Design for a sustainability approach to organic solar cell design; the use of machine learning to quantify the trade-off between performance, stability and environmental impact

Tudur Wyn David[a], Jeff Kettle*[b]

   a. School of Computer Science and Electronic Engineering, Bangor University, Dean St, Bangor, Gwynedd, Wales, UK, LL57 1UT (t.w.david@bangor.ac.uk)
   b. James Watt School of Engineering, University of Glasgow, Glasgow G12 8QQ, United Kingdom (Jeff.Kettle@glasgow.ac.uk)

Abstract

Organic Photovoltaics (OPV) are considered one of the best performing PV technologies from an environmental point of view. Many of the constituent materials possess low embodied energies, which can generally be processed and disposed of in a less energy-intensive manner than other PV technologies. There has been an enormous range of materials used in OPVs, however, identification of the optimal materials and device architectures that provide the best environmental profile within this large search space has yet to have been conducted. This is a non-trivial task, because the selection of these materials not only impacts on the environmental profile, but also on the solar cell efficiency and it's operational stability. Here, we have developed a methodology that enables rapid assessment of the trade-off between efficiency, stability, and embodied energy of an OPV using machine learning. To achieve this, a database of OPV data was used, which has been acquired from the literature between 2011-2020 and consists of 1580 device data points. Our results highlight the importance of focusing activity on particular transport layers, substrates, and active layer materials, which are discussed further in the manuscript. We demonstrate that the trained and validated models can predict, with a high degree of confidence, the efficiency, stability, and embodied energy of an OPV. The methodology set out in this work provides a means of identifying optimum device configurations in a rapid manner such that the net energy production is maximised whilst the environmental impact of OPVs is minimised. Materials which show promise towards delivering a positive net energy are PET + barrier layer, PET (substrates), NiOx, ZrOx, CrO2, ZnO, LiF and MoO3 (Transport Layers). Active layer materials which show promise for delivering a positive net energy are DRCN7T, DR3TSBDT, ZnPc, PDPP4T-2F, PCDTBT (donors) and IT-4F, C61 (Acceptors).

1. Introduction

As a result of the EU Renewable Energy Directive and the Circular Economy Action, photovoltaics (PV) should be designed to ensure their manufacturing, selection of materials and end-of-life management is undertaken in an eco-friendly manner. This has particular importance in the context of Organic Photovoltaics (OPVs) because they are at a pre-commercial stage and there is a vast array of materials that have been reported for use in their manufacturer [1–4]. As a result, the adoption of eco-design approaches is vital to maximise the environmental profile of OPVs. There is a wide panorama of environmental assessment metrics, but much research activity has been devoted to determining the '*embodied energies*' of different OPV devices, materials and their various production processes. The embodied energy (in terms of primary energy equivalent) is dependent on the production of the constituent materials and their processing; this includes the raw material extraction, the production of the material, the processing, usage and finally the disposal of the material, depending on the particular scope considered [5].

Over the past 15 years there has been a wide range of architectures and materials used and this has led to a rapid increase in power conversion efficiency (PCE), which is now exceeding 18% for a single junction device [6]. A significant proportion of this research has focused on improving the power conversion efficiency (PCE) of OPVs, and to a lesser extent, their operational stability, scalability and cost. Nevertheless, there has been an increasing focus on improving their environmental profile as well [7–9]. As a result, there is now a large body of available OPV data in the literature concerning OPV efficiency, stability and embodied energy. If this data was combined, environmental screening of OPVs could be undertaken in a manner that has not be done before; where the PCE, stability and embodied energy are optimised concurrently.

To achieve this, advanced statistical analysis such as machine learning (ML) techniques are needed. By deploying data analytical approaches in this manner, it is possible to determine which materials/processes have the greatest impact on the OPV performance, stability and embodied energy and consider the potential trade-offs between these performance metrics. This represents a potential new paradigm in OPV eco-design.

In this work, we have utilised OPV data which has been acquired from work undertaken between 2011-2020. This dataset contains data from around 1580 devices where the PCE and stability data has been reported. In addition, the embodied energy of each device has been calculated. Two principal ML techniques are utilised in order to analyse the trade-off between

PCE, stability and embodied energy of the OPV devices. Firstly, sequential minimal optimisation regression (SMOreg) has been employed to identify which materials have the greatest impact upon the embodied energy. There are a number of papers which use other ML algorithms such as gaussian process regression, random forest and gradient boosting regression tree to model OPV performance parameters [10–12]. Descriptor importance is typically used to determine a ranking of the model attributes but does not offer a ranking of the features within the attribute class [12–14]. Therefore, the method employed in this study complements the previous work by offering a means of ranking the individual materials within the various attributes. Secondly, a genetic clustering algorithm is used to find the optimum architecture which maximises the three metrics (PCE, stability and embodied energy). This multi-metric approach combining energy-based green metrics and LCA affords an extremely useful tool for evaluating the environmental impact of an OPV. This provides a methodology to rapidly scan datasets of OPV materials with associated performance and stability data, and subsequently determine which are too environmentally damaging, too unstable or lack sufficient performance without the need for exhaustive experimentation.

2. Methodology

2.1 Introduction

For the purposes of this study, a database of OPV performance and stability data was collated. This dataset consisted of data collected by Krebs et al. at the Danish Technical University (DTU) who ran the "lifetime predictor" on the Plastic photovoltaics website from 2011-2017 and additional data was added by manually scraping journal articles between 2017-2020 [15]. An additional complexity was sourcing data containing embodied energy of the individual layers and processes, discussed further in the next section.

Each device data point consisted of the following information; device architecture, testing condition (typically following an ISOS protocol [16]) and the performance and stability metrics. The latter is often not explicitly stated within the text, so this was extracted from figures using a plot digitizer. Figure 1 shows a schematic of the generic structure of the OPVs employed in this analysis with each device component defined as: Substrate (S), Electrode 1 (E1), Transport Layer 1 (TL1), Active Layer 1 (A1), Active Layer 2 (A2), Transport Layer 2 (TL2), Electrode 2 (E2), Encapsulation (Encap). The active layer is split into two components to represent the donor and polymer materials.

2.2 Embodied Energy and Net Energy Determination

Data for the individual embodied energies of the materials and processes was also sourced from the literature and used to calculate the total embodied energy of the device (E_{Emb}). The authors estimate that there are 187 papers in the literature that have applied LCA to OPVs between 2011 and 2021 (based on a Web of Science search). Nevertheless, there are additional papers in the dye sensitised solar cell and perovskite communities that have made embodied energy calculations for materials/processes and these have been used for this study. Using the embodied energy values from the literature, the database containing OPV structure and performance parameters was populated with the literature values of embodied energy for each individual material and process in order to calculate E_{Emb}. Where possible, an average of three embodied energy calculations was taken. It is estimated that 85% of data was sourced from other papers. However, a small subset of data required calculations by these authors, which was computed with Simapro using the same approaches developed by Krebs [1, 17–19]. One of the main challenges related to inconsistent units or inconsistent scales of productions, so assumptions were made in order to form a consistent unit system. For example, many sources quote embodied energy as either MJkg^{-1} or MJm^{-2}. The values quoted in MJkg^{-1} were converted to MJm^{-2} by considering the volume and density of the materials in the device using the thickness and active areas quoted in papers.

