processes

MDPI

Review

A Review of Machine Learning in Organic Solar Cells

Darya Rasul Ahmed and **Fahmi F. Muhammadsharif** *

Department of Physics, Faculty of Science and Health, Koya University, Koya KOY45, Kurdistan Region-F.R., Iraq; darya.ahmed@koyauniversity.org

* Correspondence: fahmi.fariq@koyauniversity.org or fahmi982@gmail.com

Abstract: Organic solar cells (OSCs) are a promising renewable energy technology due to their flexibility, lightweight nature, and cost-effectiveness. However, challenges such as inconsistent efficiency and low stability limit their widespread application. Addressing these issues requires extensive experimentation to optimize device performance, a process hindered by the complexity of OSC molecular structures and device architectures. Machine learning (ML) offers a solution by accelerating material discovery and optimizing performance through the analysis of large datasets and prediction of outcomes. This review explores the application of ML in advancing OSC technologies, focusing on predicting critical parameters such as power conversion efficiency (PCE), energy levels, and absorption spectra. It emphasizes the importance of supervised, unsupervised, and reinforcement learning techniques in analyzing molecular descriptors, processing data, and streamlining experimental workflows. Concludingly, integrating ML with quantum chemical simulations, alongside high-quality datasets and effective feature engineering, enables accurate predictions that expedite the discovery of efficient and stable OSC materials. By synthesizing advancements in ML-driven OSC research, the gap between theoretical potential and practical implementation can be bridged. ML can viably accelerate the transition of OSCs from laboratory research to commercial adoption, contributing to the global shift toward sustainable energy solutions.

Keywords: machine learning; organic solar cell; feature selection; photovoltaic parameter; classification algorithm

check for updates

Academic Editor: Xiong Luo

Received: 3 December 2024
Revised: 10 January 2025
Accepted: 21 January 2025
Published: 1 February 2025

Citation: Ahmed, D.R.; Muhammadsharif, F.F. A Review of Machine Learning in Organic Solar Cells. *Processes* **2025**, *13*, 393. https://doi.org/10.3390/pr13020393

1. Introduction

This section briefly explores the exciting advancements and ongoing challenges in the development of organic solar cells (OSCs), focusing on how machine learning (ML) is revolutionizing the OSC field. It elaborates on how traditional trial-and-error approaches are slow and resource-intensive, limiting progress in OSCs. That is where ML comes in, offering powerful tools to speed up material discovery and device optimization. It emphasizes the current state of OSC research, the hurdles in adopting ML, and practical strategies to bridge the gap between material science and artificial intelligence (AI), paving the way for more efficient, stable, and scalable solar technologies.

A promising avenue for the development of OSCs can be realized via the use of organic semiconductors. This is due to the unique properties of these materials such as high synthetic flexibility, which permits remarkable control over the bandgap, energy level, and carrier mobility of the active layer of OSC devices [1,2]. The active layers are commonly made of electron donor and acceptor materials. It is worth noting that recent advances in the synthesis of non-fullerene acceptors have led to significant improvements in power conversion efficiencies (PCEs) [3]. Some OSC devices have achieved PCEs exceeding 19%.

The development of OSCs offers significant potential in renewable energy technologies due to their lightweight, flexible, and semi-transparent nature. However, traditional trial-and-error methods for discovering and optimizing OSC materials are inefficient and time-consuming. These approaches are hindered by lengthy processes, complex donor/acceptor interface morphologies, and strong electron–phonon couplings, making it challenging to predict key performance metrics like PCE. Despite recent advancements in theoretical insights and experimental techniques, the expansive space of organic compounds makes material discovery labor-intensive [4] compared to the perovskite solar cells [5–7].

ML is a discipline within artificial intelligence. It provides a promising alternative to accelerate OSC research. By leveraging data-driven approaches, ML can streamline the discovery and optimization of materials, significantly reducing the reliance on serendipity and extensive experimental testing [8]. Despite these advancements, substantial improvement in PCE and stability is still necessary for OSCs to be competitive with inorganic devices and to reach market use. The current methodologies are hindered by lengthy time consumption, tedious purification stages, and strict synthesis methods that plague the trial-and-error experimental routines [2]. Additionally, predicting the PCE of OSC material components is challenging due to several obstacles such as the complex donor/acceptor (D/A) interface morphology, strong electron–phonon couplings, and strong electron–electron interactions. These features necessitate state-of-the-art theoretical approaches from quantum chemistry, statistical mechanics, and quantum dynamics for precise OSC simulations. Light-induced processes such as exciton production, migration, and dissociation, along with charge transfer to the appropriate electrodes, further complicate the development of a predictive rule [9].

Scharber's model was introduced [10] to forecast photovoltaic efficiency using the limited electronic properties of donor and acceptor materials. Despite its utility, Scharber's model struggles to incorporate critical factors like morphology and excited-state dynamics, limiting its applicability to modern OSCs. Numerous challenges impede its extension, particularly the difficulty in incorporating additional descriptors such as structural, topological, and thermodynamic factors [11].

The data-driven paradigm for material discovery is efficient and effective in leveraging pertinent information [12]. The methodical approach to this is ML, which derives insights from historical data to assist in evaluating candidates for laboratory positions. The identification of superior candidate materials for OSCs can be expedited and rendered more cost-effectively through multidimensional designs utilizing ML, density functional theory (DFT) calculations, and the available experimental data survey, as can be seen in Figure 1.

The application of ML to the complex domain of OSCs has not yet yielded particularly impressive results. The performance of OSCs is influenced by numerous factors such as solvent additives, crystallinity, molecule orientation, and processing solvents [13]. Morphological features are crucial for charge separation at the donor/acceptor interface. More research is required to make effective use of ML with photovoltaic materials.

Material degradation in organic materials, particularly non-fullerene acceptors (NFAs) and polymer donors, exhibits chemical instability when exposed to air or light. This affects the molecular structure and leads to reduced PCE over time. Thermal and mechanical instability is the flexible nature of OSCs making them susceptible to mechanical deformation and thermal stress, which can disrupt the active layer morphology and electrode interfaces. Recent strategies to address these issues include the molecular engineering of more robust materials, optimizing device architectures to enhance encapsulation, and developing predictive models for degradation pathways. These approaches aim to achieve stable, high-performing OSCs suitable for large-scale applications in building integration and wearable electronics [14,15].

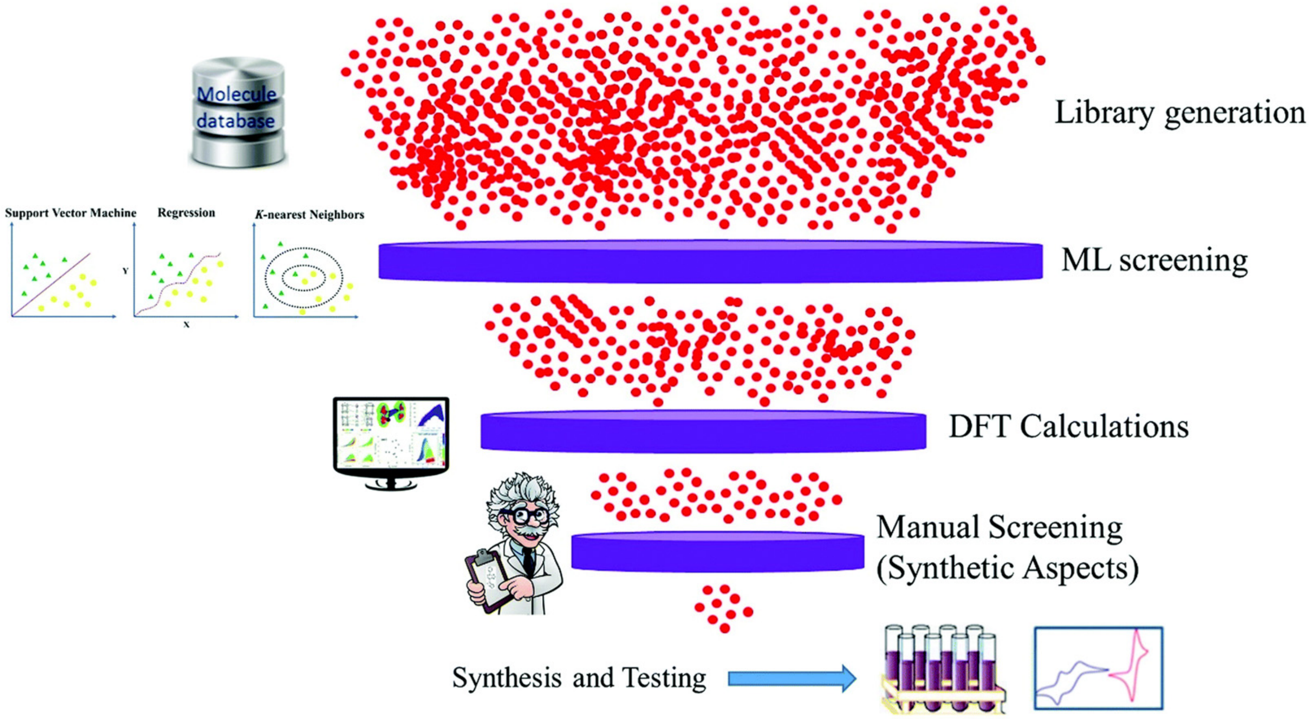


Figure 1. Computer-assisted design and screening of materials for OSCs. Reproduced with permission from [8].

ML can be utilized to analyze molecular descriptors, predict material properties, and optimize device architectures in OSCs. The motivation is to demonstrate how ML can accelerate the discovery of high-performance materials and optimize OSC designs, leading to more efficient and cost-effective solar energy solutions.

Despite growing interest in integrating ML into OSC research, progress has been limited by small datasets, inconsistent data quality, and challenges in adoption in workflow. The complexity of organic materials and the lack of generalized, explainable ML frameworks further hinder advancements. This review addresses the limitations by presenting recent developments and suggesting strategies to overcome these challenges. Additionally, it outlines the potential for ML to enhance the stability and performance of OSC devices, paving the way for their large-scale application. Also, it provides actionable insights for researchers in both ML and material sciences, bridging the gap between these disciplines. Key contributions include highlighting the need for robust ML models tailored to OSC research, discussing the importance of high-quality data and descriptor selection, and identifying how ML can significantly enhance predictive accuracy and accelerate material discovery.

2. The Use of ML in OSCs

ML is a branch of artificial intelligence that focuses on building models capable of predicting outcomes and discovering new materials based on large quantities of reliable data. These datasets, derived from experiments or computations, describe the materials' behaviors, qualities, and applications. By analyzing these data, ML techniques can uncover patterns and relationships that might not be evident through traditional methods [16].

In the context of OSCs, ML involves applying statistical models and computational methods to analyze, predict, and optimize device performance. This process includes conducting experiments with OSCs, running simulations, and employing theoretical models, followed by the application of ML algorithms to interpret the complex relationships within the datasets. The integration of ML, high-performance computing (HPC), and adequate data now presents an opportunity to streamline materials discovery [8].

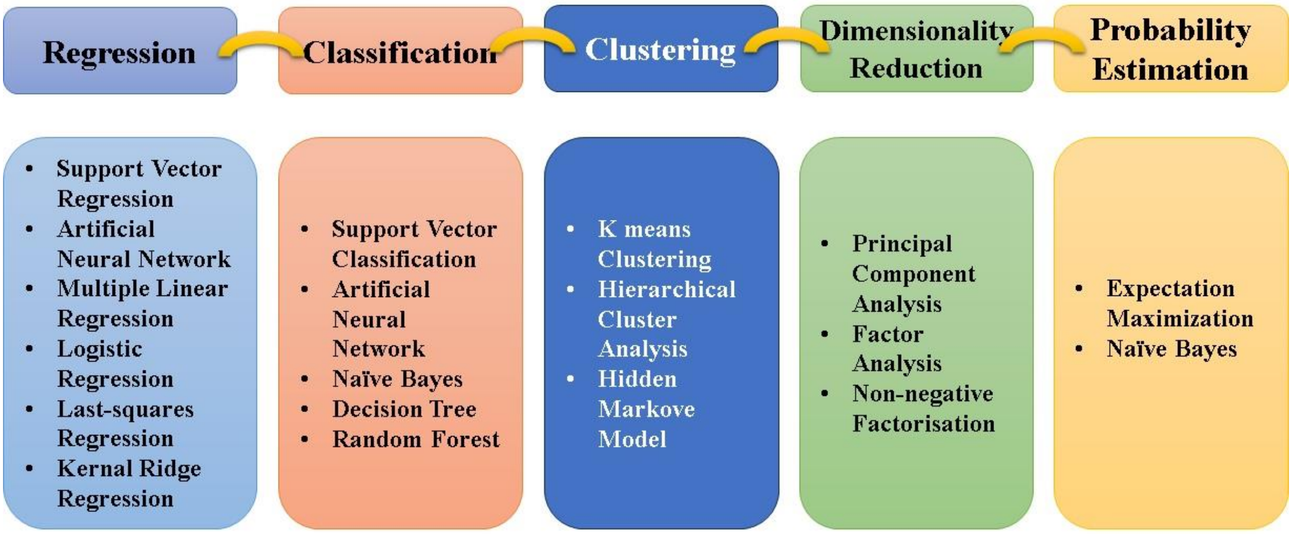
The remarkable achievements of ML in fields such as image identification and translation have aroused the interest of materials scientists [17]. These advancements demonstrate the potential of ML to gain insights into the fundamental principles governing material behavior, offering significant time savings compared to traditional quantum chemistry computations and experimental methods.

For instance, in a case study involving the design and testing of materials for OSCs, molecular dynamics simulations combined with ML techniques led to the identification of promising new material candidates. The process began with collecting data on various molecular configurations and their properties. ML models were then trained to predict the performance of these configurations, significantly narrowing down the list of potential candidates for experimental testing. This approach not only saved time but also reduced the cost associated with trial-and-error experimentation [17].

However, the success of ML in this domain is heavily dependent on the quality, size, and form of the dataset. OSCs, like many material science domains, have scattered and heterogeneous data due to the complexity of their working principles [8]. Effective data-collection strategies, preprocessing techniques, and feature selection are crucial to developing robust ML models that can make accurate predictions.

For example, recent research has employed ML algorithms to predict the PCE of OSCs based on molecular descriptors and device architecture parameters. By analyzing vast datasets, these models identified key factors influencing PCE and provided insights into optimizing material properties and device configurations for enhanced performance.

By leveraging ML techniques, researchers can accelerate the discovery of high-performance materials and optimize OSC designs, paving the way for more efficient and cost-effective solar energy solutions [8]. Choosing the right ML algorithm is crucial since it has a major impact on how well the predictions turn out. There are several ML methods, as shown in Figure 2, so it is not possible for one algorithm to always give the best prediction in every situation. Choosing the right algorithm, which is often carried out by trial and error, is vital to producing a highly successful model. Along this line, different ML algorithms that can be found in the literature were utilized in chemistry and material science applications [18–20].


Regression
- Support Vector Regression
- Artificial Neural Network
- Multiple Linear Regression
- Logistic Regression
- Last-squares Regression
- Kernal Ridge Regression

Classification
- Support Vector Classification
- Artificial Neural Network
- Naïve Bayes
- Decision Tree
- Random Forest

Clustering
- K means Clustering
- Hierarchical Cluster Analysis
- Hidden Markove Model

Dimensionality Reduction
- Principal Component Analysis
- Factor Analysis
- Non-negative Factorisation

Probability Estimation
- Expectation Maximization
- Naïve Bayes

Figure 2. Different types of ML algorithms.

The integration of ML with HPC offers a streamlined approach to materials discovery, substantially diminishing the time and expenses linked to conventional trial-and-error experimentation.

3. Steps of ML Applications

ML analysis involves four fundamental steps: sample collection (Section 3.1), data processing (Section 3.2), training of the ML algorithm (Section 3.3), and then testing (Section 3.4). Each step presents unique challenges and requires tailored strategies to ensure effective implementation. Data preprocessing focuses on cleaning, organizing, and preparing raw data for analysis. Model selection and training involve choosing the most suitable algorithm and optimizing it using the training data. Model evaluation assesses the model's performance using metrics and validation techniques. Finally, deployment integrates the trained model into real-world applications to generate actionable insights or predictions [21].

