

## Electronic Supplementary Information

### Machine learning-Assisted Development of Organic Photovoltaics via High-Throughput *In-Situ* Formulation

Na Gyeong An,<sup>ab</sup> Jin Young Kim,<sup>\*b</sup> and Doojin Vak<sup>\*a</sup>

<sup>a</sup> CSIRO Manufacturing, Clayton, Victoria, 3168 Australia

<sup>b</sup> School of Energy and Chemical Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, South Korea

E-mail: [jykim@unist.ac.kr](mailto:jykim@unist.ac.kr), [doojin.vak@csiro.au](mailto:doojin.vak@csiro.au)

## Experimental Procedures

### Materials

PM6 (PBDB-T-2F, poly[(2,6-(4,8-bis(5-(2-ethylhexyl-3-fluoro)thiophen-2-yl)-benzo[1,2-b:4,5-b']dithiophene))-alt-(5,5-(1',3'-di-2-thienyl-5',7'-bis(2-ethylhexyl)benzo[1',2'-c:4',5'-c']dithiophene-4,8-dione)]), Y6 (BTP-4F, 2,2'-((2Z,2'Z)-((12,13-bis(2-ethylhexyl)-3,9-diundecyl-12,13-dihydro-[1,2,5]thiadiazolo[3,4-e]thieno[2'',3'':4',5']thieno[2',3':4,5]pyrrolo[3,2-g]thieno[2',3':4,5]thieno[3,2-b]indole-2,10-diyl)bis(methanylylidene))bis(5,6-difluoro-3-oxo-2,3-dihydro-1H-indene-2,1-diylidene))dimalononitrile) and IT-4F (ITIC-2F, 3,9-bis(2-methylene-((3-(1,1-dicyanomethylene)-6,7-difluoro)-indanone))-5,5,11,11-tetrakis(4-hexylphenyl)-dithieno[2,3-d:2',3'-d']-s-indaceno[1,2-b:5,6-b']dithiophene) were purchased from 1-material. Commercial ZnO NPs solution (5.6% w/v in isopropanol (IPA)) was purchased from Infinity PV. ortho-DCB (o-DCB, 1, 2-dichlorobenzene, anhydrous, 99.0%) was purchased from Sigma-Aldrich.

### Rigid Device Fabrication

A custom-built 3D printer (K200, K-Clone)-based slot die coater was used for hot slot die deposition.<sup>1, 2</sup> Patterned ITO glass substrates were cleaned by ultrasonication in detergent, distilled water, acetone and IPA, then dried in an oven to remove residual washing solvents and treated under UV-ozone for 20 min to change the surface properties. The ZnO NP solution was slot-die coated by using a slot-die head with a 50  $\mu\text{m}$  shim and 13 mm coating width. The gap between the substrate and head was approximately 100  $\mu\text{m}$ , and the coating was performed at 5 mm s<sup>-1</sup> coating speed with 1  $\mu\text{l cm}^{-2}$  coating volume. The ZnO NP films were annealed at 120 °C for 10 min. For optimization of processing head/bed temperatures, solutions of PM6:IT-4F (1:1.2, w/w) and PM6:Y6 (1:1.2, w/w) in o-DCB (10 mg ml<sup>-1</sup> of donor and the relative amount of acceptors) were mixed at 90 °C overnight. The slot die coated films were used without additional thermal treatment. All printing experiments and sample preparation were carried out in air. For the fabrication of the top electrodes, a 10 nm MoO<sub>3</sub> (Sigma-Aldrich) layer and a 100 nm Ag (KJ Lesker) layer were deposited by thermal evaporation through a shadow mask. Active area of the device is 10 mm<sup>2</sup>.

### Flexible Device Fabrication via R2R Process

R2R devices were fabricated on pre-patterned transparent conducting electrode (TCE)-coated PET substrates (OPV8, 8  $\Omega \text{ sq}^{-1}$ , Mekoprint). The R2R experiment was carried out

using a commercial R2R coater (Mini Labo, Yasui Seki) combined with a customized slot die setup. A ZnO NP solution was slot die coated on the substrate at 20 cm min<sup>-1</sup> coating speed with a 15  $\mu$ l min<sup>-1</sup> solution feed rate. Based on 26 cm<sup>2</sup> min<sup>-1</sup> (1.3 cm width  $\times$  20 cm length) coating area, the wet film thickness was approximately 5.77  $\mu$ m. The film was then thermally annealed on a curved hot plate at 120 °C for about 30 sec. Active layers were fabricated from various solutions and deposition parameters at fixed a deposition temperature (head/bed : 90/130 °C) and coating speed (30 cm min<sup>-1</sup>). Detailed coating procedure is described in the section below. The films were cut to 11 cm length for metal evaporation. All printing experiments and sample preparation also were carried out under atmospheric conditions and top electrodes (MoO<sub>3</sub>/Ag) were deposited using thermal evaporator in the same manner as rigid devices. Active area of the device is 14 mm<sup>2</sup>.

### Active Layer Deposition Procedure

Type (i): Pre-mixed solutions of PM6:Y6:IT-4F with various PM6 to Y6:IT-4F (4:1) ratios were prepared in o-DCB with fixed concentration of PM6 (10 mg ml<sup>-1</sup>). The solutions were deposited in the range of 8.5–120  $\mu$ l min<sup>-1</sup> (wet film thickness: 2.18–30.8  $\mu$ m) using a programmable syringe pump (NEMESYS, CETONI GmbH), thus corresponding TDDs linearly increased. Experimental validations of the optimized formulations (BPF and BEF) were performed in the same way.

Type (ii): Solutions of PM6:Y6 and PM6:IT-4F were prepared separately. Donor:acceptor ratios were varied, but donor concentration was fixed at 10 mg ml<sup>-1</sup> in o-DCB. The solutions were injected into a Y-connector and the *in-situ* mixed solution was fed through tubing with a very small inner diameter of 0.3 mm to minimize diffusion. Total solution flow was kept constant at 36.67  $\mu$ l min<sup>-1</sup> or 59.10  $\mu$ l min<sup>-1</sup> (wet film thickness: 9.40 and 15.2  $\mu$ m) while changing relative flow rates of two pumps. The dead volume of the tubing from the Y-connector and the slot die was calculated to be about 60  $\mu$ l for our typical set-up. This means the *in-situ* mixed solution gets deposited around 1.6 min after the mixing at 36.67  $\mu$ l min<sup>-1</sup> flow rate. The composition of the film at each position was calculated based on the relative flow rate and the deposition offset originated from the dead volume of the deposition system.

Type (iii): Two solutions of PM6 and Y6:IT-4F (4:1, w/w) were prepared in o-DCB with 20 mg ml<sup>-1</sup> separately. The solutions were mixed and deposited using the same method as type (ii) experiments at various total solution flows (39, 54.6, 65, 70  $\mu$ l min<sup>-1</sup>) at a fixed coating speed of 30 cm min<sup>-1</sup> and 1.3 cm coating width. The solution flows correspond to 10, 14, 16.7, and

17.9  $\mu\text{m}$  of wet film thicknesses and 20, 28, 35.6 and 50  $\mu\text{g cm}^{-2}$  of TDD, respectively. The composition of the film at each position was also calculated by considering the relative flow rate and the deposition offset, same as with type (ii) experiments.

### **Device Characterization**

*J-V* characteristics of printed OPV devices were measured using a Keithley 2400-LV source measurement unit under inert conditions. A 150 W Xenon lamp (Newport) coupled with an AM 1.5G solar spectrum filter was used as the light source. Light was illuminated through a quartz window of the glove box and intensity was calibrated and monitored using a secondary reference cell (Hamamatsu S1133, with KG-5 filter, 2.8 mm  $\times$  2.4 mm of photosensitive area) which was pre-calibrated by a certified reference cell (PV Measurements, certified by NREL) under 1000  $\text{W m}^{-2}$  AM 1.5G illumination from an Oriel AAA solar simulator fitted with a 1000 W Xe lamp. The calibrated intensity was also regularly confirmed using encapsulated devices under the AAA solar simulator. All device data were manually collected. IPCE spectra were measured using a QEX7 quantum efficiency measurement system (PV Measurements, Inc.) under ambient conditions.

### **Film Characterization**

UV-vis absorption spectra of thin films were obtained by using a Cary 5000 spectrophotometer (Agilent, USA). For comparison with the R2R processed films, the batch-processed films were prepared by the 3D printer-based flatbed slot die coater<sup>1,2</sup> at the optimized deposition temperature (head/bed : 90/130  $^{\circ}\text{C}$ ). AFM height and phase images (2  $\mu\text{m} \times 2 \mu\text{m}$ ) were obtained using a Multimode V microscope (Veeco, USA) in tapping mode. A dimension microscope running with a Nanoscope V controller was used to obtain AFM images of organic thin films. The images were measured in high-resolution tapping mode under ambient conditions. Silicon cantilevers (Bruker) with a resonant frequency of 300 kHz were used with a rotated tip to provide a more symmetric representation of features over 200 nm. Film thicknesses were measured using a surface profiler (P6, KLA Tencor, USA). The samples were slot die coated on PET substrates by prepared by one or two programmable syringe pumps.

### **ML Modelling**

ML algorithms (random forest, gradient boosting, AdaBoost, bagging, extra trees, histogram-based gradient boosting regression) available in Scikit-Learn package ([https://scikit-](https://scikit-learn.org/)

[learn.org/](https://www.learn.org/)) for Python were employed with default modelling parameters to construct ML models. The 2218 datasets (deposition density (DD) of three materials as input and PCE as output) were divided into training and testing sets. We followed a general data split method, 80/20 rule (split 80% of the data selected into a training set and the remaining 20% sample into the testing set).<sup>3</sup> To avoid an un-uniform distribution of training and testing set split, a random sampling method was employed. A hold-out validation technique was used and repeated 10 times to confirm reproducibility and generalizability.

### Model Performance Metrics

The model performance was assessed by calculating the following metrics.

