Comparative Analyses of Data Driven Machine Learning Models for TADF Emitters

Kalishankar Bhattacharyya*

Abstract:

Thermally activated delayed fluorescence (TADF) chromophores have attracted significant attention because they can harvest singlets and triplets in organic light-emitting diodes (OLEDs), resulting in high external quantum efficiency (EQE). This work aims to use a data-driven machine-learning model to predict the relationship between EQE and essential features of TADF-based OLEDs. The study uses a set of experimental data and applies various machine-learning models to analyze the relationship between EQE and the features of the device. The Random Forest model is found to have the highest accuracy for predicting EQE for non-doped TADF emitters. The importance of feature selection and its correlation with EQE is also analyzed, providing insight into how to select the best machine-learning model for rapid material screening and device optimization for non-doped TADF materials.

Introduction:

In recent years, machine-learning (ML) has appeared as an influential complement to computational calculations because of its ability to produce first-principle results at significantly lower cost promoting a screening of extensive areas of chemical space.[1-5] It provides a relationship between the input variable to the output performance resulting in a prediction performance vis-a-vis mapping relationships. ML models can be prepared on experimental data, which implies they can learn from the existing data. In this context, several ML models, e.g., supervised and

unsupervised algorithm, has documented a successful application in many areas of chemistry, i.e., homogeneous catalysis, excited state, heterogeneous catalysis, electrocatalysis, and organic electronics.[6-13] Based on the experimental and ab initio data set, several research works have reported excellent ML models on the fast prediction of the band gap, evaluation of singlet-triplet energy gap, and fluorescence and phosphorescence rates in reference to the excited state properties.[14-18] The examples are persistent and continue to be applications in various fields.[19-22]

Thermally activated delayed fluorescence (TADF) materials have been regarded as the most promising approach to harnessing 100% efficiency theoretically from the singlet and triplet exciton in organic electronics.[23-29] In the TADF process, chromophores are first excited to the singlet state manifold and transferred to the triplet state through inter-system (ISC) crossing pathways.[30] Consequently, a reverse inter-system crossing mechanism (rISC) happens between the triplet and singlet state, followed by delayed fluorescence. In recent times, significant experimental efforts have been made to capitalize on higher EQE of over 30.7% by in-silico design of molecular configurations and device fabrications.[31, 32] Nevertheless, the higher performance of TADF-based organic-light emitting devices and an in-depth understanding of material properties based on the features representing the generalized model still need to be improved. Therefore, a design principle with good descriptors is a prerequisite to enhance the modeling of TADF materials for efficiency enhancement.[33, 34] In this context, researchers have started to apply the ML model to quantify the electronic properties to explore the new TADF emitters with the variation of structural modulations.[35-37] Guided by the ML, significant progress has been achieved for device optimization with materials selection in applying organic solar cells.

Aspuru-Guzik and co-workers have developed an ML model for accurately predicting the rISC rate over the 1.6 million chromophores data set.[38] Based on the virtual screening and *first-principle*

calculations, they have been able to predict efficient TADF materials with maximum EQE of ~22%.[38] Successful applications of ML approaches for the prediction of physical quantities of OLED materials directly promoted the evaluation of EQE values from the electronic properties of the host/guest composite.[39] Chan et al. have shown a Random Forest model that could extract correlations between various features over the 304 blue phosphorescent OLED materials.[40] Lee et al. have shown occupied and unoccupied energy level positions, singlet-triplet energy level splitting, and measured PLQY of guests can make an efficient prediction of EQE for the host/guest OLED design.[41] Based on these features, the Random Forest model predicts average correlation coefficients of r^2 = 0.85 and 0.74 for the training set and testing set, respectively.[33, 41] With a slightly different approach to constructing the molecular structure and features, Troisi and co-workers have reported several TADF emitters from the high-throughput virtual screening.[42, 43] Recently, it is reported that an ensemble-based learning approach could better predict EQE of guest chromophores used in the device.[44] A complex relationship arises between the several features leading to the deviation of importance ranking among them. Nevertheless, quantifying the critical factors between the features and the EQE is challenging and hardly recognized. One of the possible reasons is that the features considered in the analyses might alter EQE independently.

In this work, experimentally available feature properties based on the 200 data sets are assembled for ML algorithms to screen out the crucial factors of controlling TADF properties.[45] Based on the dataset, which is solely from the experimental work, ML model reveals the critical feature, i.e., photoluminescence quantum yield (PLQY) along with power efficiency (PE), current efficiency(CE), and emission wavelength, that yields particular attention for device fabrication. A systematic comparison between the several ML methods, i.e., Random Forest Regression (RF), Support Vector Machine (SVM), K- Nearest Neighbors (KNN), Gradient Boosting (GB), Extreme

Gradient Boosting (XGB), Light Gradient Boosting Machine (LGB), the Random Forest model and Gradient Boosting model demonstrates more satisfactory accuracy for the forecast of EQE for the non-doped TADF emitters. It is important to note that this work can not predict new trends of EQE of the TADF materials in line with reported work in this field, but improving the dataset and selecting the right features to lead more accurate predictions. This work highlights the potential of using machine learning to optimize the EQE of TADF materials.