Another major challenge is related to the large number of recently reported active layer materials, many of which have no reports in the literature of their embodied energy ($E_{Active1}/E_{Active2}$). In these circumstances, estimates were made based on the number of synthesis steps. Stiebitz produced a technical report where the embodied energy was considered as a function of the number of synthesis steps [20]. In the study by Stiebitz, three different polymers were studied: P3HT, PTB7 and PCDTBT and the embodied energy determined as a function of the number of synthesis steps, which were 5, 12 and 9 synthesis steps for each polymer, respectively. The embodied energy for several other polymers are additionally deterined and included in the analysis. These were MEH-PPV, PPE-PPV, DPPTPTA and PFDT2BT. These additional polymers are added to the analysis of Stiebnitz to increase the accuracy of the regression model. From this the average embodied energy per step is calculated and used to determine the embodied energy for the different polymers based on the number of synthesis steps for each material. Figure 2 illustrates the fitted plot for the embodied energy per mole of polymer as a function of synthesis steps. Information regarding the accuracy of this is limited due to few reports on the embodied energies of more recent polymers. Therefore,

this assumption does represent a limitation of the work, which could be overcome as more accurate values for embodied energy become avavilable or more robust methods of estimation are derived. In addition, the embodied energy per molar mass was determined from several similar materials found in the literature, which were not present in the database, and subsequently converted based on the molar mass of the material and the quantity used. Using this, a regression model was formed that estimated embodied energy of newer polymers and acceptor materials based upon the number of synthetic steps. The deposition process of each material is also considered in this work and values for the different process are incorporated into the embodied energy value when only the material embodied energy is given in the literature. Often, papers will explicitly state whether the values include the processing energy. When there is ambiguity, comparison with similar materials will show whether the value has been incorporated or not.

Once the embodied energy of each individual material was found, or calculated, the total embodied energy (E_{Emb}) for each OPV device in the data set can be calculated using equation 1,

$$E_{Emb} = \sum (E_{Mat} + E_{Process}), \tag{1}$$

where E_{Emb} is the total embodied energy of the OPV device including the embodied energy of the materials, E_{Mat} and the embodied energy of the processing for each material, $E_{Process}$. The embodied energy of the materials can be defined as,

$$E_{Mat} = E_{Sub} + E_{TL1} + E_{Active1} + E_{Active2} + E_{TL2} + E_{Elec2} + E_{Encap}, \tag{2}$$

where $E_{Sub}, E_{Elec1}, E_{TL1}, E_{Active1}, E_{Active2}, E_{TL2}$ and E_{Elec2} are the embodied energies of each constituent material in each layer (Substrate, Electrode 1, Transport Layer 1, Active 1, Active 2, Transport Layer 2 and Electrode 2 respectively). $E_{Process}$ incorporates the energy required for processing each material, for example spin coating, evaporation, screen printing and slot die. An approximate exponential decay curve, for the maximum power point, is fitted between T_0 (time at E_0) and T_{80}, and a linear degradation between T_{80} and T_{S80}.

Once E_{Emb} had been calculated, a method to compare the trade-off between PCE, stability and embodied energy was required. To this end, the authors developed a metric known as "Net energy" (E_{Net}) which calculated the energy yield of the solar cell operated at AM1.5G for a time period of 0 hours to T_{S80} lifetimes and subtracted the energy used in the OPV manufacture

(i.e the E_{Emb}). T_{S80} is defined as the time taken for an OPV to degrade to 80% of its stabilised initial output power, to differentiate from the stability metric T_{80}, defined as the time taken to drop to 80% of the initial power at $t = 0$ $hours$ and largely covers the 'burn-in' period. The E_{Net} is calculated during both time periods by integrating the functional power output, $P(t) = I_{MPP}(t)V_{MPP}(t) = P_0 e^{-\alpha t}$, over time,

$$E_{Net} = \left(\int_0^{T_{80}} P_0 e^{-\alpha_1 t} dt + \int_{T_{80}}^{T_{S80}} P_0 e^{-\alpha_2 t} dt \right) - (E_{Emb}), \qquad (3)$$

where $I_{MPP}(t)V_{MPP}(t)$ is the time – variation of the power output between time $t = 0$ and T_{S80}, P_0 is the power at $t = 0$ and α_1 and α_2 are fitted exponential decay constant. The numerical value for α_1 is in the range $[6 \times 10^{-5}, 5.6]$ and α_2 is in the range $[7 \times 10^{-5}, 0.092]$. Since the exponential decay curve is being approximated, the method for determining E_{Net} will possess some error. It is worth noting that only data acquired with ISOS – L standards has been included in the database, as it is impossible to estimate the light stability of OPVs measured using the ISOS-D protocols or other test conditions.

Once E_{Net} for a particular data point has been calculated, determining the optimum combination of materials to achieve the greatest E_{Net} can be achieved by adopting ML approaches; the optimisation processes can be applied to determine the best material combinations for maximising the E_{Net} output. A schematic flow chart summarising the method employed is shown in Figure 3, illustrating how the material and energy costs are acquired and the stages of analysis using ML and genetic algorithms.

2.3 Sequential Minimal Optimisation regression

The ML algorithm, sequential minimal optimisation regression (SMOreg) has been adopted in this work to rank how different materials and layers impact on the *embodied* and *net energies*. SMOreg operates by finding an optimal hyperplane which fits the data based on the response parameter being modelled. This is achieved by ensuring that the output value, y (E_{Emb} or E_{Net}), of each data entry deviates less than a required error, ε, from the fitted regression line $f(x)$. The functional form of the hyperplane is determined by optimising the weights of each attribute using variational calculus and the method of Lagrange multipliers [21]. The acquired model can subsequently be analysed in terms of the derived weights with more positive weights corresponding to attributes which will lead to an increase in either E_{Emb} or E_{Net}. whilst more

negative weights correspond to attributes which will reduce E_{Emb} or E_{Net}. More details of the SMOreg algorithm can be found in [22]. The ML algorithm takes the dataset inputs, which are represented as strings for the various material types and converts each into a binary number which the ML algorithm can read. This is automatically implemented within WEKA 3.8.

2.4 Genetic Algorithm Clustering

A genetic algorithm (GA) clustering technique was also employed as a means of determining the optimum material sets for maximising the E_{Net} output of OPVs. GAs are randomized search and optimization techniques inspired by the principals of evolution and natural selection [23].

In this work, a clustering approach is employed using the 'GenClust++' algorithm [24]. In this approach, each cluster is determined based on a similarity metric which uses the Manhattan Distance Class, where the distance is determined based on a square grid formation and the shortest path is determined from the grid lines. For categorical values, the strings are compared by considering the number of bit positions in which the two strings differ.

When applying the GA algorithm, the combination of different materials for each layer is optimised to form the entire device stack. This is achieved by clustering similar characteristics, based on E_{Net} value, whereas when using the SMOreg ML algorithm, the best attributes for enhancing the E_{Net} are found; these could all belong to a single attribute class (for example the electrode) and not take into consideration the optimisation for combining each device layer together. Therefore, even if each device layer is determined by considering the full list of weights derived using SMOreg, the identified device architecture would not be an optimised solution, as in the case of using the GenClust++ algorithm.

Results and Discussion

3.1 Data exploration

The data was initially visualised in order to better understand the data structure. Figure 4 shows the categorical violin plots for (a) E_{Emb} and (b) E_{Net} of the dataset, prior to analysis using ML. These plots intuitively show the median, interquartile range and 1.5 × interquartile range. In addition, the plot shows the kernel density estimation to illustrate the distribution of each

parameter. In the case of the E_{Net}, the data was split for positive and negative values and plotted on separate axis to allow for a logarithmic scale to be used. A positive value indicates that the energy produced by the OPV during the operation is greater than the energy used in the manufacture. Interestingly, only 15% of the device architectures are found to be positive E_{Net} generators. The reason for this is twofold; firstly, there are a number of devices with poor stability in our data set, and consequentially the energy that is generated cannot surpass the energy needed for manufacture because the device operates for too short a period of time. Secondly, many small devices are made using glass substrates so possess a high E_{Sub} . From our data, unless the T_{S80} stability is greater than 230 hours, the energy generated will never overcome the energy used in the manufacture of the glass substrate on its own (assuming an initial PCE of 8%).

In addition, Figure 5 illustrates scatter graphs between (a) E_{Emb} and PCE, (b) E_{Net} and PCE, (c) E_{Emb} and T_{80} and (d) E_{Net} and T_{S80}. From Figure 5 (a) and (c), it is observed that the E_{Emb} is similar for many of the data points. This is because many devices employ similar architectures and use similar materials. For example, 58% of devices in our dataset used a P3HT:PCBM active layer blend; 78% used an ITO electrode 1; 93% used either silver or aluminium as electrode 2 and 77% used a glass substrate. From Figure 5 (a) and (c) it can be seen that the data is grouped in two distinct regions, centred at approximately 350 MJm^{-2} and 40 MJm^{-2}. Most cell devices reside in the group near the 350 MJm^{-2} value, whilst most modules reside in the group near the 40 MJm^{-2} value. This is principally due to the materials used by the two types of devices which govern the total E_{Emb}. For Figure 5 (b) and (d), many devices are again grouped with similar E_{Net}, primarily due to the similar E_{Emb}, however, a small positive correlation can now be observed. This indicates that as E_0 and T_{S80} increase, so does the E_{Net} of the device. This is as expected; a device with a higher initial efficiency will be able to produce more energy, and a device which operates for longer will also produce more energy over its lifetime.

