



# A Machine Learning Approach for Metal Oxide Based Polymer Composites as Charge Selective Layers in Perovskite Solar Cells

Murat Onur Yildirim,<sup>[a]</sup> Elif Ceren Gok,<sup>[a]</sup> Naveen Harindu Hemasiri,<sup>[b]</sup> Esin Eren,<sup>[d, e]</sup> Samrana Kazim,<sup>[b, c]</sup> Aysegul Uygun Oksuz,<sup>[\*, e]</sup> and Shahzada Ahmad<sup>[\*, [b, c]</sup>

A library of metal oxide-conjugated polymer composites was prepared, encompassing WO<sub>3</sub>-polyaniline (PANI), WO<sub>3</sub>-poly(N-methylaniline) (PMANI), WO<sub>3</sub>-poly(2-fluoroaniline) (PFANI), WO<sub>3</sub>-polythiophene (PTh), WO<sub>3</sub>-polyfuran (PFu) and WO<sub>3</sub>-poly(3,4-ethylenedioxythiophene) (PEDOT) which were used as hole selective layers for perovskite solar cells (PSCs) fabrication. We adopted machine learning approaches to predict and compare PSCs performances with the developed WO<sub>3</sub> and its composites. For the evaluation of PSCs performance, a decision tree model

that returns 0.9656 R<sup>2</sup> score is ideal for the WO<sub>3</sub>-PEDOT composite, while a random forest model was found to be suitable for WO<sub>3</sub>-PMANI, WO<sub>3</sub>-PFANI, and WO<sub>3</sub>-PFu with R<sup>2</sup> scores of 0.9976, 0.9968, and 0.9772 respectively. In the case of WO<sub>3</sub>, WO<sub>3</sub>-PANI, and WO<sub>3</sub>-PTh, a K-Nearest Neighbors model was found suitable with R<sup>2</sup> scores of 0.9975, 0.9916, and 0.9969 respectively. Machine learning can be a pioneering prediction model for the PSCs performance and its validation.

## Introduction

Perovskite solar cells (PSCs) have been intensively investigated due to their broad light absorption (300–800 nm), high absorption coefficient, long carrier diffusion length, high charge carrier mobility and tuneable band gap.<sup>[1–2]</sup> Organic–inorganic halide PSCs have gained significant attention due to a simple fabrication process and high-power conversion efficiency (PCE).<sup>[3]</sup> The perovskite is represented by a typical formula of ABX<sub>3</sub>, where A is an organic cation such as methylammonium

(MA), formamidinium (FA), etc.; B is a metal (typically Pb); and X is a halogen anion (I, Br, Cl, or a mixture of these).<sup>[3]</sup> MA based PSCs lack desired characteristics of thermal stability, moisture-induced degradation, and hysteretic *I*–*V* behaviour, which limits its applications.<sup>[3]</sup> Mixed cation and anion based perovskites have shown advantageous properties<sup>[3–5]</sup> and mixed cation of formamidinium/methylammonium (FAMA) gave enhanced performance due to an intense light absorption, stability and reduced *J*–*V* hysteresis.<sup>[3]</sup> Further, the incorporation of cesium (Cs) can also efficaciously decrease the crystallization temperature during the annealing process and induce reliability.<sup>[3,5]</sup> Depending on the light incidence, the architect of PSCs can be either *n*–*i*–*p* or *p*–*i*–*n* type, where *n*- and *p*-type are electron and hole selective materials respectively, and *i* denotes the light harvesting layer.<sup>[6]</sup> PSCs with *p*–*i*–*n* planar architecture display negligible hysteresis, solution processability at low temperature, and the potential for scaleup using a continual coating method.<sup>[7]</sup> Fullerenes type acceptor, especially [6,6]-phenyl-C(61)-butyric acid methyl ester (PCBM), are typically used as *n*-type charge transport layer,<sup>[7]</sup> while poly(3,4-ethylenedioxythiophene):poly(styrene sulfonate) (PEDOT:PSS) as a transparent hole transport material in lieu of Spiro-OMeTAD, was employed.<sup>[8]</sup> However, PEDOT:PSS possesses undesirable features such as hygroscopic nature, inferior thermal stability and inability to block electrons,<sup>[8]</sup> thus an effective hole transport layer (HTL) is paramount. Cogal *et al.* reported graphene-based polymer composites as HTL.<sup>[9]</sup> Recently, transition metal oxides such as WO<sub>3</sub>, owing to its high work functions, high carrier mobility, and excellent thermal stability are being used as HTLs.<sup>[8]</sup> Comprehensive studies on metal oxide based HTLs suggest that hybrid organic-inorganic composites can be promising candidates as injecting carriers from perovskite absorber to electrode.<sup>[10–12]</sup>

[a] M. O. Yildirim, E. C. Gok  
Department of Industrial Engineering  
Engineering Faculty  
Suleyman Demirel University  
32260 Isparta (Turkey)

[b] N. H. Hemasiri, Dr. S. Kazim, Prof. S. Ahmad  
BCMaterials, Basque Center for Materials  
Applications and Nanostructures  
UPV/EHU Science Park, 48940 Leioa (Spain)  
E-mail: shahzada.ahmad@bcmaterials.net

[c] Dr. S. Kazim, Prof. S. Ahmad  
IKERBASQUE  
Basque Foundation for Science  
48009 Bilbao (Spain)

[d] Dr. E. Eren  
Department of Energy Technologies, Innovative Technologies  
Application and Research Center  
Suleyman Demirel University  
32260 Isparta (Turkey)

[e] Dr. E. Eren, Prof. A. U. Oksuz  
Department of Chemistry  
Faculty of Arts and Science  
Suleyman Demirel University  
32260 Isparta (Turkey)  
E-mail: ayseguluygun@sdu.edu.tr



Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cplu.202100132>

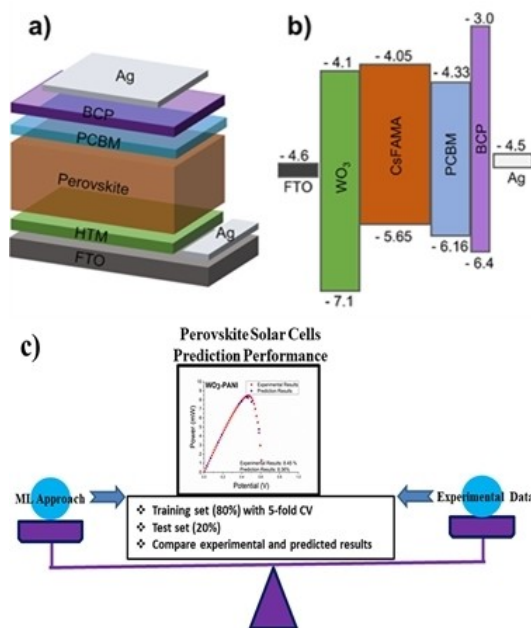


This article is part of a Special Collection on "Perovskite Materials and Devices"

