

# Machine Learning a Phosphor’s Excitation Band Position

Nakyung Lee<sup>1,2,†</sup>, Małgorzata Sójka<sup>1,2,†</sup>, Annie La<sup>1,2</sup>, Syna Sharma<sup>1,2</sup>, Seán Kavanagh<sup>3</sup>, Docheon Ahn<sup>4</sup>,  
David O. Scanlon<sup>5</sup>, Jakoah Brgoch<sup>1,2</sup>

<sup>1</sup>*Department of Chemistry, University of Houston, Houston, Texas 77204, USA*

<sup>2</sup>*Texas Center for Superconductivity, University of Houston, Houston, Texas 77204, USA*

<sup>3</sup>*Harvard University Center for the Environment, Cambridge, Massachusetts 02138, United States*

<sup>4</sup>*PLS-II Beamline Department, Pohang Accelerator Laboratory, POSTECH, Pohang 37673, Republic of Korea*

<sup>5</sup>*School of Chemistry, University of Birmingham, Birmingham, UK*

<sup>†</sup>These authors contributed equally to this work

[\\*jbrgoch@uh.edu](mailto:jbrgoch@uh.edu) (J. Brgoch)

[\\*d.o.scanlon@bham.ac.uk](mailto:d.o.scanlon@bham.ac.uk) (D. O. Scanlon)

## Abstract

Creating superior lanthanide-activated inorganic phosphors is pivotal for advancing energy-efficient LED lighting and backlit flat panel displays. The most fundamental property these luminescent materials must possess is effective absorption/excitation by a blue InGaN LED for practical conversion into white light. The  $5d_1$  excited state energy level of lanthanides, which determines the excitation peak position, is influenced by the inorganic host structure, including the local environment, crystal structure, and composition, making it challenging to predict in advance. This study introduces a new extreme gradient boosting machine learning method that quantitatively determines a phosphor’s longest (lowest energy) excitation wavelength. We focus on the  $\text{Ce}^{3+} 4f \rightarrow 5d$  transition due to its well-defined  $5d_1$  energy level observed in excitation and diffuse reflectance spectra. The model was trained on experimental data for 357  $\text{Ce}^{3+}$  cation substitution sites sourced from literature and in-house measurements and ultimately experimentally validated through the successful synthesis of a novel, blue-excited, green-emitting phosphor:  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}:\text{Ce}^{3+}$ . This compound’s excitation under commercial blue LED wavelength aligned remarkably well with the model’s predictions. These results highlight the transformative potential of data-driven approaches in expediting the discovery of blue-absorbing phosphors for next-generation LED lighting.

## 1. Introduction

Solid-state lighting, used today in applications ranging from general in-home and commercial illumination to high-resolution display technologies, was made possible by the development of blue-emitting indium gallium nitride (InGaN) LED chips.<sup>1,2</sup> Blue LEDs offer outstanding luminous efficacy, a small, durable architecture, a long usable lifetime, eco-friendly composition, and, most notably, serve as the starting point for producing energy-efficient white light. Indeed, converting this nearly monochromatic light source into a functional broad-spectrum white light is achieved by down-converting the LED’s blue ( $\sim 440$  nm to  $\sim 470$  nm) emission using one or more inorganic phosphors. The phosphors absorb and partially convert the blue light into usually yellow/green and red wavelengths, which in combination cover most of the visible spectrum, creating white light. Unfortunately, very few unique commercially produced phosphors are compatible with blue LEDs, with less than ten materials typically meeting the industry’s

strict requirements.<sup>3</sup>

Designing new viable phosphors is an active and ongoing research effort that is an intrinsically multidimensional challenge, necessitating a deep understanding of phosphor chemistry.<sup>4</sup> Phosphors are inorganic host compounds, like oxides, nitrides, or halides, typically substituted with  $\text{Ce}^{3+}$  or  $\text{Eu}^{2+}$  that acts as the luminescence center or activator.<sup>5–8</sup> These two lanthanide ions are primarily used because they possess a broad emission with selection rule allowed  $4f \leftrightarrow 5d$  electronic transitions. Depending on the interaction between the phosphor host and the activator ion, various optical responses can be achieved. Research has predominantly focused on understanding and predicting the factors controlling a phosphor's thermal quenching (temperature-dependent optical behavior) and photoluminescence quantum yield (PLQY), with the goal of identifying specific factors that support thermally robust photoluminescence and near-unity PLQYs.<sup>9–13</sup> Advances in high-throughput experimental and data-driven approaches have also improved the ability to produce desirable emission colors.<sup>14–17</sup> However, realizing phosphor's most critical property—its absorption of blue LED light—is still primarily guided by empirical understanding or costly first principles computational modeling.<sup>18–20</sup> Additionally, blue light absorption by lanthanide-substituted phosphors, especially  $\text{Ce}^{3+}$ , is relatively uncommon outside of the garnet family of materials. This has prompted researchers to instead consider UV ( $\sim 365$  nm) or violet ( $\sim 400$  nm) excited phosphors despite their limited potential for commercialization. The recent surge of phosphors published operating with these alternative LEDs underscores the need for new approaches that can spur the discovery of suitable blue-excited materials.

The position of the excitation band is set by crystal-chemical interactions between the inorganic phosphor host and the lanthanide ion. In the case of  $\text{Ce}^{3+}$  and  $\text{Eu}^{2+}$ , these interactions cause the  $5d$  orbitals to experience two well-known effects. The first concerns the host ligands (anions) stabilizing the centroid of the  $5d$  orbital energy by withdrawing electron density and reducing electric repulsion.<sup>21–25</sup> This phenomenon, called the nephelauxetic effect or centroid shift, is dominated by the electronegativities and atomic polarizabilities of the coordinating anions, among other factors. The second effect, crystal field splitting, separates the nominally degenerate  $5d$  orbitals based on the local coordination environment, such as coordination number, polyhedron volume, bond length, and coordination geometry (symmetry).<sup>26–30</sup> For example, lower coordination numbers with shorter bond lengths facilitate stronger crystal field splitting, as in octahedral or cubic environments, whereas larger sites like dodecahedral environments tend to have weaker crystal field splitting. The combined centroid shift and crystal field splitting generates a redshift ( $D$ ), which represents the energy difference between the initial free ion ( $\text{Ce}^{3+}$ : 6.17 eV; 201 nm and  $\text{Eu}^{2+}$ : 4.22 eV; 294 nm) and the position of the lowest energy  $5d$  orbital ( $5d_1$  orbital) when substituted in the inorganic host.<sup>31–34</sup>

Considering the factors influencing the  $5d_1$  energy level, and thus the excitation wavelength of the  $\text{Ce}^{3+}$  or  $\text{Eu}^{2+}$  phosphor ( $4f \rightarrow 5d_1$ ), prior efforts have employed density functional theory (DFT) approaches to understand this transition.<sup>35–37</sup> These calculations are analogs to time-dependent DFT methods routinely employed for molecules, but they remain relatively uncommon in the solid state, primarily due to their high computational cost. In extended solids, particularly  $\text{Ce}^{3+}$ -doped phosphors, the energies and relative oscillator strengths of the  $4f \rightarrow 5d$  transitions can be calculated using lanthanide-centered defect clusters within the host crystal structure. For example, prior work by our group and others have performed quantum chemical, wave-function-based complete active-space self-consistent-field (CASSCF) and second-order many-body perturbation theory (CASPT2) calculations on  $\text{Ce}^{3+}$ -centered embedded clusters.<sup>38–43</sup> These calculations have been used to understand how local coordination environments can influence a phosphor's absorption/excitation band position and reveal rare-earth substitution site preference.<sup>37</sup> The difficulty with

these calculations is that they are mainly helpful in understanding a phosphor’s excitation spectrum *after* their discovery. Moreover, their computational cost is high; large supercells and hybrid functionals are required to model the dilute lanthanide substitution. Consequently, it is impractical to use high-throughput calculations to uncover the structure-composition trends that guide lanthanide excitation band position, thereby limiting prediction or screening potential.

Here, the  $5d_1$  excited state energy level position is effectively predicted through a new supervised machine learning model leveraging the extreme gradient boosting (XGBoost) algorithm. We constructed a dataset containing experimentally measured  $5d_1$  excitation energies for  $\text{Ce}^{3+}$  in 357 unique cation substitution sites across 337 different host materials, meticulously compiled from peer-reviewed studies and in-house measurements.  $\text{Ce}^{3+}$  is the first lanthanide ion considered due to the straightforward experimental measurement of the redshift  $D$ , and therefore  $5d_1$  energy, using optical spectroscopy. A feature set was engineered to reflect the dominant factors influencing the  $4f \rightarrow 5d_1$  optical transition by incorporating information on the local coordination environment, broader host crystal structure information, and chemical composition characteristics. Recursive feature elimination (RFE) and leave-one-group-out cross-validation (LOGO-CV) were employed to ensure model robustness and avoid over-fitting. The model’s effectiveness for identifying blue-excited materials was then experimentally validated by screening for  $\text{Ce}^{3+}$  substituted phosphors with excitation band positions that match standard InGaN LEDs. A promising candidate material,  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}$  substituted with  $\text{Ce}^{3+}$ , was identified, synthesized, and characterized to confirm the model’s predictions, emphasizing the value of our new approach in guiding the discovery and design of future phosphors.

## 2. Experimental Methods

### 2.1 Prediction Model Construction

The model to predict  $5d_1$  required data collection and feature engineering, as discussed below. Target values were collected from peer-reviewed publications regardless of the energy scale and then converted into eV for model training. Some of the most important features involved our group’s previously reported machine learning prediction models for relative permittivity ( $\epsilon_r$ , dielectric constant) and centroid shift ( $\epsilon_c$ ). These models were both updated and revised to get the most reliable input features for this work.<sup>44</sup> The relative permittivity model utilized a training data set of 2,254 values extracted from the Materials Project and cross-referenced with Pearson’s Crystal Database to ensure only experimentally reported compounds were used in the training set. The range of  $\epsilon_r$  value was further limited to between 0 and 8, reflecting the typical relative permittivity range for most inorganic host structures. XGBoost was employed for the  $\epsilon_r$  model, and the model was validated through leave-one-out-cross-validation (LOO-CV)<sup>45,46</sup>, achieving an  $R^2$  of 88.7% (**Figure S1a**), comparable with the previously published model. Subsequently, the centroid shift model was updated using the revised predicted relative permittivity values, and  $\epsilon_c$  was similarly optimized with XGBoost, yielding a LOO-CV  $R^2$  of 90.1% (**Figure S1b**).

The main model in this work, predicting the  $5d_1$  energy, employed the tree-based XGBoost algorithm due to its demonstrated effectiveness in handling small and noisy datasets while mitigating overfitting. This was crucial considering the relatively modest size of the training set. Hyperparameter optimization was conducted with leave-one-group-out-cross-validation. It is necessary to group targets by composition because select targets with polymorphs or multiple substitution sites in a single phosphor host typically have very similar feature sets. LOGO-CV can prevent potential data leakage from this shared information. The optimization process encompassed 11 hyperparameters of the XGBoost model, including tree boosters

that regulate individual tree models (*e.g.*, learning rate, max depth, and subsample), overfitting regularizations (*e.g.*, L1 and L2 regularization), and the base score parameter. The models were scored using mean absolute error (MAE) to quantify average model performance and assess actual deviations in predicted values, while recursive feature elimination was implemented to address multicollinearity and further limit overfitting.<sup>47</sup> After RFE, the feature set was reduced from 124 initial features to 44 scientifically relevant features (**Table S1**). The XGBoost model was finally retrained using this refined feature set, and hyperparameters were re-optimized with analysis by LOGO-CV. All model construction and implementation were done within the Python ecosystem, leveraging Scikit-learn libraries.<sup>48</sup>

## 2.2 Predicted Phosphor Synthesis and Characteristic Measurements

$\text{Ca}_2\text{SrSc}_6\text{O}_{12}:\text{Ce}^{3+}$  selected for experimental validation was synthesized by solid-state reaction starting from  $\text{CaCO}_3$  (Alfa Aesar, 99.0%),  $\text{SrCO}_3$  (Alfa Aesar, 99%),  $\text{Sc}_2\text{O}_3$  (Thermoscientific, 99.99%), and  $\text{CeO}_2$  (Sigma Aldrich, 99.995%). Each component was weighed in the appropriate stoichiometric ratio, with 1 mol% of the lanthanide activator added. The mixed starting reagents were ground in an agate mortar and pestle using hexane as a wetting medium and then further milled for 30 min in a high-energy ball mill (Spex 800 M Mixer/Mill). The mixture was pressed into a 6 mm diameter pellet and placed on a bed of sacrificial powder in an alumina crucible. The pellet was first heated to 1100°C for 10 h with a heating and cooling rate of 3°C/min under flowing 5%  $\text{H}_2$ /95%  $\text{N}_2$  gas. The product was then ground in hexane, pelletized again, and heated a second time to 1300°C for 8 h with a heating and cooling rate of 3°C/min under flowing 5%  $\text{H}_2$ /95%  $\text{N}_2$  gas.

