Machine Learning Approaches for EV Charging Behavior: A Review

5 authors, including:

Sakib Shahriar
American University of Sharjah
14 PUBLICATIONS   **47** CITATIONS

SEE PROFILE

Ahmed Osman
American University of Sharjah
89 PUBLICATIONS   **1,471** CITATIONS

SEE PROFILE

Salam Dhou
Harvard Medical School
53 PUBLICATIONS   **164** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Special Issue "Electric Vehicle Charging Technologies and Trends" View project

Project   Power System Protection View project

Machine Learning Approaches for EV Charging Behavior: A Review

SAKIB SHAHRIAR[1], **A. R. AL-ALI**[1], **(Life Senior Member, IEEE),**
AHMED H. OSMAN[2], **(Senior Member, IEEE), SALAM DHOU**[1], **(Member, IEEE),**
AND MAIS NIJIM[3], **(Member, IEEE)**

[1]Department of Computer Science and Engineering, American University of Sharjah, Sharjah, United Arab Emirates
[2]Department of Electrical Engineering, American University of Sharjah, Sharjah, United Arab Emirates
[3]Department of Electrical Engineering and Computer Science, Texas A&M University-Kingsville, Kingsville, TX 78363, USA

Corresponding author: A. R. Al-Ali (aali@aus.edu)

ABSTRACT As the smart city applications are moving from conceptual models to development phase, smart transportation is one of smart cities applications and it is gaining ground nowadays. Electric Vehicles (EVs) are considered one of the major pillars of smart transportation applications. EVs are ever growing in popularity due to their potential contribution in reducing dependency on fossil fuels and greenhouse gas emissions. However, large-scale deployment of EV charging stations poses multiple challenges to the power grid and public infrastructure. To overcome the issue of prolonged charging time, the simple solution of deploying more charging stations to increase charging capacity does not work due to the strain on power grids and physical space limitations. Therefore, researchers have focused on developing smart scheduling algorithms to manage the demand for public charging using modeling and optimization. More recently, there has been a growing interest in data-driven approaches in modeling EV charging. Consequently, researchers are looking to identify consumer charging behavior pattern that can provide insights and predictive analytics capability. The purpose of this article is to provide a comprehensive review for the use of supervised and unsupervised Machine Learning as well as Deep Neural Networks for charging behavior analysis and prediction. Recommendations and future research directions are also discussed.

INDEX TERMS Electric vehicles, machine learning, smart city, smart transportation, big data.

I. INTRODUCTION

In recent years, climate change has become a growing concern. As such, United Nations (UN) have placed combatting climate change under one of the sustainable development goals (SDGs), with plans to jointly raise 100 billion dollars by 2020 to fight the crisis [1]. The transportation sector is responsible for over a quarter of the world's energy consumption [2]. According to UN, two thirds of the world's population is projected to live in urban areas by 2050 [3], which would inherently increase the demand for vehicles to provide urban mobility and in turn increase fossil fuel consumption and greenhouse emissions. According to a Chinese study, Electric Vehicles (EVs) could potentially provide 45% reduction in carbon emissions compared to

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

conventional internal combustion engine (ICE) vehicles after considering the energy cost of production, assembly, transportation, and usage [4]. Therefore, EVs are considered to be the frontrunners in providing clean source of transportation.

Within the smart cities context, the massive growth in EV popularity [5] can be attributed to the rapid improvements in battery technology. The latest EVs have the capacity to travel between 300-500 kms per full charge unlike the older versions which would often last less than 100 kms per full charge. The improvement in battery technology has made EVs far more usable, not only for commuting short distances but also inter-city travel. Consequently, the number of EV charging stations have grown, allowing greater flexibility for drivers to plan their drives. Furthermore, the reliability of EVs has improved considerably since the earlier days thereby offering greater consumer trust and satisfaction. The technological improvement can also be attributed to

the market competitiveness with many private and public companies taking the initiative to produce commercial EVs.

Despite the promising signs, there remains a few challenges. Firstly, most EVs take a long time to charge, therefore causing great inconvenience. In addition to this, many EV owners do not have the capacity to charge their cars at home but must rely on public charging stations. Due to the high power needs of the EVs, integrating them on massive scale will place huge constraints on the power distribution grid [6]–[8]. Un-coordinated EV charging behavior is likely to cause further degradation and instability in power distribution networks. The implications of the power constraints mean that it is virtually impossible to increase the charging station capacity to meet the increasing changing needs. Unlike gas stations for ICE vehicles, where the vehicles can get refueled in minutes, EVs often require hours to recharge. The easy solution of increasing the capacity by deploying more stations is not feasible because in addition to the power needs, there also exists a physical space limitation and the number of charging stations can only be increased to a limited number. Thus, the optimal solution is to better manage the scheduling of charging stations. Several works have focused on smart scheduling to efficiently manage the charging load using optimization problems [9], [10] and metaheuristic approach [11]. Studies have shown that energy management with regards to EV charging greatly impacts wholesale electricity market [12], adding further significance to the need for understanding charging behavior. Researchers have investigated the psychological dynamics that influence charging behavior [13], [14]. Spoelstra [15] used charging transactions data and interviews with EV drivers to provide an analysis of EV charging behavior and the factors that influence such behavior. Although the outcome of these studies provides a high-level understanding of EV charging behavior, it is important to quantify the results in order to use it for scheduling and management. Analysis of charging behavior using simulations, such as in [16], [17] contains assumptions that might not hold true in real-world scenarios. Similarly, estimation of EV behavior derived from ICE vehicle driving data [18]–[20] and synthetically generated data [21], [22] may not reflect the unpredictable charging behavior in everyday scenario. Other strategies such as multi-location charging, whereby employees are encouraged to charge at home as well as the workplace, to control the load have shown promising results [23]. However, these approaches are only suitable in theory as it is not easy to control or enforce user charging behavior.

With the immergence of big data analytics and machine learning (ML), which have revolutionized fields of natural language processing, image, audio and video recognition, the focus has shifted towards utilizing data-driven approaches [24]–[26] to solve the EV charging problem. Using historical data of charging load and user behavior, ML algorithms can be utilized to train and learn the trends and patterns from the data. After the training phase, accurate predictions can be obtained. Such predictions can then be utilized, either independently or in conjunction with other algorithms, to enhance EV charging scheduling strategies. Studies have shown that ML algorithms are capable of providing good forecasts for timeseries data [27], and can therefore be used for charging behavior predictions. Traditional approaches for the analysis of charging behavior using qualitative studies is limited by the fact that it cannot be easily integrated into practical applications such as scheduling. In contrast, using ML approach, predictions of future behavior can be obtained which can then be used by a smart scheduler for optimal scheduling. Although alternative methods such as modeling and simulation can be useful for practical applications, they contain assumptions that can limit their accuracy. On the other hand, ML-based models can make use of both historical data as well as everyday factors such as weather and traffic variables to accurately capture the trends in charging behavior.

Although there exist several review works in the literature with regards to EV charging, they do not focus on the charging behavior from a data driven approach. For instance, [28] provides a survey of optimization and mathematical modelling-based solutions to EV operations and management, including EV charging. Ahmad *et al.* [29] provided a review of various charging technologies and standards with a case study from Germany. The authors in [30] reviewed charging station location problem, which is concerned with the location and deployment of charging stations. Perhaps the most related work to the one proposed in this article is presented in [31], where the authors review scheduling, clustering, and forecasting strategies for EV charging. It also presents a review of data-driven approaches along with other approaches such as optimization and how they have been used for EV charging strategies. In contrast, the survey work in this article is solely focused on providing a review of the existing ML approaches used in analysis and prediction of EV charging behavior. The key contributions of this survey are the following:

1) It provides an overview of the various ML approaches, including supervised, unsupervised, and deep learning, as well as the common ML evaluation metrics.
2) It provides a comparison of the recent works in predicting EV charging behavior using ML-based approaches and highlights their impacts.
3) It proposes a discussion about the limitations of the existing studies and provides future research directions.

The rest of the paper is organized as follows. A background section including key concepts in ML is provided in Section II. This is followed by a brief description about the commonly available datasets for EV charging. A detailed review of ML for understanding and prediction of charging behavior is presented next in sections III, IV and V. The authors then highlight the challenges and propose future research directions in Section VI. Section VII concludes the paper.

II. BACKGROUND

This section summarizes the background information pertaining to EV charging, ML techniques, and evaluation of ML algorithms. A brief description of the common datasets that can be used for prediction of charging behavior is also presented.

