Electric Vehicle Charging Demand Prediction Using a Novel Machine Learning-Based Technique

3 authors:

Omer Can Tolun
Adana Science and Technology University

Kasım Zor
Adana Alparslan Türkeş Science and Technology University

Onder Tutsoy
Adana Science and Technology University

SDEWES2023.0175

Electric Vehicle Charging Demand Prediction Using a Novel Machine Learning-Based Technique

Ö.C. Tolun*, K. Zor, Ö. Tutsoy

Adana Alparslan Türkeş Science and Technology University, Turkey
(*octolun@atu.edu.tr)

Abstract

Recently, global warming has changed the course of fuel type in vehicles from fossil fuels into the use of electricity. The popularity of electric vehicles (EVs) has been ascending day by day due to fact that they have several advantages such as being environmentally friendly, having lower maintenance requirement, and better mileage performance in city driving over conventional vehicles. Despite the merits of the EVs, there are also a few disadvantages consisting of the integration of the EVs into the existing infrastructure and their expensiveness by means of initial investment cost. In addition to those, machine learning- (ML) based techniques are usually employed in the EVs for battery management system, drive performance, and passenger safety. This paper aims to implement an EV monthly charging demand prediction by using a novel technique based on an ensemble of Pearson correlation (PC) and analysis of variance (ANOVA) along with gated recurrent unit (GRU) networks and eXtreme Gradient Boosting (XGBoost) decision trees for the Eastern Mediterranean Region of Türkiye with respect to performance and error measures including R-squared, mean absolute percentage error (MAPE), and mean absolute error (MAE) in a benchmarking manner. According to the obtained results, a hybrid of ANOVA and XGBoost model outperformed the other models in terms of accuracy and computational time. To the best of one's knowledge, there is a significant gap in EVs charging demand prediction and this study will bridge the gap in the literature.

Electric Vehicle Charging Demand Prediction Using a Novel Machine Learning-Based Technique

Ömer Can Tolun[1*], Kasım Zor[2], Onder Tutsoy[3]
Department of Electrical and Electronic Engineering
Adana Alparslan Türkeş Science and Technology University, Adana, Turkey
e–mail: {octolun, kzor, otutsoy}@atu.edu.tr

ABSTRACT

Recently, global warming has changed the course of fuel type in vehicles from fossil fuels into the use of electricity. The popularity of electric vehicles (EVs) has been ascending day by day due to fact that they have several advantages such as being environmentally friendly, having lower maintenance requirement, and better mileage performance in city driving over conventional vehicles. Despite the merits of the EVs, there are also a few disadvantages consisting of the integration of the EVs into the existing infrastructure and their expensiveness by means of initial investment cost. In addition to those, machine learning- (ML) based techniques are usually employed in the EVs for battery management system, drive performance, and passenger safety. This paper aims to implement an EV monthly charging demand prediction by using a novel technique based on an ensemble of Pearson correlation (PC) and analysis of variance (ANOVA) along with gated recurrent unit (GRU) networks and eXtreme Gradient Boosting (XGBoost) decision trees for the Eastern Mediterranean Region of Türkiye with respect to performance and error measures including R-squared, mean absolute percentage error (MAPE), and mean absolute error (MAE) in a benchmarking manner. According to the obtained results, a hybrid of ANOVA and XGBoost model outperformed the other models in terms of accuracy and computational time. To the best of one's knowledge, there is a significant gap in EVs charging demand prediction and this study will bridge the gap in the literature.

KEYWORDS

Electric Vehicle, Charging, Demand, Prediction, Machine-Learning, Gated Recurrent Unit, eXtreme Gradient Boosting decision trees.

INTRODUCTION

Global warming assigns to the gradual increase in the Earth's average surface temperature, primarily caused by the build-up of greenhouse gases, such as carbon dioxide, in the atmosphere. Mitigating the effects of global warming requires global action to reduce greenhouse gas emissions, promote renewable energy sources, and invest in sustainable practices. Therefore, EVs are becoming increasingly popular because they offer numerous advantages, such as being environmentally friendly, requiring less maintenance, and providing better mileage performance in city driving compared to internal combustion engine (ICE) driven vehicles [1]. EVs offer various benefits; however, they also have some drawbacks, such as extended charging time, battery degradation, limited range, and the restricted availability of charging infrastructure. In addition, it is essential to consider alterations to energy demand patterns and potential impacts on electricity system operations [2]. The power system could experience significant effects due to a substantial increase in EV usage. The impact on power

* Corresponding author

plant dispatch, system peak load, and carbon emissions will be influenced by the charging mode of EVs as well as the power plant fleet [3].

