IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 June 2023, accepted 4 August 2023, date of publication 28 August 2023, date of current version 8 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3308928

TOPICAL REVIEW

Electric Vehicle Energy Demand Prediction Techniques: An In-Depth and Critical Systematic Review

FATEMEH MARZBANI[1], AHMED H. OSMAN[2], (Senior Member, IEEE), AND MOHAMED S. HASSAN[2]
[1]Department of Industrial Engineering, American University of Sharjah, Sharjah, United Arab Emirates
[2]Department of Electrical Engineering, American University of Sharjah, Sharjah, United Arab Emirates

Corresponding author: Mohamed S. Hassan (mshassan@aus.edu)

This work was supported in part by the Open Access Program from the American University of Sharjah.

ABSTRACT Accurate forecast of electric vehicle energy demand is vital for maintaining the stability and reliability of power systems. With the increasing prevalence of electric vehicles in transportation systems, the anticipating demand surges with precision in terms of timing and location is becoming ever more critical for utilities to guarantee sufficient supply. The intermittent and stochastic nature of electric vehicle electricity consumption is a significant challenge in accurately forecasting of electric vehicles demand. As a result, there is a growing field of research focused on developing models that can effectively capture and interpret such complex data. In improving the potential of accurate prediction models, conducting a comprehensive review of literature, examining current research overviews, and exploring potential expansions and extensions of models are all critical components. In this review, a comprehensive overview of prior research conducted for forecasting electric vehicle energy demand is presented, including a detailed examination of the benefits and drawbacks of the methods used. Additionally, potential gaps in the field are identified, and recommendations for future research directions are provided.

INDEX TERMS Electric vehicle charging demand, forecast, stochatic models, machine learning, nonlinear modeling, time series.

I. INTRODUCTION

Extensive consumption of fossil fuels has resulted in global warming during the past decades. As such, the Paris Agreement forged in 2015 defines the goal of a maximum two-centigrade temperature rise by the end of the century [1]. Electricity generation and transportation systems are major carbon-intensive sectors [2] and transportation systems' energy consumption exceeds a quarter of global energy utilization [3].

In comparison with conventional vehicles that are equipped with Internal Combustion Engines (ICEs), the employment of Electric Vehicles (EVs) contributes to 45% less carbon emission [4]. Hence, the goal of a low-carbon energy future and reducing greenhouse gas emissions can be achieved by the electrification of the transportation system and transforming conventional transportation systems into smart ones. Moreover, implementing the concept of smart cities and smart transportation systems will make traditional networks and services more efficient.

Within the smart cities context, smart transportation, as one of the main pillars, has recently entered the development stage after completing the conceptual step successfully. In this regard, traditional transportation systems have been recently introduced to increase the electrification of vehicles in order to comply with the previously mentioned emission standards and to be able to achieve smart cities' objectives.

Hence, rapid growth in the number of EVs has been observed since 2015. There has been a significant surge in the number of electric vehicles (EVs) on the road, with the

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

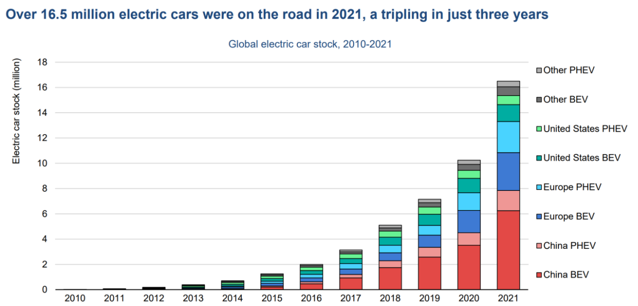

FIGURE 1. Incremental trend of available EVs on transportation networks [5].

count of EVs tripling between 2017 and 2021. Figure 1 illustrates the aforementioned increase. This number reached about 16.5 million at the end of 2021 and is expected to increase to 145 million in 2030 [5]. This anticipated large-scale integration of EVs into power systems induces both troublesome problems and opportunities for power systems management.

While EV integration into the power grid has environmental and economic benefits, it may cause many problems to the existing electric power system. Power quality and reliability issues, serious bottlenecks in the distribution network, and increased peak load demand are the most important adverse impacts of such deployment [6], [7], [8], [9]. Moreover, accurate EVEC predictions possess the capability to decrease carbon emissions considerably. Through the optimization of charging schedules, the incorporation of EVs into the electricity network is capable of reducing the overall environmental impacts. This task is achieved by promoting the use of more environmentally friendly energy sources and minimizing off-peak hours. From a wider perspective, it is of great importance to address such implications in order to achieve a more environmentally mindful transportation environment.

In this regard, a precise forecast of EV Energy Consumption (EVEC) will provide the power system operators with effective decision-making abilities for energy management of smart cities. Particularly, controlling the network of EV charging stations in a smart manner mandates an accurate forecast of EVEC [10]. Furthermore, it can contribute to enhancing the operation of charging stations and reducing maintenance costs. Additionally, the design, planning, and expansion of EV charging infrastructure are heavily dependent on the output of the EVEC forecast [11]. Therefore, it is of utmost importance to consider EVs energy demand modeling and prediction in energy system frameworks and scheduling.

Despite the fact that electric vehicles are a relatively new concept, numerous research studies have been conducted in the literature to predict their performance. These studies employ diverse methods, making comparisons complicated. To address this issue, the present paper organizes the varied approaches used for EVEC modeling through a classification framework that evaluates each approach's strengths and weaknesses in relation to the impact study's scope.

Although there are various review papers in the literature regarding EVEC prediction, they tend to focus on specific groups of data modeling techniques and fail to offer a comprehensive overview of all the available techniques from different algorithms. In contrast, the research work in this paper provides a comprehensive review of state-of-the-art techniques used for EVEC modeling and prediction. The key contributions of this survey include:

1) The present research synthesize the existing literature in order to provide the researchers a useful starting point for any future research work in the field of EVEC prediction.
2) A critical analysis of the existing literature is provided in the present review that assesses the performance, advantages, and drawbacks of the algorithms used in prior research studies.
3) It establishes the state of art by providing an up to date and novel classification of the current state of art in the field of EVEC prediction, evaluation metrics, and utilized algorithms structure.
4) It also highlight gaps in the existing literature and identify areas that need further investigation.

The organization of the paper is as follows. Section II presents a review of different types of EVs. In Section III, a classification for the available prediction techniques for energy demand of EVs is provided. Section IV reviews the linear models employed in the literature, while Section V and VI discuss the nonlinear and hybrid models, respectively. Section VII briefly discusses the evaluation criteria for predictive models. Section VIII presents a detailed discussion, and finally, Section IX concludes the paper and suggests future research directions.

II. CATEGORIES OF EVs AND MODELING OF BATTERY PARAMETERS

The market for electric vehicles is undergoing rapid evolution. Electric vehicles are available in both fully electric and hybrid models that incorporate both an electric motor and an internal combustion engine. This section provides a brief overview of the categories and EVs' battery parameter modeling.

There exist three main types of EVs: Hybrid Electric Vehicles (HEVs), Plug-in Hybrid Electric Vehicles (PHEVs), and Battery Electric Vehicles (BEVs).HEVs are powered by two different electric source, Internal Combustion Engine (ICE) and electric motor. This method enables enhanced fuel efficiency in comparison with ICEs.

Unlike HEVs, BEVs are predominantly powered by a single source and are comprised of three main technical components: electric motor, power controller, and rechargeable battery pack. The third class of EVs is PHEV. They employ batteries to operate an electric motor and are capable of recharging from an external power source [12], [13]. As reported by [5], the most commonly used type of EV across various markets is the BEV, as depicted in Fig. 2.

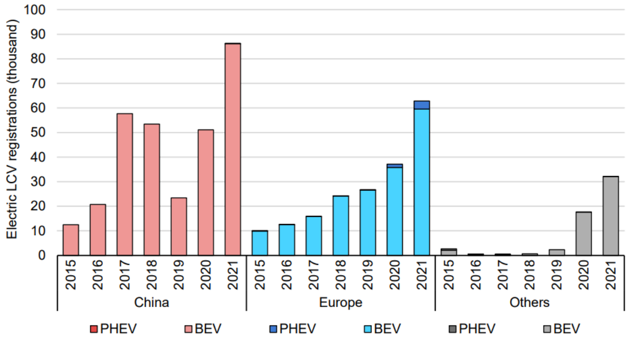

FIGURE 2. Market share of EV types 2015-2021 [5].

The two primary battery technologies dominating the market are lead-acid and Li-ion batteries. The battery capacities of various EV types conform to either a normal or Gamma distribution. These distributions are illustrated by (1) and (2) [14].

$$f(C;\alpha;\beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} C^{\alpha-1} e^{-\frac{C}{\beta}}, \tag{1}$$

$$g(C;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(C-\mu)^2}{2\sigma^2}}. \tag{2}$$

The parameters μ and σ represent the parameters of the normal distribution, while α and β represent the parameters of the gamma distribution. One can calculate the initial charge of an EV using the equations above.

III. EVEC PREDICTION MODELS

Forecasting EVs' charging demand is complex since it involves various uncertainties. These intermittencies stem from the random characteristics of EVs' driver behaviors. Accordingly, for seamless integration of EVs into the electricity network, it is vital to consider such uncertainties imposed on the grid in system modeling and management. In this respect, various research studies have been conducted in the literature to tackle the problem of providing precise EVEC prediction in power networks.

In this study, a literature review is conducted in order to provide foundation knowledge of this research topic. Accordingly, this section will look at previously developed data-driven techniques for EVEC forecast and will discuss their corresponding drawbacks and highlight the contribution of the present research.

