ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Quantum Chemistry–Machine Learning Approach for Predicting Properties of Lewis Acid–Lewis Base Adducts

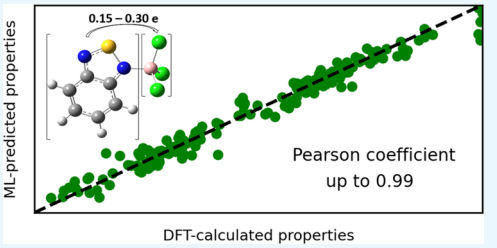
Hieu Huynh, Thomas J. Kelly, Linh Vu, Tung Hoang, Phuc An Nguyen, Tu C. Le, Emily A. Jarvis,* and Hung Phan*

Cite This: *ACS Omega* 2023, 8, 19119–19127

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Synthetic design allowing predictive control of charge transfer and other optoelectronic properties of Lewis acid adducts remains elusive. This challenge must be addressed through complementary methods combining experimental with computational insights from first principles. *Ab initio* calculations for optoelectronic properties can be computationally expensive and less straightforward than those sufficient for simple ground-state properties, especially for adducts of large conjugated molecules and Lewis acids. In this contribution, we show that machine learning (ML) can accurately predict density functional theory (DFT)-calculated charge transfer and even properties associated with excited states of adducts from readily obtained molecular descriptors. Seven ML models, built from a dataset of over 1000 adducts, show exceptional performance in predicting charge transfer and other optoelectronic properties with a Pearson correlation coefficient of up to 0.99. More importantly, the influence of each molecular descriptor on predicted properties can be quantitatively evaluated from ML models. This contributes to the optimization of a priori design of Lewis adducts for future applications, especially in organic electronics.



INTRODUCTION

Optimizing functional materials under real working conditions is essential but can be a tedious and expensive process. Accordingly, this has been assisted by computational tools for several decades. A variety of chemical properties can be predicted using current computational methods, but these often require computing the quantum mechanical wavefunction, a costly endeavor, especially for large molecules and any properties involving excited states and atypical bonding. In the last decade, machine learning (ML) is becoming a versatile computing tool to assist molecular design and optimization, together with calculations from physical laws.

ML algorithms have been successfully employed for classification, regression, clustering, or dimensionality reduction tasks of large sets of input data. Machine learning is promising to solve data bottlenecks in many problems in chemistry and materials science.[1−6] Solutions employing machine learning offer advances to screen high volumes of compounds for advanced material applications ranging from efficient organic photovoltaics[7−12] to organic light-emitting diodes[13,14] to high-temperature alloys[15] and many more.

Having eligible descriptors is critical for developing many ML models for practical applications as well as advancing fundamental understanding. Some descriptors can be computed readily from molecular structures, which can be called molecular descriptors. These descriptors, for example, the number of hydrogen atoms or molecular weight or the number of conjugated bonds, can be generated readily and economically for use in ML models.[16−18] In several reports, a variety of descriptors were obtained from the traditional quantum approach, so-called quantum descriptors.[9,19,20] Examples include frontier molecular orbital energies, electron population, and triplet states. Despite the advantage of being calculated from first principles, compared to molecular descriptors, quantum descriptors present the challenges of time- and resource-consuming calculations and the uncertainty originating from their dependency on the level of theory employed.

In addition to predicting the final performance of functional devices, a variety of fundamental properties of molecules and materials have been predicted by machine learning.[21−25] For example, machine learning can be employed to predict the energies of highest occupied molecular orbital–lowest unoccupied molecular orbital (HOMO–LUMO) orbitals,[26] lattice energies,[27] and charge transfer integrals of organic crystals.[24,27] Many such important properties are either very challenging to obtain experimentally or complicated and time-consuming to be calculated using traditional computational tools. Machine learning can be applied to rapidly screen and

Received: April 25, 2023
Accepted: May 9, 2023
Published: May 19, 2023

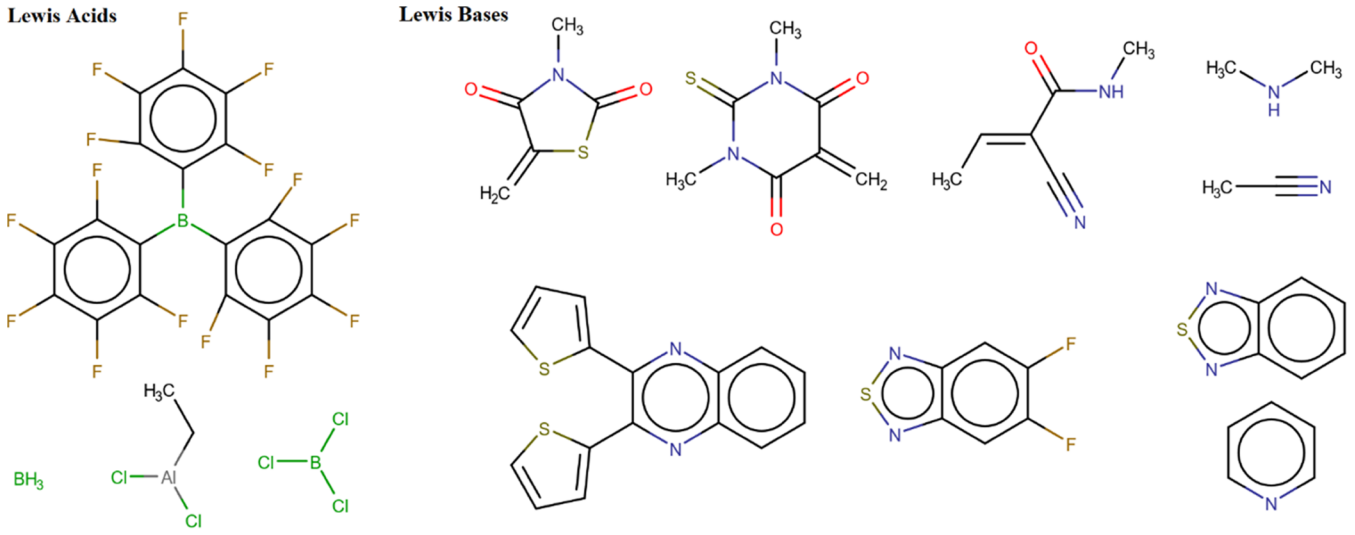


Figure 1. Chemical structures of representative Lewis acids and Lewis bases.

predict those properties. For example, intramolecular reorganization energy, which is an important property of organic semiconductors, is typically an expensive calculation using *ab initio* methods. Recently, the employment of deep neural networks or kernel ridge regression ML models significantly reduced the time and power required for computing intramolecular reorganization.[28] Machine learning has been shown, in combination with legacy quantum methods, to increase the accuracy of computed properties with a minimal increase in computational expense.[29]

In this work, we aim to employ machine learning using readily obtained molecular descriptors to predict properties that can be calculated using density functional theory (DFT) but are costly and highly sensitive to calculation details. We chose a model dataset to be the adducts of Lewis bases (LBs) and Lewis acids (LAs). In addition to showcasing classic coordinate covalent bonds, our model dataset is inspired from recent work using Lewis acids to form adducts with organic semiconductors to tune their optoelectronic properties and doping levels.[30−36] These adducts are formed by the partial electron density transfer from a semiconducting conjugated molecule or polymer, usually containing Lewis basic moieties, to external Lewis acids. Most of the molecules in these adducts have an alternating donor−acceptor motif, in which the acceptor unit contains atoms with a nonbonding pair of electrons capable of coordinating with Lewis acids. Boron-based LAs, such as BF_3, BCl_3, and $B(C_6F_5)_3$, have been widely utilized.

