IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 June 2023, accepted 21 July 2023, date of publication 24 July 2023, date of current version 1 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3298542

SURVEY

A Comprehensive Review of the Application of Machine Learning in Fabrication and Implementation of Photovoltaic Systems

SRABANTI DATTA[1], ANIK BAUL[2], GOBINDA CHANDRA SARKER[3], PINTU KUMAR SADHU[2], AND DEIDRA R. HODGES[1], (Senior Member, IEEE)

[1]Department of Electrical and Computer Engineering, Florida International University, Miami, FL 33174, USA
[2]College of Science and Engineering, Central Michigan University, Mount Pleasant, MI 48858, USA
[3]Department of Electrical and Electronic Engineering, Mymensingh Engineering College, Mymensingh 2200, Bangladesh

Corresponding author: Srabanti Datta (sdatt006@fiu.edu)

ABSTRACT Solar energy is a promising source of renewable energy, but its low efficiency, instability, and high manufacturing costs remain a big challenge. Recently, machine learning (ML) techniques have gained attention in the photovoltaic (PV) sector because of advances in computer power, tools, and data creation. The existing ML techniques used for fabrication and the different operational procedures of the PV sector have shown very impressive results with a higher degree of accuracy and precision. While previous studies have discussed ML techniques for PV fabrication or operational procedures, there is a lack of end-to-end research that covers the entire process from fabrication to implementation. In this paper, we present a comprehensive review of the application of ML in the field of solar energy, focusing on the development of new materials, enhancement of solar cell efficiency, implementation, and integration with the system, including fault detection, sizing, control, forecasting, management, and site adaptation. We evaluated more than 100 research articles, a significant proportion of which were published in the past three years. In our study investigating ML implementation in solar cell fabrication, we discovered that the Random Forest (RF), Linear Regression (LR), XGBoost, and Artificial Neural Network (ANN) algorithms are the most commonly employed techniques. Our findings demonstrate that XGBoost exhibits superior performance in optoelectronic prediction, while RF, LR, and ANN algorithms are better suited for predicting electrical parameters. Moreover, our analysis indicates recent ML research in this field explicitly emphasizes perovskite solar cells (PSCs). This work also discusses the challenges, directions, insights, and potential applications of ML for future PV system analysis.

INDEX TERMS Deep learning, machine learning, material discovery, fabrication of PV cells, perovskite solar cell, photovoltaic systems.

NOMENCLATURE

ANN	Artificial Neural Network.
ARC	Anti Reflective Coating.
ARMA	Auto Regressive Moving Average.
BORT	Bayesian Optimization based Regression Tree.
CBM	Conduction Band Minimum.
CNN	Convolutional Neural Network.
DT	Decision Tree.
EMS	Energy Management Systems.
EWT	Empirical Wavelet Transform.
FF	Fill Factor.
GAMA	General Automated Machine Learning Assistant.
GBDT	Gradient Boosted Decision Tree.
GPR	Gaussian Process Regression.
ICMMD	Intra Class Maximum Mean Discrepancy.
J_{sc}	Short Circuit Current.
KNN	K-Nearest Neighbors.
LR	Linear Regression.
LSTM	Long Short-term Memory.
LVRT	Low Voltage Ride Through.

The associate editor coordinating the review of this manuscript and approving it for publication was Liangxiu Han.

MAPE	Mean Absolute Percentage Error.
ML	Machine Learning.
MLP	Multilayer Perceptron.
MPPT	Maximum Power Point Tracking.
PCE	Power Conversion Efficiency.
PLSR	Partial least squares regression.
PSC	Perovskite Solar Cell.
PV	Photovoltaic.
PVS	Photovoltaic System.
RF	Random Forest.
RMSE	Root Mean Square Error.
SHAP	Shapley Additive Explanation.
SVD	Singular Value Decomposition.
SVM	Support Vector Machine.
SVR	Support Vector Regression.
TL	Transfer Learning.
TWSVM	Twin Support Vector Machine.
V_{oc}	Open Circuit Voltage.
VBM	Valence Band Maximum.
VMD	Variational Mode Decomposition.
XGB	Extreme Gradient Boost.

I. INTRODUCTION

The annual increase in global energy demand, driven by a growing population, necessitates a corresponding increase in energy production [1]. Traditional power generation infrastructure based on fossil fuels leaves a harmful environmental footprint. As a result, governments and decision-makers are increasingly inclined to provide funds for the research and development of renewable energy technologies [2]. Recently, PV energy generation has become a popular and highly regarded renewable energy source [3], [4]. Researchers have taken a keen interest in PV solar cells because of their ability to harness electrical energy directly from solar radiation [5]. A unit PV cell captures photons from the sunlight incident upon it and produces electron-hole pairs (EHP), which gives rise to photo-electric current as output. Individual PV cells are connected together to make a solar panel, which can either be used as a standalone unit or combined with other panels to build a large power station. Finding proper PV materials, optimizing the device structure, and developing the fabrication techniques are the three main aspects of designing a standard solar cell [6]. Figure 1 depicts various solar cells divided into different generations based on their technology and developmental stages [7].

In 1954, a team of researchers from US Bell Labs led by Pearson developed the initial crystalline silicon solar cell, which had a power conversion efficiency (PCE) of 4.5%. This incident marked a significant milestone in the application of solar electricity [8]. The PCE of modern mono-crystalline/poly-crystalline silicon solar cells in industrial settings has exceeded 20% [9], [10]. However, such silicon-based solar cells are relatively expensive to manufacture, fragile, easily damaged, pollute the environment significantly and have limited temperature tolerance. Hence, while silicon-based solar cells have been widely used and have proven to be effective, newer technologies such as perovskite solar cells (PSCs) and thin-film solar cells are being developed to address these disadvantages. Thin-film solar cells such as Cadmium Telluride (CdTe), Copper Indium Gallium Selenium (CIGS) have reached high PCE in the laboratory, but the industrial applications are limited by high production costs and environmental contamination [11], [12], [13]. The issue of high manufacturing costs can be resolved by third-generation PSCs without compromising efficiency.

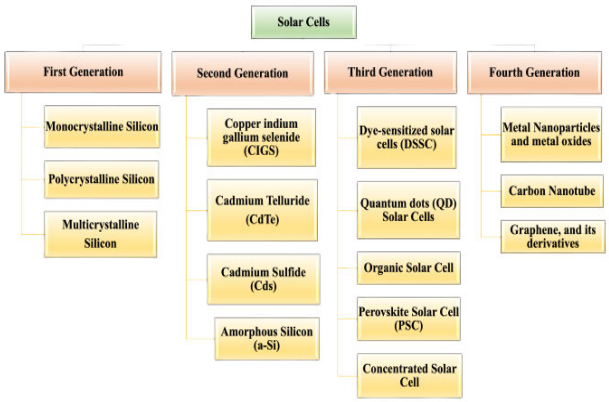

FIGURE 1. Different type of solar cells.

The fabrication of efficient and stable solar cells is fraught with various obstacles. A prevalent evaluation metric for PV technology is the PCE, which measures the ratio of solar energy input to electrical energy output. The quality of materials used in the fabrication of solar cells significantly impacts their efficiency. The performance of solar cells can be affected by the purity, crystal structure, and defect density of the materials used. Obtaining high-quality materials can be difficult and expensive, particularly for newer and more exotic materials. Additionally, the degradation behavior and environmental stability of solar cell materials are yet to fully understand, making it challenging to predict their performance under different environmental conditions, such as temperature and moisture [14].

Trial and error are time-consuming, laborious, and expensive, and it may not be feasible to test all materials and fabrication conditions. Hence ML is an emerging technology that has the potential to revolutionize the field of solar cell research [15], [16]. By using ML algorithms, it is easy to analyze large amounts of data and make predictions about the properties of photovoltaic materials, which can help improve the efficiency and stability of solar cells [17]. The application of ML in solar cell research can speed up the discovery of new suitable materials. ML can be used to analyze large amounts of data from the fabrication process, such as temperature, humidity, and pressure, and make predictions about how to optimize the process [18]. This can help to improve the efficiency and stability of solar cells by reducing defects and upgrading the uniformity of materials. Moreover, ML

algorithms can be utilized to predict the optical properties of materials, such as bandgap, absorption coefficient, and carrier mobility [19], [20]. This information can be used to optimize the design of solar cells, which can help to improve their efficiency and stability.

After the PV cells are fabricated, they are evaluated for their effectiveness and power output. The cells are sorted and assembled into modules if they meet the necessary requirements. Each module is composed of multiple PV cells that absorb electromagnetic radiation. This radiation is derived from the sun's conversion of gravitational potential energy into electromagnetic radiation as the sun acts as a gravitational energy transducer rather than a direct energy generator [21]. These modules can then be used to build solar arrays for power generation. The arrays may be mounted atop structures, within utility-scale power facilities, on the rooftop of a building, or in other suitable places where sunlight can easily reach. The produced electricity is transferred to the grid through an inverter for consumers. In many regions of the world, numerous methods exist to increase the installed PV capacity [22]. However, PV system installation still entails high costs and problems with performance that must be fixed quickly. There are still efforts being made to lower the costs of installation while also improving the effectiveness and connecting them to electrical grids. Maintenance of the modules is essential to ensure optimal output and enhance security and reliability [23]. PV systems require constant supervision since solar radiation is susceptible to unpredictable fluctuations and uncertainties [1]. Therefore, it is very important to discuss the challenges related to the operation of PV infrastructure. ML algorithms have emerged as a replacement to the conventional approaches for the solution of these challenges to enhance system efficiency, reliability, and cost-effectiveness of PV systems, similar to the fabrication process. Some of the widely researched application areas of ML in both sides are illustrated in Figure 2.

Numerous studies have been conducted to evaluate the implementation of different ML techniques addressing a range of issues related to photovoltaic system operation, control, and diagnosis. In comparison, some papers reviewed the use of ML in the fabrication stage. Hence, there is a lack of comprehensive survey that provides an end-to-end picture of the application of ML in photovoltaic research. Our contributions are-

- To our knowledge, this is the first review paper that depicts a complete picture of how ML can be utilized across the entire PV infrastructure, starting from fabrication to system operations.
- Our research considered every generation of PV technologies that, include PSCs, Silicon-based, and thin film solar cells. In contrast, previous studies in the literature mostly focused on either PSCs or Silicon-based solar cells.
- This study reviews recent ML practices in four key areas during PV cell fabrication-
  - Optimization of electrical parameters
  - Exploration of new materials for PV applications
  - Optimization of optoelectronic parameters
  - Stability analysis
- ML application in the six most critical matters relating to the operation and maintenance of solar cells are presented in this study.
  - Fault detection in PV
  - Irradiance forecasting and PV output power estimation
  - Control methodology of PV
  - PV sizing system
  - Management
  - Site selection for PV installation

The article is organized in the following manner. In section II, we have provided brief descriptions of the most commonly used ML in the field of photovoltaics. After that, in section III, some recent developments of ML application in PV fabrication are reviewed, followed by the literature review of other conventional solar cell trends regarding AI in recent years in section IV. A comprehensive discussion regarding the recent trend of ML and DL algorithms in PV operation and maintenance is provided in section V. Lastly, in section VI based on the literature reviews in this study, we have suggested some future research directions and scope of utilizing ML models in this field before concluding in section VII

II. MACHINE LEARNING MODELS

Both regression and classification tasks are achieved using the most efficient ML algorithms in every aspect of the solar cell. In the fabrication stages, the prediction of PCE, bandgap, conduction band, valence band, doping concentration, and selection of optimal materials are of significant importance. ML approaches such as Random Forest(RF), Decision Tree (DT), XGBoost (XGB), Gradient Boosted Decision Tree (GBDT) are most commonly utilized for these tasks. The use of deep learning is limited by the lack of a large dataset compared to numerous features. On the other hand, in the PV system implementation stage, Deep Learning based models such as Convolutional Neural Network (CNN), Long Short Term Memory (LSTM), and hybrid architectures are more popular due to the availability of large training samples. This section summarizes some of the most commonly used ML techniques in literature for PV research.

A. DECISION TREE (DT)

DT is one of the most simple ML algorithms. Although it is mostly popular for the classification task, it is widely applied in regression as well [24]. A decision tree consists of roots, branches, and leaves. Each branch makes a decision regarding whether a statement is true or not, and the final output is taken from the leaves. The most important requirement for DT is finding the optimum tree structure. This is achieved by selecting a splitting criterion in root and branches that best separates the input samples. This is also known as attribute

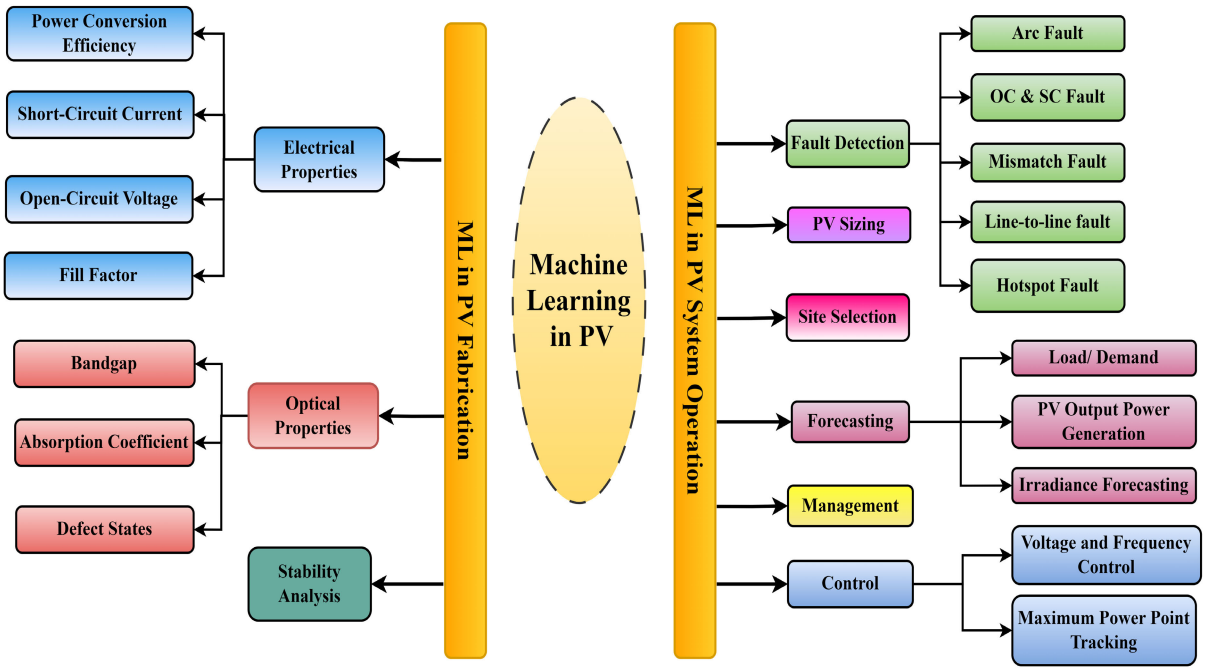

FIGURE 2. ML Applications in PV fabrication and operation.

selection measure (ASM). The most commonly used methods for ASM are information gain and gini index, which are calculated as stated in equations (1) and (3) respectively. For a node, the entropy before splitting is denoted as $Entropy(S)$, where S is the number of samples at the current node, and S_v is the number of samples of class v. Entropy for N number of features is determined by equation (2), where P_j refers to the probability of randomly selecting an element of class j. DT, based on information gain, tries to maximize this value; in contrast, gini scores are minimized [25].

$$Information\ gain = Entropy(S) - \sum(\frac{|S_v|}{|S|}Entropy(S_v)) \tag{1}$$

$$where,\ Entropy(s) = -\sum_{j=1}^{N} -P_j \times log_2(P_j) \tag{2}$$

$$gini = 1 - \sum_{j=1}^{N} p_j^2 \tag{3}$$

B. RANDOM FOREST (RF)

RF is a popular ML method that addresses the high variance of DT. It is an ensemble of hundreds of unpruned independent decision trees which are fitted on training data with the bagging method in order to capture non-linear relationships [26]. The learning process starts with creating a bootstrapped dataset by sampling randomly from the original one. Each observation can be selected multiple times. Usually, two-thirds of the training data ends up in the bootstrapped dataset, while the rest is called an ''out-of-bag'' (OOB) set and is used for evaluation purposes. If the input has a total of m number of features, then each decision tree is made by arbitrarily selecting a subset of k from m features (k ϵm). Each tree is made to grow to its full extent. For the regression analysis task, the final output $\hat{y}(x)$ is the average prediction of all trees as shown in equation (4). Where the number of trees is denoted by K and $h_i(x)$ is the output prediction if ith tree. In the case of the classification task, RF uses the majority vote of all trees as its final output.

$$\hat{y}(x) = \frac{1}{K} \times \sum_{i=1}^{K} h_i(x) \tag{4}$$

C. GRADIENT BOOSTED DECISION TREE (GBDT)

GBDT is another tree-based ensemble ML algorithm where hundreds of DTs are used as a base learner [27]. These trees are built sequentially, and each tree emphasizes the miss-classification of the previous tree. GBDT is a very popular method that is widely used in regression and classification tasks. Given a set of n input observations, (x_i, y_i) where i equals to $(1, 2, \ldots, n)$ and a differential loss function $\mathcal{L}$, GBDT first start with making an initial prediction as shown in equation (5). This equation attempts to determine an initial value γ that minimizes the sum of the loss function over all observations before initiating the training iterations.

$$F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{N} \mathcal{L}(y_i, \gamma_i) \tag{5}$$

$$r_{im} = -[\frac{\partial \mathcal{L}(y_i, F(x_i))}{\partial F(x_i)}]_{f_{(x)}=f_{m-1}(x)} \tag{6}$$

$$\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \epsilon R_{ij}} \mathcal{L}(y_i, F_{m-1}(x_i) + \gamma) \tag{7}$$

$$F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J} r_{jm} I(x \epsilon R_{jm}) \tag{8}$$

M denotes the number of trees. Then, for any tree m in the training sequence, pseudo residuals r_{im} are calculated by equation (6) over all input samples. Next, the tree is fitted to predict the residuals. The tree structure might be limited depending on the user-defined parameter called maximum depth. The output of each leaf in the tree is determined by equation (7). The prediction output of the current tree is updated based on the output of previous tree and learning rate ν multiplied by the leaf output as shown in equation (8). This process is repeated until the convergence conditions are met.

D. XGboost (XGB)

XGBoost (XGB) [28] is a powerful ML model that consists of an ensemble of decision trees. The structure of these trees is selected using a quality score that is similar to the impurity score used by other tree-based models according to greedy algorithms. It is a type of boosting algorithm, meaning each tree is developed based on the performance of the previous tree by minimizing the objective function described in equation (9). Each new tree results in smaller residuals.

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \mathcal{L}(y_i, \hat{y}i) + \sum_{j=1}^{T} \Omega(f_j) \tag{9}$$

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{10}$$

In equation (9), $\mathcal{L}$ is the loss function that measures the difference between the predicted value $\hat{y}_i$ and the actual label y_i where f_j is the j-th tree in the ensemble. XGB differs from the traditional gradient tree boosting techniques by utilizing various regularization techniques, which are represented by Ω term. The purpose of regularization is to reduce the variance of the model by introducing a small bias. This decreases the dependency on observations and helps to make accurate predictions in the long term. The number of leaf nodes is denoted by T. In equation (10), γ encourages pruning and reduces complexity in the ensemble trees, w_j is the leaf weight, and λ is another user-defined parameter that controls the magnitude of the penalty introduced by the regularization. The final output of the XGB model is determined by aggregating the output of all trees. XGBoost has become popular due to its fast performance and ability to handle a large number of features, missing values, and noisy data. The authors [28] also introduced features such as sparsity-aware split finding, parallel learning, etc., for further optimization on a larger dataset.