Prior work in this area can be divided into three main categories based on the utilized methods: linear, nonlinear, and hybrid models Fig.1. The following sections offer an overview of these classifications and their corresponding algorithms. Additionally, the literature's available research studies are also examined.

IV. LINEAR MODELS

The initial classification of EVEC prediction models includes linear techniques that employ linear functions to model and forecast time series. Autoregressive Moving Average (ARMA) (and its variants), Multiple Linea Regression (MLR), Bayesian inference models, and Principal Component Analysis (PCA) are among the most frequently addressed linear techniques for the purpose of EVEC prediction.The subsequent subsections cover the discussion of these algorithms.

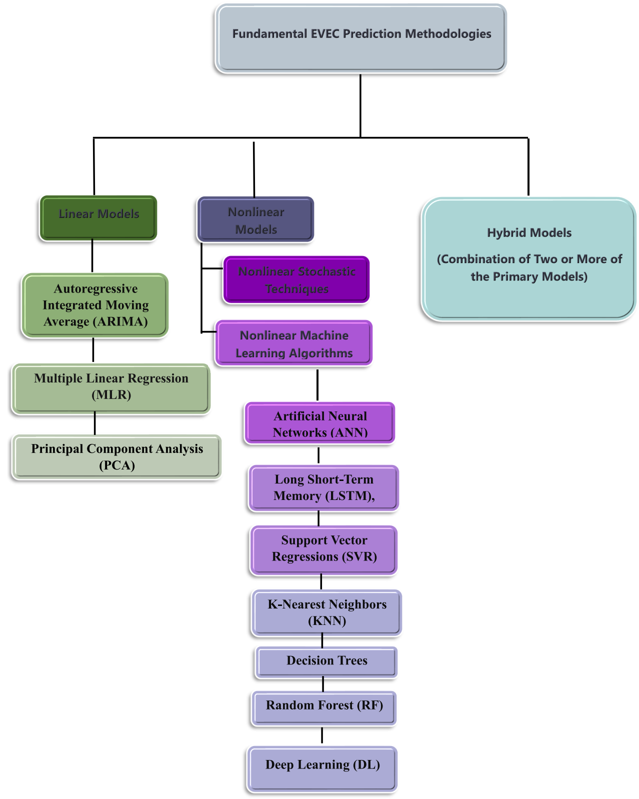

FIGURE 3. Fundamental EVEC prediction approaches.

A. AUTOREGRESSIVE MOVING AVERAGE (ARMA)

ARIMA-based techniques are founded on time series analysis and the seasonality available in the data set. These algorithms utilize the historical record of the data set to obtain future values [15]. The ARIMA mathematical formula is demonstrated in (3).

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}, \tag{3}$$

where y_t denotes the time series data at time t, ε_t is the error term at time t, ϕ_i for $i = 1, \ldots, p$ and θ_j for $j = 1, \ldots, q$ are the parameters of the auto-regressive and moving average parts of the model, respectively. Moreover, μ represents a constant.

The research carried out in [16], implements the ordinary Auto Regressive Integrated Moving Average (ARIMA) approach to provide day-ahead aggregated electricity charging demand of EV Supply Equipment (EVSEs). This model is applied to real-world data obtained from nonresidential EVSEs equipment in California. It is affirmed that larger aggregation leads to smaller prediction errors.

Similarly, in [17], the charging demand of EV parking lots is predicted using ARIMA models. This charging demand is estimated based on the driving distance and drivers' behavior. Additionally, the optimum parameters of the ARMA model are obtained using mean square error. Seasonal ARIMA (SARIMA), a more sophisticated ARIMA-based algorithm, is utilized in [18] in order to predict the electricity demand of EV charging stations. The under-investigated case study in this work is comprised of aggregated electricity consumptions of more than 2400 stations and the most efficient SARIMA parameters for both short-term and long-term forecasts are identified. The research study in [19] includes self-similarity of the EVs' demand time series by employing Fractional ARIMA (FARIMA) for short-term prediction. The results affirm the out-performance of FARIMA when compared with those obtained from ordinary ARIMA.

B. MULTIPLE LINEAR REGRESSION (MLR)

MLR refers to a group of mathematical frameworks that takes into account the empirical relationship between two or more variables. These models develop a linear relationship between independent and dependent variables [20] and are demonstrated in (4). These models have been used in several studies to assess EVs' demand prediction using various factors such as driver's behavior, battery state of charge, power terrain efficiency, and microscopic driving parameters [21], [22], [23], [24].

Article [25] performs an MLR on a real-world data set recorded on ordinary travel routes, where vehicle speed, acceleration, energy consumption rate, and battery state of charge are used as input variables. The results demonstrate the efficiency of MLR in comparison with the persistent method. Similar to the work conducted in [25], study [22] implements MLR for the purpose of predicting EVs' energy consumption. The forecasted values are used to identify an optimum scheduling framework for EV charging. Below equation describes the mathematical formula for MLR.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \epsilon, \tag{4}$$

where y is the dependent variable, $x_1, x_2, \ldots, x_p$ are the independent variables,$\beta_0, \beta_1, \ldots, \beta_p$ are the coefficients of the independent variables, and ε is the error term.

C. PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA is a statistical dimension-reduction technique and is the basis for multivariate data analysis. An orthogonal transformation is used in this method to transform the dependent variables of a data set into a completely new set of uncorrelated variables [26]. The study conducted in [27] examines the applicability of PCA to forecast EVs' charging demand of a real-world data set recorded in Kansas City for the purpose of EV infrastructure decision-making strategies. The proposed model is capable of dealing with time series of different dimensions and provides future required parking regulation by analyzing EVs' charging and parking duration.

In [28], the daily consumption of secondary EV substations is examined using a PCA model. The original data is transformed into three new orthogonal variables representing 96% of the total variance of the substations data. These PCA components are used to classify substations' data sets to standard demand profiles. While the research presented in [28] employs PCA for labeling EVs' demand profile, article [29] adopts the output of PCA model to establish a polynomial of PCA components to estimate future values of EVs' energy consumption. For this purpose, the model coefficients are set such that these weights demonstrate the importance of the corresponding variable relative to the output of the model.

Linear prediction models have advantages such as being easy to implement, having low computational complexity, and being easily interpretable. However, these models are susceptible to over fitting and are sensitive to noise. Furthermore, linear prediction models rely on the assumption of a linear relationship between dependent and independent variables, which is not always the case in real-world data. Therefore, it is crucial to develop methods that can handle the uncertainties and nonlinearity present in real-world data. Nonlinear models used for EVEC prediction are discussed in the following section.

V. NONLINEAR MODELING ALGORITHMS

The second category of EVs' load prediction methodologies is nonlinear algorithms. Although linear methods assume a linear relationship between the variables in a time series, nonlinear methods utilize nonlinear functions to map input and output data. Complex time series with nonlinear features can be effectively analyzed using these techniques, however, they may require larger input data sets compared to linear models. There are two categories into which nonlinear models can be classified: stochastic models and machine learning algorithms.

A. STOCHASTIC NONLINEAR TECHNIQUES

Nonlinear stochastic forecast and modeling methods are capable of analyzing time series data that displays nonlinear behavior. This approach involves relaxing the assumption of linearity and employing more complex models to capture the nonlinear relationships among variables. Stochastic models make an assumption that time series data is produced by a stochastic process in which the values of the variable at each time point are random variables. Nonlinear stochastic time series approaches integrate nonlinear functions into the stochastic process, enabling greater flexibility in the modeling of the data. Therefore, numerous research studies have focused on predicting EVEC using nonlinear stochastic techniques. This section provides an analysis of these research studies.

In the article [30], the probability density function methodology is utilized to model the energy demand of residential EVs. The framework proposed in this study is based on a DL structure that involves four phases, namely convolutional layers, GRU, autoregressive model, and Kernel density estimator. The study includes a case analysis of

348 residential EVs in the eastern USA, and the model is applied to this data to forecast EVEC for three different time horizons (10, 15, and 30 minutes ahead). The results obtained from the study confirm the feasibility of the proposed model.

The EV load profile is forecasted in [31] using a stochastic technique built on conditional probability distribution. In this study, the model parameters were derived from experimental data sets. The effectiveness of the approach was evaluated by utilizing EV data obtained from surveys. In addition, K-NN is employed to cluster EV usage patterns into three categories. The final variables included in the model are SOC, charging time, weekend/weekday, and EV usage mode. The case study involves EV charging data from both single and multiple regions in the UK. The results obtained demonstrate the efficacy of the model in terms of prediction accuracy.

The research outlined in [32] applies Modified Pattern-based Sequence Forecasting (MPSF) to anticipate EV charging demand 24 hours ahead. The technique is implemented on a real data set collected at UCLA campus, and its performance is compared against three commonly used prediction methods, namely SVR, K-NN, and RF. The findings suggest that MPSF surpasses the benchmark methods regarding forecast accuracy, whereas K-NN proves to be the swiftest algorithm.

Nonlinear stochastic techniques are able to account for the nonlinear relationship between independent and dependent variables, resulting in more accurate predictions compared to linear models. However, these models are demanding significant computational resources, challenging to implement and interpret, and may be susceptible to over fitting. Additionally, the effectiveness of the model can be impacted by the selection of model parameters. While nonlinear stochastic models can overcome the limitations of linear models, their main drawback lies in the selection of model parameters. However, with the increasing availability of data and advancements in computing power, machine learning algorithms have emerged as a preferred alternative. These factors have been key drivers in the shift towards the adoption of machine learning techniques for time series prediction. The upcoming subsection will discuss the application of nonlinear machine learning techniques for forecasting EV demand.

