THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA

Article

Predicting Emission Wavelengths in Benzobisoxazole-Based OLEDs with Gradient Boosted Ensemble Models

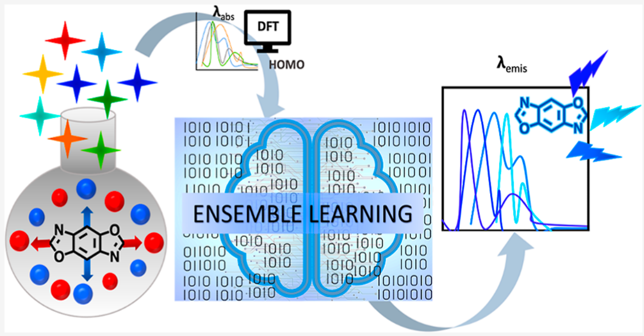
Shambhavi Tannir,[#] Yuning Pan,[#] Nathaniel Josephs, Christopher Cunningham, Nathan R. Hendrick, Annie Beckett, James McNeely, Aaron Beeler, Malika Jeffries-EL,* and Eric D. Kolaczyk*

Cite This: *J. Phys. Chem. A* 2024, 128, 6116−6123

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We demonstrate the use of gradient-boosted ensemble models that accurately predict emission wavelengths in benzobis[1,2-*d*:4,5-*d*′]oxazole (BBO) based fluorescent emitters. We have curated a database of 50 molecules from previously published data by the Jeffries-EL group using density functional theory (DFT) computed ground and excited state features. We consider two machine learning (ML) models based on (i) whole cruciform molecules and (ii) their constituent fragment molecules. Both ML models provide accurate predictions with root-mean-square errors between 30 and 36 nm, competitive with state-of-the-art deep learning models trained on orders of magnitude more molecules, and this accuracy holds even when tested on four new BBO emitters unseen by the models. We also provide an interpretable feature importance analysis and discuss the relevant relationships between DFT and changes in predicted emission wavelength.



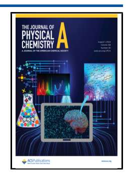
■ INTRODUCTION

Machine learning (ML) is rapidly evolving as a vital tool in accelerating the discovery of new materials.[1−3] Historically, new materials design and synthesis results from trial-and-error methods, which require several years of research, resources, and equipment. Improvements in molecular electronics structure methods, more powerful computational resources, and availability of experimental data sets allow us to cross into the next level of data-driven materials discovery.[4−6] Currently, ML and high throughput experimentation are being used successfully in drug discovery and have recently started to gain traction in materials science.[7,8]

Organic fluorescent materials have many applications spanning lighting, imaging, sensing, and display technologies.[8−10] Organic light emitting diodes (OLEDs) are gaining popularity in display and lighting technologies due to their inexpensive processing, flexibility, and energy efficiency.[11,12] The emission colors in OLEDs are traditionally due to fluorescence or phosphorescence mechanisms. In solid-state devices, direct fluorescence is statistically limited to 25% internal quantum efficiency (IQE), which restricts their external quantum efficiencies (EQE) to 5%. While phosphorescence and thermally activated delayed fluorescence (TADF) are due to mechanisms that allow the IQE to reach 100% (and EQE >5%), these materials often have poor color quality, poor resolution, and broad emissions thereby rendering them ineffective for display screens.[13,14] Additionally, some applications such as organic pump lasers and visible light communication rely on nanosecond (ns) responses which fluorescent emitters provide, unlike phosphorescence and TADF materials, which are on the microsecond (μs) to millisecond (ms) time scale.[15−18]

OLEDs can be broadly classified as red, green, or blue (RGB) emitters based on their emitted color. While several known examples of thermally stable and long-lived red and green fluorescent molecules exist, blue light materials with similar benchmarks are much more challenging to design.[11,19−23] This is mainly due to their broader energy gaps (2.8−3.1 eV) and their high energy of emission (<450 nm), which results in rapid overheating and deterioration of the devices.[24,25] Display technology is still seeking “deep-blue” materials that have CIE (Commission International de l’Éclairage) coordinates less than 0.16, 0.06 (Figure 1, right).[26,27] The Jeffries-EL group has developed multiple novel small-molecule emitters based on a benzobis[1,2-*d*:4,5-*d*′]oxazole (BBO) core, an electron-deficient ring system well-known for its thermal and oxidative stability. BBO is a conjugated molecule consisting of a central benzene ring and fused oxazole units flanking either side (Figure 1, left). The key

Received: January 4, 2024
Revised: June 18, 2024
Accepted: June 18, 2024
Published: July 15, 2024



https://doi.org/10.1021/acs.jpca.4c00077
J. Phys. Chem. A 2024, 128, 6116−6123

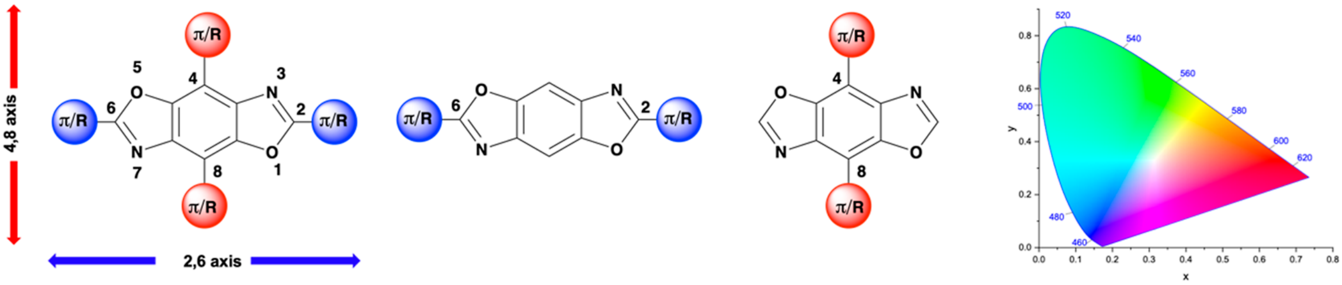


Figure 1. (Left) Schematic of "cruciform" and "fragment" benzobisoxazole core with the numbering system (p = aryl group, R = alkyl group); (Right) CIE 1931 chromaticity diagram.