3.2 Analysis of Embodied and Net energies using SMOreg

The SMOreg ML algorithm was applied to the dataset and used to analyse trends in the E_{Emb}. In this work, the data was analysed using cross – validation approach to minimise overfitting of the data. Cross – validation is a resampling procedure for training and testing an ML algorithm where several folds of train/test splits are produced. This improves the reliability of

the derived model by reducing overfitting. The regression results are shown in Figure 6 (a). The fitting parameters for the three subsets are shown in Table 1 and a strong correlation coefficient (CC) of 0.988 is achieved when using all structural components. The high correlation should not be a surprise since the E_{Emb} was calculated from LCAs using similar approaches. In relation to the model fitting, positive weights indicate materials and processes with a higher E_{Emb} and negative weights relate to materials and processes with a lower E_{Emb}. The magnitude of the weight indicating the relative importance of the factor. Table 2 provides a ranking of the fifteen most beneficial and fifteen most detrimental materials for E_{Emb} based upon the SMO algorithm weightings. In this table, the name represents the device layer and the material identified in the parentheses. Inspection of the weights show that the most beneficial attributes are *substrate*, *electrode 1* and *TL1* materials, in rank order, with the *substrate (20%)* and *electrode 1 (13%)* categories all occupying the top 50% of the weights. The most detrimental attributes are *electrode 1, transport layer 1* and *2, active layer* and *electrode 2* in rank order with *electrode 1* and *TL1* materials all occupying the worst 50% of weights.

For substrate material the most beneficial materials are PET, PET with a barrier layer, glass and quartz, in rank order. There was not enough meaningful data from other substrate materials in our dataset. Based upon the dataset, it is clear that one of the highest impact modifications OPV manufacturers can make to reduce E_{Emb} is to use a PET substrate. The main alternative to PET is to use glass, which has a much higher E_{Emb}. For electrode 1 the most beneficial materials are printed silver (Ag) or PEDOT:PSS. The second most detrimental device layers are the electrodes and 27% of the positive weightings corresponding to electrodes. The most detrimental attribute is identified as ITO for E1. For E2, the most detrimental material is evaporated Ag. The most beneficial TL1 materials are identified as CuI, ZnO and PEDOT:PSS. This provides some promise that commonly used transport layers such as ZnO and PEDOT:PSS do not carry a large environmental impact. However, it should be noted that TL1, TL2 and active materials do not have as significant an impact on overall E_{Emb} as substrate and electrode materials. It is also noteworthy that ZnO nanoparticles used as TL2 are identified as being beneficial as well as the active material IDTBRIDFBR. In rank order, the most detrimental TL1 materials are NiO_x, CrO_2 and TiO_2 and for TL2 the ranking is TiO_x, MoO_3 ZrO, LiF and V_2O_5. This is as expected due to the high manufacturing costs for many metal oxide materials.

Subsequently, the SMOreg algorithm has been applied to the calculated E_{Net} output of the OPVs, calculated using equation 3. E_{Net} considers the energy generated over the course of the OPV and subtracts the E_{Emb}. The regression result is shown in Figure 6 (b) and the fitting parameters are shown in Table 1. Using the E_{Net} produced, a good fitting is achieved for the test set with a CC of 0.838. Subsequently, the weights obtained from the SMOreg model can be analysed and ranked as the ten most beneficial and ten most detrimental for increasing the E_{Net}, as shown in Table 3. Inspection of the top 50% of weightings from the SMOreg algorithm show that the most beneficial attributes in order to achieve high E_{Net} is by adjusting *Transport Layer 1* and *2*, *Active layer 1* and *2* and *substrate*s. The transport layers occupy approximately half of the positive weights. The most detrimental attributes are *active layers 1* and *2, transport layer 1* and *2, electrode 1* and *2* in rank order.

Considering the substrate, using a PET layer coated with a barrier layer is found to be the most beneficial for improving the E_{Net} whilst using a quartz substrate proved to be the most detrimental. The former shows the benefit of using a barrier film; whilst using a barrier layer results in an increase in E_{Emb}, the E_{Net} increases more significantly since the additional barrier layer ensures that the OPV remains operational and generating energy for a longer period of time. A number of reports have used a quartz substrate in OPVs; whilst this gives the better UV transmittance, it's clearly a poor choice as a substrate given the high E_{Emb}.

For the bottom electrode (Electrode 1), a composite electrode of Chromium – Aluminium proves to be beneficial, and Ag and Au are found to be detrimental for E_{Net}. The likely reason for this is that there are a large number of poorly performing small devices (i.e. cells) in the literature. As they survive for <100 hours during light soaking, they cannot recover the energy needed to manufacture the OPVs. Furthermore, the objective of many studies is not on maximising E_{Net} and represent proof of concept for novel materials. However, as the technology matures, the findings of this study should be updated to include materials and devices focussed on E_{Net} enhancement. Nevertheless, this methodology provides a means of directing research priorities. There are some interesting characteristics; for example, the use of NiO_x also leads to an increase in the *embodied energy*, but due to the enhanced stability of devices made with NiO_x, it found to significantly improve the E_{Net} output of OPVs.

For both improving and reducing the E_{Net}, several active layer materials were identified, and these results could be used to optimise combinations of different active materials for improved E_{Net} output. Notably, the three most beneficial active layer materials are DR3TSBDT, IDFBR

and IDTBRIDFBR. In contrast, the most detrimental active material is PBDTTT-C-T, which is a widely reported material, but possesses a low stability. Other active layer materials such as PBDTTTPD are also identified as having a detrimental effect on E_{Net}.

The second transport layer (TL2) plays a prominent role governing the E_{Net} output, primarily due to its effect upon device stability. ZnO nanoparticles and MoO$_3$ are found to improve the E_{Net} when used as TL2, whilst V$_2$O$_5$ and ZrO are found to be detrimental. No top electrode (Electrode 2) materials were identified for improving the E_{Net}. However, the use of an Au for Electrode 2 is clearly detrimental and should be avoided. Finally, it should also be noted that the E_{Emb} was used as a predictive attribute for E_{Net} calculations and found to be a negative influence. This means that as the E_{Emb} increases, if the stability and PCE stay constant, the E_{Net} decreases and vice versa. This demonstrates the importance of selecting materials from the outset that possess low manufacturing costs whilst also maintaining a balance between using materials with a high energy output potential.

The analysis has been subsequently repeated using only the active layer materials as the predictive attributes for E_{Net} and discounting the other OPV layers. Whilst the biggest improvements in E_{Net} are related to substrates and E1, there is a much wider range of alternative options of active layer materials as new donor and acceptor materials are regularly being reported. Table 1 shows the fitting parameters corresponding to the regression fitting shown in Figure 6 (c), which shows a CC of 0.719. The drop in CC is to be expected as the algorithm is trying to calculate E_{Net} without many of the factors such as electrodes and transport layers which are needed for a more accurate model. Removing these factors allows for investigation of how the active layer materials impact E_{Net} alone. The CC values obtained deomstrate that a high value can be obtained for E_{Emb} whilst a moderate value can be obtained for E_{Net}; this is as expected given the nature of how these values were determined. Furthermore, these values prove to be higher or comparable to previously reported CCs when modelling OPV performance. For example, a CC of 0.79 was reported by Sahu *et* al. when predicting the PCE using of OPVs based on 13 material properties [12]; Nagasawa *et* al. reported a CC of 0.62 when predicting the PCE based on molecular fingerprints of conjugated polymers used in OPVs [25] and Wu *et* al. reported CC values of 0.54, 0.59 and 0.70 when using linear regressions, multinominal logistic regression and random forest, respectively, when modelling modelling the PCE based on various donor/acceptor pairs in non-fullerene organic solar cells [26]. Table 4 reports the results of the SMOreg weightings when considering only active layer materials, and

hence their relative effect upon E_{Net} calculations. It should be noted that the adopted method does yield an absolute percentage error when predicting the net energy, and readers using our dataset should be aware that this exists. The absolute error is measured at 36%. However, this is not unexpected given the high degree of variability observed in the stability data across devices which employ similar materials. This can be attributed to lab-to-lab variations in fabrication such as minor variations in the testing methodology, material purity, processing techniques and human error. Similar values of absolute errors have also been observed by others when analysing large datasets of OPV data using machine learning [27, 28, 29]. Additionally, the error in the model when predicting high net energies could be reduced by solely considering high performance devices; this would reduce variability and produce a model specifying in identifying high net energy materials range only. Finally, our approach allows readers to rapidly evaluate which materials to use in OPV to assess their sustainability and performance. Nevertheless, it should be noted that to verify the findings of this study, experimental work should be conducted to fabricate devices using the identified materials and their net energies determined.