B. NONLINEAR MACHINE LEARNING TECHNIQUES

The abundance of historical data records in smart city environments, coupled with the nonlinear characteristics of EVs' energy consumption, provides a basis for the application of nonlinear techniques in solving the issue of EV demand forecasting [33]. K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), Support Vector Regression (SVR), and Random Forest (RF)are among the most widely utilized machine learning models for forecasting EV energy consumption. This section includes a discussion of these methods as they apply to predict EVs' energy consumption.

1) ARTIFICIAL NEURAL NETWORKS (ANNs)

ANNs are computational-based models that are inspired by the structure of neuron connections in the human brain. These models are designed to replicate the functioning of neurons in the human brain. ANNs are proficient at identifying and capturing non-linear relationships between input and output variables through the use of non-linear approximations. Each neural network is comprised of three layers: the input, the hidden, and the output layer.The layerd structure of the artificial neural network (ANN) is depicted in Fig. 4. In this model, each input is associated with a specific weight and several activation functions are utilized to map the input variables to the output layer [34]. Literature reports on several research studies that examine the efficiency of ANNs for predicting EVs' charging demand [35], [36], [37].

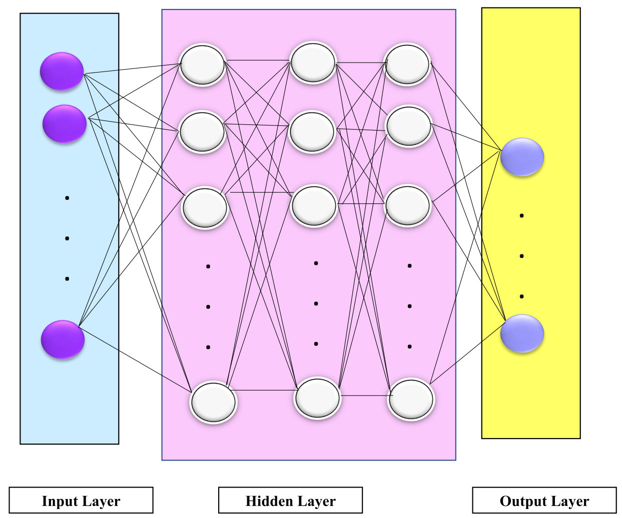

FIGURE 4. Layered structure of ANN.

In [38], ANN model is used to provide super short-term EVs' charging demand prediction. The network input includes electricity charging demand, charging session start time, and charging session end time. Root Mean Square Error (MSE) and Mean Absolute Error (MAE) are used as evaluation metrics to assess the efficiency of ANNs. The modeling technique utilized in the work carried out in [39] involves the use of three distinct variations of ANNs: conventional ANN, rough ANN, and recurrent rough ANN. Moreover, the correlation between trip duration, arrival, and departure time is considered in the modeling procedure. The results obtained indicate that the recurrent ANN exceeds the alternative methods that were considered in the study.

LSTM is a type of Recurrent Neural Network (RNN) and its origin can be traced back to 1997 [40]. Classic RNNs models suffer from a problem where the network can preserve information that should be forgotten by the network or forget information that should be remembered due to the imperfect functioning of neurons. Hence, LSTM was developed to tackle this issue by means of a structure that includes a memory cell and introduces an intelligently controlled self-circulating cycle that creates a pathway for the gradient to flow for extended periods of time [41].An LSTM cell is composed of several key elements, including a forget gate, an input gate, an output gate, and a cell state Fig. 5. In the

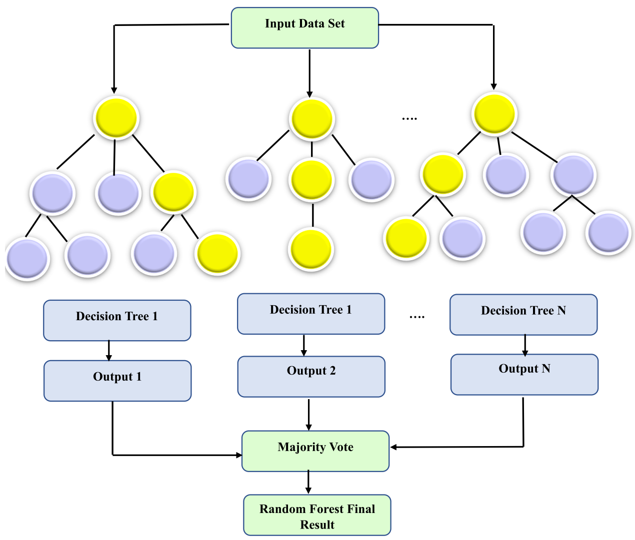

FIGURE 5. **Majority vote structure of random forest algorithm.**

literature, there exist multiple research studies that utilize LSTM for EV charging prediction [37], [42]. Following the approach taken in [42], study [36], makes use of the LSTM technique to forecast the demand for EVs in multiple time scales, specifically 15 and 30 minutes ahead. The results demonstrate the out-performance of LSTM in comparison with ANN in terms of prediction accuracy.

Models based on ANN are renowned for their capacity to capture non-linear relationships between input variables, as well as their ability to manage noisy or incomplete input data. Their primary strengths include their capacity to learn from input data and adapt the resulting model to new and unseen data through parameter fine-tuning.

One of the significant limitations of ANNs-based models is their high data requirements for the training process, which can be both time-consuming and computationally expensive. Furthermore, the effectiveness of these models is heavily influenced by the selection of activation function, hidden node count, and number of hidden layers. Additionally, ANNs are challenging to interpret due to their complex internal computations. To address the limitations mentioned above, more advanced techniques are employed for the prediction of EVEC.One such technique, as described in the following subsection, is SVR.

2) SUPPORT VECTOR REGRESSION (SVR)

SVR is an extension of the Support Vector Machine (SVM) algorithm in machine learning and It originated from statistical learning theory in the 1990s [43]. The objective of SVR is to minimize the error acquired by fitting the model to data using a hyperplane that is positioned in close proximity to each point. This is done to guarantee that all points remain within a predetermined distance from the hyperplane. SVR is employed to forecast a value using a set of input samples through the use of a kernel function and Lagrange multipliers. The anticipated value is computed by adding a bias term to a linear combination of the Lagrange multipliers and the kernel function of the input samples. The SVR equation is illustrated by

$$\hat{y} = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x_{train}, x_{test}) + b. \tag{5}$$

The variables used in Equation 1 for SVR are as follows: $\hat{y}$ represents the predicted value, α_i and α_i^* are Lagrange multipliers, $K(x_{train}, x_{test})$ denotes the kernel function, x_i represents the training samples, x_{test} denotes the input sample, and b represents the bias term.

Differing from other machine learning algorithms, SVR employs structural a nonlinear algorithm constructed upon the foundation of structural risk minimization. This empowers SVR to achieve exceptional generalization ability. SVR has been utilized in multiple research studies to forecast EV demand, and some of these studies are discussed in the following paragraph.

The effectiveness of SVR for predicting EV demand is examined by the authors in [44]. The study employs a real-time data set recorded from a charging station between 2011 and 2014. Normalized root mean square error is used as an evaluation metric, and SVR is applied to the data set. The results indicate that SVR exhibits high prediction accuracy, thereby highlighting its efficiency.

Similar to the results reported in [44], papers [45], [46], [47] also showcase the outstanding capabilities of SVR in predicting EV load. The study presented in reference [45] utilizes the Support Vector Regression (SVR) algorithm to analyze EV charging demand, taking into account various influential factors such as meteorological conditions, number of EVs, festival periods, and weeks. By detecting and correcting any flawed data points in the historical data set, this model achieves improved prediction accuracy. To evaluate the feasibility of the proposed model, it was applied to a historical data set of charging stations located in Shandong, China. Comparisons were made between the results obtained using the SVR approach and those obtained through a traditional forecasting method based on EV usage patterns. The findings affirm the feasibility and effectiveness of the SVR model.

Another SVR approach is presented in [46], where the day-ahead electricity consumption of a charging station in Indonesia is modeled and forecasted incorporating both historical charging transactions and weather data. The findings suggest that SVR surpasses other machine learning algorithms in accuracy. The work in [47] introduces a refined variant of SVR called Coarse Gaussian SVR (CG-SVR) which is utilized to analyze EV traffic in Agartala, India. The proposed CG-SVR model yields more precise results and requires less training time when compared to other machine learning techniques.

Similar to ANN-based approaches, although SVR models are capable of capturing nonlinearity in input data, they require model parameter selection and their performace is heavily dependant on the quality of input data. Unlike ANN algorithms,SVR models are easier to interpret due to the implementation of a small set of support vectors. Furthermore, SVR models require a smaller amount of training data compared to ANNs.

3) K-NEAREST NEIGHBOR (K-NN)

K-NN is a machine learning algorithm that does not require parameter estimation. This algorithm can be used to perform both classification and regression tasks. The algorithm computes the distance between data points using distance measures such as Euclidean distance and predicts the output value for a given data point by averaging the values of its k-nearest neighbors [48], [49]. The overall framework of K-NN for regression can be outlined as follows:

1) Select the number of K nearest neighbors to be taken into account.
2) Compute the pairwise distance between the input sample and all the samples in the training data set.
3) Pick the K samples from the training set that have the smallest distances to the input sample.
4) Calculate the mean value of the K nearest neighbors that were determined, and assign it as the predicted value for the input sample.

