IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 June 2024, accepted 4 July 2024, date of publication 8 July 2024, date of current version 16 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3424854

SURVEY

Spatiotemporal Deep Learning for Power System Applications: A Survey

MOHSEN SAFFARI, (Graduate Student Member, IEEE), AND MAHDI KHODAYAR, (Member, IEEE)
Department of Computer Science, The University of Tulsa, Tulsa, OK 74104, USA
Corresponding author: Mohsen Saffari (mohsen-saffari@utulsa.edu)

This work was supported by the National Science Foundation under Grant ECCS-2223628.

ABSTRACT Understanding spatiotemporal correlations in power systems is crucial for maintaining grid stability, reliability, and efficiency. By discerning connections between spatial and temporal dimensions, operators can anticipate and address issues such as congestion, voltage instability, and equipment failures. Recent advancements in power system analysis have leveraged spatiotemporal correlations through sophisticated data-driven algorithms. In this survey paper, we conduct a comprehensive examination of deep learning frameworks tailored to tackle the complexities inherent in spatiotemporal data analysis within power systems. We categorize machine learning methodologies into discriminative, generative, and reinforcement learning, providing a structured overview of their mathematical foundations, advantages, and limitations in processing dynamic power system measurements. Through empirical evaluations, we assess the performance of these methodologies across various spatiotemporal applications, including cyber attack detection, fault identification, demand response, and renewable energy forecasting, offering insights into their efficacy and applicability. Additionally, we identify emerging topics within the machine learning domain that hold promise for future endeavors in power systems analysis.

INDEX TERMS Spatiotemporal correlations, power systems applications, deep neural architectures.

ABBREVIATIONS

AC-NFTSM	Actor-Critic and Nonsingular Fast Terminal Sliding Mode.
AH-DAPE	Attention-based Hierarchical Dynamic Graph Pooling Network.
ASARSA	Adaptive SARSA.
ASU-LSTM	Appearance Similarity Updating LSTM.
AttG-BDGNets	Attention-Guided Bidirectional Dynamic Graph Networks.
B-LSTM	Bi-directional LSTM.
Bi-AnoGAN	Bidirectional Anomaly GAN.
BtM	Behind-the-Meter.
C-GAN	Conditional GAN.
CGRVAE	Convolution Graph Rough VAE.
CKF	Cubature Kalman Filter.
CNN	Convolutional Neural Networks.
ConvRLSTM	Convolutional Rough LSTM.
Cplx-STGCN	Complex-valued Spatiotemporal Graph Convolutional Network.
CSTWPP	Convolution-based Spatial-Temporal Wind Power Predictor.
CVAE	Conditional VAE.
cWGAN-GP	Conditional Wasserstein GAN with Gradient Penalty.
DCNN	Deep Convolutional Neural Network.
DDL	Deep Discriminative Learning.
DDQN	Double Deep Q-Network.
DNNs	Deep Neural Networks.
EA-MAAC	Experience Augmented Multi-Agent Actor-Critic.
ELM	Extreme Learning Machine.
ESARSA	Effective SARSA.
ESN	Echo State Network.
FDQN	Fast Deep Q-Network.

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

FTSBA	Fast Transient Stability Batch Assessment.
GAN	Generative Adversarial Network.
GC-LSTM	Graph Convolution LSTM.
GCGRU	Graph Convolution GRU.
GCLSTM	Graph Convolution LSTM.
GCNN	Graph Convolutional Neural Network.
RLSTM	Graph Rough LSTM.
GRU	Gated Recurrent Unit.
GSINN	Group Solar Irradiance Neural Network.
GSTGNN	Gating Spatiotemporal Graph Neural Network.
GWO	Gray Wolf Optimization.
KL	Kullback-Leibler.
LSTM	Long-Short Term Memory.
LSTM-AE	Long-Short Term Memory Autoencoder.
MA-OCDDPG	Multi-Agent Option-Critic Deep Deterministic Policy Gradient.
MAAAC	Multi-Agent Actor-Attention-Critic.
MAAC	Multi-Agent Actor-Critic.
MADDPG	Multi-Agent DDPG.
MADDPG	Multi-Agent Deep Deterministic Policy Gradient.
MAE	Mean Absolute Error.
MAPE	Mean Absolute Percentage Error.
MAPPO	Multi-agent PPO.
MARL	Multi-Agent RL.
MG-ASTGCN	Multi-Grained Attention-based Spatial-Temporal GCN.
MOAC	Myopic Optimization-based Actor-Critic.
MTSSCNN	Multi-Temporal-Spatial-Scale CNN.
NILM	Non-Intrusive Load Monitoring.
NREL	National Renewable Energy Laboratory.
NWP	Numerical Weather Prediction.
OPF	optimal power flow.
PCA	Principal Component Analysis.
PDF	Probability Distribution Function.
PMU	Phasor Measurement Unit.
POMDP	Partial Observable Markov Decision Process.
PPO	Proximal Policy Optimization.
PV	Photovoltaic.
R-GNN	Recurrent Graph Neural Network.
RC	Reservoir Computing.
ReLU	Rectified Linear Unit.
RF	Random Forest.
RL	Reinforcement Learning.
RMSE	Root Mean Square Error.
RNN	Recurrent Neural Networks.
RT-CPDLC	Real-Time Cyber-Power Event Detection, Location and Classification.
SAC	Spatial Auto Correlation.
SAE	Stacked Autoencoder.
SARSA	State-Action-Reward-State-Action.
SARSA-MDP	SARSA Markov Decision Process.
SCADA	Supervisory Control and Data Acquisition.
SGD	stochastic Gradient Descent.
st-GNN	Spatiotemporal Graph Neural Network.
STAN	Spatiotemporal Attention Network.
STCNN	Spatiotemporal CNN.
STDGCNN	Spatiotemporal Directed Graph CNN.
STFT	Short-Time Fourier Transform.
STGRVAE	Spatiotemporal Graph Rough VAE.
STNN	Spatiotemporal Neural Network.
STSGCNN	Spatiotemporal Synchronous GCNN.
SVM	Support Vector Machines.
SVR	Support Vector Regression.
TD3	Twin Delayed Deep Deterministic.
TFVAE	Time-Frequency VAE.
TML-CNN	Teaching and Mutual Learning CNN.
VAE	Variational Autoencoder.
VMD-CNN	Variational Mode Decomposition CNN.
VSA	Voltage Stability Analyzer.

I. INTRODUCTION

In the realm of power systems, recognizing the relationship between space and time holds significant weight. Spatial correlation, in essence, pertains to the similarity in behaviors observed across different locations within the system, while temporal correlation refers to the persistence of these behaviors over time. When we merge these two aspects, we derive the concept of spatiotemporal correlation, which essentially underscores how the combined influence of space and time impacts system dynamics. This understanding becomes particularly crucial within power distribution systems, where it serves as a linchpin for various essential tasks such as sustainable energy forecasting [1], [2], [3], [4], fault detection [5], [6], [7], [8], load management [9], [10], [11], [12], [13], state estimation [14], [15], [16], [17], and cyber security [18], [19], [20], [21]. By discerning how phenomena evolve both spatially and temporally, we can optimize distribution efficiency, bolster reliability, and fortify the power distribution system's resilience against unforeseen challenges. Hence, prioritizing the analysis and utilization of spatiotemporal correlations emerges as a cornerstone in the realm of power systems engineering.

Data-driven approaches have been crucial in studying the spatiotemporal correlations of power systems over the past decade. The effectiveness and dependability of data-driven approaches used in managing and evaluating power systems are closely connected to the representation of data, specifically the characteristics generated from the original measured data [22]. Hence, a considerable proportion of difficulties related to the utilization of conventional data-driven algorithms in power systems stems from the approaches employed for preprocessing, specifically through unsupervised dimensionality reduction techniques such as principal component

analysis (PCA) [23], independent component analysis [24], linear discriminant analysis [25], auto-encoder (AE) [26], t-distributed stochastic neighbor embedding [27]. Although these strategies are helpful, they frequently fail to consider the intricate and ever-changing spatiotemporal patterns of power systems.

Generally speaking, the deep learning approaches for analysis and understanding of spatiotemporal correlations within the power systems are categorized as:

1) Discriminative Models: These models aim to directly uncover the relationship between input spatiotemporal patterns and the model's output across various classification and regression tasks within power systems domains [28]. Recent studies have demonstrated their capacity to yield accurate performance in load monitoring [29], voltage regulation [30], and forecasting tasks [31]. Nevertheless, a notable drawback of discriminative models lies in their dependence on labeled data for training, which can be limited or costly to acquire in certain contexts. Moreover, these models may encounter challenges in capturing the underlying generative processes of the data, potentially impeding their capacity to generalize to novel scenarios.
2) Generative Models: Unlike discriminative models, generative approaches seek to encapsulate the inherent distribution of spatiotemporal correlations rather than explicitly modeling the input-output relationship [28]. By grasping the distribution of input data, generative models can produce samples closely resembling the original spatiotemporal correlations within the power systems, offering valuable insights into the intrinsic variability and uncertainty of power systems dynamics. Within power systems analysis, these models have proven effective in tasks such as cyber security in active power distribution systems [32], probabilistic wind speed forecasting [33], and anomaly detection [34] by discerning the fundamental probability distribution of observed data. However, a notable challenge with generative models lies in the computational complexity inherent in their training, particularly when confronted with high-dimensional spatiotemporal data.
3) Reinforcement Learning (RL): These models present a distinctive approach in power systems analysis, focusing on sequential decision-making processes within dynamic environments. Unlike discriminative and generative models, these models operate by learning optimal strategies through interaction with the environment, maximizing cumulative rewards over time [35]. In the context of power systems, RL models have shown promise in tasks such as adaptive power system emergency control [36], cyber-physical security assessment [37], load management [38], and optimal power flow (OPF) control [39]. By leveraging spatiotemporal correlations, RL agents can make sequential decisions that enhance system performance and resilience in real-time scenarios. However, one challenge with RL models in power systems lies in the complexity of the environment and the need for extensive training to learn effective policies.

This survey paper is organized as follows. In Section II, deep discriminative architectures with their respective mathematics are thoroughly explained. By utilizing diverse practical applications and datasets in power systems, this paper comprehensively clarifies and empirically compares different variants of these machine learning approaches. Section III explores probabilistic deep neural architectures, encompassing traditional models such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs). This section covers both the practical applications and conceptual advantages of these techniques. Additionally, Section IV investigates the widespread use of deep RL algorithms in spatiotemporal analysis of power systems operation and management. Section V delves into emerging topics and novel challenges within the domain of deep learning and introduces potential future research lines within data-driven methods for spatiotemporal analysis of power systems. Finally, Section VI provides a conclusion that synthesizes the paper's findings.

II. DEEP DISCRIMINATIVE LEARNING

Deep discriminative learning (DDL) stands as a prevalent method for processing spatiotemporal correlations within power systems, addressing various challenges concerning system resilience and reliability. In this approach, given a training dataset comprising input-output pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, the primary objective is to directly learn a function f that maps input patterns $X = (x_1, x_2, \ldots, x_n)$ to output values Y, approximated as $\hat{Y} = f(X) = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)$. The discriminative function $f(x)$ is typically trained by minimizing an error function, computed by measuring the discrepancy between the actual output values Y and the estimated outputs $\hat{Y}$. The selection of the error function hinges on the specific task at hand; for instance, cross-entropy often serves as a preferred choice for various classification tasks, such as fault detection and event classification. Over the past decade, a multitude of discriminative approaches have been utilized to tackle a wide range of research challenges in learning spatiotemporal correlations within power systems. In the subsequent subsections, we undertake a comprehensive review of the key data-driven discriminative models employed in this realm.

A. CONVOLUTION NEURAL NETWORKS

Convolutional neural networks (CNNs) play a crucial role in feature representation learning on a variety of tasks, including face recognition [40], [41], traffic scene understanding [42], [43], and medical imaging [44], [45]. As shown in Fig. 1, CNNs are composed of several convolutional layers, and each layer comprises multiple convolutional kernels to extract diverse features from the input layer. Each neuron within a feature map establishes connections with a local region of neurons in the preceding layer, known as its receptive

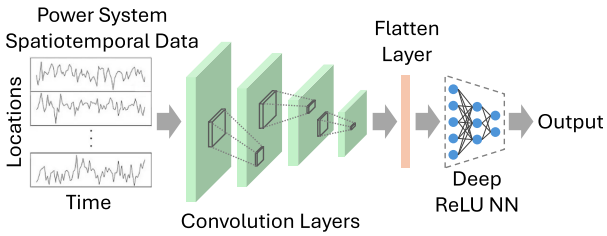

FIGURE 1. General framework of CNN architecture.

field. Through this connectivity pattern, the convolutional layer captures spatial dependencies and hierarchical features present in the input [46]. During operation, each kernel convolves across its respective receptive field, performing a weighted sum of the input values. The resulting convolved feature map is then subjected to an element-wise activation function, typically a rectified linear unit (ReLU), which introduces nonlinearities into the network [47].

The training procedure of CNNs involves optimizing the network's parameters, particularly the kernels (i.e., weights) of the convolutional layers, to minimize the error between the predicted outputs and the ground truth labels. For both classification [48], [49] and regression tasks [50], the error calculation typically involves the use of a loss function that quantifies the disparity between the predicted and actual outputs. In classification tasks, cross-entropy loss is commonly employed, while mean squared error (MSE) is often used for regression tasks. During training, CNNs utilize gradient descent optimization algorithms, such as stochastic gradient descent (SGD) [51], [52] or its variants like Adam, to iteratively update the network's weights in the direction that reduces the loss function. This process involves computing the gradient of the loss function with respect to each weight using backpropagation, which efficiently propagates the error gradient through the network layers [53]. Mathematically, the weight update rule in gradient descent can be expressed as:

$$w_{ij}^{(l)} \leftarrow w_{ij}^{(l)} - \eta \frac{\partial L}{\partial w_{ij}^{(l)}} \tag{1}$$

where $w_{ij}^{(l)}$ represents the weight connecting neuron j in layer l, η is the learning rate controlling the step size of the weight updates, and $\frac{\partial L}{\partial w_{ij}^{(l)}}$ denotes the partial derivative of the loss function L with respect to the weight $w_{ij}^{(l)}$. The chain rule of calculus is used to compute the gradients of the loss function with respect to the weights in each layer, facilitating efficient weight updates. These updates gradually refine the CNN's parameters, enabling it to learn meaningful spatiotemporal representations of the input power system measurements.

CNNs play a pivotal role in capturing spatiotemporal correlations within power systems. Zhang et al. [54] utilized CNN architecture to detect false data injection (FDI) attacks in modern power systems. Their model analyzes spatiotemporal correlations using the Cubature Kalman filter and Gaussian process regression. Subsequently, a deep CNN is designed to delineate the functional relationship between these correlations and the output. Similarly, in [55] and [56], CNN architectures are employed to extract spatiotemporal correlations for precise FDI attack detection in power systems. Furthermore, in [57], a multi-view CNN (MCNN) is devised to combat spoofing cyber attacks in distribution synchrophasors data (DSD). This study involves the extraction of spatiotemporal correlations from DSD's raw frequency measurements, achieved through the fast S transform after removing common components. The resulting correlations are then inputted into the MCNN to identify spoofing attacks.

Numerous studies have utilized CNN architecture effectively for fault detection and event classification applications. For instance, Wang et al. [58] proposed a multiscale deep CNN for analyzing Supervisory Control and Data Acquisition (SCADA) data in wind turbine fault detection. Zhang et al. [59] introduced a hybrid fault diagnosis framework for power grids, combining variational mode decomposition with deep CNN. Similarly, Li et al. [60] presented a knowledge-based CNN with support vector machines (SVM) for transformer fault analysis. Additionally, Hao and Li [61] developed a method to transform power flow data into dynamic images, using CNN for feature extraction. In a related study [62], voltage sag characterization was employed to locate faults in distribution networks using deep CNN and spatiotemporal analysis. The effectiveness of this model was evaluated on the IEEE 13-node system. Additionally, Basumallik et al. [63] presented a spatiotemporal CNN-based classification method for event categorization utilizing Phasor Measurement Unit (PMU) packet data streams. Experimental findings on the IEEE 118-bus system showcased the superior performance of their approach compared to the recurrent neural architecture (i.e., Long Short-Term Memory (LSTM)).

Moreover, CNN architectures and spatiotemporal characteristics are leveraged to address non-intrusive load modeling (NILM) challenges in power systems. For instance, Zhang et al. [64] propose a non-intrusive method for identifying the load state of a distribution network, employing a deep CNN to analyze the non-local spatiotemporal features of load on-off state switching points and temporal features. Liu et al. [65] develop a NILM method for multi-energy coupling (MEC) appliances, establishing a teaching and mutual learning framework using two deep CNNs. The efficacy of this approach is validated across five types of MEC appliances. Wu and Wang [66] introduce a concatenated CNN for capturing and analyzing ultra-short time load signals, evaluating its performance on UK-DALE and BLUED datasets. Additionally, other studies such as those by Moradzadeh et al. [67], Teixeira et al. [68], and Chen et al. [69] demonstrate the effectiveness of CNNs and spatiotemporal features in addressing the NILM problem in power systems.

Multiple studies have explored the efficiency of CCN-based models in the domain of power generation and load forecasting. For instance, Yin and Xie [70] introduced

a Multi-Temporal Spatial-Scale CNN (MTSSCNN) method designed to learn nonlinear spatiotemporal features from load time series data, resulting in accurate load forecasting outcomes. Jeong and Kim [71] introduced a space-time CNN that leverages the location and historical data of photovoltaic (PV) power generators for multiple PV forecasting. Also, Similarly, Feng et al. [72] utilized CNN architecture to provide intra-hour solar forecasting, demonstrating the superior performance of deep CNN models over shallow machine learning models with meteorological predictors. Hu et al. [73] proposed a CNN-based spatiotemporal wind power predictor, capturing space-time features among multiple wind farms for 5-30 minutes ahead of wind forecasting tasks. Additionally, a three-dimensional CNN was employed to automatically extract spatiotemporal features from numerical weather prediction (NWP) data, showcasing the significant effectiveness of CNN in capturing intrinsic spatiotemporal features for wind speed forecasting [74]. In another study by Hong and Satriani [75], a day-ahead image-based spatiotemporal wind speed forecasting framework was proposed, utilizing Taguchi's orthogonal array to design a robust two-dimensional CNN for wind speed forecasting at an offshore wind farm. Similarly, other recent studies [50], [76], [77] have highlighted the effectiveness of CNN models in renewable energy forecasting tasks.

B. RECURRENT NEURAL NETWORKS

Recurrent Neural Networks (RNNs) constitute a class of artificial neural networks particularly adept at modeling sequential data due to their inherent capacity to capture temporal dependencies within sequences. Unlike traditional deep ReLU neural networks, RNNs possess recurrent connections that allow them to maintain a memory of past inputs, thereby enabling the processing of sequences of arbitrary length (as shown in Fig. 2). This memory mechanism enables RNNs to dynamically adapt their internal states based on the current input as well as the information stored from previous time steps, making them well-suited for tasks involving time-series data or sequences with temporal dependencies. Through recurrent connections and feedback loops, RNNs can effectively model the temporal dynamics of sequential data, making them invaluable in applications such as natural language processing, speech recognition, time series prediction, and, notably, spatiotemporal modeling in power systems.

Expanding on the foundation of RNNs, specialized variants like Gated Recurrent Units (GRUs), LSTM architecture, and Reservoir Computing (RC) models have emerged as powerful tools in spatiotemporal modeling within power systems. These models tackle the limitations encountered by standard RNNs, notably the vanishing gradient problem [78], by integrating gating mechanisms that control the information flow within the network.

GRUs achieve this through two main gates: the update gate z_t and the reset gate r_t. At each time step t, the update gate

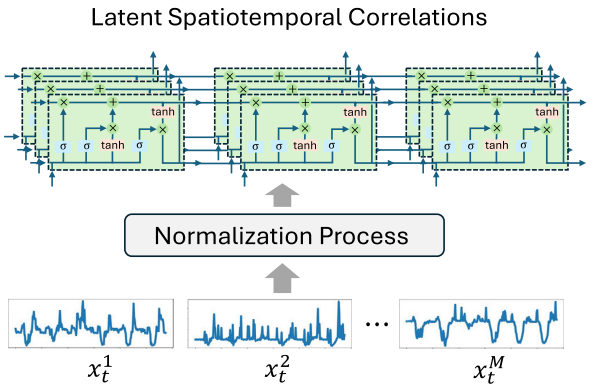

FIGURE 2. General framework of RNN architecture. Here, x_t^i shows the power systems measurements at time step t.

controls the extent to which the previous memory should be updated, while the reset gate determines how much of the past information should be discarded. The hidden state h_t is updated using these gates, and the candidate activation $\tilde{h}_t$ is computed based on the input and the previous hidden state using:

$$\begin{aligned} z_t &= \sigma(W_z \cdot [h_{t-1}, x_t]) \\ r_t &= \sigma(W_r \cdot [h_{t-1}, x_t]) \\ \tilde{h}_t &= \tanh(W \cdot [r_t \odot h_{t-1}, x_t]) \\ h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \end{aligned} \tag{2}$$

where σ is the sigmoid activation function, W_z, W_r, and W are weight matrices, x_t is the input at time step t, and $\odot$ represents element-wise multiplication.

In contrast, LSTM networks [79], [80] introduce three main gates: the input gate i_t, the forget gate f_t, and the output gate o_t. Additionally, they maintain a cell state c_t alongside the hidden state h_t. These gates and the cell state are updated as:

$$\begin{aligned} i_t &= \sigma(W_i \cdot [h_{t-1}, x_t]) \\ f_t &= \sigma(W_f \cdot [h_{t-1}, x_t]) \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t]) \\ \tilde{c}_t &= \tanh(W_c \cdot [h_{t-1}, x_t]) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \tag{3}$$

Here, W_i, W_f, W_o, and W_c are weight matrices, and $\tilde{c}_t$ represents the candidate cell state. Eq. (3) shows how information is stored, updated, and retrieved over time in LSTM units, allowing them to capture and utilize long-range dependencies more effectively than traditional RNNs or GRUs.

The RC architecture [81], [82] represents a distinctive approach within the realm of RNNs that unlike LSTM and GRU models, which involve complex internal gating mechanisms, leverages a fixed, high-dimensional, and randomly connected recurrent network called the reservoir. The primary

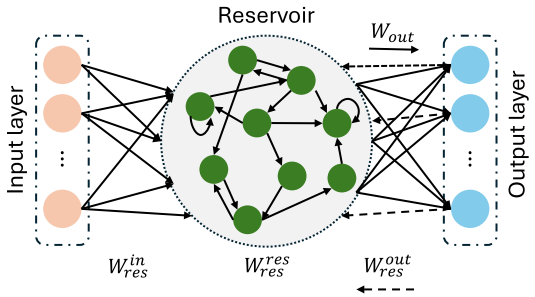

FIGURE 3. General framework of reservoir computing architecture.

innovation of RC lies in the separation of the dynamic reservoir from the output layer, where only the weights of the output layer are trained, significantly simplifying the training process. This decoupling allows RC models to efficiently capture temporal dependencies and dynamic behavior of input sequences without the computational burden associated with training the entire network, as seen in LSTM and GRU architectures. As shown in Fig. 3, the RC architecture typically comprises three main components: the input layer, the reservoir, and the output layer. The input layer maps the input signal u_t to the reservoir state space. The reservoir consists of a large number of sparsely connected neurons with fixed, randomly initialized weights W_{in} and W_{res}. The dynamic state of the reservoir x_t evolves according to the nonlinear transformation:

$$\begin{aligned} x_{t+1} &= \tanh(W_{\text{res}}^{\text{in}} u_t + W_{\text{res}}^{\text{res}} x_t + W_{\text{res}}^{\text{out}} y_{t-1}) \\ y_t &= W_{\text{out}} x_t \end{aligned} \tag{4}$$

where tanh is the activation function, and y_t is the output layer. Here, $W_{\text{res}}^{\text{in}}$ is the weight matrix that connects the input nodes to the reservoir layer, $W_{\text{res}}^{\text{res}}$ denotes the weights in the reservoir layer, and $W_{\text{res}}^{\text{out}}$ shows the matrix of feedback weights which are used to incorporate the output of the previous time step in the current state. Also, W_{out} represents the trainable weights that maps the latent high-dimensional reservoir states to the output layer.

The training procedure of the RC seeks to solve the optimization problem $\arg\min \sum_{t=1}^{T} \left\| y_t^{\text{target}} - W_{\text{out}} x_t \right\|^2$ by adjusting the output weights W_{out}. The optimal solution for this optimization problem is obtained as:

$$W_{\text{out}} = Y^{\text{target}} (XX^{\top})^{-1} X^{\top} \tag{5}$$

where Y^{target} is the given target observations, and X is the designed matrix with x_t in the tth column. Here, λ is positive parameter regularization determined by the validation set.