A. EV CHARGING

EV charging is primarily divided into three charging levels as defined by the Electric Power Research Institute. Level 1 charging provides the slowest charging rate, operating at standard 120 V/15A [32]. The charging equipment in this case is installed on the EV and power is transferred to the vehicle using a plug and cord set. Level 2 charging on the other hand uses 240 VAC and has been utilized for both private and public facilities. It provides faster charging as compared to Level 1 but requires installation of dedicated charging equipment. Level 3, also referred to as 'fast charging', utilizes 480 VAC and is typically deployed in commercial/public settings with the goal of providing 'grab and go' service similar to gas stations for ICE vehicles. Level 3 provides the quickest charging rate with vehicles being able to recharge in less than 30 minutes. Figure 1 provides an illustration of the 3 charging techniques.
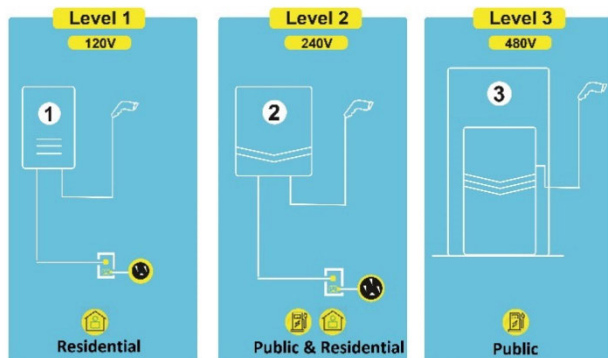

FIGURE 1. Categories of EV charging.

EV charging can also be categorized into residential charging and non-residential (commercial) charging. Typically, Level 1 and Level 2 chargers are deployed for residential purposes. Residential charging behavior is more predictable, and scheduling is therefore easier. Typically, users leave their vehicles to charge overnight or arrange charging sessions depending on their working hours. The number of vehicles using residential charging is also predictable because usually people who own EVs in a given area is likely to utilize the stations within that residential area. In contrast, the number of vehicles using non-residential charging station is unpredictable and depends on a lot of factors. For instance, traffic of a public charging facility near a shopping mall will depend on lot of factors including weather, day of the week, and mall offers. Therefore, it is perhaps more significant to understand the charging behavior of non-residential charging

facilities which are dynamic in nature to provide more precise scheduling.

B. MACHINE LEARNING AND PREDICTIVE ANALYTICS

Machine learning (ML) provides computer systems with the ability to learn from experience without the need for explicit programming. The experience in this context is the dataset that the algorithms use to train themselves on. With time the models are able to discover the underlying trends and patterns in the dataset. Upon successful learning, these models are able to make accurate predictions about the future and therefore provide predictive analytics. ML algorithms are typically categorized into supervised and unsupervised learning. Further categorization can be done depending on the type of the variable to be predicted, also known as response variable. If the response variable is continuous, the problem being solved is called a regression problem. Conversely, if the response variable is categorical, the problem is called a classification problem. Figure 2 illustrates the difference between regression and classification in the context of EV charging. The figure on the left portrays prediction of energy consumption based on charging session duration. This is a regression problem because the response variable, energy, is a continuous value. In contrast, the figure on the right portrays distinction of EV drivers who prefer to charge their cars during nighttime against those who prefer to charge during the day. In this case it is a classification problem because the variable of interest is categorical.
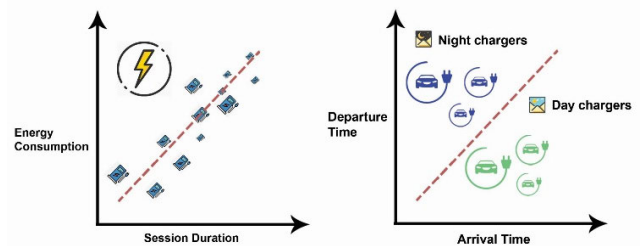

FIGURE 2. Illustration of Regression (left) and Classification (right) problems.

C. SUPERVISED LEARNING

In supervised learning, ML models are trained from labeled training dataset. As such the dataset contains both the input variables and the corresponding response variable, often called the target variable. The model iteratively learns the mapping between the input and the response variables by optimizing a given objective function. A simple example in the context of EV charging could be a dataset containing the arrival time of the vehicle, city name, and departure time of the vehicle. If the goal is to predict the departure time, the ML model will learn the relationship between the arrival time, city name (input variables) and the departure time (response variable). A discussion of all the supervised learning algorithms is beyond the scope of this work, however

the most frequently used algorithms for prediction of EV charging behavior are discussed below:

1) LINEAR REGRESSION

Linear regression (LR) can be used to model the mathematical relationship between the output variable and one or multiple input variables (multiple LR). In LR, it is assumed that there is a linear relationship between the response variable and the input features. LR can be represented by Equation 1:

$$y = b_o + b_1x_1 + b_2x_2 + \cdots + b_nx_n \qquad (1)$$

where y represents the target variable, b_o represents the y-intercept, $[x_1, x_2 \cdots x_n]$ represents the input features and $[b_1, b_2 \cdots b_n]$ represents the regression coefficients. Often gradient descent [33] method is used to find the coefficients by minimizing sum of the squared error iteratively after starting off with random values for coefficient. As the name suggests, LR is used for regression problems. LR are particularly useful when the dataset is linearly separable and the algorithm itself is very simple to implement. Overfitting [34] is a major challenge in training ML algorithms, it occurs when a given model performs exceptionally well during the training phase, by using unnecessary input features, but fails to make generalized predictions. The performance of LR can be impacted by overfitting as well as the presence of outliers.

2) DECISION TREES

A decision tree (DT) can be used for both classification and regression problems [35]. Similar to a flow chart, DTs separate complex decisions into a combination of simpler decisions using split points from the input features. The point where decisions take place is called a decision node. The points where no further split is made are called the leaf nodes. For regression problems, the average value of all the items in the leaf node is taken for prediction. For classification problems, the leaf nodes are the set of classes being predicted. DTs are simple to explain particularly using a tree diagram which can help to understand the prediction making process. However, a single DT often fails to provide good predictions and is prone to overfitting.

3) RANDOM FOREST

In random forest (RF), predictions are made by aggregating multiple decision trees. Bagging method is used in this case where the trees are created from various bootstrap sample, (i.e., sample with replacement). The aggregation for regression is done by taking the average value of the predictions by all the trees and for classification majority vote across the trees are taken [36]. RF is an example of ensemble ML, where individual ML models are first evaluated and then integrated into a single model that can often produce superior predictive performance than the individual models. The motivation behind such approach is similar to asking multiple experts about an opinion, and then taking their votes to make the final decision [37]. Similarly, a gradient boosting algorithm or XGboost [38] uses multiple DTs, with the key

difference being gradient boosting builds each tree one after another while taking the errors made by the previous trees into consideration. Both approaches greatly reduce overfitting as compared to simple DT model.

4) SUPPORT VECTOR MACHINE

A support vector machine (SVM) [39] is mainly used for classification problems but can also be used for regression in which case they are often referred to as support vector regression (SVR) [40]. SVM separates the classes with the best hyperplane that can maximize the margin between the respective classes. Using kernels such as linear, polynomial, and radial basis function (RBF), the inputs can be mapped to high dimensional feature spaces where they can be linearly separable. One of the main disadvantages of SVM is the lengthy training time. Therefore, for larger datasets SVM may not be suitable.

5) K-NEAREST NEIGHBOR

Although k-nearest neighbor (KNN) [41] can be used for both regression and classification, it is more popular for classification problems. For KNN, dedicated training phase is not required, and it is also known as a form of lazy learning. For making prediction of a new data point, a distance measure typically Euclidean distance is used to find its k nearest neighbors. Then it is assigned to the class that contains majority of the neighbors. Figure 3 illustrates this process where k is set to 3 and therefore can be also called a 3-NN algorithm. In this example, the three nearest neighbors of the new item in green are two items from the orange class and one item from the blue class. Therefore, the new item in green will be assigned to the orange class.
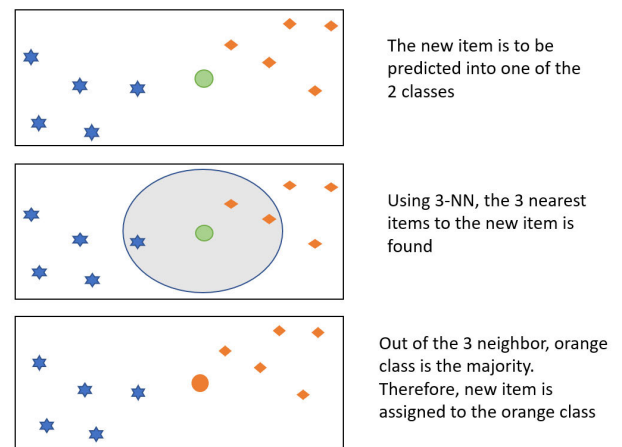

The new item is to be predicted into one of the 2 classes

Using 3-NN, the 3 nearest items to the new item is found

Out of the 3 neighbor, orange class is the majority. Therefore, new item is assigned to the orange class

FIGURE 3. Illustration of k-NN algorithm prediction.

