

Screening High-Performance Hybrid Halides Scintillators: A Comprehensive Analysis and Prediction Model

Maxim Molokeev,* Nicolay Golovnev, Andrey Zolotov, Shuai Zhang,* and Zhiguo Xia*



Cite This: *Chem. Mater.* 2025, 37, 1255–1263



Read Online

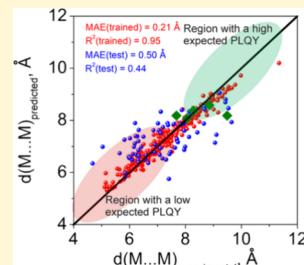
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Machine learning models were applied to predict the scintillation performances of organic–inorganic hybrid metal halides (OIMHs), focusing on their photoluminescent quantum yield (PLQY). Random Forest and Decision Tree algorithms identified the most critical structural parameter of organic molecules influencing the M···M distance between metal ions and correlated PLQY value, with an optimal distance of approximately 8 Å correlating with enhanced luminescence efficiency. This prediction was experimentally validated through the synthesis of several OIMH compounds, demonstrating strong agreement between predicted and measured PLQY values. The machine learning approach not only enabled the screening of efficient compounds but also deepened the understanding of how structural factors, such as the structure of organic molecules, govern scintillation properties. These findings underscore the potential of machine learning in accelerating the development of next-generation luminescent materials with improved performance, offering a powerful tool for future material design and optimization.



1. INTRODUCTION

The technology of X-ray detection has recently gained significant research interest and undergone rapid development, which is primarily driven by the increasing demand for its applications in various fields such as industry detection, biomedical imaging, etc.^{1–5} Two broad categories of methods are employed for X-ray detection: direct and indirect approaches. In the direct ones, X-rays are directly absorbed to generate electronic signals. However, this approach is known to have high costs and lower efficiency. On the other hand, the indirect method involves converting X-rays into ultraviolet visible (UV-vis) light using scintillators.⁶ This provides a more cost-effective solution and is easier to implement in industrial settings. Commercially available scintillators can be divided into inorganic and organic scintillators. Inorganic single-crystal scintillators such as Bi₄Ge₃O₁₂ (BGO), Lu₃A₅O₁₂:Ce³⁺ (LuAG:Ce), and CsI:TI are widely used for X-ray detection due to their high light yield (LY), suppressed light crosstalk, and strong attenuation ability to X-rays. However, the conventional preparation processes for these inorganic scintillators involve high-temperature and time-consuming. In contrast, organic–inorganic metal halides (OIMHs) scintillators offer several advantages, such as simplified synthesis process and low-cost preparation processes. As a result, there is an urgent need to develop high-performance OIMHs scintillators that are both cost-effective and easy to synthesize, to serve as potential replacements for conventional materials. Since the performance of scintillators is closely tied to their LY, and the PLQY further plays a crucial role in determining LY, optimizing these factors is essential for advancing scintillator technology.

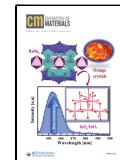
There is currently considerable research focused on enhancing the PLQY in OIMHs. For instance, it is proposed that rigid cations can maintain a stable structure through hydrogen bonds formed between organic and halogen anions.⁷ This mechanism of rigidity helps to restrict the movement of luminescence centers, thereby promoting the radiative recombination process. Furthermore, it has been observed that an optimal Mn···Mn distance is crucial for efficient emission of Mn²⁺ ions, and the organic component somehow contributes to a longer Mn···Mn distance of 8.38 Å. The distance between Mn ions significantly affects the photoluminescence properties.^{8–12} For example, a longer d(Mn···Mn) distance prevents Mn–Mn coupling, resulting in a strong green emission band.^{8,9} Conversely, a closer d(Mn···Mn) distance within a cluster or pair induces Mn–Mn ferromagnetic spin coupling during pair excitation, leading to a redshift in the emission band.⁹ Another notable finding is the competition between the closest d(Mn···Mn) distance and the energy transfer rate of Mn²⁺. The unique zero-dimensional (0D) isolated structure of OIMHs allows for a much longer d(Mn···Mn) distance across all luminescence centers of Mn²⁺, which enhances spontaneous emission. This universal regulatory strategy enables the achievement of highly efficient, narrow-band green emission.¹²

Received: November 19, 2024

Revised: January 12, 2025

Accepted: January 13, 2025

Published: January 17, 2025



Our research group previously conducted a systematic screening of various OIMHs containing ns² metal ions, such as Pb²⁺, Sn²⁺, Sb³⁺, and Bi³⁺, using Machine Learning methods.¹³ The study revealed that the PLQY of these 0D halides increases as the M···M average distance increases (M are metal ions), the M local symmetry decreases, and the MX_n polyhedron distortion increases. Among these parameters, the d(M···M) distance was found to have the most significant influence, accounting for 89% importance according to the Random Forest method. The minimal distance d(M···M) of approximately 8 Å has emerged as a critical factor in achieving high PLQY in OIMHs phosphors and scintillators. However, it remains a challenge for chemists to fully comprehend and effectively control this parameter to manipulate and optimize PLQY. Gaining a comprehensive understanding of how the structure and composition of the organic component in OIMHs influence the minimal d(M···M) distance is highly desirable. To elucidate the relationship between the organic molecule and the minimal d(M···M) distance in a series of compounds, Supervised Machine Learning methods can be employed. Explainable models, such as Decision Tree¹⁴ and Random Forest [15], are well-known “white-box” methods capable of extracting meaningful rules.^{13–18} The current research has successfully utilized these models to extract such rules.

Based on these findings, a suggested model was used to test several compounds. Ultimately, we synthesized five compounds to validate this model, and observed a strong agreement between the predicted values from the machine learning results and the various parameters of the synthesized compounds. Additionally, we proposed estimating the “average” Förster radius, d₀, from a comprehensive data set of compounds. This estimation was achieved by fitting the data using the formula PLQY = 100 × d⁶/(d⁶ + d₀⁶), where d represents the shortest M···M distance. The fitting analysis revealed that a value of d₀ = 8.0 Å effectively separates the samples into two distinct classes, characterized by high and low PLQY values, respectively.

2. EXPERIMENTAL SECTION

2.1. Sample Preparation. *Materials.* Cetylpyridinium bromide (C16PyBr, 98%, TCI), 1-Hecadecyl-3-methylimidazolium bromide (C16mimBr, 98%, TCI), hexamethonium bromide(HMeBr, 99%, Aladdin), decamethonium bromide (DMeBr, 99%, Aladdin) and triethylamine (TEA, 99%, Aladdin) are chosen as the organic cations to synthesize the Mn(II) hybrids. Manganese bromide (MnBr₂, 99%, Aladdin), Mn(CH₃COO)₂·4H₂O (99%, Aladdin), Hydrobromic acid (HBr, 48%, Macklin) and Hydriodic acid (47%, 1.5% H₃PO₂, Aladdin) were used without further purified.

