

Revolutionizing Low-Cost Solar Cells with Machine Learning: A Systematic Review of Optimization Techniques

Satyam Bhatti, Habib Ullah Manzoor, Bruno Michel, Ruy Sebastian Bonilla, Richard Abrams, Ahmed Zoha, Sajjad Hussain, and Rami Ghannam*

Machine learning (ML) and artificial intelligence (AI) methods are emerging as promising technologies for enhancing the performance of low-cost photovoltaic (PV) cells in miniaturized electronic devices. Indeed, ML is set to significantly contribute to the development of more efficient and cost-effective solar cells. This systematic review offers an extensive analysis of recent ML techniques in designing novel solar cell materials and structures, highlighting their potential to transform the low-cost solar cell research and development landscape. The review encompasses a variety of ML approaches, such as Gaussian process regression (GPR), Bayesian optimization (BO), and deep neural networks (DNNs), which have proven effective in boosting the efficiency, stability, and affordability of solar cells. The findings of this review indicate that GPR combined with BO is the most promising method for developing low-cost solar cells. These techniques can significantly speed up the discovery of new PV materials and structures while enhancing the efficiency and stability of low-cost solar cells. The review concludes with insights on the challenges, prospects, and future directions of ML in low-cost solar cell research and development.

batteries include metal poisoning for patients due to battery degradation, thus leading to malfunction in generating signals and the damage of electronic circuits.^[7]


Due to their high energy density, scavenging solar energy using photovoltaic (PV) cells has emerged as a potential and feasible solution to power miniature portable devices.^[8,9] In general, the architecture of these solar cells can be designed as regular, inverted, mesoporous, or planar structures. Furthermore, solar cells combine various materials to enable efficient photon absorption, electron transport, and electron extraction to an external circuit. This means there are vast opportunities for discovering solar cell materials and architectures.^[10] In fact, solar cell fabrication techniques involve optimizing different coating materials, thermal annealing conditions, encapsulation methods, etc., which often take place in the research laboratory.^[11] In spite

1. Introduction

Current miniature portable and implantable devices rely on batteries that need replacement and are hazardous to patients.^[1–3] Surgical removal is required when replacing batteries in implantable devices, which may be inconvenient for patients.^[4,5] Moreover, implantable biomedical devices are often powered using wires, which may cause discomfort, skin infections, and other hazards to patients.^[6] The key issues with implanting

of their benefits, solar energy harvesters still have a number of limitations, such as poor efficiency, rigidity, and stability.^[12] Despite these issues, there are a number of promising PV technologies that are working to overcome issues with high cost, efficiency, and durability, such as perovskite solar cells (PSC), organic solar cells (OSC),^[13] and dye-sensitized solar cells (DSSCs)^[14,15] The stability and efficiency of these low-cost, thin-film solar cells is still mainly poor due the effects of moisture and temperature.^[16] However, rapid progress in machine

S. Bhatti, H. U. Manzoor, A. Zoha, S. Hussain, R. Ghannam^[†]
James Watt School of Engineering
University of Glasgow
Glasgow G12 8QQ, UK
E-mail: rami.ghannam@glasgow.ac.uk

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/aesr.202300004>.

^[†]Present address: Engineering Interactive Devices Research Group, James Watt School of Engineering, University of Glasgow, Glasgow G12 8QQ, UK

© 2023 The Authors. Advanced Energy and Sustainability Research published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/aesr.202300004

B. Michel
Smart System Integration
IBM Research, IBM
Zurich Laboratory
Zurich, Switzerland

R. S. Bonilla
Department of Materials
University of Oxford
Oxford OX1 3PH, UK

R. Abrams
Onshore Renewables
AqualisBraemar LOC Group
London EC3R 6EN, UK

learning (ML) and artificial intelligence (AI) technologies suggests that they can be used to improve the performance and accelerate the discovery of these low-cost solar cells.^[17]

Moreover, the term “low-cost” solar cells generally refers to thin-film solar cells since they are less expensive to produce than conventional crystalline silicon solar cells. The production of low-cost solar cells involves depositing a thin coating of semiconductor material (organic, inorganic, or a combination of both) onto a glass or plastic substrate. Low-cost solar cells are cheaper than crystalline silicon solar cells because they use less material and do not need expensive machinery and processing techniques to make them. Therefore, low-cost solar cells are becoming increasingly popular for both large-scale solar power plants and for miniature portable electronic devices due to their affordability.^[18]

Innovation in developing new low-cost solar cells is needed, which can be achieved with the help of experimentally validated finite-element modeling using software tools such as Sentaurus TCAD.^[19] However, this is a time-consuming effort, and leveraging the power of AI can be a game changer in discovering new materials and fabrication techniques to help expedite the process of selection, design, and optimization.^[20] In fact, the literature suggests that low-cost thin-film solar cell performance can be optimized using a variety of efficient computational and statistical methods.^[21] From the systems perspective, ML algorithms can also help develop reconfigurable PV cells based on switchable CMOS addressable switches.^[4]

Additionally, conjugation is a key characteristic of organic materials, which are frequently used in such devices, and it plays a crucial part in low-cost solar cells. Conjugated polymers or tiny molecules with alternate single and double bonds frequently make up the organic components in solar cells. For the effective conversion of solar energy, conjugation enables the organic materials to absorb light in the visible region of the spectrum. An exciton, which is an excited state produced when a conjugated substance absorbs light, can be split into electrons and holes to produce an electrical current. The performance of low-cost solar technology depends highly on conjugated materials' capacity to transport these electrons and holes through the device effectively.^[22]

Furthermore, the distribution of electron density in the energy levels of materials used in solar cell architecture, known as the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) pattern, is a crucial factor that affects the solar cell's efficiency in capturing photons and producing electrical energy. Matching the HOMO and LUMO levels of different materials used in the cell is a significant challenge in solar cell design to optimize charge separation efficiency and minimize recombination, which result in energy loss and decreased efficiency.^[23,24]

To investigate the characteristics of charge carriers (electrons and holes) in solar cell materials, researchers use a method called transient decay measurement (TDM) analysis. This analysis involves monitoring the decay rate of photogenerated carriers over time following a transient pulse of light. When sunlight is absorbed by the material, it creates electron-hole pairs that produce a photocurrent in the solar cell. The TDM analysis tracks the time it takes for the photocurrent, which is related to the recombination of electron-hole pairs, to decay. During the recombination process, charge carriers combine and cancel each

other out, causing energy loss and reducing the efficiency of the solar cell.^[25]

Furthermore, in the literature, ML relates to the development and ability of the model to learn to adapt, forecast, and predict the independent variables.^[26] ML algorithms consist of three types, namely, supervised learning, unsupervised learning, and reinforcement learning.^[27] The supervised ML approach takes the input data from the user to learn from past experiences and, accordingly, trains the model.^[28] However, the unsupervised ML train model depends upon the real-time data generated and outputs depending on the information given by the user. In contrast, reinforcement learning is the subset of ML that enables an AI-driven system (also known as an agent) to learn by performing tasks and receiving feedback from its trials and errors.^[29] Herein, we discuss the various ML techniques in depth that are applied to find an optimized structure for solar cells.^[30]

Examples of ML techniques reported in the literature include linear regression, logistic regression, k-nearest neighbors (KNN), random forest (RF), etc.^[31,32] However, every problem requires a unique ML algorithm.^[33] Every algorithm has unique abilities and data requirements. For instance, due to nonlinear relations in solar cells, linear regression would not be very helpful. For logistic regression, we have to assume that factors are independent of each other, which might not be the case in solar cells. Similarly, the purpose of KNN is to locate the nearest neighbors with the best possible value. However, it is more suitable for continuous variables. So, the use of ML in optimizing solar cells depends upon the type of experiment, optimizing variables, and data type.

Since the fabrication of OSCs is cheap, most experimental work is carried out via trial and error, which does not guarantee the best performance.^[34] Instead, researchers are now turning their attention to data-driven techniques for material design and discovery.^[35] ML is one of the vital data-driven techniques that is increasing to prominence in discovering new solar cells, forecasting electrical characteristics, and performance prediction without any experimentation.^[36,37] ML uses algorithms to visualize and analyze data that has several advantages over traditional programming techniques.^[38] This article reviews the different ML algorithms used to find an optimized structure of a low-cost solar cell. The output power can be optimized for different light conditions and shading depending on the positioning of the solar cells.^[39] In our article, we discuss the integration of ML methods for designing low-cost solar cells and, consecutively, explore the literature on using different ML techniques for the advanced discovery of solar cells.

1.1. Contributions to the Literature

In our systematic review, we analyzed the role of ML in the field of solar cell design and material discovery. We conducted a systematic review of the applications of ML in the optimization, fabrication, and discovery of new photovoltaic materials. Our article is the first effort to provide a systematic review in this domain. The following are the major contributions of this article. 1) We conduct a review of 58 papers from a total of 18 380 research articles involving solar cell discovery, optimization, and fabrication using ML techniques. 2) We shortlist all ML models that can help in the discovery of new materials. 3) We review the literature

on low-cost high-performance solar cells using ML techniques. 4) Various ML techniques facilitating the discovery of solar cells were considered in the study. 5) We investigate the techniques used for the optimization of solar cells with the help of ML. 6) We highlight the challenges associated with using ML techniques for solar cell design.

1.2. State of the Art

During the past 5 years, there has been a surge in the use of ML and AI techniques for designing new solar cells.^[40,41] In this subsection, we review previously published systematic review papers on this field using ML techniques, and we discuss their limitations as well as the contributions that this review provides to the literature.

Qiuling et al.^[42] reviewed the ML techniques for only perovskite materials design and discovery. However, their review lacks a comprehensive comparison of ML techniques for other low-cost solar cells, such as organic, inorganic, hybrid, and DSSCs. Additionally, Hannes et al.^[43] discussed the challenges of ambient hybrid solar cells for IoT devices, while the article presented by Hannes et al.^[44] reveals the study on solar cell cracks using statistical parameters of electroluminescent images using ML. However, both studies presented limited ML algorithms to explore solar cell electrical characteristics.

Furthermore, Yongjie et al.^[45] reviewed recent advances in computational chemistry for OSC discovery and mentioned the density functional theory (DFT), time-dependent DFT, all-atomic molecular dynamics, and coarse-grained molecular dynamics. Although their review covered OSCs, it lacked the ML techniques to expedite the process. Next, Florian et al.^[46] reviewed the literature on designing light-harvesting devices using ML, but the review was limited to only OSCs. Likewise, a review paper presented by Sheng et al.^[47] covered only ML optimization of PCs. The studies presented by Anton et al.,^[48] Min-Hsuan et al.,^[49] and Cagla et al.^[20] explored ML approaches to discover solar cell performance analysis. However, a major drawback in these studies was that limited ML approaches were discussed and did not involve the scope for optimization as well as the fabrication of solar cells in the real environment.

Therefore, based on the above, state-of-the-art review articles on ML for solar cell discovery focused mainly on a single ML technique with a set of input data. In this work, we instead aim to systematically review the range of ML techniques for developing solar cells. These ML techniques include the procedure to preprocess the input data, various ML algorithms, optimization, and fabrication of the solar cell in a real environment. In this context, our systematic review goes beyond existing literature as it showcases how various ML techniques can be used to screen large numbers of materials for potential solar cell applications and to optimize the design of low-cost solar cells.

1.3. Organization of the Article

The rest of the article is organized as follows. The adopted methodology in reviewing the literature is discussed in Section 2, and the overall results of our systematic review in response to our research questions are presented in Section 3. In Section 4,

we discuss areas of further study, future outlook, recommendations, and open research issues. Finally, summarizing remarks are included in the conclusions section.

2. Review Methodology

In this section, we discuss our research objectives and our methodology in collecting and synthesizing the literature on ML algorithms for designing and fabricating low-cost, high-performance solar cells.

2.1. Research Objectives

The four key objectives of our systematic review article are as follows. 1) To review the range of ML techniques for designing low-cost solar cells using historical data. 2) To identify the ML techniques used specifically for the discovery of new PV materials. 3) From a device perspective, identify the specific ML and optimization techniques used for designing efficient solar cell architectures. 4) To identify ML algorithms specifically used for the fabrication of low-cost PV cells from the circuits and systems perspective.

Figure 1 maps our four research objectives and the process involved in shortlisting the research articles. Initially, we focused on extracting and preprocessing the historical data, followed by the discovery of new materials and optimization of solar cells. Finally, we reviewed the research articles that discussed the integration of ML for fabricating solar cells. Accordingly, in our systematic review, we defined these research objectives to target a set of questions that are the need for the study. Additionally, we shortlisted a set of research articles using the search engines available on Google for extracting the recent research articles published in this domain. This search was subsequently validated using the IBM Watson Studio tool.

2.2. Research Questions

Our systematic review aims to answer the following four research questions.

RQ1: What are the data-driven approaches for designing low-cost high-performance solar cells?

RQ2: How can ML algorithms facilitate the discovery of new low-cost solar cell materials?

RQ3: What are the optimization techniques used for designing an efficient low-cost solar cell architecture?

RQ4: What ML algorithms are used for fabricating low-cost solar cells from a circuits and systems perspective?

2.3. Review Protocol

For structuring our systematic review, we instigated a review protocol, and the following are the prerequisites of the adopted analogy. In this section, we discuss the search strategy, inclusion criteria, exclusion criteria, and screening mechanisms for selecting relevant research papers.