Machine learning (ML) is a set of methods that can acquire the data pattern without explicit programming and predict the imminent data with uncovered patterns.<sup>[13]</sup> It will replace the traditional trial and error method which demands longer time and resources to predict the performance and reliability of PSCs (stability, PCE, fabrication techniques and material synthesis).<sup>[14–16]</sup> For an ML based data-driven approach, it is necessary to couple experimental data with the modelling.<sup>[17]</sup> An extensive amount of data is not required and computing time is fast compared to other complicated models.<sup>[18]</sup> By employing ML technique in solar cells, material properties, optimized device architects and fabrication processes can be predicted, and data reconstruction is attracting significant interest for research and development.<sup>[18]</sup> Reports dealing with PSCs through ML approach are in scarce,<sup>[19]</sup> the majority of prediction of materials entails, electro-optical features, *J*-*V* performance, which are associated with dye-sensitized and organic solar cells.<sup>[18]</sup> Thermodynamic stability of 20,000 randomly selected materials using seven different ML methods were predicted.<sup>[20]</sup> Random forest (RF) model predicts band gap of perovskites by employing 18 descriptors and 10 stable perovskites and band gaps were elucidated.<sup>[21]</sup> Similarly, discovery of materials with an optimal band gap for single junction was predicted.<sup>[22]</sup> 300 octahedral oxyhalides with geometric and electronic data were used to sequence ML model and the band gap predictions were made on 5000 oxyhalides test data.<sup>[23]</sup> ML methods were used to develop a model that contains 333 data points from 2000 scientific articles to guide the designing of new perovskite to predict model performance.<sup>[16]</sup> Using alternating conditional expectations – ML approach, nonlinear mapping between band gap and properties of constituent elements were studied.<sup>[24]</sup> Methodology was presented to predict band gap of undiscovered 5158 hybrid perovskites for PSCs, for this, six different ML regression algorithms were constructed and implemented.

Gradient Boosting Regressor (GBR) achieve higher performance during the train-test process, and was used to predict 5158 band gap values.<sup>[25]</sup> Saidi *et al.* developed a dataset consisting of structural and band gap features of 862 halide perovskites to improve a predictive ML model that receives the complex trends and correlations of this chemical space. This study showed that a well-designed hierarchical ML approach has superior accuracy in predicting the features of halide perovskites. The root-mean-square errors for the lattice constants, octahedral angle and band gap for halide perovskites were calculated as 0.01 Å, 5°, and 0.02 eV, respectively. The hierarchical convolutional neural network (CNN) was also reported as a suitable approach in materials design.<sup>[26]</sup> In another study, a gradient boosting regressor (GBR) ML model was applied onto structural and elemental features for perovskite formation energy prediction. The bigger training set is then utilized to train a convolutional neural network model (the screening model) with the generic Magpie elemental properties with high prediction power. The root mean square error (RMSE), mean absolute error (MAE), *R*<sup>2</sup> for descriptor with Magpie element was 0.11, 0.25, 0.83, respectively. The screening model was used to filter out promising perovskite materials out

of 21,316 hypothetical perovskite structures with a large portion evidenced from the previous literature.<sup>[27]</sup> Automated identification tool via machine learning methods to situate the dominant loss via the light intensity-dependent performances as an input was made. The highest accuracy of the prediction using > 2 million simulations with a RF classifier was obtained > 82 % when utilizing the performance of all of the simulated light intensities and the mobility of the layers. The prediction of the dominant recombination using ML approach can be enhanced by adding the performances under different light intensities, doping, and mobilities.<sup>[28]</sup> Subsequently, machine learning approach has been applied to predict the properties of perovskite (such as phase stability, band gap, electronic transport features, etc.) and designing of innovative perovskite for the development of PSCs.<sup>[29]</sup> Our aim is to put forward a direct appraisal between experimental data and machine learning algorithms for PSCs. Radio frequency (rf) plasma-enhanced method was used to synthesize WO<sub>3</sub>-conjugated polymer composites as charge selective layer. A library of composites including tungsten trioxide-polyaniline (WO<sub>3</sub>-PANI), tungsten trioxide-poly(*N*-methylaniline) (WO<sub>3</sub>-PMANI), tungsten trioxide-poly(2-fluoroaniline) (WO<sub>3</sub>-PFANI), tungsten trioxide-polythiophene (WO<sub>3</sub>-PTH), tungsten trioxide-polyfuran (WO<sub>3</sub>-PFu), tungsten trioxide-poly(3,4-ethylenedioxythiophene) (WO<sub>3</sub>-PEDOT) were prepared using a rotating capacitively-coupled rf plasma process. The composites prepared through rf rotating modification are advantageous due to solvent-free, nontoxic, and well-controlled deposition.<sup>[9,30–32]</sup> The fabricated PSCs architect and energy level diagram using WO<sub>3</sub> as charge selective layer is shown (Figure 1a & b). Four different machine learning algorithms were used to build the models for UV-Vis spectra, *J*-*V* and external quantum efficiency (EQE) spectra prediction. The



**Figure 1.** a) Device architecture, b) energetic diagram of perovskite solar cells using WO<sub>3</sub> as charge selective layer and c) computing of perovskite solar cells predicted with input datasets using machine-learning approach.

values in the dataset are continuous, and the ML algorithms were built on regression models and not on classification. Hereto unreported, we report our findings on composites of WO<sub>3</sub>-conjugated polymer as hole transport layer in triple cation-based PSCs and the validation between experimental data and simulation using the machine learning model (Figure 1c).

## Results and Discussion

### Machine learning model performance

Machine learning was implemented in a systemic manner involving four steps which are (i) defining research objectives for UV-Vis spectra, EQE spectra and *J-V* spectra predictions, (ii) constructing dataset with experimental results, (iii) applying preferred machine learning algorithms and (iv) evaluating models based on the dataset in Table 1 with R<sup>2</sup> and RMSE scores.

Machine learning algorithms were applied on experimental results from seven different composite based films. In the first two models, we input wavelength (nm) to predict absorption and external quantum efficiency (EQE) respectively. Current density (mAcm<sup>-2</sup>) was predicted based on voltage (V) in the third machine learning model. In the constructed dataset, there were no missing values but there were outliers, which may affect the prediction performance. Thus, no imputation was required to fill missing values while outliers were removed from the dataset. Before building the machine learning models, the data was subject to scaled,  $\mu=0$  and  $\sigma=1$  with standard scaler. All the data points subsequently follows each other. Therefore, no shuffling was implemented to avoid misleading data.

For fitting and evaluating models, the dataset was separated into two parts as 80% train set and 20% test set. The five-fold cross validation method was applied on the train set to monitor the overfit and underfit problems and the final evaluation was made with unseen 20% test set which is usually called hold-out set.