Synthesis of (TEA)₂MnI₄. To a mixture of 0.52 g (1.08 mmol) Mn(CH₃COO)₂·4H₂O and 0.5 mL (3.60 mmol) TEA, 1.5 mL of concentrated stabilized HI was added. The resulting slightly colored yellow solution changed color to pink when heated. The solution was evaporated to a volume of ~1.0 mL, cooled slowly and left covered in the refrigerator overnight. The next day, the resulting yellow precipitate was filtered under a vacuum for 5–10 min, then it was dried in the oven for 1 day. The single crystal of (TEA)₂MnI₄ for the X-ray diffraction analysis was directly selected from the total mass of the precipitate.

Synthesis of Manganese Bromide Hybrids. Take [C16Py]₂MnBr₄ for example, manganese bromide was used as the source of Mn²⁺. After dissolving MnBr₂ and C16PyBr into a desired amount of HBr and deionized water under vigorous stirring to make sure the reagents were completely dissolved, [C16Py]₂MnBr₄ was synthesized by slow evaporation of the above solution for a few days at 50 °C before

collecting the obtained single crystals. The preparation methods of [C16mim]₂MnBr₄, HMeMnBr₄ and DMeMnBr₄, as well as other analogs with other halogen components, were similar despite C16mimBr, HMeBr and DMeBr of the corresponding manganese halides were used in the process.

2.2. Characterization. The PXRD patterns of the samples, along with the temperature-dependent PXRD measurements, were obtained using an Aeris PXRD diffractometer (PANalytical Corporation, The Netherlands) set to operate at 40 kV and 15 mA with monochromatized Cu K α radiation (λ = 1.5406 Å). Photoluminescence (PL) and photoluminescence excitation (PLE) measurements were conducted on an FLS1000 fluorescence spectrophotometer (Edinburgh Instruments Ltd., UK). The PLQYs were recorded using a Hamamatsu Quantaurus-QY Spectrometer (Model C11347-11). These PLQYs were determined using the formula $\eta_{QE} = I_S / (E_R - E_S)$, where I_S represents the luminescence emission spectra of the sample, E_R refers to the spectra of the excitation light from an empty integrated sphere, and E_S indicates the excitation spectra of the excited sample. X-ray induced radioluminescence (RL) was captured using an FLS1000 fluorescence spectrophotometer (Edinburgh Instruments Ltd., UK). The scintillator was securely affixed to the circular window of an integrating sphere equipped with an X-ray source (Amptek Mini-X tube featuring an Ag target and a maximum power output of 4 W). X-ray images were captured using a complementary metal oxide semiconductor (CMOS) camera. Modulation transfer function (MTF) measurements are instrumental in assessing the fundamental spatial resolution of an imaging system. Spatial resolution is determined at the spatial frequency where MTF equals 0.2. The MTF curve was derived using the slanted-edge method.

3. DATA AND METHOD

3.1. Data Set Description. The crystal structures of 371 organic–inorganic halides, denoted as Org^{*}MX₄, were obtained from the Crystallographic Open Database (COD). The Org refers to a small organic molecule consisting of only C, H, and N elements, while M represents a metal ion with similar radii in the range of 0.55–0.66 Å. The possible metal ions include Co²⁺, Cu²⁺, Fe²⁺, Mg²⁺, Mn²⁺, Ni²⁺, Pd²⁺, and Zn²⁺. It is important to mention that we specifically selected M²⁺ ions for this study. Additionally, considering our choice of X = Cl⁻, Br⁻, and I⁻, it is evident that we focused on Four-coordinate MX₄²⁻ anions.

It is not expected that ions such as Co²⁺ or Mg²⁺ would lead to scintillation effects. However, we included them in our training data set solely to increase its size. The selection was predicated on the similarity of their ionic radii to those of Cu and Mn, as well as their analogous crystallographic packing tendencies, which are expected to facilitate the formation of isostructural compounds. It is important to note that Machine Learning can identify outliers that do not conform to general rules. This flexibility allowed us to incorporate all these ions into a single pool without significant issues, while significantly expanding the number of samples and enhancing the overall accuracy of the model. This represents a significant advantage of our approach.

To create a training data set, 296 random samples were selected, with a Mean Average Error (MAE) of 0.21 Å. The data includes COD number, Scheme, experimental d(M···M) (Å), calculated d(M···M) (Å), and the d(M···M) difference (Å), which represents the discrepancy between experimental and calculated bond lengths. The remaining 75 samples were set aside for testing, with a MAE of 0.50 Å. The details of these sets can be found in Table S1 and Table S2. The distances between M ions were measured using Diamond software and recorded in the respective tables. The atom types of M and X

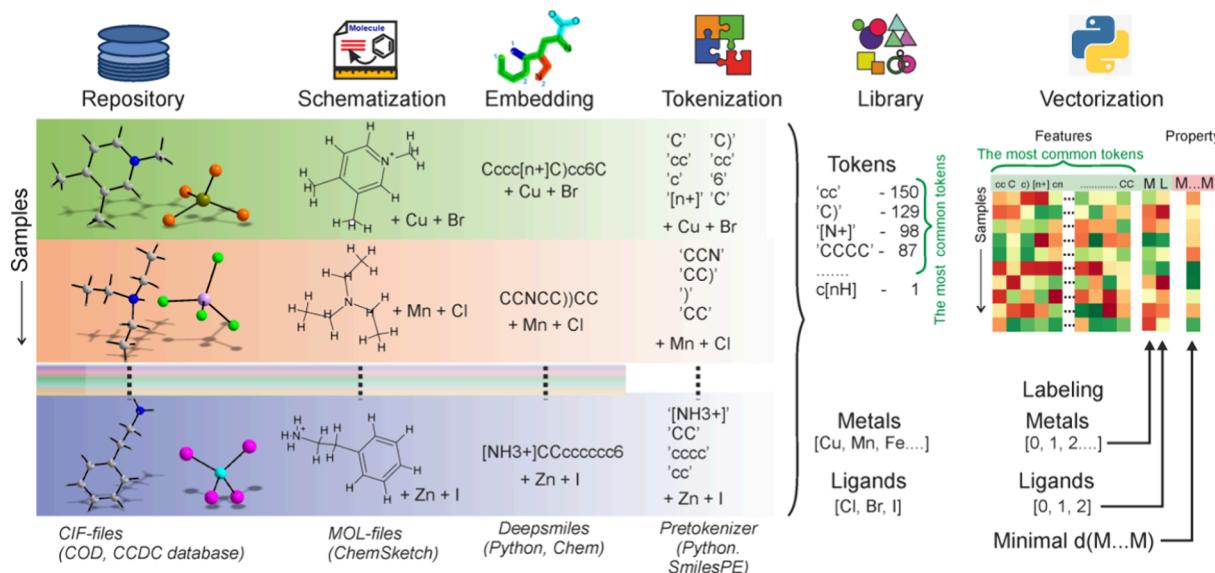


Figure 1. A scheme with a visual representation that emphasizes the crucial steps in constructing a unified representation for all small organic molecules using DeepSMILES and tokens, as well as labeled metals and ligands, by leveraging joint vectors.