D. UNSUPERVISED LEARNING AND STATISTICAL MODELS

In unsupervised learning, the training dataset is comprised of only input variables, without labeled output variables. In many practical applications, data labeling is time

consuming and costly. The goal of the ML model is to find structures or patterns within the dataset. Cluster analysis is a common example of unsupervised learning whereby the goal of the ML model is to find clusters of items that have some common elements. Unsupervised learning can be utilized to find clusters of EV behavioral patterns. Figure 4 provides an illustration of clustering in the context of EV charging. In this simple example, only 2 input features are being considered to group the items. Based on the arrival time and day of the week, we can notice 3 distinguish clusters of charging behavior.
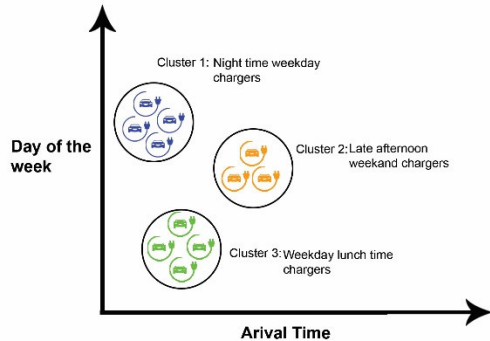

FIGURE 4. Simple illustration of clustering.

1) K-MEANS CLUSTERING

In K-Means clustering, individual data points form k clusters with each point being assigned at the beginning to k center points in a random manner. Next the data points are reassigned to the closest center based on new center calculations. The number of clusters must be known beforehand or can be calculated best on elbow method [42]. K-Means is a simple clustering algorithm but is sensitive to outliers and initial assignment. K-Means is among the most popular clustering algorithms along with Density-Based clustering (DBSCAN) [43] and hierarchical clustering [44].

2) GAUSSIAN MIXTURE MODEL

Gaussian mixture model (GMM) [45] is a probabilistic learning model that can represent normally distributed sub-populations by considering multiple normal distributions of the dataset. Although GMMs are mainly used as unsupervised learning, variations of it exist for supervised learning. For unsupervised model, prior knowledge of the subpopulations is not required. Based on the distributions of the dataset, such as binomial, poisson and exponential, various form of mixture models can be derived. Other common mixture models include beta mixture model (BMM) where the beta probability distribution is considered.

3) KERNEL DENSITY ESTIMATOR (KDE)

The shape of probability density function (PDF) must be assumed in parametric estimation methods. When this is not possible, we can use nonparametric estimation to estimate the

PDF of a continuous random variable using kernel functions. Kernel functions must be symmetrical, nonnegative and area under the function curve must be 1. Popular kernels for KDE include normal or gaussian KDE (GKDE) and diffusion-based KDE (DKDE) [46].

E. DEEP LEARNING

Deep learning (DL) is a subset of ML that utilizes artificial neural networks (ANNs). DL models, unlike ML models, contain a large amount of composition of learned functions. More specifically, using layered hierarchy of concepts, complex concepts are defined in terms of simpler concepts and more abstract representations are gathered using less abstract ones [47]. DL is an emerging technology, but it dates back to 1940s and was known by various names, such as connectionism and cybernetics, during earlier days. The recent success of DL-based models can be attributed to two main factors: 1) Availability of larger datasets to train DL models. 2) Availability of powerful computers to build and train complex models to achieve groundbreaking results [47]. DL-based models currently provide cutting-edge solutions to various areas in natural language processing (NLP) and computer vision.

One of the most common DL methods is multilayer perceptron (MLP), which is often simply referred to as ANN. MLP can use non-linear approximation given a set of input features and can be used for both regression and classification. An MLP consist of input layer which takes in the set of features, the hidden layers that learns the representations and the output layer that makes the final predictions. Figure 5 shows an MLP with 3 hidden layers for binary classification.
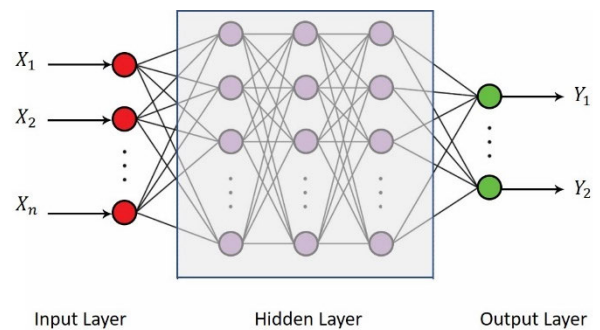

FIGURE 5. ANN with 3 hidden layers for binary classification.

For sequential and time series data, recurrent neural networks (RNN) are more suitable because they have the ability to process short-term dependencies by considering some of the previous predictions. For instance, in NLP tasks such as predicting next words of a sentence, it is likely that the previous few words greatly influence the predictions and thus they need to be considered for the next predictions. For tasks when long term dependencies needs to be considered, a RNN-based model known as long short-term memory (LSTM) [48] can be used.

F. EVALUATION OF ML MODELS
This section defines the common metrics used in the review paper to evaluate the ML models for predicting EV charging behavior.

1) EVALUATION OF REGRESSION
Assuming that the original value is represented by y and the predicted value is represented by $\bar{y}$ and n represents the groups of values in the dataset, then the following methods (Equations 2-5) are commonly used to evaluate the performance of regression models:

Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}{n}} \qquad (2)$$

Mean absolute error (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \bar{y}_i| \qquad (3)$$

Mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \bar{y}_i}{y_i} \right| * 100 \qquad (4)$$

Symmetric mean absolute percentage error (SMAPE):

$$SMAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \bar{y}_i|}{(|y_i| - |\bar{y}_i|)/2} * 100 \qquad (5)$$

If the predicted value, $\bar{y}$ is very different to the actual value y, the result of these metrics will be high. On the other hand, the lower the values of these metrics, the more accurate the models are. MAPE is problematic when the actual value y is close to 0 in the denominator and therefore creating a bias. On the contrary, SMAPE is the preferred metric for EV charging prediction because both the actual value and the predicted value is in the denominator [49].

2) EVALUATION OF CLASSIFICATION
For classification models, a confusion matrix is often used to present the classification results and can also be used to derive additional metrics. Figure 6 shows a confusion matrix for binary classifiers where Class 1 represents the positive class and Class 2 represents the negative class. True positive (TP) represents the cases where the model predicted positive class value and the actual value also belonged to the positive class and similarly true negative (TN) represents the cases where the model predicted negative class and the actual value belonged to the negative class.

	Predicted Class 1	Predicted Class 0
Actual Class 1	True Positive (TP)	False Negative (FN)
Actual Class 0	False Positive (FP)	True Negative (TN)

FIGURE 6. Confusion matrix for binary classification.

False positive (FP) represents the cases where the model predicted 'falsely' a positive class for an actual value belonging to the negative class, and False negative (FN) represents the cases where the model predicted 'falsely' a negative class where in fact the value belonged to the positive class. Based on this we can define the following metrics (Equations 6-9):

Classification Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (6)$$

True positive rate (TPR) or recall:

$$TPR = \frac{TP}{FN + TP} \qquad (7)$$

Precision:

$$Precision = \frac{TP}{TP + FP} \qquad (8)$$

F1 Score:

$$F1Score = \frac{2 * Precision * TPR}{Precision + TPR} \qquad (9)$$

3) EVALUATION OF CLUSTERING
To determine the quality of clustering, cluster validation can be done in two ways. In external validation, the ground truth or the actual labels are known and can then be compared to the cluster assignment results. Common external validation includes purity, entropy [50] and adjusted rand index (ARI) [51]. Internal validation can be used if the ground truth is not available. Common internal validation methods, such as silhouette and dunn index [50] work by ensuring items within the same cluster are highly similar and clusters are themselves dissimilar from one another.

G. EV CHARGING DATASETS
The success of good predictive ML model depends on the quality of the dataset. In this section, the commonly used datasets for studying EV charging behavior are discussed.

Two EV charging datasets were presented in [52], one of them containing about 8500 residential charging sessions and the other dataset containing more than 1 million sessions from a public charging facility in the Netherlands. The residential dataset contains charging data spanning for a year (March 2012-March 2013) along with the trip details of EVs using GPS logger. This provides a good platform for studying residential EV driving and charging behavior. The non-residential dataset spans from January 2011 until December 2015 and was collected by ElaadNL. My Electric Avenue [53] consists of driving and charging behavior of UK drivers from January 2014 to November 2015. Residential water and energy data collected by Pecan Street's water and electricity research is hosted by Dataport [54]. Although this dataset is available to public for research, it is limited to residential EV charging behavior. ACN-Data [55] is among the most recently released public dataset on EV charging, containing more than 30,000 charging

sessions collected from two non-residential charging sites in California. Additional user data such as estimated departure time and requested energy is collected through user mobile application by scanning a QR code. When a user does not use the mobile application, default values are generated for these fields, without attaching user identifier for such sessions.

H. EV CHARGING BEHAVIOR

Given a charging session, we consider session charging behavior $B_{session}$ as following:

$$B_{session}\left(t_{con}, t_{full}, t_{discon}, e\right) \qquad (10)$$

where t_{con} represents the connection time (also arrival or start time), t_{full} represents the time after which no charge was delivered, t_{discon} represents the disconnection time (also departure or end time) and e represents the energy delivered to the car during the session. Based on the above, we can define the session duration, S_{dur}, as follows:

$$S_{dur} = t_{con} - t_{discon} \qquad (11)$$

Other aspects of charging behavior are implicit in nature because they are impacted by the characteristics defined above and can include the energy consumption of the entire charging outlet, EV charging profile and charging load of the station. We have also considered works that studied specific charging behavior such as time to next plug in, whether a vehicle will charge next day and speed of charging.