For structural characterization, we recorded diffractograms (Figure 2) for the WO<sub>3</sub> and its composites prepared by rf-plasma

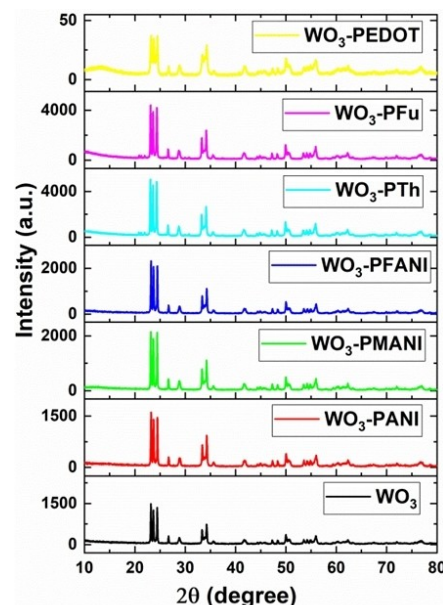


Figure 2. X-ray diffractograms for different WO<sub>3</sub> and its composites

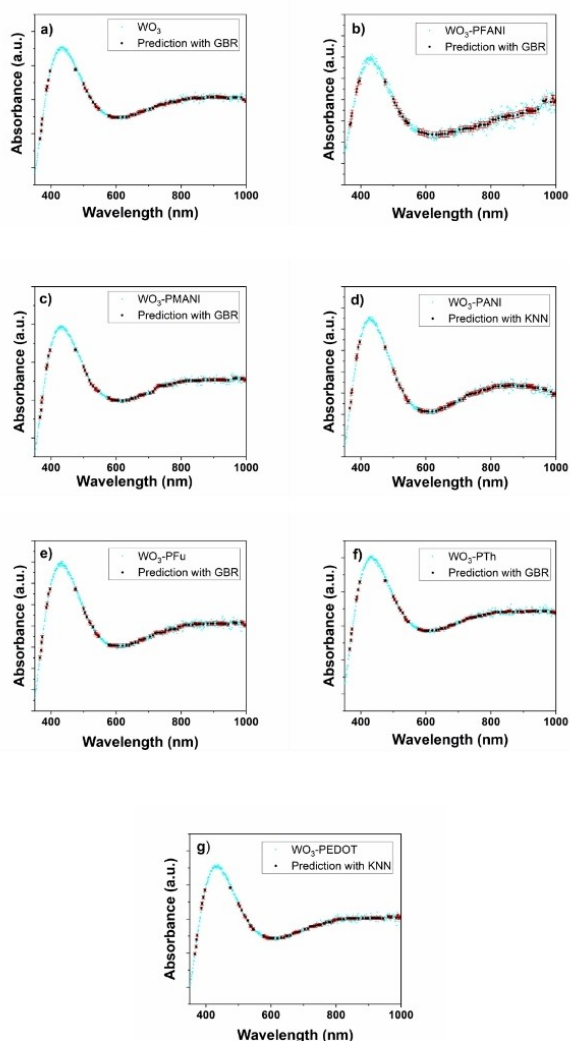
coupling. XRD pattern of WO<sub>3</sub> confirms monoclinic structure of WO<sub>3</sub> (PCPDS No. 00-024-0747). The diffraction peaks that belong to [020], [200], [120], [112], [022], [220], [122], [312], [040], [232], [402] of monoclinic phases are observed at  $2\theta = 23.58, 24.38, 26.59, 28.93, 33.26, 34.18, 35.65, 46.04, 48.24, 50.21, 56.11$ . XRD pattern of the WO<sub>3</sub> composites display similar diffractions, and due to the binding of polymer with WO<sub>3</sub>, the peak at 35°, shows increase in the intensity.

To evaluate the electrical properties of WO<sub>3</sub> and its composites, we measured electrical conductivity at room temperature by four-probe method. Variation in the conductivity value was noted due to the formation of composites and the pristine WO<sub>3</sub> gave the lowest conductivity value ( $7 \times 10^{-5}$  S/cm) among the samples, while the WO<sub>3</sub>-PANI based composites showed highest conductivity value ( $36 \times 10^{-5}$  S/cm). WO<sub>3</sub>-PMANI, WO<sub>3</sub>-PFANI, WO<sub>3</sub>-PTh, WO<sub>3</sub>-PFu and WO<sub>3</sub>-PEDOT based composites measured 29, 3, 13, 2,  $24 \times 10^{-5}$  S/cm respectively. The conductivity values are in accordance with the device photovoltaic properties, and WO<sub>3</sub>-PANI gave higher performance among them. This we ascribed due to the high charge carrier concentration built in the devices, which increase the performance.

Figure 3 displays UV-Vis absorption spectra and the bands at ca. 430 and 800 nm can be attributed to the doping level and formation of polarons and bipolarons.<sup>[33–34]</sup> The WO<sub>3</sub> based composites gave a similar response and resembled that of pristine WO<sub>3</sub> film due to the deposition of very thin polymer coating onto WO<sub>3</sub> particles during plasma polymerization process.<sup>[30–31]</sup>

We noted higher performance for the GBR and KNN than the rest of the R<sup>2</sup> methods (Table 2). Gradient boosting gave good results for WO<sub>3</sub>, WO<sub>3</sub>-PFANI, WO<sub>3</sub>-PMANI, WO<sub>3</sub>-PFu and WO<sub>3</sub>-PTh with 0.971684, 0.903560, 0.968168, 0.961903 and 0.973991 R<sup>2</sup> scores respectively. WO<sub>3</sub>-PANI and WO<sub>3</sub>-PEDOT

Table 1. Dataset Description				
Materials		UV	<i>J-V</i>	EQE
WO <sub>3</sub>	Train size	570	50	121
	Test size	114	10	25
PFU	Train size	570	18	121
	Test size	114	4	25
PFANI	Train size	570	39	121
	Test size	114	8	25
PTh	Train size	570	50	121
	Test size	114	10	25
PMANI	Train size	570	50	121
	Test size	114	10	25
PANI	Train size	570	47	121
	Test size	114	10	25
PEDOT	Train size	570	31	121
	Test size	114	7	25
Input		Wavelength	Potential	Wavelength
Output		Absorbance	Current density	EQE



**Figure 3.** UV-Vis spectra of  $\text{WO}_3$  and its composites based films using the best machine learning models.

were well predicted by KNN with the scores of  $R^2$  0.965050 and 0.970469 respectively. The predictions of best models fit into

the original dataset for each PSCs and can be deduced from Figure 3.