were transformed into numerical format using Label encoding, a method that converts categorical variables into integers for computational analysis. This encoding facilitated the examination of the data set by enabling the labels to be processed efficiently. Subsequently, the encoded labels and the shortest distances between M ions, denoted as $d(M \cdots M)$, were plotted in (Figure S1). This figure provides a visual representation of the data set, showcasing how the various atom types of M and X are distributed and correlating these distributions with the spatial arrangement of M ions. This visualization helps identify patterns and supports further analysis and conclusions. The histogram illustrates a nearly balanced distribution of M ions and $d(M \cdots M)$ distances. However, the histogram for X shows an imbalance, mainly due to a lack of samples with X = I in the COD database. Consequently, the conclusions drawn in this manuscript will primarily apply to samples with X = Cl or Br.

3.2. Embedding Data Set. One of the pivotal components in Machine Learning is embedding, which involves the transformation of complex data into numerical or vector representations that can be effectively utilized for model development. In our particular case, we employed the widely used standard SMILES representation for molecules,^{19,20} condensed into a single string line. However, we enhanced this approach by adopting the advanced DeepSMILES²¹ technique, which offers an improved version of the SMILES representation. The DeepSMILES syntax effectively addresses the challenge of unbalanced parentheses by utilizing only close parentheses, where the number of parentheses denotes the branch length. Furthermore, DeepSMILES overcomes the issue of pairing ring closure symbols by employing a single symbol at the ring closing position, where the symbol indicates the size of the ring.²¹

Initially, the crystal structure was partitioned into two distinct components: organic molecules and inorganic cations M, X (refer to Figure 1). Subsequently, the molecule was simplified and represented using a scheme, which was then saved as a MOL-file using ChemSketch software. To facilitate further analysis, the MOL-files can be conveniently and automatically converted to DeepSMILES format using a Python module available at (<https://github.com/baoilleach/>)

deepsmiles). After that all string lines were tokenized using SmilesPE pretokenizer (Smiles Pair Encoding, SPE) library in Python (<https://pypi.org/project/SmilesPE/>).²⁴ The SPE initially acquires a vocabulary of frequently occurring SMILES or DeepSMILES substrings from a substantial chemical data set (such as ChEMBL). It then proceeds to tokenize SMILES based on this acquired vocabulary, enabling effective training of deep learning models. Notably, SPE extends the conventional atom-level tokenization approach by introducing human-readable and chemically interpretable SMILES substrings as additional tokens, thereby enhancing the overall representation.²⁴

The tokens were sorted, and the top 50 most frequently occurring tokens were chosen as representative tokens for all compounds. This number is a variable parameter, and we experimented with several values: 20, 30, 40, 50, 70, 100, and 200. We found that 50 is the minimum number of tokens that does not increase the model error. Using more than 50 tokens results in the same error, while using fewer tokens leads to an increase in this value. The number of times each representative token appeared in each molecule was encoded as the corresponding numerical value, while the absence of a token was coded as zero. As a result, all molecules within the set of 371 compounds were transformed into vectors within a 50-dimensional latent space (Figure 1), as previously demonstrated.^{19,24}

3.3. Random Forest Method. In this study, the prediction tool employed is the random forest (RF), which is an ensemble method based on regression trees.¹⁴ It is crucial to highlight that Random Forest (RF) is a nonparametric Machine Learning method, which significantly reduces the risk of overfitting when working with data sets of low sample size. This feature is particularly important for our tasks, considering the limited number of samples available. The regression trees are constructed through recursive binary partitioning of the multidimensional predictor space, creating multiple “decision trees” during the training process. The output of the RF model is determined by the mode of the classes in the case of classification or the mean/average prediction in the case of regression.¹⁴

The data set is initially divided into two distinct groups: the training data set, typically comprising 70–90% of the total data set, and the test data set, containing the remaining 10–30%. The training data set is used to randomly select relevant features and samples for constructing the Decision Tree model (Figure 2). This process is repeated until an optimal ensemble

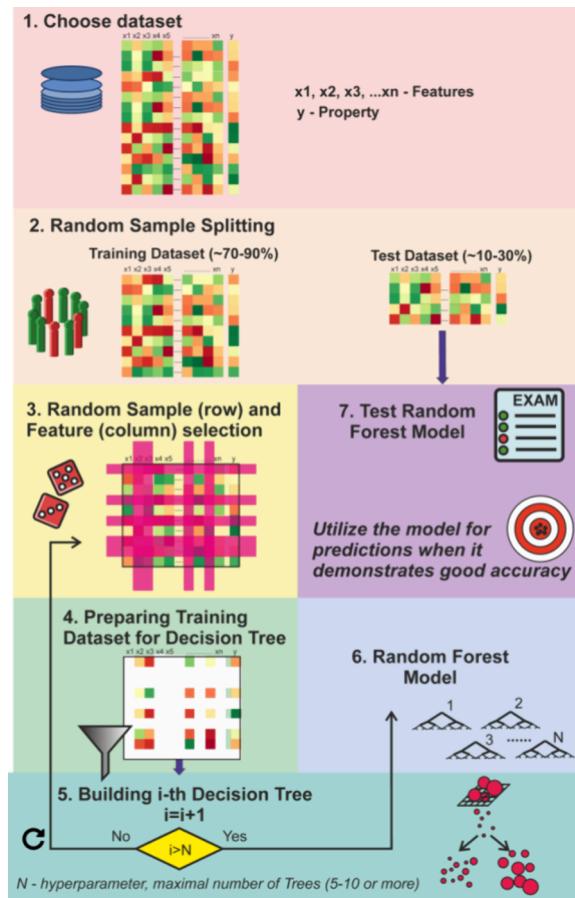


Figure 2. Primary process flow in this study with several key steps, including data set processing, model building, validation, and generating predictions using the Random Forest model.

of N trees is formed, where N represents the essential hyperparameter that requires tuning in the Random Forest method. The ensemble of trees is then evaluated using the test data set, and if it achieves a small Mean Absolute Error (MAE) or high accuracy, the model is considered reliable for making predictions.