Results and Discussion

Data visualization plays a crucial role in any ML model. This is why it is essential to examine the distribution of the data sets carefully. To gain a better understanding of the data, parity plots of the chosen features are presented in Figure 1. These plots provide a visual representation of the data and allow for easy identification of patterns and outliers. The current data set, which is based on available experimental results, contains a total of 200 data points. This information is provided in the supporting information file, which is available for further reference. The primary objective of the current ML model is to predict the EQE of the materials. To achieve this goal, the model uses seven key features as input variables. These features are carefully selected based on their relevance and importance to the prediction of EQE. In general, TADF emitters that have not been doped typically have an emission peak within a range of 420 nm to 650 nm, which corresponds to orange to red emissions. This is an interesting feature to note as it is indicative of the potential of these emitters to be used in device structures. Furthermore, the data analysis of non-doped TADF emitters has revealed that their EQE and PLQY typically falls between 8% to 20% and 40% to 80% respectively. This range is considered promising, as it suggests that these emitters can be used effectively in device structures. Overall, these data indicate that TADF emitters have the potential to be highly effective in a variety of device structures, and further analysis is warranted.

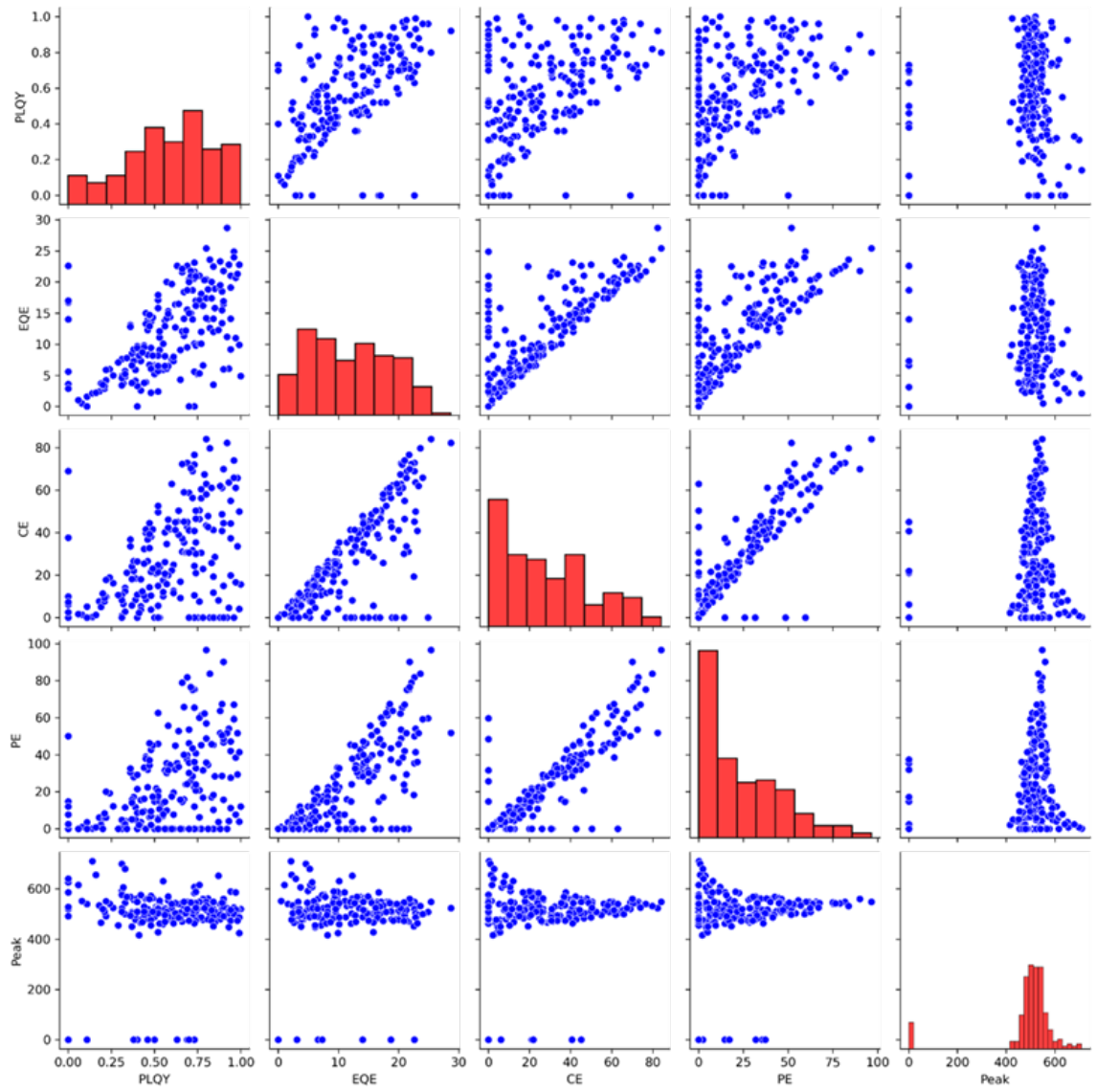

Figure 1. Initial inspection of the data set using parity plot.