The utilization of K-NN for rapid prediction of electric vehicle charging outlet demand is explored in [50]. The goal of this article is to provide an hour-ahead forecast for 20 electric vehicle charging stations at UCLA campus, based on data recorded from 2011 to 2014. Two different similarity measures are employed in the study, namely the Euclidean distance and the Time Weighted Dot Product (TWDP). The findings indicate that K-NN yields more precise predictions when TWDP is utilized in place of the Euclidean distance. The modeling approach presented in this research is utilized to develop a cellphone application for EV owners at UCLA.

In article [51], the evaluation of K-NN's performance was continued from previous work [50]. This was done by comparing K-NN with two other prediction algorithms, ARIMA and Pattern Sequence-based Forecasting (PSF). The data utilized for the case study consisted of EVs' data recorded at charging stations on the UCLA campus, which included transaction start and end time and total energy consumption. Notably, driver behavior and geographical data of the EVs were not utilized in the modeling process. The findings indicate that K-NN exhibited higher prediction accuracy compared to both ARIMA and PSF.

The study described in [52] employed k-NN to develop a hierarchical clustering algorithm for forecasting the energy consumption profiles of 15 different EVs in Beijing, China. The method involved assigning each data point as a separate cluster initially, calculating the pairwise distance between all these clusters, and subsequently merging the two closest clusters to minimize the sum of square within the clusters.

Unlike SVR, K-NN is a non-parametric approach that can be easily understood and used when only a small amount of data is available. However, the performance of these models is highly dependent on the choice of the distance parameter, K. On the other hand, while SVR models are efficient in dealing with input data of higher dimensions, K-NN models are not suitable for high-dimensional data and input data that includes missing values.

4) DEEP LEARNING (DL)

Deep Learning (DL) is a category of machine learning methods that can be traced back to 1940 [53]. However, it has only been in recent years that this algorithm has been applied successfully in various fields. This is primarily due to two factors: the recent emergence of powerful computers capable of training complex mathematical models and the availability and accessibility of more extensive data sets for DL training purposes [54], [55]. There has been extensive use of deep learning algorithms in the literature for predicting electric vehicle charging demand [38], [56], [57], [58], [59], [60], [61], [62]. The subsequent paragraphs delve into several research studies that utilize DL techniques.

The research carried out in [63], provides a one day ahead prediction of EVEC of charging stations in Caltech campus and Jet Propulsion Laboratory. The technique estimates the posterior distribution using variational inference and utilizes LSTM parameters for the prior distribution. To evaluate its effectiveness, the approach is benchmarked against SVR and MLR, with the results demonstrating that the proposed method outperforms the others.

A comparative study of DL techniques for ultrashort term (minute level) prediction of EV energy consumption (EVEC) was conducted by authors in [38].The study evaluated the feasibility of six machine learning algorithms, including ANN, LSTM, Recurrent Neural Network (RNN), Gated Recurrent Units (GRU), Stacked Auto-Encoders (SAEs), and the Bi-directional LSTM (Bi-LSTM), using historical EV data collected in Shenzhen, southern China, from July 1st, 2017 to June 30th, 2018. This data set includes electricity charging demand, charging session start time, and charging session end time. Root Mean Square Error (MSE) and Mean Absolute Error (MAE) are used as evaluation metrics to assess the efficiency of the previously mentioned methods.The results indicated that all six models were effective for super short term prediction, but LSTM performed better than the other five methods.

In [64], a DL-based forecasting method called Sequence to Sequence was developed for predicting EVEC up to one month and five months in advance. The effectiveness of this approach was tested on real-world data from 1200 charging stations in Los Angeles. In order to evaluate the performance of the Sequence-to-Sequence algorithm, conventional DL models such as ARIMA, LSTM, and XGBoost were used as benchmark models. The findings revealed that Sequence to Sequence outperformed the other models for both short- and long-term predictions, demonstrating significant improvements in accuracy.

Article [59] introduced a novel DL-based probabilistic framework for EVEC prediction. This framework is comprised of three steps. This framework consists of two main steps. Firstly, wavelet decomposition and normalization techniques were utilized to standardize the traffic flow data and divide it into different frequencies. Secondly, a Convolutional Neural Network (CNN) was employed to predict the traffic flow data, followed by a mixture model-based method to forecast the arrival rate of EVs. Finally, the predicted arrival rates were fed into a Queuing-based probabilistic model to obtain the EVEC forecast. The study was conducted using historical data collected from January to December 2014 in the U.K. The experimental results showed that the proposed model was effective in predicting EVEC.

Authors in [65] develop a novel DL-based approach for the purpose of EVEC forecast. The method decomposes the input time series into multiple sub time series using Empirical Mode Decomposition (EMD) and then applies Deep LSTM (DLSTM) to each sub time series to forecast EVEC values. The parameters of DLSTM are optimized using the Arithmetic Optimization Algorithm (AOA). The research employed data from EV charging stations in Georgia Tech, USA, as the case study. The effectiveness of the proposed model was compared with several conventional DL-based methods, and the results indicated that the proposed approach was superior to the traditional DL-based methods.

While the study in [65] proposed a novel EVEC forecast technique, the work carried out in [66] makes a comparison between four well established DL-based approaches (ANN, LSTM, GRUs, and RNNs) to predict future values of EVEC time series recorded in Morocco. The data set consisted of 2000 data points, collected during 1793 charging transactions at two three-phase charging locations (22KW and 11KW). The outcomes indicated that all four techniques were viable, with the single-layer GRU model demonstrating superior prediction performance.

DL-based techniques have gained popularity in the field of time series prediction due to their various advantages. Some of their strengths include their ability to effectively handle and process large amounts of complex and nonlinear data. They also possess automatic feature selection capabilities. However, there are some disadvantages associated with DL models, such as their difficulty in diagnosing and correcting errors, tendency to over fit, and limited interpretability.

5) DECISION TREE (DTs) AND RANDOM FOREST (RFs)

Decision trees (DTs) are a type of nonlinear machine learning method that can be used for classification and regression tasks [67], [68]. These algorithms derive their name from their tree-like structure, which consists of multiple nodes and branches This algorithm is designed to break down complex decisions into a series of simpler choices. There exist two types of nodes: decision nodes and leaf nodes. Decision nodes are those at which input data will be divided into several subsets implementing the most effective splits. Whereas a leaf node refers to a node where no further splits are possible. This process of decision-making at nodes continues until a decision node is transformed into a leaf node [69], [70]. One of the widely used algorithms for constructing decision trees can be summarized as follows:

1) Start with a root node that corresponds to the entire data set.
2) Identify the optimal attribute to split the data at the current node. This is often accomplished by computing an information gain or Gini impurity score for each attribute and selecting the one with the highest score as the splitting criterion.
3) Next, one branch is created for each possible value of the selected attribute, and the data is split accordingly.
4) Repeat steps 2 and 3 recursively for each child node until a stopping criterion is satisfied. The criterion can be defined by setting a maximum tree depth, a minimum number of instances per leaf, or other specified conditions

RFs can be seen as an extension of decision trees in which multiple decision trees are combined and applied to different subsets of the input data to perform a prediction task. Therefore, for regression tasks, the final output is calculated by averaging the predictions of all decision trees in the ensemble [71], [72]. The fundamental framework of RFs and the majority vote structure are depicted in Fig.3.

The literature indicates that both DT and RF algorithms have been utilized in multiple studies for predicting EVEC [68], [69]. This subsection provides an overview of research that has utilized these tools.

The work presented in [70] introduces a prediction framework for estimating the Electric Vehicle Energy Consumption (EVEC) of traffic data in South Korea for the year 2014. This framework consists of three phases of analysis. Initially, cluster analysis is applied to the case study to identify various traffic patterns. Subsequently, influential factors are identified using grey relation analysis. Finally, a decision tree is used in the prediction stage. The study considers several variables such as climate data, traffic data, and the state of charge (SOC) of the battery. The proposed technique is capable of forecasting EVEC for different scenarios, including weekdays, weekends, winter, and summer.

In [71], an RF-based model was employed to predict EVEC in Helsinki, Finland, with a case study focused on seven public e-bus charging sites. This approach splits input data into multiple training samples. Next, DT is applied to all the subsets and the predicted values corresponding to each subset is obtained. The output of each DT is then utilized by RF to determine the final prediction value using majority vote approach. The accuracy of RF was compared to that of SVR, and the results showed that RF provided more precise predictions. The study's findings will serve as a guide for bus electrification in Helsinki.

In a likewise manner RF is used in article [72] for distribution forecast of EVEC profile. This study takes into account the spatio-temporal coupling between EVs and charging stations for different types of available EVs in Shanghai. The prediction process involves two steps. First, an improved RF technique is introduced, where RF parameters are fine-tuned using the Harmony Memory (HM) algorithm. The model is applied to the case study to obtain predicted values. Next, a bottom-up method is used to consider individual driver charging behavior and the spatio-temporal distribution coupling between EVs and charging stations. The proposed approach's feasibility is confirmed compared to traditional RF, SVR, and Back Propagation Neural Network (BPNN).

Decision trees have several strength points, including their ability to handle both categorical and numerical variables, as well as input with missing values. They are also considered easy to interpret. However, DTs are prone to over fit and they can exhibit bias towards features that have different scales.

As RFs consist of multiple DTs, they share similar advantages to DTs, such as the capability of working with input data containing missing values and handling both categorical and numerical variables. However, RF models require significant computational power and may not be efficient when handling data sets that involve randomness.

