

REVIEW

# Energy forecasting in smart grid systems: recent advancements in probabilistic deep learning

Devinder Kaur<sup>1</sup> | Shama Naz Islam<sup>1</sup> | Md Apel Mahmud<sup>2</sup>  | Md Enamul Haque<sup>1</sup> | Zhao Yang Dong<sup>3</sup>

<sup>1</sup>School of Engineering, Deakin University, Waurn Ponds, Australia

<sup>2</sup>Faculty of Engineering and Environment, Northumbria University, Newcastle Upon Tyne, UK

<sup>3</sup>School of Electrical & Electronics Engineering Nanyang Technological University, Singapore

Correspondence

Md Apel Mahmud, Faculty of Engineering and Environment, Northumbria University, Newcastle Upon Tyne NE1 8ST, United Kingdom.  
Email: [md.a.mahmud@northumbria.ac.uk](mailto:md.a.mahmud@northumbria.ac.uk)

[Correction added on 3-October-2022, after first online publication: Affiliation number 3 is updated in this version of the paper]

## Abstract

Energy forecasting plays a vital role in mitigating challenges in data rich smart grid (SG) systems involving various applications such as demand-side management, load shedding, and optimum dispatch. Managing efficient forecasting while ensuring the least possible prediction error is one of the main challenges posed in the grid today, considering the uncertainty in SG data. This paper presents a comprehensive and application-oriented review of state-of-the-art forecasting methods for SG systems along with recent developments in probabilistic deep learning (PDL). Traditional point forecasting methods including statistical, machine learning (ML), and deep learning (DL) are extensively investigated in terms of their applicability to energy forecasting. In addition, the significance of hybrid and data pre-processing techniques to support forecasting performance is also studied. A comparative case study using the Victorian electricity consumption in Australia and American electric power (AEP) datasets is conducted to analyze the performance of deterministic and probabilistic forecasting methods. The analysis demonstrates higher efficacy of DL methods with appropriate hyper-parameter tuning when sample sizes are larger and involve nonlinear patterns. Furthermore, PDL methods are found to achieve at least 60% lower prediction errors in comparison to other benchmark DL methods. However, the execution time increases significantly for PDL methods due to large sample space and a tradeoff between computational performance and forecasting accuracy needs to be maintained.

## 1 | INTRODUCTION

Energy forecasting has a crucial role to play in planning, investment, decision making, and mitigating operational and management challenges in modern power systems popularly termed as smart grid (SG) systems [1]. With the advent of smart meters and advanced metering infrastructure (AMI) in SG, a significant increase in the bidirectional flow of energy and data is observed between the grid and end-users. Thus, a number of data analytics applications such as energy forecasting have emerged recently in the SG domain. Such applications can be highly useful for scheduling generation, implementing demand response strategies, and ensuring financial benefits through optimum bidding in the energy market [2, 3]. In the past decades, traditional statistical methods such as autoregressive integrated moving average (ARIMA) and associated

techniques have been extensively used for forecasting energy demand and generation [4]. However, in recent times due to the breakthrough of smart metering and thus high volume of data generation, statistical methods face scalability issues and can not analyze complex nonlinear data features [5].

Since the last decade, artificial intelligence (AI)-based methods involving machine learning (ML) and deep learning (DL) algorithms have attracted much attention for their ability to generate accurate forecasts in SG systems [6]. Particularly, sequence-based DL algorithms such as recurrent neural networks (RNN) and long short-term memory (LSTM) models have proven to be potential methods to deal with the nonlinear energy data features with long sequences [7]. Furthermore, SG data generation experience the problem of stochastic uncertainties when renewable generation intermittencies and variations in energy consumption behaviors are considered. Moreover,

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *IET Generation, Transmission & Distribution* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

tackling parametric or model uncertainties is another challenge faced by aforementioned forecasting methods, which are deterministic in nature and provide point forecasts. In such cases, probabilistic methods [8–10] having the ability to generate prediction intervals (PIs) highlight their effectiveness to manage uncertainties compared to point forecasting methods. Recently, probabilistic deep learning (PDL) has made its way to carry out the forecasting more efficiently by integrating deep neural networks and Bayesian inference [11, 12]. However, this approach needs to be investigated more and offers a substantial scope in modern power systems and forecasting applications.

Furthermore, by ensembling multiple methods as one hybrid approach, forecasting accuracy can be highly improved [13]. However, the model complexity could be one of the main shortcomings of such approach and thus, a tradeoff between the accuracy and computational complexity needs to be maintained. In this direction, data pre-processing may play a crucial role in improving the model performance and minimizing the forecasting error [14]. In this context, dimensionality reduction (DR) and feature extraction are the key approaches often adopted for effective data pre-processing [15, 16]. In this regard, principal component analysis (PCA) [17] and singular value decomposition (SVD) [18] have been widely used as underlying pre-processing techniques with different forecasting methods to deal with the challenges of high-order dimensionality in SG data. From DL domain, auto-encoders (AE) implemented with convolutional layers have been used as a feature extraction scheme to filter multiple dimensions in consumption data [19].

Although, there have been a significant number of research papers which surveyed ML methods for solar irradiance forecasting [20–22] and the authors in [23] reviewed state-of-the-art DL methods for renewable energy forecasting. None of the aforementioned review papers investigated advanced DL methods such as RNN and LSTM in detail for energy forecasting applications with a comparison made between the statistical, ML, and DL methods. Furthermore, to the best of our knowledge, no review paper has yet considered reviewing the recent developments in probabilistic forecasting domain, especially contributions of probabilistic deep learning for energy forecasting applications in SG systems. This motivates us to review state-of-the-art forecasting methods holistically along with the recent developments in PDL domain in an application-oriented manner.

To be specific, we make the following contributions:

- A comprehensive review of statistical, ML, DL, probabilistic, and hybrid forecasting methods is presented along with their applications in SG systems considering different time horizons. The existing pre-processing techniques to aid energy forecasting performance are also discussed.
- A number of statistical, ML, DL, and PDL techniques are implemented for a number of energy consumption datasets for load and renewable generation forecasting to compare forecasting accuracies. The impact of high variability and data sizes on different methods is also evaluated. The analysis

shows that RNN and LSTM can achieve higher accuracies with least prediction errors under deterministic forecasting methods with larger dataset sizes especially in the presence of high variability. However, the accuracy can vary with the choice of activation function and hyper-parameters tuning which needs to be appropriately selected for a given dataset.

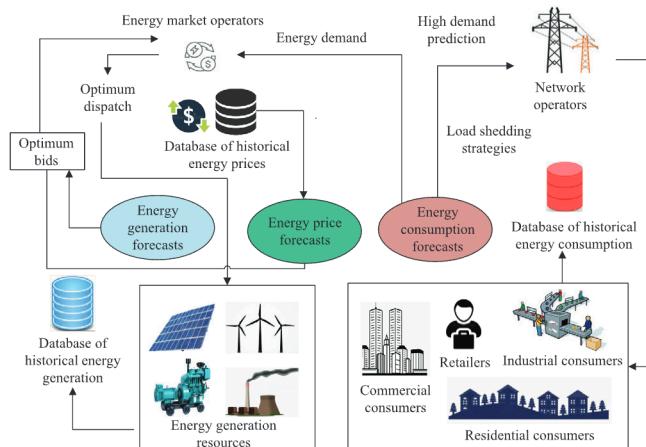
- Among probabilistic forecasting techniques, Bayesian bidirectional LSTM (BLSTM) is observed to outperform deterministic methods by exhibiting least error and tighter prediction intervals for different energy forecasting datasets. Though PDL methods can improve the forecasting accuracy, they have a higher execution time as compared to deterministic methods, and thus, a tradeoff between the model performance and computational cost is required.

The rest of the paper is organized as follows. Section 2 discusses the applications of forecasting in SG systems. Section 3 outlines the time horizon-based categorization of forecasting methods. Section 4 describes the taxonomy of the energy forecasting methods. Section 5 elaborates the different deterministic forecasting methods. The contributions related to probabilistic forecasting methods have been described in Section 6. A number of hybrid methods have been presented in Section 7. Further, data pre-processing methods are briefly discussed in Section 8. Section 9 represents a comparative case study of energy forecasting methods on two different datasets. Finally, Section 10 concludes the paper and discusses the future directions.

## 2 | APPLICATIONS OF FORECASTING IN ENERGY SYSTEMS

Energy forecasting with high accuracy and precision can provide significant insights for grid planning and reliability and has a wide range of applications including load shedding, optimum dispatch, peak load shaving etc. [24]. The data generated from SG systems is usually related to the energy consumption by end users involving residential, commercial, and industrial consumers as shown in Figure 1. Furthermore, additional data is generated from distributed energy resources (DERs) such as solar and wind power plants. The generated data is further utilized by forecasting methods to forecast energy generation, demand, and prices in SG systems. The authors in [25] surveyed various challenges and trends of demand forecasting with respect to different time horizons and regions. Similarly, in [26] authors compared existing statistical methods for demand forecasting and emphasized the significance of attaining least root-mean-square error (RMSE). Further, Kong *et al.* in [27] employed DL methods to forecast short-term energy consumption and demand for individual households using highly granular data.

With the increasing incentives for demand response programs, it has become essential to forecast electricity demand more accurately. In addition, forecasting the flexibility with



**FIGURE 1** Overview of forecasting applications for energy systems

which the users can change their energy demand has emerged as a challenging problem. Existing research has looked into developing a more accurate model for demand response calculation through feature engineering [28]. The authors in [29] has incorporated renewable generation forecasts to improve the accuracy of the energy optimisation algorithm for demand response. On the other hand, the impact of load and price forecasts on optimum demand management has been elaborated in [30]. Recently, demand flexibility prediction using ML models have been considered in [31] which takes into account the demand response potentials of electric vehicles and hot water systems.

As energy generation using renewable-based DER is more intermittent due to exogenous factors such as weather changes and user behaviors, ensuring high forecasting accuracy is a very challenging task in the grid today [32]. In this regard, [33] used vector autoregressive (VAR) to forecast solar irradiance, temperature, and wind speed for 61 locations in the United States. Similarly, Messner et al. in [34] used VAR method to forecast wind power generation based on high dimensional data.

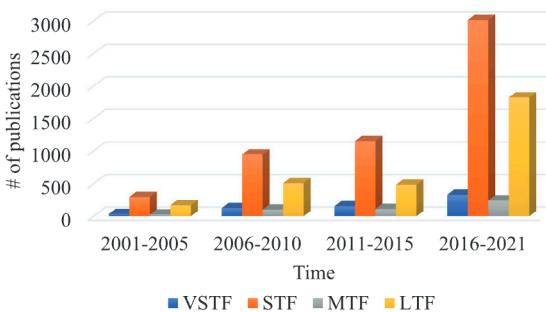
Furthermore, forecasting price spike is an important application which considers the biggest risk factor in the energy market. Recent zero or negative pricing as observed in the Australian national electricity market (NEM) can negatively impact generators [35]. On the other hand, positive high price spikes can give more profits to generators, especially during peak generation hours [36]. Yang et al. surveyed the latest trends in the decision making for electricity retailers using consumed load price forecasting [37]. Furthermore, [38] utilized a set of relevance vector machines to predict the price for individual time periods ahead-of-time and implemented micro-genetic algorithm to optimize linear regression ensemble coefficients for aggregated price forecasts. Toubeau et al. in [39] utilized PDL method to forecast wind and PV generation in return to predict the electricity prices generated from renewable DERs. Moreover, the locational marginal price for optimal scheduling of the energy storages is performed using artificial neural networks (ANN) in [40].

### 3 | TIME HORIZONS FOR ENERGY FORECASTING METHODS

The time period over which forecasts are generated is defined as the time horizon or timescale. It is one of the key parameters to categorize forecasting approaches in SG systems [41]. Depending on time horizon, energy forecasting can be broadly classified as following:

- Very short-term forecasting (VSTF): This class of forecasting involves time horizon from minutes to few hours, usually between (0-3 h) [42]. VSTF can help dealing with random changes in renewable energy generation which can be predicted only before a short period of time. It offers a wide range of applications in renewable energy resources (RES) such as wind and solar power forecasting [43, 44]. In this regard, Potter *et al.* in [45] presented a 2.5 min ahead forecasting system for Tasmanian wind farms using ANN and fuzzy logic as a hybrid approach.
- Short-term forecasting (STF): It involves energy forecasts ranging from few minutes to a few days ahead. This class plays a prime role in various grid operations such as dispatch scheduling, reliability analysis etc. [46]. Furthermore, accurate STF helps avoiding underestimation and overestimation of the demand and thus, substantially contributes to the grid reliability [47, 48].
- Medium-term forecasting (MTF): It implies to horizons ranging from few days to a few months ahead within a year [49]. MTF supports adequacy assessment, maintenance, and fuel supply scheduling in SG systems. Furthermore, it contributes to risk management using price forecasting and therefore, plays a significant role in evaluating the financial aspects of energy systems [50].
- Long-term forecasting (LTF): It involves horizons measured in months, quarters, and even years. LTF is crucial for load growth and energy generation planning operations over longer periods of time [51, 52]. LTF helps removing the impact of random fluctuations arising in shorter term and predict the longer term trends. In this context, Azad *et al.* [53] predicted the wind speed trends of two meteorological stations in Malaysia for 1 year using neural networks to manage the challenges posed by intermittent nature of wind generation.