Figure 2 presents a heat map of the Pearson correlation between various features from the studied data set, as determined by a simple statistical analysis. The heat map clearly shows that CE, PE, and PLQY have a strong correlation with EQE. On the other hand, turn-on-voltage (V_{on}), which is an important descriptor related to the stability of the device, appears to have a low correlation with EQE. Additionally, some of the features such as Peak, Tp(ηs), Td(μs) have low variance, indicating that the value of these features remains unchanged across the studied data points.

Manually determining the most important feature that correlates with EQE can be a difficult task. However, by using a data-driven ML algorithm, it is possible to identify the intrinsic features that govern EQE as prerequisites. In other words, the data-driven ML algorithm can help to extract the most informative features from the dataset, which is useful for predicting EQE.

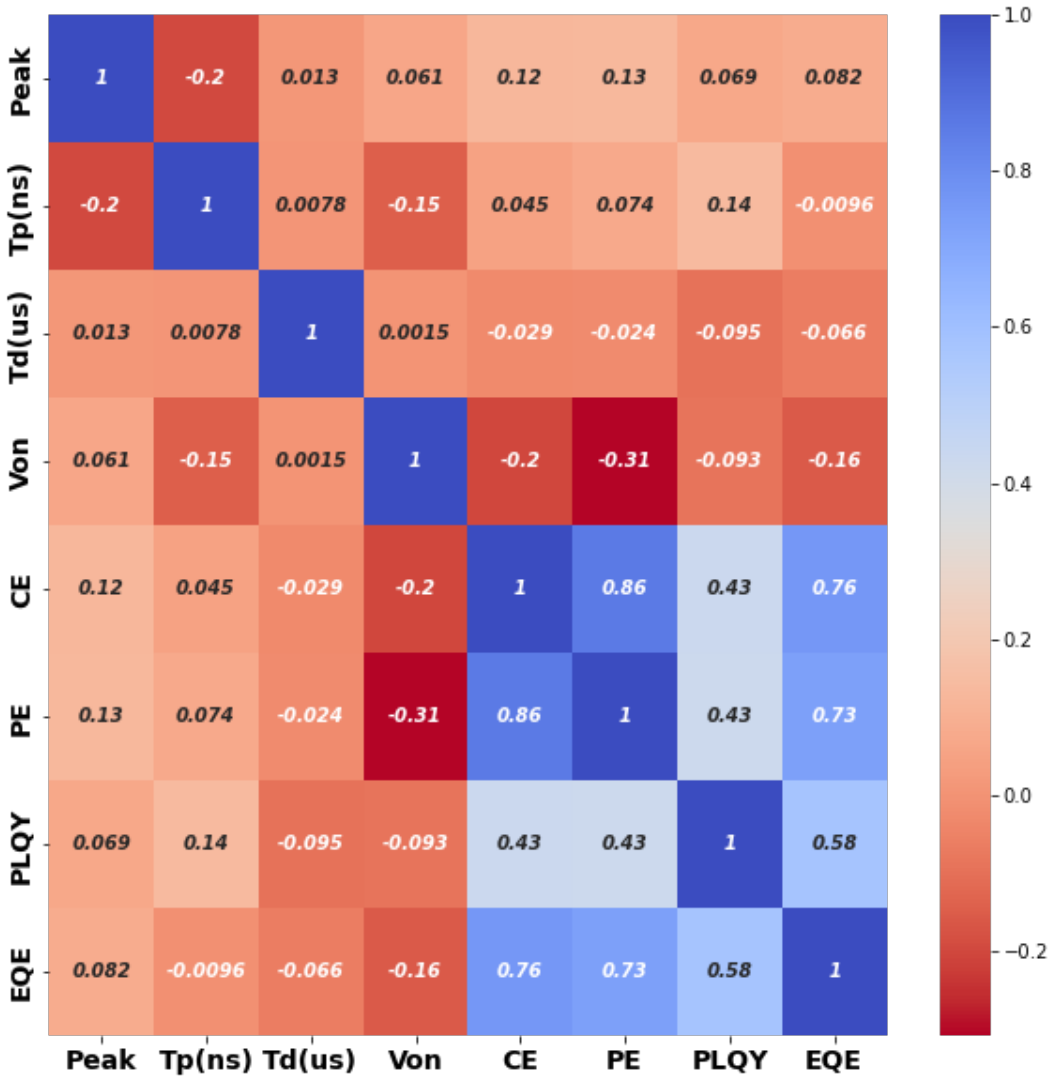

Figure 2. Correlation matrix of the material and TADF factors influencing the EQE. The positive and high Pearson correlation coefficient values (r) shown in the plot implies the strong correlation between the variables.