### EQE response prediction for perovskite solar cells based on $\text{WO}_3$ and its composites

Figure 4 illustrates the experimental and machine learning based external quantum efficiency (EQE) spectrum of triple cation based PSCs using  $\text{WO}_3$  and its composites-as HTL. It can be deduced (Figure 4), the overall EQE response of the fabricated PSCs are in the agreement with machine learning approach. We noted  $>80\%$  conversion in the EQE spectrum in the range from 400–700 nm for the  $\text{WO}_3$ -based composites as hole selective layers in PSC. Among the  $\text{WO}_3$  composites HTL-based PSCs, the higher EQE values were represented by  $\text{WO}_3$ -PANI,  $\text{WO}_3$ -PTh-based PSCs that is supported by its higher photovoltaics performance.

In contrast to the UV prediction, a single machine learning approach of random forest outperformed for all PSCs while making predictions on EQE. Each decision tree has a high variance, but low bias. Random forest, which is consist of various decision trees that are trained by a different portion of the train set, corroborates its flexibility and data adaption ability on EQE data. It constitutes a well-balanced model in terms of bias-variance trade-off. The  $R^2$  scores (Table 3), were 0.997959, 0.996145, 0.995918, 0.996925, 0.994619, 0.996411 and 0.996198 for  $\text{WO}_3$ ,  $\text{WO}_3$ -PFANI,  $\text{WO}_3$ -PMANI,  $\text{WO}_3$ -PANI,  $\text{WO}_3$ -PFu,  $\text{WO}_3$ -PTh and  $\text{WO}_3$ -PEDOT, respectively.

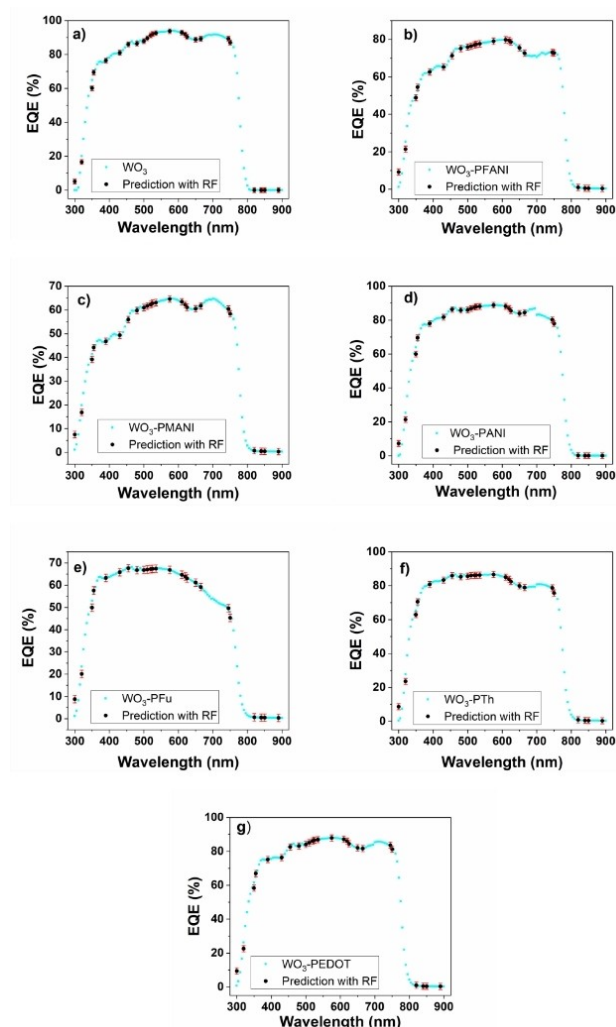
### Prediction of J-V curves based on $\text{WO}_3$ and its composites as HTL

Figure 5 shows the comparison between the machine learning approach and experimental J-V curves of PSCs based on  $\text{WO}_3$  and its composites. To evaluate the PSCs performance prediction (Table 4), we input data points from experimental results. The theoretical predictions from machine learning methods with high accuracy were supported by the experiment results.

**Table 2.** UV-Vis prediction performance of machine learning models for  $\text{WO}_3$  and its composites.

Material	Model	$R^2$	RMSE	Material	Model	$R^2$	RMSE
$\text{WO}_3$	KNN	0.968754	0.001263	$\text{WO}_3$ -PFu	KNN	0.954203	0.001323
	DT	0.959869	0.001431		DT	0.947120	0.001421
	RF	0.963235	0.001370		RF	0.949759	0.001385
	GBR	0.971684	0.001202		GBR	0.961903	0.001206
$\text{WO}_3$ -PFANI	KNN	0.896149	0.001082	$\text{WO}_3$ -PTh	KNN	0.969205	0.001259
	DT	0.833142	0.001374		DT	0.957356	0.001482
	RF	0.871240	0.001207		RF	0.963582	0.001370
	GBR	0.903560	0.001044		GBR	0.973991	0.001157
$\text{WO}_3$ -PMANI	KNN	0.964657	0.001196	$\text{WO}_3$ -PEDOT	KNN	0.970469	0.001188
	DT	0.954171	0.001362		DT	0.959529	0.001391
	RF	0.954120	0.001363		RF	0.961290	0.001360
	GBR	0.968168	0.001135		GBR	0.970416	0.001189
$\text{WO}_3$ -PANI	KNN	0.965050	0.001230				
	DT	0.950332	0.001467				
	RF	0.953313	0.001422				
	GBR	0.960530	0.001307				





**Figure 4.** EQE analysis of PSCs based on a)  $\text{WO}_3$  b)  $\text{WO}_3$ -PFANI c)  $\text{WO}_3$ -PMANI d)  $\text{WO}_3$ -PANI e)  $\text{WO}_3$ -PFu f)  $\text{WO}_3$ -PTh and g)  $\text{WO}_3$ -PEDOT using the best machine learning models and experimental results.

The highest PCE was displayed by  $\text{WO}_3$ -PANI composites and its derivatives, and 8.45% of PCE was measured, which is in

accordance with the  $\text{WO}_3$ -PANI electrical properties. The enhanced PCE performance was ascribed due to its high fill factor (FF) values owing to faster charge transfer.<sup>[35]</sup> PANI with intriguing hole extraction feature provide significant contribution onto PSCs based on  $\text{WO}_3$  composites. The introduction of methyl as electron-donating group or fluorine as electron-withdrawing group influences the performance of PSCs including  $\text{WO}_3$  composites with PANI and its derivatives. Variation in the performance of PSCs with  $\text{WO}_3$  composites can be ascribed to the difference in electron-withdrawing features and aromaticity behaviors of conjugated polymers as donor types in  $\text{WO}_3$  composites. The low performance for  $\text{WO}_3$ -PFu-based PSCs was displayed due to the poor interface, which inhibits the exciton diffusion leading to charge recombination pathways.<sup>[36]</sup>

Before comparing ML results with P-V characteristics of PSCs, train and test scores were monitored to observe if any overfitting occurred. In Table S1, train and test scores of all the models are given and no overfitting is observed. On overfitting, train scores show highly accurate results although test scores suffer from low accuracies.<sup>[37]</sup>

Table 5 shows that the decision tree is the rational approach for  $\text{WO}_3$ -PEDOT, while random forest dominates on  $\text{WO}_3$ -PFANI with 0.996836,  $\text{WO}_3$ -PMANI with 0.997616 and  $\text{WO}_3$ -PFu with 0.977237  $R^2$  score on current density prediction. KNN, on the other hand, estimated the PSCs based on  $\text{WO}_3$ ,  $\text{WO}_3$ -PANI and  $\text{WO}_3$ -PTh better with  $R^2$  scores of 0.997520, 0.991630 and 0.996970, respectively.