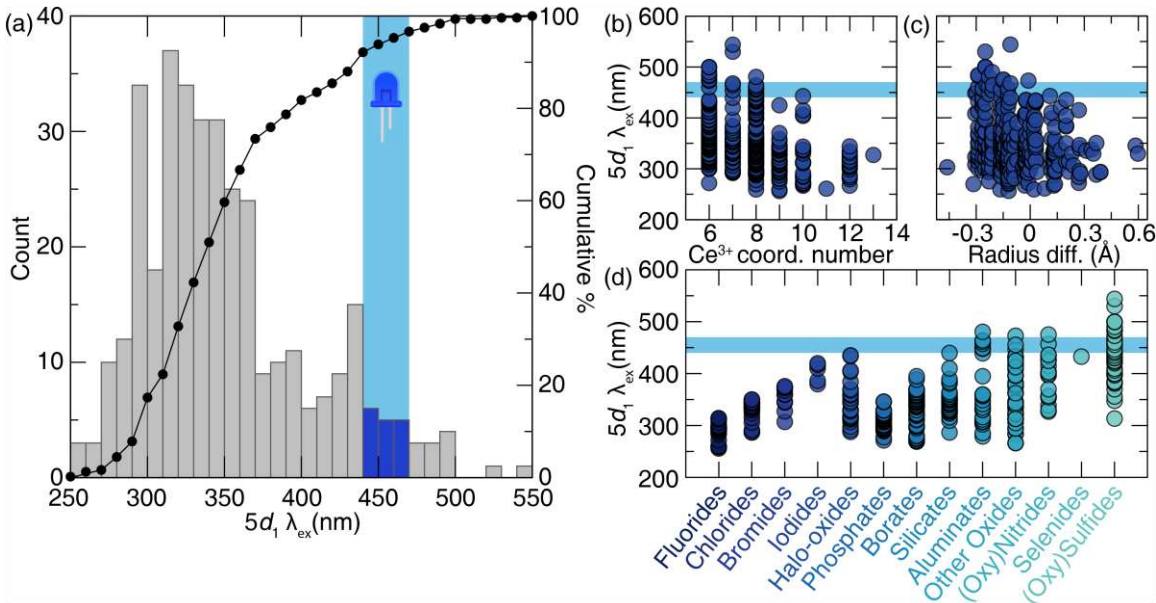
The product was characterized using powder X-ray diffraction on an X'Pert PANalytical Empyrean 3 equipped with Cu  $K\alpha$  radiation ( $\lambda = 1.54056 \text{ \AA}$ ). Additionally, high-resolution synchrotron powder X-ray diffractograms were collected at 9B beamline of PLS-II. Rietveld refinements were performed using the General Structural Analysis System II (GSAS-II) software.<sup>49</sup> The background was described using a Chebyshev-1 function, and the peak shapes were modeled using a pseudo-Voigt function. Photoluminescence measurements involved mixing the polycrystalline products in an optically transparent silicon epoxy (United Adhesives Inc., OP 4036) and depositing the combination onto a quartz slide (Chemglass). Photoluminescent excitation and emission and temperature-dependent luminescence measurements were obtained using a PTI fluorescence spectrophotometer with a 75 W xenon arc lamp for excitation. A Janis cryostat (VPF-100) was employed to establish a temperature-controlled environment from 80 K to 640 K. The photoluminescent quantum yield was determined following the method of de Mello *et al.*<sup>50</sup> using a Spectralon-coated integrating sphere (150 mm diameter, Labsphere). Photoluminescent lifetimes were measured using a 455 nm NanoLED equipped with the Horiba DeltaFlex Lifetime.

## 3. Results and Discussion

### 3.1 $5d_1$ Data Extraction and Feature Engineering

The construction of the machine learning model to predict the  $5d_1$  excitation energy commenced with the compilation of 357 experimentally measured  $5d_1$  energy levels, also reported as redshift values ( $D$ ), from the literature as well as data collected in-house.<sup>34</sup> These data represent 337 individual host structures, including seven crystal polymorphs and compounds with multiple possible crystallographically independent cation substitution sites in a single host crystal structure. Diffuse reflectance and photoluminescence excitation spectra were both used to obtain reliable  $5d_1$  target values. The distribution

of collected target values shows that the training set continuously covers the entire excitation wavelength range from UV to blue/green (**Figure 1a**). Notably, most of the population is concentrated in the UV range, with only ~4% of  $\text{Ce}^{3+}$  excitation target values falling in the typical blue LED emission range (440 nm - 470 nm), emphasizing the challenge of discovering blue-excited  $\text{Ce}^{3+}$  phosphors. The excitation wavelengths can be analyzed based on their meta-data to understand the physical and chemical properties that dictate excitation wavelength. The relationship between  $5d_1$  values and the coordination of the polyhedron occupied by  $\text{Ce}^{3+}$ , plotted in **Figure 1b**, highlights that higher  $\text{Ce}^{3+}$  coordination number (>8) rarely generates excitation in the blue range due to smaller crystal field splitting energies (and thus lower  $D$ ). A correlation between  $5d_1$  values and the radius difference between  $\text{Ce}^{3+}$  and the substituted cation further revealed that smaller sites tend to generate blue excitation due to stronger crystal field splitting (**Figure 1c**). When categorized by the host structure anion type, the target values also effectively demonstrate the trend of the nephelauxetic series<sup>21,25</sup> (**Figure 1d**), emphasizing the importance of compositional information such as electronegativity, electron affinity, and polarizability. This basic materials informatics analysis is already helpful for conceptualizing phosphor design and can also guide feature set engineering.



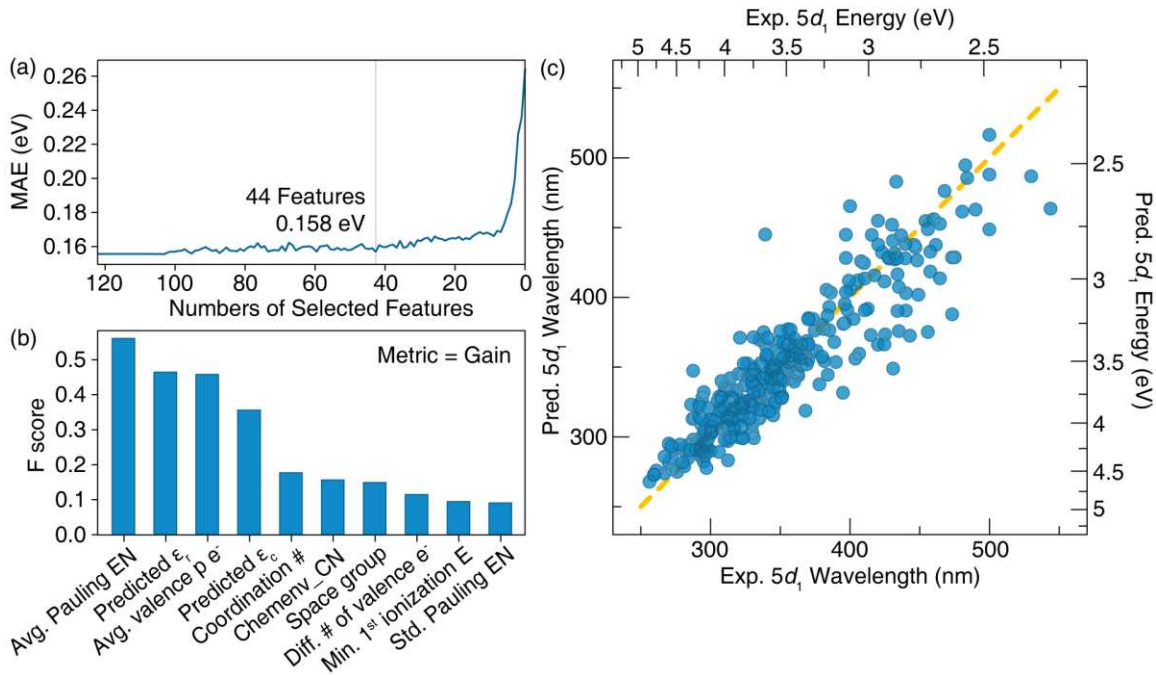
**Figure 1.** Target value distribution of the  $\text{Ce}^{3+}$ -substituted phosphors collected in the training set grouped by (a) excitation wavelength, (b)  $\text{Ce}^{3+}$  coordination number of the crystallographically independent substitution sites, (c) ionic radius difference between the substituted cation and  $\text{Ce}^{3+}$ , and (d) the host anion type. The blue-shaded region highlights conventional emission wavelengths of blue-emitting InGaN LED chips.

An optimal set of relevant features was subsequently developed to create a quantitative machine-learning model. An initial 124 features were assembled, encompassing characteristics from the local environment of the cation substitution site (13 features), host crystal structure information (14 features), and host composition (95 features), with the complete set listed in **Table S1**. The local cation site features are necessary to estimate crystal field splitting for the  $\text{Ce}^{3+}$   $5d$  orbitals. These features include the polyhedron volume, coordination number, and local symmetry. Features were also incorporated to describe the host crystal structure, which is essential for distinguishing polymorphic phosphors. These crystal structure attributes include space group number, volume of unit cell, volume per atom, volume per Z, and unit cell parameters, among other features. The cation substitution site and crystal structure information

were extracted from the Materials Project database or experimental crystallographic information files from the Inorganic Crystal Structure Database (ICSD) and Pearson’s Crystal Database (PCD).<sup>51–53</sup> In cases where the experimentally reported host crystal structure contains statistical site disorder, the *OrderDisorderedStructureTransformation* module from the Pymatgen library was used to predict a low-energy ordered arrangement from which the features were extracted.<sup>54</sup> The compositional features were finally developed by relating elemental properties to the stoichiometric composition. Each elemental feature (e.g., polarizability, electron affinity, etc.) was expanded into five variables (maximum, minimum, average, difference, and standard deviation) to capture multidimensional relationships between features and target values. Two physics-based features were also added: relative permittivity ( $\epsilon_r$ , dielectric constant) and centroid shift ( $\epsilon_c$ ) because they are understood to impact the nephelauxetic effect and, therefore, the  $5d_1$  position.<sup>44</sup> These values depend explicitly on the host material and necessitated updated versions of our group’s previously published supervised machine-learning regression models for these properties.

### 3.2 Machine Learning the $5d_1$ Energy Level

The  $5d_1$  energy level prediction model was first constructed using target values from 357 unique cation sites. XGBoost was selected for its effective overfitting regulation methods since the training set contains only a few hundred target values. During model training, LOGO-CV was implemented to ensure appropriate inclusion of the training data while mitigating potential biases associated with  $n$ -fold cross-validation. The initial model with all 124 features showed an MAE of  $\pm 0.159$  eV (RMSE = 0.160 eV) and an  $R^2$  of 84.3 % (**Figure S2**). It is worth noting that this  $\pm 0.16$  eV MAE manifests differently across the wavelength scale due to the non-linear relationship between energy and wavelength. For instance, the MAE when  $5d_1$  is between 3.50 eV and 3.65 eV corresponds to a wavelength difference of only approximately 15 nm (from  $\sim 354$  nm to  $\sim 339$  nm), whereas at lower energy, the same energy gap can result in a more substantial wavelength difference, such as a 27 nm gap between 2.60 eV ( $\sim 477$  nm) and 2.75 eV ( $\sim 450$  nm).