parameters that define OLEDs electronically are the highest occupied molecular orbital (HOMO), the lowest unoccupied molecular orbital (LUMO), and the energy gap (E_g) where E_g = |HOMO − LUMO|. Blue emitting materials have a wider band gap between 2.8 and 3.2 eV. Hence, tuning the HOMO and LUMO levels in a molecule can lead to selectivity in the desired emission output. Due to its structure, the BBO core has four points of modification, marked as positions 2, 6, 4, and 8. It has two orthogonal conjugation pathways, one through the 2,6 axis and another perpendicular 4,8 axis, passing through the central benzene ring. Advantageously, this leads to the separation of the frontier molecular orbitals (FMOs), allowing us to selectively tune their optoelectronic properties by substituting different groups at these positions.[28−37] In contrast to other conjugated materials in which structural modifications alter the LUMO, HOMO and E_g. The Jeffries-EL group has previously shown that the cross-conjugated BBO "cruciforms," where all four positions have aryl groups attached, follow molecular heredity, such that the optoelectronic properties of the cruciform or "child" molecules is the sum of the linear or "parent" BBO properties (Figure 1, left).[32] As a result, the optical and electronic properties of the "child" can be predicted by evaluating those of the parent.[38] Analysis of the cruciform compounds indicated that their HOMO levels were mostly influenced by the groups along the 4,8 axis and the LUMO levels were mostly affected by the groups along the 2,6 axis. This recently discovered heredity phenomenon, of the parent molecules combinatorial properties resulting in unique cruciform properties, also makes BBO a robust molecular template for the rational design of new materials.

Understanding a molecule's optoelectronic properties is crucial in designing new materials. Density functional theory (DFT) calculations are commonly used to gain insights into a molecule's ground state structural and electronic information and its general excited state properties, such as absorption data. However, using DFT to predict emission data is expensive and frequently inaccurate.[39,40] With potentially millions of possible molecular combinations of BBO structures, individual synthesis and computational analysis of BBO congeners is not only a costly and time-consuming endeavor but also does not effectively predict the nonproductive structural combinations. ML can bridge the gap between expensive materials synthesis and complex property prediction.

Recently ML has been used to predict TADF emitters and their optimized devices by predicting the best EQEs.[41,42] These works focus on the device performances of the RGB gamut emitting materials and source data from literature to form large data sets. Using large data sets, numbering in tens or hundreds of thousands of molecules, has led to advanced ML models propelling related fields such as drug discovery.[43,44] However, large data sets are often obtained from the literature using complex text mining algorithms, resulting in potentially redundant or even incorrect values in the data sets.[45−47]

On the other hand, in organic and materials chemistry, functional molecules have specific properties that we seek, resulting in fewer but high-quality data observations.[48−51] Furthermore, as we focus on fewer properties to be predicted (such as blue emission) and further zoom in on specific mechanisms/types of molecules (such as direct fluorescence in nonmetalated organic molecules), our sample size reduces further.[52,53] Nevertheless, due to the small, high quality data sets have been successfully used to understand underlying electronic mechanisms and geometric properties that affect a molecule's properties. For instance, Troisi et al. constructed a database of only 80 distinct nonfullerene acceptor based organic photovoltaics. They found that all the best-performing materials/molecules had small energy gaps between the LUMO and LUMO+1 levels.[54]

In this work, we report two gradient-boosted decision tree models, based on eXtreme gradient boosting (XGBoost), that accurately predict emission wavelengths of BBO-based emitters. The cruciform model (CM) is built from a cruciform BBO molecules database, while the fragment model (FM) is made from the "parents," i.e., the constituent fragments database. While the FM considers the features of the individual constituents, the predicted outcome is of their corresponding cruciform. Both models achieve accurate predictions with an approximate root-mean-square errors (RMSE) of 31 nm that is competitive with recent models trained on 1000X more observations from the literature.[55,56] We also compared the XGBoost models, CM and FM, to gradient boosting machine (GBM) and compared their performances and the details can be found in the Supporting Information, Section S3.1, Table S3. To improve our understanding of the influence of the features on predicted emission, we also performed various feature analyses and confirmatory analysis of variance tests.

■ METHOD

We describe the main steps in our model development—database curation, model selection and training, model validation and results' interpretation.

Database Curation. We curated a database from 50 cruciform BBO molecules designed by the Jeffries-EL group and their constituent fragments (29 molecules) (for structural details see—https://github.com/KolaczykResearch/Blue-BBO-ML/tree/main). The features of these molecules were calculated at the mpw3LYP/SV/CPCM(CH_2Cl_2) using ORCA,[57−60] similar to the previously benchmarked DFT methods by the Jeffries-EL group[34] (see Section S1.1 in

https://doi.org/10.1021/acs.jpca.4c00077
J. Phys. Chem. A 2024, 128, 6116−6123

Supporting Information, for details). We combined 81 experimental measurements of λ_{max}^{emis} (maximum emission wavelength), with some measurements in multiple solvents: 31 from $CHCl_3$, 15 from THF, and 35 in film.

Descriptor Information. To represent each molecule, we chose five descriptors: (i) HOMO, (ii) E_g, (iii) absorption maxima ($\lambda_{max}{}^{abs}$), and (iv) KK (or K × K, which is the square of the displacement vector between the ground and excited state coordinates based on the vertical gradient approximation)[61] from the DFT calculations as well as (v) the solvent system from the experimental measurements. The molecules' features are described in Table 1 below. Since LUMO is a linear combination of E_g and HOMO (E_g + HOMO = LUMO), we excluded LUMO from the feature set.

Table 1. Descriptions of the DFT Features of the Molecules Used in the Modeling

DFT features	descriptions
HOMO	highest occupied molecular orbital in eV
E_g	energy gap (HOMO − LUMO) in eV
λ_{max}^{abs}	wavelength at maximum absorbance in nm
KK	measure of vertical displacement
solvent	solvent used in experimental observations
LUMO*	lowest unoccupied molecular orbital in eV

As previously stated, the unique structure of BBO allows for modulation of the HOMO and LUMO levels with varying electronic groups along either axis. This consequently affects ground to excited state transitions, such as narrowing of E_g with extension of conjugation along the axes. A representative example is the BBO **26BT48BT**, shown in Figure 2, where BT refers to bithiophene groups along the 2,6 and 4,8 axes.