Notably, all detrimental attributes are the donor active material (A1) whilst the beneficial attributes are a mixture of donor and acceptor (A1 and A2). It should also be highlighted that there are only 12 positive weightings from the dataset. This is because the dataset contains active layers, such as those made with DRCN7T, PBDTTT and ZnPc, where there are relatively few literature reports of the stability, but it is relatively high, so the algorithm creates a small number of large positive weights, and a large number of negative weights. NDP2 doped DiNPB, C_6PcH_2 and SubPc are the three most detrimental materials. Significantly, some materials corresponding to recent reports since 2019 are identified as being beneficial. These include IDTT, DR3TSBDT, IDTBR and IDFBR. Non-fullerene materials such as IDTT and IT-4f are additionally identified as being beneficial. PCBM-61, which is also a commonly used material, is also found to be beneficial. These results principally allow the different possible active layers to be screened and should provide guidance for future experimental verification.

3.2 Genetic Algorithm Clustering

Clustering ML techniques have been applied to the dataset which use a genetic search method based on the 'GenClust++' method discussed earlier. The benefit of using a clustering

approach is that it allows groups of related attributes to be associated with each other based on a similarity metric derived from the E_{Net}. This provides a means of identifying realistic device architectures based on fitting to the E_{Net} value; the mechanics of the genetic clustering algorithm enforces that each device layer is optimised simultaneously for E_{Net} classification. This provides added information over the SMOreg approach by taking into consideration the feasibility of combining the different materials to form a complete device stack. The SMOreg approach considers each attribute from all device layers individually and ranks them based on their relative benefit subsequently allowing the overall ranking of the different materials to be determined.

The E_{Net} are now categorised in groups of equal distribution such that each class contained a similar number of data points. Where the same E_{Net} values overlap across different classes, modifications were made to the distribution of instances whilst maintaining as equal a distribution as possible. Subsequently, the clustering algorithm was applied to the dataset and the classes to cluster evaluation was employed. As a result, the attributes related to energy factors such as the materials and encapsulation are clustered, and each cluster is assigned to a particular E_{net} class. The results of the clustering algorithm are shown in Table 5. The raw E_{Emb} and E_{net} values were not included during the clustering process but are included in Table 5 for clarity and interpretation of the results.

In Table 5, each row corresponds to a single cluster and the column headings represent the attributes used for clustering. The rows have been ranked from lowest E_{Net} at the top of the table to highest E_{Net} at the bottom. The column containing the E_{Emb} and E_{Net} values give the calculated values which allows for assessment of the data. The results shown in Table 5 show how different combinations of materials can be clustered based upon their E_{Net}. The materials in each cluster are selected based on the stages of the genetic algorithm clustering approach where the mode material for each layer is selected as being representative of that layer in the cluster after the genetic operations of "crossover", "elitism", "mutation" and "cloning" [24]. Based upon the data in Table 5, an OPV cell with the following structure would lead to the greatest E_{Net} output:

Glass|ITO|PEDOT:PSS|PCDTBT|PC$_{71}$BM|Ca|Ag

In addition, the clustering results indicate that the OPV modules, with the highest E_{Net} are obtained with the structure:

PET|Ag|PEDOT:PSS|P3HT|PCBM|PEDOT:PSS|Ag.

It should be noted that these combinations of materials do not necessarily occur in the dataset and are derived based on the clustering algorithm, employing genetic search methods. It is also important to note that the dataset contains relatively fewer modules (13%) than cells (87%) and that the modules contained within that dataset tend not to use some of the newer low bandgap polymers or small molecules. Modules represent technologies which have progressed from laboratory conditions to commercial application. Therefore, as the number of modules tested grow and can be added to the dataset, the clustering model will give a clear indication on the viability of the modular setups and the benefit of encapsulation techniques. Furthermore, it should be noted that Ag is identified as the best bottom electrode for E_{Net} enhancement in modules. This seems to contradict the findings in Table 3, which identifies Ag as a detrimental material for E_{Net}. However, the clustering algorithm takes into account the possible combination of materials for producing the entire module, whilst the SMOreg algorthm does not. Since most modules employ Ag as the bottom electrode in a grid formation, the clustering algorithm determines this to be the optimised material option, given the other materials used such as PET substrate. It can be seen that the majority of materials identified in the clusters are common materials such as ITO, PEDOT:PSS, P3HT and PCBM. However, interesting results can be seen when looking at other materials identified and where they reside in the overall ranking. Since the SMOreg algorithm considers the overall ranking of all materials used in the dataset when modelling the E_{Net}, it is often more suited for finding less common materials since the optimisation of the entire device stack is not imposed. Comparison between the results for SMOreg and the genetic clustering additionally allows similar materials to be identified when using both methods. For example, PCDTBT is identified using the genetic algorithm clustering for the top class of E_{Net} and is also found to be a beneficial attribute under the SMOreg algorithm. The SMOreg algorithm determines PET substrates to be beneficial for *net energy* and is identified for architectures having E_{Net} classes between "L" and "O" using the genetic algorithm clustering. Additionally, MoO_3 is identified in the top 15 weights using the SMOreg algorithm and also identified for an architecture having a E_{Net} class of "Q". It is additionally noteworthy that more recent efficient active materials such as PM6:Y6 are not identified by the clustering algorithm as possessing a high E_{Net}. However, since there are a few reports for these efficient materials which display high instability, this counterbalances the

efficient reports for materials such as PM6:Y6. This leads to the algorithm placing a greater significance on the more commonly used materials which have experinced greater development and have a higher number of high stability reports. This illustrates how this method does not pin point individual, high performing, examples, but holistically examines the entire dataset. As the number of available reports increases for novel materials, the algorithm will be able to identify more promising materials.

Even though the highest E_{Net} is achieved using cells, modules are generally displaying improved E_{Net} in comparison to cells, with the E_{Net} ranging between class 'L' and class 'O'. This could be related to the encapsulation since almost all module data used encapsulation whereas many cell devices did not (45%). In addition, modules are made using less intense processing technologies such as slot die coating and tend not to use thermal evaporation, which is more energy intensive. For the substrate material, a trend can be seen where glass and quartz lead to low E_{Net} whilst PET leads to higher E_{Emb}, although this trend is not absolute. A similar trend can be seen for Electrode 1 with ITO leading to lower E_{Net} and silver leading to higher E_{Net}. Interestingly, metal oxides do not appear as often as one would expect in high E_{Net} classes, apart from ZnO. This is due to their large E_{Emb}, as compared with PEDOT:PSS. However, there is significant variance in the data. For example, non-fullerene acceptors (NFAs) display some promise, however, there is variance in the data with some NFAs displaying good stability and others displaying poor stability. ITIC is identified for classes 'N' and 'O', indicating they relate to higher E_{Net} clusters. For modules, all-inkjet processing appears as the most promising deposition technique for silver electrodes, as well as using conductive PEDOT:PSS.