Figure 2 shows the pattern of publications for last two decades within 5 year duration with respect to different time horizons in energy systems forecasting. While LTF stands second in line, most number of publications are made for STF in the period 2016–2021, making it most widely utilized forecasting category in recent times for different applications in grid planning, operations, and energy management. Appropriate selection of the forecasting horizon is crucial for model training and hyper-parameter tuning of a forecasting method. In this context, the authors in [20] and [21] emphasized to focus on the selection of relevant time horizons while building the ML model for solar irradiation forecasting. In order to obtain the publication trends in Figure 2, the databases IEEEExplore, Sciedirect



**FIGURE 2** Publication trends in energy systems with respect to time horizons

and Google scholar has been utilized. We have used the keywords “load forecasting”, “energy forecasting”, “solar power forecasting”, “wind power forecasting”, “demand forecasting” to find out the key research papers.

## 4 | TAXONOMY OF THE ENERGY FORECASTING TECHNIQUES

This section discusses how we have categorized the state-of-the-art methods and recent advancements in energy forecasting systems along with their literature and applications. We have utilized different criteria to categorize energy forecasting methods including time horizon, representation of the forecasting output (point vs distribution), and the model performance (using error metrics) used to generate the forecasts. Given that any forecasting application can be quantified based on how far in the time horizon the output can be forecast, how much the forecasts can vary or how the forecasting model learns, the defined criteria can sufficiently capture the different dimensions of forecasting problems.

According to the defined criteria, energy forecasting methods can be broadly classified into VSTF, STF, MTF and LTF, as explained in the previous section. In terms of the forecasting model output type, the methods can be grouped into deterministic and probabilistic methods. Deterministic methods focus on generating point forecasts, whereas probabilistic methods generate forecasts in terms of prediction intervals. In terms of how the forecasting models are trained to predict future demand and generation, forecasting methods can be categorized as statistical, artificial intelligence (AI)-based methods (such as ML and DL), quantile regression-based methods, and the recent probabilistic deep learning (PDL) methods, as illustrated in Figure 3. While statistical methods are traditional and simpler in form, learning based methods involve ML, DL and its variants which are considered to be more accurate but complex at the same time. Note that, the classification of these methods may have some cross-overs among different groups. For example, statistical methods have been explained under deterministic techniques, but these can also be formulated as probabilistic methods, and have been discussed under probabilistic techniques. Similarly, ML and DL methods can also be combined with probabilistic methods and have been covered under PDL methods. Furthermore, hybrid

methods and data pre-processing techniques have been categorized under the same two categories as these techniques can be utilized in either way, deterministic or probabilistic.

The following sections discuss state-of-the-art and advanced forecasting methods in detail with their literature and applications for SG systems in accordance with the categorization reflected in Figure 3.

## 5 | DETERMINISTIC TECHNIQUES

Deterministic techniques generate forecasts in terms of fixed points rather than a probability distribution. These methods are also termed as point forecasting methods.

### 5.1 | Statistical methods

Statistical methods are classical in nature, widely used for time-series forecasting especially for STF using the historical SG data [54]. These methods are based on fitting a regression model on historical data points and then validating the model by finding the difference between actual and predicted values. This subsection stresses about existing statistical methods for the applications of SG systems forecasting.

#### 5.1.1 | Autoregressive moving average model (ARMA)

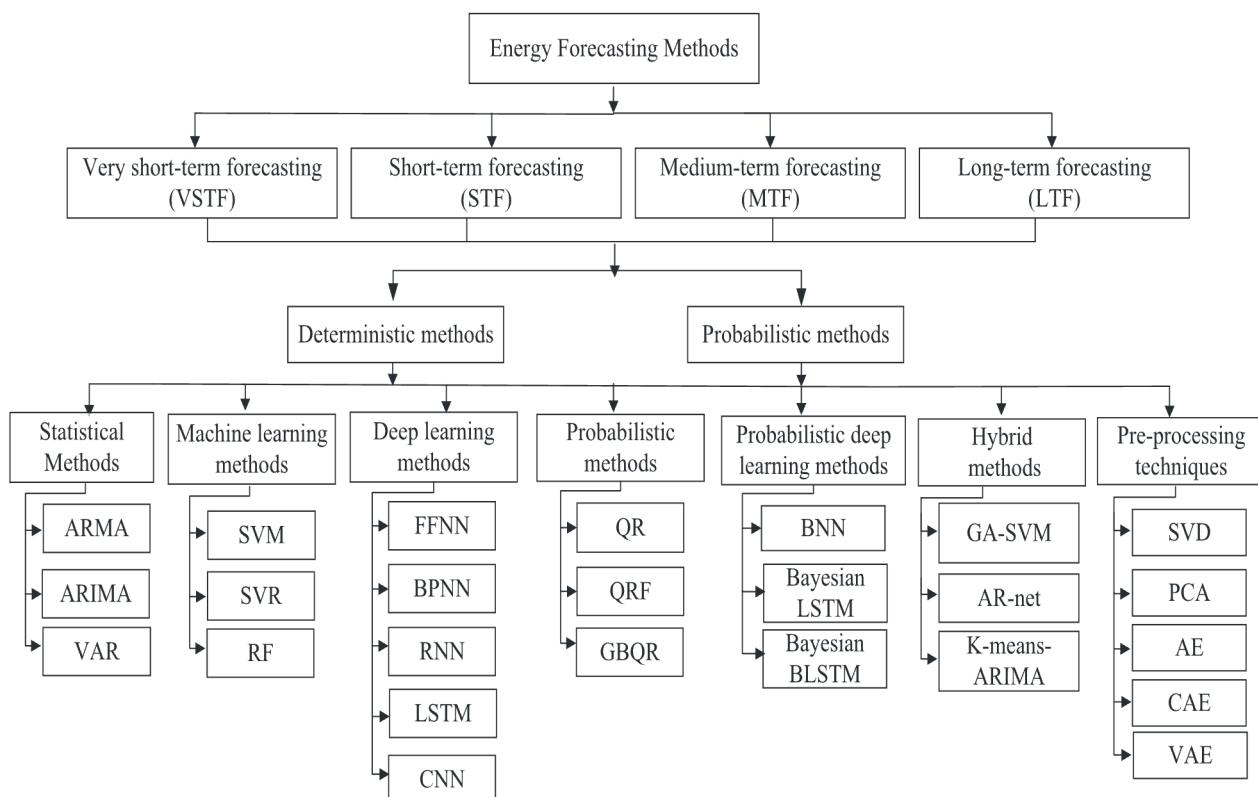
ARMA is the fundamental statistical method widely employed for time-series analysis. It is the combination of autoregressive (AR) and moving average (MA) methods. Sansa *et al.* in [55] utilized ARMA to predict the solar irradiation for a winter day with maximum 10% variation in the generation. However, ARMA can only be applied to a stationary time-series, for which mean and variance do not change over time and has an uniform distribution.

#### 5.1.2 | Autoregressive integrated moving average (ARIMA)

To deal with the non-stationary time-series, ARMA is generalized to ARIMA. ARIMA is employed in [56] for energy demand forecasting to charge the electric vehicles (EVs) using historical load data. Similarly, it has been used to forecast energy usage in [57] and [58] for control and optimization in residential microgrids. However, the ARIMA is more suitable for linear time-series as it gives relatively high RMSE values for nonlinear data and requires a high execution time for large datasets [59].

#### 5.1.3 | Vector autoregression (VAR)

VAR model is defined as an extension to the univariate autoregression, which captures linear dependencies among multiple



**FIGURE 3** Taxonomy of energy forecasting methods for smart grid systems

time-series. The authors in [33] predicted temperature, solar irradiation, and wind speed using VAR and exhibited lower RMSE values as compared to other statistical methods such as simple persistence (SP). VAR is also used in [34] for short-term wind power forecasting which is integrated with lasso estimator to recursively update the VAR parameters.

#### 5.1.4 | Autoregressive/autoregressive integrated moving average with exogenous inputs (ARMAX/ARIMAX)

ARMAX/ARIMAX models take into account external variables that influence the forecasting accuracy of a desired parameter. It combines the AR and MA processes (along with the integrated component for ARIMA) with the external time-series parameters [60]. The authors in [60] utilized temperature, humidity, irradiance duration parameters as exogenous inputs to improve the accuracy of the solar generation forecasting. On the other hand, the authors in [61] have employed satellite images as exogenous variables integrated with recursive least squares filter. The authors in [62] have developed a seasonal ARIMA framework with weather variables and seasonality as exogenous inputs for forecasting load demand. Though the aforementioned methods achieve good forecasting accuracy, the ML methods that are discussed in the following subsection are found to be more efficient and scalable to deal with nonlinear traits of SG datasets with multiple time-series [63].

## 5.2 | Machine learning (ML) methods

ML is a set of computing algorithms to learn patterns from input data and improve the prediction accuracy automatically over time. ML is broadly categorized into supervised and unsupervised learning [64]. Supervised learning involves labeled data and unsupervised learning deals with the unstructured data [65]. Following subsections discuss some of the existing ML methods and their applications for energy systems forecasting.

### 5.2.1 | Support vector machine (SVM)

SVM is widely used in ML to classify data variables into different planes with respect to margin between the data classes. For SG systems, Shi *et al.* proposed SVM to forecast one day ahead photovoltaic (PV) generation in China by classifying weather data into cloudy, clear, foggy, and rainy categories [66]. Similarly, authors in [67] employed SVM to predict socio-demographic features of users from energy consumption data with the help of convolutional filters.

### 5.2.2 | Support vector regression (SVR)

The concept of SVM adapted to regression analysis is known as SVR. It supports nonlinear traits in the prediction model and provides accurate forecasting results. Since time-series

forecasting involves regression operation, SVR is comprehensively used in the existing research for demand and generation forecasting [68–71]. The authors in [72] applied SVR for short-term load forecasting and compared the results with other competitive techniques such as ARMA and artificial neural networks. Furthermore, [73] predicted long-term energy consumption using SVR after decomposing time-series consumption data into two components to justify trend and oscillations, respectively.

### 5.2.3 | Random forest (RF)

RF is another ML algorithm which operates as an ensemble of decision trees and often utilized for classification and regression tasks. Saeedi *et al.* implemented different ML methods including decision tree, linear regression, and multilinear perceptron (MLP) to estimate PV generation output in Hawaii [74]. They claimed the superior performance by RF with highest R-squared value of 98% as compared to aforementioned ML models. Similarly, [75] employed a hybrid of fuzzy c-means clustering and RF to predict short-term load consumption after grouping the similar load profiles to reduce uncertainty in the individual consumption data.

However, energy datasets with high volume and complex patterns require extensive learning and optimization, where ML methods usually fail to perform efficiently [76]. In this context, DL methods based on artificial neural networks are emerging as a highly competent set of techniques to deal with high dimensionality and uncertainty traits of SG data [77].

## 5.3 | Deep learning (DL) methods

DL methods involve artificial neural networks (ANNs) defined as sequence of layers having  $n$  number of connected nodes termed as neurons [78]. An input from external data is fed to the network via input layer and an output is obtained by optimizing the network with learning algorithm via output layer. ANNs with more than one hidden layer are known as deep neural networks (DNN). This subsection discusses state-of-the-art DL methods employed in energy systems forecasting along with relevant literature and applications.

### 5.3.1 | Feed forward neural networks (FFNN)

Vanilla ANNs without having loops within layers are known as FFNNs. They can be single or multilayer perceptrons and are primarily used for supervised learning. Bhaskar *et al.* presented a FFNN-based wind power forecasting framework integrated with wavelet decomposition to forecast 30 h ahead wind power built on wind speed estimation [79]. Similarly, [80] investigated the use of FFNNs for short-term price forecasting for Spanish and PJM electricity markets and conducted a comparative analysis with wavelet-ARIMA and fuzzy neural networks.

### 5.3.2 | Back-propagation neural networks (BPNN)

In contrast to FFNNs, BPNNs involve a back propagation algorithm such stochastic gradient descent (SGD) to fine tune the weight parameters at each neural layer based on the error generated by the loss function [81]. BPNN is often integrated with optimization techniques or other ML models for improved performance. For example, wind speed prediction has been performed in [82] by optimizing the weight parameters of BPNN using improved particle swarm optimization. On the other hand, a flower pollination algorithm has been implemented in [83] to optimize the BPNN for wind speed forecasting. In the context of energy consumption forecasting based on socio-econometric factors, a differential evolution technique integrated with adaptive mutation is implemented to optimize the initial connection weights of BPNN [84]. Following subsections involve neural networks based on underlying BPNN algorithm.