We used the calculated data for the 50 BBO cruciform molecules and the 29 fragment molecules and the solvent measurements for the cruciform and fragment models, respectively. The distributions of these descriptors are shown in Figure 3.

ML Model Selection and Building. While deep learning models are popular for ML researchers, they typically rely on prohibitively large data sets for experimental data. Therefore, with only 50 molecules in our data set, we relied on *eXtreme Gradient Boosting (XGBoost)*.[62] XGBoost is a popular model that has exhibited cutting-edge performance on various prediction tasks. Furthermore, due to the heterogeneity of the molecules in our data set, we employed an ensemble framework to reduce the variation in our predictions.[63] We then validated our use of the ensemble learning approach by conducting a prediction stability analysis which justified our use of an ensemble framework.[64] Details of the analysis can be found in the Supporting Information (Section S3.2). We also implemented a GBM (Section S1.3, Table S4) for model comparison.[65]

We have two ensemble frameworks: cruciform framework that uses the DFT features of the BBOs as the predictors; and the fragment framework, that uses the DFT features of the horizontal and vertical fragments of the BBOs as the predictors. Each of the frameworks is an ensemble of 100 component models (weak learners), and we fit each component model using a XGBoost and GBM with λ_{max}^{emis} as the response and RMSE as the metric. We used cross-validation to select the hyperparameters for our XGBoost and GBM models, and the 4 parameters include the maximum number of iterations, maximum depth of the tree, learning rate, minimum loss reduction required for a split. Specifically, for each of the hyper-parameters, we selected some as candidate values, and we performed a grid search with a 10-fold cross-validation (we employed an 80/10/10 train/validate/test split of our complete data set).[55] This resulted in the best combination of the parameter values in terms of the averaged RMSE on validation set. Each component model was then trained and tested with a random train/validation/test split, and we reported the averaged RMSE of the 100 component models as the performance metric for the ensemble framework (Table S3). The details of the algorithm is in Section S2 in the Supporting Information.[55] We also evaluated our models using other metrics, namely MAE, MSE and MAPE, and their details can be found in Section S3.1 of the Supporting Information. The results from each of these metrics are qualitatively comparable to the RMSE results in terms of predicting the emission wavelength for new molecules.

Model Validation. Additionally, to assess the accuracy of our models on unseen data, we used our models to predict emissions for 4 BBO-based molecules that are not included in the training data set, referred to as the "holdout" BBOs. They contain the same BBO core as the molecules in our database and are functionalized by adamantyl or phenyl groups on the 2,6 axis and carbazole groups on the 4,8 axis (see github.com/KolaczykResearch/Blue-BBO-ML). The DFT properties of these 4 BBOs are in the range of the DFT values of the molecules in our training data set (see Figure 3). Therefore, we expected our models to make accurate predictions of the emission of these four molecules. A manuscript detailing their design, optoelectronic, and device features is under review elsewhere.

RESULTS & DISCUSSION

Model Training (Performance) Analysis. As a consequence of the small size of our data set, the random sampling of our train/validate/test split can have undue influence. To mitigate this concern, we employed an ensemble learning framework. Specifically, we repeated the model fitting

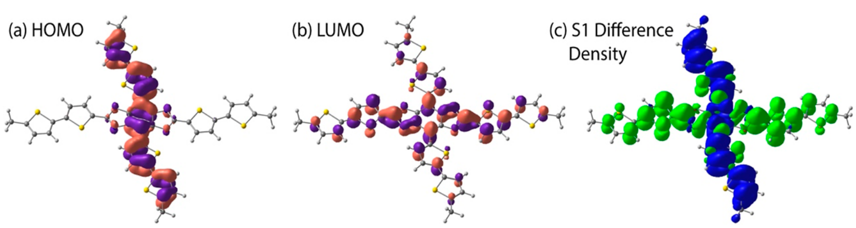


Figure 2. (a) HOMO, (b) LUMO and (c) S1 difference density for a representative cruciform (**26BT48BT**) with isosurfaces of 3×10^{-2} for (a,b) and 3×10^{-4} for (c).

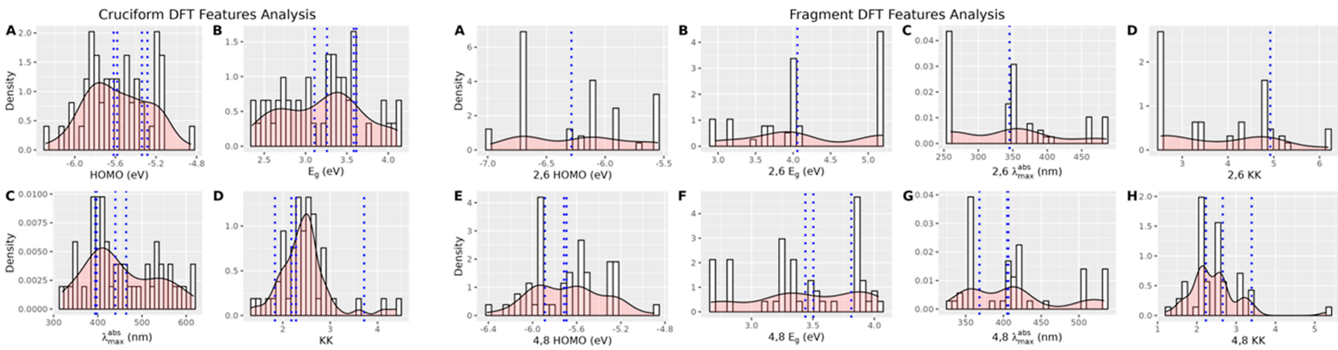


Figure 3. Distributions of the DFT features used in our models (histogram and density plot) and the DFT values of the four BBO-based molecules not included in the training set (blue dots).

procedure 100 times, obtaining 100 separate XGBoost models (eq (1)). Each XGBoost model has a prediction for the emission wavelength for the i-th molecule as $\hat{y}_i = \sum_{k=1}^{K} f_k(x_i)$, where f_k is the k-th regression tree model, x_i is the predictors for the i-th molecule, and K is the number of regression trees within each XGBoost model. We computed the RMSE on it is the corresponding test set for each component model, which we recall is a random sample of 10% of the entire data set (eq S2) and the formula for RMSE is given as $\text{RMSE} = \sqrt{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / n}$, where y_i is the true emission measurement for the i-th molecule. The RMSEs range from 19 to 51 nm, a level of variability that corroborates our use of an ensemble framework.[66] Our final model is the average of our 100 components. The detailed methodology can be found in Section S2 of Supporting Information.