3.1. Sample Collection

The initial stage involves gathering data, which can originate from theoretical models and hands-on experiments. Data cleaning or modification may be necessary to eliminate inconsistencies and noise. Data splitting for training and testing sets can significantly impact model performance. Common ratios include 60:40, 70:30, 80:20, and even 90:10. The simplest approach is to use non-overlapping datasets while maintaining record order, such as using 70% for training and 30% for testing. However, this may lead to issues if the response is not uniformly distributed. Random sampling can ensure that answer values span the whole spectrum from lowest to greatest, reducing the risk of bias [22].

One major challenge emerged while dealing with small datasets, particularly in material sciences where obtaining high-quality data can be difficult. A good rule of thumb is to have a minimum of 50 data points for a decent ML model, but some models, like neural networks, require much larger datasets. For instance, big datasets were successfully utilized for ML applications in health informatics and accelerated materials discovery using deep learning and neural network algorithms [22,23]. The dataset size may vary depending on the complexity of the model (e.g., neural networks typically require more data than simpler models like linear regression) and the problem domain.

Addressing this challenge may involve augmenting data through simulations or utilizing data from scholarly journals and databases [24,25].

3.2. Data Processing

Fresh data can reveal previously unseen patterns of a ML model, but this requires thorough data cleaning to handle missing data and outliers, enhancing model accuracy. Normalizing the scales of various descriptors is crucial for consistent analysis and effective utilization within a single method. Dimensionality reduction techniques, such as principal component analysis (PCA), discriminant analysis (LDA), and independent component analysis (ICA), are essential when dealing with more features (descriptors) than observations or when characteristics have strong correlations. These techniques reduce the feature space size, helping to identify the most important characteristics and improve visualization [8].

A common pitfall is overlooking the importance of feature engineering, which can significantly impact the model's performance. Techniques like feature selection and the creation of new features based on domain knowledge can enhance the model's predictive power. The rise in deep learning reduces reliance on manual feature engineering [22,26]. In OSCs, inadequate preprocessing of molecular descriptors might lead to poorly performing predictive models.

3.3. Model Training

In the context of OSCs, the relationship between performance and parameters is complex. The choice of algorithm significantly impacts the model's accuracy and generalizability. Each algorithm has unique benefits and drawbacks. Common ML algorithms in material sciences include classification, clustering, regression, and probability estimation. Classification and regression are typically used to predict material properties, while clustering helps group similar materials, and probability estimation aids in discovering new materials.

Choosing the right algorithm involves understanding the nature of the data and the specific problem. For example, regression models might be preferred for continuous data predictions, while classification models are suitable for categorical outcomes. It is crucial to experiment with multiple algorithms and hyperparameters to identify the best-performing model [8,27].

Several ML models have been successfully applied to OSC research, which include:

1. Support Vector Machines (SVMs): SVMs have been used for classifying OSC materials into high- and low-performing categories. For example, they can predict whether a new donor/acceptor pair will result in a device with high PCE by analyzing molecular descriptors.
2. Random Forests (RFs): RF models are commonly applied to regression tasks, such as predicting the PCE of OSCs based on input features like bandgap, chemical composition, and solvent properties. Their ability to handle high-dimensional data and provide feature importance rankings makes them valuable for identifying key parameters.
3. Neural Networks (NNs): Deep learning approaches, including feedforward neural networks, have been applied to predict OSC performance metrics. These models can capture non-linear relationships in large datasets but require careful tuning to avoid overfitting.
4. Gaussian Process Regression (GPR): GPR models are useful for predicting OSC properties when data are scarce. They provide uncertainty estimates, making them ideal for guiding experimental design and reducing the number of necessary experiments.
5. k-Means Clustering: This unsupervised learning technique groups materials with similar characteristics, which can aid in identifying novel donor/acceptor combinations or processing conditions.
6. Autoencoders: Autoencoders have been used to extract meaningful latent representations of OSC materials, enabling data-driven exploration of the chemical and morphological design space [8,28].

Selecting the most suitable ML algorithm for OSC research requires a thorough understanding of the dataset's characteristics and the problem's objectives. For instance, regression models might be preferred for predicting continuous variables like PCE, while classification models are more suitable for categorical outcomes, such as device stability (e.g., stable vs. unstable).

In a recent study [28], models of random forests and gradient boosting were employed to predict PCE and stability in OSCs. The study highlighted the importance of feature engineering and hyperparameter tuning to enhance the model performance, illustrating the iterative nature of model building.

By leveraging these ML models, researchers can accelerate the discovery and optimization of OSCs, paving the way for more efficient and cost-effective devices. Experimenting with multiple algorithms, tuning hyperparameters, and employing cross-validation are critical steps in achieving optimal performance. Additionally, incorporating domain knowledge, such as using chemically informed features (e.g., molecular fingerprints or descriptors), enhances model interpretability and effectiveness.

ML models have shown great promise in accelerating OSC research by reducing trial-and-error experiments and uncovering hidden patterns in complex datasets. Leveraging these techniques can significantly contribute to the discovery and development of high-performance materials and devices [8].

3.4. Model Testing

An effective model has strong performance in both training and testing datasets. Statistical analysis techniques, including mean squared error (MSE)—Equation (1); root mean squared error (RMSE)—Equation (2); and coefficient of determination (R^2)—Equation (3), are employed to assess model efficacy [8].

$$MSE = \sum_{i=1}^{m} \frac{1}{m}(x_i - y_i)^2 \tag{1}$$

where x_i is the predicted value and y_i is the target variable. This measures the average squared difference between the x_i and y_i values. A smaller MSE indicates that the predictions are closer to the true values, making it a key metric for evaluating model accuracy.

$$RMSE = \sqrt{MSE} \tag{2}$$

RMSE is simply the square root of the MSE. It provides an error measurement in the same unit as the original data, making it easier to interpret the magnitude of the prediction errors.

$$R^2 = 1 - \frac{MSE}{Var(y)} \tag{3}$$

where $Var(y)$ is the variance of the sample data. R^2 quantifies how well the model explains the variance of the actual data (y). It ranges from 0 to 1, where a value closer to 1 means the model captures most of the variability in the data, indicating a good fit.

One challenge is ensuring the model's generalizability to new, unseen data. Techniques such as cross-validation and bootstrapping can help assess model stability and robustness, ensuring that the model performs well beyond the initial test set. By addressing these practical challenges and implementing detailed strategies at each step, ML can be effectively utilized to advance research in OSCs, leading to more efficient and innovative solutions.

4. Types of ML Algorithm

In this section, the most important ML algorithms in the field of organic materials and OSCs are discussed. Also, their implementation strategies along with related parametric settings for each of the algorithms are elaborated.

The design and synthesis of materials with beneficial, innovative properties is a highly dynamic field in modern science, fostering considerable research in biomaterials, cell and tissue engineering, OSCs, light-emitting materials, and nanomaterials for various medical and non-medical applications. These advancements involve interdisciplinary efforts from fields including engineering, biology, physics, and chemistry. Although theoretical and computational science is making some headway, experimental science remains the primary focus. Material designers would greatly benefit from understanding how to anticipate the characteristics of new materials before synthesis and how the macroscopic features of materials relate to the microscopic properties of molecular components [29].

The development and optimization of materials with advanced and innovative properties represent a dynamic and rapidly evolving area in modern science. This field encompasses a broad range of applications, including biomaterials, cell and tissue engineering, OSCs, light-emitting materials, and nanomaterials for both medical and non-medical purposes. These advancements are driven by interdisciplinary collaborations spanning engineering, biology, physics, and chemistry. While theoretical and computational sciences have made notable progress in predicting material properties, experimental science remains at the forefront, often leading the discovery process. A key challenge lies in bridging the gap between theoretical predictions and experimental validation. To address this, material scientists seek to anticipate the characteristics of new materials prior to synthesis and establish clear connections between macroscopic material properties and the molecular-level structures that define them. This integration of predictive and experimental approaches holds great promise for accelerating innovation in material design and functionality.

ML offers powerful tools to achieve these goals, especially in material sciences, where the relationships between structure, properties, and performance are often complex and non-linear.

1. Supervised Learning: Supervised learning algorithms are trained on labeled data, meaning each training example is paired with an output label. These algorithms learn to map inputs to outputs, which is critical for predicting the properties of new materials [30,31]. They are mostly used to categorize data into predefined classes. For example, in OSCs, classification algorithms can predict whether a new material will act as a donor or acceptor based on its molecular structure [32]. In supervised learning, accuracy is a metric that measures how well a model correctly predicts the target variable, calculated as the ratio of correct predictions to the total number of predictions. It is widely used in classification tasks to evaluate performance but can be misleading for imbalanced datasets, where one class dominates; in such cases, metrics like precision (the proportion of correct positive predictions), recall (the ability to identify all actual positives), F1-score (the harmonic mean of precision and recall), or ROC-AUC (the area under the curve representing true positive versus false positive rates) are more informative. Below are the key algorithms of supervised learning.
    - Support Vector Machines (SVM): SVMs are effective in classifying materials based on their electronic properties. For instance, they can help determine which molecular structures are likely to result in high-efficiency donor or acceptor materials for OSCs [4].
    - Decision Trees and Random Forests: These algorithms identify critical structural features that determine material performance. They can be used to analyze various molecular descriptors and pinpoint which attributes are most influential in achieving high PCE [33].
    - Linear Regression: Linear regression is often used to model the relationship between molecular descriptors and PCE. For example, linear regression can help establish how changes in molecular structure affect the efficiency of OSCs [34].
    - Neural Networks: Neural networks can capture more complex, non-linear relationships between structure and efficiency. They are particularly useful in modeling the intricate dependencies between various molecular features and the overall performance of OSCs [35].
2. Unsupervised Learning: Unsupervised learning algorithms deal with data without labeled responses. They are useful for discovering hidden patterns or intrinsic structures in the data [36]. Below are the key algorithms of unsupervised learning.

- Clustering Algorithms: Clustering algorithms, such as k-Means, can group materials with similar properties, aiding in the identification of promising material families. For instance, clustering can reveal which sets of molecular structures consistently yield high-efficiency OSCs [37].
- Dimensionality Reduction Techniques: Techniques like PCA reduce the complexity of data while retaining essential patterns, which is crucial when dealing with high-dimensional datasets in material sciences. PCA can help identify the most influential factors in determining OSC performance, streamlining the design process [11].

3. Semi-supervised Learning: Semi-supervised learning strikes a balance between supervised and unsupervised methods, making it especially useful when labeled data are hard to come by but unlabeled data are abundant. Imagine having a small set of data points with labels and a much larger set without them. Semi-supervised learning uses the labeled data to guide the learning process and make sense of the unlabeled data. Techniques like self-training allow a model to start learning with labeled data, then predict labels for the unlabeled data and improve itself iteratively. Graph-based approaches also come into play, where relationships between data points are mapped to spread labels from known points to unknown ones [38].
4. Reinforcement Learning: Reinforcement learning involves training models through trial and error, using feedback from their actions. This approach can optimize material synthesis processes or experimental procedures to maximize efficiency or yield. For instance, Q-Learning and Deep Q-Networks (DQN) can optimize the sequence of synthesis steps to produce materials with desired properties efficiently, thereby refining the fabrication process of OSCs to enhance their stability and efficiency [11].

5. ML Analysis of OSCs

In this section, light is shed on ML contributions in the development of OSCs by helping researchers discover better materials, optimize device performance, and understand stability issues. Also, elaborations are given on how ML can predict organic molecules with the right properties for high efficiency, thereby suggesting new designs of materials. ML accelerates progress and makes it easier to bridge the gap between lab research and real-world applications, ultimately pushing us closer to more affordable and sustainable solar energy solutions.

The application of ML analysis significantly improves the effective screening of potential candidates for OSCs. Understanding the relationship between molecular attributes and PCE is crucial. It is essential to examine the relationship between specific device performance metrics and molecular characteristics to meet the requirements of diverse applications, such as high open-circuit voltage of solar cells for energy conversion, elevated short-circuit current V_{OC}, and solar window applications, as well as J_{SC}.

Successfully screening potential candidates for OSCs requires a thorough understanding of the relationships between molecular properties and PCEs. Equally important is the study of how molecular properties correlate with specific device performance parameters to meet the demands of particular applications. For instance, achieving high V_{OC} is critical for solar-to-fuel energy conversion, while high J_{SC} is essential for solar window applications. Understanding these correlations enables targeted material optimization tailored to specific functional requirements.

Because ML can forecast performance based on molecular parameters, it has a broad use in the field of OSC research. However, the kind of descriptors that are employed has a significant impact on how accurate an ML model's predictions can be. Descriptors play a crucial role in producing accurate predictions by acting as a translator between researchers and the database. When the goal property is not well defined, choosing candidate descriptions becomes a substantial task. In general, some aspects affect a material's properties, so choosing appropriate descriptors for a certain property is an important step before using ML. This is particularly true for microscopic descriptors, whose determination can be costly both computationally and empirically [39,40].

An effective material description must satisfy a minimum of three criteria: (i) it should provide a unique characterization of the material, (ii) it should be sensitive to the target property, and (iii) it should be easy to calculate [8]. When the target property is ambiguous, meeting these criteria becomes challenging, leading to potential setbacks in developing accurate and trustworthy ML models for OSCs. This highlights the need for a clear definition of target properties to ensure the selection of relevant and effective descriptors, ultimately improving the success of ML-driven screening processes in OSC research.

The following sections cover how ML is revolutionizing the development of OSCs. It starts by discussing how hybrid modeling combines different techniques to link molecular properties to device performance for better optimization. It then explores how ML and computational models help predict and improve OSC efficiency. The content also highlights the role of ML in discovering and designing new materials, optimizing production processes, and improving yield and device durability. Finally, it looks at how data analysis tools like pattern recognition and predictive modeling enhance our understanding of OSC performance, showcasing ML's essential role in speeding up innovation in solar technology.

1. Hybrid and Multiscale Modeling: These approaches integrate different modeling techniques to provide a comprehensive understanding of material behavior across various scales [11].
    - Atomistic or Molecular-Level Models: These models focus on the interactions at the molecular level, which are crucial for understanding the fundamental properties of materials. For instance, molecular dynamics simulations can reveal how molecular vibrations and rotations affect the electronic properties of OSCs [41].
    - Continuum or Device-Level Models: These models help in understanding how molecular-level properties translate to macroscopic device performance. For example, continuum models can simulate the charge transport properties in OSCs, providing insights into how molecular arrangements affect overall efficiency.

Combining these models helps in linking the detailed molecular structure with the overall performance of OSCs, leading to better optimization strategies. For instance, hybrid modeling can combine molecular dynamics simulations with device-level models to predict how changes at the molecular scale impact device performance.

2. Performance Prediction and Optimization: Performance prediction and optimization involve using computational models, statistical methods, or ML techniques to forecast and improve the performance of a system, device, or process.
    - Performance Prediction: In the context of OSCs, performance prediction involves using models or algorithms to estimate and forecast the characteristics and efficiency of the solar cell based on various factors. This prediction may encompass the expected PCE, short-circuit current density (J_{sc}), open-circuit voltage (V_{oc}), fill factor (*FF*), or other key metrics that quantify the effectiveness of the solar cell in converting sunlight into electricity. For example, ML models can predict how different material compositions and device architectures will perform under specific operating conditions [42].
    - Optimization Strategies: Optimization involves adjusting parameters such as material composition, device architecture, layer thicknesses, interfaces, or manufacturing processes to maximize efficiency, increase stability, or enhance other

desirable characteristics. ML algorithms can be used to identify the optimal combinations of these parameters, significantly reducing the need for extensive trial-and-error experimentation. For instance, genetic algorithms can be employed to explore a vast parameter space and find the best configuration for high-efficiency OSCs [11].