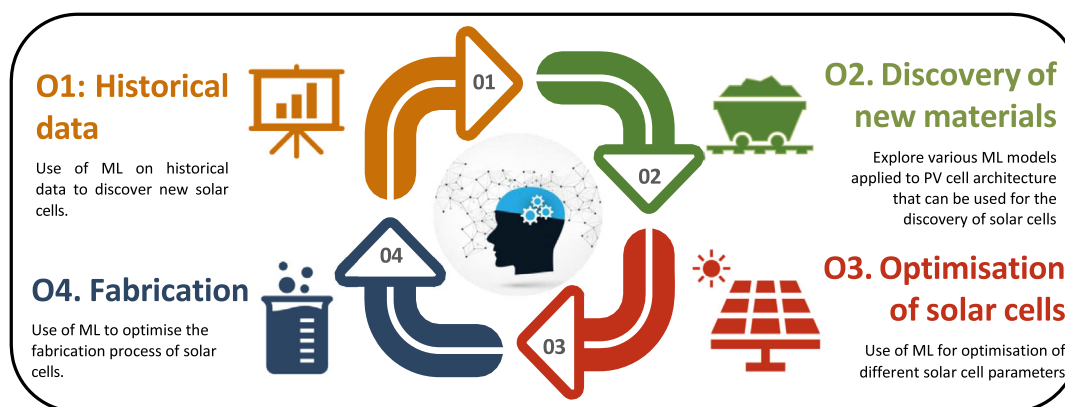


Figure 1. The objectives of our research are fourfold. The first objective, O1, involves identifying all the literature on low- cost solar cell designs using ML techniques. Our second objective, O2, involves reviewing the literature on materials discovery, whereas O3 identifies specific ML techniques used for optimizing solar cell architectures. Finally, O4 involves classifying the range of ML algorithms for designing low-cost PV cells from a circuits and systems perspective.

2.3.1. Search Strategy

Our review considered the latest research articles from major publishing houses that include IET, Science Direct, Nature, AIP, Wiley, IEEE explorer, IoP science, ACS publications, and MDPI. Our search also included non-prereviewed articles from arXiv. Thus, we performed the critical appraisal using the AACODS (Authority, Accuracy, Coverage, Objectivity, Date, Significance) checklist as an evaluation and critical appraisal tool of grey literature (publications and research created by groups not affiliated with conventional academic or commercial publishing institutions).

We begin with queering all the repositories with different research items. We defined the keywords such as “Machine Learning,” “Data-driven approach,” “PV cell architecture,” “Solar cells,” “Low-cost,” “Optimization,” and “fabrication” shown in **Table 1** for collecting our research articles. In **Figure 2**, we demonstrate the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) model showing a screening of the short-listed publications depending on our research questions. Articles were scanned based on their title and abstract as well as a full-text read of the publications. In addition, we developed search strings using Boolean operators (AND, OR) to connect these keywords.

2.3.2. Inclusion Criteria

The following are the parameters used in the inclusion criteria. 1) We included only English-language articles involving the

data-driven approaches of designing solar cells using ML techniques and were pertinent to the study issues such as poor data quantity and data quality. 2) We included the pertinent articles facilitating the discovery of only low-cost solar cells using ML methods before determining their eligibility. 3) We included comparative studies involving the optimization and robustness of solar cells designed from ML services. 4) We targeted only articles that discussed ML for solar cells, solar cell optimization, and publications on ML integration on solar cells.

2.3.3. Exclusion Criteria

The following is a list of the exclusion criteria for shortlisting the research papers based on our research objectives and targeted research questions. 1) Research articles published in languages other than English. 2) Research papers that are not available in full text. 3) Editorials, survey reviews, abstracts, and brief papers involving secondary studies are excluded. 4) Articles that did not address the integration of ML approaches with solar cells and the ones that involved the expensive manufacturing of solar cells. 5) The research articles published before 2018 were also excluded due to the unavailability of quality input data that resulted in poor implementation of ML techniques.

2.3.4. Screening Phase

Articles were further screened in two phases. In the first phase, we examined the title and the abstract of each research article to

Table 1. Keywords and their definitions used for our search from January’2018 to August’2022.

“Machine learning”	The development of computer systems to adapt and learn without being given explicit instructions by analyzing data patterns.	[142]
“Data-driven approach”	The computer can personalize the information by using data to guide its activities.	[143]
“PV cell architecture”	The different combination layers, doping, meshing, contacts, thickness, etc.	[144]
“Solar cells”	Electronic devices capable of converting solar radiation directly into electricity.	[145]
“Low-cost”	All solar cells whose manufacturing process is less expensive than traditional crys-talline silicon solar cells.	[18]
“Optimization”	To utilize a situation or resource in the greatest or most efficient way possible.	[146]
“Fabrication”	The process of creating something (solar cells) through invention or production.	[147]

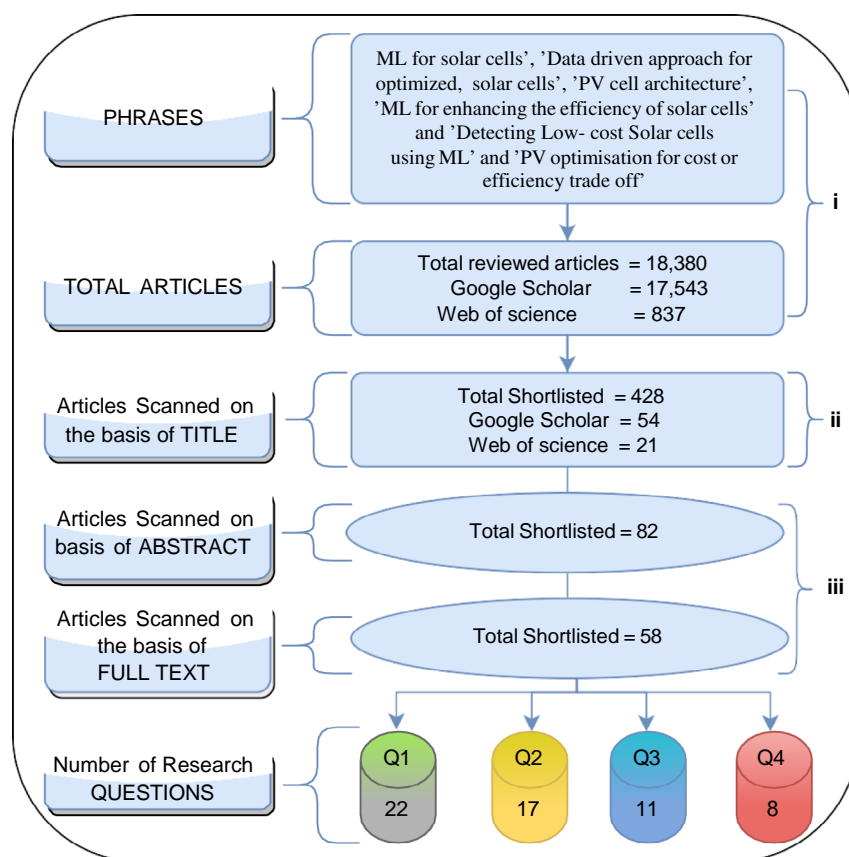


Figure 2. The PRISMA model shows the process of shortlisting the research articles, including the screening phase based on our assigned research questions from January 2018 to August 2022. The screening of the research articles was done on search engines such as Google Scholar and Web of Science. The respective combination of keywords and phrases was added to the advanced search and subsequently, the articles were shortlisted from manual screening. Further, the total research articles were manually screened based on reading the title, abstract, and full text of the research papers. Therefore, the four questions, Q1, Q2, Q3, and Q4, resulted in a total of 22, 17, 11, and 8 research articles, respectively.

check whether they satisfied our inclusion criteria. In the second phase, we further shortlisted our articles based on their full text. It is worth mentioning that the same piece of writing frequently appeared in various publications. For example, conference papers frequently appear in journals. We take into account the original writing each item was reviewed throughout the screening stage two. At least two of the contributors of this paper who were entrusted with classifying the items as either pertinent or not pertinent might require more research, as finalized until any such item is either published or the authors have a discussion tagged as relevant or not. Survey and review papers were excluded from our review. Finally, each article was carefully classified and evaluated thematically.

2.4. Review Results

In this section, we discussed the results that we obtained from shortlisting the research articles. The publication trends such as the number of articles published over a period of 5 years, the number of articles per research question, and publishing houses are discussed in detail in this section. In addition, we presented a

new state-of-the-art of approach to validate the research articles using the IBM Watson Studio.

2.4.1. Publication Trends

Based on the information presented in the title and abstract, we screened 82 manuscripts that satisfied our search criteria. Following a second screening phase, only 58 papers were relevant to our inclusion criteria.

In terms of publication trends, it appears that the majority of research articles (67%) were focused on addressing research questions RQ1 and RQ2, as demonstrated from **Figure 3a**. Moreover, only 2 articles were published in IEEE Xplore conference proceedings, as shown in **Figure 3b**. Consequently, **Figure 3c** represents the bar chart of the distribution of selected publications according to their types for each year. Based on our analysis, we can fairly comment that the maximum number of papers are published in Science Direct in 2019, followed by Wiley in 2019; however, the least number of articles are published in IET, IOP Sciences, IEEE Xplore Conferences, AIP, and Springer. Furthermore, most articles were published with Science Direct, Wiley, and ACS, as demonstrated in **Figure 3c**.

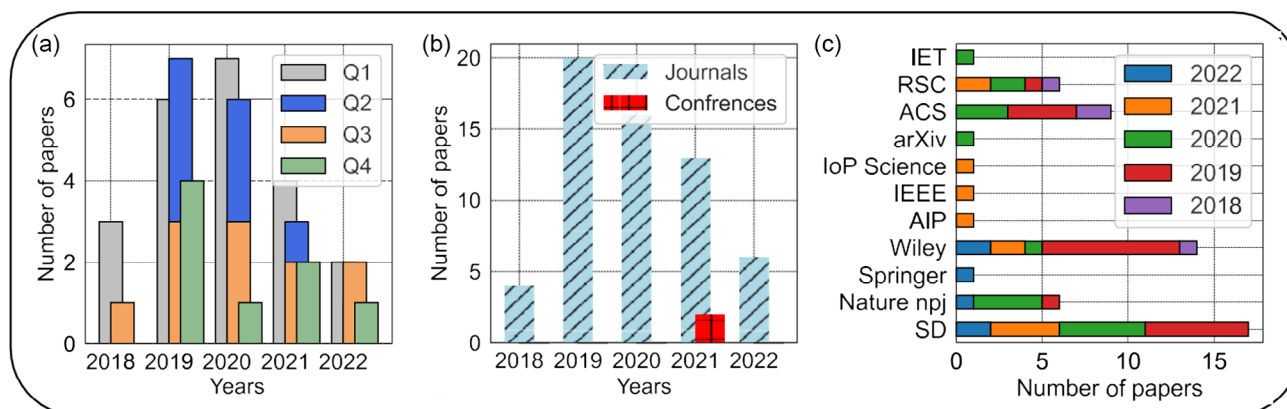


Figure 3. The figure demonstrates the publication trends for the defined research questions. a) Number of papers shortlisted as per the research questions from 2018 to 2022. b) Numerical count of research articles published in the conference or journal consecutively from 2018 to 2022 according to our shortlisted questions. c) Periodic distribution of achieved articles, research articles, and peer-reviewed publications shortlisted depending upon our research questions and research objectives according to the different publishers from 2018 to 2022.

2.4.2. Validation of Papers

Further, in order to validate the shortlisted research papers, we used IBM's Watson Studio tool which involves the process of experimentation to deployment, as well as data exploration, model development, and training. IBM Watson Studio is a data science IDE tool designed to help data scientists develop ML models. Moreover, using Watson Studio's "smart suggestions," we simplified the shortlisted papers from predictions and push models with the Watson ML platform across any cloud.

To perform this validation, details of each shortlisted paper were first tabulated in a spreadsheet. Parameters such as the year of the publication, the first author, the publisher, type of manuscript (journal or conference) were fed as input data to Watson Studio's Auto AI tool. Further, to validate the papers, we run a new project under the AutoAI experiment which allows the user to build a fully automated ML model to predict or forecast the parameter under consideration. However, we need to associate different ML or Natural language Programming services and compute the configuration of 8 vCPU and 32 GB RAM. Once the configuration to run an ML model was set, then we uploaded our data file (IBM3.csv) to our Watson Studio project. Uploading the dataset also gives us an opportunity to visualize the dataset in the form of charts and the Watson Studio tool automatically arranges the dataset to avoid any null values in the data. Therefore, once the input is provided to the Watson Studio model, it processes for the ML algorithm automatically. Thus, we set the predicted parameter, that is, the output result under consideration to be the year of the publication.

In the AutoAI experiment, the tool automatically uses various ML techniques after the analysis of the data. Here, for our model, AutoAI applied Multiple Classification prediction types and the model was optimized for root mean square error (RMSE) and run time. After the experiment is run on the Watson Studio tool, the dataset is read, split holdout (10%), read training data (90%), preprocessing, and model selection are performed. Consequently, the relationship map presented in Figure 4 describes the best feature transformers, pipelines used, and

the top algorithms. Moreover, the relationship map gives an indication of the best algorithm that is used by the AutoAI model and highlights the path in terms of pipeline. Herein, pipeline 8 achieved top position due to its least computational time and reduced path for feature transformers.

Accordingly, the progress map in Figure 5 shows the selected algorithm, hyperparameter optimization, feature engineering, and the most optimized feature transformers. Additionally, the most optimized ML model used was Snap Logistic Regression, having pipeline 8 showing the accuracy of the shortlisted papers, and finally, presenting the feature transformers such as the principal component analysis, univariate feature selection, and the product. There is a slight discrepancy in the accuracy of the model due to the fact that the research articles highlighted in different search engines, such as Google Scholar, Web of Sciences, IEEE Xplore, etc., display research articles that are out of the scope of our defined research questions in the methodology section. Also, most of the research articles are repeated at different search engines and whilst doing our manual search of the research paper, we subtracted those articles.