### 5.3.3 | Recurrent neural networks (RNN)

ANNs with directed cycles between hidden layers are defined as RNN. Contrary to FFNNs, they have a memory state and can process variable data length, thus facilitating the sequence processing in time-series. Shi *et al.* proposed RNN for STF of granular consumption data for 920 households in Ireland [76]. However, RNNs face the problem of over-fitting and may reflect false positive results for new data values. This issue can be addressed by increasing diversity and volume in individual consumption data by introducing a pooling system of neighboring households with correlated historical consumption. Similarly, RNN is used in [7] to predict short-term energy consumption values for 112 households. However, RNN suffers from the problem of vanishing and exploding gradients during the learning process with data having long sequences.

### 5.3.4 | Long short-term memory (LSTM)

LSTM is a variant of RNN which deals with vanishing/exploding gradients through its cell and gating architecture. In this direction, authors in [27] proposed a RNN-based LSTM technique to perform short-term load forecasting for residential consumers and demonstrated improved accuracy at granular and aggregated level forecasts. In [85], authors included neuro-fuzzy logic in conjunction with the LSTM and found that the technique can outperform other methods. The authors in [86] combined persistence model for predicting sunny weather and utilized this information to improve the solar power forecasting using LSTM integrated with auto-encoders. In [87], authors combined convolutional auto-encoder and K-means algorithm to extract features related to customer behavior and integrated them with LSTM models for short-term energy consumption forecasting.

### 5.3.5 | Convolutional neural networks (CNN)

CNN is a class of deep neural networks, extensively used in feature extraction and data filtering applications. In [88], CNN is used in conjunction with wavelength transform (WT) method to extract features from PV generation dataset to support effective forecasting. Furthermore, Lee *et al.* investigated CNN filters integrated with LSTM to forecast one day-ahead PV solar power in Korea, considering the local weather information alongside [89].

Authors in [90] and [91] have reviewed various DL-based energy forecasting methods and observed that the addition of more layers to a neural network can increase the forecasting accuracy. However, it sometimes overfits the model to a specific dataset as it captures the noise during training phase. To overcome this challenge, dropout [92] was proposed as a potential technique which addresses overfitting by dropping out the random neuron units. However, uncertainty representation in generation data from RES is considered as another challenging task for point forecasting methods discussed in above sections. The uncertainty generally arises due to external factors such as weather conditions and cannot be captured very well by point forecasting methods.

## 6 | PROBABILISTIC TECHNIQUES

With the integration of RES in the modern power grid, forecasting trends are shifting from point to probabilistic in regards to the future demand and generation at disaggregated levels [93]. Hong *et al.* presented a review for probabilistic methods and emphasized their importance over point forecasting with ever changing needs of power industry [94, 95].

### 6.1 | Parametric versus non-parametric approaches

This subsection identifies the existing literature for probabilistic energy forecasting, which is mainly categorized under parametric and non-parametric approaches. The authors in [9] provided a brief review of these two approaches for wind generation forecasting.

Parametric approaches assume a certain probability density function for the parameter distribution, such as normal distribution. Dowell *et al.* proposed a parametric probabilistic scheme based on Bayesian probability and sparse VAR to forecast very short-term wind power generation in Southeastern Australian wind farms with a 5 min interval [96]. The authors confirmed that their method achieves least RMSE in comparison to the standard AR and VAR methods. They further utilized the similar parametric approach in [51] to forecast long-term probabilistic horizons for load consumption. A hybrid probabilistic deterministic approach for wind generation forecasting has been developed in [97], where temporally local Gaussian processes are used to investigate forecasting errors. Furthermore, for the application of price forecasting, probabilistic methods have

been utilized by various authors and the relevant contributions are outlined in [98–100]. Though this approach simplifies the analysis and reduces computation cost, sometimes the parameter distribution may not accurately fit with a known function [9].

On the other hand, non-parametric approaches do not assume a fixed function for the probability density of the output parameters. In this approach, the predictive probability densities of the parameter are represented by a range of quantile forecasts [9]. For wind power forecasting in the presence uncertainties, the authors in [101] implemented a kernel density estimation model and represented the wind power with a number of kernel functions. An ensemble-based probabilistic forecasting model was developed in [102] which transforms meteorological data to wind power output and generates predictive distributions in a non-parametric manner. On the other hand, a non-parametric approach is considered in [103] to obtain prediction intervals as outputs of neural network models for generating wind power forecasts.

## 6.2 | State-of-the-art methods

### 6.2.1 | Quantile regression

Quantile regression is an extended version of linear regression and aims to obtain predictive quantiles [9]. Taieb *et al.* in [104] emphasized the need to predict the future demand in probabilistic format, which can further support the grid operations related to the energy generation and distribution. They employed additive quantile regression (QR) to forecast the individual energy consumption for 3696 smart houses in Ireland with 30 min interval. To support probabilistic load forecasting, Liu *et al.* proposed quantile regression averaging (QRA) to obtain PIs for consumption with 90% percentiles [10]. The authors claimed effective forecasts using probabilistic evaluation metrics such as Pinball and Winkler score.

However, the existing probabilistic approaches in energy forecasting often suffer high computational complexities and thus, more efficient methods need to be developed and reviewed considering the current energy market scenarios [105, 106].

- *Quantile regression averaging (QRA)*: It is a combination approach for forecasting to compute prediction intervals. It involves applying quantile regression to the smaller forecast models in a hybrid form. The authors in [107] performed price forecasting using QRA to improve the profitability of electricity trading. In terms of load forecasting, the authors in [108] performed QRA on point forecasts to develop a computationally inexpensive yet accurate estimate of the load variability.
- *Quantile regression Factoring (QRF)*: This approach utilises a factor model to account for unobserved factors that impacts the distribution of the observable parameters. Quantile factor analysis is performed to estimate the number of factors for each quantile [109]. The authors in [110] developed a

multi-factor QR model to demonstrate how electricity price forecasts are impacted by volatility of market prices, demand forecasts as well as oil, coal and gas prices. In [99] authors extended QRA integrated with factor models and principal component analysis (PCA) to provide interval price forecasts. After applying QRA, PCA is used to select from individual forecast models available for averaging.

### 6.2.2 | Robabilistic deep learning (PDL)

Bayesian probability incorporated with DL methods can be used to provide forecasting results in the form of PIs, contrary to traditional deep neural networks that are deterministic in nature and generate point forecasts. PDL expresses model parameters as a function of probability distributions such as normal distribution. To be precise, PDL models can predict future PIs with different percentiles that can explain the certain or uncertain factors in the energy data and hence, enables better decision making. This section outlines the significant contributions in the field of PDL to support the applications of demand, generation, and price forecasting in modern power systems.

- *Bayesian neural networks (BNN)*: The concept of Bayesian probability incorporated with ANN is defined as BNN. Yang *et al.* in [12] proposed BNN technique to forecast individual energy demands at household level after quantifying the shared uncertainties between different customer groups. In addition, a clustering-based data pooling system is presented to tackle the issue of over-fitting by increasing the data volume and diversity. The authors demonstrated lower Winkler and Pinball scores for probabilistic methods. Furthermore, a comparative study is presented with the benchmark point forecasting and probabilistic techniques such as QRA and quantile regression factoring (QRF).
- *Bayesian LSTM*: Sun *et al.* proposed a BNN integrated LSTM approach to curb the challenges posed by weather uncertainty in distributed PV generators and thereafter generated the net-load forecasts in the form of PIs with greater accuracies [77]. In addition, they improved the forecasting performance by clustering individual sub profiles based on similar energy consumption patterns prior to applying Bayesian approach. They implemented their method on a real SG dataset acquired from AusGrid involving rooftop PV generation measurements for every half an hour for a period of 3 years.
- *Bayesian BLSTM*: In a similar manner, authors in [39] proposed a PDL technique to deal with the problem of uncertainty in energy markets. The authors integrated LSTM with bidirectional RNN by enabling the propagation of training sequence forwards and backwards and proposed their method as bidirectional-LSTM (BLSTM). The proposed network is then trained to generate a Gaussian or a non-parametric predictive distribution of the dependent variables present in the energy data such as PV generation. Furthermore, Copula-based sampling is employed over predicted distributions to generate predictive scenarios.

However, probabilistic methods are computationally expensive due to large sample space. In this regard, [111] proposed dropout as one of the potential solutions which works as an approximator to make the Bayesian inference process less complex, computationally. However, the computational complexity of the Bayesian approach remains one of the main concerns and therefore, more generalized and effective solutions need to be explored in the future research.

- *Deep quantile regression*: These methods combine QR methods with deep learning models to predict the quantiles involved with the uncertainty. The authors in [112] used QR neural network to forecast load probability density in a short-term horizon. On the other hand, regional wind power forecasts using QR neural networks have been investigated in [113], where the authors used ramp functions to avoid the crossing of multiple quantiles. Similarly, LSTM integrated with penalised QR has been implemented in [114] to obtain probabilistic load forecasts in terms of PI.
- *Deep Gaussian processes*: Deep Gaussian processes involve stacking of Gaussian processes as in the layers of neural networks. Each Gaussian process layer works as a single layer neural network. These processes can learn the model uncertainty similar to Bayesian models. Deep Gaussian processes have been used for load forecasting in [115] and [116]. The authors in [115] found that this method becomes computationally intensive for larger data sets in comparison to other benchmark probabilistic deep learning methods. However, the benefit lies in the requirement for smaller number of training samples, which is critical for STF [116].

## 7 | HYBRID METHODS

Statistical methods face challenges from their inability to process big data, while AI-based methods have other shortcomings in terms of model complexity and dependence on large training datasets. Furthermore, deep neural networks struggle with overfitting, vanishing/exploding gradients, and appropriate hyper-parameter tuning. Therefore, a hybrid approach involving two or more methods could be more useful to overcome aforementioned limitations of state-of-the-art point forecasting methods [117]. In this direction, different learning and statistical methods can be integrated together along with optimization techniques to work as one holistic approach. In this context, authors in [21] presented a comparative analysis of various hybrid techniques by integrating ML methods with optimization algorithms such as genetic algorithm (GA) to enhance the model efficacy for PV generation forecasting. Various hybrid methods along with their literature work for energy forecasting are discussed as below.

### 7.1 | GA-SVM

In [118], authors combined four different state-of-the-art forecasting methods namely, ARIMA, SVM, ANN, and adaptive neurofuzzy inference system (ANFIS) with GA optimization

technique to forecast 1-h-ahead PV power generation by utilizing green energy office building data in Malaysia. Additionally, authors considered the solar irradiance, air, and the historical solar power data as the input features. They claimed high precision in the prediction outputs with minimum error of 5.64% with hybrid approach as compared to individual aforementioned methods.

## 7.2 | AR-net

As classical methods lack long-range dependencies (LRD) and neural networks are complex and lack interpretability, authors in [119] presented a simple yet efficient technique in which AR process is modeled using neural networks and named as AR-net. To be precise, AR-net mimics the traditional AR process with the FFNNs. Furthermore, the AR coefficients which are trained using least square method in existing approaches, are fitted using stochastic SGD in AR-net.

The parameters of first layer in AR-net are taken as equivalent to the classical AR-coefficients. In addition, authors discussed the equivalences between FFNNs and classic AR models to overcome scalability issues while ensuring model interpretability and simplicity. They emphasized the use of deep neural networks such as RNN and CNN to deal with scalability issues. But DL models are highly complex to interpret and understand the dynamics of models and need appropriate consideration of the suitability in each time-series application.

## 7.3 | K-means-ARIMA

In [120] authors combined k-means clustering with autoregressive neural networks for hourly solar forecasts. Similarly, authors in [121] used wavelet packet transform along with ANN for the application of wind power forecasting.

However, aforementioned hybrid methods posed a common issue of escalating the computational complexity. In addition, if one method in the hybrid combination has the poor performance, it is going to affect the overall hybrid method. However, by including an appropriate optimization algorithm to optimize model parameters, the efficacy and performance can be improved. Thus, maintaining a tradeoff between the computational complexity and performance is of utmost importance for the hybrid approach.

## 7.4 | PSO-SVR

In [122], Particle swarm optimization (PSO) algorithm is used to optimize the parameters of SVR for the application of yearly peak load and energy demand forecasting of Iran National Energy Electric Energy System. In addition, a hybrid approach is applied to reduce the forecasting error by combining ARIMA, ANN, and the proposed PS-SVR technique for long-term load forecasting.

# 8 | DATA PRE-PROCESSING TECHNIQUES FOR ENERGY FORECASTING

Data representation and pre-processing is one of the early key stages in data analytics. Processing raw data with irrelevant and redundant information can produce misleading and incorrect forecasting results. Data pre-processing involves various methods such as normalization, cleaning, transformation, feature engineering, dimensionality reduction etc. It transforms raw data to a final training dataset which can be further fed to the data processing techniques [123]. For energy systems, various authors have emphasized the use of different pre-processing techniques to improve the forecasting accuracy, as discussed in the following subsections.