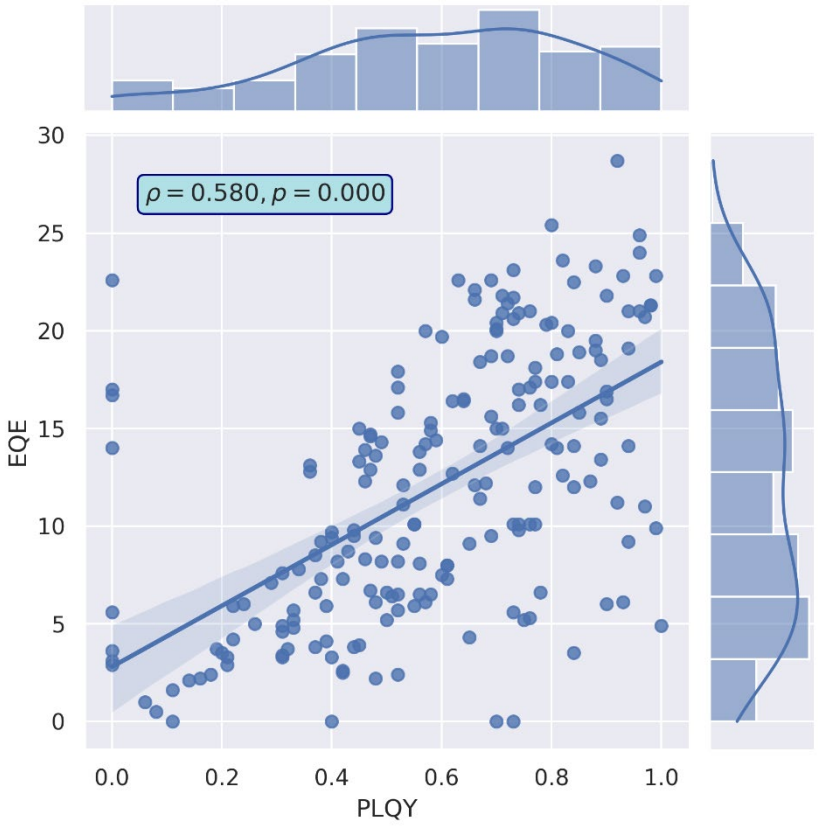

Figure 3. Regression scatter plots of EQE values versus PLQY of the constructed data set.

The computed Pearson's correlation coefficient (r) indicates only a slight correlation between PLQY and EQE, and no clear trend was identified using a simple statistical algorithm. As can be seen in the scattered plot in Figure 3, the relationship between PLQY and EQE is inversely proportional. For example, when the PLQY distribution is high, around 75%, the EQE distribution is minimal at around 15%. However, when the PLQY distribution is around 40%, the EQE distribution is at its maximum, around 5%. While some of the materials have higher efficiencies surpassing 25%, it is unexpected that the majority of the EQE distribution falls within the range of 10%. This suggests that the current statistical model needs to be revised in order to find a more accurate predictive model for device efficiency based on the available experimental data points.

Therefore, it is necessary to use a machine-learning algorithm to construct better models for the prediction of EQE, based on simple features.

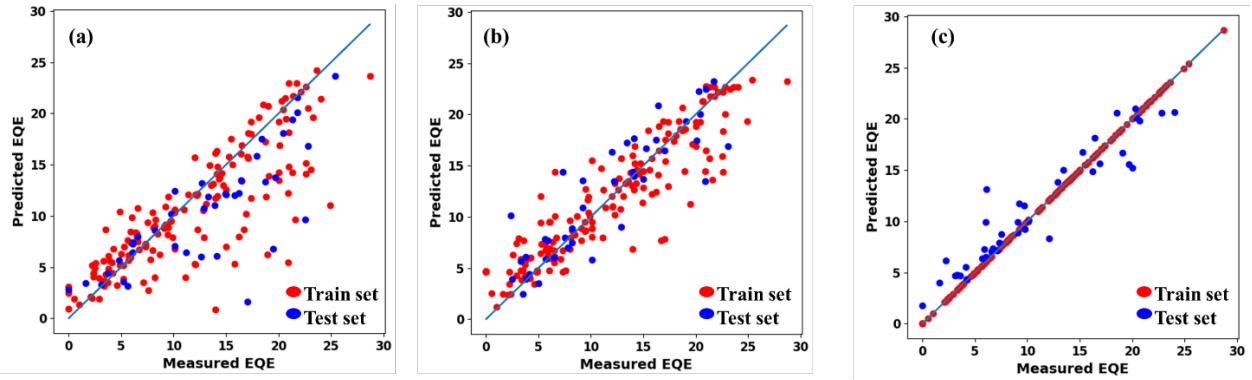

Figure 4. Comparison of measured and predicted EQE by different ML algorithms on train and test sets; (a) Support Vector, and (b) K-NN (K = 3), (c) Random Forest.