3. Results and Analysis

This section of the systematic review discusses our shortlisted research articles and how they are aligned with our research objectives and questions. Figure 6 shows the workflow of the planning (data extraction and data preprocessing), training (applying various ML techniques and comparing the model's accuracy), testing (optimization), and execution (fabricating solar cells in the laboratory) for discovering new solar cell architectures. As previously mentioned, our review focuses on low-cost solar cells such as PSCs, OSCs, and hybrids. Moreover, in Figure 6, the block data synthesis discusses the data extraction in a statistical form, Pearson's correlation coefficient matrix, solar cell architecture with layer combinations, and data preprocessing for classification problems. Whereas the second block of ML algorithms were supervised learning as well as unsupervised learning for classification, support vector regression (SVR), KNN,

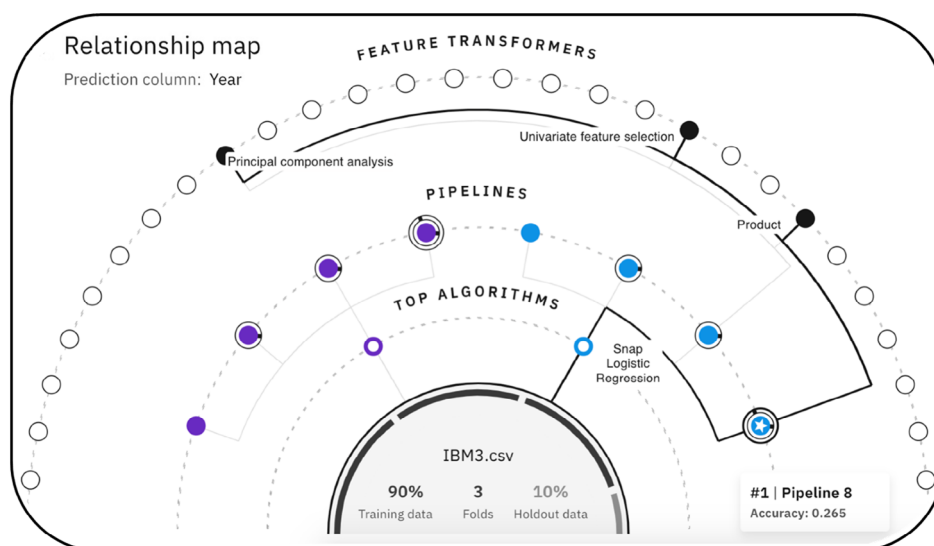


Figure 4. This figure illustrates a relationship map that predicts the number of research articles published annually using the IBM Watson Studio tool. The figure also provides valuable insights into the feature transformers, pipelines, and the top ML algorithms used to validate the shortlisted research papers. To generate this map, we manually shortlisted research articles from Google Scholar and Web of Science and provided them as input data to the IBM Watson tool. The AutoAI experiment tool then provided information on the research articles published based on our defined research questions for the study at hand. Overall, this figure highlights the efficacy of using ML algorithms and tools like IBM Watson Studio for analyzing and predicting research trends in various fields.

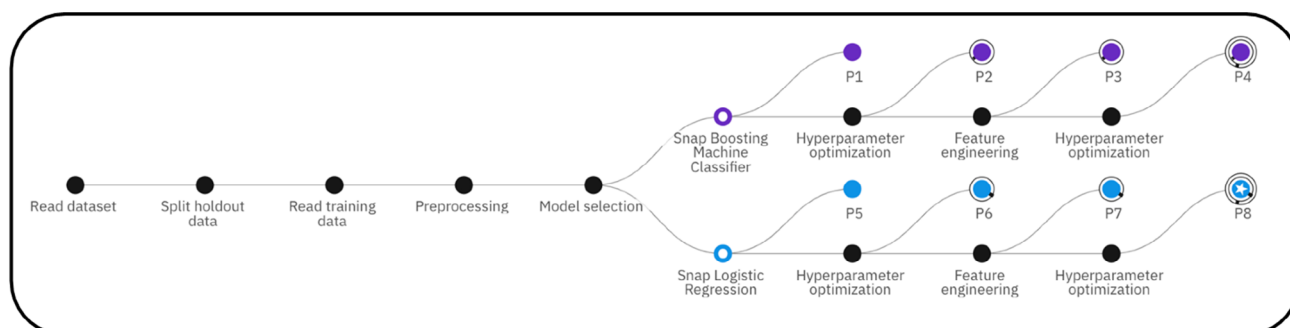


Figure 5. The pipeline representation of the ML algorithms used to validate the shortlisted papers using the IBM Watson studio tool. Accordingly, the algorithm uses ML techniques such as Snap Logistic Regression, Hyperparameter optimization, feature engineering, and another hyperparameter optimization to determine the most optimized algorithm for predicting the shortlisted research papers.

LR, support vector machine (SVM), and ANN. Furthermore, the third block discusses the optimization techniques, for instance, bandgap versus PCE curve, ternary contour plots, predicted versus calculated PCE, predicted versus ground truth curve, predicted accuracy of the ML model, and the total energy dissipation versus time curve. Finally, the fourth block discusses the fabricated solar cells.

3.1. Data-Driven Approaches for Designing Low-Cost Solar Cells

Solar cells are typically designed with specific objectives, such as reliability, affordability, efficiency, and stability. To predict the structure of low-cost solar cells, research is ongoing to gather and analyze data from previous solar cell fabrication experiments

in real-world environments. The quantity and quality of the extracted dataset are crucial to the effectiveness of ML algorithms. Based on the literature, larger input datasets generally result in higher accuracy and lower functional error values. Consequently, this section focuses on addressing RQ1.

3.1.1. Perovskite Solar Cells

Jino et al.^[50] explored the application of the gradient boost regression trees (GBRT) ML technique^[51] in the development of lead-free perovskites. They compiled a dataset comprising electronic structures of potential halide double perovskites and employed GBRT to estimate the bandgaps of these materials. Their results demonstrated that GBRT could accurately predict the materials' bandgaps, enabling the identification of promising candidates for subsequent

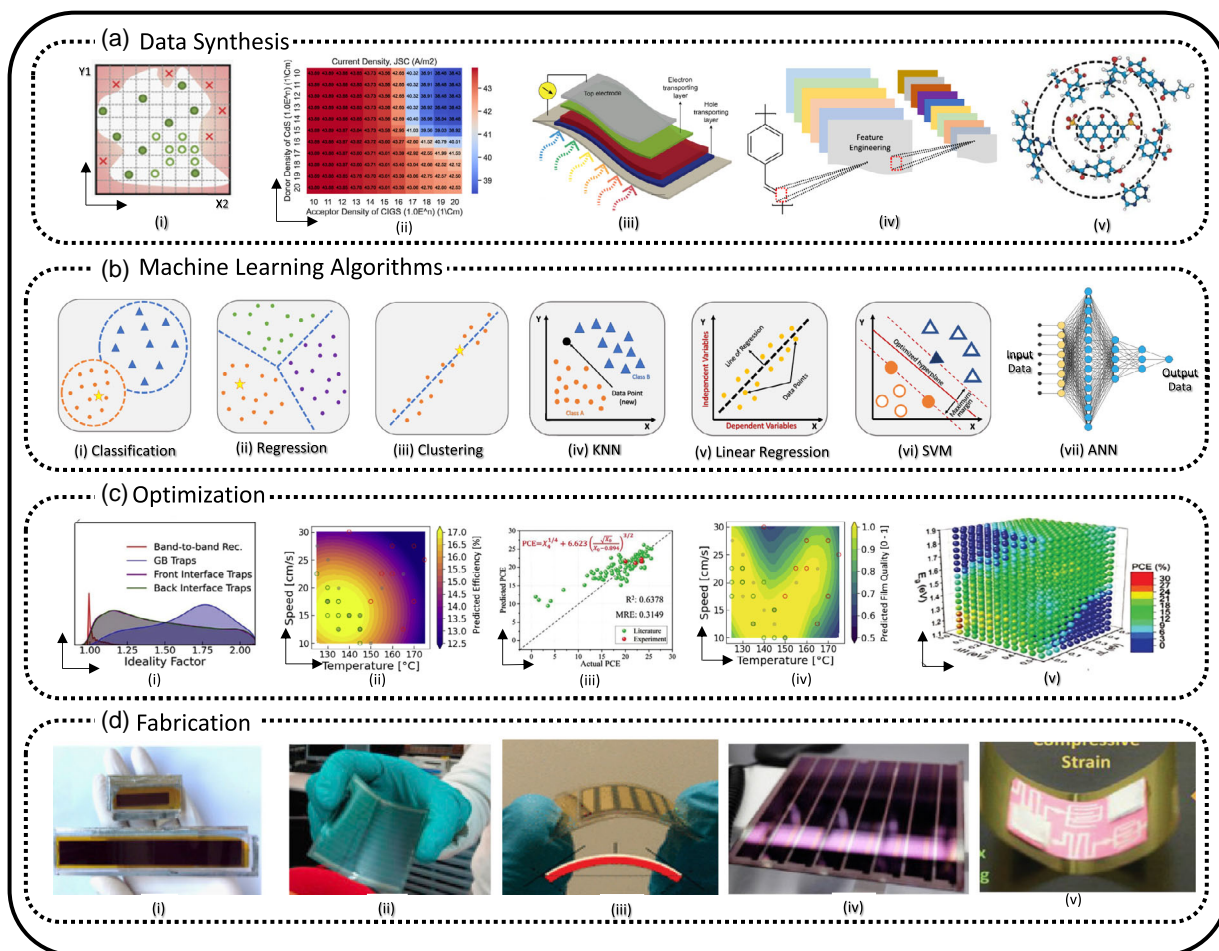


Figure 6. The figure demonstrates the general workflow of the process of discovering low-cost solar cells using ML algorithms. The block diagram is divided into four groups, (a) data synthesis, (b) ML algorithms, (c) optimization, and (d) fabrication. The first group (a) represents data synthesis, (i) Discusses the data extraction in a statistical form, Reproduced with permission.^[57] Copyright 2022, Elsevier. (ii) Pearson's correlation coefficient matrix, (iii) Solar cell architecture with layer combinations,^[56] (iv) data-preprocessing for classification problems^[57] and (v) Gradient-based extraction of data.^[58] The second group (b) represents ML algorithms, including (i) Classification, (ii) Regression, (iii) Clustering,^[59] (iv) K-nearest neighbors (KNN), (v) Linear regression, (vi) Support Vector Machine (SVM) and (vii) Artificial Neural Network (ANN). Reproduced with permission.^[60] Copyright 2022, Nature. The third group (c) discusses the optimization techniques, (i) Ideality Factor, Reproduced with permission.^[61] Copyright 2023, Wiley-VCH GmbH. (ii) Ternary contour plots, Reproduced with permission.^[62] Copyright 2022, Elsevier. (iii) Predicted versus Calculated PCE. Reproduced with permission.^[82] Copyright 2023, Wiley-VCH GmbH. (iv) Speed versus Temperature curve, Reproduced with permission, 2022, Joule, Elsevier. (v) Predicted accuracy of ML model. Reproduced with permission^[63]. Copyright 2023, Wiley-VCH GmbH. The fourth block discusses the fabricated solar cells. (vi) Reproduced with permission.^[64] Copyright 2019, Elsevier and Reproduced with permission.^[65–66] Copyright 2022, Wiley-VCH GmbH.

investigation. Initially, they generated the dataset using two space groups of the crystal structure with 540 hypothetical chemical compounds of $A_2B^{1+}B^{3+}X_6$. Finally, they conducted statistical analysis on the attributes that were chosen to determine design principles for the development of fresh lead-free perovskites.

Moreover, a study presented by Jinxin et al.^[52] showed how 333 data points from nearly 2000 peer-reviewed papers were used to build ML models for designing PSCs. Their ML models included linear regression, KNN, RF, and artificial neural networks (ANN) for building two forecasting models, material property characteristics and device performance prediction. The higher R-value proves that the expected trend is consistent with actual experiments and PSC physics. The highest theoretically computed solar cell efficiency curve depending on the solar

spectrum has a bandgap area in the range of 1.15–1.35 eV, and this bandgap region predicts a PCE of above 25%.

Moreover, Felipe et al.,^[34] demonstrated a new data-driven optimization framework to bridge the mismatch between R&D and industrial production of solar cells. Further, their framework incorporated scalable inference and technoeconomic analysis using ML approaches to predict the root cause of the underperformance in PSCs. They also compared traditional R&D optimization versus their proposed total revenue optimization framework using linear, binned, and nonlinear functions. Consequently, they presented a case study for fabricating 144 PSCs choosing 12 various combinations of dominant processes. In addition, they proposed a surrogate-based blackbox model such as Gaussian process regression (GPR) and Bayesian optimization (BO).^[53]

In a conference, Maniell et al.^[54] demonstrated how the optoelectronics properties of PSCs can be predicted using ML methods. A model was developed for testing the bandgap of new different types of PSCs, and the bandgap was capable of predicting the chemical properties and material composition. $C_{5x}MA_{1-x}PbI_3$, $CSP(I_xBr_{1-x})_3$, and $MAPb_{1-x}Sn_xI_3$ were the perovskite materials used for testing and resulted in bandgaps ranging from 1.3 to 2.3 eV. In addition, their study presented a curve showing the predicted PCE values from the ML model versus the actual PCE from fabricated samples. Moreover, another result showed that the predicted value of the fabricated $CSSnI_3$ was 1.15 eV whereas the fabricated sample had a bandgap of 1.25 eV. Finally, their research article discussed various ML models such as ANN, RF algorithm, and support vector regression.

In addition, Li et al. demonstrated how ML can accelerate the discovery and investigation of PSCs.^[55] Their algorithms were based on inverse temperature crystallization (ITC) and were used to automate the process of evaluating single crystals of metal halide perovskites, which allowed the researchers to quickly identify and perfect the conditions for the synthesis of high-quality single crystals. Using 45 organic ammonium cations, 8172 metal halide perovskite synthesis processes were carried out. The screening enhanced the number of metal halide perovskite materials by five times and resulted in designing a new combination of PSCs such as $[C_2H_7N_2][PbI_3]$ and $[C_7H_{16}N_2][PbI_4]$. In addition, to enable experiment generation and data management, they used a software pipeline called Experiment Specification, Capture and Laboratory Autonomous Technology (ESCALATE). Further, their research added 17 new materials (a 400% increase) of metal halide perovskites, which are accessible via ITC. This helped identify conditions that lead to the formation of perovskite single crystals consisting of 19 of 45 target perovskite compositions.

In 2020, Yun et al.^[36] investigated the ML lattice constants for cubic perovskite A_2XY_6 compounds. Their dataset included a broad spectrum of Fmm group perovskite halides and a total of 79 samples. With lattice constants ranging from 8.109 Å to 11.790 Å, 79 cubic perovskite compounds were investigated. The ionic radii of [K, Cs, Rb, Tl], [Ge, Mn, Ni, Pd, Pt, Si, Cr, Pd, Ir, Mo, Pb, Re, Se, Ta, Sn, Te, Ti, W, Zr, Ru, Tc, Po, U, Os, Hf], and [F, Cl, Br, I] were among those used as descriptors. The GPR was used for determining the relation between the ionic radii and the lattice constants for cubic perovskites. They used MATLAB for the computational exploration of the model and achieved CC, RMSE, and mean absolute error (MAE) of 99.72%, 65%, and 0.44%, respectively.

