Improved Prediction of Maximum EQE in TADF-based OLEDs Through Ensemble Learning

Bidhan Chandra Garain, and Swapan K Pati

Theoretical Sciences Unit,
School of Advanced Materials (SAMat),
Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR),
Bangalore 560064, India.

Email: pati@jncasr.ac.in & swapan.jnc@gmail.com
Abstract

In recent times, we have seen colossal growth in the field of Thermally Activated Delayed Fluorescence (TADF)-based OLEDs in terms of synthesis and applications in sensing and imaging. However, the device-level application is still limited to the unpredictability of external quantum efficiency (EQE). Although theoretical research involving internal quantum efficiency (IQE) and mechanistic pathways for reverse intersystem crossing (rISC) in TADF systems have been explored quite rigorously, investigation on EQE is lacking. With the emergence of data-driven analysis being the fourth paradigm of science (empirical, theoretical, and computational being the previous three), we have employed ML models on 30 features of 123 samples, availed from literature to predict the EQE\textsubscript{max}. On the one hand, the employed models capture device selectivity but are prevalent to the emissive range of chromophores. We have shown that Gradient Boosting (GB), an ensemble learning model, has been able to predict EQE\textsubscript{max} with $r^2$ score of 0.71±0.04/0.84 and a low RMSE of 4.22±0.55/2.53 for the train/test set. Considering the current state-of-the-art (SOTA), this is the best model which can predict for TADF chromophores of any emissive range and delineate the effect of device architecture. We also have carried out feature importance analysis to make this so-called black-box model interpretable. This analysis has helped to figure out essential parameters responsible for better EQE efficiency. Even the learning curve is still ascending, proving that the model can improve its prediction if more training examples are provided in the future. All the computations can be done using easily accessible cloud computations.

**Keywords:** Machine Learning, TADF, OLEDs, EQE, Ensemble learning
1. Introduction

The advancement in the research of Thermally Activated Delayed Fluorescence (TADF)-based materials has added a whole new dimension to the field of OLEDs with prodigious applications ranging from imaging to sensing.\textsuperscript{1-3} From conventional fluorescence emitters, since the ground state is in singlet state, we could only harvest one-quarter (singlet state: triplet states is 1:3) of the generated singlet excitons for emission. Whereas in TADF emitters, the other three-quarter of the total generated excitons get converted into triplet excitons, which can be again be harvested into singlet excitons through reverse-intersystem crossing (rISC) mechanism. So, theoretically, we can achieve 100\% internal quantum efficiency (IQE). Efficient rISC requires the proper balance of two contrasting parameters, low singlet-triplet gap between first excited singlet state ($S_1$) and lowest triplet state ($T_1$) and moderate spin-orbit coupling (SOC) between these two states. The traditional way to synthesize a TADF emitter is to make a perpendicular donor-acceptor (D-A) system either through a bond between donor (D) and acceptor (A) or through a spiro-conjunction between D and A. This architecture reduces the overlap between HOMO and LUMO inducing charge-transfer (CT) characteristics in the low-lying excited states ($S_1$ and $T_1$) and effectively reduces the energy gap. While this phenomenon lowers the coupling strength between these two states owing to the same symmetry of excitation and reduces the oscillator strength for the transition from $S_1$ to the ground state ($S_0$). In the last couple of decades, a large number of developments happened in the field of efficient organic TADF-based chromophores and their synthesis. With such burgeoning, alternative approaches to D-A systems have been scanned. Hatakayama \textit{et al.} have shown that nitrogen is manifests a opposite resonance effect than boron and relative para-substitution of boron and nitrogen can enhance such opposite resonance effect which can significantly separate the HOMO-LUMO.\textsuperscript{4} Devising coplanar system with high oscillator strength with low $\Delta E_{ST}$ have been shown by Chen \textit{et al.}\textsuperscript{5} Even the theoretical studies involving spin-vibronic coupling or hyperfine coupling has made a significant impact explaining rISC pathways in TADF systems.\textsuperscript{6-8} Samanta \textit{et al.} have shown the effect of LE characteristics in triplet manifolds for a better rISC rate.\textsuperscript{9}

