IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 24, 2021, accepted March 31, 2021, date of publication April 5, 2021, date of current version April 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3071269

Deep Learning in Smart Grid Technology: A Review of Recent Advancements and Future Prospects

MOHAMED MASSAOUDI[1,2], (Student Member, IEEE), HAITHAM ABU-RUB[1], (Fellow, IEEE), SHADY S. REFAAT[1], (Senior Member, IEEE), INES CHIHI[3,4], AND FAKHREDDINE S. OUESLATI[2]

[1]Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha, Qatar
[2]Laboratoire matériaux molécules et applications (LMMA) à l'IPEST, Carthage University, Tunis 2036, Tunisia
[3]Laboratory of Energy Applications and Renewable Energy Efficiency (LAPER), El Manar University, Tunis 1068, Tunisia
[4]Department of Engineering (DOE), The Faculty of Science Technology and Medicine (FSTM), University of Luxembourg, L-1359 Luxembourg City, Luxembourg

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

This work was supported in part by the National Priorities Research Program (NPRP) through the Qatar National Research Fund (a member of Qatar Foundation) under Grant NPRP10-0101-170082, in part by IBERDROLA QSTP LLC, and in part by Qatar National Library.

ABSTRACT The current electric power system witnesses a significant transition into Smart Grids (SG) as a promising landscape for high grid reliability and efficient energy management. This ongoing transition undergoes rapid changes, requiring a plethora of advanced methodologies to process the big data generated by various units. In this context, SG stands tied very closely to Deep Learning (DL) as an emerging technology for creating a more decentralized and intelligent energy paradigm while integrating high intelligence in supervisory and operational decision-making. Motivated by the outstanding success of DL-based prediction methods, this article attempts to provide a thorough review from a broad perspective on the state-of-the-art advances of DL in SG systems. Firstly, a bibliometric analysis has been conducted to categorize this review's methodology. Further, we taxonomically delve into the mechanism behind some of the trending DL algorithms. We then showcase the DL enabling technologies in SG, such as federated learning, edge intelligence, and distributed computing. Finally, challenges and research frontiers are provided to serve as guidelines for future work in the futuristic power grid domain. This study's core objective is to foster the synergy between these two fields for decision-makers and researchers to accelerate DL's practical deployment for SG systems.

INDEX TERMS Smart grid, deep learning, deep neural networks, edge computing, distributed and federated learning, power systems.

NOMENCLATURE

Abbreviations

DDL	Distributed deep learning
DL	Deep learning
DRL	Deep reinforcement learning
DRN	Deep residual network
EI	Edge intelligence
EPS	Electric power systems
FL	Federated learning
IoT	Internet of things
LSTM	Long short-term memory neural network
NN	Neural network
PVPF	Photovoltaic power forecasting

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

I. INTRODUCTION

Increased concerns about the exponential growth of electricity demand and the bulk integration of Renewable Energy Sources (RES) brought a set of new challenges on the adaptability of the traditional grid for such reality [1]. Faced with the ever-growing population and energy demand, the emergence of Smart Grid (SG) as the most convenient solution provides the necessary tools to enhance the services of the antiquated grid using Information and Communication Technologies (ICT) [2]. This ICT-enabled grid presents the most effective solution to overcome the major issues in the

outdated grid, such as poor adaptation to outliers and heterogeneous sources [2]. The next-generation Electric Power Systems (EPSs) must satisfy the users' needs in terms of high flexibility to sudden events and customer behavior, Energy Management (EM) of distributed renewable energy sources, and adoption of low-cost and easy-to-deploy technical solutions [3]–[5].

A. CHALLENGES OF SMART GRID IMPLEMENTATION

Domestic electricity consumption has risen steadily in recent years due to population growth and rapid industrialization [6]. To meet the electricity demand, the renewable energy consumption in the United States is projected to climb from 11.34 quadrillion Btu to 21.51 quadrillion Btu in 2050 with a nearly 50% increase [6]. The distributed renewable energy market flourished as the fastest-growing sector with an estimation of surpassing traditional sources in 2050 [1]. For their smooth and mature penetration, the SG requires a robust and intelligent coordination platform between the different elements of EPSs. Traditional automation based on instructive tasks and traditional methods for tedious routine operations between the utility grid parts is ineffective in dealing with unexpected situations and sustainability problems [7]. The use of conventional approaches for electrical operations through deterministic programming makes the power flow issues remain difficult to control due to the heterogeneous multi-agent practitioners for the ''energy mix'' generation of the grid [2]. Furthermore, the conventional automation techniques require manual monitoring restoration and operation regulation leading to frequent problems and downtimes, especially with the incorporation of Renewable Energy Sources (RES) [8], [9]. The SG's underpinnings tend to automatically communicate with different electrical components and deduce the future behavior of each section using extensive calculations in which DL has the main share of their effective deployment [10].

B. EMERGENCE OF DEEP LEARNING

Several milestones have been reached in presenting Machine Learning (ML) techniques for various sub-areas in SG [11]. However, shallow neural networks and sample ML models pose many challenges that make them seldom employed for complex problems in EPSs [12], [13]. These challenges broadly lie in two facts: Firstly, the nondeep-learning algorithms are ineffective for high-dimensional representations with unreasonable complexities [14], [15]. Secondly, the accuracy of simple ML models can not be improved with large amounts of data [16], [17]. To tackle these problems, the learning paradigm shifts to DL as the most dazzling flagship of ML for exploiting Big Data (BD) abundance with hierarchical feature extraction, high efficiency, and timely manner [18]. Deep Neural Networks (DNNs) have quickly ascended to the spotlight due to the improvement of computing performance and data capacity. DL paradigm has achieved great success due to its strong potential to represent learning. The performance of DL techniques is solidified based on multiple processing units to learn feature representations with several layers of abstraction [19]. Due to its wide success, there is a significant proliferation in the use of DL for EPSs to exhibit complex correlations from heterogeneous data with different formats. Notably, the complexity of SG has intensified the need for DL to make use of the massive data from smart meters and Internet of Things (IoT) devices [20]. Therefore, the SG community has been encouraged to apply DL methods to solve a range of miscellaneous and critical problems. These problems broadly include forecasting tasks, fault detection and diagnosis, cybersecurity, and prediction [21].

C. RELATED WORKS AND MOTIVATION

Despite the rising interest in DL techniques in SG, the recent review articles are scattered across sub-areas of EPSs. Furthermore, the existing body of knowledge reported so far in availability published papers lacks a critical standpoint overview for the recent methodologies that perfectly tailor DL to EPSs such as distributed DL models and edge intelligence. Pioneering relevant review articles for DL & SG applications are reported in Table 1 [1], [10], [11], [13], [21], [23]. From Table 1, it can be concluded that the reported research strategy leads to losing sight of significance in tracing the development line of the energy field. This paper comes to provide a systematic review of DL methods applied to SG to foster the synergy between these two research hotspots. Beyond reviewing the recent DL methods, their merits, and limitations, this review will elicit escalating attention on the emerging DL enabling technologies for EPSs. These technologies include Federated Learning (FL), Distributed DL (DDL), Edge Intelligence (EI), Big Data DL (BDDL), Deep Transfer Learning (DTL), and Incremental Learning (IL). Finally, a fruitful discussion on the research frontiers that intersects advanced DL and EPSs is conducted.

D. RESEARCH METHODOLOGY AND SYSTEMATIC REVIEW PROTOCOL

Starting from September 2019, the multiple-methods approach was conducted [24]. The collection of the mainstream research papers on SG/AI from Web of Science (WoS), Scopus, IEEE Xplore, Science Direct, and Google scholar was conducted as the largest databases of peer-reviewed articles. Only peer-reviewed articles written in English, providing experimental results, and having a unique identifier from the mentioned databases were taken into consideration, including reviews, research articles, patent reports, and conference proceedings. The adopted methodology for conducting this review article employs a combination of keywords categorized into three main groups, specifically, *'Deep Learning'*, *'Smart Grid'*, and *'Prediction'*.

The search methodology focuses on the recent research articles from 2015-2020 to identify the comprehensive statues of the AI applications on SG. The filtering process results in 220 research papers from 600 related papers selected based on their relevance by reading the title, abstract, conclusion,

TABLE 1. List of the related review papers.

Reference	Year	Selected keywords	One-sentence summary	Coverage			
				SG	DL	EI	DDL
[23]	2018	DL; DRL; Electric power systems; Optimization	The emergence of RL for SG	✓	✓	-	-
[20]	2019	IoT; Edge computing; Fog Computing; DDL	The integration of DL in Industrial IoT	-	✓	✓	✓
[1]	2020	SG; DRL; IoT; 5G; Cyber-security	The implementation of ML in Smart cities	✓	✓	-	-
[13]	2020	AI; power electronics; predictive maintenance	AI applications for EPSs	✓	-	-	-
[22]	2020	RES; Forecasting; ML; ensemble models	Forecasting models for buildings	✓	-	-	-
[21]	2020	SG; Grid reliability; RES	The importance of SG transition	✓	-	-	-
[11]	2020	ML; Smart dispatch; DNN; Parallel systems	ML Applications of in smart dispatch EPSs	✓	-	-	-
[10]	2021	SG; AI; EM; distributed grid intelligence	Tailoring AI in EPSs and EM	✓	-	-	-
This paper	2021	SG; DL; DDL; EPSs; federated learning	DL applications for SG	✓	✓	✓	✓

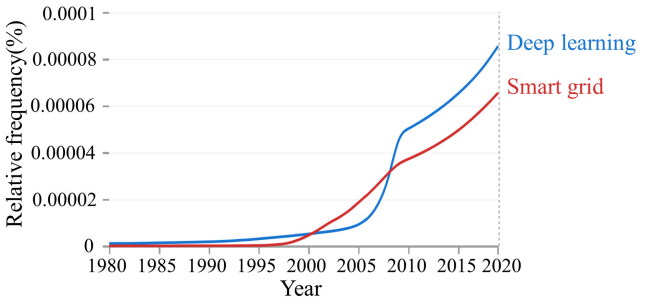

FIGURE 1. Frequency of use of terms deep learning and smart grid in books from 1980 to 2020 (Google Books Ngram Viewer, 2020).

and full text. The filtered articles are tabulated and unified to facilitate the comparative analysis and assessments according to the prediction horizon, applications, used data, error measures, AI classes, experimental setup, etc. The following criteria were applied: (i) SG and EPSs are considered. (ii) the feasibility analysis of the forecasting models is given high importance in the selection process. (iii) the evaluation of the forecasting models emphasizes the use of scale-free metrics. (iv) the future directions and perspectives taking into consideration the latest research articles to give a general standpoint of the current status of SG-based DL and future work. Fig. 1 presents a timescale variations on the frequency of use of terms SG and DL in scientific books from Google Books Ngram Viewer. It can be seen from Fig. 1 that the SG paradigm initially appeared in 1997, while DL achieves an exponential peak since 2006. This high correlation of the SG and DL lies in their complementarity to promote their applicability in real-world policies. A bibliometric analysis based on a thesaurus file from WoS website is conducted to shape the review structure, as illustrated in Fig.2. The aforementioned keywords shape the review structure. Fig.2 shows three topic clusters: DL models, SG, and enabling technologies (i.e., transfer learning). These clusters are taken into consideration to structure the content review in the next subsection.

E. STRUCTURE OF THE REVIEW

The information flow of this review is structured in a top-down manner, as illustrated in Fig.3. Section II emphasizes the ultimate need for SG in the energy hub. Section III comprehensively presents the commonly used DL methods for SG system. These DL methods essentially include Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Restricted Boltzmann Machines (RBM), Autoencoders (AE), Deep Reinforcement Learning (DRL), and Generative Adversarial Network (GAN). Section IV presents some insights on DL enabling technologies underpinning the SG paradigm. Section V introduces the possible future work and possible directions for empowering the role of DL in the SG area by emphasizing the undiscovered fields. Section VI concludes this review paper.

II. PRIMER ON SMART GRID

The SG technology presents a potentially powerful concept to sustainable energy operations. This next-generation network takes advantage of the customer action and energy stockholders to empower the energy delivery in a secure, economic and sustainable manner. A huge amount of data sources and control points in the grid cope with end-user needs toward efficient decision-making actions [23]. For realizing two-way communication, SG elements require efficient integrity to meet the desired functionalities of SG.

However, the widespread adoption of SG systems introduces several critical challenges. The complexity of SG and huge amount of data require advanced automation and information management tools. The bulk penetration of RES into the electrical grid leads to unstable and volatile power generation. This volatility requires DL-based models to address the uncertainty and intermittency of renewable EPSs [25]. Furthermore, the wide development of Advanced Metering Infrastructures (AMI) and Wide Area Monitoring Systems (WAMS) intensifies the necessity of DL-based techniques to deal with the massive data produced.

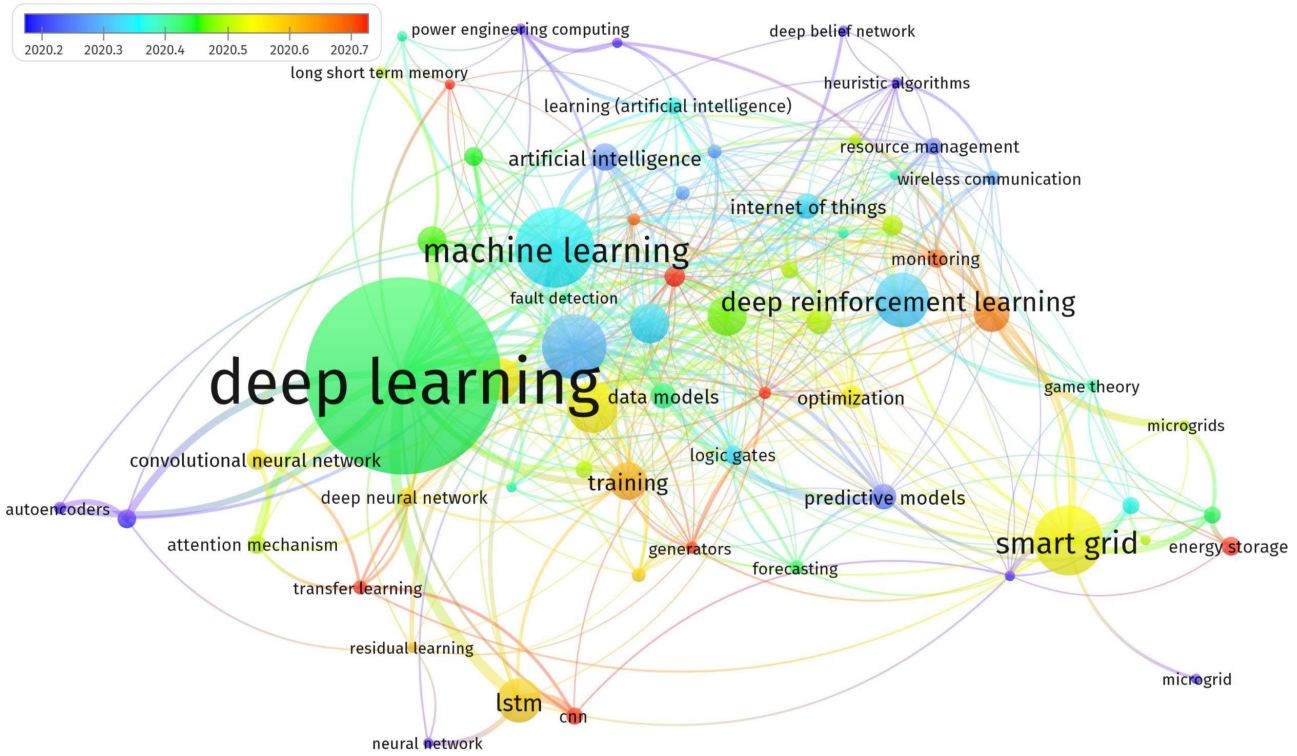

FIGURE 2. Bibliometric visualization for the author-supplied keywords, created with VOSviewer Software.