Following the discussions on the background materials, we now review the recent works utilizing ML for EV charging behavior. We will begin with a comparison of the supervised learning techniques in the next section before moving on to unsupervised learning approaches. Although deep learning can be classified into supervised and unsupervised learning, we have decided to include an independent section because of the recent success of deep learning models and the growing interest of the research community. This would be convenient for researchers who are solely interested in deep learning works.

III. SUPERVISED LEARNING FOR ANALYSIS AND PREDICTION OF CHARGING BEHAVIOR

Frendo *et al.* [56] developed regression models to predict the departure time of EVs. The models were trained on historical data containing over 100000 charging sessions spanning over 3 years. Features such as mean session duration by user ID, arrival time and weekday were among the significant predictors of departure time. The best performing model XGBoost achieved MAE of 82 minutes for departure time prediction. The predictions made by the ML models had significant impact on scheduling quality. Similarly, [57] used mean estimation to predict user behavior in terms of start time and session duration. The authors then used LR model to predict the energy consumption using session duration. Although the predictions of EV behavior was integrated into the smart charging algorithm to achieve stabilization to the power grid, the paper did not evaluate the performance of

the predictive models. SVM was used to predict arrival and departure time of EVs in [58]. The dataset used for training consists of 3 years (2012-2014) charging data of commuters using EVs in University of California, San Diego (UCSD) campus. The input features used to train the ML model included temporal features (week. day, hour) as well as previous arrival and departure times. The average MAPE for arrival time was 2.85% and for departure time was 3.7%. The proposed model demonstrated superior performance against simple persistence reference forecast. The paper failed to address SVM hyperparameter tuning which can often enhance predictions [59].

The authors in [49] utilized several ML models, including DT, K-NN and RF to predict session duration and energy consumption from two charging datasets. The first dataset contained charging sessions from University of California, Los Angeles (UCLA) campus, thus representing non-residential charging behavior. The second dataset represented residential charging data from UK EV drivers. For session duration, SVR performed the best (SMAPE 10.54%) followed by LR (SMAPE 11.05%). As for energy consumption, RF performed the best (SMAPE 8.65%) with DKDE a close second (SMAPE 8.73%). Based on the preliminary results by the various models, the authors selected SVR, RF and DKDE to form an ensembled model. The proposed ensembled model outperformed the individual models in both predictions. The SMAPE for charging duration was 10.4% and for energy consumption was 7.54%. The results of the proposed model when applied to scheduling algorithm not only reduced peak load by 27%, but also reduced charging cost by 4%. Similarly, [60] used ensemble models including RF, naive Bayes (NB) and ANNs to predict whether or not the EVs will be charged the next day in a household. The hours of the day the EVs will be charged in the next day is also predicted. Among the input features used for the predictive models included charging consumption of the last day and charging occurrence time of the last day. The combination of RF, NB, AdaBoost and GBoost algorithms provide the best performance achieving TPR of 0.996 for predicting whether the EVs will be charged the next day and accuracy of 0.724 for predicting the hours of the next day when the EVs will be charged. This study provides a different aspect of charging behavior compared to other works and is particularly useful for residential settings.

KNN was used to predict energy consumption at a charging outlet in [61] using the data from UCLA campus. The best performing model was with k set to 1 (1-NN) and the performance of the model was significantly improved by using time-weighted dot product dissimilarity measure that achieved SMAPE of 15.27%. The predictive model was integrated to a cell phone application that can predict the end time of charging and the available energy in about 1 second. Various ML algorithms were utilized in [62] with the goal of predicting the energy needs at a charging outlet in the next 24 hours. Among the ML algorithms used was pattern sequence-based forecasting (PSF) [63], which works

TABLE 1. Supervised learning for charging behavior.

Source	Charging Behavior	ML Model	Results and Impacts
[49]	Predict session duration and energy consumption from both residential and non-residential	SVR, RF and DKDE combined to form an ensembled model.	SMAPE charging duration: 10.4%, SMAPE energy consumption: 7.54%. Reduced peak load by 27%, and reduced charging cost by 4% when integrated to scheduler.
[55]	Predict session duration and energy needs for non-residential public charging space, CA, USA	Probabilistic GMM	SMAPE user duration: 12.25%, SMAPE energy consumption: 12.73%
[56]	Predict EV charging departure time	Regression including XGBoost and LR	Best result using XGB: MAE of 82 minutes for departure
[57]	Predict start time, end time, energy consumption.	LR for consumption	-
[58]	Predict arrival and departure time EVs in a university campus (non-residential)	SVMs	Average MAPE arrival time: 2.85%, departure time: 3.7%.
[60]	Predict whether the EVs will be charged the next day, and which hours they will be charged for residential dataset.	Ensembled model using RF, Naïve Bayes, AdaBoost and Gradient boosting	TPR for predicting whether the EVs will be charged: 0.996, Accuracy for predicting the hours when the EVs will be charged: 0.724
[61]	Predict energy consumption at a charging outlet in a university campus (non-residential)	KNN, Best model was with k set to 1 (1-NN)	SMAPE: 15.27%. The predictive model integrated to a mobile application can predict the end charging time and energy in 1 second
[62]	Predict the energy needs at a charging outlet in the next 24 hours for a university campus	Various ML including PSF, SVR, RF	Best result using PSF model with average SMAPE: 14.06%
[64]	Predict energy consumption of a session	PSF-based using kNN	SMAPE: 7.85%
[65]	Classify whether the driver will use fast charging	Binary log. regression	Accuracy: 0.894
[66]	Predict the time to next plug for residential charging	SVR with radial basis	MAE: 0.124 minutes, RMSE: 0.158 minutes
[67]	Predict charge profiles in workplace	XGBoost, LR and ANN	Best result using XGBoost MAE: 126 W. Addition to scheduling lead to up to 21% increase in charge.
[68]	Develop model to predict charging speed using temperature, connection time, SOC	LR	-
[69]	Predict charging capacity and daily charging times.	RF	MAPE: 9.76% 1 station. MAPE: 12.8% for groups of stations for prediction of charging load for next 15 minutes

by first classifying the days using clustering before making predictions for that day. PSF and SVR methods achieved the best performance on the UCLA dataset, with the current hour and previous hour energy being the most significant input variables. However, it must be noted that the SVR took significantly longer time for hyperparameter tuning. A similar approach was used in [64], where PSF-based method using kNN produced the best performance (SMAPE 7.85%) in predicting energy consumption of a session. Lee et al. [55] used GMMs to predict session duration and energy needs from a non-residential public charging space in California, USA. By considering the distribution of the known arrival times, the GMM model was used to predict the session duration and the energy consumption. The best model obtained SMAPE of 15.85% for session duration and SMAPE of 14.43% for energy consumption on the public charging (Caltech campus) dataset. Comparison shows that the GMM predictive model achieved significantly greater predictions when compared to user inputs, where the users were asked to estimate their departure times and energy needs. Binary logistic regression was used in [65] to classify whether or not the driver will make use of fast charging in a given day. Features such as travel time duration, driving speed, temperature and whether the driver's last trajectory included fast charging was used to develop the predictive model. The proposed model achieved superior performance compared to LR models with overall prediction accuracy of 0.894. Additionally, the following conclusions were drawn: drivers

are more likely to fast charge with increased travel duration and travel distance, and drivers who exhibit fast charging habits are more likely to use fast charging on their next day trajectories. Venticinque and Nacchia [66] used SVR with RBF kernel to predict the time to next plug. Using residential charging data, the best performing model in this work achieved MAE of 0.124 minutes and RMSE of 0.158 minutes. Frendo et al. [67] used a dataset consisting of charging processes, i.e. timeseries data of charging power, from a workplace to predict charge profiles. The best performance was achieved by XGBoost model with MAE of 126 W outperforming ANN and LR models. The result of this approach when integrated to form scheduling have resulted in up to 21% increase in energy charge for the EV. In [68], LR was used to model charging speed by considering variables such as temperature, connection time and state of charge (SOC). Some variables such as temperature were found to impact charging speed, for instance it was noted that an increase in 1 degree Celsius resulted in an increase of charging speed with 3.7 W. The model was analyzed graphically and statistically and while the charging speed is very relevant in terms of predicting the departure time, the study failed to consider the predictive performance of the model. Lu et al. [69] used RF model to predict charging capacity and the daily charging times. The proposed model outperformed SVR on training set and achieved MAPE of 9.76% on prediction of the charging load for the next 15 minutes for a single station. For a group of charging

stations, proposed RF model once again outperformed both SVR and DT models with 12.8% MAPE. Feature importance analysis showed that previous day's charge was the most important input feature. Table 1 summarizes the research that used supervised learning for EV charging behavior predictions.