Due to the remarkable generalization capabilities inherent in RNNs, numerous research endeavors have devised RNN-based frameworks to address various power system challenges, spanning from behind-the-meter (BtM) disaggregation [83], [84], [85], to cyber-physical attack detection [86], [87], [88], [89], [90] and sustainable energy and load forecasting [91], [92], [93], [94], [95], [96], [97], [98], [99]. For instance, Razavi et al. [83] proposed an LSTM-based multi-input single-output model that leverages historical data from individual households as simultaneous inputs to forecast target time series. This approach effectively captures spatial correlations among residential units indirectly. Similarly, Zhang et al. [84] introduced a method for predicting BtM PV power generation, employing an attention-LSTM neural network and transfer learning. Categorizing weather data into four types, they identify key factors influencing PV power generation and utilize LSTM to capture temporal patterns. Furthermore, Zaboli et al. [85] presented a data-driven framework integrating LSTM and stacked autoencoders (SAEs) for forecasting residential load profiles, considering PV, battery energy storage systems, and electric vehicle loads. Their approach involves minute-level data extraction, facilitating accurate predictions, and enhancing the understanding of load dynamics.

In the realm of power system cyber security and fault identification, Musleh et al. [86] present a novel method aimed at detecting FDI attacks in distribution systems. Their approach employs a spatiotemporal learning algorithm, utilizing an LSTM autoencoder to grasp normal system behaviors and identify potential FDI attacks through analysis of measurement errors. Additionally, James et al. [87] introduce a machine learning framework designed for FDI attack detection in AC state estimation. This method leverages wavelet transform and GRU neural networks to examine temporal correlations within estimated system states. Experimental validation conducted on IEEE 118- and 300-bus power systems underscores the efficacy of the proposed models, demonstrating satisfactory attack detection accuracy. Also, Yadav and Pradhan [88] introduce a novel approach that PCA with sequential deep learning to classify cyber-induced outages and natural events in power systems. Their methodology relies on PCA to extract distinctive spatiotemporal progression patterns, which are subsequently classified using an ensemble LSTM network. Moreover, Zhang et al. [100] introduced a novel RNN-based deep autoencoder aimed at detecting wind turbine faults through the analysis of spatiotemporal SCADA data. Their framework comprises a deep GRU autoencoder designed to capture spatiotemporal features, followed by the development of a support vector regression model optimized by gray wolf optimization (GWO) to effectively classify faults. Furthermore, Kim et al. [90] designed an Echo State Network (ESN) as an efficient RC architecture for detecting the FDI attacks in smart grids. Their results demonstrated that the ESN model achieved comparable attack detection accuracy to LSTM and GRU models but with significantly reduced training times when tested on a three-bus power system dataset.

In forecasting applications, Park et al. [91] introduce an ensemble-based RNN framework to estimate generation power at target PV sites. Their approach leverages historical samples from other PV sites obtained through clustering and distance-based sampling. Similarly, Jahangir et al. [94] present a method addressing uncertainty in renewable

energy and electricity price data. Their approach combines micro-clustering with bidirectional LSTM networks, categorizing data hourly to allocate distinct forecasting units and enabling investigation of past and future data. Also, Fu et al. [92] propose a multi-head self-attention network to capture spatial correlations among wind farms, while a sequence-to-sequence model captures temporal dependencies of wind power time series. Moreover, Hu et al. [101] proposed a robust ESN-based model enhanced by a quality-driven loss function for wind power prediction interval to quantify the prediction uncertainty. By evaluating their proposed model on a real wind power dataset, the authors showed the proposed model can reduce the mean prediction interval width by up to 16.69% and save up to 5 times computation time compared to LSTM. Similarly, the research study [99] introduced an RC-based model equipped with a multi-objective GWO method for deterministic and probabilistic wind power prediction. The numerical results of this study, with respect to deterministic and probabilistic metrics, showed higher performance of the proposed RC-based model than state-of-the-art methodologies.

Furthermore, in [95], a spatiotemporal PV output probability prediction method is introduced, leveraging appearance similarity updating (ASU) in conjunction with LSTM architecture. The ASU method is utilized to quantitatively assess the lead-lag relationship between the forecasting errors of each power station within a local small PV cluster. Subsequently, by incorporating the output of ASU analysis along with historical weather data, LSTM enables accurate prediction results with a 15-minute horizon for datasets originating from Jilin and Inner Mongolia. In [97], a fast variant of RC architectures with smaller tunable parameters is proposed for residential energy demand forecasting. The proposed model is evaluated on four different energy demand data sets for a 24-hour energy demand prediction with a granularity of 15 min. The results show that the proposed fast RC model improved the traditional RC architecture as well as the LSTM model in terms of training time metrics. Furthermore, Fujimoto et al. [98] introduced a deep ESN architecture for edge computing and proposed an efficient online learning scheme to keep prediction models up-to-date. Using real-world data from over 500 households, the framework demonstrated high accuracy with significantly reduced computational costs. This approach is well-suited for short-term residential demand forecasting and can enhance demand-side energy management.

C. CNN-RNN

CNN-RNN models offer a robust approach for capturing spatiotemporal features in data by seamlessly integrating CNN and RNN architectures. Mathematically, the CNN component processes the input power system measurements X using convolutional operations to extract spatial features, generating feature maps $F = \text{CNN}(X)$ that encode spatial information across the input space. These feature maps are then seamlessly integrated into the RNN component, where they are processed sequentially over time to capture temporal dependencies. Let H_t denote the hidden state of the RNN at time step t, and θ represent the parameters of the RNN. The RNN component processes the feature maps F_t over time as $H_t = \text{RNN}(F_t, H_{t-1}; \theta)$ where H_t represents the hidden state at time step t, F_t denotes the convolution feature maps at time step t, and H_{t-1} is the hidden state from the previous time step. By jointly leveraging spatial and temporal information, CNN-RNN models excel in learning how spatial features evolve and interact over time, effectively capturing complex spatiotemporal correlations within the data. Fig. 4 illustrates an example of a CNN-LSTM architecture for short-term wind speed prediction for an $N \times N$ array of wind turbines in a local area. As depicted in the figure, the correlations among the wind turbines are represented in an array, which is subsequently fed into the CNN-LSTM model at each time step t. This process extracts informative spatiotemporal features for the wind speed prediction task.

Multiple recent studies have considered CNN-RNN frameworks for spatiotemporal sustainable energy and load forecasting. For instance, in [102], a short-term spatiotemporal load forecasting framework is introduced, leveraging spatial auto-correlation (SAC) and CNN-LSTM to capture spatiotemporal characteristics of load time series sub-signals derived from discrete wavelet transform. Experimental results conducted on a dataset sourced from power substations in Tehran, Iran, demonstrate the efficiency of the proposed approach compared to recent load forecasting methodologies. Chai et al. [103] exploit the CNN-LSTM model to extract the spatiotemporal features of a synthetic PV measurement dataset recorded in 56 locations in the USA. Similarly, CNN-LSTM is employed in [104] and [105] for PV power generation forecasting task. In that work, the authors show the superior performance of CNN-LSTM compared with individual CNN and LSTM in capturing spatiotemporal correlation of PV power generation sites. More recently, Yang et al. [106] integrated a multi-scale convolutional neural network (MSCNN) with an ESN architecture for solar irradiance prediction. Their proposed model extracts task-relevant features from the 1-D time series data of multiple locations using the MSCNN. These extracted features are then fed into the ESN model to obtain a high-dimensional state space. The authors demonstrated the superiority of their spatiotemporal model over the LSTM network in terms of performance and accuracy. By integrating Rough set theory into CNN-LSTM architecture, Saffari et al. [107] propose the end-to-end convolutional rough LSTM model to extract spatiotemporal correlations among 20×20 array of wind turbines in North Carolina USA. In [108] a CNN is employed to extract spatial features of 49 wind farms with cloud point distribution, and an attention-based bidirectional LSTM architecture is devised to learn temporal correlations. Also, Chen et al. [109] proposed a multifactor CNN-LSTM model to extract spatial features relationship between the meteorological factors of wind sites in a local area in Texas, USA.

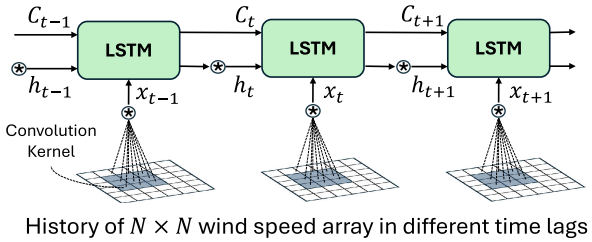

FIGURE 4. General framework for CNN-LSTM architecture to extract spatiotemporal features of wind turbines. ⊛ denotes the convolution operation.

Moreover, considerable research has delved into the CNN-RNN model's efficacy in cyber-attack and anomaly detection, as well as voltage stability assessment within power systems. For example, Kong et al. [110] utilize a CNN-GRU model to accurately discern the health state of wind turbines through spatiotemporal analysis of SCADA data. In another study, D'Angelo and Palmieri [111] introduce a SAE-based CNN-RNN framework for detecting cyber-attacks in power control systems. Leveraging two SAEs (i.e., CNN-SAE and RNN-SAE), this model adeptly captures spatial and temporal dependencies within the data, exhibiting near-perfect classification outcomes across both binary and multi-class scenarios. Additionally, Ruan et al. [112] propose a spatiotemporally coordinated cyber attack strategy targeting meteorological data used in renewable energy forecasting. By designing white- and black-box attack scenarios, they assess the performance of CNN-LSTM for energy forecasting in the presence of these attacks. Simulation results on the IEEE 39-bus benchmark underscore the substantial economic losses and system collapse induced by the proposed cyberattack strategy. Additionally, Ma et al. [113] introduced a two-layer algorithm for fault identification and localization using spatiotemporal PMU data. Initially, they transform the PMU data into 2D images utilizing Gramian angular field and Short-Time Fourier Transform (STFT). Subsequently, a CNN-LSTM architecture is utilized to extract spatiotemporal features and classify faults on the IEEE Standard New England 39-bus Test System. Additionally, to maintain the power system in a secure state, Adhikari et al. [114] propose a CNN-LSTM model for real-time assessment of short-term voltage stability. Their model harnesses the benefits of transfer learning, enabling effective operation with only a limited number of labeled samples. Experimental outcomes on IEEE 9-bus and New England 39-bus test systems demonstrate the superiority of the proposed model for online applications.

In addition to energy forecasting and cyber attack anomaly detection tasks, numerous studies have demonstrated the efficacy of CNN-RNN models in NILM. For instance, Kaselimi et al. [115] propose a deep learning architecture that integrates CNN with a recurrent property to effectively model the spatial and temporal interdependencies of power signals for energy disaggregation. By incorporating multiple channels representing various power-related variables, such as active, reactive, apparent power, and current, the model achieves enhanced performance and faster convergence times compared to existing approaches. Similarly, Zhou et al. [116] combine CNN, LSTM, and random forest models for spatiotemporal NILM. They convert one-dimensional load data into a two-dimensional matrix, extract spatial features using CNN, learn temporal dependencies with LSTM, and employ random forest to decode spatiotemporal features and output labels. Experimental validation on two datasets confirms the effectiveness of the proposed method in achieving accurate NILM using low-frequency load data. Additionally, the authors of [116] integrate a CNN-LSTM with random forest architectures to extract task-relevant features from low-frequency load data for performing NILM. Moreover, Wang et al. [117] designed an adaptive sliding window to prepare data from appliance operation characteristics and fuse shallow CNN with a two-layer nested LSTM to perform spatiotemporal load decomposition tasks.

D. GRAPH CNN

Graph-based methodologies are essential for comprehensively modeling the spatiotemporal intricacies of input time series data within power systems [118], [119]. By conceptualizing the relationships and interconnections among diverse power system components as nodes and edges within a graph structure, these approaches facilitate the capturing of both spatial and temporal dependencies ingrained within the data. Here, the dynamic graph $\mathcal{G} = \{G_t\}_{t=\bar{t}}^{\bar{t}+\tau}$ consists of multiple graph snapshots corresponding to time interval $[\bar{t}, \bar{t}+\tau]$. Each graph snapshot $G_t = (\mathcal{V}_t, \mathcal{E}_t)$ encapsulates the intricate relationships among entities in power systems. Here, $\mathcal{V}_t$ denotes the set of nodes, and $\mathcal{E}_t$ signifies the set of edges connecting nodes within $\mathcal{V}_t$ at time step t. Each node v_i is connected by edges e_{ij} originating from node v_j. The neighborhood of a given node v_i is expressed as $N_{v_i} = \{u | u \in \mathcal{V}, (v_i, u) \in \mathcal{E}_t\}$. Typically, such data is characterized by a nodal feature matrix X_t with dimensions $n \times f$ and an adjacency matrix A_t, sized $n \times n$, reflects the presence of edges: $a_{ij} = 0$ if $e_{ij} \notin \mathcal{E}_t$ and $a_{ij} = 1$ if $e_{ij} \in \mathcal{E}_t$. Dynamic graphs denoted as $G_t = (\mathcal{V}_t, \mathcal{E}_t, X_t)$, crucially employed in modeling spatiotemporal characteristics within power system analysis, the features of nodes dynamically evolve over time, enhancing the modeling capacity and analytical depth of such methodologies.

Spectral graph convolutional neural networks (GCNNs) [120] are a potent tool employed for extracting spatial features from dynamic graph inputs $\mathcal{G}_t$. Operating within the spectral domain, GCNNs utilize the eigenvalues and eigenvectors of the graph Laplacian matrix L_t, which encapsulates the structural information of the dynamic graph. As shown in Fig. 5, this spectral approach enables GCNNs to transform the input dynamic graph features X_t into the spectral domain by applying multiple convolution layers and obtain task-relevant spatiotemporal features of the input dynamic graph. Mathematically, the spectral graph convolution of

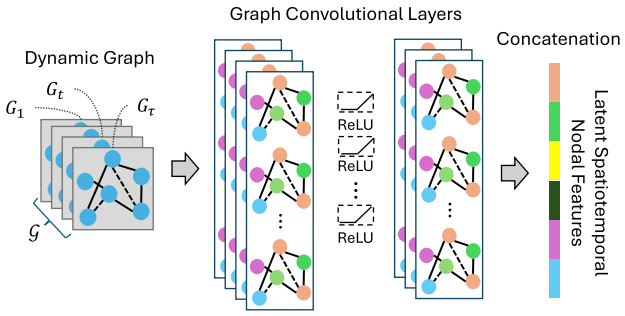

FIGURE 5. General framework of GCNN architecture.

graph G_t at time step t is computed as $\mathcal{F}_\theta * X_t = V_t \mathcal{F}_\theta V_t^T X_t = I_n - D_t^{-\frac{1}{2}} W_t D_t^{-\frac{1}{2}}$, where the parameter vector θ represents the convolutional filter parameters $\mathcal{F}_\theta = \text{diag}(\theta)$ in the frequency domain. Here, the matrix V_t contains the eigenvectors of the normalized Laplacian matrix L_t. Additionally, the expression $D_t^{-\frac{1}{2}} W_t D_t^{-\frac{1}{2}}$ denotes a normalized version of the weighted adjacency matrix W_t using the degree matrix D_t. Due to computational complexity in producing $\mathcal{F}_\theta$, Chebyshev polynomials $\{T_j\}_{j=0}^{J}$ are commonly employed to approximate the underlying filters. This approximation is expressed as $\mathcal{F}_\alpha \approx \sum_{j=0}^{J} \alpha_j T_j \left(\frac{2}{\omega_{max}} L_t - I_n\right)$, where α_j represents the j-th Chebyshev coefficient, and ω_{max} denotes the maximum eigenvalue of L. Simplifying the convolution operation with $J = 1$ and $\omega_{max} = 2$, and assuming equal and opposite magnitudes for α_0 and α_1, the simplified form of spectral GCNN can be expressed as $\mathcal{F}_\theta * X_t \approx \alpha_0 \left(I_n + D_t^{-\frac{1}{2}} W_t D_t^{-\frac{1}{2}}\right) X_t$. This streamlined approach enhances the computational efficiency of spectral graph convolution, offering a valuable framework for analyzing dynamic graph data in various applications.

The spatiotemporal feature extraction capabilities of GCNN architectures have led to their widespread adoption in addressing diverse challenges within modern power systems. The research study [121] utilized GCNN frameworks for spatiotemporal load forecasting of a local area in Hangzhou, China. Zhang et al. [122] propose a novel method for short-term solar power forecasting, emphasizing the optimization of graph structures to capture spatio-temporal correlations among neighboring PV sites. Through an analysis of geographical and weather factors, key PV sites are selected to minimize data redundancy and enhance forecasting accuracy. Leveraging complex network theory, a unique index evaluates the connectivity of the graph structure, enhancing the predictive capability of the GCNN model. Evaluation on a dataset from 20 locations in Jilin, China, demonstrates the superiority of the proposed model over CNN-LSTM and LSTM architectures. Similarly, Karimi et al. [123] propose a spatiotemporal graph neural network for forecasting PV power in large-scale PV systems. More recently, Liu et al. [124] introduce a spatiotemporal approach for ultra-short-term wind farm cluster power forecasting by analyzing fluctuation processes and partitioning wind farm cluster power based on distinct patterns. Their proposed model, employing a spatiotemporal graph neural network for pattern prediction, outperforms benchmarks on real wind farm cluster power datasets. Additionally, in [125], a wind speed predictor is developed that captures the spatial relationship of wind farms through Granger causality testing. It employs a multi-scale graph convolutional approach combined with a temporal convolution layer to extract the most influential spatiotemporal features, thereby enhancing the accuracy of wind speed forecasting tasks with a 4-hour horizon. Also, Dong et al. [126] develop a spatiotemporal GCN for wind power forecasting, utilizing a directed graph convolutional structure and a temporal convolutional network (TCN) to effectively learn spatiotemporal correlations. The directed GCNN layer enables the characterization of asymmetric spatial correlations, while the TCN layer extracts temporal features. By leveraging historical data from 15 wind farms in Australia, the model demonstrates superior accuracy compared to existing methods.

Several recent studies have proposed GCNN-based frameworks for intelligent fault identification and voltage stability assessment tasks. For instance, Tong et al. [127] developed a spatiotemporal GCNN model aimed at classifying transient faults in power transmission lines. By incorporating both graph structures and bus voltage signals, the model enables rapid fault classification by explicitly considering spatial information within sampling sequences, resulting in enhanced feature extraction capabilities. Similarly, Hu et al. [128] introduced a fault diagnostic model tailored for distribution systems. Their approach utilizes deep GCNN architecture alongside spatiotemporal convolutional blocks to extract waveform features effectively. Experimental validation conducted on IEEE 33-bus and IEEE 37-bus test systems showcased superior performance compared to regular GCNN and PCA-SVM methods, particularly under diverse fault conditions and various interference factors. Additionally, Nguyen et al. [129] proposed a comprehensive framework integrating 1D-CNN and GCNN for extracting spatiotemporal correlations from voltage measurements in microgrids. Their approach addresses fault detection, type and phase classification, and fault location. Evaluation against traditional ANN structures on the Potsdam 13-bus microgrid dataset revealed notably higher accuracy levels. In [130], an attention-based graph convolution model is introduced for monitoring the pre-fault transient stability of power systems. Experimental results conducted on IEEE 39- and IEEE 300-bus systems highlight the superiority of the devised framework, attributed to its hierarchical pooling structure and spectral unsupervised loss.

Furthermore, GCNN architectures are considered for real-time solving of non-convex OPF optimization problems in power systems. Li et al. [131] utilize graph-based neural networks to extract attention matrices to tackle OPF challenges in renewable power systems. Leveraging graph attention neural networks, the approach extracts attention

matrices for nodes and links within the power grid, effectively discerning correlations influenced by weather inputs. Comprehensive evaluations across two European renewable power system scenarios validate the method's efficacy, demonstrating superior performance compared to existing data-driven techniques. By taking advantage of GCNN and message passing technique interface, Mahto et al. [132] proposed a spatiotemporal framework for OPF solutions. In message passing, information is gathered from neighboring nodes, aggregating and updating the feature matrix. Simulation results on an IEEE 33-bus distribution network validate the superior performance of the proposed graph-based model compared with other recent deep neural networks (DNNs).

Moreover, GCNNs have been employed to tackle cyber attack detection in power systems. Qu et al. [133] focus on addressing the threat of dummy data injection attacks (DDIAs) to power system security. They highlight the challenge faced by existing detection methods due to the minimal spatiotemporal correlation between injected malicious data and legitimate data. To overcome this, they introduce temporal and spatial attention matrices aimed at capturing spatiotemporal correlations within attacks. By leveraging GCNN, they enhance dynamic correlation mining capability and computational efficiency. In a related study, Wu et al. [134] propose a spatiotemporal framework for power grid cyber-attack detection and localization, utilizing GCNN architectures in the complex-value domain. They demonstrate that complex-valued GCNNs offer higher stability in the face of perturbations in the underlying power system graph and achieve higher FDI detection accuracy.

E. GRAPH CONVOLUTION RNN

In order to address the issue of long-term dependencies in graph-structured data and alleviate the limitations of GCNN, there is a growing interest in incorporating gate mechanisms from RNNs, such as GRUs and LSTMs, into the GCNN architectures. These models extend the capabilities of traditional RNNs (i.e., GRU and LSTM) to handle sequential data associated with graphs. Graph Convolution GRU (GCGRU) extends traditional GRU networks to operate on graph-structured data, allowing for sequential modeling of graph data while capturing both temporal dependencies and the structural information encoded in the graph. Given x_t^i represents the feature vector associated with node i at each time step t, GCGRUs update the hidden state $h_{(t)}^i$ of each node i based on the features of its neighboring nodes. The update gate $z_{(t)}^i$ and reset gate $r_{(t)}^i$ are computed as sigmoid activations of linear combinations of the input features x_t^i and the previous hidden state $h_{(t-1)}^i$, while the candidate hidden state $\tilde{h}_{(t)}^i$ is obtained using the tanh activation function. The final hidden state $h_{(t)}^i$ is updated by blending the candidate hidden state with the previous hidden state based on the update gate. Mathematically, this can be expressed as:

$$
\begin{aligned}
z_t^i &= \sigma\left(W_{xz} \circledast x_t^i + W_{hz} \circledast h_{t-1}^i + b_z\right) \\
r_t^i &= \sigma\left(W_{xr} \circledast x_t^i + W_{hr} \circledast h_{t-1}^i + b_r\right) \\
\tilde{h}_t^i &= \tanh\left(W_{xh} \circledast x_t^i + W_{hh} \circledast (r_t^i \odot h_{t-1}^i) + b_h\right) \\
h_t^i &= (1 - z_t^i) \odot h_{t-1}^i + z_t^i \odot \tilde{h}_t^i
\end{aligned} \tag{6}
$$

where $\circledast$ denotes a graph convolution operation with 2D-convolution kernels W_{xz}, W_{hz}, W_{xr}, W_{hr}, W_{xh} and W_{hh}. Here, b_z, b_r, and b_h are biases in convolution operation. This formulation enables GCGRUs to capture complex spatiotemporal dependencies within graphs, making them suitable for various tasks related to power systems. Similar to GCGRU, Graph Convolution LSTM (GCLSTM) is another recurrent graph-based architecture that is an extension of the LSTM model for analyzing dynamic graphs. At each time step t, GCLSTMs update the hidden state h_t^i and memory cell c_t^i of each node i based on the features of its neighboring nodes. This update involves the computation of input, forget, and output gates i_t^i, f_t^i, o_t^i, and the candidate memory cell content $\tilde{c}_t^i$, followed by updating the memory cell and hidden state accordingly. Mathematically, this can be expressed as:

$$
\begin{aligned}
i_t^i &= \sigma\left(W_{xi} \circledast x_t^i + W_{hi} \circledast h_{t-1}^i + b_i\right) \\
f_t^i &= \sigma\left(W_{xf} \circledast x_t^i + W_{hf} \circledast h_{t-1}^i + b_f\right) \\
o_t^i &= \sigma\left(W_{xo} \circledast x_t^i + W_{ho} \circledast h_{t-1}^i + b_o\right) \\
\tilde{c}_t^i &= \tanh\left(W_{xc} \circledast x_t^i + W_{hc} \circledast h_{t-1}^i + b_c\right) \\
c_t^i &= f_t^i \odot c_{t-1}^i + i_t^i \odot \tilde{c}_t^i \\
h_t^i &= o_t^i \odot \tanh(c_t^i)
\end{aligned} \tag{7}
$$

where $W_{xi}, W_{xf}, W_{xo}, W_{xc}$ are 2D-convolution kernels (weights) that govern the influence of the input features x_t^i on the input gate i_t^i, forget gate f_t^i, output gate o_t^i, and the candidate memory cell content $\tilde{c}_t^i$, respectively, for node i at time step t. These weights control how much importance is assigned to the input features when updating the cell state and hidden state of the LSTM cell. Also, $W_{hi}, W_{hf}, W_{ho}, W_{hc}$ are weight matrices that govern the influence of the previous hidden state h_{t-1}^i on the input gate i_t^i, forget gate f_t^i, output gate o_t^i, and the candidate memory cell content $\tilde{c}_t^i$, respectively, for node i at time step t. These kernels control how much importance is assigned to the previous hidden state when updating the cell state and hidden state of the LSTM cell. Additionally, b_i, b_f, b_o, b_c are bias terms added to the input gate i_t^i, forget gate f_t^i, output gate o_t^i, and the candidate memory cell content $\tilde{c}_t^i$, respectively, for node i at time step t. The biases allow the model to learn the overall effect of each gate independent of the input and previous hidden state. Fig. 6 shows an example of a GCGRU model where the historical power system data between time steps $\bar{t} \in [t-\tau, t-1]$ are modeled as dynamic graphs $\{G_{\bar{t}}\}_{\bar{t}=t-\tau}^{\bar{t}=t-1}$. At each time step $\bar{t}$, the input graph data is fed into a spectral GCNN [120] to extract the spatial correlations. These spatial features are then input into a GRU to extract the temporal correlations. Finally, the spatiotemporal features are computed by averaging the

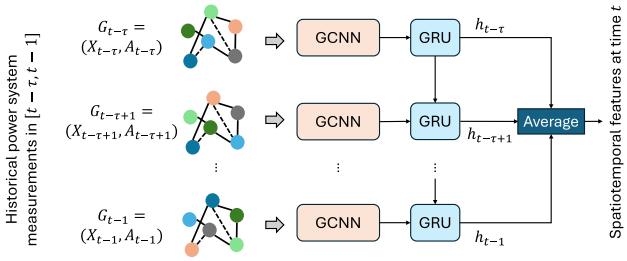

FIGURE 6. General framework for a GCGRU model for extracting spatiotemporal features.

hidden states of the GRUs across the different time steps in $[t-\tau, t-1]$.