Our recent work[37] employed electronic structure calculations to confirm the hypothesis that the changes in optical properties of parent conjugated molecules are tied to electron transfer from these molecules to Lewis acids.[30] Generally, in chemistry, this electron transfer (hereinafter called charge transfer to be consistent with previous studies) is a crucial quantum mechanical quantity. It relates not only the binding strength of a Lewis acid and a Lewis base but also the nature of the bond, for example, whether it is formed mainly by electrostatic or covalent interactions.[38] Although charge transfer in LA−LB bonds is intuitively understood *via* concepts from organic chemistry, it is too microscopically intricate to confirm experimentally.[38,39] We showed that the calculated amount of charge transfer correlated with the degree of red shift in optical absorption of the adducts for a given set of molecules.[37] In this paper, we broaden the screening and predicting power by using machine learning to predict the charge transfer and other optoelectronic properties of those adducts from molecular descriptors that can be obtained readily and inexpensively. In addition, we also obtain the relative weight of each molecular descriptor, reflecting its impact on the properties of adducts and permitting insight into the chemistry and physics associated with the molecular design.

METHODOLOGY

We designed 1016 adducts from 90 Lewis bases (LBs) and 12 Lewis acids (LAs). A majority of these LBs are acceptor moieties commonly used in high-performing donor−acceptor-based organic semiconductors,[36,40−44] while others are typical LBs in chemistry, such as NH_3, $(CH_3)_2NH$, and aniline. Most LAs are common Lewis acids and have been experimentally validated to bind with conjugated Lewis bases and cause changes in optoelectrical properties.[30,35,36,38] Some representative LBs and LAs are presented in Figure 1, and all chemical structures are given in Supporting Information (SI) Figures S1 and S2. The adducts were formed by binding one LB to one LA. Charge transfer for an individual adduct was calculated using a two-step approach successfully implemented in our previous study.[37] First, the nuclear coordinates of most adducts were optimized in DFT using the APFD[45] exchange−correlation functional with 6-311G(d,p) basis set. The adducts with LB of BI_3 were calculated with the LANL2DZ basis set due to heavy iodine atoms. The aforementioned functional was validated by comparing it to HF and two other DFT-based functionals (*i.e.*, B3LYP and CAM-B3LYP with GD3BJ dispersion). Figure S3 demonstrates the comparison conducted with adducts of NH_3 (*i.e.*, the representative for N-sp^3), pyridine (*i.e.*, the representative for N-sp^2), and acetonitrile (*i.e.*, the representative for N-sp) both in vacuum and in dichlorobenzene (DCB) using the polarizable continuum model (PCM). The degree of charge transfer is indicated to be fairly insensitive to the choice of DFT functional and impacted fairly uniformly in the presence of an implicit solvent across a representative subset of our Lewis acid−base pairs. Besides, it is noticeable that the charge transfer calculated by

https://doi.org/10.1021/acsomega.3c02822
ACS Omega 2023, 8, 19119−19127

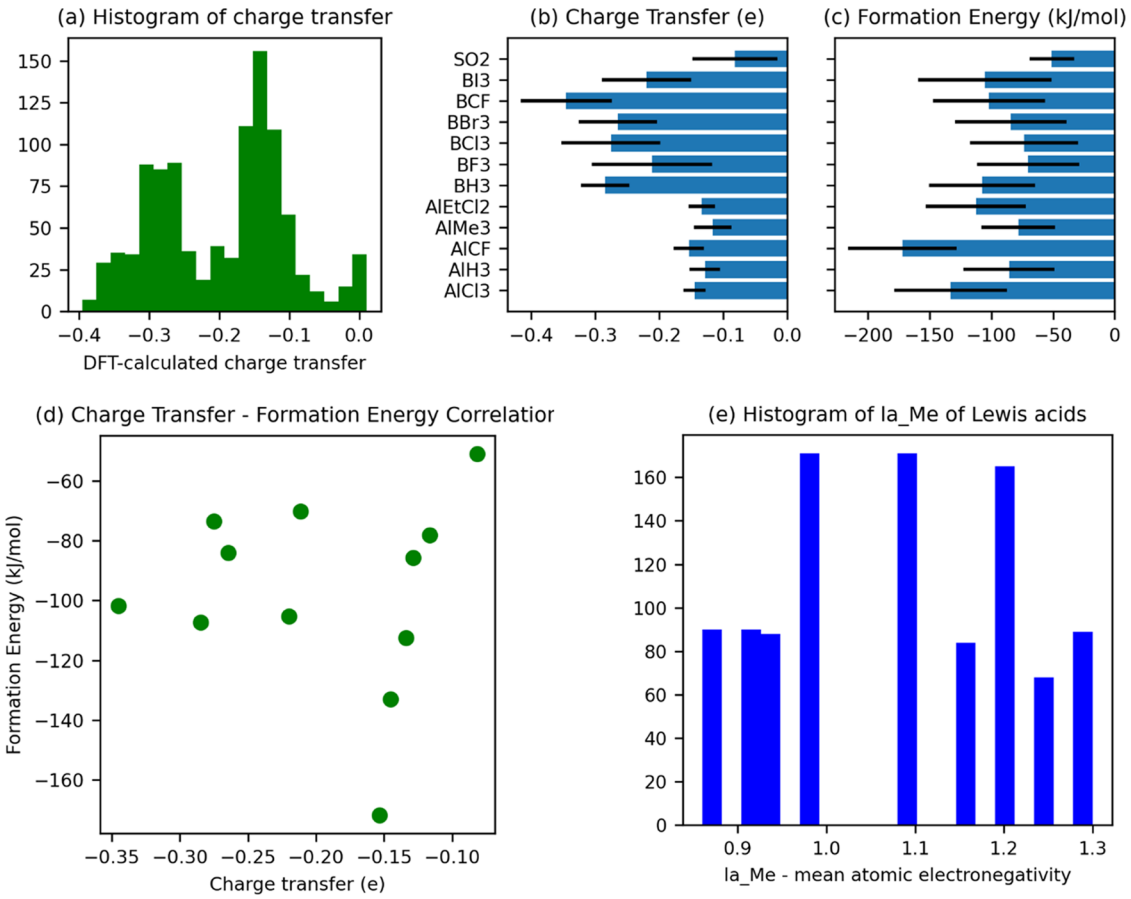


Figure 2. (a) Histogram of charge transfer of 1016 adducts; (b) the average charge transfer and (c) the average formation energy of adducts corresponding to each LA; (d) the correlation between those two properties; and (e) the histogram of one representative descriptor—the mean atomic Sanderson electronegativities scaled over carbon—of adducts.