Theoretically, a highly accurate prediction is mapped onto the coefficient of determination (r^2) value, which is pictorially closer to the red line as seen in Figure 4. Based on the data set, application of Radom Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) model to evaluate the relationship exhibits a mixed performance matrices as shown in Table1. For the RF model, r^2 of 0.99 with RMSE of 0.01 is obtained for the training set and r^2 of 0.83 with of RMSE of 7.16 is obtained for the testing set. In contrast, a very poor performance of matrices based on the SVM and K-NN (K =3) ML model is found. In SVM model, r^2 of the train and test set arises at 0.64 and 0.36 with RMSE value of 16.12 and 28.40, respectively. The K-NN with the nearest neighbor of the K=3 regression model has the training/testing RMSE of 10.25/9.87 with an r^2 value of 0.78/0.74. The fitting r^2 values demonstrate the satisfactory execution of the RF model based on the data set and shows a better performance rather than SVM and K-NN model.

Table 1. Performance evaluation of different ML algorithm on predicting EQE on the training and testing data set. r^2 and RMSE are evaluated as the mean over the 10-fold cross validation protocol.

ML model	Data	$rR2^2$	RMSE
SVM	Train (80%)	0.64	16.12
	Test (20%)	0.36	28.40
K-NN (K= 3)	Train (80%)	0.78	10.25
	Test (20%)	0.74	9.87
RF	Train (80%)	0.99	0.01
	Test (20%)	0.83	7.16
GB	Train (80%)	0.99	0.02
	Test (20%)	0.73	11.45
XGB	Train (80%)	0.98	0.52
	Test (20%)	0.82	7.65
LGB	Train (80%)	0.95	2.13
	Test (20%)	0.80	9.57

Inspired by the success rate of the RF model in the current data set, ensemble based classification model are employed further over the studied data set. The performance of matrices is tabulated in Table 1. It is seen that Gradient Boosting (GB) model has shown better performance compared to other Boosting models in this data set as seen in Figure5. Based on the 10-fold cross-validation method, GB shows an impressive r^2 value of 0.99 and 0.73 with a lower RMSE value of 0.02 and 11.45 for the train and test set, respectively. In comparison to the GB model, XGB has achieved an r^2 value of 0.98/0.82 and an RMSE of 0.52/7.65 for the train/test set. The LGB model also executed with an RMSE value of 2.13 and 9.57 for 10-fold cross-validation over the train set and test set, while the r^2 value for the train and test set arises at 0.95 and 0.80. Meanwhile, evaluation

of performance matrices of the different models shows that the GB model is preferable than any other Boosting algorithm both in training and in testing set.

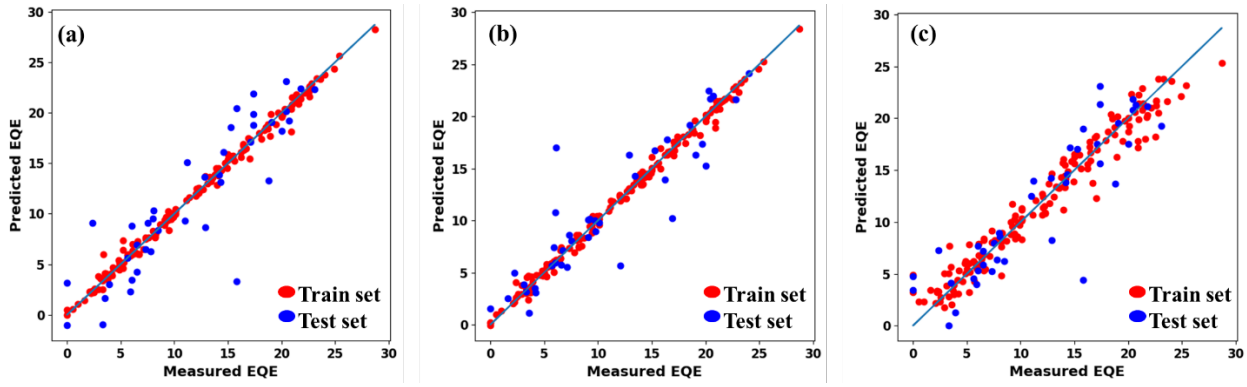

Figure 5. Comparison of true EQE and predicted values by different ML algorithms on training and test sets; (a) Gradient Boosting, (b) Extreme Gradient Boosting, (c) Light Gradient Boosting machine, algorithm.