In addition, Chenglong et al.^[56] presented a two-step ML approach for PSC design, which was based on 2006 PSC data points taken from peer-reviewed articles published between 2013 and 2020. The authors developed heuristics for high-efficiency PSC, thus, improving PCE dependent on doping of the electron transport layers (ETL). The main characteristic of their study was to determine the development of high-performance PCEs of PSCs. Their research showed that using SnO_2 and TiO_2 ETLs, mixed-cations perovskites, dimethyl sulfoxide, and dimethylformamide, as well as antisolvent treatment, led to even higher PCEs. Finally, they predicted that FA-MA-based PSC with a Cs-doped TiO_2 ETL and a Cs-FA-MA-based PSC with

S-doped SnO_2 ETL were also expected to show PCEs of up to 30.47% and 28.54%.

To expedite the identification of prospective PV cells from 2D perovskites, Hong-Jian et al.^[57] integrated atomic-level prediction with ML and DFT. Their model implemented a gradient boosting regressor (GBR), RF regressor, and an extra tree regressor (EXTR) ML for training a dataset of 2303 perovskite materials. Further, the trained model screened out 4828 materials and also pre-screened using DFT structural relaxation validation from 29 285 artificial perovskites. In fact, a maximum PCE of 30.35% and 26.03% was achieved for Sr_2VON and Ba_2VON .

Likewise, Elif et al.^[58] predicted the overall performance and bandgap in PSCs. In her analysis, she used eight different PSCs to forecast the bandgap and PCE of perovskites. Initially, they performed the bandgap estimation of perovskites from Tauc plots on UV-vis spectroscopy using the RF regression ML model with more than one decision tree and experimental approach. Later, they developed a model showing that the $J-V$ spectra predicted values for calculating the PCE. Their results showed that perovskites with bandgaps exceeding 0.99 eV could be used to model various new lead halide structure perovskites depending on the accurately predicted value of the bandgap.

Another case study presented by Xia et al.^[59] combined ML techniques with an efficient forward-inverse method to research $MAS_{nx}Pb_{1-x}I_3$ material and explored high-performance PSCs. With 14 physicochemical parameters and the Sn-Pb ratio as inputs, the E_g model of $MAS_{nx}Pb_{1-x}I_3$ was first developed for forward analysis, and the asymmetrically bowing relationship between the Sn-Pb ratio and the E_g of OMHP was used. The established NN-based models for PSC performance models showed good predictions for the data points and offered significant insights for PSC devices. Further, for the performance model, a comparison of the prediction model was made with the ML algorithms such as LR, SVR, KNR, RFR, and GBR. In fact, ML models with GBR performed best with values of R^2 , RMSE, and MAE reaching 0.9172, 0.0386, and 0.0325.

3.1.2. Organic Solar Cells

A rigorous framework involving the classification of the chemical structures in materials discovery was presented by Shinji et al.^[60] Further, the dataset of 249 Organic donor-acceptor pairs were computed based on equilibrium geometries and electronic properties such as DFT simulations. Initially, their study discussed predictions using Scharbar's model and resulted in a small-energy bandgap of 1.5 eV between the experimental and the computational energy bands. Moreover, they implemented k-NN regression for predicting OSCs characteristics and their PCEs. Finally, the study concluded that k-NN results in correlations of 0.6, which were further improved to 0.7 by implementing nonlinear kernel methods.

In addition, Harikrishna et al.^[61] investigated the PCE of OSCs using ML techniques. They developed a dataset of 280 small-molecule OSCs with 270 distinct donors. First, they analyzed the significance of orbitals in the energy conversion process and developed ML models using the characteristics of organic compounds to estimate the PCE for high-throughput virtual screening. In another study, they implemented ML methods

to study the correlations between the molecular properties and the device characteristics of an OSC.^[62] The authors designed ML methods based on 13 molecular properties as descriptors to predict the three device parameters such as V_{OC} , J_{SC} , and the fill factor. In addition, the calculations were carried out on Gaussian 09 package for a computational server having Intel Xeon 5115 CPUs. They combined multiple regression trees along with RF and GBRT to incorporate the ML methods. Further, screening of the potential compounds by these models results in high predictive ability ($r = 0.7$).^[63]

In a study by Daniele et al.,^[64] they performed computer-aided screening of polymer-based OSCs using ANN and RF models. Their dataset included 1000 experimental features, such as PCE, molecular weight of organic compounds, and various electronic properties. While the correlation coefficient of the ANN model was low, the RF model yielded higher accuracy in predictions. In another study, Min-Hsuan et al.^[65] applied RF regression to analyze nonfullerene-based OSCs, aiming to predict their overall efficiency. They compiled a dataset of 135 nonfullerene acceptor/donor pairs (117 nonfullerene acceptor materials and 30 donor materials) based on OSCs to investigate their electronic properties and device performance. Their ML model demonstrated high predictive power, achieving a coefficient of determination (R^2) of 0.85 for the training set and 0.80 for the testing set.^[66]

Furthermore, Xiaoyan et al.^[67] demonstrated an optimization technique to assess the potential of organic photovoltaic (OPV) materials and solar cell devices for industrial production. They presented an automated characterization of OPV materials, device performance, and photostability. The GPR ML technique drove the optimization method with optical absorption characteristics and indicated better prediction accuracies for PV electrical

characteristics. Moreover, the efficiency and photostability screening for 100 process conditions were completed in 70 h. They also proposed a model material system of PM6:Y6; completely automated device fabrication in air resulted in a maximum PCE of 14%.

In one of the latest papers published by Ahmad et al.,^[68] they discuss the implementation of ML to screen small-molecule donors for OSCs and molecular descriptors feed ML methods. The coauthors collected a dataset of 340 OSC devices with donors represented as small molecules, while acceptors as fullerenes for the ML-assisted pipeline suitable for small-molecule donors for Y6 (an electron acceptor). In addition, they performed ML analysis on an open-source platform called Konstanz Information Miner (KNIME). Further, for training the model, the dataset was divided into training sets, validating sets and external test sets. Also, the descriptors and experimental PCE were used as input to the ML model. They compared the result depending on various regression techniques, such as RF, LR, SVM, and k-NN, for the prediction of PCE. Using data from small donors paired with fullerenes, the SVM model was trained and showed higher prediction ability. The PCE of a few small-molecule donors linked with Y6 was predicted using their approach and developed are more than 1000 new small-molecule donors. Accordingly, the PCEs were anticipated, and the top 10 applicants with a PCE of over 13% were chosen in their study.

In addition, **Figure 7** shows the information on input data for various materials that were reviewed based on our defined research questions for three types of solar cells such as PSCs, OSCs, and hybrid. Most of the ML algorithms used in the process are highlighted to determine the resultant output in terms of electrical characteristics of reconfigurable solar cells.

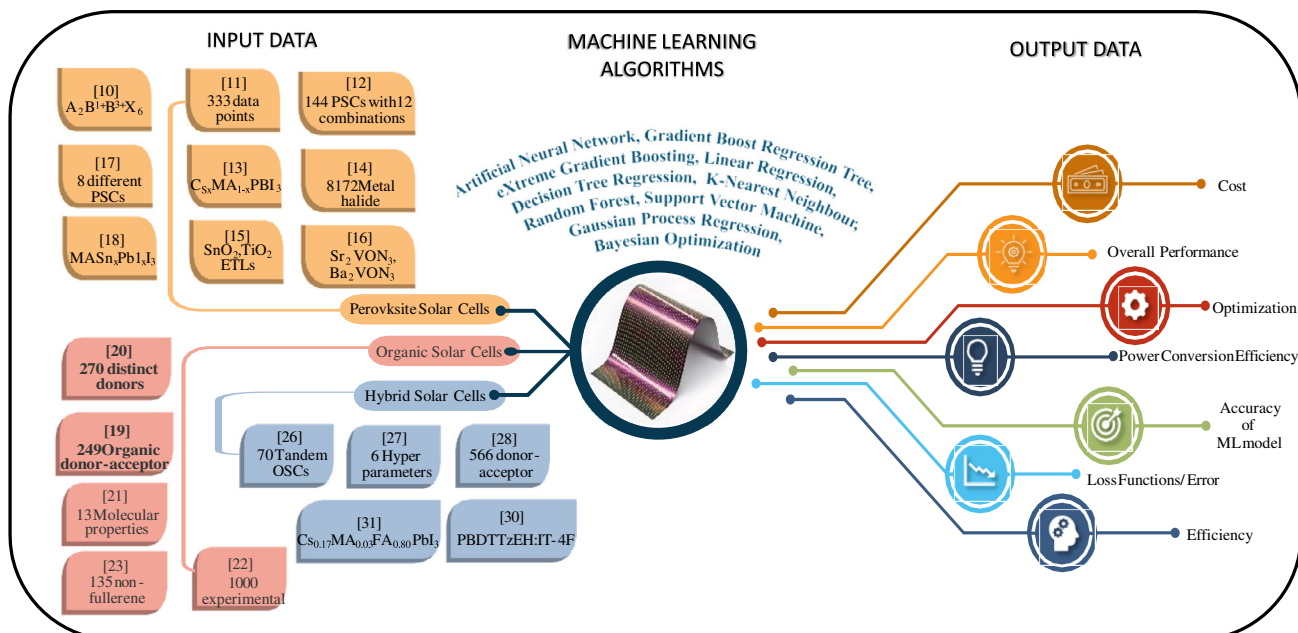


Figure 7. Information on input data for various materials that are reviewed based on our defined research questions for three types of solar cells such as PSCs, OSCs, and hybrid. The majority of ML algorithms used in the process are highlighted to determine the resultant output in terms of electrical characteristics of reconfigurable solar cells. The numbers in the box of the input data section are linked with reference numbers answering our RQ1 for PSCs, OSCs, and hybrid solar cells.

The numbers in the box of the input data section are accordingly linked with reference numbers answering research question 1 (RQ1) for PSCs, OSCs, and hybrid solar cells.

3.1.3. Hybrid Solar Cells

Another article presented by Min-Hsuan et al.^[49] investigated the performance and matching band structure for Tandem OSCs by implementing two ML methods, RF and the SVR. The ML techniques were initially developed using 70 tandem OSCs (37 conventional and 33 inverted tandem OSCs), which were used as the data points. Furthermore, to understand the structure, they calculated Pearson's correlation coefficient.

Among the two ML methods, the efficient method for forecasting solar efficiency was the RF regression having eight electronic features of selection.^[69]

Moreover, to address the stability concerns with PSCs, Tianmin et al.^[70] used a progressive ML algorithm to investigate the impact of input data by providing a reliable and accurate approach for deep mining of the hidden hybrid organic-inorganic solar cells. To predict the electronic bandgaps of hybrid organic-inorganic perovskites (HOIP), they implemented GBR, SVR, and KRR using material property. The best results from six hyperparameters were chosen. They also used DFT calculations for the chosen hybrid inorganic organic (HIO) perovskites and incorporated them into the Vienna Ab-initio simulation package (VASP). Their results show that the GBR model performs with the highest level of accuracy ($R^2 = 0.943$, $MAE = 0.203$, $MSE = 0.086$) when compared to the SVR ($R^2 = 0.826$, $MAE = 0.367$, $MSE = 0.276$) and KRR ($R^2 = 0.819$, $MAE = 0.387$, $MSE = 0.288$) models.^[71]

The effect of enhancing the descriptors using ML prediction for small-molecule-based OSCs was discussed by Zhi-Wen et al. in his study.^[72] The dataset consists of a total of 566 organic donor-acceptor (D/A) pairs found from the literature search, with 513 unique donors and 33 unique acceptors (including C_{60} , $P C_{61}BM$, $P C_{71}BM$, ITIC, IDTBR, IDIC, PDIs, etc.) among the donors. Further, they implemented ML models including the k -NN, KRR, and SVR to predict the PCE of hybrid solar cells. Also, the study examined Pearson's correlation coefficient for the combinations of descriptors, including donor molecules and device parameters.

In another study presented by Yao et al.,^[73] five different ML algorithms were used and gave 565 donor-acceptor combinations for training the dataset. Furthermore, to implement the material design and donor-acceptor pairs, the screening of non-fullerene in OSCs was performed. They used 565 donor/acceptor (D/A) combinations as training data sets in their study to assess the viability of these ML algorithms for use in directing material design and the screening of D/A pairs. Therefore, the ML techniques RF and BRT offer the best prediction capacities. Additionally, RF and BRT models are screened and estimated to be more than 32 million D/A pairs, respectively. Finally, six photovoltaic D/A couples are picked and synthesized so that their experimental and predicted PCEs can be used for critical comparison.

In an investigation presented by Kakaraparthi et al.,^[74] the co-authors used the RF model on an experimental dataset consisting

of 0.85 correlation coefficient for the ML of nonfullerene and polymer OSCs. Moreover, 200 932 conjugated polymers produced by the combinatorial coupling of acceptor and donor units were screened virtually. Additionally, a number of conjugated polymers centered on benzodithiophene and thiazolothiazole were created, produced, and studied using various alkyl chains in order to assess the efficacy of the ML model. In terms of the selection of alkyl chains, PBDTTzEH: IT-4F demonstrated a PCE of 10.10% and, thus, shows good predictions while using ML techniques.

One of the primary concerns with perovskites is their stability. As a result, Shijing et al.^[75] demonstrated how to discover the most stable organic-inorganic alloyed perovskites using a sequential learning framework. They introduced a data-fusion approach for estimating Gibbs free energy of mixing from DFT and experimentally analyzed degradation using aging tests. Moreover, they applied ML probabilistic constraints in an end-to-end BO approach to combine data from high-throughput degradation testing and first-principle simulations of phase thermodynamics. The results showed that perovskites centered at $CS_{0.17}MA_{0.03}F A_{0.80}P bI_3$ exhibit low optical change with increased temperature, moisture, and light having more than 17-fold stability improvement over $MAP bI_3$ by sampling 1.8% of the discretized $CS_xMA_yF A_{1-xy}P bI_3$ compositional space (MA, methylammonium; F A, formamidinium; P bI_3 , lead halide).