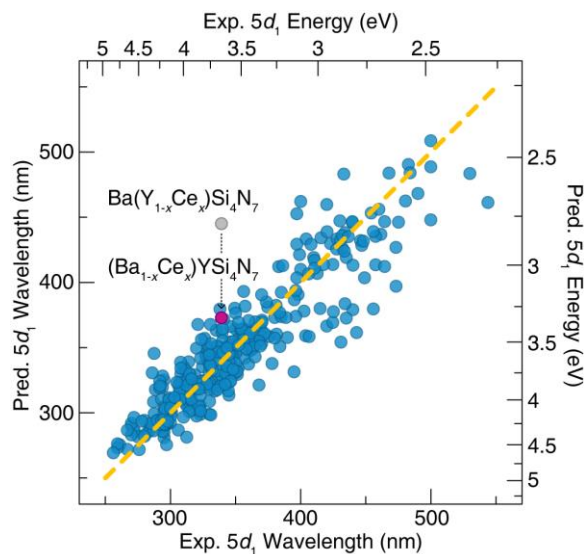
**Figure 2.** (a) RFE result with the initial XGBoost model. 44 features were chosen with minimal MAE. (b) Top 10 important features of this prediction model in terms of gain (F score). (c) Plot of experimental  $5d_1$  energy versus

predicted  $5d_1$  by the final model. The MAE of this initial model is  $\pm 0.153$  eV with  $R^2$  of 83.8 %

A notable advantage of XGBoost, and machine learning in general, is the ability to analyze feature importance, thereby facilitating the interpretation of the relationship between targets and features. Therefore, a revised model was prepared using recursive feature elimination to iteratively remove the least important feature(s) and track model performance with each refined feature set. **Figure 2a** illustrates the change in model performance starting from the complete 124 feature set. Removing the least important feature(s) in each step shows that the model performance remains nearly constant based on the MAE until there are only  $\sim 40$  features. Further feature reduction causes the model's error to exponentially increase when there are fewer than 10 features. A balance of features to performance suggests that 44 features are a reasonable number to build a reliable model. Analyzing the feature importance scores for the top 10 out of these 44 features indicates that the most critical features include the composition's average Pauling electronegativity, reflecting the host structure's nephelauxetic effect (**Figure 2b**). This is followed in importance by the predicted relative permittivity and centroid shift, corroborating the direct influence of the nephelauxetic effect on the  $5d_1$  position. The fifth most important feature, coordination number, supports the influence of crystal field splitting on the  $5d_1$  energy level. The next feature, chemenv\_CN, also relates to the coordination environment but considers geometrical properties too.<sup>55</sup> Table S1 also provides all 44 selected features, marked with bold text. The physical relevance of these significant features to the  $5d_1$  energy level supports and enhances our understanding of the underlying physical mechanisms governing the  $4f \rightarrow 5d_1$  transition.

The final operational XGBoost model re-trained using the 44 selected features exhibited an MAE of  $\pm 0.153$  eV (RMSE = 0.154 eV) and an  $R^2$  of 83.8 %, indicating satisfactory model performance (**Figure 2c**). This model shows modest improvements, but the smaller feature set reduces the likelihood of overfitting. The remaining prediction uncertainty likely stems from various sources. In some respects, it is simply a function of the limited data set size. However, it could also have originated from the raw experimental training data used to build the model. During the data cleaning process that was conducted before building the operational model, several training data problem cases were identified that could generate uncertainty. First, significant discrepancies were observed in the reported  $5d_1$  values when comparing the same compound across multiple independent publications and data collected in our lab during reproduction efforts. Sometimes the most reliable data was found in archived literature, while in-house measurements were more reliable in other instances. Secondly, publications rarely provide the excitation spectrum across different  $\text{Ce}^{3+}$  concentrations, which can affect the excitation peak position by altering lattice parameters or preferred site selectivity. For example, the excitation spectrum of  $\text{NaCaBO}_3:\text{Ce}^{3+}$  exhibited spectral broadening across six different  $\text{Ce}^{3+}$  concentrations (0.5 % to 5%), with  $\lambda_{\text{ex,max}}$  shifting between 347 nm to 355 nm,<sup>56</sup> whereas  $\text{BaSi}_7\text{N}_{10}:\text{Ce}^{3+}$  provided a  $\lambda_{\text{ex,max}}$  of 325 nm at 0.5% and 1%  $\text{Ce}^{3+}$  concentration, respectively.<sup>57</sup> However, many studies only provided one excitation spectrum at one optimal  $\text{Ce}^{3+}$  concentration. Consequently, the machine learning model did not employ the  $\text{Ce}^{3+}$  concentration as a feature, even though this information would likely improve the model's predictive accuracy. Furthermore, the  $4f \rightarrow 5d_1$  excitation peak is not always clearly reported in the literature or there can be conflicting reports. For instance, one publication reported a  $\lambda_{\text{ex,max}} = 320$  nm for  $\text{CaAl}_2\text{B}_2\text{O}_7:\text{Ce}^{3+}$ .<sup>58</sup> In contrast, another paper suggested that the strongest excitation peak can be deconvoluted into a strong peak at 310 nm ( $5d_2$ ) and a weak peak at 338 nm ( $5d_1$ ).<sup>59</sup> Data that were not abundantly unambiguous were removed from the training data set. Observations such as these show the complexity of accurately building a prediction model through experimental data and highlight the need for careful consideration of data selection in future refinements of the predictive model.





**Figure 3.** Plot of experimental  $5d_1$  energy versus predicted  $5d_1$  by the final model with fixed cation site information for  $\text{BaYSi}_4\text{N}_7:\text{Ce}^{3+}$ . This model with a hypothetical “ $(\text{Ba}_{1-x}\text{Ce}_x)\text{YSi}_4\text{N}_7$ ” has an MAE of  $\pm 0.153$  eV and an  $R^2$  of 83.8 %

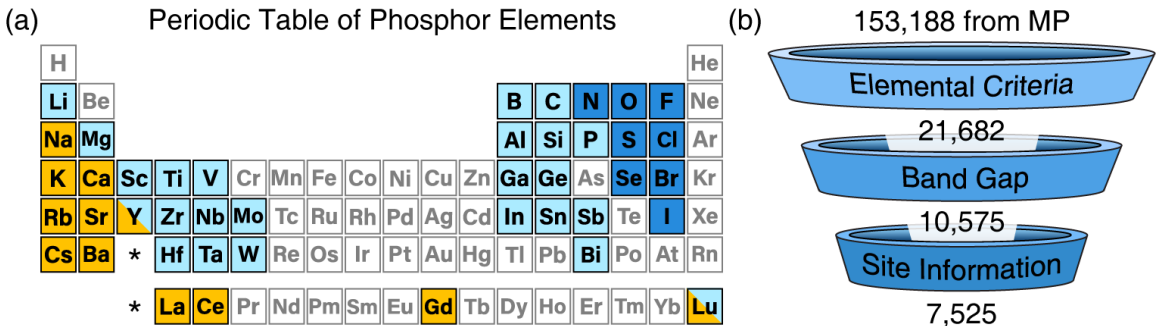
The model serves as more than just a regression prediction tool. It can also provide critical insights by evaluating reported data. For instance, during the first training phase, the data set contained  $\text{CaSnO}_3:\text{Ce}^{3+}$  with an excitation peak position of 246 nm (5.04 eV), as reported in the literature.<sup>60</sup> During early training runs, this compound was one of the most significant outliers in our trained model. Reproducing this compound in-house revealed the experimental excitation peak was closer to 378 nm (3.28 eV), which makes more crystal-chemical sense. Subsequent training iterations incorporated this updated value, and the model’s predictions aligned with experimental results, underscoring the use of data-driven techniques to refine and enhance data reliability.

Additionally, this model can aid in analyzing experimentally measured excitation spectra to help determine the most likely substitution site of an activator ion for phosphors containing multiple possible cation substitution sites. In systems with no strong energetic preference for the activator ion to substitute on a specific crystallographic site, the lanthanide may occupy multiple crystallographic positions concurrently, leading to an ensemble optical response. This phenomenon is evident in the training dataset, which includes 13 compounds reported with  $\text{Ce}^{3+}$  on multiple unique crystallographic substitution sites. Determining the exact location of lanthanide ions in the host crystal structure to deduce their coordination environment requires various complementary spectroscopic techniques. Yet, sometimes, the results may not be conclusive, and a best guess is provided. An interesting outcome following this type of substitutional assignment was observed with this model for  $\text{BaYSi}_4\text{N}_7:\text{Ce}^{3+}$ , shown as the highlighted point in **Figure 3**.<sup>61</sup> Conventional phosphor design principles suggest that the  $\text{Ce}^{3+}$  should preferentially substitute for cation sites with similar size, oxidation state, and coordination number. Some evidence is provided in the original report<sup>61</sup> supporting that  $\text{Ce}^{3+}$  may occupy the  $\text{Y}^{3+}$  site; however, when building the model with  $\text{Ce}^{3+}$  on the  $\text{Y}^{3+}$  site, a substantial discrepancy of 0.87 eV between the experimental (339 nm; 3.66 eV) and predicted (445 nm; 2.79 eV)  $5d_1$  values was found, rendering it the most significant outlier. A hypothetical model was created here that placed  $\text{Ce}^{3+}$  on the  $\text{Ba}^{2+}$  site instead. Even though they have different oxidation states,  $\text{Ce}^{3+}$  substituting for  $\text{Ba}^{2+}$  is relatively common, and this hypothetical model yielded an expected value of 373 nm (3.32 eV), closely matching the experimental measurement. This finding demonstrates that machine learning can aid in interpreting optical properties and offer valuable insights into experimental results.



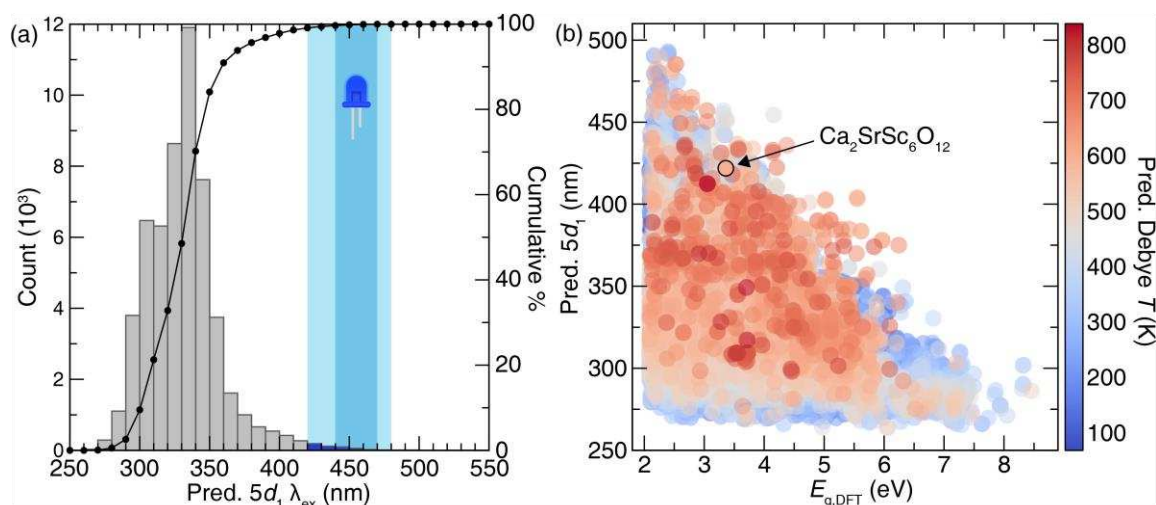
### 3.3 Screening for Ce<sup>3+</sup> Substituted Inorganic Phosphors with Target Excitation

To demonstrate the robustness, reliability, and utility of the final model for phosphor identification, the model was used to screen a wide dataset of candidate materials prior to experimentally investigating a promising candidate. All 153,188 materials listed in the Materials Project database,<sup>51</sup> including both experimentally reported and computationally predicted structures, were initially considered and then refined according to a set of heuristic design rules. The first of these rules was based on host composition. Potential phosphors were required to include the yellow-marked elements in **Figure 4**, which represent likely Ce<sup>3+</sup> substitution sites based on the literature.<sup>12</sup> These elements all have sufficiently large ionic radii and reasonable coordination numbers ranging from  $r_{\text{CN}=6}=1.01$  Å to  $r_{\text{CN}=12}=1.34$  Å to support Ce<sup>3+</sup> substitution.<sup>62</sup> The polyhedral backbone of the host was then limited to elements marked in light blue (cations) and dark blue (anions) based on common host chemistries contained in the training data. Y<sup>3+</sup> and Lu<sup>3+</sup> have been reported as both substitution sites or backbone cations due to their slightly smaller ionic radii than Ce<sup>3+</sup>. This list also excludes elements that could interfere with Ce<sup>3+</sup> photoluminescence (*e.g.*, Cr<sup>3+</sup>, Mn<sup>2+</sup>, Fe<sup>3+</sup>),<sup>63–65</sup> elements like Pb, Tl, and Cd due to sustainability and toxicity concerns, and expensive metals like Au, Pt, and Ir. After this filtering, 21,682 compounds that included at least one yellow and one dark blue-marked element, with optional light blue elements, were selected.