VI. HYBRID APPROACHES

Hybrid modeling algorithms involve combining various modeling techniques to form a unified and cohesive approach, where each model contributes to a distinct characteristic of the data. The combination of various techniques in hybrid modeling makes it a potent tool for time series analysis, enabling it to leverage the advantages of multiple methods to attain more precise and robust predictions [73]. The literature contains several studies that have developed hybrid models for EVEC prediction by combining the primary techniques discussed in the previous section [36], [44], [74], [75], [76], [77], [78], [79], [80], [81], [82]. The paragraphs below provide a discussion of some of these studies.

In a recent study presented in [83], a hybrid model was developed for EVEC prediction by combining SARIMA and DL techniques. This algorithm utilizes SARIMA to capture the linear features and identify any seasonality in the data, while LSTM is applied to the residual of the data set to take into account nonlinear characteristic of the data. The output of both SARIMA and LSTM are then combined to provide predicted values for the input time series. The study evaluated the proposed technique through three experiments conducted on a real-world hourly data set of EVEC from a charging station in Spain during 2015-2016. The research findings were compared with other techniques such as SARIMA, LSTM, Extreme Learning Machine (ELM), SVR, SARIMA-ELM, and SARIMA-SVR. The results confirmed the proposed hybrid model's efficiency and performance superiority over the other methods.

In a similar manner, by combining the multi variable residual correction Grey Model (GM) and LSTM as a unified prediction model for EVEC, the authors of [84] were able to benefit from the advantages of both approaches. The GM was used to account for the impact of different factors on EVEC, including weather conditions, electricity prices, and the number of electric vehicles. The LSTM was then employed to accurately identify influential factors, minimizing prediction errors. The study focused on EVEC data recorded in China in 2017, and the results confirmed the effectiveness of the developed technique.

Article [85] presents a hybrid model that combines LSTM and Temporal Encoder-Decoder (TED) to improve the efficiency of EV charging station prediction. The encoder is responsible for reducing the input data dimensions to retain and select only the relevant information, while the temporal dependencies between the input variables are determined by LSTM. To evaluate the effectiveness of this method, two sets of EVEC from charging stations in China and the USA are used. Classical prediction models such as ARIMA, LSTM, and Historical Average (HA) are used as benchmarks to compare the performance of the hybrid model. The results indicate that the proposed algorithm performs better than traditional techniques.

The article [35] presents a novel forecasting approach that integrates artificial neural networks (ANNs), recurrent neural networks (RNNs), and Q-learning. The Q-learning model is constructed based on ANN and RNN, where it utilizes the output of both ANN and RNN as its input. The research investigated three distinct EV charging scenarios, namely coordinated, uncoordinated, and smart. The case study data set was generated using Keras software. Results indicate that the proposed hybrid model can improve the prediction accuracy of both ANN and RNN by 50%.

Hybrid models combine the advantages of various models, which leads to improved efficiency and makes them more accurate and reliable compared to individual methods. Additionally, they demonstrate better robustness to noise and outliers. There are also some disadvantages associated with hybrid models for time series prediction. One of the major disadvantages of these models is their computational cost since they utilize multiple technique. Another possible drawback is the challenge of selecting the most suitable combination of models and determining the appropriate weight for each model.

Table 1 illustrates the hybrid techniques proposed by various research studies along with the benchmark algorithms and case studies investigated.

VII. EVALUATION CRITERIA FOR PREDICTIVE MODELS

There are multiple metrics available to assess the effectiveness of time series predictions techniques [86]. This section describes the commonly used evaluation metrics in the papers discussed in the previous sections.

Mean Absolute Error (MAE) is the most commonly forecast evaluation measure used by research studies conducted in the field of EVEC prediction. It is based on the average of the obtained error and is demonstrated in below equation.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{Y}_i - Y_i|, \tag{6}$$

TABLE 1. Comparison of various models for EVEC forecasting: performance evaluation on benchmarks and case studies.

Article Number	Proposed Model	Benchmarks	Case Studies
[74]	GAN-RF	SVR,BPNN	Charging Station, Xian Yang (China) June 2019
[36]	LSTM-DL	LSTM, ANN	Charging Station, Shenzhen (China) 2017-2018
[75]	KL-VMD-IWOA-LSSVM	BPNN, WOA-LSSVM IWOA-LSSVM KL-VMD KL-VMD-LSSVM IWOA-LSSVM KL-VMD-IWOA-LSSVM	Charging Station, Shanghai (China)-2017
[76]	Hybrid Kernel Density Estimator (HKDE)	Gaussian and Diffusion-based KDE	EV Users at UCLA Campus
[88]	LSTM-QEM	LOCF,NOCB EM,k-NN	KEPCO 206-2018
[77]	GA-PSO-SVR	SVR-PSO, SVR-GA	Simulated Data
[79]	ESG	DT, RF, K-NN	Aichi Prefecture, Japan 2012-2013
[80]	Time Series-Machine Learning	DHR BSTS STLM	Korea Environment Corporation
[81]	Hybrid Niche Immunity Lion Algorithm and Convolutional Neural Network	SVM-CNN	Charging Station, Beijing 2016-2017
[94]	SARIMA-DL	SARIMA, LSTM SARIMA-ELM SARIMA-SVR	Charging Station in Spain 2015-2016
[82]	EMGM-LSTM	LSTM, EMGM	Unknown Location-2017
[83]	TED-LSTM	LSTM,LSTM-Encoder, Historical Average	Charging Stations in China-USA
[84]	Q-learning Technique ANN and RNN	Q-Learning, ANN, RNN	Simulated Data

where n is the number of samples, $\hat{Y}i$ is the predicted value for sample i, and Y_i is the true value for sample i.

Mean Absolute Percentage Error (MAPE) measures the percentage of the error. The equation for MAPE is given as

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{Y}_i - Y_i}{Y_i}\right|, \tag{7}$$

where n is the number of samples, Y_i is the true value for sample i, and $\hat{Y}i$ is the predicted value for sample i.

RMSE is a widely used performance measure in the literature for EVEC time series prediction. Unlike some other metrics, RMSE assign more weights to larger errors when compared with smaller ones. The mathematical equation for calculating RMSE is given by

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}, \tag{8}$$

where sample numbers is presented by n, $\hat{Y}i$ is the predicted value for the i_{th} sample, and Y_i is the true value for sample i.

The coefficient of determination, commonly known as R-squared, is another popular performance metric used for time series prediction. This metric refers to the goodness of the fit and measures the amount of variance in the dependent variable that can be explained by the independent variables. While in the case of MAE, MAPE, and RMSE, smaller values indicate more precise predictions, a perfect fit is indicated by an R^2 value of 1, while an R^2 value of 0 indicates no correlation between the variables. The following equation explains the concept of R^2:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}, \tag{9}$$

where n is the number of samples, Y_i is the true value for sample i, $\hat{Y}i$ is the predicted value for sample i, and $\bar{Y}$ is the mean of the true values.

The precision of a prediction technique can be determined using the Mean Estimation Deviation (MED), which measures the average absolute difference between the predicted values and the original data. The subsequent mathematical equation can be utilized to obtain the MED value:

TABLE 2. Summary of prediction evaluation measures used by hybrid models.

Article Number	RMSE	MAPE	MAE	R^2	MED
[74]	✓	✓			
[36]	✓		✓		
[75]	✓	✓	✓		
[76]		✓			✓
[88]		✓			
[44]	✓				
[77]	✓	✓	✓	✓	✓
[79]	✓	✓	✓	✓	✓
[80]	✓	✓	✓		
[81]	✓	✓	✓		
[82]		✓			
[83]	✓	✓	✓		
[84]	✓				

$$MED = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{10}$$

where original data is depicted by Y_i, $\hat{Y}i$ represents the predicted value, and n is the total number of observations. Table 2 provides a summary of the evaluation metrics used by papers discussed in Section VI.

VIII. DISCUSSION

The escalating demand for sustainable and efficient transportation, coupled with the increasing adoption of EVs has brought significant attention to the field of EV energy demand prediction in recent years. The precise forecast of EV energy demand is a vital aspect in planning efficient charging infrastructure, optimizing energy management systems, and ensuring the consistent functionality of electricity grids.

The present review paper is focused on the latest progressions in EVEC forecasting, with a particular emphasis on the utilization of linear, nonlinear, and hybrid modeling techniques. This section delves into the significant findings of our review and examines their potential implications for future research and practical applications in the realm of electric vehicle technology and energy management.

A variety of modeling and prediction algorithms have been proposed in the literature for EVEC forecast, ranging from straightforward linear models to more intricate hybrid and nonlinear models. Due to their simplicity and interpretability linear models, including ARMA, MLP, and PCA are frequently employed. These models operate on the assumption that there exists a linear correlation between the variables. However, this assumption could restrict their ability to capture non-linear relationships, which in turn could decrease their efficiency in terms of prediction precision. Nevertheless, recent studies and experiments indicate that by adding extra features and non-linear transformations, the accuracy of linear models for EV energy demand prediction can be enhanced.

Both stochastic and machine learning nonlinear models have become more prevalent in recent years owing to their capability to represent complicated non-linear connections between input and output variables. The fusion of machine learning, artificial intelligence, and advanced big data analysis techniques is restructuring EVEC prediction models. Appreciating their central significance is vital to grasp the dynamic terrain of EVEC prediction, enabling a thorough evaluation of the present state of the field. These models have demonstrated encouraging outcomes in precise EVEC prediction, particularly when trained on comprehensive and diverse data sets. However, they suffer from encountering over fitting and require meticulous tuning of hyperparameters in order to avoid any probable bias and variance problems.