Several recent studies have utilized GCRNNs to address sustainable energy generation and load forecasting challenges. For example, Khodayar and Wang [135] introduced a GCLSTM-based model for wind speed forecasting, employing a scalable graph convolutional deep learning architecture to effectively capture spatial and temporal wind features. Integration of Rough Set Theory [136], [137] enhances the model's robustness, as demonstrated by simulation results showcasing superior performance over shallow architectures and state-of-the-art models in wind speed prediction accuracy. Similarly, Jiao et al. [138] applied the GCLSTM model to solar irradiance forecasting in distributed PV systems, leveraging a GCNN to extract critical features and an LSTM to capture temporal correlations. Furthermore, to accurately forecast load consumption, Arastehfar et al. [139] introduced a graph convolution LSTM to extract spatiotemporal information from users with similar consumption patterns. Experimental results conducted on the Low Carbon London [140] and Customer Behavior Trials [141] datasets showcase the superior performance of the proposed model compared to traditional LSTM and deep ReLU neural networks.

In the spatiotemporal BtM disaggregation task, Khodayar and Wang [142] introduced a novel approach employing a spatiotemporal GCLSTM autoencoder to capture intricate space-time correlations among residential units. This was further augmented by a spatiotemporal graph dictionary learning technique, effectively enhancing the sparsity of latent spatiotemporal correlations. Similarly, Saffari et al. [143] harnessed the power of dynamic graph modeling and deep DL models to discern the most crucial spatiotemporal features. Diverging from the methodology of [142], they employed a capsule neural network (CapsNet) rather than a deep ReLU neural network for decoding the latent sparse representation and estimating load and PV power generation values at each time step t. Experimental findings, based on real-world energy disaggregation datasets, underscore the superiority of these spatiotemporal GCLSTM models, enhanced with DL and CapsNet, over recent deep learning baselines. Another noteworthy contribution is the geometric GCLSTM method for BtM PV forecasting, as proposed by the research study [144], leveraging insights derived from a limited number of sensors (i.e., pyranometers, satellite irradiation images, and power meters) within a distribution system. Additionally, Saffari et al. [145] introduced an attention-based GCGRU method, integrating deep extreme learning machine (ELM) [146] as the decoder for load and PV estimation.

Moreover, recent research has harnessed the capabilities of GCRNN models in addressing diverse classification problems within power systems analysis. For instance, Ahmed et al. [147] propose a deep spatiotemporal graph learning method for cyber-power event identification and localization at the distribution level. The introduced GCGRU autoencoder model utilizes physical measurements from PMUs and cyber data from communication networks. Experimental results on two different test systems, spanning multiple cyber-power system events, vividly demonstrate the high classification accuracy of the proposed model. Additionally, a spatiotemporal fault diagnostic framework is developed in [148], where a GCLSTM is engineered to extract spatiotemporal features from voltage measurement units installed at critical buses. The authors evaluate their model's performance on IEEE 123-node systems to showcase its superiority over GCNN, LSTM, and CNN architectures. Furthermore, Liu et al. [149] have devised a spatiotemporal framework by combining GCNN and GRU architectures for transient stability assessment of power grids. The proposed gating spatiotemporal graph neural network (GSTGNN), augmented with a weighted cross-entropy loss function, is employed to extract and fuse crucial spatiotemporal features, effectively addressing the voltage assessment task at hand. Furthermore, Presekal et al. [150] present a spatiotemporal framework that integrates GLSTM for feature analysis and a deep CNN for time-series classification-based online cyber attack situational awareness aimed at enhancing power grid resilience. Simulation results underscore the model's significant ability to identify active attack locations compared to recent deep learning models.

F. PERFORMANCE COMPARISON AND DISCUSSION

In this section, we compare deep discriminative approaches applied to various classification and regression tasks within spatiotemporal power system analysis. We evaluate different methodologies for classification applications based on precision, recall, and F1 scores. For regression-based applications, we use root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) as key performance metrics. Tables 1 and 2 present the quantitative outcomes achieved by different spatiotemporal discriminative architectures when addressing classification- and regression-based challenges in power system analysis, respectively. Across diverse applications, both CNN-based and RNN-based methods demonstrate comparable accuracy performances, as illustrated in the tables. For example, in the task of FDI attack detection within the power systems, the DCNN-CKF [54] model, which integrates spatiotemporal correlations through a Cubature Kalman filter before training

a deep CNN model, achieves an accuracy of 68.233%. Similarly, the LSTM-AE [86] model exhibits a slightly superior F1 score of 69.687% on the same dataset. Comparable performances are observed when comparing results obtained from CNN- and RNN-based models. Notably, CNN-RNN hybrid models display enhanced accuracy compared to their individual counterparts. For instance, in the NILM application on the UK-DALE dataset, CNN-GRU [110] surpasses GRU-GWO-SVR [100] and VMD-CNN [59] by margins of 6.365% and 3.817%, respectively. Similarly, in PMU event classification and voltage stability assessment, models such as STFT-CNN-LSTM [113] and TempCNN-LSTM [114] outperform STCNN [63] and FTSBA [151] methods by margins of 3.976% and 4.889% in terms of F1 score, respectively. This superior performance of CNN-RNN architecture compared to individual CNN and RNN architectures is attributed to their ability to simultaneously capture complex spatiotemporal correlations by leveraging the strengths of both CNN and RNN frameworks.

As illustrated in Table 1, graph-based spatiotemporal approaches demonstrate superior performance compared to other baseline methods. For instance, in voltage stability assessment on the New England 39-bus system with 10 generators, AH-DAPE [130] outperforms TempCNN-LSTM [114] and FTSBA [151] by margins of 3.689% and 8.578%, respectively, in terms of F1 score. Similarly, in the fault identification within the IEEE 123-bus system, Cplx-STGCN [134] outperforms CNN-RNN-SAE [111] by 3.242% in terms of the F1 score metric. This higher classification performance is attributed to the utilization of graph data structures and spectral graph convolutions, which enable the modeling of intricate spatiotemporal correlations.

Furthermore, the GCRNN models enhance the classification accuracy of GCNN architectures across various spatiotemporal tasks. Notably, as depicted in the table, the AttG-BDGNets [152], leveraging an attention-based graph LSTM model, outperforms Spectral GCN [153] by 4.125%. Similarly, RT-CPDLC [147] improves the F1 score of GCN by 4.857% in the PMU event classification task. This heightened generalization capacity of GRCNN over GCNN stems from its utilization of both recurrent and convolutional operations for mining dynamic graphs derived from power system measurements.

Table 2 compares the performance of recent deep discriminative architectures for regression-based tasks, including sustainable energy and electrical load forecasting. As shown in the table, both CNN- and RNN-based models show comparable forecasting accuracy. For instance, STAN [92] that captures spatiotemporal correlations using a multi-head attention-based sequence-to-sequence model, slightly improve the Mean Absolute Percentage Error (MAPE) of CSTWPP [73] that applies 2D convolutional layers on spatiotemporal wind speed feature map. Similarly, in PV power forecasting, one can observe that the ASU-LSTM [95] reduces the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) of STCNN [63] by 0.432 and 0.538, respectively. Also, the SAC-ConvLSTM [102] that leverages the advantages of both CNN and RNN architectures outperforms B-LSTM [94] and MTSSCNN baselines by 1.728% and 1.964%. Moreover, the graph-based models show superior performance in regression applications as well. Due to the shown results in the table, the STSGCNN [121] outperforms SAC-ConvLSTM [102] and B-LSTM [94] by 1.558% and 3.355% in terms of MAPE, respectively. Similarly, the GRLSTM [135] that utilizes robust Rough LSTM to process spatiotemporal wind speed graph-based data decrease the MAPE of ConvRLSTM [107] and STAN [92] by 2.398% and 4.322%, respectively. The superiority of graph-based models in power system analysis stems from their capacity to explicitly represent intricate interdependencies among system components as a graph structure. By harnessing graph convolution operations, these models adeptly gather information from adjacent nodes, facilitating the extraction of spatiotemporal features crucial for analyzing power system behavior over time.

G. ADVANTAGES AND DISADVANTAGES OF DEEP DISCRIMINATIVE MODELS

In the realm of spatiotemporal analysis in power systems, CNNs [54], [59], [63], [65], [70], [71] offer notable advantages in capturing spatial dependencies within data, making them particularly adept at processing information from various sensors distributed across the power grid. By leveraging shared weights and local receptive fields, CNNs efficiently extract spatial features, enabling robust identification of patterns such as fault detection and classification of anomalies within the grid topology. However, CNNs may struggle to effectively model temporal dependencies inherent in time-series data, a critical aspect in power system analysis where the temporal dynamics of voltage fluctuations, load demand, and renewable energy generation play pivotal roles. The RNNs [92], [94], [95], [100], [154], on the other hand, excel in capturing sequential dependencies over time, making them well-suited for modeling temporal dynamics in power systems. Their ability to maintain internal state representations allows for the propagation of information across time steps, facilitating accurate forecasting of future grid states and dynamic event prediction. Nonetheless, RNNs may encounter challenges in handling long-term dependencies and suffer from the vanishing gradient problem, limiting their effectiveness in capturing complex spatiotemporal interactions. To mitigate these challenges, recent research has explored attention mechanisms in RNNs, offering an alternative approach to capture long-range dependencies in spatiotemporal data within power systems. By dynamically weighting the importance of different input features at each time step, attention mechanisms enable RNNs to focus on relevant information while effectively filtering out noise and irrelevant data.

The CNN-RNN architectures [107], [110], [111], [116] offer a compelling solution to address the limitations of individual CNN and RNN models in the spatiotemporal

TABLE 1. Deep Discriminative Architectures Across Different Classification-based Power Systems Applications. For each application, the best results are shown in bold and the second best results are underlined.

Application	Category	Model	Dataset	Performance Metric		
				Precision (%)	Recall (%)	F1-Score (%)
FDI Attack Detection	CNN	DCNN-CKF [54]	IEEE 300-bus system	75.128	62.498	68.233
	RNN	LSTM-AE [86]		66.123	73.658	69.687
	CNN-RNN	CNN-RNN-SAE [111]		74.259	71.193	72.694
	GCNN	Cplx-STGCN [134]		78.946	73.147	75.936
	GRNN	GC-LSTM [150]		**85.946**	**80.687**	**83.234**
Fault Identification	CNN	VMD-CNN [59]	IEEE 123-bus system	75.146	78.159	76.623
	RNN	GRU-GWO-SVR [100]		80.989	68.248	74.075
	CNN-RNN	CNN-GRU [110]		78.257	82.749	80.440
	GCNN	STGCN [128]		80.743	**85.127**	82.877
	GRNN	R-GCN [148]		**85.329**	84.725	**85.026**
NILM	CNN	TML-CNN [65]	UK-DALE	50.198	61.456	55.259
	RNN	DL-LSTM [154]		49.178	63.175	55.305
	CNN-RNN	CNN-LSTM-RF [116]		55.117	57.842	56.447
	GCNN	Spectral GCN [153]		61.752	63.478	62.603
	GRNN	AttG-BDGNets [152]		**64.449**	**69.175**	**66.728**
Event Classification	CNN	STCNN [63]	IEEE New England 39-bus system	69.023	75.140	71.952
	CNN-RNN	STFT-CNN-LSTM [113]		71.519	80.897	75.919
	GCNN	GCN [127]		73.811	**89.932**	81.078
	GRNN	RT-CPDLC [147]		**85.576**	86.297	**85.935**
Voltage Stability Assessment	CNN	FTSBA [151]	IEEE New England 10-generator 39-bus system	82.371	81.944	82.157
	CNN-RNN	TempCNN-LSTM [114]		85.184	88.991	87.046
	GCNN	AH-DAPE [130]		90.388	**91.085**	90.735
	GCRNN	GSTGNN [149]		**93.680**	90.651	**92.141**

analysis of power systems. By combining the strengths of both CNNs and RNNs, these architectures provide a comprehensive framework capable of capturing both spatial and temporal dependencies within the data. CNNs serve as feature extractors, effectively capturing spatial patterns from grid sensor data, while RNNs handle the sequential nature of temporal data, capturing dynamic dependencies over time. This synergistic approach enables CNN-RNN architectures to effectively model the complex interactions between spatial and temporal dimensions in power system data, leading to improved performance in tasks such as load forecasting, fault detection, and event classification. Furthermore, CNN-RNN architectures mitigate the vanishing gradient problem encountered by standalone RNNs, as the CNN component pre-processes the data, reducing the burden of long-term temporal dependencies on the RNN. However, CNN-RNN architectures may introduce additional complexity and computational overhead compared to standalone models, requiring careful architecture design and parameter tuning.

The GCNNs [121], [122], [123], [124], [126] as a powerful tool for leveraging graph data structures in the spatiotemporal modeling of power systems, offering several advantages over traditional CNN, RNN, and CNN-RNN approaches. Firstly, GCNNs inherently exploit the graph structure of the power grid, allowing for the propagation of information between neighboring nodes in the graph. This enables GCNNs to capture spatial dependencies more effectively compared to traditional CNNs, which typically operate on regular grid-like structures and may struggle to capture the irregular connectivity of the power grid. Additionally, GCNNs can naturally handle varying graph topologies and dynamic changes in network configurations, addressing the limitations of RNNs and CNN-RNNs in modeling temporal dynamics in power systems. However, it's worth noting that GCNNs may still face challenges in capturing long-range dependencies within the graph, particularly in large-scale power systems with complex network structures.

To address the challenges of capturing long-range dependencies within GCNN structures, the GCRNNs extend the capabilities of GCNNs by incorporating recurrent connections, allowing for the propagation of information not only between neighboring nodes but also across multiple time steps. By integrating recurrent connections into the graph convolutional framework, GCRNNs [138], [139], [140], [142], [143] can effectively capture both spatial and temporal dependencies within the power grid graph. This enables GCRNNs to model the dynamic interactions between

TABLE 2. Deep Discriminative Architectures Across Different Regression-based Power Systems Applications. For each application, the best results are shown in bold and the second best results are underlined.

Application	Category	Model	Dataset	Performance Metric		
				RMSE	MAE	MAPE (%)
Wind Power Forecasting	CNN	CSTWPP [73]	Wind Integration National Dataset	2.769	2.634	7.596
	RNN	STAN [92]		2.549	2.129	6.869
	CNN-RNN	ConvRLSTM [107]		1.856	1.665	4.945
	GCNN	STDGCN [125]		1.297	1.196	4.069
	GRNN	GRLSTM [135]		**1.064**	**0.852**	**2.547**
PV Power Generation forecasting	CNN	STCNN [71]	Solar Integration National Dataset	1.652	1.463	8.605
	RNN	ASU-LSTM [95]		1.114	1.031	8.369
	CNN-RNN	CNN-LSTM [105]		**1.032**	0.978	6.641
	GCNN	st-GNN [123]		1.166	0.862	5.905
	GRNN	GSINN [138]		1.079	**0.716**	**4.764**
Electrical Demand Forecasting	CNN	MTSSCNN [70]	Low Carbon London	2.445	2.268	9.102
	RNN	B-LSTM [94]		2.237	1.912	8.200
	CNN-RNN	SAC-ConvLSTM [102]		1.897	1.661	6.403
	GCNN	STSGCN [121]		1.522	1.327	4.845
	GRNN	GCLSTM [139]		**1.239**	**1.113**	**3.902**

different components of the grid over time, providing a more comprehensive understanding of spatiotemporal phenomena such as voltage fluctuations, load dynamics, and renewable energy integration. Moreover, GCRNNs offer flexibility in handling varying graph topologies and dynamic changes in network configurations, making them well-suited for real-world applications in power system analysis and forecasting. Despite these advantages, GCRNNs may introduce additional complexity and computational overhead compared to standalone GCNs, requiring careful optimization and parameter tuning.

III. DEEP GENERATIVE LEARNING

Generative modeling offers an alternative paradigm to address spatiotemporal correlation learning in power system analysis. Unlike discriminative modeling, which focuses on learning the conditional distribution of output given input $p(Y|X)$, generative models aim to capture the joint distribution of X and Y and then map that distribution to $P(Y|X)$ for a supervised task. By modeling the joint distribution, generative models can potentially offer a deeper understanding of the underlying data generating process, facilitating tasks such as data synthesis, anomaly detection, and uncertainty quantification [155], [156], [157]. Various GAN- and VAE-based approaches have been explored in the context of power system analysis to capture complex spatiotemporal correlations for probabilistic and deterministic tasks. In this section, we explore the details of these generative models and review the spatiotemporal approaches that employ them for solving different problems in power systems analysis and operation.

A. VARIATIONAL AUTOENCODER

VAEs [158], [159], [160] are a class of generative models that leverage the variational inferance techniques to learn latent representations of high-dimensional data. The VAE framework comprises two NNs: an encoder $q_\phi(z|x)$ parametrized by ϕ, responsible for mapping the input data (i.e., power system measurements) x to a lower-dimensional latent representation z, and a decoder $p_\theta(x|z)$ parametrized by θ, which reconstructs the original data x from the latent representation z. In a VAE, the encoder q_ϕ maps the input data to a normal distribution by parameterizing the mean (μ) and variance (σ^2) of a multivariate Gaussian distribution in the latent space. Mathematically, the encoder outputs two vectors, one for the mean (μ) and another for the variance (σ^2). These vectors are then used to parameterize the Gaussian distribution $q_\phi(z|x) \sim \mathcal{N}(\mu, \sigma^2 I)$. The decoder $p_\theta(x|z)$ reconstructs the input data from the latent representation by mapping the latent variables $z \sim \mathcal{N}(\mu, \sigma^2 I)$ back to the original data x. In a VAE, the decoder $p_\theta(x|z)$ takes a sample from the latent space $\mathcal{N}(\mu, \sigma^2 I)$ as input and generates a synthetic data samples $\tilde{x} \sim p(x)$ where $p(x)$ is the probability distribution on the original data [161].

In training VAEs, the objective is to optimize the model parameters to best approximate the underlying data distribution by maximizing the Evidence Lower Bound (ELBO) which serves as a surrogate objective for the intractable true likelihood of the input data x. Mathematically, the ELBO can be expressed as $\mathcal{L}(\theta, \phi; x) = \mathbb{E}_{z\sim q_\phi(z|x)}[\log p_\theta(x|z)] - D_{KL}(q_\phi(z|x)||p(z))$ where D_{KL} denotes the Kullback-Leibler divergence between the approximate posterior $q_\phi(z|x)$ and the prior distribution $p(z)$. By maximizing the ELBO,

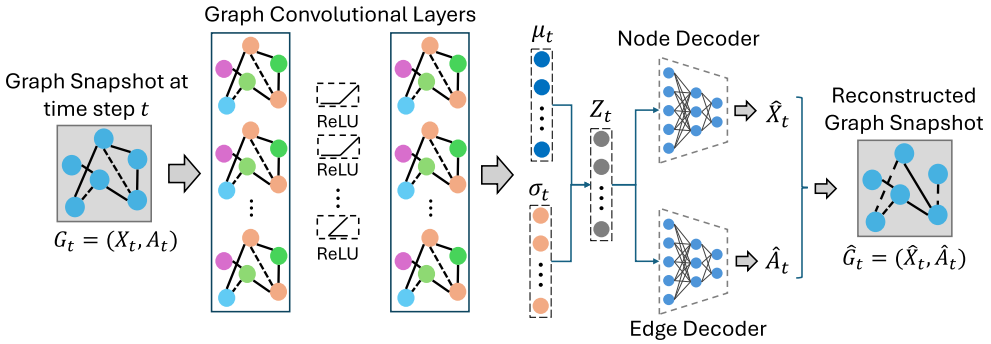

FIGURE 7. General framework of variational graph convolutional autoencoder.

VAEs learn to encode meaningful latent representations while simultaneously generating realistic data samples [162]. Figure 7 illustrates a variant of the graph VAE designed for learning the most significant node- and edge-based features of the input graph. As depicted, the VAE encoder (i.e., GCN), maps the input graph to a latent space distribution. Sampling from this distribution allows for the reconstruction of the input graph using two distinct decoders for node and edge features.

The VAEs have found extensive applications across various domains within power systems, serving as a powerful tool for unsupervised feature extraction and as a supervised method for estimating discrete labels or continuous variables. For instance, Pan et al. [163] introduced a VAE-based approach for load profile generation, which integrates deep CNNs as both encoder and decoder components. Saffari et al. [31] present a robust generative solution by integrating GCNN, VAE, and Rough Set theory for PV power forecasting. Their model effectively learns the probability distribution functions (PDF) of each PV site, enhancing the accuracy of future value predictions. Similarly, in [164], a graph convolutional VAE architecture is proposed to learn the continuous nodal PDF of arbitrary graphs representing solar irradiance. Experimental results conducted on the National Solar Radiation Database demonstrate superior performance in probabilistic radiation prediction across geographically distributed irradiance data. More recently, Ma et al. [165] integrate VAE and generalized regression neural network models to introduce a spatiotemporal generative autoencoder for probabilistic wind forecasting. Their experiments, conducted on the Global Energy Forecasting Competition 2014 dataset, highlight the significant performance gains achieved by the proposed model under diverse weather conditions.

Additionally, in load monitoring, Khodayar et al. [166] introduce a generative LSTM framework tailored to address the uncertainty inherent in power resource monitoring. Their approach involves learning the continuous PDF of load parameters from intricate temporal variations in measurements using developed VAE. Through numerical experiments conducted on the 68-bus New England and New York Interconnect System, they demonstrate the effectiveness of the proposed VAE-based model across various probabilistic estimation metrics. Furthermore, Regan et al. [167] present an attention-based VAE incorporating LSTM and dictionary learning [168] modules for NILM. Their study illustrates how generative modeling coupled with attention mechanisms enhances deep learning's comprehension of the spatiotemporal correlations among NILM features. Lastly, Zheng et al. [169] propose a multi-scale load forecasting algorithm employing VAE and LSTM architectures to model sequential data and accurately predict electricity consumption.

In the realm of fault identification and cybersecurity within power systems, various spatiotemporal methodologies have leveraged the VAE framework. Recently, Wang et al. [170] employed attention-based GRU and VAE models for unsupervised locational FDI attack detection in state estimation within smart grids. Their experimental results across multiple power system measurements demonstrate significant performance in their proposed model. Furthermore, Mylonas et al. [171] utilized a conditional VAE to characterize the PDF of accumulated fatigue in wind turbines using historical SCADA data. Additionally, Aftabi et al. [172] integrated RNNs, VAEs, and deep ReLU networks to develop a comprehensive generative framework for the detection, diagnosis, and localization of cyberattacks on smart grids. Through the evaluation of a networked power transmission system, they establish the superiority of their proposed model over traditional model-based attack detection methods.

B. GENERATIVE ADVERSARIAL NETWORK

GAN frameworks [34], [173], [174], [175] have emerged as a powerful tool in the field of spatiotemporal generative modeling, offering a novel approach to learning realistic data distributions. As shown in Fig. 8, GANs comprise two deep neural networks [176]: the generator G_ϕ parameterized by ϕ and the discriminator D_θ parameterized by θ, which are involved in a minimax game. The generator takes random noise z from a prior distribution $p(z)$ and generates synthetic data samples $\hat{x} = G(z)$. Simultaneously, the discriminator receives both real data samples x from the true data distribution $p_{\text{data}}(x)$ and generated samples $\hat{x}$, and aims to distinguish

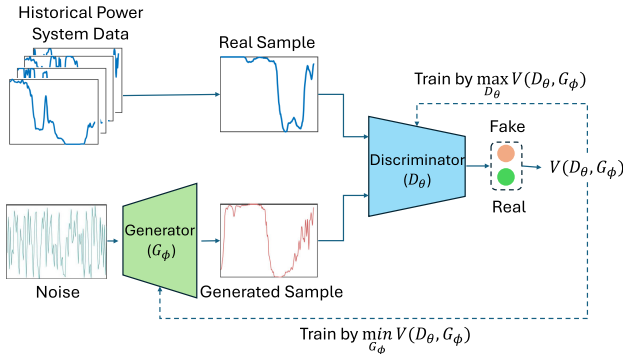

FIGURE 8. General framework of the GAN architecture.

between them by assigning high probabilities to real data and low probabilities to generated data. Through adversarial training, the generator learns to generate increasingly realistic samples, while the discriminator improves its ability to distinguish between real and fake data. This iterative process results in the generation of high-quality synthetic data that follows the underlying distribution of input data $p(X)$ [177].