## 8.1 | Singular value decomposition (SVD)

Dimensionality reduction and feature engineering are two highly regarded pre-processing techniques. Kaur *et al.* in [7] and [14] used SVD as the base technique to decompose high dimensional data into lower dimensions prior to energy forecasting. The authors employed tensors and SVD matrix decomposition to achieve dimensionality reduction. Similarly, authors in [124] presented a data decomposition and compression method for user load profiles based on k-means clustering and SVD (k-SVD) method. Firstly, a sparse coding technique is used over load profiles to compress the data by exploiting the sparsity present in load profiles. Then, k-SVD is proposed to decompose and extract the partial usage patterns (PUPs) in it. They claimed that their technique outperforms other existing data compression methods namely discrete wavelet transform (DWT) and standalone k-means clustering.

## 8.2 | Principal component analysis (PCA)

PCA is a classical dimensionality reduction technique used to transform the higher data dimensions into lower representations in the form of orthogonal matrix known as principal components. It is proposed in [17] to eliminate the features from a PV generation dataset that has little or no significance in the forecasting output. It allows the proposed method to make more precise predictions. Authors compared their results with differential evolution (DE) and particle swarm optimization (PSO) and observed least RMSE values using their method.

## 8.3 | Auto-encoders (AE)

AEs are type of neural networks used to encode high dimensional data into lower representations using the encoding layer. The performance of AEs is checked with the help of decoding layer by monitoring the reconstruction ratio. Furthermore, there

**TABLE 1** Existing state-of-the-art forecasting methods: A comparative summary

Sr. no.	Methods	References	Advantages	Disadvantages
M1	Statistical	[33, 34, 46, 54, 56–59]	Simple and low computational cost	Not reliable for big data and nonlinear data components in energy datasets
M2	Machine learning (ML)	[21, 22, 64, 66, 68–72, 74, 75, 123, 133–135]	Efficient and capable to deal with big energy datasets	Not reliable for heterogeneous and long sequence data problems, generate point forecasts
M3	Deep learning (DL)	[5, 7, 23, 27, 35, 43, 53, 67, 76, 78–80, 90–92, 136, 137]	Efficient for long and nonlinear data patterns	Overfitting, require appropriate hyper-parameter tuning, generate point forecasts
M4	Probabilistic methods	[9, 10, 47, 51, 93, 94, 96, 104–106]	Provide prediction intervals (PIs)	High computational complexity
M5	Probabilistic deep learning (PDL)	[11, 12, 39, 77, 111, 138]	Provide PIs, reliable, and efficient	Limited literature and high computational cost
M6	Hybrid (ensemble)	[13, 21, 49, 118, 119, 121, 139, 140]	High accuracy and scalable	Application specific, high computational complexity
M7	Pre-processing techniques	[16, 19, 124, 125, 128–131]	Improve forecasting performance	Application specific

exist different variants of AEs such as sparse auto-encoders (SAE). A special class of SAE is used in [125] as a framework to identify the over-voltage to classify the faults and determine the power quality disturbances. Furthermore, AEs are used in [86] with LSTM to improve PV estimation accuracy in variable weather conditions. The authors claimed to reduce the impact of weather uncertainty using encode-decoder framework for day-ahead PV forecasting for multiple locations.

#### 8.4 | Convolutional auto-encoders (CAE)

AEs incorporated with the convolutional layers of CNN are termed as CAEs. In [19], authors emphasized the role of data compression and dimensionality reduction using CAEs from the perspective of data storage and transmission. They proposed a feature extraction technique based on CAEs to capture daily and seasonal variations by representing 8640 dimensional space into 100-dimensional vector. The authors reported 19–40% decline in the reconstruction error using their technique and a 130% increase in compression ratio as compared to standard methods. Furthermore, Shao *et al.* used CAE integrated with LSTM to perform multi-step STLF involving energy, time, and user behavior components. Firstly, they combined CAE and k-means to identify user behavior and then employed LSTM to train time and energy components for STLF with 10% improvement in overall prediction performance [87].

#### 8.5 | Variational auto-encoders (VAE)

VAEs are relatively a new class of AEs in energy systems which employ the concept of variational inference and Bayesian optimization to carry out the encoding and decoding operations [126, 127]. VAE integrated with recurrent neural layers and its variants have been recently explored for the applica-

**TABLE 2** Energy forecasting surveys from past 5 years

Publication Year	References	M1	M2	M3	M4	M5	M6	M7
2016	[141]	✓	✓	✓	✗	✗	✗	✗
2017	[90]	✓	✓	✓	✗	✗	✗	✗
2018	[142]	✗	✗	✓	✓	✗	✗	✗
2019	[6]	✓	✓	✓	✓	✗	✗	✗
2020	[143]	✗	✗	✓	✓	✓	✗	✗
2020	[144]	✗	✓	✓	✓	✗	✓	✗
2021	[145]	✗	✓	✓	✓	✗	✓	✓
2022	[146]	✓	✓	✓	✓	✗	✓	✓
2022	Proposed survey	✓	✓	✓	✓	✓	✓	✓

M1: statistical; M2: ML; M3: DL; M4: probabilistic; M5: PDL; M6: Hybrid; M7: pre-processing.

tion of anomaly detection and to make the forecasting process more efficient [128–131]. To reduce the dimensionality arising from time lags considered in renewable generation forecasting, VAEs have been implemented in [132] before performing forecasts with BLSTM. VAE combined with Bayesian BLSTM has been employed in [15] to carry out dimensionality reduction of model parameters and forecasting, respectively, for solar generation dataset. The authors reported lower reconstruction and RMSE errors using their technique along with uncertainty quantification in model parameters of BLSTM.

Table 1 presents a comparative summary of aforementioned forecasting methods along with relevant references, advantages, and disadvantages of each category. Furthermore, Table 2 reflects the surveys published in the last five years with respect to number of methods reviewed in each. It can be observed that no survey to the best of our knowledge has attempted to review all categories of methods comprehensively. We attempt to propose a more comprehensive review especially including probabilistic approaches.

## 9 | CASE STUDY

### 9.1 | Comparative analysis of benchmark methods

This section presents a case study for comparative analysis of various forecasting methods discussed in above sections using two different datasets. To be specific, RNN and LSTM are chosen as benchmark DL models due to their superior performance reported in existing papers. For similar reasons, ARIMA, SVR, Bayesian ANN, and BLSTM are chosen as representatives of commonly used statistical, ML, and PDL methods, respectively.

#### 9.1.1 | Data description

In order to evaluate the performance of forecasting methods, two energy datasets from the AEP [147] and Victorian energy consumption benchmark [148] are taken. AEP dataset consists of hourly energy consumption values from October 2004 to August 2018. Victorian dataset includes half hourly energy consumption values from April 2012 to March 2014 for 25 Victorian houses. To implement point forecasting methods, 5 time lags (equivalent to 5 h for AEP dataset and 2.5 h for Victorian dataset) are considered. For PDL methods, 24 time lags (equivalent to 24 h for AEP dataset and 12 h for Victorian dataset) are considered. Results are implemented in python environment using Keras and Tensorflow libraries.

#### 9.1.2 | Evaluation metrics

RMSE and MAE are the most commonly used evaluation metrics to evaluate the point forecasting methods [39],[77]. The RMSE represented by  $\epsilon$  and MAE represented by  $\rho$  are defined using following equations:

$$\epsilon = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_{pred}(t) - Y_{act}(t))^2}, \quad (1)$$

$$\rho = \frac{1}{n} \sum_{t=1}^n |Y_{pred(t)} - Y_{act(t)}|, \quad (2)$$

where  $Y_{pred}(t)$  and  $Y_{act}(t)$  are the predicted and actual values at time stamp  $t$  and  $n$  is the total number of samples, respectively. Thus, the main objective for forecasting model is to minimize the error in predicted values as:

$$\min_{\theta} \epsilon, \rho, \quad (3)$$

where  $\theta$  represents the model parameters for each forecasting method such as weights and biases or lag coefficients. Furthermore, average Pinball loss is an important evaluation metric for probabilistic methods [12, 77]. So, to evaluate Bayesian ANN and Bayesian BLSTM, average Pinball loss is computed focusing on the sharpness and consistency of the approximated

distribution. In this regard, least value of Pinball loss is more desirable. For actual data values ( $Y_{act}$ ) and predictions at  $t^{th}$  timestamp ( $\hat{Y}_{t,q}$ ), Pinball loss over percentile  $q \in [0, 1]$  is formulated as:

$$Pinball(Y_{act}, \hat{Y}_{t,q}, q) = \begin{cases} (Y_{act} - \hat{Y}_{t,q})q & Y_t \geq \hat{Y}_{t,q} \\ (\hat{Y}_{t,q} - Y_{act})(1 - q) & Y_{act} \leq \hat{Y}_{t,q} \end{cases}. \quad (4)$$

Furthermore, for confidence  $(1 - \gamma) \times 100$ , Winkler score is computed using:

$$Winkler = \begin{cases} \delta & lb_t \leq y_t \geq ub_t \\ \delta + 2(lb_t - y_t)/\gamma & lb_t \geq y_t \\ \delta + 2(y_t - ub_t)/\gamma & ub_t \leq y_t \end{cases}, \quad (5)$$

where  $lb_t$  and  $ub_t$  represent the lower and upper bounds of probabilistic forecasts at interval  $t$ , respectively. And,  $\delta = ub_t - lb_t$  is the PI width at  $t$ .

Furthermore, skill score in the terms of Brier Score (BS) [149] is also computed to evaluate the error in probabilistic predictions defined as:

$$BS = \frac{1}{n} \sum_{t=1}^{t=n} (f_t - y_t)^2, \quad (6)$$

where  $f_t$  stands for estimated forecasting values at interval  $t$  using approximated posterior distribution and  $n$  is the number of samples form testing set.

#### 9.1.3 | Result discussion

As shown in Table 3, both datasets are further divided into two cases each to evaluate the impact of different sample sizes. For the Victorian energy, 2 and 1 year of consumption sample size are considered. While for AEP, 10 and 1 year of consumption data are considered. Then, ARIMA, SVR, RNN, LSTM, CNN, Bayesian ANN, and Bayesian BLSTM are trained with 70% of the total samples for each case and tested for remaining 30%. For DL methods, 40 neural units per layer, 100 epochs, and batch size of 1000 are chosen. *Adam* is used as an optimizer to minimize the loss function of mean squared error (MSE). Additionally, RNN and LSTM methods are considered with two different variations in the form of activation functions namely, hyperbolic tangent function (*tanh*), and rectified linear unit (*ReLU*). Activation functions play an important role in activating the hidden layer in neural networks. Also, SVR is modeled with radial basis function (*rbf*) and *linear* kernel. Then, RMSE and MAE values are computed on the testing dataset using (1) and (2), respectively.

As demonstrated in the table, Bayesian BLSTM performs similar to LSTM (with *tanh*) for larger training datasets and stationary time-series, that is, with 10 years of data for AEP.

**TABLE 3** Comparative analysis on evaluation metrics for Victorian and AEP case study

Sr. no.	Method	Victorian energy consumption dataset				AEP dataset			
		2 years		1 year		10 years		1 year	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
1	ARIMA	—	—	0.0753	0.0566	—	—	0.0429	0.0370
2	SVR-rbf	0.0813	0.0795	0.0792	0.0605	0.0363	0.0298	0.0415	0.0331
3	SVR-linear	0.0676	0.0651	0.0749	0.0562	0.0310	0.0242	0.0362	0.0285
4	RNN-tanh	0.0736	0.0559	0.0774	0.0579	0.0193	0.0141	<b>0.0371</b>	<b>0.0285</b>
5	RNN-relu	<b>0.0631</b>	<b>0.0463</b>	<b>0.0701</b>	<b>0.0519</b>	0.0290	0.0228	0.0527	0.0424
6	LSTM-tanh	0.0805	0.0608	0.0851	0.0637	<b>0.0189</b>	<b>0.0136</b>	0.0472	0.0368
7	LSTM-relu	0.0689	0.0516	0.0804	0.0599	0.0239	0.0183	0.0551	0.0430
8	CNN	0.0665	0.0495	0.0796	0.0597	0.0197	0.0142	0.0445	0.0342
9	Bayesian ANN	0.0481	0.0417	0.0438	0.0405	0.0198	0.0154	0.0564	0.0445
10	Bayesian BLSTM	<b>0.0251</b>	<b>0.0281</b>	<b>0.0305</b>	<b>0.0260</b>	<b>0.0168</b>	<b>0.0126</b>	<b>0.0167</b>	<b>0.0121</b>

Bold values represent the best performing methods in the form of least errors for each case of dataset.

