

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Convergence of Photovoltaic Power Forecasting and Deep Learning: State-of-Art Review

MOHAMED MASSAOUDI^{1,2}, (Member, IEEE), INES CHIH^{3,4,5}, HAITHAM ABU-RUB¹, (Fellow, IEEE), SHADY S. REFAAT¹, (Senior Member, IEEE), and FAKHREDDINE S. OUESLATI⁶

¹Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha, Qatar

²Laboratoire Matériaux Molécules et Applications (LMMA) à l'IPEST, Carthage University, Tunis, Tunisia

³Laboratory of Energy Applications and Renewable Energy Efficiency (LAPER), El Manar University, Tunis 1068, Tunisia

⁴University of Luxembourg, The Faculty of Science, Technology and Medicine, Department of Engineering, Luxembourg

⁵National Engineering School of Bizerta, Carthage University, Tunis 7080, Tunisia

⁶National Engineering School of Carthage, Carthage University, Tunis, Tunisia

Corresponding author: Mohamed Massaoudi (e-mail: mohamed.massaoudi@qatar.tamu.edu)

This publication was made possible by the National Priorities Research Program (NPRP) grant [NPRP10-0101-170082] from the Qatar National Research Fund (a member of Qatar Foundation). The Open Access Funding was provided by Qatar National Library.

ABSTRACT Deep learning (DL)-based PV Power Forecasting (PVPF) emerged nowadays as a promising research direction to intelligentize energy systems. With the massive smart meter integration, DL takes advantage of the large-scale and multi-source data representations to achieve a spectacular performance and high PV forecastability potential compared to classical models. This review article taxonomically dives into the nitty-gritty of the mainstream DL-based PVPF methods while showcasing their strengths and weaknesses. Firstly, we draw connections between PVPF and DL approaches and show how this relation might cross-fertilize or extend both directions. Then, fruitful discussions are conducted based on three classes: discriminative learning, generative learning, and deep reinforcement learning. In addition, this review analyzes recent automatic architecture optimization algorithms for DL-based PVPF. Next, the notable DL technologies are thoroughly described. These technologies include federated learning, deep transfer learning, incremental learning, and big data DL. After that, DL methods are taxonomized into deterministic and probabilistic PVPF. Finally, this review concludes with some research gaps and hints about future challenges and research directions in driving the further success of DL techniques to PVPF applications. By compiling this study, we expect to help aspiring stakeholders widen their knowledge of the staggering potential of DL for PVPF.

INDEX TERMS Photovoltaic power forecasting, deep learning, big data, discriminative learning, generative learning, deep reinforcement learning.

Nomenclature

Abbreviations

AE Auto-Encoders
 AM Attention Mechanism
 BDDL Big Data Deep Learning
 BM Boltzmann Machines
 CDQN Continuous DQN
 CNN Convolutional Neural Network
 CWGAN-GP Conditional Wasserstein GAN with gradient penalty
 DBN Deep Belief Networks
 DCGAN Deep Convolutional GAN

DCIGN Deep Convolutional Inverse Graphics Network
 DDPG Deep Deterministic Policy Gradient
 DDQN Double Deep Q Network
 DL Deep Learning
 DPVPF Deterministic photovoltaic Power Forecasting
 DQN Deep Q Network
 DTL Distributed Transfer Learning
 GAN Generative Adversarial Network
 GRU Gated Recurrent Unit
 LSTM Long-Short Term Memory
 MLP Multilayer Perceptron
 NTM Neural Turing Machine

PPVPF Probabilistic photovoltaic Power Forecasting
 PVPF Photo-Voltaic Power Forecasting
 RBM Restricted Boltzmann Machines
 ResNet Residual Neural Network
 RNN Recurrent Neural Network
 SAE Sparse Autoencoders
 SG Smart Grid
 SimGAN Simulated GAN
 SOM Self Organizing Map
 VAE Variational Autoencoders
 WGANGP Wasserstein GAN with Gradient Penalty

I. INTRODUCTION

THE rapid expansion of Distributed Energy Resources (DERs) is driven by the vast exploitation of carbon-intensive energy sources and climate change concerns that threaten human survival and social progress [1]. Among all alternative sources, solar energy, specifically, Photovoltaic (PV) solar energy, has been getting the highest interest globally in the modern electricity grid, with estimates to satisfy a quarter of electricity needs by 2050 [2]. PV power plants' deployment merits include inexhaustibility in PV systems supply, long-life span, and excellent economic viability in the mid and long term [3]. However, the discontinuity and time-varying behavior of PV power flow bring into question the reliability and efficiency of PV systems [4]. Moreover, the sudden weather changes threaten the unit commitment and affect the demand and supply balance [5]. Therefore, PV Power Forecasting (PVPF) is a crucial factor for reliable power supply as it significantly reduces the sensitivity of energy systems to weather intermittency [6]. Consequently, the futuristic Smart Grid (SG) paradigm has considerably spurred the adoption of accurate PVPF techniques.

In this context, the energy community has been focusing on developing effective forecasting techniques to meet various technical challenges [7], [8]. With computer hardware and software development, forecasting models take advantage of High-Performance Computing (HPC) to achieve higher effectiveness. PVPF plays a vital role in handling a series of risk assessments and solving risk decision-making issues for an uninterruptible energy supply. PVPF can be conducted directly by predicting the PVPF [9], or indirectly by predicting the environmental factors encompassing the most relevant frequencies originating from weather conditions, such as solar irradiation (Fig. 1). Obviously, solar energy is presented as the most significant and critical parameter in concluding the characteristics of the solar units [10], [11]. Next, the predicted output is employed to deduce the PVPF via a predetermined mathematical model. However, it has been reported that direct PVPF leads to more accurate results than indirect PVPF [12], [13]. Nevertheless, this review article has considered both direct and indirect PVPF models.

Conceptually, the determination of PVPF lies in 1) Physical methods, 2) statistical methods 3) AI methods, and 4) Hybrid methods as the combination between them, as illustrated in Fig. 1. The physical models establish the math-

ematical formulas for the PV Power Generation (PVPF) equipment to conduct a deterministic closed-form solution for PVPF [14]. Physical models employ Numerical Weather Prediction (NWP) or ground measurement devices that meet the appropriate calibrated service facilities [15]. In [16], a physical model based on NWP has been adopted for solar irradiation uncertainty forecasting. It has been empirically proven that the suitable selection of the modeling window length is critical for predicting the confidence intervals. However, the proposed model has poor anti-interference capabilities reflecting the unsatisfactory prediction performance [16]. On the other hand, statistical forecasting is carried out through extensive numerical patterns analysis based on statistical theory. Statistical algorithms require a data set acquisition to build their domain knowledge since they neglect the investigated physical process. Moreover, statistical and physical models were not significant enough to be effective with unsatisfactory accuracy in numerous non-trivial problems such as Renewable Energy (RE) forecasting and weather forecasting. AI, specifically Machine Learning (ML) techniques consist of advanced complex approaches to acquire knowledge expertise and lead to accurate results and better generalization capabilities. Although ML is a very promising domain for power systems due to the abundance of computational resources and high-resolution databases, ML techniques have only been accorded to a few considerations compared to statistical and physical techniques in PV systems [17].

Deep Learning (DL) is considered an evolution of ML comprising multiple cornerstone-like models. More broadly, DL has been given a significant emphasis in academic circles for the last decade but only recently has broken into the industrial world for application-oriented research. Artificial Neural Networks (ANN) with multiple layers (hence called "deep") of interconnected neurons have sprung up and sparked a renewed interest in the research community, resulting in a plethora of research papers. Non-deep learning methods comprise one to three operational layers, whereas DL methods stacks multiple layers (more than three) of simple modules hierarchically. Elaborately, DL is advantageous to classical ML methods in distinguishing and learning multiple complexity levels due to three principle factors [18].

Firstly, the classical models heavily rely on a generation of hand-crafted features to track data patterns. This task necessitated manual design and feature learning, which are often labor-intensive and ad-hoc [19]. Fortunately, DL methods can intelligently learn from parse data representation using a general learning process [18]. This eliminates the need for domain expertise and hard core feature extraction adopted in handed engineering- and shallow learning-based models. More specifically, feature extraction-based DL is deducted automatically and optimally configured using an end-to-end pipeline to promote faster learning without being told to do so explicitly [20]. Secondly, traditional ML techniques such as Random Forest (RF) and Decision Trees (DT) might not handle multidimensional data [19]. For such models, the

training time becomes terribly low for deep varying level sizes [19]. Furthermore, the efficiency of these models can be unsatisfactory since the variable correlations are neglected [19]. Fortunately, DL methods can efficiently be fueled by massive amounts of data with a high level of complexity and multidimensionality to predict nonlinear behaviors accurately. Thus, DL models can achieve an outstanding predictive performance without the need for pre-defined relationships. Big Data technology uses DL to process a large pool of datasets and offers a potential solution to overcoming the set problem. Thirdly, DL models can hold and store more information within the neurons than the basic ANN model [18]. This can allow learning distributed representation (many-to-many relationships between types of representations), enabling generalization to new combinations of values not explicitly shown in learning data [18]. The former factors made the deployment of DL models more application-oriented than ML, strengthening the pervasive adoption rate by the advanced manufacturer.

A. THE DIRE NECESSITY OF DL IN PVPF

High-precision PVPF can potentially promote the grid's accommodation of PVPG by alleviating the negative impacts uncertainties on the utility grid. However, it is quite challenging to achieve satisfactory results with the classic prediction models. Recently, DL has become a research hotspot for its excellent ability to handle nonlinear Time Series (TS) energy data [21]. Thus, the marriage of PVPF and DL gives an impetus to build more sustainable and robust energy management paradigms [20]. DL methods have been successfully used in solar irradiance and solar power production forecasting. Notably, Deep Neural Network (DNN) architectures provide capabilities to learn hierarchical features from the data set while providing a more efficient representation than shallow models and improving generalization potential [22]. Hence, by eliminating the unpredictability factor, the research community tends to make great strides towards accurate forecasts and reliable decision-making [22].

With the whispered adoption of Advanced Metering Infrastructure (AMI), massive amounts of stored information with a variety of data types, complex relationships, and explosive growth from PV stations will be continuously generated, resulting in big or fast/real-time data streams. DL models can handle the big amounts of data generated from weather stations with Big Data Deep Learning (BDDL) to produce accurate results. DL models take advantage of the increase of computational power of Graphical Processing Units (GPUs) to carry out massive data on stream for high-quality forecasts. DL independently extracts features as an efficient big data-driven analytic scheme to process insufficient quality data that contains noise, heterogeneous data. DL models can efficiently handle the complexity, diversity, and integrity data conundrums that encounter meteorological data integration to improve the steadiness and security of power dispatch [20].

B. REVIEW NOVELTY AND CONTRIBUTIONS

The rising interest in DL underpinning PVPF systems intensified the need for a taxonomic review to summarize the most recent development in PVPF [23]. Fig. 2 illustrates the milestones of the AI development from the early attempts until the emergence of DL in 2010.

This paper's primary motivation comes from providing a unifying overview of the DL methods related to the PVPF applications. This article seeks to foster the synergy between the PVPF systems and DL methods. The main contribution is to enable further work both by industry and academia to speed up the practical adoption of DL techniques for PVPF. A bibliometric and network analysis on the PVPF topic was conducted to organize the data in a more reader-friendly form from Web of Science (WoS) core collection database and VOSviewer software [24]. The VOSviewer software was employed to reveal the thematic content of the articles set based on the identification of the keywords. Keywords included by authors of the articles and occurred more than three times in the WOS core database from 2015-2021 were exported into Comma-Separated Values (CSV) format and enrolled in the final analysis. Of the 150 keywords, the initial search identified that 50 met the threshold. The keyword combinations are employed in the systematic review protocol to provide a broad overview of research trends in DL techniques in PVPF systems. Fig.3 presents the mapping analysis of the commonly-occurring term with the VOSviewer.

Using rigorous bibliometric indicators, Fig.3 shows that the node covering the widest and most noticeable area is the fields of "deep learning" and "prediction" with a smaller size. From Fig.3, the emergent research topics are classified into two core clusters: DL models and enabling technologies such as Deep Transfer Learning (DTL). These clusters are devoted to organizing the paper content. To do so, the recent DL architectures applied to PVPF and selected from the period 2015 to 2021 have been analyzed. Furthermore, the development of Deterministic PVPF (DPVPPF) and Probabilistic PVPF (PPVPPF) were presented. In this work, we focus on reviewing the current signs of progress and pointing out potential future directions of DL for PVPF.

Some existing works have studied PVPF and AI, listed in Table 1 with a brief description of their related topics and the differences with this review [22], [25]–[29]. In [30], a comprehensive review of RE forecasting methods has been conducted with a particular emphasis on wind and solar energy. Specific focus of this review reports a growing interest of studying DL techniques for forecasting applications regarding their inherent feature extraction capabilities. However, the review coverage includes both wind and solar resources, which may lead to loose contributions and explanations, especially when discussing the forecasting architectures. In [6], Renewable Energy Sources (RES) forecasting methods have been reviewed. In the RES context, the authors provided the common understandings and promising research insights, including hierarchical forecasting, probabilistic forecasting, and forecast combination. Additionally,

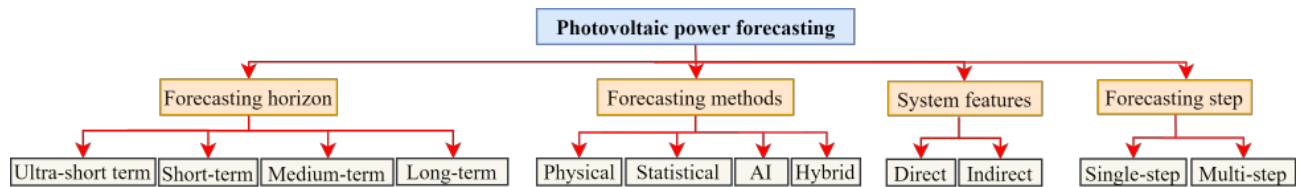


FIGURE 1: PVPF classification.

