predicted ε_c model serves as a starting point to generate the binding energy of electrons at the top of the valence band. This information is not only of interest for luminescence but also for many other disciplines of science such as photo-catalytic splitting of water, battery potentials, valence band offsets in semiconductor hetero-junction or in core shell particles among applications. Further tuning and improvement of these models for ε_c has the potential to make a broad impact on assisting the development of numerous functional inorganic materials.

SUPPLEMENTARY MATERIAL

See the supplementary material for the feature set of the relative permittivity model (Table S1); centroid shift training lables (Table S2); centroid shift model statistics of ten randomly generated test sets (Table S3); relative permittivity data distribution (Fig. S1); and the plot of parameter optimization for cost and epsilon (Fig. S2).

ACKNOWLEDGMENTS

The authors acknowledge the National Science Foundation (NSF) (No. DMR 18-47701). This work was also supported by the Welch Foundation (No. E-1981) and the Texas Center for Superconductivity at the University of Houston (TcSUH). The authors declare no competing financial interest.

DATA AVAILABILITY

The training and test datasets, as well as the codes used to generate the models in this work are openly available in https://github.com/BrgochGroup/CentroidShiftPredictor, Ref. 46.

REFERENCES

[1]J. K. Kim and E. F. Schubert, Opt. Express **16**, 21835 (2008).
[2]S. Pimputkar, J. S. Speck, S. P. DenBaars, and S. Nakamura, Nat. Photonics **3**, 180 (2009).
[3]N. C. George, K. A. Denault, and R. Seshadri, Annu. Rev. Mater. Res. **43**, 481 (2013).
[4]C. Jørgensen, *Modern Aspects of Ligand Field Theory* (North-Holland, Amsterdam, 1971).
[5]P. Dorenbos, J. Lumin. **135**, 93 (2013).
[6]M. Mikami and N. Kijima, Opt. Mater. **33**, 145 (2010).
[7]P. Dorenbos, Phys. Rev. B **65**, 235110 (2002).
[8]P. Dorenbos, Phys. Rev. B **62**, 15640 (2000).
[9]P. Dorenbos, Phys. Rev. B **62**, 15650 (2000).
[10]P. Dorenbos, Phys. Rev. B **64**, 125117 (2001).
[11]P. Dorenbos, J. Lumin. **99**, 283 (2002).
[12]P. Dorenbos, J. Alloys Compd. **341**, 156 (2002).
[13]C. A. Morrison, J. Chem. Phys. **72**, 1001 (1980).
[14]B. F. Aull and H. P. Jenssen, Phys. Rev. B **34**, 6640 (1986).
[15]F. J. DiSalvo and S. J. Clarke, Curr. Opin. Solid State Mater. Sci. **1**, 241 (1996).
[16]T. Wang, Q. Xiang, Z. Xia, J. Chen, and Q. Liu, Inorg. Chem. **55**, 2929 (2016).
[17]T. Suehiro, N. Hirosaki, and R.-J. Xie, ACS Appl. Mater. Interfaces **3**, 811 (2011).
[18]J. W. H. Van Krevel, J. W. T. Van Rutten, H. Mandal, H. T. Hintzen, and R. Metselaar, J. Solid State Chem. **165**, 19 (2002).
[19]Y. Q. Li, A. C. A. Delsing, G. De With, and H. T. Hintzen, Chem. Mater. **17**, 3242 (2005).
[20]L. Zhang, J. Zhang, X. Zhang, Z. Hao, H. Zhao, and Y. Luo, ACS Appl. Mater. Interfaces **5**, 12839 (2013).
[21]S. Schmiechen, H. Schneider, P. Wagatha, C. Hecht, P. J. Schmidt, and W. Schnick, Chem. Mater. **26**, 2712 (2014).
[22]W.-Y. Huang, F. Yoshimura, K. Ueda, Y. Shimomura, H.-S. Sheu, T.-S. Chan, C.-Y. Chiang, W. Zhou, and R.-S. Liu, Chem. Mater. **26**, 2075 (2014).
[23]S. Ishihara, M. Tachiki, and T. Egami, Phys. Rev. B **49**, 16123 (1994).
[24]S. H. Wemple and M. DiDomenico, Phys. Rev. B **3**, 1338 (1971).
[25]A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, APL Mater. **1**, 011002 (2013).
[26]G. Pilania, X.-Y. Liu, and Z. Wang, J. Mater. Sci. **54**, 8361 (2019).
[27]Y. Zhuo, S. Hariyani, E. Armijo, Z. Abolade Lawson, and J. Brgoch, ACS Appl. Mater. Interfaces **12**, 5244 (2020).
[28]S. Ganguly, C. S. Kong, S. R. Broderick, and K. Rajan, Mater. Manuf. Process. **28**, 726 (2013).
[29]B. J. M. Webb-Robertson, K. F. Ferris, and D. M. Jones, IEEE Trans. Nucl. Sci. **55**, 1210 (2008).

[30]H. Jin, H. Zhang, J. Li, T. Wang, L. Wan, H. Guo, and Y. Wei, J. Phys. Chem. Lett. **11**, 3075 (2020).
[31]G. Pilania, K. J. McClellan, C. R. Stanek, and B. P. Uberuaga, J. Chem. Phys. **148**, 241729 (2018).
[32]Y. Zhuo, A. Mansouri Tehrani, A. O. Oliynyk, A. C. Duke, and J. Brgoch, Nat. Commun. **9**, 4377 (2018).
[33]P. Dorenbos, Phys. Rev. B **85**, 165107 (2012).
[34]I. Petousis, W. Chen, G. Hautier, T. Graf, T. D. Schladt, K. A. Persson, and F. B. Prinz, Phys. Rev. B **93**, 115151 (2016).
[35]I. Petousis, D. Mrdjenovich, E. Ballouz, M. Liu, D. Winston, W. Chen, T. Graf, T. D. Schladt, K. A. Persson, and F. B. Prinz, Sci. Data **4**, 160134 (2017).
[36]S. Baroni, P. Giannozzi, and A. Testa, Phys. Rev. Lett. **58**, 1861 (1987).
[37]O. Villars and K. Cenzual, *Pearson's Crystal Data: Crystal Structure Database for Inorganic Compounds* (ASM International, OH, 2007).
[38]H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, Adv. Neural Inf. Process. Syst. **1**, 155 (1996).
[39]F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É Duchesnay, J. Mach. Learn. Res. **12**, 2825 (2011).
[40]S. Wang, Z. Song, Y. Kong, Z. Xia, and Q. Liu, J. Lumin. **200**, 35 (2018).
[41]T. Chen and C. Guestrin, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, Inc., New York, NY, 2016), pp. 785–794.
[42]J. W. M. Verweij, M. T. Cohen-Adad, D. Bouttet, H. Lautesse, B. Moine, and C. Pédrini, Chem. Phys. Lett. **239**, 51 (1995).
[43]M. V. Hoffman, J. Electrochem. Soc. **118**, 1508 (1971).
[44]L. Pauling, *The Nature of the Chemical Bond*, 3rd ed. (Cornell University Press, Ithaca, NY, 1960).
[45]R. D. Shannon, Acta Crystallogr. Sect. A **32**, 751 (1976).
[46]CentroidShiftPredictor website, https://github.com/BrgochGroup/CentroidShiftPredictor.