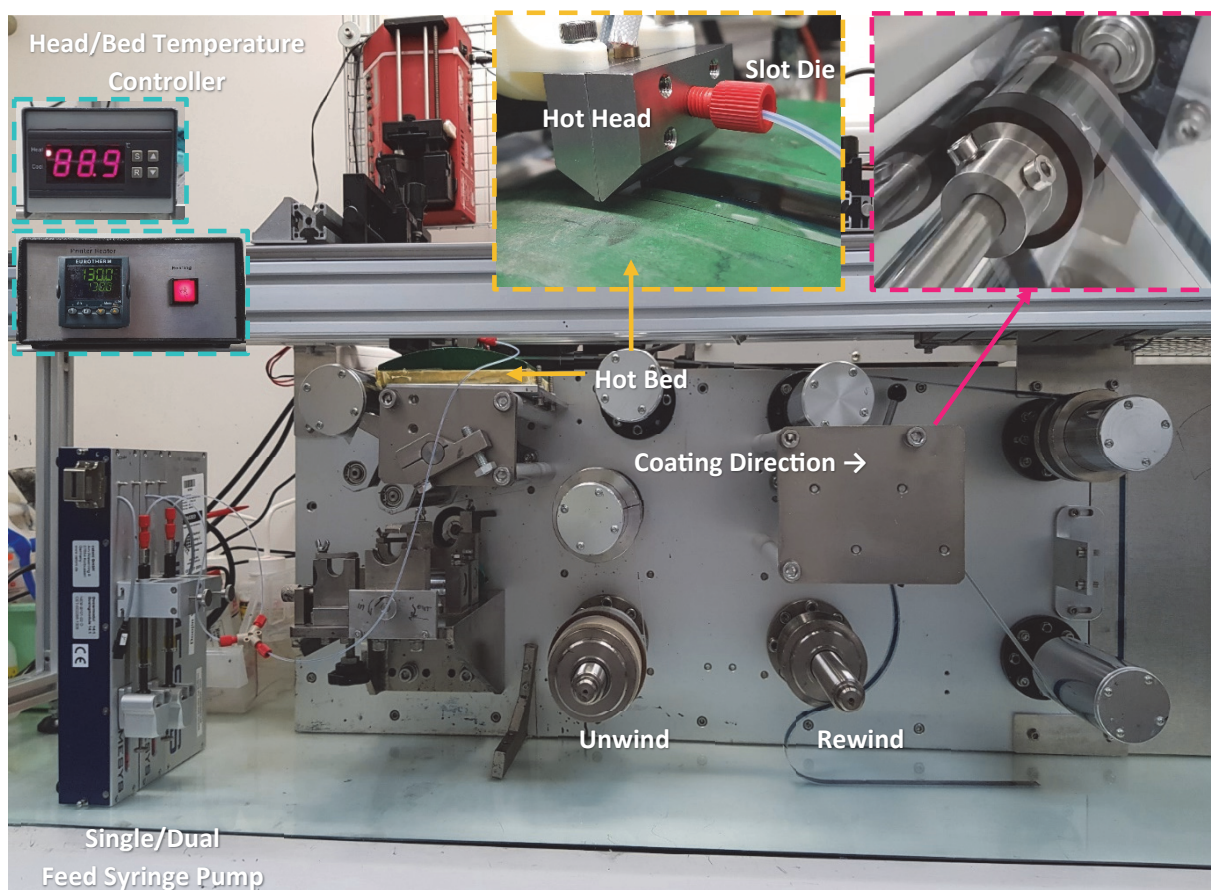
$$\text{Mean Absolute Percentage Error (MAPE)} = \frac{100}{N} \sum_{k=1}^N \frac{|P_k - M_k|}{M_k} \quad (\text{Equation S1})$$

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\frac{\sum_{k=1}^N (M_k - P_k)^2}{N}} \quad (\text{Equation S2})$$

$$\text{Coefficient of Determination (R}^2\text{)} = 1 - \frac{\sum_{k=1}^N (M_k - P_k)^2}{\sum_{k=1}^N (M_k - \bar{M})^2} \quad (\text{Equation S3})$$

$$\text{Variance} = \sum_{k=1}^N \frac{(P_k - \bar{P})^2}{N} \quad (\text{Equation S4})$$

For Eq. S1–S4,<sup>3, 4</sup>  $M_k$  is the measured and  $P_k$  is the predicted value and  $\bar{M}$  is mean of measured and  $\bar{P}$  is mean of predicted values.  $N$  is total number of data trained into machine learning. MAPE can be problematic as it can cause division-by-zero errors. Therefore, for each model building trial, 0% PCE data were excluded to obtain MAPE from the remaining data.



**Figure S1.** Equipment for hot deposition R2R slot die coating.

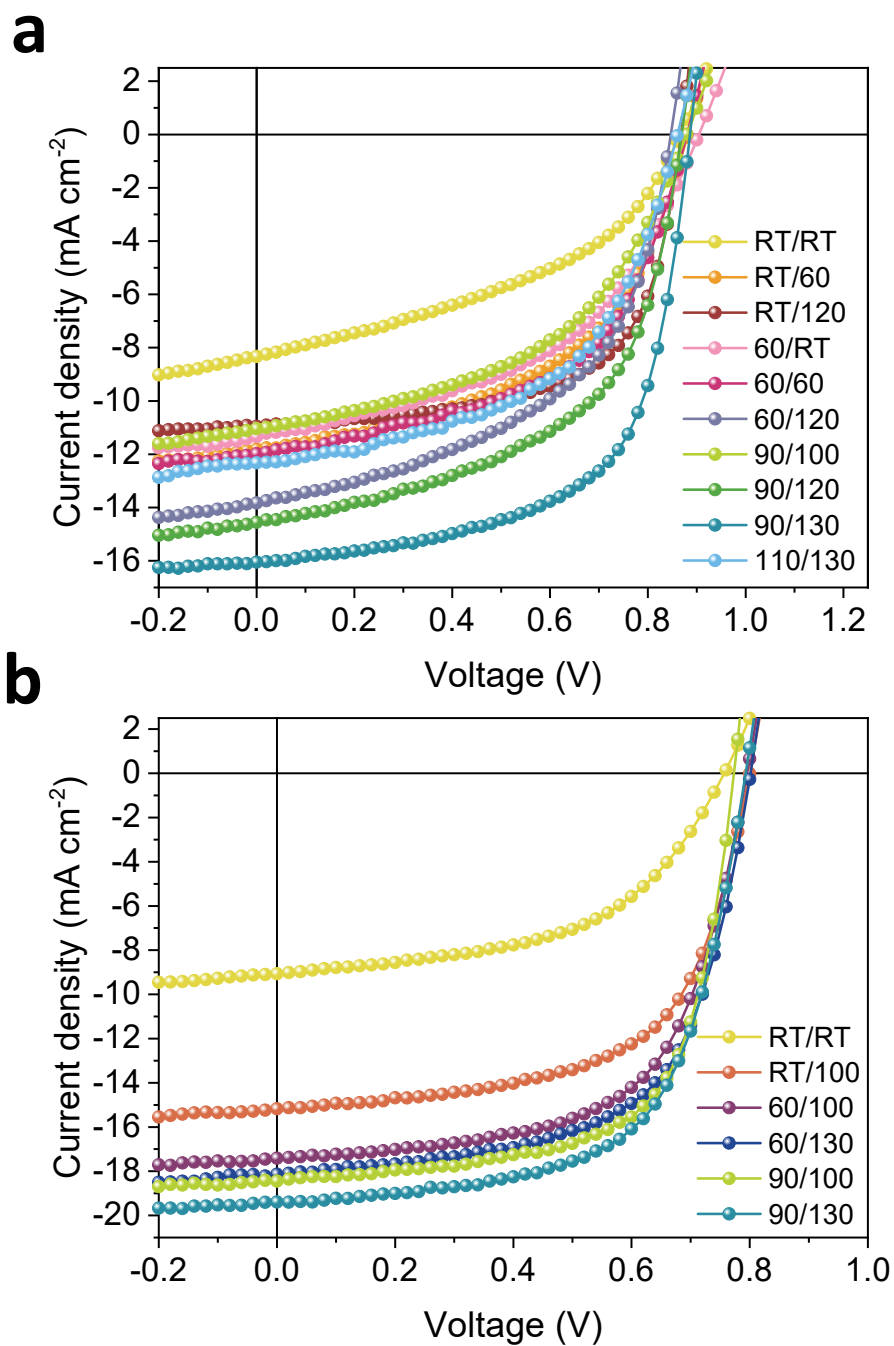
**Table S1.** Summary of photovoltaic parameters of slot die coated OPV devices on glass substrates at variable processing temperatures.