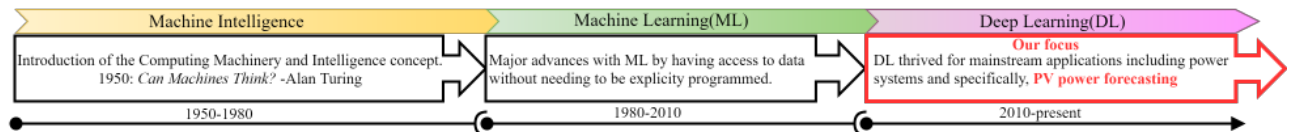


FIGURE 2: Deep Learning: timeline and evolution

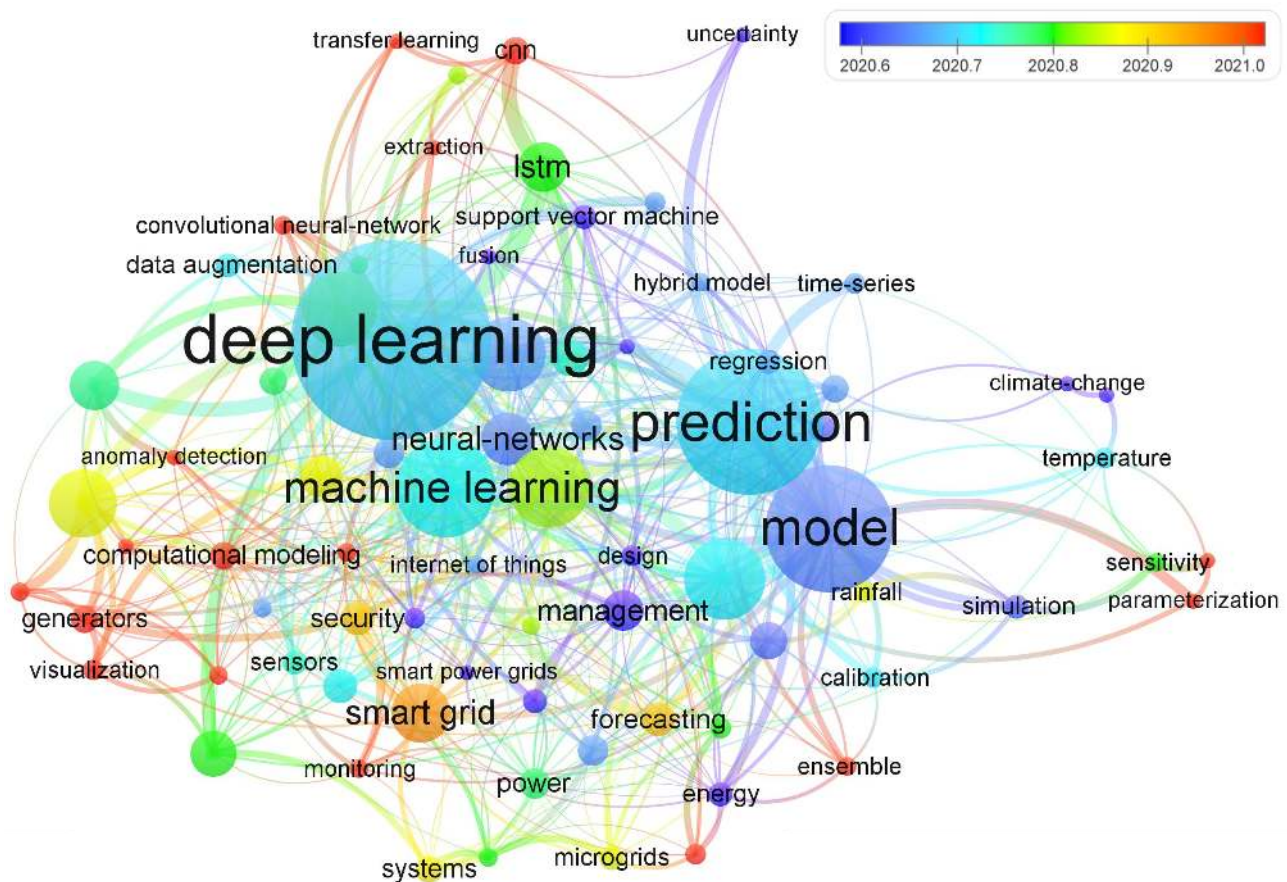


FIGURE 3: Bibliometric analysis for the author-supplied keywords; The size of nodes presents the frequency of recurrence. The connections between the nodes illustrate their co-occurrence in the same article. When the distance between the keywords is short, the keywords co-occur more frequently with each other.

some helpful recommendations and common research pitfalls for publishing high-quality journals were provided. Nonetheless, the AI techniques adapted for RES forecasting have not been discussed. More importantly, these works focus exclusively on the RES spectrum and do not mainly focus on PVPF. Therefore, the key characteristics of PV variability

and how DL can solve the PV limitations are scarce. In [31], the authors have reported the typical policies related to solar forecasting for grid penetration. However, this work has not been extended to intensively investigate DL forecasting techniques. In [32], the concept of overarching thinking is introduced to contradicts the basic classification of predictive

methods into statistical, ML, or NWP. Moreover, the main post-processing methods for PVPF were reviewed to enhance the goodness of the forecasts. However, an in-depth analysis of regression models is not provided. Paper [22] conducted a review on direct PVPF with a special focus on statistical and ML models provided in the literature. The authors of this review classified the data-driven techniques into persistence methods, statistical approaches, ML approaches, and hybrid techniques. Nevertheless, DL methods' investigation, their potential benefits, and shortcomings are not discussed in detail [22]. In [25], the authors have been limited to reviewing probabilistic forecasting for electricity consumption and PVPF. This review reported that all the forecasting engines essentially depend on the extreme forecasting scenarios, leading to poor computability scalability in the existing ML methods [25]. Therefore, every model should be customized to well perform a forecasting task. However, the DL models architectures were not explored elaborately. Besides, the implementation factors of key issues of DL approaches were not outlined. The authors in [29] have extensively studied the BD models for PVPF. Thirty-eight papers were deeply analyzed to enlist the most relevant ML models. It can be pointed out the Extreme learning machines achieved an excellent accuracy-computational time tradeoff [29]. Nonetheless, the DL paradigms and their executions were not covered. Besides, the notable techniques for PVPF deployment were not explored explicitly. All the related works partially cover the aims of this work. However, the related works paid less attention to results based on DL methods. In contrast, we limit this holistic review to the DL-based PVPF, leaving aside shallow ML and physical methods-based PVPF. Therefore, this study provides insights not previously fully covered or evaluated by other reviews [33].

This review gives a particular emphasis on the application of DL methods to the PVPF. To the best of our knowledge, different from the previous works, this is the first initiative to give a bird's eye view on the applications of DL for PVPF, which is not adequately addressed in the existing literature. To fill this gap, this review focuses on the use of DL for PVPF applications. The main contributions of this review are expounded as follows:

- First, we derive taxonomies for PVPF based on various criteria such as the forecasting horizon, forecasting models, system features, and forecasting range.
- Second, the DL models-based PVPF are systematically classified into learning-related. A comprehensive and complete review of different algorithms applied in the case of PVPF is provided to give critical insights into their strengths and limitations. We aim to allow the reader to readily distinguish the efficacy gaps at a glimpse. Further, an evaluation and discussion of the role of meta-heuristics in carrying out the functions required in DL within the PVPF-realm have been conducted.
- Third, the enabling technologies for DL-based PVPF

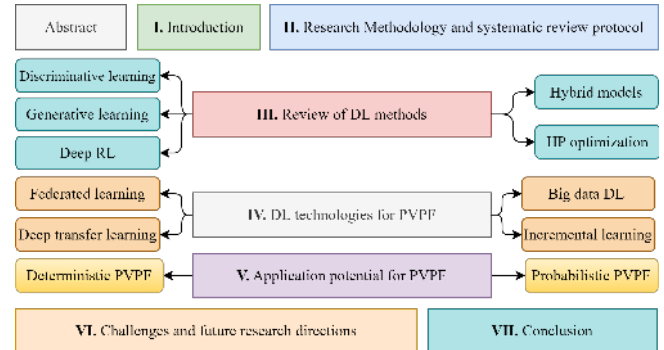


FIGURE 4: Diagrammatic view of the organization of the paper.

were rigorously reviewed in a more comprehensive and applicant-oriented manner, such as federated learning, transfer learning, and BDDL, where previous works are summarized logically.

- Fourth, pioneering works related to deterministic and PPVPF have been deeply investigated.
- Fifth, a critical view over the existing research challenges are presented, and future directions in PVPF studies to the deployment of competent, scalable, and computationally effective algorithms based on DL are discussed.

C. REVIEW STRUCTURE

The rest of this paper is structured as portrayed in Fig. 4. Concretely, section II presents the review methodology. Section III describes the popular taxonomies of PVPF techniques. Section IV comprehensively investigated the DL methods for PVPF. Section V presents the possible enabling DL techniques for PVPF. Section VI discusses the significant applications of DL techniques for PVPF. These applications cover DPVPF and PPVPF. In section VII, the possible future directions for empowering the PVPF performance by emphasizing the undiscovered fields have been presented. Section VIII concludes this review paper.

II. RESEARCH METHODOLOGY AND SYSTEMATIC REVIEW PROTOCOL

To benefit reading, extensive searches have been performed to fetch the most relevant content. For instance, the time horizon is one of the essential tools to classify PVPF techniques. Depending on the time domain, there are four distinguished forecasting horizons as illustrated in Fig. 1; specifically, Ultra-Short-Term (USTF) from seconds to one hour [34], Short-Term (STF) with the prediction period from hours to one day, Medium-Term spans up to a month ahead, and long-Term predictions for a month to a year [12]. With the aim of covering the largest number of articles regarding the review topic, possible variations were also employed for this selection. Hence, the search string utilized was: 'Deep learning' AND 'Photovoltaic power' AND 'Forecasting' OR

TABLE 1: Summary and comparison of the related reviews on PVPF with our review paper

Literature	Year	Selected keywords	One-sentence summary	Scope				
				DL	DTL	BDDL	DPVPF	PPVPF
[22]	2018	PVPF; AI; ML	An overview of Analysis of ML models for PVPF	-	-	-	✓	-
[25]	2018	Probabilistic forecasting	A systematic literature review of PPVPF	-	-	-	-	✓
[26]	2019	PVPF; ML	A milestone Review of AI methods for PVPF	-	-	-	✓	-
[29]	2019	Data mining; PVPF	A review on BD models for PVPF	✓	✓	✓	-	-
[27]	2020	PV; PVPF; AI; ML	Insights to the application of ML for PVPF	✓	-	-	✓	-
[28]	2020	PVPF; CNN; LSTM	A comprehensive review on ML models for PVPF	✓	-	-	✓	-
This paper	2021	PVPF; DL; SG	A high-level review on DL methods for PVPF	✓	✓	✓	✓	✓

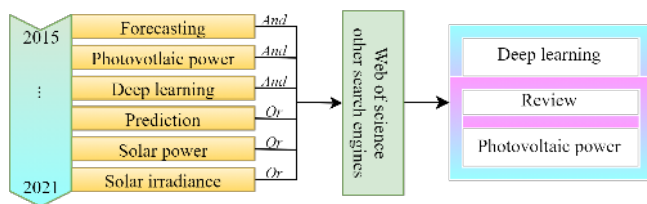


FIGURE 5: Search methodology based on keywords combinations using Boolean operators AND/OR. The specific query searched was: (Prediction OR Forecasting) AND (Deep learning) AND (Solar power OR Solar irradiance OR Photovoltaic power).

'Prediction' OR 'Solar power' OR 'Solar irradiance' OR 'deep neural networks' as shown in Fig.5.

The publications on each platform based on the keywords were made on June 15, 2019, totaling 350 articles. The identification of the relevant research works is conducted according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) [35]. We identified the main academic research databases, including Google Scholar, IEEE Xplore, Science Direct, and Nature. The identification process of relevant papers from the early stage of the selection to the final selected publications has been displayed in Fig. 6.

To achieve a more complete and inclusive understanding, this review paper contributes to the existing research papers by answering the following five Research Questions (RQ):

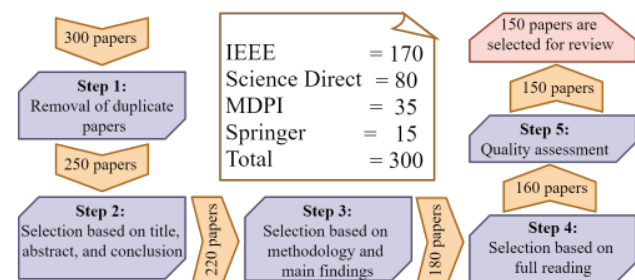


FIGURE 6: Flowchart of the paper selection process with respect to PRISMA protocol.

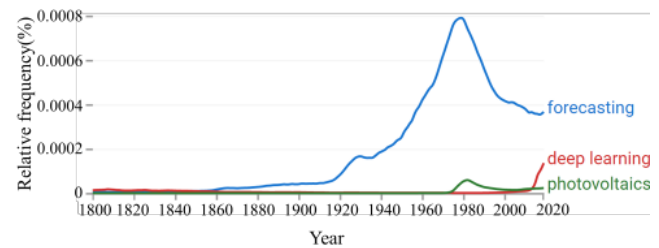


FIGURE 7: Frequency of contrasting words "deep learning", "forecasting", and "photovoltaics" from Google Books Ngram Viewer results (1800-2020).

RQ1: What is the popular taxonomies for PVPF?; RQ2: What are the most up-to-date DL methods for PVPF?; RQ3: What are the DL methods for deterministic and probabilistic PVPF?; RQ4: How Big data and transfer learning can enhance the PVPF accuracy?; RQ5: What are the research frontiers and future research directions?

The answers to these questions consider multiple sources of information to ensure the accuracy and objectivity of the main findings of this paper. The inclusion criteria were as follows: (1) articles published in English-language documents that were published between January 1, 2015, and August 01, 2021; (2) articles used a particular DL model for PVPF;

Fig. 7 presents a timescale variation on the frequency of use of terms "Forecasting", "Deep learning", and "Photovoltaics" in scientific books from Google Books Ngram Viewer. It can be seen from Fig. 7 that the popularity of DL has significantly increased in the last years, while the forecasting paradigm had decreased since 1980. This result shows that forecasting applications must take advantage of the increasing DL trend, especially for PV systems.

III. OVERVIEW OF PVPF AND DL

To help readers better distinguish between the emerging learning paradigms applied to PVPF, DL methods were classified into three classes: discriminative learning, generative learning, and Deep Reinforcement Learning (DRL). A brief review of the knowledge and understanding of these three concepts is presented. We, then briefly discuss the potential of well-established mature DL structures in PVPF. Further-

more, we cursorily investigated several Hyperparameter Optimization (HO) used for DL in PV systems.

A. DISCRIMINATIVE LEARNING

This subsection mainly introduces common DL models for PVPG estimations: Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Generative Adversarial Network (GAN).

1) MLPs

MLP, also named Fully Connected Network (FCN) is the quintessential DNN model [36]. Conceptually, this MLP consists of three fully connected layers in a feed-forward architecture, specifically, input layer, numbers of intermediate (hidden) layers, and output layer, as shown in Fig. 8(a) [8], [37]. By increasing the networkability, the hidden layers computationally reveal underlying patterns of data at deep levels of abstraction. With the use of nonlinear/linear activation functions, the MLP configuration can resolve complex mappings between a set of observations and response variables. The MLP mechanism is described as follows [38]:

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (1)$$

$$y_k = \phi(u_k + b_k) \quad (2)$$

where x_j , w_{kj} and b_k denote the neurons' input, synaptic weight function, and bias term, respectively. ϕ denotes the activation function. As a supervised learning algorithm, the MLP network employs backpropagation to minimize the cost function. Moreover, two common training methods are used as part of MLP: Feed-forward Back-Propagation, and Levenberg–Marquardt. The optimal selection of the training algorithm can enhance the convergence speed and the model accuracy.