**TABLE 4** Pinball score by PDL forecasting methods

Sr. no.	Method	Victorian (2 years)	Victorian (1 year)	AEP (10 years)	AEP (1 year)
1	Bayesian ANN	0.0019	0.0210	0.0050	0.0150
2	Bayesian BLSTM	0.0016	0.0180	0.0030	0.0070

However, for non-stationary and smaller training samples, Bayesian BLSTM provides the best results by achieving the least error value, as reflected in Table 3. It should be noted that from point forecasting methods, RNN with *Relu* gives second best results in terms of lower error values. In addition, Bayesian BLSTM provides PIs for future predictions and forecasts with more accuracy in relation to the ground truth, as reflected in Figure 6. RNN and LSTM have less error as compared to the statistical ARIMA and traditional ML methods such as SVR. For AEP dataset, LSTM is performing better with larger samples. For smaller sample size, RNN demonstrates less error and computational cost. Models with *tanh* function perform better with uniform data, such as the AEP dataset. However, with higher variability and seasonal trends, that is, for the Victorian dataset, *Relu* performs better as it sparsely activates neurons rather than activating all neurons at the same time. Furthermore, datasets with more samples tend to train better and yield greater accuracy. In addition, ARIMA takes a longer time to train for larger datasets and thus, fails to generate the output. SVR with *linear* kernel has less error compared to SVR with *rbf* kernel. When dataset has linearly separable features, *linear* kernel will be sufficient to achieve superior performance.

Tables 4 and 5 display the Pinball and Winkler scores, respectively, given by Bayesian ANN and Bayesian BLSTM for different data sizes. It can be inferred from the values that in case of more data and hence larger training set, scores are lower indicating the improvement in accuracy. Also, Bayesian

**TABLE 5** Winkler score by PDL forecasting methods

Sr. no.	Method	Victorian (2 years)	Victorian (1 year)	AEP (10 years)	AEP (1 year)
1	Bayesian ANN	0.7264	0.7834	0.1273	0.2341
2	Bayesian BLSTM	0.5432	0.6865	0.0703	0.0902

**TABLE 6** Brier score for PDL forecasting methods

Sr. no.	Method	Victorian (2 years)	Victorian (1 year)	AEP (10 years)	AEP (1 year)
1	Bayesian ANN	0.0624	0.0734	0.0117	0.0141
2	Bayesian BLSTM	0.0432	0.0435	0.0103	0.0102

BLSTM outperforms Bayesian ANN with the help of extensive training capability of bidirectional layer. Similarly, the BS has been computed for the two probabilistic methods across the different dataset sizes in Table 6. The performance comparison shows that Bayesian BLSTM achieves a lower BS for all considered scenarios.

Figure 4 reflects the comparison graphs of errors observed by forecasting methods considered in Table 3 with respect to each sample size. It can be inferred from the figures that for point forecasting, standard DL methods provide least error irrespective of the data size and variability. However for PDL forecasting, Bayesian BLSTM outperforms all the other comparative methods and generate PIs.

Furthermore, Figure 5 shows the actual and predicted energy consumption values using RNN and LSTM on each dataset. The results are plotted for activation functions which produce superior performance for a given algorithm. Normalized energy consumption values are plotted over half hourly and per hour intervals for Victorian and AEP dataset, respectively.

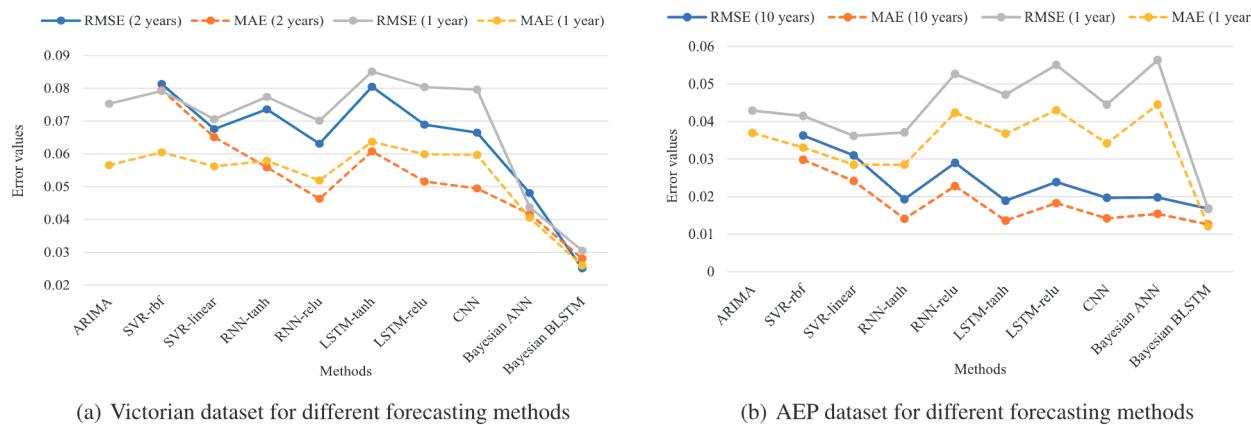


FIGURE 4 Comparative analysis of forecasting methods based on evaluation metrics for different datasets with different sample sizes

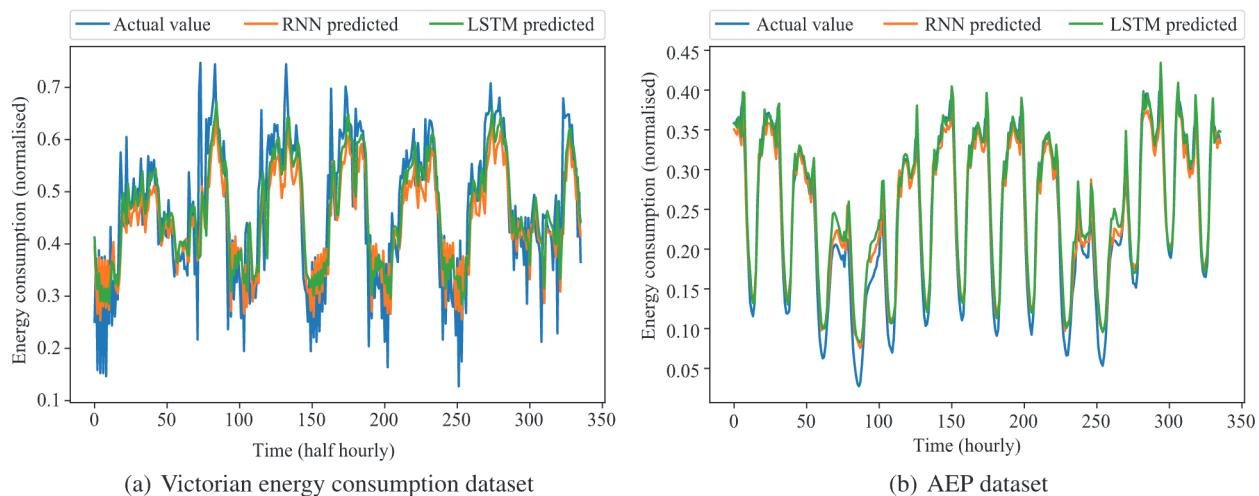


FIGURE 5 Predicted versus actual values for RNN and LSTM with one week and two weeks of test data for the Victorian and AEP dataset, respectively

From Figure 5a, it is inferred that the Victorian dataset has higher variability leading to more uncertainty in the energy consumption and hence, the performance of the models is slightly poorer. To deal with its non-stationary nature, logarithmic transformation is adapted. However, in Figure 5b, AEP dataset is uniformly distributed for which predicted values show greater similarity with actual values and thus, less errors. Furthermore, Figure 6 represents a comparison graph of actual energy consumption values (ground truth) with predictive mean over 90% and 50% prediction PIs generated by Bayesian BLSTM on both the datasets. PIs reflect the intervals for future probabilities on different percentiles which enable the PDL methods to quantify for uncertainties. Figure 6a clearly shows wider PIs as compared to Figure 6b as there is more variability in Victorian consumption values than AEP.

In addition, a comparative plot for the execution time taken by each method is presented in Figure 7. Note that, the reported execution time precisely belongs to the training process of each forecasting method. Time taken for error computation and parameter tuning is not included. As shown in the figure, PDL

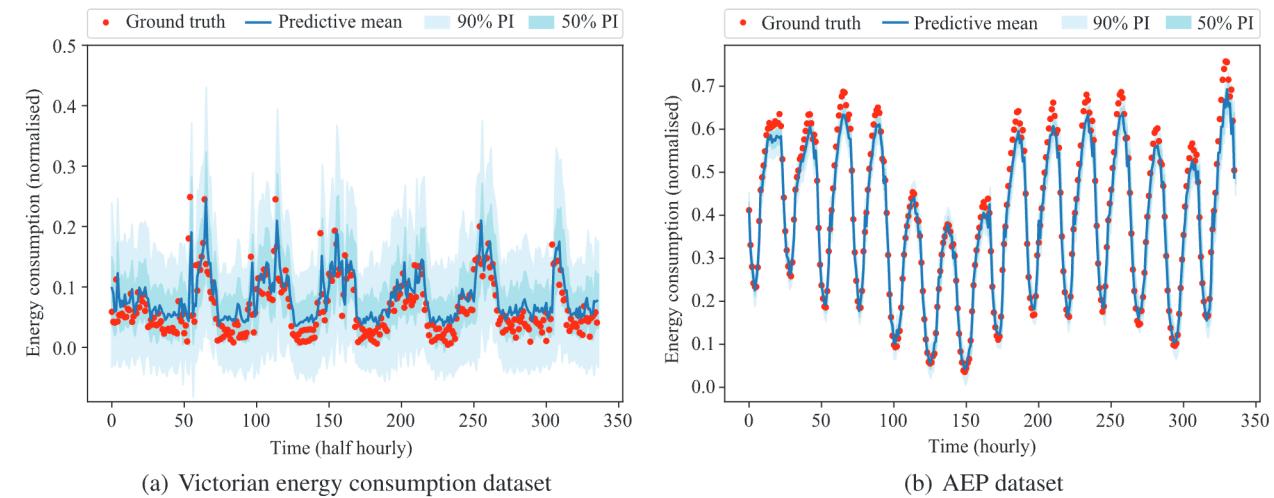
methods take more time compared to point forecasting methods. There is tradeoff involved between the time complexity and forecasting accuracy for the PDL methods, which needs to be maintained.

## 9.2 | Comparative analysis of PDL methods for renewable generation forecasting

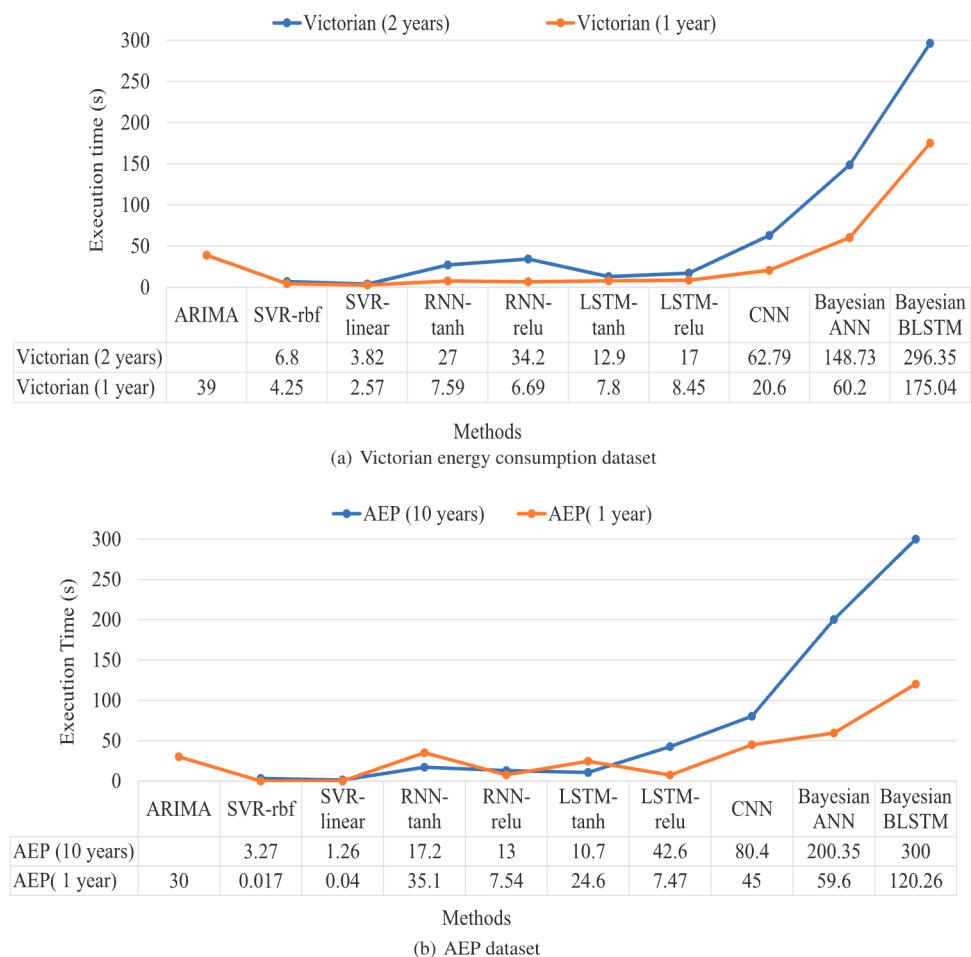
The following two subsections represent the extended case study of probabilistic methods for solar and wind generation data. It is important to note that the hyper parameter settings for all the three cases is kept same with the optimization function as Adam.