Table 2 summarizes the average RMSE of each model component on its corresponding test set and the average

Table 2. RMSE of the Predicted Emission for the Cruciform and Fragment Model for the BBO Molecules in the Training Set and the New BBO Molecules Using XGBoost[a]

	cruciform RMSE Avg & SD (nm)	fragment RMSE Avg & SD (nm)
training BBOs	31.60 ± 7.69	31.08 ± 7.71
holdout BBOs	35.30	31.50

[a]Avg = average; SD = standard deviation.

RMSE on the four holdout molecules. The fragment model performs on par with the cruciform model (31.08 vs 31.60 nm) on the cross-validated results and is comparable to the holdout BBOs (31.50 vs 35.30 nm respectively). Notably, the fragment model is significantly less expensive due to the total time required to compute the DFT properties of the data sets, which we report in Table 3. This suggests that the fragment model offers competitive predictions at a much lower computational cost when considering small databases with few features.

Table 3. Time to Run DFT Calculations on the BBO Cruciform (BBO-c) and the Fragment (BBO-f) Molecules[a]

cruciform Avg (SD) (hr/BBO-c)	fragment Avg (SD) (hr/BBO-f)
24.70 (41.76)	10.95 (21.9)

[a]Avg = average; SD = standard deviation.

Result Interpretation with Accumulated Local Effects Plots. A major drawback of many modern ML methods is that they lack interpretability. A recent method to disentangle the effects of predictors in "black box" models is the *accumulated local effects* (ALE) plot.[67] Although similar approaches have been used in the past, such as marginal plots and partial dependence plots,[65] they suffer from ignoring correlation in the features.[68] ALE plots, on the other hand, are explicitly designed to estimate how marginal changes in feature space affect predictions on observations in that small interval. This is particularly important for DFT features with known physical relationships, such as E_g, HOMO, and λ_{max}^{abs}. Since we used an ensemble learning framework, our ALE plots are computed by averaging over the ALE plots of each model component. To study the feature effects, we investigated a natural generalization of ALE plots to ensemble models by averaging over the ALE plots of the model components. The detailed methodology can be found in Section S4.2 of the Supporting Information. The ensemble ALE plots for the E_g, HOMO, and λ_{max}^{abs} in Figure 4 were generated for the cruciform (**C**) features as well as the 2,6-Fragment (**26F**) and 4,8-Fragment (**48F**). Essentially, a positive slope in the ALE plot indicates that an increase in the feature on the x-axis would lead to an increase in the prediction, and vice versa for negative slopes. These are computed based on predictions for similar molecules, hence they capture a notion of local dependency. In addition, the absolute value of the slope measures the strength of the effect at the value of the feature on the x-axis. The flat region suggests that the changes in the feature on the x-axis within that region have little impact on the prediction. Hence, our analysis shows a clear order of influence where E_g > HOMO > λ_{max}^{abs} i.e. changes in E_g result in adjustments of the predicted outcome more than changes in HOMO and absorption wavelengths do. When we considered the E_g feature for the **26F**, the **48F**, and the **C**, they all have an inverse relationship with the predicted values (Figure 4a−c). For example, in **C**, an increase of 1 eV in the E_g results in a shift of 56 nm in the model's predicted emission. Similarly, for **26F** and **48F**, the decrease in prediction would be 12 and 41 nm, respectively.

We hypothesized that there is an axial dominance in play where substitution along the 4,8 axes of the fragment has a more significant effect on the cruciform molecules' properties, as most of the 48 fragments are sterically distorted in comparison to the primarily planar 26 fragments (Figure 1 left, github.com/KolaczykResearch/Blue-BBO-ML for structural details of each fragment). Our feature analysis indicates a negligible effect of the **26F** HOMO on the output (Figure 4d).