| Active Layer           | Head (°C) | Bed (°C) | $J_{sc}$ (mA cm <sup>-2</sup> ) | $V_{oc}$ (V)         | FF                   | PCE (%)             |
|------------------------|-----------|----------|---------------------------------|----------------------|----------------------|---------------------|
| PM6:IT-4F <sup>a</sup> | RT        | RT       | 8.33<br>(7.98±0.20)             | 0.86<br>(0.87±0.01)  | 0.42<br>(0.42±0.003) | 3.02<br>(2.89±0.06) |
|                        |           | 60       | 11.8<br>(10.9±0.60)             | 0.88<br>(0.88±0.00)  | 0.50<br>(0.49±0.01)  | 5.25<br>(4.65±0.34) |
|                        |           | 120      | 11.0<br>(11.1±0.65)             | 0.86<br>(0.84±0.02)  | 0.64<br>(0.55±0.05)  | 6.00<br>(5.16±0.52) |
|                        | 60        | RT       | 11.3<br>(11.9±0.32)             | 0.90<br>(0.90±0.004) | 0.48<br>(0.44±0.02)  | 4.89<br>(4.72±0.10) |
|                        |           | 60       | 12.0<br>(11.3±0.42)             | 0.88<br>(0.88±0.004) | 0.53<br>(0.56±0.02)  | 5.81<br>(5.68±0.08) |
|                        |           | 120      | 13.8<br>(11.4±1.14)             | 0.84<br>(0.85±0.01)  | 0.52<br>(0.55±0.02)  | 6.00<br>(5.43±0.27) |
|                        | 90        | 100      | 11.0<br>(11.0±0.12)             | 0.88<br>(0.88±0.01)  | 0.48<br>(0.48±0.004) | 4.66<br>(4.62±0.06) |
|                        |           | 120      | 14.6<br>(13.7±0.53)             | 0.86<br>(0.87±0.01)  | 0.55<br>(0.54±0.01)  | 6.85<br>(6.44±0.29) |
|                        |           | 130      | 16.1<br>(15.4±0.58)             | 0.88<br>(0.88±0.01)  | 0.63<br>(0.64±0.01)  | 8.84<br>(8.55±0.21) |
|                        | 110       | 130      | 12.3<br>(12.2±0.38)             | 0.86<br>(0.86±0.01)  | 0.53<br>(0.52±0.02)  | 5.80<br>(5.50±0.28) |
| PM6:Y6 <sup>b</sup>    | RT        | RT       | 9.07<br>(8.45±0.41)             | 0.76<br>(0.76±0.00)  | 0.52<br>(0.53±0.01)  | 3.56<br>(3.38±0.12) |
|                        |           | 100      | 15.2<br>(13.7±0.74)             | 0.80<br>(0.79±0.01)  | 0.61<br>(0.62±0.01)  | 7.38<br>(6.64±0.38) |
|                        | 60        | 100      | 17.4<br>(16.9±0.45)             | 0.80<br>(0.78±0.01)  | 0.61<br>(0.62±0.01)  | 8.53<br>(8.23±0.25) |
|                        |           | 130      | 18.1<br>(18.1±0.38)             | 0.80<br>(0.79±0.01)  | 0.62<br>(0.61±0.01)  | 9.02<br>(8.87±0.12) |
|                        | 90        | 100      | 18.4<br>(18.1±0.38)             | 0.78<br>(0.78±0.01)  | 0.65<br>(0.64±0.02)  | 9.33<br>(8.96±0.31) |
|                        |           | 130      | 19.4<br>(19.1±0.37)             | 0.80<br>(0.79±0.01)  | 0.62<br>(0.63±0.01)  | 9.69<br>(9.55±0.08) |

\* All statistical data were calculated from 6 devices.

<sup>a</sup> Active Layer 1 = PM6:IT-4F (1:1.2, w/w) in 10 mg ml<sup>-1</sup> DCB

<sup>b</sup> Active Layer 2 = PM6:Y6 (1:1.2, w/w) in 10 mg ml<sup>-1</sup> DCB



**Figure S2.** Photovoltaic performance of devices.  $J$ - $V$  characteristics of slot die coated OPV devices on glass substrates at variable processing temperatures for (a) PM6:IT-4F and (b) PM6:Y6.

## Supplementary Note 1. (Total) Deposition Density ( $\mu\text{g cm}^{-2}$ )

The slot die coating technique is classified as a pre-metered method where all ink injected into the slot die head is coated onto a substrate with no loss of material,<sup>5</sup> hence the thickness of films is precisely controlled by the solution flow. Therefore, the method makes it possible to quantify compositional information (e.g. component ratio or thickness of materials) at a given coating position. In this study, new terms were employed; DD and TDD to refer to the quantitative amount deposited of each material (e.g. PM6 amount) and their sum (e.g. sum of PM6, Y6 and IT-4F amounts) per unit area (in  $\mu\text{g cm}^{-2}$  unit), respectively. Thus, they provide indirect thickness information and the composition ratio of materials at each coating position.

The solution flows were controlled linearly by using programmable syringe pumps and the coating speed was fixed (Fig. S3a). Therefore, the deposition volume of each solution at each position can be calculated (Fig. S3; and Fig. 3b and c) using the following parameters and procedure.

- 1)  $P$  (sec): Time (Period) from the start to the end of flow ramp.
- 2)  $A_0$  (or  $B_0$ ) ( $\mu\text{l min}^{-1}$ ): Initial flow rate of solution A or B.
- 3)  $A_F$  (or  $B_F$ ) ( $\mu\text{l min}^{-1}$ ): Final flow rate of solution A or B. In case of a single feed, the initial and final flow rates of solution A are denoted as  $A'_0$  or  $A'_F$ , respectively.
- 4)  $S$  ( $\text{cm min}^{-1}$ ): Coating speed (Line speed).
- 5)  $x$  (cm): Instantaneous coating position.
- 6)  $I$  (cm): Initial coating position or deposition offset originating from the dead volume of the dual-feed deposition system. It is zero for type (i) experiment and non-zero for type (ii) and (iii) experiments.
- 7)  $F$  ( $\mu\text{l min}^{-1}$ ): Instantaneous flow rate of a solution.
- 8)  $C$  ( $\text{mg ml}^{-1}$ ): Concentration of a donor/acceptor/blend solution.
- 9)  $W$  (cm): Coating width, which is fixed at 1.3 cm in this study.
- 10)  $T$  ( $\mu\text{g cm}^{-2}$ ): Instantaneous (total) deposition density at specific coating position.

The calculation procedures to determine TDD or DD were divided into four steps in the following manner.

- 1)  $F$  depending on coating position

If we only consider  $F$  in a type (i) experiment and the coating distance during a given period is denoted as  $\alpha$ ,  $F$  can be represented as following (Eq.S4).

$$F = \frac{(A'_F - A'_O)x}{\alpha} + A'_O \quad (\text{Equation S5})$$

2) Substituting  $\alpha$  with the pre-defined parameters

During a given period, the coating distance can be calculated at constant line speed.

$$\alpha = P(S) \times S \text{ (cm min}^{-1}\text{)} \times \frac{1 \text{ min}}{60 \text{ sec}} = \frac{PS}{60} \text{ (cm)} \quad (\text{Equation S6})$$

3)  $F$  at given coating position with the pre-defined parameters

We can re-write equation (1) as below. The flow rate profiles of solution A and B in type (ii) and (iii) experiments can be also obtained in the same manner.

① Type (i) experiment

$$F = \frac{60(A'_F - A'_O)x}{PS} + A'_O \quad (\text{Equation S7})$$

② Type (ii) and (iii) experiment

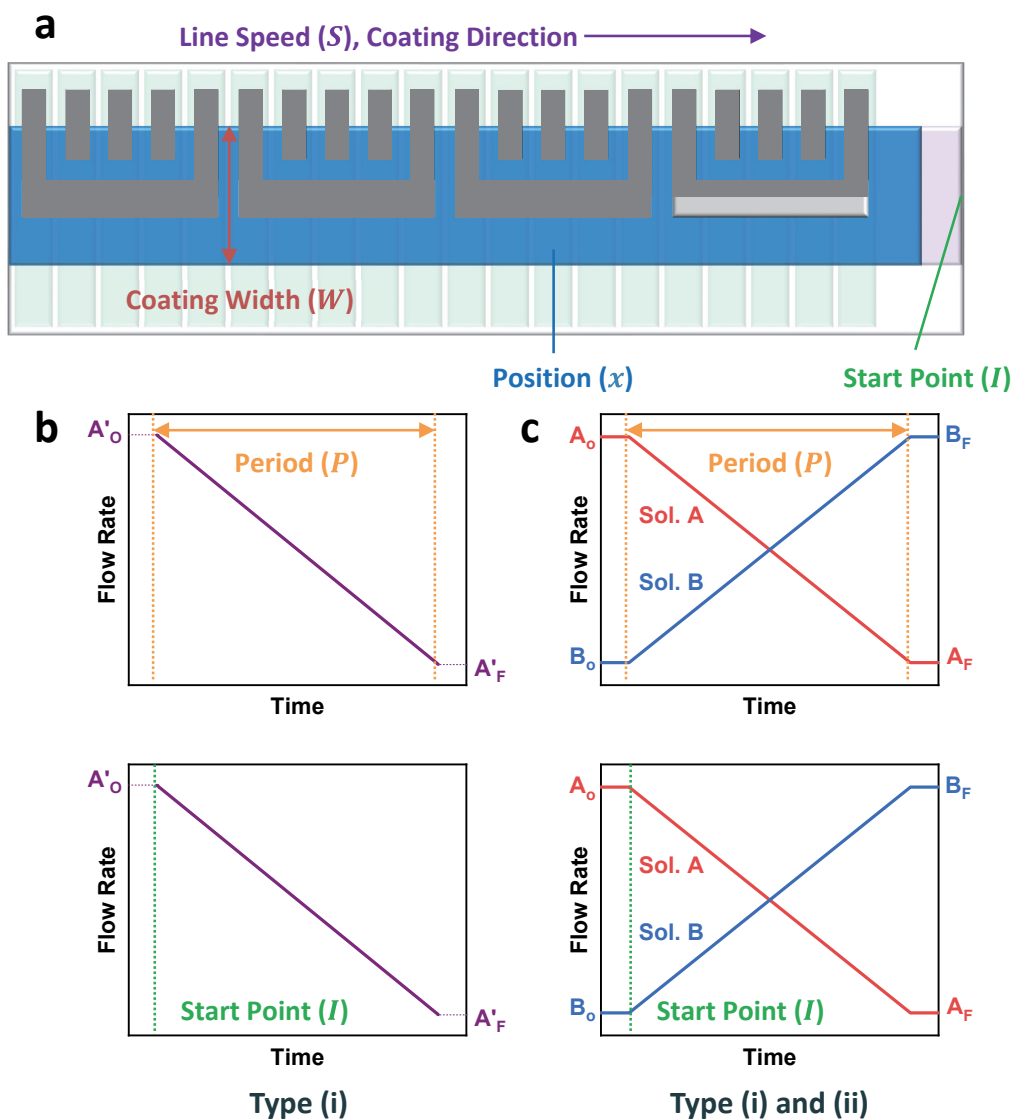
$$\text{Solution A} \begin{cases} x \leq I, F = A_O \\ x \geq I, F = \frac{60(A_F - A_O)}{PS}(x - I) + A_O \end{cases} \quad (\text{Equation S8})$$

$$\text{Solution B} \begin{cases} x \leq I, F = A_F \\ x \geq I, F = -\frac{60(A_F - A_O)}{PS}(x - I) + A_F \end{cases} \quad (\text{Equation S9})$$