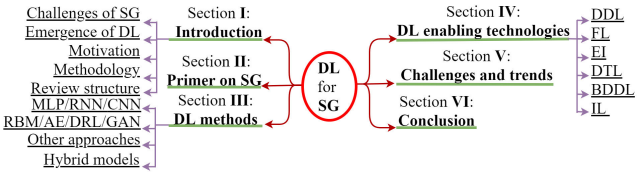

FIGURE 3. The information flow of the paper.

III. REVIEW OF DL METHODS

Recently, DL arouses a greater attraction at a far faster pace than even a decade ago. This appears quite obviously regarding the massive number of technologies where DL fingerprint is meaningfully marked [26]. Particularly, utility-scale plants have been remarkably expanding around the world during the last decade [27]. Due to the fast growth in global installed capacity, utility-scale generating facilities face multiple challenges in terms of performance monitoring, power losses, faults and failures, large complexity, and big size across hundreds of acres of land. Stakeholders and researchers tend to find scalable solutions in such huge plants to address these challenges [27]. Traditionally methods of monitoring utility-scale projects become too costly in the utility-scale. DL techniques contribute toward filling these gaps by self-monitoring and self-healing, automated diagnostics provider. Several studies report the increasing role of DL algorithms in revolutionizing utility-scale systems [28]. Fascinated by the enticing popularity of DL, the following subsection is devoted to tackling the popular DL methods to SG.

A. MLP

Multilayer Perceptron (MLP) is a DNN with densely connected layers to acquire the strong fitting ability for nonlinear systems [29], [30]. Over the past decades, neural networks have achieved a significant evolution from the simple McCulloch-Pitts Neuron to more complex MLP structures, as shown in Fig. 4 [31]. Regarding Fig. 4, Neural networks architectures have passed by radical transformations to enhance learning performance [32]. Operational Neural Networks (ONNs) and Self-ONNs present a diversification of the conventional MLP that introduces embedding nonlinear patch-wise transformations for designing more compact networks with improved prediction capability [33]. The objective of these futuristic variants of MLP networks is to boost generality potential with less network complexity and minimal training data [34]. For formalizing the dynamics of MLPs model of K hidden layers $\{h^1, h^K\}$,$\{f \in \mathbb{R}^{d_i}, i \in [1, K]\}$ is the nonlinear function described as $f(x, \theta) = \sum_{i=1}^{k} w_i \phi(x, J_i)$. J_i denotes the weight from the input layer to the hidden node i and w_i denote the weight from the hidden node i to the output layer. Here, $\phi(i)$ denotes the activation function, while $\phi(x, J_i) = \phi(J_i^T x)$ is the output of the ith hidden neuron. $\phi = \{J_1, \ldots, J_k, w_1, \ldots, w_k\}$ denotes all the model parameters.

Paper [37] proposed MLP based Double Least Absolute Shrinkage and Selection Operator (dLASSO-MLP) model for measuring quality variables of soft sensors in industrial processes. The dLASSO-MLP model retrieves the redundant hidden nodes to avoid overfitting. An Iterative Residual blocks Based DNN (IRBDNN) model has been introduced

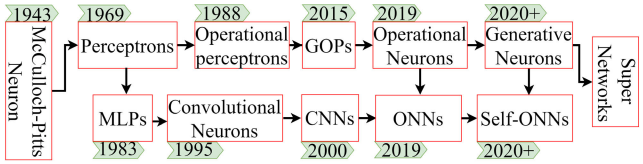

FIGURE 4. Timescale evolution of Artificial Neural Networks with ONNs: Operational Neuron Networks, GOPs: Generalized Operational Perceptrons, +: under improvement. Super Networks belongs to Super AI, which meets the technological singularity (These concepts are purely speculative for the time of writing this review and may not exist in the future) [35], [36].

to predict individual occupants' short-term strength requirements by exploring the underlying Spatio-temporal correlations of customers behaviors [38]. It has been concluded from the implementation of IRBDNN that the spatial correlations between different appliances used in a household and the iterative residual blocks can boost the model performance [38]. However, the data representation in communication networks can potentially increase the accuracy of the IRBDNN model [38]. The authors in [39] employed MLP for malicious attack detection of SG networks. The proposed solution provides a 99% accuracy over 10000 simulations. In [40], the authors have developed an MLP model for STLF based on smart meter data. According to the simulation results, MLP achieved better performance than conventional ML techniques [40]. However, the training time is relatively slow [40]. MLP reveals several features that promote its implementation for nonlinear problems of single and multi-complex tasks such as distributed representation and computation, mapping capabilities, powerful generalization, and high-speed information processing [41]. There are several advantages of MLP model, especially in higher-dimensional settings, still, the loophole lies in algorithm complexity, long-training time for large MLPs, and higher computational burden [41]. By stacking multiple layers, constructing and training this DNN model can be computationally expensive [42].

B. RNNs

RNN is designed for sequential time series data where the output of the network is fed back to the input as illustrated in Fig. 5(a). The recursive processing of RNN contains hidden layers to with feedback loop to provide a useful information about the past states [43]. For a sequence of input $x_t = (x_1, \ldots, x_T) \in \mathbb{R}^{N\times T}$ and output vectors $y_t = (y_1, \ldots, y_T) \in \mathbb{R}^{M\times T}$, the hidden states are $h_t = (h_1, \ldots, h_T) \in \mathbb{R}^{H\times T}$. Let's consider h_t and h_{t-1} the hidden state at time t and $t-1$, respectively. Here, h_t can be written as: $h_t = f_h(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$, where f_h denotes the nonlinear function such as the sigmoid function σ. W_{xh},W_{hh} and W_{hy} represent the input, hidden, and output weights, respectively. b_h, and b_y denote the hidden and output bias. The network output y_t can be written as $y_t = W_{hy}h_t + b_y$. Fig.5.(a) shows the architecture of the proposed model.

In [44], an intrusion detection system-based RNN has been introduced for detecting network attacks and fraudulent

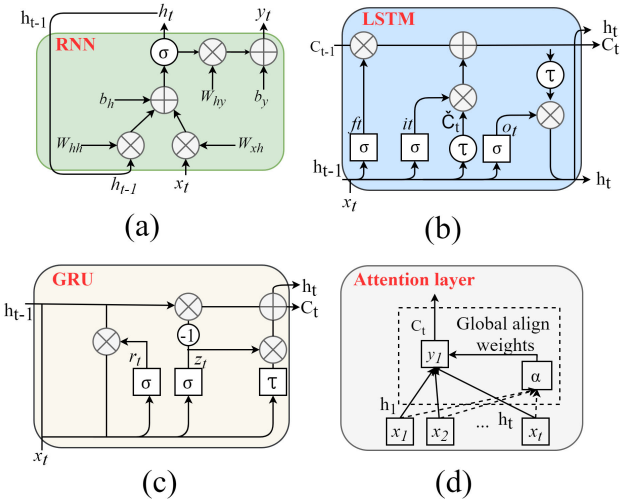

FIGURE 5. Structure of (a) RNN (b) LSTM (c) GRU, and (d) AL.

transactions in the blockchain-based energy network. The SG attacks are avoided by generating blocks with short signatures and hash function [44]. The RNN model has achieved an overall accuracy rate of 98.23%. The merit of RNN lies in its internal memory. However, RNNs are prone to the vanishing and exploding gradient dilemma [45]. For real EPSs applications, this shortage makes RNN usually replaced with two types of memory gated structures: Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) [46].

1) LSTM

The major success of LSTM model is accorded to its excellent ability in temporal feature extraction from input data $x_1, \ldots, x_T$ with $0 \leqslant t \leqslant T$. The LSTM mechanism consists of the deployment of three memory gates, namely, input gate $i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$, forget gates $f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$, and output gate $o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$ as shown in Fig. 5(b) [45], [47]. Where x_t, σ, and c_t denote the input sample at time t, the activation function, and the memory unit, respectively. (b_f,b_i,b_o) and (W_f,W_i,W_o) stands for the bias and weight matrix for each gate, respectively. The symbol $\odot$ is the corresponding multiplication of the elements [45]. A comparative study of some popular nonlinear activation functions is presented in Table 5.

The authors in [49] employed LSTM and an aggregation function based on Choquet integral for Solar irradiation forecasting. The proposed architecture provides a clear information about the largest consistency among the conflicting forecasting results by aggregating different LSTM networks [49]. The simulation results based on six datasets collected from different regions in Finland demonstrated the performance superiority of the proposed approach compared to four standalone LSTM models with different configurations [49]. A STLF method based-LSTM and Multivariable Linear Regression (MLR), named LSTM-MLR, is proposed to capture the time series variations of STLF using ensemble empirical-mode decomposition [50]. However, the

TABLE 2. Nonlinear activation functions.

Function	Advantages	Drawbacks
ReLU [48]	-Suitable for hidden layers -Less time and complexity	-Dying neuron problem
Sigmoid [48]	-Supports backpropagation -Suitable for output neurons	-Vanishing gradient -Computationally costly
Softmax [46]	-Useful for multiclass classification -Suitable for output neurons	-Can not be used for hidden layers
ELU [46]	-Fast -Suitable for hidden layers	-Exploding gradient
Leaky Relu [48]	-Prevents dying neuron -Computationally proficient	-Exploding gradient
Tanh [48]	-Supports backpropagation -Has a smooth slope	-Vanishing gradient -Slow convergence
Softplus [46]	-High sparsity	-Incomplete saturation
Swish [48]	-Self-gated function -Fast	-Less popular

computational complexity can limit the proposed LSTM-MLR from the practical adoption in real power systems. In [51], a parallel LSTM-CNN model has been proposed for STLF. Reference [52] introduced a hierarchical dilated LSTM model for mid-term electric load forecasting. Multilayer LSTM is equipped with dilated recurrent skip connections and a spatial shortcut path from lower layers for better forecastability and universality potential [52]. This model was based on the winning submission to the M4 forecasting competition for monthly data in 2018 [52]. In [53], a rapid islanding detection method-based LSTM classifier has been proposed, demonstrating the applicability of LSTM for classification problems. From these studies, it can be concluded that LSTM model dramatically improved the State-Of-The-Art (SOTA) of RNN in time series prediction. This improvement is due to its strong ability to learn long-tailed temporal dependencies using memory units and customized gates [45]. However, the LSTM structure has limited potential in learning spatial features.

2) GRU

GRU network has been proposed to alleviate the computational burden from the LSTM architecture by using only two gates instead [45]. These gates are a reset gate and an update gate (Fig. 5(c)) [45]. The update gate has the same functionality as the forget gate of an LSTM, which decides if the information is useful or forgotten. The reset gate is used to decide if the information would be saved or removed. Paper [54] used a Residual GRU (ResGRU) with CNN for fault diagnosis in PV arrays. The proposed ResGRU model has a strong anti-interference ability to effectively identify single and hybrid faults with an accuracy of 98.61%. Unfortunately, the ResGRU model's computational complexity poses the problem of massive waveform storage, especially with online fault diagnoses [54]. In [55], Particle Swarm Optimization (PSO) method is used to tune GRU model, which shows that GRU is very sensible to the initial state of parameters. A Fault type classification method based-GRU has been presented, which makes use of discrete wavelet transform for data pre-processing [56]. Despite the high flexibility, GRU architecture is prone to poor spatial feature representation and high computational efficiency, making its trustworthy practical implementation questionable. [31].

3) ATTENTION MECHANISM

Inspired by the selective visual mode of human beings, the attention-based RNN locates discriminative regions of interest. The Attention Mechanism (AM) operates by attributing different weights for different features at different time steps as shown in Fig.5(d). The Attention Layer (AL) focuses on a discrete aspect of the information or lessens the interference of irrelevant or noisy information brought by the global features. The AM can be classified into kinds of attention: deterministic Soft Attention (SA) and stochastic Hard Attention (HA) [57]. SA is a fully differentiable deterministic mechanism that can be plugged into an existing system. Alternatively, instead of using all the hidden states in HA as an input for the decoding, the system samples a hidden state y_i with the probabilities s_i. Compared to HA, SA is easier to implement [57]. Two variant attention mechanisms are introduced: Self-Attention mechanism (SAM) and Multi-Head Attention Mechanism (MHM) [58]. The self-attention is employed to capture spatial dimensions in the nonlocal feature dependence, which is hard to be extracted by convolution kernels. Meanwhile, MHM employs multiple self-attention mechanisms. Extensive studies found that MHM is more efficient than SAM for selective attention.

Attention-based models have been frequently used in EPSs. For instance, paper [58] employed SAM and Multi-Task Learning for Photovoltaic Power Forecasting (PVPF). Despite the fundamental role of SAM in enhancing the RNN-based models' performance, a massive number of additional learn-able hyperparameters have been added into the prediction system, which requires extensive tuning, making it unsatisfactory for real-world scenarios [58]. In [59], the authors used a multi-head CNN-RNN architecture for anomaly detection. From the simulation results, the type of anomalies in the learning library for fault identifications are diagnosed, specifically, point, context-specific, and collective [59]. In real applications, the multi-head CNN-RNN model could face other unknown types of fault samples, and at this time, the proposed model may fail [59]. Paper [60] introduced a gated dual attention unit (GDAU) for predicting bearing Remaining Useful Life (RUL). The GDAU model combines AM and GRU to improve the prediction accuracy and convergence speed [60].

4) BIDIRECTIONAL MECHANISM

Bidirectional mechanism (BiM) is proposed for uni-directional-based schemes to capture both past and future semantic information from the forward and backward directions simultaneously [61]. The bidirectional modeling uses two separately computed directions of information processing: one forward neural network scanning the sequence from left to right and one backward network reading the sequence in the other direction [61]. Assuming the input of time t is the embedding layer w_t,at time $t-1$, the output of the forward hidden unit is $\overleftarrow{\mathrm{h}}_{t-1}$,and the output of the backward hidden unit is $\overrightarrow{\mathrm{h}}_{t+1}$, Then the output of the backward and of the hidden unit at time t is equal as follow: $\overrightarrow{\mathrm{h}}_t = L(w_t, \overrightarrow{\mathrm{h}}_{t-1}, c_{t?1}$, $\overleftarrow{\mathrm{h}}_t = L(w_t, \overleftarrow{\mathrm{h}}_{t+1}, c_{t+1})$ where $L()$ denotes the hidden layer operation of the LSTM hidden layer [61]. The forward output vector is $\overrightarrow{\mathrm{h}}_t \in \mathbb{R}^{1\times H}$ and the backward output vector is $\overleftarrow{\mathrm{h}}_t \in \mathbb{R}^{1\times H}$, and they should be combined to obtain the text feature. It should be clarified that H denotes the number of hidden layer cells:$H_t = \overrightarrow{\mathrm{h}}_t \| \overleftarrow{\mathrm{h}}_t$. Owing to the advantages of BiM, it has been used for forecasting applications. The authors in [62] employed Bidirectional LSTM (BiLSTM) for wind speed interval prediction. A sequence-to-sequence mechanism based on the bidirectional GRU for type recognition and time location of combined power quality disturbance has been introduced [63]. Unfortunately, the details about computational time and the network size were missing, limiting the model generalization. A feature-attention mechanism-based BiGRU method for RUL prediction for electrical assets has been proposed [64]. Despite achieving STOA performance, the high complexity of the proposed model, the hyperparameter optimization may lack convergence guarantees [64].