E. MULTILAYER PERCEPTRON NETWORK (MLP)

An MLP is a feed-forward neural network consisting of multiple layers of nodes, where the output of one layer serves as the input to the next. The basic building block of an MLP is the artificial neuron, also called perceptron [29]. Each perceptron in a layer receives inputs from the previous layer, performs a weighted sum of these inputs, and then applies an activation function to generate its output. The activation function introduces non-linearity, allowing it to model complex relationships between inputs and outputs. The number of neurons in each layer and the number of layers themselves are design parameters that can be adjusted to control the capacity of the network. The parameters of an MLP, including the weights and biases of each neuron, are learned from the training data using an optimization algorithm such as backpropagation. This algorithm uses stochastic gradient descent to minimize a loss function that measures the difference between the network's predicted outputs and the true outputs.

$$p = \phi(\sum_{j} w_j x_j + b) \tag{11}$$

Equation (11) shows the mathematical representation of a single perceptron unit p, where the inputs to the unit, weights, and biases are denoted by x_j, w_j and b, respectively. ϕ is the activation function. If we consider L hidden layers, then we can write the MLP network as shown in equation (12). Where x and y represent the input and output of the model. the output of any layer l is denoted by $h^{(l)}$. $W^{(l)}$ is the weight matrix of that particular layer. Different layers can have different activation function, which is denoted by $\phi^{(l)}$.

$$\begin{aligned} h^{(i)} &= \phi^{(i)}(W^{(i)}x + b^{(i)}) \\ h^{(l)} &= \phi^{(l)}(W^{(l)}h^{(l-1)} + b^{(l)}) \\ y &= \phi^{(f)}(W^{(f)}h^{(l)} + b^{(f)}) \end{aligned} \tag{12}$$

F. LONG SHORT-TERM MEMORY (LSTM)

Long Short-Term Memory (LSTM) is a special type of recurrent neural network (RNN) architecture designed to address the problem of vanishing gradients [30], [31] in traditional RNNs, which can make it difficult to learn long-term dependencies in sequential data. LSTMs introduce a memory cell state that allows them to carry distant past information across many timesteps. It operates utilizing three gating mechanisms that allow the network to selectively forget, remember or update information from previous time steps. Equations (13)-(18) describe an LSTM unit for a single time instance [32].

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{13}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{14}$$

$$\tilde{C} = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{15}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C} \tag{16}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{17}$$

$$h_t = o_t * \tanh(C_t) \tag{18}$$

where x_t, h_{t-1} represents the input sequence component of the current timestep and the hidden state from the previous timestep, respectively. They forget input and output gates are

denoted by f_t,i_t, and O_t, respectively. The output of these gates is taken through the application of a sigmoid activation function that outputs between 0 and 1. The forget gate f_t controls how much of the previous cell state should be forgotten. Then, the current cell state C_t is updated by i_t along with $\tilde{C}$, where $\tilde{C}_t$ is the candidate cell state. W_f,W_i,W_c, and W_o denote the weight matrices, which are used to optimize gate behavior. These matrices are learned during the training of LSTM on input samples. b_f,b_i and b_o are called biases.

G. CONVOLUTIONAL NEURAL NETWORK (CNN)

CNNs are one of the most popular and efficient classes of DL algorithms, most commonly used for computer vision tasks. Due to its ability to automatically learn hierarchical representations and local patterns of the input data, it is also applied in tasks such as fault detection, PV load, irradiance forecasting, etc. CNN works by performing convolution and pooling operations over the input sequence. Depending on the type of input, CNN can be 1-D, 2-D, or 3-D. The convolution operation extracts local features from the input image, while the pooling operation reduces the dimensionality of the features and makes them more robust to small spatial variations. Pooling downsamples the feature map by taking the maximum or average value within each local neighborhood which is a user parameter. The learning ability of CNN can be increased by stacking multiple CNN and pooling layers on top of each other similar to MLP architecture. For instance, if we have L convolutional layers, then the first convolution operation of multivariate input sequence z can be written as shown in equation (19) [32].

$$C_{1,j} = ReLU(Z^* h_{1,j} + b_{1,j}) \tag{19}$$

here, (*) represents the convolution operation. The output feature map of the first convolutional layer is denoted by $C_{1,j}$, where $h_{1,j}$ is the J^{th} convolution kernel and $b_{1,j}$ is the bias term. The activation function shown in the equation is a rectified linear unit (ReLU) to adopt non-linearity in the input. The output feature map is then fed into a subsequent convolutional layer, where it is further processed to extract higher-level features. Thus, the I^{th} convolutional layer C_I producing output feature space J^{th} can be formulated as equation (20).

$$C_{I,j} = ReLU(C^*_{I-1,j} h_{I,j} + b_{I,j}) \tag{20}$$

III. ML IN PEROVSKITE SOLAR CELL FABRICATION

In recent years PSCs have attracted significant attention because of their low cost [5], high absorption coefficient [33], [34], tunable bandgap [35], long carrier lifetime and diffusion length [36], [37], high defect tolerance, low exciton energy, and high carrier mobility [38]. The configuration of PSCs is illustrated in Figure 3, which consists of an absorber layer, also referred to as the perovskite layer. This layer is positioned between the hole transport layer (HTL) and the electron transport layer (ETL). Researchers have made significant strides in improving the PCE of PSCs, which has increased from 3.8 % to 25.7 % because of extensive research efforts and the superior properties of these cells [39], [40]. This rapid advancement has made PSCs a practical and attractive alternative to traditional solar technology [41], [42].

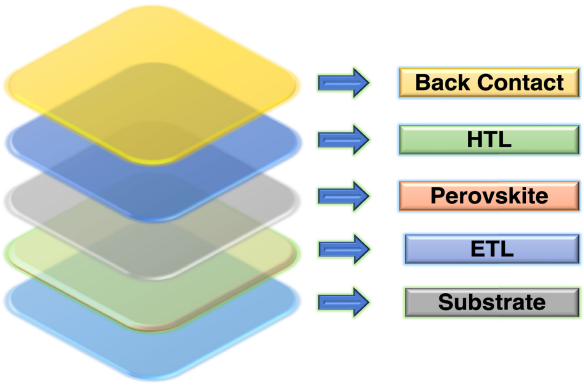

FIGURE 3. Structure of PSC.

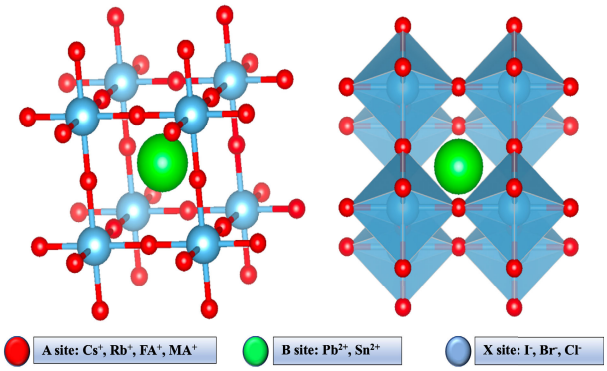

FIGURE 4. Crystal structure of perovskite material.

While PSCs have become very efficient within a relatively short period, they still face several obstacles before becoming viable commercial technology. The perovskite material has the general molecular structure of ABX_3 where A is an organic or inorganic monovalent cation, B is a divalent metal ion, and X is a halogen anion [43], [44]. The crystal structure of perovskite material is depicted in Figure 4 Cesium (Cs^+), Rubidium (Rb^+), Methylammonium ($MA^+/CH^3NH_3^+$), Formamidinium ($FA^+/NH_2CHNH_2^+$) are commonly used at A site; lead (Pb^+), Tin (Sn^+) are usually used at B site; and Iodide (I^-), Bromide (Br^-), Chloride (Cl^-)are normally used at X site [39]. In essence, there are infinite combinations and possibilities for perovskites in solar cell applications. One of the most notable characteristics of perovskite materials is their tunable bandgap, which allows them to absorb a wide range of photons. By changing the composition of the perovskite, it is possible to adjust its bandgap, which makes it possible to tune the absorption properties of the material. To achieve the highest efficiency in PSCs, it is crucial to

find a perovskite material with an optimal band gap that matches the solar spectrum. Perovskites have the ability to adjust their bandgap within a range of 1.5 eV to 3.2 eV. However, the most efficient band gap for a perovskite solar cell is around 1.6 eV. This particular band gap allows for efficient absorption of a wide range of photons present in the solar spectrum [43]. Therefore, it is important to tune the composition of perovskites to optimize their properties. However, relying solely on trial and error for such a complex scenario is neither realistic nor logical.

The structural instability of PSCs is another major challenge that poses a significant obstacle to their commercialization, as it can cause degradation of the cells over time. This instability can be caused by a variety of factors, including temperature, humidity, and exposure to light. The efficiency and lifespan of solar cells can be negatively impacted by the deterioration of perovskite materials. Another issue that contributes to the instability of PSCs is the use of Pb, which is a toxic substance [45] and can potentially harm the environment [46]. The use of Pb in PSCs is a concern because it can leach out of the cells over time and contaminate the surrounding environment [47]. To address these instability issues, researchers have been working on developing new perovskite materials that are more stable and less toxic [48], [49]. This includes developing lead-free perovskite materials, using other materials to stabilize the perovskite structure, and using encapsulation methods to protect the solar cells from the environment.

In recent years, ML techniques have been used to predict the bandgap, PCE, and stability of PSCs with high accuracy, which can significantly accelerate the development and optimization of PSCs. The implementation of ML in the fabricaontion process of PSCs has been demonstrated in Figure 5. Numerous investigations have shown that the predictive ability of ML methods can be utilized to effectively estimate the bandgap, PCE, and stability of PSCs. This section presents a concise summary of the application of ML techniques in the fabrication of PSCs.

A. ML IN THE PREDICTION OF ELECTRICAL AND OPTOELECTRONICS PROPERTIES

Lu et al. [50] proposed that using a combination of ML, experimental, and density functional theory (DFT) methods in a pipeline can give reliable and accurate insights into the physical laws that govern the preparation of perovskite solar cells (PSCs). Their ML model and Shapley Additive Explanations (SHAP) showed that A-site cations have the most significant impact on the PCE, and regulating cation components can enhance crystallization and reduce defects in PSCs.

Y. Liu and colleagues [51] developed and evaluated a total of 49 ML models using 814 real experimental data from published literature. These models utilized 7 different ML algorithms to predict the band gap, Conduction band minimum (CBM), and valence band maximum (VBM) of perovskites, as well as the electrical properties of PSCs. XGBoost models were found to be the best for predicting the bandgap, CBM, and VBM, while RF models were the most effective for predicting the electrical properties. Eight chemical compositions (MA, FA, CS, Pb, Sn, BR, Cl, I) were used as features for the bandgap, CBM, and VBM models, and a set of 13 features (MA, FA, Cs, Pb, Sn, Br, Cl, I, Bandgap, energy level alignments, electron mobility, and hole mobility) were used for the electrical parameter models. Using the prediction models and SHAP theory, they found that the perovskite material was the main factor affecting the PCE of PSCs. It was suggested that increasing the FA content in the perovskite material would increase the PCE of PSCs.

Gok et al. [43] utilized an RF ML method to predict the bandgap and PCE of halide using eight different halide perovskite compositions. They obtained the bandgap values by analyzing Tauc plots and predicted the performance of the solar cells using J-V spectra. RbCsFAMAPI, CsFAMAPI, CsFAPI, FAPI, MAPI, MAPI-Cl, FAPI+MAPBr, and FAMAPI-Br were the eight different perovskite compositions that were used as absorber layers in this work. In their model, only the absorber layer was taken into consideration, but the hole transport layers and the electron transport layers were not taken into consideration. According to their model, the FAPI-based PSCs had the lowest PCE of 15 %, while FAMAPI-Br had the highest value of 19.3%.

PSCs typically exhibit optimal efficiency when their bandgap falls within the range of 1.1 eV to 1.8 eV [52]. Yan et al. [40] represented XGBoost and RF algorithms to predict some properties of five new compositions of perovskites composed of $(FAPbI_3)_x$ $(MAPbBr_{2.8}Cl_{0.2})_{1-x}$ with bandgaps below 1.60 eV for experimental guidance to develop high-performance solar cells. Their model predicted the bandgap, J_{SC} and V_{OC} for the five perovskites. The highest J_{SC} and V_{OC} values were found for the $(FAPbI_3)_{0.99}(MAPbBr_{2.8}Cl_{0.2})_{0.01}$ composition. The authors also analyzed XRD patterns and Time-measured Photo Luminescence (TRPL) spectra for the five perovskites. Based on the supporting information, they determined that $(FAPbI_3)_{0.95}(MAPbBr_{2.8}Cl_{0.2})_{0.05}$ shows the longest photo-generated carrier lifetime and has been selected as the champion perovskite composition with a predicted bandgap of 1.55 eV. Finally, they designed and simulated a PSC with Glass $/FTO/SnO_2/(FAPbI_3)_{0.95}(MAPbBr_{2.8}Cl_{0.2})_{0.05}/$ *Spiro-OMeTAD/Au* configuration and achieved 22.5 % of PCE with 82.4 % FF, 24.6 mA/cm^2 J_{SC} and V_{OC}.

In a recent study by Jiang et al. [53], a GPR ML technique was employed to optimize the doping concentration of KI in $MAPbI-3$ solar cells. Doping with alkali metals is a well-established method for reducing defect density in perovskite films, owing to their inherent stability against oxidation and decomposition. The GPR model was trained on KI doping concentration and voltage as input features to predict the J_{SC} and PCE at various KI doping concentrations.

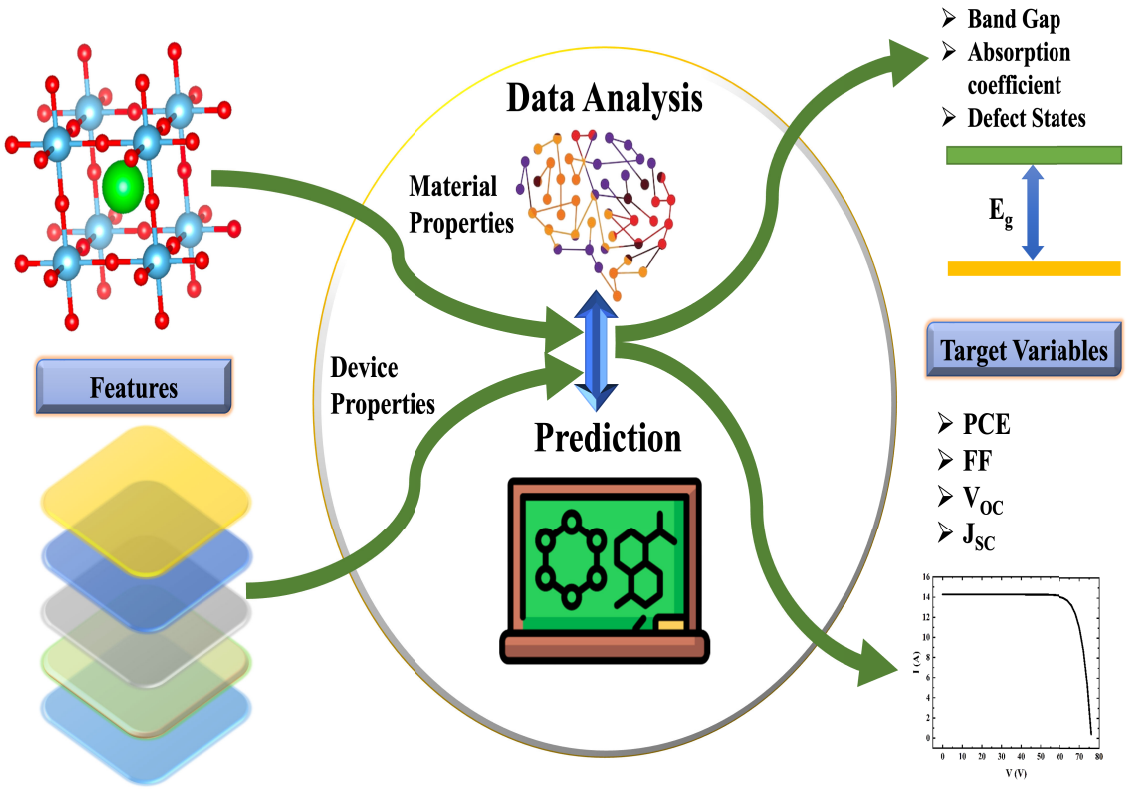

FIGURE 5. ML application in PV fabrication.

Following two rounds of prediction, the highest PCE of 20.91 % was achieved with a high FF of 80 % in $MAPbI_3$ solar cells doped with three % KI.

Yilmaz et al. [54] used ML for the prediction of bandgap and PCE utilizing three different algorithms. In their research paper, the authors utilized a range of characteristics to predict bandgap, including perovskite-type, structure, and layer thickness, along with the optoelectronic, chemical, and physical properties of both cations and anions to serve as descriptors for predicting bandgap. The authors incorporated bandgap, among other descriptors, to accurately predict the properties of PCE, Electron Transport Material (ETM), and Hole Transport Material (HTM), as well as their corresponding HOMO-LUMO values. XGBoost was found to be the best algorithm for predicting bandgap. According to their work, the most influential descriptors for accurately predicting bandgap were identified as the thickness of the inorganic layer, the radius of the anion, and the 2D radius of the cation. For PCE prediction, random forest models performed better for both regular and inverted cell structures. The PCE prediction analysis showed that for regular structures, the inorganic layer thickness, 2D cation radius, and bandgap have the most significant influence on PCE, while for inverted structures, the most effective effects are caused by the inorganic layer thickness, 2D cation radius, and conduction band energy of HTL martial. The bandgap predictive model shows an inverse relationship between the bandgap and the inorganic layer thickness and a nearly inverse linear relation between the PCE and the bandgap of 2D cations.

Hybrid PSCs use an HTM To facilitate the extraction and transportation of photogenerated holes from the perovskite layer to the electrode. The use of an HTM reduces the energy barrier between the perovskite and electrode, leading to an increase in the efficiency of the charge separation process. Additionally, the HTM also plays a crucial role in suppressing charge recombination at the perovskite/electrode interface, which is another important factor that affects the overall performance of the PSCs [55], [56], [57]. In a recent research article by Yildirim et al. [58], an ML was proposed to predict the impact of HTMs with different substitution groups on both the optical bandgap and the performance of solar cells. The authors aimed to develop a predictive tool that could be used to optimize the selection of HTMs for photovoltaic applications using both the R.F. model and Automatic machine learning (AutoML) framework, General Automated Machine Learning Assistant (GAMA). AutoML is a rapidly developing field within the realm of machine learning. It aims to address the challenge of finding the most optimal combination of data processing, learning algorithm, and hyperparameters to achieve the highest performance on

a given dataset within a specified computational budget. AutoML represents a novel approach to streamlining the machine learning process by automating the selection and configuration of key components [59]. They discovered that the model generated by GAMA was much more consistent and responsive compared to R.F. It was determined that the implementation of AutoML would provide greater benefit for tasks and datasets of increased complexity. The prediction methodology employed by the GAMA system was found to be a pioneering, efficient, and rapid approach in the field of HTM screening and was deemed an effective and reliable method.