IV. UNSUPERVISED AND STATISTICAL LEARNING FOR ANALYSIS AND PREDICTION OF CHARGING BEHAVIOR

Other works have utilized unsupervised learning, mainly clustering techniques to identify EV charging behavior. GMM was used in [70] to find 13 distinct clusters of charging behavior for non-residential charging. Charging sessions containing information about start time, connection duration, distance between two sessions and hours between sessions were considered as features. Distinction between daytime and overnight charging was found to be the largest distinction between the types of charging sessions. For instance, some of the clusters contained users who charge during daytime, within the same charging station with medium charging duration, and narrow medium hours between sessions. The clustering result was evaluated using ARI for clusters 7-13, with all except one (ARI of 0.54) ARI value being below 0.6, and therefore indicating good general agreement. GMM was also used in [71] to create EV profiles that captured charging behavior by considering number of charging events, start time and SOC. The EV charging profiles generated were then validated with average charging demand. Flammini *et al.* [72] used charging transaction data and developed a BMM to represent the multi-modal probability distributions of variables such as connection time and idle time. The proposed model showed good fit when compared to the empirical data graphically and the following conclusions were made after analysis: 25% of the total energy is supplied in the weekend, significant differences were noticed for plug-in and plug-out profiles among weekdays and weekends, 50% of the recharges last for less than 4 hours and the idle time on average lasts for 4 hours. While the results provide good insight into charging behavior, the proposed model was not validated for predictive performance. DBSCAN clustering was used in [73] to find 3 clusters of EV charging behavior based on arrival and departure times. The first cluster, named charge near home, contained sessions with most arrival times during afternoon and evening and most departures the next morning. The second cluster, named charge near work, contained sessions with arrivals in the morning and departures in the evening. The final cluster, named park to charge, contained sessions scattered throughout the day with short idle time (i.e. they charge quick and leave). The authors also provided qualitative and graphical analysis using violin plots to explain charging behavior but failed to provide performance evaluation of the clustering. The work in [74] used K means clustering with Euclidian distance cost function to categorize user charging behavior into 4 groups, using mean and standard deviation of arrival and departure times as well as the Pearson correlation coefficient between

stay duration and energy consumption. The authors did not provide evaluation for clustering, but the labels generated by clustering were then used by ANN to classify user behavior. A similar approach was noted in [75] where K-means clustering was used to find 3 clusters of charging behavior using Euclidean distance measure. The cluster evaluation was not performed but the results were used by K-NN algorithm for classification and the accuracy of classification was 97.9 with area under ROC curve (AUC) value of 0.994. The authors in [66] used k-means clustering with squared Euclidean distance cost function to predict energy/time requirements of next charge using features such as hour of day, day of week and charging time. Using the elbow method, 6 clusters were found, and the reported silhouette score was 0.7. Based on the clustering result, KNN was used to classify future instance of data into these clusters, obtaining precision of 0.5 and recall of 0.47. Gerossier *et al.* [76] used hierarchical clustering with Ward linkage method to understand EV charging behavior of 46 residential EVs. The clustering result indicates 4 group of behaviors which were night and morning chargers (which makes up more than 50% of the sample), evening chargers when people usually return from work, charging sessions scattered throughout the day and late evening chargers. The performance of clustering was not evaluated, but the results were used to forecast the charging load using RF model that achieved comparable results (MAE of 4.9 kW) to the benchmark gradient boosting method. Expectation maximization (EM) algorithm was used in [77] to find 4 clusters of charging behavior. Then, a mixture model was used to predict EV behavior and simulation results showed that as prediction error increases, the cost reduction and savings decreases. In [78], K means clustering was used to find patterns in EV charging profiles of 3 UK counties. To find the optimum number of clusters, Davies–Bouldin evaluation criterion was used. Although plots were used to display cluster centroids and provide a graphical analysis of daily charging demand, the paper did not provide a performance evaluation of the clustering.

Non-parametric statistical estimation such as KDE and DKDE can also be used to model user charging behavior. Khaki *et al.* [79] used DKDE to predict session duration and energy consumption. Using graphical plot of mean estimation deviation (MED) for comparison, the proposed DKDE method was superior compared to GKDE. In [80], GKDE was used to divide charging behavior into charging that ends withing the same day (intra-day) and charging that carries on to the next day (inter-day). Comparing correlation between charging start time, session duration and energy using Pearson correlation coefficient and Kendall rank correlation showed that correlation can only be noticed if the classifications into the 2 groups, *i.e.,* intra-day and inter-day, were made. A Hybrid estimator that uses both GKDE and DKDE was proposed in [81] to predict charging session duration and energy consumption. Comparison of MED shows that accuracy of prediction is better using the hybrid model than the individual models. A time series-based

TABLE 2. Unsupervised and statistical learning for charging behavior.

Source	Charging Behavior	ML Model	Results
[66]	Cluster charging behavior. Based on the clustering result, classify future instance of data into these clusters	K-means clustering with squared Euclidean distance. KNN for classification	6 clusters were found, Silhouette score: 0.7. For classification, Precision: 0.5, Recall: 0.47.
[70]	Find distinct clusters of user charging behavior for non-residential charging	GMM	ARI value above 0.6 for all except 1 cluster
[71]	Create EV profiles that captures charging behavior from existing data.	GMM	-
[72]	Model charging behavior by representing the multi-modal probability distributions of variables such as connection time and idle time using transaction data	BMM	25% of total energy supplied in weekend, significant differences noticed for plug-in and plug-out times for weekdays and weekends, 50% of the recharges last for less than 4 hours and idle time lasts for 4 hours.
[73]	Find clusters of EV charging behavior based on arrival and departure times.	DB clustering	3 distinct clusters found. Cluster evaluation not provided.
[74]	Find cluster of EV charging behavior	K means clustering with Euclidian distance	4 distinct clusters found. Cluster evaluation not provided.
[75]	Find cluster of EV charging behavior to provide labels for each cluster, then use classification for future sessions to predict which cluster it belongs to	K means clustering with Euclidian distance, K-NN for classification	3 clusters found and the cluster evaluation was not provided. Classification accuracy: 97.9 with Area Under ROC Curve (AUC): 0.99
[76]	Find cluster of EV charging behavior of 46 residential EVs	Hierarchical clustering with Ward linkage method	4 clusters found and cluster evaluation was not provided. Result of clustering used to forecast charging load using RF model MAE: 4.9 kW
[77]	Find cluster of EV charging behavior and predict future behavior	EM for Clustering and mixture model to predict future behavior	4 clusters of charging behavior. Simulation result: As prediction error increases, the cost reduction and savings decrease.
[78]	Find patterns in EV charging profiles of 3 UK counties	K means clustering, Davies–Bouldin evaluation criterion used to find number of clusters	5 clusters for Leicestershire county, 6 clusters for Nottinghamshire county and 3 clusters for West Midlands. Cluster evaluation was not provided.
[79]	Predict session duration and energy consumption	Statistical, DKDE model	Using graphical plot MED, the proposed DKDE method was superior compared to GKDE
[80]	Distinguish charging behavior into intra-day and inter-day charging	Statistical, GKDE model	Correlation comparisons show that correlation can only be noticed if the classifications are made into the 2 groups
[81]	Predict charging session duration and energy consumption	Hybrid estimator using GKDE and DKDE	Stay duration MED: 0.75 hour, Energy consumption MED: 1.68 kWh
[82]	Predict the parking lot charging demand using expected arrival and departure times	Time series-based forecasting, ARIMA	Charging load MAPE: 1.44%. Potential saving $770k yearly for 6-bus system & $240M for IEEE-24 bus system if proposed method is used to forecast charging load.

forecasting method, ARIMA, has been used in [82] to predict the parking lot charging demand using expected arrival and departure times. The proposed model which decouples the daily charging demand of EV parking area from the seasonally changing load profile outperformed regular ARIMA achieving MAPE of 1.44%. The author claimed a potential saving of $770k annually for the 6-bus system and $240M for the IEEE-24 bus system if the proposed method is used to forecast charging load profiles. Table 2 summarizes the research that used unsupervised and statistical learning for charging behavior predictions and analysis.

V. DEEP LEARNING FOR ANALYSIS AND PREDICTION OF CHARGING BEHAVIOR

DL models have been extensively and successfully used in past studies for electric load forecasting [83]–[85]. Load forecasting is crucial for energy management and operation of electric utilities. In the context of EV charging, short term charging load forecasting is particularly important for smart scheduling. Analysis and prediction of short-term

charging load can be considered as an implicit analysis of charging behavior because it depends to a large extent on the arrival and departure time of EV drivers. Zhu *et al.* [86] used multiple RNN-based models to predict the hourly charging load of a public charging station in China. Several DL-based models including RNN, LSTM and gated recurrent units (GRU) were used for prediction. GRU model using 1 hidden layer provided the best performance with normalized RMSE (NRMSE) of 2.89%. Other studies [87] have looked at even higher resolution of time series data, and provided EV charging load forecasting for super short term (minutely data). A comparative study of various DL models showed that LSTM performs better than conventional ANNs by reducing the forecasting error by more than 30%. LSTM model obtained the best performance with MAE of 0.29 kW and RMSE of 0.44 kW. Using graphical and qualitative analysis of the same dataset, certain conclusions were made about charging behavior. Firstly, for working days during rainy season, the charging load peaks at 11 pm, starts to decrease at 3 pm and then increases again at 8 am. Holidays

also have low charging load during the day as more drivers do not travel on holidays.