The training of GANs is framed as a minimax game involving the generator and discriminator networks. The generator aims to reduce the dissimilarity between the distribution of generated samples $p(G(z))$ and real samples $p(x)$. Meanwhile, the discriminator attempts to maximize the discrepancy between the distribution of real samples and that of fake samples. The GAN's objective function can be formulated as follows [178]:

$$\min_{G_\phi} \max_{D_\theta} V(D_\theta, G_\phi) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D_\theta(x)] + \mathbb{E}_{z\sim p(z)}[\log(1 - D_\theta(G_\phi(z)))] \quad (8)$$

where $V(D_\theta, G_\phi)$ represents the value function that the discriminator seeks to maximize and the generator aims to minimize. This objective encourages the generator to produce samples that are indistinguishable from real data, while the discriminator aims to correctly classify real and fake (or generated) samples. Through iterative optimization using techniques such as SGD or its variants, the generator and discriminator learn to improve their respective abilities, leading to the generation of increasingly realistic data samples [178].

In recent years, various GAN architectures have emerged to tackle complex challenges within the spatiotemporal analysis of power systems, particularly scenario generation. This critical task involves creating diverse hypothetical scenarios to evaluate the behavior and performance of the power grid under different conditions, encompassing factors such as load fluctuations, weather patterns, equipment malfunctions, and market dynamics [179], [180], [181], [182], [183]. For example, Chen et al. [181] developed a conditional GAN (CGAN) tailored for solar scenario generation, integrating user-defined labels during training to enable event-based scenario creation. These scenarios demonstrate statistical consistency with historical data and effectively capture spatiotemporal correlations across multiple locations. Similarly, Yang et al. [182] introduced a CGAN for generating scenarios related to PV power generation, employing it in the design of hybrid energy storage systems. A common challenge in GAN training is mode collapse, where the generator network produces repetitive or limited sample variations. Addressing this issue, Li et al. [183] proposed a federated learning framework based on least square GANs for renewable scenario generation. This framework learns a shared global model and generates scenarios in a privacy-preserving manner.

Moreover, GAN architectures find application in spatiotemporal classification tasks within power systems analysis, notably in event detection, fault identification, and FDI attack detection. For instance, Zheng et al. [184] present a two-stage framework for synthesizing PMU data, employing GANs for data augmentation to enhance event classification accuracy. Their model integrates GANs for data generation and incorporates neural ordinary differential equations to improve model explainability. Similarly, Cheng et al. [34] propose a bidirectional GAN-based algorithm for real-time event classification using streaming PMU data validated on a large-scale dataset from the Eastern Interconnection of the US. Additionally, Wu et al. [185] leverage GAN architecture for conductor galloping monitoring, a critical task for power system safety. Their work introduces a curve reconstruction method using CGANs to fully synthesize transmission line galloping curves, demonstrating accurate reconstruction with minimal sensor usage. Moreover, in order to tackle imbalanced dataset challenges in FDI attack detection, a GAN framework is devised in [173], with a GRU serving as the generator and a transformer neural network acting as the discriminator. This framework aims for the precise classification of various FDI attack scenarios on the IEEE 118-bus system.

Numerous studies have employed GANs in spatiotemporal power systems data generation, owing to their effectiveness in synthesizing data. For instance, Yang et al. [186] tackle the power system data recovery problem by proposing a GAN-based architecture employing LSTM neural networks as the generator and discriminator for data generation tasks. Their results demonstrate the efficacy of the GAN-based method in effectively recovering lost data. Additionally, to address missing data in wind turbine data collection, Hu et al. [187] introduce a spatiotemporal GAN architecture capable of capturing historical decay and feature correlations among wind turbines under varying environmental conditions. Experimental findings highlight the significant performance of the proposed GAN model in data imputation tasks. Moreover, Song et al. [188] introduced ProfileSR-GAN, a two-stage GAN-based framework designed to tackle the issue of load profile super-resolution. ProfileSR-GAN aims to restore high-frequency components from low-resolution load profiles to generate high-resolution load profiles. Moreover, research by Hu et al. [189], and Silva et al. [190] underscores the notable performance of

GAN-based models in generating load profiles. These studies collectively emphasize the utility of GANs in enhancing spatiotemporal data generation capabilities within power systems analysis.

Several state-of-the-art methodologies leverage GAN architectures for sustainable energy and load forecasting tasks. For instance, Wen et al. [191] devised a GAN-based generative approach specifically for regional solar forecasting across an entire geographical area. Their method employs GANs to learn the temporal variation of spatial solar irradiance maps, enabling accurate prediction of future SIM steps. Similarly, Yuan et al. [192] introduced an enhanced GAN model with guaranteed convergence to precisely capture the uncertainty of solar and wind power resources, thereby improving forecasting accuracy. They applied this model to learn the intrinsic spatiotemporal patterns of multiple-site renewable energy systems, demonstrating significant performance gains in forecasting both wind and solar power scenarios using real-world datasets. In a similar vein, Wei et al. [193] developed a GAN-based model tailored for a large-scale hydro-wind-solar hybrid system, aiming to capture the intricate spatiotemporal relationships between wind farms and PV plants. Their numerical experiments conducted on a renewable energy base in southwest China showcased the superior performance of the proposed generative model in generating high-quality scenarios. Moreover, in [175], a conditional GAN-based architecture is developed for residential load forecasting. The proposed method utilizes LSTM and CNN architectures for the generator and discriminator networks, respectively, to capture both spatial and temporal correlations. Experimental results from this study demonstrate significant performance enhancement compared to discriminative architectures.

C. PERFORMANCE COMPARISON AND DISCUSSION

Similar to discriminative approaches, we conducted a comprehensive evaluation of various state-of-the-art generative techniques for analyzing spatiotemporal power systems, encompassing both classification and regression tasks. Table 3 presents the experimental findings of these models across FDI attack detection, fault identification, and voltage stability assessment scenarios. Notably, the GAN-based approaches demonstrated superior accuracy compared to VAE-based models across diverse applications. As shown in Table 3, GAN-based architectures exhibit superior generalization capabilities compared to VAE-based approaches. For example, in voltage stability assessment, the CycleGAN [174], leveraging adversarial training and cycle-consistency loss functions, outperforms VSA-VAE [194] by 3.095% based on the F1-score metric. Likewise, GAN-based models demonstrate enhancements of 4.285% and 1.816% over VAE-based models in FDI attack detection and fault identification, respectively. Moreover, Table 4 illustrates how GANs enhance performance in regression-based applications. For instance, in load profile generation, the ProfileSR-GAN [188] outperforms the Load-VAE [163] framework by 0.333 and 0.562 in terms of RMSE and MAE metrics, respectively. Similarly, in wind forecasting, the GAN-CLSTM [195] reduces the forecasting MAPE of STGRVAE [196] by 1.723% by training CNN and RNN architectures in an adversarial fashion. Similar improvements of GANs over VAEs are observable across various applications detailed in Tables 4 and 3. This superiority can be attributed to their implicit modeling of the latent spatiotemporal space through the generator network, enabling GANs to capture intricate and nonlinear relationships in power system spatiotemporal measurements more effectively compared to VAE-based models, which rely on an explicit latent space representation.

D. ADVANTAGES AND DISADVANTAGES OF DEEP GENERATIVE MODELS

Deep generative models [172], [173], [191], [197] offer distinctive advantages compared to discriminative models in the context of spatiotemporal modeling in power systems analysis. Unlike discriminative models, which focus on learning the conditional PDF of the target given the input, generative models aim to capture the underlying data distribution directly. This enables generative models to generate new samples that closely resemble the training data, facilitating data augmentation and synthetic data generation for tasks with limited labeled data availability. Additionally, deep generative models inherently capture the high-dimensional and nonlinear nature of the data distribution, allowing for more flexible and expressive representations of complex spatiotemporal patterns within power system measurements. Furthermore, generative models can provide insights into the latent structure of the data, enabling the discovery of hidden variables or features that may be critical for understanding system behavior. While generative models may pose challenges in training stability and convergence, advancements in training techniques and architectures have mitigated many of these issues, making them increasingly viable for spatiotemporal modeling in power systems.

In spatiotemporal modeling of power systems, the VAEs [31], [171], [172], [196] offer distinct advantages and disadvantages. VAEs provide a probabilistic framework for generative modeling, facilitating uncertainty quantification crucial for the inherent unpredictability in power systems. Additionally, VAEs inherently perform dimensionality reduction, aiding in managing the high-dimensional nature of spatiotemporal power data. However, VAEs often suffer from producing blurry and noisy outputs and may struggle to capture complex dependencies and structures within the data in applications with the limited number of data points.

The GAN architectures [182], [188], [189], [191] offer several advantages of VAEs in spatiotemporal modeling of power systems. GANs excel in generating high-fidelity, sharp samples, potentially capturing the nuanced dynamics of power systems more accurately compared to VAEs. The adversarial training framework of GANs encourages the model to produce realistic samples by competing with

TABLE 3. Deep Generative Architectures Across Different Classification-based Power Systems Applications. For each application, the best results are shown in bold.

Application	Category	Model	Dataset	Performance Metric		
				Precision	Recall	F1-Score
FDI Attack Detection	VAE	VAE-RNN [172]	IEEE 300-bus system	84.493	78.125	81.184
	GAN	GAN-GRU [173]		**88.739**	**82.432**	**85.469**
Fault Identification	VAE	CVAE [171]	IEEE 123-bus system	83.158	**85.816**	84.466
	GAN	Bi-AnoGAN [34]		**88.406**	84.258	**86.282**
Voltage Stability Assessment	VAE	VSA-VAE [194]	IEEE Standard New England 39-Bus Test System	87.294	95.469	91.199
	GAN	CycleGAN [174]		**91.772**	**96.959**	**94.294**

TABLE 4. Deep Generative Architectures Across Different Regression-based Power Systems Applications. For each application, the best results are shown in bold.

Application	Category	Model	Dataset	Performance Metric		
				RMSE	MAE	MAPE
Wind Forecasting	VAE	STGRVAE [196]	Wind Integration National Dataset	1.978	1.353	3.931
	GAN	GAN-CLSTM [195]		**1.164**	**0.752**	**2.208**
PV forecasting	VAE	CGRVAE [31]	Solar Integration National Dataset	1.918	1.723	4.274
	GAN	Multiscale-GAN [191]		**1.053**	**0.891**	**3.015**
PV Scenario Generation	VAE	CNN-VAE [197]	IEEE New England 39-bus system	1.769	1.305	5.281
	GAN	C-GAN [182]		**1.092**	**0.821**	**4.710**
Load Forecasting	VAE	TFVAE-LSTM [169]	IEEE New England 10-generator 39-bus system	1.815	1.593	5.061
	GAN	cWGAN-GP [175]		**1.211**	**1.184**	**3.151**
Load Profile Generation	VAE	Load-VAE [163]	IEEE New England 10-generator 39-bus system	1.807	1.520	3.116
	GAN	ProfileSR-GAN [188]		**1.474**	**0.958**	**2.870**

a discriminator network, leading to sharper outputs with fine details. However, GANs require careful training and may suffer from mode collapse [177], [198], where they fail to capture the full diversity of the data distribution, especially in complex and high-dimensional spatiotemporal data like power systems. Conversely, VAEs provide a probabilistic generative model that inherently supports uncertainty quantification, which is essential for addressing the inherent unpredictability in power systems. Additionally, VAEs perform better dimensionality reduction, aiding in managing the high-dimensional nature of spatiotemporal power data. Therefore, the choice between GANs and VAEs for spatiotemporal analysis of power systems depends on the specific modeling requirements, considering the trade-offs between model fidelity, uncertainty quantification, and computational efficiency.

IV. DEEP REINFORCEMENT LEARNING

Reinforcement learning approaches [35], [199], [200] have attracted significant attention in the realm of spatiotemporal learning within power systems, offering a dynamic framework to optimize system operation and management. As shown in Fig. 9, the system's behavior is modeled as a Markov Decision Process (MDP) [201], where the state of the system at each time step encapsulates relevant spatiotemporal information such as power generation, consumption, and grid topology. Formally, the state s_t at time t can be represented as a vector comprising features characterizing the system's spatial and temporal dynamics, denoted as $s_t = (x_t, y_t, t)$, where x_t and y_t denote the spatial coordinates and t represents the temporal dimension. The action a_t taken by the RL agent influences the system's state transition from s_t to s_{t+1}, thereby impacting the subsequent evolution of the power system.

Mathematically, the RL agent aims to learn an optimal policy π^* that dictates the selection of actions to maximize the cumulative reward over time. The reward function $R(s_t, a_t, s_{t+1})$ quantifies the immediate desirability of transitioning from state s_t to s_{t+1} by executing action a_t. Typically, in power systems, the reward function is designed to

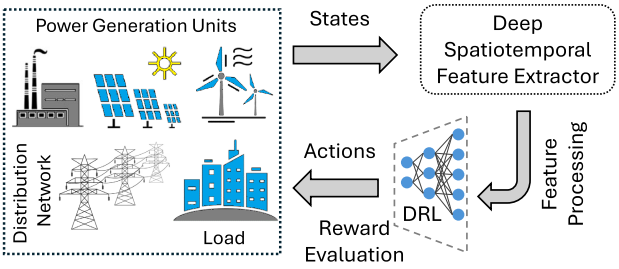

FIGURE 9. General framework of deep reinforcement learning architecture.

incentivize actions leading to improved system performance, such as minimizing transmission losses, maintaining voltage stability, and enhancing renewable energy integration. Thus, the RL agent's objective is to iteratively explore and exploit the state-action space to learn an optimal policy that maximizes the expected cumulative reward, formulated as [35]:

$$\pi^* = \arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t, s_{t+1})\right] \tag{9}$$

where γ denotes the discount factor accounting for the long-term consequences of actions. Through this iterative learning process, RL techniques offer promising avenues for spatiotemporal learning in power systems, enabling adaptive and efficient decision-making in complex and dynamic environments.

Generally speaking, the RL methods can be categorized into three main classes [202]: policy-based, value-based, and model-based methods. Policy-based methods directly learn the optimal policy, a mapping from states to actions, without explicitly computing value functions of the states. These methods typically involve training a neural network to approximate the policy and using techniques like policy gradients to update the network parameters. On the other hand, value-based methods focus on estimating the value of state-action pairs, aiming to maximize the expected return. Finally, model-based methods involve learning a model of the environment dynamics, enabling the agent to plan by simulating possible future states and actions. In this section, we aim to explore the intricacies of widely used RL algorithms, assessing their performance and applicability to spatiotemporal learning in power systems [203]. We seek to provide insights into their effectiveness and suitability for addressing the complexities inherent in power system optimization and management through empirical evaluations and comparative analyses.

A. Q-LEARNING

Q-learning [204], [205] is a fundamental reinforcement learning algorithm that forms the basis for many advanced techniques used in the spatiotemporal analysis of power systems. At its core, Q-learning aims to learn the optimal action-value function $Q(s_t, a_t)$, where s_t represents the state of the environment and a_t denotes the action taken by the agent at the time step t. The key idea behind Q-learning is the iterative update of the action-value function using the Bellman equation:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\left[r_{t+1} + \gamma \max_{a} Q(s_{t+1}, a) - Q(s_t, a_t)\right] \tag{10}$$

where $\alpha \in [0, 1]$ is the learning rate, r_{t+1} is the immediate reward obtained by taking action a_t in state s_t, and $\gamma \in [0, 1)$ is the discount factor. Q-learning iteratively improves the action-value function by minimizing the temporal difference error between the predicted and target values. While Q-learning is effective in simple environments with discrete action spaces, it may struggle in high-dimensional or continuous action spaces commonly encountered in power system analysis.

Deep Q-Networks (DQNs) [206] address the limitations of traditional Q-learning by employing deep neural networks to approximate the action-value function $Q(s, a; \theta)$, parameterized by neural network weights θ. DQN enables the handling of high-dimensional state spaces (e.g., spatiotemporal power system data) by learning a nonlinear mapping from states to action values. The key innovation of DQN lies in its use of experience replay and target networks to stabilize training and improve sample efficiency. Experience replay consists of retaining agent encounters (state, action, reward, next state) within a replay buffer and drawing mini-batches from this buffer for neural network training. Meanwhile, target networks are used to compute target values for training, reducing the impact of moving targets on learning stability. The DQN loss function is defined as the mean squared error between the predicted and target values:

$$\mathcal{L}(\theta) = \mathbb{E}_{(s,a,r,s')\sim\mathcal{D}}\left[\left(r + \gamma \max_{a'} Q(s', a'; \phi) - Q(s, a; \theta)\right)^2\right]$$
$$\theta \leftarrow \theta - \alpha \nabla_\theta L(\theta) \tag{11}$$

where $\mathcal{D}$ denotes the replay buffer, ϕ represents the parameters of the target network, and γ is the discount factor.

Double Deep Q-Network (DDQN) [207] extends the DQN framework by mitigating overestimation bias, which can occur when using the maximum action-value estimate for both action selection and evaluation. DDQN achieves this by decoupling the action selection and evaluation processes, employing separate online and target networks for each. The online network is used to select actions, while the target network is used to evaluate actions, reducing the likelihood of overestimation. Mathematically, DDQN involves updating the target network parameters ϕ less frequently than the online network parameters θ, thus stabilizing training and improving convergence. The Q-learning update rule for DDQN is similar to DQN, but it utilizes the action selection network to compute the target Q-value. Mathematically, the

update rule for DDQN can be represented as [208]:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \cdot (r_{t+1} + \gamma \cdot Q(s_{t+1}, \text{argmax}_{a'} Q(s_{t+1}, a'; \theta); \phi) - Q(s_t, a_t)) \tag{12}$$

where $\arg\max_{a'} Q(s', a'; \theta)$ represents the action selected by the online network. The loss function for DDQN remains similar to DQN, computed as the mean squared error between the predicted Q-value and the target Q-value, where the target Q-value is calculated using the target network.

Several recent studies have developed spatiotemporal Q-learning-based frameworks in diverse power system applications. For instance, Babar et al. [209] introduced micromodels for agile demand response, facilitating precise monitoring, learning, and scheduling of demand flexibility by multiple agents. Similarly, Lu et al. [210] utilized multi-agent Q-learning in an hour-ahead demand response algorithm for home energy management systems, demonstrating reduced user energy bills and dissatisfaction costs. Moreover, Xu et al. [211] proposed a multi-agent reinforcement learning (MARL) framework integrating Q-learning and ELM neural networks for home energy management systems. Their model schedules the energy consumption of various home appliances based on electricity price trends obtained from the ELM neural network. Additionally, a recent study [212] introduced a double Q-learning framework for voltage stability control, augmenting the MARL model with a GCNN to capture topology changes and spatiotemporal correlations in nodal features. Experimental results on the IEEE 39-bus system affirm the superior control performance of the GCN DDQN model amidst grid topology changes.

Recent studies [213], [214] have leveraged deep Q-networks for load-shedding applications. Zhang et al. [213] introduced a deep Q-network designed to determine real-time optimal load-shedding strategies for maintaining power system stability, utilizing spatiotemporal information extracted by a CNN-LSTM architecture. Similarly, [214] proposed an improved double deep Q-network for the load shedding problem, incorporating a graph neural network for spatiotemporal feature extraction. Simulation results conducted on modified IEEE 39- and 300-bus systems underscore the efficacy of the proposed spatiotemporal Q-learning-based framework in offering both economical and reliable control strategies.

Furthermore, recent research efforts have explored the application of Q-learning techniques to enhance the cyber security of power systems. For instance, Liu et al. [215] proposed a DQN-based framework to evaluate the cyber-physical security of power systems, particularly focusing on the challenges posed by the intermittent generation of renewable energy sources. Their approach models the states of power systems as partially observable MDP (POMDP), enabling a more comprehensive assessment of system vulnerabilities. Similarly, Li and Wu [216] addressed the issue of low-latency cyber attack detection in smart grids by developing a DQN within the framework of an improved MDP. Notably, their proposed DQN incorporates a meticulously designed reward function, allowing for flexible trade-offs between detection delays and accuracy.

B. SARSA

The State–Action–Reward–State–Action (SARSA) [217] algorithm stands as a pivotal tool in the realm of reinforcement learning, particularly when applied to decision-making processes in spatiotemporal power network analysis. It operates by iteratively updating the value of state-action pairs based on the agent's experience within the environment. At each time step t, the agent observes the current state of the network, selects an action according to its policy, receives a reward, transitions to a new state, and takes another action. Unlike Q-learning, the SARSA updates the value of the action taken rather than the best action for the next state. This distinction makes SARSA particularly suitable for scenarios where the agent's actions directly influence subsequent states. Mathematically, the update rule for SARSA can be expressed as:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)\right] \tag{13}$$

where $Q(s_t, a_t)$ represents the value of taking action a_t in state s_t, while r_{t+1} denotes the immediate reward obtained upon transitioning to the next state s_{t+1}. The learning rate α and discount factor γ regulate the magnitude of updates and the importance of future rewards, respectively. Here, the term $Q(s_{t+1}, a_{t+1})$ embodies the value of the action taken under the current policy, guiding the agent's learning process towards optimal decision-making [218].

In the context of spatiotemporal power network analysis, SARSA emerges as a powerful technique for addressing critical challenges, including voltage control [219], energy management [220], demand response [221], and cybersecurity of power systems [222]. Tousi et al. [219] designed a MARL framework where each agent is equipped with voltage control devices, and a multi-agent SARSA algorithm is proposed for training these agents to have an acceptable voltage profile in all nodes of the power system. More recently, the SARSA algorithm is applied in [221] to address demand response in industrial multi-energy microgrids with a variety of sustainable energy resources. Experimental results demonstrate the superior performance of the proposed adaptive SARSA framework compared to recent demand response models. In energy management, Aljohani and Mohammed [220] utilized the SARSA algorithm to learn the maximum travel policy for electric vehicles, treating them as agents, with the defined optimal behavior of the agent serving as the reward function. Simulation results show a slight performance improvement of the proposed approach compared to the double deep Q-network algorithm. Additionally, Kurt et al. [222] demonstrated the effectiveness of the SARSA algorithm in detecting FDI attacks, modeled

as a POMDP problem. Numerical investigations demonstrate the effectiveness of the proposed detection scheme in the reliable identification of cyber-attacks in the smart grid.

C. ACTOR-CRITIC

Actor-Critic is a reinforcement learning framework that combines both policy-based and value-based methods [223]. It consists of two neural networks: the "actor" learns the policy to select actions, while the "critic" evaluates these actions by estimating their value. Here, the "actor" component represents the policy function, denoted as $\pi(s, a; \theta)$, which maps state s to actions a with parameters θ. This policy function guides decision-making by selecting actions that maximize expected rewards. In the context of power systems, actions might correspond to adjustments in generation, transmission, or distribution to optimize performance metrics such as stability, reliability, or efficiency. Meanwhile, the "critic" component evaluates the action-value function, denoted as $Q(s, a; w)$, which estimates the expected return when taking action a in state s, parameterized by w. The action-value function provides feedback to the actor by assessing the quality of chosen actions, enabling refinement of the policy over time. Mathematically, the actor updates its parameters θ by gradient ascent to maximize the expected return, while the critic updates its parameters w through temporal-difference learning to minimize the temporal difference error between predicted and actual returns, as shown in the equations below:

$$\begin{aligned} \Delta\theta &= \alpha \nabla_\theta \log \pi(s, a; \theta) Q(s, a; w) \\ \Delta w &= \beta \delta \nabla_w Q(s, a; w) \end{aligned} \tag{14}$$

Here, α and β represent learning rates, ∇_θ and ∇_w denote gradients with respect to the actor and critic parameters, respectively, and δ signifies the temporal difference error, calculated as the difference between the observed and predicted returns. Through iterative updates based on these equations, the Actor-Critic framework learns to navigate the spatiotemporal dynamics of power systems, continually improving control policies to adapt to changing conditions and optimize system performance.

Several recent studies have employed actor-critic frameworks and developed improved versions for various spatiotemporal challenges in power systems. For instance, Hu et al. [224] proposed an experience-augmented multi-agent actor-critic algorithm enhanced by an attention mechanism to learn high-quality spatiotemporal policies for voltage control at the distribution level. Additionally, in [225], a Gumbel-softmax soft actor-critic algorithm is proposed for real-time dynamic network reconfiguration and Volt-VAR control. Also, Bakakeu et al. [226] present a Multi-Agent Actor-Critic (MAAC) framework tailored for optimizing energy utilization within a heterogeneous cluster of electric machines equipped with energy generation and storage capabilities in a microgrid environment. Moreover, Mu et al. [227] have addressed the voltage control problem by formulating it as a decentralized POMDP and leveraging the graph-based MAAC framework. Simulation results on IEEE 33- and 141-bus systems show the effectiveness of the proposed model in learning spatiotemporal correlations of distribution networks by integrating GCN into the MAAC framework.

Furthermore, Mazare [228] employs the actor-critic framework to enhance the cyber-security of wind farms against FDI attacks, as well as match and mismatch disturbances. By integrating a fixed-time observer-based sliding mode control mechanism into the actor-critic framework, the proposed model enhances both the convergence time and the steady-state accuracy compared to recent studies. Additionally, Gassi and Baysal [229] address energy management in microgrids with sustainable energy resources by integrating a DNN-based actor-critic framework and a linear programming myopic optimization model. Also, a robust actor-critic augmented with a heuristic mechanism is proposed in [230] to address the automatic generation control challenge exacerbated by the disturbances stemming from the stochastic nature of renewable energies. Moreover, Gu and Huang [231] develop a robust multi-agent actor-attention-critic (MAAAC) framework for the reactive power optimization process in an active distribution network under the high permeability of distributed generation. The experimental results of this study on IEEE 33- and IEEE 123-node networks reveal the superiority of the MAAAC framework in extracting task-relevant spatiotemporal features and the high accuracy of the proposed MAAAC model under varying degrees of data uncertainties.