### 9.2.1 | Solar case study

This subsection includes the performance evaluation of probabilistic methods implemented on solar generation data taken



**FIGURE 6** Prediction intervals (PIs) with ground truth and predictive mean using Bayesian BLSTM



**FIGURE 7** Comparative analysis of execution time for respective forecasting method

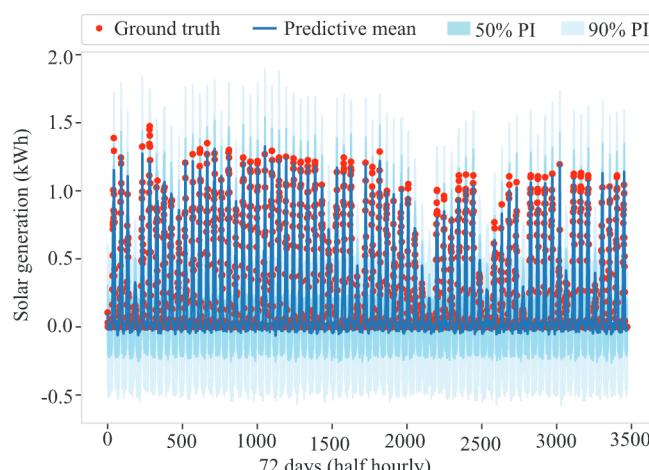
from Ausgrid [150]. The data has been acquired every 15 min for 1 year over the year 2012–2013.

Table 7 includes the numerical results for prediction errors by Bayesian ANN and Bayesian BLSTM. It can be seen from

the results that BLSTM outperforms the ANN when integrated with Bayesian probabilistic method. Furthermore, Figure 8 presents the prediction intervals for solar case study on the test data.

**TABLE 7** Solar forecasting by PDL forecasting methods

Sr. no.	Method	RMSE	MAE	Pinball	Winkler	Brier
1	Bayesian ANN	0.4553	0.3990	0.7902	10.643	5.8601
2	Bayesian BLSTM	0.0998	0.0641	0.1424	0.3860	0.0244

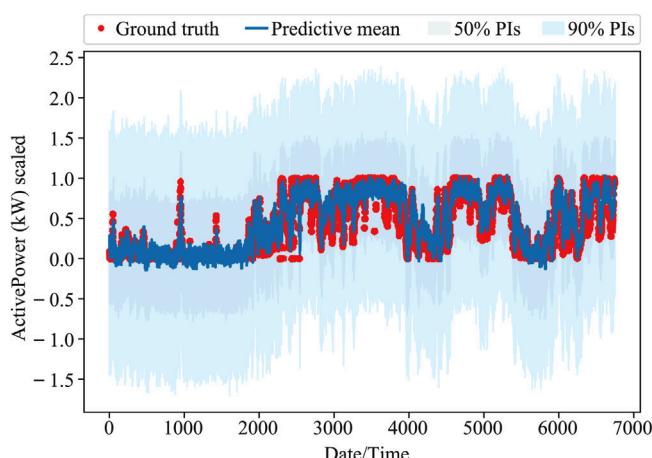
**FIGURE 8** Prediction intervals (PIs) with ground truth and predictive mean using Bayesian BLSTM for solar data**TABLE 8** Wind forecasting results for PDL forecasting methods

Sr. no.	Method	RMSE	MAE	Pinball	Winkler	Brier
1	Bayesian ANN	0.1489	0.3421	0.3410	4.231	2.124
2	Bayesian BLSTM	0.1396	0.1021	0.1807	3.4882	1.286

### 9.2.2 | Wind case study

This section elaborates the wind case study carried on wind generation data. The dataset is acquired from wind turbines in Turkey [151] with an interval of 10 min from 1 January, 2018 to 31 December 2018. The data features include wind speed (m/s), active power (kW), theoretical power cover (kWh), and wind direction. The dataset contains total 50,530 number of samples and 5 features as discussed above. The active power generation feature is used a target variable to predict wind power for the short-term horizon, that is, for the next 10 min. The data is normalized between the range of (0,1) before training using Bayesian BLSTM on the past one day's wind generation data, which is 144 number of samples/lags (Table 8).

Table 9 demonstrates the numerical results for wind forecasting using probabilistic methods namely, Bayesian ANN and BLSTM, respectively. Bayesian BLSTM exhibits significant improvement in pinball score and moderate improvements in the other scores when compared with Bayesian ANN. Furthermore, Figure 9 illustrates the comparison graph between the prediction intervals at percentile 50 and 90 with predictive mean and ground truth using Bayesian BLSTM for wind data.

**FIGURE 9** Prediction intervals (PIs) with ground truth and predictive mean using Bayesian BLSTM for wind data

## NOMENCLATURE

PDL	Probabilistic deep learning
DL	Deep learning
SG	Smart grid
ML	Machine learning
AEP	American electric power
AMI	Advanced metering infrastructure
ARIMA	Autoregressive integrated moving averages
AI	Artificial intelligence
RNN	Recurrent neural networks
LSTM	Long short-term memory
PI	Prediction intervals
DR	Dimensionality reduction
PCA	Principal component analysis
SVD	Singular value decomposition
AE	Autoencoders
BLSTM	Bidirectional long short-term memory
DER	Distributed energy resources
RMSE	Root mean-square error
VAR	Vector autoregressive
NEM	National energy market
ANN	Artificial neural networks
RES	Renewable energy sources
PV	Photo-voltaic
VSTF	Very short-term forecasting
STF	Short-term forecasting
MTF	Medium-term forecasting
LTf	Long-term forecasting

## AUTHOR CONTRIBUTIONS

Devinder Kaur: Conceptualization, data curation, formal analysis, investigation, methodology, resources, software, validation, visualization, writing - original draft, writing - review and editing. Shama Naz Islam: Conceptualization, methodology, project administration, resources, supervision, writing - original draft, writing - review and editing. Md Apel Mahmud: Project administration, resources, supervision, writing - original draft, writing - review and editing. Enamul Haque: Resources,

writing - original draft, writing - review and editing, Zhao Yang Dong (GE): Resources, writing - original draft, writing - review and editing.

## CONFLICT OF INTEREST

The authors have declared no conflict of interest.

## CONCLUSION AND FUTURE DIRECTIONS

In this paper, a comprehensive review of classical and advanced forecasting methods is presented for modern energy systems. A number of statistical, AI-based, probabilistic, and hybrid methods are discussed in detail with respect to their applications in energy systems. In addition, impact of data pre-processing techniques on forecasting performance is also highlighted. After conducting a comparative case study on two different datasets, it is inferred that DL methods with appropriate activation function and hyper-parameter tuning yield higher forecasting accuracy than the traditional statistical and ML methods. However, uncertainty in the energy data from exogenous factors is a major challenge that can be tackled more efficiently with probabilistic methods. To support this, we implemented Bayesian ANN and Bayesian BLSTM as PDL forecasting techniques and the latter outperformed all the other comparative methods by demonstrating least error values. However, it exhibit high computational cost in the terms of time complexity and processing units. So, probabilistic forecasting techniques need to be explored more in the future work and potential solutions to reduce computational cost need to be proposed in the energy domain. Moreover, forecasting multi-dimensional demand and generation while ensuring high accuracy is another issue that requires substantial attention.

## ORCID

Md Apel Mahmud  <https://orcid.org/0000-0002-5302-5338>

## REFERENCES

- Gungor, V.C., Sahin, D., Kocak, T., Ergut, S., Buccella, C., Cecati, C., et al.: Smart grid technologies: Communication technologies and standards. *IEEE Trans. Ind. Inf.* 7(4), 529–539 (2011)
- Çimen, H., Çetinkaya, N., Vasquez, J.C., Guerrero, J.M.: A microgrid energy management system based on non-intrusive load monitoring via multitask learning. *IEEE Trans. Smart Grid* (2020)
- Islam, S.N., Baig, Z., Zeadally, S.: Physical layer security for the smart grid: Vulnerabilities, threats, and countermeasures. *IEEE Trans. Ind. Inf.* 15(12), 6522–6530 (2019)
- Agoua, X.G., Girard, R., Kariniotakis, G.: Short-term spatio-temporal forecasting of photovoltaic power production. *IEEE Trans. Sustainable Energy* 9(2), 538–546 (2017)
- Jahangir, H., Tayarani, H., Gougheri, S.S., Golkar, M.A., Ahmadian, A., Elkamel, A.: Deep learning-based forecasting approach in smart grids with micro-clustering and bi-directional LSTM network. *IEEE Trans. Ind. Electron.* 68(9), 8298–8309 (2021)
- Wang, Y., Chen, Q., Hong, T., Kang, C.: Review of smart meter data analytics: Applications, methodologies, and challenges. *IEEE Trans. Smart Grid* 10(3), 3125–3148 (2018)
- Kaur, D., Kumar, R., Kumar, N., Guizani, M.: Smart grid energy management using rnn-lstm: A deep learning-based approach. In: 2019 IEEE Global Communications Conference (GLOBECOM), pp. 1–6. IEEE, Piscataway (2019)
- Kaur, D., Islam, S.N., Mahmud, M., et al.: A bayesian deep learning technique for multi-step ahead solar generation forecasting. *arXiv preprint arXiv:220311379* (2022)
- Zhang, Y., Wang, J., Wang, X.: Review on probabilistic forecasting of wind power generation. *Renew. Sustain. Energy Rev.* 32, 255–270 (2014)
- Liu, B., Nowotarski, J., Hong, T., Weron, R.: Probabilistic load forecasting via quantile regression averaging on sister forecasts. *IEEE Trans. Smart Grid* 8(2), 730–737 (2015)
- Ali, M., Prakash, K., Pota, H.: A bayesian approach based on acquisition function for optimal selection of deep learning hyperparameters: A case study with energy management data. *Sci. Proc. Ser. 2(1)*, 22–27 (2020)
- Yang, Y., Li, W., Gulliver, T.A., Li, S.: Bayesian deep learning-based probabilistic load forecasting in smart grids. *IEEE Trans. Ind. Inf.* 16(7), 4703–4713 (2019)
- Torabi, M., Hashemi, S., Saybani, M.R., Shamshirband, S., Mosavi, A.: A hybrid clustering and classification technique for forecasting short-term energy consumption. *Environ. Prog. Sustain. Energy* 38(1), 66–76 (2018)
- Kaur, D., Aujla, G.S., Kumar, N., Zomaya, A.Y., Perera, C., Ranjan, R.: Tensor-based big data management scheme for dimensionality reduction problem in smart grid systems: Sdn perspective. *IEEE Trans. Knowl. Data Eng.* 30(10), 1985–1998 (2018)
- Kaur, D., Islam, S.N., Mahmud, M., et al.: A vae-based bayesian bidirectional lstm for renewable energy forecasting. *arXiv preprint arXiv:210312969* (2021)
- Manjili, Y.S., Vega, R., Jamshidi, M.M.: Data-analytic-based adaptive solar energy forecasting framework. *IEEE Syst. J.* 12(1), 285–296 (2017)
- Huang, Y.C., Huang, C.M., Chen, S.J., Yang, S.P.: Optimization of module parameters for pv power estimation using a hybrid algorithm. *IEEE Trans. Sustain. Energy* 11(4), 2210–2219 (2020)
- de Souza, J.C.S., Assis, T.M.L., Pal, B.C.: Data compression in smart distribution systems via singular value decomposition. *IEEE Trans. Smart Grid* 8(1), 275–284 (2015)
- Ryu, S., Choi, H., Lee, H., Kim, H.: Convolutional autoencoder based feature extraction and clustering for customer load analysis. *IEEE Trans. Power Syst.* 35(2), 1048–1060 (2019)
- Voyant, C., Notton, G., Kalogirou, S., Nivet, M.L., Paoli, C., Motte, F., et al.: Machine learning methods for solar radiation forecasting: A review. *Renewable Energy* 105, 569–582 (2017)
- Akhter, M.N., Mekhilef, S., Mokhlis, H., Shah, N.M.: Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. *IET Renewable Power Gener.* 13(7), 1009–1023 (2019)
- Kumar, D.S., Yagli, G.M., Kashyap, M., Srinivasan, D.: Solar irradiance resource and forecasting: a comprehensive review. *IET Renewable Power Gener.* 14(10), 1641–1656 (2020)
- Wang, H., Lei, Z., Zhang, X., Zhou, B., Peng, J.: A review of deep learning for renewable energy forecasting. *Energy Convers. Manage.* 198, 111799 (2019)
- Hernandez, L., Baladron, C., Aguiar, J.M., Carro, B., Sanchez-Esguevillas, A.J., Lloret, J., et al.: A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings. *IEEE Commun. Surv. Tutorials* 16(3), 1460–1495 (2014)
- Rossi, B., Chren, S.: Smart grids data analysis: A systematic mapping study. *IEEE Trans. Ind. Inf.* 16(6), 3619–3639 (2019)
- Hsiao, Y.H.: Household electricity demand forecast based on context information and user daily schedule analysis from meter data. *IEEE Trans. Ind. Inf.* 11(1), 33–43 (2014)
- Kong, W., Dong, Z.Y., Jia, Y., Hill, D.J., Xu, Y., Zhang, Y.: Short-term residential load forecasting based on lstm recurrent neural network. *IEEE Trans. Smart Grid* 10(1), 841–851 (2017)
- Sha, H., Xu, P., Lin, M., Peng, C., Dou, Q.: Development of a multi-granularity energy forecasting toolkit for demand response baseline calculation. *Appl. Energy* 289, 116652 (2021)
- Albogamy, F.R., Hafeez, G., Khan, I., Khan, S., Alkhannash, H.I., Ali, F., et al.: Efficient energy optimization day-ahead energy forecasting in smart grid considering demand response and microgrids. *Sustainability* 13(20), (2021)