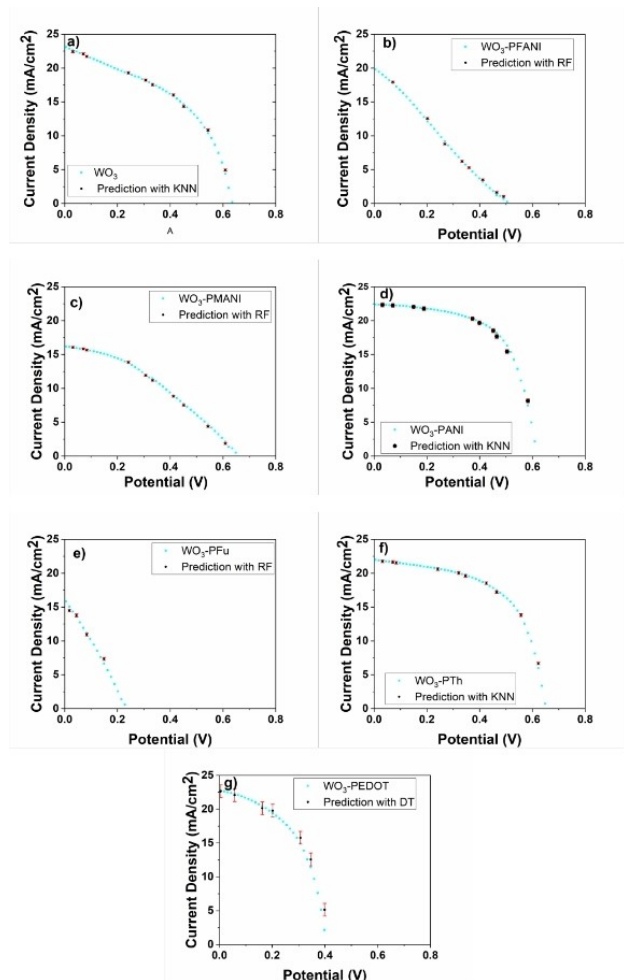
The power-voltage characteristics estimated from both experimentally and theoretically displayed similar trend as of  $J$ - $V$  characteristics. The output power increases linearly in the low voltage range, achieved maxima and then shows a steep decline at the higher voltage. Similar to the PCE, the power-voltage characteristics (Figure 6) display a point of maximum output power.<sup>[38]</sup>

## Conclusion

We report a library of  $\text{WO}_3$  and its composites prepared by using rotating radio frequency plasma-method. The derived

**Table 3.** EQE spectrum prediction performance of machine learning models under  $\text{WO}_3$  and its composites.

Material	Model	$R^2$	RMSE	Material	Model	$R^2$	RMSE
$\text{WO}_3$	KNN	0.990173	3.605384	$\text{WO}_3$ -PFu	KNN	0.977822	3.822153
	DT	0.995983	2.305150		DT	0.993764	2.026741
	RF	0.997959	1.643105		RF	0.994619	1.882729
	GBR	0.996133	2.261715		GBR	0.993904	2.003832
$\text{WO}_3$ -PFANI	KNN	0.982739	3.955444	$\text{WO}_3$ -PTh	KNN	0.980721	4.653979
	DT	0.995269	2.070804		DT	0.994656	2.450234
	RF	0.996145	1.869365		RF	0.996411	2.008077
	GBR	0.995239	2.077400		GBR	0.994693	2.441717
$\text{WO}_3$ -PMANI	KNN	0.984498	3.012688	$\text{WO}_3$ -PEDOT	KNN	0.983318	4.349175
	DT	0.995880	1.553186		DT	0.995432	2.275811
	RF	0.995918	1.546000		RF	0.996198	2.076264
	GBR	0.995902	1.548957		GBR	0.995526	2.252323
$\text{WO}_3$ -PANI	KNN	0.984992	4.223487				
	DT	0.995445	2.326741				
	RF	0.996925	1.911781				
	GBR	0.995383	2.342688				



**Figure 5.** Performance of the best machine learning models compared to measured *J-V* curves of PSCs based on a)  $\text{WO}_3$ , b)  $\text{WO}_3$ -PFANI c)  $\text{WO}_3$ -PMANI d)  $\text{WO}_3$ -PANI e)  $\text{WO}_3$ -PFu, f)  $\text{WO}_3$ -PTh and g)  $\text{WO}_3$ -PEDOT.

composites were then employed as hole transport materials for perovskite solar cell fabrication. The influence of  $\text{WO}_3$  and its composites on the photovoltaic performance was validated with machine learning methods. Machine learning approaches

are paramount to uncover concealed patterns and decipher the relationships of the variables. By adopting the modeling of UV-Vis spectra, performance-related curves such as *J-V* and EQE can be deducted to validate solar cell performance. We demonstrated pathways to build machine learning models for solar cell parameters such as EQE, *J-V*, and UV prediction. This will reduce associated laboratory cost and the presented ML models will accelerate the advancement in perovskite solar cells. Four machine-learning methodologies, from simple to complex, were adopted to predict three key characteristics. K nearest neighbors and gradient boosting gave a better estimation of UV data while K nearest neighbors, random forest and decision tree computed successful *J-V* measurements. Furthermore, random forest was noted to be suitable and outperformed other models for EQE prediction. During the modelling, five-fold cross validation was applied on 80% training data set and evaluation of the models made on 20% test data. To calculate the power output of solar cells, both experimental and predicted *J-V* datasets were utilized. Our predicted results are in agreement with the experimental results, and validate the potential of machine learning models for predicting performance with promising  $R^2$  and RMSE scores. Arguably, machine learning models can drastically reduce experimental process, workforce, and associated material design expenses with quick performance prediction of perovskite solar cells and charge selective layers.