Pioneering work is presented in [39], where it proves that the typical MLP network is perfectly tailored for a day-ahead PVPF. More importantly, the developed method could offer prominent enhancement for one day-ahead PVPF results, especially with TanhAxon activation function and Levenberg–Marquardt learning rule. However, hyperparameter tuning for a deep MLP involves a lot of complexity in modeling, which could be a challenging issue. Such observations are confirmed in [40], wherein the MLP performance mainly hinges on offline training and the hyperparameters' suitability which require significant human labor for fine-tuning. From the authors' work in [40], a two-state model-based MLP and knowledge-based Neural network (KBNN) were designed for offline and online on-site deployment. Here, KBNN is used to avoid data shortage by including prior analytical prediction equations. Hence, the KBNN significantly the MLP model with the insufficiency of labeled data. The approximate MAPE is 11%. In [41], a DL-based mapping model between the concurrent sky image and surface irradiation has been introduced. Aside from the hybrid model, the sky images were clustered in the preprocessing

stage using K-means clustering based on Convolutional Autoencoder (CAE) to enhance the feature representation of high-dimensional data. Despite the high complexity cost, the mapping modeling of surface irradiance boosts the forecasting model performance. In [42], an MLP model has been proposed for solar radiation forecasting. Four types of uncertainties were presented; errors from meters' measurements, sarcastically of weather data, ML uncertainties, and errors due to the forecasting range. A reliability index has been comprehensively introduced to assess the goodness of the forecasts [42]. Although the model generalization to complex cases, the MLP model is restricted to reveal patterns among sequential samples such as PV generation data. This is mainly due to the fact that this configuration does not save the previous information in an internal memory [43]. Consequently, the Time Series (TS) data are trained independently, which may lead to poor accuracy [43].

2) CNNs

CNNs, popularly termed as ConvNet, are a popular class of feed-forward networks designed to process grid-like structured data [28]. The core of the CNN network was conceptualized based on three principal elements: convolution, pooling computations, and Fully Connected Layers (FCL), as shown in Fig. 8(b). These elements lead to a spectacular features extraction capacity and robust feature representation [44]. The objective of the pooling layers is to merge semantically similar features into a single one by applying a specific function. This allows the pooling layer to reduce the feature map dimension, accelerating the system convergence [44]. While the convolutional layer extracts local features from contiguous data. The layered structure of CNN comprises one-dimensional CNN, two-dimensional CNN, and three-dimensional CNN. For PVPF, one-dimensional CNN is essentially used to process sequential data [28]. The fully connected layers are usually the last few layers and are used to summarize information.

In [45], the authors introduced a hybrid framework by integrating a hybrid CNN-Long-Short Term Memory (LSTM) network-based PVPF model. Using the hybrid paradigm, CNN automatically filters out noise and extracts the valuable features, while LSTM efficiently handles sequential inputs. More specifically, the authors have sought to exploit the Multiple Relevant and Target variables Prediction Pattern (MRTPP) method to optimize the distribution of the input features. This promotes the forecasting engine's efficiency in capturing the nonlinear variation of PVPG for multi-step prediction. However, the convolutional layer in CNN has a convolutional kernel of constant size and a limited receptive domain, which is limited to local feature learning. Whereas in [46], the authors proposed a specialized CNN for 15-min ahead minutely-averaged PVPF. The proposed model provides accurate predictions represented by a 5.7% forecast skill without intensive hand-engineered features as input. However, the major limitation of CNN can not be fully suitable for capturing time dependencies.

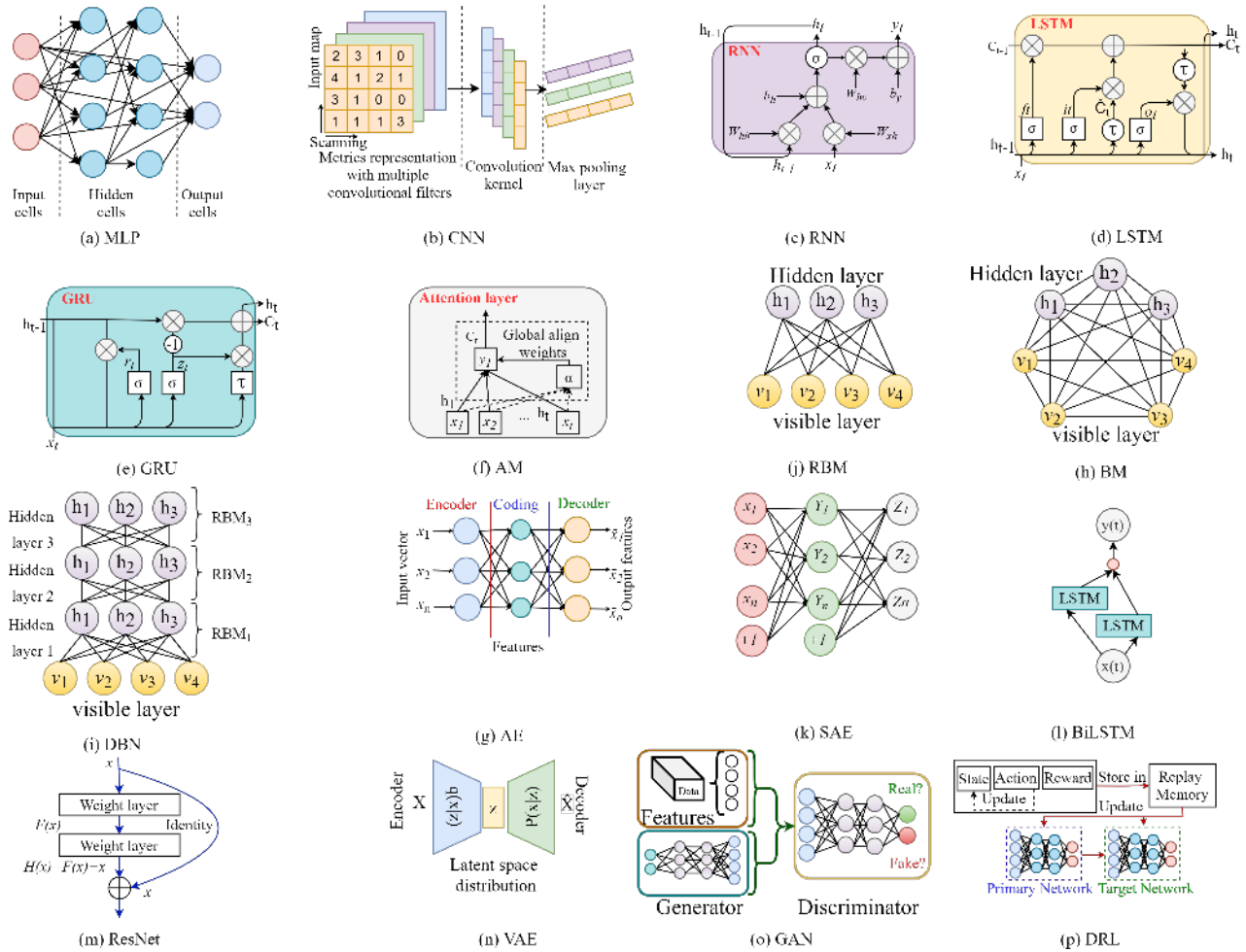


FIGURE 8: Typical structure and functions of typical MLP, CNN, RNN, LSTM, GRU, AM, RBM, BM, DBN, AE, SAE, BiLSTM, ResNet, VAE, GAN, and DRL (a) MLP architecture with a naive structure comprising two hidden layers (b) CNN with four convolutional layers (c) Schematic of a RNN block (d) Underlying gating structure of LSTM (e) Underlying gating structure of GRU (f) Schematic of an AM block (g) Operating principle of an RBM where the neurons form a bipartite graph (h) The conceptual structure of an BM (i) The DBN architecture stacking multiple RBMs structures hierarchically (j) The inner structure of an AE (k) The schematic diagram of an SAE (l) The inner structure of an BiLSTM (m) ResNet mechanism as a CNN that increases the depth of a network by stacking convolutional layers and using skip connections to increase the receptive field (n) The inner structure of an VAE (o) The conceptual structure of an GAN, which creates lifelike artifacts from a target distribution (p) The architecture of an DRL, where the initial state are fed to the primary network.

3) RNNs

RNNs are a special type of DNN based on control theory composed of a chain of neurons whose output is connected not only to the next layer but also to feedback connection, as shown in Fig. 8(c) [47]. By sequencing TS data as an input vector, the RNN cell allows the underlying information to persist until feeding it back to the next prediction. Thus, the RNN provides a quicker implementation and fast training. The TS data passes through a cell in a sequential vector, at each step the cell output value is concatenated with the next time step data, and the output value of the cell serves as input for the next time step. The process is repeated until the last time step data. However, the most common drawback of the

vanilla RNN model is the limited capability of handling long-term dependencies [47].

For TS data, RNNs based models form the core of most sophisticated fancy TS applications, which allows them to be perfectly tailored for PVPF and widely depicted in pioneering articles. To name a few, the authors in [48] proposed a deep RNN to predict solar irradiation accurately. It is worth mentioning that five RNN categories were rigorously described, specifically, the standard RNN, Deep RNN, stacked RNN, Deep RNN with shortcut connections, and Deep RNN with deep output layer. Compared with other benchmarks, using realistic data from natural resources in Canada, the RNN showed better accuracy when processing TS high-level

features. In [49], an LSTM-Convolutional network (LSTMC) has been adopted for PVPF. The proposed approach results in RMSE of 0.621kW for short-term PVPF. Surprisingly, it can be remarked that LSTMC model outperforms CNN-LSTM. This paper demonstrates that extracting temporal correlations first and then spatial correlations using this processing order could improve the forecasting effectiveness. Despite the high accuracy of the proposed model, it is concluded that a large number of training samples is extremely needed for achieving generalizability.

LSTM network is a special chain-like structure with memory cells proposed by Hochreiter and Schmidhuber [50]. The LSTM is a particular type of RNN that overcomes the notorious vanishing gradient phenomenon in modeling long-range temporal dependencies [51]. A typical LSTM cell mainly consists of input gate i_t , forget gate f_t , output gate o_t , and one control gate c_t to manage the information flow, as shown in Fig. 8(d). Concretely, these gates are used to control the update, maintenance, and deletion of information contained in cell status. The operations on the memory block are managed using adaptive multiplicative gates. The LSTM gates, hidden outputs, and cell states are computed as following [52]:

$$I_t = \sigma(W_I x_t + U_i h_{t-1} + b_I) \quad (3)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (4)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (5)$$

$$\tilde{c} = f_t \odot c_{t-1} \quad (6)$$

$$h_t = o_t \odot \tanh(c_t) \quad (7)$$

where x_t and c_t is the input sample at time t and the memory unit, respectively. (b_f, b_I, b_o) , (W_f, W_I, W_o) , and (U_f, U_I, U_o) stands for the bias terms, weight vector, and input weights for each gate respectively. The symbol \odot denotes Hadamard product. h_{t-1} is the hidden layer for the respective gates x in the current timestamp. σ activation function denotes the element-wise calculation. The LSTM has an excellent potential to process time-based information. For instance, the authors in [53] employed an LSTM-based to perform an hourly PVPF. The particle swarm optimization (PSO) algorithm was employed to adjust the load dispatch. The simulation results from the proposed model exhibit that it considerably increases the accurateness of prediction a Mean Absolute Percentage Error (MAPE) error of 15.87%. However, the computational cost is high compared to MLP model.

Gated Recurrent Unit (GRU) is considered as one of the preferred single TS forecasters [54]. By using the recurrent connections, the GRU architecture permits the network to access the historical information. The GRU is a kind variant of RNN that has gates that modulate the flow of information inside the unit, as shown in Fig. 8(e). A typical GRU cell is composed of only two gates, the reset gate $r_t = \sigma(w^r x_t + U^r h_{t-1})$ and the update gate $z_t = \sigma(w^z x_t + U^z h_{t-1})$ [54]. σ is a smooth and differentiable function; b_z, W_z , and U_z

are the bias, the input constant of the update gate (z), and the previous activation constant, respectively. Nonetheless, it is distinguished for its gate reduction strategy to accelerate the learning process without lowering the performance. Paper [55] employed GRU for PVPF. In their study, various processing blocks were built based on the characteristics of each block to promote the proposed approach accuracy. Specifically, the Pearson coefficient is exploited to rank the feature inputs according to their relevance to the PVPG. Next, K-means clustering is used to group the training data according to the similarities of input patterns. These groups are utilized to generate an averaged PVPF output. The proposed GRU architecture demonstrated its expertise in capturing temporal dependencies. However, the average training time is 365.40 seconds, which is painfully slow compared to statistical techniques such as Auto-regressive Integrated Moving Average (only 3.66 seconds). Bidirectional LSTM (BiLSTM) is an improvement to one-way RNN where the forward and backward hidden layers are combined to access both the preceding and succeeding data [56]. Bidirectional Mechanism (BM) is a way of learning the information from both directions, as shown in Fig. 8(l) [56]. In a nutshell, BiLSTM can handle the sequential modeling challenge better than conventional LSTM by acquiring the forward and reverse information from the cyclic feedbacks. For instance, the authors in [57] proposed a BiLSTM model to model nonlinear time dynamics for PVPF, which helps in boosting the model performance significantly. The proposed model can accurately detect meteorological changes over time. Meanwhile, the real-world implementation of the BiLSTM model requires memory-bandwidth-bound computation, which compromises their application ability due to high computation and storage. To combat this challenge to a large extent, Attention Mechanism (AM) is associated in the RNNs structures for better generalization ability by mimicking the attention of the human brain [58].

AM puts more focus on the input sequence's core elements that affect the quality of the forecasts to learn the information in the input sequence better. The principle idea of the AM in DL is also to neglect the irrelevant data to the current task and only to select the information that is more critical to the current mission, as shown in Fig. 8(f). AM can be categorized into spatial attention, channel attention, and self-attention [59]. AM allows the forecasting models to pay more attention to useful features, so AM is widely used in RNNs. The computation of AM is initialized by a Query (Q) and Keys (K) as $f(Q, K_i) = Q^T K_i$. The Softmax activation function is used for weights standardization as [59]:

$$a_i = \text{Softmax}(f(Q, K_i)) = \frac{\exp(f(Q, K_i))}{\sum_{j=1}^L \exp(f(Q, K_j))} \quad (8)$$

The attention value is obtained by calculating the sum $\text{Attention}(Q, K, V) = \sum_i^L a_i * \text{Value}_i$, where L is the size of the input sequence. In [60], the authors proposed a Convolutional self-Attention Based LSTM (CA-LSTM) for PVPF. The self AM is a special form of AM, which better

captures the syntactic and semantic information from the row of TS data. The proposed model aims to fully use the features of long sequence inputs, achieving an overall MAPE of 10%. Thus, the AM successfully improves the traditional LSTM performance with a lower MAPE of 17%. Reference [61] proposed an LSTM-based Temporal AM (TA-LSTM) for solar generation forecasting. The proposed model employs partial autocorrelation to follow the input lag. The TA-LSTM produces a Root Mean Square Error (RMSE) of 0.26kW, which seems to prove its high competitiveness compared to the classic LSTM in PVPF. Meanwhile, the hyperparameter list is increased, which leads to extensive hyperparameter tuning. Despite the high suitability of this network for TS processing of the gated RNN architectures, their massive storage and computation requirements hinder their application ability, particularly for PVPF.

B. GENERATIVE LEARNING

In this section, we review the state-of-the-art DL architectures. These ubiquitous DL architectures are Auto-Encoders (AEs), Restricted Boltzmann Machines (RBMs), and Deep Belief Networks (DBNs), and Generative Adversarial Network, respectively.