**Figure 4.** (a) The target elements for candidate phosphor host compositions. Elements forming the polyhedral backbone of the host structure are marked with the cation as light blue and the anion as dark blue. The possible Ce<sup>3+</sup> substitution site cations are marked in yellow. (b) Schematic diagram of selecting possible host structures from Materials Project.

In addition to elemental criteria, further screening was applied to ensure the suitability of the host compounds for Ce<sup>3+</sup> photoluminescence. Specifically, to prevent (thermal) ionization-induced temperature-dependent quenching of the Ce<sup>3+</sup> emission, the bandgap ( $E_g$ ) should be larger than the energy of target excitation wavelengths. Given that this work targets 450 nm excitation (2.76 eV), only host structures with calculated electronic  $E_{g,\text{DFT}}$  exceeding 2.1 eV were retained. This threshold accounts for the well-known underestimation of band gaps at the PBE level (which are available in the Materials Project). Considering this condition, the candidate set was further reduced to 10,575 compounds, ensuring a more reliable selection of potential hosts.<sup>66</sup> Next, compounds present in the training set or exhibiting improper crystal chemistry (*e.g.*, abnormal coordination numbers not purged in earlier down selection) were removed, reducing the dataset to 7,525 possible host crystal structures. Of these candidate hosts, around 90% have multiple crystallographically-distinct likely substitution sites for Ce<sup>3+</sup>, expanding the final set of potential lanthanide substitutions to 54,885 possible locations for Ce<sup>3+</sup>, for which their 5d<sub>1</sub> energy was predicted.



**Figure 5.** (a) Distribution of predicted excitation wavelengths for the 54,885  $\text{Ce}^{3+}$ -substituted cation sites of possible host structures. The blue-shaded region highlights the InGaN LED emission range. (b) A plot mapping the DFT-calculated band gap (PBE-level), predicted Debye temperature, and the  $5d_1$  prediction. The black-edged data point demonstrates  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}:\text{Ce}^{3+}$ .

Plotting these 54,885 potential  $\text{Ce}^{3+}$ -substituted cation sites from the 7,525 unique host candidates in **Figure 5a** reveals that most  $\text{Ce}^{3+}$  phosphors are excited in the UV range, agreeing with the literature. Only 494 cation sites ( $\sim 0.9\%$ ) in 211 unique host structures ( $2.8\%$ ) were predicted to have blue excitation, highlighting the challenge of discovering blue-excited  $\text{Ce}^{3+}$ -based phosphors. Plotting the predicted excitation wavelengths against the host anion type showed that the most promising candidates are nitride/selenide/sulfide-containing compounds or garnet-based oxide crystal structures with gallium or germanium (**Figure S3**). These are classically understood to have the most significant redshifts and compatibility with blue LEDs.<sup>8,21,24</sup>

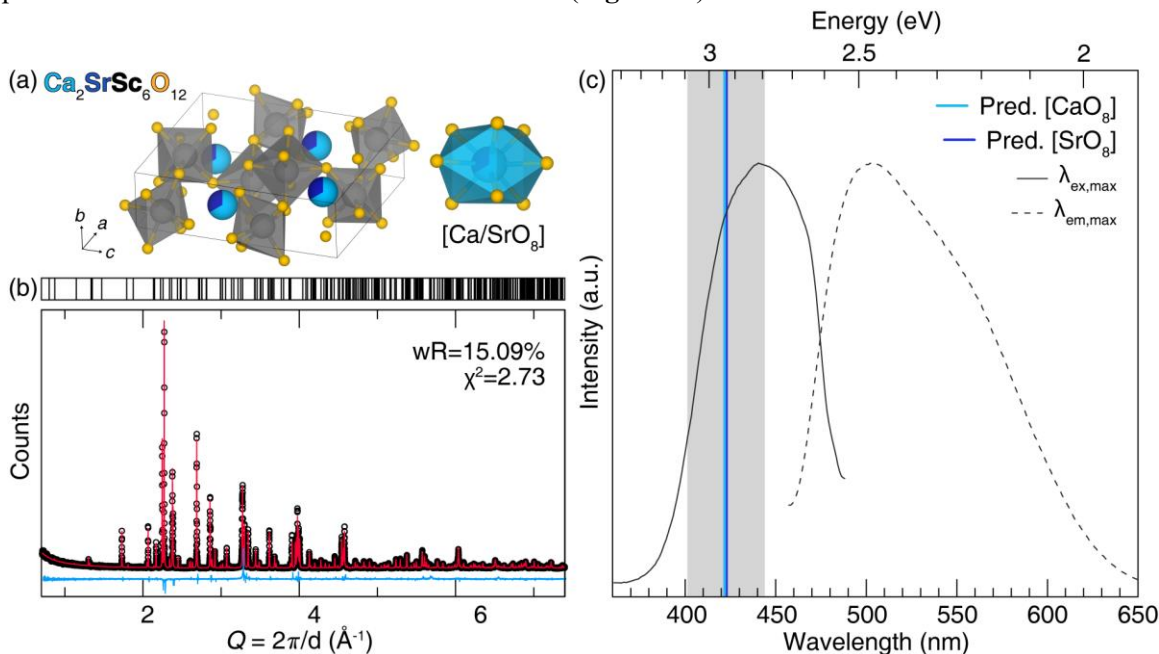
Selecting a blue-excited phosphor from this carefully screened list appears straightforward once the machine-learning predictions are made. However, compatibility with a blue LED chip doesn't guarantee practicality. To further accelerate phosphor host selection, this model was paired with additional proxies for phosphor performance as part of our selection pipeline. **Figure 5b** presents a sorting diagram of predicted  $5d_1$  values versus the host material's band gap calculated at the DFT(PBE) level of theory ( $E_{g,\text{DFT}}$ ). A wide bandgap is crucial for thermally robust  $\text{Ce}^{3+}$  photoluminescence. An interesting trend is observed here. There is a negative correlation between excitation wavelength and bandgap.  $\text{Ce}^{3+}$  substituted compounds with very wide bandgaps ( $\sim 6$  eV) are unlikely to be useful with blue LED chips due to high ionicity, which leads to a small centroid shift and weak crystal field splitting. Therefore, finding materials with sufficiently long wavelength excitation for industry applications will require a careful balance with the host material's bandgap. A second proxy that is well-regarded for correlating with a phosphor's photoluminescent quantum yield (PLQY) is the host structure's Debye temperature.<sup>10,12</sup> Phosphors with relatively high Debye temperatures and, thus, greater structural rigidity tend to exhibit higher PLQYs. Our group previously developed a machine learning model to predict the Debye temperature of host structures.<sup>5</sup> By incorporating this model into our analysis, as shown by the color of the data points in **Figure 5b**, we observe that many materials within the blue LED excitation range are predicted to have moderate to low Debye temperatures ( $\leq 500$  K). However, some data points exhibit Debye temperatures greater than 600 K, indicating the potential to discover thermally robust, high photoluminescence quantum yield (PLQY) novel phosphors

within this prediction set. Upon further analysis, many of the materials in this range are associated with garnet-type crystal structures or are nitrides, which are common blue-excited  $\text{Ce}^{3+}$  phosphor hosts. The list of predicted  $E_{\text{g,DFT}}$ , Debye  $T$ , and  $5d_1$  levels is available on GitHub (<https://github.com/BrgochGroup>).

After a thorough review of potential phosphor host compounds from this screened set, a novel composition was identified for synthesis.  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}:\text{Ce}^{3+}$  stood out due to its unique structure compared to other candidates while still suggesting impressive performance potential ( $5d_1 = 2.94$  eV or 422 nm, an  $E_{\text{g,DFT}} = 3.39$  eV, and a Debye temperature = 594 K).

### 3.4 Experimental Validation of the $5d_1$ Model

The experimental validation of the selected potential inorganic phosphor involved solid-state synthesis.  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}$  crystallizes in the orthorhombic space group  $Pnma$  (no. 62) with a single crystallographically distinct statistically-mixed  $[(\text{Ca}/\text{Sr})\text{O}_8]$  substitution site (**Figure 6a**).<sup>67</sup> Synchrotron powder X-ray diffraction was collected for  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}:\text{Ce}^{3+}$  and analyzed using the Rietveld refinement method starting from the published  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}$  structure file (PCD #1934108).  $\text{Ce}^{3+}$  was omitted from the refinement due to its low concentration. The diffractogram verified the sample's purity (**Figure 6b**), while the refined lattice parameters provided in **Tables S2** and **S3** agree with the original structure. The  $5d_1$  energy prediction was performed using the experimentally refined structure, accounting for site disorder with pymatgen. This yielded a prediction of 422 nm (2.94 eV) for  $\text{Ce}^{3+}$  substitution on both the  $[\text{CaO}_8]$  and  $[\text{SrO}_8]$  sites, closely matching the original prediction. Experimentally measuring the photoluminescence excitation spectrum showed the peak falling at a maximum ( $\lambda_{\text{max}}$ ) of 440 nm (2.82 eV), deviating by only 18 nm (0.12 eV) from the predicted value—well within the model's MAE (**Figure 6c**).

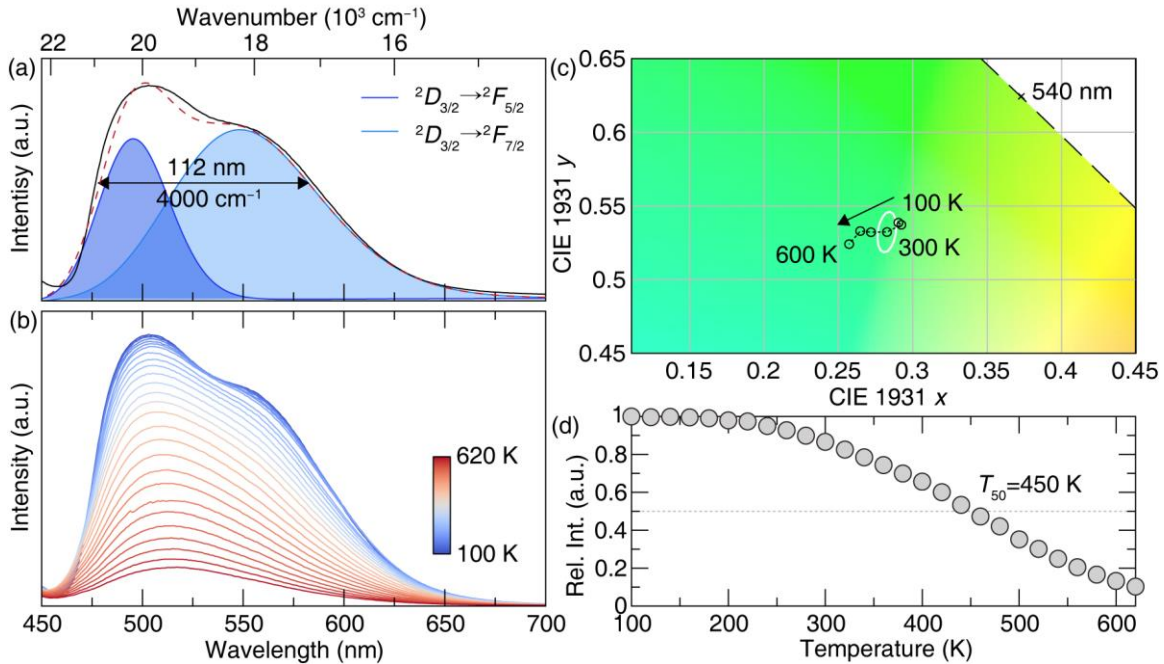


**Figure 6.** (a) Crystal structure of  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}$  (PCD #1934108) and  $\text{Ca}^{2+}/\text{Sr}^{2+}$  site for the prediction. (b) Rietveld refinement of  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}:\text{Ce}^{3+}$  starting from PCD #1934108. (c) Normalized excitation and emission spectrum of  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}:\text{Ce}^{3+}$  with predicted  $5d_1$  energy in wavelength scale of  $[\text{CaO}_8]$  and  $[\text{SrO}_8]$  site. The predicted  $5d_1$  of  $[\text{CaO}_8]$  is 422 nm, while  $[\text{SrO}_8]$  is 422 nm. The experimental excitation peak position is 440 nm, and the emission peak is 503 nm.