To build the RF model, we utilized a self-written python script called RandomForest.py, implemented in Python 3.6 programming language.²² The program extensively employed standard libraries such as numpy, pandas, sklearn, matplotlib, and mpl_toolkits. To account for the stochastic nature of this machine-learning algorithm, we performed ten repetitions of cross-validation and averaged the performance across them. Each time, the data were randomly split into two data sets: one for training (80% of the total data) and another for testing (20% of the total data). The number of trees, denoted as N , is a hyperparameter that can be fine-tuned in the Random Forest method. We started with $N = 3$ and incrementally increased it by one, continuing this process until the cross-validation MAE value no longer decreased. This iterative approach revealed

that an optimal N value was 10. Another crucial hyperparameter is the maximum depth of the trees. In our study, we chose not to restrict tree depth, allowing the trees to expand until further splitting was not feasible due to the limited number of samples. The MAE values for the training ($\text{MAE} = 0.21 \pm 0.05 \text{ \AA}$) and test data sets ($\text{MAE} = 0.50 \pm 0.05 \text{ \AA}$) were obtained as the results of the 5-fold cross-validation. The comparison of all experimental and calculated $d(\text{M}\cdots\text{M})$ values are presented in Table S1, S2. The low error indicates that the main correlation between the experimental features and the primary properties has been revealed (Figure 3).

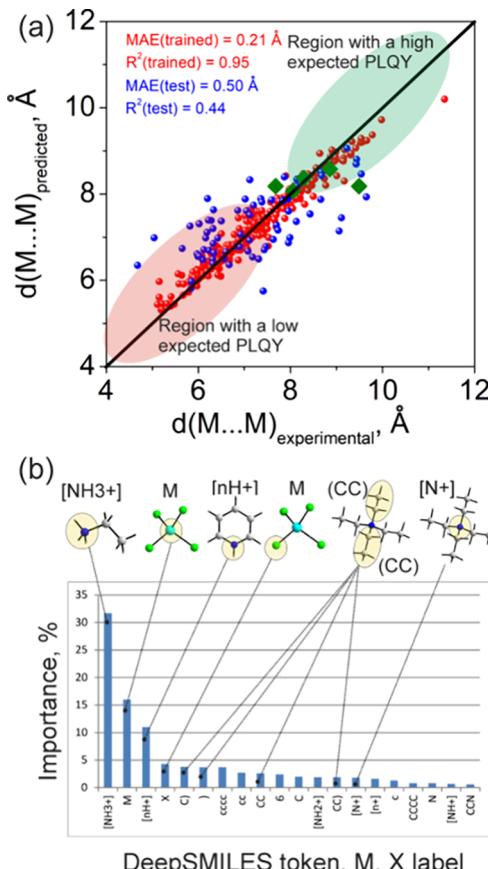


Figure 3. Comparison between predicted and experimentally observed shortest $d(\text{M}\cdots\text{M})$ distances in OIMH compounds is depicted (a). The red dots represent the training data set, while the blue dots represent the test data set. Compounds that potentially exhibit high scintillation properties are highlighted within a green zone. These compounds can be easily distinguished from those with potentially low PLQY due to the relatively small prediction error of $\pm 0.5 \text{ \AA}$. Importance of first 20 feature parameters on $d(\text{M}\cdots\text{M})$ obtained by Random Forest method (b). The tokens '[NH3+]', '[nH+]', and 'C') along with the presence of M with X are identified as the most significant feature parameters. Some of the most important tokens have been depicted and highlighted in actual molecules.

As a result, the molecule scheme of organic part composed of C, N, and H atoms, along with specific metal M and halogen X (Cl, Br, I) definitions, enables the prediction of the lowest $d(\text{M}\cdots\text{M})$ distance with the average error $\pm 0.5 \text{ \AA}$. Even a small error can be sufficient to differentiate between compounds with a distance of 6 \AA and low scintillation properties, and compounds with distances of $9\text{--}10 \text{ \AA}$ that exhibit high scintillation characteristics. Consequently, compounds exhibit-

ing high $d(M\cdots M)$ values can be readily identified, and it is these particular compounds that are expected to demonstrate a high PLQY in scintillation (Figure 3a).

It should be noted that RF allows measuring the importance of the feature after training. The selected value is permuted among the training data and the error is computed on this perturbed data set. The importance score for the selected feature is computed by averaging the difference in error before and after the permutation over all trees.¹⁵ The score is normalized by the standard deviation of these differences. Features which produce large values for this score are ranked as more important than features which produce small values (Figure 3b). It is evident that the tokens '[NH3+]', '[nH+]', 'C', and the type of M and X have the most substantial impact, accounting for a total influence of 66.8%, on the shortest $d(M\cdots M)$ distance. The number of nitrogen atoms in the aliphatic portion associated with '[NH3+]' and the aromatic fragments related to '[nH+]' tokens within the molecules significantly influence this distance. Moreover, the presence of aliphatic branched chains associated with 'C)', 'CC)', and long chains associated with 'CCCC' token also exert control over the $d(M\cdots M)$ distance. At the current stage, it is not possible to distinguish between positive and negative influences; only their importance has been revealed. To uncover the underlying rules, it is crucial to visualize the relationships among these influences. Building a Decision Tree is the most convenient and effective way to achieve this. In the next section, we will delve into the construction of a Decision Tree to gain deeper insights.

3.4. Decision Tree Method. In the previous chapter, we successfully developed an RF prediction model and determined the most influential feature parameters that contribute to the desired concrete properties. The focus of this chapter is to present the rules derived from the RF model, providing an interpretation of the predictions. Decision Tree (DT) is widely recognized as one of the best methods for obtaining interpretability.^{14,16,23} It is worth emphasizing that DT is a nonparametric Machine Learning method, which mitigates the risk of overfitting when working with data sets of low sample size. This characteristic is particularly significant for our task, ensuring reliable and meaningful interpretations of the predictions.