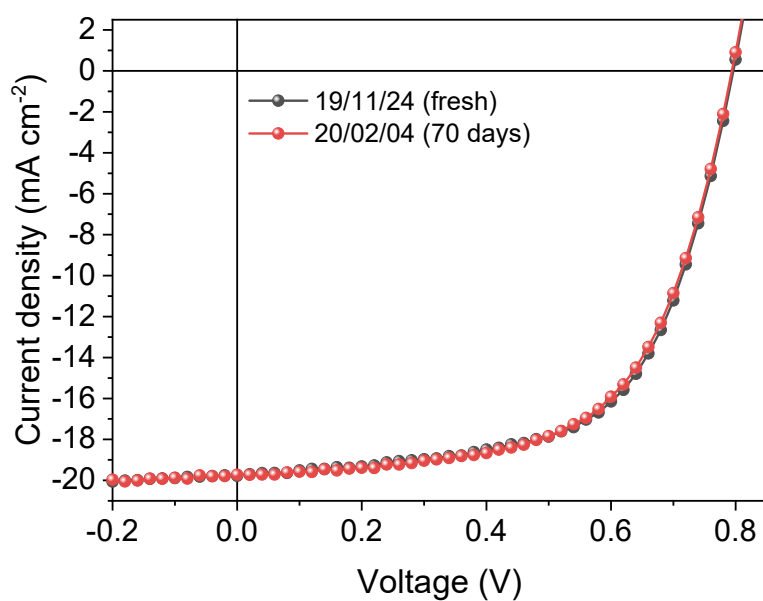
As the flow profiles of the two solutions are symmetrical relative to each other,  $A_F$  and  $A_O$  can be replaced with  $B_O$  and  $B_F$ .

4)  $T$  at given  $C$ ,  $S$  and  $W$

$$\therefore T = \frac{C \text{ (mg ml}^{-1}\text{)} \times F \text{ (}\mu\text{l min}^{-1}\text{)}}{S \text{ (cm min}^{-1}\text{)} \times W \text{ (cm)}} \times \frac{1000 \mu\text{g}}{1 \text{ mg}} \times \frac{1 \text{ mL}}{1000 \mu\text{L}} = \frac{CF}{SW} \text{ (}\mu\text{g cm}^{-2}\text{)} \quad (\text{Equation S10})$$



**Figure S3.** Schematic illustration of R2R device and schematic graph for type (i)–(iii) experiments. (a) Schematic illustration of R2R device being coated with a given line speed and coating direction. Flow rate change depending on time and position at (b) type (i) (single feed) and (c) type (ii) and (iii) experiments (dual feed).



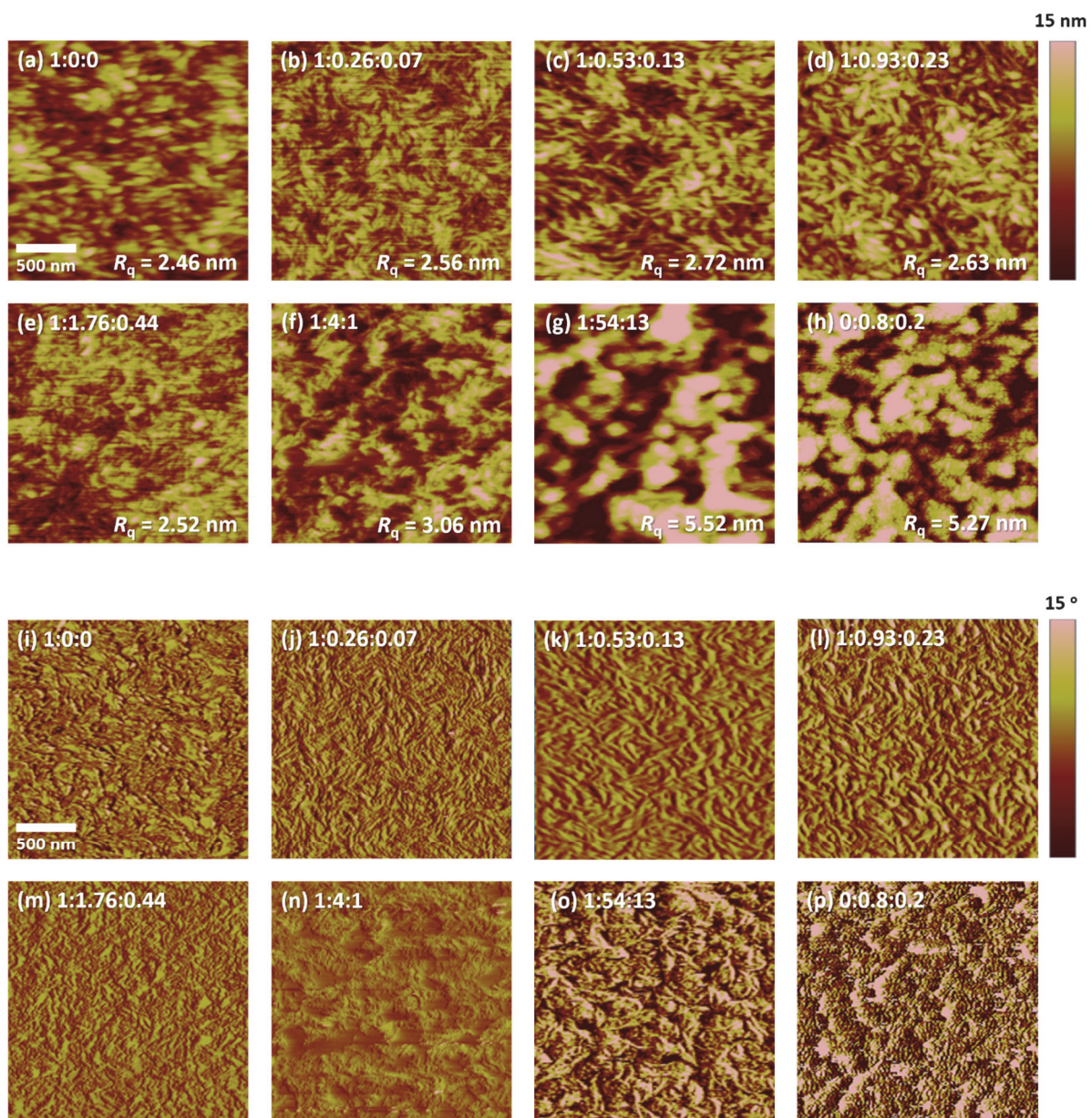
**Figure S4.** Photovoltaic performance of devices under nitrogen atmosphere storage.  $J$ – $V$  characteristics of R2R produced OPV device for fresh and stored 70 days.

**Table S2.** Summary of photovoltaic parameters of R2R produced OPV devices for fresh and stored 70 days.

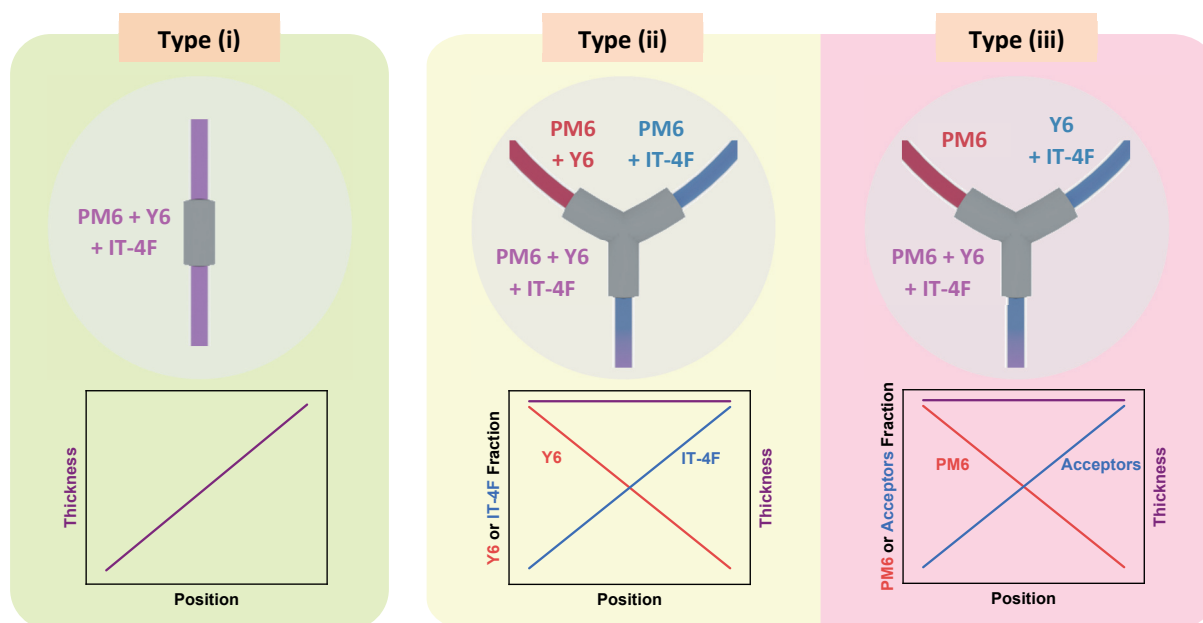
| PM6:Y6:IT-4F<br>(w/w) | Date               | $J_{sc}$<br>(mA cm <sup>-2</sup> ) | $V_{oc}$<br>(V) | FF    | PCE<br>(%) |
|-----------------------|--------------------|------------------------------------|-----------------|-------|------------|
| 1:0.93:0.23           | 19/11/24 (fresh)   | 19.80                              | 0.80            | 0.612 | 9.69       |
|                       | 20/02/04 (70 days) | 19.72                              | 0.80            | 0.607 | 9.58       |