https://doi.org/10.1021/acs.jpca.4c00077
J. Phys. Chem. A 2024, 128, 6116−6123

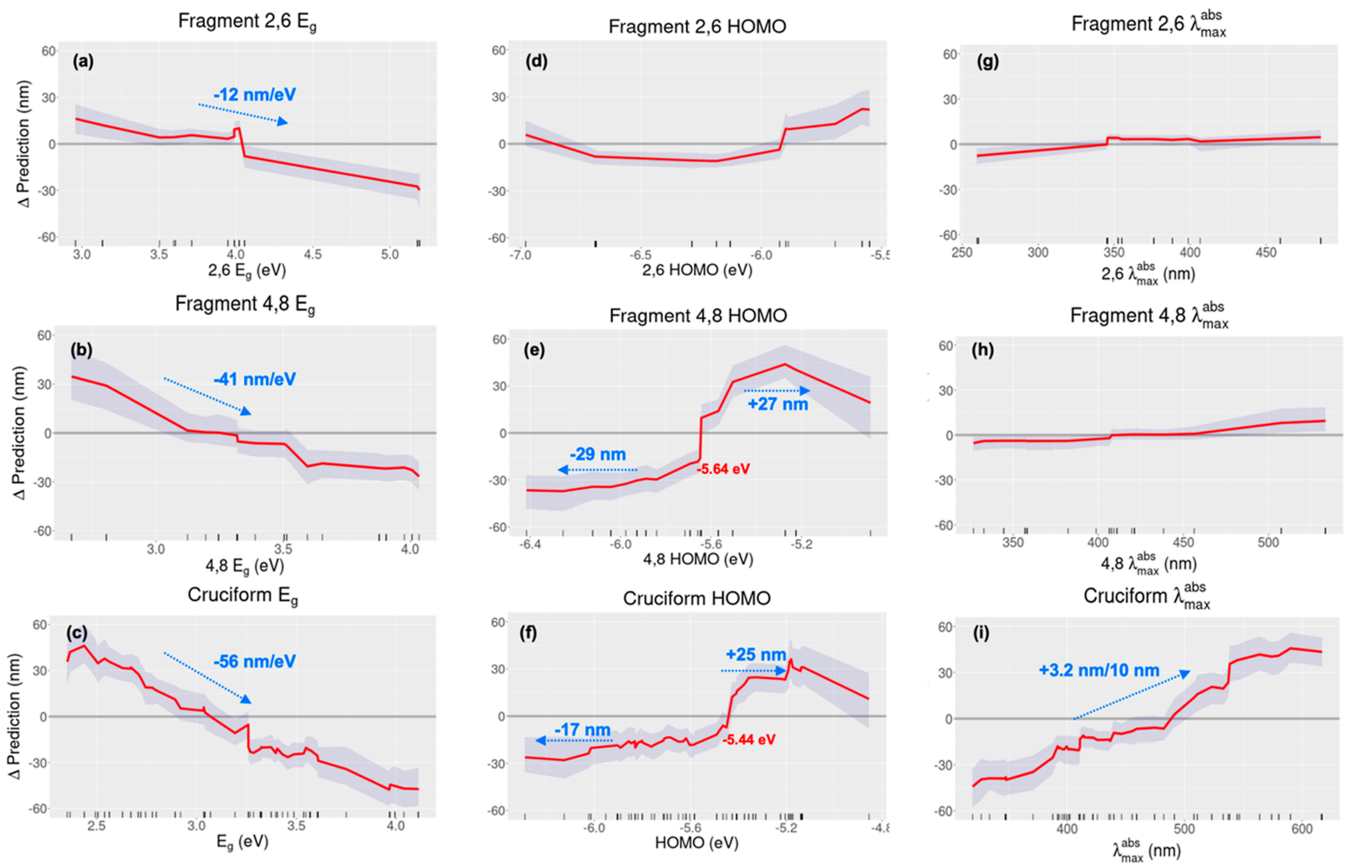


Figure 4. Accumulated local effects of the EG (a−c), HOMO (d−f), $\lambda_{max}{}^{abs}$ (g−i) individually on the prediction. The ribbon (gray) represents the 90% confidence bands for the ALE effects at different observations, and the mean effect is in red.

In contrast, the **48F**'s HOMO levels affect the emission output (Figure 4e). We found a "threshold value," which is the value when either positive or negative adjustments must be made to the predictions depending on the direction. The HOMO threshold value for the **48F** feature effect is at −5.64 eV, above which (right arrow) the output is positively adjusted by 27 nm. Conversely, when the HOMO is under −5.64 eV (left arrow), the outcome is adjusted by subtracting 29 nm. Finally, when we considered the **C** HOMO features, we found it followed a similar adjustment pattern to the **48F**. However, the threshold point here was −5.44 eV with a positive adjustment of 35 nm on the right and a negative adjustment of 17 nm on the left (Figure 4f).

Interestingly, in our analysis of the λ_{max}^{abs} feature, the fragment features had little effect on the final prediction (Figure 4g,h). However, the λ_{max}^{abs} feature for the **C** suggests a positive adjustment of 3.2 nm in the output for every 10 nm increase in the input data (Figure 4i). The ALE studies of the KK and solvent did not significantly affect the predicted emissions (Figure S2). Bivariate ALE plots can provide insight into relationships between two features and their sum/overall effect on the predicted outcome. However, our bivariate feature analysis indicates a negligible impact on predictions when we consider the interaction between features such as the HOMO and λ_{max}^{abs}, E_g and HOMO, E_g, and $\lambda_{max}{}^{abs}$, etc. Details of these plots can be found in the Supporting Information (Figures S3−S7, Table S7).

■ CONCLUSIONS

In this work, we have utilized five features (E_g, HOMO, $\lambda_{max}{}^{abs}$, KK, solvent) that play a role in the emission properties of a molecule. To overcome the small data challenge of our database, we employed gradient-boosted learning models in an ensemble framework. We compared a model trained on features of a whole cruciform and one trained on features of the cruciforms' constituent fragments. Both models showed remarkable accuracy, as evidenced by RMSEs of 30−36 nm for the predicted emissions, which are highly comparable to other models with 1000s of molecules/inputs in their training sets. We built our database from cruciform molecules and their axis fragment molecules. As expected, it was much faster to calculate DFT features for the structurally simpler fragments as there are fewer components than there are combinations that define the cruciforms ($n_1 + n_2$ versus $n_1 \times n_2$ with n_1 being the 2,6 fragments and n_2 the 4,8 fragments). However, it remains to be seen if this simpler model is competitive on larger data sets.

As we hoped, our model could predict emissions for molecules not part of its training and we predicted the emission wavelengths for four new/held-out molecules, showing remarkably accurate RMSEs between 31 and 36 nm. Feature effects are highly consequential in understanding how ML models generate their predictions. We have utilized ALE analyses to determine the most essential features and how they affect predicted emissions. We found that the E_g and $\lambda_{max}{}^{abs}$ features play significant roles in predicting emissions, and their variation results in a linear adjustment in the outcome. In contrast, we have identified threshold values for HOMO that

https://doi.org/10.1021/acs.jpca.4c00077
J. Phys. Chem. A 2024, 128, 6116−6123

determine whether a positive or negative adjustment is needed. We believe these relationships could be of independent interest beyond our abilities to predict emissions accurately.

Encouraged by the promising results of our small database, we are currently working to extend them by adding more molecules. To this end, we are expanding our database with the latest designed molecules via a novel high throughput experimentation platform, allowing for faster screening of potential BBO-based emitters. This synthetic technique has been explored in the discovery of novel molecules with catalysis and pharmaceutical applications[69,70] and has the potential to aid organic electronics design, too. A more extensive database will inform us on how to improve our ML algorithm, and coupled with interpretable models, they will pave the way for more sophisticated learning methods and, eventually, AI-guided OLED design.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.4c00077.

Model details, database, XGBoost and GBM methods, ensemble learning, details of the model results, details of the model interpretation (PDF)

AUTHOR INFORMATION

Corresponding Authors

Eric D. Kolaczyk − *Department of Mathematics and Statistics, Boston University, Boston, Massachusetts 02215, United States; Department of Mathematics and Statistics, McGill University, Montreal, QC H3A 0G4, Canada*; Email: eric.kolacyzk@mcgill.ca

Malika Jeffries-EL − *Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States; Division of Material Science and Engineering, Boston University, Boston, Massachusetts 02215, United States*; orcid.org/0000-0002-9134-4938; Email: malikaj@bu.edu

Authors

Shambhavi Tannir − *Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States*; orcid.org/0000-0003-3878-4798

Yuning Pan − *Department of Mathematics and Statistics, Boston University, Boston, Massachusetts 02215, United States*

Nathaniel Josephs − *Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, United States*

Christopher Cunningham − *Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States*

Nathan R. Hendrick − *Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States*

Annie Beckett − *Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States*

James McNeely − *Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States*

Aaron Beeler − *Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States*; orcid.org/0000-0002-2447-0651

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.4c00077

Author Contributions

[#]S.T. and Y.P. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to thank the National Science Foundation (DMS-2141384) for financial support of this work. We also thank the Boston University Supercomputing Centre (BU-SCC) for supercomputer allocation for this work.