Hartree−Fock, which includes the full exchange with no self-interaction and no correlation potential, is in close agreement with three different levels of functional choice. Due to the inclusion of an empirical dispersion model in APFD as well as its use and agreement with experimental trends seen in our prior study,[37] this relatively recent functional was employed here for geometry optimization. After optimization, the nonbonding adducts were eliminated from the dataset. Then, charge transfer was calculated using atomic partial charges from NBO population analysis.[46,47]

Molecular descriptors were calculated using the Dragon package.[48] Four groups of descriptors from Dragon were selected based on the higher level of insights by which they can inform the molecular design. They are constitutional descriptors (molecular composition information such as molecular weight—MW or mean atomic Sanderson electronegativity scaled to C—Me), atom-centered fragments (Ghose−Crippen descriptors defined for hydrogen atoms, carbon atoms, and heteroatoms such as the number of $=CH_2$ fragment—C-015), functional group count descriptors (count descriptors of various functional groups such as the number of nonaromatic conjugated $C(sp^2)$—nCconj or number of imides—$nN(CO)_2$), and molecular properties (such as Moriguchi octanol−water partition coeff. (logP)—MLOGP). Molecular descriptors were independently calculated for the LB set and the LA set and then combined to create 141 descriptors (Table S1) to build ML models and predict the aforementioned charge transfer and other properties of the adducts.

The ML models in this study/research were chosen based on their versatility and applicability in chemistry and materials science. They belong to linear-based models (linear regression (LR) and ridge linear regression (RIDGE)), support vector machine regression (SVR), k-nearest neighbor regression (KNN), artificial neural network (ANN), and decision-tree-ensemble-based models (random forest (RF) and gradient boosting (GB)). 20% of the dataset was selected as the test set using the stratified shuffle split function of the Scikit-learn Python module, which randomly selects the test set while keeping the histogram of both the training set and test set similar to the overall dataset (Figure S5).[49,50] This splitting approach, often called stratified sampling in statistics, is commonly used in machine learning to avoid significant sampling bias toward certain groups of the predicted values.[50] Prior to training models, grid searches were also performed to optimize the model hyperparameters. The models are validated by two methods. First, they are used to predict the test set, and then, the Pearson correlation coefficients (r) and root mean squared error (RMSE) between ML-predicted results and DFT-calculated ones can be calculated.[50] The ML models were also validated by cross-validation algorithm with stratified-shuffle-split as the splitter resulting in 30 data points of r and RMSE for testing sets.

https://doi.org/10.1021/acsomega.3c02822
ACS Omega 2023, 8, 19119−19127

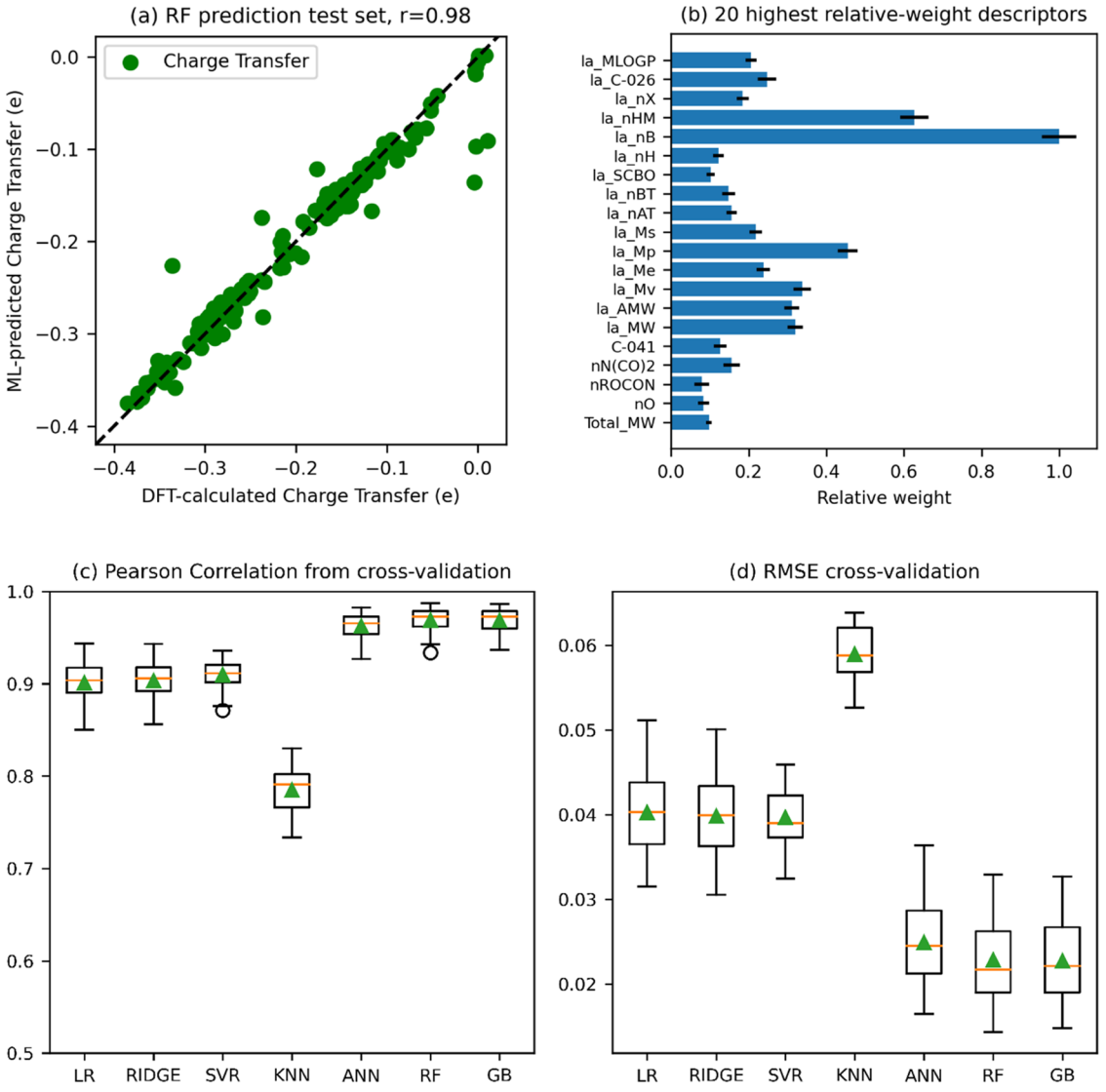


Figure 3. (a) RF-predicted charge transfers *versus* DFT-calculated charge transfers for 204 adducts in the test set. The diagonal line is simply a plot of $y = x$ as a helpful visual guide; (b) relative weights of 20 descriptors with the highest relative weights from the random forest model; (c) Pearson correlation coefficients (r) and (d) root mean squared error (RMSE) from the cross-validation algorithm, where the dataset was split into training and testing sets with a ratio of 8:2 30 times, resulting in 30 data points of r and RMSE for each model. In boxplots (c, d), the line in the box represents the median, the green triangle represents the arithmetic mean, the box covers from the first quartile to the third quartile of the data, and the whisker extends from the lowest data point within the distance of 1.5 times the interquartile range below the first quartile to the highest data point within the same distance above the third quartile, and the circle represents the outliers.