Photonic curing [60] is a process used in the manufacture of photovoltaic devices. It is a method of applying light energy to a material to initiate or accelerate a chemical reaction. In the context of photovoltaic devices, photonic curing is typically used to cure or harden photovoltaic materials, such as photovoltaic polymers, which are the basis for some types of flexible and lightweight solar panels. The photonic curing process can be performed using U.V. light, visible light, or laser light, and it allows for the efficient and precise hardening of photovoltaic materials. This helps to ensure that the photovoltaic device has the necessary physical and electrical properties for optimal performance and durability [61]. In a study conducted by Xu et al. [62], the application of Bayesian Optimization (B.O.) was explored as a means of optimizing the PCE of photonically cured $MAPbI_3PSCs$ on ITO-coated willow glass (W.G.). The B.O. framework was utilized with four input variables, namely the concentration of $MAPbI_3$, additive $C.H._2I_2$ volume, pulse voltage, and pulse length. The results of this optimization led to a PCE of 11.42%, which was achieved using the device configuration of $WG/ITO/NiO/MAPbI_3/P.C._{61}BM/BCP/A$. To gain a deeper understanding of the factors affecting device performance, the authors conducted a SHAP analysis. The analysis revealed that higher levels of $C.H._2I_2$ and lower concentrations of $MAPbI_3$ were associated with improved device performance.

Cai et al. [63] proposed an ML approach based on a forward-reverse framework to determine the relationship between important parameters and photovoltaic performance in high-profile organic metal halide perovskite (OMHP) materials ($MASn_xPb_{1-x}I_3$). The objective of the forward analysis procedure aims to develop the bandgap and device performance models using ML, whereas the reverse engineering approach is for predicting the optimization parameters of mixed Sn-Pb perovskites and experimental realization. According to their proposed E_g model, an asymmetrical bowing relationship has been found between E_g and the Sn-Pb composition ratio. Following that, they developed an NN-based performance model that considered the energy level of the OMHP and carrier transport layers. The highest PCE of single-junction PSCs has been predicted at about 1.35 eV of V_{OC}. Finally, using reverse engineering, an optimum Sn-Pb composition ratio of approximately 0.6 was obtained for high-performance perovskite solar cells, which was further verified by experiments.

Toxicity is a barrier to large-scale commercial production and the photovoltaic field use of Pb-based halide perovskites [64], [65]. Tin (Sn) PSCs are the most promising Pb-free alternatives [66]. Sn-based PSCs are still in the early phases of research and will require a significant amount of time and effort to achieve an optimal structure. T. Bak et al. [67] proposed an ML model to speed up the discovery of Pb-free Sn PSCs with optimal structure and PCE. They used DNN for predicting and recommending the optimal Sn-perovskite composition. $FTO/PEDOT:PSS/EDA_{0.01}PEA_{0.07}Cs_{0.03}FA_{0.51}MA_{0.38}SnI_3/ICBA/BCP/Ag$ is their recommended configuration for the Pb-free Sn-based PSCs. They also analyzed that when they fabricated Sn PSCs in alternative combinations from the recommendation, the device performance dropped significantly, indicating that their proposed DNN model was successful in determining the optimal device configuration.

Table 1 presents a concise overview of the survey on research papers published in the last two years that have utilized ML approaches for accurately predicting the electrical and optoelectronic properties of PSCs. These research papers signify the growing interest in applying cutting-edge ML techniques to address key challenges in the field of PV devices and offer important insights into the potential of ML-based methods for advancing the development of efficient and cost-effective solar cell architecture.

B. ML IN THE PREDICTION OF STABILITY

Çagla et al. investigated the impact of storage conditions, cell manufacturing materials, and deposition techniques on both the stability [49] and efficiency [68] of the devices. Their results have been summarized in Table 2. They also analyzed the ambient conditions on stability and summarized that high humidity, high temperature, and high illumination have negative effects on solar cell stability.

Yilmaz et al. [54] employed an ARM-based ML technique to identify the key descriptor that influences the stability of 2D/3D PSCs. The findings indicate that layered perovskites are more stable than those with passivated and mixed structures. Among the commonly used 2D cations, BA demonstrated greater stability compared to PEA, although PEA had a higher average power conversion efficiency. Mixed 3D cation perovskites, particularly those containing FA, showed higher stability in 2D/3D structures compared to those composed of MA only.

Mammeri et al. [5] used an extremely randomized trees technique for stability prediction of perovskite solar cells. They found out the impact of each PSC layer on the stability. They trained 1050 samples of perovskite devices with different materials, deposition techniques, and storage conditions using the extremely randomized trees method. The results indicated that using hydrophobic materials in PSC layers

TABLE 1. Recent research review on the prediction of electrical & optical properties during fabrication using ML.

Ref	Research	Methods	Results	Year
[50]	Predict the PCE from the experimental parameters	LR, ANN, XGBoost, GBDT, RF	A-site cation is crucial to getting highly efficient PSCs	2023
[51]	Predict the bandgap, CBM, VBM, and electrical properties of PSCs	XGBoost, RF, KNN, SVR, LR, GBDT, MLP	The most important influence on PCE is the features of the perovskite itself, especially the addition of FA^{+}	2022
[43]	Investigate the influence of different perovskite compositions on optoelectronic parameters and device performance.	RF	R^2 score is greater than 0.99 for the prediction model	2022
[40]	Predict five unexplored $(FAPbI_3)_x$ $(MAPbBr_{2.8}$ $Cl_{0.2})_{1-x}$ with low bandgaps towards to high PCE	XGBoost, RF	Highest PCE of 22.5% has been achieved for the planar $(FAPbI_3)_{0.95}(MAPbBr_{2.8}$ $Cl_{0.2})_{0.05}$ PSCs	2022
[53]	Find out the optimal doping concentration for $MAPbI_3$ PSCs	GPR	Highest PCE of 20.91% and 80% of FF has been found from 3% of KI doped $MAPbI_3$ PSCs	2020
[54]	Predict the bandgap and PCE depending on perovskite structure and their physical, chemical, and optoelectronic properties.	RF, XGBoost, ANN	• Thickness of the inorganic layer, the radius of the anion, and the 2D cation are found to be the most important parameters for bandgap predictions • Bandgap, inorganic layer thickness, and the radius of 2D cation are the most important parameters for PCE predictions	2022
[58]	Predict the suitability of HTMs and predict the influence of HTMs on the optical bandgap and PSCs performance.	GAMA, RF	GAMA performed well with 0.0542 ± 0.0470 RMSE score in predicting HTM	2023
[62]	Optimize the PCE of photonically cured $MAPbI_3PSCs$	BO	Higher level of CH_2I_2 additive and lower concentration of $MAPbI_3$ improves device performance.	2023
[63]	Establish the relationship between key parameters and photovoltaic performances in mixed Sn-Pb PSCs.	LR, SVR, KNN, RF, GBR, ANN, BO	Sn:Pb composition ratio near 0.6 has been obtained for PSCs with high performance.	2022
[67]	Accelerate the discovery of optimal Pb free Sn PSCs	DNN	MAPE of J_{sc}, V_{OC}, and FF have been obtained 10.93%, 17.69%, and 13.83% respectively.	2022

enhances the stability of the device. Additionally, perovskite layers with multi-cation and 2D/3D crystal structures provide long-term stability. The study found that for regular cells, using $TiO_2/m-TiO_2$ as ETL, 2D/3D perovskite as absorber layer, P_3HT as HTL, and LiTFSi + TBP as HTL second layer and Carbon as back contact improved the device stability with DMF + DMSO as precursor solution and Chlorobenzene as an anti-solvent solution. For inverted cells, stability was improved by using BCP a PCBM, $MAPBI_{3-x}Clx$, NiO and DEA, and Aluminum back contact.

Hu et al. [6] used ML to design a regular high-performance PSC (with the structure of $glass/TiO_2/FA_{0.85}$ $MA_{0.1}Cs_{0.05}Pb(I_{0.97}Br_{0.03})_3/spiro-OMeTAD/Au$) with 23.4 % PCE and the best long-term ambient stability. Here ML was employed to create models that relate stability and efficiency (including V_{OC}, J_{SC}, and FF) to five different factors,

i.e., bandgap, defect density, grain size, fluorescence lifetime and surface roughness. Four ML algorithms, including PLSR, ridge regression, SVR, and GPR, were used to predict the best map between the features and the target properties. Among all the four algorithms, SVR performed the best. According to their findings, the bandgap has the greatest impact on PCE and JSC, while the surface roughness and grain size have the greatest impact on device stability. They designed several annealing temperatures (60,90,110,120°C) to change the grain size in perovskite thin films and various organic compounds (ortho-, meta-, para-methoxy phenylethylamine iodized salts) to vary the surface roughness. The best device stability was achieved by the largest grain size and the least surface roughness.

Low-dimensional (LD) perovskite, also known as the capping layer, which is deposited on top of the perovskite absorber, is used to increase the environmental stability of perovskite solar cells. Hartono et al. [69] presented an ML approach to optimize the capping layer for $MAPbI_3$. They took into consideration 21 organic salts as capping-layer materials, including salts based on iodine and bromine, with various sizes, branches, and chemical properties. According to the RF model and SHAP values, the low number of hydrogen-bond donors and small topological polar surface area (TPSA) of the organic capping layer are the topmost factors for determining device stability. According to this study, the most stable capping layer is Phenyltriethylammonium iodide (PTEAI), which does indeed have a low number of hydrogen-bond donors and a small TPSA). PTEAI successfully extends the $MAPbI_3$ stability lifetime by 4 ± 2 times over bare $MAPbI_3$ and 1.3 ± 0.3 times over state-of-the-art octylammonium bromide (OABr).

Table 2 outlines a brief summary of our survey on research papers published in the past three years, which utilize ML techniques to analyze various factors that can impact the stability of PSCs.

IV. ML IN THE FABRICATION OF OTHER CONVENTIONAL SOLAR CELLS

Abadi et al. [71] presented a novel method for predicting device performance and identifying the structure-property correlations in organic photovoltaics (OPVs) through a combination of ML and Taguchi Design Experiments (TODE) [72]. They created a database of 240 small-molecule organic solar cells with 220 different donor materials blended with two fullerene-based acceptors ($PC_{61}BM$ and $PC_{71}BM$). The model used ten quantum chemical features related to energy conversion as descriptors. The TODE and ANN algorithms were combined to obtain useful and constructive design guidelines. The model predicts that reducing the optical band gap increases the performance of the organic solar cell by increasing the number of photons absorbed by the device. The model also identifies the optimal values of or Δ_L (Energetic difference of LUMO and LUMO+1) and Δ_H (Energetic difference of HOMO and HOMO-1) for each optical band gap. It has been observed that with low optical band gap organic solar cell materials, lower values of Δ_L and Δ_H result in higher PCE. The prediction model states that having Δ_L and Δ_H nearly equal to 0 leads to good PCE values.

Zhu et al. [73] employed ML techniques to investigate the paramount variables that affect the device performance of the $Cu(In_{1-x}Ga_x)Se_2(CIGS)$ solar cells as well as the underlying correlations. From previous research and experiments, they selected 15 input features, including CGI, GGI ratio, thickness of the CIGS layer, highest substrate temperature during fabrication, fabrication method, alkali treatment, substrate type, and others for the ML algorithms. RF performed best at predicting the efficiency of CIGS solar cells, among other algorithms. Based on the results of the RF algorithm, it has been evaluated that alkali post-deposition treatment (PDT), CIGS thickness, buffer layer thickness, CGI, and GGI gradient are the most significant features that have the greatest impact on device performance. In conclusion, they recommended a co-evaporation fabrication method for CIGS with CdS as the buffer layer. Also suggested that the ideal CIGS thickness is 2-2.3 micrometers, and the ideal substrate temperature is 540-600°C for highly efficient CIGS solar cells with CGI ratio adjusted from 0.87-0.95 and GGI ratio from 0.37-0.40.

Salman et al. [74] developed and implemented an ANN-based ML model to extract the actual Cu doping profiles that result from the process of diffusion annealing and cool-down in the fabrication sequence of CdTe solar cells. They used PVRD-FASP [75] solver to generate a big dataset. The goal of this work is to mimic the Cu diffusion process implemented in the PVRD-FASP solver with the ANN ML model. The main purpose of this study was to replicate the Cu diffusion process carried out by the PVRD-FASP solver using the ANN ML model. The results obtained from the PVRD-FASP solver simulation are in good agreement with the values predicted by the ANN algorithm. It was also found that creating Cu doping profiles using the PVRD-FASP solver is a lengthy process, whereas using the ANN algorithm is almost instantaneous. This makes the ANN algorithm a useful tool in the development of optimal doping profiles for CdTe solar cells.

W.B Xiao and colleagues [76] utilized ANN-based ML models to predict the PCE of three types of silicon solar cells: mono-crystalline, multicrystalline, and amorphous crystalline. The training data was obtained by characterizing experimental solar cell samples under various conditions of incident light intensity and ambient temperature. The prediction results were evaluated in terms of correlation coefficient and mean of square error (MSE). The prediction results were found to be consistent with traditional physics-based studies of solar cells. The authors also suggested the optimal number of hidden layers in the ANN architecture for each of the three solar cell technologies and discussed the significance of the number of hidden layers in determining the quality of predictions. From their analysis, in the case of multi- and amorphous crystalline cells utilizing 3 or 4 hidden layer units

TABLE 2. Recent research review on PV stability using ML.

Ref	Research	Methods	Results	Year
[49]	Find out the effects of cell manufacturing materials, deposition technique, storage condition, and ambient condition on cell stability.	ARM, Decision trees	The cells stored under low humidity were found to be more stable. Mixed cation perovskites and multi-spin coating technique improve device stability	2020
[54]	Identify the descriptors leading to high stability of 2D/3D PSCs.	ARM	Layered perovskite structures are more stable	2020
[5]	Identify the influence of several materials and each individual perovskite layer on device stability	Extra Trees Classifier	Hydrophobic materials in PSC layers improved device stability. Multi-cation and 2D/3D crystal structures as absorber layers give long-term stability.	2023
[6]	Identify the influence of five factors, including grain size, defect density, bandgap, fluorescence lifetime, and surface roughness, on device stability.	PLSR, Ridge Regression, SVR, GPR	Surface roughness, and grain size are the most influential to the long-term stability	2022
[69]	Optimize the capping layer for $MAPbI_3$ to get better stability.	RF	PTEAI extends the $MAPbI_3$ by 4 ± 2 times over bare $MAPbI_3$	2020
[70]	Find the relationship of Reproducibility, Hysteresis, and Stability in PSCs	ARM, Pooled Variance	The materials and deposition techniques that have favorable outcomes in terms of low hysteresis and high reproducibility also exhibit advantageous effects on achieving high PCE and long-term stability.	2020

led to a high correlation coefficient and low MSE. Regarding single crystalline cells, the best results were obtained using eight hidden layer units.

Anti-reflection coatings (ARC) are used on the surface of solar cells to reduce the amount of light reflected back into the atmosphere, increasing the amount of light absorbed by the cell. This results in higher efficiency and power output from the solar cell [77]. In a recent study, Shivam et al. [78] proposed using ANN to predict the optimal thickness of the ARC coating based on the voltage and current generated by the solar cell. The authors utilized PC1D, a 1-D numerical simulation tool, to create the training database for their ANN model. The results showed that increasing the thickness of the ARC coating can lead to the greater power output from the solar cell.

Kaya and Hajimirza [79] developed a Neural Network (NN) surrogate model to predict the optical performance of thin film multi-layered amorphous-silicon-based solar cells using Neural Network (NN) technique. The training dataset was created using finite-difference time-domain (FDTD) simulation methods. The authors found that the NN-based methodology was accurate and much faster than FDTD simulations in predicting optical absorption as a function of incident light wavelength and cell geometry. This work explores the fundamental reasons behind the ability of an ML-based surrogate model to approximate the results of a device physics-based simulation model. The resulting optimization solution suggests a significant improvement in external quantum efficiency compared to bare silicon and a random design.

Wasmer et al. [80] proposed an ensemble of decision trees with feature sub-sampling as the ML model to understand the PCE of mass-produced Q. ANTUM solar cells [81] based on p-type Czochralski Silicon (Cz-Si) wafer. The main goals of their work were to understand the global hourly time trends of solar cell efficiency observed in a 1-week dataset and to understand the impact of different features of solar cell production on the temporal evolution of produced solar cell performance. The training data was collected from 500,000 solar cells using inline measurements over one week. The total number of features was 329, which we clustered into groups using hierarchical clustering based on their similarity with respect to their impact on a cell performance metric. The ML model that was developed in this work was utilized to make predictions of solar cell efficiency with respect to time, using time as one of the features. The authors proposed that this approach could be applied to real production lines since the prediction accuracy was extremely high.

V. ML IN PVS OPERATION AND MAINTENANCE

This section evaluates the application of various ML techniques to address various issues related to the PV system's post-fabrication phases in order to integrate it with the system successfully.

A. FAULT DETECTION IN PVS

Photovoltaic system (PVS) is operated in an outdoor environment since it is required to be exposed to solar radiation to generate electrical energy. However, like many electrical systems, PVS is also subjected to numerous faults due to external outdoor conditions and aging during its life cycle [82]. This section provides a brief overview of various faults that can take place in PVS and their detection trend in recent years.

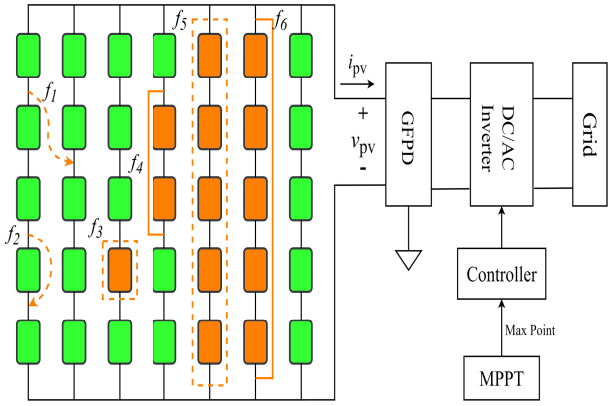

FIGURE 6. Different type of fault in PV array.