## Experimental Section

Machine learning (ML) model was developed for PSCs fabricated with  $\text{WO}_3$  and its composites ( $\text{WO}_3$ -PANI,  $\text{WO}_3$ -PMANI,  $\text{WO}_3$ -PFANI,  $\text{WO}_3$ -PTh,  $\text{WO}_3$ -PFu,  $\text{WO}_3$ -PEDOT) as charge selective layers. These composites were prepared using single-step radio frequency (rf) rotating plasma method.<sup>[30–32]</sup>

## Perovskite solar cells fabrication

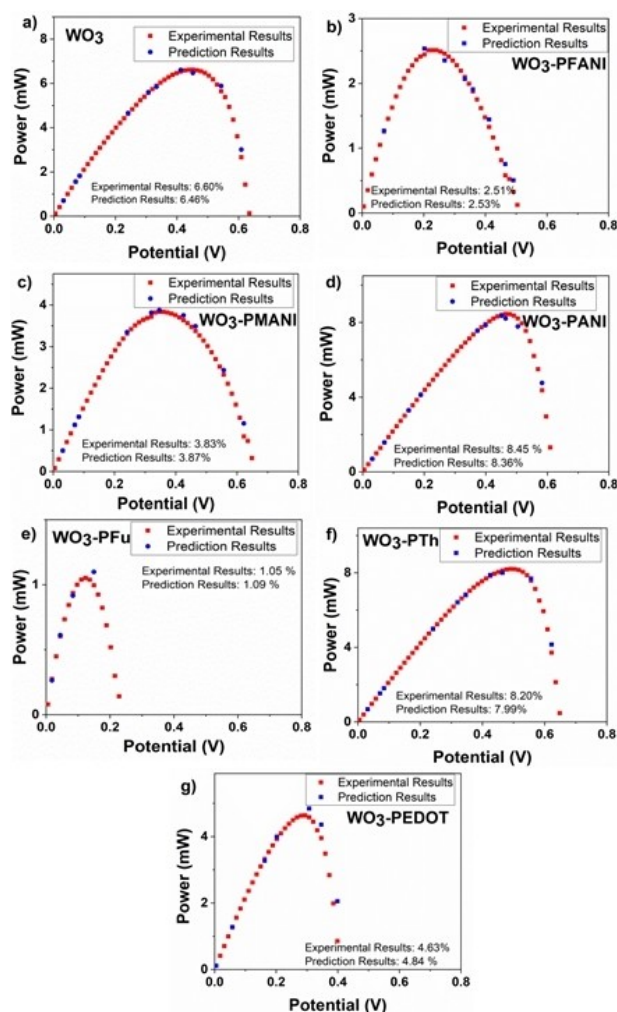
**Materials:** The perovskite precursors were purchased from Dyesol except  $\text{PbI}_2$  and  $\text{CsI}_2$  procured from Tokyo Chemical Industry (TCI) and were employed as such. [60]PCBM > 99.5% and Bathocuproine (BCP) were purchased from Solenne BV and TCI respectively.

**Table 4.** *J-V* Curve prediction performance of machine learning models under  $\text{WO}_3$  and its composites.

Material	Model	$R^2$	RMSE	Material	Model	$R^2$	RMSE
$\text{WO}_3$	KNN	0.997520	0.000024	$\text{WO}_3$ -PFu	KNN	0.873937	0.000102
	DT	0.981538	0.000066		DT	0.931167	0.000075
	RF	0.996859	0.000027		RF	0.977237	0.000043
	GBR	0.981542	0.000066		GBR	0.931164	0.931164
$\text{WO}_3$ -PFANI	KNN	0.993196	0.000040	$\text{WO}_3$ -PTh	KNN	0.996970	0.000023
	DT	0.990572	0.000048		DT	0.973945	0.000068
	RF	0.996836	0.000028		RF	0.996301	0.000026
	GBR	0.990569	0.000048		GBR	0.973931	0.000068
$\text{WO}_3$ -PMANI	KNN	0.996967	0.000024	$\text{WO}_3$ -PEDOT	KNN	0.813980	0.000267
	DT	0.993357	0.000036		DT	0.965626	0.000115
	RF	0.997616	0.000021		RF	0.938070	0.000154
	GBR	0.993350	0.000036		GBR	0.965616	0.000115
$\text{WO}_3$ -PANI	KNN	0.991630	0.000035				
	DT	0.957747	0.000079				
	RF	0.987814	0.000042				
	GBR	0.957777	0.000079				

**Table 5.** PCE value derived from experimental data and predicted from machine learning approach for WO<sub>3</sub> and its composites as HTL in PSCs.

Samples	PCE from experimental data [%]	PCE from ML Model [%]
WO <sub>3</sub>	6.60	6.46
WO <sub>3</sub> -PANI	8.45	8.36
WO <sub>3</sub> -PMANI	3.83	3.87
WO <sub>3</sub> -PFANI	2.51	2.53
WO <sub>3</sub> -PTh	8.20	7.99
WO <sub>3</sub> -PFu	1.05	1.09
WO <sub>3</sub> -PEDOT	4.63	4.84

**Figure 6.** A comparison between the machine learning simulation and measured *P-V* characteristics of PSCs based on a) WO<sub>3</sub> b) WO<sub>3</sub>-PFANI c) WO<sub>3</sub>-PMANI d) WO<sub>3</sub>-PANI e) WO<sub>3</sub>-PFu f) WO<sub>3</sub>-PTh and g) WO<sub>3</sub>-PEDOT.

Clevios PV Al 4083 PEDOT:PSS was acquired from Heraeus Germany and used after filtration through a 0.45  $\mu$ m PVDF filter.

**HTM solution preparation:** WO<sub>3</sub> and WO<sub>3</sub>/polymer solution was carried out by stirring 5 mg of materials in 1 ml of pure ethanol for around 6 h and then ultra-sonicating for 2–3 min. The solutions were then subjected to ultracentrifugation for 15 min at 3000 rpm and the supernatant was collected and filtered using 0.45  $\mu$ m PTFE filter for device fabrication.

**Device fabrication:** The laser-etched FTO-coated glasses (TEC15) were used as substrates, were thoroughly cleaned prior to device fabrication via ultra-sonication with Hellmanex II solution for 30 min, washing with deionized water and ethanol followed by ultrasonication in acetone, ethanol and isopropanol for 20 min each step. The substrates were dried using compressed air and treated by UV-ozone for 30 min prior to use. The filtered WO<sub>3</sub> and WO<sub>3</sub>/polymer composite solutions were spin-coated atop of FTO substrates at 3000 rpm for 30 s with 1500 rpm/s acceleration, annealed at 120 °C for 5 min, and repeated with the same spin coating parameters followed by annealing at 120 °C for 30 min. After cooling down to 80 °C the samples were transferred (to avoid moisture absorption to the HTM layer) to the argon-filled glovebox under controlled moisture and oxygen conditions (H<sub>2</sub>O level: < 1 ppm and O<sub>2</sub> level: < 10 ppm). Triple-cation-perovskite precursor solution was prepared containing CsI (0.10 M), FAI (1.05 M), PbI<sub>2</sub> (1.24 M), MABr (0.12 M) and PbBr<sub>2</sub> (0.12 M) in an anhydrous solvent mixture of *N,N*-dimethylformamide (DMF) and dimethylsulfoxide (DMSO) with 4:1 vol ratio. The perovskite precursor solution was then spin-coated in a two-step spin-coating program set at 1000 rpm and 6000 rpm for 10 and 30 s, respectively. During the second step of spinning, 112  $\mu$ L of chlorobenzene was dripped at 10 s before ending the program followed by annealing at 100 °C for 1 h for perovskite crystallization. After cooling down to room temperature ETM layer was deposited atop of the perovskite layer using 12 mg/ml PCBM in chlorobenzene at 1200 rpm for 20 s and vacuum dried for 5 min. Finally bathocuproine (BCP) was deposited atop of ETM layer at 5000 rpm for 40 s followed by evaporating Ag (80 nm, < 1 Å/s) in a thermal evaporator under low vacuum conditions (10<sup>-7</sup> Torr). For control devices PEDOT:PSS was spin-coated on FTO substrates at 5000 rpm for 30 s and annealed at 125 °C for 20 min followed by transferring to the glove box at 80 °C during the cooling process.