The process of constructing a Decision Tree begins with an initial step of organizing the data set according to a specific feature parameter, denoted as "i." This sorting operation is crucial as it establishes the basis for the subsequent division of the data set into two distinct subsets, labeled Y1 and Y2. The division is performed based on a threshold value, referred to as " m_i ," which serves as the decision boundary for splitting the data (Figure S2). This threshold is carefully chosen to optimize the separation of the data into the two subsets. Once the data set is divided, the primary objective becomes the minimization of a calculated metric that evaluates the homogeneity or purity of the resulting subsets. This metric can be either the normalized sum of dispersions (D) or entropies (E), which are quantitative measures of the subsets' variability or disorder. The dispersions (D) are calculated using the formula $D = N_1 \cdot D_1 + N_2 \cdot D_2$, where N_1 and N_2 represent the number of elements in subsets Y1 and Y2, respectively, and D_1 and D_2 are the dispersions within those subsets. Similarly, entropies (E) are computed as $E = N_1 \cdot E_1 + N_2 \cdot E_2$, with E_1 and E_2 representing the entropy values of the subsets. The goal is to find a split that results in the lowest possible value of D or E,

thus ensuring that each subset is as homogeneous as possible. Figure S2 visually represents this process, illustrating how the decision tree construction hinges on these calculations to effectively partition the data. This step is repeated for all feature parameters, and the parameter with the lowest D or E value is selected. The rule " $i < m_i$ " is used to construct the upper portion of the Decision Tree. Y1 and Y2 may be further subdivided using the same procedure until the complete Decision Tree is formed.

The resulting Decision Tree can be utilized to predict outcomes by passing data parameters from the root to the internal nodes and ultimately to the terminal node. To build the DT model, we employed a self-written python script called `DecisionTree.py`, implemented in Python 3.6 programming language.²¹ The program utilized standard libraries such as numpy, pandas, sklearn, matplotlib, and mpl_toolkits. To simplify the task, the DT models with a depth of 3 were utilized. The DT models that yield the best data segregation are presented in Figure 4. One can observe that out of the total

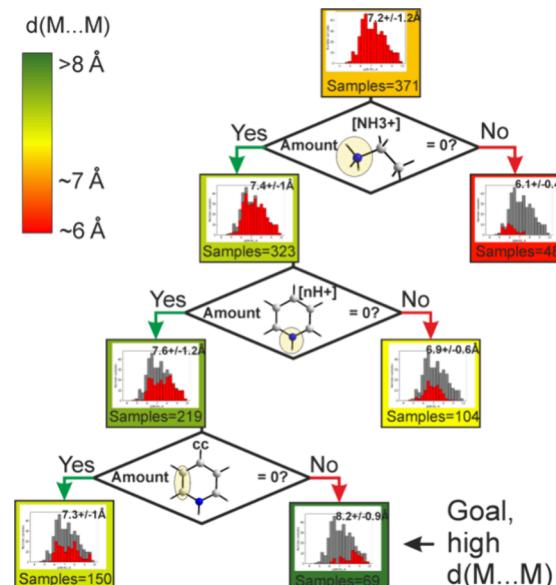


Figure 4. Decision Tree illustrated the rules that differentiate 69 compounds with a large $d(M\cdots M)$ value of $8.2 \pm 0.9 \text{ \AA}$ from the remaining 302 compounds. Each group is accompanied by a histogram in red color, which highlights representative samples within the group. Additionally, a separate histogram in gray color showcases the initial sample distribution. Groups with small average $d(M\cdots M)$ values are represented in red color box, while groups with high values are depicted in green. Molecules that result in a high $d(M\cdots M)$ value should not contain the tokens '[NH3+]' or '[nH+', but should include the 'cc' token. Molecule fragments corresponding to each token are also depicted.

371 samples, 48 samples contain the '[NH3+]' token, making them distinct from the remaining 323 samples. Interestingly, these 48 samples exhibit a relatively low average $d(M\cdots M)$ value of 6.1 \AA . Therefore, based on our findings, we can conclude that the presence of a terminal -NH₃ group in a molecule somehow inhibits the formation of large M···M contacts. The next rule, '[nH+] = 0', separates 104 samples with a low $d(M\cdots M)$ value of 6.9 \AA from 219 samples with a high $d(M\cdots M)$ value of 7.4 \AA . This finding suggests that the presence of an N–H group in the aromatic part of the molecule also inhibits the formation of large M···M contacts.

The third important rule concerned about presence of aromatic C–C bonds in the molecule. The 69 samples that possess the ‘[NH3+]’ and ‘[nH+]’ tokens, along with the additional ‘cc’ token, are distinctly separated from the remaining 150 samples with low $d(M\cdots M)$ values (Figure 4). These three rules appear to be the most effective, and extracting additional rules would require a larger sample size. It should be noted that discrete parameter $X = (\text{Cl}, \text{Br}, \text{I})$ was not explicitly included in the modeling rules. This implies that the presence of I, Br, or Cl may not be a critical factor significantly affecting the $d(M\cdots M)$ distance. Consequently, our model may still be applicable for predicting compounds with $X = \text{I}$, despite the imbalance in their representation within the data set. Notably, we chose to halt further division as it resulted in the splitting of the 69 samples into two groups: one with a significantly larger quantity of 67 samples ($d(M\cdots M) = 8.2 \pm 0.8 \text{ \AA}$) and another with a much smaller quantity of 2 samples ($d(M\cdots M) = 10.7 \pm 0.5 \text{ \AA}$). These two samples, which exhibit remarkably high $d(M\cdots M)$ values, differ from the rest based on an additional fourth rule: ‘Amount of ‘C’ > 3?’ If the answer is yes, then we have identified these two samples. The token ‘C’ was previously discussed in the previous chapter and is associated with the presence of aliphatic branched chains, as depicted in Figure 3b.

4. RESULTS AND DISCUSSION

We have made the discovery that the presence of a terminal -NH₃ group inhibits the formation of large M···M contacts. This inhibition can be attributed to the formation of short N–H···X (X = Cl, Br, I) hydrogen bonds, which compact organic molecules into a tightly constructed 3D network, attract MX₄ cation groups toward each other, consequently resulting in a lower $d(M\cdots M)$ value. Interestingly, we also found that N–H groups within the aromatic rings of molecules exhibit a similar inhibitory effect on large M···M contacts, suggesting that the underlying mechanism is likely the same. The nature of the next rule, which involves the existence of aromatic C–C bonds in the molecule, remains less clear. However, it is of high importance as it separates samples with high $d(M\cdots M)$ values. One possibility is that molecules with aromatic rings tend to construct supramolecular constructions, and this steric effect may bring M ions further apart, resulting in a larger M···M distance. Despite the lack of complete clarity regarding this rule, it serves as a crucial factor in effectively categorizing these samples based on their $d(M\cdots M)$ distances.