RESULTS AND DISCUSSION

The data analysis before employing ML models is presented in Figures 2 and S4. The DFT-calculated charge transfer distribution of over 1000 adducts is shown in a histogram in Figure 2a, with two apparent peaks at around −0.15 and −0.30 electrons. This number indicates the amount of electron transfer (*i.e.*, loss, hence the negative sign) from Lewis bases to Lewis acids upon the formation of adducts. Figure 2b reveals that the average absolute value of charge transfer in adducts with boron-based LAs (∼0.3) is noticeably higher than aluminum-based LAs (∼0.15), corresponding to the two aforementioned peaks in the histogram. The small peak near 0 in the histogram in Figure 2a is the result of charge transfer for SO_2 adducts. The average charge transfer is −0.20 e with a standard deviation of 0.09 e. This range of charge transfer in LALB adducts is consistent with previous computational and experimental studies.[38,51] In addition to charge transfer, we extracted other physiochemical properties of the adducts from the same DFT calculations in order to get more insights into the coordination bonds and showcased the applicability of machine learning in predicting a wide range of properties. One such important property is the formation energy, which indicates the binding strength of Lewis acids and bases energetically. Figure 2c shows the average formation energy for each LA, which is the difference in total energy of the product (adducts) and reactant (LA plus LB). The variation of formation energy among different LAs is discernably different from that of charge transfer. For example, the adducts of AlCF and $AlCl_3$ have the higher formation energy but lower charge transfer compared to those of BCF and BCl_3, respectively. On the other hand, BH_3 has comparable formation energy but much higher charge transfer compared to those of AlH_3. The uncorrelated behavior of the properties of over 1000 adducts is indicated in Figure 2d, which is consistent with the

https://doi.org/10.1021/acsomega.3c02822
ACS Omega 2023, 8, 19119−19127

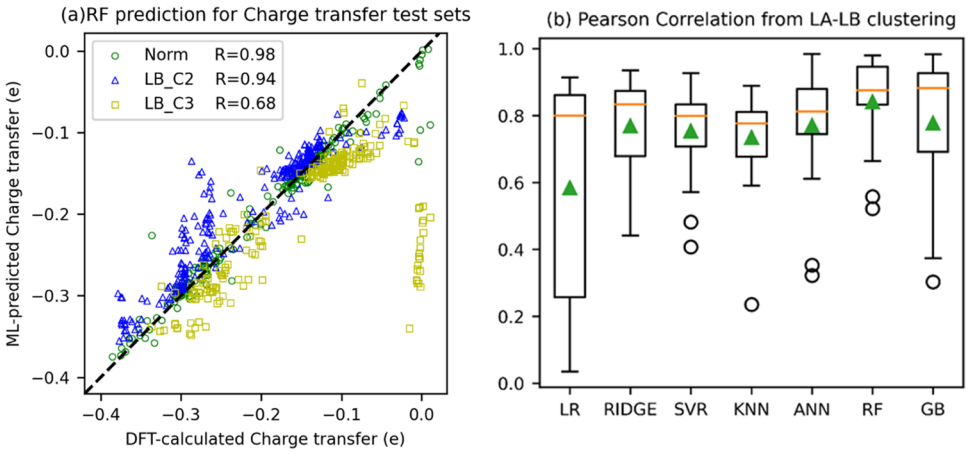


Figure 4. (a) RF-predicted charge transfers *versus* DFT-calculated charge transfers for test sets of stratified shuffle selection (norm) and leave-one-cluster-out (LOCO) predictions of LB cluster 2 (LB_C2) and cluster 3 (LB_C3). (b) Pearson correlation boxplots analogous to Figure 3c,d of seven ML models from all LOCO predictions when clustering based on LA and LB structures.

experimental values in previous studies.[38,52−54] Most descriptors are discretely distributed, with an example of Me of LAs shown in Figure 2e. Compared to the continuous distributions of charge transfer, one can deduce that the charge transfer is not well-correlated with any of the descriptors, as demonstrated in Figure S4.

Figure 3 demonstrates the exceptional performance of the ML models in predicting charge transfer. Figure 3a plots the charge transfer in LALB adducts predicted by the RF model *versus* that calculated by DFT. These are the charge transfers of the 204 adducts in the test set, which is not included in the training set to build the model. Similar plots of ML-predicted *versus* DFT-calculated charge transfers for all other models are presented in Figures S6−S8. Cross-validation across all machine learning models can accurately predict charge transfer with exceptionally high correlation coefficient r (~0.73−0.99, Figure 3c) and small RMSE (~0.06−0.01, Figure 3d), which demonstrates the reliability of ML models over different choices of training and test sets. Among these, ANN, RF, and GB perform the best with r around 0.97 and RMSE around 0.02 e, which is about 10% of the second peak in the charge transfer histogram. The fact that ANN, RF, and GB models are more accurate than LR, RIDGE, and KNN models might be attributed to aforementioned low correlations between charge transfer and the descriptors related to the difference in their distributions (continuous *versus* discrete).

In order to inform the molecular design toward adducts with desirable charge transfer, we evaluate the influence of each descriptor in determining the predicted outcome from ML models, which is the charge transfer in this case. To that end, we extract the feature importance from RF and GB models (with 30-time cross-validation) using available algorithms in Scikit-learn and calculate the relative weight of each descriptor by dividing its feature importance by the highest feature importance. All molecular descriptors with their relative weights are given in Table S1 for both RF and GB models. The relative weights are calculated by dividing the feature importance of each descriptor with the highest value (*i.e.*, the feature importance of la_nB). The 20 descriptors (out of 141) with the highest relative weights from the RF model are plotted in Figure 3b. Interestingly, among those 20 descriptors, the number of high-weight descriptors for LAs and LBs is 15 and 4, respectively (total_MW is the total molecular weight of LAs and LBs). On average, the relative weights of all LA and LB descriptors from the RF model are 0.171 ± 0.212 and 0.018 ± 0.025, respectively, and those from the GB model are 0.042 ± 0.180 and 0.005 ± 0.014, respectively. It implies that, at least for this dataset, the molecular descriptors of LAs are more significant than those of LBs in determining the charge transfer. With 11 LAs out of 12 containing either B or Al to bind with LBs, both GB and RF models properly capture the two highest weight descriptors as the number of boron atoms (la_nB) and the number of heavy atoms—aluminum in this case (la_nHM). Apart from la_nB and la_nHM, other notable LA descriptors are the mean atomic polarizability (la_Mp), mean atomic van der Waals volume (la_Mv), and molecular weight of LA (la_MW). Notable LB descriptors are the number of double bonds of carbon and heteroatom (C-041) and the number of imides ($nN(CO)_2$).