3. Materials Discovery and Design: Materials discovery and design involve the systematic search, identification, and development of new materials or the optimization of existing materials with desired properties for specific applications.
    - Property Prediction and Screening: ML models can predict the properties of potential materials, allowing researchers to screen large databases and identify promising candidates quickly. For example, predictive models can estimate the electronic properties of new organic molecules, aiding in the discovery of high-performance materials for OSCs [37,43].
    - Database Mining and High-Throughput Screening: ML algorithms can mine existing databases of materials to identify patterns and correlations that may not be apparent through traditional analysis. High-throughput screening techniques can rapidly evaluate a vast number of materials, accelerating the discovery process [44].
    - Structure–Property Relationships: Understanding the relationships between molecular structure and material properties is crucial for designing new materials. ML can help elucidate these relationships, guiding the rational design of materials with desired characteristics.
    - Design and Synthesis: Once promising materials are identified, ML can aid in optimizing the synthesis processes to ensure reproducibility and scalability. For example, ML models can suggest optimal reaction conditions to synthesize high-purity materials efficiently.
4. Process and Manufacturing Optimization: Process and manufacturing optimization in the context of OSCs involves improving and refining the procedures, techniques, and production methods used in fabricating these photovoltaic devices [31].
    - Process Control and Standardization: ML can be used to develop standardized protocols that ensure consistent quality and performance of OSCs. For example, ML algorithms can monitor production processes in real time, adjusting parameters to maintain optimal conditions.
    - Yield Improvement: By analyzing production data, ML can identify factors that influence yield and suggest modifications to improve it. This can lead to higher efficiency and lower costs in OSC manufacturing.
    - Scaling Production and Cost Reduction: ML techniques can optimize manufacturing processes to make them more scalable and cost-effective. For instance, predictive models can help in planning resource allocation and minimizing waste.
    - Robustness and Reliability: ML can enhance the robustness and reliability of OSCs by identifying and mitigating factors that lead to device degradation. This can result in longer-lasting and more stable solar cells.
5. Pattern Recognition and Data Analysis: Pattern recognition and data analysis involve the systematic process of identifying meaningful patterns, structures, or relationships within datasets, enabling the extraction of valuable insights or information [37].
    - Data Collection and Preprocessing: Efficient data collection and preprocessing are crucial for ensuring high-quality inputs for ML models. This includes cleaning data, handling missing values, and normalizing data to make it suitable for analysis.

- Exploratory Data Analysis (EDA): EDA techniques help in understanding the underlying patterns and distributions in the data. Visualization tools can provide insights into how different variables interact and influence OSC performance.
- Feature Extraction and Selection: Identifying the most relevant features or descriptors is essential for building accurate ML models. Techniques like PCA can reduce the dimensionality of the data, focusing on the most significant variables.
- Clustering and Classification: Clustering algorithms can group similar data points, helping to identify patterns in material properties. Classification algorithms can categorize materials based on their predicted performance.
- Regression and Prediction: Regression techniques can model the relationships between variables, providing predictions for new data points. These predictions can guide the development of new materials and the optimization of OSCs.
- Anomaly Detection and Outlier Analysis: Identifying anomalies and outliers in the data can reveal potential issues or novel phenomena that warrant further investigation. This can lead to new discoveries and improvements in OSC technology.
- Correlation and Relationship Analysis: Understanding the correlations and relationships between different variables helps in identifying key factors that influence OSC performance. This knowledge can inform the design and optimization of new materials [44].

It is worth acknowledging the transformative potential of ML in identifying material properties, optimizing synthesis processes, and predicting device performance metrics. For instance, the ability to accurately model donor/acceptor interactions using molecular descriptors has streamlined material discovery. Additionally, advanced algorithms like neural networks and random forests have shown potential in predicting efficiency metrics with higher accuracy than traditional methods. Such insights not only clarify the theoretical capabilities of ML but also demonstrate its pivotal role in accelerating the innovation and optimization of OSC technologies.

5.1. Molecular Descriptors

Molecular descriptors, which describe a molecule's physical and chemical characteristics, are derived from the molecular structure of a compound. They vary in complexity from more basic properties like charge distribution to more intricate ones like the number of a particular atom. Thousands of different categories of molecular descriptors exist, ranging from zero-dimensional (0D) to three-dimensional (3D) ones [45].

Atomic number, atom type, and molecular weight are examples of molecular information that is described using 0D descriptors, which do not imply topology or atom connection. One-dimensional descriptors provide counts and types of chemical fragments. On the other hand, topological and topo-chemical molecular properties are defined by 2D descriptors. Lastly, geometrical information is captured by 3D descriptors, which also contain conformational information like partial surface charges and molecule volume. The majority of the molecule's properties must be provided by an ideal expression, which should also be devoid of unnecessary details.

Different representations of the same molecule can capture a wide range of chemical details, often at varying levels of complexity. Figure 3 showcases some of these different forms. Molecular descriptors, which are straightforward and quick to compute, enable the rapid assessment of a large number of materials.

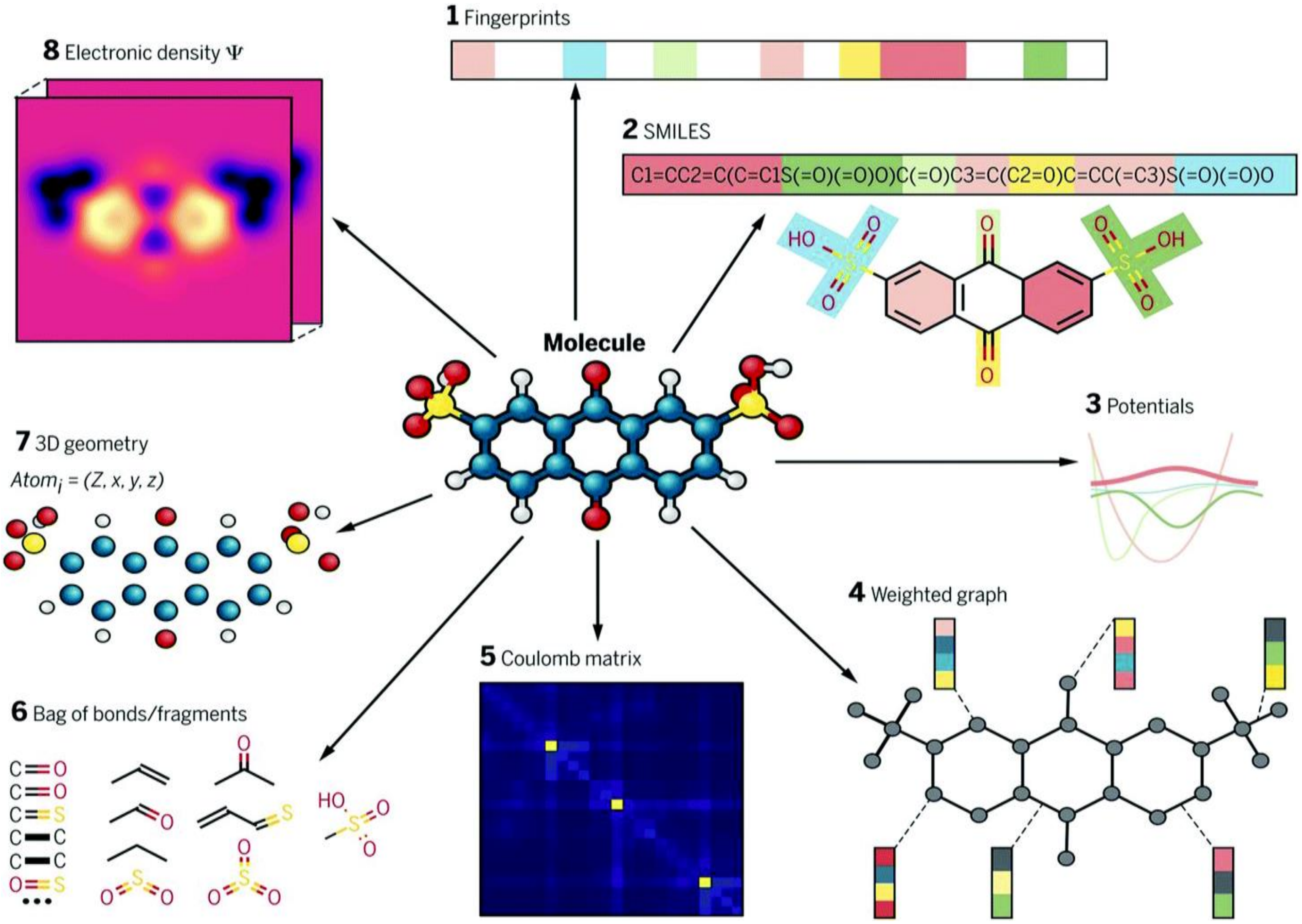


Figure 3. Different types of molecular representations applied to one molecule, AQDS, which is used in the construction of organic redox-flow batteries. Clockwise from top: (1) A fingerprint vector that quantifies presence or absence of molecular environments; (2) SMILES strings that use simplified text encodings to describe the structure of a chemical species; (3) potential energy functions that could model interactions or symmetries; (4) a graph with atom and bond weights; (5) Coulomb matrix; (6) bag of bonds and bag of fragments; (7) 3D geometry with associated atomic charges; and (8) electronic density. Reproduced with permission from [46].

Pereira et al. created a dataset including 111,000 molecules and trained a ML model utilizing RF methodology [47]. By using this model, they forecasted the LUMO and HOMO with an error of less than 0.16 eV, without employing any DFT computations. This can accelerate the high-throughput screening of organic semiconductors for solar cells. Sui et al. created a series of innovative acceptors derived from multi-conformational bistricyclic aromatic (BAE) compounds [48].

They have forecasted their PCEs utilizing a ML model constructed from experimental data via a cascaded support vector machine (CasSVM). The CasSVM model is an innovative two-tier network (see Figure 4), comprising three subset SVM models that produce J_{SC}, V_{OC}, and *FF* as outputs in the first tier. The second level was employed to determine the correlation between the outputs of the first level and the final endpoint PCE. The most established CasSVM model has forecasted the PCE value of OPVs with a mean absolute error (MAE) of 0.35 (%), representing about 10% (3.89%) of the average PCE. The R^2 value was 0.96. This methodology can be highly beneficial for experimental chemists to evaluate probable candidates prior to synthesis.

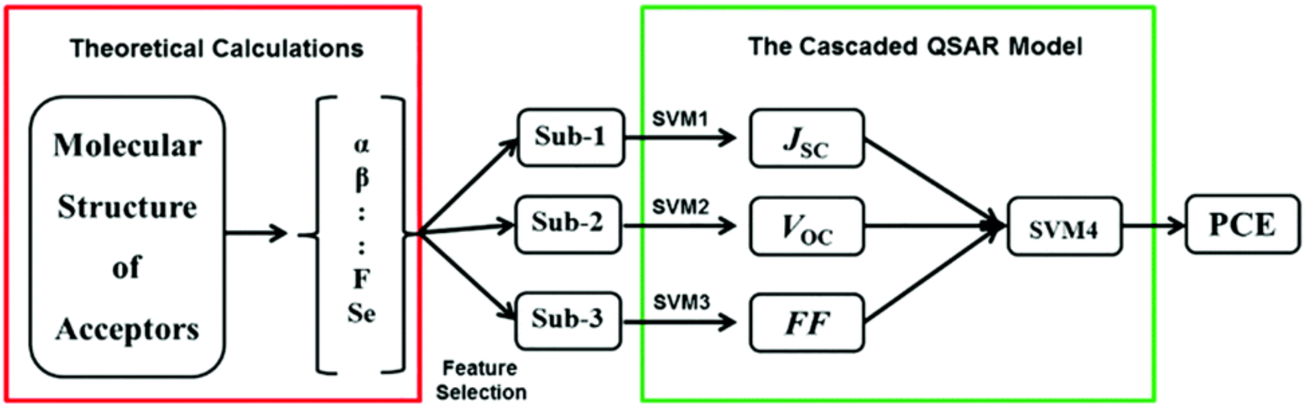


Figure 4. Structure of the cascaded QSAR model based on a cascaded support vector machine (CasSVM) framework for predicting J_{SC}, V_{OC}, FF, and PCE in organic photovoltaic devices. Adapted with permission from [48].

Molecular descriptors are numerical values derived from the molecular structure of a compound, describing its physical, chemical, and geometric characteristics. They range from simple properties like atomic numbers to complex features like molecular topology and electronic distribution. Thousands of categories of molecular descriptors exist, broadly classified into zero-dimensional (0D), one-dimensional (1D), two-dimensional (2D), and three-dimensional (3D) descriptors.

- Zero-dimensional descriptors provide basic molecular information, such as atomic number, molecular weight, and atom types, without including topological or connectivity information.
- One-dimensional descriptors capture counts and types of chemical fragments, representing molecular composition.
- Two-dimensional descriptors include topological and topo-chemical properties, reflecting atom connectivity and chemical bonding patterns.
- Three-dimensional descriptors capture geometric and conformational information, such as molecular volume, surface area, and partial charges.

The choice of molecular descriptors significantly influences the performance of ML models in predicting molecular properties. Pereira et al. created a dataset of 111,000 molecules and trained an RF model to predict LUMO and HOMO energy levels. By utilizing 2D and 3D molecular descriptors, the model achieved an error of less than 0.16 eV, eliminating the need for computationally intensive DFT calculations [47].

5.2. Comparison of Prediction Accuracies

ML plays a key role in improving the performance of OSCs through the use of molecular descriptors. By looking at different types of descriptors, researchers are making strides in predicting energy levels, estimating PCE, and speeding up the screening process for new materials. The following points outline recent studies that demonstrate how combining molecular descriptors and ML is unlocking new potential in OSC development.

1. Descriptors for Energy Level Predictions: When predicting LUMO and HOMO, studies have demonstrated that 3D descriptors generally outperform 2D descriptors due to their inclusion of geometric information. The RF model achieved an MAE of 0.16 eV using combined 2D and 3D descriptors, whereas models relying solely on 2D descriptors showed an MAE of approximately 0.24 eV [47].

2. Descriptors for PCE Predictions: Sui et al. developed a cascaded support vector machine (CasSVM) model using a combination of 0D, 1D, and 2D descriptors to predict key device parameters such as J_{SC}, V_{OC}, and FF, which were then correlated to PCE [48,49]. Their model achieved an MAE of 0.35% for PCE predictions, corresponding to about 10% of the average PCE value (3.89%). In contrast, earlier models that excluded 2D descriptors showed higher MAE values, often exceeding 0.50%. The R^2 value of 0.96 for the CasSVM model highlights its superior predictive capability when utilizing a diverse set of molecular descriptors.
3. High-Throughput Screening: Omar et al. compared 0D and 3D descriptors for high-throughput screening of organic semiconductors [50]. They found that 3D descriptors, incorporating molecular volume and partial charges, improved the classification accuracy of high- versus low-performing materials by 15% compared to 0D descriptors alone.

5.3. Efficiency Versus Complexity

While 3D descriptors generally provide higher predictive accuracy, they are computationally more expensive to calculate. In contrast, 0D and 1D descriptors are faster to compute and are sufficient for initial screening. An ideal molecular descriptor balances informativeness and computational efficiency, capturing essential molecular properties without introducing extraneous details.

Molecular descriptors enable rapid assessment of large libraries of materials, accelerating the identification of promising candidates for OSCs. By leveraging diverse descriptors, experimental chemists can better evaluate potential compounds before synthesis, thereby optimizing the design process.

5.4. Molecular Fingerprints

Molecular fingerprints are computerized representations of chemical structures that exclude precise structural features such as coordinates. They are utilized to query databases and discern similarities among compounds. Multiple methodologies are available for transforming a molecular structure into a digital representation, such as key-based fingerprints, circular fingerprints, and topological or path-based fingerprints, each encompassing additional subtypes. This review provides a broad overview of molecular fingerprints and their applications in OSCs, so readers who are interested in exploring detailed methodologies or specific implementations may refer to these references [51,52], which delve deeper into the technical aspects of fingerprint generation and their computational frameworks.

In recent years, organic photovoltaics have seen widespread use of non-fullerene acceptors [53–56]. In 2017, Aspuru-Guzik and his team compiled a dataset of over 51,000 non-fullerene acceptors. These acceptors were based on various compounds, including benzothiadiazole (BT), diketopyrrolopyrroles (DPPs), perylene diimides (PDIs), tetraazabenzodifluoranthenes (BFIs), and fluoranthene-fused imides, sourced from the Harvard Clean Energy Project (HCEP) [57].