D. DDPG

Inspired by the actor-critic framework, deep deterministic policy gradient (DDPG) combines the strengths of deep neural networks with policy gradient methods, enabling agents to effectively navigate the high-dimensional action spaces inherent in power system control [232]. The DDPG employs a deterministic policy gradient approach to learn continuous action policies, making it particularly well-suited for problems characterized by continuous control, such as power flow optimization and voltage control. In the DDPG, the actor network is trained to maximize the expected cumulative reward by directly adjusting the policy parameters, while the critic network learns to estimate the value function, capturing the expected future rewards associated with state-action pairs. The update equations for the actor and critic networks can be expressed as follows:

$$\begin{aligned} \nabla_{\theta^\mu} J &\approx \mathbb{E}\left[\nabla_{\theta^\mu} Q(s, a|\theta^Q)|_{s=s_t, a=\mu(s_t|\theta^\mu)}\right] \\ \nabla_{\theta^Q} J &\approx \mathbb{E}\left[\frac{\partial Q(s, a|\theta^Q)}{\partial a} \frac{\partial a}{\partial \theta^Q}|_{s=s_t, a=\mu(s_t)}\right] \end{aligned} \tag{15}$$

where, J represents the expected cumulative reward, $\mu(s_t|\theta^\mu)$ denotes the policy function parameterized by θ^μ, and $Q(s, a|\theta^Q)$ denotes the action-value function parameterized by θ^Q. The gradients with respect to the actor and critic

parameters (θ^{μ} and θ^{Q}) are computed to update the respective network weights.

In the context of spatiotemporal analysis of power systems, the DDPG offers a promising approach to addressing different challenges such as voltage control [233], [234], [235] in modern power systems. Li et al. [233] developed a spatiotemporal framework by integrating attention-based graph convolution and the DDPG algorithm for voltage fluctuation control in distribution networks with renewable energy resources. They have validated the performance of the proposed model on the modified IEEE 33, 69, and 128-bus systems. Also, Wang et al. [234] have employed the DDPG algorithm for multi-agent voltage control formulated as a Markov Game, utilizing a heuristic method to partition agents. Additionally, in [235], a Multi-Agent DDPG (MADDPG) framework is designed for solving the Volt/Var problem formulated as a POMDP. In the proposed method, the spatiotemporal uncertainties associated with PV power generation and loads are represented through stochastic programming as scenarios within the MADDPG algorithm. Simulations carried out on the IEEE 123-node system under both PV-peak and load-peak scenarios affirm the superior performance of the DDPG-based approach.

Moreover, the application of the DDPG algorithm extends to other challenging spatiotemporal tasks within power systems. For instance, Jendoubi and Bouffard [236] devised a multi-agent hierarchical DDPG approach for scheduling the operation of controllable devices within electric networks. Their experimental findings, based on one-hour resolution load and PV data, indicate that the developed DDPG approach outperforms other control strategies. Additionally, Chengqing et al. [237] utilized attention-based GCGRU and the DDPG algorithm to construct a spatiotemporal wind power prediction model. Similarly, Zhang et al. [238] applied the MADDPG algorithm for spatiotemporal fault diagnosis and protection strategies in power systems, demonstrating superior performance compared to conventional SARSA and DDPG algorithms. Furthermore, Li et al. [239] tackled the OPF problem as a multi-objective optimization challenge, developing a spatiotemporal DDPG-based algorithm to dynamically search for OPF solutions. Their experimental evaluations on IEEE 33-, 69-, and 118-bus systems underscored the proposed framework's effectiveness in enhancing power system robustness amidst voltage fluctuations arising from renewable energy resource uncertainty.

E. PPO

Proximal Policy Optimization (PPO) plays an important role in deep RL advancement, particularly in addressing challenges associated with policy optimization in complex environments like power systems with their multi-dimensional control tasks [240]. Originating from the family of policy gradient methods, the PPO maintains a fundamental objective of enhancing sample efficiency and stability while learning optimal policies. While DDPG employs an actor-critic architecture with continuous action spaces, PPO operates directly on the policy space, simplifying implementation and training procedures. This algorithm operates by iteratively optimizing a parameterized policy function, denoted as π_θ, where θ represents the policy parameters. The central objective of PPO revolves around maximizing the expected cumulative reward, commonly formulated as the expected return $J(\theta) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t r_t]$, given a policy parameterization θ. Here, $\mathbb{E}$ denotes the expectation operator, γ represents the discount factor, and r_t signifies the immediate reward at time step t [241], [242].

The PPO enhances policy optimization by introducing a surrogate objective function that constrains policy updates to prevent drastic policy changes. The surrogate objective, $L^{\text{CLIP}}(\theta)$, is formulated as the clipped probability ratio between the new and old policies, weighed by an advantage function $A(s, a)$:

$$L^{\text{CLIP}}(\theta) = \mathbb{E}[\min(r_t(\theta) * A_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon) * A_t)] \tag{16}$$

where $r_t(\theta)$ represents the ratio of the probabilities of actions taken under the new and old policies. Here, the parameter ϵ serves as a hyperparameter controlling the extent of clipping. Moreover, PPO incorporates an entropy term into its objective function to promote exploration and prevent premature convergence to suboptimal policies. The overall objective function of PPO is a weighted combination of the clipped surrogate objective and the entropy term:

$$L(\theta) = \mathbb{E}[L^{\text{CLIP}}(\theta) - c * H(\pi_\theta)],$$

where $H(\pi_\theta)$ denotes the entropy of the policy distribution and c represents the coefficient governing the trade-off between exploration and exploitation.

In the domain of spatiotemporal analysis within power systems, the PPO algorithm has attracted notable attention in recent studies. For instance, Shi et al. [240] introduce a Multi-Agent PPO (MAPPO) algorithm tailored for local power grids amidst the uncertainties posed by sustainable energy sources and the flexibility inherent in electric vehicle scheduling. Their approach incorporates a GAN architecture to enrich the training data with diverse scheduling scenarios. Also, Wu et al. [243] develop a spatiotemporal graph-based MAPPO framework to train online controller policies for the optimal adjustment of distributed energy resource setpoints. Their experimental validation on the 1-minute resolution Pecan Street dataset demonstrates higher robustness compared with other benchmark methods.

Furthermore, Liang et al. [244] demonstrate the efficacy of PPO in the frequency regulation of wind turbines across multiple local farms. Their methodology frames the cooperative frequency control problem within each farm as a decentralized POMDP, subsequently employing MAPPO to address it. Likewise, Zhou et al. [245] tackle the distributed generator rescheduling challenge by formulating it as a decentralized POMDP, proposing a MAPPO algorithm to devise an optimal

rescheduling strategy that bolsters the resilience of distributed systems. Additionally, Zhang et al. [246] showcase the superior performance of MAPPO over multi-agent deep Q-learning in an energy-adaptive monitoring system tailored for smart farms.

F. PERFORMANCE COMPARISON AND DISCUSSION

Table 5 presents the numerical outcomes of various RL techniques applied to diverse tasks. As depicted, both Q-learning and SARSA-based methodologies exhibit comparable performance. For instance, in voltage control applications, GC-DDQN [212] marginally outperforms SARSA-MDP [219] by a difference of 0.046 in per unit (pu) average cumulative voltage violation. However, SARSA models demonstrate superior performance over Q-learning methods in demand response and FDI attack detection applications. For instance, in FDI attack detection, FDQN [216] outperforms FDI-SARSA [222] by 1.91%. The similar performances of SARSA and Q-learning-based strategies stem from similarities in their learning processes and reward functions. However, due to Q-learning's greedy policy improvement strategy, we observed a faster convergence rate than SARSA-based algorithms across various applications.

In the table, one can observe the superior performance of actor-critic-based methodologies compared to SARSA and Q-learning approaches. For instance, in applications such as energy management and cyber attack detection, MOAC [229] and AC-NFTSM [228] demonstrate higher performance over ELM Q-learning [211] and FDI-SARSA [222] by 2.971% and 2.637% respectively. A similar pattern persists across various other applications, suggesting the enhanced generalization capability of actor-critic models in comparison to SARSA and Q-learning. This superior performance of actor-critic-based approaches can be attributed to the dual-network architecture employed for policy and value function learning, which facilitates more efficient and rapid learning. Additionally, the table illustrates the better performance of DDPG-based models over actor-critic frameworks. For instance, in solving the constraint OPF optimization problem, MG-ASTGCN [233] outperforms Cplx-STGCN TD3 [247] by 7.93%. Similarly, in an energy management scenario, MA-OCDDPG [236] reduces the total annual cost by 10.47% compared to the MAOC [229] model. This higher performance of DDPG-based methods is attributable to the utilization of experience buffers, which enable efficient learning from past agent experiences, and the inherent robustness of the model in handling uncertain measurements in power systems.

Moreover, Table 5 presents the superiority of PPO-based frameworks over other deep RL approaches, particularly in demand response and cyber attack detection applications. For instance, in demand response scenarios, GAN-MAPPO [240] demonstrates notable improvements in operational cost compared to MG-ASTGCN [233] and A-Q-learning [209] by 17.0$ and 42.42$, respectively, based on data from the steel powder manufacturing dataset. Similarly, MAPPO-UM [246] exhibits superior performance in FDI attack detection on the IEEE 39-bus system, surpassing CDDPG [248] and AC-NFTSM [228] by 2.50% and 3.564% in terms of F1 score, respectively. This enhanced efficacy of PPO is attributed to its adeptness in managing policy changes, facilitated by the incorporation of a clipped objective function, which restricts the magnitude of policy updates per iteration, thereby ensuring stability and efficiency in training.

G. ADVANTAGES AND DISADVANTAGES OF DEEP RL ALGORITHMS

Deep RL has emerged as a promising approach for spatiotemporal analysis within power systems, offering a range of methodologies, each with distinct advantages and limitations. The SARSA [219], [220], [221], [222], a classic deep RL algorithm, presents a straightforward implementation and convergence guarantees under specific conditions, making it particularly suitable for tasks characterized by clear episodes. However, SARSA often exhibits slow convergence and high variability in the learning process, resulting in lower sample efficiency compared to other methods. Additionally, its on-policy nature limits its effectiveness in certain spatiotemporal applications such as OPF optimization, energy management in microgrids, and frequency regulation within power systems. On the other hand, Q-learning approaches [211], [214], [215], [216] offer ease of implementation and is model-free, handling stochastic environments effectively and demonstrating flexibility through its off-policy nature. Nonetheless, Q-learning is susceptible to variability in learning outcomes and may converge sluggishly to accurate Q-values. Furthermore, it is prone to overestimating action values and requires careful selection of learning rates to mitigate sensitivity issues.

Actor-critic architectures [224], [226], [228], [230], [231], [250], [251] have gained prominence for their effectiveness in continuous action spaces, offering better convergence compared to SARSA and Q-learning in application such as demand response optimization, voltage regulation, and power system restoration. By enabling independent learning of policy and value functions, actor-critic methods exhibit enhanced adaptability within spatiotemporal analysis of power systems. Nevertheless, their efficacy is contingent upon meticulous tuning of multiple hyperparameters, and instability in learning policy and value functions remains a concern. Moreover, performance variability persists across different architectural and algorithmic choices. On the other hand, DDPG-based approaches [233], [236], [237], [239] show enhanced stability through the incorporation of target networks, particularly excelling in high-dimensional continuous action spaces within power systems. Despite its advantages, DDPG frameworks often experience a high variance in the learning process, leading to slower convergence and a tendency to overestimate action values.

The PPO-based approaches [240], [244], [245], [246] introduce advancements aimed at mitigating the challenges inherent in policy gradient methods [233], [239], [248].

TABLE 5. Deep Reinforcement Learning Architectures Across Different Power Systems Applications. For each application, the best results are shown in bold and the second best results are underlined.

Application	Deep RL Algorithm	Model	Dataset	Performance metric	Result
Voltage Control	SARSA	SARSA-MDP [219]	IEEE 123-bus system	Average Cumulative Voltage Violation (pu)	0.467
	DQN	GC-DDQN [212]			0.421
	Actor-Critic	EA-MAAC [224]			0.393
	DDPG	MADDPG [249]			**0.381**
Energy Management	SARSA	ESARSA [220]	Pecan Street	Total Annual Cost ($)	1313.36
	Q Learning	ELM Q-Learning [211]			1303.85
	Actor-Critic	MOAC [229]			1274.34
	DDPG	MA-OCDDPG [236]			**1140.86**
OPF	DDPG	MG-ASTGCN [239]	IEEE 30-Bus System	Normalized Time and Setpoint Satisfaction Function	**0.453**
	Actor-Critic	Cplx-STGCN TD3 [247]			0.492
Demand Response	SARSA	ASARSA [221]	PJM Electricity Market	Operation Cost ($)	171.32
	Q Leaning	A-Q-Learning [209]			183.74
	Actor-Critic	MAAC [231]			162.68
	DDPG	MG-ASTGCN [233]			158.32
	PPO	GAN-MAPPO [240]			**141.32**
Cyber	SARSA	FDI-SARSA [222]	IEEE 39-Bus System	F1 Score (%)	90.786
	DQN	FDQN [216]			88.876
	Actor-Critic	AC-NFTSM [228]			93.423
	DDPG	CDDPG [248]			94.487
	PPO	MAPPO-UM [246]			**96.987**

By specifically addressing the issue of high variance, the PPO enhances training stability through the utilization of a clipped objective function, thereby ensuring more consistent learning dynamics. This stability contributes to PPO's reputation for achieving good sample efficiency and robustness, which are essential characteristics in the context of power system analysis where data may be limited or noisy. However, like other deep reinforcement learning algorithms, PPO demands careful tuning of hyperparameters to achieve optimal performance. Furthermore, its efficacy may vary depending on the specifics of the underlying environment and task, necessitating thorough experimentation and adaptation to ensure reliable results within the context of spatiotemporal analysis in power systems.

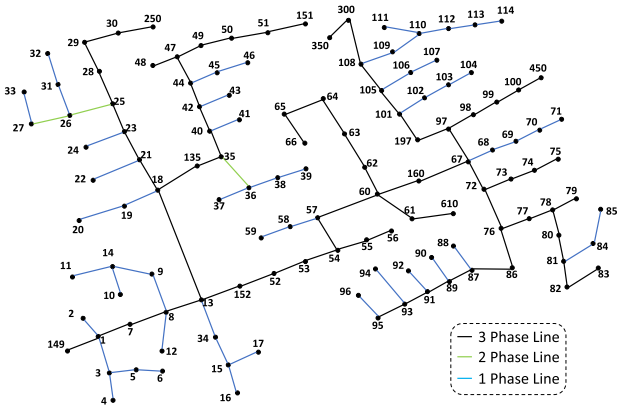

FIGURE 10. IEEE 123-bus system diagram.

V. DATASETS

In Sections III, IV, and V, we presented the numerical results of various deep data-driven models applied to various spatiotemporal tasks in the power system domain. This section provides a comprehensive overview of different datasets utilized in our experiments.

- **IEEE 123-bus system:** For fault identification, we utilized simulated data from the IEEE 123-bus system generated by DigSilent PowerFactory [252]. Fig. 10 provides a general diagram of the IEEE 123-bus system. In our experiments, we defined several fault classes, including single-phase-to-ground, double-phase-to-ground, two-phase faults, and three-phase faults. Each fault was initiated at 0.1 seconds and cleared at 0.105. The dataset comprises a total of 26, 200 fault samples.
- **IEEE 300-bus system:** We utilized simulated data from the IEEE 300-bus system for the FDI attack detection task as a benchmark model representing a large-scale power network. We simulated various scenarios where false data is injected into different measurement points across the network. These scenarios include both random and targeted attacks on bus voltages

and line flows. The considered dataset consists of comprehensive records, including normal operational states and multiple attack scenarios, amounting to 50, 000 samples.

- **IEEE New England 39-bus system:** This simulated dataset is employed for the event classification task. It includes 1000 samples, each encapsulating detailed measurements and system states such as bus voltages, angles, and power flows across different network components. Each sample is annotated with event types, and the total number of samples in this study is considered to be 35, 000.
- **IEEE New England 10-generator 39-bus system:** For the voltage stability assessment application, we simulate a dataset from the IEEE New England 10-generator 39-bus system that includes time-series data of voltage magnitudes and angles for all 39 buses, generator outputs, and power flows, recorded at one-minute intervals over one year, totaling 525,600 data points per variable. The dataset incorporates spatial metadata such as bus coordinates, line connectivity, and varied load profiles to reflect realistic operational conditions. Contingencies and disturbances are synthetically introduced to provide diverse stability events.
- **UK-DALE [253]:** This dataset is utilized for NILM application in the power system that comprises detailed electricity usage data from five households in the UK. It includes high-frequency recordings of aggregate power consumption at 1 Hz and sub-metered appliance-level data at varying frequencies, ranging from 6 seconds to 1 minute, collected over several years. This results in 16 millions of data points per household, providing extensive temporal resolution and granularity.
- **Wind Integration National Dataset [254]**: To perform spatiotemporal wind speed prediction, we utilized the Wind Integration National Dataset provided by the National Renewable Energy Laboratory (NREL). This dataset comprises high-resolution wind speed data recorded at 5-minute intervals over a ten-year period, spanning more than 120,000 geographic locations across the United States, resulting in over 12 billion data points. Additionally, the dataset includes essential meteorological variables such as temperature, pressure, and humidity, which are crucial for accurate forecasting. For our experiments, we used historical spatiotemporal wind speed data from 2007 to 2012 to train the models and evaluate their performance on the data from 2013.
- **Solar Integration National Dataset [255]**: This dataset, provided by the NREL, is employed for spatiotemporal PV forecasting. The dataset includes high-resolution solar irradiance data recorded at 30-minute intervals over ten years, covering over 100, 000 geographic locations across the United States. The dataset also encompasses critical meteorological variables such as temperature, cloud cover, and humidity, essential for accurate PV forecasting. Spatial metadata includes coordinates and elevation of each location, enhancing the spatial analysis. In our experiments, we split the dataset by seasons. For each season, we consider 80% of samples as training and validation sets and 20% as testing set. The entire seasonal training and testing sets contain 21, 024 and 5, 256 samples.
- **Low Carbon London [256]:** The Low Carbon London dataset, provided by UK Power Networks, is utilized for spatiotemporal demand forecasting. This dataset includes energy consumption readings from 5, 567 London households, collected at half-hour intervals from November 2011 to February 2014, resulting in approximately 167 million rows. The dataset records energy consumption in kWh, unique household identifiers, dates, and times. In our experiments, we utilized data from 2011 to 2013 as the training set and data from 2014 as the test set.
- **PJM Electricity Market [257]:** Similar to [210], we employed the PJM electricity market dataset for spatiotemporal demand response in home energy management that includes price and energy data from the PJM electricity market. The dataset spans from January 1, 2016, to February 21, 2017, for training purposes, and the model predictions cover February 22-28, 2017. It contains high-frequency recordings of electricity prices and energy consumption data at half-hour intervals.
- **Pecan Street [258]**: The Pecan Street dataset, employed for spatiotemporal energy management, provides detailed electricity consumption and generation data from over 1,000 residential units. This dataset includes high-frequency recordings at one-minute intervals, capturing variables such as energy usage, solar generation, and appliance-level consumption. For our experiments, we focused on data from 100 residential units, utilizing 80% of each month's samples for training and validation, while the remaining 20% constituted our test set.

VI. FUTURE DEVELOPMENT

Over the last decade, significant efforts have been made to integrate spatiotemporal feature learning into various aspects of power system analysis. While remarkable progress has been achieved, there remains ample opportunity to enhance the efficacy of spatiotemporal approaches by harnessing state-of-the-art machine learning techniques. This section introduces these advanced methodologies, elucidating their potential to further refine the performance of such approaches.

A. NORMALIZING FLOW

As discussed in this paper, GANs and VAEs have demonstrated impressive performance results on challenging tasks within spatiotemporal analysis of power systems. GANs often struggle with mode collapse, where the generator produces similar samples, limiting their ability to capture the diverse and nuanced patterns present in power system

data. Additionally, training GANs can be unstable, requiring careful tuning of hyperparameters and architectural choices to achieve desirable results [259], [260]. VAEs, on the other hand, face challenges in generating high-fidelity samples due to the variational lower bound used in the training objective, which may lead to blurry or less realistic outputs. Moreover, VAEs often struggle to capture long-range dependencies and complex structures in spatiotemporal data, limiting their effectiveness in modeling the intricate dynamics of power systems [261], [262].

To address these limitations, the adoption of the normalizing flow framework presents a compelling alternative for spatiotemporal analysis of power systems. Normalizing flows [263] are a class of generative models used to learn complex probability distributions. At their core, they transform a simple base distribution, such as a Gaussian, into a more complex distribution through a series of invertible transformations. These transformations are typically parameterized by neural networks, allowing for flexibility in modeling complex distributions. Normalizing flows offer several advantages, including exact likelihood evaluation and invertibility, which facilitate more accurate modeling of complex distributions without sacrificing fidelity or interpretability. By employing invertible transformations parameterized by neural networks, normalizing flows can effectively capture the intricate spatiotemporal dependencies and heterogeneity present in power system data. Furthermore, the explicit likelihood estimation provided by normalizing flows enables principled uncertainty quantification, essential for robust decision-making in power system operations and planning [264]. Leveraging the flexibility and expressiveness of normalizing flows, researchers can enhance the performance of spatiotemporal analysis tasks such as demand forecasting, anomaly detection, and grid optimization, ultimately contributing to the development of more efficient and resilient power systems.

B. PHYSICS-INFORMED

Conventional machine learning frameworks often face challenges due to the inherent complexity and dynamic nature of power system data, which includes a multitude of interdependent variables and physical constraints. Traditional machine learning models may struggle to capture the underlying physics and causal relationships governing power system behavior, leading to suboptimal performance and limited interpretability [131], [135], [265]. Additionally, conventional approaches typically rely on large amounts of labeled data for training, which may be scarce or costly to obtain in the context of power systems.

In response to these challenges, there is growing interest in adopting physics-informed machine learning frameworks for spatiotemporal analysis of power systems [266], [267]. By integrating domain knowledge and physical principles into the learning process, physics-informed models offer several key advantages over conventional approaches. These models can effectively capture the underlying physics and constraints of power system dynamics, leading to more accurate and interpretable predictions. Furthermore, physics-informed machine learning enables the incorporation of prior knowledge and constraints into the learning process, reducing the need for extensive labeled data and enhancing the generalization capabilities of the models. Recent studies have shown the merits of these models in a variety of fields, including fluid dynamics [268], biomedical engineering [269], and power system [270]. By leveraging the complementary strengths of machine learning and physics-based modeling, researchers can develop robust and reliable solutions for various spatiotemporal analysis tasks in power systems, including load forecasting, fault detection, and renewable energy integration.

C. EXPLAINABLE AI

While different ML frameworks have been applied to different power system applications, most of the developed frameworks suffer from a lack of interpretability. Explainable artificial intelligence (XAI) refers to the capability of artificial intelligence systems to provide transparent, interpretable, and understandable explanations for their decisions or predictions [271], [272]. XAI assists humans in comprehending the process by which a machine algorithm generates its output. It aids in assessing the correctness, fairness, and transparency of models, thereby facilitating AI-assisted decision-making. XAI plays a crucial role in fostering trust and confidence among organizations when utilizing AI models. In the context of spatiotemporal power system analysis, where complex interactions among various components, such as generators, transmission lines, and loads, occur over time and space, the need for XAI becomes significant. Interpretability in AI models allows power system operators and engineers to comprehend the reasoning behind AI-driven insights, facilitating informed decision-making and enhancing trust in AI-based solutions [273].

Three stages of explainability can be considered. (1) Before modeling, techniques such as visualization, domain-based and model-based feature engineering [274], [275], data summarization, and exploratory data analysis [276] pave the way by providing insights into the data's characteristics and relationships. (2) Within the model architecture, mechanisms like self-attention [277] and multi-head attention [278] enhance interpretability by allowing the model to focus on relevant features and relationships, promoting modularity and sparsity to simplify understanding. (3) Post-modeling, interpretability is further refined through prediction-level methods such as feature importance analysis [274], accumulated local effects (ALE) plots [279], individual conditional expectation (ICE) [280], and partial dependence plots (PDP) [281] to illuminate the impact of features on predictions.

Additionally, techniques such as feature importance analysis, SHapley Additive exPlanations (SHAP) values [273], Local Interpretable Model-agnostic Explanations (LIME) [282], and surrogate models contribute to the enhancement

of explainability [276]. Feature importance analysis assesses the contribution of each input variable to the model's predictions, providing insights into the factors driving the outcomes. SHAP values offer a game-theoretic approach to quantifying the impact of features on predictions, facilitating a deeper understanding of the model's behavior. LIME generates local interpretations by approximating the model's behavior around specific instances, aiding in understanding predictions at an individual level. Surrogate models, which are simpler models trained to mimic the behavior of the primary model, offer a more interpretable representation of the underlying decision logic. Integrating these methodologies enables practitioners to develop models that deliver accurate predictions and provide transparent explanations, fostering trust and facilitating informed decision-making across diverse applications.