With the evolution of data-driven analysis, researchers have started using ML-based models. Aspuru-Guzik \textit{et al.} have used neural nets and subsequent quantum-chemistry calculations for accurate prediction of rISC rate ($k_{ISC}$) and from high throughput virtual screening they have been able to screen TADF emitter with EQE_{max}~22 \%.\textsuperscript{10} Meftahi \textit{et al.} have predicted the power conversion efficiency (PCE) using BRANNLP method with an $r^2$ score of 0.72.\textsuperscript{11} Chan \textit{et al.} have predicted the current efficiency (CE) for blue phosphorescent materials from various device descriptors like layer thickness and triplet energy of electron transport layer through various machine learning algorithms, like XGBoost (Extreme gradient boosting), Random Forest (RF) and k-Nearest Neighbor (kNN).\textsuperscript{12} RF has shown the highest efficiency with $r^2$ score of 0.67/0.73 for train/test datasets. Troisi \textit{et al.} have also done a virtual screening of 700 TADF emitters out of 40000 systems for quantum chemistry calculations and predicted 125 highly efficient emitters among them, whose structures significantly deviate from the D-A architecture.\textsuperscript{13} Shu \textit{et al.}
also have used the high throughput virtual screening (HTVS) method for scanning better TADF emitters.\textsuperscript{14} However, to the best of our knowledge, none of these models emphasize the relationship between EQE\textsubscript{max} and device architecture. Recently, Lee has proposed an RF-based ML approach to predict EQE\textsubscript{max} only for green TADF emitters with $r^2$ score of 0.85/0.74 for train/test dataset with RMSE of 4.16/4.72 respectively using just host and guest parameters.\textsuperscript{15} Importance of the similarity matrix between host and guest has been emphasized in this work.

Generally, in the OLED field particularly in TADF-based OLEDs, there are two steps. The first one is to synthesize a chemical system (GUEST) that will have a low $\Delta E_{ST}$ and thus it can manifest a high PLQY or internal quantum efficiency. Most of the previously mentioned works try to fill this void (GUEST). The other part comes next, which is to engineer the device architecture from the chemical systems, which is the crucial part for the overall performance of the OLED devices. This work does not survey the molecular structure part. That is beyond the scope of this work. But rather it is centered around the device architecture part. From the literature, it is evident that a molecule with small $\Delta E_{ST}$ or high PLQY does not exhibit a high external quantum efficiency (EQE). The output parameter not only depends on the previous two parameters but rather on the overall device architecture. Finding the optimum device architecture is a trial and error-based method and extremely expensive because of both instrumental requirements (like spin coating or chemical vapor deposition) and the cost of the materials of HTL and ETL. Although, interestingly over the years, many devices have been developed. So, with data availability and massive growth of machine learning (ML) in recent years, ML-based prediction for EQEmax for any TADF emitters irrespective of their emissive range (or CIE value) has still been lacking.

Our work in this article probes on the engineering aspect of the TADF-based OLED architectures rather than the chemical aspect of molecular structure of GUEST through ML-based approach. We have carried out this work to optimize with machine learning algorithms what the optimum requirements of the device architecture would be to give best EQEmax. Device structure beyond host and guest makes a significant impact on EQEmax like various energy levels of hole and electron transport layer and their thickness. Certainly, a change in device fabrication leads to a certain change in EQEmax. The work on the detailed aspect of the optimum molecular structures of GUEST through ML-based approach is beyond the scope of this article.

In this work, we have compiled our own database from the existing device information for TADF-based OLEDs available in the literature and employed ensemble learning (Bagging and Boosting) based models for improved predictions of EQEmax. Among the models, Gradient Boosting (GB) has shown the best performance with an $r^2$ score of 0.71/0.84 for train/test set and RMSE of 4.22/2.53 respectively which is good considering the applicability of the model to all TADF emitters along with device selectivity with close range metrics compared to the models with high selectivity to chromophores with specific emissive range. After the model implementations, we have checked the feature importance score to analyse the effects of the input features to give insights.
We do hope that this work will make a noteworthy impact on further research on TADF-based OLEDs. All the descriptors can be obtained experimentally and learning curve is still rising to indicate this model will perform even better with increasing data points. The computation is easily accessible to experimentalists since all the computations can be done using free cloud computation (Google Colab).