In order to have confidence that the derived device architectures are experimentally feasible, the results from the clustering calculation has been compared with the literature. All of the device architectures found in Table 5 have been reported previously in the literature. This illustrates how the algorithm is capable of determining realistic configurations based on the dataset. For example, the device architecture determined in cluster 19 (highest E_{Net} architecture) has been previously reported for outdoor operation and a T_{80} of approximately 410 hours was achieved [16]. In contrast, the lifetime of the architecture determined in cluster 1 (lowest E_{Net} architecture) was tested under continuous 1 Sun illumination and a T_{80} of approximately 4 hours was observed [17]. Therefore, as the dataset increases in complexity and size, it is feasible that this method could be used to determine architectures which have not

been previously tested whilst still having confidence in the feasibility of the derived device structure.

Conclusion

A large database of OPV performance, stability, and embodied energies has been collected for analysis using machine learning approaches. $\mathbf{E_{Net}}$ has been calculated for each device based on the initial efficiency, T_{80} and T_{S80} and $\mathbf{E_{Emb}}$. Machine learning techniques are then applied to analyse the $\mathbf{E_{Net}}$ by applying the SMOreg algorithm using the structural components of each device as the predictive attributes. This provided a model giving the weights, and therefore relative significance of the different attributes in the improvement or deterioration of the $\mathbf{E_{Net}}$. The SMOreg algorithm was additionally used to model the PCE and stability (T_{80} and T_{S80}). For these two parameters the CC was 0.78, 0.67 and 0.75 respectively. These CC values represent moderately high values and illustrate that the SMOreg algorithm is still capable of producing relatively accurate models for these parameters. The CC for these were additionally checked since they correspond to the parameters used when deriving $\mathbf{E_{Net}}$. In addition, a clustering algorithm, using a genetic search method, was employed in order to determine he optimum device configuration. This provides a way of rapidly scanning datasets of OPV material performance and stability data and identify which materials are too environmentally damaging to consider for application in future devices and modules. A relatively high absolute error is identified when applying these ML algorithms. However, this can be expected due to high level of variability in the stability data, which is dependent on lab-to-lab processing and testing approaches.

They key results from the analysis are:

1) Transport Layer 1 and 2, Active layer 1 and 2 and substrates have the greatest impact on increasing the $\mathbf{E_{Net}}$

2) Active layers 1 and 2, transport layer 1 and 2, electrode 1 and 2 could have the greatest influence on negatively impacting $\mathbf{E_{Net}}$

3) Commonly used materials (other than active layers) that show promise towards delivering positive $\mathbf{E_{Net}}$: PET + barrier layer, PET (substrates), NiOx, ZrOx, CrO2 (TL1), ZnO, LiF, MoO3 (TL2)

4) Commonly used materials (other than active layers) that show negative weighting and whose use should be minimised in OPV manufacturer: Quartz, Glass (substrate); V_2O_5, ZrO (TL2); Ag (evaporated), Au (E2)

5) Commonly used active layer materials that show promise towards delivering positive E_{Net}: DRCN7T, DR3TSBDT, ZnPc, PDPP4T-2F, PCDTBT (donors); IT-4F, C61 (Acceptors)

6) Commonly used active layer materials that show negative weighting and whose use should be minimised in OPV manufacture: MEH-PPV, CuPc, Pentacene, PSBTBT, PCPDTBT, PPE-PPV (donors)

7) Modules generally show better E_{Net} values than cells due to their greater stability, lower wastage of material during manufacture and lower processing energy

In summary, the analysis method used in this work provides a means of identifying optimum device configurations that allow manufacturers to optimise OPV performance, stability and embodied energy. As the technology matures and material sets grow in complexity, this method could be extended to rapidly identify the best device architectures for development without the need for exhaustive, time – intensive experimentation.

Acknowledgements

References

[1]   B. Azzopardi, C. J. M. Emmott, A. Urbina, F. C. Krebs, J. Mutale, and J. Nelson, "Economic assessment of solar electricity production from organic-based photovoltaic modules in a domestic environment," *Energy Environ. Sci.*, vol. 4, no. 10, pp. 3741–3753, Oct. 2011.

[2]   C. J. Mulligan, C. Bilen, X. Zhou, W. J. Belcher, and P. C. Dastoor, "Levelised cost of electricity for organic photovoltaics," *Sol. Energy Mater. Sol. Cells*, vol. 133, pp. 26–31, Feb. 2015.

[3]   F. Machui *et al.*, "Cost analysis of roll-to-roll fabricated ITO free single and tandem organic solar modules based on data from manufacture," *Energy Environ. Sci.*, vol. 7, no. 9, pp. 2792–2802, Aug. 2014.

[4]   M. Mrinalini, N. Islavath, S. Prasanthkumar, and L. Giribabu, "Stipulating Low Production Cost Solar Cells All Set to Retail…!," *Chem. Rec.*, vol. 19, no. 2–3, pp. 661–674, Feb. 2019.

[5]   J. H. Wong, M. Royapoor, and C. W. Chan, "Review of life cycle analyses and embodied energy requirements of single-crystalline and multi-crystalline silicon

photovoltaic systems," *Renewable and Sustainable Energy Reviews*, vol. 58. Elsevier Ltd, pp. 608–618, 01-May-2016.

[6]   Q. Liu *et al.*, "18% Efficiency organic solar cells," *Sci. Bull.*, vol. 65, no. 4, pp. 272–275, Feb. 2020.

[7]   M. D. Chatzisideris, P. K. Ohms, N. Espinosa, F. C. Krebs, and A. Laurent, "Economic and environmental performances of organic photovoltaics with battery storage for residential self-consumption," *Appl. Energy*, vol. 256, p. 113977, Dec. 2019.

[8]   C. L. Chochos *et al.*, "Current status, challenges and future outlook of high performance polymer semiconductors for organic photovoltaics modules," *Progress in Polymer Science*, vol. 91. Elsevier Ltd, pp. 51–79, 01-Apr-2019.

[9]   M. L. Parisi, S. Maranghi, L. Vesce, A. Sinicropi, A. Di Carlo, and R. Basosi, "Prospective life cycle assessment of third-generation photovoltaics at the pre-industrial scale: A long-term scenario approach," *Renew. Sustain. Energy Rev.*, vol. 121, p. 109703, Apr. 2020.

[10]   X. Du *et al.*, "Elucidating the Full Potential of OPV Materials Utilizing a High-Throughput Robot-Based Platform and Machine Learning," *Joule*, vol. 5, no. 2, pp. 495–506, Feb. 2021.

[11]   W. Sun *et al.*, "Machine learning–assisted molecular design and efficiency prediction for high-performance organic photovoltaic materials," *Sci. Adv.*, vol. 5, no. 11, p. eaay4275, Nov. 2019.

[12]   H. Sahu, W. Rao, A. Troisi, and H. Ma, "Toward Predicting Efficiency of Organic Solar Cells via Machine Learning and Improved Descriptors," *Adv. Energy Mater.*, vol. 8, no. 24, p. 1801032, Aug. 2018.

[13]   M.-H. Lee, "Insights from Machine Learning Techniques for Predicting the Efficiency of Fullerene Derivatives-Based Ternary Organic Solar Cells at Ternary Blend Design," *Adv. Energy Mater.*, vol. 9, no. 26, p. 1900891, Jul. 2019.

[14]   W. Sun *et al.*, "Artificial Intelligence Designer for Highly-Efficient Organic Photovoltaic Materials," *J. Phys. Chem. Lett.*, vol. 12, no. 36, pp. 8847–8854, Sep. 2021.

[15]   T. W. David, H. Anizelli, T. J. Jacobsson, C. Gray, W. Teahan, and J. Kettle, "Enhancing the stability of organic photovoltaics through machine learning," *Nano Energy*, vol. 78, Dec. 2020.

[16]   M. O. Reese *et al.*, "Consensus stability testing protocols for organic photovoltaic materials and devices," *Sol. Energy Mater. Sol. Cells*, vol. 95, no. 5, pp. 1253–1267, 2011.

[17]   N. Espinosa and F. C. Krebs, "Life cycle analysis of organic tandem solar cells: When are they warranted?," *Sol. Energy Mater. Sol. Cells*, vol. 120, no. PART B, pp. 692–700, Jan. 2014.