1) RBMs and DBN

The RBM (Restricted Boltzmann Machine) network is energy-based stochastic neural networks, as shown in Fig. 8(j). In a nutshell, this variant of Boltzmann Machines (BM) has node connections both within layers and between layers (Fig. 8(h)). The RBM model has the potential to learn the input probability distribution in supervised/unsupervised learning. The RBM architecture has two levels with symmetrical connections between them, one is the visible layer v , which contains the input and output, and the other is the hidden layer h with n units [18]. The visible and hidden units that follows a joint distribution can be expressed as:

$$P(v, h) = \frac{\exp(-E(v, h))}{z} \quad (9)$$

$$E(v, h) = v'wh + \alpha'v + b'h \quad (10)$$

and $z = \sum_{v, h} \exp(-E(v, h))$. Deep Belief Networks (DBN) is an unsupervised greedy learning algorithm with a stacked RBM units as shown in Fig. 8(i) [18]. The DBN performs layer-wise training to learn probability distribution of the input vectors. The DBN processing consists of using layer-by-layer unsupervised pre-training to select the suitable initial parameters and supervised fine-tuning mechanism to rebuild training samples by tuning the parameters [62]. The RBM layers were seen as feature extractors to generate a high-dimensional abstraction of the inner relationships of the data.

In [63], the authors proposed an Integrating gray data preprocessor and DBN for day-ahead PVPF. These contributions are completed in [62], where the authors proposed a day-ahead global solar radiation forecasting using functional DBN. The performance enhancement of the developed technique relies on an embedding clustering layer and knowledge

functions from empirical models. These processing units demonstrate a sophisticated and elongated iterative fashion, thus improving the model robustness for longer time dependencies [62]. Although the DBN model demonstrated its efficiency in various forecasting tasks, the DBN architecture is prone to model structure and parameter optimization challenges.

2) Autoencoders

Autoencoder (AE) architecture is one of the most groundbreaking unsupervised learning models that learn characteristics from unlabeled data representation [9]. AEs are loosely inspired by the way the human brain works. Typical components of the approach are the encoder and the decoder [20]. By minimizing the reconstruction error between the input data at the encoding layer and its reconstruction at the decoding layer. There are many types of AEs, and the most commonly used ones are; Stacked Autoencoders (StAE), Denoising Autoencoder (DAE), Sparse Autoencoders (SAE), CAE, and Denoising CAE (DCAE), and Variational Autoencoders (VAE) [20]. An AE-LSTM network is proposed for day-ahead PVPF for the next day at 15-min interval [64], with a normalized RMSE of 4.56%. The proposed AE-LSTM model jointly exploits the feature extraction of AE and the sequential TS forecasting engine of LSTM. However, the trans day weather volatility is poorly predicted by the formal predictor. The authors of [65] established an AE-driven DL model-based PVPF method to overcome the stochastic behavior of PV power output, which achieves an optimal $R^2 = 99.5\%$. Despite the ability of the proposed model to provide accurate one-step and multi-step forecasting results, the VAE shown in Fig. 8(n) is prone to the vanishing latent variable problem.

3) GAN

GAN is an unsupervised pre-trained network consisting of two competing neural networks: the generator $G(z)$ and the discriminator $D(x)$, as presented in Fig. 8(o). By learning the real data x distribution, G generates realistic scenarios until they cannot be distinguished anymore from real data. The fake data is generated from random noise using Gaussian distribution. This operation is conducted by deliberately introducing feedback at the back-fed input cell from the input noisy variables $p_z(z)$. And D correctly distinguishes whether the input data comes from the true data $p_{data}(x)$ or the generator. The two models are optimized simultaneously by updating the network weights in an alternating manner. The hyperparameters are tuned based on optimizing the loss and varying the randomness. The objective function formulated as [66]:

$$\begin{aligned} \min_G \max_D V(D, G) = & E_{x \sim p_{data}(z)} [\log(D(x))] \\ & + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \end{aligned} \quad (11)$$

GAN has gradually attracted prominence in the PVPF domain, especially for data augmentation purposes. For instance, the authors in [66] employ Wasserstein GAN with

Gradient Penalty (WGANGP) for weather classification-based PVPF. The WGANGP was utilized for data augmentation purposes by producing synthetic data that follows the same heteroscedasticity of original data. The newly generated data is fed to the CNN model to improve its performance by ameliorating the feature representation. A series of experiments on 33 meteorological weather types were conducted and proved the effectiveness of the method by comparing it with other methods. However, the major caveat of the GAN model lies in the fact that the GAN training is relatively unstable.

C. DEEP REINFORCEMENT LEARNING

Recently, DRL has been introduced as a combination of DL (DL), and Reinforcement Learning (RL) to better cope with the dynamic changes of the unsteady PV environment [67]. By bridging DL and RL, DRL shows its great potential in handling complex tasks and high scalability to suit complicated and unfamiliar environments, particularly in power systems [68]. As a goal-oriented learning method from the environment feedback, the traditional RL extends its potential to store high-space actions and states with an intuitive hierarchical feature extraction ability and nonlinear approximation ability of DL architectures, as shown in Fig. 8(p). Consequently, DL models intersection state information from sequential PV data [69]. From the DRL scheme, the agent continuously interacts with the state environment over a series of trial-and-error processes to shape the optimal strategies [70]. Several methods for the Q values updates were provided in the literature, such as State-Action-Reward-State-Action (SARSA) and Asynchronous-Advantage-Actor-Critic (A3C) and Deep Q Net (DQN) [68]. In RL, the environment can be modeled as a Markov Decision Process (MDP) expressed as (S, A, P, R, γ) . Here, S , A , P , and R denote the discrete states in the environment, a finite set of actions provided for the agent, the state transition probability matrix, and the reward function, respectively. γ presents the discount factor utilized to quantify the importance of the future and present rewards. Despite the ample opportunities of DRL, efforts to employs the DRL in PVPF have been scarcely found in the literature.

Using the powerful representation ability of DNNs, the value function is fitted to optimize the explosion of continuous and integer state-action space problems. Mainly, DRL provides high-dimensional input or large action sets to solve intractable problems using self-adjustments and optimization solutions. However, current DRL techniques are dependent on massive training data and expensive computational requirements, which may be unacceptable in practical PV advertising platforms. DRL models were roughly divided into two main settings: Model-based, and Model-free-based. The following subsections describe these concepts.

1) Model-Based

Model-Based DRL (MDRL) expects that the agent understands the system dynamics and how the system crosses

from one state to another one and how rewards are generated. This MDRL methods have been effective in terms of data-efficiency, transferability, and universality [70]. However, MDRL is computationally expensive and ineffective in rapidly varied environments [70]. MDRL has been proposed to solve the optimal action-selection policy [71]. MBRL employs an internal model to approximate the environment, and the control behavior can be learned through this model. It has been reported that model-based approaches are more efficient than model-free approaches. However, MDRL needs to save the state transition matrix and employing Dynamic Programming (DP) algorithms leading to massive calculation requirements [72]. This approach is not always practicable, especially in complicated paradigms where the agent has limited to no knowledge about its environment. Authors in [73] applied an MDRL-based MuZero algorithm to solve the scheduling problem of distributed microgrid, particularly for PV systems. The proposed approach combines the Monte-Carlo tree search method with a learned NN to efficiently learn a network model. However, Despite the high sample efficiency of the proposed model, the model design is complicated, especially for large PV systems.

2) Model-Free Based

Model-Free DRL (MFDRL) conducts the optimal policy without explicitly learning the model of the environment. It can be achieved by using three approaches: Value-Based RL (VBRL), Policy-Based DRL (PDRL), and Actor-Critic (AC) based [80]. VDRL is a prominent learning method to deal with high-dimensional state space and discrete or continuous action spaces in optimization problems [81]. On the other hand, PDRL architecture guarantees better convergence and keeps relatively high efficiency in high-dimensional or continuous action space [81]. However, agents require millions of time steps to learn tasks from many iterative systems. Since updates occur in small steps, agents may under-explore their environments or under-develop strategies, leading to exploration shortcomings in some cases. AC is a fusion of policy-based and value-based models to constitute an end-to-end learning paradigm from perception to action [82]. Asynchronous advantage actor-critic (A3C) and deep deterministic policy gradient, and twin delayed deep deterministic policy gradient (TD3) were the standard representations of AC method [81].

MFDRL is widely depicted for optimization and control of PVPG in research works. To name a few, in [83], the authors adopted a novel strategy that brings together DQN and CNN to cope with the uncertainties of an isolated microgrid. Concisely, the proposed DRL model optimizes the sum of diesel generators' generation cost and the penalty of non-served power demand. However, the curse of dimensionality persists with the DQN. A combination of Policy Dynamics based Win or learn Fast-Policy Hill Climbing (PDWoLF-PHC) and Back Propagation Neural Network (BPNN) network has been adopted to tackle the RES uncertainties for fast-response regulation units [84]. The proposed model optimizes the

TABLE 2: Comparison between popular DL algorithms

Model	Ref.	Advantage	Weakness
LSTM	[54]	- Effectively model dynamics in sequential data - Flexible input dimensions	- Lack of spatial data modeling - Difficult HO
MLPs	[41]	- Able to reveal the highly nonlinear relationships - Easy-to-implement	- Insufficiency for high-dimensional input data representation - Poor performance for spatio-temporal patterns
GRU	[54], [74], [75]	- Less computational cost compared to LSTM - Less vulnerable to the gradient explosion dilemma	- Limited applicability - Lack of spatial data modeling
CNN	[76]	- Efficient spatial feature extraction - Distributed deployment	- Expensive computational requirements - Limited temporal modeling potential
AE	[9], [76]	- Does not require labeled data - Robustness and noise-resilient architecture	- Time-consuming pre-training process - Prone to gradient explosion
GNN	[77]	- Efficient for graph-structured information modeling - Explicitly extract high-level representations	- Require extensive work in building a graph for non-graph data - Prone to time complexity issue
RBM	[63]	- Can learn high-dimensional feature patterns - flexible and useful for pre-training	- Low robustness for noisy data - Sensitive to parameter initialization
ResNet	[78]	- Powerful representational ability - High compelling performance with multiple layers	- Increased complexity and low explainability potential - Prone to over-fitting
GAN	[77]	- Does not require Monte Carlo approximation for training - Does not a deterministic bias	- Low interpretability and explainability - Unstable training
DRL	[79]	- Unsupervised training - Able to scale tasks with high-dimensional state and action spaces	- Slow convergence and long computation time - May has convergence issues in practice
DBN	[63]	- Unsupervised greedy learning - Less computationally expensive	- Heavy HO - Difficult implementation

coordinated control for the source-grid-load. Despite the high efficiency of the proposed model for automatic generation control, its deployment requires high exploration costs in a multi-area interconnected grid. A multi-agent Double Deep Q Network and an Action Discovery (DDQN-AD) has been proposed for distributed RES management [85]. However, the proposed model is limited to homogeneous agents to work effectively. Paper [86] proposed an automatic generation control-based AC strategy. The proposed method relies on DQN to follow an isolated microgrid paradigm's interaction agent-environment without the need for RE forecasts. However, the proposed system can only have deterministic policies, limiting its feasibility in practical power grids. To sum up, DRL is limited for optimization and control tasks, enhancing the prediction systems' efficiency. Nonetheless, the forecasting problems remain unsolved entirely via DRL, which requires more investigations.

D. HYBRID MODELS-BASED

The efficiency of stand-alone DL models can be unsatisfactory in PVPF in different case scenarios due to inappropriate HO, bias-overfitting conundrum, and unbearable complexity in both computational and spatial dimensions. To bridge that gap, the combination of two or more cross-discipline methods (a.k.a hybrid models) is commonly proposed to forecast PVPF with an improved performance than the single DL models [7]. This performance enhancement refers to the fact that the single DL models have their strengths and limitations, as reported in Table 2. Specifically, DL has some

limitations, including the lack of interpretability with DL outputs that we cannot even fathom how they are generated yet, extensive computation requirements, and the need for massive data to efficiently perform the desired task. Hybrid models are often preferred for solving the insurmountable PVPF problems to eliminate or reduce the shortcomings of single models by combining them with another model in order to obtain impressive results [87]. Fig. 9 shows a comprehensive distribution of the reviewed papers in this review according to the forecasting method.

As remarked in Fig. 9, the hybrid methods are by far more deployed than the rest of DL models. For instance, a Conv-GRU model has been proposed to predict the PVPF accurately [88]. The proposed model provides a high versatility to deal with the nonlinear behaviors to provide an accurate PV output. A CNN-LSTM model has been proposed for PVPF [89]. However, the extraction of positional and temporal representation in the PV output requires explicit recognition of patterns and regularities in data, challenging to compute due to the massive computational burden in real-life application. An AE-LSTM model has been proposed [9]. A DBN-based Auto-Regressive has been proposed for nonlinear TS modeling [90], which provides decent performance. But, the algorithm is fragile when faced with the PV volatile behavior when applied to different locations and not suitable for PVPF. An innovative USTF method has been depicted in [34]. The authors' work consists of implementing of the underlying Local Sensitive Hash algorithm (LSH). The used taxonomy considers four weather conditions: clear, cloudy, rainy, and

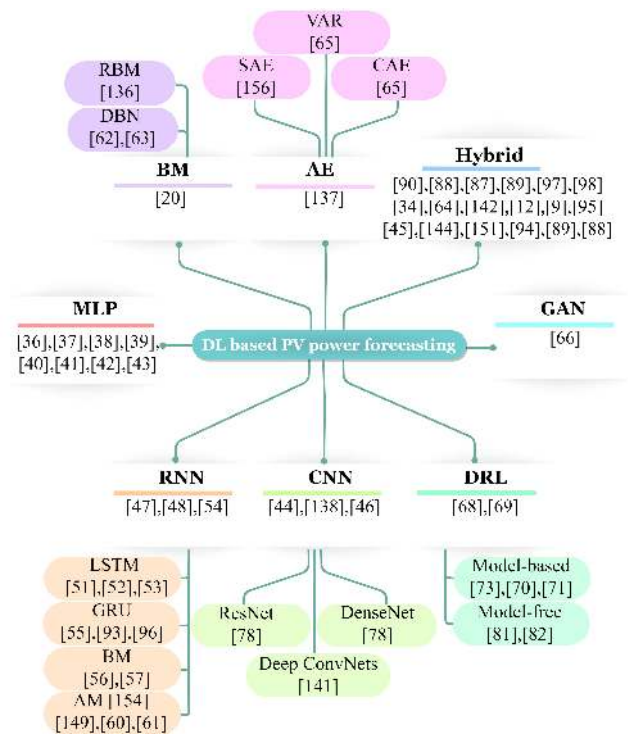


FIGURE 9: Distribution of the selected papers per DL method.

snowy weather. LSH profoundly investigates the coupling correlated weather features. The methodology adopted for LSH system classifies the PV power segments and generates a PPF output. In [34], the authors exploited a hybrid method for an accurate hourly PV power prediction based on a gradient-descent Back-Propagation method (BP), Schema Frog Leaping Algorithm (SFLA), and ANN named BP-SFLA-ANNs model. Subsequently, they adopted BP-SFLA-ANNs model consists of using SFLA to mediate between BP and ANNs models. BP model provides the values of the primary hyperparameters of ANNs to let the SFLA start from this initial selection to further search for more suitable parameters of a typical ANNs. The interaction between SFLA and the BP led to a superior ANN accuracy and less computational burden compared to an SFLA-ANNs without the initial tuning of BP. Further applications of hybrid models in PVPF are listed in Table 3.