To leverage the benefits of multiple approaches while overcoming their limitations, hybrid models that integrate two or more of the primary models have been suggested. When compared to individual methods, hybrid techniques have exhibited improved performance.For instance, these models can integrate linear and nonlinear approaches to capture both linear and nonlinear dependencies between input and output variables. In a similar manner, a hybrid method can merge a stochastic and machine learning models to provide future values of EVEC, harnessing the advantages of both models.

Upon reviewing the research studies presented in this paper, it was discovered that there are five key limitations to the works conducted in the literature. These limitations are concisely outlined in the following paragraph.

Real-world data availability is a significant limitation in many studies. While some studies use simulated data sets, it is crucial to apply prediction techniques to real case studies in order to ensure the feasibility of the forecasting approach for real-world applications. Limited availability of data on consumer behavior and usage patterns due to the novelty of electric vehicles can pose a challenge in accurately predicting demand and developing effective models for electric vehicle adoption. Furthermore, accessing the existing data can also be a challenge since it may be owned and controlled by different organizations and companies. For researchers who do not have the required resources or connections, this can pose a barrier to accessing the necessary data.

It is notable that the majority of studies that used real case studies did not offer access to the data used. Providing links to the employed data sets would facilitate other researchers' ability to conduct comprehensive comparisons of various modeling techniques applied to the same data sets. Therefore, we recommend citing the source of the data used for future reference.

The geographical location at which the data sets are collected is another constraint of the utilized case studies. Geographic factors, such as climate, infrastructure, and cultural norms, can cause significant variation in consumer behavior and usage patterns. Hence, models based on data measured at a particular geographic area may not accurately predict EV energy demand in other regions. The case studies examined in the majority of the reviewed research papers were limited to specific geographic areas. Conducting prediction approaches on electric vehicle data sets collected from different geographic regions can increase the validity and feasibility of the research findings.

The reviewed paper lacks sensitivity analysis regarding the inclusion of various input variables in the prediction model. Developing accurate prediction models for electric vehicle energy demand requires sensitivity analysis as an essential aspect. Through conducting an analysis on the impact of different input variables on the model's output, researchers can identify the most significant and influential variables affecting energy demand. This information can be used to adjust the model accordingly for improved accuracy. It is recommended for researchers to incorporate sensitivity analysis of the model input variables since it is an essential tool to ensure the reliability and effectiveness of electric vehicle energy demand prediction model.

A further limitation identified in these reviewed papers is that the performance of the utilized techniques is mainly evaluated based on their accuracy and prediction error, without taking into account the execution time of the models. However, when these models are applied in online applications by electricity system operators, the execution time becomes a critical factor due to the large amount of data to be processed. Future researchers should address this limitation by comparing their proposed techniques not only based on prediction accuracy but also on their execution time, to benchmark methods. This comparison will add value to their work, making it more reliable and suitable for real-time applications where the efficient processing of large amounts of data is crucial.

The fourth limitation observed in the reviewed papers is the lack of justification for the selection of a specific prediction method. The methods are being applied to the data set without carrying out a thorough time series analysis of the case study data.Our recommendation to tackle this problem is to conduct a comprehensive time series analysis that can identify the best approach for predicting electric vehicle energy demand. It's crucial to avoid the risks of implementing unnecessarily complex methods when simpler ones could suffice, or using overly simplistic methods when more sophisticated ones may be required.

The ultimate constraint of the studies analyzed in this research is that those utilizing machine learning methods fail to address the two significant issues of over fitting and under fitting in machine learning-based models. Over fitting arises when a model is excessively complex and conforms too closely to the training data, resulting in a low training error but a high testing error. Conversely, under fitting occurs when a model is too simplistic and is unable to capture the underlying patterns in the data, leading to elevated errors in both training and testing. Achieving a suitable trade-off between model complexity and performance is vital to prevent over fitting or under fitting. It is highly significant to report both training and testing errors to ensure that the model demonstrates acceptable robustness that can generalize well to new data without over fitting.

IX. CONCLUSION

To summarize, this paper has explored several methods for forecasting electric vehicle energy consumption, encompassing linear, nonlinear, and hybrid models. The studies analyzed in this paper have shown that every method has its advantages and limitations with regard to precision, computational complexity, and interpret-ability. In general, the models that combine linear and nonlinear methods, known as hybrid models, demonstrated superior performance in predicting accuracy when compared to other models. Nonetheless, the selection of the optimal model depends on the unique attributes of the data and the objectives of the prediction task. This paper has addressed various restrictions of these models and provided suggestions for further research investigations.

Future research endeavors may encompass a variety of domains encompassing technological advancements, refining methodologies, and considering social and environmental aspects. Integrating blockchain and IoT applications enhances precision and real-time adaptability. Therefore, forthcoming research should explore the integration of these technologies within this field. Additionally, future investigations should prioritize proposing methodologies to augment spatial and temporal resolution in prediction models, accommodating diverse EV charging patterns. Moreover, the inclusion of uncertainty analysis in future studies is imperative, as quantifying uncertainties can yield more robust insights for grid management and energy policy. Furthermore, environmental impact assessments are recommended to be undertaken by researchers in this domain, evaluating the environmental consequences of distinct EV energy management strategies, including grid load distribution and emissions reduction potential. Notably, the reviewed literature lacks consideration of user psychology. Hence, future research could delve into User Behavior Modeling, focusing on comprehending EV users' preferences and habits to enable personalized energy demand prediction models. Additionally, Dynamic Pricing Strategies hold significance in practical applications. To enhance research applicability, future studies are advised to explore dynamic pricing schemes that align EV charging with grid conditions, promoting off-peak charging and bolstering demand response initiatives.

ACKNOWLEDGMENT

This article represents the opinions of the author(s) and does not mean to represent the position or opinions of the American University of Sharjah.

REFERENCES

[1] C. Breyer, S. Khalili, E. Rantanen, and D. Bogdanov, "Solar photovoltaic capacity demand for a fully sustainable transport sector—How to fulfil the Paris agreement by 2050," in *Proc. IEEE 7th World Conf. Photovoltaic Energy Convers. (WCPEC), A Joint Conf. 45th IEEE PVSC, 28th PVSEC, 34th EU PVSEC)*, Jun. 2018, pp. 1632–1637.

[2] X. Zhang, J. Xie, R. Rao, and Y. Liang, "Policy incentives for the adoption of electric vehicles across countries," *Sustainability*, vol. 6, no. 11, pp. 8056–8078, Nov. 2014.

[3] IEA. (2020). *Key World Energy Statistics 2020*. Paris, France: License CC BY 4.0. [Online]. Available: https://www.iea.org/reports/key-world-energy-statistics-2020

[4] X. Zhang, F. Gao, X. Gong, Z. Wang, and Y. Liu, "Comparison of climate change impact between power system of electric vehicles and internal combustion engine vehicles," in *Advances in Energy and Environmental Materials*. Singapore: Springer, 2018, pp. 739–747.

[5] IEA. (2022). *Global EV Outlook 2022*. Accessed: Mar. 18, 2023. [Online]. Available: https://www.iea.org/data-and-statistics/data-product/global-ev-outlook-2022

[6] V. Slednev, P. Jochem, and W. Fichtner, "Impacts of electric vehicles on the European high and extra high voltage power grid," *J. Ind. Ecol.*, vol. 26, no. 3, pp. 824–837, Jun. 2022, doi: 10.1111/jiec.13216.

[7] L. Xiaojun, Z. Xiaodong, Z. Jun, and H. Wenli. (2018). *Energy Consumption and Environmental Impact Analysis of a Novel Vacuum Sewage Collection System: A Case Study in China*. [Online]. Available: https://iopscience.iop.org/article/10.1088/1755-1315/199/5/052033

[8] X. Wu, X. Hu, X. Yin, and S. J. Moura, "Stochastic optimal energy management of smart home with PEV energy storage," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 2065–2075, May 2018, doi: 10.1109/TSG.2016.2606442.

[9] D. Toquica, P. M. De Oliveira-De Jesus, and A. I. Cadena, "Power market equilibrium considering an EV storage aggregator exposed to marginal prices—A bilevel optimization approach," *J. Energy Storage*, vol. 28, Apr. 2020, Art. no. 101267, doi: 10.1016/j.est.2020.101267.

[10] H. J. Feng, L. C. Xi, Y. Z. Jun, Y. X. Ling, and H. Jun, "Review of electric vehicle charging demand forecasting based on multi-source data," in *Proc. IEEE Sustain. Power Energy Conf. (iSPEC)*, Nov. 2020, pp. 139–146.

[11] G. Nalcaci, A. Özmen, and G. W. Weber, "Long-term load forecasting: Models based on MARS, ANN and LR methods," *Central Eur. J. Oper. Res.*, vol. 27, no. 4, pp. 1033–1049, Dec. 2019.

[12] Z. Chen, R. Xiong, K. Wang, and B. Jiao, "Optimal energy management strategy of a plug-in hybrid electric vehicle based on a particle swarm optimization algorithm," *Energies*, vol. 8, no. 5, pp. 3661–3678, Apr. 2015.

[13] M. Hannan, "Hybrid electric vehicles and their challenges," *Renew. Sustain. Energy Rev.*, vol. 29, pp. 135–150, Jan. 2014.

[14] Y. Mu, J. Wu, N. Jenkins, H. Jia, and C. Wang, "A spatial–temporal model for grid impact analysis of plug-in electric vehicles," *Appl. Energy*, vol. 114, pp. 456–465, Feb. 2014.

[15] B. K. Nelson, "Time series analysis using autoregressive integrated moving average (ARIMA) models," *Academic Emergency Med.*, vol. 5, no. 7, pp. 739–744, Jul. 1998.