3.2. ML to Facilitate the Discovery of Solar Cells (Q2)

This section discusses the research articles and peer-reviewed journals related to the discovery of solar cells using ML techniques.

3.2.1. Discovery of Organic Structures

Tianmin et al.^[76] presented a goal-oriented approach to expedite the identification of hybrid organic-inorganic perovskites (HOIPs) suitable for photovoltaic applications from a pool of 230 808 HOIP candidates. They integrated ML techniques with density functional theory (DFT) calculations. After applying charge neutrality and stability criteria, 686 orthorhombic-like HOIPs with suitable bandgaps were selected and further screened using ML. The ensemble learning approach employed three ML models—GBR, SVR, and KRR—to predict the bandgaps of 38 086 HOIP candidates. Ultimately, DFT calculations confirmed 132 stable, nontoxic orthorhombic-like HOIPs (devoid of Cd, Pb, and Hg) with appropriate bandgaps for solar cell applications.

Oleksandr et al.^[77] employed ML in a feedback loop to learn from experimental data, recommend exploration of experimental parameters, and identify areas in the synthetic parameter space that would allow for highly monodispersed PbS quantum dots. Their findings revealed that a method that yields a record-large bandgap (611 nm exciton) of PbS nanoparticles with a well-defined excitonic absorption peak half-width at half-maximum (HWHM) of 145 meV which enables nucleation to prevail over growth by incorporating a growth-inhibiting precursor (oleylamine). They also enhanced monodispersity at longer wavelengths with HWHM values of 55 meV at 950 nm and 24 meV at 1500 nm, surpassing the best-reported values of 75 and 26 meV, respectively.

Double chalcogenide perovskites were investigated in a study presented by Michael et al.^[78] to find new photovoltaic absorbers that can take the place of CH₃NH₃PbI₃. ML approaches were used to categorize materials as potential photovoltaic absorbers using information from the periodic table, thus avoiding unnecessary computation due to the wide range of possible compounds. On the created data set, a random forest method obtains a crossvalidation accuracy of 86.4%. Traditional and statistical approaches are used to identify over 450 potential alternatives, with Ba₂AlNbS₆, Ba₂GaNbS₆, Ca₂GaNbS₆, Sr₂InNbS₆, and Ba₂SnHfS₆ emerging as the most promising options when thermodynamic stability, kinetic stability, and optical absorption are taken into account.

Nastaran et al.^[79] in a study showed that ML techniques were used by computationally intensive DFT simulations to quickly and precisely estimate the properties of OPV materials. One-hot descriptors, OPV PCE, open-circuit potential (V_{oc}), short-circuit density (J_{sc}), HOMO energy, LUMO energy, and the HOMO–LUMO gap were all quantified in the study. With a standard error of 0.5 for a percentage of PCE for both the training and test sets, the most reliable and predictive models were able to predict PCE. Their methodology helps to expedite the design of OPVs for use in green energy applications by prescreening possible donor and acceptor materials.

An ML framework introduced by Noor et al.^[80] involved optimizing the capping layer of perovskite degradation. They featured 21 organic halide salts, used them as capping layers on (MAPbI₃) films, aged them rapidly, and implemented supervised ML and Shapley values to identify factors determining stability. They discovered a correlation between higher MAPbI₃ film stability and organic molecules' limited number of hydrogen bonding donors and tiny topological polar surface area. Phenyltriethylammonium iodide (PTEAI), the best organic halide, successfully increases the stability lifespan of MAPbI₃ by 4.2 times over bare MAPbI₃ and 1.3–0.3 times over cutting-edge octylammonium bromide (OABr).

Zhilong et al.^[81] developed a goal-oriented approach that leverages ML to accelerate ab initio predictions of undiscovered spinels in the periodic table. Utilizing this method, they successfully identified eight spinels with direct bandgaps and room-temperature thermal stability out of 3,880 unknown spinels (CaAl₂O₄, CaGa₂O₄, SnGa₂O₄, CaAl₂S₄, CaGa₂S₄, CaAl₂Se₄, CaGa₂Se₄, CaAl₂T_e₄). They created a semiconductor classification model based on the XGBoost technique, which demonstrated a strong structure–property relationship, a high prediction accuracy of 91.2%, and low computational cost of just a few milliseconds. This proposed goal-oriented strategy facilitates the design of a broad range of energy materials, reducing the research time required for spinel screening by almost 3.4 years.

The accuracy for predicting the bandgap of an OSC is a vital factor in terms of the characterization of solar cell devices. Accordingly, Yiming et al.^[82] used ML algorithms to predict the performance of different architectures for the compound ABX₃ type in PSCs. Also, they gathered 227 experimental datasets consisting of the bandgap of perovskites extracted from recently published 1254 publications. For their model, they used ML methods such as RF, XGBoost, LR, *k*-NN, SVR, and multi-layer perceptron (MLP). Their prediction analysis from ML

models showed that B-site metal and the X-site halogen ion have a significant impact on bandgaps of the ABX₃-type perovskites from SHAP explanations.

Muhammad et al.^[83] did the critical analysis of the small-molecule donors for OSCs such as fullerene using the ML methods. In order to train the ML model, they used molecular descriptors as an input and consecutively, they implemented a number of ML techniques to measure the best ML algorithm for the desired outcome. The dataset used in the study consists of 250 OSCs having a combination of acceptors and donors as fullerenes (*P* C₆₁BM and *P* C₇₁BM). They used the platforms like Konstanz Information Miner (KNIME) and Weka platforms to implement the ML model and thus, the RF model resulted in the best predictive model with Pearson's coefficient as 0.93. Finally, to determine the most efficient materials, the PCE value for the small-molecular donor was predicted.

3.2.2. Discovery of Hybrid Halide Structures

With multiple newly developed, computationally economical, and high-performing (Pearson's correlation coefficient = 0.7–0.8) ML models employing pertinent descriptors, Harikrishna et al.^[84] carried out high-throughput virtual screening of 10 170 candidate compounds, assembled from 32 distinct building blocks. Furthermore, to create effective molecules, crucial building elements are recognized, and new design principles are implemented. Additionally, 126 candidates are suggested for synthesis and device fabrication with theoretically projected efficiency >8%.

Moreover, Shohei et al.^[85] devised a rapid material search scheme based on materials informatics for PSC materials following the existence of viable alternative perovskites **Table 2**. In fact, more than 28 million double-perovskite-like compounds were screened using this method. Additionally, five organic–inorganic tin–halide perovskites as well as 17 potassium-, sodium-, and ammonium-based tin–halide perovskites were among the 24 most promising possibilities found. Promising solar cell materials included two perovskites based on transition metals.

Further, Lifei et al.^[86] constructed *N*-annulated perylene sensitizers and put forth one goal-directed approach that combined quantum chemical analysis with data mining approaches. Using MLR to build the robust quantitative structure–property relationship (QSPR) model, they were able to identify the key characteristics using genetic algorithm (GA). The potential dyes were then created using the model's recommendations. The proposed molecules' overall PCEs were anticipated by the model to be 15.7%, up 22.0% from reference dyes C₂₈₁.

For the electrical characteristics of metal halide perovskites (MHPs), which possess a vast materials design space in the billions range, Wissam et al.^[87] employed CNN to create a predictive model. Furthermore, they demonstrated that as compared to simple techniques, a well-designed hierarchical ML strategy offers a higher degree of predictability in terms of MHP features. The bandgaps for the MHPs' lattice constants, octahedral angle, and RMSE were all calculated using the hierarchical ML scheme, and the corresponding RMSE values were 0.01 eV, 5 degrees, and 0.01.

Table 2. Literature discussing the ML for facilitating the discovery of solar cells.

230 808 HOIP, 686 orthorhom-bic	GBRT, SVR, KRR	Nontoxic orthorhombic-like HOIPs (free of Cd, Pb, Hg)	[65]
Double chalcogenide per-ovskites, 450 alternatives	ML with DFT simulations	Ba ₂ AlNbS ₆ , Ba ₂ GaNbS ₆ , Ca ₂ GaNbS ₆ , and Ba ₂ SnHfS ₆	[67]
Organic photovoltaic materials, one hot descriptor	Intensive DFT simulations	Design of OPVs' prescreening possible donor and acceptor materials	[68]
21 organic halide salts	Supervised ML and Shapley values	Phenyltriethylammonium iodide (PTEAI)	[69]
3880 unknown spinels	XGBoost method	CaAl ₂ O ₄	[70]
227 experimental dataset	RF, XGBoost, LR, k-NN, SVR and MLP	P C ₆₁ BM and P C ₇₁	[71]
250 OSCs dataset	RF model	ABX ₃ -type perovskites	[72]
28 million double-perovskite	- - -	17 sodium-, potassium-, and ammonium-based tin-halide perovskites	[73]
N-annulated perylene sensitiz-ers	MLR and QSPR model	C ₂₈₁	[74]
Metal halide perovskites (MHPs)	CNN	0.01 eV, 5 degrees, and 0.01	[75]
10 000 candidates	QSPR	eight promising organic dyes	[76]
Lead-free halide perovskite material	RF, RR, SVR, and GBRT	Lead-free halide double PSC	[77]
78 400 DHOIPs	Integrating ML techniques	19 promising ones, HSE06 calculations	[78]

Yaping et al.^[88] combined ML with computational quantum chemistry results in the establishment of an accurate, reliable, and interpretable QSPR model. The predictive model was used to perform virtual screening and assess synthetic accessibility in order to discover novel, efficient, and easily synthesized organic dyes for DSSCs. Finally, out of almost 10 000 candidates, eight promising organic dyes with high PCE and synthetic accessibility were eliminated.

Moreover, Zongmei et al.^[89] investigated the discovery of PSC materials via ML stability and calculated the bandgap of lead-free halide perovskite materials. They performed a comparative analysis of four different ML techniques such as the RF, ridge regression, SVR, and the GBR tree. Among these four ML techniques, XGBoost gave the highest predictive performance i.e., R²:0.9935 and MAE:0.0126 in terms of thermodynamic stability, and accordingly, the RF gave the highest predictive performance, that is, R²:0.9410 and MAE:0.1492 for bandgap analysis of the lead-free halide double PSCs. Moreover, their study showed an interesting result that XBoost performs best when considering the thermodynamic stability and electronegativity's linear correlation.

Furthermore, Jialu et al.^[90] demonstrated that the discovery of double-hybrid organic-inorganic perovskites (DHOIPs) can be accelerated by integrating ML techniques, high-throughput screening, and DFT. In contrast to other studies, the anisotropy of organic cations of DHOIPs was first assessed, and then the properties were predicted using an ML technique using low-level calculations to predict the properties of DHOIPs accurately. From 78 400 DHOIPs, 19 promising ones with suitable bandgaps for solar cells were selected and verified using HSE06 calculations.

John et al.^[91] investigated that bias, temperature, light, and H₂O, O₂, and air pressure affected device performance and recovery. They first talked about important studies that assess the 3 R cycle's capabilities of perovskites and how ML algorithms may help determine the best values for each operating parameter. They then looked at perovskite dynamics and degradation, highlighting the difficulties in understanding this 3 R cycle. Finally, they suggested an ML paradigm with a shared knowledge

library for improving long-term performance and forecasting device performance recovery.

3.2.3. Discovery of Solar Cells Using Natural Language Processing

In another study, a framework related to the high-throughput synthesis of the PSCs was discussed with ML image recognition used for automated characterization by Jeffrey et al.^[92] Perovskite single-crystal synthesis was carried out at high throughput, and the results were identified using convolutional neural network (CNN)-based image recognition. Also, they quickly created 96 distinct crystallization environments using a protein drop setter and then examined the crystals. On the other hand, a CNN was used to determine if crystals had been produced using a dataset of 7,000 photographs. Then, a larger dataset of 25 000 photos was employed with this classifier. The first synthesis of (3 - P LA)₂P bCl₄ was then achieved after they employed ML modeling to predict the ideal conditions for synthesizing a novel perovskite single crystal.

A study presented by Lei et al.^[93] showed ML techniques based on natural language processing (NLP) to predict the properties of solar cell materials, which were then examined using first-principle calculations. The aim of the study was to reduce the amount of human interaction and enable computers (without supervision) to learn the latent knowledge about solar cell materials depending on the textual data and generate predictions about the composition of solar cells. The first-principles calculations were used to determine the projected material's density of states, UV-vis absorption spectra, as well as band structures in order to assess their suitability for photovoltaic applications. The formula and targeted keywords for solar cells were represented as vectors in the ML process, which facilitated the successful relationship extraction of the materials and their applications. The ML model was validated using first-principles calculations on the unusual solar cell materials included in the list, and the projected candidates, such as AS₂O₅, have good electrical and optical characteristics that are suitable for solar cell applications.

3.3. ML for Solar Cell Optimization

The focus of this section is RQ3, which involves examining the optimization techniques used with ML algorithms to develop optimized and reconfigurable solar cells. The technical research articles that showed experimental work for implementing the ML algorithms for discovering the optimized solar cells is included.

Moreover, **Figure 8** displays multiple layered internal architectures of solar cells and the necessary chemical components for creating reconfigurable solar cells. Specifically, Figure 8a depicts the perovskite's chemical structure with a carbon composition, whereas Figure 8b shows the arrangement of the chemical components in a solar cell and Figure 8c shows the various layers of a solar cell that have been sliced for clarity in depicting the solar cell architecture. Finally, Figure 8d showcases the outer layer of a solar cell, including Ag, BCP, PCBM, perovskite, Poly-TPD, ITO, and glass.

3.3.1. Donor/Acceptor Ratio for Higher PCE

Most scientific advancements in the field of materials have been produced experimentally, frequently using one variable at a time testing. However, neither are the properties of materials- based systems straightforward nor related.^[94] Authors in another study,^[95] claim that the optimization of OSCs has a high level of complexity due to the high complexity and interconnectivity of different components. Changing one component can have an unforeseen impact on other components. Hence ML can play a vital role in the optimization process of OSCs. They used P DCT BT : P C₇₁ solar cell and observed the effect of donor/acceptor ratio, total concentration, spin speed, and additive volume on PCE(%). The authors applied SVM using the radial basis function. They conducted two sets of experiments, where they used optimized results of the first experiment in the second experiment and found a significant increase in PCE of fabricated

devices.^[35] In the first set of the experiment, only 3 out of 15 devices were above the threshold (PCE 6.3%); however, in the second, all 13 devices produced PCE above the threshold.