2. Computational Methodology

In this work, we have collected 123 device information of 123 different chromophores and from there 30 features have been selected. They are (1) HOMO energy of GUEST (2) -LUMO energy of GUEST (3) -HOMO energy of HOST (4) -LUMO energy of hole injection layer (HIL) (5) ΔE_{ST} of GUEST (6) -HOMO energy of 1st hole transport layer (HTL1) (7) HIL thickness (8) -HOMO energy of 1st hole transport layer (HTL1) (9) -LUMO energy of 1st hole transport layer (HTL1) (10) HTL1 thickness (11) -HOMO energy of 2nd hole transport layer (HTL2) (12) -LUMO energy of 2nd hole transport layer (HTL2) (13) HTL2 thickness (14) -HOMO energy of 3rd hole transport layer (HTL3) (15) -LUMO energy of 3rd hole transport layer (HTL3) (16) HTL3 thickness (17) total HTL thickness (HTL1 + HTL2 + HTL3) (18) emissive layer (EML) thickness (19) T_{1} energy of 1st electron transport layer (ETL1) (20) -HOMO of ETL1 (21) -LUMO of ETL1 (22) ETL1 thickness (23) -HOMO energy of ETL2 (24) -LUMO energy of ETL2 (25) ETL2 thickness (26) total ETL thickness (ETL1 + ETL2) (27) Work function of the cathode (28) Cathode thickness (29) doping concentration (30) PLQY in HOST matrix. In the case of the absence of HTL3 material, the thickness is 0 and HOMO-LUMO energy is taken to be of HLT2 material. The same approach is also taken when ETL1 material is absent.

We have considered 85% of the total data as a training dataset (104) and 15% as a test dataset (19). Further, we did 4-fold cross-validation on train dataset (4 * 26 + 19 = 104 + 19 = 123) while evaluating any ML model. A detailed description on the cross-validation can be found in SI. We have performed all the calculations using Google Colab.

2.1 Metrics

Pearson correlation: Correlation coefficients measure the nature and strength of the relationships between two variables. One such important correlation coefficient is the Pearson correlation of Pearson r, which is defined as,

\[
r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}
\]

where x_i is the input variables, \( \bar{x} \) is the mean value of input variables, y_i is the output variables and \( \bar{y} \) is the mean value of the output variables.

r² score: it is regression score function determines how good a given fit is, it is defined as,
\[ r^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \]

where \( \hat{y}_i \) is the predicted value for the \( i \)th sample, \( y_i \) is the true value and \( \bar{y} \) is again the mean value of the output variables.

**RMSE value:** RMSE is the standard deviation of the residuals. This is a more quantitative metric compared \( r^2 \) score. This is defined as,

\[ RMSE(y, \hat{y}) = \sqrt{\frac{1}{n-1} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2} \]

### 2.2 Machine Learning Models:

Initially, we have started with employing Support Vector Regression (SVR), K-NN regression and Decision Tree (DT) as very simple ML models.

1) SVM: SVM is very simple, and it easily captures non-linear relationships because of its kernel-trick algorithm. So, according to our initial assessment, this model should have been able to capture the non-linearity of the model. Even we have used the polynomial kernel in SVM method. Mathematical formulations are discussed and is given in SI (Machine Learning Models).

2) Decision Tree: Although DT is generally used for the classification problems, here we have used it for regression problem. From SVM implementation, we have understood that although the relationships are non-linear in nature, but these not so straightforward. So, we have proceeded through a logic-based approach. Since, we had 30 features, we have assessed that it would be helpful to grow trees downward and big tree will be fruitful in terms of results and interpretation. Mathematical formulations are discussed and is given in SI (Machine Learning Models).