D. DOMAIN ADAPTATION

Conventional machine learning frameworks often face challenges in the spatiotemporal analysis of power systems due to domain shifts and data heterogeneity [56], [72], [76], [92]. These frameworks typically rely on labeled training data collected from specific domains or conditions, which may not fully represent the diverse operating scenarios and environmental conditions encountered in real-world power systems. As a result, models trained on one set of data may struggle to generalize to unseen domains or adapt to changes in operating conditions, leading to poor performance and limited applicability. Moreover, conventional machine learning approaches may require large amounts of labeled data for each target domain, which can be impractical or costly to obtain in the context of power system analysis. Additionally, these models may fail to leverage valuable information from related domains or historical data, further hindering their ability to capture the complex spatiotemporal dynamics of power systems.

To address these challenges, domain adaptation, and transfer learning frameworks offer a promising approach for enhancing the spatiotemporal analysis of power systems. These frameworks aim to leverage knowledge from related domains or auxiliary data sources to improve model generalization and adaptation to new domains or operating conditions. By learning transferable representations from source domains, domain adaptation methods enable models to generalize better to target domains with limited labeled data.

Deep domain adaptation techniques can be broadly categorized into three main categories: (1) Discrepancy-based methods [283], [284] aim to minimize the distributional difference between the source and target domains by directly measuring the dissimilarity between their feature distributions. (2) Reconstruction-based techniques [285], [286] focus on reconstructing the input data from the learned representations and leveraging autoencoders or generative models to encourage domain-invariant representations. (3) Adversarial-based approaches [287], [288] introduce domain adversarial learning, where a domain discriminator is trained to distinguish between source and target domain samples, while the feature extractor aims to fool this discriminator by learning domain-invariant representations. These models address domain shifts by aligning feature distributions, encouraging domain-invariant representations, facilitating knowledge transfer from a labeled source domain to an unlabeled or sparsely labeled target domain. Multiple recent studies have developed robust domain adaptation techniques in different applications including traffic scene understanding [42], [43], medical imaging [289], [290], fault diagnosis [291], [292], etc. By incorporating domain adaptation and transfer learning techniques into spatiotemporal analysis of power systems, researchers can mitigate the challenges associated with domain shifts, data scarcity, and model generalization, ultimately leading to more robust and effective predictive models for power system operations and planning.

E. FEDERATED LEARNING

The conventional machine learning framework faces significant challenges in the spatiotemporal analysis of power systems due to privacy concerns and data decentralization. Traditional machine learning methods often require centralizing sensitive data from various sources for model training, which can be impractical or raise privacy issues in the context of power systems, where data comes from diverse geographical locations [3], [64], [70], [83]. Federated learning [293] addresses these challenges by allowing model training to be performed locally on distributed data sources (such as smart meters, sensors, and power generators) without sharing raw data. Instead, only model updates or aggregated information is exchanged between devices or nodes. This decentralized approach not only preserves data privacy and security but also enables the analysis of spatiotemporal patterns across the power system while respecting regulatory constraints. Furthermore, federated learning can improve the robustness and generalization of models by leveraging the diversity of data across different locations and time periods. It allows for the incorporation of local insights and variations into the learning process, leading to more accurate predictions and better adaptation to changing conditions within the power system.

Generally, federated learning approaches can be categorized into three main classes [294]: (1) horizontal federated learning, where individual power grid operators can collaborate without sharing sensitive data, pooling their resources to collectively train models that capture spatiotemporal patterns across different regions. This allows for the creation of robust predictive models capable of forecasting power demand, identifying anomalies, and optimizing grid operations while maintaining data privacy. (2) vertical federated learning enables collaboration between different entities within the power system, such as utilities and renewable energy providers, facilitating the integration of diverse data sources for more comprehensive analyses. and (3) federated transfer

learning that facilitates transferring the insights gained from one region or aspect of the power grid, accelerates model refinement and improves performance across the entire system [295], [296]. Several recent studies have shown practical advantages of federated learning frameworks in healthcare systems [297], [298], natural language processing [299], [300], and the Internet of Things [301], [302]. Utilizing federated learning techniques, future research endeavors in the spatiotemporal analysis of power systems can derive actionable insights from extensive and diverse spatiotemporal data, all while prioritizing data privacy and security.

VII. CONCLUSION

This survey paper provides a comprehensive overview of the applications of deep learning algorithms in spatiotemporal analysis of power systems. By categorizing deep machine learning frameworks into discriminative, generative, and reinforcement learning, we have presented a structured examination of various methodologies, their mathematical formulations, and their respective advantages and limitations. By exploring each category, we have elucidated how different frameworks address the complexities inherent in spatiotemporal data analysis within power systems. Through empirical evaluations, we have scrutinized the performance of these methods across diverse spatiotemporal applications, offering insights into their efficacy and applicability. Furthermore, our discussion extends beyond current practices, as we have identified emerging topics within the realm of machine learning that hold promise for future endeavors in the deep spatiotemporal analysis of power systems. In essence, this survey serves as a valuable resource for researchers and practitioners alike, offering a comprehensive understanding of the state-of-the-art methodologies while also pointing towards exciting avenues for future research and development in the intersection of deep machine learning and power systems operation and analysis.

REFERENCES

[1] Y. Hao, X. Wang, J. Wang, and W. Yang, "A new perspective of wind speed forecasting: Multi-objective and model selection-based ensemble interval-valued wind speed forecasting system," *Energy Convers. Manag.*, vol. 299, Jan. 2024, Art. no. 117868.

[2] J. Wang, X. Niu, L. Zhang, Z. Liu, and X. Huang, "A wind speed forecasting system for the construction of a smart grid with two-stage data processing based on improved ELM and deep learning strategies," *Exp. Syst. Appl.*, vol. 241, May 2024, Art. no. 122487.

[3] L. Massidda, F. Bettio, and M. Marrocu, "Probabilistic day-ahead prediction of PV generation. A comparative analysis of forecasting methodologies and of the factors influencing accuracy," *Sol. Energy*, vol. 271, Mar. 2024, Art. no. 112422.

[4] D.-S. Lee and S.-Y. Son, "PV forecasting model development and impact assessment via imputation of missing PV power data," *IEEE Access*, vol. 12, pp. 12843–12852, 2024.

[5] A. Cano, P. Arévalo, D. Benavides, and F. Jurado, "Integrating discrete wavelet transform with neural networks and machine learning for fault detection in microgrids," *Int. J. Electr. Power Energy Syst.*, vol. 155, Jan. 2024, Art. no. 109616.

[6] M. M. Zaben, M. Y. Worku, M. A. Hassan, and M. A. Abido, "Machine learning methods for fault diagnosis in AC microgrids: A systematic review," *IEEE Access*, vol. 12, pp. 20260–20298, 2024.

[7] M. Khosravi, M. Trik, and A. Ansari, "Diagnosis and classification of disturbances in the power distribution network by phasor measurement unit based on fuzzy intelligent system," *J. Eng.*, vol. 2024, no. 1, Jan. 2024, Art. no. e12322.

[8] S. Awasthi, G. Singh, and N. Ahamad, "Classifying electrical faults in a distribution system using K-Nearest neighbor (KNN) model in presence of multiple distributed generators," *J. Inst. Eng. (India), Ser. B*, vol. 105, no. 3, pp. 621–634, Jun. 2024.

[9] T. Chugh, K. Tyagi, and R. Seth, "Machine learning techniques for non-intrusive load monitoring for enhanced energy predictions," in *Proc. IEEE 14th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2024, pp. 0322–0328.

[10] B. Kommey, E. Tamakloe, J. J. Kponyo, E. T. Tchao, A. S. Agbemenu, and H. Nunoo-Mensah, "An artificial intelligence-based non-intrusive load monitoring of energy consumption in an electrical energy system using a modified K-Nearest neighbour algorithm," *IET Smart Cities*, Jan. 2024.

[11] H. Rafiq, P. Manandhar, E. Rodriguez-Ubinas, O. A. Qureshi, and T. Palpanas, "A review of current methods and challenges of advanced deep learning-based non-intrusive load monitoring (NILM) in residential context," *Energy Buildings*, vol. 305, Feb. 2024, Art. no. 113890.

[12] M. Cordeiro-Costas, D. Villanueva, P. Eguía-Oller, M. Martínez-Comesaña, and S. Ramos, "Load forecasting with machine learning and deep learning methods," *Appl. Sci.*, vol. 13, no. 13, p. 7933, Jul. 2023.

[13] S. S. Fazlhashemi, M. E. Khodayar, M. Sedighizadeh, and M. Khodayar, "Decentralized robust operation of the unbalanced microgrids in distribution networks: A convex relaxation approach," *Electric Power Syst. Res.*, vol. 229, Apr. 2024, Art. no. 110087.

[14] Y. Zhang, J. Zhao, D. Shi, and S. Chung, "Deep reinforcement learning-enabled adaptive forecasting-aided state estimation in distribution systems with multi-source multi-rate data," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Feb. 2024, pp. 1–5.

[15] D. A. Weiss, A. M. Borsa, A. Pala, A. J. Sederberg, and G. B. Stanley, "A machine learning approach for real-time cortical state estimation," *J. Neural Eng.*, vol. 21, no. 1, Feb. 2024, Art. no. 016016.

[16] Y. He, S. Chai, Z. Xu, C. S. Lai, and X. Xu, "Power system state estimation using conditional generative adversarial network," *IET Gener., Transmiss. Distrib.*, vol. 14, no. 24, pp. 5823–5833, Dec. 2020.

[17] Y. Huang, Q. Xu, C. Hu, Y. Sun, and G. Lin, "Probabilistic state estimation approach for AC/MTDC distribution system using deep belief network with non-Gaussian uncertainties," *IEEE Sensors J.*, vol. 19, no. 20, pp. 9422–9430, Oct. 2019.

[18] S. Y. Diaba, M. Shafie-Khah, and M. Elmusrati, "Cyber security in power systems using meta-heuristic and deep learning algorithms," *IEEE Access*, vol. 11, pp. 18660–18672, 2023.

[19] T. Mazhar, H. M. Irfan, S. Khan, I. Haq, I. Ullah, M. Iqbal, and H. Hamam, "Analysis of cyber security attacks and its solutions for the smart grid using machine learning and blockchain methods," *Future Internet*, vol. 15, no. 2, p. 83, Feb. 2023.

[20] N. Tatipatri and S. L. Arun, "A comprehensive review on cyber-attacks in power systems: Impact analysis, detection, and cyber security," *IEEE Access*, vol. 12, pp. 18147–18167, 2024.

[21] M. M. Rahman, M. Ali, A. Rahman, and W. Sun, "A real-time cyber-physical HIL testbed for cybersecurity in distribution grids with DERs," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Feb. 2024, pp. 1–5.

[22] M. Khodayar, G. Liu, J. Wang, and M. E. Khodayar, "Deep learning in power systems research: A review," *CSEE J. Power Energy Syst.*, vol. 7, no. 2, pp. 209–220, Mar. 2021.

[23] S. Chahboun and M. Maaroufi, "Principal component analysis and machine learning approaches for photovoltaic power prediction: A comparative study," *Appl. Sci.*, vol. 11, no. 17, p. 7943, Aug. 2021.

[24] P. K. Ray, A. Mohanty, and T. Panigrahi, "Power quality analysis in solar PV integrated microgrid using independent component analysis and support vector machine," *Optik*, vol. 180, pp. 691–698, Feb. 2019.

[25] G. Singh, Y. Pal, and A. K. Dahiya, "Classification of power quality disturbances using linear discriminant analysis," *Appl. Soft Comput.*, vol. 138, May 2023, Art. no. 110181.

[26] S. Tasnim, A. Rahman, A. M. T. Oo, and M. Enamul Haque, "Autoencoder for wind power prediction," *Renewables, Wind, Water, Sol.*, vol. 4, no. 1, pp. 1–11, Dec. 2017.

[27] S. Afrasiabi, M. Afrasiabi, B. Behdani, M. Mohammadi, M. S. Javadi, G. J. Osório, and J. P. S. Catalão, "Photovoltaic array fault detection and classification based on T-distributed stochastic neighbor embedding and robust soft learning vector quantization," in *Proc. IEEE Int. Conf. Environ. Electr. Eng. IEEE Ind. Commercial Power Syst. Eur.*, Sep. 2021, pp. 1–5.

[28] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[29] H. Hwang and S. Kang, "Nonintrusive load monitoring using an LSTM with feedback structure," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[30] T. Wu, I. L. Carreño, A. Scaglione, and D. Arnold, "Spatio-temporal graph convolutional neural networks for physics-aware grid learning algorithms," *IEEE Trans. Smart Grid*, vol. 14, no. 5, pp. 4086–4099, Sep. 2023.

[31] M. Saffari, M. Khodayar, S. M. J. Jalali, M. Shafie-Khah, and J. P. S. Catalão, "Deep convolutional graph rough variational auto-encoder for short-term photovoltaic power forecasting," in *Proc. Int. Conf. Smart Energy Syst. Technol. (SEST)*, Sep. 2021, pp. 1–6.

[32] J. Xie, A. Rahman, and W. Sun, "Bayesian GAN-based false data injection attack detection in active distribution grids with DERs," *IEEE Trans. Smart Grid*, vol. 15, no. 3, pp. 3223–3234, May 2024.

[33] R. Yuan, B. Wang, Z. Mao, and J. Watada, "Multi-objective wind power scenario forecasting based on PG-GAN," *Energy*, vol. 226, Jul. 2021, Art. no. 120379.

[34] Y. Cheng, N. Yu, B. Foggo, and K. Yamashita, "Online power system event detection via bidirectional generative adversarial networks," *IEEE Trans. Power Syst.*, vol. 37, no. 6, pp. 4807–4818, Nov. 2022.

[35] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017.

[36] Q. Huang, R. Huang, W. Hao, J. Tan, R. Fan, and Z. Huang, "Adaptive power system emergency control using deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1171–1182, Mar. 2020.

[37] X. Liu and C. Konstantinou, "Reinforcement learning for cyber-physical security assessment of power systems," in *Proc. IEEE Milan PowerTech*, Jun. 2019, pp. 1–6.

[38] A. Sheikhi, M. Rayati, and A. M. Ranjbar, "Dynamic load management for a residential customer; reinforcement learning approach," *Sustain. Cities Soc.*, vol. 24, pp. 42–51, Jul. 2016.

[39] Z. Yan and Y. Xu, "Real-time optimal power flow: A Lagrangian based deep reinforcement learning approach," *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 3270–3273, Jul. 2020.

[40] M. S. E. Saadabadi, S. R. Malakshan, H. Kashiani, and N. M. Nasrabadi, "CCFace: Classification consistency for low-resolution face recognition," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Sep. 2023, pp. 1–10.

[41] P. C. Neto, F. Boutros, J. R. Pinto, M. Saffari, N. Damer, A. F. Sequeira, and J. S. Cardoso, "My eyes are up here: Promoting focus on uncovered regions in masked face recognition," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2021, pp. 1–5.

[42] M. Saffari and M. Khodayar, "Low-rank sparse generative adversarial unsupervised domain adaptation for multitarget traffic scene semantic segmentation," *IEEE Trans. Ind. Informat.*, vol. 20, no. 2, pp. 2564–2576, Feb. 2024.

[43] M. Saffari, M. Khodayar, and S. M. J. Jalali, "Sparse adversarial unsupervised domain adaptation with deep dictionary learning for traffic scene classification," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 7, no. 4, pp. 1139–1150, Aug. 2023.

[44] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical image analysis using convolutional neural networks: A review," *J. Med. Syst.*, vol. 42, no. 11, pp. 1–13, Nov. 2018.

[45] D. R. Sarvamangala and R. V. Kulkarni, "Convolutional neural networks in medical image understanding: A survey," *Evol. Intell.*, vol. 15, no. 1, pp. 1–22, Mar. 2022.

[46] S. M. J. Jalali, M. Khodayar, S. Ahmadian, M. K. Noman, A. Khosravi, S. M. S. Islam, F. Wang, and J. P. S. Catalão, "A new uncertainty-aware deep neuroevolution model for quantifying tidal prediction," in *Proc. IEEE Ind. Appl. Soc. Annu. Meeting (IAS)*, Oct. 2021, pp. 1–6.

[47] S. M. J. Jalali, S. Ahmadian, M. Khodayar, A. Khosravi, M. Shafie-Khah, S. Nahavandi, and J. P. S. Catalão, "An advanced short-term wind power forecasting framework based on the optimized deep neural network models," *Int. J. Electr. Power Energy Syst.*, vol. 141, Oct. 2022, Art. no. 108143.

[48] M. Saffari, M. Khodayar, M. S. E. Saadabadi, A. F. Sequeira, and J. S. Cardoso, "Maximum relevance minimum redundancy dropout with informative kernel determinantal point process," *Sensors*, vol. 21, no. 5, p. 1846, Mar. 2021.

[49] I. Niazazari, R. Jalilzadeh Hamidi, H. Livani, and R. Arghandeh, "Cause identification of electromagnetic transient events using spatiotemporal feature learning," *Int. J. Electr. Power Energy Syst.*, vol. 123, Dec. 2020, Art. no. 106255.

[50] Z. Gan, C. Li, J. Zhou, and G. Tang, "Temporal convolutional networks interval prediction model for wind speed forecasting," *Electric Power Syst. Res.*, vol. 191, Feb. 2021, Art. no. 106865.

[51] M. Khodayar, J. Wang, and Z. Wang, "Deep generative graph distribution learning for synthetic power grids," 2019, *arXiv:1901.09674*.

[52] S. M. J. Jalali, S. Ahmadian, B. Nakisa, M. Khodayar, A. Khosravi, S. Nahavandi, S. M. S. Islam, M. Shafie-Khah, and J. P. S. Catalão, "Solar irradiance forecasting using a novel hybrid deep ensemble reinforcement learning algorithm," *Sustain. Energy, Grids Netw.*, vol. 32, Dec. 2022, Art. no. 100903.

[53] J. Regan and M. Khodayar, "A triplet graph convolutional network with attention and similarity-driven dictionary learning for remote sensing image retrieval," *Exp. Syst. Appl.*, vol. 232, Dec. 2023, Art. no. 120579.

[54] G. Zhang, J. Li, O. Bamisile, D. Cai, W. Hu, and Q. Huang, "Spatio-temporal correlation-based false data injection attack detection using deep convolutional neural network," *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 750–761, Jan. 2022.

[55] G. Zhang, J. Li, O. Bamisile, Z. Zhang, D. Cai, and Q. Huang, "A data driven threat-maximizing false data injection attack detection method with spatio-temporal correlation," in *Proc. IEEE/IAS Ind. Commercial Power Syst. Asia*, Jul. 2021, pp. 318–325.

[56] M. Lu, L. Wang, Z. Cao, Y. Zhao, and X. Sui, "False data injection attacks detection on power systems with convolutional neural network," *J. Phys., Conf.*, vol. 1633, no. 1, Sep. 2020, Art. no. 012134.

[57] W. Qiu, Q. Tang, Y. Wang, L. Zhan, Y. Liu, and W. Yao, "Multi-view convolutional neural network for data spoofing cyber-attack detection in distribution synchrophasors," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3457–3468, Jul. 2020.

[58] H. Wang, H. Wang, G. Jiang, Y. Wang, and S. Ren, "A multiscale spatio-temporal convolutional deep belief network for sensor fault detection of wind turbine," *Sensors*, vol. 20, no. 12, p. 3580, Jun. 2020.

[59] Q. Zhang, W. Ma, G. Li, J. Ding, and M. Xie, "Fault diagnosis of power grid based on variational mode decomposition and convolutional neural network," *Electric Power Syst. Res.*, vol. 208, Jul. 2022, Art. no. 107871.

[60] Z. Li, Z. Jiao, and A. He, "Knowledge-based convolutional neural networks for transformer protection," *CSEE J. Power Energy Syst.*, vol. 7, no. 2, pp. 270–278, Mar. 2021.

[61] B. Hao and T. Li, "Fault diagnosis using deep convolutional neural network with dynamic computer-visualised power flow," in *Proc. IEEE 3rd Conf. Energy Internet Energy Syst. Integr. (EI)*, Nov. 2019, pp. 749–753.

[62] S. Turizo, G. Ramos, and D. Celeita, "Voltage sags characterization using fault analysis and deep convolutional neural networks," *IEEE Trans. Ind. Appl.*, vol. 58, no. 3, pp. 3333–3341, May 2022.

[63] S. Basumallik, R. Ma, and S. Eftekharnejad, "Packet-data anomaly detection in PMU-based state estimator using convolutional neural network," *Int. J. Electr. Power Energy Syst.*, vol. 107, pp. 690–702, May 2019.

[64] Z. Zhang, Y. Li, J. Duan, Y. Duan, Y. Guo, Y. Cao, and C. Rehtanz, "A non-intrusive load state identification method considering non-local spatiotemporal feature," *IET Gener., Transmiss. Distrib.*, vol. 16, no. 4, pp. 792–803, Feb. 2022.

[65] H. Liu, C. Liu, H. Zhao, H. Tian, J. Liu, and L. Tian, "Non-intrusive load monitoring method for multi-energy coupling appliances considering spatio-temporal coupling," *IEEE Trans. Smart Grid*, vol. 14, no. 6, pp. 4519–4529, Nov. 2023.

[66] Q. Wu and F. Wang, "Concatenate convolutional neural networks for non-intrusive load monitoring across complex background," *Energies*, vol. 12, no. 8, p. 1572, Apr. 2019.

[67] A. Moradzadeh, B. Mohammadi-Ivatloo, M. Abapour, A. Anvari-Moghaddam, S. G. Farkoush, and S.-B. Rhee, "A practical solution based on convolutional neural network for non-intrusive load monitoring," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 10, pp. 9775–9789, Oct. 2021.

[68] F. C. Ferraz, R. V. A. Monteiro, R. F. S. Teixeira, and A. S. Bretas, "A Siamese CNN+ KNN based classification framework for non-intrusive load monitoring," *J. Control, Automat. Elect. Syst.*, vol. 34, no. 4, pp. 842–857, 2023.

[69] T. Chen, H. Qin, X. Li, W. Wan, and W. Yan, "A non-intrusive load monitoring method based on feature fusion and SE-ResNet," *Electronics*, vol. 12, no. 8, p. 1909, Apr. 2023.

[70] L. Yin and J. Xie, "Multi-temporal-spatial-scale temporal convolution network for short-term load forecasting of power systems," *Appl. Energy*, vol. 283, Feb. 2021, Art. no. 116328.

[71] J. Jeong and H. Kim, "Multi-site photovoltaic forecasting exploiting space-time convolutional neural network," *Energies*, vol. 12, no. 23, p. 4490, Nov. 2019.

[72] C. Feng, J. Zhang, W. Zhang, and B.-M. Hodge, "Convolutional neural networks for intra-hour solar forecasting based on sky image sequences," *Appl. Energy*, vol. 310, Mar. 2022, Art. no. 118438.

[73] T. Hu, W. Wu, Q. Guo, H. Sun, L. Shi, and X. Shen, "Very short-term spatial and temporal wind power forecasting: A deep learning approach," *CSEE J. Power Energy Syst.*, vol. 6, no. 2, pp. 434–443, Jun. 2020.

[74] K. Higashiyama, Y. Fujimoto, and Y. Hayashi, "Feature extraction of NWP data for wind power forecasting using 3D-convolutional neural networks," *Energy Proc.*, vol. 155, pp. 350–358, Nov. 2018.

[75] Y.-Y. Hong and T. R. A. Satriani, "Day-ahead spatiotemporal wind speed forecasting using robust design-based deep learning neural network," *Energy*, vol. 209, Oct. 2020, Art. no. 118441.

[76] S. Sun, Y. Liu, Q. Li, T. Wang, and F. Chu, "Short-term multi-step wind power forecasting based on spatio-temporal correlations and transformer neural networks," *Energy Convers. Manag.*, vol. 283, May 2023, Art. no. 116916.

[77] A. Uselis, M. Lukoševičius, and L. Stasytis, "Localized convolutional neural networks for geospatial wind forecasting," *Energies*, vol. 13, no. 13, p. 3440, Jul. 2020.

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

[79] M. Cui, M. Khodayar, C. Chen, X. Wang, Y. Zhang, and M. E. Khodayar, "Deep learning-based time-varying parameter identification for system-wide load modeling," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6102–6114, Nov. 2019.

[80] S. M. J. Jalali, S. Ahmadian, M. Khodayar, A. Khosravi, V. Ghasemi, M. Shafie-Khah, S. Nahavandi, and J. P. S. Catalão, "Towards novel deep neuroevolution models: Chaotic levy grasshopper optimization for short-term wind speed forecasting," *Eng. Comput.*, vol. 38, no. 3, pp. 1787–1811, Aug. 2022.

[81] Z.-M. Zhai, L.-W. Kong, and Y.-C. Lai, "Emergence of a resonance in machine learning," *Phys. Rev. Res.*, vol. 5, no. 3, Aug. 2023, Art. no. 033127.

[82] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, "Recent advances in physical reservoir computing: A review," *Neural Netw.*, vol. 115, pp. 100–123, Jul. 2019.