However, computational complexity is one of the main weaknesses of the hybrid models due to using of two or more techniques. Thus, the accuracy improvement should not compromise the computational complexity of mixed models. The performance of a mixed model depends on the performance of a single model.

E. HYPERPARAMETER OPTIMIZATION OF DL ARCHITECTURES

The ever-increasing complexity of the newly developed DL methods has raised an emerging resurgence of research on

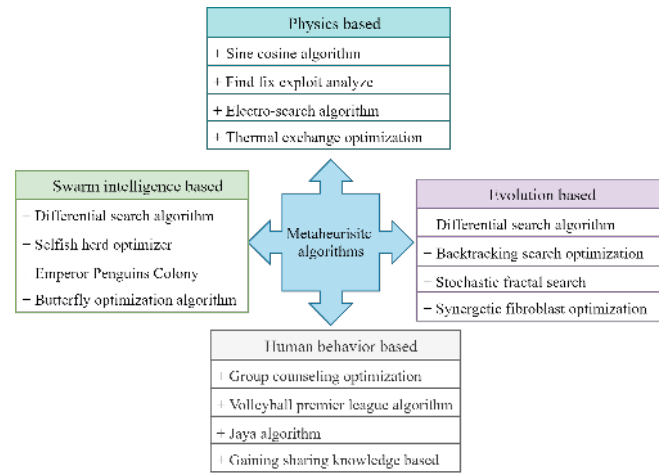


FIGURE 10: Meta-heuristic algorithms with Sine cosine algorithm [99], Find fix exploit analyze [100], Electro-search algorithm [101], Selfish heard algorithm [102], Emperor Penguins colony [103], Butterfly optimization algorithm [104], Group counseling optimization [105], Volleyball premier league algorithm [106], Jaya algorithm [107], Gaining sharing knowledge based [108], Differential search algorithm [109], Backtracking search optimization [110], Stochastic fractal search [111], Synergistic fibroblast optimization [112].

HO. A wide range of hyper-tuning techniques was adopted to support DL algorithms or provide an alternative for specific optimization tasks. Automated hyperparameter selection is an essential step to save the rare resources of human expertise and notorious efforts. Meta-heuristic techniques offer the adequate tools to provide an optimal or near-optimal configuration of DL models due to their efficiency and scalability for various complex applications. Meta-heuristic algorithms can be divided into four distinguished categories, specifically, evolution-based, swarm intelligence-based, physics-based, and human behavior-based, as shown in Fig. 10.

For instance, Ant Colony Optimization algorithm has been proposed for model tuning to accurately predict the PVPF [113]. Paper [114] developed a CNN and a Salp Swarm Algorithm (SSA) for PVPF. For different types of weather, five CNN regression models are designed. Consequently, the prediction engine is easy-to-implement even if the knowledge of the hyperparameters was limited. Paper [56] adopted a combination of BiLSTM, Sine Cosine Algorithm (SCA), and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) for solar irradiance forecasting. Similarly, the authors in [115] established an improved BiLSTM model with Genetic Algorithm (GA). However, the proposed version of the BiLSTM model is susceptible to the random weight initialization.

TABLE 3: Related hybrid DL models.

Methods	Ref.	Error metrics	Lowest error	Data set
VAE	[65]	RMSE, MAE, R^2 , EV	$R^2 = 99.50\%$	National Institute of Standard and Technology grid-connected plant in Algeria
CNN-LSTM	[91]	MAPE, RMSE, MAE	RMSE=0.13kW	PV inverter installed in Haenam, South Korea
PC-LSTM	[92]	MSE, MAE, R^2	$R^2 = 91\%$	Different areas in Australia
NARX-LSTM	[12]	nMAE, nRMSE, R^2	$R^2 = 99.89\%$	Desert Knowledge Australia Solar Center Urbana-Champaign solar farm-Flyover, USA
GRU-RNN	[93]	NRMSE, MAE, MAPE	MAPE=1.27%	the arid Northern Cape region, South Africa
MC CNN	[94]	RMSE, MAPE, R^2	MAPE= 8.639 %	PV plants at 164 sites in South Korea
ConvLSTM	[95]	RMSE, MAE, MAPE	RMSE=6.64kW	PV plant in Morocco
GRU	[96]	MAPE, RMSE	RMSE=67.29W	Solar Radiation Database for Phoenix International Airport in Phoenix, Arizona.
DenseNet	[78]	MSE, MAE, MSLE, MASE, MHE	MAE=0.152kW	Desert Knowledge Data from Solar Center, Australia
ResNet	[78]	MSE, MAE, MSLE, MASE, MHE	MAE=0.180kW	Desert Knowledge Australia Solar Center
Modified LSTM	[97]	MAE, RMSE	RMSE=0.55kW	Ansan, Gyeonggi-do, Korea
VMD-ARMA-DBN	[98]	NRMSE, NMAE, TIC	NMAE=1.033%	PV power plant in China-Yunnan

IV. DEEP LEARNING MODELS USED IN PVPF APPLICATIONS

In this section, the notable techniques for PVPF have been discussed, including DTL, BDDL, incremental learning, on-line learning, and federated learning.

A. FEDERATED LEARNING BASED PVPF

As DL models based on a central server are suffering from critical privacy-intrusive and security challenges, without explicit awareness of the users, these concerns hold particularly strong. Concerns about data awareness in PV systems may lead end-users to be progressively unwilling to send their potentially private and personal data to data centers, raising the problem about how DL algorithms will be trained [116]. Federated Learning (FL) is regarded as privacy-preserving collaborative learning that gained a whispered emphasis from academics and practitioners due to its significant contributions [117]. This emergent technology permits various parties ideally, mobile devices, to cooperatively train a DL model on their combined sets, without any participants having to reveal their local data to a centralized server.

For the energy hub, FL seems to be the missing puzzle piece in the widespread adoption of SE. For instance, the authors in [118] applied a secure FL method based on Bayesian LSTM Model to address solar irradiation prediction. The proposed model proves its efficiency in solving the problem of sparse samples by using a distributed training framework. Meanwhile, the missing and faulty data can considerably hinder the training of the FL model. FL is still in the early stages of development. There are many works to do and a few structural achievements in the approval process that still need to happen before the researchers could leverage the FL technology to its full potential.

B. DEEP TRANSFER LEARNING IN PVPF

To further supplement the staggering potential of DL, Deep Transfer Learning (DTL) based on Transfer Learning (TL) and DL has been proposed to overcome the insufficiency problem of the actual training data sets convergence issues and isolated learning shortcomings [119]. DTL employs the pre-existing knowledge acquired from a DNN model for a particular task to solve related ones [120]. Concretely, the sensor measurements are hardly sufficient and sparsely generated [119]. Instead of learning any new task from scratch, DTL is devoted to avoiding much expensive data-labeling efforts using cross-domain data sets [121]. Furthermore, the HO-based DL algorithms could be conducted using TL [122]. This learning framework operates by transferring the knowledge gained by a DNN model in handling a task (source problem) to solve another related task. We define a source domain \mathcal{D}_s and a target domain \mathcal{D}_t with $\mathcal{D}_s \neq \mathcal{D}_t$. A learning and source tasks $(\mathcal{T}_s, \mathcal{T}_t)$ with $\mathcal{T}_s \neq \mathcal{T}_t$ constitute a DTL framework if η_t is a nonlinear function represented by a DNN. The source and target domains is formulated as [121]:

$$\mathcal{D}_s = \{(x_j^s, y_i^s)\}_{i=1}^M \quad (12)$$

$$\mathcal{D}_t = \{(x_j^t, y_j^t)\}_{j=1}^N \quad (13)$$

With M and N are labeled samples in \mathcal{D}_s and \mathcal{D}_t , respectively; X_s and X_t denote the feature spaces of source and target domains, respectively. x_i denotes the data instance and y_i is the related class label. In [123], a Shared-Optimized-Layer LSTM (SOL-LSTM) network has been proposed for PVPF. The rationale of the proposed model design is to combine a Sequential Model-Based Global Optimization (SMBO) with the LSTM network. DTL can solve the data insufficiency of the newly build PV plant by pre-training the hyperparameters in a similar source domain and fine-tuning in the target domain. However, with the increase of

the data's volume, the SOL-LSTM performance significantly decreases. In [122], a DNN has been proposed to tackle nonlinear weather uncertainties for RES farms. However, the DTL model seeks repeated access and preprocessing of a potentially massive data set of source tasks to establish the necessary knowledge base for the downstream target task. This requirement may not be feasible in large PV systems due to the lack of data-intensive computing resources. Therefore, it is of great significance to merge DTL and BD solutions to solve the data problems in large-scale PV systems.

Although the DTL would significantly enhance the performance of the DL-based PVPF, the research work on the DTL use for PV systems is relatively limited contrary to other areas such as fault detection in PV systems [124]. This may be related to the essential need for data sets with minimum reasonable similarity.

C. BIG DATA DEEP LEARNING FOR PVPF AND CHALLENGES

As known, DL models are data-hungry, requiring a large data to work effectively [125], [126]. Reciprocally, DL scales well with growing amounts of data [127]. However, the computational efficiency of the actual calculators is considerably limited, especially with the trend of increasing DNN size. To bridge this gap, it is, thereby, generally not an option to employ Big Data technologies for model training, especially with the continuous flow of data pouring from monitoring systems and smart devices [128]. Big Data and Analytics provide the means to predict the weather conditions and PVPF for PV systems to work at peak efficiency. Data mining can truly be beneficial for enhancing the PVPF in a complex ground-level infrastructure.

In [124], the authors have proposed a deep Feed-Forward Neural Network (FFNN) with big data set from Australia. A H2O package has been employed to combine non-deep and deep learning methods with Apache Spark cluster-computing framework. The simulation results proved that the lagged information does not need more than the previous 24 hours of historical information to provide and accurate results. However, the grid search algorithm tuning of the trained model may produce heavy computation requirements. Consequently, an adequate selection of optimization algorithms will be of utmost importance to establish cost-efficient prediction tools.

In [126], the authors introduced a big data solution by combining used physical and dynamical theories and intelligent algorithms to solve big data problems. However, decent documentation of the used software (*Sun4CastTM*) is missing. Paper [129] proposed an annual rooftop solar irradiation analysis based Spark-based fuzzy partitioning LSTM model. The proposed solution employs rooftop characteristics, horizontal solar irradiation, visibility of the sky, and shading factor. Hence, the simulations prove that ensemble models represented by RF overcome NN symbolized by extreme learning machine ensembles with an accuracy of 92%. The authors in [130] applied a big data forecasting tool

based to solve the PVPF problem, where, Pyspark package is employed to implement the neural network model. The simulation results produce an average RMSE=0.03 MW with fast convergence. Unfortunately, the computational complexity disables the model from the transition from proof of concept to production. Whereas a massive amount of engineering is needed to deploy it in production. Big Data frameworks applied to PV systems can contour several limitations such as data privacy and security, multisource data integration, real-time data processing to ensure that the data clearly conveys what they need the DL to learn for real-world PV plants [131].

D. INCREMENTAL AND ONLINE LEARNING

Under the umbrella of DL, incremental and online Learning has emerged as a continuous evolving scheme to improve the universality of the prediction engines for accurate regional forecasting. Hence, the difference between these two algorithms is quite challenging. Online Learning (OL) dynamically trains or adapts the model using each incoming data point at each time step, without saving [132]. Thus, OL is used to handle large volumes of streaming data transmitted at high velocity. Incremental Learning (IL) provides a fast remodeling from batch learning of data at different time intervals, and has the capability to integrate new knowledge on-the-fly of the predefined model if the network deems to be expanded. The incremental samples can be fed from the available Samples (SIL) or even unseen classes (CIL).

Authors [133] established an online PVPF method to handle concept drift. A model-agnostic online forecasting (MAOF)-based LSTM model is used, generating an MAPE=23.59%. However, the model produces a serious performance degradation in particular case scenarios leading to stability issues. The work in [134] applied an incremental learning model for solar irradiance based-Regression Enhanced Incremental Self-Organising Neural Network (RE-SOINN). The proposed model is trained incrementally as new data come in progressively. This architecture avoids the tedious retraining process of DL models.

V. APPLICATION POTENTIAL IN PV POWER FORECASTING

With the increasing spatiotemporally coupled uncertainties in PVPF, PVPF becomes a desperate need to ensure grid stability and weaken the uncertainty of solar PV power, hence paving the way towards a large-scale economic deployment of RES in the electricity grid. Conceptually, PVPF can be broadly classified into two folds, point PVPF and interval PVPF.

A. POINT PV POWER FORECASTING

Point forecasting, so-called deterministic forecasting, is widely regarded as essential for optimal power system management. Point forecasting models have been thoroughly researched over the years, and the trend of developing more accurate forecasts is still booming [135]. The average PVPF

$P_{t+k|t}$, is estimated to be produced from a PV system during a specific time period, for the PVPF forecast made at time step t , for a look-ahead time, $t+k$, if it would function under an equivalent constant PV power. The time horizon T , for which the prediction is generated, defines the total length of the forecast period. Deterministic PVPF provides accurate and specific future values [136], [137]. Further, these methods are easy-to-use, deploy and evaluate using score metrics such as RMSE, MAE, R, and MAPE [33]. Unsteady PVPF threatens energy generation. However, spot forecasts do not include the uncertainties around the mean value. Therefore, their results can be unreliable and misleading in particular scenarios [12].

Table 4 presents several exemplary applications of deterministic DL methods for PVPF. For instance, the authors in [144] adopted a LSTM-CNN model. It produced the most accurate forecasts over single LSTM and CNN, with a forecasting skill of 37%–45%. But, the computational burden is ten times longer than the standalone LSTM. A deep Extreme Learning Machine (ELM) has been applied in [145]. The proposed model incorporated Enhanced Colliding Bodies Optimization (ECBO), Variational Mode Decomposition (VMD), and a ELM algorithms. However, the proposed model does not shed light on uncertainty abstraction and reasoning. Therefore the forecasting engine is found inefficient in dealing with meteorological data pervaded with uncertainty.

B. PROBABILISTIC PV POWER FORECASTING

With the increasing PV power uncertainties, Probabilistic Deep Learning (PDL) has become the de-facto solution to lessen negative impacts on power system reliability and economic efficiency from stochastic PV generation [146]. PPVPF can provide an estimated interval where various possible PVPF values for a specific time are generated to quantify the intrinsic uncertainties associated with point forecasts [146], [147]. PPVPF draws excellent attention to balancing authorities for its ability to provide prediction interval, quantile, density, or conditional probability distribution of future predicted power [148]. From the literature, the most overwhelming PPVPF methods lie in conditional quantile regression (QR) and conditional expectile regression [149]. For instance, An Improved quantile CNN has been proposed for indirect PVPF to compute consistent quantile estimates [150]. The simulation results indicate that two-stage training strategy has a positive influence on enhancing forecasting accuracy. A deep QR-CNN-based Wavelet Transform (WT) has been exploited to model DPVPF and PPVPF [151]. The CNN-WT efficiently provides a wider view of a prediction. The proposed model is tested using TS data collected by Elia, Belgium's transmission system operator. The erage Coverage Error (ACE) obtained varies from -1.02 to 0.43 for a prediction horizon of 15 min.