REFERENCES

(1) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science* **2018**, *361* (6400), 360−365.

(2) Sun, W.; Li, M.; Li, Y.; Wu, Z.; Sun, Y.; Lu, S.; Xiao, Z.; Zhao, B.; Sun, K. The Use of Deep Learning to Fast Evaluate Organic Photovoltaic Materials. *Adv. Theory Simul.* **2019**, *2* (1), 1800116.

(3) Correa-Baena, J.-P.; Hippalgaonkar, K.; van Duren, J.; Jaffer, S.; Chandrasekhar, V. R.; Stevanovic, V.; Wadia, C.; Guha, S.; Buonassisi, T. Accelerating Materials Development via Automation, Machine Learning, and High-Performance Computing. *Joule* **2018**, *2* (8), 1410−1420.

(4) Sun, W.; Zheng, Y.; Yang, K.; Zhang, Q.; Shah, A. A.; Wu, Z.; Sun, Y.; Feng, L.; Chen, D.; Xiao, Z.; et al. Machine Learning-Assisted Molecular Design and Efficiency Prediction for High-Performance Organic Photovoltaic Materials. *Sci. Adv.* **2019**, *5* (11), 1−8.

(5) Saal, J. E.; Oliynyk, A. O.; Meredig, B. Machine Learning in Materials Discovery: Confirmed Predictions and Their Underlying Approaches. *Annu. Rev. Mater. Res.* **2020**, *50*, 49−69.

(6) Moosavi, S. M.; Jablonka, K. M.; Smit, B. The Role of Machine Learning in the Understanding and Design of Materials. *J. Am. Chem. Soc.* **2020**, *142* (48), 20273−20287.

(7) Zeng, J.; Tao, Y.; Giese, T. J.; York, D. M. Modern Semiempirical Electronic Structure Methods and Machine Learning Potentials for Drug Discovery: Conformers, Tautomers, and Protonation States. *J. Chem. Phys.* **2023**, *158* (12), 124110.

(8) Xiao, F.; Gao, H.; Lei, Y.; Dai, W.; Liu, M.; Zheng, X.; Cai, Z.; Huang, X.; Wu, H.; Ding, D. Guest-host doped strategy for constructing ultralong-lifetime near-infrared organic phosphorescence materials for bioimaging. *Nat. Commun.* **2022**, *13*, 186. (2022)

(9) Pode, R. Organic Light Emitting Diode Devices: An Energy Efficient Solid State Lighting for Applications. *Renewable Sustainable Energy Rev.* **2020**, *133* (July 2019), 110043.

(10) Wu, X.; Mao, S.; Chen, J.; Huang, J. Strategies for Improving the Performance of Sensors Based on Organic Field-Effect Transistors. *Adv. Mater.* **2018**, *30* (17), 1705642.

(11) Hong, G.; Gan, X.; Leonhardt, C.; Zhang, Z.; Seibert, J.; Busch, J. M.; Bräse, S. A Brief History of OLEDs—Emitter Development and Industry Milestones. *Adv. Mater.* **2021**, *33* (9), 2005630.

(12) Carter, C. M.; Cho, J.; Glanzer, A.; Kamcev, N.; O'Carroll, D. M. Cost, Energy and Emissions Assessment of Organic Polymer Light-Emitting Device Architectures. *J. Cleaner Prod.* **2016**, *137*, 1418−1431.

(13) Wei, Q.; Fei, N.; Islam, A.; Lei, T.; Hong, L.; Peng, R.; Fan, X.; Chen, L.; Gao, P.; Ge, Z. Small-Molecule Emitters with High Quantum Efficiency: Mechanisms, Structures, and Applications in OLED Devices. *Adv. Opt. Mater.* **2018**, *6* (20), 1800512.

(14) Eng, J.; Thomas, J. P. Open Questions on the Photophysics of Thermally Activated Delayed Fluorescence. *Commun. Chem.* **2021**, 91.

(15) Kuehne, A. J. C.; Gather, M. C. Organic Lasers: Recent Developments on Materials, Device Geometries, and Fabrication Techniques. *Chem. Rev.* **2016**, *116* (21), 12823−12864.