[18]   N. Espinosa *et al.*, "Life cycle assessment of ITO-free flexible polymer solar cells prepared by roll-to-roll coating and printing," in *Solar Energy Materials and Solar Cells*, 2012, vol. 97, pp. 3–13.

[19]   F. C. Krebs, N. Espinosa, M. Hösel, R. R. Søndergaard, and M. Jørgensen, "25th anniversary article: Rise to power - OPV-based solar parks," *Advanced Materials*, vol. 26, no. 1. John Wiley & Sons, Ltd, pp. 29–39, 08-Jan-2014.

[20]   P. Stiebitz, "Hyperspectral Polymer Solar Cells, Integrated Power for Microsystems," Golden, CO (United States), May 2014.

[21]   A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3. Kluwer Academic Publishers, pp. 199–222, Aug-2004.

[22]   U. Maulik and S. Bandyopadhyay, "Genetic algorithm-based clustering technique,"

Pattern Recognit., vol. 33, no. 9, pp. 1455–1465, Sep. 2000.

[23]   C. H. Peters, I. T. Sachs-Quintana, J. P. Kastrop, S. Beaupré, M. Leclerc, and M. D. McGehee, "High Efficiency Polymer Solar Cells with Long Operating Lifetimes," *Adv. Energy Mater.*, vol. 1, no. 4, pp. 491–494, Jul. 2011.

[24]   M. Z. Islam, V. Estivill-Castro, M. A. Rahman, and T. Bossomaier, "Combining K-MEANS and a genetic algorithm through a novel arrangement of genetic operators for high quality clustering," *Expert Syst. Appl.*, vol. 91, pp. 402–417, Jan. 2018.

[25]   S. Nagasawa, E. Al-Naamani, and A. Saeki, "Computer-Aided Screening of Conjugated Polymers for Organic Solar Cell: Classification by Random Forest," *J. Phys. Chem. Lett.*, vol. 9, no. 10, pp. 2639–2646, May 2018.

[26]   Y. Wu, J. Guo, R. Sun, and J. Min, "Machine learning for accelerating the discovery of high-performance donor/acceptor pairs in non-fullerene organic solar cells," *npj Comput. Mater.*, vol. 6, no. 1, pp. 1–8, Dec. 2020.

[27]   H. Sahu and H. Ma, 2019. Unraveling correlations between molecular properties and device parameters of organic solar cells using machine learning. *The journal of physical chemistry letters*, *10*(22), pp.7277-7284.

[28]   W. Li, R. Jacobs and D. Morgan, 2018. Predicting the thermodynamic stability of perovskite oxides using machine learning models. *Computational Materials Science*, *150*, pp.454-463.

[29]   Ç. Odabaşı and R. Yıldırım, 2019. Performance analysis of perovskite solar cells in 2013–2018 using machine-learning tools. *Nano Energy*, *56*, pp.770-791.
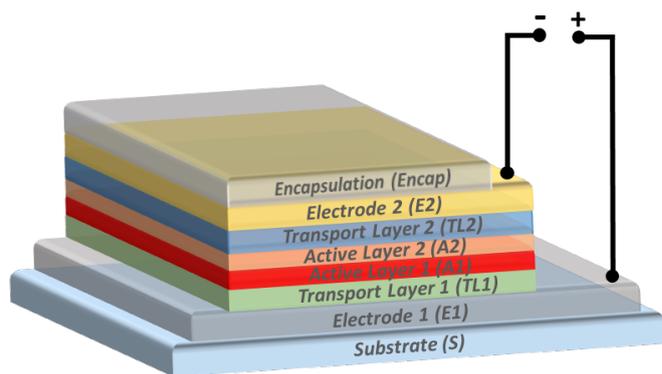
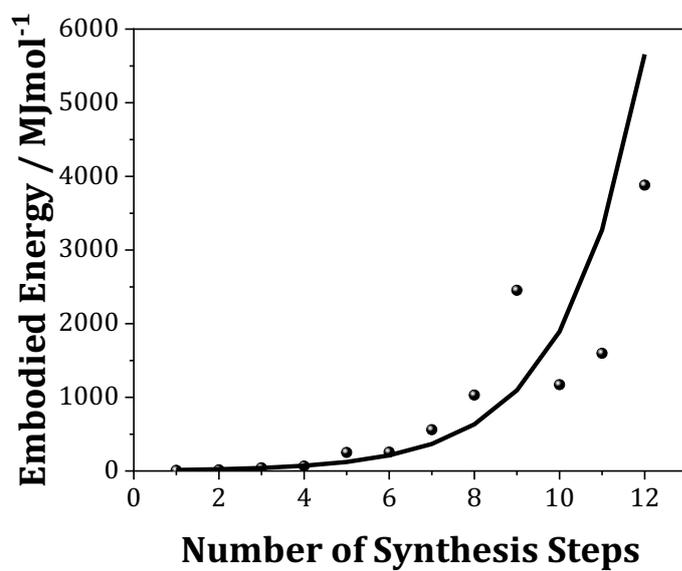Figure 1: Device architecture employed in dataset.