The remaining optical properties were analyzed with a blue LED to assess the phosphor's practical

performance. Collecting the room temperature photoluminescence emission ( $\lambda_{\text{ex}} = 440 \text{ nm}$ ) spectrum revealed the phosphor exhibits a light green color ( $\lambda_{\text{em,max}}=503 \text{ nm}$ ) with an extremely broad ( $112 \text{ nm}$ ;  $4,000 \text{ cm}^{-1}$ ) *fwhm*, spanning from  $450 \text{ nm}$  to  $650 \text{ nm}$  (**Figure 7a**). The emission spectrum at  $100 \text{ K}$  was deconvoluted into two distinct peaks ( $\lambda_{\text{max},1}=497 \text{ nm}$ ;  $20,134 \text{ cm}^{-1}$ ,  $\lambda_{\text{max},2}=553 \text{ nm}$ ;  $18,074 \text{ cm}^{-1}$ ) separated by  $\approx 2,000 \text{ cm}^{-1}$  as expected for the spin-orbit-coupled  $5d_1 \rightarrow {}^2F_{5/2}, {}^2F_{7/2}$  transitions of  $\text{Ce}^{3+}$ .<sup>68,69</sup> Measuring the temperature-dependent photoluminescence, plotted in **Figure 7b**, shows the two peaks merge due to thermal broadening as the temperature increases. The emission peak also shifted slightly from  $503 \text{ nm}$  at  $100 \text{ K}$  to  $517 \text{ nm}$  at  $650 \text{ K}$ . These changes lead to a shift in the CIE coordinates falling outside the 3-step MacAdam ellipse across the entire temperature range, indicating the change in emission color as a function of temperature is noticeable to the standard human observer (**Figure 7c**). Analyzing the change in emission intensity over the specified temperature range, as plotted in **Figure 7d**, reveals that the optical response of the phosphor was only slightly affected from  $100 \text{ K}$  to room temperature, with a decrease of approximately 5% before the thermal performance of the phosphor is impacted. The  $T_{50}$ , defined here as the temperature at which the luminescence intensity reaches 50% of its maximum value, was found to be  $450 \text{ K}$ , surpassing the current expectations set by the U.S. Department of Energy ( $423 \text{ K}$ ).<sup>3</sup>

The measured PLQY is slightly low at 33(2) %. However, it is essential to emphasize that this synthesis is unoptimized, as it lacks post-processing, fluxes/mineralizers, or other conventional methods commonly employed to enhance phosphor performance. The lower PLQY could also be attributed to charge-compensating defects resulting from  $\text{Ce}^{3+}$  substituting for  $\text{Ca}^{2+}/\text{Sr}^{2+}$ . These defects are known to quench the PLQY, yet aliovalent substitutions are common in literature. To improve the performance of this phosphor, strategies such as co-substitution with charge-compensating ions or additional chemical modifications to minimize the detrimental defects and enhance the PLQY, ultimately optimizing the material for practical use. Finally, photoluminescence lifetime measurement was collected for completeness and yielded a decay time of  $37.8(6) \text{ ns}$  (**Figure S4**). Overall, the performance of  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}:\text{Ce}^{3+}$  aligns with the expected performance based on the calculated Debye temperature and (PBE-level) band gap,<sup>12</sup> and there is excellent agreement with the new  $5d_1$  model created here, ensuring blue-LED compatibility.



**Figure 7.** (a) Deconvoluted emission spectrum of  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}:\text{Ce}^{3+}$  at 100 K under 450 nm excitation. (b) Temperature-dependent emission spectra measured from 100 K to 620 K. (c) The emission spectra from 100 K to 600 K with 100 K interval, plotted in CIE 1931 space, showing chromatic stability of  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}:\text{Ce}^{3+}$ . (d) Relative integrated emission intensity as a function of temperature.

#### 4. Conclusion

This study introduced a supervised machine-learning approach for predicting the  $5d_1$  excitation energy levels in  $\text{Ce}^{3+}$  phosphors. The XGBoost model was trained on a diverse dataset comprising 357 experimentally measured  $5d_1$  energy levels from 337 different host structures, achieving a mean prediction accuracy of  $\pm 0.15$  eV. This high level of performance was accomplished through feature engineering, leave-one-group-out cross-validation techniques, and recursive feature elimination, resulting in a refined set of 44 scientifically relevant features. Feature importance analysis provided further insights into the physical mechanisms determining the  $5d_1$  energy levels, including average Pauling electronegativity, relative permittivity, and centroid shifts, among others. Beyond its capabilities for linear regression prediction, this  $5d_1$  model also demonstrated effectiveness as a tool for analyzing the excitation spectra of multi-cation site phosphors and for interpreting outlier detection. Finally, this new machine learning model allowed the rapid prediction of more than 50,000 possible phosphor  $5d_1$  energies, leading to the successful identification, synthesis, and characterization of a novel, blue-excited phosphor, green-emitting composition:  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}:\text{Ce}^{3+}$ . This discovery was the ultimate validation of the model's effectiveness in practical phosphor discovery.

#### ORCID

Nakyung Lee: 0000-0001-5048-1816

Małgorzata Sójka: 0000-0001-9346-8929

Seán Kavanagh: 0000-0003-4577-9647

Docheon Ahn: 0000-0003-1814-5418

David Scanlon: 0000-0001-9174-8601

Jakoah Brgoch: 0000-0002-1406-1352

#### Notes

The authors declare no competing financial interest.

#### 5. Acknowledgments

The authors thank the U.S. National Science Foundation (DMR-2349319), the Welch Foundation (E-2181), the Texas Center for Superconductivity at the University of Houston (TcSUH), and the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2024-00419831) for supporting this work.

#### 6. Data Availability

The  $\text{Ce}^{3+}$   $5d_1$  raw training data, feature generation code, and prediction model GitHub ([https://github.com/BrgochGroup/Ce\\_5d1\\_Prediction](https://github.com/BrgochGroup/Ce_5d1_Prediction)) along with a permanent version of record at Zenodo (<https://doi.org/10.5281/zenodo.14872504>).

#### 7. Supporting Information



The Supporting Information is available free of charge at <https://doi.org/10.1038/nmat4270>.

- Plots of the revised relative permittivity, centroid shift models, and the initial 5d1 prediction model before recursive feature elimination, list of feature names for the initial and after RFE feature sets, predicted value distribution by the host anion type, Rietveld refined results and lifetime measurement at room temperature for  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}:\text{Ce}^{3+}$ .

## 8. References

- (1) Pust, P.; Schmidt, P. J.; Schnick, W. A Revolution in Lighting. *Nat. Mater.* **2015**, *14* (5), 454–458. <https://doi.org/10.1038/nmat4270>.
- (2) Pimpulkar, S.; Speck, J. S.; DenBaars, S. P.; Nakamura, S. Prospects for LED Lighting. *Nat. Photonics* **2009**, *3* (4), 180–182. <https://doi.org/10.1038/nphoton.2009.32>.
- (3) Department of Energy. 2022 Solid-State Lighting R&D Opportunities. <https://www.energy.gov/eere/ssl/articles/doe-publishes-2022-solid-state-lighting-rd-opportunities>
- (4) Hariyani, S.; Sójka, M.; Setlur, A.; Brgoch, J. A Guide to Comprehensive Phosphor Discovery for Solid-State Lighting. *Nat. Rev. Mater.* **2023**, *8* (11), 759–775. <https://doi.org/10.1038/s41578-023-00605-6>.
- (5) Höppe, H. A. Recent Developments in the Field of Inorganic Phosphors. *Angew. Chem. Int. Ed.* **2009**, *48* (20), 3572–3582. <https://doi.org/10.1002/anie.200804005>.
- (6) Xie, R.-J.; Hirotsaki, N. Silicon-Based Oxynitride and Nitride Phosphors for White LEDs—A Review. *Sci. Technol. Adv. Mater.* **2007**, *8* (7–8), 588. <https://doi.org/10.1016/j.stam.2007.08.005>.
- (7) Li, G.; Tian, Y.; Zhao, Y.; Lin, J. Recent Progress in Luminescence Tuning of  $\text{Ce}^{3+}$  and  $\text{Eu}^{2+}$ -Activated Phosphors for P-WLEDs. *Chem. Soc. Rev.* **2015**, *44* (23), 8688–8713. <https://doi.org/10.1039/C4CS00446A>.
- (8) Xia, Z.; Meijerink, A.  $\text{Ce}^{3+}$ -Doped Garnet Phosphors: Composition Modification, Luminescence Properties and Applications. *Chem. Soc. Rev.* **2017**, *46* (1), 275–299. <https://doi.org/10.1039/C6CS00551A>.
- (9) Kim, Y. H.; Arunkumar, P.; Kim, B. Y.; Unithrattil, S.; Kim, E.; Moon, S.-H.; Hyun, J. Y.; Kim, K. H.; Lee, D.; Lee, J.-S.; Im, W. B. A Zero-Thermal-Quenching Phosphor. *Nat. Mater.* **2017**, *16* (5), 543–550. <https://doi.org/10.1038/nmat4843>.
- (10) Denault, K. A.; Brgoch, J.; Klob, S. D.; Gaultois, M. W.; Siewenie, J.; Page, K.; Seshadri, R. Average and Local Structure, Debye Temperature, and Structural Rigidity in Some Oxide Compounds Related to Phosphor Hosts. *ACS Appl. Mater. Interfaces* **2015**, *7* (13), 7264–7272. <https://doi.org/10.1021/acsami.5b00445>.
- (11) Zhuo, Y.; Hariyani, S.; Armijo, E.; Abolade Lawson, Z.; Brgoch, J. Evaluating Thermal Quenching Temperature in  $\text{Eu}^{3+}$ -Substituted Oxide Phosphors via Machine Learning. *ACS Appl. Mater. Interfaces* **2020**, *12* (5), 5244–5250. <https://doi.org/10.1021/acsami.9b16065>.
- (12) Zhuo, Y.; Mansouri Tehrani, A.; Oliynyk, A. O.; Duke, A. C.; Brgoch, J. Identifying an Efficient, Thermally Robust Inorganic Phosphor Host via Machine Learning. *Nat. Commun.* **2018**, *9* (1), 4377. <https://doi.org/10.1038/s41467-018-06625-z>.
- (13) Yeh, C.-W.; Chen, W.-T.; Liu, R.-S.; Hu, S.-F.; Sheu, H.-S.; Chen, J.-M.; Hintzen, H. T. Origin of Thermal Degradation of  $\text{Sr}_{2-x}\text{Si}_5\text{N}_8:\text{Eu}_x$  Phosphors in Air for Light-Emitting Diodes. *J. Am. Chem. Soc.* **2012**, *134* (34), 14108–14117. <https://doi.org/10.1021/ja304754b>.
- (14) Jin, M.; Zeng, Z.; Fu, H.; Wang, S.; Yin, Z.; Zhai, X.; Zhang, Q.; Du, Y. Strain-Negligible  $\text{Eu}^{2+}$  Doping Enabled Color-Tunable Harsh Condition-Resistant Perovskite Nanocrystals for Superior Light-Emitting Diodes. *JACS Au* **2023**, *3* (1), 216–226. <https://doi.org/10.1021/jacsau.2c00593>.
- (15) Jiang, L.; Jiang, X.; Zhang, Y.; Wang, C.; Liu, P.; Lv, G.; Su, Y. Multiobjective Machine Learning-Assisted Discovery of a Novel Cyan–Green Garnet: Ce Phosphors with Excellent Thermal Stability. *ACS Appl. Mater. Interfaces* **2022**, *14* (13), 15426–15436. <https://doi.org/10.1021/acsami.2c02698>.
- (16) Pust, P.; Weiler, V.; Hecht, C.; Tücks, A.; Wochnik, A. S.; Henß, A.-K.; Wiechert, D.; Scheu, C.; Schmidt, P. J.; Schnick, W. Narrow-Band Red-Emitting  $\text{Sr}[\text{LiAl}_3\text{N}_4]:\text{Eu}^{2+}$  as a next-Generation LED-Phosphor Material. *Nat. Mater.* **2014**, *13* (9), 891–896. <https://doi.org/10.1038/nmat4012>.
- (17) Wang, L.; Xie, R.-J.; Suehiro, T.; Takeda, T.; Hirotsaki, N. Down-Conversion Nitride Materials for Solid State Lighting: Recent Advances and Perspectives. *Chem. Rev.* **2018**, *118* (4), 1951–2009. <https://doi.org/10.1021/acs.chemrev.7b00284>.
- (18) Stephan, M.; Zachau, M.; Gröting, M.; Karplak, O.; Eyert, V.; Mishra, K. C.; Schmidt, P. C. A Theoretical Investigation of  $4f \rightarrow 5d$  Transition of Trivalent Rare Earth Ions in Fluorides and Complex Oxides. *J. Lumin.* **2005**, *114* (3), 255–266. <https://doi.org/10.1016/j.jlumin.2005.01.008>.