## Supplementary Note 2. AFM Measurement

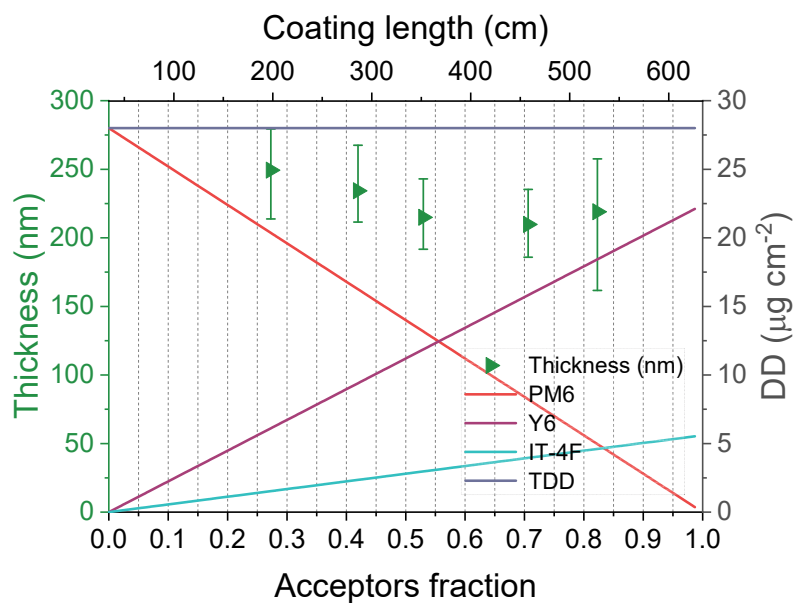
AFM height and phase images were collected at selected composition films to gain insight into the morphological difference modulated by mixing donor and acceptor only solutions. The height and phase images of a donor-only composition (1:0:0) film, as shown in Fig. S5a–i, clearly exhibited nanoscale aggregation on its surface and formed smooth films with root-mean-square (RMS) of 2.46 nm. Fig. S5b–e correspond to the height images of 1:0.26:0.07–1:1.76:0.44. Their domains seem to be needle-like in shape, which is slightly different from that of the 1:0:0 film. Nevertheless, nanoscale granular surfaces shown in the donor-only film were also observed. It implies that the crystallization of the donor primarily determines phase separation of the bulk-heterojunction film with a composition of up to 1:2.2 donor:acceptors ratio. The RMS values of the films were found to be 2.46–2.72 nm, which are almost the same as that of the donor-only film. These phenomena are similar to previous reports.<sup>6, 7</sup> Height images of 1:4:1 and acceptor-dominant composition (1:54:13) films (Fig. S5f–g), show irregular and bumpy surfaces due to an excessive amount of acceptors clumping together, lead to an increasing domain size and RMS of 3.06 and 5.52 nm, respectively. The apparent phase separation between donor and acceptors was observed in their phase images, as shown in Fig. S5n–o, it would have triggered a reduction in their device performance. A 0:0.8:0.2 film in Fig. S5h–p exhibited a similar large phase separation, as well as morphology and RMS values with acceptor-dominant composition films in Fig. S5g–o.



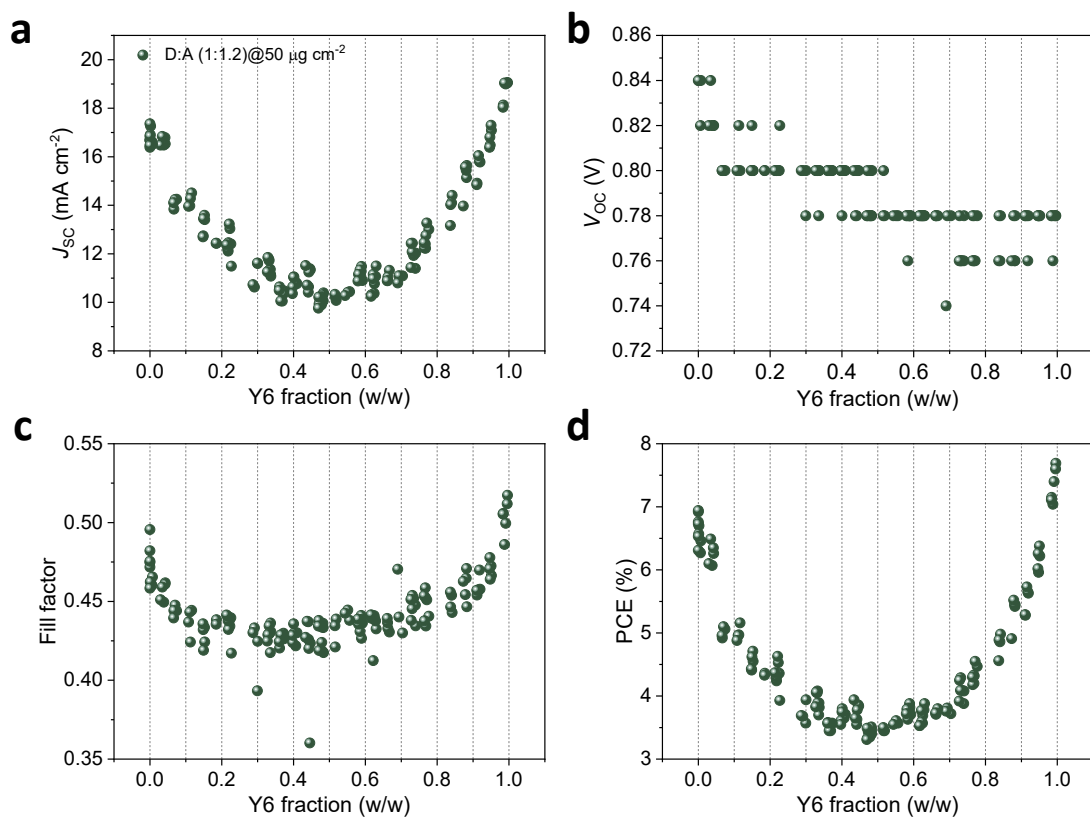
**Figure S5.** AFM height and phase images. AFM. (a)–(h) height and (i)–(p) corresponding phase images of dual-feed slot die coated films on top of Poly(ethyl benzene-1,4-dicarboxylate) (PET)/ZnO nanoparticles (ZnO NPs) at various D:A<sub>1</sub>:A<sub>2</sub> ratios. All images are 2  $\mu\text{m}$   $\times$  2  $\mu\text{m}$ .



**Figure S6.** Graphical representation of three experiment types. Green, yellow and pink boxes illustrate the solution conditions and schematic changes of thickness and ratios depending on coating position, which correspond to the three experiment types.



**Figure S7.** Film thickness measurements. Thicknesses of dual-feed R2R slot die coated films at various D:A<sub>1</sub>:A<sub>2</sub> ratios (Fig. 1c, type (iii) experiment). Top x-axis represents physical positions of each device relative to the start of the coating.



**Figure S8.** Photovoltaic parameters of R2R produced OPV devices. (a)  $J_{sc}$ , (b)  $V_{oc}$ , (c) FF and (d) PCE of the OPV devices depending on Y6:IT-4F ratios at PM6:acceptors = 1:1.2 (w/w) (fixed TDD at 50  $\mu\text{g cm}^{-2}$ ).

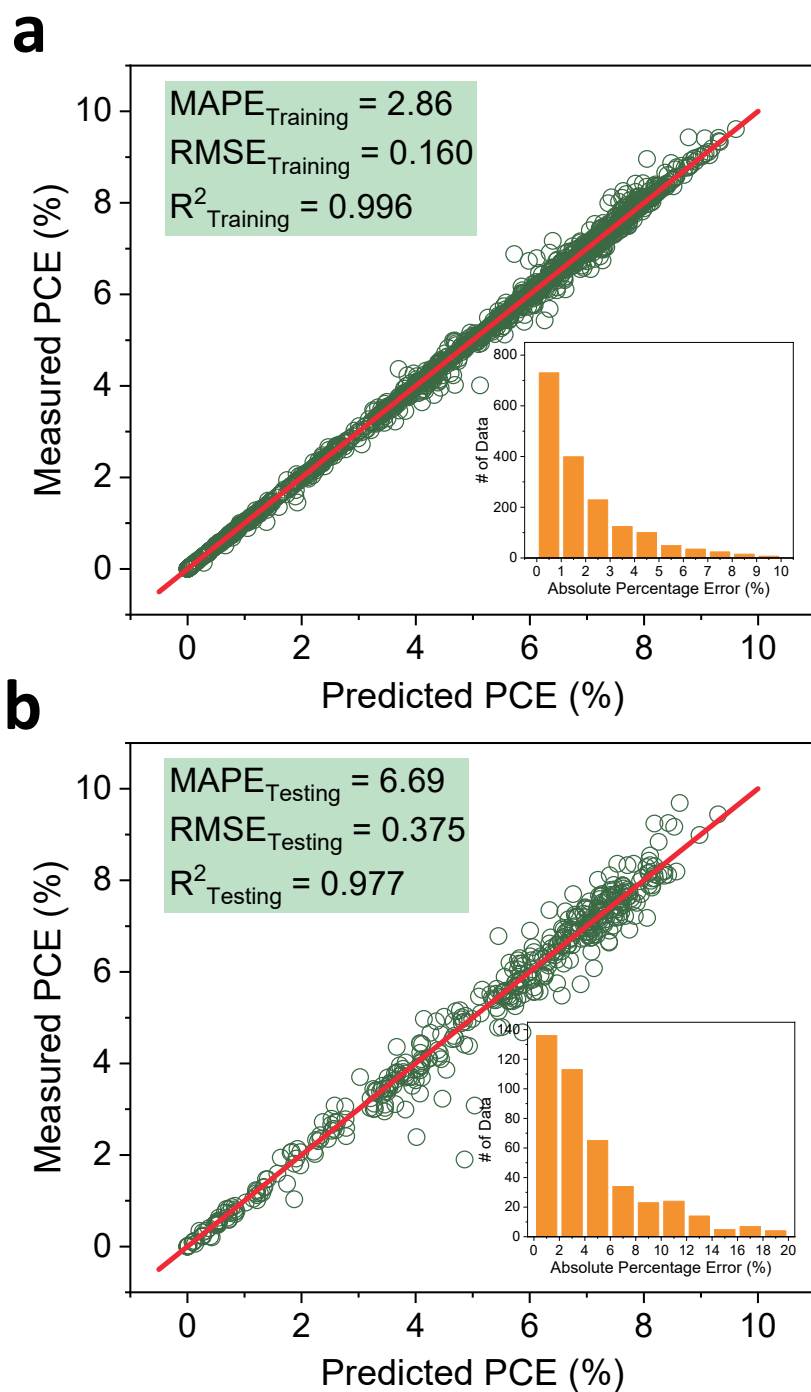
### Supplementary Note 3. Validation of ML Models