(16) Samuel, I. D. W.; Turnbull, G. A. Organic Semiconductor Lasers. *Chem. Rev.* **2007**, *107* (4), 1272−1295.
(17) Minotto, A.; Haigh, P. A.; Łukasiewicz, Ł. G.; Lunedei, E.; Gryko, D. T.; Darwazeh, I.; Cacialli, F. Visible Light Communication with Efficient Far-Red/near-Infrared Polymer Light-Emitting Diodes. *Light: Sci. Appl.* **2020**, *9* (1), 70.
(18) López-Fraguas, E.; Arredondo, B.; Vega-Colado, C.; del Pozo, G.; Najafi, M.; Martín-Martín, D.; Galagan, Y.; Sánchez-Pena, J. M.; Vergaz, R.; Romero, B. Visible Light Communication System Using an Organic Emitter and a Perovskite Photodetector. *Org. Electron.* **2019**, *73*, 292−298.
(19) Jin, S.-H. Synthesis and Electroluminescent Properties of Poly(p-Phenylene-Vinylene)s with 3′,3′-Diheptyl-3,4-Propylenedioxythiophene Pendant Group for Light-Emitting Diode Applications *Organic Light Emitting Materials and Devices XI*; SPIE, 2007; Vol. *6655*, p 665507.
(20) Ha, J. M.; Hur, S. H.; Pathak, A.; Jeong, J.-E.; Woo, H. Y. Recent Advances in Organic Luminescent Materials with Narrowband Emission. *NPG Asia Mater.* **2021**, *13* (1), 53.
(21) Oda, S.; Sugitani, T.; Tanaka, H.; Tabata, K.; Kawasumi, R.; Hatakeyama, T. Development of Pure Green Thermally Activated Delayed Fluorescence Material by Cyano Substitution. *Adv. Mater.* **2022**, *34* (32), 2201778.
(22) Chen, C.-T. Evolution of Red Organic Light-Emitting Diodes: Materials and Devices. *Chem. Mater.* **2004**, *16* (23), 4389−4400.
(23) Zou, Y.; Hu, J.; Yu, M.; Miao, J.; Xie, Z.; Qiu, Y.; Cao, X.; Yang, C. High-Performance Narrowband Pure-Red OLEDs with External Quantum Efficiencies up to 36.1% and Ultralow Efficiency Roll-Off. *Adv. Mater.* **2022**, *34* (29), 2201442.
(24) Monkman, A. Why Do We Still Need a Stable Long Lifetime Deep Blue OLED Emitter? *ACS Appl. Mater. Interfaces* **2022**, *14* (18), 20463−20467.
(25) Setzer, T.; Friederich, P.; Meded, V.; Wenzel, W.; Lennartz, C.; Dreuw, A. Meltdown! Local Heating by Decaying Excited Host Positive Polarons Triggers Aggregation Quenching in Blue PhOLEDs. *ChemPhysChem* **2018**, *19* (21), 2961−2966.
(26) Yang, X.; Xu, X.; Zhou, G. Recent Advances of the Emitters for High Performance Deep-Blue Organic Light-Emitting Diodes. *J. Mater. Chem. C* **2015**, *3* (5), 913−944.
(27) International Commission on Illumination *Commission Internationale de l'eclairage*; University Press: Cambridge, 1932.
(28) Mike, J. F.; Makowski, A. J.; Jeffries-EL, M. An Efficient Synthesis of 2, 6-Disubstituted Benzobisoxazoles: New Building Blocks for Organic Semiconductors. *Org. Lett.* **2008**, *10* (21), 4915−4918.
(29) Tlach, B. C.; Tomlinson, A. L.; Morgan, K. D.; Collins, C. R.; Zenner, M. D.; Jeffries-EL, M. Effect of Extended Conjugation on the Optoelectronic Properties of Benzo[1,2-d:4,5-d′]bisoxazole Polymers. *Aust. J. Chem.* **2014**, *67*, 711−721.
(30) Chavez III, R.; Cai, M.; Tlach, B.; Wheeler, D. L.; Kaudal, R.; Tsyrenova, A.; Tomlinson, A. L.; Shinar, R.; Shinar, J.; Jeffries-EL, M. Benzobisoxazole Cruciforms: A Tunable, Cross-Conjugated Platform for the Generation of Deep Blue OLED Materials. *J. Mater. Chem. C* **2016**, *4* (17), 3765−3773.
(31) Intemann, J. J.; Hellerich, E. S.; Ewan, M. D.; Tlach, B. C.; Speetzen, E. D.; Shinar, R.; Shinar, J.; Jeffries-EL, M. Investigating the Impact of Conjugation Pathway on the Physical and Electronic Properties of Benzobisoxazole-Containing Polymers. *J. Mater. Chem. C* **2017**, *5*, 12839−12847.
(32) Wheeler, D. L.; Diodati, A. V.; Tomlinson, L.; Jeffries-EL, M. Evaluating the Role of Molecular Heredity in the Optical and Electronic Properties of Cross-Conjugated Benzo[1,2 - D:4,5 - d ′]Bisoxazoles? *ACS Omega* **2020**, *5* (21), 12374−12384.
(33) Burney-Allen, A. A.; Shaw, J.; Wheeler, D. L.; Diodati, L.; Duzhko, V.; Tomlinson, A. L.; Jeffries-el, M. Benzobisoxazole Cruciforms: A Cross-Conjugated Platform for Designing Tunable Donor/Acceptor Materials. *Asian J. Org. Chem.* **2021**, *10*, 215−223.
(34) Wheeler, D. L.; Tannir, S.; Smith, E.; Tomlinson, A. L.; Jeffries-EL, M. A Computational and Experimental Investigation of Deep-Blue Light-Emitting Tetraaryl-Benzobis[1,2-d:4,5-d′]Oxazoles. *Mater. Adv.* **2022**, *3*, 3842−3852.
(35) Wheeler, D. L.; Fisher, L.; Friederich, P.; Cunningham, C.; Muthike, A. K.; Aspuru-Guzik, A.; Goodson, T.; Jeffries-EL, M. Carbazole-Substituted Benzobisoxazoles: Near-UV Fluorescent Emitters and Ambipolar Hosts for Organic Light-Emitting Diodes. *J. Mater. Chem. C* **2023**, *11*, 211−222.
(36) Tlach, B. C.; Tomlinson, A. L.; Ryno, A. G.; Knoble, D. D.; Drochner, D. L.; Krager, K. J.; Jeffries-EL, M. Influence of Conjugation Axis on the Optical and Electronic Properties of Aryl-Substituted Benzobisoxazoles. *J. Org. Chem.* **2013**, *78* (13), 6570−6581.
(37) Wheeler, D. L.; Diodati, A. V.; Tomlinson, A. L.; Jeffries-EL, M. Evaluating the Role of Molecular Heredity in the Optical and Electronic Properties of Cross-Conjugated Benzo[1,2-d:4,5-D′]-Bisoxazoles. *ACS Omega* **2020**, *5* (21), 12374−12384.
(38) Xu, B.; Mu, Y.; Mao, Z.; Xie, Z.; Wu, H.; Zhang, Y.; Jin, C.; Chi, Z.; Liu, S.; Xu, J.; et al. Achieving Remarkable Mechanochromism and White-Light Emission with Thermally Activated Delayed Fluorescence through the Molecular Heredity Principle. *Chem. Sci.* **2016**, *7* (3), 2201−2206.
(39) Ghosh, S.; Bhattacharyya, K. Origin of the Failure of Density Functional Theories in Predicting Inverted Singlet-Triplet Gaps. *J. Phys. Chem. A* **2022**, *126* (8), 1378−1385.
(40) Körzdörfer, T.; Brédas, J. L. Organic Electronic Materials: Recent Advances in the DFT Description of the Ground and Excited States Using Tuned Range-Separated Hybrid Functionals. *Acc. Chem. Res.* **2014**, *47* (11), 3284−3291.
(41) Bhattacharyya, K. Comparative Analyses of Data Driven Machine Learning Models for TADF Emitters. *ChemRxiv* **2023**, chemrxiv:2023-zj3tv.
(42) Garain, B. C.; Pati, S. K. Improved Prediction of Maximum EQE in TADF-Based OLEDs Through Ensemble Learning. *ChemRxiv* **2022**, chemrxiv:2022-636c2.
(43) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; et al. PubChem Substance and Compound Databases. *Nucleic Acids Res.* **2016**, *44* (D1), D1202−D1213.
(44) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; et al. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40* (D1), D1100−D1107.
(45) Rodrigues, T. The Good, the Bad, and the Ugly in Chemical and Biological Data for Machine Learning. *Drug Discovery Today: Technol.* **2019**, *32*−*33*, 3−8.
(46) Cai, J.; Luo, J.; Wang, S.; Yang, S. Feature Selection in Machine Learning: A New Perspective. *Neurocomputing* **2018**, *300*, 70−79.
(47) Kononova, O.; He, T.; Huo, H.; Trewartha, A.; Olivetti, E. A.; Ceder, G. Opportunities and Challenges of Text Mining in Materials Research. *iScience* **2021**, *24* (3), 102155.
(48) Wu, Y.; Guo, J.; Sun, R.; Min, J. Machine Learning for Accelerating the Discovery of High-Performance Donor/Acceptor Pairs in Non-Fullerene Organic Solar Cells. *npj Comput. Mater.* **2020**, *6* (1), 120−128.
(49) Zhang, Y.; Ling, C. A Strategy to Apply Machine Learning to Small Datasets in Materials Science. *npj Comput. Mater.* **2018**, *4* (1), 25.
(50) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K. R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9* (8), 3404−3419.
(51) Yamada, H.; Liu, C.; Wu, S.; Koyama, Y.; Ju, S.; Shiomi, J.; Morikawa, J.; Yoshida, R. Predicting Materials Properties with Little Data Using Shotgun Transfer Learning. *ACS Cent. Sci.* **2019**, *5* (10), 1717−1730.
(52) Kim, S.; Cho, D.; Kim, S.; Kang, S.; Koh, E.; Kang, I.; Kim, H.; Mi Cho, Y.; Kim, Y. 25-3: Machine-Learning-Assisted Materials