E.g.: If HOMO energy of HTL1 is lower than Y + PLQY is higher than X, what can be the value by a node of the tree.

3) K-NN: This model is fully non-parametric. The reason for choosing the model is to get better performance. Since the output value is calculated from the nearby data points by taking the average of the nearby input values.

\[ \hat{y} = \frac{1}{K} \sum_{i=1}^{K} y_k \]

Further moved towards an ensemble-based model with DT as a weak learner. Although SVM and kNN are nowadays getting used as base learner but are not well documented. In ensemble-based learning models, we have employed Random Forest (RF), AdaBoost, Gradient Boosting (GB), XGBoost (XGB) and LightGBM models.
be used for ensemble learning, but the state-of-the-art works which included the SVM as base learner for ensemble learning did not perform up to the mark. Even kNN as base learner for ensemble learning is not also well documented. Only DT has been hugely successful as a base learner for ensemble learning methods. Both bagging and boosting methods use the growing and pruning techniques of ensemble of trees for better performance than a single tree can provide. An explanation of weak learner vs. ensemble learner and a schematic representation of success of ensemble learning is given in SI (Figure S9).

First, we have used AdaBoost. It is the simplest boosting algorithm. It uses stumps (a decision tree with one node and two leaves) as its base learner. Generally, stumps are weak in predicting because of their depth. But, in AdaBoost, a sample is initially chosen considering an equal probability of choice for all samples. Later, based on the performance of a stump next sample is chosen with a different probability. This procedure goes on. In the end, each stump is associated with different weights based on the outcome results and weighted mean is taken as the final output.

The idea behind Gradient Boosting and Extreme Gradient Boosting is same. Both the methods try to reduce the loss function with the addition of new trees. But addition of the trees is guided by the derivative of loss function by previous results, which is called the gradient. In GB, only 1st order term is only considered, while in XGB, both 1st and 2nd order derivatives are considered. As a result, overfitting is an issue. So subsequently, regularization terms are also added. A more detailed methodology is given in SI (Machine Learning Models).

Both the methods mentioned above require a large computational cost if they require frequent retraining. Light Gradient Boosting Machine (LightGBM) is the most recent modification of the gradient boosting algorithm. Apart from GBDT, it adds gradient-based one-sided sampling (GOSS) and elusive feature bundling (EFB) which reduces the overall computational cost. It is nearly 20 times faster than conventional GBDT. Random Forest makes an ensemble of DTs, each of them capable of a prediction but all the outcomes are pooled (taking the mean in case of regression) for final prediction. While RF does a bootstrapping of datasets and gives an aggregated output. The base learner is the DT. Each time a row sampling (some examples are taken) and column sampling (some features are taken), which is basically called bagging takes place with replacement, and they are fed to a DT and output is evaluated.

**Hyperparameter Tuning:**

We have carried out hyperparameter tuning using GridSearchCV algorithm for the models except for RF. For RF, we have used RandomSearchCV algorithm. All the hyperparameters are listed in Table S1. All the implemented codes, raw dataset and preprocessed dataset can be found at https://github.com/BidhanChandra/ML_EQE_max_Prediction

3. **Results and Discussions**

3.1 Exploratory Data Analysis (EDA)
EDA helps to understand the pattern within the data, preprocess the data for the ML part and find relevant parameters which influence the overall model efficiency, which includes calculation of Pearson correlation and visualization of heatmap between output and inputs also among the inputs. Fig. 1 (a) represents the correlation heatmap between variables and as well as with the output variable with a correlation value greater than 0.20. The Full correlation heatmap is shown in Figure S1. PLQY shows the highest positive correlation of 0.73 with the EQE\text{max} while ΔE_{ST} shows a negative correlation of -0.39 with EQE\text{max}, which is quite expected. Note that, this kind of correlation was not seen in earlier scientific literature, most likely due to the repetitive usage of the same input with varied output. Fig. 1 (b) represents further details into the relationship of PLQY with EQE\text{max}. This figure suggests that although there is a one-to-one correlation between PLQY with EQE\text{max}, a high PLQY does not guarantee high EQE\text{max}. Now, to analyze the effect of HOMO-LUMO offset, we find that even though there are systems with PLQY equal to 100%, like SpiroAc-TRZ, but high EQE\text{max} ~37% become possible because of a high horizontal dipole ratio of 83%. High horizontal dipole ratio ensures good exciton transfer from HOST to GUEST. To find the actual effect of HOMO-LUMO offset, we argue that devices where LUMO energy of HOST – LUMO energy of GUEST) is high, EQE\text{max} is generally high if PLQY is large. This energy ordering further prevents any exciton transfer to the HOST from GUEST and improves the efficiency.