To regulate the DFT methods for calculating the HOMO and LUMO values of new non-fullerene acceptors, a dataset of 94 experimentally reported molecules was used. Instead of the commonly used linear regression, they opted for Gaussian process regression due to the lack of a linear trend. They applied the Scharber model to estimate the PCE of OSCs, focusing on non-fullerene acceptors and the standard electron donor material, poly[N-90-heptadecanyl-2,7-carbazole-alt-5,5-(40,70-di-2-thienyl 20,10,30-benzothiadiazole)] (PCDTBT). The DFT-calculated HOMO and LUMO values of the acceptors, along with the experimentally reported values for PCDTBT, were inputs for the Scharber model. To validate the PCE predictions of the Scharber model, they compared them with 49 experimentally reported values, finding only a weak correlation (r = 0.43 and R^2 = 0.11) [57].

Predicting the PCE and specific device properties is crucial. To enhance a particular property, it is essential to understand the relationship between that property and the molecular descriptors. This connection helps identify which molecular features influence the property, allowing for targeted improvements [58]. For instance, most high-performing OSC devices exhibit lower open-circuit voltages (V_{OC}). In bulk heterojunction (BHJ) OSCs, charge separation is generally associated with considerable voltage losses due to the additional energy necessary to dissociate excitons into free carriers.

This voltage loss in high-performance OSCs is generally around 0.6 V, which is approximately 0.2–0.3 V higher than the losses observed in silicon (c-Si) and gallium arsenide (GaAs) solar cells [59]. Non-fullerene acceptors exhibiting extended thin-film absorption and appropriate energy levels can facilitate an optimal balance between V_{OC} and J_{SC} [60]. Their structural adaptability enables significant modulation of absorption and molecular energy levels. ML can markedly expedite the identification of appropriate materials.

By predicting specific parameters, it can further enhance the (PCE). Aspuru-Guzik and his team calibrated the open-circuit voltage (V_{OC}) and short-circuit current density (J_{SC}) values, which were calculated using the Scharber model and available experimental data, based on structural similarity. They derived information from the molecular graph utilizing enhanced connectivity fingerprints and employed a Gaussian process. This calibration technique reduced the functional dependence of the computed properties, enabling high-throughput virtual screening.

In 2019, Sun et al. collected a dataset of 1719 donor materials [39]. The researchers experimented with various inputs, including seven types of molecular fingerprints, two types of descriptors, ASCII strings, and images. They classified donor materials into two categories based on their PCE: "low" and "high". The models developed using fingerprints exhibited the best performance, achieving an 86.76% accuracy in predicting the PCE class. To validate the ML results, they synthesized ten donor materials, and the model accurately classified eight of these molecules. The experimental results closely matched the predicted outcomes. However, this study's practical value is limited because categorizing PCE into just two broad categories (0–2.9% and 3–14.6%) is much simpler than predicting the PCE of individual semiconductors with precision.

In the same year, Nagasawa et al. extracted 2.3 million molecules from the Harvard Clean Energy Project database [40]. Out of the dataset, 1000 molecules were initially chosen based on their calculated PCE. The researchers used MACCS fingerprints and the extended connectivity fingerprint (ECFP6) key to train their ML model. Through random forest (RF) screening, they further narrowed down the selection to 149 molecules, as shown in Figure 5. However, the RF method's accuracy for predicting PCE was only 48%. They ultimately selected one polymer for its synthetic feasibility, but the solar cell device made from this polymer had a PCE of 0.53%, significantly lower than the RF prediction of 5.0–5.8%.

This disparity can be attributed to two primary factors. The RF model was first trained on PCE values derived from the Scharber model, which exhibits suboptimal performance. The structures of polymer donors documented in the literature are more intricate than those of the semiconductors in the HCEP database. Notwithstanding these features, the predictive accuracy of the RF model for PCE remains inadequate. Consequently, the ML model must enhance its accuracy, and various materials should be synthesized for empirical validation.

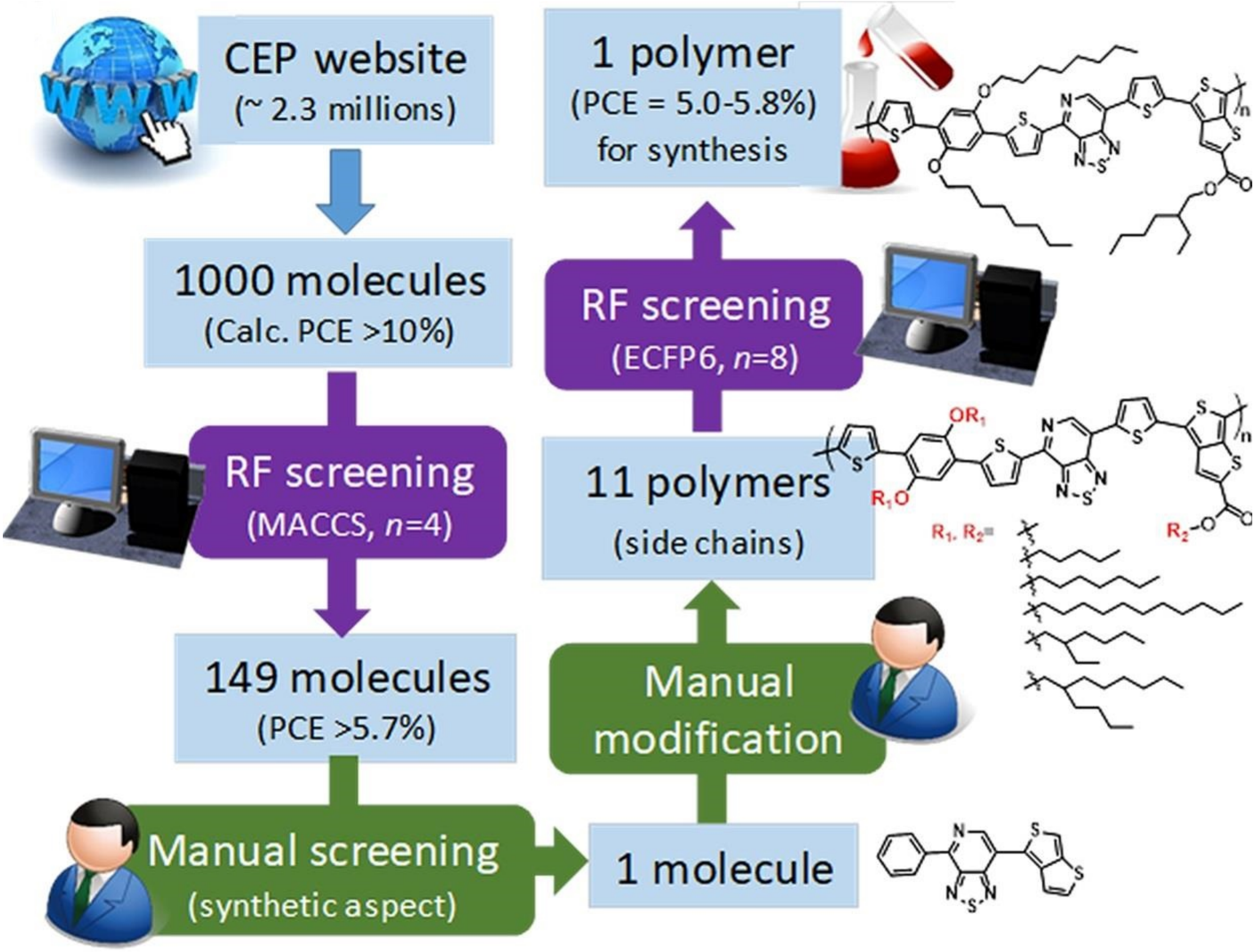


Figure 5. Scheme of polymer design by combining RF screening and manual screening/modification. Reproduced with permission from [40].

Schmidt and colleagues assembled a dataset including 3989 monomers and developed a model utilizing a grammar variational autoencoder (GVA) [61]. Even without knowing the precise locations of individual atoms, the trained model can calculate the LUMO and lowest optical transition energies. Furthermore, conformations with the required LUMO and optical gap energies can be synthesized using this approach. Deep neural network (DNN) predictions were more accurate than grammar variational autoencoder (GVA) predictions; however, forecasting the LUMO still requires density functional theory (DFT) calculations to find the atomic locations. Therefore, it is not possible to bypass DFT calculations when using the DNN model.

When compared to neural networks trained on molecular fingerprints, SMILES, Chemception, and molecular graphs, their suggested models performed better. Peng and Zhao utilized convolutional neural networks (CNNs) to develop models for generating and predicting the properties of non-fullerene acceptors. These models aid in the design and analysis of these materials, leveraging the power of CNNs to identify and optimize key characteristics [62]. Peng and Zhao used various molecular descriptors, including extended-connectivity fingerprints, Coulomb matrices, molecular graphs, bag-of-bonds, and SMILES strings, to construct their models. The depth of the convolutional layers in their CNNs influenced the diversity of the generated (NFAs). In order to confirm the compounds that were predicted, they used quantum chemistry computations. They employed an attention method to decipher the outcomes of feature extraction using dilated convolution layers in their prediction model. They concluded that graph-based representations of molecules were more effective than string-based representations.

In most experimental studies, donor and acceptor materials for OSCs are optimized separately. However, optimizing only one component at a time limits the exploration of potential combinations. Troisi used ML to investigate whether these components should be optimized individually or if simultaneous optimization would yield better results [63]. They took molecular fingerprints as their starting point and searched the literature for combinations of 262 donors (D) and 76 acceptors (A). Despite the tiny dataset, they achieved a high accuracy (r = 0.78) by predicting the PCE of BHJ solar cells using these donor/acceptor combinations. The most promising combination was recommended for experimental testing.

The study by Wu et al. is notable for its comprehensive approach and transformative impact on organic photovoltaic research. It stands out due to the scale of its analysis, utilizing 565 donor/acceptor pairs as training data and screening an unprecedented 432 million pairings for PCE predictions. The use of advanced ML models, such as boosted regression trees (BRT) and random forests (RF), achieving prediction accuracies of 0.71 and 0.70, respectively, highlights the methodological rigor [64,65]. The experimental validation of six selected pairings, with results closely aligning with model predictions, underscores the reliability of their workflow [8]. Focusing on non-fullerene acceptors (NFAs) from the high-performing Y6 series, the study exemplifies how ML can accelerate material discovery and optimize device performance, marking a significant advancement in the field. The entire workflow of the study is illustrated in Figure 6.

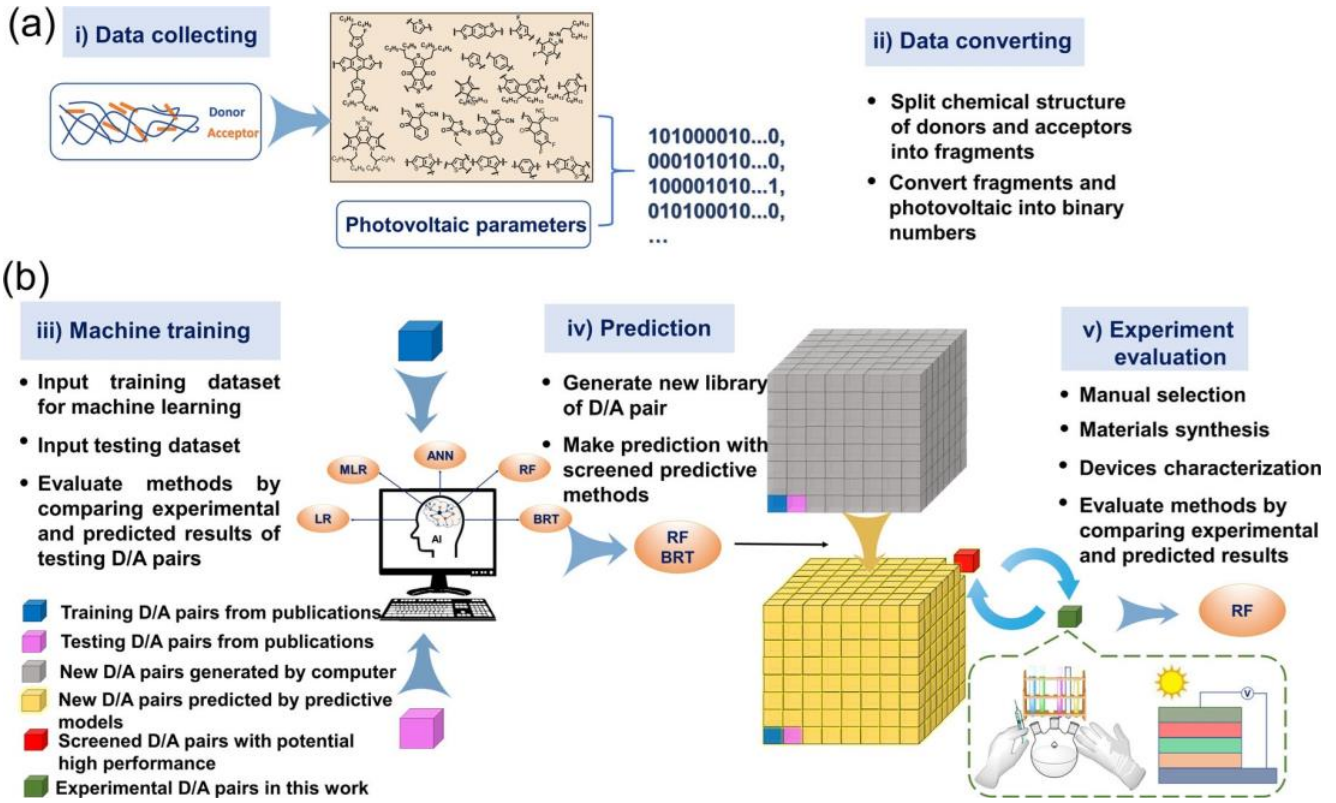

(a)
i) Data collecting
Donor
Acceptor
Photovoltaic parameters
101000010...0,
000101010...0,
100001010...1,
010100010...0,
...
ii) Data converting
• Split chemical structure of donors and acceptors into fragments
• Convert fragments and photovoltaic into binary numbers
(b)
iii) Machine training
• Input training dataset for machine learning
• Input testing dataset
• Evaluate methods by comparing experimental and predicted results of testing D/A pairs
MLR
ANN
RF
LR
BRT
AI
RF
BRT
iv) Prediction
• Generate new library of D/A pair
• Make prediction with screened predictive methods
v) Experiment evaluation
• Manual selection
• Materials synthesis
• Devices characterization
• Evaluate methods by comparing experimental and predicted results
RF
Training D/A pairs from publications
Testing D/A pairs from publications
New D/A pairs generated by computer
New D/A pairs predicted by predictive models
Screened D/A pairs with potential high performance
Experimental D/A pairs in this work

Figure 6. Workflow of building, application, and evaluations of ML methods. (**a**) Scheme of collecting experimental data and converting chemical structure to digitized data. (**b**) Scheme of machine training, predicting, and method evaluation. Reproduced with permission from [64].

5.5. Images

ML has made significant strides in image recognition by identifying features within complex backgrounds and associating them with specific outputs. To put this skill to use, Sun and colleagues trained a deep neural network to detect and automatically categorize chemical structures; this allowed them to estimate the PCE of OSCs [66]. The researchers used unaltered images of chemical structures for their model, which was both fast and low in computational cost, making it feasible to run on a personal computer. This approach achieved an accuracy of 91.02% in predicting the PCE of donor materials. The workflow of this study is illustrated in Figure 7.