- (19) Qin, X.; Liu, X.; Huang, W.; Bettinelli, M.; Liu, X. Lanthanide-Activated Phosphors Based on 4f-5d Optical Transitions: Theoretical and Experimental Aspects. *Chem. Rev.* **2017**, *117* (5), 4488–4527. <https://doi.org/10.1021/acs.chemrev.6b00691>.
- (20) Xia, Z.; Ma, C.; Molokeev, M. S.; Liu, Q.; Rickert, K.; Poeppelmeier, K. R. Chemical Unit Cosubstitution and Tuning of Photoluminescence in the  $\text{Ca}_2(\text{Al}_{1-x}\text{Mg}_x)(\text{Al}_{1-x}\text{Si}_{1+x})\text{O}_7\text{:Eu}^{2+}$  Phosphor. *J. Am. Chem. Soc.* **2015**, *137* (39), 12494–12497. <https://doi.org/10.1021/jacs.5b08315>.
- (21) Jørgensen, Chr. K. Fractional Charges, Integral Oxidation States and the Nephelauxetic Effect in the Five Transition Groups. *Helv. Chim. Acta* **1967**, *50* (S1), 131–146. <https://doi.org/10.1002/hlca.19670500911>.
- (22) Tchougréeff, A. L.; Dronskowski, R. Nephelauxetic Effect Revisited. *Int. J. Quantum Chem.* **2009**, *109* (11), 2606–2621. <https://doi.org/10.1002/qua.21989>.
- (23) Dorenbos, P. Relating the Energy of the Xe 5d<sub>1</sub> Configuration of Ce<sup>3+</sup> in Inorganic Compounds with Anion Polarizability and Cation Electronegativity. *Phys. Rev. B* **2002**, *65* (23), 235110. <https://doi.org/10.1103/PhysRevB.65.235110>.
- (24) Dorenbos, P. Lanthanide 4f-Electron Binding Energies and the Nephelauxetic Effect in Wide Band Gap Compounds. *J. Lumin.* **2013**, *136*, 122–129. <https://doi.org/10.1016/j.jlumin.2012.11.030>.
- (25) Dorenbos, P. The Nephelauxetic Effect on the Electron Binding Energy in the 4f<sup>n</sup> Ground State of Lanthanides in Compounds. *J. Lumin.* **2019**, *214*, 116536. <https://doi.org/10.1016/j.jlumin.2019.116536>.
- (26) Dorenbos, P. Crystal Field Splitting of Lanthanide 4f<sub>n-1</sub>5d-Levels in Inorganic Compounds. *J. Alloys Compd.* **2002**, *341* (1), 156–159. [https://doi.org/10.1016/S0925-8388\(02\)00056-7](https://doi.org/10.1016/S0925-8388(02)00056-7).
- (27) Dorenbos, P. 5d-level energies of and the crystalline environment. III. Oxides containing ionic complexes. *Phys. Rev. B* **2001**, *64*, 125117. <https://journals.aps.org/prb/abstract/10.1103/PhysRevB.64.125117>.
- (28) Dorenbos, P. 5d-Level Energies of Ce<sup>3+</sup> and the Crystalline Environment. I. Fluoride Compounds. *Phys. Rev. B* **2000**, *62* (23), 15640–15649. <https://doi.org/10.1103/PhysRevB.62.15640>.
- (29) Dorenbos, P. 5d-Level Energies of Ce<sup>3+</sup> and the Crystalline Environment. IV. Aluminates and “Simple” Oxides. *J. Lumin.* **2002**, *99* (3), 283–299. [https://doi.org/10.1016/S0022-2313\(02\)00347-2](https://doi.org/10.1016/S0022-2313(02)00347-2).
- (30) Newman, D. J. Theory of Lanthanide Crystal Fields. *Adv. Phys.* **1971**, *20* (84), 197–256. <https://doi.org/10.1080/00018737100101241>.
- (31) Dorenbos, P. Relation between Eu<sup>2+</sup> and Ce<sup>3+</sup> f ↔ d-Transition Energies in Inorganic Compounds. *J. Phys. Condens. Matter* **2003**, *15* (27), 4797. <https://doi.org/10.1088/0953-8984/15/27/311>.
- (32) Dorenbos, P. Electronic Structure and Optical Properties of the Lanthanide Activated RE<sub>3</sub>(Al<sub>1-x</sub>Ga<sub>x</sub>)<sub>5</sub>O<sub>12</sub> (RE=Gd, Y, Lu) Garnet Compounds. *J. Lumin.* **2013**, *134*, 310–318. <https://doi.org/10.1016/j.jlumin.2012.08.028>.
- (33) Dorenbos, P. Energy of the First 4f<sup>n</sup> → 4f<sup>n-1</sup>5d Transition of Eu<sup>2+</sup> in Inorganic Compounds. *J. Lumin.* **2003**, *104* (4), 239–260. [https://doi.org/10.1016/S0022-2313\(03\)00078-4](https://doi.org/10.1016/S0022-2313(03)00078-4).
- (34) Dorenbos, P. Ce<sup>3+</sup> 5d-Centroid Shift and Vacuum Referred 4f-Electron Binding Energies of All Lanthanide Impurities in 150 Different Compounds. *J. Lumin.* **2013**, *135*, 93–104. <https://doi.org/10.1016/j.jlumin.2012.09.034>.
- (35) Gracia, J.; Seijo, L.; Barandiarán, Z.; Curulla, D.; Niemansverdriet, H.; van Gennip, W. Ab Initio Calculations on the Local Structure and the 4f–5d Absorption and Emission Spectra of Ce<sup>3+</sup>-Doped YAG. *J. Lumin.* **2008**, *128* (8), 1248–1254. <https://doi.org/10.1016/j.jlumin.2007.12.027>.
- (36) Ning, L.; Wang, Y.; Wang, Z.; Jin, W.; Huang, S.; Duan, C.; Zhang, Y.; Chen, W.; Liang, H. First-Principles Study on Site Preference and 4f → 5d Transitions of Ce<sup>3+</sup> in Sr<sub>3</sub>AlO<sub>4</sub>F. *J. Phys. Chem. A* **2014**, *118*, 6, 986–992. <https://pubs.acs.org/doi/10.1021/jp410542y>.
- (37) Wen, J.; Ning, L.; Duan, C.-K.; Zhan, S.; Huang, Y.; Zhang, J.; Yin, M. First-Principles Study on Structural, Electronic, and Spectroscopic Properties of γ-Ca<sub>2</sub>SiO<sub>4</sub>:Ce<sup>3+</sup> Phosphors. *J. Phys. Chem. A* **2015**, *119* (29), 8031–8039. <https://doi.org/10.1021/acs.jpca.5b02873>.
- (38) Zhong, J.; Hariyani, S.; Zhuo, Y.; Zhao, W.; Liu, X.; Wen, J.; Brgoch, J. Combining Experiment and Computation to Elucidate the Optical Properties of Ce<sup>3+</sup> in Ba<sub>5</sub>Si<sub>8</sub>O<sub>21</sub>. *Phys. Chem. Chem. Phys.* **2020**, *22* (4), 2327–2336. <https://doi.org/10.1039/C9CP05576B>.
- (39) Zhong, J.; Zhao, W.; Zhuo, Y.; Yan, C.; Wen, J.; Brgoch, J. Understanding the Blue-Emitting Orthoborate Phosphor NaBaBO<sub>3</sub>:Ce<sup>3+</sup> through Experiment and Computation. *J. Mater. Chem. C* **2019**, *7* (3), 654–662. <https://doi.org/10.1039/C8TC05629C>.
- (40) Ning, L.; Wu, C.; Li, L.; Lin, L.; Duan, C.; Zhang, Y.; Seijo, L. First-Principles Study on Structural Properties and 4f → 5d Transitions of Locally Charge-Compensated Ce<sup>3+</sup> in CaF<sub>2</sub>. *J. Phys. Chem. C* **2012**, *116*, 34, 18419–18426. <https://pubs.acs.org/doi/10.1021/jp305593h>.
- (41) Wen, J.; Ning, L.; Duan, C.-K.; Chen, Y.; Zhang, Y.; Yin, M. A Theoretical Study on the Structural and Energy