[16] M. Pertl, F. Carducci, M. Tabone, M. Marinelli, S. Kiliccote, and E. C. Kara, "An equivalent time-variant storage model to harness EV flexibility: Forecast and aggregation," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 1899–1910, Apr. 2019.

[17] M. H. Amini, A. Kargarian, and O. Karabasoglu, "ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation," *Electr. Power Syst. Res.*, vol. 140, pp. 378–390, Nov. 2016.

[18] L. Buzna, P. De Falco, S. Khormali, D. Proto, and M. Straka, "Electric vehicle load forecasting: A comparison between time series and machine learning approaches," in *Proc. 1st Int. Conf. Energy Transition Medit. Area (SyNERGY MED)*, May 2019, pp. 1–5.

[19] N. Korolko, Z. Sahinoglu, and D. Nikovski, "Modeling and forecasting self-similar power load due to EV fast chargers," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1620–1629, May 2016.

[20] P. L. Bartlett, P. M. Long, G. Lugosi, and A. Tsigler, "Benign overfitting in linear regression," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 48, pp. 30063–30070, Dec. 2020.

[21] A. Gerossier, R. Girard, and G. Kariniotakis, "Modeling and forecasting electric vehicle consumption profiles," *Energies*, vol. 12, no. 7, p. 1341, Apr. 2019. [Online]. Available: https://www.mdpi.com/1996-1073/12/7/1341

[22] Y. Xiong, C.-C. Chu, R. Gadh, and B. Wang, "Distributed optimal vehicle grid integration strategy with user behavior prediction," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2017, pp. 1–5.

[23] T. H. C. Tat and P. Fränti, "Real-time electric vehicle load forecast to meet timely energy dispatch," in *Proc. IEEE Int. Conf. Service Oper. Logistics, Informat. (SOLI)*, Jul. 2018, pp. 148–153.

[24] C. De Cauwer, W. Verbeke, T. Coosemans, S. Faid, and J. Van Mierlo, "A data-driven method for energy consumption prediction and energy-efficient routing of electric vehicles in real-world conditions," *Energies*, vol. 10, no. 5, p. 608, May 2017. [Online]. Available: https://www.mdpi.com/1996-1073/10/5/608

[25] R. Zhang and E. Yao, "Electric vehicles' energy consumption estimation with real driving condition data," *Transp. Res. D, Transp. Environ.*, vol. 41, pp. 177–187, Dec. 2015.

[26] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*. New York, NY, USA: Springer, 2013.

[27] C. Chen, Y. Song, X. Hu, and I. G. Guardiola, "Analysis of electric vehicle charging behavior patterns with function principal component analysis approach," *J. Adv. Transp.*, vol. 2020, Nov. 2020, Art. no. 8850654.

[28] A. Bosisio, A. Berizzi, A. Vicario, A. Morotti, B. Greco, G. Iannarelli, and D.-D. Le, "A method to analyzing and clustering aggregate customer load profiles based on PCA," in *Proc. 5th Int. Conf. Green Technol. Sustain. Develop. (GTSD)*, Nov. 2020, pp. 41–47.

[29] X. Qi, G. Wu, K. Boriboonsomsin, and M. J. Barth, "Data-driven decomposition analysis and estimation of link-level electric vehicle energy consumption under real-world traffic conditions," *Transp. Res. D, Transp. Environ.*, vol. 64, pp. 36–52, Oct. 2018.

[30] A. J. Jahromi, M. Mohammadi, S. Afrasiabi, M. Afrasiabi, and J. Aghaei, "Probability density function forecasting of residential electric vehicles charging profile," *Appl. Energy*, vol. 323, Oct. 2022, Art. no. 119616.

[31] C. Crozier, T. Morstyn, and M. Mcculloch, "A stochastic model for uncontrolled charging of electric vehicles using cluster analysis," 2019, *arXiv:1907.09458*.

[32] M. Majidpour, C. Qiu, P. Chu, R. Gadh, and H. R. Pota, "A novel forecasting algorithm for electric vehicle charging stations," in *Proc. Int. Conf. Connected Vehicles Expo (ICCVE)*, Nov. 2014, pp. 1035–1040.

[33] H. Li, J. Zhu, Y. Zhou, D. Feng, K. Zhang, and B. Shen, "Review of load forecasting methods for electric vehicle charging station," in *Proc. IEEE/IAS Ind. Commercial Power Syst. Asia (I&CPS Asia)*, Jul. 2022, pp. 1833–1837.

[34] M. A. Nielsen, *Neural Networks and Deep Learning*. San Francisco, CA, USA: Determination Press, 2015.

[35] M. Dabbaghjamanesh, A. Moeini, and A. Kavousi-Fard, "Reinforcement learning-based load forecasting of electric vehicle charging station using *Q*-learning technique," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 4229–4237, Jun. 2021.

[36] J. Zhu, Z. Yang, Y. Chang, Y. Guo, K. Zhu, and J. Zhang, "A novel LSTM based deep learning approach for multi-time scale electric vehicles charging load prediction," in *Proc. IEEE Innov. Smart Grid Technol.-Asia (ISGT Asia)*, May 2019, pp. 3531–3536.

[37] T. Unterluggauer, K. Rauma, P. Järventausta, and C. Rehtanz, "Short-term load forecasting at electric vehicle charging sites using a multivariate multi-step long short-term memory: A case study from Finland," *IET Electr. Syst. Transp.*, vol. 11, no. 4, pp. 405–419, Dec. 2021.

[38] J. Zhu, Z. Yang, M. Mourshed, Y. Guo, Y. Zhou, Y. Chang, Y. Wei, and S. Feng, "Electric vehicle charging load forecasting: A comparative study of deep learning approaches," *Energies*, vol. 12, no. 14, p. 2692, Jul. 2019.

[39] H. Jahangir, H. Tayarani, A. Ahmadian, M. A. Golkar, J. Miret, M. Tayarani, and H. O. Gao, "Charging demand of plug-in electric vehicles: Forecasting travel behavior based on a novel rough artificial neural network approach," *J. Cleaner Prod.*, vol. 229, pp. 1029–1044, Aug. 2019.

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

[41] B. Lindemann, T. Müller, H. Vietz, N. Jazdi, and M. Weyrich, "A survey on long short-term memory networks for time series prediction," *Proc. CIRP*, vol. 99, pp. 650–655, May 2021.

[42] N. Kumar, D. Kumar, and P. Dwivedi, "Load forecasting for EV charging stations based on artificial neural network and long short term memory," in *Advanced Network Technologies and Intelligent Computing*. Cham, Switzerland: Springer, 2022, pp. 444–454.

[43] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines," in *Proc. Adv. Neural Inf. Process. Syst.*, 1997, vol. 28, no. 7, pp. 779–784.

[44] X. Liao, X. Kang, M. Li, and N. Cao, "Short term load forecasting and early warning of charging station based on PSO-SVM," in *Proc. Int. Conf. Intell. Transp., Big Data Smart City (ICITBS)*, Jan. 2019, pp. 305–308.

[45] Q. Sun, J. Liu, X. Rong, M. Zhang, X. Song, Z. Bie, and Z. Ni, "Charging load forecasting of electric vehicle charging station based on support vector regression," in *Proc. IEEE PES Asia–Pacific Power Energy Eng. Conf. (APPEEC)*, Oct. 2016, pp. 1777–1781.

[46] D. A. Renata, K. Fauziah, P. Aji, A. Larasati, H. Halidah, D. P. Tasurun, Y. Astriani, and Riza, "Modeling of electric vehicle charging energy consumption using machine learning," in *Proc. Int. Conf. Adv. Comput. Sci. Inf. Syst. (ICACSIS)*, Oct. 2021, pp. 1–6.

[47] D. Adhya, A. Pal, A. K. Chakraborty, and A. Bhattacharya, "Machine learning application for prediction of EV charging demand for the scenario of Agartala, India," in *Proc. 4th Int. Conf. Energy, Power Environ. (ICEPE)*, Apr. 2022, pp. 1–5.

[48] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Amer. Statistician*, vol. 46, no. 3, pp. 175–185, Aug. 1992.

[49] R. L. Alvarez-Llames, R. Pérez-Chacón, M. Martínez-Ballesteros, A. Troncoso, and F. Martínez-Álvarez, "A nearest neighbours-based algorithm for big time series data forecasting," in *Hybrid Artificial Intelligent Systems* (Lecture Notes in Computer Science), vol. 9648, F. Martínez-Álvarez, A. Troncoso, H. Quintián, and E. Corchado, Eds. Cham, Switzerland: Springer, 2016.

[50] M. Majidpour, C. Qiu, P. Chu, R. Gadh, and H. R. Pota, "Fast prediction for sparse time series: Demand forecast of EV charging stations for cell phone applications," *IEEE Trans. Ind. Informat.*, vol. 11, no. 1, pp. 242–250, Feb. 2015.

[51] M. Majidpour, C. Qiu, P. Chu, R. Gadh, and H. R. Pota, "Modified pattern sequence-based forecasting for electric vehicle charging stations," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Nov. 2014, pp. 710–715.

[52] Q. Huang, Y. Chen, Z. Sun, J. Cao, J. Zhao, H. Ouyang, P. Zhu, D. Wang, and Y. Liu, "Forecasting of the electric vehicles' charging amount of electricity based on curves clustering," in *Proc. 13th Int. Conf. Natural Comput., Fuzzy Syst. Knowl. Discovery (ICNC-FSKD)*, Jul. 2017, pp. 2424–2428.