1) TYPES OF FAULTS

1) Line-to-Line Fault: This fault arises from accidentally shorting two points of different potentials in a PV array [83]. It could occur in a single string or between two lines, as shown by f_1 and f_2 in Figure 6. This faulty condition might be attributed to mechanical damage, degradation in insulation, cable aging, other external factors, etc. Detection of this fault in low irradiance is challenging, and failure to detect might cause fire hazards in large PV stations.
2) Arc Fault: Arc fault in photovoltaics is caused by the junction and conductor discontinuity due to damaged insulation. Arc faults could be series or parallel. It can also be either DC or AC depending on the location of the fault with respect to the inverter. Parallel arc fault current is higher than the series, thus can easily trigger circuit breakers. The difficulty in detection comes from the low-energy dc series arc fault. Early detection of arc faults is extremely vital as this can lead to a hazardous situation.
3) Hot spot fault: This is a type of mismatch fault arises from the imbalance of power distribution in a PV module that creates a localized overheated region. The cells within this high-temperature region are forced into reverse bias and act as a load and consume the energy of other cells [84]. This fault is caused by shading over a group of cells due to plants, vegetation, material defect in PV cells or mechanical damage.
4) Mismatch Fault: A Mismatch fault takes place when one cell or a group of cells has different electrical properties in a PV module. This fault may be short lasting due to conditions such as partial shading since solar modules are susceptible to shading. Permanent mismatch faults can be attributed to defects in PV cells, degradation etc. [85]
5) Open and Short Circuit Fault: Open circuits in PV modules can be attributed to factors such as old power cables, overheating, poor connections, loose junction box connectors, etc. This faulty condition can be linked to other operational safety concerns like arcing, hot spots, etc. Short circuits, on the contrary, are caused by bad connections between modules, manufacturing defects, contamination of cell surface, etc. Open circuit (f_3, f_5) and short circuit (f_4, f_6) condition is shown in Figure 6.
6) Other Faults: Several other faults can disrupt regular operation and prevent photovoltaics from obtaining optimal performances. An accidental connection between internal wiring with ground wire can cause a line to ground fault. A bypass diode is a safety device that has several functions in PVS. It reduces the shading effect and prevents reverse current during hot spot fault. Faults in bypass diode can be caused by a short-circuit or open-circuit of the diode. Cracks in protective glass, degradation in cells and defects in anti-reflective coating happen over the long-running operation of solar modules

2) RECENT FAULT DETECTION TREND IN PVS

1) Machine Learning Models: Hussain et al. [86] studied open circuit faults using solar irradiance and total power as features. Open circuit fault is categorized into low fault with less than 30% PV modules disconnected, high fault with disconnected modules greater than 30%, and string fault when an entire string is disconnected. The input labelling is done using agglomerative hierarchical clustering algorithm before feeding to the supervised ML approaches, thus utilizing the advantage of unsupervised model. The authors have employed several ml models, including KNN, DT, SVM and RF. The models achieved 100% accuracy on their dataset when it did not have any noise or missing value. However, it reduced to 94% for Gaussian Naïve Bayes (NB) with noisy and missing data in a 15-year-old PVS. To detect series arc fault (SAF), the authors in this article [87] proposed twin support vector machine (TWSVM). SAF introduces fluctuations in the dc bus current. The authors argued that traditional time domain analysis is insufficient to differentiate the stable arc state from

regular operation. Hence denoising is first performed through hankel singular value decomposition (SVD) to eliminate switching frequency and background noise. Then, empirical wavelet transform (EWT) is applied to obtain the modal component in both the time and frequency domains. To achieve effective segmentation in frequency spectrum, mathematical morphology is applied to improve the EWT. Based on the composite multiscale permutation entropy (CMPE) of each modal component of EWT, another study proposes TWSVM [88] classifier and the hyperparameters are optimized using the salp swarm optimization algorithm [89]. The arc fault data was taken considering several working conditions, such as dynamic shading, MPPT adjustment, wind blowing and inverter startup. The proposed model achieved 98.10% accuracy. However, the applying EWT results in a slower speed, specifically 246ms. Cai and Wai [90] combined an optimized variational mode decomposition (VMD) with support vector machine (SVM) to detect dc arc fault. This method decomposes the input currents in the time domain into multiple sub-signals known as intrinsic mode functions (IMFs) of different frequency components [91]. To reduce the number of iterations for VMD to converge, the parameters of VMD are optimized considering the categorical correlation of features of each input sample. The extracted IMFs are input to the SVM. The training of SVM is performed to search for nonlinear hyperplanes. The parameters of SVM are optimized by the particle swarm optimization algorithm [92] due to its shortest search time. Additional comparison with grid search and genetic algorithm is also carried out. The proposed scheme can detect arc fault in different scenarios such as inverter startup, wind blowing and shadow occlusions achieving an accuracy of 98.21%. The method can detect parallel arc faults as well. Adhya et al. [93] employed light gradient boosting (LGBM), categorical boosting (CatBoost) and extreme gradient boosting (XGBoost) algorithms to detect line-to-line, open circuit, partial shading, and degradation fault in PV arrays. They used six features for the task. The features include the ratio of current, voltage and power at faulty condition to the normal operation at the maximum power point, hourly yield and efficiency of PV array and the ratio of array capture loss at normal to faulty condition. The Grid search algorithm is utilized to find the optimal hyperparameters. The ML models are compared with the random forest classifier, and the LGBM model is shown to achieve 99.996% fault detection and 99.745% fault classification accuracy. Xu et al. [94] proposed fuzzy c-mean clustering (FCM) algorithm based on the current-voltage (j-v) characteristic curve and fill factor. The faults considered in the study are short-circuit, open circuit, degradation and partial shading. The model is further validated by the intra-class maximum mean discrepancy (ICMMD) method. Based on the value of ICMMD, reclassification is performed to enhance the accuracy of FCM. The proposed model obtained 100% accuracy on a smaller dataset in a simulated environment. However, dependency on the I-V curve makes the framework offline.

2) Deep Learning Models: ML algorithms heavily rely on manual feature engineering, which can be time consuming and requires domain expertise [95]. As the input dimensionality increase, ML models struggle to capture complex patterns. On the contrary, DL models with their hierarchical architecture learns intricate patterns and generally more scalable and capable of handling large feature space. Most commonly used DL models in fault detection are MLP and variants of CNN. They are mostly modeled as classification task. In [96], the authors argued that using a single channel 1-D CNN tends to misclassify PV system faults due to environmental interference and inverter regulation. They proposed using both time and frequency domains to enhance feature extraction ability. Hankel SVD is used on input currents to eliminate the interference of inverter switching frequency. Then FFT decomposition is applied to the filtered data to obtain frequency domain signal. Both time and frequency domain signals are processed in parallel by 2 CNN models. Their output is concatenated and used for binary classification. The proposed scheme can precisely detect low-energy arc faults with 96.67% accuracy. Hong and Pula [97] proposed a three-dimensional CNN for fault detection from AC and DC power data. The faults considered in this study are line-to-line, open-circuit, and short-circuit faults. The input measurements are first transformed into a 2-D image using gramian angular summation fields (GASF) [98] and a stack of 20 such images is chosen to be the input of the CNN module. The measurements were taken at the rate of 1536 Hz. The proposed method is compared with machine learning algorithms such as decision tree, random forest, k nearest neighbor and support vector machines. 3-D CNN outperforms these machine learning algorithms along with 2D CNN, achieving 96.43% accuracy. Mustafa et al. [99] proposed a multioutput classification scheme using Bi-LSTM for fault identification and location detection. Using line voltages, system voltage, current, irradiance, and temperature as features for each time step, three faults namely line-to-line, line-to-ground and open-circuit faults are considered for the classification task. Ten strings with ten modules per string PVS is designed using PSCAD software to generate a large amount of training and validation data for the proposed algorithm under various temperatures and solar irradiance. They showed BiLSTM to be superior compared

to RF, 1D CNN, and LSTM with 99.94% fault classification and 99.54% location detection accuracy. The article in [100] presents an interesting approach where the average output power reduction in other words, severity in case a fault occurs along with its type is determined. The authors proposed a stacked gated recurrent unit (GRU) network in a multi-output manner. One output head is a classifier trained with a cross-entropy loss function that predicts fault type. The types of faults considered are open circuit, short-circuit, aging, partial shading, dust accumulation, and potential induced degradation (PID). PID refers to the large voltage difference between PV cells and array frames. In contrast to this research, this fault is not typically studied in the literature. The other output head predicts the severity of the fault trained using the mean absolute error loss function. The hourly input features namely, irradiance, temperature, zenith angle, current and voltage are structured so that the model using the past 24 hours' data, can predict the fault type and it's severity every hour. The trained model is evaluated using 5-fold cross-validation and compared to CatBoost algorithm. The trained model outperforms the CatBoost method obtaining 96.9% accuracy in fault type prediction and 0.67 MAE when tested with dedicated sensor readings of irradiance and temperature. However, the accuracy drops to 86.4% and 2.09 MAE when tested with readings provided by satellite. This outcome is natural since dedicated sensors can provide precise readings compared to a satellite for a particular area. The proposed approach is also tested with climate data that is different from the climate the model is trained with. Also, the capability to predict unknown fault types is also tested. In both cases, the model is able to provide satisfactory results. The model can predict unknown faults for which the model is not trained with, by selecting a threshold in the severity estimation and considering it as a binary classification problem. Voutsinas et al. [101] employed MLP to detect short-circuit, open-circuit, and mismatch faults. The proposed model structure consists of a regression and a classification output head. The regression head also predicts the fault type by outputting a number between 0 and 4. The proposed algorithm obtained a classification accuracy of 93.4% and mean absolute error of 0.305. Monitoring large-scale PV plants using UAVs increases efficiency, reduces time consumption and human error. The authors in this article [102] proposed a CNN model based on RGB images of PV modules that are taken through UAVs. They tested their algorithm on visual faults included burn marks, delamination, snail trail, glass breakage and discoloration. Due to the limited availability of training images, the authors applied augmentation techniques to increase training samples such as rotating, adding noise, flipping, warping and blurring. The proposed model is able to obtain classification accuracy from 91% to 100% on the tested fault conditions. The authors also showed that the framework outperforms pre-trained models such as VGG-16 and ResNet-50. Lu et al. [103] developed a dual channel CNN (dcCNN), which can automatically extract critical features and classify them into different fault categories. The type of faults considered in this study is line-to-line, partial shading, and open circuit. The sequential current and voltage data from PVS are first normalized and transformed into an electrical time series graph (ETSG) and taken as input to the dcCNN model. Discrete wavelet transform (DWT) is applied to detect fault features that are not obvious in time domain measurements, such as line-to-line fault at low irradiance. The model architecture consists of two stages; feature extractor and classifier. In the first stage, two ETSG input for current and voltage are fed to two separate CNN modules for extracting relevant features. Inspired by the attention mechanism in the classifier stage, the authors proposed using trainable weights to train the model to automatically select necessary features for the classification task. The proposed algorithm was able to achieve 99.6% accuracy outperforming convolutional autoencoder (1D-CAE), deep residual network (1D-ResNet), and CNN without the trainable feature selector. They have also conducted case studies to detect faults under complex conditions such as low irradiance and fault impedance. However, the accuracy the proposed model is reduced to 82.52% still outperforming other models. Uneven dust accumulation cause thermal imbalance in different regions in PV panels leading to lower power efficiency and reduced lifetime. Fan et al. [104] proposed a deep residual neural network-based (DRNN) method to detect uneven dust concentration in PV panels. Raw training images were subjected to processing such as perspective transformation, silver grid lines removal, and equivalent segmentation. Finally, the dust distribution of the segments is clustered using the k means algorithm. The proposed DRNN technique achieved 78.7% R^2 score, outperforming the MLP model.

3) Transfer learning Based Models: DL models need to be retrained before applying to a new dataset. Additionally, DL methods require a large amount of labeled data for better generalization capability. However, in the real world, quality samples on abnormal conditions are challenging to obtain [105]. Thus TL approach is recommended. TL refers to the process where DL algorithms are pre-trained on a large dataset and later fine-tuned on a limited number of observations. By leveraging the learned representations of pre-trained models, TL often leads to improved model performance compared to training from scratch. For instance, Shihavuddin et al. [106] used pre-trained EfficientDet

and YOLOv5 models to detect surface damage in solar panels and wind turbines. The dataset consists of optical and infrared images, which are fined tuned to detect damaged areas in bounding boxes. While intersection over union (IoU) is used to determine the overlap between ground truth and predicted bounding boxes, mean average precision (MAP) is used as the final evaluation matrix. The proposed framework detected surface damage in solar panels with 86.7% MAP. Sung et al. [95] introduced a transfer learning-based approach for detecting low-energy dc series arc faults. The algorithm consists of two stages. A 1-d CNN module is pre-trained to predict the presence of arc faults from input currents. The knowledge stored within the CNN module is transferred to an LSTM network for multi-class classification. Their proposed method can detect no arc fault, low energy arc fault, and stable arc fault condition. The algorithm shows 95.8% accuracy outperforming other machine learning models, namely SVM, RF, Light GBM, XGBoost and deep learning model namely 2D CNN and 1D CNN. It is able to detect arc fault within 63 milliseconds. The authors in [105] discussed the correlation between the availability of large datasets and model performance on automatic fault detection. AI models generally perform better with sufficiently high-quality training data. Even then, training models in a lab environment might cause inconsistent accuracy in real scenarios. Hence, the authors proposed lightweight transfer learning, where CNN is trained on the source dataset and later can be fine-tuned on the target datasets. The source dataset is created using photovoltaic power profile evaluation software and an arc generator according to UL-1699B-2018 standard. The target domain dataset is created using four to twelve mono-crystalline solar cells. The authors presented a mechanism for fault data augmentation, where the wasserstein generative adversarial network with gradient penalty (WGAN-GP) is trained to generate fault samples from a latent vector. The proposed algorithm achieved 96.4% arc detection accuracy trained only on 20 arcing samples on the target domain, where an increase in accuracy (98.7%) is observed when the arcing samples are 300. The proposed technique outperforms other transfer learning approaches, such as deep adaptation - deep convolutional GAN [107], transfer RNN, and deep transfer learning with VGG16 [108]. Guo et al. [109] employed transfer learning-based LSTM for detecting different fault scenarios, namely shading, hot-spot and short-circuit. The process relies on the difference between the anticipated power of defective conditions and the actual power of normal states. The factors affecting power generation for photovoltaics, such as solar irradiance, temperature, and humidity are first classified using adaptive k means clustering algorithm. These predictions along with time series data of normal operating conditions are used to pre-train LSTM neural network. Finally, the LSTM model is fine-tuned considering different fault scenarios separately. Ahmed et al. [110] utilized seven layered convolutional neural network models based on infrared images to classify into three PV health categories: healthy, hotspot and faulty. The trained model was able to obtain 96% accuracy. Transfer learning is applied by utilizing the trained model by fine-tuning to predict five class categories. The extended categories include bird drop, single cell defect, patchwork, string defect, and shading and is shown to achieve 97.63% accuracy. The framework is compared with pre-trained networks such as SqueezeNet, ShuffleNet and GoogleNet. GoogleNet. While the proposed algorithm outperforms them and requires less execution time, GoogleNet achieves almost similar 97.62% validation accuracy.

B. ML FOR IRRADIANCE FORECASTING AND PV OUTPUT POWER ESTIMATION

Designing a solar PV system requires precise measurements of diffuse horizontal irradiance. Weather conditions like solar irradiance and temperature greatly influence the energy production from PV systems. This causes production levels of this energy source to vary, which poses a challenge for power companies in balancing electricity production and consumption while using PV systems. To address this, numerous machine learning (ML) algorithms have been employed to predict solar irradiation and the output power from PV systems.

Miranda et al. [111] developed a regression model where the input features were the measure of global horizontal irradiance and a geographic coordinate. The dataset was prepared from six locations for 11 years at half-hourly intervals. In the cleaning process of the dataset, rejection of data in the range $0.015 <$ clearness index < 1 was considered, whereas for sunset and sunrise, it was 10. In the pre-processing stage, seven variables were calculated by conducting feature engineering. The ANN model showed better performance compared to other models. However, the proposed methodology used separate models for different locations. It is required to have a single model with a combination of different zones and horizons. Asghar et al. [112] proposed regularization-based LSTM architecture to predict irradiance. The feature engineering of the proposed network is a combination of Pearson correlation, feature normalization, and 1-day lag irradiance. The Pearson auto correlation determines the link between input and output. The authors found that air temperature and one-day lag are strongly associated with the output variable, so these were used as input features. Next, a one-day delay is applied to the normalized irradiance. To avoid overfitting, dropout regularization was adopted in the LSTM. By reducing data dimensionality, the Pearson correlation resolves memory issues and limits computational complexity.

TABLE 3. Recent research review on fault detection in photovoltaic system.

Type	Ref	Faults	Methods	Results	Year
Machine learning	[86]	Open-circuits	KNN, DT, SVM, NB, RF, MLP	100% (No noise and missing value) 94% (with nosie)	2022
	[87]	series arc fault	Hankel SVD (denoising) IEWT+TWSVM	98.10%	2021
	[90]	arc fault	OVMD + SVM	98.21%	2022
	[93]	line to line, open circuit, degradation, partial shading	LGBM, CatBoost, XGBoost	99.996% (detection), 99.745% (classification)	2022
	[94]	short circuit, open circuit, degradation, partial shading	Fuzzy C- Mean Clustering +ICMMD	100%	2022
Deep Learning	[96]	arc fault	Hankel SVD + FFT + Multi input Inception CNN	96.67%	2022
	[97]	line to line, open circuit and short circuit fault	GAF (GASF) + 3D CNN	96.43%	2022
	[99]	Line to line, line to ground, open circuit	Multioutput BiLSTM	99.94% (classification) 99.54% (location)	2023
	[100]	open circuit, short circuit, aging, dust accumulation, potential induced degradation (PID), partial shading and fault severity	Multioutput GRU	96.9% (fault type), 0.67 MAE (Severity)	2022
	[101]	Short circuit, open circuit, mismatch fault	Multioutput MLP	93.4% and 0.305 MAE	2022
	[102]	Visual faults - Burn marks, delamination, snail trail, glass breakage and discoloration	CNN	91% to 100%	2021
	[103]	line to line, partial shading and open circuit	Dual channel CNN (dc-CNN)	99.6%	2021
	[104]	Uneven dust accumulation detection	DRNN	78.7% R^2 score	2022
Transfer Learning	[106]	Surface damage detection	EfficientDet, YOLOv5	0.861 MAP	2021
	[95]	Series arc fault	TL–LEDarcNet (Transfer learning CNN + LSTM)	95.80%	2022
	[105]	Arc fault	Transfer learning 1D CNN + fault data augmentation with WGAN-GP	96.4% to 98.7%	2021
	[109]	Shading, Hot-Spot and Short Circuit	k mean clustering + Transfer learning LSTM	4.32% RMSE	2022
	[110]	Hotspot, bird drop, single cell defect, patchwork, string defect, and shading	Transfer learned CNN	97.63%	2021

The result of the proposed method was measured against the ANN method, and the developed method showed the better accuracy. Abdel-Nasser et al. [113] combined LSTM and Choquet integral to forecast solar irradiance. The LSTM serves the purpose of achieving accuracy and the Choquet integral does the aggregation of the input parameters. Choquet integral will integrate the predictions of n number of LSTM models. After aggregating, the Choquet integral will provide the final forecast. The methodology was tested using six different real-time datasets. As the proposed methodology did not consider any data type, the process can be used for different kinds of forecasting.

Kartini et al. [114] developed a novel hybrid method, combining modified decomposition and a feed-forward NN model. The method was used to predict an hour ahead global horizontal irradiance. Depending on the weather data, the value of modified decomposition varied. The input and output of the model was group and used for the feed-forward NN. When the neural network is optimized, a sigmoid activation function is employed to link each of the several layers

TABLE 4. Recent research review on irradiance forecasting of PV using ML.