Figure 2 represents the regression plot with the distribution of EQE\text{max} with PLQY, ΔE_{ST}, -LUMO energy of HTL1 and -HOMO energy of ETL2. As we have mentioned earlier, a
high value of PLQY does not guarantee a high value of EQE_{max}. PLQY has a very high distribution in the range of 80-100, but we found a very small distribution. Rather, PLQY distribution in the range of 60-80 produced a large distribution of EQE_{max} in the range of 10-20. ΔE_{ST} < 100 meV always favors rISC and as a result, most of the high EQE_{max} systems have high distribution over the range 0.0-0.1 eV. However, there is also light distribution when ΔE_{ST} > 100 meV. We believe that those systems may employ the spin-vibronic coupling mechanism to harness the excitons. The -LUMO energy of HTL1 is inversely correlated with the EQE_{max}. Higher LUMO energy is better for improved EQE_{max}. On the other hand, -HOMO energy of ETL2 is positively correlated with EQE_{max}, so that lower the HOMO energy, better is the EQE_{max}. HTL1 is connected to HIL/anode (in case of absence of HIL) and ETL2 is connected to the cathode. Lower HOMO energy contributes to the high triplet energy of ETL2, which is favorable for electron transfer through the overall ETL layer and ceasing dissipation of excitons.

3.2 Machine Learning

For the initial analysis of the data, we have first employed simple ML algorithms (Support Vector Regression, K-Nearest Neighbor and Decision Trees) to train the data and have tested the employed model against test data. The details for choosing such models are explained in the SI. The $r^2$ score and the corresponding RMSE values for the models for both train and set are reported in Table- 1.

![Figure 2](image_url)

Figure 2 Regression plot of EQE_{max} vs (a) PLQY (b) ΔE_{ST} (c) -LUMO energy of HTL1 (d) -HOMO energy of ETL2.
We have then employed an SVM-based regression method with a polynomial kernel of degree 3. The polynomial kernel is expressed as,

\[ K(x, x') = (\gamma(x, x') + r)^d \]

where d is the degree of the polynomial.

Table 1: Performance metrics of SVR, DT and K-NN

<table>
<thead>
<tr>
<th>ML Model</th>
<th>Reason for choice</th>
<th>(r^2) score(_{train})</th>
<th>(r^2) score(_{test})</th>
<th>RMSE(_{train})</th>
<th>RMSE(_{test})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support Vector Regression (SVR)</td>
<td>Polynomial Regression</td>
<td>0.63±0.05</td>
<td>0.34</td>
<td>4.72±0.41</td>
<td>5.09</td>
</tr>
<tr>
<td>Decision Tree (DT)</td>
<td>Logic based approach</td>
<td>0.41±0.15</td>
<td>-0.06</td>
<td>5.99±1.05</td>
<td>6.46</td>
</tr>
<tr>
<td>K-NN (K=3) regression</td>
<td>Nonparametric method</td>
<td>0.39±0.04</td>
<td>0.62</td>
<td>6.08±0.54</td>
<td>3.88</td>
</tr>
</tbody>
</table>

With SVR, we have got a cross-validated \(r^2\) score of 0.63±0.04 for train set with a mean RMSE of 4.59, while for test set, we obtained \(r^2\) score of 0.34 with RMSE of 5.09. It is quite clear that it highly overfits and thus performs very poorly for the test set. Although, in the high EQE\(_{max}\) region, it performs well. However, in the low EQE\(_{max}\) region, it fails quite miserably, even predicting a negative EQE\(_{max}\) value, as evident from the Fig. S2. To extract further learning capability of SVR with more data, we have performed learning curve analysis. From Fig. S2, we can see that the improvement of the model stopped after 70 training examples and it reached the saturation. So, further improvement with increasing data points is not possible.