Furthermore, a challenge in studying chemically combinatorial datasets is that machine learning extrapolability cannot be assessed accurately using the (even stratified shuffle) random splitting for obtaining training and testing sets. It is suggested that the so-called leave-one-cluster-out (LOCO) method should be used to provide more insights into the extrapolability of machine learning models.[55] In order to address this issue, the charge transfer dataset was clustered to perform LOCO predictions. First, the dataset is grouped into five different clusters based on LB structures (Table S2). Figure 4a shows the RF-predicted charge transfer values of LOCO predictions with the highest and the lowest Pearson correlations and those of "normal" predictions based on stratified shuffle split (Figure 3a). Similar graphs of the remaining ML models are demonstrated in Figure S9, and the Pearson correlation boxplots of LB-structure-based LOCO predictions of seven ML models are presented in Figure S11a. The results indicate that the LOCO predictions present an average of 16.21% lower Pearson correlation values compared to the stratified shuffle selection. This reduction, which is smaller than those reported in prior studies,[55,56] confirms the robustness of our models in predicting charge transfer for Lewis bases that have structure motifs not presented in the training set. It is also noticeable that all ML models demonstrate poor performance in predicting charge transfer very close to zero.

https://doi.org/10.1021/acsomega.3c02822
ACS Omega 2023, 8, 19119−19127

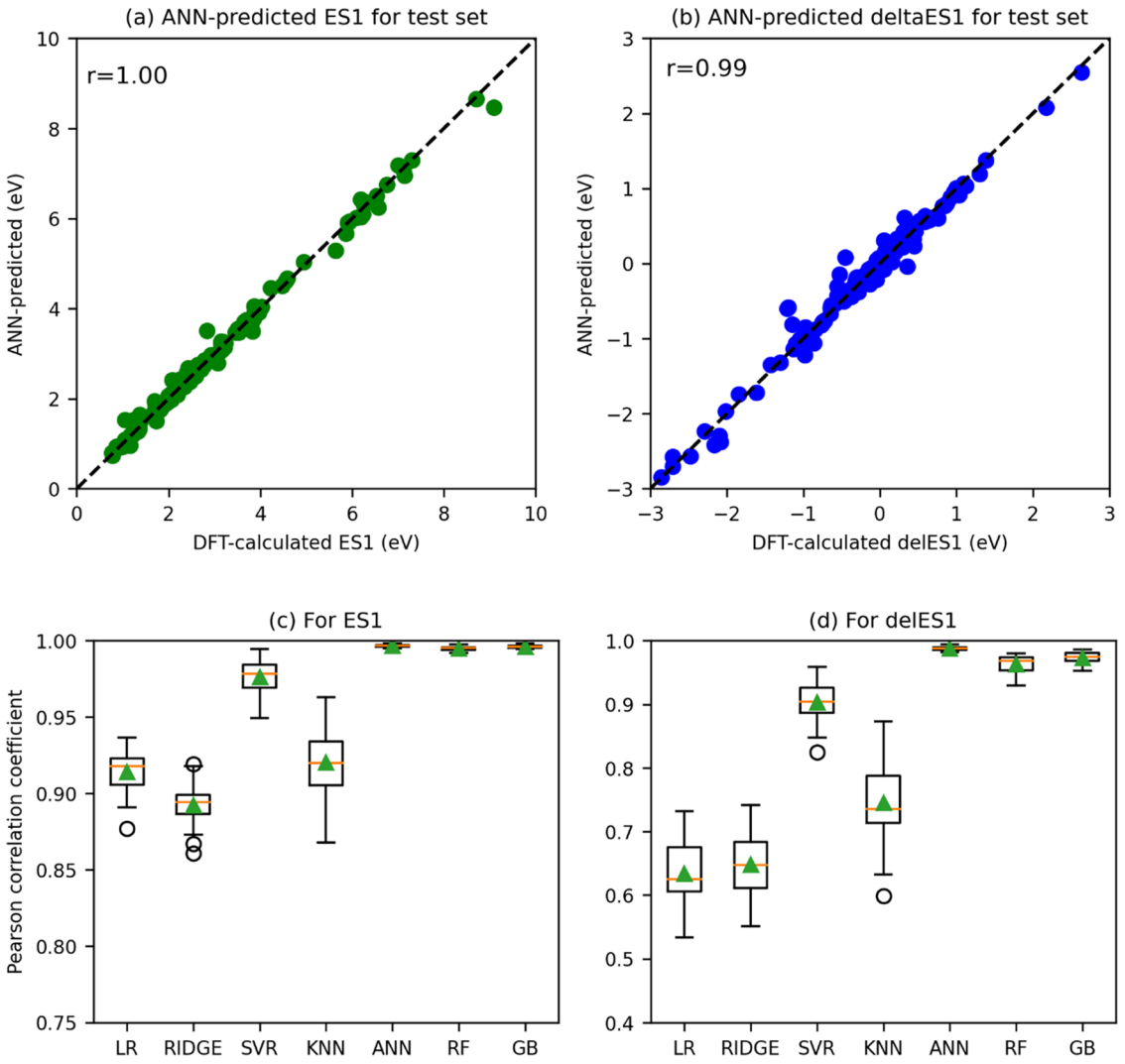


Figure 5. Performance of artificial neural network ML models in predicting the TD-DFT-calculated first excited states of LALB adducts (ES1) (a) and the shift of the first excited states from Lewis bases to adducts (equals to adducts minus LBs—delES1) (b). Pearson correlation coefficients of all models for ES1 (c) and delES1 (d) calculated from the cross-validation algorithm aforementioned.

For LA clustering, the dataset is divided into 12 clusters corresponding to 12 LAs (each cluster is named by the representative LA) to conduct LOCO predictions. The similar graphs and boxplots of LA-structure-based LOCO predictions are illustrated in Figures S10 and S11b, which can be taken as supporting evidence for the lower accuracy in LOCO predictions compared to normal prediction, especially for BCF and SO_2 clusters. For BCF, high DFT-calculated charge transfer adducts cannot be predicted well, which might result from the significantly high number of fluorine atoms in BCF compared to all other LAs in the training set, where all BCF adducts are excluded. In the case of SO_2, the fact that the sulfur atom is not presented in the LOCO training set descriptors (none of the other LAs has sulfur) is hypothesized to result in the inferior LOCO predictions. Compared to all other clusters, the BCF and SO_2 results are statistically determined to be outliers. With the exclusion of the outliers, the average reduction of Pearson coefficients with LA-structure-based LOCO predictions is 13.33%. Finally, the Pearson correlations of both LA-structure-based and LB-structure-based LOCO prediction were combined for each ML model and are demonstrated in Figure 4b. The boxplots show the reliability of all ML models with average Pearson correlation values all above 0.7 except for the LR model. Similar to the normal prediction, RF is among the most stable ML model with the highest performance (*i.e.*, 0.84 ± 0.14 in Pearson correlation) in the LOCO predictions.