[83] S. E. Razavi, A. Arefi, G. Ledwich, G. Nourbakhsh, D. B. Smith, and M. Minakshi, "From load to net energy forecasting: Short-term residential forecasting for the blend of load and PV behind the meter," *IEEE Access*, vol. 8, pp. 224343–224353, 2020.

[84] J. Zhang, L. Hong, S. N. Ibrahim, and Y. He, "Short-term prediction of behind-the-meter PV power based on attention-LSTM and transfer learning," *IET Renew. Power Gener.*, vol. 18, no. 3, pp. 321–330, Feb. 2024.

[85] A. Zaboli, V.-N. Tuyet-Doan, Y.-H. Kim, J. Hong, and W. Su, "An LSTM-SAE-based behind-the-meter load forecasting method," *IEEE Access*, vol. 11, pp. 49378–49392, 2023.

[86] A. S. Musleh, G. Chen, Z. Yang Dong, C. Wang, and S. Chen, "Spatio-temporal data-driven detection of false data injection attacks in power distribution systems," *Int. J. Electr. Power Energy Syst.*, vol. 145, Feb. 2023, Art. no. 108612.

[87] J. J. Q. Yu, Y. Hou, and V. O. K. Li, "Online false data injection attack detection with wavelet transform and deep neural networks," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3271–3280, Jul. 2018.

[88] R. Yadav and A. K. Pradhan, "PCA-LSTM learning networks with Markov chain models for online classification of cyber-induced outages in power system," *IEEE Syst. J.*, vol. 15, no. 3, pp. 3948–3957, Sep. 2021.

[89] W. Xu and F. Teng, "A deep learning based detection method for combined integrity-availability cyber attacks in power system," 2020, *arXiv:2011.01816*.

[90] K. Kim, H. Sasahara, and J.-I. Imura, "Cyberattack detection in smart grids based on reservoir computing," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 971–976, 2023.

[91] K. Park, J. Yim, H. Lee, M. Park, and H. Kim, "Real-time solar power estimation through RNN-based attention models," *IEEE Access*, vol. 12, pp. 62502–62510, 2023.

[92] X. Fu, F. Gao, J. Wu, X. Wei, and F. Duan, "Spatiotemporal attention networks for wind power forecasting," in *Proc. Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2019, pp. 149–154.

[93] Y. Cui, J. Zhang, and W. Zhong, "Short-term photovoltaic output prediction method based on similar day selection with grey relational theory," in *Proc. IEEE Innov. Smart Grid Technol.*, May 2019, pp. 792–797.

[94] H. Jahangir, H. Tayarani, S. S. Gougheri, M. A. Golkar, A. Ahmadian, and A. Elkamel, "Deep learning-based forecasting approach in smart grids with microclustering and bidirectional LSTM network," *IEEE Trans. Ind. Electron.*, vol. 68, no. 9, pp. 8298–8309, Sep. 2021.

[95] M. Yang, K. Wang, X. Su, M. Ma, G. Wu, and D. Huang, "Short-term photovoltaic output probability prediction method considering the spatio-temporal-condition dependence of prediction error," *CSEE J. Power Energy Syst.*, early access, Apr. 20, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10106199

[96] M. Dorado-Moreno, L. Cornejo-Bueno, P. A. Gutiérrez, L. Prieto, C. Hervás-Martínez, and S. Salcedo-Sanz, "Robust estimation of wind power ramp events with reservoir computing," *Renew. Energy*, vol. 111, pp. 428–437, Oct. 2017.

[97] K. Brucke, S. Schmitz, D. Köglmayr, S. Baur, C. Räth, E. Ansari, and P. Klement, "Benchmarking reservoir computing for residential energy demand forecasting," *Energy Buildings*, vol. 314, Jul. 2024, Art. no. 114236.

[98] Y. Fujimoto, M. Fujita, and Y. Hayashi, "Deep reservoir architecture for short-term residential load forecasting: An online learning scheme for edge computing," *Appl. Energy*, vol. 298, Sep. 2021, Art. no. 117176.

[99] J. Wang, T. Niu, H. Lu, W. Yang, and P. Du, "A novel framework of reservoir computing for deterministic and probabilistic wind power forecasting," *IEEE Trans. Sustain. Energy*, vol. 11, no. 1, pp. 337–349, Jan. 2020.

[100] Y. Zhang, Y. Han, C. Wang, J. Wang, and Q. Zhao, "A dynamic threshold method for wind turbine fault detection based on spatial–temporal neural network," *J. Renew. Sustain. Energy*, vol. 14, no. 5, Sep. 2022, Art. no. 053304.

[101] J. Hu, Y. Lin, J. Tang, and J. Zhao, "A new wind power interval prediction approach based on reservoir computing and a quality-driven loss function," *Appl. Soft Comput.*, vol. 92, Jul. 2020, Art. no. 106327.

[102] R. Jalalifar, M. R. Delavar, and S. F. Ghaderi, "SAC-ConvLSTM: A novel spatio-temporal deep learning-based approach for a short term power load forecasting," *Exp. Syst. Appl.*, vol. 237, Mar. 2024, Art. no. 121487.

[103] S. Chai, Z. Xu, Y. Jia, and W. K. Wong, "A robust spatiotemporal forecasting framework for photovoltaic generation," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5370–5382, Nov. 2020.

[104] K. Wang, X. Qi, and H. Liu, "Photovoltaic power forecasting based LSTM-convolutional network," *Energy*, vol. 189, Dec. 2019, Art. no. 116225.

[105] Q. Dai, X. Huo, Y. Hao, and R. Yu, "Spatio-temporal prediction for distributed PV generation system based on deep learning neural network model," *Frontiers Energy Res.*, vol. 11, Jun. 2023, Art. no. 1204032.

[106] D. Yang, T. Li, Z. Guo, and Q. Li, "Multi-scale convolutional echo state network with an effective pre-training strategy for solar irradiance forecasting," *IEEE Access*, vol. 12, pp. 13442–13452, 2024.

[107] M. Saffari, M. Williams, M. Khodayar, M. Shafie-Khah, and J. P. S. Catalão, "Robust wind speed forecasting: A deep spatio-temporal approach," in *Proc. IEEE Int. Conf. Environ. Electr. Eng. IEEE Ind. Commercial Power Syst. Eur.*, Sep. 2021, pp. 1–6.

[108] G. Yu, C. Liu, B. Tang, R. Chen, L. Lu, C. Cui, Y. Hu, L. Shen, and S. M. Muyeen, "Short term wind power prediction for regional wind farms based on spatial–temporal characteristic distribution," *Renew. Energy*, vol. 199, pp. 599–612, Nov. 2022.

[109] Y. Chen, S. Zhang, W. Zhang, J. Peng, and Y. Cai, "Multifactor spatio-temporal correlation model based on a combination of convolutional neural network and long short-term memory neural network for wind speed forecasting," *Energy Convers. Manag.*, vol. 185, pp. 783–799, Apr. 2019.

[110] Z. Kong, B. Tang, L. Deng, W. Liu, and Y. Han, "Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units," *Renew. Energy*, vol. 146, pp. 760–768, Feb. 2020.

[111] G. D'Angelo and F. Palmieri, "A stacked autoencoder-based convolutional and recurrent deep neural network for detecting cyberattacks in interconnected power control systems," *Int. J. Intell. Syst.*, vol. 36, no. 12, pp. 7080–7102, Dec. 2021.

[112] J. Ruan, C. Yang, Q. Wang, S. Wang, G. Liang, J. Zhao, and J. Qiu, "Assessment of spatiotemporally coordinated cyberattacks on renewable energy forecasting in smart energy system," *Appl. Energy*, vol. 347, Oct. 2023, Art. no. 121470.

[113] H. Ma, X. Lei, Z. Li, S. Yu, B. Liu, and X. Dong, "Deep-learning based power system events detection technology using spatio-temporal and frequency information," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 13, no. 12, pp. 545–556, Jun. 2023.

[114] A. Adhikari, S. Naetiladdanon, and A. Sangswang, "Real-time short-term voltage stability assessment using combined temporal convolutional neural network and long short-term memory neural network," *Appl. Sci.*, vol. 12, no. 13, p. 6333, Jun. 2022.

[115] M. Kaselimi, E. Protopapadakis, A. Voulodimos, N. Doulamis, and A. Doulamis, "Multi-channel recurrent convolutional neural networks for energy disaggregation," *IEEE Access*, vol. 7, pp. 81047–81056, 2019.

[116] X. Zhou, S. Li, C. Liu, H. Zhu, N. Dong, and T. Xiao, "Non-intrusive load monitoring using a CNN-LSTM-RF model considering label correlation and class-imbalance," *IEEE Access*, vol. 9, pp. 84306–84315, 2021.

[117] J. Wang, J. Wang, T. Bai, R. Zhang, Y. Ding, L. Yang, and S. Zhang, "Non-intrusive load disaggregation based on multiple optimization of appliance features and CNN-NLSTM model," *J. Electric Power Sci. Technol.*, vol. 38, no. 1, pp. 146–153, 2023.

[118] M. Khodayar, A. F. Bavil, and M. E. Khodayar, "Deep attention GRU-GRBM with dropout for fault location in power distribution networks," in *Proc. IEEE Texas Power Energy Conf. (TPEC)*, Feb. 2024, pp. 1–6.

[119] M. Khodayar, S. Mohammadi, M. E. Khodayar, J. Wang, and G. Liu, "Convolutional graph autoencoder: A generative deep neural network for probabilistic spatio-temporal solar irradiance forecasting," *IEEE Trans. Sustain. Energy*, vol. 11, no. 2, pp. 571–583, Apr. 2020.

[120] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[121] Q. Yu and Z. Li, "Correlated load forecasting in active distribution networks using spatial–temporal synchronous graph convolutional networks," *IET Energy Syst. Integr.*, vol. 3, no. 3, pp. 355–366, Sep. 2021.

[122] M. Zhang, Z. Zhen, N. Liu, H. Zhao, Y. Sun, C. Feng, and F. Wang, "Optimal graph structure based short-term solar PV power forecasting method considering surrounding spatio-temporal correlations," *IEEE Trans. Ind. Appl.*, vol. 59, no. 1, pp. 345–357, Jan. 2023.

[123] A. M. Karimi, Y. Wu, M. Koyuturk, and R. H. French, "Spatiotemporal graph neural network for performance prediction of photovoltaic power systems," in *Proc. AAAI Conf. Artif. Intell.*, May 2021, vol. 35, no. 17, pp. 15323–15330.

[124] X. Liu, Y. Zhang, Z. Zhen, F. Xu, F. Wang, and Z. Mi, "Spatio-temporal graph neural network and pattern prediction based ultra-short-term power forecasting of wind farm cluster," *IEEE Trans. Ind. Appl.*, vol. 60, no. 1, pp. 1794–1803, Jan. 2024.

[125] Z. Li, L. Ye, Y. Zhao, M. Pei, P. Lu, Y. Li, and B. Dai, "A spatiotemporal directed graph convolution network for ultra-short-term wind power prediction," *IEEE Trans. Sustain. Energy*, vol. 14, no. 1, pp. 39–54, Jan. 2023.

[126] X. Dong, Y. Sun, Y. Li, X. Wang, and T. Pu, "Spatio-temporal convolutional network based power forecasting of multiple wind farms," *J. Modern Power Syst. Clean Energy*, vol. 10, no. 2, pp. 388–398, Mar. 2022.

[127] H. Tong, R. C. Qiu, D. Zhang, H. Yang, Q. Ding, and X. Shi, "Detection and classification of transmission line transient faults based on graph convolutional neural network," *CSEE J. Power Energy Syst.*, vol. 7, no. 3, pp. 456–471, May 2021.

[128] J. Hu, W. Hu, J. Chen, D. Cao, Z. Zhang, Z. Liu, Z. Chen, and F. Blaabjerg, "Fault location and classification for distribution systems based on deep graph learning methods," *J. Modern Power Syst. Clean Energy*, vol. 11, no. 1, pp. 35–51, Jan. 2023.

[129] B. L. H. Nguyen, T. Vu, T.-T. Nguyen, M. Panwar, and R. Hovsapian, "1-D convolutional graph convolutional networks for fault detection in distributed energy systems," in *Proc. IEEE 1st Ind. Electron. Soc. Annu. Online Conf. (ONCON)*, Dec. 2022, pp. 1–6.

[130] J. Huang, L. Guan, Y. Chen, S. Zhu, L. Chen, and J. Yu, "A deep learning scheme for transient stability assessment in power system with a hierarchical dynamic graph pooling method," *Int. J. Electr. Power Energy Syst.*, vol. 141, Oct. 2022, Art. no. 108044.

[131] C. Li, A. Kies, K. Zhou, M. Schlott, O. E. Sayed, M. Bilousova, and H. Stöcker, "Optimal power flow in a highly renewable power system based on attention neural networks," *Appl. Energy*, vol. 359, Apr. 2024, Art. no. 122779.

[132] D. K. Mahto, V. K. Saini, A. Mathur, R. Kumar, and S. Verma, "MPGCN-OPF: A message passing graph convolution approach for optimal power flow for distribution network," in *Proc. IEEE Int. Conf. Power Electron., Drives Energy Syst. (PEDES)*, Dec. 2022, pp. 1–6.

[133] Z. Qu, Y. Dong, Y. Li, S. Song, T. Jiang, M. Li, Q. Wang, L. Wang, X. Bo, J. Zang, and Q. Xu, "Localization of dummy data injection attacks in power systems considering incomplete topological information: A spatio-temporal graph wavelet convolutional neural network approach," *Appl. Energy*, vol. 360, Apr. 2024, Art. no. 122736.

[134] T. Wu, A. Scaglione, and D. Arnold, "Complex-value spatiotemporal graph convolutional neural networks and its applications to electric power systems AI," *IEEE Trans. Smart Grid*, vol. 15, no. 3, pp. 3193–3207, May 2024.

[135] M. Khodayar and J. Wang, "Spatio-temporal graph deep neural network for short-term wind speed forecasting," *IEEE Trans. Sustain. Energy*, vol. 10, no. 2, pp. 670–681, Apr. 2019.

[136] M. Khodayar, O. Kaynak, and M. E. Khodayar, "Rough deep neural architecture for short-term wind speed forecasting," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2770–2779, Dec. 2017.

[137] M. Khodayar and M. Teshnehlab, "Robust deep neural network for wind speed prediction," in *Proc. 4th Iranian Joint Congr. Fuzzy Intell. Syst. (CFIS)*, Sep. 2015, pp. 1–5.

[138] X. Jiao, X. Li, D. Lin, and W. Xiao, "A graph neural network based deep learning predictor for spatio-temporal group solar irradiance forecasting," *IEEE Trans. Ind. Informat.*, vol. 18, no. 9, pp. 6142–6149, Sep. 2022.

[139] S. Arastehfar, M. Matinkia, and M. R. Jabbarpour, "Short-term residential load forecasting using graph convolutional recurrent neural networks," *Eng. Appl. Artif. Intell.*, vol. 116, Nov. 2022, Art. no. 105358.

[140] J. R. Schofield, R. Carmichael, S. Tindemans, M. Bilton, M. Woolf, and G. Strbac, "Low carbon London project: Data from the dynamic time-of-use electricity pricing trial, 2013," *UK Data Service, Social Netw.*, vol. 7857, no. 2015, pp. 7851–7857, 2015.

[141] Irish Social Science Data Archive, 2009–2010. (2012). *Cer Smart Metering Project—Electricity Customer Behaviour Trial*. [Online]. Available: https://www.ucd.ie/issda/data/commissionforenergyregulationcer/

[142] M. Khodayar, G. Liu, J. Wang, O. Kaynak, and M. E. Khodayar, "Spatiotemporal behind-the-meter load and PV power forecasting via deep graph dictionary learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 10, pp. 4713–4727, Oct. 2021.

[143] M. Saffari, M. Khodayar, M. E. Khodayar, and M. Shahidehpour, "Behind-the-meter load and PV disaggregation via deep spatiotemporal graph generative sparse coding with capsule network," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 20, 2023, doi: 10.1109/TNNLS.2023.3280078.

[144] C. Qin, A. K. Srivastava, A. Y. Saber, D. Matthews, and K. Davies, "Geometric deep-learning-based spatiotemporal forecasting for inverter-based solar power," *IEEE Syst. J.*, vol. 17, no. 3, pp. 3425–3435, Sep. 2023.

[145] M. Saffari, M. Khodayar, and M. E. Khodayar, "Deep recurrent extreme learning machine for behind-the-meter photovoltaic disaggregation," *Electr. J.*, vol. 35, no. 5, Jun. 2022, Art. no. 107137.

[146] M. Saffari, M. Khodayar, and M. Teshnehlab, "Random weights rough neural network for glaucoma diagnosis," in *Proc. 17th Int. Conf. Natural Comput., Fuzzy Syst. Knowl. Discovery (ICNC-FSKD)*. Guiyang, China: Springer, Jul. 2021, pp. 534–545.

[147] A. Ahmed, S. Basumallik, A. Gholami, S. K. Sadanandan, M. H. N. Namaki, A. K. Srivastava, and Y. Wu, "Spatio-temporal deep graph network for event detection, localization and classification in cyber-physical electric distribution system," *IEEE Trans. Ind. Informat.*, vol. 20, no. 2, pp. 2397–2407, Feb. 2024.

[148] B. L. H. Nguyen, T. V. Vu, T.-T. Nguyen, M. Panwar, and R. Hovsapian, "Spatial–temporal recurrent graph neural networks for fault diagnostics (don't short) in power distribution systems," *IEEE Access*, vol. 11, pp. 46039–46050, 2023.

[149] J. Liu, C. Yao, and L. Chen, "Time-adaptive transient stability assessment based on the gating spatiotemporal graph neural network and gated recurrent unit," *Frontiers Energy Res.*, vol. 10, Apr. 2022, Art. no. 885673.

[150] A. Presekal, A. Stefanov, V. S. Rajkumar, and P. Palensky, "Attack graph model for cyber-physical power systems using hybrid deep learning," *IEEE Trans. Smart Grid*, vol. 14, no. 5, pp. 4007–4020, Sep. 2023.

[151] R. Yan, G. Geng, Q. Jiang, and Y. Li, "Fast transient stability batch assessment using cascaded convolutional neural networks," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 2802–2813, Jul. 2019.

[152] Z. Wang and X. Zhao, "AttG-BDGNets: Attention-guided bidirectional dynamic graph IndRNN for non-intrusive load monitoring," *Information*, vol. 14, no. 7, p. 383, Jul. 2023.

[153] Y. Jie, Q. Yajuan, W. Lihui, and L. Yu, "Non-intrusive load decomposition based on graph convolutional network," in *Proc. IEEE 5th Int. Electr. Energy Conf. (CIEEC)*, May 2022, pp. 1941–1944.

[154] M. Khodayar, J. Wang, and Z. Wang, "Energy disaggregation via deep temporal dictionary learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1696–1709, May 2020.

[155] S. M. J. Jalali, M. Khodayar, S. Ahmadian, M. Shafie-Khah, A. Khosravi, S. M. S. Islam, S. Nahavandi, and J. P. S. Catalão, "A new ensemble reinforcement learning strategy for solar irradiance forecasting using deep optimized convolutional neural network models," in *Proc. Int. Conf. Smart Energy Syst. Technol. (SEST)*, Sep. 2021, pp. 1–6.

[156] M. Khodayar, M. Saffari, M. Williams, and S. M. J. Jalali, "Interval deep learning architecture with rough pattern recognition and fuzzy inference for short-term wind speed forecasting," *Energy*, vol. 254, Sep. 2022, Art. no. 124143.

[157] M. Khodayar, J. Wang, and M. Manthouri, "Interval deep generative neural network for wind speed forecasting," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3974–3989, Jul. 2019.

[158] C. Doersch, "Tutorial on variational autoencoders," 2016, *arXiv:1606.05908*.

[159] T. Cemgil, S. Ghaisas, K. Dvijotham, S. Gowal, and P. Kohli, "The autoencoding variational autoencoder," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 15077–15087.

[160] Y. Pu, Z. Gan, R. Henao, X. Yuan, C. Li, A. Stevens, and L. Carin, "Variational autoencoder for deep learning of images, labels and captions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.

[161] M. S. Mahmud, J. Z. Huang, and X. Fu, "Variational autoencoder-based dimensionality reduction for high-dimensional small-sample data classification," *Int. J. Comput. Intell. Appl.*, vol. 19, no. 1, Mar. 2020, Art. no. 2050002.

[162] C. Guo, J. Zhou, H. Chen, N. Ying, J. Zhang, and D. Zhou, "Variational autoencoder with optimizing Gaussian mixture model priors," *IEEE Access*, vol. 8, pp. 43992–44005, 2020.

[163] Z. Pan, J. Wang, W. Liao, H. Chen, D. Yuan, W. Zhu, X. Fang, and Z. Zhu, "Data-driven EV load profiles generation using a variational auto-encoder," *Energies*, vol. 12, no. 5, p. 849, Mar. 2019.

[164] M. Khodayar, S. Mohammadi, M. Khodayar, J. Wang, and G. Liu, "Convolutional graph auto-encoder: A deep generative neural architecture for probabilistic spatio-temporal solar irradiance forecasting," 2018, *arXiv:1809.03538*.

[165] L. Ma, L. Huang, and H. Shi, "A novel spatial–temporal generative autoencoder for wind speed uncertainty forecasting," *Energy*, vol. 282, Nov. 2023, Art. no. 128946.

[166] M. Khodayar and J. Wang, "Probabilistic time-varying parameter identification for load modeling: A deep generative approach," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1625–1636, Mar. 2021.

[167] J. Regan, M. Saffari, and M. Khodayar, "Deep attention and generative neural networks for nonintrusive load monitoring," *Electr. J.*, vol. 35, no. 5, Jun. 2022, Art. no. 107127.

[168] M. Khodayar, M. E. Khodayar, and S. M. J. Jalali, "Deep learning for pattern recognition of photovoltaic energy generation," *Electr. J.*, vol. 34, no. 1, Jan. 2021, Art. no. 106882.

[169] K. Zheng, P. Li, S. Zhou, W. Zhang, S. Li, L. Zeng, and Y. Zhang, "A multi-scale electricity consumption prediction algorithm based on time-frequency variational autoencoder," *IEEE Access*, vol. 9, pp. 90937–90946, 2021.

[170] Y. Wang, Y. Zhou, J. Ma, and Q. Jin, "A locational false data injection attack detection method in smart grid based on adversarial variational autoencoders," *Appl. Soft Comput.*, vol. 151, Jan. 2024, Art. no. 111169.

[171] C. Mylonas, I. Abdallah, and E. Chatzi, "Conditional variational autoencoders for probabilistic wind turbine blade fatigue estimation using supervisory, control, and data acquisition data," *Wind Energy*, vol. 24, no. 10, pp. 1122–1139, Oct. 2021.

[172] N. Aftabi, D. Li, and P. Ramanan, "A variational autoencoder framework for robust, physics-informed cyberattack recognition in industrial cyber-physical systems," 2023, *arXiv:2310.06948*.

[173] B. Zhou, X. Li, T. Zang, Y. Cai, J. Wu, and S. Wang, "The detection of false data injection attack for cyber–physical power systems considering a multi-attack mode," *Appl. Sci.*, vol. 13, no. 19, p. 10596, Sep. 2023.

[174] L. Zhu and D. J. Hill, "Data/model jointly driven high-quality case generation for power system dynamic stability assessment," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5055–5066, Aug. 2022.

[175] Y. Su, Q. He, J. Chen, and M. Tan, "A residential load forecasting method for multi-attribute adversarial learning considering multi-source uncertainties," *Int. J. Electr. Power Energy Syst.*, vol. 154, Dec. 2023, Art. no. 109421.

[176] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, Jan. 2018.

[177] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[178] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 27, 2014, pp. 1–9.

[179] H. Li, Z. Ren, M. Fan, W. Li, Y. Xu, Y. Jiang, and W. Xia, "A review of scenario analysis methods in planning and operation of modern power systems: Methodologies, applications, and challenges," *Electric Power Syst. Res.*, vol. 205, Apr. 2022, Art. no. 107722.

[180] J. Li, J. Zhou, and B. Chen, "Review of wind power scenario generation methods for optimal operation of renewable energy systems," *Appl. Energy*, vol. 280, Dec. 2020, Art. no. 115992.

[181] Y. Chen, Y. Wang, D. Kirschen, and B. Zhang, "Model-free renewable scenario generation using generative adversarial networks," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3265–3275, May 2018.

[182] X. Yang, H. He, J. Li, and Y. Zhang, "Toward optimal risk-averse configuration for HESS with CGANs-based PV scenario generation," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 3, pp. 1779–1793, Mar. 2021.

[183] Y. Li, J. Li, and Y. Wang, "Privacy-preserving spatiotemporal scenario generation of renewable energies: A federated deep generative learning approach," *IEEE Trans. Ind. Informat.*, vol. 18, no. 4, pp. 2310–2320, Apr. 2022.

[184] X. Zheng, B. Wang, D. Kalathil, and L. Xie, "Generative adversarial networks-based synthetic PMU data creation for improved event classification," *IEEE Open Access J. Power Energy*, vol. 8, pp. 68–76, 2021.

[185] D. Wu, H. Cao, D. Li, and S. Yang, "Energy-efficient reconstruction method for transmission lines galloping with conditional generative adversarial network," *IEEE Access*, vol. 8, pp. 17310–17319, 2020.