Figure 3. Schematic Representations of the variations of ensemble-based learning, their nature and used model under each class in this work
In fact, the DT has also shown poor performance with the cross-validated $r^2$ score of 0.41 with a high variance of 0.15 for train set, while for test set it is -0.06 (Fig. S3(a)). Optimized DT has a maximum depth of 2. Because of such structure of the DT, outputs are parallel to X-axis with 4 distinct values (mean value of the samples of each leaf node). Basically, based on a given condition, a constant value has been assigned in each case. Such poor performance is also reflected in the RMSE values of 5.98±1.05/6.46 for train and test set. With max. depth of 2, it has been decisive based on two input parameters: PLQY and LUMO energy of HTL1 with feature importance of 0.89 and 0.11 (Fig. S3(b)). Now moving towards K-NN, we have scanned the optimum number of neighbors for regression and with neighbors=3, the model has performed the best (Fig. S4(b)). With train set, it has produced cross-validated $r^2$ score of 0.39±0.04 and 0.62 for test set and mean RMSE of 6.08±0.54/3.88 over train/test set.

As, we can see, all the above three models could not perform up to the expectations. We thus moved towards ensemble-based learning, where we have used DT as a base learner (weak learner) model and through repetitive training, we have got better performance from these ensemble-based learning models. Figure 3 shows a schematic representation of the variations of ensemble-based learning. Generally, there are two major types of ensemble-based learning methods: (1) Bagging (2) Boosting. Bagging is parallel in nature and RF is an example of a bagging algorithm. With RF, we have got cross-validated $r^2$ score of 0.65 with a high variance of 0.12 for train set and for test, set we have got very a high $r^2$ score of 0.74, which corresponds to mean RMSE of 4.59±0.37/3.18 for train/test set. To get an insight into the employed model, we have plotted feature importance plot. It selects PLQY and ΔEST as the most important features with scores of 0.28 and 0.10 respectively. Note that, it has shown very little acceleration in the learning curve. In fact, the $r^2$ score for validation has improved by 0.02 going from 70 to 78 training examples.

![Figure 4](image)

Figure 4 Comparison of GB, XGB, AdaBoost, LightGBM and RF with (a) $r^2$ score (b) RMSE. Blue bar indicates train set result (cross-validated); orange bar indicates test set result. Black line indicates the variance.

Furthermore, we have focused on boosting algorithms and have employed GB, AdaBoost, XGB and LightGBM model (Fig. S5-S8). Among them, GB has performed the best even surpassing the XGB efficiency (Figure 4). GB has provided a cross-validated $r^2$
Figure 5: Train/ test metrics (r^2 score and RMSE) of GB Model. For train set, the evaluated metrics is the mean of cross-validation score of 0.71, with a variance of 0.04 for train set. While for the test set, the r^2 score is 0.84. The model has also produced a low RMSE of 4.22±0.55/2.53 for train/test set (Figure 5). A comparison of our work with related ML-based works on OLEDs is shown in Figure 7, which shows that RMSE value for GB is 4.22 which is very close to previously reported 4.16. The best part is our model is not limited to green TADF OLEDs rather general to all emissive range. Interestingly, XGB has performed with r^2 score of 0.70±0.03/0.70 and RMSE of 4.30±0.46/3.44 for train/test set (Fig. S8).