C. CNNs

The CNN layers were initially used to capture the semantic correlations of underlying spatial features between slice-wise representations by convolution operations in multiple-dimensional data [65]. The feature mapping of CNN contains k filters spatially repartitioned into different channels [66]. The pooling operation is applied to shrink the width and height of the feature map. The convolutional layer y_i is calculated as $y_i = f_a(\sum_i k_{ij}x_i)$ [61]. where x_i, k_i,and y_i denote the feature input, the convolutional kernel, and the hidden layer of the ith iteration, respectively while f_a represents the activation function. The pooling layer equation for $j \times k$ metric dimension is given as $y_{ijk} = max(x_{i,j+p,k+q})$. where p and q denote the vertical and horizontal index in local neighborhood. The fully connected layer υ is generated as follows $\upsilon = f_a(w^{\upsilon}y + b)$ where b and w denote the bias, weight matrix [67].

CNNs analyze the hidden patterns using pooling layers for scaling, shared weights for memory reduction, and filters for capturing the semantic correlations by convolution operations in multiple-dimensional data [68]. Thus, CNN architecture acquires a strong potential in understanding spatial features [69], [70]. Despite the CNN potential, CNN model suffers from its disability in capturing special features [70]. The authors in [71] implemented GRU for Long-term load forecasting. In [72], a Temporal Convolutional Network (TCN) has been associated with Light Gradient Boosting Machine (LGBM) to address the issue of unsatisfying accuracy in STLF. As an enhanced variant of CNN, TCN model associates a sequence of dilated causal convolutions and residual connections for better effectiveness for stacking deep layers [73]. A CNN with Squeeze-and-Excitation modules (CNN-SE) has been proposed for STLF [74]. In the CNN-SE architecture, SE mitigates the redundancy caused by the massive number of input channels while the CNN model aggregates micrometeorological data from different acquisition sites [74]. The proposed model provides a SOTA multi-dimensional analysis using the squeeze-and-excitation block [74]. In [75], a CNN-based locational detection algorithm has been proposed for multi-label classification of false data injection attack. For better accuracy, a Bad Data Detector (BDD) has been employed to refine data quality on IEEE-14 and 118 bus systems [75]. However, the training difficulty increases with the depth and number of CNN layers [56]. Therefore, the number of hidden layers of CNN is preferred to be small, which can decrease its performance [56], [75]. From the previous CNNs applications, the CNN model contributes to the prediction system with efficient spatial feature extraction and distributed implementation. However, the downsides of CNN include a lack of temporal data modeling and long training time [76].

D. RBMs

The Deep Belief Network (DBN) is a deep feedforward neural network with input h^0 and L computational layers h^i; $(i = 1, 2, \ldots, L)$. Each layer h^i is an RBM layer RBM^i. The RBM model comprises visible units $v \in \{0, 1\}^{g_v}$ and hidden units $h \in \{0, 1\}^{g_h}$ with g_v and g_h denote the number of visible and hidden units, respectively. The shallow Boltzmann machine (BM), shown in Fig.6(a), is found ineffective for its poor learning potential and high calculation complexities. Therefore, RBMs restrict BMs without linking the hidden-hidden and visible-visible connections among units on the same layer, as shown in Fig.6(b). The RBM is a two-layer training model in the construction of a DBN (Fig.6(c)).

The RBM layer employs a generative graphical model that encodes the probability density function of its input

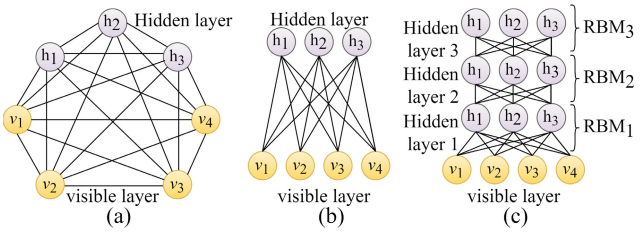

FIGURE 6. Structure of (a) BM (b) RBM (c) DBN, respectively.

layer h^{i-1} into its latent feature vector h^i. The RBM uses an energy function $E(v, h)$ with inputs modeled with the Gaussian function. The energy function equation can be written as follows:

$$E(v,h) = -\sum_{i=1}^{g_v}\sum_{j=1}^{g_h} W_{ij}\frac{v_i}{\sigma_i}h_j - \sum_{i=1}^{g_v}\frac{(v_i-a_i)^2}{2\sigma_i^2} - \sum_{i=1}^{gh} b_j h_j \quad (1)$$

with W_{ij} denote the weight between visible and hidden units, σ_i denote the Gaussian standard deviation of the visible units, and a_i and b_j denote the bias terms. For DBN with multiple RBMs, it can be built using the layer-wise greedy pretraining through the above procedure for each RBM.

In [77], the authors employed an improved DBN for STLF using meteorological data, load demand data, and demand-side management data. The improvement methodology consists of a virtue of processing units, specifically, Hankel matrix and gray relational analysis for correlation analysis, Gauss-Bernoulli RBMs for identifying the probability density of data, Bernoulli-Bernoulli RBMs for processing binary data, and mixed pre-training and GA for parameter optimization. The proposed model in [77] is found outperforming other benchmarks with a MAPE=6.07%. A DL model-based Factored Conditional RBM (FCRBM) has been firstly introduced for STLF [78]. The proposed FCRBM algorithm conducts dimensionality reduction and hyperparameter optimization using modified mutual information and genetic wind-driven algorithm [78]. A Conditional DBN (CDBN) has been proposed for detecting false data injection attacks in real-time [79]. According to the simulation experiments using IEEE 118-bus and 300-bus test systems, the CDBN provide a powerful capability of learning high-level representations of raw input data [79]. From the obtained results, a general conclusion can be drawn that giving higher importance to demand-side management data and the electricity price data enhances the prediction accuracy of the proposed model [79]. However, the proposed methodology requires four different kinds of data to operate. This condition is difficult to satisfy in practical industrial applications [79].

E. AUTOENCODERS

An Autoencoder (AE) is an auto-associative feedforward Neural Network (NN) [80]. It can learn effective representations from the raw input in an unsupervised manner [61]. The AE is essentially introduced for feature extraction or dimensionality reduction using three elements: encoder, coding layer, and decoder, as illustrated in Fig.7(a). Let's consider $x_{(k)} = (x_1, x_2, \ldots, x_N)^T, x \in \mathbb{R}^m$ and $y_{(k)} = (y_1, y_2, \ldots, y_N)^T, y \in \mathbb{R}^m$ the normalized unlabeled input vector and the ouptut vector, respectively. The AE maps the input vector $\hat{x} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_N)^T$ as: $y_{(k)} = \sigma(W_{(k)}x_{(k-1)} + b_{(k)})$ and $\hat{x}_{(k)} = \sigma(\hat{W}_{(k)}y_{(k-1)} + \hat{b}_{(k)})$. Here, $x_{(k)}$ and $x_{(k-1)}$ denote the inputs of the kth layer and $(k-1)$ layer, respectively. $W_{(k)}$ and $\hat{W}_{(k)}$ denote the weights while $b_{(k)}$ and $\hat{b}_{(k)}$ denote the bias. The classic AEs have several bottlenecks in terms of overfitting, poor generalization ability, and sensibility to noise inference. To suppress the adverse effects of auto-encoding architecture, several models have been proposed, including Sparse AE (SAE) (Fig.7(b)), Stacked AE (StAE), Variational AE (VAE), Stacked Denoising AE (SDAE), Stacked Contractive AE (SCAE), and Convolutional AE (ConvAE) [81]. Stacked AE consists of stacking multiple AEs layer by layer to better find the encoding-decoding scheme than AE by mapping the input vector towards a lower-dimensional manifold. The AE is trained by minimizing the squared reconstruction error $L(x, \hat{x}) = \frac{1}{2}\sum_x ||x - \hat{x}||^2$. However, Stacked AEs are prone of overfitting in case of a high risk copying the input to the output with any feature learning.

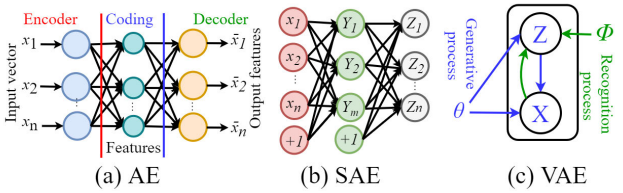

FIGURE 7. Inner structure of a)AE, b)SAE c)VAE.

To fill this gap, SDAE has been proposed to create a noise corruptions in feature extraction [82]. The DAE is trained in an unsupervised bottom-to-up manner to reconstruct the best clean input x from a corrupted version of input x'. The SDAE acquires a more flexible mapping than the classical AE. VAE is an NN-based generative probabilistic graphical model. The core idea of VAE consists of computing an efficient Latent Space Regularization (LSR) to enable generative process. Using a variational inference for latent representation learning, the probability density distribution of the LSR of a VAE typically matches that of the training data much closer than an original AE. Suppose that Z and $p_\theta(X\,Z)$ denote the latent variable and the variational approximation of the intractable posterior, respectively. The generator network $g(Z, \theta) = p_\theta(X\ Z)p(Z)$ approximates the generative process $p_\theta(X) = p_\theta(X\ Z)p_\theta(Z)$ as illustrated in Fig.7(c).

Paper [83] used stacked AE to extract the relevant features such as land use maps for spatial load forecasting. A data-driven bottom-up spatial and temporal STLF approach is conducted and demonstrated its generalization potential on larger areas, and diverse regions [83]. A Joint Latent VAR (JLVAR)-based monitoring method for gearbox failure detection of wind turbines has been proposed and proved its feasibility on Supervisory Control and Data Acquisition (SCADA) systems [84]. For electricity price forecasting, the authors proposed SDAE, RS-SDAE, a Random Sample Consensus (RANSAC) with Stochastic Neighbor Embedding (SNE), and SDAE for obtaining a robust representation of feature inputs [85]. Due to the hybridization approach, the additional hyperparameters dramatically decrease the model's learning efficiency [85]. In [86], an unsupervised DL scheme-based cyberattack detection system for transmission protective relays has been proposed using a 1-dimensional convolutional-based AE. A comparative study of AE and its variants is presented in Table 3.

TABLE 3. AE variants with their advantages and restrictions.

Model and Ref.	Advantages	Restrictions
SDAE [87]	-Improves the stability and robustness of AE -The repair ability minimizes the loss function	-Prone to the vanishing of errors -Problems with memorizing the training data
SAE [88]	-The sparsity penalty prevents overfitting. - Uses a limited number of hidden nodes	-The scale information is not considered -potential spectral distortion and noisy results
StAE [61]	-Efficient in handling high-dimensional data -Final encoding layer is compact and fast	-Higher chances of overfitting. -Poor feature representation from noisy data
SCAE [89]	-Robust encoding representation -Noise removal without knowing noise sensibility	-Difficult selection of activation functions -Low generalization for complex problems
ConvAE [90]	-Ability to extract the local features -Efficient noise reduction	-Problems with learning meaningful features -Regularization mechanism is necessary
VAE [91]	-Significant control of latent distribution -High continuity in the latent space	-Computationally expensive -Low expressiveness in capturing the true posterior distribution

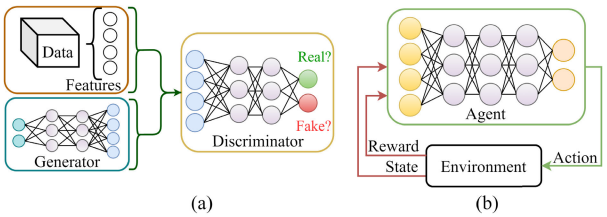

FIGURE 8. DL models with (a)GAN structure (b)DRL structure.

F. DEEP REINFORCED MODELS AND DEEP UNSUPERVISED LEARNING

1) DEEP GENERATIVE MODELS (DGM)

GAN has gained tremendous interest as a mainstream super-resolution model. GAN uses a Generator Network (GN) and Discriminator Network (DN) to create synthetic data that follows the same distribution from the original data set as shown in Fig. 8(a) [92].

More specifically, the GN mimics the data distribution using noise vectors to confuse the discriminator in differentiating between the fake image and the real image. The GN is trained to boost the probability of fooling the DN by making the new data indistinguishable from the real one. The discriminator's role is to be trained to distinguish generated fake image created by the generator from the original image following the two-player zero-sum game [93]. For discrete data, GAN is frequently employed for Data Augmentation (DA). For example, paper [50] proposed a Signals Augmented Self-Taught Learning (SASLN) network for Fault Diagnosis of Wind turbine Generation (FDWG). Here, GAN is used for signal augmentation to be fed to SLN model [50]. A fusion between CNN and GAN resulting in Wasserstein GAN with Gradient Penalty (WGANGP) model has been proposed to increase the size of data for PVPF [92]. Thirty-three meteorological weather types are reclassified into ten weather types [92]. The WGANGP was used to synthesize new, realistic training data samples by simulating input samples for improved training data sets. Despite the good performance generalization, the CNN-WGANGP may misclassify some confusing samples into a particular weather type, leading to lower prediction accuracy [92].

In [94], a Conditional Wasserstein GAN with gradient penalty (CWGAN-GP) has been introduced to model the uncertainties and the variations of the load. However, the CWGAN-GP lacks unexplained behavior of the network and specific rule for the determination of inner mechanism [94]. A GAN-based super-resolution reconstruction method for low-frequency electrical measurement data has been proposed [95]. The proposed architecture employs a generator based on DRN and a discriminator based on CNN to provide high reconstruction accuracy while sacrificing the computational complexity [95]. Reference [96] employs Wasserstein GAN with gradient penalty to capture the real distribution of the electricity consumption data for electricity theft detection. Reference [97] employs GAN for power loss mitigation of active distribution networks. Despite the large popularity of GANs, the GAN training and assessment is quite challenging and unstable. Moreover, the applicability of GAN is limited to non-discrete data such as computer vision and image recognition applications.

2) DEEP REINFORCEMENT LEARNING

DRL is the combination of DNNs and RL [98]. The basic idea behind DRL consists of assigning rewards and punishments for an agent to shape its policies. Here, the goal of DRL is to maximize the computing reward functions with reasonable actions (a_i), as illustrated in Fig. 8(b). The DRL mechanism consists of generating an autonomous agent that can navigate the search space and provide an optimal policy of actions. The DNN represents a large number of states (s_i) and approximating the action values to learn the best action choices over a set of states through the interaction with the environment [99]. DRL can be classified into value-based models, such as Deep Q-learning (DQL), Double DQL, Duel DQL, and policy-gradient-based models such as Deep Deterministic Policy Gradient (DDPG) and Asynchronous Advantage Actor Critic (A3C) [100], [101].