To get insight into the importance of features to predict the EQE of non-doped TADF emitters, the RF model is taken as a further features evaluation. Surprisingly, CE and PE are the most dominating factor for the EQE prediction with the highest relative importance score of ~0.30. While, PLQY and Peak place second and third rank with a score value of ~0.17 and ~0.07, respectively (see supporting information file). It is important to note that emission wavelength is a crucial feature for the prediction of PLQY which has a smaller effect on EQE prediction based on the RF model. Similarly, other features have a negligible impact. In addition to the feature correlation in the RF model, very similar trends are observed in the GB model (see supporting information file). Similarity performance in both RF and GB models stresses that ensemble learning would be a reliable approach for the data-driven modeling of organic electronics.

Ensemble learning techniques, such as Random Forest and Gradient Boosting Regression, are characterized by their low bias and low variance properties. This means that they are able to

effectively capture the underlying patterns within the training data and have a reduced risk of overfitting, even when the size of the training set is relatively small. The Random Forest Regression algorithm is based on the principle of Bagging, while the Gradient Boosting Regression method follows the principles of Boosting. As a result of this, these ensemble techniques possess excellent generalization capabilities.

Conclusions

This contribution aims to use a data-driven framework to unbiasedly compare the experimental features of efficient TADF materials. Machine learning algorithms are used to analyze and model key characteristics controlling the EQE of a TADF device. The analysis shows that four features - PE, CE, PLQY, and emission wavelength - strongly correlate with EQE. The Random Forest and Gradient Boosting models show the best performance accuracy for predicting EQE for non-doped TADF emitters. Feature importance and correlation analysis are performed to identify the key factors for EQE prediction. The results suggest that ensemble-based learning can be used to identify promising non-doped TADF materials for further study and to understand the relationship between material properties and device performance.

Computational Methods

In this work, python programming language was used in the google-colab and scikit-learn packages has been employed for the simulation of all ML algorithms.[46, 47] Total 200 data set points were collected from the literature and the entire data sets are given in supporting information file. During the model fitting, data sets were divided into 80% and 20% as a training and test data set and10-fold cross-validation was performed to evaluate all ML model. The success of the ML

model has been quantified by the measurement of the r^2 i.e. regression score function and root-mean square error (RMSE) value using the following equations:

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - y_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - y_i)^2}$$

where Y_i and y_i represent the ith value of the measured experimental data set and ith value of the predicted data set based on ML model. $\bar{Y}$ and n indicate the average value obtained from experimental data set, and total number of dataset points, respectively.

Linear Regression is a fundamental technique that is utilized to establish a linear relationship between a set of predictor variables and a target variable (EQE) in a given dataset. However, this simple model can fall short in its ability to accurately capture the underlying patterns within the data, which is why more advanced methods are often employed to improve the predictions. One such method is the K-Nearest Neighbour Regression algorithm, which is a distance-based approach that attempts to map the features to the EQE by analyzing the closest training examples. Another powerful method is Support Vector Machine Regression, which uses a hyperplane to fit a model that can capture more complex, non-linear patterns. Furthermore, I employed a variety of ensemble models such as Random Forest Regression, Gradient Boost Regression, Extreme Gradient Boosting Regression, and Light Gradient Boosting Machine Regression, which are tree-based models that leverage multiple decision trees and take an average of the predictions to improve the overall predictions. Additionally, I have provided the results from 100 cross-validated distributions of training and prediction RMSE for each of the algorithms applied.

ASSOCIATED CONTENT

Supporting Information. Additoonl result and discussions and date sets can be found at supporting information file. Code is available at: https://github.com/qckb

AUTHOR INFORMATION

Corresponding Author

Kalishankar Bhattacharyya* Max-Planck-Institut für Kohlenforschung, Department of Molecular Theory and Spectroscopy, Kaiser-Wilhelm-Platz 1, 45470, Mülheim an der Ruhr, Germany.

Present Address: Max-Planck-Institut für Polymer Research, Mainz-55128, Germany.

E-mail: ksb@kofo.mpg.de

Funding Sources

KB acknowledge the Alexander von Humboldt Foundation for postdoctoral fellowship.

Notes: The authors declare no competing financial interests.

ACKNOWLEDGMENT

The Author thanks Mr. Sayan Dutta and Dr. Indranil Bhattacharjee for fruitful discussions.

REFERENCES

(1) Keith, J. A.; Vassilev-Galindo, V.; Cheng, B.; Chmiela, S.; Gastegger, M.; Müller, K.-R.; Tkatchenko, A. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chem. Rev.* **2021**, *121* (16), 9816-9872.
(2) Westermayr, J.; Gastegger, M.; Schütt, K. T.; Maurer, R. J. Perspective on integrating machine learning into computational chemistry and materials science. *J. Chem. Phys.* **2021**, *154* (23), 230903.