30. Chan, S.C., Tsui, K.M., Wu, H.C., Hou, Y., Wu, Y.C., Wu, F.F.: Load/price forecasting and managing demand response for smart grids: Methodologies and challenges. *IEEE Signal Process Mag.* 29(5), 68–85 (2012)
31. Hu, J., Zhou, H., Zhou, Y., Zhang, H., Nordströmd, L., Yang, G.: Flexibility prediction of aggregated electric vehicles and domestic hot water systems in smart grids. *Engineering* 7(8), 1101–1114 (2021)
32. Voyant, C., Motte, F., Notton, G., Fouilloy, A., Nivet, M.L., Duchaud, J.L.: Prediction intervals for global solar irradiation forecasting using regression trees methods. *Renew. Energy* 126, 332–340 (2018)
33. Liu, Y., Roberts, M.C., Sioshansi, R.: A vector autoregression weather model for electricity supply and demand modeling. *J. Mod. Power Syst. Clean Energy* 6(4), 763–776 (2018)
34. Messner, J.W., Pinson, P.: Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting. *Int. J. Forecasting* 35(4), 1485–1498 (2019)
35. Pino, R., Parreno, J., Gomez, A., Priore, P.: Forecasting next-day price of electricity in the spanish energy market using artificial neural networks. *Eng. Appl. Artif. Intell.* 21(1), 53–62 (2008)
36. Zhao, J.H., Dong, Z.Y., Li, X., Wong, K.P.: A framework for electricity price spike analysis with advanced data mining methods. *IEEE Trans. Power Syst.* 22(1), 376–385 (2007)
37. Yang, J., Zhao, J., Luo, F., Wen, F., Dong, Z.Y.: Decision-making for electricity retailers: A brief survey. *IEEE Trans. Smart Grid* 9(5), 4140–4153 (2017)
38. Alamaniotis, M., Bargiolas, D., Bourbakis, N.G., Tsoukalas, L.H.: Genetic optimal regression of relevance vector machines for electricity pricing signal forecasting in smart grids. *IEEE Trans. Smart Grid* 6(6), 2997–3005 (2015)
39. Toubeau, J.F., Bottieau, J., Vallée, F., De-Grève, Z.: Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets. *IEEE Trans. Power Syst.* 34(2), 1203–1215 (2018)
40. Liu, M., Quilumba, F.L., Lee, W.J.: Dispatch scheduling for a wind farm with hybrid energy storage based on wind and lmp forecasting. *IEEE Trans. Ind. Appl.* 51(3), 1970–1977 (2014)
41. Soman, S.S., Zareipour, H., Malik, O., Mandal, P.: A review of wind power and wind speed forecasting methods with different time horizons. In: North American Power Symposium 2010, pp. 1–8. IEEE, Piscataway (2010)
42. Hyndman, R.J., Athanasopoulos, G.: Forecasting: principles and practice. OTexts (2018)
43. Guan, C., Luh, P.B., Michel, L.D., Wang, Y., Friedland, P.B.: Very short-term load forecasting: wavelet neural networks with data pre-filtering. *IEEE Trans. Power Syst.* 28(1), 30–41 (2012)
44. Guan, C., Luh, P.B., Michel, L.D., Chi, Z.: Hybrid kalman filters for very short-term load forecasting and prediction interval estimation. *IEEE Trans. Power Syst.* 28(4), 3806–3817 (2013)
45. Potter, C.W., Negnevitsky, M.: Very short-term wind forecasting for tasmanian power generation. *IEEE Trans. Power Syst.* 21(2), 965–972 (2006)
46. López, J.C., Rider, M.J., Wu, Q.: Parsimonious short-term load forecasting for optimal operation planning of electrical distribution systems. *IEEE Trans. Power Syst.* 34(2), 1427–1437 (2018)
47. Fan, S., Hyndman, R.J.: Short-term load forecasting based on a semi-parametric additive model. *IEEE Trans. Power Syst.* 27(1), 134–141 (2011)
48. Ribeiro, G.T., Mariani, V.C., dos Santos Coelho, L.: Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting. *Eng. Appl. Artif. Intell.* 82, 272–281 (2019)
49. Amjadi, N., Daracepour, A.: Midterm demand prediction of electrical power systems using a new hybrid forecast technique. *IEEE Trans. Power Syst.* 26(2), 755–765 (2010)
50. Wang, Y., Zhou, Z., Botterud, A., Zhang, K.: Optimal wind power uncertainty intervals for electricity market operation. *IEEE Trans. Sustainable Energy* 9(1), 199–210 (2017)
51. Hong, T., Wilson, J., Xie, J.: Long term probabilistic load forecasting and normalization with hourly information. *IEEE Trans. Smart Grid* 5(1), 456–462 (2013)
52. Xu, W.t., Wang, Y.l., Li, T., Tang, Q., Zhang, J.f., Shen, L., et al.: Long term intelligent load forecasting method considering the expectation of power market transaction. In: 2017 29th Chinese Control And Decision Conference (CCDC), pp. 2310–2315. IEEE, Piscataway (2017)
53. Azad, H.B., Mekhilef, S., Ganapathy, V.G.: Long-term wind speed forecasting and general pattern recognition using neural networks. *IEEE Trans. Sustain. Energy* 5(2), 546–553 (2014)
54. Deb, C., Zhang, F., Yang, J., Lee, S.E., Shah, K.W.: A review on time series forecasting techniques for building energy consumption. *Renewable Sustain. Energy Rev.* 74, 902–924 (2017)
55. Sansa, I., Boussaada, Z., Mazigh, M., Bellaaj, N.M.: Solar radiation prediction for a winter day using arma model. In: 2020 6th IEEE International Energy Conference (ENERGYCon), pp. 326–330. IEEE, Piscataway (2020)
56. Amini, M., Karabasoglu, O., Ilić, M.D., Boroojeni, K.G., Iyengar, S.: Arima-based demand forecasting method considering probabilistic model of electric vehicles' parking lots. In: 2015 IEEE Power & Energy Society General Meeting, pp. 1–5. IEEE, Piscataway (2015)
57. Amirhosseini, D.Z., Sabzehgar, R., Rasouli, M.: Power management and optimization for a residential smart microgrid using stochastic methods. In: 2018 9th IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG), pp. 1–6. IEEE, Piscataway (2018)
58. Elmouatamid, A., Ouladsine, R., Bakhouya, M., Zine Dine, K., Khairdar, M., et al.: A control strategy based on power forecasting for micro-grid systems. In: 2019 IEEE International Smart Cities Conference (ISC2), pp. 735–740. IEEE, Piscataway (2019)
59. Kedrowski, R., Nelson, J., Nair, A.S., Ranganathan, P.: Short-term seasonal energy forecasting. In: 2018 IEEE International Conference on Electro/Information Technology (EIT), pp. 0696–0700. IEEE, Piscataway (2018)
60. Li, Y., Su, Y., Shu, L.: An armax model for forecasting the power output of a grid connected photovoltaic system. *Renew. Energy* 66, 78–89 (2014)
61. Marchesoni Acland, F., Alonso Suárez, R.: Intra-day solar irradiation forecast using rls filters and satellite images. *Renew. Energy* 161, 1140–1154 (2020)
62. Elamin, N., Fukushige, M.: Modeling and forecasting hourly electricity demand by sarimax with interactions. *Energy* 165, 257–268 (2018). <https://www.sciencedirect.com/science/article/pii/S0360544218319297>
63. Twanabasu, S.R., Bremsdal, B.A.: Load forecasting in a smart grid oriented building. (2013)
64. Ma, Z., Zhang, C., Qian, C.: Review of machine learning in power system. In: 2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia), pp. 3401–3406. IEEE, Piscataway (2019)
65. Kotsiantis, S.B., Zaharakis, I., Pintelas, P.: Supervised machine learning: A review of classification techniques. *Emerg. Artif. Intell. Appl. Comp. Eng.* 160(1), 3–24 (2007)
66. Shi, J., Lee, W.J., Liu, Y., Yang, Y., Wang, P.: Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Trans. Ind. Appl.* 48(3), 1064–1069 (2012)
67. Wang, Y., Chen, Q., Gan, D., Yang, J., Kirschen, D.S., Kang, C.: Deep learning-based socio-demographic information identification from smart meter data. *IEEE Trans. Smart Grid* 10(3), 2593–2602 (2018)
68. Hu, Q., Zhang, S., Yu, M., Xie, Z.: Short-term wind speed or power forecasting with heteroscedastic support vector regression. *IEEE Trans. Sustainable Energy* 7(1), 241–249 (2015)
69. Yang, H.T., Huang, C.M., Huang, Y.C., Pai, Y.S.: A weather-based hybrid method for 1-day ahead hourly forecasting of pv power output. *IEEE Trans. Sustain. Energy* 5(3), 917–926 (2014)
70. Jiang, H., Zhang, Y., Muljadi, E., Zhang, J.J., Gao, D.W.: A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization. *IEEE Trans. Smart Grid* 9(4), 3341–3350 (2016)
71. Ren, Y., Suganthan, P.N., Srikanth, N.: A novel empirical mode decomposition with support vector regression for wind speed forecasting. *IEEE Trans. Neural Networks Learn. Syst.* 27(8), 1793–1798 (2014)