In [149], the authors propose a Robust Self-Attention Multi-horizon (RSAM) model for PPVPF using QR. However, the QR generates a non-differentiable loss function that

threatens the model stability and robustness. The proposed model employs a self-attention-based transformer model. However, the problem of crossing quantile curves is frequently observed, particularly when considering a dense set of quantiles or using a small data set. Furthermore, the DNN is naturally deterministic and limited to PPVPF. Therefore, Bayesian probability is often integrated with DL models to provide prediction intervals associated with forecasted point values [152], [153]. Substantial research has shown that PPVPF is scarcely investigated compared to deterministic PVPF. A number of DL techniques have been exploited in the literature for PPVPF. For instance, paper [154] presents a Robust Self-Attention Multi-horizon (RSAM). The proposed model indicates an 18.60% improvement compared to the conventional LSTM. A Traditional Encoder Single Deep Learning (TESDL) framework has been proposed, which provides a 27% improvement in accuracy factor [155]. Reference [156] exploited an SAE and Lower Upper Bound Estimation Method for PPVPF. It was found that the wind speed, weather temperature, weather relative humidity, global horizontal radiation, and diffuse horizontal radiation can effectively predict PV energy production. Despite their importance, PPVPF may ignore the interdependence shape of forecast errors among look-ahead timesteps, and may lose their potential in practical use in the time-dependent and multi-stage decision-making processes, such as the trading strategies design in a multi-market environment.

VI. CHALLENGES AND FUTURE RESEARCH DIRECTIONS

The last section concludes this review with rigorous investigations and guidelines for future studies. The main findings and the research frontiers of this study are enlisted as following:

- There is a desperate need to convince the PV experts that DL concept is efficient and satisfactory to gain acceptance by operators and stakeholders. This high-tech concept needs to overcome several weaknesses to win broader acceptance and confidence from the energy hub. This poor DL integration can be explained by various reasons from the industrial perspective. For instance, DL limitations lie in poor generalization potential in learning evolving operating conditions. with the high complexity of the atmospheric condition, designing an ideal DL method is beyond the bounds of possibility. To build PV industry trust, DL models have to overcome the lack of representativeness of the train sets and the potential adversarial attacks.
- The time resolution for a large extent of the research works emphasizes the STF-based PVPF. However, the MTF and LTF are essential for energy trading, strategic planning, and degradation-rate-impacted energy potentials of PV panels. Nevertheless, the performance validity expires in a more extended period with increased error values compared to STF forecasts.

TABLE 4: Selected recent works on DL-based PVPF.

Method	Ref.	Benchmarks	Inputs	Data set source
CNN	[138]	SVM ¹ , RF, DT, MLP, LSTM	Temperature, Solar radiation, Historical PV power	PV station in Taiwan
FSEL	[139]	LSTM, ARIMA, VAR	Installation price, Customer price, Temperature, Average sunshine, Population density, Treasury yield	local PV installation in Jeollanam-do and Gwangju
Self-attention	[140]	SVM, DNN, ELM, CNN, LSTM	Time information, Weather information and Historical power generation	Power generation data set published by the University of Massachusetts Amherst in 2017
Deep ConvNets	[141]	PM, ResNet	Solar radiation, Sky images	Folsom, CA, USA -NREL Golden, CO Data Set
AE-LSTM-PM	[64]	PM, FFNN, DNN, LSTM, AE-LSTM	Irradiance, Solar Zenith angle, Temperature, Minute index, Cloud cover, Day index	SOA university Bhubaneswar, Odisha, India
Copula function-LSTM	[142]	LSTM, SVM	Temperature, Wind speed, Wind direction, Cloud cover ratio, Radiation, Surface pressure, Precipitation, Humidity	PV power station data from China and the US.
LSTM	[143]	LSTM-RNN, BPNN ² , SVM, PM	Historical PV power	Desert Knowledge Australia Solar Center (DKASC)
ResNet, DenseNet	[78]	SVR, ES, MLP, CNN, QR	Historical PV power, Numerical weather prediction, Diffuse and global horizontal radiation	Desert Knowledge Australia Solar Center (DKASC)

¹ Support Vector Machines ² Back-propagation Neural Network

- The hybrid models possess immense potentialities in PV paradigms. Consequently, the number of papers that address the hybridization of DL is ever-growing. However, the model hybridization entails an elevated computational burden. Almost half of the reviewed papers for forecasting applications incorporate hybrid models. It can be deduced from the related works that GRU and LSTM architectures have been frequently deployed in the model fusion due to their suitability in time series data. The hybrid models' high accuracy should not compromise their reasonable complexity.
- The shortage of PV skilled professionals and experts to deploy DL techniques presents a severe problem that impedes the vast deployment of these techniques. The major concern for PV practitioners is the lack of clear guidance rules for algorithms' structure and HP tuning. Therefore, finding near-optimal solutions can be a cumbersome problem for non-qualified operators. As DL jobs are in high demand, matching DL and PV technology landscape is relatively limited. More importantly, it is quite hard to find qualified man-labor in both domains of interest. Getting sufficient professional knowledge in both DL and PV technologies requires personal initiative due to the lack of resources in these infancy subjects.
- The explainability and interpretability of the proposed PV systems is a severe challenge for their practical feasibility. DL models operate in a "black-box" fashion, impeding their whispered adoption due to the lack of explainability. The poor visibility of model performance may lead to manufacturing problems especially due to safety-critical concerns. Nevertheless, interpretable DL methods have scarcely been applied and tested in PVPF,

- where the transparency and understandability of the decision logic of the forecasting engine are not guaranteed.
- The data mining and big data analytics are essential for the cost-effectiveness of predictive modeling in PV systems. The data accumulated from weather stations are processed in a continuous flow or stream with various formats, sizes, and variability. Big data analytics is a means to improve data power stability. However, most research papers do not shed light on the utilization of DL models in actual PV plants.
- PVPF is recently proposed to mitigate energy uncertainties. However, the proposed methods may face several implementation issues, as reported in Table 2. In fact, most of the proposed methods were still in the proof-of-concept stage without passing to real-world applications. The implementation barrier lies in the laborious DL adjustment in terms of storage, dimension, search capacity, and convergence settings for actual standalone or grid-connected PV systems.
- The available DL models were commonly validated in the Standard Test Conditions (STC). Nonetheless, very little work has been done for real-world validation, where the performance accuracy dramatically decreases for real PV systems.
- Data privacy awareness presents the key enabler for the integrity and reliability of forecasting systems. There is a pressing need to protect data privacy from vulnerabilities and cyber intrusions for promoting the DL deployment. For instance, false data detection tools are needed to preserve the forecasting engine and achieve effective decision-making. FL-based PVPF presents a prospective direction towards a Secure and Resilient grid [118]. However, the studies aiming to cover data-

driven cybersecurity technologies-based PVPF are exceedingly rare in the literature.

- The proposed techniques often operate for a specific time frame over a specific geographical area. For instance, DL models can outperform the benchmarks for the forecasting situation under scrutiny. But, this model could not perform equally in other PV areas with different topographies and weather patterns. In fact, DL is hitherto inefficient in regional PVPF and limited to the technical characteristics of PV plants. Designing general models that could transfer learning from local PV plant to another is a potential solution towards model universality.
- The online prediction tools are seldom investigated for PV plants despite their importance in real PV systems. Online prediction effectively adapts to newly incoming information. More concretely, offline DL methods fail with unpredictable conditions during the process. Online PVPF can obtain near-optimal predictions and promotes model stability. This requires an engagement of sufficient storage space of the infrastructure and an adequate frequency for model updates.
- Another challenge is pre and post-processing of data. Data preprocessing and error post-processing is a serious concern, especially with the sheer size of data. The unfathomable amounts of data lead to noise and imprecise knowledge problems such that they can be difficult to surmount. Data preprocessing usually includes data normalization, faulty and missing data filtering, data resolution adjustment, data augmentation correlation analysis data clustering, and graph constriction. While, post-processing procedures usually include various pruning routines, rule filtering, or even knowledge integration. Paper [32] taxonomized post-processing techniques for solar forecasting into four classes: deterministic-to-deterministic, probabilistic-to-deterministic, deterministic-to-probabilistic, and probabilistic-to-probabilistic post-processing. In this paper, the authors reported that post-processing is vital for consistency, quality, and value in the PVPF context [32].
- The integration of satellite and ground-based measurements is rarely being studied. The reason for this shortage may refer to the limited access to satellite images which prohibits the data collection. Despite satellite-derived irradiance datasets provide spatial diversity, the obtained accuracy is relatively low in the presence of poor datasets. Designing a data-light DL systems is a potential solution to overcome the large data requirements barrier and achieve satisfactory results.
- Standardizing forecast evaluation towards an universal functional form is mandatory to facilitate the selection of the suitable model among others. Quantifying predictability presents a tiresome task due to a large number of error metrics. The diversified score criteria may inhibit attaining a statistical consensus of the model's

goodness. Unifying the score metrics is essential to gain the industry and end-users acceptability in real-world problems. The standardized criteria alleviate the costs of the prediction system and the economic impact. Although error metrics standardization seems an intuitively appealing task, research works are limited towards that goal.

- Hierarchical TS Forecasting (HTSF) is deemed suitable to achieve excellent performance in PVPF through explanatory variables. However, the use of HTSF is limited. HTSF follows a hierarchical aggregation structure at different levels by reconciling incoherent forecasts according to their proximity from individual TS. Therefore, the relationships within the hierarchy are preserved. This entails a problem of coherency at different granularity levels of the time-varying observations. Therefore, the adjustment between upper and lower levels is vital to ensure the consistency of the forecast. DL is particularly well-positioned to predict the TS data of all nodes in the hierarchy and reconcile them [157]. However, to the authors' best knowledge, HTSF is still not applied to PV systems.
- Forecasting with multimodal and multilevel information fusion is scarcely discussed. Multimodal learning allows learners to merge the information from different sources. In the PV context, the multimodal data may include sky images and cloud motion speed records. These heterogeneous data from different modalities present complementary information from multiple sources. Information fusion from different modalities with strong end-to-end governance standards can significantly enhance the prediction accuracy, boosting interest towards concepts to model in this area.
- Smart meters (SM) sensing may bring rigorous challenges to forecasting accuracy, especially with their short service life span. A meter failure can bring a plethora of problems for simple causes such as internet loss, software flaws, and hardware malfunctions. However, manually checking all SM on a regular basis can be labor-intensive. DL can efficiently work for early detection of inaccurate SM, towards longer-lasting SM. The careful investigation of the data reliability and sensing tampering early on helps in producing more robust predictions.
- The performance of DL methods depends on the accessibility of abundant quality of PV data to meet power quality standards. However, these models run slowly and have narrow boundaries of the frequency domain division in the production environment. The DL deployment lies in three major aspects: portability, scalability, and computational cost. Operationalizing and robustifying DL are still tedious tasks that mandate fruitful research in this direction.

VII. CONCLUSION

With the inevitable emergence of the Smart Grid, DL-as-a-service plays an essential and indispensable role in the bulk penetration of Photovoltaic (PV) energy across efficient PV Power Forecasting (PVPF) systems. This paper provides a comprehensive review of the recent PVPF involving DL. We took a deeper dive into the well-known architectures for PVPF. Three types of emerging learning methods are classified, specifically, discriminative learning, generative learning and deep reinforcement learning. The DL methods have their own merits and numerous shortcomings, which may be covered by optimal hyperparameter tuning. Different PVPF strategies concerning time horizons have been described in the study. A vast majority of case studies from the literature demonstrated that hybridization and assembling straighten DL techniques leading to better accuracy and high robustness. It is hoped that this review paper would help researchers and practitioners to improve forecasting accuracy through moving DL models from the nascent stage to real-world applications and to come up with more precise PV energy forecasts.