The authors in [102] proposed a Multi-Agent DRL (MADRL) model for EV charging stations with Energy Storage Systems (ESSs) and photovoltaic (PV) systems. The cooperative MADRL model-based on CommNet provides active and intelligent energy management while handling

real-time dynamic data [102]. Paper [103] introduced a Cyber-Attack Recovery Strategy for SG-based DRL. The proposed DRL has been employed for re-closing the tripped transmission lines at the optimal re-closing time [103]. Unfortunately, the proposed DRL may suffer from parameter instability during training [103]. In [104], the authors proposed a Multi-Agent DRL-based Volt-Var Optimization (VVO) framework for unbalanced distribution power systems. The actions are attributed to different agents to mitigate the action dimension of each agent [104]. The simulation results proved that DQL can be used with continuous actions [104]. A deep Q-network (DQN) based approach has been used for accurate STLF [105]. In [106], a cyberphysical vulnerability assessment approach based DQN has been introduced, which requires sufficient power system data acquired to correctly identify the contingencies. From the previous studies, DRL models have some bottlenecks, such as the lack of compatibility with continuous action spaces and slow policy convergence.

G. OTHER APPROACHES

Beyond the cited DL methods, Capsule Networks (CN) and Deep Spiking neural network (SNN) are employed for SG applications. CN are an advanced form of CNN comprising a group of vector neurons, primary capsules layer, convolution layer, and digit capsules layer [107]. Capsules in CN identify spatial patterns between the lower-level entities and applies a dynamic routing algorithm to recognize these relationships. From the authors' work in [108], a weight-shared capsules network is employed to further supplement the generalization capability of the original CN. Nevertheless, the proposed model is dependent on the quality of labeled data to perform fault diagnosis of machinery. This data is difficult to acquire, especially when the faults happen rarely.

SNN was introduced as one of the third generation ANN [109]. SNNs are considered as efficient models for temporal coding in neurons where the neurons interact with other nodes through excitatory or inhibitory spikes. Furthermore, SNNs can support huge parallel processing using neurons cluster, which significantly accelerates the execution time [110]. The proposed forecasting system in [110] associates SNN with a group search optimizer for automatic hyperparameter tuning to perform probabilistic forecasting with excellent performance and a short training time of 1.31 seconds.

H. HYBRID MODELS-BASED

DL algorithms have their strengths and weakness in terms of hyperparameter settings, data exploration of the computational burden [9]. Table 4 reports the advantages and downsides of DL models. According to Table 4, it can be remarked that the reported downsides of DL methods impede them from becoming canonical approaches in the power industry. Each DL method has the characteristics that make it tailored to a specific application of the SG area better than the other methods. To overcome the DL shortcomings,

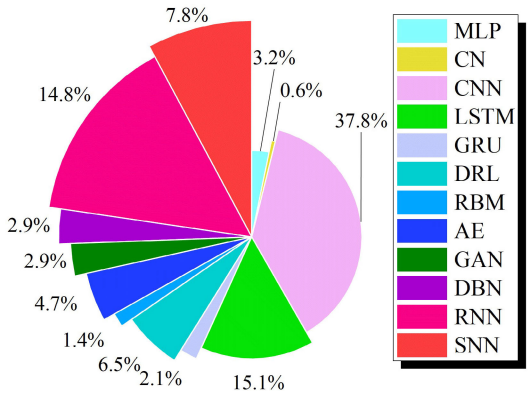

FIGURE 9. Pictorial representation of the DL Models applied to SG in 2015-2021 (WoS, 2021).

hybrid models have been extensively proposed based on single DL models to tackle EPSs bottlenecks, as shown in Fig. 10. For the sake of following the share of DL methods usage for SG systems limited to 01/01/2015-21/01/2021, several popular DL models have been coupled with "Smart grid" and searched on Google scholar research engine. As a result, a pictorial representation of the most popular DL techniques correlated to SG with their frequency of use is illustrated in Fig. 9.

We can see that CNN and RNN models have a high applicability and universality potential rather than the newly developed models. Paper [111] proposed a hybrid method for wind energy forecasting. The proposed method combines DBN and Support Vector Regression (SVR). DBN is a supervised DL technique having l-layers and parabolic hidden cells [111]. The proposed model outstrips individual models' accuracy (SVR and DBN). However, the proposed method is time-consuming [111]. The authors in [112] combine Long Short-Term Memory (LSTM) and CNN for power demand forecasting. In [92], the authors used GAN and CNN for a day ahead PVPF. Wasserstein GAN with Gradient Penalty (WGANGP) is proposed to classify thirty-three meteorological weather types for generating synthetic training data. Authors in [113] employ CNN, and GRU models to learn spatio-temporal representations for probabilistic wind power forecasting. From the loss versus epoch variation, it worth saying that the convergence of the model was reached in the first ten iterations, which may lead to serious overfitting problems [113]. In [114], a cybersecurity diagnosis and localization method using hybridization of AE, RNN, LSTM, and DNN has been proposed. Despite the model's universality potential to cope with other networked industrial control systems, it has been found that this hybrid design can not detect unknown attack/fault types [114]. A novel Graph Neural Network (GNN) based framework, combining graph convolutional network (GCN) and LSTM model has been proposed for multi-task multi-task transient stability classification [115]. The hybrid design of GNN is found able to effectively analyze complex spatio-temporal patterns in the IEEE 39 Bus system and IEEE 300 Bus system [115]. The fly in the ointment is that the GNN-based models require

TABLE 4. Mainstream DL architectures with their advantages and limitations.

Model	Ref.	Advantages	Limitations
MLP	[45], [71], [117]	- Reveal the highly nonlinear relationships - Easy to implement	- Inefficient for high-dimensional input space - Slow learning and difficult parameter tuning
LSTM	[45], [71], [117]	- Able to capture abstract temporal features - Alleviate the vanishing gradient problems	- Poor spatial features representation - Difficult implementation
GRU	[45], [71], [117]	- Fast convergence with minimum parameters - Capture the temporal information of data	- Poor spatial features representation - Difficult parameter tuning
CNN	[70]	- Ability at capturing spatial correlations - High generalization potential	- High computational cost - Difficult parameter tuning
AE	[54]	- Do not need unlabeled data for training - Excellent performance for feature abstractions	- Extensive processing time and fine-tuning - Training can suffer from the vanishing of the errors
CN	[107], [118], [119]	- Overcome the deficiency of max-pooling - High robustness to affine transformation	- Large number of parameters - Low generalization ability
RBM	[120]	- Can learn a complex probability distribution - Powerful representation capability	- Low robustness for noisy data - Limited applications due to the binary units
GNN	[121], [122]	- Powerful in learning on graph data - End-to-end training fashion	- Poor performance with different class labels - Low accuracy with dissimilar features
DRL	[123]–[126]	- Unsupervised learning potential - High effectiveness with complex problems	- Overfitting problems - High complexity
DRN	[127]	- Robustness against the vanishing gradients - Powerful with spatio-temporal dependencies	- Problem of diminishing feature reuse - Long training time
DBN	[123]–[126]	- Unsupervised training - High effectiveness for feature extractions	- Disable to process multi-dimensional - Training can be very slow and inefficient

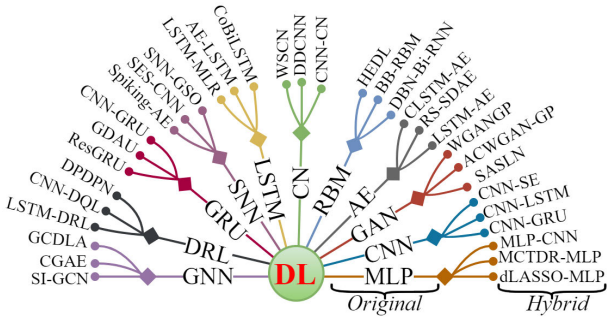

FIGURE 10. Hybrid deep learning models [116].

to compute not only the input data feature but also topological information, which can cause poor generalization potential [115].

IV. POTENTIAL DL ENABLING TECHNOLOGIES

The enabling technologies behind the deployment of DL are tackled in this section, including DDL, FL, EI, DTL, BDDL, and IL.

A. DISTRIBUTED DEEP LEARNING

DL achieved a quantum leap for making use of complex algorithms to achieve a SOTA performance. However, training large DL models may come with an incredible number of iterations, heavy hyper-parameter optimization, enormous computing time, and burden. With the existing non-distributed computing approach, the calculation of billions of parameters for DL models can be terribly slow and expansive. Here, DDL comes to fill this gap using High-Performance Computing (HPC) for training large networks [128]. HPC strategy employs parallel programming paradigms with multiple distributed data centers or edge devices to significantly alleviate the impractical computation burden and afford fast turnaround time for model training [129], [130]. Thus, Distributing the heavy-duty analytics and computations make DL more productive and exhibit the desired behavior whilst lessening the burden on the cloud. Multi-Central Processing Unit clusters and Tensor Processing Unit (TPU) provide the required tools for scaling up the training to achieve an effective computing performance.

Several DL libraries that allow multi-node computing can be used, including Horovod, distributed Tensorflow. Parallel Computing (PC) illustrated by Data Parallelism (DP) (Fig.11(a)) or Model Parallelism (MP) (Fig.11(b)), or Layer Pipelining (LP) (Fig.11(c)). DP divides the data into partitions according to nodes, while MP shares the calculations of different nodes of DL model by different computers. Compared to DP, MP is an ambiguous concept in terms the model repartition method not fixed on a clear basis. In [131], the authors employed a distributed deep AE for Phasor Measurement Unit (PMU) detection. Paper [132] employed a distributed DRL for load scheduling in residential SG. In [133], the authors propose a cloud-based DDL framework for phishing and Botnet attack detection and mitigation. The proposed solution is composed of security mechanisms working cooperatively: Distributed CNN (DCNN) method and a cloud-based temporal LSTM network model. In [134], Double DQL-Based Distributed operation has been proposed

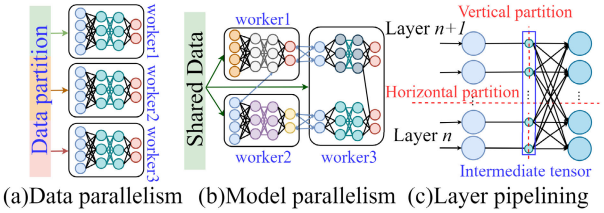

FIGURE 11. Distributed training of DNN with PC.

for managing the operation of a community battery energy storage system. A DDL-based IoT/Fog network attack detection system has been introduced in [135]. The proposed system proves that distributed attack detection using DDL can achieve better performance centralized algorithms in terms of accuracy using the sharing of parameters [135].

B. FEDERATED LEARNING

FL method employs distributed training process over end devices equipped with AI chips and siloed data centers on edge [136]. Thus, IoT devices compute model training with localized data and its storage capabilities instead of transferring data to central computing facilities. FL provides a suitable solution to lessen the communication costs, privacy concerns and the adverse effects of data centralization [137]. This method provides a privacy-preserving mechanism, which can be extremely beneficial, especially for highly distributed systems.

Smart cities sensing make use of three classes of FL: namely, horizontal, vertical, and transfer FL described in [136]. In [138], a probabilistic solar irradiation forecasting based on Bayes LSTM-NN has been proposed with a FL scheme shown in Fig.12. In [139], a FL mechanism-based DRL has been proposed for EM of multiple smart houses. The weakness of this study is that the proposed algorithm may suffer from overestimating the state-action values [139]. Paper [140] shed light on DeepFed, a federated DL scheme for Intrusion Detection. The proposed model proved that FL-based solutions can trade off between model performance and privacy concerns. Reference [141] proposed a GAN-based synthetic feeder generation mechanism to ingest power system distribution feeder models using a device-as-node representation. The disadvantage is that this model does not reduce the dimension of the data, so the computational burden of the proposed architecture is relatively high [141]. From these studies, the challenging points can be summarized in two major points: 1) The limited bandwidth of wireless communication of the current IoT devices. Thus, collaborative learning can be heavy. 2) The FL process relies on collaborative computing from IoT devices that must fully participate in the learning process. In case of a sudden interruption, before the learning process converges, the disconnected devices could significantly affect the learning quality.

C. EDGE INTELLIGENCE

DL workloads witnessed significant growth with the large availability of voluminous data and hardware and software

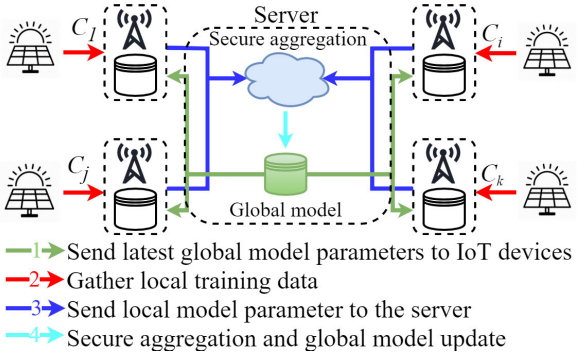

FIGURE 12. Schematic diagram of the FL scheme [138].

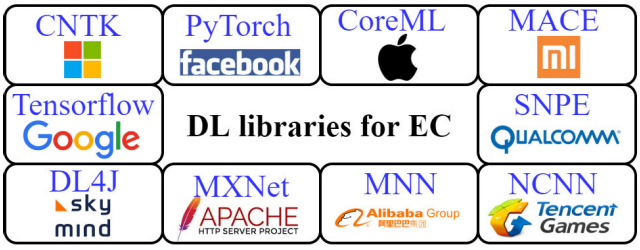

FIGURE 13. Popular DL libraries for edge computing.

improvement. Consequently, it is estimated that a single SG system produces 22 gigabytes of data per day for two million users [142]. Alternatively, on-board processing and Cloud Computing (CC) will be insufficient to store and possess the petabytes of data from multiple embedded power generator sensors and their communication with home sensors and appliances. To remedy these weaknesses, traditional cloud-based computing has been replaced by EC and Fog Computing (FC), leading to edge intelligence (EI) [143]–[145]. EI is the merge of EC and AI to push both data and intelligence to analytic platforms [146]. Big tech companies have put forth leading projects to demonstrate the advantages of Edge Computing (EC) in paving the last mile of DL [143]. EI coordinates a set of connected edge IoT devices that are used for data collection, caching, processing, and analysis based on AI [147]. DL has the required potential to automatically investigate the data from edge devices for quick real-time predictive decision-making [148]. The authors in [149] introduced an edge-cloud integrated solution using RL. The proposed solution tracks the demand response with high exactitude. Fig 13 presents the most popular DL libraries for EC [116]. Furthermore, Table 5 provides a comparative study of mostly used computing types for SG engineering research.

D. DEEP TRANSFER LEARNING

Despite the high potential of DL in representation learning, it can be pinpointed that the performance of DL dramatically degrades with small sample size. Furthermore, DL models have limited use as they require extensive measurements and recording burden to create a huge number of learning examples. Particularly, EPSs may not be in the same

TABLE 5. Advantages and restrictions of popular deep learning libraries for edge computing.