The relation between EVs and ML can be described as a symbiotic one, where ML techniques are used to optimize the performance of EVs and to overcome some of the challenges associated with their use. One of the primary areas where ML is used in EVs is battery management [4]. ML algorithms can be used to monitor and analyse data from the battery to predict its remaining capacity, identify potential faults, and optimize its performance. This allows the EV to operate more efficiently and extend the battery's lifespan. Another area where ML is used in EVs is in the development of autonomous driving capabilities [5]. ML algorithms are used to train the vehicle to recognize and respond to different road conditions, traffic patterns, and hazards, making driving safer and more efficient. In addition, ML algorithms can be used to optimize the energy consumption of EVs by predicting traffic patterns and adjusting the vehicle's speed and route to minimize energy consumption [6]. This can reduce the overall carbon footprint of the EV and make it more environmentally friendly. Finally, ML algorithms can be used to predict when parts will need to be replaced in the EV, reducing downtime and maintenance costs [7]. Overall, the relationship between EVs and ML is a mutually beneficial one, where ML techniques are used to optimize the performance of EVs and overcome some of the challenges associated with their use, making them more efficient, safer, and cost-effective. In addition to the ML algorithms, statistical methods can also be used to analyse and optimize various aspects of EVs, including battery management, performance analysis, charging optimization, range estimation, and customer segmentation.

In the literature, there have been a number of studies undertaken to investigate the issue of predicting the charging demand of the EVs. Yi et al. proposed a study containing time-series forecasting for the demand of commercial EV charging by utilizing a deep learning approach-Sequence to Sequence (Seq2Seq). The results indicated that long short-term memory (LSTM) and Seq2seq produce adequate prediction performance for one-step ahead [8]. Sumanasena et al. intended to create a designing and evaluating for a comprehensive artificial intelligence (AI) framework that provides solutions for the deficiencies in electrical vehicle infrastructure (EVI). This AI framework's contribution to addressing the emerging challenges of distributed energy resources in EV adoption has been validated by the empirical evaluation conducted on a real-world EVI case study [9]. Amini et al. presented a methodology for forecasting electricity demand based on historical data provided by EV parking lots. For forecasting medium-term demand, an auto-regressive integrated moving average (ARIMA) model is utilized. Being founded on the simulation results, it appeared that the proposed method for forecasting electricity demand has a high degree of accuracy [10]. Kim and Kim compared different methods for forecasting electricity consumption due to the rise in popularity of EVs. The paper presents a comparison of various modelling techniques, including trigonometric exponential smoothing state space, ARIMA, ANN, and LSTM modelling [11]. A great deal of research has been conducted on EVs, since they are more environmentally friendly and efficient than ICE driven vehicles.

Hu et al. proposed a method which was consisting deep-learning for predicting the demand for electric vehicle charging in the short term, which predicts 5 minutes in advance the future charging demand for a charging station. As a result of the paradigm of machine theory of mind, the proposed model takes into consideration specific historical charging habits and current charging demand variations. Two case studies on real EV charging demand datasets confirm the model's superiority over existing models [12]. Arias and Bae studied big data analytics to estimate the demand for electric vehicle charging. An analysis of cluster has been used to categorize traffic patterns and to identify influential factors a relational analysis has been utilized. Furthermore, a decision tree has been used to determine which factors should be considered for classification. Essentially, the proposed electric vehicle charging demand model could provide a basis for future research on how charging electric vehicles affects the electric grid [13]. Xydas et al. proposes a short-term load forecast model utilizing an AI technique and support vector machines. By comparing the method with Monte Carlo forecasting, the accuracy of the method is evaluated [14]. Moon et al. conducted a study that examined the variations in EV charging demand depending on the preferences of consumer for EVs, charging times, and types of electric vehicle supply equipment and sheds light on the important aspects to consider when building the infrastructure for EVs. The results showed that consumers prefer charging in the evenings. According to the study, policy makers should consider key political implications for taking pre-emptive measures before making significant changes to electricity supply infrastructure [15]. This paper proposes a ensemble algorithm consisting of the statistical and ML techniques to predict the charging demand of the EVs. The key contributions of the paper can be summarized as:

- The dataset that is utilized in this paper is attained by the authors via PlugShare and the Energy Market Regulatory Authority (EMRA).
- PC and ANOVA have been selected as feature selection methods that are in the class of statistical methods. EV charging demand prediction has been performed for the determined region by hybridizing GRU networks and XGBoost separately with these feature selection methods.
- The outputs obtained after this hybridization process have been compared with the auto-seasonal regressive integrated moving average exogenous (SARIMAX) method in the literature for the purpose of performance evaluation.

Following the introduction including a brief review of the literature, the paper is organized in the following manner: The material and methods section provides information on the data set information, as well as the fundamentals of PC and ANOVA used for feature selection, and the foundations of GRU networks and XGBoost. The results and discussion section describes the performance and results of the GRU networks and XGBoost and compares their outcomes. Finally, the conclusion section summarizes the findings and outlines future perspectives.

MATERIAL AND METHODS

In this section, data set information, the bases of the feature selection methods and the fundamentals of the prediction of charging demand algorithms are introduced.

Material

One of the main challenges facing EVs is their limited driving range, which can be a significant obstacle for long-distance travel [16]. Unlike traditional ICE driven vehicles, EVs rely on batteries, which need to be recharged regularly. This means that EV drivers need to plan their trips carefully to ensure that they have enough range to reach their destination and find a charging station along the way. Therefore, the driving range of the EV has been limited to 200 km when obtaining the charger stations and the historical check-in records. In other words, the region of the data set is selected the Eastern Mediterranean Region of Turkey that is shown in Figure 1. The major factor in the preference of the Eastern Mediterranean Region of Turkey is the intensity of vehicle mobility caused by summer holiday activity in coastal cities. The region also offers high quality tourism services, which attract many tourists to the area. This makes the area an ideal place for conducting research on the mobility of vehicles. As well, the extensive industrial activity in the area has further enhanced the importance of the site where the study will be conducted.

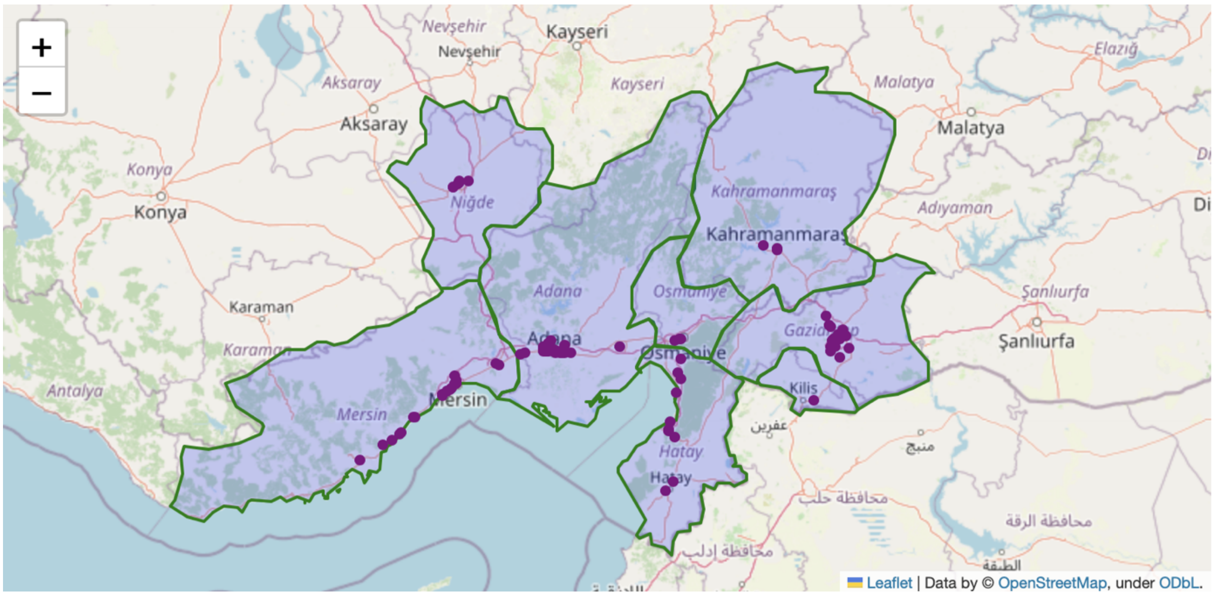

Figure 1. The selected region for the dataset and the distribution of the charger station [17].