- Spectral Properties of  $\text{Ce}^{3+}$  Ions Doped in Various Fluoride Compounds. *J. Phys. Chem. C* **2012**, *116* (38), 20513–20521. <https://doi.org/10.1021/jp306357d>.
- (42) Finley, J.; Malmqvist, P.-Å.; Roos, B. O.; Serrano-Andrés, L. The Multi-State CASPT2 Method. *Chem. Phys. Lett.* **1998**, *288* (2), 299–306. [https://doi.org/10.1016/S0009-2614\(98\)00252-8](https://doi.org/10.1016/S0009-2614(98)00252-8).
  - (43) Andersson, Kerstin.; Malmqvist, P. Aake.; Roos, B. O.; Sadlej, A. J.; Wolinski, Krzysztof. Second-Order Perturbation Theory with a CASSCF Reference Function. *J. Phys. Chem.* **1990**, *94* (14), 5483–5488. <https://doi.org/10.1021/j100377a012>.
  - (44) Zhuo, Y.; Hariyani, S.; You, S.; Dorenbos, P.; Brgoch, J. Machine Learning *5d*-Level Centroid Shift of  $\text{Ce}^{3+}$  Inorganic Phosphors. *J. Appl. Phys.* **2020**, *128* (1), 013104. <https://doi.org/10.1063/5.0012434>.
  - (45) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System *KDD'16*. **2016**, 785–794 <https://dl.acm.org/doi/10.1145/2939672.2939785> (accessed 2025-01-13).
  - (46) Efron, B.; Tibshirani, R. Improvements on Cross-Validation: The .632+ Bootstrap Method *JASA*. **1997**, *92*, 438, 548–560. <https://www.jstor.org/stable/2965703?origin=crossref&seq=1>.
  - (47) Li, D.; Li, K.; Xia, Y.; Dong, J.; Lu, R. Joint Hybrid Recursive Feature Elimination Based Channel Selection and ResGCN for Cross Session MI Recognition. *Sci. Rep.* **2024**, *14* (1), 23549. <https://doi.org/10.1038/s41598-024-73536-z>.
  - (48) Pedregosa, F.; Pedregosa, F.; Varoquaux, G.; Varoquaux, G.; Org, N.; Gramfort, A.; Gramfort, A.; Michel, V.; Michel, V.; Fr, L.; Thirion, B.; Thirion, B.; Grisel, O.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Dubourg, V.; Vanderplas, J.; Passos, A.; Tp, A.; Cournapeau, D. Scikit-Learn: Machine Learning in Python. *Mach. Learn. PYTHON*.
  - (49) Toby, B. H.; Von Dreele, R. B. GSAS-II: The Genesis of a Modern Open-Source All Purpose Crystallography Software Package. *J. Appl. Crystallogr.* **2013**, *46* (2), 544–549. <https://doi.org/10.1107/S0021889813003531>.
  - (50) de Mello, J. C.; Wittmann, H. F.; Friend, R. H. An Improved Experimental Determination of External Photoluminescence Quantum Efficiency. *Adv. Mater.* **1997**, *9* (3), 230–232. <https://doi.org/10.1002/adma.19970090308>.
  - (51) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1* (1), 011002. <https://doi.org/10.1063/1.4812323>.
  - (52) Zagorac, D.; Müller, H.; Ruehl, S.; Zagorac, J.; Rehme, S. Recent Developments in the Inorganic Crystal Structure Database: Theoretical Crystal Structure Data and Related Features. *J. Appl. Crystallogr.* **2019**, *52* (5), 918–925. <https://doi.org/10.1107/S160057671900997X>.
  - (53) Pearson's Crystal Data: Crystal Structure Database for Inorganic Compounds. *Choice Reviews Online*. **2008**, *45*, 45-3800-45–3800. <https://doi.org/10.5860/CHOICE.45-3800>.
  - (54) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (Pymatgen): A Robust, Open-Source Python Library for Materials Analysis. *Comput. Mater. Sci.* **2013**, *68*, 314–319. <https://doi.org/10.1016/j.commatsci.2012.10.028>.
  - (55) Waroquiers, D.; George, J.; Horton, M.; Schenk, S.; Persson, K. A.; Rignanese, G.-M.; Gonze, X.; Hautier, G. ChemEnv: A Fast and Robust Coordination Environment Identification Tool. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **2020**, *76* (Pt 4), 683–695. <https://doi.org/10.1107/S2052520620007994>.
  - (56) Zhang, X.; Song, J.; Zhou, C.; Zhou, L.; Gong, M. High Efficiency and Broadband Blue-Emitting  $\text{NaCaBO}_3$ :  $\text{Ce}^{3+}$  Phosphor for NUV Light-Emitting Diodes. *J. Lumin.* **2014**, *149*, 69–74. <https://doi.org/10.1016/j.jlumin.2014.01.012>.
  - (57) Li, Y. Q.; Delsing, A. C. A.; Metslaar, R.; de With, G.; Hintzen, H. T. Photoluminescence Properties of Rare-Earth Activated  $\text{BaSi}_7\text{N}_{10}$ . *J. Alloys Compd.* **2009**, *487* (1), 28–33. <https://doi.org/10.1016/j.jallcom.2009.08.019>.
  - (58) You, H.; Hong, G. Luminescence and Energy Transfer Phenomena of Several Rare Earth Ions in the  $\text{CaAl}_2\text{B}_2\text{O}_7$ . *Mater. Res. Bull.* **1997**, *32* (6), 785–790. [https://doi.org/10.1016/S0025-5408\(97\)00042-1](https://doi.org/10.1016/S0025-5408(97)00042-1).
  - (59) Yang, H.; Li, C.; He, H.; Zhang, G.; Qi, Z.; Su, Q. Luminescent Properties of  $\text{RE}^{3+}$ -Activated  $\text{CaAl}_2\text{B}_2\text{O}_7$  ( $\text{RE}=\text{Tb}, \text{Ce}$ ) in VUV-Visible Region. *J. Lumin.* **2007**, *124* (2), 235–240. <https://doi.org/10.1016/j.jlumin.2006.03.008>.
  - (60) Chen, Y.-X.; Tsai, M.-H.; Tseng, I.-W.; Yen, H.-C.; Huang, C.-L. The Photoluminescence of Single-Phase Warm White-Light-Emitting Luminescence Using  $\text{CaSnO}_3$ :  $\text{Ce}^{3+}$ /  $\text{Mn}^{4+}$ /  $\text{Dy}^{3+}$  Phosphors. *J. Asian Ceram. Soc.* **2021**, *9* (3), 1055–1066. <https://doi.org/10.1080/21870764.2021.1936764>.
  - (61) Li, Y. Q.; de With, G.; Hintzen, H. T. Synthesis, Structure, and Luminescence Properties of  $\text{Eu}^{2+}$  and  $\text{Ce}^{3+}$  Activated  $\text{BaYSi}_4\text{N}_7$ . *J. Alloys Compd.* **2004**, *385* (1), 1–11. <https://doi.org/10.1016/j.jallcom.2004.04.134>.
  - (62) Shannon, R. D. Revised Effective Ionic Radii and Systematic Studies of Interatomic Distances in Halides and Chalcogenides. *Acta Crystallogr. A* **1976**, *32* (5), 751–767. <https://doi.org/10.1107/S0567739476001551>.

- (63) Zhao, F.; Song, Z.; Liu, Q. Advances in Chromium-Activated Phosphors for Near-Infrared Light Sources. *Laser Photonics Rev.* **2022**, *16* (11), 2200380. <https://doi.org/10.1002/lpor.202200380>.
- (64) Adachi, S.  $\text{Mn}^{4+}$  vs  $\text{Mn}^{2+}$ —A Comparative Study as Efficient Activator Ions in Phosphor Materials: A Review. *J. Lumin.* **2023**, *263*, 119993. <https://doi.org/10.1016/j.jlumin.2023.119993>.
- (65) Sun, B.; Zhang, H.; Wang, Y. Research Progress on  $\text{Fe}^{3+}$ -Activated near-Infrared Phosphor. *Rev. Adv. Mater. Sci.* **2024**, *63* (1). <https://doi.org/10.1515/rams-2023-0160>.
- (66) Brgoch, J.; DenBaars, S. P.; Seshadri, R. Proxies from Ab Initio Calculations for Screening Efficient  $\text{Ce}^{3+}$  Phosphor Hosts. *J. Phys. Chem. C* **2013**, *117* (35), 17955–17959. <https://doi.org/10.1021/jp405858e>.
- (67) Kwestroo, W.; Langereis, C.; Nabben, H. Alkaline Earth Scandates. *Mater. Res. Bull.* **1982**, *17* (5), 641–646. [https://doi.org/10.1016/0025-5408\(82\)90047-2](https://doi.org/10.1016/0025-5408(82)90047-2).
- (68) Dieke, G. H. Spectra and Energy Levels of Rare Earth Ions in Crystals, **1968**. <https://www.osti.gov/biblio/4826175>.
- (69) Wegh, R. T.; Meijerink, A.; Lamminmäki, R.-J.; Jorma Hölsä. Extending Dieke's Diagram. *J. Lumin.* **2000**, *87–89*, 1002–1004. [https://doi.org/10.1016/S0022-2313\(99\)00506-2](https://doi.org/10.1016/S0022-2313(99)00506-2).

## Supporting Information

### Machine Learning a Phosphor's Excitation Band Position

Nakyung Lee<sup>1,2,†</sup>, Małgorzata Sójka<sup>1,2,†</sup>, Annie La<sup>1,2</sup>, Syna Sharma<sup>1,2</sup>, Seán Kavanagh<sup>3</sup>, Docheon Ahn<sup>4</sup>,  
David O. Scanlon<sup>5,\*</sup>, Jakoah Brgoch<sup>1,2,\*</sup>

<sup>1</sup>*Department of Chemistry, University of Houston, Houston, Texas 77204, USA*

<sup>2</sup>*Texas Center for Superconductivity, University of Houston, Houston, Texas 77204, USA*

<sup>3</sup>*Harvard University Center for the Environment, Cambridge, Massachusetts 02138, United States*

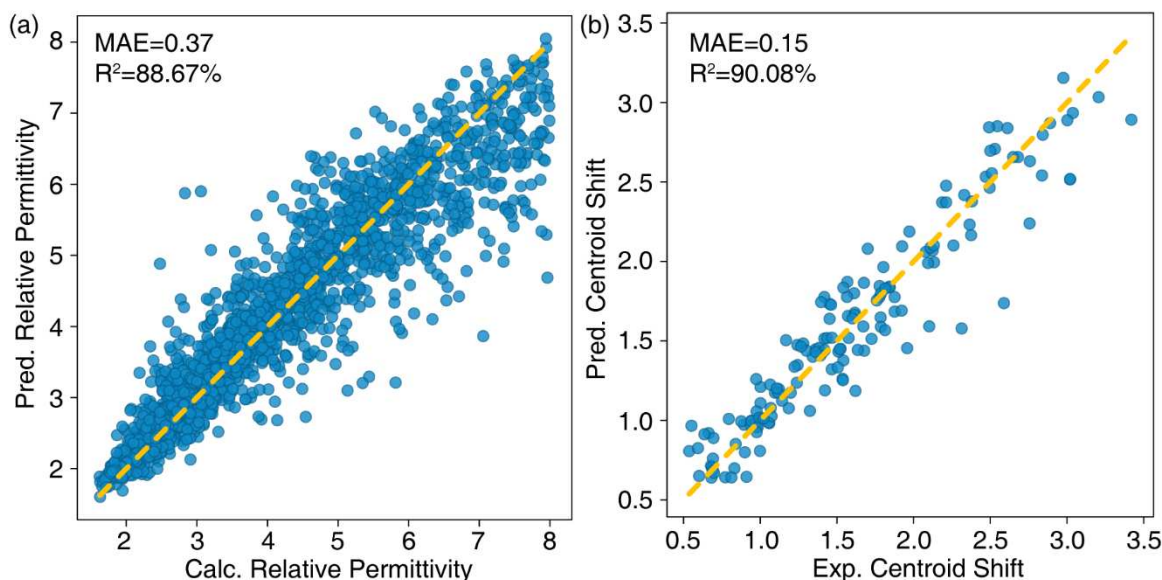
<sup>4</sup>*PLS-II Beamline Department, Pohang Accelerator Laboratory, POSTECH, Pohang 37673, Republic of Korea*

<sup>5</sup>*School of Chemistry, University of Birmingham, Birmingham, UK*

<sup>†</sup>These authors contributed equally to this work

[\\*jbrgoch@uh.edu](mailto:jbrgoch@uh.edu) (J. Brgoch)

[\\*d.o.scanlon@bham.ac.uk](mailto:d.o.scanlon@bham.ac.uk) (D. O. Scanlon)

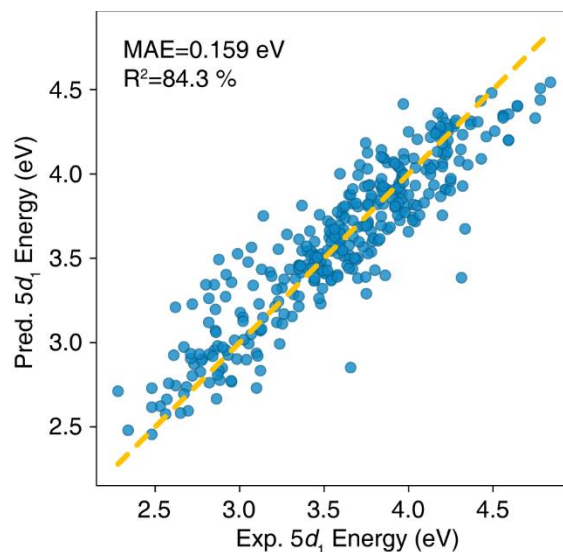


**Figure S1.** Leave-One-Out cross-validation of (a) the reconstructed relative permittivity prediction model and (b) the reconstructed centroid shift prediction model.