REFERENCES

- [1] I.-Y. Joo and D.-H. Choi, "Distributed optimization framework for energy management of multiple smart homes with distributed energy resources," *IEEE Access*, vol. 5, pp. 15 551–15 560, 2017.
- [2] G. M. Wilson, M. Al-Jassim, W. K. Metzger, S. W. Glunz, P. Verlinden, G. Xiong, L. M. Mansfield, B. J. Stanbery, K. Zhu, Y. Yan et al., "The 2020 photovoltaic technologies roadmap," *Journal of Physics D: Applied Physics*, vol. 53, no. 49, p. 493001, 2020.
- [3] E. Pursiheimo, H. Holttinen, and T. Koljonen, "Inter-sectoral effects of high renewable energy share in global energy system," *Renewable Energy*, vol. 136, pp. 1119–1129, 2019.
- [4] M. Massaoudi, S. S. Refaat, H. Abu-Rub, I. Chihi, and F. S. Oueslati, "A hybrid bayesian ridge regression-cwt-catboost model for pv power forecasting," in *2020 IEEE Kansas Power and Energy Conference (KPEC)*. IEEE, 2020, pp. 1–5.
- [5] B. K. Sahu, "A study on global solar pv energy developments and policies with special focus on the top ten solar pv power producing countries," *Renewable and Sustainable Energy Reviews*, vol. 43, pp. 621–634, 2015.
- [6] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy forecasting: A review and outlook," *IEEE Open Access Journal of Power and Energy*, 2020.
- [7] M. Massaoudi, S. S. Refaat, H. Abu-Rub, I. Chihi, and F. S. Oueslati, "Pls-cnn-bilstm: An end-to-end algorithm-based savitzky-golay smoothing and evolution strategy for load forecasting," *Energies*, vol. 13, no. 20, p. 5464, 2020.
- [8] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub, "A novel stacked generalization ensemble-based hybrid lgbm-xgb-mlp model for short-term load forecasting," *Energy*, vol. 214, p. 118874, 2020.
- [9] A. Gensler, J. Henze, B. Sick, and N. Raabe, "Deep learning for solar power forecasting—an approach using autoencoder and lstm neural networks," in *2016 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, 2016, pp. 002 858–002 865.
- [10] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569–582, 2017.
- [11] S. Boubaker, M. Benganem, A. Mellit, A. Lefza, O. Kahouli, and L. Kolsi, "Deep neural networks for predicting solar radiation at hail region, saudi arabia," *IEEE Access*, vol. 9, pp. 36 719–36 729, 2021.
- [12] M. Massaoudi, I. Chihi, L. Sidhom, M. Trabelsi, S. S. Refaat, H. Abu-Rub, and F. S. Oueslati, "An effective hybrid narx-lstm model for point and interval pv power forecasting," *IEEE Access*, vol. 9, pp. 36 571–36 588, 2021.
- [13] G. M. Yagli, D. Yang, and D. Srinivasan, "Automatic hourly solar forecasting using machine learning models," *Renewable and Sustainable Energy Reviews*, vol. 105, pp. 487–498, 2019.
- [14] J. Alonso-Montesinos and F. J. Battles, "Solar radiation forecasting in the short- and medium-term under all sky conditions," *Energy*, vol. 83, pp. 387–393, 2015.
- [15] V. Kallio-Myers, A. Riihelä, P. Lahtinen, and A. Lindfors, "Global horizontal irradiance forecast for finland based on geostationary weather satellite data," *Solar Energy*, vol. 198, pp. 68–80, 2020.
- [16] A. Murata, H. Ohtake, and T. Oozeki, "Modeling of uncertainty of solar irradiance forecasts on numerical weather predictions with the estimation of multiple confidence intervals," *Renewable energy*, vol. 117, pp. 193–201, 2018.
- [17] L. Gigoni, A. Betti, E. Crisostomi, A. Franco, M. Tucci, F. Bizzarri, and D. Mucci, "Day-ahead hourly forecasting of power generation from photovoltaic plants," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 2, pp. 831–842, 2017.
- [18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [19] M. Massaoudi, I. Chihi, L. Sidhom, M. Trabelsi, S. S. Refaat, and F. S. Oueslati, "Enhanced random forest model for robust short-term photovoltaic power forecasting using weather measurements," *Energies*, vol. 14, no. 13, p. 3992, 2021.
- [20] M. Massaoudi, H. Abu-Rub, S. S. Refaat, I. Chihi, and F. S. Oueslati, "Deep learning in smart grid technology: A review of recent advancements and future prospects," *IEEE Access*, vol. 9, pp. 54 558–54 578, 2021.
- [21] M. A. F. Lima, P. C. Carvalho, L. M. Fernández-Ramírez, and A. P. Braga, "Improving solar forecasting using deep learning and portfolio theory integration," *Energy*, vol. 195, p. 117016, 2020.
- [22] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. Idris, W. Van Deventer, B. Horan, and A. Stojcevski, "Forecasting of photovoltaic power generation and model optimization: A review," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 912–928, 2018.
- [23] C. Croonenbroeck and G. Stadtmann, "Renewable generation forecast studies—review and good practice guidance," *Renewable and Sustainable Energy Reviews*, vol. 108, pp. 312–322, 2019.
- [24] N. J. Van Eck and L. Waltman, "Software survey: Vosviewer, a computer program for bibliometric mapping," *scientometrics*, vol. 84, no. 2, pp. 523–538, 2010.
- [25] D. W. Van der Meer, J. Widén, and J. Munkhammar, "Review on probabilistic forecasting of photovoltaic power production and electricity consumption," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1484–1512, 2018.
- [26] M. N. Akhter, S. Mekhilef, H. Mokhlis, and N. M. Shah, "Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques," *IET Renewable Power Generation*, vol. 13, no. 7, pp. 1009–1023, 2019.
- [27] A. Mellit, A. Massi Pavan, E. Ogliari, S. Leva, and V. Lughi, "Advanced methods for photovoltaic output power forecasting: A review," *Applied Sciences*, vol. 10, no. 2, p. 487, 2020.
- [28] R. Ahmed, V. Sreeram, Y. Mishra, and M. Arif, "A review and evaluation of the state-of-the-art in pv solar power forecasting: Techniques and optimization," *Renewable and Sustainable Energy Reviews*, vol. 124, p. 109792, 2020.
- [29] G. de Freitas Viscondi and S. N. Alves-Souza, "A systematic literature review on big data for solar photovoltaic electricity generation forecasting," *Sustainable Energy Technologies and Assessments*, vol. 31, pp. 54–63, 2019.
- [30] J. Devaraj, R. Madurai Elavarasan, G. Shafiullah, T. Jamal, and I. Khan, "A holistic review on energy forecasting using big data and deep learning models," *International Journal of Energy Research*, 2021.
- [31] D. Yang, W. Li, G. M. Yagli, and D. Srinivasan, "Operational solar forecasting for grid integration: Standards, challenges, and outlook," *Solar Energy*, vol. 224, pp. 930–937, 2021.
- [32] D. Yang and D. van der Meer, "Post-processing in solar forecasting: Ten overarching thinking tools," *Renewable and Sustainable Energy Reviews*, vol. 140, p. 110735, 2021.
- [33] M. Massaoudi, S. S. Refaat, and H. Abu-Rub, *On the Pivotal Role of Artificial Intelligence Toward the Evolution of Smart Grids*. John Wiley Sons, Ltd, 2021, ch. 15, pp. 359–420. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119422464.ch15>

- [34] M. Yang and X. Huang, "Ultra-short-term prediction of photovoltaic power based on periodic extraction of pv energy and lsh algorithm," *IEEE Access*, vol. 6, pp. 51 200–51 205, 2018.
- [35] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and P. Group, "Preferred reporting items for systematic reviews and meta-analyses: the prisma statement," *PLoS medicine*, vol. 6, no. 7, p. e1000097, 2009.
- [36] A. Moncada, W. Richardson, and R. Vega-Avila, "Deep learning to forecast solar irradiance using a six-month utsa skyimager dataset," *Energies*, vol. 11, no. 8, p. 1988, 2018.
- [37] U. Hasson, S. A. Nastase, and A. Goldstein, "Direct fit to nature: An evolutionary perspective on biological and artificial neural networks," *Neuron*, vol. 105, no. 3, pp. 416–434, 2020.
- [38] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub, "A novel stacked generalization ensemble-based hybrid lgbm-xgb-mlp model for short-term load forecasting," *Energy*, vol. 214, p. 118874, 2021.
- [39] R. M. Ehsan, S. P. Simon, and P. Venkateswaran, "Day-ahead forecasting of solar photovoltaic output power using multilayer perceptron," *Neural Computing and Applications*, vol. 28, no. 12, pp. 3981–3992, 2017.
- [40] Q. Liu and Q.-J. Zhang, "Accuracy improvement of energy prediction for solar-energy-powered embedded systems," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 6, pp. 2062–2074, 2015.
- [41] Z. Zhen, J. Liu, Z. Zhang, F. Wang, H. Chai, Y. Yu, X. Lu, T. Wang, and Y. Lin, "Deep learning based surface irradiance mapping model for solar pv power forecasting using sky image," *IEEE Transactions on Industry Applications*, vol. 56, no. 4, pp. 3385–3396, 2020.
- [42] C. Voyant, G. Nottot, C. Darras, A. Fouilloy, and F. Motte, "Uncertainties in global radiation time series forecasting using machine learning: The multilayer perceptron case," *Energy*, vol. 125, pp. 248–257, 2017.
- [43] K. Mahmud, S. Azam, A. Karim, S. Zobaed, B. Shanmugam, and D. Mathur, "Machine learning based pv power generation forecasting in alicia springs," *IEEE Access*, vol. 9, pp. 46 117–46 128, 2021.
- [44] Y. Sun, G. Szűcs, and A. R. Brandt, "Solar pv output prediction from video streams using convolutional neural networks," *Energy & Environmental Science*, vol. 11, no. 7, pp. 1811–1818, 2018.
- [45] J. Qu, Z. Qian, and Y. Pei, "Day-ahead hourly photovoltaic power forecasting using attention-based cnn-lstm neural network embedded with multiple relevant and target variables prediction pattern," *Energy*, vol. 232, p. 120996, 2021.
- [46] Y. Sun, V. Venugopal, and A. R. Brandt, "Short-term solar power forecast with deep learning: Exploring optimal input and output configuration," *Solar Energy*, vol. 188, pp. 730–741, 2019.
- [47] M. Massaoudi, I. Chihi, L. Sidhom, M. Trabelsi, S. S. Refaat, and F. S. Oueslati, "Performance evaluation of deep recurrent neural networks architectures: Application to pv power forecasting," in *2019 2nd International Conference on Smart Grid and Renewable Energy (SGRE)*. IEEE, 2019, pp. 1–6.
- [48] A. Alzahrani, P. Shamsi, C. Dagli, and M. Ferdowsi, "Solar irradiance forecasting using deep neural networks," *Procedia Computer Science*, vol. 114, pp. 304–313, 2017.
- [49] K. Wang, X. Qi, and H. Liu, "Photovoltaic power forecasting based lstm-convolutional network," *Energy*, vol. 189, p. 116225, 2019.
- [50] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [51] M. AlKandari and I. Ahmad, "Solar power generation forecasting using ensemble approach based on deep learning and statistical methods," *Applied Computing and Informatics*, 2020.
- [52] Y. Jung, J. Jung, B. Kim, and S. Han, "Long short-term memory recurrent neural network for modeling temporal patterns in long-term power forecasting for solar pv facilities: Case study of south korea," *Journal of Cleaner Production*, vol. 250, p. 119476, 2020.
- [53] L. Wen, K. Zhou, S. Yang, and X. Lu, "Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting," *Energy*, vol. 171, pp. 1053–1065, 2019.
- [54] M. Massaoudi, I. Chihi, L. Sidhom, M. Trabelsi, S. S. Refaat, and F. S. Oueslati, "Performance evaluation of deep recurrent neural networks architectures: Application to pv power forecasting," in *2019 2nd International Conference on Smart Grid and Renewable Energy (SGRE)*. IEEE, 2019, pp. 1–6.
- [55] Y. Wang, W. Liao, and Y. Chang, "Gated recurrent unit network-based short-term photovoltaic forecasting," *Energies*, vol. 11, no. 8, p. 2163, 2018.
- [56] T. Peng, C. Zhang, J. Zhou, and M. S. Nazir, "An integrated framework of bi-directional long-short term memory (bilstm) based on sine cosine algorithm for hourly solar radiation forecasting," *Energy*, vol. 221, p. 119887, 2021.
- [57] H. Sharadga, S. Hajimirza, and R. S. Balog, "Time series forecasting of solar power generation for large-scale photovoltaic plants," *Renewable Energy*, vol. 150, pp. 797–807, 2020.
- [58] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
- [59] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.
- [60] D. Yu, W. Choi, M. Kim, and L. Liu, "Forecasting day-ahead hourly photovoltaic power generation using convolutional self-attention based long short-term memory," *Energies*, vol. 13, no. 15, p. 4017, 2020.
- [61] C. Pan, J. Tan, D. Feng, and Y. Li, "Very short-term solar generation forecasting based on lstm with temporal attention mechanism," in *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*. IEEE, 2019, pp. 267–271.
- [62] H. Zang, L. Cheng, T. Ding, K. W. Cheung, M. Wang, Z. Wei, and G. Sun, "Application of functional deep belief network for estimating daily global solar radiation: A case study in china," *Energy*, vol. 191, p. 116502, 2020.
- [63] G. W. Chang and H.-J. Lu, "Integrating gray data preprocessor and deep belief network for day-ahead pv power output forecast," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 1, pp. 185–194, 2018.
- [64] Y. Zhang, C. Qin, A. K. Srivastava, C. Jin, and R. K. Sharma, "Data-driven day-ahead pv estimation using autoencoder-lstm and persistence model," *IEEE Transactions on Industry Applications*, vol. 56, no. 6, pp. 7185–7192, 2020.
- [65] A. Dairi, F. Harrou, Y. Sun, and S. Khadraoui, "Short-term forecasting of photovoltaic solar power production using variational auto-encoder driven deep learning approach," *Applied Sciences*, vol. 10, no. 23, p. 8400, 2020.
- [66] F. Wang, Z. Zhang, C. Liu, Y. Yu, S. Pang, N. Duić, M. Shafie-Khah, and J. P. Catalão, "Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting," *Energy conversion and management*, vol. 181, pp. 443–462, 2019.
- [67] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski et al., "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.
- [68] D. Zhang, X. Han, and C. Deng, "Review on the research and practice of deep learning and reinforcement learning in smart grids," *CSEE Journal of Power and Energy Systems*, vol. 4, no. 3, pp. 362–370, 2018.
- [69] T. T. Nguyen, N. D. Nguyen, and S. Nahavandi, "Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications," *IEEE transactions on cybernetics*, vol. 50, no. 9, pp. 3826–3839, 2020.
- [70] L. Xi, L. Zhou, Y. Xu, and X. Chen, "A multi-step unified reinforcement learning method for automatic generation control in multi-area interconnected power grid," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 2, pp. 1406–1415, 2020.
- [71] A. Nagabandi, G. Kahn, R. S. Fearing, and S. Levine, "Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 7559–7566.
- [72] D. Wu, X. Chen, X. Yang, H. Wang, Q. Tan, X. Zhang, J. Xu, and K. Gai, "Budget constrained bidding by model-free reinforcement learning in display advertising," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018, pp. 1443–1451.
- [73] H. Shuai and H. He, "Online scheduling of a residential microgrid via monte-carlo tree search and a learned model," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1073–1087, 2020.
- [74] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1539–1548, 2017.
- [75] M. Dong and L. Grumbach, "A hybrid distribution feeder long-term load forecasting method based on sequence prediction," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 470–482, 2019.