In addition, the total charger station value is 92 for this region that is also shown in Figure 1. The historical check-in column of the data set is acquired from a PlugShare. It is a popular mobile and web application that helps EV drivers find available charging stations near their location. The app provides real-time information on charging station availability, location, and type, as well as user reviews and ratings [18]. The data set is obtained with a resolution of daily in this paper. The data set from these charging stations cover the period between 2018 and 2022, with 270 columns totalling the historical check-ins. After that resolution of the data set is converted from daily to monthly and the column of the data set which is used for prediction of the charging demand is also converted to 54. Seven features of the data set are then concatenated with the historical check-in column. The distribution of historical check-in records for each charger location are shown in Figure 2.

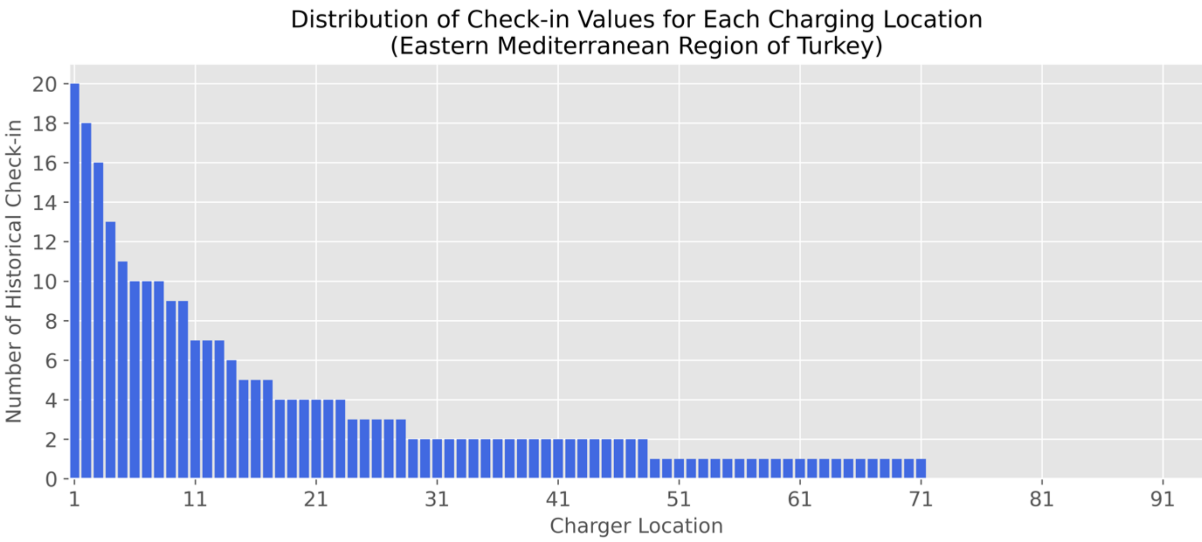

Figure 2. Distribution of check-in values for each charger location.

Check-in predictor of the data set is previous monthly check-in (P_MC) obtained from the charger stations. External predictors namely, electricity (E), gasoline (G), diesel (D), and liquefied petroleum gas (LPG) are obtained from EMRA [19]. Calendar predictors that are years (Y) and month (M) are attained by utilising datetime package in Python programming language. As shown in Figure 3, the methodology followed in the paper is described. Next sub-section presents the methods of the paper.