**Device characterization:** Current density–voltage (*J-V*) curves were performed using an AAA Oriel solar simulator (Newport) producing 1 sun AM1.5G, and were recorded by applying an external potential bias to the devices. The generated photocurrent was recorded at a scan rate of 10 mV/s (pre-sweep delay: 10 s) with the help of Keithley 2400 source meter and 0.09 cm<sup>2</sup> black metal mask as the active area. IPCE measurements were carried out using a 150 W xenon lamp attached to a Bentham PVE300 motorized 1/4 m monochromator.

The electrical conductivity of WO<sub>3</sub> and its composites was measured in a standard four-probe method using PCIDAS6014 current source, a voltmeter and a temperature controller at room temperature. Dry powders were pressed into pellets using a steel die having 13-mm diameter in a hydraulic press under a pressure of 700 MPa. X-ray diffraction (XRD) analyses were made in powder form and performed on Bruker D8 Advance diffractometer with CuK $\alpha$  radiation ( $\lambda$ : 1.54 Å).

## Prediction models

Machine learning can be classified into three main categories such as supervised, unsupervised and reinforcement learning.<sup>[13]</sup> Supervised learning is applied when the specific target labels are known. Unsupervised learning, however, clusters the data since there is no information about the target labels in the dataset. Reinforcement learning utilizes punishment and reward system to improve the model over the iterations.

We adopted supervised learning roof since the target values, which are continuous and known. K Nearest Neighbours (KNN), Decision Tree (DT), Random Forest (RF) and Gradient Boosting (GBR), are briefly represented (Figure 7) as regressors and are preferred

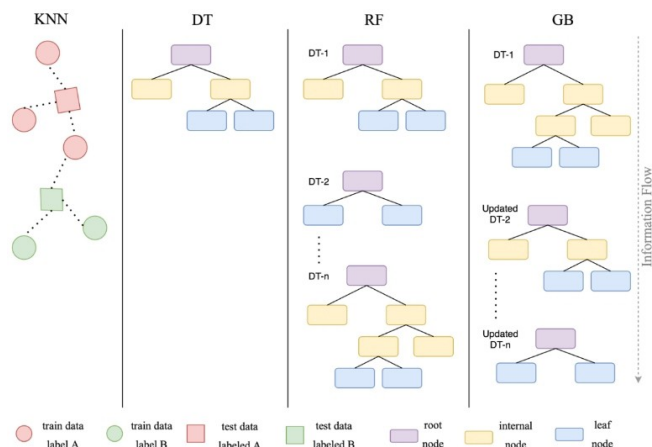


Figure 7. Illustration of machine learning models.

machine learning algorithms. KNN and GBR were selected to unravel the influence of model complexity on the predicted results. Besides these algorithms, tree-based structures which represent flexibility, are simple but can be complex when having various hyperparameters, were applied. KNN algorithms work on defined  $k$  numbers of nearest neighbours to make prediction about the output.<sup>[39]</sup> Although it is one of the simplest machine learning approaches, it performs well since its working principle is based on similar observations and behaves in a similar way. Observations which need to be predicted in the test data set receive a class label according to the  $k$  value and assigning a proper  $k$  value is essential to avoid over-fit and under-fit problems in the model. DT creates little subspaces which are not overlapped and these regions are represented by the corresponding nodes.<sup>[39]</sup> To shorten the learning process and achieve a high-accuracy tree, it's critical to decide which attribute in the data will serve as the root node and inner node of the tree. Features that contribute most to the model selected on initial nodes to generate proper trees. Information spread through internal nodes with basic Yes/No or True/False questions where the tree starts with the root node and finalizes with the leaf nodes. Combination of  $N$  number of trees which is user-defined creates the RF where each tree has a single vote on an input vector.<sup>[41]</sup> It has a lot more modular framework and a function to improve model efficiency due to its structure as it incorporates the basic structure of decision trees. Each tree in the algorithm is made up of randomly chosen data. Subsequently, each tree is unique in its own right and makes its own decisions. After acquiring the training data to construct a forest structure, the Random Forest algorithm attempts the test data for each tree structure it generates. Final prediction value is determined after the majority voting process. Another tree based yet complex algorithm is GBR, it experienced by each tree it built. The logic behind this algorithm is to achieve higher accuracy after each iteration by reducing the loss function.<sup>[41]</sup> The previous tree is used to create the next tree structure by reducing the error rate. The algorithm continues to run until it reaches the defined error rate or the number of iterations.

### Performance metrics

$R^2$  and Root Mean Squared Error (RMSE) are known and applicable metrics to evaluate the performance of regression models. There are some key differences between these two metrics. RMSE gives equal attention to each data point where  $R^2$  is more sensitive to the outliers while calculating the prediction error. Equations respect to  $R^2$  and RMSE are given in [Equation (1)] and [Equation (2)].

$$RMSE = \left( \frac{1}{n} \sum_{i=1}^n (y_i - y_{pred})^2 \right)^{1/2} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_{pred})^2}{\sum_{i=1}^n (y_i - y_{mean})^2} \quad (2)$$

Instead of traditional train and test split, the data set is divided into  $K$  pieces and while  $K-1$  train set is created in each iteration in  $k$ -fold cross validation which promotes objectivity. After deriving the performance metrics for each iteration, the final metric of the model constitutes by the average of  $K$  piece performance metrics.

### Acknowledgements

This work received funding from the European Union H2020 Programme under a European Research Council Consolidator grant [MOLEMAT, 726360]. PARASOL (RTI2018-102292-B-I00), Spanish ministry of Science and Innovation.

### Conflict of Interest

The authors declare no conflict of interest.