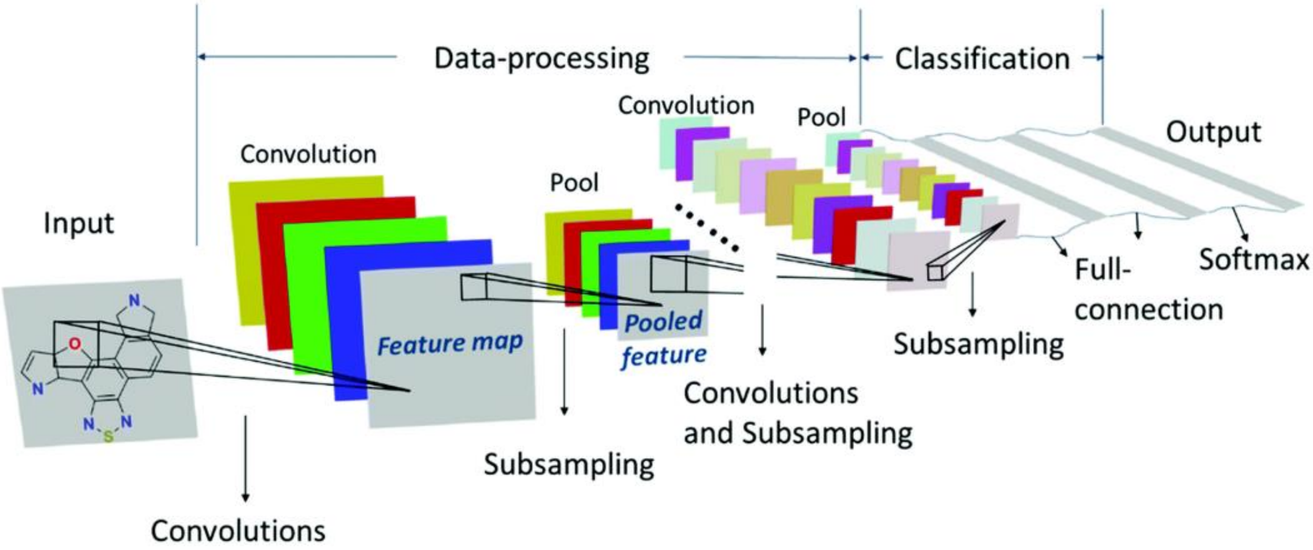


Figure 7. Structure of the convolutional neural network (CNN). Reproduced with permission from [66].

However, this study has several limitations. Firstly, the ML model was trained using data from the HCEP, but the molecules reported in the literature are generally more complex than those in the HCEP database. Secondly, the Scharber model's PCE estimates were based on energy levels calculated using DFT, which are not always accurate. The performance of OSCs is influenced by many factors, including the materials in the active layer, solubility, solvent additives, crystallinity, and molecular orientation. Only images of chemical structures are used as the input does not provide realistic results. Molecular descriptors, which provide more detailed information about the molecules, are a better option compared to just using pictures of the structures.

5.6. Microscopic Properties

Optical gap, charge-carrier mobility, ionization potential, electron affinity, and hole–electron binding energy are some of the microscopic features of organic materials that determine the efficiency of OSCs. When contrasted with more basic topological descriptors, these microscopic descriptors offer a more grounded view of solar cell applications. However, computing or experimentally determining these microscopic properties can be costly and time-consuming.

To address this, Ma and colleagues used 13 microscopic properties as descriptors to train a model for predicting PCE. They utilized a dataset of 270 small molecules for this purpose. This approach aims to enhance the accuracy of PCE predictions by incorporating detailed microscopic properties, despite the higher computational and experimental costs involved [67]. PCE was predicted using a variety of methods, such as artificial neural networks, gradient boosting, and random forest. The gradient boosting model stood out among the rest, achieving an amazing R-value of 0.79. Unfortunately, these models rely on computationally expensive characteristics like excited state and polarizability. Massive, high-throughput virtual screening of possible compounds is hindered by this hefty price tag.

Ma and colleagues utilized random forest (RF) and gradient boosting regression tree (GBRT) algorithms to predict key device characteristics such as (V_{OC}), (J_{SC}), and (*FF*) based on microscopic properties. They found a strong correlation between JSC (r = 0.78) and *FF* (r = 0.73) with PCE, indicating these factors are reliable predictors of efficiency. However, V_{OC} showed a very weak correlation with PCE (r = 0.15), which aligns with findings from recent studies [40]. The J_{SC} and *FF* are found to be poorly correlated (r = 0.33), with almost no correlation between V_{OC} and J_{SC} (r = −0.18) as well as V_{OC} and *FF* (r = −0.09).

The impact of various descriptors on ML models' prediction abilities was studied by Trois and colleagues. Using information from 566 donor/acceptor pairs retrieved from the literature, they trained k-Nearest Neighbors (k-NN), Kernel Ridge Regression (KRR), and Support Vector Regression (SVR) models. To improve the accuracy of these ML models in predicting the performance of OSCs, our investigation sought to identify the most effective descriptors [68]. The research made use of both spatial (topological) and temporal (physical) characteristics, including energy levels, molecule size, light absorption, and mixing characteristics. The ML models benefited greatly from the structural descriptors. Some physical parameters correlated strongly with PCE, but these did not improve the model's predictive capability as the structural descriptors already included this information.

When developing organic semiconductors, a number of building pieces are utilized to construct push–pull conjugated systems. These building blocks include electron-deficient, electron-rich, and p-spacer units. In order to screen 10,000 compounds made from 32 distinct building blocks, Ma and colleagues used ML algorithms. Their research set out to deduce how the molecules' characteristics are impacted by the type and configuration of these building pieces. Using their ground and excited states, we were able to calculate their descriptive properties. They found 126 possible candidates with efficiency predictions above 8% using ANN and gradient boosting regression trees (GBRT) models. This method was effective in finding OSC candidates through screening.

With a Pearson's coefficient (r) of 0.68, the ML model trained by Padula et al. to forecast device parameters outperformed the Scharber model [11]. In OSCs, the thermodynamics of mixing the materials in the active layer dictates how the film morphology evolves. Charge transfer and light harvesting are both impacted by this evolution, which in turn affects the device's stability and performance [69,70]. Investigating the connection between the characteristics of molecular interactions and the phase behavior of thin films is crucial. To achieve this goal, Perea et al. investigated the phase evolution of fullerenes and polymers using the ANN model in conjunction with the Flory–Huggins solution theory [71]. Solubility parameters were predicted using the surface charge distribution and the ANN model. To characterize the stability of polymer–fullerene blends, a figure of merit was developed, which is combined with solubility characteristics (Figure 8).

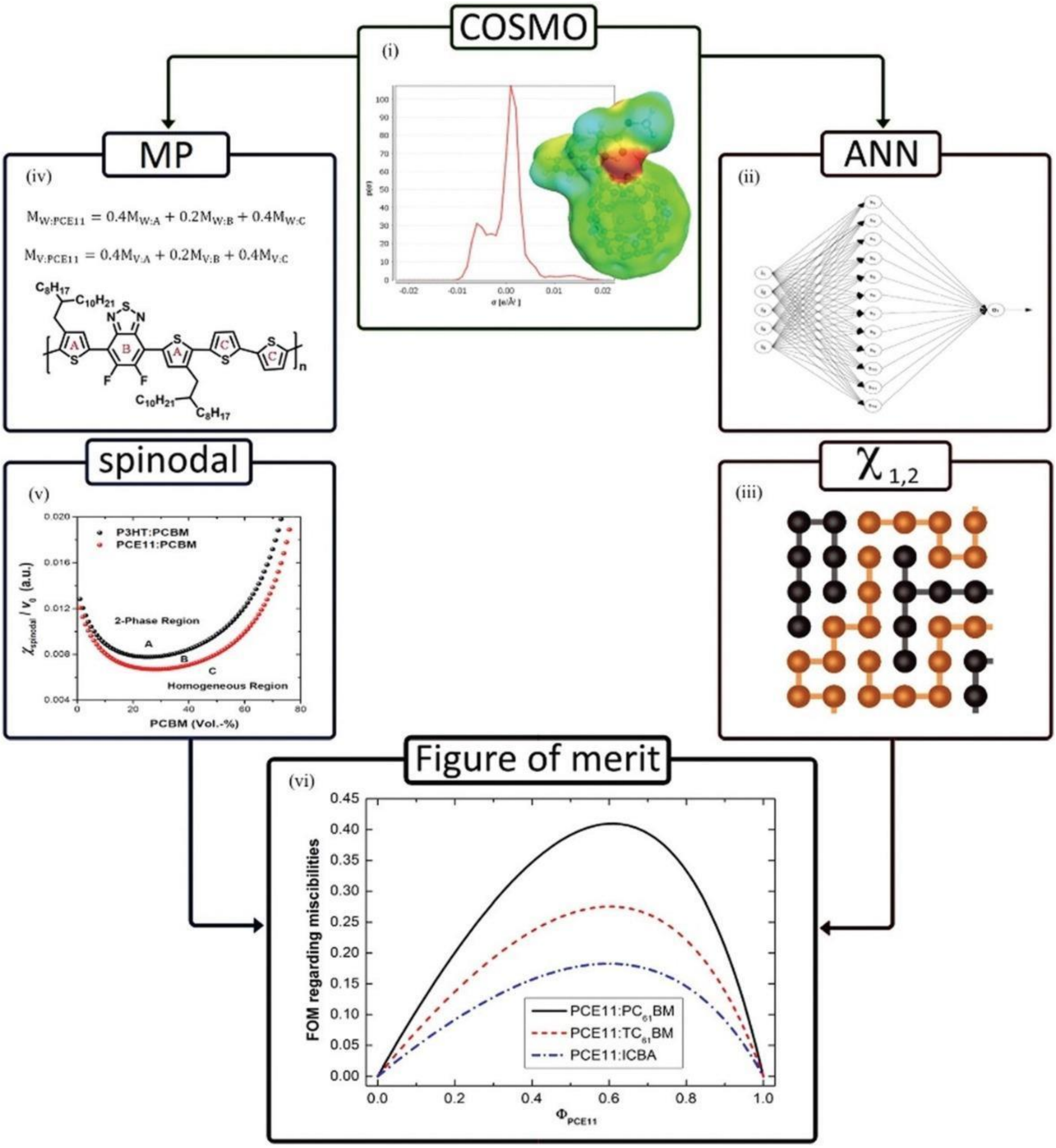


Figure 8. Computational flowchart describing the routine for determining the relative stability capable of describing the microstructure of polymer–fullerene blends. (**i**) Creation of the s-profile from the conductor-like screening model (COSMO); (**ii**) s-moments as extracted from COSMO are fed into an artificial neural network (ANN) to determine Hansen solubility parameters (HSPs); (**iii**) HSPs are used to calculate the qualitative Flory–Huggins interaction parameters (χ1,2); (**iv**) implementation of moiety-monomer-structure properties (reduced molar volumes/weights); (**v**) spinodal demixing diagrams resulting from polymer blend theory; and (**vi**) figure of merit (FoM) defined as the ratio of the Flory–Huggins intermolecular parameter and the spinodal diagram forms the basis of a relative stability metric. Reproduced with permission from [71].

5.7. Energy Levels

The performance of OSCs is significantly influenced by the energy levels of the donor and acceptor materials. When there is a mismatch in these energy levels, it can lead to substantial energy loss due to radiative recombination, which in turn reduces the PCE of the OSCs [72]. In 2017, Aspuru Guzik's group investigated millions of molecular motifs using 150 million DFT calculations [73]. PCE was predicted using Scharber's model [10] and the calculated energy level was used as input. Candidates with a PCE of more than 10% were identified.

Automatic thiophene-based polymer production from donor and acceptor units, orbital level calculation using Hückel-based models, and photovoltaic characteristic evaluation were all reported by Imamura et al. in 2017 [74]. PCE was calculated using Scharber's model, but its performance is very poor [11,57,75]. Molecular descriptors and microscopic properties of semi-conductors were totally ignored. With a training set R^2 of 0.85 and a testing set R^2 of 0.80, Min-Hsuan Lee demonstrated excellent prediction accuracy using random forest (RF) modeling on a database including 4100 bulk heterojunction solar cells [76].

As discussed earlier, various examples of ML applications in binary solar cells have been highlighted. However, ternary OSCs generally exhibit better performance than binary ones. One of the main issues with binary OSCs is their limited light harvesting capability due to the narrow absorption range of organic semiconductors. In contrast, ternary OSCs include a third component, which can be either a donor or an acceptor. This additional component not only enhances photon harvesting by serving as an extra absorber but also contributes to achieving a more favorable morphology [77]. The operation of ternary solar cells is more intricate than that of binary solar cells, making the identification of optimal third components for ternary solar cells a tough endeavor [78,79]. Min-Hsuan Lee has developed a ML model for ternary solar cells utilizing random forest, gradient boosting, k-Nearest Neighbors (k-NN), Linear Regression, and Support Vector Regression. The LUMO value of the donor (D1) exhibited a significant linear connection with PCE (r = −0.55), but the correlations of other markers with PCE were minor [80]. The V_{OC} value has a strong correlation with the donor's HOMO (r = −0.54) and LUMO (r = −0.54), indicating that the donor's energy levels require additional examination to elucidate the origin of V_{OC} in ternary OSCs. The Random Forest model exhibited the highest R^2 score (0.77 on the test set) across all ML approaches. In a separate work, he developed the ML model to forecast the voltage of operation characteristics of fullerene derivative-based ternary OSCs. The descriptions were identical to those in a prior study [81]. The Random Forest model exhibited an R^2 score of 0.77. Both investigations utilized only the energy levels of organic semiconductors as descriptors, neglecting other chemical descriptors and the influence of thin film shape. Enhancing the efficiency of OSCs necessitates the development of a hybrid modeling framework that integrates thin-film features, including the optimal ratio of the three components, and fabrication parameters, such as annealing temperature and solvent additives. By controlling these variables, we may improve charge generation and minimize voltage loss, hence increasing total device efficiency [82]. Theoretical analysis of the morphology of the three components is much more complex than that of two components.

Theoretical analysis of the morphology in ternary OSCs presents greater challenges compared to binary systems due to the added complexity of the third component. This third component, often a donor or acceptor, introduces additional interactions that affect photon harvesting, charge separation, and transport mechanisms. Morphological studies require advanced modeling to understand how the three components interact at a molecular level and influence the efficiency of the device.

For instance, the interplay between energy levels and a thin-film structure in ternary OSCs demands the integration of experimental data and computational simulations to predict and optimize device performance effectively [82]. These complexities underscore the need for hybrid modeling frameworks that incorporate both structural and energetic descriptors to achieve meaningful insights and guide experimental designs. By carefully controlling these variables, it is possible to enhance charge generation and reduce voltage loss, thereby improving the overall efficiency of the device [82]. A theoretical analysis of the morphology becomes significantly more complex when dealing with three components compared to two components.