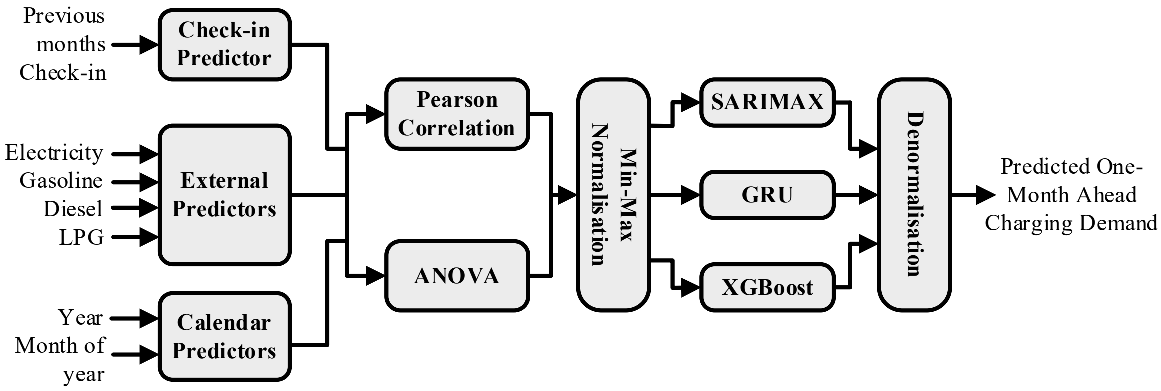

Figure 3. The methodology followed in the paper.

Methods

In this section, the bases of the feature selection methods and the fundamentals of the prediction of charging demand algorithms are presented. Feature selection is the process of selecting a subset of relevant features or variables from a larger set of features that are available in a dataset. It is an important step in many data analysis tasks as it can improve the accuracy, interpretability, and efficiency of models, while avoiding overfitting and other potential issues. After that, SARIMAX, GRU and XGBoost algorithms are applied to the obtained data set for the charging demand prediction step.

<u>*Pearson correlation.*</u> It also known as the Pearson's r or Pearson product-moment correlation coefficient, is a measure of the linear relationship between two parameters [20]. It is denoted by the symbol 'r', and it ranges from -1 to +1. The formula for calculating the Pearson correlation coefficient is as follows:

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\sum_{i=1}^{n}\left(y_i - \bar{y}\right)}} \tag{1}$$

where x and y are the two parameters, $\bar{x}$ and $\bar{y}$ are the mean of the two parameters, n is the number of the each parameter. Based on the Pearson correlation coefficient, the heat map obtained for each parameter is shown in Figure 4. If two variables have a positive correlation, then an increase in one variable is associated with an increase in the other variable, while a negative correlation means that an increase in one variable is associated with a decrease in the other variable. A correlation of zero means that there is no relationship between the variables. It has several properties, including that it is symmetric (i.e., the correlation between x and y is the same as the correlation between y and x), and it is sensitive to outliers and non-linear relationships between variables [21]. In addition, parameters of correlation coefficients less than 0.1 are undervalued to observe the most correlated parameters in the following figure.

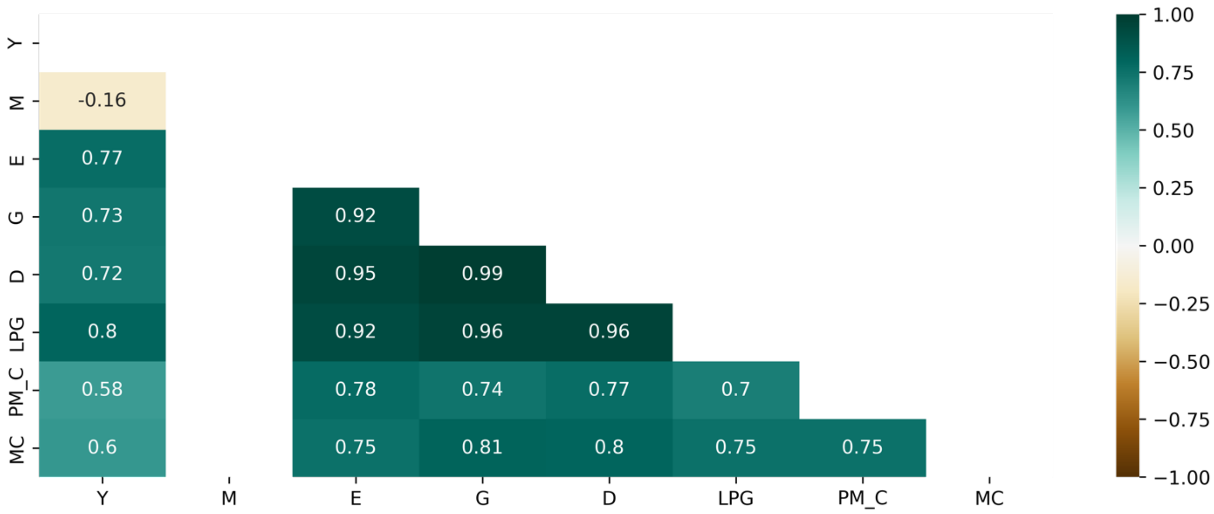

Figure 4. The Pearson correlation heat map for each parameter.