Library	Advantages	Disadvantages
CNTK [150]	-Supports data parallelism -Enhanced optimized algorithms	-Architectural Edge limitation -Android/iOS incompatibility
PyTorch [151]	-Less plug and play -Architectural Edge support	-Incompatible with DSP -Does not support Mobile GPU
Tensorflow [152]	-Rich DL library -Supports high/low level APIs	-Code complexity -Difficult in debugging
DL4J [153]	-Supports java coding -Provide inbuilt GPU support -Scalable on Hadoop	-Incompatible with iOS -Does not support DSP/FPGA
MXNet [154]	-High portability and scalability -Works with multiple languages	-Incompatible with Android/iOS -Training support limitation
CoreML [155]	-Low-level technology stack -Compatible with iOS/Android	-Does not support FPGA/GPU
SNPE [156]	-High flexibility -Efficient edge support	-Does not support FPGA/GPU -Training support limitation
NCNN [157]	-Very lightweight and easy to install -Impressive speed	-Does not support FPGA/GPU -Training support limitation
MNN [158]	-Fast computing -Handcrafted ARM assembly optimization	-Does not support FPGA/GPU -Training support limitation
Paddle-Mobile [159]	-Supports CUDA and cuDNN -Provide a training support	-Does not support DSP/GPU -Training support limitation
MACE [160]	-Android/iOS compatibility -Mobile GPU support	-Training support limitation
FANN [161]	-Easy to use and well documented -ARM architecture support	-Does not support FPGA/GPU -Android/iOS incompatibility

dimensional space and may not have the same distribution, which makes achieving the desired accuracy and turnaround time a challenging problem [162]. In other words, DL techniques are expected to provide salable and generalized solutions solutions across a diverse range of mainstream EPSs applications [163], [164]. In [15], the authors claimed that the generalization ability is a severe dilemma for the employment of DL techniques in wind and solar energy prediction. Reference [165] reports that one of the common issues of the prediction techniques resumes in the poor model universality, which can be referred to the limited data sources. For instance, authors in [166] reports that the lack of data has been known as one of the common failure reasons in

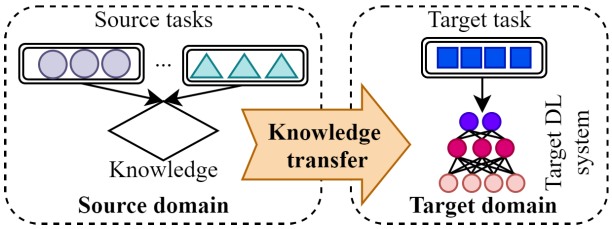

FIGURE 14. Principle of Deep Transfer learning.

load forecasting. To bridge this gap, Deep Transfer Learning (DTL) is well-positioned for alleviating expensive data-labeling efforts using cross-domain datasets [167], [168]. This learning framework operates by transferring the knowledge gained by a DNN model in handling a task (source problem) to solve another related task, as shown in Fig. 14.

A given a source domain D_s and a learning task T_s, a target domain D_t and a learning task T_t, DTL aims to achieve a better learning performance of the target predictive nonlinear function $r_t()$ in D_t using the knowledge in D_s and T_s, where $D_s \neq D_t$, or $T_s \neq T_t$. DTL methods can be classified into four classes based on source and target domains and tasks: (i)mapping transfer, (ii)instance transfer, (iii)network-transfer, and (iv)adversarial-transfer. In [169], the authors address the PVPF using DTL and LSTM network. From the simulation results, DTL can be unnecessary or inefficient when the data in the target domain are satisfactory [169]. In [170], a model transfer DL has been proposed for PV systems. The limitation of this approach is that the tasks require some reasonable similarity to successfully perform DTL [169]. Paper [50] employed a SLN as a particular form for DTL for FDWG. To work effectively, DTL requires some datasets with minimum reasonable similarity [50]. Paper [166] employed DTL and meta-learning using deep neural networks [166]. The proposed solution successfully realizes the model extension in the new target domain [166]. From the authors work, the selection of the optimal configuration for the prediction model is a complex problem that reduces the practical adoption of this solution in real-world problem with the high complexity of electricity-consumption data patterns [166].

E. BIG DATA DEEP LEARNING

Enormous volumes of information are quickly expanding in the PV systems with the persistent use of sensors, Distributed Computing (DC), and advanced information and communication systems. With the sheer size of data, DL methods have become dramatically more prominent for their unparalleled potential to efficiently learn discriminant features using BD technologies. BD computing is classified into two folds: (1) batch processing of massive information of on-disk data with no time constraints. (2) streaming processing of in-memory data in a real-time or short period of time [171]. Several computing frameworks have been proposed to compute BD, such as Hadoop, ComMapReduce, Dryad, Piccolo, and Spark; such systems have the capabilities to scale up DL.

Data mining can truly be beneficial for enhancing the PVPF [171]. Based on this reference, the weather information for the actual day is irrelevant for the prediction system. Paper [172] presents an accurate PVPF method using Apache Flume, Spark, Hadoop Distributed File System (HDFS), and Hive. In [173], the authors introduced Sun4Cast^{TM}, a BD-based renewable power forecasting solution. Paper [174] proposed a PVPF based Spark-based fuzzy partitioning LSTM model. The objective of this study consists of estimating the temporal variability of the production problem of Switzerland [174].

F. INCREMENTAL LEARNING

With the ever-changing environment, Incremental Learning (IL) helps in learning from data streams. IL aims to accommodate new patterns without compromising the loss of historical knowledge [175]. Typically, IL operates with new data without forgetting the knowledge which rising stability-plasticity trade-off. Plasticity operates when the model acquires new information constantly, whereas the updated model's stability stands for maintaining the acquired knowledge [176]. Paper [176] introduced an Incremental Deep Convolutional Computation Model for BD. For processing a large quantity of online data, The incremental knowledge can be saved into the idle neurons with a large probability [176]. In [175], an Adaptive Incremental DL Scheme has been proposed to tackle the hyperparameter online tuning by maintaining high cost-efficient processing. However, the proposed algorithm lacks parameter convergence guarantees [175].

V. CHALLENGES AND FUTURE RESEARCH DIRECTIONS

The development of DL for SG is currently having some opportunities as well as some breakthroughs. To our best knowledge, some concrete points can be mentioned as follows:

A. BETTER PERFORMANCE WITH SMALL DATA

Typically, DL models require large amounts of data for learning to be effective. For particular SG applications, such large training datasets are not publicly available, difficult to collect, costly, and possibly problematic due to privacy regulation [177]. Although data augmentation and big synthetic training datasets techniques can partially cover the lack of large labeled datasets, it remains cumbersome to fully satisfy the requirements for training by hundreds or thousands, if not fewer, high-quality data points. Supervised training of DL models with small datasets is prone to overfitting [178]. DTL and IL can provide additional information to enhance the data representation and learning process. Paper [179] proposed a VAE-based DGM model for overcoming the limited data set with bearing fault diagnosis. In a similar application, the authors in [180] proposed a Stacked SAE for gearbox fault diagnosis, which achieves a SOTA performance with small labeled data. Few-Shot Learning (FSL) has been proposed to tackle the data scarcity for intelligent fault diagnosis [181].

B. TAILORING QUANTUM DEEP LEARNING IN SMART GRID

Quantum Deep Learning (QDL) can provide a huge breakthroughs to power systems [182]. Despite the inherent potential of QDL, there are few pieces of evidence of its practical application in power systems. Paper [183] reveal that the electric power grid can significantly make use of Quantum computing to tackle the EPSs challenges. QDL employs Big data to speed up the training and make the computation process more efficient [184]. Despite substantial efforts in industry and academia, no error-corrected qubits have been built so far. Therefore, QDL can still far from actual EPSs applications.

C. DEMOCRATIZING DL THROUGH GOVERNMENTAL POLICIES

In the last few years, many countries such as USA and China prepared a strategic roadmap to speed up AI diffusion on a larger scale, especially for the energy market [185]. As a result of the extensive investments of this high technological expertise, it has been strongly remarked the significant increase in the share of scientific papers in the AI-based SG landscape. Despite the recent waves of AI democratization on an international level, the adaptation of such high technology is now at a low level of spread in the developed countries. This is due to the high financial cost of the redefinition of the traditional grid infrastructure and adding the necessary flexibility to be perfectly tailored for the SG paradigm.

D. PRIVACY AND SECURITY ISSUES IN DL

Cyber-physical SG systems have a massive data flow which makes the wonder how the security of the grid in the long term is ensured. In any SC operation, the network stores the new data via cloud storage. The size of the generated data from the grid operations is huge and increases each day in the order of thousands of zettabytes (10^{21}) [186]. From the dynamic grid operations, DL methods are prone to cyber attacks resulting in wrong predictions and control failures. The attacks that threaten the DL privacy fall into Model Extraction Attack (MEA) (duplicating model parameters) and Model Inversion Attack (MIA) (stealing sensitive information) [187]. DL's privacy and security issues against adversarial and poisoning attacks are vital bottlenecks scarcely studied and require a deep investigation.

E. LACK OF SKILLED MANPOWER

DL in its current form is relatively still new and complex technology. Thus, there is a severe shortage of competent Data Scientists (DS) with the required skill-sets in DL to turn data into actionable insights. In 2018, it was reported that there was a 50 percent shortage in DS supply jobs versus demand [188]. According to reference [189], it is assumed that 11.5 Million data science jobs will be created by 2026. From the World Economic Forum (WEF), 20% of the market jobs could have the fingerprint of data science creating

TABLE 6. Computing types with their advantages and shortages.

Type	Brief description	Advantages	Limitations
FC	FC offers access to computational resources as a service for real-time large-scale data analysis	-Enhanced security and privacy -Supporting the densely distributed data sources	-Mobility support issue -High latency in service delivery
CC	CC refers to client-server computing where resources are managed in a centralized fashion	-Highly accessible service -Networked backups prevent data loss from Hardware failures	-Costly bandwidth additions and low responsiveness -Low security and privacy
EC	EC enables data to be processed and transferred at the edge of a network where the data is collected	-Improved performance and high analytics capabilities -Less expensive route to scalability and versatility	-Saving bandwidth can lead to overlooking important information -Very costly with security problems
PC	PC is a type of computation where the calculations or processes are executed simultaneously to increase the speed and efficiency	-High performance and time saving -Improved speed of processing	-PC is hard to program and implement -Communication and synchronization in PC is challenging
DC	DC is a computation type in which networked computers communicate and coordinate a common program	-Shared resources and increased scalability -Reliability and high fault tolerance	-High computational resources cost -Difficult troubleshooting -Security issues

133 million jobs by 2022 [190]. Consequently, skilled workforce shortages are looming on the horizon and becoming a serious constraint, especially with the increased applicability of data science in many sectors such as the energy industry [191], [192]. A shortfall of skilled DS manpower has reflected the attractive annual salaries that can reach $135,776 for entry DS per annum in the United States [188]. According to McKinsey research, data science skills are something that the job market worldwide desperately needs [188]. Data scientists often need a combination of domain experience as well as in-depth knowledge of science, technology, and mathematics. There is no denying the fact that mastering each of these domains is somewhat elusive. Beyond promoting data science learning with university scholarships, free online courses, and incentives, job recruiters and top tech companies suggest jobs conversion from other fields towards data science to meet the increased demand [193]. Further, DL systems are intricate black boxes making decisions that are not easily interpretable from a human perspective. To cover this gap, several researchers proposed an explainable DL, i.e., an understandable internal DL architecture for human ease [194]. For example, in [195], an explainable DNN (xDNN) program has been introduced to facilitate the DL interpretability and make it easy to learn for practitioners and engineers [196]. Automatic DL is also considered as a promising solution to implement DL models without complexities [197]–[199].

F. COMMUNICATION INFRASTRUCTURES, PROTOCOLS, AND INVESTMENTS

With decision-makers increasingly seeing DL as a key enabler for thriving the futuristic SG, a fear of missing out this huge DL potential is globally spreading due to poor communication infrastructure [200]. In the last few years, numerous nations such as the USA and China prepared a strategic roadmap to speed up the DL diffusion on a larger scale through investment, protocols, advanced communication infrastructure, and risk management. The use of DL in government must take into account privacy and security, compatibility with legacy systems, and evolving workloads. The global data science market is expected to grow at a compound annual growth rate of 42.2% from 2020 to 2027 to reach USD 733.7 billion [201]. As DL's applicability to the next generation of technology grows, many decision-makers are worried that they will be left behind and not share in the gains. However, DL workloads have specific requirements from the underlying infrastructure, which can be resumed in three dimensions: scalability, portability, and timing [202]. Scalability reflects the ability to support a massive amount of data [202], [203]. Portability refers to the flexibility of the workload to be transportable across core, edge, and endpoint deployments [204]. Timing describes analyzing streaming databases in a real-time or near-real-time manner by involving advanced computing technologies such as DL accelerators [205], [206]. Despite the recent waves of DL democratization on an international level, the adaptation of the required technology is now at a low level of spread in several developed countries. This is due to the high financial cost of the redefinition of the traditional power grid and adding the necessary flexibility to the communication infrastructure to be perfectly tailored for DL techniques.

VI. CONCLUSION AND FUTURE DIRECTIONS

In the last decade, Deep Learning (DL) has become promising dawn for Smart Grids (SG). Inspired by the recent computational neuroscience discoveries, this review has comprehensively discussed the mainstream DL approaches in power system applications. Furthermore, several state-of-the-art paradigms are highlighted, including distributed DL, edge intelligence, and federated learning. DL's major applications in SG include energy forecasting, fault detection, cybersecurity awareness, prediction, and optimization to meet the

technical requirements for safe and secured power system operations.This paper makes the following contributions:

- The outlines of this paper have investigated the research landscape about DL approaches applied in the SG paradigm and analyze the extent of accuracy, feasible scenarios, and limitations.
- This review has pointed out DL's computational limits for power systems and how to lessen the computational power requirements by introducing DL enabling technologies such as distributed DL and federated learning.
- Finally, the emerging challenges and requirements and future directions of SG and DL models have been addressed.

Many of the reported research works are being conducted on SG applications, which still in the early stage of development. Some DL architectures, such as long short-term memory and deep convolutional neural network, have been heavily deployed to resolve dissimilar issues for a wide range of power engineering applications, including state of charge prediction, energy optimization, and power grid resiliency within the SG paradigm. Despite being recently introduced, generative adversarial network and deep reinforcement learning have been extensively included in multiple research works as efficient tools for modeling multifaceted problems that are considered vital to SG efficiency. Critical challenges and future research trends were depicted to draw fruitful discussions of the limitations of DL methodologies in the SG domain. In summary, DL enabling techniques highly require advanced communication and computation to speed up their practical adoption for SG systems rather than standing on the conceptual level. The future search directions have been investigated, which require interdisciplinary work to overcome the emerging bottlenecks for a flourishing SG market. This study's future work will give a particular focus on Explainable DL (XDL) algorithms for SG systems to shed light on the ample opportunities of XDL for transparent and understandable learning paradigms.

REFERENCES

[1] Z. Ullah, F. Al-Turjman, L. Mostarda, and R. Gagliardi, "Applications of artificial intelligence and machine learning in smart cities," *Comput. Commun.*, vol. 154, pp. 313–323, Mar. 2020.

[2] R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S. D. Langhans, M. Tegmark, and F. F. Nerini, "The role of artificial intelligence in achieving the sustainable development goals," *Nature Commun.*, vol. 11, no. 1, 2020, Art. no. 233.

[3] M. Albano, L. L. Ferreira, and L. M. Pinho, "Convergence of smart grid ICT architectures for the last mile," *IEEE Trans. Ind. Informat.*, vol. 11, no. 1, pp. 187–197, Feb. 2015.

[4] S. S. Refaat and H. Abu-Rub, "Implementation of smart residential energy management system for smart grid," in *Prov. IEEE Energy Convers. Congr. Expo. (ECCE)*, Sep. 2015, pp. 3436–3441.