Ref	Research	Methods	Results	Year
[111]	half-hourly horizontal solar irradiance prediction	ANN, RF, EBT, EMP	RMSE ranging from 5.86% to 9.36% w/m^2 and R^2 score of 0.9974 to 0.9983	2021
[112]	one day ahead solar irradiance forecasting	Pearson Correlation-LSTM	MAE 28.3 W and R^2 0.85	2022
[113]	Irradiance forecasting with six different sites and three solar profiles	LSTM-Choquet Integral	RMSE from 19.72 to 30.33 for six different sites	2021
[114]	An hour ahead horizontal irradiance forecasting based on weather data	MD-FNN	MAPE error of 1%	2022
[115]	One-hour, one-day and one week ahead PV power forecasting in two seasons	CNN, CNN-LSTM, ARMA	RMSE score in Summer: • 0.046 KW (ARMA -1h) • 0.051 KW (CNN -1d) • 0.045 KW (CNN-LSTM -1w) RSME Score in Winter: • 0.188 KW (ARMA -1h) • 0.036 KW (CNN -1d) • 0.1 KW (CNN-LSTM -1w)	2020
[116]	solar irradiance forecasting	Convolutional GNN-LSTM	RMSE Score of 0.0052	2022
[117]	5 minute, 15 minute, 1 hour and 3 hour ahead PV power generation prediction	LSTM	Normalized RMSE and R^2 score of- • 5 minute ahead forecast: 98.0% and 0.988 • 15 minute ahead forecast: 96.6% and 0.949 • 1 hour ahead forecast: 94.8% and 0.963 • 3 hour ahead forecast: 92.9% and 0.927	2021

together. To predict solar power, Pavithra et al. [118] proposed a method where average temperature, surface pressure, wind speed, and humidity were used to prepare the dataset. Levenberg-Marquardt algorithm was used for solar power forecasting due to its high speed of operation with respect to other algorithms. In the irradiance, it was found that 6.5-7.5 $kW - hr/m^2$ occurred most of the time. They used ANN to train the model. To get maximum power point irrespective of loads, the data of the model was used in the PV array. While applying the data in the PV panels, the J-V characteristic curve shows the maximum power at a particular voltage. After that, LSTM was utilized to analyze the time series data and predict future input parameters. Suresh et al. used different CNN architectures (conventional CNN, multi-CNN, CNN-LSTM) to provide better forecasts [115]. The training time of CNN-LSTM is more than twice of other CNN models. They collected solar data from a PV panel of the Wroclaw University of Science and Technology. The input parameters were irradiation, wind speed, and temperature. At each 15-minute interval, the data was collected. The CNN used 175,200*4 matrix (1)-D structure) where the time step count is 175,200 and four columns are input features (irradiation, wind speed, ambient temperature, and PV module temperature). The efficiency of the CNN models was validated using the Autoregressive Moving Average (ARMA) model. The advantage of the ARMA model is that it has less computation cost and it can forecast faster than CNN once it best fits. They implemented one-hour, day, and week-ahead forecasting. In the case of an hour-ahead prediction, all models performed similarly, and for the day ahead, CNN-simple and CNN-LSTM performed similarly. CNN-LSTM performed comparatively better than other models when tested for the week-ahead forecasting. Jiao et al. [116] used a Graph Neural Network (GNN) and LSTM to forecast irradiance. The convolution GNN identifies the required features and LSTM gathers the temporal correlations. The graph matrix will be the input of the GNN layers and the recurrent LSTM layers will be cascaded to produce output. The computation cost of the proposed method is $O(max(E, W))$, where E is the edge number of the graph and W is the edge number in the LSTM network. On the other hand, Ahn et al. [117] aimed to forecast short-term power output based on weather data. As weather can impact power fluctuations for a short interval. They have proposed a deep RNN-based

model. In this study, weather data, for instance, solar radiation, temperature, humidity, wind speed, and PV power are collected in real-time and used as input features. The data were normalized using min-max scaling. Instead of batch normalization, layer normalization was applied, improving the training and test computation time. Their proposed model showed better results for five minutes ahead power generation forecast. However, the prediction capability is lower for an hour and three hours ahead forecasting. The authors emphasized on using more weather features like cloud images, dust sensors, etc. for better accuracy. It is evident that DL models such as CNN, LSTM ANN, and ML models, for instance, RF, ARMA, are the most commonly used techniques for this purpose. This is due to their better handling capability of sequential or time series data. These models typically perform well for short-term forecasts. However, as the forecasting horizon increases, the accuracy of the predictions may deteriorate. Additionally, these models are sensitive toward outliers and missing values, which is very common when dealing with real-world PV station data. Table 4 presents a brief overview of the articles and proposed algorithms reviewed in this section.

C. ML IN THE CONTROL METHODOLOGY OF PV

Proper integration of PV systems with the electricity grid requires sophisticated control methods. This section examines various control techniques that utilize machine learning (ML) to address the issues of maximum power point tracking (MPPT) in electric PV systems.

Agrawal et al. [119] used a transformer-based model for predicting the maximum power point. The model was trained using multidimensional time series input features. The dataset contains weather data from 50 locations. The proposed method shows better accuracy (more than 99.5% for both average and peak power). In the future, authors will consider multiple data sources with more comprehensive weather features and conditions to enhance the robustness of the proposed model. Cao et al. [120] offered a reinforcement learning-based PV voltage control for a distribution system. Centralized offline training and decentralized online execution are the two stages of the proposed method. From historical data, the actor and the critic in the training stage learned the optimal control strategy. In the next step, the actor will take actions based on the parameters. In the proposed method, no additional information can be used after training as it is done and agents will make decisions based on local information. The study in [121], proposed another reinforcement learning-based voltage control system of PV. The proposed method involved PV inverter's reactive control without any data between PV inverters. The proposed mechanism used reward state considering connecting point voltage, action state considering reactive power output, and reward state considering functional control of the network. By observing the voltage, the actor calculates the reactive power. Using the output, Critic updates the value function. The updated value function works as an evaluation point to update the probability policy by the actor. The proposed method was validated using simulation. However, the scalability and robustness need to be evaluated further as the method was tested for a low number of clients.

Zaidan et al. [122] proposed a three-phase multilevel PV-grid neural point clamp inverter that operates in the current control mode. ANN-based MPPT controls the current reference of the controller. There are three layers in the ANN MPPT. The temperature and the irradiance are the input features of the ANN model and the maximum power point voltage is the model's output. The PV module of MATLAB was used to generate the dataset. The proposed method's simulation result showed unity power factor. This study in [123] suggested a machine learning-based PV system's maximum power prediction method. PV Voltage, output voltage, PV current, irradiance, and temperature were used as the input features and the proportional integration value for MPPT was the output feature of the supervised machine learning classification model. The data was collected using MATLAB Simulink and was processed into time-series form. To avoid overfitting, the load current was discarded in the processing stage. The model is dependent on generated data. The model needs to test using real PV data. Mahesh et al. [124] developed a regression (linear and nonlinear) machine learning algorithm to control MPPT. Likewise, [123] data was collected using a MATLAB Simulink MPPT model. The model predicts maximum power and corresponding voltage where irradiation and temperature were input features of the model. The output of the model will help to determine the duty cycle of the boost converter. For both linear and nonlinear regression models, the accuracy was just over 95% which needs to be increased. In the future, the authors will consider different partial shading patterns to check the robustness of the proposed method. The research in [125] presented a machine learning-based MPPT identification with a pre-existing perturb and observer method. To locate MPPT, the multivariate regression model used irradiation, temperature, and humidity as input features. The model produced accurate results for high power range (900 to 1000) compared low power range (0 to 100). The training time of the model was high.

Mushen et al.'s proposed method for MPPT is based on the gradient ascent method that considers first-order derivative to update parameters [126]. Temperature and irradiance will act on the PV module, which will produce voltage and current. The gradient ascent method will use these to calculate the parameters of PV models and the duty cycle correlated with MPPT. The duty cycle is fed to the converter to identify the MPP. The method considers that temperature is constant. The method showed 68% power improvement compared to the system without gradient ascent. To obtain the MPPT, a deep learning-based method was developed by Rafeeq Ahmed et al. [127]. Weather conditions such as light intensity and temperature were considered the determining factors. The technique involved two-stage tracking to find out the MPPT. Initial data will be provided to the control and measurement

unit, which will share the data with the MPP point search unit. The search unit will return its findings and share the inputs with the ANN module. ANN module will calculate and share its result with the control and result unit. As the ANN module is interfering with the result of the search unit, it is not clear the purpose of the search unit. However, the result of the MPPT is 98%. In another research [128], the authors proposed an ANN-based method combining the BP neural network and ELMAN neural network to produce better results in different weather conditions. The data was collected from National Renewable Energy Laboratory, and the min-max method was applied to normalize the data to keep the data in the range of 0 to 1. There were two types of data. ELMAN neural network performed better for data of small fluctuations, and BP neural network performed better for large fluctuations. They combined both neural networks to have a better result for every situation. Kofinas et al. [129] proposed a reinforcement learning-based approach for MPPT control of PV. The method allowed tracking and adjusting a PV's maximum power even without previous knowledge. Despite having the PV source dynamics or having the characteristics of the PV source predefined, the proposed method solved the MPPT control problem. The algorithm's goal is to learn the system's optimum configuration depending on PV source performance. It is necessary to define a Markov Decision Process (MDP) model of the source behavior to apply a reinforcement learning strategy. MDP is a set of tuples of a finite set of states of operating point, a finite set of actions to change the state, transition function, and reward function for correlating actions. An MDP creates a sequential model considering a series of state changes based on executed actions to find the optimal policy for maximum payoff. The best state action integration is used to export a policy. The model needs to be optimized and state space can be reduced. Mahesh et al. [130] also proposed an MPPT control mechanism. They offered a decision-tree machine learning algorithm to develop the model. The dataset was prepared using the PV panel technical parameters. The model used irradiance and temperature as input features to predict maximum power and voltage. The model's predicted value determines the duty cycle to produce maximum power by driving the PV panel. The tracking efficiency of the work is around 94%. The authors considered more partial shading effects as future scope. Table 5 lists the cases in which ML has been used in the control system of PV. In the case of grid-tied PVS, abnormalities such as transient grid sags and solar irradiation flickering can cause the system to go offline. A simultaneous shutdown of PVS may result in unstable situations or outages in the grid. So, adopting a fault ride-through scheme is suggested to remain connected and assist in the voltage recovery [131]. khan et al. [132] proposed an LVRT scheme using neural network and Finite Control Set Model Predictive Control (FCS-MPC). The NN acts as a fault detector during voltage sags and signals the FCS-MPC controller to work to stabilize the system. The proposed mechanism successfully restored voltage within the time limit requirements indicated in the grid codes [133].

D. ML IN PV SIZING SYSTEM

Appropriate sizing of a solar PV system is essential to ensure the quality and consistency of a power supply and optimize the financial life-cycle savings. Different ML approaches have been used to determine the appropriate amount of panels, battery storage capacity, tilt, and azimuth angles. The work in [134] has developed the SolarMapper tool for mapping small-scale solar arrays at a large scale. It was made utilizing CNN, which offers pixel-by-pixel labeling for input data. Two significant experiments are run in this study. The Duke California Solar Array dataset was used in the training part, which comprises 16000 manually identified solar arrays and 400 km^2) images. The performance metric showed that the model was precise, with a precision of 0.76. The researchers assessed the PV array's installed surface area based on the segmentation procedure before calculating the solar capacity. The existing PV array capacity was then predicted using a simplified linear regression. By calculating the parameters for each array using color imagery, the model was able to attain a correlation coefficient of 0.91. The authors of [135] proposed a topology reconfiguration method to optimize PV arrays using ANNs. Numerous topologies were considered in the study, including series-parallel topology, parallel topology, bridge link topology, honeycomb topology, and total cross tied. The simulation results demonstrate that the proposed technique finds the best topologies to deploy the PV panels with better accuracy (98%) than previous methods. He has demonstrated the comparison with other Ml techniques like support vector machine(SVM), Naive Bayes(NB), and k-nearest neighbor(KNN).

In [136], a deep neural network (DNN) method is presented for calculating PV size, tilt, and azimuth solely from meter data. This method can predict PV size with an error of 2.09% in a data set with fixed tilt and azimuth values and 3.98% in a data set with variable tilt and azimuths. The mean absolute percentage errors of PV tilt and azimuth are 10.1 % and 2.8 %, respectively. This proposed strategy is more effective than the benchmark linear regression approach. However, in simulations, the same irradiance measurement data were used to generate several of the PV profiles, and they used synthetic PV generated data for the evaluation, whereas the results in the literature to which they are being compared are based on actual PV generation profiles. In order to meet a home load in a region without electricity, Vijay and Saravanan [137] presented a new approach using a Bayesian optimization-based regression tree (BORT) technique for calculating the ideal size of a PV system. The proposed research assesses the forecasted global horizontal irradiance (GHI) values and the effectiveness of the Bayesian optimization was evaluated in comparison to other existing approaches. According to the results of the analysis, the BORT algorithm calculates GHI with a lower mean square error than the current machine

TABLE 5. Recent research review on control methodology of PV using ML.

Ref	Research	Methods	Results	Year
[119]	Maximum power point tracking with multivariate time series approach	Transformer	MAPE Score of 0.47%, power efficiency of 99.54% and peak power efficiency of 99.98%	2022
[120]	Developed MADRL algorithm for voltage regulation usign PV inverters and tested in IEEE 123-bus system	Attention Based MADDPG Algorithm	Average Voltage Deviation on- • test data 0.12% • IEEE 123-bus system 0.12%	2020
[123]	ML to predict proportional integral value that is used to find duty cycle for maximum power	• Ensemble boosted tree • Ensemble bagged trees • Ensemble subspaces • Fine, Medium & Coarse KNN • Fine, Medium & Coarse Tree	Best three model with accuracy- • Ensemble boosted tree 98.1% • Ensemble subspaces 97% • Fine Tree 96%	2022
[124]	Predicted P_{mp} & V_{mp} to determine duty cycle for MPPT boost converter based on irradiance and temperature	Linear & Non-Linear Regression Model	LR - • RMSE 0.202 (P_{mp}), 0.0314 (V_{mp}) • Avg. Efficiency (95.21 - 98.41)% NLR - • RMSE 0.0206 (P_{mp}), 0.124 (V_{mp}) • Avg. Efficiency (95.32 - 98.45)%	2022
[125]	Reduced the response time of P & O method	Modified P & O Algorithm with Multivariate Regression ML Method	• MAPE 0.2% • average efficiency in detecting MPP 99.8%	2019
[126]	Studied MPPT by determining optimal duty cycle	Gradient Ascent	• Response time 10s • PV power increment 19% • PV power increment 68% under shading	2021
[127]	MPPT by predicting the reference voltage at different weather condition	BPNN	98% accuracy	2022
[128]	PV Power generation based on several weather parameters	BPNN, ELMAN NN	• SMAPE for BPNN 1.64% • SMAPE for ELMAN NN 1.02%	2021
[129]	MPPT using reinforcement learning without relying on any labeled initial knowledge	RLMPPT	Under varying load and environment condition achieves a very fast response and low power error	2017
[130]	MPPT achieved by predicting maximum power and voltage	DT	• Tracking efficiency more than 93.93% • Response time 0.16s • Settling time 0.27s	2022
[132]	LVRT	FCS-MPC Controller, NN fault classifier	• Voltage sags detection time 2ms • Recognition accuracy 98.6%	2022

learning techniques. The optimal design of a PV system was effectively carried out using the predicted GHI. Table 6 contains a summary of the models that have been studied in this section.

E. PV MANAGEMENT

Consumers are finding Energy Management Systems (EMS) increasingly valuable for lowering their electricity costs and improving efficiency, thanks to the integration of smart grid

TABLE 6. Recent research review on sizing, management and site selection of PV using ML System.

Type	Ref	Research	Methods	Results	Year
Sizing Methodologies	[134]	mapping small-scale solar arrays at a large scale	CNN	0.76% (precision) 0.91% (correlation coefficient)	2019
	[135]	finds the best topologies to deploy the PV panels with a better degree of accuracy	ANN	98%	2022
	[137]	to meet a home load in a region without electricity	BORT	52.57 (RMSE) 2763.7 (MSE) 20.985 (MAE)	2022
	[136]	calculating PV size, tilt, and azimuth	DNN	2.09 % (Error for fixed data) 3.98 % (Error for variable data)	2020
PV Management	[138]	using a central EMS to reduce the grid power consumption	LR RT ANN	0.0013 (RMSE for battery power validation) 0.407 (RMSE for DG power validation)	2023
	[139]	a novel demand response program (DRP) for renewable based microgrids (MGs)	GAN	0.6272 (RMSE for forecasting of PV system output power) 0.5831 (RMSE for forecasting of tidal system output power)	2023
	[78]	a multi-objective predictive EMS management plan for grid-connected PV-battery HES	CNN	93.08 % (coefficient of determination of energy output), and 97.25 % (coefficient of determination of electric load)	2021
	[140]	forecasting-based optimization for residential EMSs, sensitivity analyses for seasonality and forecasting	gradient-boosted decision tree	0.36 ($rRMSE_{avg}$ of PV for Prosumer 1), 0.664 ($rRMSE_{avg}$ of load forecasts for Prosumer 1), 0.411 ($rRMSE_{avg}$ of PV for Prosumer 2), 0.701 ($rRMSE_{avg}$ of load forecasts for Prosumer 2),	2023
Site Selection	[141]	point-to-point connection between meteorological station observations and satellite data	RF	0.89 (R^2 for weekly forecasting), 0.83 (R^2 for daily forecast)	2020

technologies, including smart meters and demand-response algorithms. The primary aim of an Energy Management System is to arrange the energy flow within the system in real-time, by reducing a pre-determined objective function, while preserving a secure, dependable, and safe system operation. It is possible to increase the effectiveness, stability, and performance of solar systems and lower maintenance and operation costs by utilizing machine learning algorithms in photovoltaic EMS. This section will review some of the ML methods used in PV management.

The authors of [78] proposed a multi-objective predictive EMS management plan for (HES) grid-connected PV-battery hybrid energy systems (HES). As one of the three levels of the control approach, they primarily used CNNs to estimate energy production and electric load. Using hourly data on energy and load, the suggested technique was evaluated for both fixed and variable power costs. In simulations, high coefficient of determination was found in the simulation result for the predictions of energy output and electric load, which were 93.08 %, and 97.25%, respectively. This new model considerably reduced carbon dioxide emissions and power costs compared to homes without hybrid energy systems and homes with hybrid energy systems but no energy management plan. Regarding trade-offs between financial profitability, computational burden, and privacy, Müller et al. [140] comprehensively analyzed forecasting-based optimization for residential EMSs. The study's base consists of two PV battery systems owned by actual prosumers, each with an EMS, a stationary storage system, and rooftop solar power. They also investigated the sensitivity analyses for seasonality and forecasting horizon. Gradient-boosted decision trees were used for forecasting purposes. $rRMSE_{avg}$ of PV for Prosumer 1 and 2 were 0.36 and 0.411, respectively, whereas the $rRMSE_{avg}$ of loads were 0.604 and 0.701, accordingly. The results indicated that, compared to a situation without a battery, the maximum potential profit in the two prosumer scenarios over 14 months was 466 and 555 euros. This model can be applied to new systems without an extensive data record, over 90 % of the potential financial gain was attained at a significantly lower computational complexity than in previous cases with comparable savings. This scenario was resistant to changes in weather model inputs and did not have a dependency on complex pricing forecasts.

In [138], a centralized EMS was used in combination with diesel generators sets (DGs), battery storage systems (BSS) and PV systems to reduce the grid power consumption and

also to make the most of the use of hybrid power sources in the building. They compared different modules, like linear regression (LR), regression trees (RT), and artificial neural networks (ANN) to predict the most accurate outcomes. Compared to the other algorithms, LR reduced the total grid power by up to 7346.103 KWh, and coarse tree by up to 7414.583 KWh. Regarding reliably forecasting standard grid power in a distribution network, it was apparent that Linear Regression (LR) and Regression Coarse Tree (RCT) performed better than other methods. The outcomes were validated using Matlab simulation software. The technique presented in [139] proposed a novel demand response program (DRP) for renewable-based microgrids (MGs), which considers tidal units and solar energy in the power systems. They developed a predictive scheduling structure that was optimal for renewable MGs. The system modeled and predicted different renewable energy sources (RESs) of PV and tidal units. Grey wolf optimization algorithm (GWOA) and generative adversarial network (GAN) were combined in a novel optimization method that successfully solved the issue. The IEEE experimental grid simulation results showed that the suggested GAN layout outperformed the SVR, AR, and ANN in predicting the output power of the tidal and PV agents by providing lower RMSE and AAPE quantities. Table 6 provides information on the papers that have reported on estimating PV output power and forecasting irradiance.