ANOVA. It is a statistical method used to test for differences between two or more groups of data. Specifically, it is used to determine if the means of the groups are significantly different from each other [22]. It computes F-value for each feature in the data set, given a set of target variables. This function is commonly used for performing ANOVA feature selection, where we want to select the most relevant features for predicting a target variable based on their variance. F-value measures the ratio of the between-group variance to the within-group variance, which indicates how well the feature separates the different target variable groups [23]. As shown in Figure 5, the highest F-scores indicates that they are most important features for predicting the target variable.

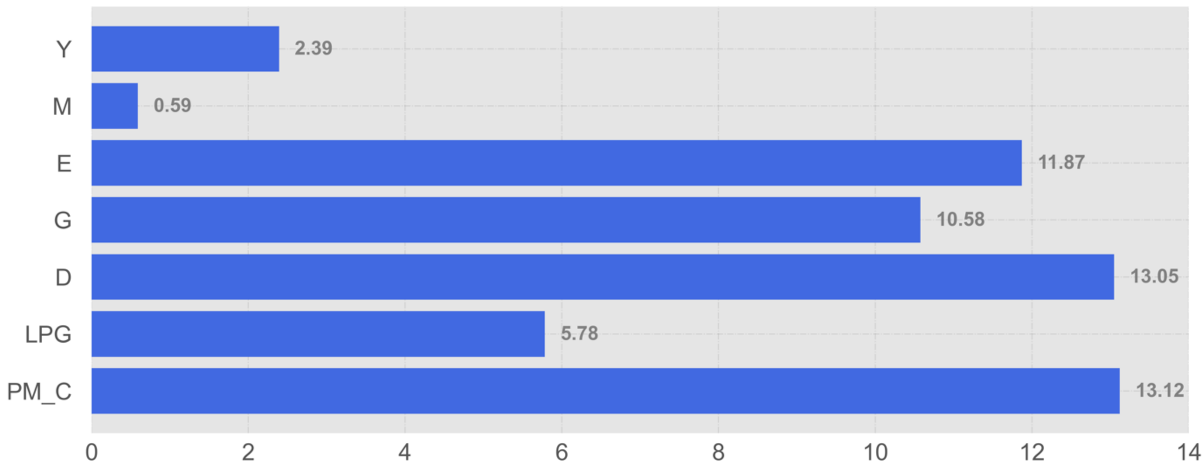

Figure 5. F-scores of each feature.

SARIMAX. It is a time series analysis and forecasting technique that extends the popular ARIMA model. It allows for the inclusion of additional exogenous variables that may affect the time series being analysed. The model is based on the idea that a time series can be decomposed into three components: trend, seasonality, and residual. The model takes into account the autocorrelation and moving average components of the time series, as well as any seasonal patterns that may be present. The model is specified using four main parameters that are seasonal period (s), autoregressive order (p), order of differencing (d), and moving average order (q). In this paper, auto-ARIMA is used to obtain the values of model (p,d,q) and the model values is $(2,2,1)$.

GRU Networks. This network is a type of Recurrent Neural Network (RNN) architecture that is used to process sequential data such as time series or natural language data. GRU was first introduced in 2014 by Cho et al. as a simpler alternative to the LSTM network, but with comparable performance [24]. GRU is designed to overcome the vanishing gradient problem that affects many RNNs, which occurs when the gradients used to update the weights during backpropagation become extremely small over time, making it difficult for the network to learn from long-term dependencies in the data. The cell mechanisms of GRU networks are illustrated in Figure 6. In addition, GRU uses a gating mechanism to selectively update or forget information at each time step. The update and reset gates allow the network to selectively remember or forget certain information over time, which can help it to model longer-term dependencies in the data [25].