A LSTM-based model was also used in [88] to forecast charging load for 30 and 15 minutes intervals. The proposed model outperformed ANN model with RMSE of 12.3 kWh and MAE of 5.5 kWh for 15-minute interval, and RMSE of 28.3 kWh and MAE of 16.9 kWh for 30-minute interval. It was also noted that as the time interval decreases, the prediction error decreases because there are more data points available for training. Zhang *et al.* [89] used convolutional neural network (CNN) to estimate traffic flow and mixture model to estimate arrival rates. The results were then used by a queuing model to forecast EV charging load obtaining MAPE of 3.21%.

Deep generative method is used in [90] to classify charging profiles of EV. A charging profile can simply be considered as the distribution of charging arrival and departure times of EVs. The proposed method achieved better performance than benchmark hidden markov model (HMM) by obtaining overall accuracy of 0.98 and F1 score of 0.8. In [74], the authors first used clustering technique and hand-labelling from data visualization to obtain labels of charging behavior. ANN consisting of 3 hidden layers (optimal after hyperparameter tuning) was then trained using backpropagation to make predictions on a class of EV charging behavior. The model achieved average accuracy of 78% on test set. Table 3 summarizes the research that used deep learning for charging behavior predictions.

TABLE 3. Deep learning for charging behavior.

Source	Approach	DL Model	Results
[86]	Predict the hourly charging load of a public charging station in China	RNN-based models	Best result using GRU NRMSE: 2.89%.
[87]	Provide EV charging load forecasting for super short term (minutely data)	DL models including ANN, LSTM, Autoencoders	Best result using LSTM MAE: 0.29 kW, RMSE: 0.44 kW
[88]	Forecast charging load for 30-minutes and 15-minutes intervals	LSTM	15-minute data RMSE: 12.3 kWh, MAE: 5.5 kWh 30-minute data RMSE: 28.3 kWh, MAE: 16.9 kWh
[89]	Estimate traffic flow and consequently estimate EV arrival rates.	Convolutional Neural Network (CNN)	Forecast of EV charging load-MAPE: 3.21%
[90]	Classify charging profiles of EV	Deep generative model	Accuracy: 0.98, F1-score: 0.8
[74]	Predict charging behavior where the labels were obtained using clustering	ANN	Accuracy: 78%

VI. RECOMMENDATIONS AND FUTURE WORK

There are several challenges in using ML for EV charging behavior predictions. Firstly, as described in section II G,

there is a lack of datasets required to train ML models. More specifically, there are only two well-known EV charging dataset that are available publicly for researchers. Many other datasets are owned by commercial companies. Also, the available datasets only represent the charging behavior of the geographical area they were recorded in. Therefore, ML models developed using these datasets will not apply well to every part of the world. EVs are getting popular across the globe and therefore more data needs to be collected for major cities that can help to characterize charging behavior of specific cities. Initiatives must be taken by researchers to provide the datasets used by their studies and make them online for further research and validation. Furthermore, only few of the reviewed works integrated the predictive models for smart scheduling. Scheduling remains a key area in the management of EV charging and both long-term (day ahead, weak ahead) and short-term (minute ahead, hour ahead) ML predictions and their implications on scheduling need to be considered. Most of the previous works considered the use of historical data with few variables such as arrival time, departure time and energy consumption for training ML models. To better categorize and predict charging behavior, higher dimensional data that consider variables like traffic, weather and local area events need to be utilized. For instance, the usage of weather data could indicate the charging behavior in the incident of heavy rainfall or snow which would otherwise not be considered by the models. Furthermore, a comprehensive cluster analysis that provides a comparative study and evaluation of all clustering techniques is needed. Cluster analysis provided excellent characterization of consumer behavior in domains like coffee shops [91] and therefore has the potential to better characterize charging behavior. In section 4, it was noted that a lot of studies that used clustering failed to evaluate the performance of the cluster algorithms quantitatively and this needs to be addressed in future works. Besides supervised and unsupervised learning, reinforcement learning (RL) [92] is the third category of ML, that allows the model to learn a behavior through trial-and-error interactions with the environment through notions of reward and punishment. There are a few recent works utilizing RL-based concepts for scheduling of EV charging [93]–[96] and therefore has great potential for further research. In summary, the following areas can be explored for future research:

- Utilization of short-term and long-term ML predictions into scheduling.
- The use of high dimensional dataset with input features such as traffic and weather conditions for model training.
- A comprehensive cluster analysis of EV charging dataset.
- Reinforcement learning for EV scheduling.

VII. CONCLUSION

This article provided a comprehensive survey on the use of machine learning for EV charging behavior analysis and prediction. The common supervised and unsupervised ML

algorithms used for prediction of EV charging behavior were defined. The survey provided a comparative analysis of various works using supervised, unsupervised, and deep learning methods for EV charging prediction. The key challenges were also discussed which include the lack of public charging datasets and the lack of high dimensional data. Recommendations for future work based on the existing work were also provided. The research directions that this article provide include the need for a comprehensive cluster analysis and the use of reinforcement learning for EV scheduling.

ACKNOWLEDGMENT

REFERENCES

[1] *Global indicator framework for the Sustainable Development Goals and Targets of the 2030 Agenda for Sustainable Development*. New York, NY, USA: U. G. Assembly, 2017.

[2] *Key World Energy Statistics 2018–Analysis*. IEA. Accessed: Jun. 1, 2020. [Online]. Available: https://www.iea.org/reports/key-world-energy-statistics-2019

[3] (May 16, 2018). *68% of the World Population Projected to Live in Urban Areas by 2050, Says UN, UN DESA | United Nations Department of Economic and Social Affairs*. Accessed: Jun. 1, 2020. [Online]. Available: https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html

[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] *Global EV Outlook 2019–Analysis*. IEA. Accessed: Jun. 1, 2020. [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2019

[6] A. Ramanujam, P. Sankaranarayanan, A. Vasan, R. Jayaprakash, V. Sarangan, and A. Sivasubramaniam, "Qantifying the impact of electric vehicles on the electric grid: A simulation based case-study," in *Proc. 8th Int. Conf. Future Energy Syst.*, New York, NY, USA, May 2017, pp. 228–233, doi: 10.1145/3077839.3077854.

[7] M. D. Galus, M. Zima, and G. Andersson, "On integration of plug-in hybrid electric vehicles into existing power system structures," *Energy Policy*, vol. 38, no. 11, pp. 6736–6745, Nov. 2010, doi: 10.1016/j.enpol.2010.06.043.

[8] K. Clement-Nyns, E. Haesen, and J. Driesen, "The impact of charging plug-in hybrid electric vehicles on a residential distribution grid," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 371–380, Feb. 2010.

[9] Z. Moghaddam, I. Ahmad, D. Habibi, and Q. V. Phung, "Smart charging strategy for electric vehicle charging stations," *IEEE Trans. Transp. Electrific.*, vol. 4, no. 1, pp. 76–88, Mar. 2018.

[10] M. S. Kuran, A. Carneiro Viana, L. Iannone, D. Kofman, G. Mermoud, and J. P. Vasseur, "A smart parking lot management system for scheduling the recharging of electric vehicles," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2942–2953, Nov. 2015.

[11] J. García-Álvarez, M. A. González, and C. R. Vela, "Metaheuristics for solving a real-world electric vehicle charging scheduling problem," *Appl. Soft Comput.*, vol. 65, pp. 292–306, Apr. 2018, doi: 10.1016/j.asoc.2018.01.010.

[12] A. Foley, B. Tyther, P. Calnan, and B. Ó Gallachóir, "Impacts of electric vehicle charging under electricity market operations," *Appl. Energy*, vol. 101, pp. 93–102, Jan. 2013, doi: 10.1016/j.apenergy.2012.06.052.

[13] T. Franke and J. F. Krems, "Understanding charging behaviour of electric vehicle users," *Transp. Res. F: Traffic Psychol. Behav.*, vol. 21, pp. 75–89, Nov. 2013, doi: 10.1016/j.trf.2013.09.002.

[14] L. Hu, J. Dong, and Z. Lin, "Modeling charging behavior of battery electric vehicle drivers: A cumulative prospect theory based approach," *Transp. Res. C, Emerg. Technol.*, vol. 102, pp. 474–489, May 2019, doi: 10.1016/j.trc.2019.03.027.

[15] J. Spoelstra, "Charging behaviour of Dutch EV drivers," M.S. thesis, Dept. Sci. Innov. Manage., Utrecht Univ., Utrecht, The Netherlands, 2014.

[16] R. Bi, J. Xiao, V. Viswanathan, and A. Knoll, "Influence of charging behaviour given charging station placement at existing petrol stations and residential car park locations in Singapore," *Procedia Comput. Sci.*, vol. 80, pp. 335–344, 2016.