https://doi.org/10.1021/acs.jpca.4c00077
J. Phys. Chem. A 2024, 128, 6116−6123

Discovery of Blue Emitter for More Efficient and Durable OLED Device. *SID Int. Symp. Dig. Tech. Pap.* **2021**, *52*, 314−316.

(53) Sun, W.; Zheng, Y.; Yang, K.; Zhang, Q.; Shah, A. A.; Wu, Z.; Sun, Y.; Feng, L.; Chen, D.; Xiao, Z.; et al. Machine Learning-Assisted Molecular Design and Efficiency Prediction for High-Performance Organic Photovoltaic Materials. *Sci. Adv.* **2019**, *5* (11), 1−9.

(54) Kuzmich, A.; Padula, D.; Ma, H.; Troisi, A. Trends in the electronic and geometric structure of non-fullerene based acceptors for organic solar cells. *Environ. Sci.* **2017**, *10*, 395−401.

(55) Joung, J. F.; Han, M.; Hwang, J.; Jeong, M.; Choi, D. H.; Park, S. Deep Learning Optical Spectroscopy Based on Experimental Database: Potential Applications to Molecular Design. *JACS Au* **2021**, *1* (4), 427−438.

(56) Jeong, M.; Joung, J. F.; Hwang, J.; Han, M.; Koh, C. W.; Choi, D. H.; Park, S. Deep Learning for Development of Organic Optoelectronic Devices: Efficient Prescreening of Hosts and Emitters in Deep-Blue Fluorescent OLEDs. *npj Comput. Mater.* **2022**, *8* (1), 147.

(57) Neese, F. The ORCA Program System. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2* (1), 73−78.

(58) Neese, F. Software Update: The ORCA Program System, Version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8* (1), No. e1327.

(59) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA Quantum Chemistry Program Package. *J. Chem. Phys.* **2020**, *152* (22), 224108.

(60) Neese, F. Software Update: The ORCA Program System—Version 5.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2022**, *12* (5), No. e1606.

(61) de Souza, B.; Neese, F.; Izsák, R. On the Theoretical Prediction of Fluorescence Rates from First Principles Using the Path Integral Approach. *J. Chem. Phys.* **2018**, *148* (3), 34104.

(62) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Min.*; Association for Computing Machinery, 2016, pp 784−795.

(63) Hastie, T.; Tibshirani, R.; Friedman, J.; Hastie, T., Tibshirani, R., Friedman, J., Eds.; Springer New York: New York, NY, 2009; pp 605−624.*Ensemble Learning BT - The Elements of Statistical Learning: Data Mining, Inference, and Prediction*

(64) Hintze, J. L.; Nelson, R. D. Violin Plots: A Box Plot-Density Trace Synergism. *Am. Stat.* **1998**, *52* (2), 181−184.

(65) Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29* (5), 1189−1232.

(66) Thiesen, R.; Kim, H.; Yang, Y.; Hodgkinson, L.; Mahoney, M. When are Ensembles Really Effective? *Advances in Neural Information Processing Systems 36*; NeurIPS Proceedings, 2023; Vol. *36*, pp 15015−15026.

(67) Apley, D. W.; Zhu, J. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2020**, *82* (4), 1059−1086.

(68) Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*; 2nd ed.; Lulu.com, 2023.

(69) Chatterjee, S.; Guidi, M.; Seeberger, P. H.; Gilmore, K. Automated Radial Synthesis of Organic Molecules. *Nature* **2020**, *579* (7799), 379−384.

(70) Buitrago Santanilla, A.; Regalado, E. L.; Pereira, T.; Shevlin, M.; Bateman, K.; Campeau, L.-C.; Schneeweis, J.; Berritt, S.; Shi, Z.-C.; Nantermet, P.; et al. Nanomole-Scale High-Throughput Chemistry for the Synthesis of Complex Molecules. *Science* **2015**, *347* (6217), 49−53.

https://doi.org/10.1021/acs.jpca.4c00077
J. Phys. Chem. A 2024, 128, 6116−6123