Given the high accuracy of ML models in predicting charge transfer, we used them to predict other DFT-calculated properties of the adducts employing the same set of descriptors. One such property is the first excited state (hereinafter called ES1) of the adducts and the other is the shift in the first excited states of the adducts from the first excited states of Lewis bases (hereinafter called delES1). ES1 and delES1 are desirable fundamental properties in designing LALB adducts as light emitters or light absorbers in LEDs, solar cells, and other applications. All excited states were computed by time-dependent DFT (TD-DFT) with the same level of theory as the one used for geometry optimization. The average and standard deviation values of ES1 are 2.927 and 1.509 eV, respectively, and those of delES1 are −0.197 and 0.816 eV, respectively. Histograms of ES1 and delES1 are given in Figure S12. The negative delES1 indicates the red shift in the adducts compared to LBs, which is consistent with the

https://doi.org/10.1021/acsomega.3c02822
ACS Omega 2023, 8, 19119−19127

donor−acceptor molecular orbital hypothesis.[30,37] It should be noted here that the calculated oscillator strength of all ES1 is very low, which can be attributable to the low vertical excitation overlap of the HOMOs and LUMOs. Similar to charge transfer, we can train ML models, notably, ANN, RF, and GB, to predict ES1 and delES1 with exceptional levels of accuracy. This is demonstrated in Figure 5. Especially, for ES1, r varies narrowly around 0.96 for all models. The same descriptors are also shown to predict two other fundamental properties of the adducts—the HOMOs and the formation energy with a high level of accuracy (Figures S13 and S14). As has been noted by others, it is expected that accurate DFT prediction of the HOMO−LUMO gap may be quite sensitive to calculation details, including particulars of exchange−correlation functionals coupled with how solvent models are employed, especially when extrapolating to larger complexes and the solid state.[57] Although we see encouraging trends in our preliminary studies of ES1 and delES1 results here, a comprehensive study of these issues remains for future studies.

CONCLUSIONS

In summary, we demonstrate a powerful yet facile alternative methodology to DFT calculations in approaching quantum chemistry by employing machine learning analysis. With the utilization of readily obtained molecular descriptors, machine learning presents the capability of predicting DFT-calculated charge transfer and other (advanced) physical and chemical properties of the Lewis acid−Lewis base adducts. The prediction of charge transfers, first excited states, red shifts in the first excited states, HOMOs, and formation energy of the adducts show a high level of accuracy for a wide range of machine learning models from linear regression to decision-tree regression, especially noteworthy is the exceptional accuracy for ANN and ensemble models like RF and GB. Even with leave-one-cluster-out testing, our ML models are shown to have a relatively high level of accuracy in predicting the DFT charge transfer of most Lewis adduct clusters, whose structural motifs are absent in the training chemical space. We also analyze the feature importance that influences the prediction of charge transfer for RF and GB models, which might provide insights for molecular design toward specific applications. In a broader context, our approach is promising to accurately and economically screen and predict a variety of fundamental properties of molecules that influence the performance of functional devices, such as solar cells and LEDs.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c02822.

LALB_codes (ZIP)

Chemical structures of 90 Lewis bases in this study; Hartree−Fock-calculated charge transfers *versus* charge transfer calculated by DFT functionals (B3LYP, CAM-B3LYP with GD3BJ dispersion, and APFD); relative feature importance (Relative_FI) and its one standard deviation (Error_FI) of all molecular descriptors used in training random forest (RF) and Gradient Boosting (GB) models for predicting charge transfer; plots of DFT-calculated charge transfers and two representative descriptors; LB clustering rule based on the hybridization and the functional groups; Pearson correlation boxplots; the codes used in this study can also be accessed freely at a GitHub repository after the paper is published (PDF)

AUTHOR INFORMATION

Corresponding Authors

Emily A. Jarvis − *Loyola Marymount University, Los Angeles, California 90045, United States*; Email: emily.jarvis@lmu.edu

Hung Phan − *Fulbright University Vietnam, Ho Chi Minh 72908, Vietnam; Soka University of America, Aliso Viejo, California 92656, United States*; orcid.org/0000-0001-5774-0390; Email: hung.phan@fulbright.edu.vn

Authors

Hieu Huynh − *Fulbright University Vietnam, Ho Chi Minh 72908, Vietnam*

Thomas J. Kelly − *Loyola Marymount University, Los Angeles, California 90045, United States*

Linh Vu − *Fulbright University Vietnam, Ho Chi Minh 72908, Vietnam*

Tung Hoang − *Independent Researcher, Palo Alto, California 94303, Unites States*

Phuc An Nguyen − *Fulbright University Vietnam, Ho Chi Minh 72908, Vietnam*

Tu C. Le − *School of Engineering, STEM College, RMIT University, Melbourne, Victoria 3000, Australia*; orcid.org/0000-0003-3552-8211

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c02822

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

All the graphs were conducted by Matplotlib.[58] H.P. acknowledges the support of Fulbright University Vietnam *via* Professional Development Fund and the funding for high-performance computing server and LMU Seaver College of Science and Engineering computing resources.

REFERENCES

(1) Schmidt, J.; Marques, M. R. G.; Botti, S.; Marques, M. A. L. Recent Advances and Applications of Machine Learning in Solid-State Materials Science. *Npj Comput. Mater.* **2019**, *5*, No. 83.

(2) Chen, C.; Zuo, Y.; Ye, W.; Li, X.; Deng, Z.; Ong, S. P. A Critical Review of Machine Learning of Energy Materials. *Adv. Energy Mater.* **2020**, *10*, No. 1903242.

(3) Cichos, F.; Gustavsson, K.; Mehlig, B.; Volpe, G. Machine Learning for Active Matter. *Nat. Mach. Intell.* **2020**, *2*, 94−103.

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

(5) Chen, A.; Zhang, X.; Zhou, Z. Machine Learning: Accelerating Materials Development for Energy Storage and Conversion. *InfoMat* **2020**, *2*, 553−576.

(6) Saeki, A.; Kranthiraja, K. A High Throughput Molecular Screening for Organic Electronics via Machine Learning: Present Status and Perspective. *Jpn. J. Appl. Phys.* **2019**, *59*, No. SD0801.

(7) 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*, No. 120.