Figure 2. Variation in polymer E_{Emb} as a function of the number of synthesis steps based upon data from 12 recent papers of LCA of donor or acceptor polymers (2012-2021).
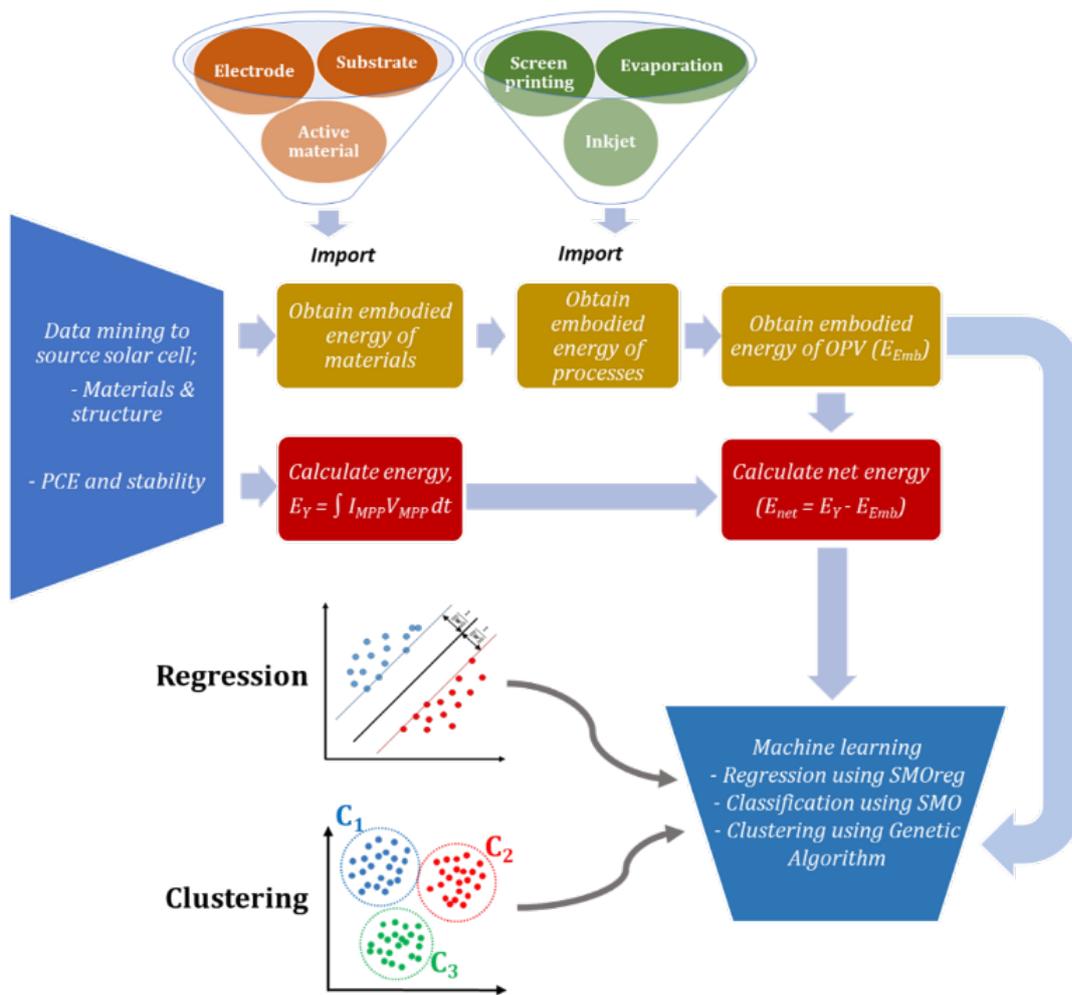
Figure 3: Schematic of how material and energy costs are acquired and stages of analysis using ML genetic algorithms.
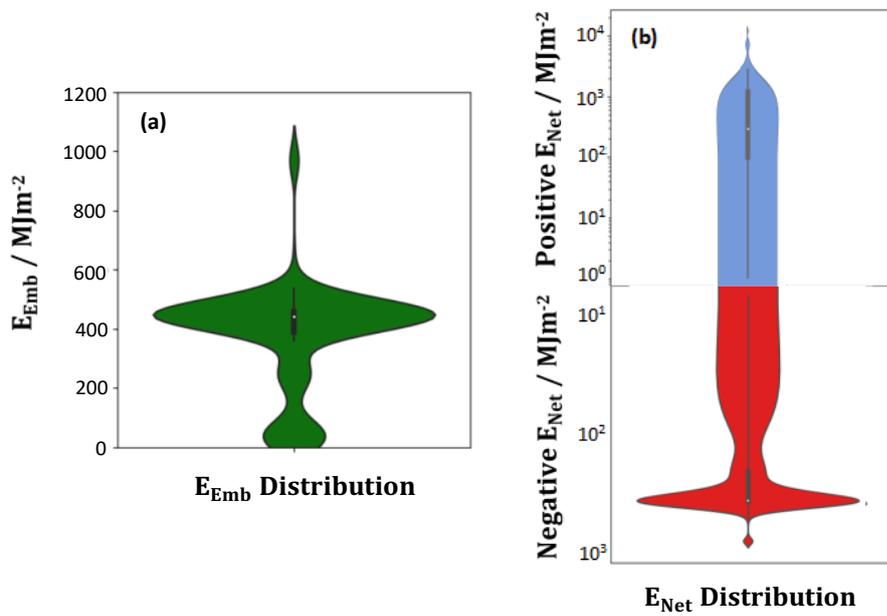
Figure 4: Categorical distribution for (a) $\mathbf{E_{Emb}}$ and (b) $\mathbf{E_{Net}}$ where 15% of studies considered display a positive value.
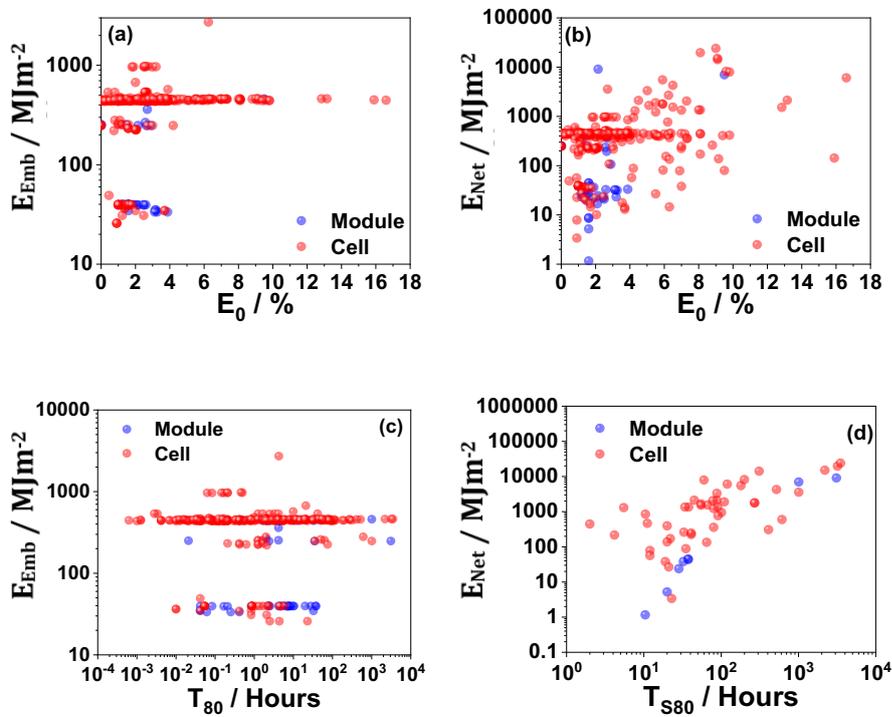
Figure 5: Distributions of (a) $\mathbf{E_{Emb}}$ as a function of E_0, (b) $\mathbf{E_{Net}}$ as a function of E_0, (c) $\mathbf{E_{Emb}}$ as a function of T_{80} and (d) $\mathbf{E_{Net}}$ as a function of T_{S80}.

Figure 6: (a) Predicted E_{Emb} as a function of reference E_{Emb} using SMOreg algorithm for all structural attributes. (b) Predicted E_{Net} as a function of reference E_{Net} using SMOreg algorithm

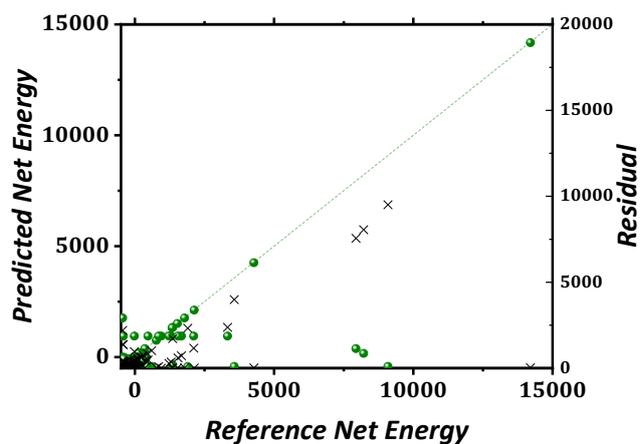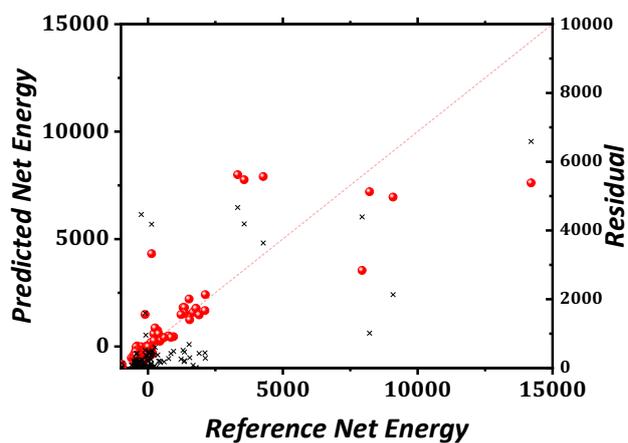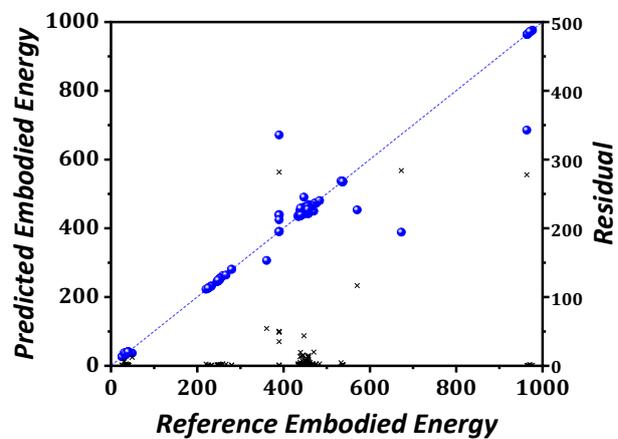for all structural attributes. (c) Predicted E_{Net} as a function of reference E_{Net} using SMOreg algorithm for only active material attributes. Regression results are shown as circles whilst residuals are shown are shown as crosses. Absolute values for the residuals were taken.