We used the open-source Scikit-learn library and explored available algorithms: RF, gradient boosting, AdaBoost, bagging, extra trees (ET), and histogram-based gradient boosting regression. The algorithms are ensemble methods, which combine weak estimators to generate a strong learner in order to improve generalizability and robustness against outliers.<sup>8</sup> We also tried some non-ensemble methods and found the predicted values were far from the experimental result. Therefore, only ensemble methods were investigated further by comparing the mean absolute percentage error (MAPE), root mean square error (RMSE) and coefficient of determination ( $R^2$ ) of the algorithms. Each algorithm was trained by using randomly selected 80% of 2218 datasets. Then the trained model was used to predict the PCE of the training dataset. MAPE, RMSE and  $R^2$  were calculated from experimental PCE of the training data and predicted PCEs. The rest (20% of unused datasets, called a testing set) was used to predict PCE and then the three performance metrics were also calculated. The process was repeated 10 times, so-called 10 times trial hold-out validation, and their best predictive powers are summarized in Table S3. Among the six algorithms, RF, ET and bagging algorithms showed good prediction accuracy with similar performance metrics. It is because ET is a modification of RF algorithm.<sup>9</sup> We found that RF is an enhancement of bagging that can improve variable selection<sup>10</sup> and has been widely used due to its robustness to outliers and high performance with non-linear high-dimensional data.<sup>11</sup> It was successfully applied in various fields such as lithium-ion battery capacity estimation,<sup>12</sup> wheat biomass estimation<sup>13</sup> and magnetic resonance image synthesis<sup>11</sup> etc. Therefore, we chose the RF algorithm to analyze the OPV data.

Fig. S9a shows a training data set and its highest performance metrics of the RF model (Table S4). Experimentally measured PCEs are used as the X value and predicted PCEs are used as the Y value. For the perfect model, all green dots would be located on the redline, i.e.  $X = Y$ . The figure shows that most data points are in line with the ideal line and the performance metrics are exceptional compared to previous reports. However, the performance metrics of the training set can be overestimated due to overfitting.<sup>14, 15</sup> Therefore, the validation results of the testing data set are more important since the performance metrics were obtained from unknown data. The results are shown in Fig. S9b. As expected, the data points are more scattered than those of the training set, and the performance metrics are inferior. However, the performance metrics are still exceptional, and most of the data are within 6 % error range. The reason would be the consistently collected dataset. Previous works used data collected from multiple reports in the literature.

**Table S3.** Summary of averaged model performance metrics (MAPE, RMSE and  $R^2$ ) at best trial obtained by different ML algorithms. 10 times trial hold-out validation method (80% and 20% of 2218 dataset were used for training and testing data, respectively) was used to select an algorithm for further study.

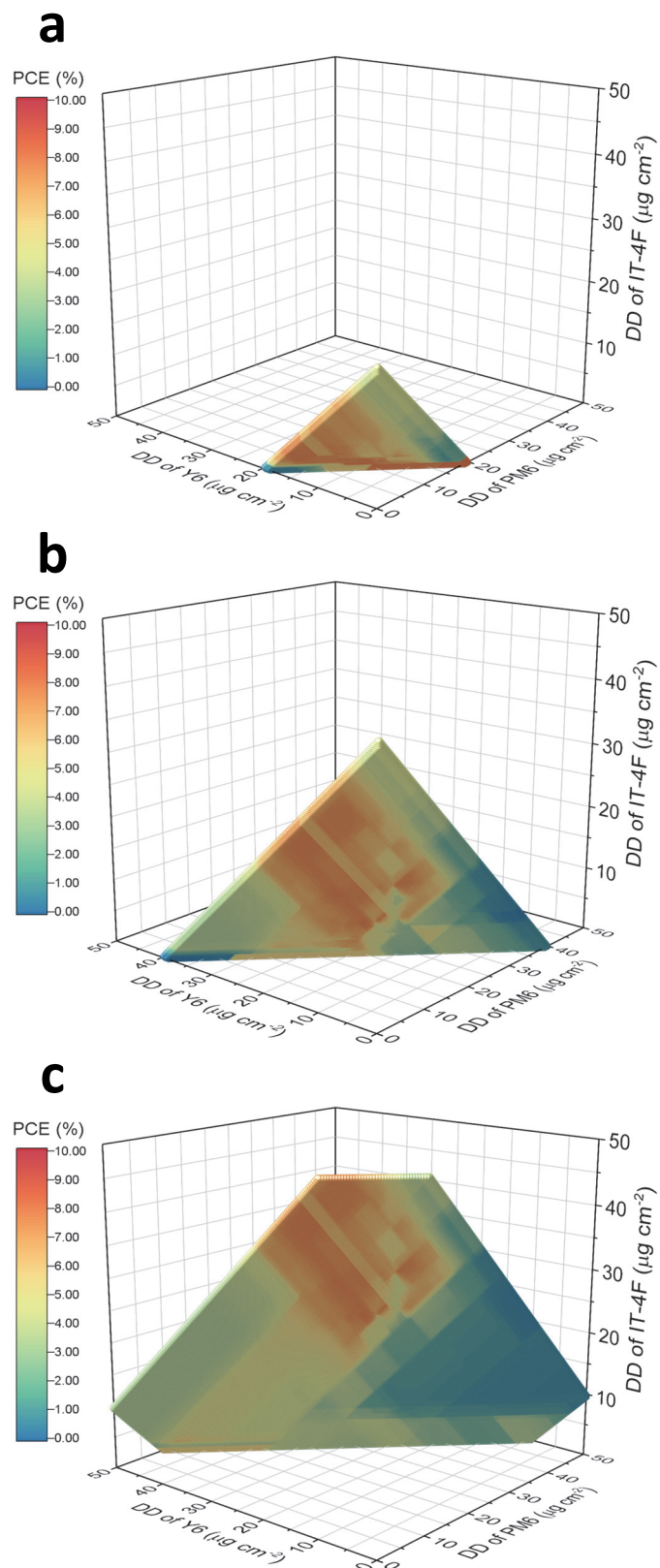
| <b>ML Model</b>                   | <b>Data Type</b> | <b>MAPE</b> | <b>RMSE</b> | <b><math>R^2</math></b> |
|-----------------------------------|------------------|-------------|-------------|-------------------------|
| Random forest                     | Training         | 2.86        | 0.160       | 0.996                   |
|                                   | Testing          | 6.69        | 0.375       | 0.977                   |
| Gradient boosting                 | Training         | 19.4        | 0.692       | 0.918                   |
|                                   | Testing          | 21.9        | 0.728       | 0.907                   |
| AdaBoost                          | Training         | 86.0        | 1.08        | 0.799                   |
|                                   | Testing          | 75.2        | 1.08        | 0.795                   |
| Bagging                           | Training         | 3.25        | 0.188       | 0.994                   |
|                                   | Testing          | 6.54        | 0.387       | 0.975                   |
| Extra trees                       | Training         | 0.044       | 0.035       | 1.00                    |
|                                   | Testing          | 5.95        | 0.370       | 0.976                   |
| Histogram-based gradient boosting | Training         | 11.4        | 0.381       | 0.975                   |
|                                   | Testing          | 12.4        | 0.496       | 0.959                   |



**Figure S9.** Model performance metrics and absolute percentage error distributions of training and testing data. Linear correlation between the measured and predicted efficiency of (a) training and (b) testing data in the RF model. Red line is represented as a reference for fitting the results. Various model performance metrics (MAPE, RMSE and  $R^2$ ) are summarized in green box. Inset graph shows the population distribution of absolute percentage errors for both training and testing data, respectively.

**Table S4.** Summary of averaged model performance metrics (MAPE, RMSE and  $R^2$ ) at best trial obtained by different ML algorithms. 10 times trial hold-out validation method (80% and 20% of 2218 dataset were used for training and testing data, respectively) was used to select an algorithm for further study.

| <b>Trial</b> | <b>Data Type</b> | <b>MAPE</b> | <b>RMSE</b> | <b><math>R^2</math></b> |
|--------------|------------------|-------------|-------------|-------------------------|
| 1            | Training         | 2.86        | 0.160       | 0.996                   |
|              | Testing          | 6.69        | 0.375       | 0.977                   |
| 2            | Training         | 2.91        | 0.165       | 0.995                   |
|              | Testing          | 7.10        | 0.416       | 0.971                   |
| 3            | Training         | 2.79        | 0.160       | 0.996                   |
|              | Testing          | 8.18        | 0.503       | 0.955                   |
| 4            | Training         | 2.88        | 0.162       | 0.996                   |
|              | Testing          | 6.65        | 0.424       | 0.966                   |
| 5            | Training         | 2.89        | 0.163       | 0.995                   |
|              | Testing          | 9.30        | 0.470       | 0.962                   |
| 6            | Training         | 3.21        | 0.159       | 0.996                   |
|              | Testing          | 8.25        | 0.396       | 0.971                   |
| 7            | Training         | 2.98        | 0.162       | 0.995                   |
|              | Testing          | 8.02        | 0.406       | 0.972                   |
| 8            | Training         | 2.84        | 0.157       | 0.996                   |
|              | Testing          | 7.06        | 0.499       | 0.956                   |
| 9            | Training         | 2.89        | 0.162       | 0.995                   |
|              | Testing          | 6.12        | 0.384       | 0.973                   |
| 10           | Training         | 2.63        | 0.162       | 0.995                   |
|              | Testing          | 9.70        | 0.420       | 0.969                   |
| Average      | Training         | 2.89        | 0.161       | 0.996                   |
|              | Testing          | 7.71        | 0.429       | 0.967                   |



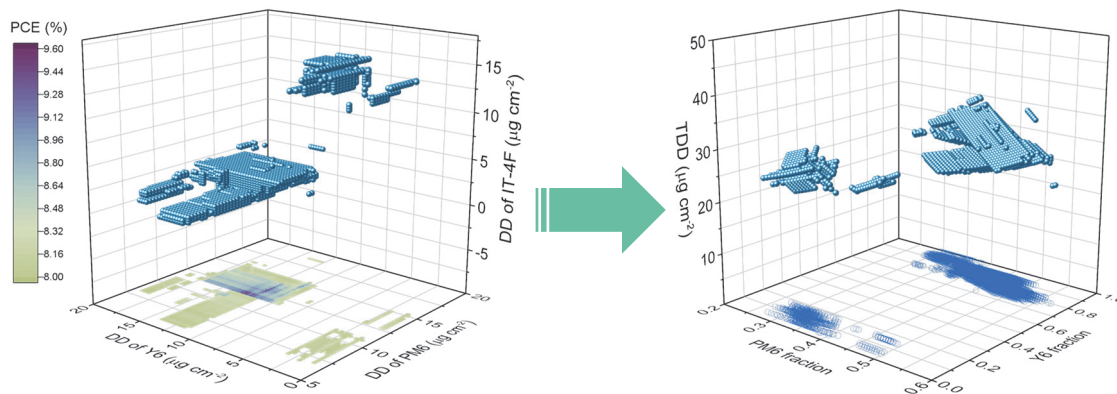
**Figure S10.** Predicted PCEs of various thickness regions. Predicted PCE variations at TDD of (a)  $\sim 20$ , (b)  $\sim 40$  and (c)  $\sim 60 \mu\text{g cm}^{-2}$ , which correspond to thin, middle and thick film regions. TDD was obtained from the sum of DD for PM6, Y6 and IT-4F, respectively.