(3) Ramakrishnan, R.; von Lilienfeld, O. A. Machine learning, quantum chemistry, and chemical space. *Rev. Comput. Chem.* **2017**, *30*, 225-256.
(4) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559* (7715), 547-555.
(5) Dral, P. O. Quantum chemistry in the age of machine learning. *J. Phys. Chem. Lett.* **2020**, *11* (6), 2336-2347.
(6) Hachmann, J.; Olivares-Amaya, R.; Atahan-Evrenk, S.; Amador-Bedolla, C.; Sánchez-Carrera, R. S.; Gold-Parker, A.; Vogt, L.; Brockway, A. M.; Aspuru-Guzik, A. The Harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid. *J. Phys. Chem. Lett.* **2011**, *2* (17), 2241-2251.
(7) Hueffel, J. A.; Sperger, T.; Funes-Ardoiz, I.; Ward, J. S.; Rissanen, K.; Schoenebeck, F. Accelerated dinuclear palladium catalyst identification through unsupervised machine learning. *Science* **2021**, *374* (6571), 1134-1140.
(8) Reid, J. P.; Sigman, M. S. Holistic prediction of enantioselectivity in asymmetric catalysis. *Nature* **2019**, *571* (7765), 343-348. DOI: 10.1038/s41586-019-1384-z From NLM PubMed-not-MEDLINE.
(9) Kitchin, J. R. Machine learning in catalysis. *Nat. Catal.* **2018**, *1* (4), 230-232.
(10) Esterhuizen, J. A.; Goldsmith, B. R.; Linic, S. Interpretable machine learning for knowledge generation in heterogeneous catalysis. *Nat. Catal.* **2022**, *5* (3), 175-184.
(11) Westermayr, J.; Marquetand, P. Machine learning for electronically excited states of molecules. *Chem. Rev.* **2020**, *121* (16), 9873-9926.
(12) Dral, P. O.; Barbatti, M.; Thiel, W. Nonadiabatic excited-state dynamics with machine learning. *J. Phys. Chem. Lett.* **2018**, *9* (19), 5660-5663.
(13) Mai, H.; Le, T. C.; Chen, D.; Winkler, D. A.; Caruso, R. A. Machine learning for electrocatalyst and photocatalyst design and discovery. *Chem. Rev.* **2022**, *122* (16), 13478-13515.
(14) Zhuo, Y.; Mansouri Tehrani, A.; Brgoch, J. Predicting the band gaps of inorganic solids by machine learning. *J. Phys. Chem. Lett.* **2018**, *9* (7), 1668-1673.
(15) Palkovits, S. A Primer about Machine Learning in Catalysis–A Tutorial with Code. *ChemCatChem* **2020**, *12* (16), 3995-4008.
(16) Liu, X.; Wang, X.; Gao, S.; Chang, V.; Tom, R.; Yu, M.; Ghiringhelli, L. M.; Marom, N. Finding predictive models for singlet fission by machine learning. *Npj Comput. Mater.* **2022**, *8* (1), 1-10.
(17) Li, P.; Wang, Z.; Li, W.; Yuan, J.; Chen, R. Design of Thermally Activated Delayed Fluorescence Materials with High Intersystem Crossing Efficiencies by Machine Learning-Assisted Virtual Screening. *J. Phys. Chem. Lett.* **2022**, *13* (42), 9910-9918.
(18) Shi, H.; Jing, W.; Liu, W.; Li, Y.; Li, Z.; Qiao, B.; Zhao, S.; Xu, Z.; Song, D. Key Factors Governing the External Quantum Efficiency of Thermally Activated Delayed Fluorescence Organic Light-Emitting Devices: Evidence from Machine Learning. *ACS omega* **2022**, *7* (9), 7893-7900.
(19) Palkovits, R.; Palkovits, S. Using artificial intelligence to forecast water oxidation catalysts. *ACS Catal.* **2019**, *9* (9), 8383-8387.
(20) Yılmaz, B.; Yıldırım, R. Critical review of machine learning applications in perovskite solar research. *Nano Energy* **2021**, *80*, 105546.
(21) Li, J.; Pradhan, B.; Gaur, S.; Thomas, J. Predictions and strategies learned from machine learning to develop high-performing perovskite solar cells. *Advanced Energy Materials* **2019**, *9* (46), 1901891.