3.3.2. Conductivity Optimization of Solar Cells

SVM regression was used in another study^[96] for the optimization of p - CZS/ n - Si, p - CZS/ p^+ n - Si heterogeneous solar cells. SVM was implemented with a radial-based function using Scikit-learn^[97] in python. They used tenfold crossvalidation to tackle the problem of over-fitting. They predicted the figure of merit (FOM) from film conductivity and optical transmission in the desired transmission range. Optimization results show that FOM increased from 14.8 to 173 μ . Furthermore, current density increased from 11.8 to 17.9 mA Cm^{-2} for p - CZS/ n - Si solar cells and from 13.8 to 18.0 mA Cm^{-2} p - CZS/ p^+ n - Si for solar cells. The authors claimed their approach is valid for any general application to any material synthesis process with multiple parameters.^[98]

3.3.3. Selection of Donor/Acceptor Pairs

From 2010 to 2017, 320 organic donor and acceptor pairs (heterojunction solar cells) were reported in the literature. These 320 donors and acceptors can make 19 912 combinations. Authors in another study^[99] applied distance-based ML techniques KNN and SVM to optimize PCE. They provided a list of unexplored donor and acceptor combinations that can be helpful in the future in fabricating highly efficient solar cells. The use of back propagation neural network, deep neural network, SVM, and the random forest is reported in another study^[100] to predict highly efficient OSCs. The dataset contained 1719 realistic donor materials of OSCs. The authors used images, ASCII strings, and fingerprints as input, and concluded that fingerprints with 1000 bits can provide higher conversion efficiency. The authors also proposed ten new materials.

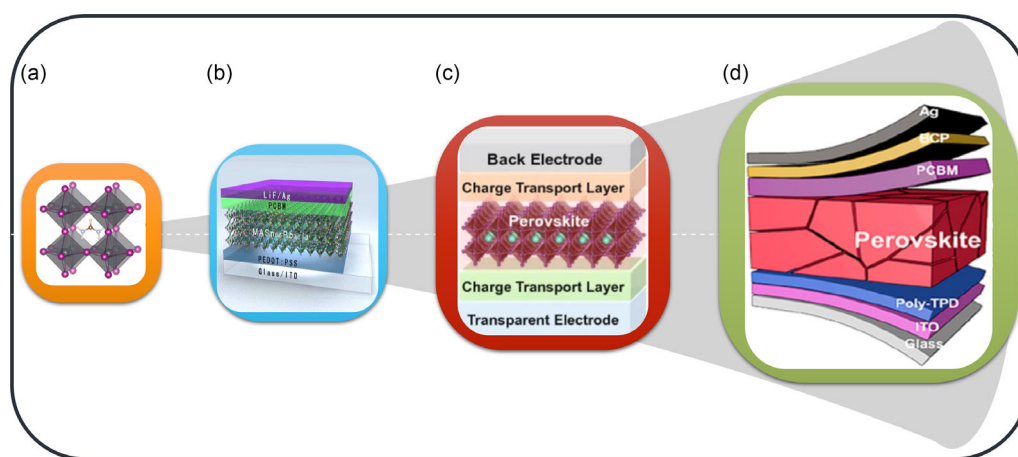


Figure 8. The figure displays multiple layered internal architectures of solar cells and the necessary chemical components for creating reconfigurable solar cells. Specifically, a) the perovskite's chemical structure with a carbon composition Reproduced with permission.^[139] Copyright 2023, MDPI. b) the arrangement of the chemical components in a solar cell, Reproduced with permission.^[148] Copyright 2021, Wiley-VCH GmbH. c) the various layers of a solar cell that have been sliced for clarity in depicting the solar cell architecture Reproduced with permission.^[140] Copyright 2021, Springer Nature, and d) the outer layer of a solar cell, including Ag, BCP, PCBM, Perovskite, Poly-TPD, ITO, and glass Reproduced with permission.^[141] Copyright 2020, Elsevier.

3.3.4. Stability Optimization

Stability is a good indicator of the life span of a solar cell. Multiple parameters can affect the stability of OSCs. Authors in another study^[101] optimized these parameters using sequential minimal optimization regression on a dataset obtained from the website of Danish Technical University (DTU).^[102] Authors have presented shortlisted layer-wise materials with the highest weights in sequential minimal optimization regression. These materials are the most influential materials governing the stability and performance of OPV devices.^[103]

3.3.5. Copper Content Optimization in CdTe Solar Cells

Cu is essential in CdTe solar cells as back contact and doping agent. Diffusion depth optimization of Cu resulted from diffusion annealing, and cool down in the fabrication of CdTe solar cell was reported in another study.^[104] ANN predicts data generated from software simulation using the Keras library in python. ANN was fed with temperature and duration of diffusing process time. Results show that the predicted and actual depths are only $0.009\mu\text{m}$ apart.

3.3.6. Optimization of Diode Model for Solar Cell Simulations

A bioinspired modified spotted hyena optimization algorithm was implemented in another study,^[105] to compare one diode model, two diode modes, and three diode model solar cells in MATLAB. The authors obtained I-V and P-V curves. They found that the three-diode model is the most accurate model.

3.3.7. Optimization of Spray Plasma Processing

Optimization is a common theme in materials research when synthesizing a particular material or determining the ideal processing conditions to obtain the desired attribute. The difficulties emerge from the fact that there are several parameters whose weights might influence the outcomes. Additionally, gathering experimental data takes time and money. Authors in ref. [106], presented the work of ref. [107], where BO was used to optimize the rapid plasma process. The authors used six different parameters as input that affect PCE: linear speed of spray, substrate temperature, the flow rate of precursor, gas flow rate into plasma nozzle, the height of plasma nozzle, and plasma duty cycle, while some other parameters were kept constant such as precursor formulation, concentration, etc. The optimization result showed that PCE increased from 15% to 17 %.

3.4. ML for the Efficient Fabrication of Solar Cells

Most research articles cover various ML algorithms used to fabricate PSCs effectively. However, in this section, our emphasis is on RQ4, which examines the most optimal ML algorithms that have proven effective in identifying efficient techniques for fabricating PSCs.

PSCs are cheap to fabricate and as a result most researchers fabricate these low-cost solar cells by trial and error. Also, fabricating a solar cell consists of a large percentage of permutations

and combinations of various physical parameters such as materials used, doping layers, the thickness of the different layers, meshing, contacts, bulkiness, etc. In addition, solutions-based techniques for fabricating solar cells require less time to manufacture. However, they exhibit stability concerns. Therefore, we review the ML methods for designing a reconfigurable PSC.

Liu et al.^[108] demonstrated how ML can be used to develop a sequential learning architecture that produces PSCs. The study used rapid spray plasma processing (RSP) method to develop open-air perovskite devices and implemented various methods for optimizing the process. The researchers achieved an efficiency improvement of 18.5% by screening only 100 process scenarios, despite limited experimental budget. This was mainly due to three key innovations, 1) prior experimental data was used as a probabilistic constraint to enable flexible knowledge transfer between experimental processes, 2) subjective human observations and ML insights were combined when selecting the next experiments, and 3) an adaptive strategy that used BO was employed to identify the region of interest before conducting local exploration for high-efficiency devices. These innovative approaches enabled the researchers to rapidly identify the optimal conditions for producing PSCs and resulted in a significant improvement in efficiency.

Another research article presented by Vincent et al.^[109] discussed a quick and simple tool for identifying the primary losses in PSCs. To comprehend the light intensity dependency of the open-circuit voltage and how it relates to the main recombination mechanism, their model used large-scale drift-diffusion simulations. The ML algorithm was developed using more than 2 million simulations and resulted in a prediction accuracy of up to 82%.

In their research, Xabier et al.^[110] used big data to discover OSC materials, such as non-fullerene acceptors and low-bandgap donor-based polymers. They examined computational methods for selecting promising chemicals from online libraries and outlined key high-throughput experimental screening and characterization techniques for OSCs. Their work achieved unparalleled data generation rates, enhancing big data preparedness, and applied ML algorithms to identify quantitative structure-activity relationships and extract molecular design insights for OPV.

Aaron et al.^[111] combined design of experiments (DOE) and ML approaches to optimize small-molecule OPV cells based on DRCN5T donors and nonfullerene acceptors like ITIC, IT-M, and IT-4 F. They determined the optimal experimental processing parameters for PCE using ML-generated PCE landscape maps. Cagla et al.^[112] investigated the influence of manufacturing materials, deposition techniques, and storage conditions on the stability of 404 organolead halide PSCs using a dataset from 181 publications. They employed association rule mining and decision tree-based ML methods for their analysis.

Nahdia et al.^[113] proposed a method to analyze the material and device performance including the experimental, modeling, and ML techniques. Moreover, they also included various manufacturing conditions for the measurement of device performance by providing a set of electrical as well as electronic device characteristics that result in a large and efficient improvement for the respective solar energy harvesting devices. Following, they considered some of the key mechanical properties such as

annealing temperature, surfactant selection, and charge carrier dynamics in OSCs. Similarly, another study presented by Bart et al.^[114] consisted of the predictions that are related to the bandgap of organic crystal structures using the ML techniques. The two cutting-edge models combined yield a MAE of 0.388 eV, or 13% of the average bandgap of 3.05 eV, for the ensemble. The bandgap for 260 092 materials in the Crystallography Open Database (COD) is predicted using the trained models.

Fan et al.^[115] presented the ML-assisted designing and fabrication of solar cells. The elements can be divided into four sub-categories: data measurement, material properties, optimization of device architectures, and optimization of fabrication processes. The typical types of ML techniques discussed involve ANN, GA, PSO, SA, RF, etc. Among them, ANN and GA are the two ML techniques that are most frequently used.

4. Open Research Issues and Future Outlook

In this section, we highlight some of the key insights that the authors have notably found interesting and consecutively, present the future outlook of the potential research incorporating ML and the discovery of new materials to develop re-configurable solar cells. In addition, this section also includes the limitations and pitfalls of the ongoing research that need to be addressed for developing efficient, robust, and stable solar cell architectures.

According to our review, few articles were published in the domain of using ML for fabricating solar cells. Furthermore, our study revealed that input data was clustered around PSCs, OSCs, and hybrid solar cells. Furthermore, most research used the ANN, GBRT, XGBoost, EXTR, LR, DTR, KNN, RF, SVM, SVR, GPR, and BO algorithms to determine output characteristics such as cost, PCE, the accuracy of the ML model, loss function, and error. Lastly, ML was used for optimizing the following solar cell parameters: donor/acceptor ratio, conductivity, donor/acceptor materials, stability optimization, copper content optimization, and spray plasma processing.

4.1. Limitations

Although there are numerous advantages of using ML for solar cell discovery, there are several open issues. From our systematic review, we came across multiple challenges that need to be addressed with regard to the discovery of new low-cost solar cells. Key among these challenges are as follows.

4.1.1. Vulnerability of the Input Data

As previously mentioned, the majority of low-cost solar cells are fabricated by trial and error in a research environment, leading to high vulnerability in terms of input data.^[116] Therefore, as a necessary step, all the ML algorithms should undergo model validation.^[117] Another key issue is data scarcity in the field of data-driven solar materials science.^[118] Also, text mining and picture recognition are too considered as solutions for overcoming these primary problems involving poor quantity and quality of the datasets.^[119]

4.1.2. Stability of Thin-Film Solar Cells

One of the key concerns in designing low-cost solar cells in the real environment is the stability of organic, inorganic, and hybrid solar cells due to the different compositions of chemical components. These solar cells are very unstable and have a short life period.^[120] Previously, studies have shown that solar cell efficiency and stability are inversely proportional. Additionally, addressing the critical aspects of stability such as thermal, moisture, and chemical composition is essential^[121]

4.1.3. Unreliable Forecasts

A significant concern when using ML algorithms in solar cell discovery is the potential for imprecise predictions and results from the ML models.^[122] Generally, ML algorithms provide confidence intervals for the estimated and anticipated values related to solar cells. However, the predicted values for the discovery of solar cells seem to approach up to a maximum of 95% using the GPR and BO using the probability distribution, which sometimes proves to result in the poor fabrication of solar cells. Therefore, the ML models' prediction models need to be classified properly to avoid such discrepancies.^[123]

4.1.4. Rigorously Fabricating Solar Cells in Labs

The researchers are rigorously fabricating solar cells depending upon the hit and trial methods, which waste a lot of time, resources, and materials. In addition, if the researchers follow the same procedure in the upcoming years, it is noted that it will further delay the discovery of new materials used to fabricate solar cells.^[124] Moreover, using the permutation and combinations of different layers, electrical characteristics, and other components required to design the solar cells and fabricate solar cells in the laboratory will lead to other consequences which can be avoided with the use of ML techniques and AI integration.^[125]

4.1.5. Data Scarcity and Ineffective Data Analysis

First, it is noted from the study that there is a lack of data availability and, thus, poor data analysis. Second, it is advised to integrate feature engineering, modeling, and domain technical expertise to increase the effectiveness of the created ML model. In parallel, validation experiments should be run to verify the analytical outcomes of the ML model, such as the high-performing prediction candidate. Only a few research studies have used experiments to validate their forecasted materials.^[126]

4.2. Future Outlook

The future goals and prospective outlook for discovering new low-cost solar cells are mentioned below. Initially, there was a large room for data collection and monitoring to provide input to ML models. Moreover, the extracted data needs feature scaling and data-preprocessing to be used effectively in ML algorithms. Therefore, an appropriate data selection technique must be used to interpolate or extrapolate the data depending on various dependent and independent variables in feature selection. In

addition, since ML and AI techniques have recently gained significant importance, adversarial robust ML techniques will play a vital role in forecasting and predicting the design of solar cell architectures.^[82,127–130]

Moreover, ML can aid in predicting the performance of solar cells, leading to the development of dependable and cost-effective solar cells. By predicting the performance of solar cells before production, manufacturers can save resources and avoid producing poorly performing cells. Additionally, ML is being utilized to create new materials for cost-effective solar cells. By analyzing large amounts of data from various sources, ML can identify materials with desired characteristics for solar cells, reducing the cost and time spent on experimentation and speeding the process of developing new materials.