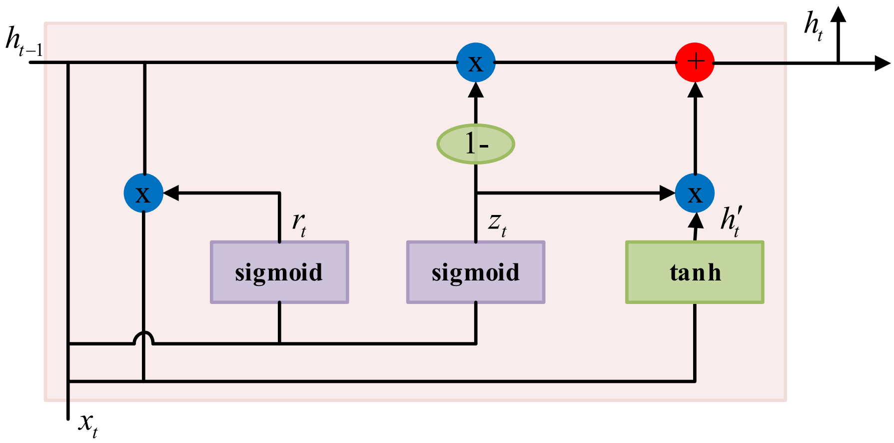

Figure 6. GRU cell structure.

GRU networks differ from LSTM networks in that they combine the cell state and hidden state into a single variable, h_t, while LSTM networks keep them separate. The mathematical operations used by GRU networks to perform this merging are as follows:

$$z_t = \sigma\left(\omega_z \cdot [h_{t-1}, x_t] + b_z\right) \tag{2}$$

$$r_t = \sigma\left(\omega_r \cdot [h_{t-1}, x_t] + b_r\right) \tag{3}$$

$$h'_t = \tanh\left(\omega_h \cdot [r_t \odot h_{t-1}, x_t] + b_h\right) \tag{4}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot h'_t \tag{5}$$

where x_t is input vectors, h_t is hidden layer vectors, b_h, b_r, and b_z are bias vectors, ω_h, ω_r, and ω_z are parameter matrices.

Equations (2) and (3) are used to receive inputs from the previous hidden state and the current data point, followed by processing these inputs with corresponding weights. The resulting output is then filtered for unnecessary information using a non-linear activation function, tanh, as described in equation (4). This section of the network is commonly known as the reset gate. The network then updates the hidden states using a set of mathematical operations, as stated in equation (5), and transmits the updated hidden state to the next cell. These steps are repeated throughout the training phase [26].

XGBoost. Decision trees are a popular and powerful gradient boosting algorithm that is used for supervised machine learning tasks, such as regression, classification, and ranking. It is designed to be highly scalable, efficient, and flexible, making it a popular choice for a wide range of applications. It works by iteratively training decision trees on the residual errors of the previous tree. It uses gradient descent optimization to minimize a loss function, which measures the difference between the predicted and actual values. The algorithm is designed to handle both sparse and dense data and includes various regularization techniques, such as L1 and L2 regularization, to prevent overfitting [27]. The purpose of Figure 7 is to provide a better understanding of the tree structure of XGBoost.

The algorithm is used for supervised learning problems, where predicting the target label y with using training data x. Training loss and regularization term comprise the objective function of XGBoost as follows:

$$obj(f) = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \Omega(f) \tag{6}$$

where n is the number of training examples, L indicates the training loss function, y_i is the real values, $\hat{y}_i$ is the estimated values and the last one Ω is a regularization term. As a result of the sum of all the trees, the final outcome is estimated as shown in equation 7.

$$\hat{y}_i = \sum_{k=1}^{j} f_k(x_i) = F_j(x_i) \tag{7}$$

where j is the number of the trees. The objective function of the j^{th} tree is obtained by substituting the equation (7) into (6), yields

$$obj(f_k) = \sum_{i=1}^{n} L\left(y_i, \hat{y}_i^{j-1} + f_j(x_i)\right) + \Omega(f_k) + C \tag{8}$$

Taking into account the regularization term of the first $j-1$ tree, the constant in the equation is derived. Equation (8) is converted to equation (9) using the Taylor expansion as follows:

$$obj(f_k) = \sum_{i=1}^{n} \left[L\left(y_i, \hat{y}_i^{j-1}\right) + g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i) \right] + \Omega(f_k) + C \tag{9}$$

where $g_i = \partial_{\hat{y}_i^k} L\left(y_i, \hat{y}_i^{j-1}\right)$ and $h_i = \partial_{\hat{y}_i^k}^2 L\left(y_i, \hat{y}_i^{j-1}\right)$. The regularization term is calculated using the following formula to reduce the complexity of the model.