[186] D. Yang, M. Ji, Y. Lv, M. Li, and X. Gao, "Generative adversarial network-based data recovery method for power systems," *Appl. Math. Nonlinear Sci.*, vol. 9, no. 1, Jan. 2024.

[187] X. Hu, Z. Zhan, D. Ma, and S. Zhang, "Spatiotemporal generative adversarial imputation networks: An approach to address missing data for wind turbines," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–8, 2023.

[188] L. Song, Y. Li, and N. Lu, "ProfileSR-GAN: A GAN based super-resolution method for generating high-resolution load profiles," *IEEE Trans. Smart Grid*, vol. 13, no. 4, pp. 3278–3289, Jul. 2022.

[189] Y. Hu, Y. Li, L. Song, H. P. Lee, P. J. Rehm, M. Makdad, E. Miller, and N. Lu, "MultiLoad-GAN: A GAN-based synthetic load group generation method considering spatial–temporal correlations," *IEEE Trans. Smart Grid*, vol. 15, no. 2, pp. 2309–2320, Mar. 2024.

[190] W. N. Silva, L. H. T. Bandória, B. H. Dias, M. C. de Almeida, and L. W. de Oliveira, "Generating realistic load profiles in smart grids: An approach based on nonlinear independent component estimation (NICE) and convolutional layers," *Appl. Energy*, vol. 351, Dec. 2023, Art. no. 121902.

[191] H. Wen, Y. Du, X. Chen, E. G. Lim, H. Wen, and K. Yan, "A regional solar forecasting approach using generative adversarial networks with solar irradiance maps," *Renew. Energy*, vol. 216, Nov. 2023, Art. no. 119043.

[192] C. Jiang, Y. Chen, Y. Mao, Y. Chai, and M. Yu, "Forecasting spatio-temporal renewable scenarios: A deep generative approach," 2019, *arXiv:1903.05274*.

[193] H. Wei, Z. Hongxuan, D. Yu, W. Yiting, D. Ling, and X. Ming, "Short-term optimal operation of hydro-wind-solar hybrid system with improved generative adversarial networks," *Appl. Energy*, vol. 250, pp. 389–403, Sep. 2019.

[194] H. Yang, R. C. Qiu, X. Shi, and X. He, "Unsupervised feature learning for online voltage stability evaluation and monitoring based on variational autoencoder," *Electric Power Syst. Res.*, vol. 182, May 2020, Art. no. 106253.

[195] S. Parri, V. Kosana, and K. Teeparthi, "A hybrid GAN based autoencoder approach with attention mechanism for wind speed prediction," in *Proc. 22nd Nat. Power Syst. Conf. (NPSC)*, Dec. 2022, pp. 224–229.

[196] S. M. J. Jalali, M. Khodayar, A. Khosravi, G. J. Osório, S. Nahavandi, and J. P. S. Catalão, "An advanced generative deep learning framework for probabilistic spatio-temporal wind power forecasting," in *Proc. IEEE Int. Conf. Environ. Electr. Eng. IEEE Ind. Commercial Power Syst. Eur.*, Sep. 2021, pp. 1–6.

[197] Y. Qi, W. Hu, Y. Dong, Y. Fan, L. Dong, and M. Xiao, "Optimal configuration of concentrating solar power in multienergy power systems with an improved variational autoencoder," *Appl. Energy*, vol. 274, Sep. 2020, Art. no. 115124.

[198] D. Saxena and J. Cao, "Generative adversarial networks (GANs): Challenges, solutions, and future directions," *ACM Comput. Surv.*, vol. 54, no. 3, pp. 1–42, Apr. 2022.

[199] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "A brief survey of deep reinforcement learning," 2017, *arXiv:1708.05866*.

[200] P. Ladosz, L. Weng, M. Kim, and H. Oh, "Exploration in deep reinforcement learning: A survey," *Inf. Fusion*, vol. 85, pp. 1–22, Sep. 2022.

[201] J. Filar and K. Vrieze, *Competitive Markov Decision Processes*. Berlin, Germany: Springer, 2012.

[202] H. Byeon, "Advances in value-based, policy-based, and deep learning-based reinforcement learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, pp. 348–354, 2023.

[203] H. Zhang and T. Yu, "Taxonomy of reinforcement learning algorithms," in *Deep Reinforcement Learning: Fundamentals, Research and Applications*. Singapore: Springer, 2020, pp. 125–133.

[204] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, pp. 279–292, May 1992.

[205] B. Jang, M. Kim, G. Harerimana, and J. W. Kim, "Q-learning algorithms: A comprehensive classification and applications," *IEEE Access*, vol. 7, pp. 133653–133667, 2019.

[206] P. Wolf, C. Hubschneider, M. Weber, A. Bauer, J. Härtl, F. Dürr, and J. M. Zöllner, "Learning how to drive in a real world simulation with deep Q-networks," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 244–250.

[207] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 30, 2016, pp. 1–7.

[208] L. Wang, W. Mao, J. Zhao, and Y. Xu, "DDQP: A double deep Q-learning approach to online fault-tolerant SFC placement," *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 1, pp. 118–132, Mar. 2021.

[209] M. Babar, P. H. Nguyen, V. Cuk, I. G. Kamphuis, M. Bongaerts, and Z. Hanzelka, "The evaluation of agile demand response: An applied methodology," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6118–6127, Nov. 2018.

[210] R. Lu, S. H. Hong, and M. Yu, "Demand response for home energy management using reinforcement learning and artificial neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6629–6639, Nov. 2019.

[211] X. Xu, Y. Jia, Y. Xu, Z. Xu, S. Chai, and C. S. Lai, "A multi-agent reinforcement learning-based data-driven method for home energy management," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3201–3211, Jul. 2020.

[212] R. R. Hossain, Q. Huang, and R. Huang, "Graph convolutional network-based topology embedded deep reinforcement learning for voltage stability control," *IEEE Trans. Power Syst.*, vol. 36, no. 5, pp. 4848–4851, Sep. 2021.

[213] J. Zhang, Y. Luo, B. Wang, C. Lu, J. Si, and J. Song, "Deep reinforcement learning for load shedding against short-term voltage instability in large power systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 8, pp. 4249–4260, Aug. 2023.

[214] Y. Pei, J. Yang, J. Wang, P. Xu, T. Zhou, and F. Wu, "An emergency control strategy for undervoltage load shedding of power system: A graph deep reinforcement learning method," *IET Gener., Transmiss. Distrib.*, vol. 17, no. 9, pp. 2130–2141, May 2023.

[215] 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.

[216] Y. Li and J. Wu, "Low latency cyberattack detection in smart grids with deep reinforcement learning," *Int. J. Electr. Power Energy Syst.*, vol. 142, Nov. 2022, Art. no. 108265.

[217] D. Zhao, H. Wang, K. Shao, and Y. Zhu, "Deep reinforcement learning with experience replay based on SARSA," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2016, pp. 1–6.

[218] H. Jiang, R. Gui, Z. Chen, L. Wu, J. Dang, and J. Zhou, "An improved Sarsa(λ) reinforcement learning algorithm for wireless communication systems," *IEEE Access*, vol. 7, pp. 115418–115427, 2019.

[219] M. R. Tousi, S. H. Hosseinian, and M. B. Menhaj, "A multi-agent-based voltage control in power systems using distributed reinforcement learning," *Simulation*, vol. 87, no. 7, pp. 581–599, Jul. 2011.

[220] T. M. Aljohani and O. Mohammed, "A real-time energy consumption minimization framework for electric vehicles routing optimization based on SARSA reinforcement learning," *Vehicles*, vol. 4, no. 4, pp. 1176–1194, Oct. 2022.

[221] N. Mughees, M. H. Jaffery, A. Mughees, E. A. Ansari, and A. Mughees, "Reinforcement learning-based composite differential evolution for integrated demand response scheme in industrial microgrids," *Appl. Energy*, vol. 342, Jul. 2023, Art. no. 121150.

[222] M. N. Kurt, O. Ogundijo, C. Li, and X. Wang, "Online cyber-attack detection in smart grid: A reinforcement learning approach," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5174–5185, Sep. 2019.

[223] I. Grondman, L. Busoniu, G. A. D. Lopes, and R. Babuska, "A survey of actor-critic reinforcement learning: Standard and natural policy gradients," *IEEE Trans. Syst., Man, Cybern., C*, vol. 42, no. 6, pp. 1291–1307, Nov. 2012.

[224] D. Hu, Z. Ye, Y. Gao, Z. Ye, Y. Peng, and N. Yu, "Multi-agent deep reinforcement learning for voltage control with coordinated active and reactive power optimization," *IEEE Trans. Smart Grid*, vol. 13, no. 6, pp. 4873–4886, Nov. 2022.

[225] R. Wang, X. Bi, and S. Bu, "Real-time coordination of dynamic network reconfiguration and volt-VAR control in active distribution network: A graph-aware deep reinforcement learning approach," *IEEE Trans. Smart Grid*, vol. 15, no. 3, pp. 3288–3302, May 2024.

[226] J. Bakakeu, S. Baer, H.-H. Klos, J. Peschke, M. Brossog, and J. Franke, "Multi-agent reinforcement learning for the energy optimization of cyber-physical production systems," in *Artificial Intelligence in Industry 4.0*. Cham, Switzerland: Springer, 2021, pp. 143–163.

[227] C. Mu, Z. Liu, J. Yan, H. Jia, and X. Zhang, "Graph multi-agent reinforcement learning for inverter-based active voltage control," *IEEE Trans. Smart Grid*, vol. 15, no. 2, pp. 1399–1409, Mar. 2024.

[228] M. Mazare, "Reinforcement learning-based fixed-time resilient control of nonlinear cyber physical systems under false data injection attacks and mismatch disturbances," *J. Franklin Inst.*, vol. 360, no. 18, pp. 14926–14938, Dec. 2023.

[229] K. Bio Gassi and M. Baysal, "Improving real-time energy decision-making model with an actor-critic agent in modern microgrids with energy storage devices," *Energy*, vol. 263, Jan. 2023, Art. no. 126105.

[230] L. Xi, J. Wu, Y. Xu, and H. Sun, "Automatic generation control based on multiple neural networks with actor-critic strategy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 6, pp. 2483–2493, Jun. 2021.

[231] Y. Gu and X. Huang, "A reactive power optimization partially observable Markov decision process with data uncertainty using multi-agent actor-attention-critic algorithm," *Int. J. Electr. Power Energy Syst.*, vol. 147, May 2023, Art. no. 108848.
[232] T. Tiong, I. Saad, K. T. K. Teo, and H. B. Lago, "Deep reinforcement learning with robust deep deterministic policy gradient," in *Proc. 2nd Int. Conf. Electr., Control Instrum. Eng. (ICECIE)*, Nov. 2020, pp. 1–5.
[233] J. Li, R. Zhang, H. Wang, Z. Liu, H. Lai, and Y. Zhang, "Deep reinforcement learning for voltage control and renewable accommodation using spatial–temporal graph information," *IEEE Trans. Sustain. Energy*, vol. 15, no. 1, pp. 249–262, Jan. 2024.
[234] S. Wang, J. Duan, D. Shi, C. Xu, H. Li, R. Diao, and Z. Wang, "A data-driven multi-agent autonomous voltage control framework using deep reinforcement learning," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4644–4654, Nov. 2020.
[235] H. Liu, C. Zhang, Q. Chai, K. Meng, Q. Guo, and Z. Y. Dong, "Robust regional coordination of inverter-based volt/var control via multi-agent deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 12, no. 6, pp. 5420–5433, Nov. 2021.
[236] I. Jendoubi and F. Bouffard, "Multi-agent hierarchical reinforcement learning for energy management," *Appl. Energy*, vol. 332, Feb. 2023, Art. no. 120500.
[237] Y. Chengqing, Y. Guangxi, Y. Chengming, Z. Yu, and M. Xiwei, "A multi-factor driven spatiotemporal wind power prediction model based on ensemble deep graph attention reinforcement learning networks," *Energy*, vol. 263, Jan. 2023, Art. no. 126034.
[238] T. Zhang, J. Liu, H. Wang, Y. Li, N. Wang, and C. Kang, "Fault diagnosis and protection strategy based on spatio-temporal multi-agent reinforcement learning for active distribution system using phasor measurement units," *Measurement*, vol. 220, Oct. 2023, Art. no. 113291.
[239] J. Li, R. Zhang, H. Wang, Z. Liu, H. Lai, and Y. Zhang, "Deep reinforcement learning for optimal power flow with renewables using graph information," 2021, *arXiv:2112.11461*.
[240] W. Shi, D. Zhang, X. Han, X. Wang, T. Pu, and W. Chen, "Coordinated operation of active distribution network, networked microgrids, and electric vehicle: A multi-agent PPO optimization method," *CSEE J. Power Energy Syst.*, early access, Apr. 20, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10106205
[241] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*.
[242] Y. Gu, Y. Cheng, C. L. P. Chen, and X. Wang, "Proximal policy optimization with policy feedback," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 7, pp. 4600–4610, Jul. 2022.
[243] T. Wu, A. Scaglione, and D. Arnold, "Reinforcement learning using physics inspired graph convolutional neural networks," in *Proc. 58th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2022, pp. 1–8.
[244] Y. Liang, X. Zhao, and L. Sun, "A multiagent reinforcement learning approach for wind farm frequency control," *IEEE Trans. Ind. Informat.*, vol. 19, no. 2, pp. 1725–1734, Feb. 2023.
[245] Z.-C. Zhou, Z. Wu, and T. Jin, "Deep reinforcement learning framework for resilience enhancement of distribution systems under extreme weather events," *Int. J. Electr. Power Energy Syst.*, vol. 128, Jun. 2021, Art. no. 106676.
[246] Q. Zhang, Y. Mahajan, I.-R. Chen, D. S. Ha, and J.-H. Cho, "An attack-resilient and energy-adaptive monitoring system for smart farms," in *Proc. GLOBECOM-IEEE Global Commun. Conf.*, Dec. 2022, pp. 2776–2781.
[247] T. Wu, A. Scaglione, and D. Arnold, "Constrained reinforcement learning for predictive control in real-time stochastic dynamic optimal power flow," *IEEE Trans. Power Syst.*, vol. 39, no. 3, pp. 5077–5090, May 2024.
[248] 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.
[249] D. Cao, W. Hu, J. Zhao, Q. Huang, Z. Chen, and F. Blaabjerg, "A multi-agent deep reinforcement learning based voltage regulation using coordinated PV inverters," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 4120–4123, Sep. 2020.
[250] A. H. Alobaidi, S. S. Fazlhashemi, M. Khodayar, J. Wang, and M. E. Khodayar, "Distribution service restoration with renewable energy sources: A review," *IEEE Trans. Sustain. Energy*, vol. 14, no. 2, pp. 1151–1168, Apr. 2023.
[251] A. H. Alobaidi, M. Khodayar, A. Vafamehr, H. Gangammanavar, and M. E. Khodayar, "Stochastic expansion planning of battery energy storage for the interconnected distribution and data networks," *Int. J. Electr. Power Energy Syst.*, vol. 133, Dec. 2021, Art. no. 107231.
[252] DIgSILENT GmbH. (2024). *DIgSILENT—Power System Analysis Software & Consulting*. [Online]. Available: https://www.digsilent.de/en/
[253] UKERC Energy Data Centre. (2024). *Ukerc Energy Data Centre*. [Online]. Available: https://ukerc.rl.ac.uk/cgi-bin/index.pl
[254] National Renewable Energy Laboratory. (2024). *Wind Integration National Dataset Toolkit*. [Online]. Available: https://www.nrel.gov/grid/wind-toolkit.html
[255] National Renewable Energy Laboratory. (2024). *Solar Integration National Dataset (SIND) Toolkit*. [Online]. Available: https://www.nrel.gov/grid/sind-toolkit.html
[256] Greater London Authority. (2014). *Smartmeter Energy Use Data in London Households*. [Online]. Available: https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london
[257] PJM Interconnection. (2024). *PJM Data Miner*. [Online]. Available: https://dataminer2.pjm.com/list
[258] Pecan Street. (2024). *Pecan Street—Research Data*. [Online]. Available: https://www.pecanstreet.org/
[259] S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," 2015, *arXiv:1511.06349*.
[260] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proc. Adv. In Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.
[261] S. Bond-Taylor, A. Leach, Y. Long, and C. G. Willcocks, "Deep generative modelling: A comparative review of VAEs, GANs, normalizing flows, energy-based and autoregressive models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7327–7347, Nov. 2022.
[262] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, "Normalizing flows for probabilistic modeling and inference," *J. Mach. Learn. Res.*, vol. 22, no. 57, pp. 1–64, 2021.
[263] D. J. Rezende and S. Mohamed, "Variational inference with normalizing flows," in *Proc. Int. Conf. Mach. Learn. (ICML)*, vol. 37, 2015, pp. 1530–1538.
[264] I. Kobyzev, S. J. D. Prince, and M. A. Brubaker, "Normalizing flows: An introduction and review of current methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 3964–3979, Nov. 2021.
[265] M. Khodayar and J. Regan, "Deep neural networks in power systems: A review," *Energies*, vol. 16, no. 12, p. 4773, Jun. 2023.
[266] J. Stiasny, G. S. Misyris, and S. Chatzivasileiadis, "Physics-informed neural networks for non-linear system identification for power system dynamics," in *Proc. IEEE Madrid PowerTech*, Jun. 2021, pp. 1–6.
[267] M. Saffari, M. Khodayar, and M. E. Khodayar, "Physics-informed graph capsule generative autoencoder for probabilistic AC optimal power flow," *IEEE Trans. Emerg. Topics Comput. Intell.*, early access, Apr. 2, 2024, doi: 10.1109/TETCI.2024.3377671.
[268] H. Wessels, C. Weißenfels, and P. Wriggers, "The neural particle method—An updated Lagrangian physics informed neural network for computational fluid dynamics," *Comput. Methods Appl. Mech. Eng.*, vol. 368, Aug. 2020, Art. no. 113127.
[269] M. Sarabian, H. Babaee, and K. Laksari, "Physics-informed neural networks for brain hemodynamic predictions using medical imaging," *IEEE Trans. Med. Imag.*, vol. 41, no. 9, pp. 2285–2303, Sep. 2022.
[270] G. S. Misyris, A. Venzke, and S. Chatzivasileiadis, "Physics-informed neural networks for power systems," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2020, pp. 1–5.
[271] F. K. Došilovic, M. Brcic, and N. Hlupic, "Explainable artificial intelligence: A survey," in *Proc. 41st Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2018, pp. 210–215.
[272] P. P. Angelov, E. A. Soares, R. Jiang, N. I. Arnold, and P. M. Atkinson, "Explainable artificial intelligence: An analytical review," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 11, no. 5, p. e1424, 2021.
[273] C. Xu, Z. Liao, C. Li, X. Zhou, and R. Xie, "Review on interpretable machine learning in smart grid," *Energies*, vol. 15, no. 12, p. 4427, Jun. 2022.

[274] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Definitions, methods, and applications in interpretable machine learning," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 44, pp. 22071–22080, Oct. 2019.

[275] M. Macktoobian, A. K. N. Tehrani, and M. Khodayar, "Morphological reconfiguration monitoring for homogeneous self-reconfigurable robots," in *Proc. 14th Int. Conf. Electr. Eng., Comput. Sci. Autom. Control (CCE)*, Oct. 2017, pp. 1–4.

[276] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," *Electronics*, vol. 8, no. 8, p. 832, Jul. 2019.

[277] P. Shaw, J. Uszkoreit, and A. Vaswani, "Self-attention with relative position representations," 2018, *arXiv:1803.02155*.

[278] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[279] C. Molnar, *Interpretable Machine Learning*. Abu Dhabi, United Arab Emirates: Lulu, 2020.

[280] A. Goldstein, A. Kapelner, J. Bleich, and E. Pitkin, "Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation," *J. Comput. Graph. Statist.*, vol. 24, no. 1, pp. 44–65, Jan. 2015.

[281] B. Greenwell, "PDP: An R package for constructing partial dependence plots," *R J.*, vol. 9, no. 1, p. 421, 2017.

[282] S. Mishra, B. L. Sturm, and S. Dixon, "Local interpretable model-agnostic explanations for music content analysis," in *Proc. ISMIR*, vol. 53, 2017, pp. 537–543.

[283] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. 32nd Int. Conf. Mach. Learn.*, vol. 37, Jul. 2015, pp. 97–105.

[284] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Unsupervised domain adaptation with residual transfer networks," in *Proc. Adv. In Neural Inf. Process. Syst.*, 2016, pp. 1–9.

[285] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 513–520.

[286] P. Wei, Y. Ke, and C. K. Goh, "Feature analysis of marginalized stacked denoising autoenconder for unsupervised domain adaptation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 5, pp. 1321–1334, May 2019.

[287] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2962–2971.

[288] H. Tang and K. Jia, "Discriminative adversarial domain adaptation," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 4, pp. 5940–5947.

[289] C. Hognon, P.-H. Conze, V. Bourbonne, O. Gallinato, T. Colin, V. Jaouen, and D. Visvikis, "Contrastive image adaptation for acquisition shift reduction in medical imaging," *Artif. Intell. Med.*, vol. 148, Feb. 2024, Art. no. 102747.

[290] O. Crystal, A. Khademi, A. R. Moody, P. J. Maralani, and S. E. Black, "Domain adaptation using silver standard masks for lateral ventricle segmentation in flair MRI," in *Proc. Med. Imag. Deep Learn.*, 2024, pp. 1895–1909.

[291] Q. Qian, J. Luo, and Y. Qin, "Adaptive intermediate class-wise distribution alignment: A universal domain adaptation and generalization method for machine fault diagnosis," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 21, 2024, doi: 10.1109/TNNLS.2024.3376449.

[292] Q. Wang, Y. Xu, S. Yang, J. Chang, J. Zhang, and X. Kong, "A domain adaptation method for bearing fault diagnosis using multiple incomplete source data," *J. Intell. Manuf.*, vol. 35, no. 2, pp. 777–791, Feb. 2024.

[293] J. Konečný, H. Brendan McMahan, F. X. Yu, P. Richtárik, A. Theertha Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2016, *arXiv:1610.05492*.

[294] N. Gholizadeh and P. Musilek, "Distributed learning applications in power systems: A review of methods, gaps, and challenges," *Energies*, vol. 14, no. 12, p. 3654, Jun. 2021.

[295] X. Cheng, C. Li, and X. Liu, "A review of federated learning in energy systems," in *Proc. IEEE/IAS Ind. Commercial Power Syst. Asia*, Jul. 2022, pp. 2089–2095.

[296] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," *Knowl.-Based Syst.*, vol. 216, Mar. 2021, Art. no. 106775.

[297] V. A. Patel, P. Bhattacharya, S. Tanwar, R. Gupta, G. Sharma, P. N. Bokoro, and R. Sharma, "Adoption of federated learning for healthcare informatics: Emerging applications and future directions," *IEEE Access*, vol. 10, pp. 90792–90826, 2022.

[298] T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. C. Paschalidis, and W. Shi, "Federated learning of predictive models from federated electronic health records," *Int. J. Med. Informat.*, vol. 112, pp. 59–67, Apr. 2018.

[299] X. Zhu, J. Wang, Z. Hong, and J. Xiao, "Empirical studies of institutional federated learning for natural language processing," in *Proc. Findings Assoc. Comput. Linguistics*, 2020, pp. 625–634.

[300] B. Nagy, I. Hegedus, N. Sándor, B. Egedi, H. Mehmood, K. Saravanan, G. Lóki, and Á. Kiss, "Privacy-preserving federated learning and its application to natural language processing," *Knowl.-Based Syst.*, vol. 268, May 2023, Art. no. 110475.

[301] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, "Federated learning for Internet of Things: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1622–1658, 3rd Quart., 2021.

[302] T. Zhang, L. Gao, C. He, M. Zhang, B. Krishnamachari, and A. S. Avestimehr, "Federated learning for the Internet of Things: Applications, challenges, and opportunities," *IEEE Internet Things Mag.*, vol. 5, no. 1, pp. 24–29, Mar. 2022.

MOHSEN SAFFARI (Graduate Student Member, IEEE) received the B.Sc. degree in electrical engineering from Shahrood University of Technology, in 2014, and the M.Sc. degree in electrical engineering from the K. N. Toosi University of Technology, Iran, in 2016. He is currently pursuing the Ph.D. degree in computer science with The University of Tulsa, Tulsa, OK, USA. His research interests include theories and applications of machine learning in computer vision and energy systems. He serves as a Reviewer for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE, and *Neural Processing Letters* (NEPL).

MAHDI KHODAYAR (Member, IEEE) received the B.Sc. degree in computer engineering and the M.Sc. degree in artificial intelligence from the K. N. Toosi University of Technology, Tehran, Iran, in 2013 and 2015, respectively, and the Ph.D. degree in electrical engineering from Southern Methodist University, Dallas, TX, USA, in 2020. In 2017, he was a Research Assistant with the College of Computer and Information Science, Northeastern University, Boston, MA, USA. He is currently an Assistant Professor with the Department of Computer Science, The University of Tulsa, Tulsa, OK, USA. His main research interests include machine learning and statistical pattern recognition. He is focused on deep learning, sparse modeling, and spatiotemporal pattern recognition. He has served as a Reviewer for many reputable journals, including IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, and IEEE TRANSACTIONS ON POWER SYSTEMS.