(22) Mazurenko, S.; Prokop, Z.; Damborsky, J. Machine learning in enzyme engineering. *ACS Catal.* **2019**, *10* (2), 1210-1223.
(23) Uoyama, H.; Goushi, K.; Shizu, K.; Nomura, H.; Adachi, C. Highly efficient organic light-emitting diodes from delayed fluorescence. *Nature* **2012**, *492* (7428), 234-238.
(24) Tanaka, H.; Shizu, K.; Miyazaki, H.; Adachi, C. Efficient green thermally activated delayed fluorescence (TADF) from a phenoxazine–triphenyltriazine (PXZ–TRZ) derivative. *Chem Comm* **2012**, *48* (93), 11392-11394.
(25) Adachi, C.; Baldo, M. A.; Thompson, M. E.; Forrest, S. R. Nearly 100% internal phosphorescence efficiency in an organic light-emitting device. *J. Appl. Phys.* **2001**, *90* (10), 5048-5051.
(26) Sekitani, T.; Nakajima, H.; Maeda, H.; Fukushima, T.; Aida, T.; Hata, K.; Someya, T. Stretchable active-matrix organic light-emitting diode display using printable elastic conductors. *Nat. Mater.* **2009**, *8* (6), 494-499.
(27) Bhattacharjee, I.; Acharya, N.; Bhatia, H.; Ray, D. Dual emission through thermally activated delayed fluorescence and room-temperature phosphorescence, and their thermal enhancement via solid-state structural change in a carbazole-quinoline conjugate. *J. Phys. Chem. Lett.* **2018**, *9* (11), 2733-2738.
(28) Bhattacharyya, K. Can TDDFT render the electronic excited states ordering of Azine derivative? A closer investigation with DLPNO-STEOM-CCSD. *Chem. Phys. Lett.* **2021**, *779*, 138827.
(29) 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.
(30) Hirata, S. Recent advances in materials with room-temperature phosphorescence: photophysics for triplet exciton stabilization. *Adv. Opt. Mater.* **2017**, *5* (17), 1700116.
(31) Zhao, G.; Liu, D.; Wang, P.; Huang, X.; Chen, H.; Zhang, Y.; Zhang, D.; Jiang, W.; Sun, Y.; Duan, L. Exceeding 30% External Quantum Efficiency in Non-doped OLEDs Utilizing Solution Processable TADF Emitters with High Horizontal Dipole Orientation via Anchoring Strategy. *Angew. Chem. Int. Ed.* **2022**, *61* (45), e202212861.
(32) Zhang, Y. L.; Ran, Q.; Wang, Q.; Liu, Y.; Hänisch, C.; Reineke, S.; Fan, J.; Liao, L. S. High-efficiency red organic light-emitting diodes with external quantum efficiency close to 30% based on a novel thermally activated delayed fluorescence emitter. *Adv. Mater.* **2019**, *31* (42), 1902368.
(33) Lee, M.-H. Identification of host–guest systems in green TADF-based OLEDs with energy level matching based on a machine-learning study. *Phys. Chem. Chem. Phys.* **2020**, *22* (28), 16378-16386.
(34) Tu, C.; Huang, W.; Liang, S.; Wang, K.; Tian, Q.; Yan, W. Combining machine learning and quantum chemical calculations for high-throughput virtual screening of thermally activated delayed fluorescence molecular materials: the impact of selection strategy and structural mutations. *RSC advances* **2022**, *12* (48), 30962-30975.
(35) Friederich, P.; Fediai, A.; Kaiser, S.; Konrad, M.; Jung, N.; Wenzel, W. Toward design of novel materials for organic electronics. *Adv. Mater.* **2019**, *31* (26), 1808256.
(36) Pollice, R.; dos Passos Gomes, G.; Aldeghi, M.; Hickman, R. J.; Krenn, M.; Lavigne, C.; Lindner-D'Addario, M.; Nigam, A.; Ser, C. T.; Yao, Z. Data-driven strategies for accelerated materials design. *Acc. Chem. Res.* **2021**, *54* (4), 849-860.
(37) Miyake, Y.; Saeki, A. Machine Learning-Assisted Development of Organic Solar Cell Materials: Issues, Analyses, and Outlooks. *J. Phys. Chem. Lett.* **2021**, *12* (51), 12391-12401.

(38) 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. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nat. Mater.* **2016**, *15* (10), 1120-1127.
(39) 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* (26), 1900891.
(40) Chan, C. Y.; Cui, L. S.; Kim, J. U.; Nakanotani, H.; Adachi, C. Rational molecular design for deep-blue thermally activated delayed fluorescence emitters. *Adv. Funct. Mater.* **2018**, *28* (11), 1706023.
(41) Lee, M.-H. Robust random forest based non-fullerene organic solar cells efficiency prediction. *Org. Electron.* **2020**, *76*, 105465.
(42) Zhao, K.; Omar, Ö. H.; Nematiaram, T.; Padula, D.; Troisi, A. Novel thermally activated delayed fluorescence materials by high-throughput virtual screening: going beyond donor–acceptor design. *J. Mater. Chem. C* . **2021**, *9* (9), 3324-3333.
(43) 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* (18), 7777-7787.
(44) Garain, B. C.; Pati, S. K. Improved Prediction of Maximum EQE in TADF-based OLEDs Through Ensemble Learning. **2022**.
(45) Shi, Y.-Z.; Wu, H.; Wang, K.; Yu, J.; Ou, X.; Zhang, X. Recent progress in thermally activated delayed fluorescence emitters for nondoped organic light-emitting diodes. *Chem. Sci.* **2022**, *13*, 3625-3651.
(46) Varoquaux, G.; Buitinck, L.; Louppe, G.; Grisel, O.; Pedregosa, F.; Mueller, A. Scikit-learn: Machine learning without learning the machinery. *GetMobile: Mobile Computing and Communications* **2015**, *19* (1), 29-33.
(47) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825-2830.