[17] P. Liu and J. Yu, "Charging behavior characteristic simulation of plug-in electric vehicles for demand response," in *Proc. UKACC 11th Int. Conf. Control (CONTROL)*, Aug. 2016, pp. 1–6.

[18] I. Lampropoulos, G. M. A. Vanalme, and W. L. Kling, "A methodology for modeling the behavior of electricity prosumers within the smart grid," in *Proc. IEEE PES Innov. Smart Grid Technol. Conf. Eur. (ISGT Europe)*, Oct. 2010, pp. 1–8.

[19] J. Munkhammar, J. Widen, P. Grahn, and J. Ryden, "A Bernoulli distribution model for plug-in electric vehicle charging based on time-use data for driving patterns," in *Proc. IEEE Int. Electric Vehicle Conf. (IEVC)*, Dec. 2014, pp. 1–7.

[20] P. Grahn, K. Alvehag, and L. Soder, "PHEV utilization model considering Type-of-Trip and recharging flexibility," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 139–148, Jan. 2014.

[21] J. Soares, N. Borges, M. A. Fotouhi Ghazvini, Z. Vale, and P. B. de Moura Oliveira, "Scenario generation for electric vehicles' uncertain behavior in a smart city environment," *Energy*, vol. 111, pp. 664–675, Sep. 2016, doi: 10.1016/j.energy.2016.06.011.

[22] J. Brady and M. O'Mahony, "Modelling charging profiles of electric vehicles based on real-world electric vehicle charging data," *Sustain. Cities Soc.*, vol. 26, pp. 203–216, Oct. 2016, doi: 10.1016/j.scs.2016.06.014.

[23] O. Frendo, S. Karnouskos, N. Gaertner, O. Kipouridis, K. Rehman, and N. Verzano, "Charging strategies and implications for corporate electric vehicle fleets," in *Proc. IEEE 16th Int. Conf. Ind. Informat. (INDIN)*, Jul. 2018, pp. 466–471.

[24] 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, doi: 10.1016/j.apenergy.2016.08.080.

[25] Y. Cao, H. Song, O. Kaiwartya, B. Zhou, Y. Zhuang, Y. Cao, and X. Zhang, "Mobile edge computing for Big-Data-Enabled electric vehicle charging," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 150–156, Mar. 2018.

[26] J. Lee, G.-L. Park, Y. Han, and S. Yoo, "Big data analysis for an electric vehicle charging infrastructure using open data and software," in *Proc. 8th Int. Conf. Future Energy Syst.*, New York, NY, USA, May 2017, pp. 252–253, doi: 10.1145/3077839.3081670.

[27] N. K. Ahmed, A. F. Atiya, N. E. Gayar, and H. El-Shishiny, "An empirical comparison of machine learning models for time series forecasting," *Econ. Rev.*, vol. 29, nos. 5–6, pp. 594–621, Aug. 2010.

[28] Z.-J.-M. Shen, B. Feng, C. Mao, and L. Ran, "Optimization models for electric vehicle service operations: A literature review," *Transp. Res. B, Methodol.*, vol. 128, pp. 462–477, Oct. 2019, doi: 10.1016/j.trb.2019.08.006.

[29] A. Ahmad, Z. A. Khan, M. Saad Alam, and S. Khateeb, "A review of the electric vehicle charging techniques, standards, progression and evolution of EV technologies in Germany," *Smart Sci.*, vol. 6, no. 1, pp. 36–53, Jan. 2018, doi: 10.1080/23080477.2017.1420132.

[30] Y. Zhang, X. Liu, T. Zhang, and Z. Gu, "Review of the electric vehicle charging station location problem," in *Dependability in Sensor, Cloud, and Big Data Systems and Applications*. Singapore: Springer, 2019, pp. 435–445.

[31] A. S. Al-Ogaili, T. J. Tengku Hashim, N. A. Rahmat, A. K. Ramasamy, M. B. Marsadek, M. Faisal, and M. A. Hannan, "Review on scheduling, clustering, and forecasting strategies for controlling electric vehicle charging: Challenges and recommendations," *IEEE Access*, vol. 7, pp. 128353–128371, 2019, doi: 10.1109/ACCESS.2019.2939595.

[32] K. Morrow, D. Karner, and J. Francfort, "Plug-in hybrid electric vehicle charging infrastructure review," U.S. Dept. Energy, US Dept. Energy-Vehicle Technol. Program, Contract No. 58517, 2008, vol. 34.

[33] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*. [Online]. Available: http://arxiv.org/abs/1609.04747

[34] D. M. Hawkins, "The problem of overfitting," *J. Chem. Inf. Comput. Sci.*, vol. 44, no. 1, pp. 1–12, Jan. 2004, doi: 10.1021/ci0342472.

[35] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*. Boca Raton, FL, USA: CRC Press, 1984.

[36] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[37] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.*, vol. 6, no. 3, pp. 21–45, 3rd Quart., 2006.

[38] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[39] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, Jun. 1998, doi: 10.1023/A:1009715923555.

[40] M. Awad and R. Khanna, "Support Vector Regression," in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, M. Awad R. Khanna, Eds. Berkeley, CA, USA: Apress, 2015, pp. 67–80.

[41] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.

[42] T. M. Kodinariya and P. R. Makwana, "Review on determining number of cluster in K-means clustering," *Int. J.*, vol. 1, no. 6, pp. 90–95, 2013.

[43] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, and others, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. KDD*, 1996, vol. 96, no. 34, pp. 226–231.

[44] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, Sep. 1967, doi: 10.1007/BF02289588.

[45] C. E. Rasmussen, "The infinite Gaussian mixture model," in *Proc. Adv. Neural Inf. Process. Syst.*, 2000, pp. 554–560.

[46] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," *Ann. Statist.*, vol. 38, no. 5, pp. 2916–2957, Oct. 2010.

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

[48] A. Graves, "Long short-term memory," in *Supervised Sequence Labelling With Recurrent Neural Networks*, A. Graves, Ed. Berlin, Germany: Springer, 2012, pp. 37–45.

[49] Y.-W. Chung, B. Khaki, T. Li, C. Chu, and R. Gadh, "Ensemble machine learning-based algorithm for electric vehicle user behavior prediction," *Appl. Energy*, vol. 254, Nov. 2019, Art. no. 113732, doi: 10.1016/j.apenergy.2019.113732.

[50] E. Rendón, I. Abundez, A. Arizmendi, and E. M. Quiroz, "Internal versus external cluster validation indexes," *Int. J. Comput. Commun.*, vol. 5, no. 1, pp. 27–34, 2011.

[51] J. M. Santos and M. Embrechts, "On the use of the adjusted rand index as a metric for evaluating supervised classification," in *Artificial Neural Networks (ICANN)*. Berlin, Germany: Springer-Verlag, 2009, pp. 175–184.

[52] C. Develder, N. Sadeghianpourhamami, M. Strobbe, and N. Refa, "Quantifying flexibility in EV charging as DR potential: Analysis of two real-world data sets," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Nov. 2016, pp. 600–605.

[53] J. D. Cross and R. Hartshorn, "My Electric Avenue: Integrating electric vehicles into the electrical networks," in *Proc. IET Conf.*, Jan. 2016, pp. 2–12.

[54] *Pecan Street Data*. Pecan Street Inc. Accessed: Jun. 12, 2020. [Online]. Available: https://www.pecanstreet.org/dataport/about/

[55] Z. J. Lee, T. Li, and S. H. Low, "ACN-data: Analysis and applications of an open EV charging dataset," in *Proc. 10th ACM Int. Conf. Future Energy Syst.*, New York, NY, USA, Jun. 2019, pp. 139–149, doi: 10.1145/3307772.3328313.

[56] O. Frendo, N. Gaertner, and H. Stuckenschmidt, "Improving smart charging prioritization by predicting electric vehicle departure time," *IEEE Trans. Intell. Transp. Syst.*, early access, Apr. 30, 2020, doi: 10.1109/TITS.2020.2988648.

[57] 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.

[58] Z. Xu, "Forecasting electric vehicle arrival & departure time on UCSD campus using support vector machines," Ph.D. dissertation, Dept. Eng. Sci., Univ. California, San Diego, CA, USA, 2017.

[59] K. Duan, S. S. Keerthi, and A. N. Poo, "Evaluation of simple performance measures for tuning SVM hyperparameters," *Neurocomputing*, vol. 51, pp. 41–59, Apr. 2003, doi: 10.1016/S0925-2312(02)00601-X.

[60] S. Ai, A. Chakravorty, and C. Rong, "Household EV charging demand prediction using machine and ensemble learning," in *Proc. IEEE Int. Conf. Energy Internet (ICEI)*, May 2018, pp. 163–168.

[61] 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.

[62] 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.

[63] N. Bokde, M. W. Beck, F. Martínez Álvarez, and K. Kulat, "A novel imputation methodology for time series based on pattern sequence forecasting," *Pattern Recognit. Lett.*, vol. 116, pp. 88–96, Dec. 2018, doi: 10.1016/j.patrec.2018.09.020.

[64] M. Majidpour, "Time series prediction for electric vehicle charging load and solar power generation in the context of smart grid," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. California, Los Angeles, CA, USA, 2016.