- [76] H. Chai, Z. Zhen, K. Li, F. Wang, P. Dehghanian, M. Shafie-khah, and J. P. Catalão, "Convolutional auto-encoder based sky image prediction model for minutely solar pv power forecasting," in *2020 IEEE Industry Applications Society Annual Meeting*. IEEE, 2020, pp. 1–7.
- [77] M. Khodayar, G. Liu, J. Wang, O. Kaynak, and M. E. Khodayar, "Spatiotemporal behind-the-meter load and pv power forecasting via deep graph dictionary learning," *IEEE transactions on neural networks and learning systems*, 2020.
- [78] H. Zang, L. Cheng, T. Ding, K. W. Cheung, Z. Wei, and G. Sun, "Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning," *International Journal of Electrical Power & Energy Systems*, vol. 118, p. 105790, 2020.
- [79] H. Hua, Y. Qin, C. Hao, and J. Cao, "Optimal energy management strategies for energy internet via deep reinforcement learning approach," *Applied energy*, vol. 239, pp. 598–609, 2019.
- [80] Y. Ji, J. Wang, J. Xu, X. Fang, and H. Zhang, "Real-time energy management of a microgrid using deep reinforcement learning," *Energies*, vol. 12, no. 12, p. 2291, 2019.
- [81] L. Yu, S. Qin, M. Zhang, C. Shen, T. Jiang, and X. Guan, "A review of deep reinforcement learning for smart building energy management," *IEEE Internet of Things Journal*, 2021.
- [82] C.-L. Liu, C.-C. Chang, and C.-J. Tseng, "Actor-critic deep reinforcement learning for solving job shop scheduling problems," *Ieee Access*, vol. 8, pp. 71 752–71 762, 2020.
- [83] D. Domínguez-Barbero, J. García-González, M. A. Sanz-Bobi, and E. F. Sánchez-Ubeda, "Optimising a microgrid system by deep reinforcement learning techniques," *Energies*, vol. 13, no. 11, p. 2830, 2020.
- [84] L. Xi, L. Zhou, L. Liu, D. Duan, Y. Xu, L. Yang, and S. Wang, "A deep reinforcement learning algorithm for the power order optimization allocation of age in interconnected power grids," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 3, pp. 712–723, 2020.
- [85] L. Xi, L. Yu, Y. Xu, S. Wang, and X. Chen, "A novel multi-agent ddqn-ad method-based distributed strategy for automatic generation control of integrated energy systems," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2417–2426, 2019.
- [86] L. Xi, J. Wu, Y. Xu, and H. Sun, "Automatic generation control based on multiple neural networks with actor-critic strategy," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 6, pp. 2483–2493, 2020.
- [87] K. S. Garud, S. Jayaraj, and M.-Y. Lee, "A review on modeling of solar photovoltaic systems using artificial neural networks, fuzzy logic, genetic algorithm and hybrid models," *International Journal of Energy Research*, vol. 45, no. 1, pp. 6–35, 2021.
- [88] M. Abdel-Basset, H. Hawash, R. K. Chakraborty, and M. Ryan, "Pv-net: An innovative deep learning approach for efficient forecasting of short-term photovoltaic energy production," *Journal of Cleaner Production*, vol. 303, p. 127037, 2021.
- [89] M. Tovar, M. Robles, and F. Rashid, "Pv power prediction, using cnn-lstm hybrid neural network model. case of study: Temixco-morelos, méxico," *Energies*, vol. 13, no. 24, p. 6512, 2020.
- [90] W. Xu, H. Peng, X. Zeng, F. Zhou, X. Tian, and X. Peng, "Deep belief network-based ar model for nonlinear time series forecasting," *Applied Soft Computing*, vol. 77, pp. 605–621, 2019.
- [91] W. Lee, K. Kim, J. Park, J. Kim, and Y. Kim, "Forecasting solar power using long-short term memory and convolutional neural networks," *IEEE Access*, vol. 6, pp. 73 068–73 080, 2018.
- [92] X. Luo, D. Zhang, and X. Zhu, "Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge," *Energy*, vol. 225, p. 120240, 2021.
- [93] A. Du Plessis, J. Strauss, and A. Rix, "Short-term solar power forecasting: Investigating the ability of deep learning models to capture low-level utility-scale photovoltaic system behaviour," *Applied Energy*, vol. 285, p. 116395, 2021.
- [94] J. Heo, K. Song, S. Han, and D.-E. Lee, "Multi-channel convolutional neural network for integration of meteorological and geographical features in solar power forecasting," *Applied Energy*, vol. 295, p. 117083, 2021.
- [95] A. Agga, A. Abbou, M. Labbadi, and Y. El Houm, "Short-term self consumption pv plant power production forecasts based on hybrid cnn-lstm, convlstm models," *Renewable Energy*, vol. 177, pp. 101–112, 2021.
- [96] J. Wojtkiewicz, M. Hosseini, R. Gottumukkala, and T. L. Chambers, "Hour-ahead solar irradiance forecasting using multivariate gated recurrent units," *Energies*, vol. 12, no. 21, p. 4055, 2019.
- [97] N. Son and M. Jung, "Analysis of meteorological factor multivariate models for medium-and long-term photovoltaic solar power forecasting using long short-term memory," *Applied Sciences*, vol. 11, no. 1, p. 316, 2021.
- [98] T. Xie, G. Zhang, H. Liu, F. Liu, and P. Du, "A hybrid forecasting method for solar output power based on variational mode decomposition, deep belief networks and auto-regressive moving average," *Applied Sciences*, vol. 8, no. 10, p. 1901, 2018.
- [99] S. Mirjalili, "Sca: a sine cosine algorithm for solving optimization problems," *Knowledge-based systems*, vol. 96, pp. 120–133, 2016.
- [100] A. H. Kashan, R. Tavakkoli-Moghaddam, and M. Gen, "Find-fix-finish-exploit-analyze (f3ea) meta-heuristic algorithm: An effective algorithm with new evolutionary operators for global optimization," *Computers & Industrial Engineering*, vol. 128, pp. 192–218, 2019.
- [101] A. Tabari and A. Ahmad, "A new optimization method: Electro-search algorithm," *Computers & Chemical Engineering*, vol. 103, pp. 1–11, 2017.
- [102] F. Fausto, E. Cuevas, A. Valdivia, and A. González, "A global optimization algorithm inspired in the behavior of selfish herds," *Biosystems*, vol. 160, pp. 39–55, 2017.
- [103] S. Harifi, M. Khalilian, J. Mohammadzadeh, and S. Ebrahimnejad, "Emperor penguins colony: a new metaheuristic algorithm for optimization," *Evolutionary Intelligence*, vol. 12, no. 2, pp. 211–226, 2019.
- [104] S. Arora and S. Singh, "Butterfly optimization algorithm: a novel approach for global optimization," *Soft Computing*, vol. 23, no. 3, pp. 715–734, 2019.
- [105] M. Eita and M. Fahmy, "Group counseling optimization," *Applied Soft Computing*, vol. 22, pp. 585–604, 2014.
- [106] R. Moghdani and K. Salimifard, "Volleyball premier league algorithm," *Applied Soft Computing*, vol. 64, pp. 161–185, 2018.
- [107] R. Rao, "Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems," *International Journal of Industrial Engineering Computations*, vol. 7, no. 1, pp. 19–34, 2016.
- [108] A. W. Mohamed, A. A. Hadi, and A. K. Mohamed, "Gaining-sharing knowledge based algorithm for solving optimization problems: a novel nature-inspired algorithm," *International Journal of Machine Learning and Cybernetics*, pp. 1–29, 2019.
- [109] P. Civicioglu, "Transforming geocentric cartesian coordinates to geodetic coordinates by using differential search algorithm," *Computers & Geosciences*, vol. 46, pp. 229–247, 2012.
- [110] —, "Backtracking search optimization algorithm for numerical optimization problems," *Applied Mathematics and computation*, vol. 219, no. 15, pp. 8121–8144, 2013.
- [111] H. Salimi, "Stochastic fractal search: a powerful metaheuristic algorithm," *Knowledge-Based Systems*, vol. 75, pp. 1–18, 2015.
- [112] T. Dhivyaprabha, P. Subashini, and M. Krishnaveni, "Synergistic fibroblast optimization: a novel nature-inspired computing algorithm," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 7, pp. 815–833, 2018.
- [113] M. Pan, C. Li, R. Gao, Y. Huang, H. You, T. Gu, and F. Qin, "Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization," *Journal of Cleaner Production*, vol. 277, p. 123948, 2020.
- [114] H. Aprillia, H.-T. Yang, and C.-M. Huang, "Short-term photovoltaic power forecasting using a convolutional neural network–salp swarm algorithm," *Energies*, vol. 13, no. 8, p. 1879, 2020.
- [115] H. Zhen, D. Niu, K. Wang, Y. Shi, Z. Ji, and X. Xu, "Photovoltaic power forecasting based on ga improved bi-lstm in microgrid without meteorological information," *Energy*, vol. 231, p. 120908, 2021.
- [116] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.
- [117] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 2031–2063, 2020.
- [118] X. Zhang, F. Fang, and J. Wang, "Probabilistic solar irradiation forecasting based on variational bayesian inference with secure federated learning," *IEEE Transactions on Industrial Informatics*, 2020.
- [119] Y. Tian, L. Sehovac, and K. Grolinger, "Similarity-based chained transfer learning for energy forecasting with big data," *IEEE Access*, vol. 7, pp. 139 895–139 908, 2019.

- [120] M. W. Akram, G. Li, Y. Jin, X. Chen, C. Zhu, and A. Ahmad, "Automatic detection of photovoltaic module defects in infrared images with isolated and develop-model transfer deep learning," *Solar Energy*, vol. 198, pp. 175–186, 2020.
- [121] B. Maschler and M. Weyrich, "Deep transfer learning for industrial automation," *IEEE Industrial Electronics Magazine*, 2021.
- [122] Q. Hu, R. Zhang, and Y. Zhou, "Transfer learning for short-term wind speed prediction with deep neural networks," *Renewable Energy*, vol. 85, pp. 83–95, 2016.
- [123] S. Zhou, L. Zhou, M. Mao, and X. Xi, "Transfer learning for photovoltaic power forecasting with long short-term memory neural network," in *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, 2020, pp. 125–132.
- [124] J. F. Torres, A. Troncoso, I. Koprinska, Z. Wang, and F. Martínez-Álvarez, "Big data solar power forecasting based on deep learning and multiple data sources," *Expert Systems*, vol. 36, no. 4, p. e12394, 2019.
- [125] K. Moharm, "State of the art in big data applications in microgrid: A review," *Advanced Engineering Informatics*, vol. 42, p. 100945, 2019.
- [126] S. E. Haupt and B. Kosović, "Variable generation power forecasting as a big data problem," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 2, pp. 725–732, 2016.
- [127] A. Zainab, A. Ghayeb, D. Syed, H. Abu-Rub, S. S. Refaat, and O. Bouhali, "Big data management in smart grids: Technologies and challenges," *IEEE Access*, vol. 9, pp. 73 046–73 059, 2021.
- [128] J. Á. G. Ordiano, A. Bartschat, N. Ludwig, E. Braun, S. Waczowicz, N. Renkamp, N. Peter, C. Düpmeier, R. Mikut, and V. Hagenmeyer, "Concept and benchmark results for big data energy forecasting based on apache spark," *Journal of Big Data*, vol. 5, no. 1, pp. 1–11, 2018.
- [129] A. Walch, R. Castello, N. Mohajeri, and J.-L. Scartezini, "Big data mining for the estimation of hourly rooftop photovoltaic potential and its uncertainty," *Applied Energy*, vol. 262, p. 114404, 2020.
- [130] S.-V. Oprea and A. Băra, "Ultra-short-term forecasting for photovoltaic power plants and real-time key performance indicators analysis with big data solutions. two case studies-pv agigea and pv giurgiu located in romania," *Computers in Industry*, vol. 120, p. 103230, 2020.
- [131] C. Tu, X. He, Z. Shuai, and F. Jiang, "Big data issues in smart grid—a review," *Renewable and Sustainable Energy Reviews*, vol. 79, pp. 1099–1107, 2017.
- [132] J. Thorey, C. Chaussin, and V. Mallet, "Ensemble forecast of photovoltaic power with online crps learning," *International Journal of Forecasting*, vol. 34, no. 4, pp. 762–773, 2018.
- [133] H. Lee, J.-G. Lee, N.-W. Kim, and B.-T. Lee, "Model-agnostic online forecasting for pv power output," *IET Renewable Power Generation*, 2021.
- [134] B. K. Puah, L. W. Chong, Y. W. Wong, K. Begam, N. Khan, M. A. Juman, and R. K. Rajkumar, "A regression unsupervised incremental learning algorithm for solar irradiance prediction," *Renewable Energy*, vol. 164, pp. 908–925, 2021.
- [135] H. Wang, Y. Liu, B. Zhou, C. Li, G. Cao, N. Voropai, and E. Barakhtenko, "Taxonomy research of artificial intelligence for deterministic solar power forecasting," *Energy Conversion and Management*, vol. 214, p. 112909, 2020.
- [136] S. Ogawa and H. Mori, "A gaussian-gaussian-restricted-boltzmann-machine-based deep neural network technique for photovoltaic system generation forecasting," *IFAC-PapersOnLine*, vol. 52, no. 4, pp. 87–92, 2019.
- [137] M. Khodayar, S. Mohammadi, M. E. Khodayar, J. Wang, and G. Liu, "Convolutional graph autoencoder: A generative deep neural network for probabilistic spatio-temporal solar irradiance forecasting," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 571–583, 2019.
- [138] C.-J. Huang and P.-H. Kuo, "Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting," *IEEE Access*, vol. 7, pp. 74 822–74 834, 2019.
- [139] H. Eom, Y. Son, and S. Choi, "Feature-selective ensemble learning-based long-term regional pv generation forecasting," *IEEE Access*, vol. 8, pp. 54 620–54 630, 2020.
- [140] Y. Ju, J. Li, and G. Sun, "Ultra-short-term photovoltaic power prediction based on self-attention mechanism and multi-task learning," *IEEE Access*, vol. 8, pp. 44 821–44 829, 2020.
- [141] H. Wen, Y. Du, X. Chen, E. Lim, H. Wen, L. Jiang, and W. Xiang, "Deep learning-based multi-step solar forecasting for pv ramp-rate control using sky images," *IEEE Transactions on Industrial Informatics*, 2020.
- [142] S. Han, Y.-h. Qiao, J. Yan, Y.-q. Liu, L. Li, and Z. Wang, "Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network," *Applied Energy*, vol. 239, pp. 181–191, 2019.
- [143] F. Wang, Z. Xuan, Z. Zhen, K. Li, T. Wang, and M. Shi, "A day-ahead pv power forecasting method based on lstm-rnn model and time correlation modification under partial daily pattern prediction framework," *Energy Conversion and Management*, vol. 212, p. 112766, 2020.
- [144] P. Kumari and D. Toshniwal, "Long short term memory-convolutional neural network based deep hybrid approach for solar irradiance forecasting," *Applied Energy*, vol. 295, p. 117061, 2021.
- [145] Q. Li, X. Zhang, T. Ma, C. Jiao, H. Wang, and W. Hu, "A multi-step ahead photovoltaic power prediction model based on similar day, enhanced colliding bodies optimization, variational mode decomposition, and deep extreme learning machine," *Energy*, vol. 224, p. 120094, 2021.
- [146] H. Panamtash, Q. Zhou, T. Hong, Z. Qu, and K. O. Davis, "A copula-based bayesian method for probabilistic solar power forecasting," *Solar Energy*, vol. 196, pp. 336–345, 2020.
- [147] K. Doubleday, V. V. S. Hernandez, and B.-M. Hodge, "Benchmark probabilistic solar forecasts: Characteristics and recommendations," *Solar Energy*, vol. 206, pp. 52–67, 2020.
- [148] M. Sun, C. Feng, and J. Zhang, "Probabilistic solar power forecasting based on weather scenario generation," *Applied Energy*, vol. 266, p. 114823, 2020.
- [149] J.-F. Toubreau, J. Bottieau, Y. Wang, and F. Vallee, "Interpretable probabilistic forecasting of imbalances in renewable-dominated electricity systems," *IEEE Transactions on Sustainable Energy*, 2021.
- [150] Q. Huang and S. Wei, "Improved quantile convolutional neural network with two-stage training for daily-ahead probabilistic forecasting of photovoltaic power," *Energy Conversion and Management*, vol. 220, p. 113085, 2020.
- [151] H. Wang, H. Yi, J. Peng, G. Wang, Y. Liu, H. Jiang, and W. Liu, "Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network," *Energy conversion and management*, vol. 153, pp. 409–422, 2017.
- [152] J.-F. Toubreau, J. Bottieau, F. Vallée, and Z. De Grève, "Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1203–1215, 2018.
- [153] M. Sun, T. Zhang, Y. Wang, G. Strbac, and C. Kang, "Using bayesian deep learning to capture uncertainty for residential net load forecasting," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 188–201, 2019.
- [154] S. Sharda, M. Singh, and K. Sharma, "Rsam: Robust self-attention based multi-horizon model for solar irradiance forecasting," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 2, pp. 1394–1405, 2020.
- [155] R. Chang, L. Bai, and C.-H. Hsu, "Solar power generation prediction based on deep learning," *Sustainable Energy Technologies and Assessments*, vol. 47, p. 101354, 2021.
- [156] C. Pan and J. Tan, "Probabilistic prediction of solar generation based on stacked autoencoder and lower upper bound estimation method," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.
- [157] A. Sagheer, H. Hamdoun, and H. Youness, "Deep lstm-based transfer learning approach for coherent forecasts in hierarchical time series," *Sensors*, vol. 21, no. 13, p. 4379, 2021.