F. SITE SELECTION FOR PV INSTALLATION

Recently, the concept of ''site adaptation'' has emerged in respect to solar energy installations. The reliability and overall quality of the solar radiation data collected during the solar resource evaluation process determine the profitability of any proposed solar power facility at that location. The most frequent solutions to this issue involve statistical techniques that adapt satellite estimations to in-situ data [142]. Only a few works have used ML techniques to tackle this issue.

To create a model that could precisely depict the point-to-point connection between meteorological station observations and satellite data, Narvaez et al. [141] evaluated different ML regression techniques. The input variables were the satellite-based data, while the estimated in-situ global horizontal irradiance measurements were the output. The author used ANN, RF, AdaBoost, and LR models for the evaluation. Quantile mapping, a traditional site adaption technique was used to analyze the outcomes of the various methods. It was evident that almost every ML model beat quantile mapping when comparing the RMSE, R^2, and MAE measures of the models. The random forest model performed the best across all benchmarks. The daily and weekly forecasts produced by the random forest model have minimum R^2 values of 0.89 and 0.83, accordingly. According to the results, site adaptation using machine learning models outperformed conventional methods by up to 38%.

VI. DISCUSSION AND RECOMMENDATION

The previous sections explore numerous ML strategies used for the fabrication phase and across different areas of PV systems. This part provides an overview of the current research directions and highlights some innovative methods that should be considered when evaluating the latest developments.

Our survey of ML applications in solar cell fabrication revealed that the RF, LR, XGBoost, and ANN algorithms are the most widely utilized techniques. These findings are visually presented in Figure 7, which provides a clear overview of the percentage distribution of each algorithm's usage. These results serve as valuable guidance for researchers and practitioners seeking to optimize the implementation of ML in solar cell fabrication. We also found that XGBoost outperforms other algorithms in optoelectronic prediction, while RF, LR, and ANN algorithms are better suited for predicting electrical parameters.

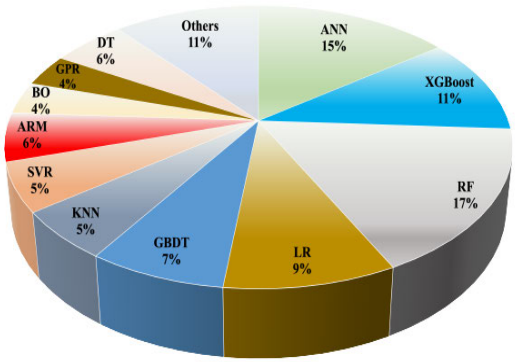

FIGURE 7. Statistical analysis for different algorithm used in the fabrication of PV cells.

Based on the results obtained from our survey, we have compiled the findings for solar cell fabrication in the following summary:

- In recent years, there has been a notable shift in research emphasis within the field of solar cell fabrication, with increasing attention directed toward PSCs in conjunction with the implementation of ML techniques.
- For ABX_3 perovskite A site cations have the most significant impacts on PSCs performance.
- FA^+ from the A site cation plays a substantial role in PSCs performance. We have observed from our reviews that combining FA^+ with other cations from the A site results in improved device performance. However, to achieve the best results, the percentage of FA^+ cations should be higher than the other cations.
- introduction of mixed cations (multi cation and 2D/3D crystal structure) as absorber layer in PSCs has emerged as a promising strategy for enhancing their performance.
- Bandgap around 1.55eV - 1.60eV would have been the best suited range for perovskite materials.

- Hole Transport Material should also been taken into consideration as it plays a crucial role in suppressing charge recombination at the perovskite interface.

While straightforward conventional methods still hold significance, many researchers are using NN to track the maximum power point of a PV system. Although NN exhibits superior performance compared to traditional methods, they still have limitations like complex optimization of hyper-parameters and need for large amount of data. Most papers that worked with MPPT utilized Matlab as their programming environment. However, there is a requirement to create additional standard scenarios using open-source programming languages like Python. In fault detection, our attention is drawn toward the lack of solid experimental results from testbeds and the evaluation of advanced ML techniques, such as vision transformers. These cutting-edge models have the potential to outperform other conventional methods like CNNs.

A critical challenge with PV systems is forecasting future energy generation. Currently, most forecasting algorithms rely on recurrent units. However, it is essential to assess more recent approaches to time-series prediction, such as transformer topologies, which focus entirely on attention mechanisms. For PV power forecasting, it is recommended to use ensemble algorithms, when working with little data and algorithms like SVR, LSTM and ELM (Extreme Learning Machine) are recommended for researchers with an ample amount of data. In our opinion, more work has to be done to create new platforms for EMS in which academics can evaluate their algorithms like OpenAI Gym. Investigating a wider range of data processing approaches is advisable because ML strategies for site adaptation are currently under development. Depending on the local environment, dust, snow accumulation, bird droppings, and atmospheric pollution can significantly reduce efficiency and power output. Therefore, routine cleaning can provide optimal working conditions. Some automated cleaning techniques include brush cleaning, Heliotex cleaning, electrostatic cleaning, robotic cleaning, etc [143]. However, the absence of research on ML applications in PV panel cleaning is particularly noteworthy. Future researchers may address this. The interconnection between several subsystems is a crucial aspect of using ML prediction tools. Additionally, the use of wireless networks that connect to robots cleaning the surfaces of the modules is growing in popularity.

While ML holds great promise for enhancing solar cell fabrication processes, several challenges must be overcome:

- Solar cell fabrication research heavily relies on experimental data, often scattered across various publications and not readily accessible in a structured format. Researchers face difficulties in obtaining datasets that accurately capture the complex interplay of process parameters, material properties, and cell performance metrics under specific lab conditions. This lack of data availability hampers developing and deploying ML models tailored for solar cell fabrication.
- In addition to integrating existing data, researchers need to focus on collecting lab-specific datasets that capture the nuances of their experimental setups and conditions. Collaborations between research institutions, industry partners, and data scientists can help to establish data collection protocols, encourage data sharing, and foster a culture of open science. Such efforts can provide researchers with valuable datasets that accurately represent real-world fabrication processes.
- According to the discussion, different models used in this field are not interpretable, require a large amount of training data, and are computationally inefficient to implement in real life. Most of the research considers a fixed environment, as there is no certainty that those models could quickly adapt to new solar systems with different characteristics and climatic conditions. Therefore, further study is needed to confirm that an algorithm trained on one PV system can successfully execute in another method to ensure the technology's widespread use.

Overcoming these challenges will pave the way for more efficient, reliable, and sustainable solar energy solutions.

VII. CONCLUSION

This study provides an in-depth survey of the recent ML application in photovoltaic research. We have reviewed over 100 articles dating from 2020 to 2022. The review discussed critical areas for ML application in both PV cell fabrication and PV system optimization in an end-to-end manner. Numerous difficulties relating to the fabrication,installation and analysis of PV systems could be solved using ML. Additionally, we have identified a number of scopes that are possible areas of exploration for future work in this area -

- DL is not normally used in the fabrication process. We have identified that the need for a large and quality dataset is the main limiting factor against utilizing deep learning architecture in the fabrication processes.
- RL-based approach is the most effective in MPPT tasks. Studies at this stage mainly utilize the dataset available through simulations and often lack sufficient generalization to apply to real-world scenarios. Further study is needed to confirm that an algorithm trained on one PV system can successfully track MPP in another method to ensure the technology's widespread use. Enhancing the speed and accuracy of MPP recognition, creating control schemes that maximize power generation in PV systems, and looking at the potential to increase PV inverter efficiency should be the main priorities.
- The study explores prevalent faults that significantly hamper the performance, durability, and safety of PVS. Moreover, various ML methodologies employed by researchers to mitigate these faults are examined. It is evident that, different studies mostly focused on different set of faults, rather than adopting a more unified approach.

- ML agents that are trained on a particular system configuration often necessitate retraining when applied to a different PVS. Nonetheless, the availability of adequate abnormal condition data for proper training may be limited, leading to potential performance issues. Additionally, since the models only rely on historical data with known fault type, they may not recognize new or unknown faults. To address this challenge, the establishment of a large benchmark dataset comprising diverse fault scenarios is essential. Consequently, we suggest knowledge transfer approach, utilizing this extensive dataset, would prove to be effective in enhancing fault detection performance across various system configurations.
- Although ML has proved effectiveness in optimizing system performances, certain aspects, including PV panel recycling, reuse after decommissioning, and fault ride-through schemes, are crucial factors. Regrettably, the available literature on ML approaches related to these specific topics is lacking. Consequently, further research is necessary to explore and develop ML-based methodologies to address these important areas.
- While existing literature presents models with remarkable accuracies in personal computing workspaces, the practical deployment of ML agents in real-world PVS (Physical Visual Systems) is often neglected, indicating a significant research gap. Addressing this aspect deserves increased attention and investigation. Pipelines like TinyML or EdgeAI can be explored, which hold potential for facilitating the integration of ML agents into PVS applications.

The data in the industry at the moment seem to be inconsistent. However, it is anticipated that the issue of insufficient data will be solved by the introduction of internet of things solutions, the use of multiple sensors, the integration of drone-provided video streams for maintenance, and the use of natural language processing techniques. The article can be used as a reference for further research into PV technologies to design and manufacture efficient PV cell structure as well as to guarantee safe and reliable operation. In summary, this work will be beneficial for readers, researchers, and engineers working in the field of PV systems, providing a complete insight starting from fabrication to operation.

REFERENCES

[1] M. Abdel-Basset, H. Hawash, R. K. Chakrabortty, and M. Ryan, "PV-Net: An innovative deep learning approach for efficient forecasting of short-term photovoltaic energy production," *J. Cleaner Prod.*, vol. 303, Jun. 2021, Art. no. 127037, doi: 10.1016/j.jclepro.2021.127037.

[2] B. Darbari and S. Rashidi, "Performance analysis for single slope solar still enhanced with multi-shaped floating porous absorber," *Sustain. Energy Technol. Assessments*, vol. 50, Mar. 2022, Art. no. 101854, doi: 10.1016/j.seta.2021.101854.

[3] H. Sharadga, S. Hajimirza, and R. S. Balog, "Time series forecasting of solar power generation for large-scale photovoltaic plants," *Renew. Energy*, vol. 150, pp. 797–807, May 2020, doi: 10.1016/j.renene.2019.12.131.

[4] K. Wang, X. Qi, and H. Liu, "A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network," *Appl. Energy*, vol. 251, Oct. 2019, Art. no. 113315, doi: 10.1016/j.apenergy.2019.113315.

[5] M. Mammeri, L. Dehimi, H. Bencherif, and F. Pezzimenti, "Paths towards high perovskite solar cells stability using machine learning techniques," *Sol. Energy*, vol. 249, pp. 651–660, Jan. 2023, doi: 10.1016/j.solener.2022.12.002.

[6] Y. Hu, X. Hu, L. Zhang, T. Zheng, J. You, B. Jia, Y. Ma, X. Du, L. Zhang, J. Wang, B. Che, T. Chen, and S. Liu, "Machine-learning modeling for ultra-stable high-efficiency perovskite solar cells," *Adv. Energy Mater.*, vol. 12, no. 41, Nov. 2022, Art. no. 2201463, doi: 10.1002/aenm.202201463.

[7] J. Pastuszak and P. Węgierek, "Photovoltaic cell generations and current research directions for their development," *Materials*, vol. 15, no. 16, p. 5542, Aug. 2022, doi: 10.3390/ma15165542.

[8] D. Zhou, T. Zhou, Y. Tian, X. Zhu, and Y. Tu, "Perovskite-based solar cells: Materials, methods, and future perspectives," *J. Nanomater.*, vol. 2018, pp. 1–15, Jan. 2018, doi: 10.1155/2018/8148072.

[9] S. Mathew, A. Yella, P. Gao, R. Humphry-Baker, B. F. E. Curchod, N. Ashari-Astani, I. Tavernelli, U. Rothlisberger, M. K. Nazeeruddin, and M. Grätzel, "Dye-sensitized solar cells with 13% efficiency achieved through the molecular engineering of porphyrin sensitizers," *Nature Chem.*, vol. 6, no. 3, pp. 242–247, Mar. 2014, doi: 10.1038/nchem.1861.

[10] D. Sherman, J. Marquez, Y. Ramirez, M. Urbina, A. Meza, and D. Hodges, "Monocrystalline silicon solar cell simulation with reduced absorber thickness and efficiency exceeding 25%," in *Proc. IEEE 48th Photovoltaic Spec. Conf. (PVSC)*, Jun. 2021, pp. 2640–2642, doi: 10.1109/PVSC43889.2021.9518818.

[11] B. Parida, S. Iniyan, and R. Goic, "A review of solar photovoltaic technologies," *Renew. Sustain. Energy Rev.*, vol. 15, no. 3, pp. 1625–1636, Apr. 2011, doi: 10.1016/j.rser.2010.11.032.

[12] M. K. Rao, D. N. Sangeetha, M. Selvakumar, Y. N. Sudhakar, and M. G. Mahesha, "Review on persistent challenges of perovskite solar cells' stability," *Sol. Energy*, vol. 218, pp. 469–491, Apr. 2021, doi: 10.1016/j.solener.2021.03.005.

[13] J. Marquez, Y. Ramirez, G. Rodriguez, M. Urbina, D. Sherman, A. Meza, and D. Hodges, "Timeline for successful commercialization of thin-film perovskite solar cells," in *Proc. IEEE 48th Photovoltaic Spec. Conf. (PVSC)*, Jun. 2021, pp. 2426–2429, doi: 10.1109/PVSC43889.2021.9518634.

[14] L. Meng, J. You, and Y. Yang, "Addressing the stability issue of perovskite solar cells for commercial applications," *Nature Commun.*, vol. 9, no. 1, p. 5265, Dec. 2018, doi: 10.1038/s41467-018-07255-1.

[15] X. Yang, L. Li, Q. Tao, W. Lu, and M. Li, "Rapid discovery of narrow bandgap oxide double perovskites using machine learning," *Comput. Mater. Sci.*, vol. 196, Aug. 2021, Art. no. 110528, doi: 10.1016/j.commatsci.2021.110528.

[16] G. Zhang, J. Yuan, Y. Mao, and Y. Huang, "Two-dimensional Janus material $MoS_{2(1-x)}Se_{2x}$ (0<x<1) for photovoltaic applications: A machine learning and density functional study," *Comput. Mater. Sci.*, vol. 186, Jan. 2021, Art. no. 109998, doi: 10.1016/j.commatsci.2020.109998.

[17] T. J. Jacobsson, "An open-access database and analysis tool for perovskite solar cells based on the FAIR data principles," *Nature Energy*, vol. 7, no. 1, pp. 107–115, Dec. 2021, doi: 10.1038/s41560-021-00941-3.

[18] M. O. Yildirim, E. C. Gok, N. H. Hemasiri, E. Eren, S. Kazim, A. U. Oksuz, and S. Ahmad, "A machine learning approach for metal oxide based polymer composites as charge selective layers in perovskite solar cells," *ChemPlusChem*, vol. 86, no. 5, pp. 785–793, May 2021, doi: 10.1002/cplu.202100132.

[19] K.-G. Lim, S. Ahn, Y.-H. Kim, Y. Qi, and T.-W. Lee, "Universal energy level tailoring of self-organized hole extraction layers in organic solar cells and organic–inorganic hybrid perovskite solar cells," *Energy Environ. Sci.*, vol. 9, no. 3, pp. 932–939, 2016, doi: 10.1039/C5EE03560K.

[20] N. J. Jeon, H. Na, E. H. Jung, T.-Y. Yang, Y. G. Lee, G. Kim, H.-W. Shin, S. L. Seok, J. Lee, and J. Seo, "A fluorene-terminated hole-transporting material for highly efficient and stable perovskite solar cells," *Nature Energy*, vol. 3, no. 8, pp. 682–689, Jul. 2018, doi: 10.1038/s41560-018-0200-6.

[21] A. McEvoy, T. Markvart, and L. Castaner, "Principles of solar cell operation," in *Practical Handbook of Photovoltaics Fundamentals and Applications*, 2nd ed. Waltham, MA, USA: Elsevier, 2011, pp. 7–10. [Online]. Available: https://shopp.elsevier.com/books/practical-handbook-of-photovoltaics/mcevoy/978-0-12-385934-1

[22] U. Akyol, D. Akal, and A. Durak, "Estimation of power output and thermodynamic analysis of standard and finned photovoltaic panels," *Energy Sources A, Recovery, Utilization, Environ. Effects*, vol. 45, pp. 1–20, May 2021, doi: 10.1080/15567036.2021.1928337.

[23] L. B. Bosman, W. D. Leon-Salas, W. Hutzel, and E. A. Soto, "PV system predictive maintenance: Challenges, current approaches, and opportunities," *Energies*, vol. 13, no. 6, p. 1398, Mar. 2020, doi: 10.3390/en13061398.

[24] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning," *Decis. Anal. J.*, vol. 3, Jun. 2022, Art. no. 100071, doi: 10.1016/j.dajour.2022.100071.

[25] S. B. Kotsiantis, "Decision trees: A recent overview," *Artif. Intell. Rev.*, vol. 39, no. 4, pp. 261–283, Apr. 2013, doi: 10.1007/s10462-011-9272-4.

[26] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[27] H. Rao, X. Shi, A. K. Rodrigue, J. Feng, Y. Xia, M. Elhoseny, X. Yuan, and L. Gu, "Feature selection based on artificial bee colony and gradient boosting decision tree," *Appl. Soft Comput.*, vol. 74, pp. 634–642, Jan. 2019, doi: 10.1016/j.asoc.2018.10.036.

[28] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[29] R. Grosse, *Lecture 5: Multilayer Perceptrons*. Toronto, On, Canada: Univ. Toronto, 2019.

[30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[31] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 6, no. 2, pp. 107–116, Apr. 1998, doi: 10.1142/S0218488598000094.

[32] A. Baul, G. C. Sarker, P. K. Sadhu, V. P. Yanambaka, and A. Abdelgawad, "XTM: A novel transformer and LSTM-based model for detection and localization of formally verified FDI attack in smart grid," *Electronics*, vol. 12, no. 4, p. 797, Feb. 2023, doi: 10.3390/electronics12040797.

[33] N.-G. Park, "Organometal perovskite light absorbers toward a 20% efficiency low-cost solid-state mesoscopic solar cell," *J. Phys. Chem. Lett.*, vol. 4, no. 15, pp. 2423–2429, Jul. 2013, doi: 10.1021/jz400892a.

[34] Z. Xiao, C. Bi, Y. Shao, Q. Dong, Q. Wang, Y. Yuan, C. Wang, Y. Gao, and J. Huang, "Efficient, high yield perovskite photovoltaic devices grown by interdiffusion of solution-processed precursor stacking layers," *Energy Environ. Sci.*, vol. 7, no. 8, pp. 2619–2623, 2014, doi: 10.1039/C4EE01138D.

[35] W. Zhang, G. E. Eperon, and H. J. Snaith, "Metal halide perovskites for energy applications," *Nature Energy*, vol. 1, no. 6, May 2016, Art. no. 16048, doi: 10.1038/nenergy.2016.48.