It is worth mentioning that the Random Forest method also identifies the ‘[NH3+]’, ‘[nH+]’ and ‘cc’ tokens as significant feature parameters. However, interestingly, the metal type label ‘M’ and halogen type label ‘X’, which are deemed highly important for prediction using Random Forest, exhibit lower importance in the Decision Tree. This suggests that these rules may appear at a deeper depth (around 4–5) in the trees and represent additional rules that are not globally applicable. In section 3.4, we uncovered a rule at depth = 4 of Decision Tree: the presence of more than three ‘C’ tokens in the molecule is necessary to achieve a very high $d(M\cdots M)$ value of 10.7 Å. Interestingly, Random Forest also identifies the importance of the parameter ‘C’, along with ‘CC’ and ‘CCCC’, which are associated with the presence of aliphatic branched chains. The nature of this effect is easily comprehensible as steric effects may cause the MX₄ groups to move apart, leading to larger $d(M\cdots M)$ distances. Looking at the situation from a different perspective, considering that we are examining organic

molecules composed of carbon (C), hydrogen (H), and nitrogen (N) atoms, it is expected that nitrogen (N) should be present. However, we have discovered that the presence of ‘[NH3+]’ and ‘nH+’ in the DeepSMILES representation of the molecule should be avoided. As a result, the only remaining possibilities are the presence of ‘[NH2+]’ or ‘[N+]’. While these were not explicitly identified as rules, they can be considered as complementary rules to the overall analysis.

All the extracted rules have been consolidated in Figure S3, and listed here: 1) the presence of a terminal -NH₃ group inhibits large M···M contacts; 2) molecules with aromatic C–C bonds tend to have larger $d(M\cdots M)$ distances; 3) more than three ‘C’ tokens in a molecule can lead to very high $d(M\cdots M)$ values, such as 10.7 Å. They can be utilized independently, without relying on a “black box” Machine Learning model, to screen and predict compounds with high $d(M\cdots M)$ distances. Although the focus of this study is not solely on distance, it is noteworthy that numerous studies have established a correlation between high $d(M\cdots M)$ values and enhanced photoluminescent quantum yields (PLQY) in phosphors and scintillators.^{7,12,13} Other structural parameters may influence quantum yield also, but large M···M distance is advantageous for scintillation, and it can be considered a beneficial factor. Therefore, the suggested rules should be regarded as a pathway for chemists to develop high-intensity scintillators. By following these rules, chemists can effectively screen and select compounds that are more likely to exhibit desirable properties for scintillation applications.

It is important to note that additional properties of scintillator materials can be effectively modeled and predicted using Machine Learning, as shown in our current study. However, this endeavor necessitates the collection of a substantial number of samples, typically exceeding 100. Currently, no database exists with such a comprehensive collection of scintillators and their documented properties. In this study, we concentrated on predicting M···M distances and were fortunate to amass an adequate number of structures. It was beneficial that the M···M distance is correlated with the quantum yield of scintillators.

Using Machine Learning algorithms, specifically Random forest and Decision Tree models, the synthesis of five Four-coordinate Mn-based hybrid materials was guided. We conducted ground truth experiments to validate the effectiveness of these rules. These materials are [TEA]₂MnI₄, [HMe]MnBr₄, [C16Py]₂MnBr₄, [C16mim]₂MnBr₄ and [DMe]MnBr₄. Figure 5a–e show the crystal structure of the

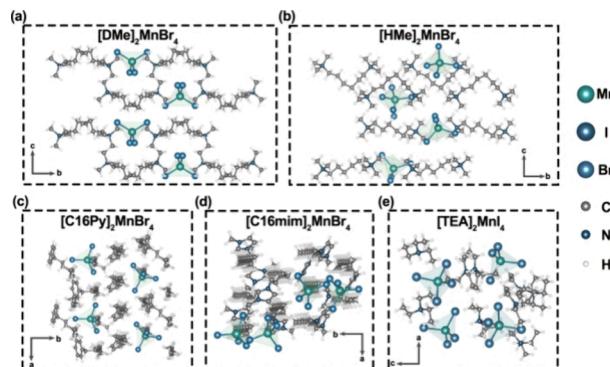


Figure 5. Crystal structure of (a) [DMe]MnBr₄ (b) [C16Py]₂MnBr₄ (c) [C16mim]₂MnBr₄ (d) [HMe]MnBr₄ (e) and [TEA]₂MnI₄.

five hybrids. The five compounds consist of isolated Mn tetrahedra and organic cations, each with different Mn···Mn distances. Table S3 shows the relationship between Mn···Mn distance and PLQY, revealing that all five compounds have relatively long Mn···Mn distances and correspondingly high PLQY values. These five hybrid materials were converted into DeepSMILES models to serve as experimental materials for correlating predicted structures with scintillation and luminescent properties through machine learning. Their scintillation properties were thoroughly characterized.

To evaluate the scintillation performance of these five compounds, Figure 6a presents the fluorescence spectra

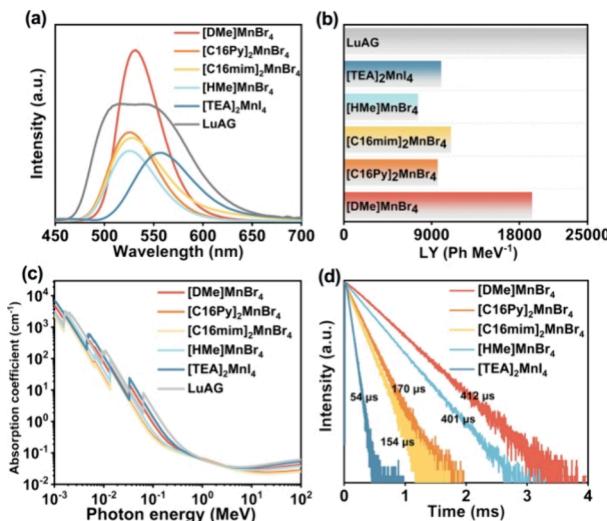


Figure 6. (a) Comparison of the X-ray fluorescence spectra of five hybrid materials with LuAG; (b) Comparison of the light yield of five hybrid materials; (c) Photovoltaic absorption coefficients of five hybrid materials; (d) Comparison of the fluorescence lifetimes of five hybrid materials.

excited by X-ray of the five hybrid materials compared with LuAG. State-of-the-art radiation detection devices require scintillators with high LY to minimize radiation exposure and improve detection contrast. The precise and standardized definition of the LY of a scintillator is as follows:

$$LY = (10^6 / \beta E_g) SQ \quad (1)$$

where E_g represents the bandgap of the scintillator, S and Q denote the QY during the transfer and luminescence stages, respectively, and β is a phenomenological parameter typically ranging from 2 to 3 for most materials. Figure 6b shows the LY calculated from Figure 6a, indicating that $[DMe]MnBr_4$ has the highest LY. Figure 6c displays the X-ray absorption coefficients of the five hybrid materials compared with LuAG, revealing that the X-ray absorption of $[DMe]MnBr_4$ is comparable to that of LuAG. Figure 6d illustrates the fluorescence lifetimes of the five compounds: $[DMe]MnBr_4$, $[HMe]MnBr_4$, $[C16Py]_2MnBr_4$, $[C16mim]_2MnBr_4$, and $[TEA]_2MnI_4$, with fluorescence lifetimes of 412 μs , 401 μs , 170 μs , 154 μs , and 54 μs , respectively, these short lifetimes make them suitable for scintillator applications.