Generally, models with high performance are considered as black box. Therefore, we have performed again feature importance analysis for interpretability of the employed models. Feature importance score examines the effect of a certain feature on the reduction in the criterion used to split nodes in DTs. In the case of XGB, PLQY has the highest score of 0.25 and -LUMO energy of HTL1 has the second-highest score of 0.10. While -LUMO energy of ETL1 and -HOMO energy of ETL2 ranks 3rd and 4th with scores of 0.097 and 0.078 respectively. Similarly, for GB, PLQY ranks first with feature importance score of 0.54 and -LUMO energy of GUEST being 2nd with a score of 0.07 (Figure 6(a)). -LUMO energy of GUEST has positive correlation with the output, which indicates a lower LUMO energy is good for EQE_{max}. A low LUMO energy is good to hinder the reverse exciton
transfer to the HOST and ceasing further dissipation of exciton. $\Delta E_{\text{ST}}$ ranks 3rd with feature importance score of 0.054. Lower $\Delta E_{\text{ST}}$ obviously guarantees an efficient rISC, which is important for better EQE$_{\text{max}}$. Apart from good performance, it still has better potential. Learning curve for both GB and XGB is still ascending. The validation for $r^2$ score has increased by 0.06 (0.65 to 0.71) when the number of training examples increased from 70 to 78 (Fig. 6(b)) for GB. The main point of this work is that with so many models employed, if one must rely on the best score among the employed models, learning curve for GB has the highest potential among the models. However, training more examples is beyond the scope of this work.

![Figure 7: Comparison of our work related to organic photovoltaic system (a) r2 score (b) RMSE](image)

### 4. Conclusions

In this work, we have developed ML models for the prediction of EQE$_{\text{max}}$ in TADF-based OLEDs, which are general to any TADF-based chromophore irrespective of its emission range and selective to device architecture. We have created a database of 123 TADF devices consisting of 30 features, which can be obtained from experiments (devoid of any further theoretical calculations). Initial data analysis has shown that among the features PLQY, $\Delta E_{\text{ST}}$, -LUMO energy of HTL1 and -HOMO energy of ETL2 manifested a high correlation with EQE$_{\text{max}}$, with $|r| > 0.20$. We have employed very simple machine learning models involving SVR, K-NN and DT. These models have failed miserably to predict any better output. Subsequently, we have employed ensemble-based learning models, which work on the gradual improvement of the weak learning models like DT. We have used both bagging (RF) and boosting (GB, AdaBoost, XGB, LightGBM) algorithms. Note that, while the bagging is parallel in nature and the boosting works on a sequential basis. Finally, among the employed ensemble-based models, GB has performed the best with $r^2$ score of $0.71 \pm 0.04/0.84$ and RMSE of $4.22 \pm 0.55/2.53$ for train/test set, which is quite good with respect to the current state-of-the-art (SOTA). Previously predicted model had a RMSE value of 4.16, but that model was selective to green TADF based OLEDs. Being a tree-based model, it has given high importance to PLQY and -LUMO energy of GUEST with feature importance scores of 0.54 and 0.07. Even the learning curve is still upwards, which proves the model can perform better with more train data points, however, that is beyond the scope of this work. The main advantage of this work is that the
experimentalists can easily deploy this model with features readily accessible to them with accessible cloud computation facility and further help the models by improving the database. This work will surely be a huge assistance in the field of TADF-based OLEDs.

5. Supplementary Materials
Supplementary materials contain the figures for full correlation heatmap, SVR, DT, K-NN, RF, AdaBoost and LightGBM, XGB Results and their descriptions (pdf).

6. Acknowledgement
BCG thanks CSIR and JNCASR for fellowship and SKP acknowledges research support from the J. C. Bose Fellowship, Science and Engineering Research Board (SERB), the Government of India.

ORCID
Bidhan Chandra Garain: 0000-0003-2258-0002
Prof. Swapan K. Pati: 0000-0002-5124-7455

Corresponding Author
Prof. Swapan K. Pati: pati@jncasr.ac.in / swapan.jnc@gmail.com

Conflicts of Interest
The authors declare no competing financial interest.

References
18. C. Duan, J. Li, C. Han, D. Ding, H. Yang, Y. Wei and H. Xu, *Chemistry of Materials*, 2016, **28**, 5667-5679.


Graphic TOC