[36] A. A. Zhumekenov, M. I. Saidaminov, M. A. Haque, E. Alarousu, S. P. Sarmah, B. Murali, I. Dursun, X.-H. Miao, A. L. Abdelhady, T. Wu, O. F. Mohammed, and O. M. Bakr, "Formamidinium lead halide perovskite crystals with unprecedented long carrier dynamics and diffusion length," *ACS Energy Lett.*, vol. 1, no. 1, pp. 32–37, Apr. 2016, doi: 10.1021/acsenergylett.6b00002.

[37] S. D. Stranks, G. E. Eperon, G. Grancini, C. Menelaou, M. J. P. Alcocer, T. Leijtens, L. M. Herz, A. Petrozza, and H. J. Snaith, "Electron-hole diffusion lengths exceeding 1 micrometer in an organometal trihalide perovskite absorber," *Science*, vol. 342, no. 6156, pp. 341–344, Oct. 2013, doi: 10.1126/science.1243982.

[38] Q. Jiang, Y. Zhao, X. Zhang, X. Yang, Y. Chen, Z. Chu, Q. Ye, X. Li, Z. Yin, and J. You, "Surface passivation of perovskite film for efficient solar cells," *Nature Photon.*, vol. 13, no. 7, pp. 460–466, Jul. 2019, doi: 10.1038/s41566-019-0398-2.

[39] M. P. U. Haris, S. Kazim, and S. Ahmad, "Low-temperature-processed perovskite solar cells fabricated from presynthesized $CsFAPbI_3$ powder," *ACS Appl. Energy Mater.*, vol. 4, no. 3, pp. 2600–2606, Feb. 2021, doi: 10.1021/acsaem.0c03160.

[40] W. Yan, Y. Liu, Y. Zang, J. Cheng, Y. Wang, L. Chu, X. Tan, L. Liu, P. Zhou, W. Li, and Z. Zhong, "Machine learning enabled development of unexplored perovskite solar cells with high efficiency," *Nano Energy*, vol. 99, Aug. 2022, Art. no. 107394, doi: 10.1016/j.nanoen.2022.107394.

[41] N. Ali, S. Rauf, W. Kong, S. Ali, X. Wang, A. Khesro, C. P. Yang, B. Zhu, and H. Wu, "An overview of the decompositions in organometal halide perovskites and shielding with 2-dimensional perovskites," *Renew. Sustain. Energy Rev.*, vol. 109, pp. 160–186, Jul. 2019, doi: 10.1016/j.rser.2019.04.022.

[42] D. Lin, T. Zhang, J. Wang, M. Long, F. Xie, J. Chen, B. Wu, T. Shi, K. Yan, W. Xie, P. Liu, and J. Xu, "Stable and scalable 3D-2D planar heterojunction perovskite solar cells via vapor deposition," *Nano Energy*, vol. 59, pp. 619–625, May 2019, doi: 10.1016/j.nanoen.2019.03.014.

[43] E. C. Gok, M. O. Yildirim, M. P. U. Haris, E. Eren, M. Pegu, N. H. Hemasiri, P. Huang, S. Kazim, A. U. Oksuz, and S. Ahmad, "Predicting perovskite bandgap and solar cell performance with machine learning," *Sol. RRL*, vol. 6, no. 2, Feb. 2022, Art. no. 2100927, doi: 10.1002/solr.202100927.

[44] L. Valerio, A. De La Rosa, V. Rodriguez, C. Enriquez, A. Telles, Y. Ramirez, D. Rivera, J. Hierro, L. Bustamante, X. Tong, and D. Hodges, "Characterization and analysis of $FA_xCs_{(1-x)}$ $Pb(I_yBr_{(1-y)})_3$ perovskite solar cells with thickness controlled transport layers for performance optimization," *AIP Adv.*, vol. 9, no. 10, Oct. 2019, Art. no. 105006, doi: 10.1063/1.5123400.

[45] C. Liu, J. Fan, H. Li, C. Zhang, and Y. Mai, "Highly efficient perovskite solar cells with substantial reduction of lead content," *Sci. Rep.*, vol. 6, no. 1, Oct. 2016, Art. no. 35705, doi: 10.1038/srep35705.

[46] A. Hossain, M. M. Hasan, M. S. Rahman, and M. A. M. Hossain, "Fully lead-free all perovskite tandem solar cell with improved efficiency: Device simulation using SCAPS-1D," in *Proc. IEEE Region Symp. (TENSYMP)*, Jun. 2020, pp. 1221–1224, doi: 10.1109/TENSYMP50017.2020.9230927.

[47] J. Li, H.-L. Cao, W.-B. Jiao, Q. Wang, M. Wei, I. Cantone, J. Lü, and A. Abate, "Biological impact of lead from halide perovskites reveals the risk of introducing a safe threshold," *Nature Commun.*, vol. 11, no. 1, p. 310, Jan. 2020, doi: 10.1038/s41467-019-13910-y.

[48] K. Miyano, N. Tripathi, M. Yanagida, and Y. Shirai, "Lead halide perovskite photovoltaic as a model p–i–n diode," *Accounts Chem. Res.*, vol. 49, no. 2, pp. 303–310, Jan. 2016, doi: 10.1021/acs.accounts.5b00436.

[49] Ç. Odabaşı and R. Yıldırım, "Machine learning analysis on stability of perovskite solar cells," *Sol. Energy Mater. Sol. Cells*, vol. 205, Feb. 2020, Art. no. 110284, doi: 10.1016/j.solmat.2019.110284.

[50] Y. Lu, D. Wei, W. Liu, J. Meng, X. Huo, Y. Zhang, Z. Liang, B. Qiao, S. Zhao, D. Song, and Z. Xu, "Predicting the device performance of the perovskite solar cells from the experimental parameters through machine learning of existing experimental results," *J. Energy Chem.*, vol. 77, pp. 200–208, Feb. 2023, doi: 10.1016/j.jechem.2022.10.024.

[51] Y. Liu, W. Yan, S. Han, H. Zhu, Y. Tu, L. Guan, and X. Tan, "How machine learning predicts and explains the performance of perovskite solar cells," *Sol. RRL*, vol. 6, no. 6, Jun. 2022, Art. no. 2101100, doi: 10.1002/solr.202101100.

[52] Z. Guo and B. Lin, "Machine learning stability and band gap of lead-free halide double perovskite materials for perovskite solar cells," *Sol. Energy*, vol. 228, pp. 689–699, Nov. 2021, doi: 10.1016/j.solener.2021.09.030.

[53] S. Jiang, C.-C. Wu, F. Li, Y.-Q. Zhang, Z.-H. Zhang, Q.-H. Zhang, Z.-J. Chen, B. Qu, L.-X. Xiao, and M.-L. Jiang, "Machine learning (ML)-assisted optimization doping of KI in $MAPbI_3$ solar cells," *Rare Met.*, vol. 40, no. 7, pp. 1698–1707, Jul. 2021, doi: 10.1007/s12598-020-01579-y.

[54] B. Yılmaz, Ç. Odabaşı, and R. Yıldırım, "Efficiency and stability analysis of 2D/3D perovskite solar cells using machine learning," *Energy Technol.*, vol. 10, no. 3, Mar. 2022, Art. no. 2100948, doi: 10.1002/ente.202100948.

[55] P. Huang, A. Hernández, S. Kazim, J. Follana-Berná, J. Ortiz, L. Lezama, Á. Sastre-Santos, and S. Ahmad, "Asymmetrically substituted phthalocyanines as dopant-free hole selective layers for reliability in perovskite solar cells," *ACS Appl. Energy Mater.*, vol. 4, no. 9, pp. 10124–10135, Sep. 2021, doi: 10.1021/acsaem.1c02039.

[56] X. Ding, H. Wang, C. Chen, H. Li, Y. Tian, Q. Li, C. Wu, L. Ding, X. Yang, and M. Cheng, "Passivation functionalized phenothiazine-based hole transport material for highly efficient perovskite solar cell with efficiency exceeding 22%," *Chem. Eng. J.*, vol. 410, Apr. 2021, Art. no. 128328, doi: 10.1016/j.cej.2020.128328.

[57] Y. Tian, L. Tao, C. Chen, H. Lu, H. Li, X. Yang, and M. Cheng, "Facile synthesized fluorine substituted benzothiadiazole based dopant-free hole transport material for high efficiency perovskite solar cell," *Dyes Pigments*, vol. 184, Jan. 2021, Art. no. 108786, doi: 10.1016/j.dyepig.2020.108786.

[58] M. O. Yildirim, E. C. G. Yildirim, E. Eren, P. Huang, M. P. U. Haris, S. Kazim, J. Vanschoren, A. U. Oksuz, and S. Ahmad, "Automated machine learning approach in material discovery of hole selective layers for perovskite solar cells," *Energy Technol.*, vol. 11, no. 1, Jan. 2023, Art. no. 2200980, doi: 10.1002/ente.202200980.

[59] F. Hutter, L. Kotthoff, and J. Vanschoren, "Hyperparameter optimization," in *Automated Machine Learning*, 1st ed., Cham, Switzerland: Springer, 2019, pp. 3–34, doi: 10.1007/978-3-030-05318-5.
[60] W. Xu, T. B. Daunis, R. T. Piper, and J. W. P. Hsu, "Effects of photonic curing processing conditions on $MAPbI_3$ film properties and solar cell performance," *ACS Appl. Energy Mater.*, vol. 3, no. 9, pp. 8636–8645, Aug. 2020, doi: 10.1021/acsaem.0c01243.
[61] J. Troughton, M. J. Carnie, M. L. Davies, C. Charbonneau, E. H. Jewell, D. A. Worsley, and T. M. Watson, "Photonic flash-annealing of lead halide perovskite solar cells in 1 ms," *J. Mater. Chem. A*, vol. 4, no. 9, pp. 3471–3476, 2016, doi: 10.1039/C5TA09431C.
[62] W. Xu, Z. Liu, R. T. Piper, and J. W. P. Hsu, "Bayesian optimization of photonic curing process for flexible perovskite photovoltaic devices," *Sol. Energy Mater. Sol. Cells*, vol. 249, Jan. 2023, Art. no. 112055, doi: 10.1016/j.solmat.2022.112055.
[63] X. Cai, F. Liu, A. Yu, J. Qin, M. Hatamvand, I. Ahmed, J. Luo, Y. Zhang, H. Zhang, and Y. Zhan, "Data-driven design of high-performance $MASn_xPb_{1-x}I_3$ perovskite materials by machine learning and experimental realization," *Light, Sci. Appl.*, vol. 11, no. 1, p. 234, Jul. 2022, doi: 10.1038/s41377-022-00924-3.
[64] W. Ke and M. G. Kanatzidis, "Prospects for low-toxicity lead-free perovskite solar cells," *Nature Commun.*, vol. 10, no. 1, p. 965, Feb. 2019, doi: 10.1038/s41467-019-08918-3.
[65] D. R. Hodges, L. V. Frias, A. D. L. Rosa, A. I. Leyva, and X. Tong, "Synchrotron and optical probing of mixed lead halide perovskites for photovoltaics," *Proc. SPIE*, vol. 11474, pp. 26–34, Aug. 2020, doi: 10.1117/12.2569674.
[66] M. T. Islam, M. R. Jani, S. Rahman, K. M. Shorowordi, S. S. Nishat, D. Hodges, S. Banerjee, H. Efstathiadis, J. Carbonara, and S. Ahmed, "Investigation of non-pb all-perovskite 4-T mechanically stacked and 2-T monolithic tandem solar devices utilizing SCAPS simulation," *Social Netw. Appl. Sci.*, vol. 3, no. 4, p. 504, Mar. 2021, doi: 10.1007/s42452-021-04487-7.
[67] T. Bak, K. Kim, E. Seo, J. Han, H. Sung, I. Jeon, and I. D. Jung, "Accelerated design of high-efficiency lead-free tin perovskite solar cells via machine learning," *Int. J. Precis. Eng. Manuf.-Green Technol.*, vol. 10, no. 1, pp. 109–121, Jan. 2023, doi: 10.1007/s40684-022-00417-z.
[68] Ç. Odabaşı and R. Yıldırım, "Performance analysis of perovskite solar cells in 2013–2018 using machine-learning tools," *Nano Energy*, vol. 56, pp. 770–791, Feb. 2019, doi: 10.1016/j.nanoen.2018.11.069.
[69] N. T. P. Hartono, J. Thapa, A. Tiihonen, F. Oviedo, C. Batali, J. J. Yoo, Z. Liu, R. Li, D. F. Marrón, M. G. Bawendi, T. Buonassisi, and S. Sun, "How machine learning can help select capping layers to suppress perovskite degradation," *Nature Commun.*, vol. 11, no. 1, p. 4172, Aug. 2020, doi: 10.1038/s41467-020-17945-4.
[70] Ç. Odabaşı and R. Yıldırım, "Assessment of reproducibility, hysteresis, and stability relations in perovskite solar cells using machine learning," *Energy Technol.*, vol. 8, no. 12, Dec. 2020, Art. no. 1901449, doi: 10.1002/ente.201901449.
[71] E. A. J. Abadi, H. Sahu, S. M. Javadpour, and M. Goharimanesh, "Interpretable machine learning for developing high-performance organic solar cells," *Mater. Today Energy*, vol. 25, Apr. 2022, Art. no. 100969, doi: 10.1016/j.mtener.2022.100969.
[72] Y. A. Ramirez, A. D. L. Rosa, C. H. Enriquez, D. A. Rivera, V. M. Rodriguez, A. J. Telles, L. V. Frias, and D. R. Hodges, "High-voltage hybrid organic-inorganic perovskite solar cells," in *Proc. IEEE 48th Photovoltaic Spec. Conf. (PVSC)*, Jun. 2021, pp. 2303–2306, doi: 10.1109/PVSC43889.2021.9518561.
[73] C. Zhu, W. Liu, Y. Li, X. Huo, H. Li, K. Guo, B. Qiao, S. Zhao, Z. Xu, H. Zhao, and D. Song, "Key factors governing the device performance of CIGS solar cells: Insights from machine learning," *Sol. Energy*, vol. 228, pp. 45–52, Nov. 2021, doi: 10.1016/j.solener.2021.09.031.
[74] G. Salman, S. M. Goodnick, A. R. Shaik, and D. Vasileska, "Machine learning for optimal copper doping profile design in CdTe solar cells," in *Proc. IEEE 48th Photovoltaic Spec. Conf. (PVSC)*, Jun. 2021, pp. 540–543, doi: 10.1109/PVSC43889.2021.9518455.
[75] A. R. Shaik, D. Brinkman, I. Sankin, C. Ringhofer, D. Krasikov, H. Kang, B. Benes, and D. Vasileska, "PVRD-FASP: A unified solver for modeling carrier and defect transport in photovoltaic devices," *IEEE J. Photovolt.*, vol. 9, no. 6, pp. 1602–1613, Nov. 2019, doi: 10.1109/JPHOTOV.2019.2937238.
[76] W. Xiao, G. Nazario, H. Wu, H. Zhang, and F. Cheng, "A neural network based computational model to predict the output power of different types of photovoltaic cells," *PLoS ONE*, vol. 12, no. 9, Sep. 2017, Art. no. e0184561, doi: 10.1371/journal.pone.0184561.
[77] C. S. Solanki and H. K. Singh, *Anti-reflection Coatings with Textured Surface for c-Si Solar Cells*. Singapore: Springer, 2017, pp. 99–114, doi: 10.1007/978-981-10-4771-8_6.
[78] K. Shivam, J.-C. Tzou, and S.-C. Wu, "A multi-objective predictive energy management strategy for residential grid-connected PV-battery hybrid systems based on machine learning technique," *Energy Convers. Manage.*, vol. 237, Jun. 2021, Art. no. 114103, doi: 10.1016/j.enconman.2021.114103.
[79] M. Kaya and S. Hajimirza, "Rapid optimization of external quantum efficiency of thin film solar cells using surrogate modeling of absorptivity," *Sci. Rep.*, vol. 8, no. 1, p. 8170, May 2018, doi: 10.1038/s41598-018-26469-3.
[80] S. Wasmer, K. Hübener, and B. Klöter, "Explaining the efficiencies of mass-produced p-type Cz–Si solar cells by interpretable machine learning," *Sol. RRL*, vol. 6, no. 5, May 2022, Art. no. 2100477, doi: 10.1002/solr.202100477.
[81] S. Wanka, D. Rychtarik, J. Müller, S. Geissler, P. Kappe, M. Spallek, U. vom Bauer, C. Ludwig, and P. Wawer, "Tra.Q—Laser marking for single wafer identification—Production experience from 100 million wafers," in *Proc. 37th IEEE Photovoltaic Spec. Conf.*, Jun. 2011, pp. 1101–1104, doi: 10.1109/PVSC.2011.6186144.
[82] Y.-Y. Hong and R. A. Pula, "Methods of photovoltaic fault detection and classification: A review," *Energy Rep.*, vol. 8, pp. 5898–5929, Nov. 2022, doi: 10.1016/j.egyr.2022.04.043.
[83] M. K. Alam, F. Khan, J. Johnson, and J. Flicker, "A comprehensive review of catastrophic faults in PV arrays: Types, detection, and mitigation techniques," *IEEE J. Photovolt.*, vol. 5, no. 3, pp. 982–997, May 2015, doi: 10.1109/JPHOTOV.2015.2397599.
[84] S. Deng, Z. Zhang, C. Ju, J. Dong, Z. Xia, X. Yan, T. Xu, and G. Xing, "Research on hot spot risk for high-efficiency solar module," *Energy Proc.*, vol. 130, pp. 77–86, Sep. 2017, doi: 10.1016/j.egypro.2017.09.399.
[85] Y. Hu, W. Cao, J. Ma, S. J. Finney, and D. Li, "Identifying PV module mismatch faults by a thermography-based temperature distribution analysis," *IEEE Trans. Device Mater. Rel.*, vol. 14, no. 4, pp. 951–960, Dec. 2014, doi: 10.1109/TDMR.2014.2348195.
[86] M. Hussain, H. Al-Aqrabi, and R. Hill, "Statistical analysis and development of an ensemble-based machine learning model for photovoltaic fault detection," *Energies*, vol. 15, no. 15, p. 5492, Jul. 2022, doi: 10.3390/en15155492.
[87] W. Gao and R.-J. Wai, "Series arc fault detection of grid-connected PV system via SVD denoising and IEWT-TWSVM," *IEEE J. Photovolt.*, vol. 11, no. 6, pp. 1493–1510, Nov. 2021, doi: 10.1109/JPHOTOV.2021.3098376.
[88] R. Khemchandani and S. Chandra, "Twin support vector machines for pattern classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 905–910, May 2007, doi: 10.1109/TPAMI.2007.1068.
[89] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, "Salp swarm algorithm: A bio-inspired optimizer for engineering design problems," *Adv. Eng. Softw.*, vol. 114, pp. 163–191, Dec. 2017, doi: 10.1016/j.advengsoft.2017.07.002.
[90] X. Cai and R.-J. Wai, "Intelligent DC arc-fault detection of solar PV power generation system via optimized VMD-based signal processing and PSO–SVM classifier," *IEEE J. Photovolt.*, vol. 12, no. 4, pp. 1058–1077, Jul. 2022, doi: 10.1109/JPHOTOV.2022.3166919.
[91] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014, doi: 10.1109/TSP.2013.2288675.
[92] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. ICNN Int. Conf. Neural Netw.*, Nov. 1995, pp. 1942–1948, doi: 10.1109/ICNN.1995.488968.
[93] D. Adhya, S. Chatterjee, and A. K. Chakraborty, "Performance assessment of selective machine learning techniques for improved PV array fault diagnosis," *Sustain. Energy, Grids Netw.*, vol. 29, Mar. 2022, Art. no. 100582, doi: 10.1016/j.segan.2021.100582.
[94] L. Xu, Z. Pan, C. Liang, and M. Lu, "A fault diagnosis method for PV arrays based on new feature extraction and improved the fuzzy C-mean clustering," *IEEE J. Photovolt.*, vol. 12, no. 3, pp. 833–843, May 2022, doi: 10.1109/JPHOTOV.2022.3151330.
[95] Y. Sung, G. Yoon, J.-H. Bae, and S. Chae, "TL–LEDarcNet: Transfer learning method for low-energy series DC arc-fault detection in photovoltaic systems," *IEEE Access*, vol. 10, pp. 100725–100735, 2022, doi: 10.1109/ACCESS.2022.3208115.
[96] X. Chen, W. Gao, C. Hong, and Y. Tu, "A novel series arc fault detection method for photovoltaic system based on multi-input neural network," *Int. J. Electr. Power Energy Syst.*, vol. 140, Sep. 2022, Art. no. 108018, doi: 10.1016/j.ijepes.2022.108018.