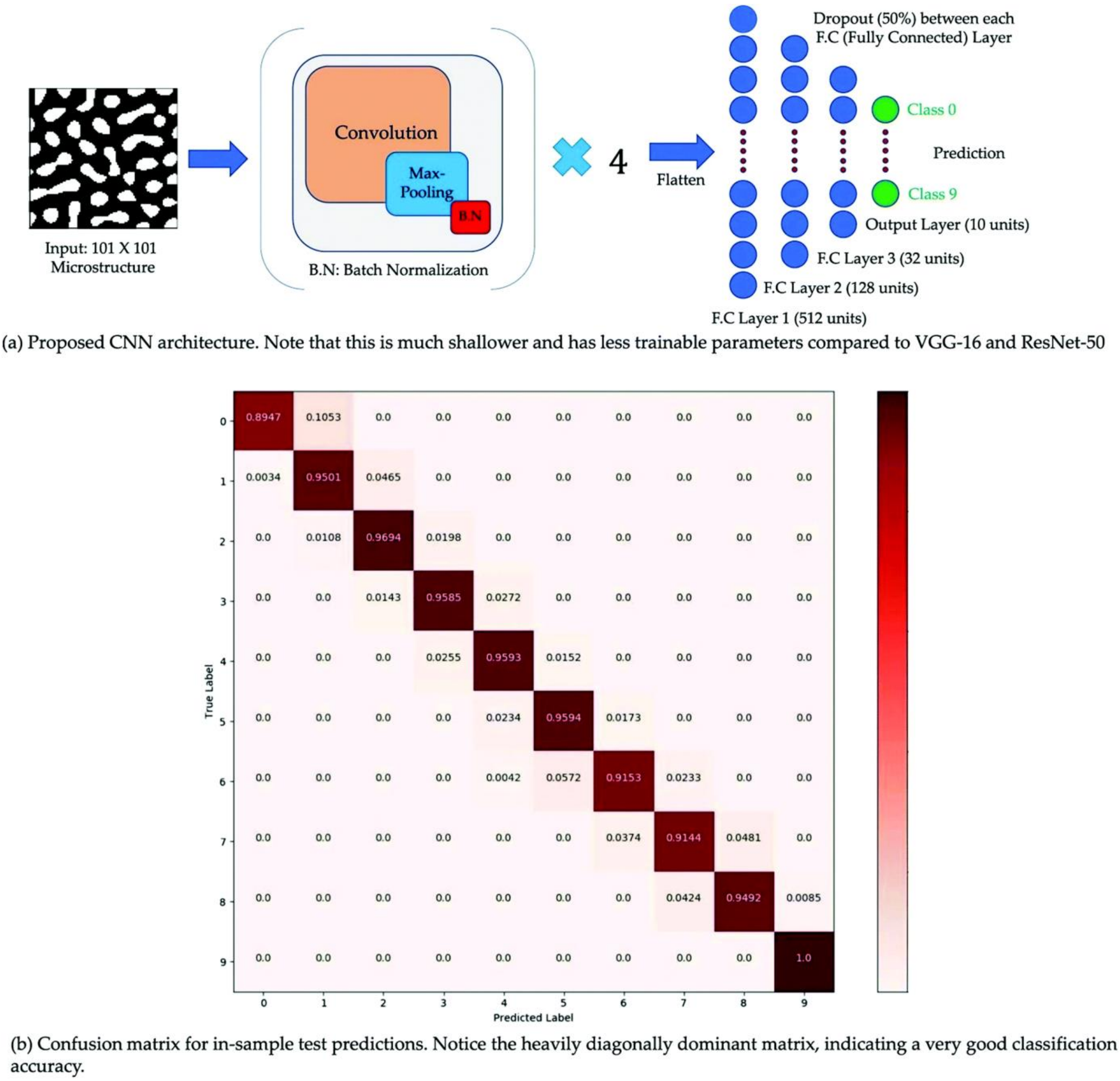
Tandem OSCs are known for their superior PCE. These cells consist of two subcells, designed to extend the range of photon response and minimize both transmission and thermalization losses. This dual-layer architecture enhances the overall efficiency by capturing a broader spectrum of light and reducing energy losses that typically occur in single-junction solar cells [83]. Developing a correlation between the efficiency and physical properties of active layer materials in OSCs is particularly challenging due to the vast diversity of organic materials available. This diversity results in a multitude of potential candidate materials, making the task more complex. To address this issue, Min-Hsuan Lee employed ML algorithms to predict the efficiency of tandem OSCs and identify optimal bandgap combinations for these devices. This approach helps streamline the selection process, making it more efficient and effective in finding high-performing material combinations [84]. Random forest regression was employed to predict the efficiency of tandem OSCs using energy levels as input data. The findings suggest that optimizing the energy offset in the LUMO level between the donor and acceptor materials can significantly enhance electron transfer and overall device performance. This optimization is crucial for improving the efficiency and effectiveness of the solar cells.

5.8. Simulated Properties

The efficiency of OSCs is largely governed by the morphology of the film. To further enhance PCE, it is crucial to have a thorough understanding of this film morphology. In addition to experimental methods, mathematical simulations can also be employed to explore the film's structure and analyze how various parameters affect it. This combination of experimental and computational approaches provides a comprehensive understanding that can lead to significant improvements in solar cell performance.

These simulations generally consist of two primary phases: the representation phase and the mapping phase. During the representation phase, a mathematical foundation is established to produce microstructures. In the mapping phase, the created microstructures are correlated with a certain desired attribute. The application of graph theory in the analysis of the microstructure of OSCs is gaining traction, since it offers a reliable approach to elucidating the correlation between microstructure and device performance [85–87]. For example, Ganapathy Subramanian and colleagues used a graph-based approach to study morphology descriptors in OSCs. They analyzed multiple mechanisms, including photon absorption, exciton diffusion, charge separation, and charge transport, offering a comprehensive assessment of how these elements affect the efficiency and performance of OSCs [88]. A strong association was shown between the graph-based technique and the computationally demanding method. In a separate investigation, they employed CNN to correlate film morphology with a short-circuit current (J_{SC}) [89].

They resolved the thermodynamically consistent Cahn–Hilliard equation for binary phase separation via an in-house finite element library [90]. A total of B65000 morphologies were generated with J_{SC}, and it evaluated each morphology using the excitonic drift-diffusion equation [91]. CNN using morphologies as input and J_{SC} as output showed a classification accuracy of 80% (Figure 9).



(a) Proposed CNN architecture. Note that this is much shallower and has less trainable parameters compared to VGG-16 and ResNet-50

(b) Confusion matrix for in-sample test predictions. Notice the heavily diagonally dominant matrix, indicating a very good classification accuracy.

Figure 9. (**a**) Simple sketch of CNN architecture and (**b**) confusion matrix. Reproduced with permission from [89].

Majeed et al. used the Shockley–Read–Hall-based drift-diffusion model to simulate current/voltage (*JV*) curves [92]. A total of 20,000 devices were produced, and electrical parameters including carrier trapping rates, energy disorder, trap densities, recombination time constants, and parasitic resistances were computed. Simulated data were employed to train the neural network.

After training the model, it was applied to the investigation of charge carrier dynamics in several famous OSC devices to determine the effect of surfactant choice and annealing temperature. The solubility of the materials in the active layer of an OSC in a specific solvent plays a crucial role in determining the film morphology, which in turn affects the device's performance. Risko and colleagues tackled this by calculating the free energy of mixing using molecular dynamics (MD) simulations. They also employed Bayesian statistics to further refine these calculations. This method provides a quick and efficient way to study a wide variety of solvents and solvent additives, helping to optimize the performance of the solar cells.

6. Challenges and Future Prospects

6.1. Data Infrastructure

The ML model for screening OPV compounds is often trained using data from the Harvard Clean Energy Project (HCEP). However, the complexity of molecules reported in the literature usually far exceeds that of the HCEP. This disparity can lead to inaccurate ML predictions. Training an ML model effectively requires a vast amount of data. In areas like image recognition, data availability is not an issue, with millions of input datasets available. In contrast, OSCs have data only in the hundreds or thousands. The accuracy of ML models reportedly improves as the number of data points (molecules) increases [23,66,84].

The ML model for screening OSC compounds is often trained using data from the Harvard Clean Energy Project (HCEP). However, the complexity of molecules reported in the literature usually far exceeds that of the HCEP, potentially leading to inaccurate ML predictions. Training an ML model effectively requires a substantial amount of data. In areas like image recognition, data availability is abundant, with millions of input datasets. In contrast, OSCs have data only in the hundreds or thousands.

Studies have demonstrated that increasing the number of data points significantly enhances ML model performance. For instance, a 10% improvement in predictive accuracy was reported in [23] when dataset size was increased from 1000 to 5000 molecules. Similarly, Sun et al. observed that doubling the dataset size improved their regression model's R^2 value by 15%. These findings underscore the importance of expanding datasets in the context of OSCs for improving the accuracy and reliability of ML predictions [39].

Including massive amounts of data in ML models trained using power conversion process descriptors is challenging because DFT computations are computationally intensive. A potential solution for small datasets is meta-learning, which involves learning from both within and across problems. Another viable approach for dealing with sparse data are a Bayesian framework. Striking a balance between data accessibility and model predictive power requires a two-pronged approach. The degree of freedom (DoF) of the model can mediate the influence of data size on model precision, potentially leading to a relationship between precision and DoF. This concept is theoretically grounded in the statistical bias-variance trade-off [93]. A significant negative point is the lack of high-quality, extensive datasets specifically tailored for OSCs. This shortage impedes the development of highly accurate ML models, as the limited and less complex data often fail to capture the intricacies of real-world OSC materials.

6.2. Descriptor Selection

A crucial step in ML modeling is the selection of molecular descriptors. While fingerprints and molecular descriptors are simple and quick to compute, they are not ideal for modeling OSCs. Understanding photovoltaic processes requires precise quantum computations on a small scale, which are prohibitively costly for rapid virtual screening on a broad scale. There needs to be an appropriate compromise between precision and quickness. Developing a new generation of descriptors specifically for organic semiconductors is critically needed, along with accurate and conveniently accessible fingerprints.

6.3. Multidimensional Design

Many models account for power PCE by correlating chemical structures but fail to consider miscibility and film morphology. Applying the theories of Flory and Huggins might enhance ML approaches. Prediction accuracy could be improved by including data from grazing-incidence small-angle X-ray scattering (GISAXS), atomic force microscopy (AFM), transmission electron microscopy (TEM), and grazing-incidence wide-angle X-ray scattering (GIWAXS) [94].

Despite ML's apparent mastery of everyday image analysis, results from these methods vary significantly. It is exceedingly difficult to perform ML analysis on images generated by the aforementioned techniques because, unlike everyday photos, they contain unique characters. Microscope images are often associated with high-level noise and aberrations [95]. Experimental images also correlate highly with the physicochemical characteristics of materials, mixing conditions, and experimental settings, adding complexity. Additionally, the physical significance of images captured using various techniques varies, necessitating different analytical approaches.

Due to multiple images for a single compound under varying experimental circumstances, accumulating a comprehensive dataset is laborious. Implementing automatic image extraction and sorting is challenging, necessitating human intervention. Analysis and specification of tasks, such as data label decision-making or target property selection, will constitute the second stage. The *FF* values are heavily affected by the active layer's morphology [96]. Thus, choosing *FF* as the objective instead of PCE might be more practical. The correlation between *FF* and other components can then be determined using PCE. Another consideration is whether to use classification or regression. Classification may be more appropriate for smaller datasets, and vice versa. Training the model and extracting patterns to provide a forecast will be the third stage. Experimental validation is the final stage. Linking visuals with performance is an uphill battle but ultimately rewarding [97].

Many ML models correlate chemical structures with PCE but fail to consider critical factors like miscibility and film morphology. Incorporating the theories of Flory and Huggins could potentially improve ML approaches. Prediction accuracy might also benefit from integrating data derived from advanced characterization techniques such as grazing-incidence small-angle X-ray scattering (GISAXS), atomic force microscopy (AFM), transmission electron microscopy (TEM), and grazing-incidence wide-angle X-ray scattering (GIWAXS) [94]. However, applying ML to these methods presents significant challenges. For example:

- GISAXS: Useful for probing nanoscale morphology with a resolution typically around 1–100 nm. Data processing often involves advanced fitting procedures to extract domain spacing and orientation information.
- AFM: Provides surface morphology details at a resolution of ~1 nm but requires noise reduction techniques to mitigate surface irregularities.
- TEM: Offers atomic to nanoscale resolution (~0.1 nm) but demands complex sample preparation and interpretation.
- GIWAXS: Captures crystallographic information with sub-nanometer resolution, requiring extensive data modeling to distinguish between amorphous and crystalline phases.

Despite the potential of these methods, image-based analysis remains challenging due to the high noise levels, aberrations, and the specialized nature of data compared to everyday photographs [95]. Images generated by these techniques are also highly dependent on physicochemical characteristics, mixing conditions, and experimental settings, necessitating tailored analytical approaches.

Compiling a comprehensive dataset for a single compound under varying experimental conditions is a labor-intensive process. Automated image extraction and sorting systems struggle to account for the variability in experimental conditions and data quality, requiring substantial human intervention. Tasks such as data labeling and target property selection represent a critical intermediate stage in the process.

Given the influence of the active layer's morphology on FF values [96], selecting FF as a primary objective instead of PCE could simplify the modeling process. Correlating FF with other parameters, including PCE, provides additional insights. For smaller datasets, classification methods may be more effective, whereas regression models are better suited for larger datasets. The final stages involve training the model to extract meaningful patterns and making performance forecasts, followed by experimental validation. While linking visual data to performance metrics is an intricate task, it holds significant promise for advancing OSC design [97]. Establishing clear resolution limits and data-processing protocols for GISAXS, AFM, TEM, and GIWAXS is essential for ensuring the reliability and interpretability of ML-based predictions.

6.4. Experimental Validation

The use of ML in OSC research is increasing, as noted in the literature. High-throughput screening is expected to continue progressing. Typically, materials are screened using heuristic rules, but these rules do not guarantee that materials can be synthesized, as their synthesis techniques are not always known. Collaboration with experimental professionals is essential to enhance the accuracy of machine predictions. Once candidates are identified by ML, a manual examination based on synthetic aspects is recommended, followed by experimental validation. However, the number of cases where experiments validate ML predictions is relatively small. Sun et al. confirmed ML findings by synthesizing ten donor materials, with eight compounds correctly categorized by the model [39]. Nagasawa et al. found a PCE of 0.53% in their OSC device fabrication and donor synthesis, significantly lower than the RF forecast of 5.0–5.8% [40]. Wu et al. manufactured six donor/acceptor pairs, with most devices exhibiting a PCE close to the predicted values [64].

In the realm of OSCs, ML has emerged as a pivotal tool for predicting material performance and guiding experimental efforts. While several studies have demonstrated the potential of ML in this field, the translation of predictions into experimental validations remains limited. Zhang et al. synthesized ten donor materials based on ML predictions, with eight compounds correctly categorized by the model, highlighting the promise of ML in guiding material selection [33].

Nagasawa et al. reported a significant discrepancy between ML predictions and experimental outcomes. The fabricated OSC device achieved a PCE of 0.53%. This was notably lower than the 5.0–5.8% range predicted by the random forest model [40]. Wu et al. fabricated six donor–acceptor pairs, with most devices exhibiting PCEs close to the predicted values, demonstrating the potential of ML in accurately forecasting OSC performance [33]. Zhang et al. constructed a database of 397 donor–acceptor pairs and trained various ML models, including random forest and gradient boosting regression trees, to predict PCE. The random forest model exhibited the highest accuracy and stability. Subsequently, they designed 20 non-fullerene acceptor molecules and, based on ML predictions, identified several candidates with predicted PCEs exceeding 12% when paired with P3HT as the donor [33]. Paul et al. developed an ensemble deep neural network architecture, SINet, leveraging both SMILES and InChI molecular representations to predict the highest occupied molecular orbital (HOMO) values of donor molecules. By employing transfer learning from a large dataset, they built robust predictive models applicable to smaller datasets, enhancing the reliability of ML predictions in OSC research [98]. Osterrieder et al. introduced an autonomous optimization platform combining Bayesian optimization with experimental fabrication and characterization. Their system efficiently navigated a four-dimensional parameter space, optimizing the composition and processing conditions of a ternary OSC system. This approach underscores the potential of integrating ML with automated experimentation to accelerate OSC development [99]. These studies underscore the potential of ML in OSC research, particularly when integrated with experimental validation. However, challenges persist in ensuring the accuracy of predictions and their practical applicability. Collaborative efforts between computational scientists and experimentalists are essential to refine ML models and enhance their predictive capabilities, ultimately accelerating the development of high-performance OSCs.

6.5. Development of Better Software

Most current ML technologies require programming skills, limiting their use to individuals with extensive knowledge of data science and computer programming. However, these individuals often lack a deep understanding of the fundamental processes involved. This gap occasionally leads to misinterpretation of results. While OSCs are a hot topic among experimental scientists, they typically lack training in ML. To address this issue, developing user-friendly software with intuitive graphical user interfaces for material specialists is beneficial. This way, experts can harness the full potential of data-driven research without worrying about complex syntax or esoteric tuning settings. Figure 10 shows how ML can play a transformative role in solving the current issues that exist in OSCs by making research and development faster and more efficient. It helps scientists identify the key material properties, molecular features, and fabrication steps that have the biggest impact on how well these solar cells work. By analyzing large datasets, ML can predict which material combinations or manufacturing methods are likely to produce the best performance and device stability, saving time and resources. Beyond that, ML speeds up the discovery of new materials, like better donor/acceptor polymers, by predicting important properties such as energy efficiency and charge movement.