[65] Y. Yang, Z. Tan, and Y. Ren, "Research on factors that influence the fast charging behavior of private battery electric vehicles," *Sustainability*, vol. 12, no. 8, p. 3439, Apr. 2020.

[66] S. Venticinque and S. Nacchia, "Learning and prediction of E-car charging requirements for flexible loads shifting," in *Internet and Distributed Computing Systems*. Cham, Switzerland: Springer Nature, 2019, pp. 284–293.

[67] O. Frendo, J. Graf, N. Gaertner, and H. Stuckenschmidt, "Data-driven smart charging for heterogeneous electric vehicle fleets," *Energy AI*, vol. 1, Aug. 2020, Art. no. 100007, doi: 10.1016/j.egyai.2020.100007.

[68] J. Mies, J. Helmus, and R. van den Hoed, "Estimating the charging profile of individual charge sessions of electric vehicles in The Netherlands," *World Electr. Vehicle J.*, vol. 9, no. 2, p. 17, Jun. 2018.

[69] Y. Lu, Y. Li, D. Xie, E. Wei, X. Bao, H. Chen, and X. Zhong, "The application of improved random forest algorithm on the prediction of electric vehicle charging load," *Energies*, vol. 11, no. 11, p. 3207, Nov. 2018.

[70] J. R. Helmus, M. H. Lees, and R. van den Hoed, "A data driven typology of electric vehicle user types and charging sessions," *Transp. Res. C, Emerg. Technol.*, vol. 115, Jun. 2020, Art. no. 102637, doi: 10.1016/j.trc.2020.102637.

[71] J. Quirós-Tortós, A. N. Espinosa, L. F. Ochoa, and T. Butler, "Statistical representation of EV charging: Real data analysis and applications," in *Proc. Power Syst. Comput. Conf. (PSCC)*, Jun. 2018, pp. 1–7.

[72] M. G. Flammini, G. Prettico, A. Julea, G. Fulli, A. Mazza, and G. Chicco, "Statistical characterisation of the real transaction data gathered from electric vehicle charging stations," *Electr. Power Syst. Res.*, vol. 166, pp. 136–150, Jan. 2019, doi: 10.1016/j.epsr.2018.09.022.

[73] N. Sadeghianpourhamami, N. Refa, M. Strobbe, and C. Develder, "Quantitive analysis of electric vehicle flexibility: A data-driven approach," *Int. J. Electr. Power Energy Syst.*, vol. 95, pp. 451–462, Feb. 2018, doi: 10.1016/j.ijepes.2017.09.007.

[74] Y. Xiong, B. Wang, C.-C. Chu, and R. Gadh, "Electric vehicle driver clustering using statistical model and machine learning," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2018, pp. 1–5.

[75] Y. Shen, W. Fang, F. Ye, and M. Kadoch, "EV charging behavior analysis using hybrid intelligence for 5G smart grid," *Electronics*, vol. 9, no. 1, p. 80, Jan. 2020.

[76] A. Gerossier, R. Girard, and G. Kariniotakis, "Modeling and forecasting electric vehicle consumption profiles," *Energies*, vol. 12, no. 7, p. 1341, Apr. 2019, doi: 10.3390/en12071341.

[77] Y. Xiong, B. Wang, C.-C. Chu, and R. Gadh, "Vehicle grid integration for demand response with mixture user model and decentralized optimization," *Appl. Energy*, vol. 231, pp. 481–493, Dec. 2018, doi: 10.1016/j.apenergy.2018.09.139.

[78] E. Xydas, C. Marmaras, L. M. Cipcigan, N. Jenkins, S. Carroll, and M. Barker, "A data-driven approach for characterising the charging demand of electric vehicles: A UK case study," *Appl. Energy*, vol. 162, pp. 763–771, Jan. 2016, doi: 10.1016/j.apenergy.2015.10.151.

[79] B. Khaki, Y.-W. Chung, C. Chu, and R. Gadh, "Nonparametric user behavior prediction for distributed EV charging scheduling," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2018, pp. 1–5.

[80] Z. Chen, Z. Zhang, J. Zhao, B. Wu, and X. Huang, "An analysis of the charging characteristics of electric vehicles based on measured data and its application," *IEEE Access*, vol. 6, pp. 24475–24487, 2018.

[81] 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.

[82] 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, doi: 10.1016/j.epsr.2016.06.003.

[83] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, Feb. 2001.

[84] K. Metaxiotis, A. Kagiannas, D. Askounis, and J. Psarras, "Artificial intelligence in short term electric load forecasting: A state-of-the-art survey for the researcher," *Energy Convers. Manage.*, vol. 44, no. 9, pp. 1525–1534, 2003, doi: 10.1016/S0196-8904(02)00148-6.

[85] K. Amarasinghe, D. L. Marino, and M. Manic, "Deep neural networks for energy load forecasting," in *Proc. IEEE 26th Int. Symp. Ind. Electron. (ISIE)*, Jun. 2017, pp. 1483–1488.

[86] J. Zhu, Z. Yang, Y. Guo, J. Zhang, and H. Yang, "Short-term load forecasting for electric vehicle charging stations based on deep learning approaches," *Appl. Sci.*, vol. 9, no. 9, p. 1723, Apr. 2019.

[87] 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, doi: 10.3390/en12142692.

[88] 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.

[89] 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.*, early access, Apr. 2, 2020, doi: 10.1109/TCYB.2020.2975134.

[90] S. Wang, L. Du, J. Ye, and D. Zhao, "A deep generative model for non-intrusive identification of EV charging profiles," *IEEE Trans. Smart Grid*, early access, May 27, 2020, doi: 10.1109/TSG.2020.2998080.

[91] I. A. Zualkernan, M. Pasquier, S. Shahriar, M. Towheed, and S. Sujith, "Using BLE beacons and machine learning for personalized customer experience in smart Cafés," in *Proc. Int. Conf. Electron., Inf., Commun. (ICEIC)*, Jan. 2020, pp. 1–6.

[92] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, no. 1, pp. 237–285, Jan. 1996.

[93] Z. Wan, H. Li, H. He, and D. Prokhorov, "Model-free real-time EV charging scheduling based on deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5246–5257, Sep. 2019.

[94] N. Mhaisen, N. Fetais, and A. Massoud, "Real-time scheduling for electric vehicles Charging/Discharging using reinforcement learning," in *Proc. IEEE Int. Conf. Informat., IoT, Enabling Technol. (ICIoT)*, Feb. 2020, pp. 1–6.

[95] J. Lee, E. Lee, and J. Kim, "Electric vehicle charging and discharging algorithm based on reinforcement learning with data-driven approach in dynamic pricing scheme," *Energies*, vol. 13, no. 8, p. 1950, Apr. 2020.

[96] H. Li, Z. Wan, and H. He, "Constrained EV charging scheduling based on safe deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2427–2439, May 2020.

A. R. AL-ALI (Life Senior Member, IEEE) received the B.Sc. degree in electrical engineering (EE) from Aleppo University, Syria, in 1979, the M.S. degree from the Polytechnic Institute of New York University, USA, in 1986, and the Ph.D. degree in electrical engineering and a minor in computer science from Vanderbilt University, Nashville, TN, USA, in 1990. From 1991 to 2000, he was a Faculty Member with the Department of EE, King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia. Since 2000, he has been working as a Professor of computer engineering with the American University of Sharjah, United Arab Emirates. His research interests include embedded systems, cyber physical systems, and the Internet of Things (IoT) and IIoT applications in smart cities.

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.

SALAM DHOU (Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science from the Jordan University of Science and Technology, Irbid, Jordan, in 2004 and 2007, respectively, and the Ph.D. degree in electrical and computer engineering from Virginia Commonwealth University, Richmond, VA, USA, in 2013. She has worked as a Postdoctoral Research Fellow with the Department of Radiation Oncology, Harvard Medical School, Boston, MA, USA, between 2013 and 2016. She is currently an Assistant Professor with the Department of Computer Science and Engineering and the Biomedical Engineering Graduate Program, American University of Sharjah, Sharjah, United Arab Emirates. Her research interests include machine learning and data mining, computer vision, and medical imaging and informatics.

SAKIB SHAHRIAR received the B.S. degree in computer engineering from the American University of Sharjah, United Arab Emirates, in 2018, where he is currently pursuing the M.S. degree in computer engineering. He is also a Research Assistant with the American University of Sharjah. His research interests include machine learning, big data analytics, and deep learning.

MAIS NIJIM (Member, IEEE) received the B.S. degree in computer science from Princess Sumaya University for Technology, Amman, Jordan, the M.S. degree in computer science from New Mexico State University, in 2004, and the Ph.D. degree in computer science from the New Mexico Institute of Mining and Technology, in 2007. She has been an Associate Professor with the Department of Electrical Engineering and Computer Science, Texas A&M University-Kingsville. Her research interests include machine learning, cyber physical systems, cybersecurity, and wireless sensor networks. She is an Editor of the *International Journal of Sensor Networks*.

• • •