[97] Y.-Y. Hong and R. A. Pula, "Detection and classification of faults in photovoltaic arrays using a 3D convolutional neural network," *Energy*, vol. 246, May 2022, Art. no. 123391, doi: 10.1016/j.energy.2022.123391.
[98] Z. Wang and T. Oates, "Imaging time-series to improve classification and imputation," 2015, *arXiv:1506.00327*.
[99] Z. Mustafa, A. S. A. Awad, M. Azzouz, and A. Azab, "Fault identification for photovoltaic systems using a multi-output deep learning approach," *Expert Syst. Appl.*, vol. 211, Jan. 2023, Art. no. 118551, doi: 10.1016/j.eswa.2022.118551.
[100] J. Van Gompel, D. Spina, and C. Develder, "Satellite based fault diagnosis of photovoltaic systems using recurrent neural networks," *Appl. Energy*, vol. 305, Jan. 2022, Art. no. 117874, doi: 10.1016/j.apenergy.2021.117874.
[101] S. Voutsinas, D. Karolidis, I. Voyiatzis, and M. Samarakou, "Development of a multi-output feed-forward neural network for fault detection in photovoltaic systems," *Energy Rep.*, vol. 8, pp. 33–42, Nov. 2022, doi: 10.1016/j.egyr.2022.06.107.
[102] N. V. Sridharan and V. Sugumaran, "Convolutional neural network based automatic detection of visible faults in a photovoltaic module," *Energy Sources, A, Recovery, Utilization, Environ. Effects*, pp. 1–16, Mar. 2021. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/15567036.2021.1905753, doi: 10.1080/15567036.2021.1905753.
[103] X. Lu, P. Lin, S. Cheng, G. Fang, X. He, Z. Chen, and L. Wu, "Fault diagnosis model for photovoltaic array using a dual-channels convolutional neural network with a feature selection structure," *Energy Convers. Manage.*, vol. 248, Nov. 2021, Art. no. 114777, doi: 10.1016/j.enconman.2021.114777.
[104] S. Fan, Y. Wang, S. Cao, B. Zhao, T. Sun, and P. Liu, "A deep residual neural network identification method for uneven dust accumulation on photovoltaic (PV) panels," *Energy*, vol. 239, Jan. 2022, Art. no. 122302, doi: 10.1016/j.energy.2021.122302.
[105] S. Lu, R. Ma, T. Sirojan, B. T. Phung, and D. Zhang, "Lightweight transfer nets and adversarial data augmentation for photovoltaic series arc fault detection with limited fault data," *Int. J. Electr. Power Energy Syst.*, vol. 130, Sep. 2021, Art. no. 107035, doi: 10.1016/j.ijepes.2021.107035.
[106] A. Shihavuddin, M. R. A. Rashid, M. H. Maruf, M. A. Hasan, M. A. U. Haq, R. H. Ashique, and A. A. Mansur, "Image based surface damage detection of renewable energy installations using a unified deep learning approach," *Energy Rep.*, vol. 7, pp. 4566–4576, Nov. 2021, doi: 10.1016/j.egyr.2021.07.045.
[107] S. Lu, T. Sirojan, B. T. Phung, D. Zhang, and E. Ambikairajah, "DA-DCGAN: An effective methodology for DC series arc fault diagnosis in photovoltaic systems," *IEEE Access*, vol. 7, pp. 45831–45840, 2019, doi: 10.1109/ACCESS.2019.2909267.
[108] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019, doi: 10.1109/TII.2018.2864759.
[109] H. Guo, S. Hu, F. Wang, and L. Zhang, "A novel method for quantitative fault diagnosis of photovoltaic systems based on data-driven," *Electr. Power Syst. Res.*, vol. 210, Sep. 2022, Art. no. 108121, doi: 10.1016/j.epsr.2022.108121.
[110] W. Ahmed, A. Hanif, K. D. Kallu, A. Z. Kouzani, M. U. Ali, and A. Zafar, "Photovoltaic panels classification using isolated and transfer learned deep neural models using infrared thermographic images," *Sensors*, vol. 21, no. 16, p. 5668, Aug. 2021, doi: 10.3390/s21165668.
[111] E. Miranda, J. F. G. Fierro, G. Narváez, L. F. Giraldo, and M. Bressan, "Prediction of site-specific solar diffuse horizontal irradiance from two input variables in Colombia," *Heliyon*, vol. 7, no. 12, Dec. 2021, Art. no. e08602, doi: 10.1016/j.heliyon.2021.e08602.
[112] E. Asghar, M. Hill, and C. Lynch, "A regularization-based long short-term memory architecture design for solar irradiance prediction," in *Proc. Int. Conf. Electr., Comput. Energy Technol. (ICECET)*, Jul. 2022, pp. 1–7, doi: 10.1109/ICECET55527.2022.9872779.
[113] M. Abdel-Nasser, K. Mahmoud, and M. Lehtonen, "Reliable solar irradiance forecasting approach based on choquet integral and deep LSTMs," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1873–1881, Mar. 2021, doi: 10.1109/TII.2020.2996235.
[114] U. T. Kartini, B. Suprianto, I. G. P. A. Buditjahjanto, L. Anifah, and M. N. Adiwana, "Optimalization global horizontal irradiance based on weather data using hybrid model modified decomposition FeedForward neural network," in *Proc. 5th Int. Conf. Vocational Educ. Electr. Eng. (ICVEE)*, Sep. 2022, pp. 125–129, doi: 10.1109/ICVEE57061.2022.9930391.
[115] V. Suresh, P. Janik, J. Rezmer, and Z. Leonowicz, "Forecasting solar PV output using convolutional neural networks with a sliding window algorithm," *Energies*, vol. 13, no. 3, p. 723, Feb. 2020, doi: 10.3390/en13030723.
[116] X. Jiao, X. Li, D. Lin, and W. Xiao, "A graph neural network based deep learning predictor for spatio-temporal group solar irradiance forecasting," *IEEE Trans. Ind. Informat.*, vol. 18, no. 9, pp. 6142–6149, Sep. 2022, doi: 10.1109/TII.2021.3133289.
[117] H. K. Ahn and N. Park, "Deep RNN-based photovoltaic power short-term forecast using power IoT sensors," *Energies*, vol. 14, no. 2, p. 436, Jan. 2021, doi: 10.3390/en14020436.
[118] C. Pavithra, M. R. Divya, M. K. Bavithra, and M. A. Jeyashree, "Machine learning for solar power forecasting," *Math. Statistician Eng. Appl.*, vol. 71, no. 3, pp. 1574–1591, Jun. 2022.
[119] P. Agrawal, H. O. Bansal, A. R. Gautam, O. P. Mahela, and B. Khan, "Transformer-based time series prediction of the maximum power point for solar photovoltaic cells," *Energy Sci. Eng.*, vol. 10, no. 9, pp. 3397–3410, Sep. 2022, doi: 10.1002/ese3.1226.
[120] D. Cao, W. Hu, J. Zhao, Q. Huang, Z. Chen, and F. Blaabjerg, "A multi-agent deep reinforcement learning based voltage regulation using coordinated PV inverters," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 4120–4123, Sep. 2020, doi: 10.1109/TPWRS.2020.3000652.
[121] S. Takayama and A. Ishigame, "Autonomous decentralized control of distribution network voltage using reinforcement learning," *IFAC-PapersOnLine*, vol. 51, no. 28, pp. 209–214, 2018, doi: 10.1016/j.ifacol.2018.11.703.
[122] A. N. Zaidan, M. A. M. Radzi, S. Shafie, N. Azis, and M. A. A. Mohd, "Multilevel PV-grid NPC inverter based artificial neural network maximum power point tracking technique," in *Proc. IEEE Ind. Electron. Appl. Conf. (IEACon)*, Oct. 2022, pp. 152–156, doi: 10.1109/IEACon55029.2022.9951857.
[123] A. Agarwala, N. Kumar, and pp. Dubey, "Machine learning based maximum power prediction for photovoltaic system," *Indian J. Pure Appl. Phys. (IJPAP)*, vol. 60, no. 10, pp. 892–898, Oct. 2022, doi: 10.56042/ijpapp.v60i10.62197.
[124] V. Mahesh, S. Meyyappan, and A. RamakoteswaraRao, "Maximum power point tracking with regression machine learning algorithms for solar PV systems," *Int. J. Renew. Energy Res. (IJRER)*, vol. 12, no. 3, pp. 1327–1338, Sep. 2022.
[125] M. Memaya, C. B. Moorthy, S. Tahiliani, and S. Sreeni, "Machine learning based maximum power point tracking in solar energy conversion systems," *Int. J. Smart Grid Clean Energy*, vol. 8, no. 6, pp. 662–669, Jul. 2019, doi: 10.12720/sgce.8.6.662-669.
[126] H. Muhsen and I. Tanninah, "Analysis and simulation of maximum power point tracking based on gradient ascent method," in *Proc. 12th Int. Renew. Eng. Conf. (IREC)*, Apr. 2021, pp. 1–5, doi: 10.1109/IREC51415.2021.9427806.
[127] K. Rafeeq Ahmed, F. Sayeed, K. Logavani, T. J. Catherine, S. Ralhan, M. Singh, R. T. Prabu, B. B. Subramanian, and A. Kassa, "Maximum power point tracking of PV grids using deep learning," *Int. J. Photoenergy*, vol. 2022, pp. 1–7, May 2022, doi: 10.1155/2022/1123251.
[128] X. Wang, Y. Wang, S. Wang, K. Shang, D. Su, and Z. Cheng, "A combination method for PV output prediction using artificial neural network," in *Proc. IEEE/IAS Ind. Commercial Power Syst. Asia (I&CPS Asia)*, Jul. 2021, pp. 205–211, doi: 10.1109/ICPSAsia52756.2021.9621582.
[129] P. Kofinas, S. Doltsinis, A. I. Dounis, and G. A. Vouros, "A reinforcement learning approach for MPPT control method of photovoltaic sources," *Renew. Energy*, vol. 108, pp. 461–473, Aug. 2017, doi: 10.1016/j.renene.2017.03.008.
[130] P. V. Mahesh, S. Meyyappan, and R. R. Alla, "Maximum power point tracking using decision-tree machine-learning algorithm for photovoltaic systems," *Clean Energy*, vol. 6, no. 5, pp. 762–775, Oct. 2022, doi: 10.1093/ce/zkac057.
[131] V. S. B. Kurukuru, A. Haque, M. A. Khan, S. Sahoo, A. Malik, and F. Blaabjerg, "A review on artificial intelligence applications for grid-connected solar photovoltaic systems," *Energies*, vol. 14, no. 15, p. 4690, Aug. 2021, doi: 10.3390/en14154690.
[132] M. A. Khan, A. Haque, V. S. B. Kurukuru, and S. Mekhilef, "Advanced control strategy with voltage sag classification for single-phase grid-connected photovoltaic system," *IEEE J. Emerg. Sel. Topics Ind. Electron.*, vol. 3, no. 2, pp. 258–269, Apr. 2022, doi: 10.1109/JESTIE.2020.3041704.

[133] M. Shafiullah, S. D. Ahmed, and F. A. Al-Sulaiman, "Grid integration challenges and solution strategies for solar PV systems: A review," *IEEE Access*, vol. 10, pp. 52233–52257, 2022, doi: 10.1109/ACCESS.2022.3174555.
[134] J. M. Malof, B. Li, B. Huang, K. Bradbury, and A. Stretslov, "Mapping solar array location, size, and capacity using deep learning and overhead imagery," Feb. 2019, *arXiv:1902.10895*.
[135] S. Kamal, P. S. Ramapraba, A. Kumar, B. C. Saha, M. Lakshminarayana, S. S. Kumar, A. Gopalan, and K. G. Erko, "Optimization of solar panel deployment using machine learning," *Int. J. Photoenergy*, vol. 2022, pp. 1–7, May 2022, doi: 10.1155/2022/7249109.
[136] K. Mason, M. J. Reno, L. Blakely, S. Vejdan, and S. Grijalva, "A deep neural network approach for behind-the-meter residential PV size, tilt and azimuth estimation," *Sol. Energy*, vol. 196, pp. 260–269, Jan. 2020, doi: 10.1016/j.solener.2019.11.100.
[137] M. Vijay and M. Saravanan, "Solar irradiance forecasting using Bayesian optimization based machine learning algorithm to determine the optimal size of a residential PV system," in *Proc. Int. Conf. Sustain. Comput. Data Commun. Syst. (ICSCDS)*, Apr. 2022, pp. 744–749, doi: 10.1109/ICSCDS53736.2022.9761011.
[138] Z. Asghar, K. Hafeez, D. Sabir, B. Ijaz, S. S. H. Bukhari, and J.-S. Ro, "RECLAIM: Renewable energy based demand-side management using machine learning models," *IEEE Access*, vol. 11, pp. 3846–3857, 2023, doi: 10.1109/ACCESS.2023.3235209.
[139] Q. Li, Z. Cui, Y. Cai, Y. Su, and B. Wang, "Renewable-based microgrids' energy management using smart deep learning techniques: Realistic digital twin case," *Sol. Energy*, vol. 250, pp. 128–138, Jan. 2023, doi: 10.1016/j.solener.2022.12.030.
[140] N. Müller, M. Marinelli, K. Heussen, and C. Ziras, "On the trade-off between profitability, complexity and security of forecasting-based optimization in residential energy management systems," *Sustain. Energy, Grids Netw.*, vol. 34, Jun. 2023, Art. no. 101033, doi: 10.1016/j.segan.2023.101033.
[141] G. Narvaez, L. F. Giraldo, M. Bressan, and A. Pantoja, "Machine learning for site-adaptation and solar radiation forecasting," *Renew. Energy*, vol. 167, pp. 333–342, Apr. 2021, doi: 10.1016/j.renene.2020.11.089.
[142] J. Polo, S. Wilbert, J. A. Ruiz-Arias, R. Meyer, C. Gueymard, M. Súri, L. Martín, T. Mieslinger, P. Blanc, I. Grant, J. Boland, P. Ineichen, J. Remund, R. Escobar, A. Troccoli, M. Sengupta, K. P. Nielsen, D. Renne, N. Geuder, and T. Cebecauer, "Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets," *Sol. Energy*, vol. 132, pp. 25–37, Jul. 2016, doi: 10.1016/j.solener.2016.03.001.
[143] J. F. Derakhshandeh, R. AlLuqman, S. Mohammad, H. AlHussain, G. AlHendi, D. AlEid, and Z. Ahmad, "A comprehensive review of automatic cleaning systems of solar panels," *Sustain. Energy Technol. Assessments*, vol. 47, Oct. 2021, Art. no. 101518, doi: 10.1016/j.seta.2021.101518.

SRABANTI DATTA received the Bachelor of Science degree in electrical and electronic engineering from the Rajshahi University of Engineering and Technology (RUET), in 2014. She is currently pursuing the Ph.D. degree with Florida International University (FIU), where she is working on research related to renewable energy, photovoltaic solar cells, and machine learning. Upon completing the bachelor's degree, she joined the Electrical and Electronic Engineering Department, Mymensingh Engineering College, as a full-time Faculty Member. She is a highly motivated and accomplished researcher passionate about renewable energy, photovoltaic solar cells, and machine learning. She is a dedicated researcher who is committed to making a positive impact in the field of renewable energy and photovoltaic solar cells. Her work promises to make significant contributions to developing renewable energy sources and reducing greenhouse gas emissions. Her research interest includes developing novel solutions for enhancing the performance of photovoltaic solar cells using machine learning techniques.

ANIK BAUL received the Bachelor of Science degree in electrical and electronics engineering from the Rajshahi University of Engineering and Technology (RUET), Bangladesh. He is currently pursuing the Master of Science degree in electrical and computer engineering with Central Michigan University, USA. He is a promising researcher in the field of electrical and computer engineering. He is a Graduate Research Assistant. Before his academic pursuits, he has been a Sub Divisional Engineer with the Bangladesh Power Development Board (BPDB), since 2015. In this role, he was responsible for designing and implementing electrical systems and managing the operation and maintenance of power generation. He demonstrated his expertise in various research projects, contributing to the development of innovative solutions to complex engineering problems. His research interests include machine learning, smart grids, and cyber security. He has also published several research papers in reputed journals and presented his work at various conferences.

GOBINDA CHANDRA SARKER was born in Dhaka, Bangladesh, in 1998. He received the B.S. degree in electrical and electronic engineering from the Mymensingh Engineering College, affiliated with the University of Dhaka, in 2023. His strong passion for AI and robotics enabled him to work on various research projects. Throughout his academic endeavors, he has published several research papers in reputed journals and conferences. His research interests include robotics, artificial general intelligence, machine learning, and deep learning algorithm application and development.

PINTU KUMAR SADHU received the Bachelor of Science degree in electrical and electronic engineering from the Khulna University of Engineering and Technology, Bangladesh, in 2010, and the Master of Science degree in electrical and computer engineering from Central Michigan University, USA, in 2022. He was a Specialist with Grameenphone Ltd., Bangladesh, from 2012 to 2020. He was with the Hardware Assisted Security Systems Laboratory, Central Michigan University. He is the author of seven peer-reviewed journal articles and two preprints. Moreover, five papers are also published in peer-reviewed conference proceedings. His research interests include providing hardware-assisted security for the Internet of Things, the Internet of Vehicles, machine learning, and embedded systems.

DEIDRA R. HODGES (Senior Member, IEEE) is currently the Department Chair and an Associate Professor with the Department of Electrical and Computer Engineering, Florida International University. She is an exceptional Leader in photovoltaics (PV) and solar energy research with extensive experience in PV and X- and gamma-ray radiation detectors with National Security. She is highly focused on advancing renewable energy, sustainability, nuclear materials, and extreme photon sensing. Her contributions include the support and development of the pixilated cadmium zinc telluride (CZT) gamma detector with BNL and highly efficient thin-film mixed perovskite halides photovoltaics. She has achieved perovskite solar cell power conversion efficiencies greater than 21%, approaching the world record efficiency of 25.2%. Her network spans the Department of Energy government laboratories, including the Brookhaven National Laboratory, the Kansas City National Security Center, Honeywell FM&T, and the National Renewable Energy Laboratory, collaborating with scientists and as a user of facilities.