[5] S. C. Müller, H. Georg, J. J. Nutaro, E. Widl, Y. Deng, P. Palensky, M. U. Awais, M. Chenine, A. Davoudi, and A. Mehrizi-Sani, "Interfacing power system and ICT simulators: Challenges, state-of-the-art, and case studies," *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 14–24, Jan. 2018.

[6] U. Energy and I. Administration. (2021). *Annual Energy Outlook 2021(AEO2021)*. Accessed: Feb. 25, 2021. [Online]. Available: https://www.eia.gov

[7] A. K. Ozcanli, F. Yaprakdal, and M. Baysal, "Deep learning methods and applications for electrical power systems: A comprehensive review," *Int. J. Energy Res.*, vol. 44, no. 9, pp. 7136–7157, Jul. 2020.

[8] O. Ellabban, H. Abu-Rub, and F. Blaabjerg, "Renewable energy resources: Current status, future prospects and their enabling technology," *Renew. Sustain. Energy Rev.*, vol. 39, pp. 748–764, Nov. 2014.

[9] A. A. Mamun, M. Sohel, N. Mohammad, M. S. H. Sunny, D. R. Dipta, and E. Hossain, "A comprehensive review of the load forecasting techniques using single and hybrid predictive models," *IEEE Access*, vol. 8, pp. 134911–134939, 2020.

[10] S. S. Ali and B. J. Choi, "State-of-the-art artificial intelligence techniques for distributed smart grids: A review," *Electronics*, vol. 9, no. 6, p. 1030, Jun. 2020.

[11] L. Yin, Q. Gao, L. Zhao, B. Zhang, T. Wang, S. Li, and H. Liu, "A review of machine learning for new generation smart dispatch in power systems," *Eng. Appl. Artif. Intell.*, vol. 88, Feb. 2020, Art. no. 103372.

[12] M. Massaoudi, A. Darwish, S. S. Refaat, H. Abu-Rub, and H. A. Toliyat, "UHF partial discharge localization in gas-insulated switchgears: Gradient boosting based approach," in *Proc. IEEE Kansas Power Energy Conf. (KPEC)*, Jul. 2020, pp. 1–5.

[13] S. Zhao, F. Blaabjerg, and H. Wang, "An overview of artificial intelligence applications for power electronics," *IEEE Trans. Power Electron.*, vol. 36, no. 4, pp. 4633–4658, Apr. 2021.

[14] A. Mellit, A. Massi Pavan, E. Ogliari, S. Leva, and V. Lughi, "Advanced methods for photovoltaic output power forecasting: A review," *Appl. Sci.*, vol. 10, no. 2, p. 487, Jan. 2020.

[15] S. Shamshirband, T. Rabczuk, and K.-W. Chau, "A survey of deep learning techniques: Application in wind and solar energy resources," *IEEE Access*, vol. 7, pp. 164650–164666, 2019.

[16] Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, and C. Wang, "Machine learning and deep learning methods for cybersecurity," *IEEE Access*, vol. 6, pp. 35365–35381, 2018.

[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[18] D. Syed, A. Zainab, S. S. Refaat, H. Abu-Rub, and O. Bouhali, "Smart grid big data analytics: Survey of technologies, techniques, and applications," *IEEE Access*, early access, Nov. 27, 2020, doi: 10.1109/ACCESS.2020.3041178.

[19] J. Wang, X. Chen, F. Zhang, F. Chen, and Y. Xin, "Building load forecasting using deep neural network with efficient feature fusion," *J. Mod. Power Syst. Clean Energy*, vol. 9, no. 1, pp. 160–169, 2021.

[20] F. Liang, W. Yu, X. Liu, D. Griffith, and N. Golmie, "Toward edge-based deep learning in industrial Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4329–4341, May 2020.

[21] M. Ourahou, W. Ayrir, B. EL Hassouni, and A. Haddi, "Review on smart grid control and reliability in presence of renewable energies: Challenges and prospects," *Math. Comput. Simul.*, vol. 167, pp. 19–31, Jan. 2020.

[22] D. Zhang, X. Han, and C. Deng, "Review on the research and practice of deep learning and reinforcement learning in smart grids," *CSEE J. Power Energy Syst.*, vol. 4, no. 3, pp. 362–370, Sep. 2018.

[23] T. Ahmad, H. Zhang, and B. Yan, "A review on renewable energy and electricity requirement forecasting models for smart grid and buildings," *Sustain. Cities Soc.*, vol. 55, Apr. 2020, Art. no. 102052.

[24] B. K. Sovacool, J. Axsen, and S. Sorrell, "Promoting novelty, rigor, and style in energy social science: Towards codes of practice for appropriate methods and research design," *Energy Res. Social Sci.*, vol. 45, pp. 12–42, Nov. 2018.

[25] M. Mohamed-Seghir, A. Krama, S. S. Refaat, M. Trabelsi, and H. Abu-Rub, "Artificial intelligence-based weighting factor autotuning for model predictive control of grid-tied packed U-cell inverter," *Energies*, vol. 13, no. 12, p. 3107, Jun. 2020.

[26] B. W. Wirtz, J. C. Weyerer, and C. Geyer, "Artificial intelligence and the public sector—Applications and challenges," *Int. J. Public Admin.*, vol. 42, no. 7, pp. 596–615, 2019.

[27] J. T. Rand, L. A. Kramer, C. P. Garrity, B. D. Hoen, J. E. Diffendorfer, H. E. Hunt, and M. Spears, "A continuously updated, geospatially rectified database of utility-scale wind turbines in the united states," *Sci. Data*, vol. 7, no. 1, pp. 1–12, Dec. 2020.

[28] X. Chen, Y. Du, H. Wen, L. Jiang, and W. Xiao, "Forecasting-based power ramp-rate control strategies for utility-scale PV systems," *IEEE Trans. Ind. Electron.*, vol. 66, no. 3, pp. 1862–1871, Mar. 2019.

[29] 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, Jan. 2021, Art. no. 118874.
[30] 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, Feb. 2020.
[31] M. Massaoudi, S. S. Refaat, and H. Abu-Rub, "On the pivotal role of artificial intelligence toward the evolution of smart grids: A review of advanced methodologies and applications," in *Smart Grid Enabling Technology*. Hoboken, NJ, USA: Wiley, 2021.
[32] D. T. Tran, S. Kiranyaz, M. Gabbouj, and A. Iosifidis, "Heterogeneous multilayer generalized operational perceptron," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 3, pp. 710–724, Mar. 2020.
[33] S. Kiranyaz, T. Ince, A. Iosifidis, and M. Gabbouj, "Operational neural networks," *Neural Comput. Appl.*, vol. 2, pp. 1–24, Mar. 2020.
[34] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, Apr. 2021, Art. no. 107398.
[35] T. Veniat and L. Denoyer, "Learning Time/Memory-efficient deep architectures with budgeted super networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3492–3500.
[36] G. Hurlburt, "Superintelligence: Myth or pressing reality?" *IT Prof.*, vol. 19, no. 1, pp. 6–11, Jan. 2017.
[37] Y. Fan, B. Tao, Y. Zheng, and S.-S. Jang, "A data-driven soft sensor based on multilayer perceptron neural network with a double LASSO approach," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 3972–3979, Jul. 2020.
[38] Y. Hong, Y. Zhou, Q. Li, W. Xu, and X. Zheng, "A deep learning method for short-term residential load forecasting in smart grid," *IEEE Access*, vol. 8, pp. 55785–55797, 2020.
[39] K. Hamedani, L. Liu, R. Atat, J. Wu, and Y. Yi, "Reservoir computing meets smart grids: Attack detection using delayed feedback networks," *IEEE Trans. Ind. Informat.*, vol. 14, no. 2, pp. 734–743, Feb. 2018.
[40] S. Hosein and P. Hosein, "Load forecasting using deep neural networks," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Apr. 2017, pp. 1–5.
[41] N. Ayoub, F. Musharavati, S. Pokharel, and H. A. Gabbar, "ANN model for energy demand and supply forecasting in a hybrid energy supply system," in *Proc. IEEE Int. Conf. Smart Energy Grid Eng. (SEGE)*, Aug. 2018, pp. 25–30.
[42] X. Pan, T. Zhao, M. Chen, and S. Zhang, "DeepOPF: A deep neural network approach for security-constrained DC optimal power flow," *IEEE Trans. Power Syst.*, early access, Sep. 24, 2020, doi: 10.1109/TPWRS.2020.3026379.
[43] M. N. Fekri, H. Patel, K. Grolinger, and V. Sharma, "Deep learning for load forecasting with smart meter data: Online adaptive recurrent neural network," *Appl. Energy*, vol. 282, Jan. 2021, Art. no. 116177.
[44] M. A. Ferrag and L. Maglaras, "DeepCoin: A novel deep learning and blockchain-based energy exchange framework for smart grids," *IEEE Trans. Eng. Manag.*, vol. 67, no. 4, pp. 1285–1297, Nov. 2020.
[45] 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 *Proc. 2nd Int. Conf. Smart Grid Renew. Energy (SGRE)*, Nov. 2019, pp. 1–6.
[46] 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. 36571–36588, 2021.
[47] M. Alazab, S. Khan, S. S. R. Krishnan, Q.-V. Pham, M. P. K. Reddy, and T. R. Gadekallu, "A multidirectional LSTM model for predicting the stability of a smart grid," *IEEE Access*, vol. 8, pp. 85454–85463, 2020.
[48] S. Sharma, "Activation functions in neural networks," *Towards Data Sci.*, vol. 6, pp. 1–7, Sep. 2017.
[49] M. Abdel-Nasser, K. Mahmoud, and M. Lehtonen, "Reliable solar irradiance forecasting approach based on choquet integral and deep LSTMs," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1873–1881, Mar. 2021.
[50] Z. Li, T. Zheng, Y. Wang, Z. Cao, Z. Guo, and H. Fu, "A novel method for imbalanced fault diagnosis of rotating machinery based on generative adversarial networks," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 3500417.
[51] B. Farsi, M. Amayri, N. Bouguila, and U. Eicker, "On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach," *IEEE Access*, vol. 9, pp. 31191–31212, 2021.
[52] G. Dudek, P. Pelka, and S. Smyl, "A hybrid residual dilated LSTM and exponential smoothing model for midterm electric load forecasting," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jan. 8, 2021, doi: 10.1109/TNNLS.2020.3046629.
[53] A. A. Abdelsalam, A. A. Salem, E. S. Oda, and A. A. Eldesouky, "Islanding detection of microgrid incorporating inverter based dgs using long short-term memory network," *IEEE Access*, vol. 8, pp. 106471–106486, 2020.
[54] W. Gao, R.-J. Wai, and S.-Q. Chen, "Novel pv fault diagnoses via sae and improved multi-grained cascade forest with string voltage and currents measures," *IEEE Access*, vol. 8, pp. 133144–133160, 2020.
[55] A. Ullah, N. Javaid, O. Samuel, M. Imran, and M. Shoaib, "CNN and GRU based deep neural network for electricity theft detection to secure smart grid," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, Jun. 2020, pp. 1598–1602.
[56] M.-F. Guo, N.-C. Yang, and W.-F. Chen, "Deep-learning-based fault classification using Hilbert–Huang transform and convolutional neural network in power distribution systems," *IEEE Sensors J.*, vol. 19, no. 16, pp. 6905–6913, Aug. 2019.
[57] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2048–2057.
[58] 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. 44821–44829, 2020.
[59] M. Canizo, I. Triguero, A. Conde, and E. Onieva, "Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study," *Neurocomputing*, vol. 363, pp. 246–260, Oct. 2019.
[60] Y. Qin, D. Chen, S. Xiang, and C. Zhu, "Gated dual attention unit neural networks for remaining useful life prediction of rolling bearings," *IEEE Trans. Ind. Informat.*, early access, Jun. 2, 2020, doi: 10.1109/TII.2020.2999442.
[61] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, H. Abu-Rub, and F. S. Oueslati, "Short-term electric load forecasting based on data-driven deep learning techniques," in *Proc. 46th Annu. Conf. IEEE Ind. Electron. Soc.*, Oct. 2020, pp. 2565–2570.
[62] A. Saeed, C. Li, M. Danish, S. Rubaiee, G. Tang, Z. Gan, and A. Ahmed, "Hybrid bidirectional LSTM model for short-term wind speed interval prediction," *IEEE Access*, vol. 8, pp. 182283–182294, 2020.
[63] Y. Deng, L. Wang, H. Jia, X. Tong, and F. Li, "A sequence-to-sequence deep learning architecture based on bidirectional GRU for type recognition and time location of combined power quality disturbance," *IEEE Trans. Ind. Informat.*, vol. 15, no. 8, pp. 4481–4493, Aug. 2019.
[64] H. Liu, Z. Liu, W. Jia, and X. Lin, "Remaining useful life prediction using a novel feature-attention-based end-to-end approach," *IEEE Trans. Ind. Informat.*, vol. 17, no. 2, pp. 1197–1207, Feb. 2021.
[65] C.-L. Liu, W.-H. Hsaio, and Y.-C. Tu, "Time series classification with multivariate convolutional neural network," *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4788–4797, Jun. 2019.
[66] X. Tao, D. Zhang, Z. Wang, X. Liu, H. Zhang, and D. Xu, "Detection of power line insulator defects using aerial images analyzed with convolutional neural networks," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 50, no. 4, pp. 1486–1498, Apr. 2020.
[67] J. Choung, S. Lim, S. H. Lim, S. C. Chi, and M. H. Nam, "Automatic discontinuity classification of wind-turbine blades using a-scan-based convolutional neural network," *J. Mod. Power Syst. Clean Energy*, vol. 9, no. 1, pp. 210–218, 2021.
[68] A. Bagheri, I. Y. H. Gu, M. H. J. Bollen, and E. Balouji, "A robust transform-domain deep convolutional network for voltage dip classification," *IEEE Trans. Power Del.*, vol. 33, no. 6, pp. 2794–2802, Dec. 2018.
[69] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1606–1615, Apr. 2018.
[70] 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, Oct. 2020.
[71] M. Dong and L. Grumbach, "A hybrid distribution feeder long-term load forecasting method based on sequence prediction," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 470–482, Jan. 2020.

[72] Y. Wang, J. Chen, X. Chen, X. Zeng, Y. Kong, S. Sun, Y. Guo, and Y. Liu, "Short-term load forecasting for industrial customers based on TCN-LightGBM," *IEEE Trans. Power Syst.*, early access, Oct. 1, 2020, doi: 10.1109/TPWRS.2020.3028133.

[73] Y. Yang, J. Zhong, W. Li, T. A. Gulliver, and S. Li, "Semisupervised multilabel deep learning based nonintrusive load monitoring in smart grids," *IEEE Trans. Ind. Informat.*, vol. 16, no. 11, pp. 6892–6902, Nov. 2020.

[74] L. Cheng, H. Zang, Y. Xu, Z. Wei, and G. Sun, "Probabilistic residential load forecasting based on micrometeorological data and customer consumption pattern," *IEEE Trans. Power Syst.*, early access, Jan. 14, 2021, doil: 10.1109/TPWRS.2021.3051684.

[75] S. Wang, S. Bi, and Y.-J.-A. Zhang, "Locational detection of the false data injection attack in a smart grid: A multilabel classification approach," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8218–8227, Sep. 2020.