Since low-cost solar cell fabrication in a research laboratory is cheap, most researchers tend to retrospectively appreciate the performance of their design after first fabricating the solar cell by trial and error. Instead, we believe it is more beneficial to perform these predictions using robust ML algorithms, which will help design and fabricate more efficient solar cells. Adopting this approach will expedite the solar cell design process. There is also space for research related to the generalized explanations of data extraction and interpretation and to achieve more accurate ML models. In general, the accuracy of the ML model depends on the input data. Researchers across the globe should target to extract sufficient data and make it available online to help the scientific community discover low-cost, high-performance solar cells.

5. Conclusions

In conclusion, this comprehensive review evaluated a broad range of ML techniques for optimizing the performance of low-cost solar cells for miniaturized electronic devices. We short-listed 58 research articles from a pool of 18 380 research publications that met our inclusion criteria and aimed to answer our research questions. Our review indicates that a significant proportion of research focuses on data-driven approaches and ML techniques for discovering low-cost solar cells, with a third of publications targeting ML algorithms in the fabrication process.

Our systematic review suggests that ML techniques have the potential to accelerate the discovery of new solar materials and architectures. Future research can expand on these findings by exploring and developing new ML techniques for solar cell optimization. Additionally, it is essential to address the scalability and sustainability of low-cost solar cell technologies to enable large-scale commercialization. Ultimately, the application of ML techniques in solar energy can revolutionize the industry and pave the way for a cleaner and more sustainable future.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

artificial intelligence, energy harvesting, machine learning, materials discovery, nanofabrication, optimizations, photovoltaics, wearable devices

Received: February 3, 2023
Revised: March 31, 2023
Published online: August 23, 2023

- [1] S. Mukherjee, S. Suleman, R. Pilloton, J. Narang, K. Rani, *Sensors* **2022**, 22, 4228.
- [2] K. Zhao, Z. Yang, S. Liu, *Adv. Funct. Mater.* **2019**, 29, 1904014.
- [3] Y. Liu, H. Khanbareh, M. A. Halim, A. Feeney, X. Zhang, H. Heidari, R. Ghannam, *Nano Select* **2021**, 2, 1459.
- [4] J. Zhao, R. Ghannam, K. O. Htet, Y. Liu, M.-K. Law, V. A. Roy, B. Michel, M. A. Imran, H. Heidari, *Adv. Healthcare Mater.* **2020**, 9, 2000779.
- [5] M. A. Hannan, S. Mutashar, S. A. Samad, A. Hussain, *Biomed. Eng. Online* **2014**, 13, 1.
- [6] R. Das, F. Moradi, H. Heidari, *IEEE Trans. Biomed. Circuits Syst.* **2020**, 14, 343.
- [7] W. G. De Voogt, *Am. J. Cardiol.* **1999**, 83, 187.
- [8] E. Stach, B. DeCost, A. G. Kusne, J. Hattrick-Simpers, K. A. Brown, K. G. Reyes, J. Schrier, S. Billinge, T. Buonassisi, I. Foster, C. P. Gomes, *Matter* **2021**, 4, 2702.
- [9] M. A. Wahba, A. S. Ashour, R. Ghannam, *IEEE Access* **2020**, 8, 170336.
- [10] M. Haroon, M. R. S. A. Janjua, *J. Comput. Electron.* **2022**, 21, 40.
- [11] R. E. Kumar, A. Tiihonen, S. Sun, D. P. Fenning, Z. Liu, T. Buonassisi, *Matter* **2022**, 5, 1353.
- [12] P. Bhuvaneswari, R. Balakumar, V. Vaidehi, P. Bala-muralidhar, in *2009 First Inter. Conf. Computational Intelligence, Communication Systems and Networks*, IEEE, Piscataway, NJ **2009**, pp. 57–61.
- [13] T. Hatem, Z. Ismail, M. G. Elmahgary, R. Ghannam, M. A. Ahmed, S. O. Abdellatif, *IEEE Trans. Electron Devices* **2021**, 68, 5319.
- [14] T. Hatem, M. G. Elmahgary, R. Ghannam, M. A. Ahmed, S. O. Abdellatif, *J. Mater. Sci.: Mater. Electron.* **2021**, 32, 25318.
- [15] M. Jacoby, *Chem. Eng. News* **2016**, 94, 30.
- [16] M. C. Beard, J. M. Luther, A. J. Nozik, *Nat. Nanotechnol.* **2014**, 9, 951.
- [17] R. Ghannam, P. V. Klaine, M. Imran, *Power Systems*, Springer Singapore **2019**, pp. 121–142.
- [18] H. J. Snaith, *J. Phys. Chem. Lett.* **2013**, 4, 3623.
- [19] M. Haroon, A. A. Al-Saadi, M. R. S. A. Janjua, *J. Phys. Org. Chem.* **2022**, 35, e4314.
- [20] Ç. Odabaşı, R. Yildirim, *Nano Energy* **2019**, 56, 770.
- [21] B. Yilmaz, R. Yildirim, *Nano Energy* **2021**, 80, 105546.
- [22] H. D. Pham, Z. Wu, L. K. Ono, S. Manzhos, K. Feron, N. Motta, Y. Qi, P. Sonar, *Adv. Electron. Mater.* **2017**, 3, 1700139.
- [23] A. Mahmood, J. Yang, J. Hu, X. Wang, A. Tang, Y. Geng, Q. Zeng, E. Zhou, *J. Phys. Chem. C* **2018**, 122, 29122.
- [24] H. Zhou, L. Yang, S. C. Price, K. J. Knight, W. You, *Angew. Chem.* **2010**, 122, 8164.
- [25] F. Chen, M.-H. Liu, R.-Q. Piao, D.-L. Zhang, Y. Wang, *Opt. Mater.* **2023**, 136, 113439.
- [26] I. El Naqa, M. J. Murphy, *Machine Learning in Radiation Oncology*, Springer International Publishing, Switzerland **2015**, pp. 3–11.
- [27] H. Wang, Z. Lei, X. Zhang, B. Zhou, J. Peng, *Deep Learn* **2016**, 98.
- [28] J. W. Shavlik, T. Dietterich, T. G. Dietterich, *Readings in Machine Learning*, Morgan Kaufmann **1990**.
- [29] F. Provost, R. Kohavi, *Mach. Learn.* **1998**, 30, 127.
- [30] A. Mahmood, J. Hu, A. Tang, F. Chen, X. Wang, E. Zhou, *Dyes Pigm.* **2018**, 149, 470.
- [31] M. I. Jordan, T. M. Mitchell, *Science* **2015**, 349, 255.
- [32] S. Shalev-Shwartz, S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, Cambridge **2014**.
- [33] D. H. Wolpert, W. G. Macready, *IEEE Trans. Evol. Comput.* **1997**, 1, 67.

- [34] F. Oviedo, Z. Ren, X. Hansong, S. I. P. Tian, K. Zhang, M. Layurova, T. Heumueller, N. Li, E. Birgersson, S. Sun, B. Mayurama, *arXiv pre-print arXiv:2004.13599* **2020**.
- [35] A. Mahmood, A. Irfan, J.-L. Wang, *Chin. J. Polym. Sci.* **2022**, *40*, 870.
- [36] L. Zhang, M. He, S. Shao, *Nano Energy* **2020**, *78*, 105380.
- [37] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature* **2018**, *559*, 547.
- [38] N. Parikh, M. Karamta, N. Yadav, M. M. Tavakoli, D. Prochowicz, S. Akin, A. Kalam, S. Satapathi, P. Yadav, *J. Energy Chem.* **2022**, *66*, 74.
- [39] M. Workman, D. Z. Chen, S. M. Musa, in *2020 Int. Conf. Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, IEEE, Piscataway, NJ **2020**, pp. 1–5.
- [40] H. R. Parikh, Y. Buratti, S. Spataru, F. Villebro, G. A. D. Reis Benatto, P. B. Poulsen, S. Wendlandt, T. Kerekes, D. Sera, Z. Hameiri, *Appl. Sci.* **2020**, *10*, 8834.
- [41] D. Bash, F. H. Chenardy, Z. Ren, J. J. Cheng, T. Buonassisi, R. Oliveira, J. N. Kumar, K. Hippal-gaonkar, *Digital Discovery* **2022**, *1*, 139.
- [42] Q. Tao, P. Xu, M. Li, W. Lu, *npj Comput. Mater.* **2021**, *7*, 1.
- [43] H. Michaels, I. Benesper, M. Freitag, *Chem. Sci.* **2021**, *12*, 5002.
- [44] H. Wagner-Mohnsen, P. P. Altermatt, in *2021 Int. Conf. Numerical Simulation of Opto-Electronic Devices (NUSOD)*, IEEE, Piscataway, NJ **2021**, pp. 51–52.
- [45] Y. Cui, P. Zhu, X. Liao, Y. Chen, *J. Mater. Chem. C* **2020**, *8*, 15920.
- [46] F. Häse, L. M. Roch, P. Friederich, A. Aspuru-Guzik, *Nat. Commun.* **2020**, *11*, 1.
- [47] S. Jiang, C.-C. Wu, F. Li, Y.-Q. Zhang, Z.-H. Zhang, Q.-H. Zhang, Z.-J. Chen, B. Qu, L.-X. Xiao, M.-L. Jiang, *Rare Met.* **2021**, *40*, 1698.
- [48] A. O. Oliynyk, J. M. Buriak, *Chem. Mater.* **2019**, *31*, 8243.
- [49] M.-H. Lee, *Energy Technol.* **2020**, *8*, 1900974.
- [50] J. Im, S. Lee, T.-W. Ko, H. W. Kim, Y. Hyon, H. Chang, *npj Comput. Mater.* **2019**, *5*, 1.
- [51] C. Persson, P. Bacher, T. Shiga, H. Madsen, *Sol. Energy* **2017**, *150*, 423.
- [52] J. Li, B. Pradhan, S. Gaur, J. Thomas, *Adv. Mater.* **2019**, *9*, 1901891.
- [53] P. Nikolaidis, S. Chatzis, *Int. J. Electr. Power Energy Syst.* **2021**, *130*, 106930.
- [54] M. Workman, D. Zhi Chen, S. M. Musa, in *2020 Int. Conf. Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, IEEE, Piscataway, NJ **2020**, pp. 1–5.
- [55] Z. Li, M. A. Najeeb, L. Alves, A. Z. Sherman, V. Shekar, P. Cruz Parrilla, I. M. Pendleton, W. Wang, P. W. Nega, M. Zeller, J. Schrier, *Chem. Mater.* **2020**, *32*, 5650.
- [56] C. She, Q. Huang, C. Chen, Y. Jiang, Z. Fan, J. Gao, *J. Mater. Chem. A* **2021**, *9*, 25168.
- [57] H.-J. Feng, P. Ma, *Appl. Phys. Lett.* **2021**, *119*, 231902.
- [58] E. C. Gok, M. O. Yildirim, M. P. Haris, E. Eren, M. Pegu, N. H. Hemasiri, P. Huang, S. Kazim, A. Uygün Oksuz, S. Ahmad, *Sol. RRL* **2022**, *6*, 2100927.
- [59] X. Cai, F. Liu, A. Yu, J. Qin, M. Hatamvand, I. Ahmed, J. Luo, Y. Zhang, H. Zhang, Y. Zhan, *Light: Sci. Appl.* **2022**, *11*, 1.
- [60] S. Nagasawa, E. Al-Naamani, A. Saeki, *J. Phys. Chem. Lett.* **2018**, *9*, 2639.
- [61] H. Sahu, W. Rao, A. Troisi, H. Ma, *Adv. Mater.* **2018**, *8*, 1801032.
- [62] H. Sahu, H. Ma, *J. Phys. Chem. Lett.* **2019**, *10*, 7277.
- [63] M. R. S. A. Janjua, *J. Phys. Chem. Solids* **2022**, *171*, 110996.
- [64] D. Padula, J. D. Simpson, A. Troisi, *Mater. Horiz.* **2019**, *6*, 343.
- [65] M.-H. Lee, *Org. Electron.* **2020**, *76*, 105465.
- [66] M. R. S. A. Janjua, *Spectrochim. Acta, Part A* **2022**, *279*, 121487.
- [67] X. Du, L. Lüer, T. Heumueller, J. Wagner, C. Berger, T. Osterrieder, J. Wortmann, S. Langner, U. Vongsaysy, M. Bertrand, N. Li, *Joule* **2021**, *5*, 495.
- [68] A. Irfan, M. Hussien, M. Y. Mehboob, A. Ahmad, M. R. S. A. Janjua, *Energy Technol.* **2022**, *10*, 2101096.
- [69] M. R. S. A. Janjua, *Inorg. Chem.* **2021**, *60*, 10837.
- [70] T. Wu, J. Wang, *ACS Appl. Mater. Interfaces* **2020**, *12*, 57821.
- [71] M. R. S. A. Janjua, *Chem. Eur. J.* **2021**, *27*, 4197.
- [72] Z.-W. Zhao, M. del Cueto, Y. Geng, A. Troisi, *Chem. Mater.* **2020**, *32*, 7777.
- [73] Y. Wu, J. Guo, R. Sun, J. Min, *npj Comput. Mater.* **2020**, *6*, 1.
- [74] K. Kranthiraja, A. Saeki, *Adv. Funct. Mater.* **2021**, *31*, 2011168.
- [75] S. Sun, A. Tihihonen, F. Oviedo, Z. Liu, J. Thapa, Y. Zhao, N. T. P. Hartono, A. Goyal, T. Heumueller, C. Batali, A. Encinas, *Matter* **2021**, *4*, 1305.
- [76] T. Wu, J. Wang, *Nano Energy* **2019**, *66*, 104070.
- [77] O. Voznyy, L. Levina, J. Z. Fan, M. Askerka, A. Jain, M.-J. Choi, O. Ouellette, P. Todorović, L. K. Sagar, E. H. Sargent, *ACS Nano* **2019**, *13*, 11122.
- [78] M. L. Agiorgousis, Y.-Y. Sun, D.-H. Choe, D. West, S. Zhang, *Adv. Theor. Simul.* **2019**, *2*, 1800173.
- [79] N. Meftahi, M. Klymenko, A. J. Christofferson, U. Bach, D. A. Winkler, S. P. Russo, *npj Comput. Mater.* **2020**, *6*, 1.
- [80] N. T. P. Hartono, J. Thapa, A. Tihihonen, F. Oviedo, C. Batali, J. J. Yoo, Z. Liu, R. Li, D. F. Marrón, M. G. Bawendi, T. Buonassisi, S. Sun, *Nat. Commun.* **2020**, *11*, 1.
- [81] Z. Wang, H. Zhang, J. Li, *Nano Energy* **2021**, *81*, 105665.
- [82] Y. Liu, W. Yan, H. Zhu, Y. Tu, L. Guan, X. Tan, *Org. Electron.* **2022**, *101*, 106426.
- [83] M. R. S. A. Janjua, A. Irfan, M. Hussien, M. Ali, M. Saqib, M. Sulaman, *Energy Technol.* **2022**, *10*, 2200019.
- [84] H. Sahu, F. Yang, X. Ye, J. Ma, W. Fang, H. Ma, *J. Mater. Chem. A* **2019**, *7*, 17480.
- [85] S. Kanno, Y. Imamura, M. Hada, *Phys. Rev. Mater.* **2019**, *3*, 075403.
- [86] L. Ju, M. Li, L. Tian, P. Xu, W. Lu, *Mater. Today Commun.* **2020**, *25*, 101604.
- [87] W. A. Saidi, W. Shadid, I. E. Castelli, *npj Comput. Mater.* **2020**, *6*, 1.
- [88] Y. Wen, L. Fu, G. Li, J. Ma, H. Ma, *Sol. RRL* **2020**, *4*, 2000110.
- [89] Z. Guo, B. Lin, *Sol. Energy* **2021**, *228*, 689.
- [90] J. Chen, W. Xu, R. Zhang, *J. Mater. Chem. A* **2022**, *10*, 1402.
- [91] J. M. Howard, E. M. Tennyson, B. R. Neves, M. S. Leite, *Joule* **2019**, *3*, 325.
- [92] J. Kirman, A. Johnston, D. A. Kuntz, M. Askerka, Y. Gao, P. Todorović, D. Ma, G. G. Privé, E. H. Sargent, *Matter* **2020**, *2*, 938.
- [93] L. Zhang, M. He, *J. Appl. Phys.* **2022**, *131*, 064902.
- [94] A. Mahmood, A. Irfan, J.-L. Wang, *J. Mater. Chem. A* **2022**, *10*, 4170.
- [95] B. Cao, L. A. Adutwum, A. O. Oliynyk, E. J. Lubner, C. Olsen, A. Mar, J. M. Buriak, *ACS Nano* **2018**, *12*, 7434.
- [96] L. Wei, X. Xu, . Gurudayal, J. Bullock, J. W. Ager, *Chem. Mater.* **2019**, *31*, 7340.
- [97] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., *J. Mach. Learn. Res.* **2011**, *12*, 2825.
- [98] A. Mahmood, A. Irfan, J.-L. Wang, *Chem. Eur. J.* **2022**, *28*, e202103712.
- [99] D. Padula, A. Troisi, *Adv. Mater.* **2019**, *9*, 1902463.
- [100] W. Sun, Y. Zheng, K. Yang, Q. Zhang, A. A. Shah, Z. Wu, Y. Sun, L. Feng, D. Chen, Z. Xiao, S. Lu, *Sci. Adv.* **2019**, *5*, eaay4275.
- [101] T. W. David, H. Anizelli, T. J. Jacobsson, C. Gray, W. Teahan, J. Kettle, *Nano Energy* **2020**, *78*, 105342.
- [102] T. W. David, H. Anizelli, P. Tyagi, C. Gray, W. Tea-Han, J. Kettle, *IEEE J. Photovoltaics* **2019**, *9*, 1768.
- [103] A. Mahmood, J.-L. Wang, *J. Mater. Chem. A* **2021**, *9*, 15684.
- [104] G. Salman, S. M. Goodnick, A. R. Shaik, D. Vasileksa, in *2021 IEEE 48th Photovoltaic Specialists Conf. (PVSC)*, IEEE, Piscataway, NJ **2021**, pp. 0540–0543.
- [105] M. Gafar, R. A. El-Sehiemy, H. M. Hasanien, A. Abaza, *J. Ambient Intell. Hum. Comput.* **2022**, *1*.
- [106] J. W. Hsu, W. Xu, *Matter* **2022**, *5*, 1334.