## Supplementary Note 4. Best Printable Formulation Prediction

Manufacturing of OPVs by industrial printing methods demands a robust formulation that shows consistently high performance regardless of thickness variation, preferably in the range of 200 nm or thicker. We first filtered compositions with high PCE ( $> 8\%$  was chosen in this case) from the whole prediction data (8,000,000 compositions) obtained from ML (Fig. S11). The axes of deposition density of each material were then converted to the relative amount of materials so that the z-axis became the thickness (total deposition density) of the film.

All x-y coordinates (PM6 fraction in total solid as x, Y6 fraction in acceptors as y) were then rounded to two decimal places with closest even numbers. The data points with the same x-y coordinates were then counted and shown in Fig. S12. We discovered some coordinates showed the same number of data. The coordinate with higher variance (as shown in Table S5) was considered to be preferable as it also means wider TDD range. Therefore, the x-y coordinate with the highest count and the higher variance was selected as the BPF.

It is noted that predicted PCEs from each trial were not identical. However, the PCE difference was minor so it did not significantly change the result of BPF and BEF.

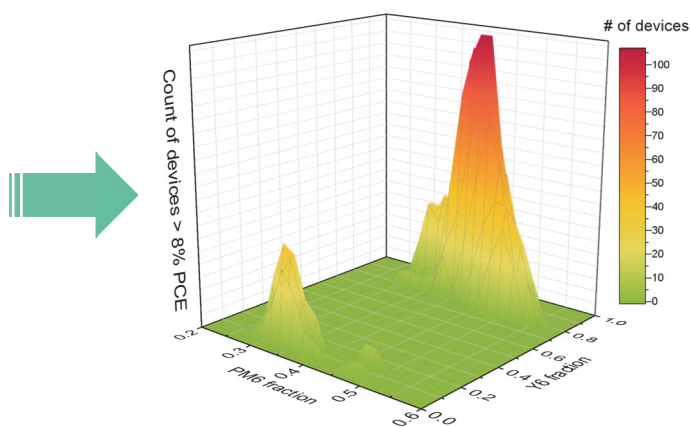


**Figure S11.** Procedure for filtering and converting axes  $>8\%$  PCE data. Representation of filtered data to be PCE  $>8\%$  from the whole prediction data and converted axes to PM6 and Y6 fraction.

| PM6 Fraction | Y6 Fraction | Corresponding PCE |
|--------------|-------------|-------------------|
| 0.264706     | 0.106667    | 8.0317            |
| 0.271845     | 0.053333    | 8.0496            |
| 0.278846     | 0.053333    | 8.0739            |
| 0.285714     | 0.053333    | 8.0704            |
| 0.27451      | 0.108108    | 8.0403            |
| 0.271845     | 0.106667    | 8.1036            |
| 0.281553     | 0.108108    | 8.0646            |

*Rounding to  
2 decimal places  
with closest even  
numbers*

| PM6 Fraction | Y6 Fraction | Corresponding PCE |
|--------------|-------------|-------------------|
| 0.26         | 0.10        | 8.0317            |
| 0.28         | 0.06        | 8.0496            |
| 0.28         | 0.06        | 8.0739            |
| 0.28         | 0.06        | 8.0704            |
| 0.28         | 0.10        | 8.0403            |
| 0.28         | 0.10        | 8.1036            |
| 0.28         | 0.10        | 8.0646            |



**Figure S12.** Procedure of counting number of data at each PM6 and Y6 fraction coordinates. Procedure rounding filtered data to generate overlapping coordinates and counting the number of data at each PM6 and Y6 fraction coordinates. It is shown as a 3D surface plot of the number of data with PCE > 8% depending on PM6 and Y6 fraction.

**Table S5.** Summary of top BPFs obtained from Fig. 4b and c.

| Label | PM6<br>Fraction | Y6<br>Fraction | Counts | Variance |
|-------|-----------------|----------------|--------|----------|
| 1     | 0.42            | 0.88           | 106    | 11.0     |
| 2     | 0.44            | 0.88           | 106    | 8.17     |
| 3     | 0.42            | 0.84           | 102    | 4.55     |
| 4     | 0.40            | 0.88           | 95     | 12.8     |
| 5     | 0.42            | 0.86           | 87     | 5.39     |
| 6     | 0.44            | 0.82           | 87     | 4.63     |
| 7     | 0.44            | 0.84           | 85     | 5.14     |
| 8     | 0.40            | 0.84           | 83     | 3.16     |
| 9     | 0.44            | 0.86           | 81     | 4.70     |
| 10    | 0.46            | 0.82           | 78     | 5.20     |

**\* Predicted Best Printable Formulation**

PM6:Y6:IT-4F (w/w)  
 $= 0.42:(1-0.42)*0.88:(1-0.42)*0.12$   
 $= 0.42:0.5104:0.0696$   
 $\approx \mathbf{1:1.22:0.17}$

**Table S6.** Summary of top BEFs obtained from Fig. 4b.

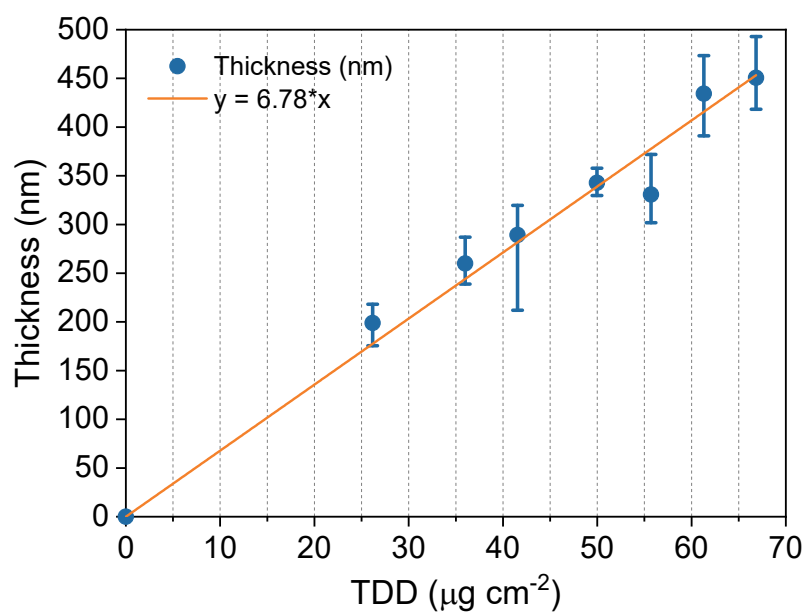
| PM6<br>( $\mu\text{g cm}^{-2}$ ) | Y6<br>( $\mu\text{g cm}^{-2}$ ) | IT-4F<br>( $\mu\text{g cm}^{-2}$ ) | PCE<br>(%) |
|----------------------------------|---------------------------------|------------------------------------|------------|
| 12.00                            | 13.00                           | 3.25                               | 9.4046     |
| 11.75                            | 13.00                           | 3.25                               | 9.3985     |
| 11.75                            | 12.75                           | 3.25                               | 9.3631     |
| 12.00                            | 12.75                           | 3.25                               | 9.3419     |
| 13.00                            | 12.00                           | 3.00                               | 9.3199     |
| 12.75                            | 12.00                           | 3.00                               | 9.2900     |
| 11.75                            | 12.25                           | 3.25                               | 9.2683     |
| 12.00                            | 12.25                           | 3.25                               | 9.2486     |
| 13.25                            | 12.00                           | 3.00                               | 9.2461     |
| 11.75                            | 12.50                           | 3.25                               | 9.2424     |

**\* Predicted Best Efficiency Formulation**

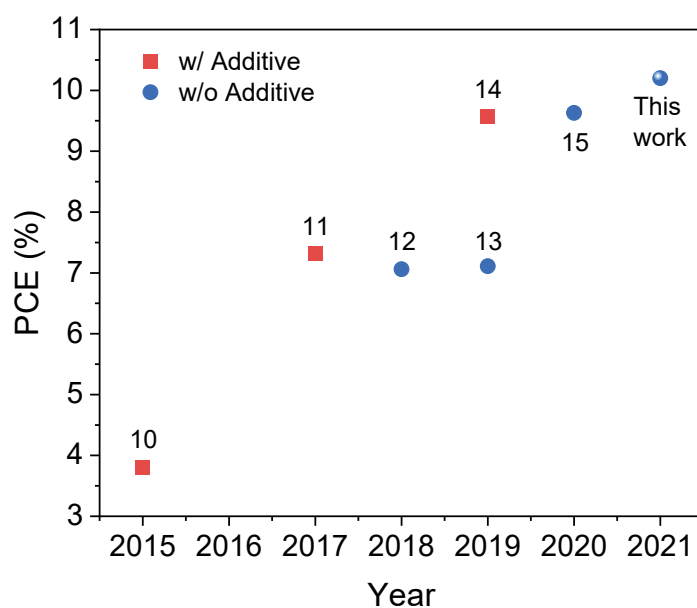
PM6:Y6:IT-4F (w/w)  
 $= 12:13:3.25$   
 $\approx \mathbf{1:1.08:0.27}$

**\* Optimized Total Deposition Density by Machine Learning**

$= (12+13+3.25) \mu\text{g/cm}^2$   
 $= \mathbf{28.25 \mu\text{g/cm}^2}$



**Figure S13.** Film thickness measurements. Film thicknesses of slot die coated R2R devices based on BEF depending on TDD.