[76] E. Balouji and O. Salor, "Classification of power quality events using deep learning on event images," in *Proc. 3rd Int. Conf. Pattern Recognit. Image Anal. (IPRIA)*, Apr. 2017, pp. 216–221.

[77] X. Kong, C. Li, F. Zheng, and C. Wang, "Improved deep belief network for short-term load forecasting considering demand-side management," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1531–1538, Mar. 2020.

[78] G. Hafeez, K. S. Alimgeer, and I. Khan, "Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid," *Appl. Energy*, vol. 269, Jul. 2020, Art. no. 114915.

[79] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2505–2516, Sep. 2017.

[80] H. Zhao, H. Liu, W. Hu, and X. Yan, "Anomaly detection and fault analysis of wind turbine components based on deep learning network," *Renew. Energy*, vol. 127, pp. 825–834, Nov. 2018.

[81] G. Dong, G. Liao, H. Liu, and G. Kuang, "A review of the autoencoder and its variants: A comparative perspective from target recognition in synthetic-aperture radar images," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 3, pp. 44–68, Sep. 2018.

[82] T. Wang, Y. He, B. Li, and T. Shi, "Transformer fault diagnosis using self-powered RFID sensor and deep learning approach," *IEEE Sensors J.*, vol. 18, no. 15, pp. 6399–6411, Aug. 2018.

[83] C. Ye, Y. Ding, P. Wang, and Z. Lin, "A data-driven bottom-up approach for spatial and temporal electric load forecasting," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 1966–1979, May 2019.

[84] L. Yang and Z. Zhang, "Wind turbine gearbox failure detection based on SCADA data: A deep learning-based approach," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 3507911.

[85] L. Wang, Z. Zhang, and J. Chen, "Short-term electricity price forecasting with stacked denoising autoencoders," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2673–2681, Jul. 2017.

[86] Y. M. Khaw, A. A. Jahromi, M. F. M. Arani, S. Sanner, D. Kundur, and M. Kassouf, "A deep learning-based cyberattack detection system for transmission protective relays," *IEEE Trans. Smart Grid*, early access, Nov. 25, 2020, doi: 10.1109/TSG.2020.3040361.

[87] W. Dong, H. Sun, Z. Li, J. Zhang, and H. Yang, "Short-term wind-speed forecasting based on multiscale mathematical morphological decomposition, K-means clustering, and stacked denoising autoencoders," *IEEE Access*, vol. 8, pp. 146901–146914, 2020.

[88] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan, and X. Chen, "Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2416–2425, Nov. 2018.

[89] W. Wang, X. Du, D. Shan, R. Qin, and N. Wang, "Cloud intrusion detection method based on stacked contractive auto-encoder and support vector machine," *IEEE Trans. Cloud Comput.*, early access, Jun. 9, 2020, doi: 10.1109/TCC.2020.3001017.

[90] S. Ryu, H. Choi, H. Lee, and H. Kim, "Convolutional autoencoder based feature extraction and clustering for customer load analysis," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1048–1060, Mar. 2020.

[91] R. Jiao, K. Peng, and J. Dong, "Remaining useful life prediction of lithium-ion batteries based on conditional variational autoencoders-particle filter," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8831–8843, Nov. 2020.

[92] F. Wang, Z. Zhang, C. Liu, Y. Yu, S. Pang, N. Duić, M. Shafie-khah, and J. P. S. Catalão, "Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting," *Energy Convers. Manage.*, vol. 181, pp. 443–462, Feb. 2019.

[93] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F.-Y. Wang, "Generative adversarial networks: Introduction and outlook," *IEEE/CAA J. Automatica Sinica*, vol. 4, no. 4, pp. 588–598, 2017.

[94] Y. Wang, G. Hug, Z. Liu, and N. Zhang, "Modeling load forecast uncertainty using generative adversarial networks," *Electr. Power Syst. Res.*, vol. 189, Dec. 2020, Art. no. 106732.

[95] F. Li, D. Lin, and T. Yu, "Improved generative adversarial network-based super resolution reconstruction for low-frequency measurement of smart grid," *IEEE Access*, vol. 8, pp. 85257–85270, 2020.

[96] A. Aldegheishem, M. Anwar, N. Javaid, N. Alrajeh, M. Shafiq, and H. Ahmed, "Towards sustainable energy efficiency with intelligent electricity theft detection in smart grids emphasising enhanced neural networks," *IEEE Access*, vol. 9, pp. 25036–25061, 2021.

[97] D. Li, J. Tan, Q. Yu, and Z. Li, "Deep-learning-assisted power loss management for active distribution networks," *Electr. J.*, vol. 34, no. 1, Jan. 2021, Art. no. 106888.

[98] D. Cao, W. Hu, J. Zhao, G. Zhang, B. Zhang, Z. Liu, Z. Chen, and F. Blaabjerg, "Reinforcement learning and its applications in modern power and energy systems: A review," *J. Mod. Power Syst. Clean Energy*, vol. 8, no. 6, pp. 1029–1042, 2020.

[99] Z. Zhang, D. Zhang, and R. C. Qiu, "Deep reinforcement learning for power system applications: An overview," *CSEE J. Power Energy Syst.*, vol. 6, no. 1, pp. 213–225, 2019.

[100] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1928–1937.

[101] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double *Q*-learning," in *Proc. AAAI Conf. Artif. Intell.*, 2016, vol. 30, no. 1, pp. 2094–2100.

[102] M. Shin, D.-H. Choi, and J. Kim, "Cooperative management for PV/ESS-enabled electric vehicle charging stations: A multiagent deep reinforcement learning approach," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3493–3503, May 2020.

[103] F. Wei, Z. Wan, and H. He, "Cyber-attack recovery strategy for smart grid based on deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2476–2486, May 2020.

[104] C. Zhang, R. Li, H. Shi, and F. Li, "Deep learning for day-ahead electricity price forecasting," *IET Smart Grid*, vol. 3, no. 4, pp. 462–469, 2020.

[105] R.-J. Park, K.-B. Song, and B.-S. Kwon, "Short-term load forecasting algorithm using a similar day selection method based on reinforcement learning," *Energies*, vol. 13, no. 10, p. 2640, May 2020.

[106] X. Liu, J. Ospina, and C. Konstantinou, "Deep reinforcement learning for cybersecurity assessment of wind integrated power systems," *IEEE Access*, vol. 8, pp. 208378–208394, 2020.

[107] R. Huang, J. Li, W. Li, and L. Cui, "Deep ensemble capsule network for intelligent compound fault diagnosis using multisensory data," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 5, pp. 2304–2314, May 2020.

[108] R. Huang, J. Li, S. Wang, G. Li, and W. Li, "A robust weight-shared capsule network for intelligent machinery fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6466–6475, Oct. 2020.

[109] W. Maass, "To spike or not to spike: That is the question," *Proc. IEEE*, vol. 103, no. 12, pp. 2219–2224, Dec. 2015.

[110] H. Wang, W. Xue, Y. Liu, J. Peng, and H. Jiang, "Probabilistic wind power forecasting based on spiking neural network," *Energy*, vol. 196, Apr. 2020, Art. no. 117072.

[111] Y. Tao and H. Chen, "A hybrid wind power prediction method," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Jul. 2016, pp. 1–5.

[112] M. Kim, W. Choi, Y. Jeon, and L. Liu, "A hybrid neural network model for power demand forecasting," *Energies*, vol. 12, no. 5, p. 931, Mar. 2019.

[113] M. Afrasiabi, M. Mohammadi, M. Rastegar, and S. Afrasiabi, "Advanced deep learning approach for probabilistic wind speed forecasting," *IEEE Trans. Ind. Informat.*, vol. 17, no. 1, pp. 720–727, Jan. 2021.

[114] D. Li, P. Ramanan, N. Gebraeel, and K. Paynabar, "Deep learning based covert attack identification for industrial control systems," in *Proc. 19th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2020, pp. 438–445.

[115] J. Huang, L. Guan, Y. Su, H. Yao, M. Guo, and Z. Zhong, "Recurrent graph convolutional network-based multi-task transient stability assessment framework in power system," *IEEE Access*, vol. 8, pp. 93283–93296, 2020.

[116] *Deep learning in Smart Grid*. Accessed: Feb. 30, 2021. [Online]. Available: https://github.com/Mohamedmassaoudi/DL-in-SG

[117] 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 Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018.

[118] L. Chen, N. Qin, X. Dai, and D. Huang, "Fault diagnosis of high-speed train bogie based on capsule network," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6203–6211, Sep. 2020.

[119] Z. Zhu, G. Peng, Y. Chen, and H. Gao, "A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis," *Neurocomputing*, vol. 323, pp. 62–75, Jan. 2019.

[120] C.-Y. Zhang, C. P. Chen, M. Gan, and L. Chen, "Predictive deep Boltzmann machine for multiperiod wind speed forecasting," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1416–1425, 2015.

[121] J. Li, X. Li, and D. He, "A directed acyclic graph network combined with CNN and LSTM for remaining useful life prediction," *IEEE Access*, vol. 7, pp. 75464–75475, 2019.

[122] Q. Lu, X. Liao, H. Li, and T. Huang, "Achieving acceleration for distributed economic dispatch in smart grids over directed networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 3, pp. 1988–1999, Jul. 2020.

[123] H. Li, Z. Wan, and H. He, "Real-time residential demand response," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 4144–4154, Sep. 2020.

[124] Z. Xia, S. Xue, J. Wu, Y. Chen, J. Chen, and L. Wu, "Deep reinforcement learning for smart city communication networks," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 4188–4196, Jun. 2021.

[125] Z. Wang, H. He, Z. Wan, and Y. Sun, "Coordinated topology attacks in smart grid using deep reinforcement learning," *IEEE Trans. Ind. Informat.*, vol. 17, no. 2, pp. 1407–1415, Feb. 2021.

[126] Y. Wang, H. Tan, Y. Wu, and J. Peng, "Hybrid electric vehicle energy management with computer vision and deep reinforcement learning," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 3857–3868, Jun. 2021.

[127] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-term load forecasting with deep residual networks," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3943–3952, Jul. 2019.

[128] W. Wen, C. Xu, F. Yan, C. Wu, Y. Wang, Y. Chen, and H. Li, "Terngrad: Ternary gradients to reduce communication in distributed deep learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1509–1519.

[129] T.-E. Huang, Q. Guo, and H. Sun, "A distributed computing platform supporting power system security knowledge discovery based on online simulation," *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1513–1524, May 2017.

[130] A. Zainab, D. Syed, A. Ghrayeb, H. Abu-Rub, S. S. Refaat, M. Houchati, O. Bouhali, and S. B. Lopez, "A multiprocessing-based sensitivity analysis of machine learning algorithms for load forecasting of electric power distribution system," *IEEE Access*, vol. 9, pp. 31684–31694, 2021.

[131] J. Wang, D. Shi, Y. Li, J. Chen, H. Ding, and X. Duan, "Distributed framework for detecting PMU data manipulation attacks with deep autoencoders," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4401–4410, Jul. 2019.

[132] H.-M. Chung, S. Maharjan, Y. Zhang, and F. Eliassen, "Distributed deep reinforcement learning for intelligent load scheduling in residential smart grids," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2752–2763, Apr. 2021.

[133] G. De La Torre Parra, P. Rad, K.-K.-R. Choo, and N. Beebe, "Detecting Internet of Things attacks using distributed deep learning," *J. Netw. Comput. Appl.*, vol. 163, Aug. 2020, Art. no. 102662.

[134] V.-H. Bui, A. Hussain, and H.-M. Kim, "Double deep *Q*-learning-based distributed operation of battery energy storage system considering uncertainties," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 457–469, 2019.

[135] A. A. Diro and N. Chilamkurti, "Distributed attack detection scheme using deep learning approach for Internet of Things," *Future Gener. Comput. Syst.*, vol. 82, pp. 761–768, May 2018.

[136] J. C. Jiang, B. Kantarci, S. Oktug, and T. Soyata, "Federated learning in smart city sensing: Challenges and opportunities," *Sensors*, vol. 20, no. 21, p. 6230, Oct. 2020.

[137] H. Li, K. Ota, and M. Dong, "Learning IoT in edge: Deep learning for the Internet of Things with edge computing," *IEEE Netw.*, vol. 32, no. 1, pp. 96–101, Jan. 2018.

[138] X. Zhang, F. Fang, and J. Wang, "Probabilistic solar irradiation forecasting based on variational Bayesian inference with secure federated learning," *IEEE Trans. Ind. Informat.*, early access, Nov. 4, 2020, doi: 10.1109/TII.2020.3035807.

[139] S. Lee and D.-H. Choi, "Federated reinforcement learning for energy management of multiple smart homes with distributed energy resources," *IEEE Trans. Ind. Informat.*, early access, Nov. 3, 2020, doi: 10.1109/TII.2020.3035451.

[140] B. Li, Y. Wu, J. Song, R. Lu, T. Li, and L. Zhao, "DeepFed: Federated deep learning for intrusion detection in industrial cyber-physical systems," *IEEE Trans. Ind. Informat.*, early access, Sep. 11, 2020, doi: 10.1109/TII.2020.3023430.

[141] M. Liang, Y. Meng, J. Wang, D. Lubkeman, and N. Lu, "FeederGAN: Synthetic feeder generation via deep graph adversarial nets," *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 1163–1173, Mar. 2020.

[142] J. Baek, Q. H. Vu, J. K. Liu, X. Huang, and Y. Xiang, "A secure cloud computing based framework for big data information management of smart grid," *IEEE Trans. Cloud Comput.*, vol. 3, no. 2, pp. 233–244, Apr. 2015.

[143] L. Lyu, K. Nandakumar, B. Rubinstein, J. Jin, J. Bedo, and M. Palaniswami, "PPFA: Privacy preserving fog-enabled aggregation in smart grid," *IEEE Trans. Ind. Informat.*, vol. 14, no. 8, pp. 3733–3744, Aug. 2018.

[144] S. Bera, S. Misra, and J. J. P. C. Rodrigues, "Cloud computing applications for smart grid: A survey," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 5, pp. 1477–1494, May 2015.

[145] D. A. Chekired, L. Khoukhi, and H. T. Mouftah, "Fog-computing-based energy storage in smart grid: A cut-off priority queuing model for plug-in electrified vehicle charging," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3470–3482, May 2020.

[146] E. Li, Z. Zhou, and X. Chen, "Edge intelligence: On-demand deep learning model co-inference with device-edge synergy," in *Proc. Workshop Mobile Edge Commun.*, 2018, pp. 31–36.

[147] Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, "Edge intelligence: Paving the last mile of artificial intelligence with edge computing," *Proc. IEEE*, vol. 107, no. 8, pp. 1738–1762, Aug. 2019.

[148] X. Wang, Y. Han, V. C. M. Leung, D. Niyato, X. Yan, and X. Chen, "Convergence of edge computing and deep learning: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 869–904, 2nd Quart., 2020.

[149] X. Zhang, D. Biagioni, M. Cai, P. Graf, and S. Rahman, "An edge-cloud integrated solution for buildings demand response using reinforcement learning," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 420–431, Jan. 2021.