72. Kavousi.Fard, A., Samet, H., Marzbani, F.: A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. *Expert Syst. Appl.* 41(13), 6047–6056 (2014)
73. Ghelardoni, L., Ghio, A., Anguita, D.: Energy load forecasting using empirical mode decomposition and support vector regression. *IEEE Trans. Smart Grid* 4(1), 549–556 (2013)
74. Saeedi, R., Sadanandan, S.K., Srivastava, A., Davies, K., Gebremedhin, A.: An adaptive machine learning framework for behind-the-meter load/pv disaggregation. *IEEE Trans. Ind. Inf.* (2021)
75. Liu, F., Dong, T., Hou, T., Liu, Y.: A hybrid short-term load forecasting model based on improved fuzzy c-means clustering, random forest and deep neural networks. *IEEE Access* (2021)
76. Shi, H., Xu, M., Li, R.: Deep learning for household load forecasting - a novel pooling deep rnn. *IEEE Trans. Smart Grid* 9(5), 5271–5280 (2017)
77. Sun, M., Zhang, T., Wang, Y., Strbac, G., Kang, C.: Using bayesian deep learning to capture uncertainty for residential net load forecasting. *IEEE Trans. Power Syst.* 35(1), 188–201 (2019)
78. Li, Y., Xu, Z., Wang, X., Wang, X.: A bibliometric analysis on deep learning during 2007–2019. *Int. J. Mach. Learn. Cybern.* 11(12), 2807–2826 (2020)
79. Bhaskar, K., Singh, S.N.: Awnn-assisted wind power forecasting using feed-forward neural network. *IEEE Trans. Sustain. Energy* 3(2), 306–315 (2012)
80. Pindoriya, N., Singh, S., Singh, S.: An adaptive wavelet neural network-based energy price forecasting in electricity markets. *IEEE Trans. Power Syst.* 23(3), 1423–1432 (2008)
81. Jiao, R., Huang, X., Ma, X., Han, L., Tian, W.: A model combining stacked auto encoder and back propagation algorithm for short-term wind power forecasting. *IEEE Access* 6, 17851–17858 (2018)
82. Zhang, Y., Chen, B., Zhao, Y., Pan, G.: Wind speed prediction of ipso-bp neural network based on lorenz disturbance. *IEEE Access* 6, 53168–53179 (2018)
83. Qu, Z., Mao, W., Zhang, K., Zhang, W., Li, Z.: Multi-step wind speed forecasting based on a hybrid decomposition technique and an improved back-propagation neural network. *Renew. Energy* 133, 919–929 (2019)
84. Zeng, Y.R., Zeng, Y., Choi, B., Wang, L.: Multifactor-influenced energy consumption forecasting using enhanced back-propagation neural network. *Energy* 127, 381–396 (2017)
85. Motepe, S., Hasan, A.N., Stopforth, R.: Improving load forecasting process for a power distribution network using hybrid ai and deep learning algorithms. *IEEE Access* 7, 82584–82598 (2019)
86. Zhang, Y., Qin, C., Srivastava, A.K., Jin, C., Sharma, R.K.: Data-driven day-ahead pv estimation using autoencoder-lstm and persistence model. *IEEE Trans. Ind. Appl.* 56(6), 7185–7192 (2020)
87. Shao, X., Kim, C.S.: Multi-step short-term power consumption forecasting using multi-channel lstm with time location considering customer behavior. *IEEE Access* 8, 125263–125273 (2020)
88. Wang, H., Yi, H., Peng, J., Wang, G., Liu, Y., Jiang, H., et al.: Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. *Energy Convers. Manage.* 153, 409–422 (2017)
89. Lee, W., Kim, K., Park, J., Kim, J., Kim, Y.: Forecasting solar power using long-short term memory and convolutional neural networks. *IEEE Access* 6, 73068–73080 (2018)
90. Almalqa, A., Edwards, G.: A review of deep learning methods applied on load forecasting. In: 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 511–516. IEEE, Piscataway (2017)
91. Zhang, D., Han, X., Deng, C.: Review on the research and practice of deep learning and reinforcement learning in smart grids. *CSEE J. Power Energy Syst* 4(3), 362–370 (2018)
92. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15(1), 1929–1958 (2014)
93. Gneiting, T., Katzfuss, M.: Probabilistic forecasting. *Annu. Rev. Stat. Appl.* 1, 125–151 (2014)
94. Hong, T., Fan, S.: Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecasting* 32(3), 914–938 (2016)
95. Hong, T., et al.: Short term electric load forecasting. (2010)
96. Dowell, J., Pinson, P.: Very-short-term probabilistic wind power forecasts by sparse vector autoregression. *IEEE Trans. Smart Grid* 7(2), 763–770 (2015)
97. Yan, J., Li, K., Bai, E.W., Deng, J., Foley, A.M.: Hybrid probabilistic wind power forecasting using temporally local gaussian process. *IEEE Trans. Sustain. Energy* 7(1), 87–95 (2016)
98. Nowotarski, J., Weron, R.: Computing electricity spot price prediction intervals using quantile regression and forecast averaging. *Comput. Stat.* 30(3), 791–803 (2015)
99. Maciejowska, K., Nowotarski, J., Weron, R.: Probabilistic forecasting of electricity spot prices using factor quantile regression averaging. *Int. J. Forecast.* 32(3), 957–965 (2016)
100. Weron, R.: Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecasting* 30(4), 1030–1081 (2014)
101. Bessa, R.J., Miranda, V., Botterud, A., Wang, J., Constantinescu, E.M.: Time adaptive conditional kernel density estimation for wind power forecasting. *IEEE Trans. Sustain. Energy* 3(4), 660–669 (2012)
102. Pinson, P., Madsen, H.: Ensemble-based probabilistic forecasting at horns rev. *Wind Energy* 12(2), 137–155 (2009)
103. Khosravi, A., Nahavandi, S.: Combined nonparametric prediction intervals for wind power generation. *IEEE Trans. Sustainable Energy* 4(4), 849–856 (2013)
104. Taieb, S.B., Husler, R., Hyndman, R.J., Genton, M.G.: Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. *IEEE Trans. Smart Grid* 7(5), 2448–2455 (2016)
105. Wang, Y., Zhang, N., Tan, Y., Hong, T., Kirschen, D.S., Kang, C.: Combining probabilistic load forecasts. *IEEE Trans. Smart Grid* 10(4), 3664–3674 (2018)
106. Aprillia, H., Yang, H.T., Huang, C.M.: Statistical load forecasting using optimal quantile regression random forest and risk assessment index. *IEEE Trans. Smart Grid* (2020)
107. Uniejewski, B., Weron, R.: Regularized quantile regression averaging for probabilistic electricity price forecasting. *Energy Economics* 95, 105121 (2021)
108. Liu, B., Nowotarski, J., Hong, T., Weron, R.: Probabilistic load forecasting via quantile regression averaging on sister forecasts. *IEEE Trans. Smart Grid* 8(2), 730–737 (2017)
109. Chen, L., Dolado, J.J., Gonzalo, J.: Quantile factor models. *Econometrica* 89(2), 875–910 (2021)
110. Bunn, D., Andresen, A., Chen, D., Westgaard, S.: Analysis and forecasting of electricity price risks with quantile factor models. *Energy J.* 37, 101–122 (2016)
111. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: International Conference on Machine Learning, pp. 1050–1059. (2016)
112. He, Y., Xu, Q., Wan, J., Yang, S.: Short-term power load probability density forecasting based on quantile regression neural network and triangle kernel function. *Energy* 114, 498–512 (2016)
113. Yu, Y., Yang, M., Han, X., Zhang, Y., Ye, P.: A regional wind power probabilistic forecast method based on deep quantile regression. *IEEE Trans. Ind. Appl.* 57(5), 4420–4427 (2021)
114. Duan, Y.: A novel interval energy-forecasting method for sustainable building management based on deep learning. *Sustainability* 14(14), (2022)
115. AlGabalawy, M., Hosny, N.S., Adly, A.R.: Probabilistic forecasting for energy time series considering uncertainties based on deep learning algorithms. *Electr. Power Syst. Res.* 196, 107216 (2021)
116. Cao, D., Zhao, J., Hu, W., Zhang, Y., Liao, Q., Chen, Z., et al.: Robust deep gaussian process-based probabilistic electrical load forecasting against anomalous events. *IEEE Trans. Ind. Inf.* 18(2), 1142–1153 (2022)
117. Hajirahimi, Z., Khashei, M.: Hybrid structures in time series modeling and forecasting: A review. *Eng. Appl. Artif. Intell.* 86, 83–106 (2019)
118. Wu, Y.K., Chen, C.R., Abdul.Rahman, H.: A novel hybrid model for short-term forecasting in pv power generation. *Int. J. Photoenergy* 2014, 569249 (2014)
119. Triebe, O., Laptev, N., Rajagopal, R.: Ar-net: A simple auto-regressive neural network for time-series. arXiv preprint arXiv:191112436 (2019)

120. Benmouiza, K., Cheknane, A.: Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models. *Energy Convers. Manage.* 75, 561–569 (2013)
121. Liu, H., Tian, H.q., Pan, D.f., Li, Y.f.: Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks. *Appl. Energy* 107, 191–208 (2013)
122. Kazemzadeh, M.R., Amjad, A., Amraee, T.: A hybrid data mining driven algorithm for long term electric peak load and energy demand forecasting. *Energy* 204, 117948 (2020)
123. Zhu, Z., Zhou, D., Fan, Z.: Short term forecast of wind power generation based on svm with pattern matching. In: 2016 IEEE International Energy Conference (ENERGYCON), pp. 1–6. IEEE, Piscataway (2016)
124. Wang, Y., Chen, Q., Kang, C., Xia, Q., Luo, M.: Sparse and redundant representation-based smart meter data compression and pattern extraction. *IEEE Trans. Power Syst.* 32(3), 2142–2151 (2016)
125. Chen, K., Hu, J., He, J.: A framework for automatically extracting over-voltage features based on sparse autoencoder. *IEEE Trans. Smart Grid* 9(2), 594–604 (2016)
126. Zhang, C., Bütepage, J., Kjellström, H., Mandt, S.: Advances in variational inference. *IEEE Trans. Pattern Anal. Mach. Intell.* 41(8), 2008–2026 (2018)
127. Ibrahim, M.R., Haworth, J., Lipani, A., Aslam, N., Cheng, T., Christie, N.: Variational-lstm autoencoder to forecast the spread of coronavirus across the globe. *PLoS One* 16(1), e0246120 (2021)
128. Pereira, J., Silveira, M.: Unsupervised anomaly detection in energy time series data using variational recurrent autoencoders with attention. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 1275–1282. IEEE, Piscataway (2018)
129. Dairi, A., Harrou, F., Sun, Y., Khadraoui, S.: Short-term forecasting of photovoltaic solar power production using variational auto-encoder driven deep learning approach. *Appl. Sci.* 10(23), 8400 (2020). <https://www.mdpi.com/2076-3417/10/23/8400>
130. Bachhav, P., Todisco, M., Evans, N.: Latent representation learning for artificial bandwidth extension using a conditional variational auto-encoder. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7010–7014. IEEE, Piscataway (2019)
131. Wang, X., Cui, P., Du, Y., Yang, Y.: Variational autoencoder based fault detection and location method for power distribution network. In: 2020 8th International Conference on Condition Monitoring and Diagnosis (CMD), pp. 282–285. IEEE, Piscataway (2020)
132. Kaur, D., Islam, S.N., Mahmud, M.A.: A variational autoencoder-based dimensionality reduction technique for generation forecasting in cyber-physical smart grids. In: 2021 IEEE International Conference on Communications Workshops. IEEE, Piscataway (2021)
133. Fathi, S., Srinivasan, R., Fenner, A., Fathi, S.: Machine learning applications in urban building energy performance forecasting: A systematic review. *Renewable Sustainable Energy Rev.* 133, 110287 (2020)
134. Xu, W.t., Wang, Y.l., Li, T., Tang, Q., Zhang, J.f., Shen, L., et al.: Long term intelligent load forecasting method considering the expectation of power market transaction. In: 2017 29th Chinese Control And Decision Conference (CCDC), pp. 2310–2315. IEEE, Piscataway (2017)
135. Percy, S.D., Aldeen, M., Berry, A.: Residential demand forecasting with solar-battery systems: A survey-less approach. *IEEE Trans. Sustainable Energy* 9(4), 1499–1507 (2018)
136. Fang, X., Yuan, Z.: Performance enhancing techniques for deep learning models in time series forecasting. *Eng. Appl. Artif. Intell.* 85, 533–542 (2019)
137. Wang, X., Zhao, Y., Pourpanah, F.: Recent advances in deep learning. *Int. J. Machine Learn. Cybern.* 11, 747–750 (2020)
138. Zhang, W., Quan, H., Srinivasan, D.: An improved quantile regression neural network for probabilistic load forecasting. *IEEE Trans. Smart Grid* 10(4), 4425–4434 (2018)
139. Zhang, Y., Xu, Y., Dong, Z.Y.: Robust ensemble data analytics for incomplete pmu measurements-based power system stability assessment. *IEEE Trans. Power Syst.* 33(1), 1124–1126 (2017)
140. Zhang, R., Dong, Z.Y., Xu, Y., Meng, K., Wong, K.P.: Short-term load forecasting of australian national electricity market by an ensemble model of extreme learning machine. *IET Gener. Transm. Distrib.* 7(4), 391–397 (2013)
141. Nagaraja, Y., Devaraju, T., Kumar, M.V., Madichetty, S.: A survey on wind energy, load and price forecasting:(forecasting methods). In: 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp. 783–788. IEEE, (2016)
142. Kabir, H.D., Khosravi, A., Hosen, M.A., Nahavandi, S.: Neural network-based uncertainty quantification: A survey of methodologies and applications. *IEEE Access* 6, 36218–36234 (2018)
143. Quan, H., Khosravi, A., Yang, D., Srinivasan, D.: A survey of computational intelligence techniques for wind power uncertainty quantification in smart grids. *IEEE Trans. Neural Networks Learn. Syst.* 31(11), 4582–4599 (2019)
144. Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., Zareipour, H.: Energy forecasting: A review and outlook. *IEEE Open Access Journal of Power and Energy* 7, 376–388 (2020)
145. Aslam, S., Herodotou, H., Mohsin, S.M., Javaid, N., Ashraf, N., Aslam, S.: A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renewable Sustainable Energy Rev.* 144, 110992 (2021)
146. Ahmad, T., Madonski, R., Zhang, D., Huang, C., Mujeeb, A.: Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: Key developments, challenges, and future research opportunities in the context of smart grid paradigm. *Renewable Sustainable Energy Rev.* 160, 112128 (2022)
147. Available from: <https://www.kaggle.com/robikscube/hourly-energy-consumption>
148. Available from: <https://data.gov.au/data/dataset/electricity-consumption-benchmarks>
149. Brier, G.W.: Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* 78(1), 1–3 (1950)
150. Available from: <https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Solar-home-electricity-data>
151. Available from: <https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset>

**How to cite this article:** Kaur, D., Islam, S.N., Mahmud, M.A., Haque, M.E., Dong, Z.Y.: Energy forecasting in smart grid systems: recent advancements in probabilistic deep learning. *IET Gener. Transm. Distrib.* 16, 4461–4479 (2022).

<https://doi.org/10.1049/gtd2.12603>