**Figure S14.** Efficiency chart of R2R processed OPV devices in recent reports.

**Table S7.** Summary of photovoltaic parameters of the following R2R processed OPV devices in recently reported literature.

| Active layer                             | Additive | Year | $J_{sc}$<br>(mA cm <sup>-2</sup> ) | $V_{oc}$<br>(V) | FF   | PCE<br>(%) | Ref.      |
|------------------------------------------|----------|------|------------------------------------|-----------------|------|------------|-----------|
| PBDTTTz-4:PC <sub>71</sub> BM            | O        | 2015 | 17.80                              | 13.10           | 0.48 | 3.80       | 16        |
| PTB7:PC <sub>71</sub> BM                 | O        | 2017 | 15.91                              | 0.76            | 0.63 | 7.32       | 17        |
| PPDT2FBT:PC <sub>71</sub> BM             | X        | 2018 | 15.61                              | 0.74            | 0.60 | 7.06       | 1         |
| PBDB-T:ITIC                              | X        | 2019 | 14.20                              | 0.86            | 0.58 | 7.11       | 2         |
| PTB7-Th:PC <sub>71</sub> BM<br>:COi8DFIC | O        | 2019 | 23.50                              | 0.68            | 0.60 | 9.57       | 18        |
| PBF-QxF:Y6                               | X        | 2020 | 20.39                              | 0.76            | 0.62 | 9.63       | 19        |
| PBDB-T-2F:Y6:IT-4F                       | X        | 2021 | 21.6                               | 0.78            | 0.60 | 10.2       | This work |

**Table S8.** Summarized photovoltaic parameters and TDDs with estimated thickness range of the R2R processed OPVs obtained from two independent batches of BEF and BPF. The whole dataset at each formulation were sub-grouped into 7  $\mu\text{g cm}^{-2}$  of TDD intervals (which corresponds to 40–50 nm of film thickness) then the photovoltaic parameters which belong to same sub-group were averaged at each sub-group.

| Condition | TDD Range ( $\mu\text{g cm}^{-2}$ )<br>(Cal. Thickness, nm) | $J_{\text{sc}}$<br>( $\text{mA cm}^{-2}$ ) | $V_{\text{oc}}$<br>(V)    | FF                        | PCE<br>(%)                |
|-----------|-------------------------------------------------------------|--------------------------------------------|---------------------------|---------------------------|---------------------------|
| BEF       | 14–21<br>(90–140)                                           | 18.8<br>(17.8 $\pm$ 1.04)                  | 0.80<br>(0.79 $\pm$ 0.01) | 0.66<br>(0.64 $\pm$ 0.02) | 9.87<br>(8.93 $\pm$ 0.54) |
|           | 21–28<br>(140–190)                                          | 21.4<br>(19.1 $\pm$ 1.07)                  | 0.80<br>(0.79 $\pm$ 0.01) | 0.59<br>(0.62 $\pm$ 0.01) | 10.2<br>(9.34 $\pm$ 0.52) |
|           | 28–35<br>(190–240)                                          | 21.6<br>(21.1 $\pm$ 0.71)                  | 0.80<br>(0.78 $\pm$ 0.01) | 0.60<br>(0.59 $\pm$ 0.02) | 10.2<br>(9.78 $\pm$ 0.36) |
|           | 35–42<br>(240–280)                                          | 20.3<br>(20.0 $\pm$ 0.26)                  | 0.78<br>(0.77 $\pm$ 0.01) | 0.57<br>(0.54 $\pm$ 0.03) | 9.02<br>(8.31 $\pm$ 0.52) |
|           | 42–49<br>(280–330)                                          | 19.9<br>(20.2 $\pm$ 0.38)                  | 0.76<br>(0.76 $\pm$ 0.01) | 0.48<br>(0.42 $\pm$ 0.02) | 7.27<br>(6.44 $\pm$ 0.38) |
|           | 49–56<br>(330–380)                                          | 20.2<br>(18.3 $\pm$ 1.19)                  | 0.76<br>(0.74 $\pm$ 0.03) | 0.44<br>(0.40 $\pm$ 0.03) | 6.75<br>(5.45 $\pm$ 0.82) |
|           | 16–23<br>(110–160)                                          | 18.6<br>(18.1 $\pm$ 0.49)                  | 0.78<br>(0.77 $\pm$ 0.01) | 0.63<br>(0.63 $\pm$ 0.01) | 9.20<br>(8.79 $\pm$ 0.24) |
|           | 23–30<br>(160–200)                                          | 19.4<br>(19.7 $\pm$ 0.83)                  | 0.78<br>(0.77 $\pm$ 0.01) | 0.61<br>(0.58 $\pm$ 0.03) | 9.15<br>(8.76 $\pm$ 0.16) |
| BPF       | 30–37<br>(200–250)                                          | 20.5<br>(20.2 $\pm$ 0.58)                  | 0.76<br>(0.76 $\pm$ 0.01) | 0.55<br>(0.52 $\pm$ 0.02) | 8.56<br>(7.93 $\pm$ 0.41) |
|           | 37–44<br>(250–300)                                          | 20.8<br>(20.6 $\pm$ 0.25)                  | 0.76<br>(0.76 $\pm$ 0.00) | 0.50<br>(0.47 $\pm$ 0.02) | 7.88<br>(7.45 $\pm$ 0.27) |
|           | 44–51<br>(300–340)                                          | 21.3<br>(20.5 $\pm$ 0.41)                  | 0.76<br>(0.76 $\pm$ 0.00) | 0.44<br>(0.42 $\pm$ 0.02) | 7.12<br>(6.48 $\pm$ 0.36) |
|           | 51–58<br>(340–390)                                          | 20.2<br>(18.9 $\pm$ 0.91)                  | 0.76<br>(0.74 $\pm$ 0.03) | 0.41<br>(0.38 $\pm$ 0.01) | 6.22<br>(5.38 $\pm$ 0.49) |

\* All statistical data were calculated from 20 devices.

\* Thickness was estimated by the calibrated equation (thickness = 6.78 \* TDD) obtained from Fig. S7.

## Supplementary References

1. S. Song, K. T. Lee, C. W. Koh, H. Shin, M. Gao, H. Y. Woo, D. Vak and J. Y. Kim, *Energy Environ. Sci.*, 2018, **11**, 3248–3255.
2. S.-I. Na, Y.-H. Seo, Y.-C. Nah, S.-S. Kim, H. Heo, J.-E. Kim, N. Rolston, R. H. Dauskardt, M. Gao, Y. Lee and D. Vak, *Adv. Funct. Mater.*, 2019, **29**, 1805825.
3. M.-H. Lee, *Adv. Energy Mater.*, 2019, **9**, 1900891.
4. D. L. J. Alexander, A. Tropsha and D. A. Winkler, *J. Chem. Inf. Model.*, 2015, **55**, 1316–1322.
5. F. C. Krebs, *Sol. Energy Mater. Sol. Cells*, 2009, **93**, 394–412.
6. W. Zhao, D. Qian, S. Zhang, S. Li, O. Inganäs, F. Gao and J. Hou, *Adv. Mater.*, 2016, **28**, 4734–4739.
7. S. Zhang, Y. Qin, J. Zhu and J. Hou, *Adv. Mater.*, 2018, **30**, 1800868.
8. C. Zhang and Y. Ma, *Ensemble Machine Learning*, Springer, New York, 1st edn., 2012.
9. P. Geurts, D. Ernst and L. Wehenkel, *Machine Learning*, 2006, **63**, 3–42.
10. N. Altman and M. Krzywinski, *Nat. Methods*, 2017, **14**, 933–934.
11. A. Jog, A. Carass, S. Roy, D. L. Pham and J. L. Prince, *Med. Image Anal.*, 2017, **35**, 475–488.
12. Y. Li, C. Zou, M. Berecibar, E. Nanini-Maury, J. C. W. Chan, P. van den Bossche, J. Van Mierlo and N. Omar, *Appl. Energy*, 2018, **232**, 197–210.
13. L. a. Wang, X. Zhou, X. Zhu, Z. Dong and W. Guo, *Crop. J.*, 2016, **4**, 212–219.
14. S. Langner, F. Häse, J. D. Perea, T. Stubhan, J. Hauch, L. M. Roch, T. Heumueller, A. Aspuru-Guzik and C. J. Brabec, *Adv. Mater.*, 2020, **32**, 1907801.
15. M.-H. Lee, *Org. Electron.*, 2020, **76**, 105465.
16. M. Helgesen, J. E. Carlé, G. A. dos Reis Benatto, R. R. Søndergaard, M. Jørgensen, E. Bundgaard and F. C. Krebs, *Adv. Energy Mater.*, 2015, **5**, 1401996.
17. Y.-C. Huang, H.-C. Cha, C.-Y. Chen and C.-S. Tsao, *Prog. Photovolt.*, 2017, **25**, 928–935.
18. J. Lee, Y.-H. Seo, S.-N. Kwon, D.-H. Kim, S. Jang, H. Jung, Y. Lee, H. Weerasinghe, T. Kim, J. Y. Kim, D. Vak and S.-I. Na, *Adv. Energy Mater.*, 2019, **9**, 1901805.
19. M. H. Yang, S.-J. Ko, N. G. An, D. R. Whang, S.-H. Lee, H. Ahn, J. Y. Kim, D. Vak, S. C. Yoon and D. W. Chang, *J. Mater. Chem. A*, 2020, **8**, 25208–25216.