[53] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[54] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[55] B. Lim and S. Zohren, "Time-series forecasting with deep learning: A survey," *Phil. Trans. Roy. Soc. A*, vol. 379, no. 2200, 2021, Art. no. 20200209.

[56] J. Zhu, Y. Li, P. Li, and K. Zhou, "Short-term load forecasting for electric vehicle charging stations based on deep learning approaches," *Appl. Sci.*, vol. 9, no. 9, p. 1858, Apr. 2019.

[57] T. Hu, K. Liu, and H. Ma, "Probabilistic electric vehicle charging demand forecast based on deep learning and machine theory of mind," in *Proc. IEEE Transp. Electrific. Conf. Expo (ITEC)*, Jun. 2021, pp. 795–799.

[58] H. Jahangir, S. S. Gougheri, B. Vatandoust, M. A. Golkar, A. Ahmadian, and A. Hajizadeh, "Plug-in electric vehicle behavior modeling in energy market: A novel deep learning-based approach with clustering technique," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 4738–4748, Nov. 2020.

[59] X. Zhang, K. W. Chan, H. Li, H. Wang, J. Qiu, and G. Wang, ''Deep-learning-based probabilistic forecasting of electric vehicle charging load with a novel queuing model,'' *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 3157–3170, Jun. 2021.

[60] M. B. Arias and S. Bae, ''Electric vehicle charging demand forecasting model based on big data technologies,'' *Appl. Energy*, vol. 183, pp. 327–339, Dec. 2016.

[61] X. Shen, H. Zhao, Y. Xiang, P. Lan, and J. Liu, ''Short-term electric vehicles charging load forecasting based on deep learning in low-quality data environments,'' *Electr. Power Syst. Res.*, vol. 212, Nov. 2022, Art. no. 108247.

[62] Q. Gao, T. Zhu, W. Zhou, G. Wang, T. Zhang, Z. Zhang, M. Waseem, S. Liu, C. Han, and Z. Lin, ''Charging load forecasting of electric vehicle based on Monte Carlo and deep learning,'' in *Proc. IEEE Sustain. Power Energy Conf. (iSPEC)*, Nov. 2019, pp. 1309–1314.

[63] D. Zhou, Z. Guo, Y. Xie, Y. Hu, D. Jiang, Y. Feng, and D. Liu, ''Using Bayesian deep learning for electric vehicle charging station load forecasting,'' *Energies*, vol. 15, no. 17, p. 6195, Aug. 2022.

[64] Z. Yi, X. C. Liu, R. Wei, X. Chen, and J. Dai, ''Electric vehicle charging demand forecasting using deep learning model,'' *J. Intell. Transp. Syst.*, vol. 26, no. 6, pp. 690–703, Nov. 2022.

[65] J. Shanmuganathan, A. A. Victoire, G. Balraj, and A. Victoire, ''Deep learning LSTM recurrent neural network model for prediction of electric vehicle charging demand,'' *Sustainability*, vol. 14, no. 16, p. 10207, Aug. 2022.

[66] M. Boulakhbar, M. Farag, K. Benabdelaziz, T. Kousksou, and M. Zazi, ''A deep learning approach for prediction of electrical vehicle charging stations power demand in regulated electricity markets: The case of Morocco,'' *Cleaner Energy Syst.*, vol. 3, Dec. 2022, Art. no. 100039.

[67] J. R. Quinlan, ''Induction of decision trees,'' *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986.

[68] Y.-Y. Song and Y. Lu, ''Decision tree methods: Applications for classification and prediction,'' *Shanghai Arch. Psychiatry*, vol. 27, no. 2, pp. 130–135, Apr. 2015.

[69] S. B. Kotsiantis, ''Decision trees: A recent overview,'' *Artif. Intell. Rev.*, vol. 39, no. 4, pp. 261–283, Apr. 2013.

[70] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, ''An introduction to decision tree modeling,'' *J. Chemometrics*, vol. 18, no. 6, pp. 275–285, Jun. 2004.

[71] A. Liaw and M. Wiener, ''Classification and regression by randomForest,'' *R News*, vol. 2, no. 3, pp. 18–22, Nov. 2002.

[72] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, ''A comparison of random forest variable selection methods for classification prediction modeling,'' *Expert Syst. Appl.*, vol. 134, pp. 93–101, Nov. 2019.

[73] Z. Hajirahimi and M. Khashei, ''Hybridization of hybrid structures for time series forecasting: A review,'' *Artif. Intell. Rev.*, vol. 56, no. 2, pp. 1201–1261, Feb. 2023.

[74] W. Gang, L. Wu, and G. Xuan, ''A load forecasting method of electric vehicles charging station group based on GAN-RF model,'' in *Proc. IEEE 5th Conf. Energy Internet Energy Syst. Integr. (EI2)*, Oct. 2021, pp. 3076–3079.

[75] F. Ru, X. Yang, H. Zou, L. Zhang, and X. Xu, ''Electric vehicle charging station load prediction based on IWOA-LSSVM combined model with improved variational mode decomposition,'' *J. Phys., Conf. Ser.*, vol. 2320, no. 1, Aug. 2022, Art. no. 012007.

[76] Y.-W. Chung, B. Khaki, C. Chu, and R. Gadh, ''Electric vehicle user behavior prediction using hybrid kernel density estimator,'' in *Proc. IEEE Int. Conf. Probabilistic Methods Appl. Power Syst. (PMAPS)*, Jun. 2018, pp. 1–6.

[77] B. Lee, H. Lee, and H. Ahn, ''Improving load forecasting of electric vehicle charging stations through missing data imputation,'' *Energies*, vol. 13, no. 18, p. 4893, Sep. 2020, doi: 10.3390/en13184893.

[78] K. Lu, W. Sun, C. Ma, S. Yang, Z. Zhu, P. Zhao, X. Zhao, and N. Xu, ''Load forecast method of electric vehicle charging station using SVR based on GA-PSO,'' *IOP Conf. Ser., Earth Environ. Sci.*, vol. 69, Jun. 2017, Art. no. 012196.

[79] X. Wang, ''Electric vehicle charging load prediction with improved kernel extreme learning machine,'' *Proc. SPIE*, vol. 12557, Aug. 2022, Art. no. 122572O.

[80] I. Ullah, K. Liu, T. Yamamoto, M. Zahid, and A. Jamal, ''Electric vehicle energy consumption prediction using stacked generalization: An ensemble learning approach,'' *Int. J. Green Energy*, vol. 18, no. 9, pp. 896–909, Jul. 2021, doi: 10.1080/15435075.2021.1881902.

[81] Y.-E. Jeon, S.-B. Kang, and J.-I. Seo, ''Hybrid predictive modeling for charging demand prediction of electric vehicles,'' *Sustainability*, vol. 14, no. 9, p. 5426, Apr. 2022.

[82] Y. Li, Y. Huang, and M. Zhang, ''Short-term load forecasting for electric vehicle charging station based on niche immunity lion algorithm and convolutional neural network,'' *Energies*, vol. 11, no. 5, p. 1253, May 2018.

[83] F. Ren, C. Tian, G. Zhang, C. Li, and Y. Zhai, ''A hybrid method for power demand prediction of electric vehicles based on SARIMA and deep learning with integration of periodic features,'' *Energy*, vol. 250, Jul. 2022, Art. no. 123738.

[84] J. Feng, J. Yang, Y. Li, H. Wang, H. Ji, W. Yang, and K. Wang, ''Load forecasting of electric vehicle charging station based on grey theory and neural network,'' *Energy Rep.*, vol. 7, pp. 487–492, Nov. 2021.

[85] M. D. Eddine and Y. Shen, ''A deep learning based approach for predicting the demand of electric vehicle charge,'' *J. Supercomput.*, vol. 78, no. 12, pp. 14072–14095, Aug. 2022, doi: 10.1007/s11227-022-04428-0.

[86] Q. Gu, Q. Dai, H. Yu, and R. Ye, ''Integrating multi-source transfer learning, active learning and metric learning paradigms for time series prediction,'' *Appl. Soft Comput.*, vol. 109, Sep. 2021, Art. no. 107583.

FATEMEH MARZBANI received the B.Sc. degree (Hons.) in electrical engineering from Shiraz University, Iran, in 2011, and the M.Sc. degree in electrical engineering from the American University of Sharjah, United Arab Emirates, in 2015, where she is currently pursuing the Ph.D. degree in engineering systems management with a specialization in the smart cities track. Her research interests include smart grids, power systems management, renewable energies, machine learning, and big data analysis. Her work strives to contribute to the advancement of intelligent and sustainable energy systems.

AHMED H. OSMAN (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Helwan University, Cairo, Egypt, in 1991 and 1996, respectively, and the Ph.D. degree in electrical engineering from the University of Calgary, Calgary, AB, Canada, in 2003. From 2004 to 2008, he was an Assistant Professor with the Department of Electrical and Computer Engineering, University of Calgary. He is currently a Professor with the Department of Electrical Engineering, American University of Sharjah, Sharjah, United Arab Emirates. His research interests include power system analysis and power system protection.

MOHAMED S. HASSAN received the M.Sc. degree in electrical engineering from the University of Pennsylvania, Philadelphia, in 2000, and the Ph.D. degree in electrical and computer engineering from the University of Arizona, USA, in 2005. He is currently a Full Professor in electrical engineering with the American University of Sharjah. In addition to electric vehicles, he has recently actively taken part in multiple projects involving fields including free space optical communications, demand response, and smart grids. His research interests include multimedia communications and networking, wireless communications, cognitive radios, resource allocation, and performance evaluation of wired and wireless networks, particularly next-generation wireless systems.