Metric	E_{Emb}	E_{Net}	E_{Net} (Active Only)
Correlation coefficient	0.988	0.838	0.719
Mean Absolute Error (MJm^{-2})	5.19	175	218
Root Mean Squared Error (MJm^{-2})	27.2	681	859
Relative Absolute Error	4.04%	36.6%	45.6%

Table 1: Fitting parameters for SMOreg applied to the dataset to predict E_{Emb}, E_{Net} and E_{Net} using only active layer materials as predictive attributes. A cross – validation approach is used for each case.

Best 15 Attributes		Worst 15 Attributes	
Name	Weight	Name	Weight
S(PET)	-0.2359	E2(CPP PEDOT:PSS)	0.0019
S(Barrier layer & PET)	-0.2343	A1(PBDTTT)	0.002
E1(Ag)	-0.1046	A2(ITIC)	0.0021
A2(IDTBRIDFBR)	-0.0674	TL2(Ca)	0.0029
TL2(ZnO NP)	-0.0559	A1(PCPDTBT)	0.004
TL1(PEDOT:PSS - High conductive)	-0.0164	A1(CuPc)	0.0041
TL1(CuI)	-0.0156	A2(C$_{60}$)	0.0067
TL1(ZnO)	-0.0151	TL2(ZrO)	0.0106
TL1(PEDOT:PSS)	-0.0148	TL2(MoO$_3$)	0.0109
TL1(Graphene Oxide)	-0.0139	TL1(TiO$_2$)	0.0231
TL1(AlQ3)	-0.0035	E1(MIM)	0.074
E2(Al)	-0.0023	E1(Cr – Aluminium)	0.1073
TL2(LiF)	-0.0022	TL1(CrO$_2$)	0.1073
TL2(BPhen)	-0.0013	TL1(NiOx)	0.1164
A2(PCBM)	-0.0005	E1(ITO)	0.1271

Table 2: Best 15 attributes and worst 15 attributes for E_{Emb} minimisation.

All structural attributes			
Best 15 Attributes		**Worst 15 Attributes**	
Name	Weight	Name	Weight
A1(DR3TSBDT)	0.0299	A1(PTB7)	-0.0093
TL2(ZnO NP)	0.026	A1(PDTSTTz-4)	-0.0096
A2(IDFBR)	0.0205	E2(Au)	-0.0107
A2(IDTBRIDFBR)	0.0182	A1(CN-P3HT)	-0.0117
S(PET + barrier layer)	0.0146	TL1(Ethanolamine)	-0.0132
A1(PCDTBT)	0.014	E1(Au)	-0.0147
TL1(NiOx)	0.0122	E1(Ag)	-0.0149
TL2(LiF)	0.0099	TL2(V_2O_5)	-0.0166
TL2(MoO_3)	0.008	A1(PBDTTTPD)	-0.0215
S(PET)	0.0074	A1(PDPP4T-2F)	-0.0226
TL1(ZrO_2)	0.0066	A2(IEICO-4FP2FBTT-H)	-0.0268
TL1(CrO_2)	0.0058	TL2(ZrO)	-0.0389
A1(ZnPc)	0.0046	S(Quartz)	-0.0448
A2(C_{60})	0.0034	Embodied Energy	-0.156
E2(Ag)	0.0018	A1(PBDTTT-C-T)	-0.2277

Table 3: Best 15 attributes and worst 15 attributes for E_{Net} maximisation.

Active Only			
Best 15 Attributes		**Worst 15 Attributes**	
Name	Weight	Name	Weight
A1(DRCN7T)	0.9232	A1(Pentacene)	-0.0362
A2(ZnPc)	0.1272	A1(CuPc)	-0.0384
A1(PDPP4T-2F)	0.0761	A1(TDCV-TPA)	-0.0382
A1(PCDTBT)	0.0547	A1(P3MHOCT)	-0.038
A1(IDTT)	0.0421	A1(MEH-PPV)	-0.0365
A2(IT-4F)	0.0232	A1(Pentacene)	-0.0362
A2(C61)	0.0101	A1(PININE:DTBT)	-0.0387
A1(PBDB-T)	0	A1(PSBTBT)	-0.0391
A1(DR3TSBDT)	0	A1(PCPDTBT)	-0.0391
A2(DF-DPB)	-0.0012	A1(CN-P3HT)	-0.0395
A1(PBDTTT-EFT)	-0.0019	A1(PPE-PPV)	-0.0402
A2(IDTBR)	-0.0092	A1(Tetracene)	-0.0403
A2(IDFBR)	-0.0114	A1(SubPc)	-0.0403
A1(DRCN5T)	-0.0142	A1($C6PcH_2$)	-0.0707
A1(IDTBRIDFB)	-0.02	A1(NDP2 doped DiNPB)	-0.173

Table 4: Best 15 attributes and worst 15 attributes for E_{Net} maximisation using only active layer materials.

Type	Encapsulation	Substrate	Electrode 1	Transport Layer 1	Active Layer 1	Active Layer 2	Transport Layer 2	Electrode 2	Embodied Energy	Net energy	Net energy (Class)	Net energy (Magnitude)
Cell	None	Quartz	ITO	MoO3	C6PcH$_2$	PCBM-61	None	Al	967.025	-966.589	A	
Cell	Rigid	Glass	Cr|Aluminum	CrO2	P3HT	PCBM-61	PEDOT:PSS	Au	536.61	-514.915	A	
Cell	None	Glass	ITO	MoO3	P3HT	PCBM-61	LiF	Al	459.76	-454.625	B	
Cell	Rigid	Glass	ITO	None	CuPc	C$_{60}$	Bathocuproine	Al	450.88	-449.426	C	
Cell	None	Glass	ITO	PEDOT:PSS	P3HT	PCBM-61	Ca	Ag	446.09	-445.799	D	
Cell	None	Glass	ITO	ZnO	PTB7	PCBM-71	MoO$_3$	Ag	456.355	-440.313	C	
Cell	Rigid	Glass	ITO	PEDOT:PSS	P3HT	PCBM-61	LiF	Al	442.68	-438.433	E	
Cell	None	Glass	ITO	PEDOT:PSS	P3HT	PCBM-61	None	Al	438.16	-437.52	G	
Cell	None	Glass	ITO	None	P3HT	PCBM	None	Al	434.74	-434.769	H	
Cell	Rigid	Glass	ITO	PEDOT:PSS	P3HT	PCBM-61	PEDOT:PSS	Al	441.87	-423.047	I	
Cell	Rigid	Glass	ITO	ZnO	P3HT	PCBM-61	PEDOT:PSS	Ag	440.58	-401.959	J	
Module	Flexible	PET	ITO	ZnO	P3HT	PCBM-61	HTLSolar	Ag	253.168	-250.22	L	
Cell	Rigid	Glass	Ag	PEDOT:PSS	P3HT	PCBM-61	LiF	Al	225.218	-221.719	M	
Cell	Rigid	PET	Ag	PEDOT:PSS	P3HT	PCBM	PEDOT:PSS	Ag	39.586	-37.44	N	
Module	Yes	PET	Ag	ZnO	PBDB-T	ITIC	HTLSolar	CPP:PEDOT:PSS	33.44	-32.343	N	
Cell	None	PET	Ag	PEDOT:PSS	P3HT	ICBA	AZO	Al	30.868	-27.52	O	
Module	Flexible	PET	Ag	PEDOT:PSS	P3HT	PCBM	PEDOT:PSS	Ag	39.586	-21.151	O	
Cell	Rigid	Glass	ITO	ZnO	P3HT	PCBM-61	MoO$_3$	Ag	455.89	117.3728	Q	
Cell	Rigid	Glass	ITO	PEDOT:PSS	PCDTBT	PCBM-71	Ca	Ag	444.91	897.9285	R	

Increasing Net energy (arrow down in final column)

Table 5: Clustered attributes based on E_{Net} output. Clusters ranked from low (A) to high (R). Each row represents a cluster and an optimised device architecture for each E_{Net} class.