**Keywords:** conjugated polymers • machine learning • perovskite solar cells • tungsten trioxide

- [1] Y. Chen, S. Yang, P. Chen, L. Lin, Z. Wu, H. Hsu, *Superlattices Microstruct.* **2019**, *130*, 569–577.
- [2] M. P. U. Haris, M. Pegu, S. Kazim, S. Ahmad, *Emergent Mater.* **2020**, *3*, 751–778.
- [3] M. Zhang, J. S. Yun, Q. Ma, J. Zheng, C. F. J. Lau, X. Deng, J. Kim, D. Kim, J. Seidel, M. A. Green, S. Huang, A. W. Y. Ho-Baillie, *ACS Energy Lett.* **2017**, *2*, 438–444.
- [4] C. Liu, W. Kong, W. Li, H. Chen, D. Li, W. Wang, B. Xu, C. Cheng, A. K. Y. Jen, *J. Mater. Chem. A* **2019**, *7*, 15880–15886.
- [5] M. Saliba, T. Matsui, J. Seo, K. Domanski, J. Correa-Baena, N. K. Mohammad, S. M. Zakeeruddin, W. Tress, A. Abate, A. Hagfeldt, M. Grätzel, *Energy Environ. Sci.* **2016**, *9*, 1989–1997.
- [6] H. Pan, X. Zhao, X. Gong, H. Li, N. H. Ladi, X. L. Zhang, W. Huang, S. Ahmad, L. Ding, Y. Shen, M. Wang, Y. Fu, *Mater. Horiz.* **2020**, *7*, 2276–2291.
- [7] Y. Zheng, J. Kong, H. Di, W. Shi, L. McMillon-Brown, H. E. Katz, J. Yu, A. Taylor, *Nanoscale* **2018**, *10*, 11342–11348.
- [8] Z. Li, *Chem. Lett.* **2015**, *44*, 1140–1141.
- [9] S. Cogal, L. Calio, G. Celik Cogal, M. Salado, S. Kazim, L. Oksuz, S. Ahmad, A. Uygun Oksuz, *Polym. Bull.* **2018**, *75*, 4531–4545.
- [10] H. Yi, D. Wang, L. Duan, F. Haque, C. Xu, Y. Zhang, G. Conibeer, A. Uddin, *Electrochim. Acta.* **2019**, *319*, 349–358.
- [11] Y. Guo, H. Lei, L. Xiong, B. Li, G. Fang, *J. Mater. Chem. A* **2018**, *6*, 2157–2165.
- [12] R. Remya, P. T. G. Gayathri, B. Deb, *Mater. Chem. and Phys.* **2020**, *255*, 123584.
- [13] K. P. Murphy, *Machine Learning A Probabilistic Perspective*, The MIT Press, Cambridge, Massachusetts **2012**.
- [14] W. Li, R. Jacobs, D. Morgan, *Comput. Mater. Sci.* **2018**, *150*, 454–463.
- [15] C. Odabasi, R. Yildirim, *Nano Energy* **2019**, *56*, 770–791.
- [16] J. Li, B. Pradhan, S. Gaur, J. Thomas, *Adv. Energy Mater.* **2019**, *9*, 1901891.
- [17] E. C. Gok, M. O. Yildirim, E. Eren, A. Uygun Oksuz, *ACS Omega* **2020**, *5*, 23257–23267.
- [18] F. Li, X. Peng, Z. Wang, Y. Zhou, Y. Wu, M. Jiang, M. Xu, *Energy Environ. Mater.* **2019**, *2*, 280–291.



- [19] B. Yilmaz, R. Yildirim, *Nano Energy* **2020**, 105546.
- [20] J. Schmidt, J. Shi, P. Borlido, L. Chen, S. Botti, M. A. L. Marques, *Chem. Mater.* **2017**, 29, 5090–5103.
- [21] K. Takahashi, L. Takahashi, I. Miyazato, Y. Tanaka, *ACS Photonics* **2018**, 5, 771–775.
- [22] R. Jacobs, G. Luo, D. Morgan, *Adv. Funct. Mater.* **2019**, 29, 1804354.
- [23] X.-Y. Ma, J. P. Lewis, Q.-B. Yan, G. Su, *J. Phys. Chem. Lett.* **2019**, 10, 6734–6740.
- [24] V. Gladkikh, D. Y. Kim, A. Hajibabaei, A. Jana, C. W. Myung, K. S. Kim, *J. Phys. Chem. C* **2020**, 124, 8905–8918.
- [25] S. Lu, Q. Zhou, Y. Ouyang, Y. Guo, Q. Li, J. Wang, *Nat. Commun.* **2018**, 9, 1–8.
- [26] W. A. Saidi, W. S. I. E. Castelli, *npj Comput. Mater.* **2020**, 6, 36.
- [27] X. Li, Y. Dan, R. Dong, Z. Cao, C. Niu, Y. Song, S. Li, J. Hu, *Appl. Sci.* **2019**, 9, 5510.
- [28] V. M. Le Corre, T. S. Sherkar, M. Koopmans, L. J. A. Koster, *Cell Rep. Phys. Sci.* **2021**, 2, 100346.
- [29] Q. Tao, P. Xu, M. Li, W. Lu, *npj Comput. Mater.* **2021**, 7, 23.
- [30] E. Eren, C. Alver, G. Yurdabak Karaca, E. Uygun, A. Uygun Oksuz, *Synth. Met.* **2018**, 235, 115–124.
- [31] E. Eren, C. Alver, G. Yurdabak Karaca, E. Uygun, L. Oksuz, A. Uygun Oksuz, *Electroanalysis* **2018**, 30, 2099–2109.
- [32] E. Eren, C. Alver, A. Uygun Oksuz, *Anadolu Univ. J. of Sci. and Technology A – Appl. Sci. and Eng.* **2018**, 19, 468–483.
- [33] B. Zou, S. Gong, Y. Wang, X. Liu, *J. Nanomater.* **2014**, 813120.
- [34] V. Chaudhary, A. Kaur, *RSC Adv.* **2015**, 5, 73535–73544.
- [35] H. Zheng, X. Xu, S. Xu, G. Liu, S. Chen, X. Zhang, T. Chen, X. Pan, *J. Mater. Chem. C* **2019**, 7, 4441–4448.
- [36] M. C. Stefan, C. Bulumulla, J. Du, K. E. Washington, R. Kularatne, H. Nguyen, M. C. Biewer, *J. Mater. Chem. A* **2017**, 5, 2473–2477.
- [37] T. Hastie, R. Tibshirani, J. H. Friedman, *Elements of Statistical Learning*, Springer **2001**.
- [38] S. Chander, A. Purohit, A. Sharma, S. P. Nehra, M. S. Dhaka, *Energy Rep.* **2015**, 1, 175–180.
- [39] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, The MIT Press **2015**.
- [40] F. Rodriguez-Galia, B. Ghimire, J. Rogan, M. Chica-Olmo, J. Rigol-Sanchez, *ISPRS J. Photogramm. Remote Sens.* **2012**, 67, 93–104.
- [41] R. Blagus, L. Lusa, *Comput. Stat. Data Anal.* **2017**, 113, 19–37.

---

Manuscript received: March 19, 2021

Revised manuscript received: May 1, 2021

Accepted manuscript online: May 3, 2021