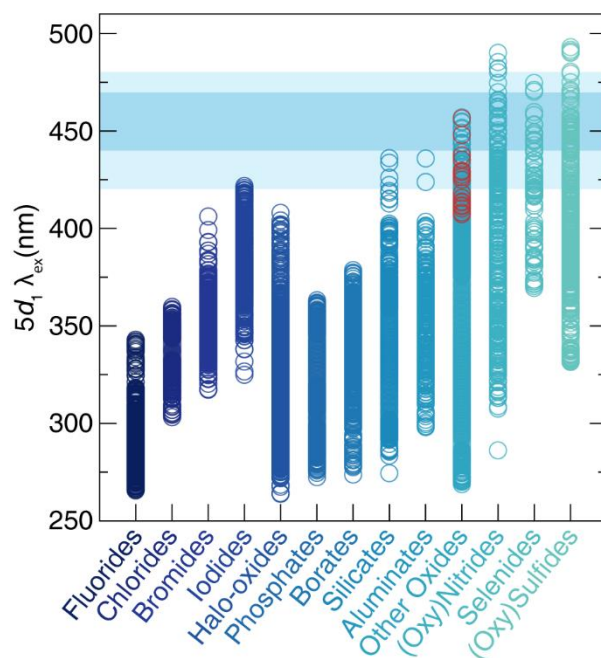
**Table S1.** Initial feature set for the  $5d_1$  prediction model. The final 44 features via RFE for the final model marked with bold font.

| Index     | Feature Name                           | Index     | Feature Name   |
|-----------|--|-----------|--|
| <b>0</b>  | <b>Predicted centroid shift</b>        | <b>24</b> | <b>b</b>   |
| <b>1</b>  | <b>Predicted relative permittivity</b> | 25        | c  |
| 2         | Polyhedron volume                      | 26        | alpha  |
| <b>3</b>  | <b>Coordination number</b>             | 27        | beta   |
| 4         | Point group                            | <b>28</b> | <b>gamma</b>   |
| <b>5</b>  | <b>Cation ionic radius</b>             | 29        | Avg. atomic weight                                       |
| <b>6</b>  | <b>Dopant ionic radius</b>             | <b>30</b> | <b>Avg. Mendeleev number</b>                             |
| 7         | Ionic radii difference                 | 31        | Avg. crystal radius                                      |
| 8         | Min. metal ligand bond length          | <b>32</b> | <b>Avg. Pauling electronegativity</b>                    |
| 9         | Max. metal ligand bond length          | <b>33</b> | <b>Avg. number of valence electrons</b>                  |
| 10        | Mean. Metal ligand bond length         | <b>34</b> | <b>Avg. Gilmore number of valence electron</b>           |
| 11        | Distortion index                       | 35        | Avg. valence s   |
| 12        | Cation site Madelung potential         | <b>36</b> | <b>Avg. valence p</b>                                    |
| 13        | CSM                                    | <b>37</b> | <b>Avg. valence d</b>                                    |
| <b>14</b> | <b>Chemenv_CN</b>                      | <b>38</b> | <b>Avg. 1<sup>st</sup> ionization potential (kJ/mol)</b> |
| <b>15</b> | <b>Space group number</b>              | <b>39</b> | <b>Avg. polarizability (Å<sup>3</sup>)</b>               |
| 16        | Crystal system                         | 40        | Avg. electron affinity (kJ/mol)                          |
| 17        | Polar axis                             | <b>41</b> | <b>Avg. density of element (g/ml)</b>                    |
| 18        | Inversion center                       | <b>42</b> | <b>Avg. specific heat (J/g·K)</b>                        |
| <b>19</b> | <b>Unit cell volume</b>                | 43        | Avg. heat of fusion (kJ/mol)                             |
| <b>20</b> | <b>Unit cell volume per Z</b>          | <b>44</b> | <b>Avg. heat of vaporization (kJ/mol)</b>                |
| 21        | Unit cell volume per atom              | 45        | Avg. thermal conductivity (W/m·K)                        |
| 22        | Density                                | 46        | Avg. heat atomization                                    |
| <b>23</b> | <b>a</b>                               | 47        | Avg. cohesive energy                                     |

| Index     | Feature Name  | Index      | Feature Name   |
|-----------|---|------------|--|
| 48        | Diff. atomic weight                                 | 86         | Min. atomic weight                                       |
| 49        | Diff. Mendeleev number                              | 87         | Min. Mendeleev number                                    |
| 50        | Diff. crystal radius                                | 88         | Min. crystal radius                                      |
| 51        | Diff. Pauling electronegativity                     | 89         | Min. Pauling electronegativity                           |
| <b>52</b> | <b>Diff. number of valence electrons</b>            | 90         | Min. number of valence electrons                         |
| <b>53</b> | <b>Diff. Gilmore number of valence electron</b>     | 91         | Min. Gilmore number of valence electron                  |
| 54        | Diff. valence s                                     | 92         | Min. valence s   |
| 55        | Diff. valence p                                     | 93         | Min. valence p   |
| 56        | Diff. valence d                                     | 94         | Min. valence d   |
| 57        | Diff. 1 <sup>st</sup> ionization potential (kJ/mol) | <b>95</b>  | <b>Min. 1<sup>st</sup> ionization potential (kJ/mol)</b> |
| <b>58</b> | <b>Diff. polarizability (Å<sup>3</sup>)</b>         | 96         | Min. polarizability (Å <sup>3</sup> )                    |
| <b>59</b> | <b>Diff. electron affinity (kJ/mol)</b>             | 97         | Min. electron affinity (kJ/mol)                          |
| 60        | Diff. density of element (g/ml)                     | 98         | Min. density of element (g/ml)                           |
| <b>61</b> | <b>Diff. specific heat (J/g·K)</b>                  | 99         | Min. specific heat (J/g·K)                               |
| 62        | Diff. heat of fusion (kJ/mol)                       | 100        | Min. heat of fusion (kJ/mol)                             |
| <b>63</b> | <b>Diff. heat of vaporization (kJ/mol)</b>          | <b>101</b> | <b>Min. heat of vaporization (kJ/mol)</b>                |
| 64        | Diff. thermal conductivity (W/m·K)                  | 102        | Min. thermal conductivity (W/m·K)                        |
| 65        | Diff. heat atomization                              | <b>103</b> | <b>Min. heat atomization</b>                             |
| 66        | Diff. cohesive energy                               | 104        | Min. cohesive energy                                     |
| 67        | Max. atomic weight                                  | <b>105</b> | <b>Std. atomic weight</b>                                |
| 68        | Max. Mendeleev number                               | 106        | Std. Mendeleev number                                    |
| <b>69</b> | <b>Max. crystal radius</b>                          | <b>107</b> | <b>Std. crystal radius</b>                               |
| 70        | Max. Pauling electronegativity                      | <b>108</b> | <b>Std. Pauling electronegativity</b>                    |
| 71        | Max. number of valence electrons                    | <b>109</b> | <b>Std. number of valence electrons</b>                  |
| 72        | Max. Gilmore number of valence electron             | 110        | Std. Gilmore number of valence electron                  |
| 73        | Max. valence s                                      | 111        | Std. valence s   |
| 74        | Max. valence p                                      | <b>112</b> | <b>Std. valence p</b>                                    |
| 75        | Max. valence d                                      | <b>113</b> | <b>Std. valence d</b>                                    |
| 76        | Max. 1 <sup>st</sup> ionization potential (kJ/mol)  | 114        | Std. 1 <sup>st</sup> ionization potential (kJ/mol)       |
| 77        | Max. polarizability (Å <sup>3</sup> )               | 115        | Std. polarizability (Å <sup>3</sup> )                    |
| 78        | Max. electron affinity (kJ/mol)                     | <b>116</b> | <b>Std. electron affinity (kJ/mol)</b>                   |
| 79        | Max. density of element (g/ml)                      | 117        | Std. density of element (g/ml)                           |
| 80        | Max. specific heat (J/g·K)                          | <b>118</b> | <b>Std. specific heat (J/g·K)</b>                        |
| <b>81</b> | <b>Max. heat of fusion (kJ/mol)</b>                 | 119        | Std. heat of fusion (kJ/mol)                             |
| <b>82</b> | <b>Max. heat of vaporization (kJ/mol)</b>           | 120        | Std. heat of vaporization (kJ/mol)                       |
| 83        | Max. thermal conductivity (W/m·K)                   | 121        | Std. thermal conductivity (W/m·K)                        |
| 84        | Max. heat atomization                               | <b>122</b> | <b>Std. heat atomization</b>                             |
| 85        | Max. cohesive energy                                | 123        | Std. cohesive energy                                     |



**Figure S2.** Plot of experimental  $5d_1$  energy versus predicted  $5d_1$  by the initial  $5d_1$  prediction model with 124 features.



**Figure S3.** Distribution of predicted excitation wavelengths of  $\text{Ce}^{3+}$  in the 54,885 cation sites screened from Materials Project, grouped by the host anion type. Red marked values are from the garnet structure type with gallium or germanium. The entire prediction set of excitation wavelength values is available on GitHub (<https://github.com/BrgochGroup>).

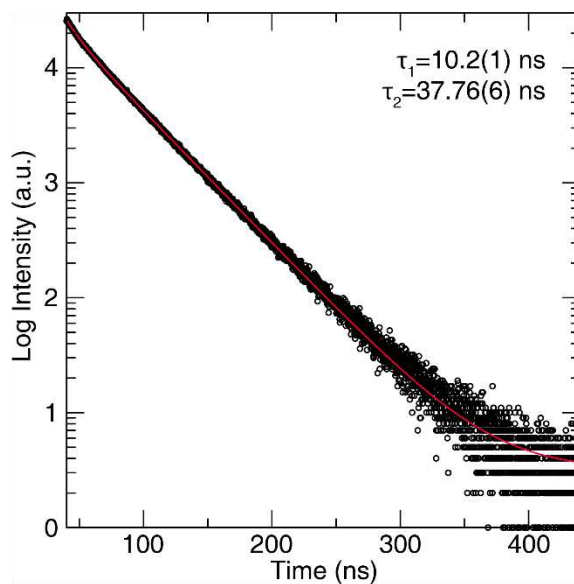
**Table S2.** Rietveld refinement data and statistics for  $\text{Ca}_2\text{SrSc}_6\text{O}_{12}$

|                         |   |
|-------------------------|---|
| Refined formula         | $\text{Ca}_2\text{SrSc}_6\text{O}_{12}$ |
| Radiation $\lambda$ (Å) | 1.546                                   |
| $2\theta$ range (°)     | 10-120                                  |
| Temperature (K)         | 295                                     |

|                          |                 |
|--------------------------|-----------------|
| Crystal system           | Orthorhombic    |
| Space group; <i>Z</i>    | <i>Pnma</i> ; 4 |
| <i>a</i>                 | 9.54741         |
| <i>b</i>                 | 3.16048         |
| <i>c</i>                 | 11.18849        |
| Volume (Å <sup>3</sup> ) | 337.606         |
| <i>wR</i>                | 15.089          |
| $\chi^2$                 | 2.73            |

**Table S3.** The refined atomic positions for Ca<sub>2</sub>SrSc<sub>6</sub>O<sub>12</sub>

| Atom | Wyck. Pos. | <i>x</i>  | <i>y</i> | <i>z</i>  | U <sub>iso</sub> | Occ.     |
|------|------------|-----------|----------|-----------|------------------|----------|
| Ca1  | 4c         | 0.2448(2) | 1/4      | 0.6535(1) | 0.0025(4)        | 0.644(5) |
| Sr1  | 4c         | 0.2448(2) | 1/4      | 0.6535(1) | 0.0025(4)        | 0.356(5) |
| Sc1  | 4c         | 0.0797(1) | 1/4      | 0.3939(1) | 0.0005(5)        | 1.000    |
| Sc2  | 4c         | 0.5705(2) | 1/4      | 0.6138(2) | 0.0013(5)        | 1.000    |
| O1   | 4c         | 0.0786(6) | 1/4      | 0.0832(5) | 0.011(2)         | 1.000    |
| O2   | 4c         | 0.2899(6) | 1/4      | 0.3276(5) | 0.009(2)         | 1.000    |
| O3   | 4c         | 0.3738(5) | 1/4      | 0.0219(5) | 0.003(2)         | 1.000    |
| O4   | 4c         | 0.4746(7) | 1/4      | 0.7858(5) | 0.015(2)         | 1.000    |



**Figure S4.** The lifetime measurement of Ca<sub>2</sub>SrSc<sub>6</sub>O<sub>12</sub>:Ce<sup>3+</sup> at room temperature under 455 nm excitation, fitted with bi-exponential decay.