(8) Lee, M.-H. Insights from Machine Learning Techniques for Predicting the Efficiency of Fullerene Derivatives-Based Ternary Organic Solar Cells at Ternary Blend Design. *Adv. Energy Mater.* **2019**, *9*, No. 1900891.
(9) Sahu, H.; Rao, W.; Troisi, A.; Ma, H. Toward Predicting Efficiency of Organic Solar Cells via Machine Learning and Improved Descriptors. *Adv. Energy Mater.* **2018**, *8*, No. 1801032.
(10) Sahu, H.; Ma, H. Unraveling Correlations between Molecular Properties and Device Parameters of Organic Solar Cells Using Machine Learning. *J. Phys. Chem. Lett.* **2019**, *10*, 7277−7284.
(11) Majeed, N.; Saladina, M.; Krompiec, M.; Greedy, S.; Deibel, C.; MacKenzie, R. C. I. Using Deep Machine Learning to Understand the Physical Performance Bottlenecks in Novel Thin-Film Solar Cells. *Adv. Funct. Mater.* **2020**, *30*, No. 1907259.
(12) Nagasawa, S.; Al-Naamani, E.; Saeki, A. Computer-Aided Screening of Conjugated Polymers for Organic Solar Cell: Classification by Random Forest. *J. Phys. Chem. Lett.* **2018**, *9*, 2639−2646.
(13) Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D.-G.; Wu, T.; Markopoulos, G.; Jeon, S.; Kang, H.; Miyazaki, H.; Numata, M.; Kim, S.; Huang, W.; Hong, S. I.; Baldo, M.; Adams, R. P.; Aspuru-Guzik, A. Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach. *Nat. Mater.* **2016**, *15*, 1120−1127.
(14) Janai, M. A. B.; Woon, K. L.; Chan, C. S. Design of Efficient Blue Phosphorescent Bottom Emitting Light Emitting Diodes by Machine Learning Approach. *Org. Electron.* **2018**, *63*, 257−266.
(15) Ong, S. P. Accelerating Materials Science with High-Throughput Computations and Machine Learning. *Comput. Mater. Sci.* **2019**, *161*, 143−150.
(16) Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. Quantitative Structure−Property Relationship Modeling of Diverse Materials Properties. *Chem. Rev.* **2012**, *112*, 2889−2919.
(17) Le, T. C.; Ballard, M.; Casey, P.; Liu, M. S.; Winkler, D. A. Illuminating Flash Point: Comprehensive Prediction Models. *Mol. Inf.* **2015**, *34*, 18−27.
(18) Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics*; John Wiley & Sons, Ltd., 2009.
(19) Lee, M.-H. Machine Learning for Understanding the Relationship between the Charge Transport Mobility and Electronic Energy Levels for N-Type Organic Field-Effect Transistors. *Adv. Electron. Mater.* **2019**, *5*, No. 1900573.
(20) Zhao, Z.-W.; del Cueto, M.; Geng, Y.; Troisi, A. Effect of Increasing the Descriptor Set on Machine Learning Prediction of Small Molecule-Based Organic Solar Cells. *Chem. Mater.* **2020**, *32*, 7777−7787.
(21) Li, W.; Liu, J.; Li, L.; Hu, L.; Su, Z.-M.; Chen, G. Machine Learning Corrections for DFT Noncovalent Interactions. In *Computational Materials, Chemistry, and Biochemistry: From Bold Initiatives to the Last Mile: In Honor of William A. Goddard's Contributions to Science and Engineering*; Shankar, S.; Muller, R.; Dunning, T.; Chen, G. H., Eds.; Springer Series in Materials Science; Springer International Publishing: Cham, 2021; pp 183−212.
(22) Migliaro, I.; Cundari, T. R. Density Functional Study of Methane Activation by Frustrated Lewis Pairs with Group 13 Trihalides and Group 15 Pentahalides and a Machine Learning Analysis of Their Barrier Heights. *J. Chem. Inf. Model.* 2020. DOI: 10.1021/acs.jcim.0c00862.
(23) Jeong, W.; Stoneburner, S. J.; King, D.; Li, R.; Walker, A.; Lindh, R.; Gagliardi, L. Automation of Active Space Selection for Multireference Methods via Machine Learning on Chemical Bond Dissociation. *J. Chem. Theory Comput.* **2020**, *16*, 2389−2399.
(24) Wang, C.-I.; Braza, M. K. E.; Claudio, G. C.; Nellas, R. B.; Hsu, C.-P. Machine Learning for Predicting Electron Transfer Coupling. *J. Phys. Chem. A* **2019**, *123*, 7792−7802.
(25) Sifain, A. E.; Lubbers, N.; Nebgen, B. T.; Smith, J. S.; Lokhov, A. Y.; Isayev, O.; Roitberg, A. E.; Barros, K.; Tretiak, S. Discovering a Transferable Charge Assignment Model Using Machine Learning. *J. Phys. Chem. Lett.* **2018**, *9*, 4495−4501.
(26) Pereira, F.; Xiao, K.; Latino, D. A. R. S.; Wu, C.; Zhang, Q.; Aires-de-Sousa, J. Machine Learning Methods to Predict Density Functional Theory B3LYP Energies of HOMO and LUMO Orbitals. *J. Chem. Inf. Model.* **2017**, *57*, 11−21.
(27) Musil, F.; De, S.; Yang, J.; E Campbell, J.; M Day, G.; Ceriotti, M. Machine Learning for the Structure−Energy−Property Landscapes of Molecular Crystals. *Chem. Sci.* **2018**, *9*, 1289−1300.
(28) Atahan-Evrenk, S.; Atalay, F. B. Prediction of Intramolecular Reorganization Energy Using Machine Learning. *J. Phys. Chem. A* **2019**, *123*, 7855−7863.
(29) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087−2096.
(30) Welch, G. C.; Coffin, R.; Peet, J.; Bazan, G. C. Band Gap Control in Conjugated Oligomers via Lewis Acids. *J. Am. Chem. Soc.* **2009**, *131*, 10802−10803.
(31) Zalar, P.; Kuik, M.; Henson, Z. B.; Woellner, C.; Zhang, Y.; Sharenko, A.; Bazan, G. C.; Nguyen, T.-Q. Increased Mobility Induced by Addition of a Lewis Acid to a Lewis Basic Conjugated Polymer. *Adv. Mater.* **2014**, *26*, 724−727.
(32) Panidi, J.; Paterson, A. F.; Khim, D.; Fei, Z.; Han, Y.; Tsetseris, L.; Vourlias, G.; Patsalas, P. A.; Heeney, M.; Anthopoulos, T. D. Remarkable Enhancement of the Hole Mobility in Several Organic Small-Molecules, Polymers, and Small-Molecule:Polymer Blend Transistors by Simple Admixing of the Lewis Acid p-Dopant B(C6F5)3. *Adv. Sci.* **2018**, *5*, No. 1700290.
(33) Yan, H.; Chen, J.; Zhou, K.; Tang, Y.; Meng, X.; Xu, X.; Ma, W. Lewis Acid Doping Induced Synergistic Effects on Electronic and Morphological Structure for Donor and Acceptor in Polymer Solar Cells. *Adv. Energy Mater.* **2018**, *8*, No. 1703672.
(34) Li, Y.; Meng, H.; Li, Y.; Pang, B.; Luo, G.; Huang, J. Adjusting the Energy Levels and Bandgaps of Conjugated Polymers via Lewis Acid−Base Reactions. *New J. Chem.* **2018**, *42*, 18961−18968.
(35) Yurash, B.; Cao, D. X.; Brus, V. V.; Leifert, D.; Wang, M.; Dixon, A.; Seifrid, M.; Mansour, A. E.; Lungwitz, D.; Liu, T.; Santiago, P. J.; Graham, K. R.; Koch, N.; Bazan, G. C.; Nguyen, T.-Q. Towards Understanding the Doping Mechanism of Organic Semiconductors by Lewis Acids. *Nat. Mater.* **2019**, *18*, 1327−1334.
(36) Bridges, C. R.; Baumgartner, T. Lewis Acids and Bases as Molecular Dopants for Organic Semiconductors. *J. Phys. Org. Chem.* **2020**, *33*, No. e4077.
(37) Phan, H.; Kelly, T. J.; Zhugayevych, A.; Bazan, G. C.; Nguyen, T.-Q.; Jarvis, E. A.; Tretiak, S. Tuning Optical Properties of Conjugated Molecules by Lewis Acids: Insights from Electronic Structure Modeling. *J. Phys. Chem. Lett.* **2019**, *10*, 4632−4638.
(38) Jonas, V.; Frenking, G.; Reetz, M. T. Comparative Theoretical Study of Lewis Acid-Base Complexes of BH3, BF3, BCl3, AlCl3, and SO2. *J. Am. Chem. Soc.* **1994**, *116*, 8741−8753.
(39) Mebs, S.; Grabowsky, S.; Förster, D.; Kickbusch, R.; Hartl, M.; Daemen, L. L.; Morgenroth, W.; Luger, P.; Paulus, B.; Lentz, D. Charge Transfer via the Dative N−B Bond and Dihydrogen Contacts. Experimental and Theoretical Electron Density Studies of Small Lewis Acid−Base Adducts. *J. Phys. Chem. A* **2010**, *114*, 10185−10196.
(40) Collins, S. D.; Ran, N. A.; Heiber, M. C.; Nguyen, T.-Q. Small Is Powerful: Recent Progress in Solution-Processed Small Molecule Solar Cells. *Adv. Energy Mater.* **2017**, *7*, No. 1602242.
(41) Zhang, G.; Zhao, J.; Chow, P. C. Y.; Jiang, K.; Zhang, J.; Zhu, Z.; Zhang, J.; Huang, F.; Yan, H. Nonfullerene Acceptor Molecules for Bulk Heterojunction Organic Solar Cells. *Chem. Rev.* **2018**, *118*, 3447−3507.
(42) Wang, Y.; Lee, J.; Hou, X.; Labanti, C.; Yan, J.; Mazzolini, E.; Parhar, A.; Nelson, J.; Kim, J.-S.; Li, Z. Recent Progress and Challenges toward Highly Stable Nonfullerene Acceptor-Based Organic Solar Cells. *Adv. Energy Mater.* **2021**, *11*, No. 2003002.