[150] *Microsoft Cognitive Toolkit (CNTK), An Open Source Deep-Learning Toolkit*. Accessed: Feb. 15, 2021. [Online]. Available: https://github.com/microsoft/CNTK

[151] *PyTorch: Tensors and Dynamic Neural Networks in Python With Strong GPU Acceleration*. Accessed: Feb. 30, 2021. [Online]. Available: https://github.com/pytorch/

[152] *TensorFlow: A System for Large-Scale Machine Learning|USENIX*. Accessed: Feb. 30, 2021. [Online]. Available: https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi

[153] *DeepLearning4j: Open-Source Distributed Deep Learning for the JVM, Apache Software Foundation License 2.0*. Accessed: Feb. 12, 2021. [Online]. Available: https://deeplearning4j.org/

[154] *Lightweight, Portable, Flexible Distributed/Mobile Deep Learning with Dynamic, Mutation-aware Dataflow Dep Scheduler; for Python, R, Julia, Scala, Go, Javascript and More*. Accessed: Feb. 30, 2021. [Online]. Available: https://github.com/apache/incubator-mxnet

[155] *Core ML: Integrate Machine Learning Models Into Your App*. Accessed: Feb. 14, 2021. [Online]. Available: https://developer.apple.com/documentation/coreml

[156] *Qualcomm Neural Processing SDK for AI*. Accessed: Feb. 30, 2021. [Online]. Available: https://developer.qualcomm.com/software/qualcomm-neural-processing-sdk

[157] *NCNN is a High-Performance Neural Network Inference Framework Optimized for the Mobile Platform*. Accessed: Feb. 30, 2021. [Online]. Available: https://github.com/Tencent/ncnn

[158] X. Jiang, H. Wang, Y. Chen, Z. Wu, L. Wang, B. Zou, Y. Yang, Z. Cui, Y. Cai, T. Yu, C. Lv, and Z. Wu, "MNN: A universal and efficient inference engine," in *Proc. MLSys*, 2020, pp. 1–13.

[159] *Multi-Platform Embedded Deep Learning Framework*. Accessed: Feb. 10, 2021. [Online]. Available: https://github.com/PaddlePaddle/Paddle-Lite

[160] *MACE is a Deep Learning Inference Framework Optimized for Mobile Heterogeneous Computing Platforms*. Accessed: Feb. 30, 2021. [Online]. Available: https://github.com/XiaoMi/mace

[161] X. Wang, M. Magno, L. Cavigelli, and L. Benini, "FANN-on-MCU: An open-source toolkit for energy-efficient neural network inference at the edge of the Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4403–4417, May 2020.

[162] S.-B. Lin, "Generalization and expressivity for deep nets," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 5, pp. 1392–1406, May 2019.

[163] E. Azarkhish, D. Rossi, I. Loi, and L. Benini, "Neurostream: Scalable and energy efficient deep learning with smart memory cubes," *IEEE Trans. Parallel Distrib. Syst.*, vol. 29, no. 2, pp. 420–434, Feb. 2018.

[164] H. Karimipour, A. Dehghantanha, R. M. Parizi, K.-K. R. Choo, and H. Leung, "A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids," *IEEE Access*, vol. 7, pp. 80778–80788, 2019.

[165] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy forecasting: A review and outlook," *IEEE Open Access J. Power Energy*, vol. 7, pp. 376–388, 2020.

[166] E. Lee and W. Rhee, "Individualized short-term electric load forecasting with deep neural network based transfer learning and meta learning," *IEEE Access*, vol. 9, pp. 15413–15425, 2021.

[167] B. Maschler and M. Weyrich, "Deep transfer learning for industrial automation," *IEEE Ind. Electron. Mag.*, early access, Jan. 18, 2021, doi: 10.1109/MIE.2020.3034884.

[168] Z. Zhou, Y. Xiang, H. Xu, Z. Yi, D. Shi, and Z. Wang, "A novel transfer learning-based intelligent nonintrusive load-monitoring with limited measurements," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 2500508.

[169] S. Zhou, L. Zhou, M. Mao, and X. Xi, "Transfer learning for photovoltaic power forecasting with long short-term memory neural network," in *Proc. IEEE Int. Conf. Big Data Smart Comput.*, Feb. 2020, pp. 125–132.

[170] 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," *Sol. Energy*, vol. 198, pp. 175–186, Mar. 2020.

[171] J. F. Torres, A. Troncoso, I. Koprinska, Z. Wang, and F. M. Álvarez, "Big data solar power forecasting based on deep learning and multiple data sources," *Expert Syst.*, vol. 36, no. 4, Aug. 2019.

[172] S.-V. Oprea and A. Bára, "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," *Comput. Ind.*, vol. 120, Sep. 2020, Art. no. 103230.

[173] S. E. Haupt and B. Kosovic, "Variable generation power forecasting as a big data problem," *IEEE Trans. Sustain. Energy*, vol. 8, no. 2, pp. 725–732, Apr. 2017.

[174] A. Walch, R. Castello, N. Mohajeri, and J.-L. Scartezzini, "Big data mining for the estimation of hourly rooftop photovoltaic potential and its uncertainty," *Appl. Energy*, vol. 262, Mar. 2020, Art. no. 114404.

[175] S.-H. Kim, C. Lee, and C.-H. Youn, "An accelerated edge cloud system for energy data stream processing based on adaptive incremental deep learning scheme," *IEEE Access*, vol. 8, pp. 195341–195358, 2020.

[176] P. Li, Z. Chen, L. T. Yang, J. Gao, Q. Zhang, and M. J. Deen, "An incremental deep convolutional computation model for feature learning on industrial big data," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1341–1349, Mar. 2019.

[177] R. Razavi-Far, E. Hallaji, M. Farajzadeh-Zanjani, M. Saif, S. H. Kia, H. Henao, and G.-A. Capolino, "Information fusion and semi-supervised deep learning scheme for diagnosing gear faults in induction machine systems," *IEEE Trans. Ind. Electron.*, vol. 66, no. 8, pp. 6331–6342, Aug. 2019.

[178] M. Xia, T. Li, L. Xu, L. Liu, and C. W. de Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 101–110, Feb. 2018.

[179] S. Zhang, F. Ye, B. Wang, and T. G. Habetler, "Semi-supervised bearing fault diagnosis and classification using variational autoencoder-based deep generative models," *IEEE Sensors J.*, vol. 21, no. 5, pp. 6476–6486, Mar. 2021.

[180] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Gearbox fault diagnosis using a deep learning model with limited data sample," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6263–6271, Oct. 2020.

[181] C. Liu, C. Qin, X. Shi, Z. Wang, G. Zhang, and Y. Han, "TScatNet: An interpretable cross-domain intelligent diagnosis model with antinoise and few-shot learning capability," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 3506110.

[182] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol. 549, pp. 195–202, Sep. 2017.

[183] R. Eskandarpour, K. J. B. Ghosh, A. Khodaei, A. Paaso, and L. Zhang, "Quantum-enhanced grid of the future: A primer," *IEEE Access*, vol. 8, pp. 188993–189002, 2020.

[184] A. Ajagekar and F. You, "Quantum computing for energy systems optimization: Challenges and opportunities," *Energy*, vol. 179, pp. 76–89, Jul. 2019.

[185] R. Cioffi, M. Travaglioni, G. Piscitelli, A. Petrillo, and F. De Felice, "Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions," *Sustainability*, vol. 12, no. 2, p. 492, Jan. 2020.

[186] C. Kacfah Emani, N. Cullot, and C. Nicolle, "Understandable big data: A survey," *Comput. Sci. Rev.*, vol. 17, pp. 70–81, Aug. 2015.

[187] B. Wang and N. Z. Gong, "Stealing hyperparameters in machine learning," in *Proc. IEEE Symp. Secur. Privacy*, May 2018, pp. 36–52.

[188] S. Sachin Kumar and P. Aithal, "How lucrative & challenging the boundary less opportunities for data scientists?" *Int. J. Case Stud. Bus.*, vol. 4, no. 1, pp. 223–236, 2020.

[189] M. A. Al-Khasawneh, "Data science tools and applications," in *Advances in Academic Research and Development*. New Delhi, India: Integrated Publications, 2020, ch. 3, p. 35.

[190] Y. K. Dwivedi, L. Hughes, E. Ismagilova, G. Aarts, C. Coombs, T. Crick, Y. Duan, R. Dwivedi, J. Edwards, A. Eirug, and V. Galanos, "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *Int. J. Inf. Manage.*, vol. 27, Aug. 2019, Art. no. 101994.

[191] P. Mikalef and J. Krogstie, "Investigating the data science skill gap: An empirical analysis," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Apr. 2019, pp. 1275–1284.

[192] A. De Mauro, M. Greco, M. Grimaldi, and G. Nobili, "Beyond data scientists: A review of big data skills and job families," in *Proc. IFKAD*, 2016, pp. 1844–1857.

[193] *Data Science for Undergraduates: Opportunities and Options*. National Academies Press, Washington, DC, USA, 2018.

[194] J. Choo and S. Liu, "Visual analytics for explainable deep learning," *IEEE Comput. Graph. Appl.*, vol. 38, no. 4, pp. 84–92, 2018.

[195] P. Angelov and E. Soares, "Towards explainable deep neural networks (xDNN)," *Neural Netw.*, vol. 130, pp. 185–194, Oct. 2020.

[196] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.

[197] T. Chen, T. Moreau, Z. Jiang, L. Zheng, and E. Yan, "*TVM*: An automated end-to-end optimizing compiler for deep learning," in *Proc. 13th Symp. Oper. Syst. Design Implement.*, 2018, pp. 578–594.

[198] D. Rutqvist, D. Kleyko, and F. Blomstedt, "An automated machine learning approach for smart waste management systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 384–392, Jan. 2020.

[199] Z. Liu, Z. Xu, S. Rajaa, M. Madadi, J. C. J. Junior, S. Escalera, A. Pavao, S. Treguer, W.-W. Tu, and I. Guyon, "Towards automated deep learning: Analysis of the autodl challenge series 2019," in *Proc. NeurIPS Competition Demonstration Track*, 2020, pp. 242–252.

[200] A. Dundar, J. Jin, B. Martini, and E. Culurciello, "Embedded streaming deep neural networks accelerator with applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 7, pp. 1572–1583, Jul. 2017.

[201] I. A. Moonesar and R. Dass, "Artificial intelligence in health policy—A global perspective," *Global J. Comput. Sci. Technol.*, vol. 1, pp. 1–7, Feb. 2021.

[202] R. Mayer and H.-A. Jacobsen, "Scalable deep learning on distributed infrastructures: Challenges, techniques, and tools," *ACM Comput. Surv.*, vol. 53, no. 1, pp. 1–37, May 2020.

[203] C. Wang, L. Gong, Q. Yu, X. Li, Y. Xie, and X. Zhou, "DLAU: A scalable deep learning accelerator unit on FPGA," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 36, no. 3, pp. 513–517, Mar. 2017.

[204] A. Mazaheri, J. Schulte, M. W. Moskewicz, F. Wolf, and A. Jannesari, "Enhancing the programmability and performance portability of GPU tensor operations," in *Proc. Eur. Conf. Parallel Process.* Poland: Springer, 2019, pp. 213–226.

[205] L.-W. Kim, "DeepX: Deep learning accelerator for restricted Boltzmann machine artificial neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1441–1453, May 2018.
[206] C. Wang, L. Gong, A. Wang, X. Li, P. C. K. Hung, and Z. Xuehai, "SOLAR: Services-oriented deep learning architectures," *IEEE Trans. Services Comput.*, vol. 14, no. 1, pp. 262–273, Feb. 2017.

MOHAMED MASSAOUDI (Student Member, IEEE) received the degree in energy engineering from the National Engineering School of Monastir, ENIM, Tunisia, in 2018. He is currently pursuing the Ph.D. degree with INSAT, Tunisia. He is also a Student Worker with the Texas A&M University at Qatar. His research interests include machine learning and deep learning techniques for energy management and big data analytics in smart grid systems. He has served as a Reviewer for some conferences and journals, including the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS and the IEEE TRANSACTIONS ON POWER SYSTEMS.

HAITHAM ABU-RUB (Fellow, IEEE) received the two Ph.D. degrees.

He has worked at many universities in many countries including Poland, Palestine, USA, Germany, and Qatar. Since 2006, he has been with Texas A&M University at Qatar, where he has served for five years as the Chair for the Electrical and Computer Engineering Program. He has been also serving as the Managing Director for the Smart Grid Center. He has published more than 400 journal and conference papers, five books and six book chapters. His main research interests include electric drives, power electronic converters, renewable energy, and smart grid. He was the recipient of many national and international awards and recognitions, the American Fulbright Scholarship, the German Alexander von Humboldt Fellowship, and many others. He has supervised many research projects on smart grid, power electronics converters, and renewable energy systems.

SHADY S. REFAAT (Senior Member, IEEE) received the B.A.Sc., M.A.Sc., and Ph.D. degrees in electrical engineering from Cairo University, Giza, Egypt, in 2002, 2007, and 2013, respectively. He has worked in industry for more than 12 years, as an Engineering Team Leader, a Senior Electrical Engineer, and an Electrical Design Engineer, on various electrical engineering projects. He is currently an Assistant Research Scientist with the Department of Electrical and Computer Engineering, Texas A&M University at Qatar. He has published more than 95 journal and conference papers. His research interests include electrical machines, power systems, smart grid, big data, energy management systems, reliability of power grids and electric machinery, fault detection, and condition monitoring and development of fault-tolerant systems. He has also participated and led several scientific projects over the last eight years. He has successfully realized many potential research projects. He is also a member of The Institution of Engineering and Technology (IET) and the Smart Grid Center–Extension in Qatar (SGC-Q).

INES CHIHI received the Ph.D. and University Habilitation degrees from the National Engineering School of Tunis, Tunisia (ENIT), in 2013 and 2019, respectively. She is currently an Associate Professor of automation and industrial computing with the National Engineering School of Bizerta, Tunisia (ENIB) and a member of the Laboratory of Energy Applications and Renewable Energy Efficiency (LAPER), University Tunis El Manar. Her research interests include intelligent modeling, control, and fault detection of complex systems with unpredictable behaviors.

FAKHREDDINE S. OUESLATI received the Ph.D. and Habilitation degrees in physics to direct research from the Faculty of Sciences of Tunis, University of Tunis El Manar, Tunisia, in 2002 and 2012, respectively. He was a Civil Engineer with the National Engineering School of Tunis (ENIT), University of Tunis El Manar, in 1990. He worked as a Project Engineer with AFH, Tunis, in 1990 and 2002. He worked as an Assistant Professor with ISSAT in 2002 and 2013, the Higher Institute of Applied Sciences and Technology of Gabes, INRST, the National Institute of Scientific and Technical Research, Borj Cedria, the University of Carthage, Tunisia, ESTI, the Higher School of Technology and Computer, and the University of Carthage, Tunis, Tunisia. He was an Associate Professor in 2013 and 2018. Since 2018, he was a Professor with the National Engineering School of Carthage (ENICarthage), University of Carthage. His research interests include materials science, energy systems, pollution, and renewable energies, with expertise in multi-component multi-phase convection-diffusion problems. Currently responsible for the axis of the flow and transport phenomena in the Laboratory of Energy and Heat and Mass Transfer (LETTM), Faculty of Sciences of Tunis (FST), Tunisia. He has several international publications, is a member of several boards of directors of scientists and conference organizing committees. He has been a visiting professor in several universities and international and prestigious organizations. He has chaired many international conferences. He was a regular reviewer for journals and international research projects and an invited professor by many institutes.