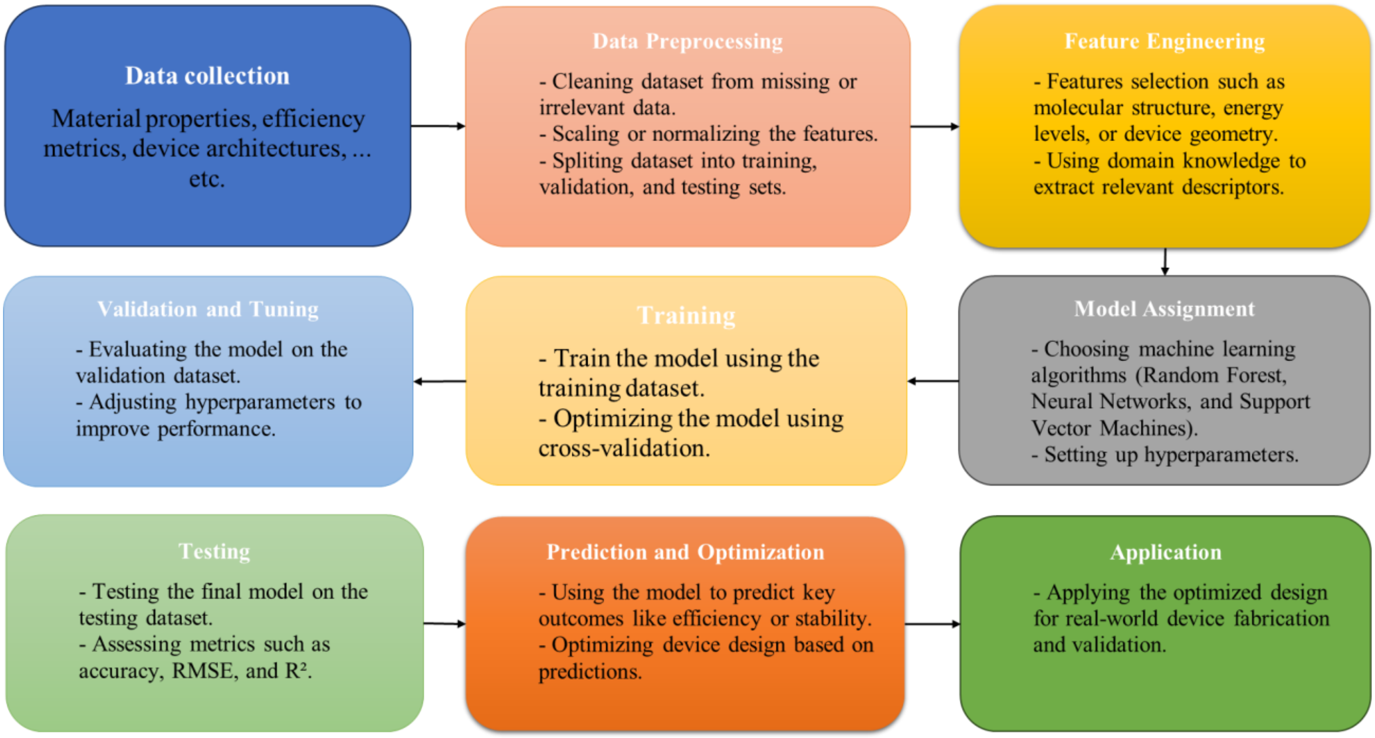

Data collection
Material properties, efficiency metrics, device architectures, … etc.
Data Preprocessing
- Cleaning dataset from missing or irrelevant data.
- Scaling or normalizing the features.
- Spliting dataset into training, validation, and testing sets.
Feature Engineering
- Features selection such as molecular structure, energy levels, or device geometry.
- Using domain knowledge to extract relevant descriptors.
Model Assignment
- Choosing machine learning algorithms (Random Forest, Neural Networks, and Support Vector Machines).
- Setting up hyperparameters.
Training
- Train the model using the training dataset.
- Optimizing the model using cross-validation.
Validation and Tuning
- Evaluating the model on the validation dataset.
- Adjusting hyperparameters to improve performance.
Testing
- Testing the final model on the testing dataset.
- Assessing metrics such as accuracy, RMSE, and R^2.
Prediction and Optimization
- Using the model to predict key outcomes like efficiency or stability.
- Optimizing device design based on predictions.
Application
- Applying the optimized design for real-world device fabrication and validation.

Figure 10. A flowchart revealing the main contribution of ML in tackling the current issues in OSCs.

7. Conclusions

ML has emerged as a powerful tool in the field of OSCs, demonstrating significant potential to predict key parameters such as energy levels, absorption spectra, and PCE. By utilizing diverse inputs like molecular fingerprints, microscopic properties, and simulated features, ML enables researchers to rapidly identify and optimize effective organic semiconductors. However, the progress of ML in OSC research is not without challenges. The quality and diversity of available datasets often limit the reliability of predictions, and the complex operational principles of OSCs further complicate the training of robust models. Despite these obstacles, the growing focus on ML reflects its importance in addressing these issues. Open-source tools and data-sharing initiatives are paving the way for more integrated approaches, helping to overcome data heterogeneity and improve predictive accuracy.

This review successfully highlighted the transformative role of ML in reshaping how OSC materials are discovered and optimized, providing solutions to some of the most persistent challenges in the field. While ML is not yet a replacement for traditional experimental methods, its ability to enhance efficiency and accuracy is undeniable. As technology continues to advance, the integration of ML in OSC research will likely surpass the limitations of trial-and-error approaches, accelerating the journey toward high-performance, cost-effective, and stable solar energy solutions.

Author Contributions: Conceptualization, F.F.M.; methodology, D.R.A.; writing—original draft preparation, D.R.A.; writing—review and editing, F.F.M.; supervision, F.F.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: The authors are grateful to the Department of Physics at Koya University for their administrative support.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ML	Machine Learning
OSCs	Organic Solar Cells
PCE	Power Conversion Efficiency
HOMO	Highest Occupied Molecular Orbital
LUMO	Lowest Unoccupied Molecular Orbital
BHJ	Bulk Heterojunction
HCEP	Harvard Clean Energy Project
DoF	Degree of Freedom
D/A	Donor/Acceptor
DFT	Density Functional Theory
NFAs	Non-Fullerene Acceptors
HPC	High-Performance Computing
PCA	Principal Component Analysis
DA	Discriminant Analysis
ICA	Independent Component Analysis
V_{oc}	Open-circuit Voltage
J_{sc}	Short-circuit Current Density
FF	Fill Factor
RF	Random Forest
LR	Linear Regression
BRT	Boosted Regression Trees
GBRT	Gradient Boosting Regression Tree
ANN	Artificial Neural Networks

k-NN	k-Nearest Neighbors
KRR	Kernel Ridge Regression
SVR	Support Vector Regression
DNN	Deep Neural Network
CNNs	Convolutional Neural Networks
BAE	Bistricyclic Aromatic Compounds
CasSVM	Cascaded Support Vector Machine
MAE	Mean Absolute Error
Si	Silicon
GaAs	Gallium Arsenide
GVA	Grammar Variational Autoencoder
MD	Molecular Dynamics
AFM	Atomic Force Microscopy
TEM	Transmission Electron Microscopy
GIWAXS	Grazing Incidence Wide-Angle X-ray Scattering

References

1. Du, X.; Heumueller, T.; Gruber, W.; Classen, A.; Unruh, T.; Li, N.; Brabec, C.J. Efficient polymer solar cells based on non-fullerene acceptors with potential device lifetime approaching 10 years. *Joule* **2019**, *3*, 215–226. [CrossRef]
2. Wan, X.; Li, C.; Zhang, M.; Chen, Y. Acceptor–donor–acceptor type molecules for high performance organic photovoltaics–chemistry and mechanism. *Chem. Soc. Rev.* **2020**, *49*, 2828–2842. [CrossRef] [PubMed]
3. Zhou, X.; Tang, W.; Bi, P.; Liu, Z.; Lu, W.; Wang, X.; Hao, X.; Wong, W.-K.; Zhu, X. Enhanced light-harvesting of benzodithiophene conjugated porphyrin electron donors in organic solar cells. *J. Mater. Chem. C* **2019**, *7*, 380–386. [CrossRef]
4. Wang, H.; Feng, J.; Dong, Z.; Jin, L.; Li, M.; Yuan, J.; Li, Y. Efficient screening framework for organic solar cells with deep learning and ensemble learning. *npj Comput. Mater.* **2023**, *9*, 200. [CrossRef]
5. Danladi, E.; Dogo, D.S.; Michael, S.U.; Uloko, F.O.; Salawu, A.A.O. Recent advances in modeling of perovskite solar cells using scaps-1d: Effect of absorber and etm thickness. *East Eur. J. Phys.* **2021**, *4*, 5–17. [CrossRef]
6. Hussain, S.S.; Riaz, S.; Nowsherwan, G.A.; Jahangir, K.; Raza, A.; Iqbal, M.J.; Sadiq, I.; Hussain, S.M.; Naseem, S. Numerical Modeling and Optimization of Lead-Free Hybrid Double Perovskite Solar Cell by Using SCAPS-1D. *J. Renew. Energy* **2021**, *2021*, 6668687. [CrossRef]
7. Saha, N.; Brunetti, G.; Armenise, M.N.; Carlo, A.D.; Ciminelli, C. Modeling Highly Efficient Homojunction Perovskite Solar Cells With Graphene-TiO_2 Nanocomposite as the Electron Transport Layer. *IEEE J. Photovolt.* **2023**, *13*, 705–710. [CrossRef]
8. Mahmood, A.; Wang, J.-L. Machine learning for high performance organic solar cells: Current scenario and future prospects. *Energy Environ. Sci.* **2021**, *14*, 90–105. [CrossRef]
9. Wadsworth, A.; Moser, M.; Marks, A.; Little, M.S.; Gasparini, N.; Brabec, C.J.; Baran, D.; McCulloch, I. Critical review of the molecular design progress in non-fullerene electron acceptors towards commercially viable organic solar cells. *Chem. Soc. Rev.* **2019**, *48*, 1596–1625. [CrossRef] [PubMed]
10. Scharber, M.C.; Mühlbacher, D.; Koppe, M.; Denk, P.; Waldauf, C.; Heeger, A.J.; Brabec, C.J. Design rules for donors in bulk-heterojunction solar cells—Towards 10% energy-conversion efficiency. *Adv. Mater.* **2006**, *18*, 789–794. [CrossRef]
11. Padula, D.; Simpson, J.D.; Troisi, A. Combining electronic and structural features in machine learning models to predict organic solar cells properties. *Mater. Horiz.* **2019**, *6*, 343–349. [CrossRef]
12. Gu, G.H.; Noh, J.; Kim, I.; Jung, Y. Machine learning for renewable energy materials. *J. Mater. Chem. A* **2019**, *7*, 17096–17117. [CrossRef]
13. Rodríguez-Martínez, X.; Pascual-San-José, E.; Campoy-Quiles, M. Accelerating organic solar cell material's discovery: High-throughput screening and big data. *Energy Environ. Sci.* **2021**, *14*, 3301–3322. [CrossRef] [PubMed]
14. Ding, P.; Yang, D.; Yang, S.; Ge, Z. Stability of organic solar cells: Toward commercial applications. *Chem. Soc. Rev.* **2024**, *53*, 2350–2387. [CrossRef] [PubMed]
15. Wang, Y.; Luke, J.; Privitera, A.; Rolland, N.; Labanti, C.; Londi, G.; Lemaur, V.; Toolan, D.T.; Sneyd, A.J.; Jeong, S. The critical role of the donor polymer in the stability of high-performance non-fullerene acceptor organic solar cells. *Joule* **2023**, *7*, 810–829. [CrossRef]
16. França, R.P.; Monteiro, A.C.B.; Arthur, R.; Iano, Y. An overview of deep learning in big data, image, and signal processing in the modern digital age. In *Trends in Deep Learning Methodologies*; Academic Press: Cambridge, MA, USA, 2021; pp. 63–87.
17. Dong, S.; Wang, P.; Abbas, K. A survey on deep learning and its applications. *Comput. Sci. Rev.* **2021**, *40*, 100379. [CrossRef]

18. Cova, T.; Pais, A. Deep learning for deep chemistry: Optimizing the prediction of chemical patterns. *Front. Chem.* **2019**, *7*, 809. [CrossRef] [PubMed]
19. Schleder, G.R.; Padilha, A.C.; Acosta, C.M.; Costa, M.; Fazzio, A. From DFT to machine learning: Recent approaches to materials science—A review. *J. Phys. Mater.* **2019**, *2*, 032001. [CrossRef]
20. Zhou, T.; Song, Z.; Sundmacher, K. Big data creates new opportunities for materials research: A review on methods and applications of machine learning for materials design. *Engineering* **2019**, *5*, 1017–1026. [CrossRef]
21. Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2022.
22. Kim, K.G. Book review: Deep learning. *Healthc. Inform. Res.* **2016**, *22*, 351–354. [CrossRef]
23. Pyzer-Knapp, E.O.; Li, K.; Aspuru-Guzik, A. Learning from the harvard clean energy project: The use of neural networks to accelerate materials discovery. *Adv. Funct. Mater.* **2015**, *25*, 6495–6502. [CrossRef]
24. Pyzer-Knapp, E.O.; Pitera, J.W.; Staar, P.W.; Takeda, S.; Laino, T.; Sanders, D.P.; Sexton, J.; Smith, J.R.; Curioni, A. Accelerating materials discovery using artificial intelligence, high performance computing and robotics. *npj Comput. Mater.* **2022**, *8*, 84. [CrossRef]
25. Taffese, W.Z.; Espinosa-Leal, L. Unveiling non-steady chloride migration insights through explainable machine learning. *J. Build. Eng.* **2024**, *82*, 108370. [CrossRef]
26. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
27. Alwadai, N.; Khan, S.U.-D.; Elqahtani, Z.M.; Ud-Din Khan, S. Machine learning assisted prediction of power conversion efficiency of all-small molecule organic solar cells: A data visualization and statistical analysis. *Molecules* **2022**, *27*, 5905. [CrossRef] [PubMed]
28. Hußner, I.; Lazarides, R.; Symes, W.; Richter, E.; Westphal, A. Reflect on your teaching experience: Systematic reflection of teaching behaviour and changes in student teachers' self-efficacy for reflection. *Z. Für Erzieh.* **2023**, *26*, 1301–1320. [CrossRef]
29. Lombardo, D.; Calandra, P.; Pasqua, L.; Magazù, S. Self-assembly of organic nanomaterials and biomaterials: The bottom-up approach for functional nanostructures formation and advanced applications. *Materials* **2020**, *13*, 1048. [CrossRef] [PubMed]
30. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
31. Alvarez-Gonzaga, O.A.; Rodriguez, J.I. Machine learning models with different cheminformatics data sets to forecast the power conversion efficiency of organic solar cells. *arXiv* **2024**, arXiv:2410.23444.
32. Chen, G.; Tang, D.-M. Machine Learning as a "Catalyst" for Advancements in Carbon Nanotube Research. *Nanomaterials* **2024**, *14*, 1688. [CrossRef]
33. Zhang, C.-R.; Li, M.; Zhao, M.; Gong, J.-J.; Liu, X.-M.; Chen, Y.-H.; Liu, Z.-J.; Wu, Y.-Z.; Chen, H.-S. Machine learning study on organic solar cells and virtual screening of designed non-fullerene acceptors. *J. Appl. Phys.* **2023**, *134*, 153104. [CrossRef]
34. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386. [CrossRef] [PubMed]
35. Goodfellow, I. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
36. Ain, A. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666.
37. Sahu, H.; Yang, F.; Ye, X.; Ma, J.; Fang, W.; Ma, H. Designing promising molecules for organic solar cells via machine learning assisted virtual screening. *J. Mater. Chem. A* **2019**, *7*, 17480–17488. [CrossRef]
38. Greenstein, B.L.; Hutchison, G.R. Screening efficient tandem organic solar cells with machine learning and genetic algorithms. *J. Phys. Chem. C* **2023**, *127*, 6179–6191. [CrossRef]
39. Sun, W.; Zheng, Y.; Yang, K.; Zhang, Q.; Shah, A.A.; Wu, Z.; Sun, Y.; Feng, L.; Chen, D.; Xiao, Z. Machine learning–assisted molecular design and efficiency prediction for high-performance organic photovoltaic materials. *Sci. Adv.* **2019**, *5*, eaay4275. [CrossRef] [PubMed]
40. Nagasawa, S.; Al-Naamani, E.; Saeki, A. Computer-aided screening of conjugated polymers for organic solar cell: Classification by random forest. *J. Phys. Chem. Lett.* **2018**, *9*, 2639–2646. [CrossRef]
41. Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Elsevier: Amsterdam, The Netherlands, 2023.
42. Afzal, M.A.F.; Hachmann, J. High-throughput computational studies in catalysis and materials research, and their impact on rational design. In *Handbook on Big Data and Machine Learning in the Physical Sciences: Volume 1. Big Data Methods in Experimental Materials Discovery*; World Scientific: Singapore, 2020; pp. 1–44.
43. Butler, K.T.; Davies, D.W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547–555. [CrossRef] [PubMed]
44. Jain, A.; Ong, S.P.; Hautier, G.; Chen, W.; Richards, W.D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002. [CrossRef]
45. Vo, A.H.; Van Vleet, T.R.; Gupta, R.R.; Liguori, M.J.; Rao, M.S. An overview of machine learning and big data for drug toxicity evaluation. *Chem. Res. Toxicol.* **2019**, *33*, 20–37. [CrossRef] [PubMed]

46. Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361*, 360–365. [CrossRef] [PubMed]
47. Pereira, F.; Xiao, K.; Latino, D.A.; Wu, C.; Zhang, Q.; Aires-de-Sousa, J. Machine learning methods to predict density functional theory B3LYP energies of HOMO and LUMO orbitals. *J. Chem. Inf. Model.* **2017**, *57*, 11–21. [CrossRef] [PubMed]
48. Sui, M.-Y.; Yang, Z.-R.; Geng, Y.; Sun, G.-Y.; Hu, L.; Su, Z.-M. Nonfullerene acceptors for organic photovoltaics: From conformation effect to power conversion efficiencies prediction. *Sol. RRL* **2019**, *3*, 1900258. [CrossRef]
49. Chattopadhyay, J.; Srivastava, N. *Application of Nanomaterials in Chemical Sensors and Biosensors*; CRC Press: Boca Raton, FL, USA, 2021.
50. Omar, Ö.H.; Del Cueto, M.; Nematiaram, T.; Troisi, A. High-throughput virtual screening for organic electronics: A comparative study of alternative strategies. *J. Mater. Chem. C* **2021**, *9*, 13557–13583. [CrossRef]
51. Cereto-Massagué, A.; Ojeda, M.J.; Valls, C.; Mulero, M.; Garcia-Vallvé, S.; Pujadas, G. Molecular fingerprint similarity search in virtual screening. *Methods* **2015**, *71*, 58–63. [CrossRef] [PubMed]
52. Pattanaik, L.; Coley, C.W. Molecular representation: Going long on fingerprints. *Chem* **2020**, *6*, 1204–1207. [CrossRef]
53. Mahmood, A.; Hu, J.-Y.; Xiao, B.; Tang, A.; Wang, X.; Zhou, E. Recent progress in porphyrin-based materials for organic solar cells. *J. Mater. Chem. A* **2018**, *6*, 16769–16797. [CrossRef]
54. Zhang, C.; Song, X.; Liu, K.K.; Zhang, M.; Qu, J.; Yang, C.; Yuan, G.Z.; Mahmood, A.; Liu, F.; He, F. Electron-Deficient and Quinoid Central Unit Engineering for Unfused Ring-Based A1–D–A2–D–A1-Type Acceptor Enables High Performance Nonfullerene Polymer Solar Cells with High Voc and PCE Simultaneously. *Small* **2020**, *16*, 1907681. [CrossRef] [PubMed]
55. Mahmood, A.; Tang, A.; Wang, X.; Zhou, E. First-principles theoretical designing of planar non-fullerene small molecular acceptors for organic solar cells: Manipulation of noncovalent interactions. *Phys. Chem. Chem. Phys.* **2019**, *21*, 2128–2139. [CrossRef] [PubMed]
56. Liu, K.-K.; Xu, X.; Wang, J.-L.; Zhang, C.; Ge, G.-Y.; Zhuang, F.-D.; Zhang, H.-J.; Yang, C.; Peng, Q.; Pei, J. Achieving high-performance non-halogenated nonfullerene acceptor-based organic solar cells with 13.7% efficiency via a synergistic strategy of an indacenodithieno [3, 2-b] selenophene core unit and non-halogenated thiophene-based terminal group. *J. Mater. Chem. A* **2019**, *7*, 24389–24399. [CrossRef]
57. Lopez, S.A.; Sanchez-Lengeling, B.; de Goes Soares, J.; Aspuru-Guzik, A. Design principles and top non-fullerene acceptor candidates for organic photovoltaics. *Joule* **2017**, *1*, 857–870. [CrossRef]
58. Xie, Y.; Wang, W.; Huang, W.; Lin, F.; Li, T.; Liu, S.; Zhan, X.; Liang, Y.; Gao, C.; Wu, H. Assessing the energy offset at the electron donor/acceptor interface in organic solar cells through radiative efficiency measurements. *Energy Environ. Sci.* **2019**, *12*, 3556–3566. [CrossRef]
59. Linderl, T.; Zechel, T.; Brendel, M.; Moseguí González, D.; Müller-Buschbaum, P.; Pflaum, J.; Brütting, W. Energy Losses in Small-Molecule Organic Photovoltaics. *Adv. Energy Mater.* **2017**, *7*, 1700237. [CrossRef]
60. Zhang, J.; Liu, W.; Zhang, M.; Liu, Y.; Zhou, G.; Xu, S.; Zhang, F.; Zhu, H.; Liu, F.; Zhu, X. Revealing the critical role of the HOMO alignment on maximizing current extraction and suppressing energy loss in organic solar cells. *IScience* **2019**, *19*, 883–893. [CrossRef] [PubMed]
61. Jørgensen, P.B.; Mesta, M.; Shil, S.; García Lastra, J.M.; Jacobsen, K.W.; Thygesen, K.S.; Schmidt, M.N. Machine learning-based screening of complex molecules for polymer solar cells. *J. Chem. Phys.* **2018**, *148*, 241735. [CrossRef]
62. Peng, S.-P.; Zhao, Y. Convolutional neural networks for the design and analysis of non-fullerene acceptors. *J. Chem. Inf. Model.* **2019**, *59*, 4993–5001. [CrossRef] [PubMed]
63. Padula, D.; Troisi, A. Concurrent optimization of organic donor–acceptor pairs through machine learning. *Adv. Energy Mater.* **2019**, *9*, 1902463. [CrossRef]
64. Wu, Y.; Guo, J.; Sun, R.; Min, J. Machine learning for accelerating the discovery of high-performance donor/acceptor pairs in non-fullerene organic solar cells. *npj Comput. Mater.* **2020**, *6*, 120. [CrossRef]
65. Li, Y.; Cai, Y.; Xie, Y.; Song, J.; Wu, H.; Tang, Z.; Zhang, J.; Huang, F.; Sun, Y. A facile strategy for third-component selection in non-fullerene acceptor-based ternary organic solar cells. *Energy Environ. Sci.* **2021**, *14*, 5009–5016. [CrossRef]
66. Sun, W.; Li, M.; Li, Y.; Wu, Z.; Sun, Y.; Lu, S.; Xiao, Z.; Zhao, B.; Sun, K. The use of deep learning to fast evaluate organic photovoltaic materials. *Adv. Theory Simul.* **2019**, *2*, 1800116. [CrossRef]
67. Sahu, H.; Rao, W.; Troisi, A.; Ma, H. Toward predicting efficiency of organic solar cells via machine learning and improved descriptors. *Adv. Energy Mater.* **2018**, *8*, 1801032. [CrossRef]
68. Zhao, Z.-W.; del Cueto, M.; Geng, Y.; Troisi, A. Effect of increasing the descriptor set on machine learning prediction of small molecule-based organic solar cells. *Chem. Mater.* **2020**, *32*, 7777–7787. [CrossRef]
69. Ye, L.; Zhao, W.; Li, S.; Mukherjee, S.; Carpenter, J.H.; Awartani, O.; Jiao, X.; Hou, J.; Ade, H. High-efficiency nonfullerene organic solar cells: Critical factors that affect complex multi-length scale morphology and device performance. *Adv. Energy Mater.* **2017**, *7*, 1602000. [CrossRef]

70. Duong, D.T.; Walker, B.; Lin, J.; Kim, C.; Love, J.; Purushothaman, B.; Anthony, J.E.; Nguyen, T.Q. Molecular solubility and hansen solubility parameters for the analysis of phase separation in bulk heterojunctions. *J. Polym. Sci. Part B Polym. Phys.* **2012**, *50*, 1405–1413. [CrossRef]
71. Perea, J.D.; Langner, S.; Salvador, M.; Sanchez-Lengeling, B.; Li, N.; Zhang, C.; Jarvas, G.; Kontos, J.; Dallos, A.; Aspuru-Guzik, A. Introducing a new potential figure of merit for evaluating microstructure stability in photovoltaic polymer-fullerene blends. *J. Phys. Chem. C* **2017**, *121*, 18153–18161. [CrossRef]
72. Yuan, J.; Zhang, H.; Zhang, R.; Wang, Y.; Hou, J.; Leclerc, M.; Zhan, X.; Huang, F.; Gao, F.; Zou, Y. Reducing voltage losses in the A-DA′ DA acceptor-based organic solar cells. *Chem* **2020**, *6*, 2147–2161. [CrossRef]
73. Hachmann, J.; Olivares-Amaya, R.; Jinich, A.; Appleton, A.L.; Blood-Forsythe, M.A.; Seress, L.R.; Román-Salgado, C.; Trepte, K.; Atahan-Evrenk, S.; Er, S. Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry–the Harvard Clean Energy Project. *Energy Environ. Sci.* **2014**, *7*, 698–704. [CrossRef]
74. Imamura, Y.; Tashiro, M.; Katouda, M.; Hada, M. Automatic high-throughput screening scheme for organic photovoltaics: Estimating the orbital energies of polymers from oligomers and evaluating the photovoltaic characteristics. *J. Phys. Chem. C* **2017**, *121*, 28275–28286. [CrossRef]
75. Pyzer-Knapp, E.O.; Simm, G.N.; Guzik, A.A. A Bayesian approach to calibrating high-throughput virtual screening results and application to organic photovoltaic materials. *Mater. Horiz.* **2016**, *3*, 226–233. [CrossRef]
76. Lee, M.-H. Robust random forest based non-fullerene organic solar cells efficiency prediction. *Org. Electron.* **2020**, *76*, 105465. [CrossRef]
77. Yue, Q.; Liu, W.; Zhu, X. n-Type molecular photovoltaic materials: Design strategies and device applications. *J. Am. Chem. Soc.* **2020**, *142*, 11613–11628. [CrossRef] [PubMed]
78. Gao, J.; Gao, W.; Ma, X.; Hu, Z.; Xu, C.; Wang, X.; An, Q.; Yang, C.; Zhang, X.; Zhang, F. Over 14.5% efficiency and 71.6% fill factor of ternary organic solar cells with 300 nm thick active layers. *Energy Environ. Sci.* **2020**, *13*, 958–967. [CrossRef]
79. Liu, T.; Ma, R.; Luo, Z.; Guo, Y.; Zhang, G.; Xiao, Y.; Yang, T.; Chen, Y.; Li, G.; Yi, Y. Concurrent improvement in J sc and V oc in high-efficiency ternary organic solar cells enabled by a red-absorbing small-molecule acceptor with a high LUMO level. *Energy Environ. Sci.* **2020**, *13*, 2115–2123. [CrossRef]
80. Lee, M.H. Insights from machine learning techniques for predicting the efficiency of fullerene derivatives-based ternary organic solar cells at ternary blend design. *Adv. Energy Mater.* **2019**, *9*, 1900891. [CrossRef]
81. Lee, M.-H. A Machine Learning–Based Design Rule for Improved Open-Circuit Voltage in Ternary Organic Solar Cells. *Adv. Intell. Syst.* **2020**, *2*, 1900108. [CrossRef]
82. Zhou, Z.; Xu, S.; Song, J.; Jin, Y.; Yue, Q.; Qian, Y.; Liu, F.; Zhang, F.; Zhu, X. High-efficiency small-molecule ternary solar cells with a hierarchical morphology enabled by synergizing fullerene and non-fullerene acceptors. *Nat. Energy* **2018**, *3*, 952–959. [CrossRef]
83. Liu, G.; Jia, J.; Zhang, K.; Jia, X.e.; Yin, Q.; Zhong, W.; Li, L.; Huang, F.; Cao, Y. 15% efficiency tandem organic solar cell based on a novel highly efficient wide-bandgap nonfullerene acceptor with low energy loss. *Adv. Energy Mater.* **2019**, *9*, 1803657. [CrossRef]
84. Lee, M.-H. Performance and matching band structure analysis of tandem organic solar cells using machine learning approaches. *Energy Technol.* **2020**, *8*, 1900974. [CrossRef]
85. Du, P.; Zebrowski, A.; Zola, J.; Ganapathysubramanian, B.; Wodo, O. Microstructure design using graphs. *npj Comput. Mater.* **2018**, *4*, 50. [CrossRef]
86. Pfeifer, S.; Pokuri, B.S.S.; Du, P.; Ganapathysubramanian, B. Process optimization for microstructure-dependent properties in thin film organic electronics. *Mater. Discov.* **2018**, *11*, 6–13. [CrossRef]
87. Noruzi, R.; Ghadai, S.; Bingol, O.R.; Krishnamurthy, A.; Ganapathysubramanian, B. NURBS-based microstructure design for organic photovoltaics. *Comput. -Aided Des.* **2020**, *118*, 102771. [CrossRef]
88. Wodo, O.; Tirthapura, S.; Chaudhary, S.; Ganapathysubramanian, B. A graph-based formulation for computational characterization of bulk heterojunction morphology. *Org. Electron.* **2012**, *13*, 1105–1113. [CrossRef]
89. Pokuri, B.S.S.; Ghosal, S.; Kokate, A.; Sarkar, S.; Ganapathysubramanian, B. Interpretable deep learning for guided microstructure-property explorations in photovoltaics. *npj Comput. Mater.* **2019**, *5*, 95. [CrossRef]
90. Wodo, O.; Ganapathysubramanian, B. Modeling morphology evolution during solvent-based fabrication of organic solar cells. *Comput. Mater. Sci.* **2012**, *55*, 113–126. [CrossRef]
91. Kodali, H.K.; Ganapathysubramanian, B. Computer simulation of heterogeneous polymer photovoltaic devices. *Model. Simul. Mater. Sci. Eng.* **2012**, *20*, 035015. [CrossRef]
92. Majeed, N.; Saladina, M.; Krompiec, M.; Greedy, S.; Deibel, C.; MacKenzie, R.C. Using deep machine learning to understand the physical performance bottlenecks in novel thin-film solar cells. *Adv. Funct. Mater.* **2020**, *30*, 1907259. [CrossRef]
93. Zhang, Y.; Ling, C. A strategy to apply machine learning to small datasets in materials science. *npj Comput. Mater.* **2018**, *4*, 25. [CrossRef]
94. Mahmood, A.; Wang, J.L. A review of grazing incidence small-and wide-angle x-ray scattering techniques for exploring the film morphology of organic solar cells. *Sol. RRL* **2020**, *4*, 2000337. [CrossRef]

95. Jones, L.; Nellist, P.D. Identifying and correcting scan noise and drift in the scanning transmission electron microscope. *Microsc. Microanal.* **2013**, *19*, 1050–1060. [CrossRef] [PubMed]
96. Zawodzki, M.; Resel, R.; Sferrazza, M.; Kettner, O.; Friedel, B. Interfacial morphology and effects on device performance of organic bilayer heterojunction solar cells. *ACS Appl. Mater. Interfaces* **2015**, *7*, 16161–16168. [CrossRef] [PubMed]
97. Pokuri, B.S.S.; Stimes, J.; O'Hara, K.; Chabinyc, M.L.; Ganapathysubramanian, B. GRATE: A framework and software for GRaph based Analysis of Transmission Electron Microscopy images of polymer films. *Comput. Mater. Sci.* **2019**, *163*, 1–10. [CrossRef]
98. Paul, A.; Jha, D.; Al-Bahrani, R.; Liao, W.-k.; Choudhary, A.; Agrawal, A. Transfer learning using ensemble neural networks for organic solar cell screening. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8.
99. Osterrieder, T.; Schmitt, F.; Lüer, L.; Wagner, J.; Heumüller, T.; Hauch, J.; Brabec, C.J. Autonomous optimization of an organic solar cell in a 4-dimensional parameter space. *Energy Environ. Sci.* **2023**, *16*, 3984–3993. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.