https://doi.org/10.1021/acsomega.3c02822
ACS Omega 2023, 8, 19119−19127

(43) Karki, A.; Gillett, A. J.; Friend, R. H.; Nguyen, T.-Q. The Path to 20% Power Conversion Efficiencies in Nonfullerene Acceptor Organic Solar Cells. *Adv. Energy Mater.* **2021**, *11*, No. 2003441.
(44) Lim, D.-H.; Ha, J.-W.; Choi, H.; Yoon, S. C.; Lee, B. R.; Ko, S.-J. Recent Progress of Ultra-Narrow-Bandgap Polymer Donors for NIR-Absorbing Organic Solar Cells. *Nanoscale Adv.* **2021**, *3*, 4306−4320.
(45) Austin, A.; Petersson, G. A.; Frisch, M. J.; Dobek, F. J.; Scalmani, G.; Throssell, K. A Density Functional with Spherical Atom Dispersion Terms. *J. Chem. Theory Comput.* **2012**, *8*, 4989−5007.
(46) Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theor. Chim. Acta* **1977**, *44*, 129−138.
(47) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, *8*, 527−541.
(48) *Talete* SRL, TALETE SRL Dragon *for Windows 2.2, 2007*. http://www.talete.mi.it/products/dragon_description.htm (accessed June 29, 2021).
(49) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(50) *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition [Book]*. https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/ (accessed March 24, 2020).
(51) Anderson, W. P. A Lewis Acid-Base Computational Chemistry Exercise for Advanced Inorganic Chemistry. *J. Chem. Educ.* **2000**, *77*, 209.
(52) McCoy, R. E.; Bauer, S. H. Energetics of the Boranes. I. The Heats of Reaction of Diborane with the Methylamines, and of Tetramethyldiborane with Trimethylamine; the Dissociation Energy of Diborane1. *J. Am. Chem. Soc.* **1956**, *78*, 2061−2065.
(53) Anderson, G. A.; Forgaard, F.; Haaland, A.; et al. On the Molecular Structure of the Complex Trimethylaluminium-Trimethylamine, (CH3)3AlN(CH3)3. *Acta Chem. Scand.* **1972**, *26*, 1947−1954.
(54) LaBarge, M. S.; Matos, J.; Hillig, K. W.; Kuczkowski, R. L. Microwave Spectrum and Structure of the Trimethylamine-Sulfur Dioxide Charge-Transfer Complex. *J. Am. Chem. Soc.* **1987**, *109*, 7222−7223.
(55) Zahrt, A. F.; Henle, J. J.; Denmark, S. E. Cautionary Guidelines for Machine Learning Studies with Combinatorial Datasets. *ACS Comb. Sci.* **2020**, *22*, 586−591.
(56) Meredig, B.; Antono, E.; Church, C.; Hutchinson, M.; Ling, J.; Paradiso, S.; Blaiszik, B.; Foster, I.; Gibbons, B.; Hattrick-Simpers, J.; Mehta, A.; Ward, L. Can Machine Learning Identify the next High-Temperature Superconductor? Examining Extrapolation Performance for Materials Discovery. *Mol. Syst. Des. Eng.* **2018**, *3*, 819−825.
(57) Zheng, Z.; Brédas, J. L.; Coropceanu, V. Description of the Charge Transfer States at the Pentacene/C60 Interface: Combining Range-Separated Hybrid Functionals with the Polarizable Continuum Model. *J. Phys. Chem. Lett.* **2016**, *7*, 2616−2621.
(58) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90−95.

https://doi.org/10.1021/acsomega.3c02822
ACS Omega 2023, 8, 19119−19127