To demonstrate the linear response range of the $[DMe]MnBr_4$ scintillator to X-ray dose rates, the emission intensity at different dose rates was measured, as shown in Figure 7a. Figure 7b displays the linear relationship between SNR and

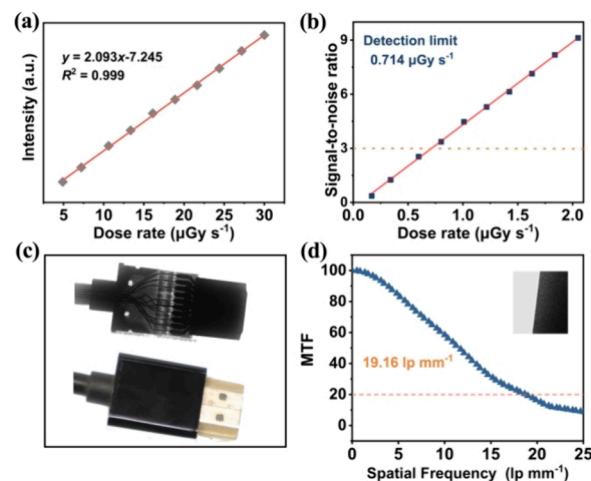


Figure 7. (a) Comparison of the X-ray fluorescence spectra of the five hybrid materials with LuAG; (b) Comparison of the light yields of the five hybrid materials; (c) Comparison of the photoelectric absorption coefficients of the five hybrid materials; (d) Comparison of the fluorescence lifetimes of the five hybrid materials.

dose rate, with the detection limit determined to be approximately 0.714 $\mu Gy\ s^{-1}$ when the signal-to-noise ratio (SNR) is 3. This detection limit is over 30 times lower than the 5.5 $\mu Gy\ s^{-1}$ dose rate required for X-ray medical diagnostics. The $[DMe]MnBr_4$ scintillator, combined with PDMS, was used to image a data cable, as shown in Figure 7c. The photographs depict the data cable under X-ray and natural light, clearly revealing the internal metal wires. To further quantify the resolution, the MTF curve was obtained by imaging a 1 mm thick aluminum sheet with sharp edges using the slanted edge method. As shown in Figure 7d, the spatial resolution of this scintillator (@MTF = 0.2) is 19.16 lp mm^{-1} . This scintillator can produce high-quality X-ray images. From this, we can see that hybrid materials with larger M···M distances have higher light yields, which is consistent with the predicted results.

5. CONCLUSIONS

In conclusion, a predictive model using Random Forest was developed based on 371 OIMH samples, achieving a low error margin of 0.5 Å for $d(M\cdots M)$ values. This high accuracy was attained using molecular sketches or DeepSMILES representations in combination with M and X types of OIMH compounds. The model facilitates efficient screening for high-scintillation OIMH compounds within structural databases. Furthermore, an Explainable Artificial Model based on Decision Tree identified key rules for predicting high $d(M\cdots M)$ distances and potential scintillation properties. We found that terminal -NH₃ and N-H groups and absence of aromatic C-C bonds hinder large M···M contacts, while [N+] tokens, aliphatic chains, increase $d(M\cdots M)$ distances. These findings informed the synthesis and characterization of five high-scintillation compounds, with $[DMe]MnBr_4$ exhibiting the highest LY and demonstrating strong application potential. This work paves the way for the prediction of scintillation properties in hybrid materials and their expanded applications.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.chemmater.4c03162>.

Training data set details (Table S1), test data set details (Table S2), the data on Mn···Mn distance and PLQY in some compounds (Table S3), histogram illustrating the distribution of different labels (Figure S1), the flowchart containing several main components (Figure S2), and the extracted rules (Figure S3) ([PDF](#))

■ AUTHOR INFORMATION

Corresponding Authors

Maxim Molokeev — *Siberian Federal University, Krasnoyarsk 660041, Russia; Laboratory of Crystal Physics, Kirensky Institute of Physics, Federal Research Center KSC SB RAS, Krasnoyarsk 660036, Russia; Laboratory of Theory and Optimization of Chemical and Technological Processes, University of Tyumen, Tyumen 625003, Russia;*
 orcid.org/0000-0002-8297-0945; Email: msmolokeev@mail.ru

Shuai Zhang — *The State Key Laboratory of Luminescent Materials and Devices, Guangdong Provincial Key Laboratory of Fiber Laser Materials and Applied Techniques, School of Physics and Optoelectronics, South China University of Technology, Guangzhou 510641, China;*
 orcid.org/0000-0001-9723-9832; Email: shuaizhang@scut.edu.cn

Zhiguo Xia — *The State Key Laboratory of Luminescent Materials and Devices, Guangdong Provincial Key Laboratory of Fiber Laser Materials and Applied Techniques, School of Physics and Optoelectronics, South China University of Technology, Guangzhou 510641, China;*
 orcid.org/0000-0002-9670-3223; Email: xiazg@scut.edu.cn

Authors

Nicolay Golovnev — *Siberian Federal University, Krasnoyarsk 660041, Russia*

Andrey Zolotov — *Siberian Federal University, Krasnoyarsk 660041, Russia*

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.chemmater.4c03162>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Russian Science Foundation grant (24-43-00006), the National Natural Science Foundations of China (22361132525 and 52425206), and the Guangdong Provincial Science & Technology Project (2023A0505050084).