- [107] M. A. Gelbart, J. Snoek, R. P. Adams, *arXiv preprint arXiv:1403.5607* **2014**.
- [108] Z. Liu, N. Rolston, A. C. Flick, T. W. Colburn, Z. Ren, R. H. Dauskardt, T. Buonassisi, *Joule* **2022**, 6, 834.
- [109] V. M. Le Corre, T. S. Sherkar, M. Koopmans, L. J. A. Koster, *Cell Rep. Phys. Sci.* **2021**, 2, 100346.
- [110] X. Rodriguez-Martinez, E. Pascual-San-José, M. Campoy-Quiles, *Energy Environ. Sci.* **2021**, 14, 3301.
- [111] A. Kirkey, E. J. Lubner, B. Cao, B. C. Olsen, J. M. Buriak, *ACS Appl. Mater. Interfaces* **2020**, 12, 54596.
- [112] Ç. Odabaşı, R. Yıldırım, *Sol. Energy Mater. Sol. Cells* **2020**, 205, 110284.
- [113] N. Majeed, M. Saladina, M. Krompiec, S. Greedy, C. Deibel, R. C. MacKenzie, *Adv. Funct. Mater.* **2020**, 30, 1907259.
- [114] B. Olsthoorn, R. M. Geilhufer, S. S. Borysov, A. V. Balatsky, *Adv. Quantum Technol.* **2019**, 2, 1900023.
- [115] F. Li, X. Peng, Z. Wang, Y. Zhou, Y. Wu, M. Jiang, M. Xu, *Energy Environ. Mater.* **2019**, 2, 280.
- [116] S.-G. Kim, J.-Y. Jung, M. K. Sim, *Sustainability* **2019**, 11, 1501.
- [117] A. J. Trappey, P. P. Chen, C. V. Trappey, L. Ma, *Appl. Sci.* **2019**, 9, 1478.
- [118] B. Sanchez-Lengeling, A. Aspuru-Guzik, *Science* **2018**, 361, 360.
- [119] C. Lee, S. Lee, G.-U. Kim, W. Lee, B. J. Kim, *Chem. Rev.* **2019**, 119, 8028.
- [120] M.-G. Ju, M. Chen, Y. Zhou, J. Dai, L. Ma, N. P. Padture, X. C. Zeng, *Joule* **2018**, 2, 1231.
- [121] E. G. Dada, J. S. Bassi, H. Chiroma, A. O. Adetunmbi, O. E. Ajibuwa, *Heliyon* **2019**, 5, e01802.
- [122] T. Huuhtanen, A. Jung, in *2018 IEEE Data Science Workshop (DSW)*, IEEE, Piscataway, NJ **2018**, pp. 66–70.
- [123] C. Zhan, J. Yao, *Chem. Mater.* **2016**, 28, 1948.
- [124] X. Zhang, S. Li, T. He, B. Yang, T. Yu, H. Li, L. Jiang, L. Sun, *Energy* **2019**, 174, 1079.
- [125] N. Gasparini, A. Salleo, I. McCulloch, D. Baran, *Nat. Rev. Mater.* **2019**, 4, 229.
- [126] L. Faes, S. K. Wagner, D. J. Fu, X. Liu, E. Korot, J. R. Ledsam, T. Back, R. Chopra, N. Pontikos, C. Kern, G. Moraes, *Lancet Digital Health* **2019**, 1, e232.
- [127] Y. Rong, Y. Hu, A. Mei, H. Tan, M. I. Saidaminov, S. I. Seok, M. D. McGehee, E. H. Sargent, H. Han, *Science* **2018**, 361, eaat8235.
- [128] G. Wang, F. S. Melkonyan, A. Facchetti, T. J. Marks, *Angew. Chem., Int. Ed.* **2019**, 58, 4129.
- [129] J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R. S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A. M. Brockway, A. Aspuru-Guzik, *J. Phys. Chem. Lett.* **2011**, 2, 2241.
- [130] E. O. Pyzer-Knapp, J. W. Pitera, P. W. Staar, S. Takeda, T. Laino, D. P. Sanders, J. Sexton, J. R. Smith, A. Curioni, *npj Comput. Mater.* **2022**, 8, 1.
- [131] W. Chen, Y. Wu, Y. Yue, J. Liu, W. Zhang, X. Yang, H. Chen, E. Bi, I. Ashraful, M. Grätzel, L. Han, *Science* **2015**, 350, 944.
- [132] A. Mahmood, J.-L. Wang, *Energy Environ. Sci.* **2021**, 14, 90.
- [133] J. Peng, E. C. Jury, P. Dönnies, C. Ciurtin, *Front. Pharmacol.* **2021**, 12, 2667.
- [134] A. Hashemi, M. Basafa, A. Behravan, *Sci. Rep.* **2022**, 12, 1.
- [135] L. Zhang, M. He, *J. Phys.: Condens. Matter* **2021**, 34, 095902.
- [136] I. Mathews, S. N. R. Kantareddy, S. Sun, M. Layurova, J. Thapa, J.-P. Correa-Baena, R. Bhattacharyya, T. Buonassisi, S. Sarma, I. M. Peters, *Adv. Funct. Mater.* **2019**, 29, 1904072.
- [137] J. J. Yoo, S. Wiegold, M. C. Sponseller, M. R. Chua, S. N. Bertram, N. T. P. Hartono, J. S. Tresback, E. C. Hansen, J.-P. Correa-Baena, V. Bulović, T. Buonassisi, S. Sik Shin, M. G. Bawendi, *Energy Environ. Sci.* **2019**, 12, 2192.
- [138] I. Mathews, S. N. Kantareddy, T. Buonassisi, I. M. Peters, *Joule* **2019**, 3, 1415.
- [139] S.-H. Chan, Y.-H. Chang, M.-C. Wu, *Front. Mater.* **2019**, 6, 57.
- [140] T. Wu, Z. Qin, Y. Wang, Y. Wu, W. Chen, S. Zhang, M. Cai, S. Dai, J. Zhang, J. Liu, Z. Zhou, *Nano-Micro Lett.* **2021**, 13, 1.
- [141] B. Li, Y. Xiang, K. I. Jayawardena, D. Luo, Z. Wang, X. Yang, J. F. Watts, S. Hinder, M. T. Sajjad, T. Webb, H. Luo, I. Marko, H. Li, S. A. J. Thomson, R. Zhu, G. Shao, S. J. Sweeney, S. Ravi, P. Silva, W. Zhang, *Nano Energy* **2020**, 78, 105249.
- [142] Z.-H. Zhou, *Machine Learning*, Springer Nature, Singapore **2021**.
- [143] H.-W. Ng, S. Winkler, in *2014 IEEE Int. Conf. Image Processing (ICIP)*, IEEE, Piscataway, NJ **2014**, pp. 343–347.
- [144] D. Venkataraman, S. Yurt, B. H. Venkatraman, N. Gavvalapalli, *J. Phys. Chem. Lett.* **2010**, 1, 947.
- [145] S. Fonash, *Solar Cell Device Physics*, Academic Press, Inc, New York **2012**.
- [146] E. K. Chong, S. H. Zak, *An Introduction to Optimization*, Vol. 75, John Wiley & Sons, Inc., Hoboken, NJ **2013**.
- [147] A. Schulz, A. Shamir, D. I. Levin, P. Sitthi-Amorn, W. Matusik, *ACM Trans. Graphics* **2014**, 33, 1.
- [148] R. M. I. Bandara, S. M. Silva, C. C. L. Underwood, K. D. G. I. Jayawardena, R. A. Sporea, S. R. P. Silva, *Energy Environ. Mater.* **2022**, 5, 370–400.

Satyam Bhatti received a B.Tech. in electrical engineering from Punjab Engineering College, India, and a master's degree (MSc) in electrical engineering for sustainable and renewable energy from the University of Nottingham, UK, in 2019 and was awarded Developing Solutions Masters Scholarship from the University of Nottingham. He completed an M.Phil. in electronics and electrical engineering from the University of Glasgow, UK, and is working as a research assistant. His research interests are in the fields of renewable energy, solar cells, machine learning, wearables, and nanoscale engineering.

Sajjad Hussain received his master's degree in wireless communications from Supelec, Gif-sur-Yvette, in 2006, and Ph.D. in signal processing and communications from the University of Rennes 1, Rennes, France, in 2009. He is a reader (associate professor) in electronics and electrical engineering at the University of Glasgow, U.K. He served previously with the Electrical Engineering Department, Capital University of Science and Technology, Islamabad, Pakistan, as an associate professor. He has over 10 years of academic experience with roles as the principal/coinvestigator in several funded projects. He has authored/coauthored over 60 journal and conference publications along with one authored book and one edited book. His research interests include 5G self-organizing networks, industrial wireless sensor networks, machine learning for wireless communications, and technology-enhanced student learning techniques. He has been an associate editor for IEEE ACCESS, since 2019. He is a fellow of the Higher Education Academy and a senior fellow of Recognising Excellence in Teaching.

Rami Ghannam received the B.Eng. in electronic engineering from King's College London, the D.I.C. and M.Sc. from Imperial College London, and the Ph.D. from the University of Cambridge in 2007. He is currently an associate professor of electronic and nanoscale engineering at the University of Glasgow. He held previous industrial positions at Nortel Networks and IBM Research GmbH. His research interests include energy harvesters, interactive devices, and engineering education. He has published over 100 articles in these areas, including five patents and one edited book. He is also a senior fellow of Glasgow's RET scheme and serves as Scotland's Regional Chair for the IEEE Education Society. He also serves as the chair for the IET Pedagogy in Engineering and Technology TN. He has been awarded the Siemens Prize.