■ REFERENCES

- (1) Chen, Q.; Wu, J.; Ou, X.; Huang, B.; Almutlaq, J.; Zhumekenov, A. A.; Guan, X.; Han, S.; Liang, L.; Yi, Z.; Li, J.; Xie, X.; Wang, Y.; Li, Y.; Fan, D.; Teh, D. B. L.; All, A. H.; Mohammed, O. F.; Bakr, O. M.; Wu, T.; Bettinelli, M.; Yang, H.; Huang, W.; Liu, X. All-Inorganic Perovskite Nanocrystal Scintillators. *Nature* **2018**, *561*, 88.
- (2) Maddalena, F.; Tjahjana, L.; Xie, A.; Arramel; Zeng, S.; Wang, H.; Coquet, P.; Drozdowski, W.; Dujardin, C.; Dang, C.; Birowosuto, M. D. Inorganic, Organic, and Perovskite Halides with Nanotechnology for High-Light Yield X- and γ -ray Scintillators. *Crystals* **2019**, *9*, 88.
- (3) Zhang, F.; Zhou, Y.; Chen, Z.; Wang, M.; Ma, Z.; Chen, X.; Jia, M.; Wu, D.; Xiao, J.; Li, X.; Zhang, Y.; Shi, Z.; Shan, C. Thermally Activated Delayed Fluorescence Zirconium-Based Perovskites for Large-Area and Ultraflexible X-ray Scintillator Screens. *Adv. Mater.* **2022**, *34*, 2204801.
- (4) Gandini, M.; Villa, I.; Beretta, M.; Gotti, C.; Imran, M.; Carulli, F.; Fantuzzi, E.; Sassi, M.; Zaffalon, M.; Brofferio, C.; Manna, L.; Beverina, L.; Vedda, A.; Fasoli, M.; Gironi, L.; Brovelli, S. Efficient, Fast and Reabsorption-Free Perovskite Nanocrystal-based Sensitized Plastic Scintillators. *Nat. Nanotechnol.* **2020**, *15*, 462.
- (5) Yang, X.; Huang, Y.; Wang, X.; Li, W.; Kuang, D. A-Site Diamine Cation Anchoring Enables Efficient Charge Transfer and Suppressed Ion Migration in Bi-Based Hybrid Perovskite Single Crystals. *Angew. Chem., Int. Ed.* **2022**, *61*, No. e202204663.
- (6) Zhou, F.; Li, Z.; Lan, W.; Wang, Q.; Ding, L.; Jin, Z. Halide Perovskite, a Potential Scintillator for X-Ray Detection. *Small Methods* **2020**, *4* (10), 2000506.
- (7) He, Z.; Wei, J.; Zhang, Z.; Luo, J.; Kuang, D. Manganese-Halide Single-Crystal Scintillator Toward High-Performance X-Ray Detection and Imaging: Influences of Halogen and Thickness. *Adv. Opt. Mater.* **2023**, *11*, 2300449.
- (8) Li, C.; Lai, J. a.; Wu, D.; Wang, Q.; Bai, Y.; He, P.; An, K.; Tang, X. Lead-Free Scintillators based on Pyridine Manganese Halide for X-ray Imaging. *J. Lumin.* **2022**, *250*, No. 119130.
- (9) Peng, H.; Zou, B.; Guo, Y.; Xiao, Y.; Zhi, R.; Fan, X.; Zou, M.; Wang, J. Evolution of the Structure and Properties of Mechanochemically Synthesized Pyrrolidine Incorporated Manganese Bromide Powders. *J. Mater. Chem. C* **2020**, *8*, 6488–6495.
- (10) Bai, X.; Zhong, H.; Chen, B.; Chen, C.; Han, J.; Zeng, R.; Zou, B. Pyridine-Modulated Mn⁺ Emission Properties of C₁₀H₁₂N₂MnBr₄ and C₅H₆NMnBr₃ Single Crystals. *J. Phys. Chem. C* **2018**, *122*, 3130–3137.
- (11) Jiang, C.; Fu, H.; Han, Y.; Li, D.; Lin, H.; Li, B.; Meng, X.; Peng, H.; Chu, J. Tuning the Crystal Structure and Luminescence of Pyrrolidinium Manganese Halides via Halide Ions. *Cryst. Res. Technol.* **2019**, *54*, 1800236.
- (12) Zhou, G.; Liu, Z.; Huang, J.; Molokeev, M. S.; Xiao, Z.; Ma, C.; Xia, Z. Unraveling the Near-Unity Narrow-Band Green Emission in Zero-Dimensional Mn²⁺-based Metal Halides: a Case Study of (C₁₀H₁₆N)₂Zn_{1-x}Mn_xBr₄ Solid Solutions. *J. Phys. Chem. Lett.* **2020**, *11*, 5956–5962.
- (13) Molokeev, M. S.; Su, B.; Aleksandrovsky, A. S.; Golovnev, N. N.; Plyaskin, M. E.; Xia, Z. Machine Learning Analysis and Discovery of Zero-Dimensional ns₂Metal Halides toward Enhanced Photoluminescence Quantum Yield. *Chem. Mater.* **2022**, *34*, 537–546.
- (14) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32.
- (15) Biau, G. Analysis of a Random Forests Model. *J. Mach. Learn. Res.* **2012**, *13*, 1063–1095.
- (16) Custode, L. L.; Iacca, G. Evolutionary Learning of Interpretable Decision Trees. *IEEE Access* **2023**, *11*, 6169–6184.
- (17) Motaev, K.; Molokeev, M.; Sultanov, B.; Kharitonsev, V.; Matigorov, A.; Palianov, M.; Azarapin, N.; Elyshev, A. Application of Machine Learning to Fischer-Tropsch Synthesis for Cobalt Catalysts. *Ind. Eng. Chem. Res.* **2023**, *62*, 20658–20666.
- (18) Endzhievskaya, I. G.; Endzhievskiy, A. S.; Galkin, M. A.; Molokeev, M. S. Machine Learning Methods in Assessing the Effect of Mixture Composition on the Physical and Mechanical Characteristics of Road Concrete. *J. Build. Eng.* **2023**, *76* (107248), No. 107248.
- (19) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparragirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, *4*, 268–276.
- (20) Pinheiro, G. A.; Mucelini, J.; Soares, M. D.; Prati, R. C.; Da Silva, J. L. F.; Quiles, M. G. Machine Learning Prediction of Nine Molecular Properties based on the SMILES Representation of the

QM9 Quantum-Chemistry Dataset. *J. Phys. Chem. A* **2020**, *124*, 9854–9866.

(21) O'Boyle, N.; Dalke, A. DeepSMILES: an Adaptation of SMILES for Use in Machine-Learning of Chemical Structures. *ChemRxiv*. **2018**.

(22) Zhu, R. Q.; Zeng, D. L.; Kosorok, M. R. Reinforcement Learning Trees. *J. Am. Stat. Assoc.* **2015**, *110*, 1770–1784.

(23) Richert, W. *Building Machine Learning Systems with Python*, 2013; Packt Publishing Ltd.

(24) Li, X.; Fourches, D. SMILES Pair Encoding: a Data-Driven Substructure Tokenization Algorithm for Deep Learning. *J. Chem. Inf. Model.* **2021**, *61*, 1560–1569.