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \tag{10}$$

where T is the number of the leaves, ω is the weight of the leaves, default values for γ and λ coefficients are 1 and 0.

In order to understand this algorithm, it is necessary to observe the intuition behind gradient boosting. By using the equation (7), the recurrence relation can be expressed as follows:

$$\begin{gathered} F_0(x_i) = f_0(x_i) \\ \vdots \\ F_k(x_i) = F_{k-1}(x) + \Delta_m(x) \end{gathered} \tag{11}$$

This boosting approach may be helpfully viewed as a golfer whacking a golf ball toward the hole at y, but only getting as far as $f_0(x_i)$. The golfer re-evaluates the direction and distance to the hole at each step, tapping the ball with decreasing force, and gradually moving the ball closer to the hole. As shown in Figure 7, five strokes are required to reach the hole, y, including two strokes, Δ_2 and Δ_3, that overshot the target.

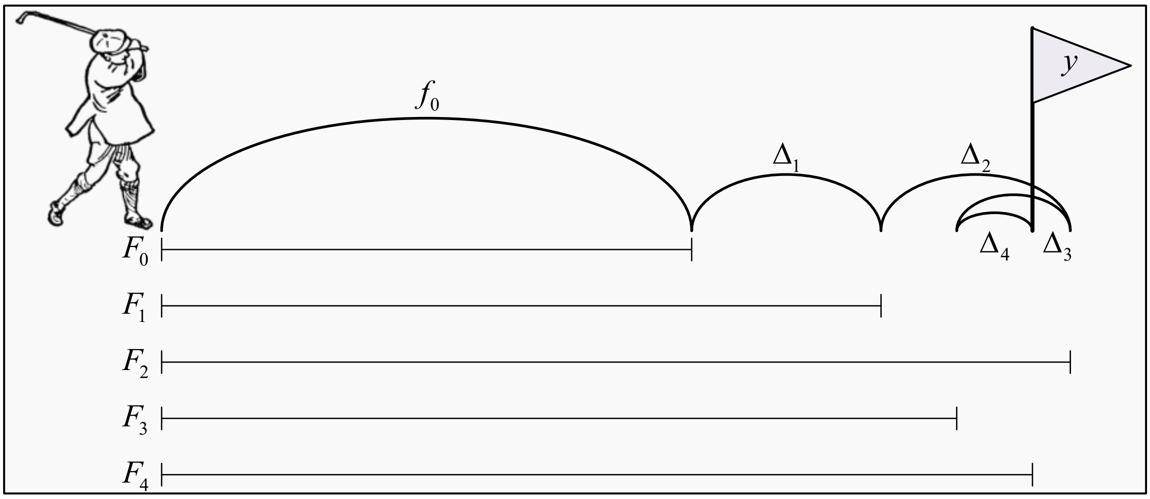

Figure 7. The illustration of the gradient boosting approach.

The golfer calculates the appropriate nudge after the initial stroke by comparing y with the first approximation, $y - f_0(x_i)$. In most cases, this difference is referred to as a residual or residual vector. The best way to approach gradient boosting is to think of it as a vector pointing from the current prediction, $\hat{y}_i = F_k(x_i)$, to the actual value, y. Nex section introduces the result and discussion of the paper.

RESULTS AND DISSCUSSION

In this study, all computations were performed on a Macintosh computer equipped with OS version 10.15.2, a 2.4GHz processor (Intel Core i5), and 8GB of memory. The Python programming language was used for all computing tasks using Spyder as an integrated development environment. In addition, pdarima, tensorflow, xgboost, sklearn packages were used to implement SARIMAX, GRU networks, XGBoost decision trees, and ANOVA respectively. The algorithms were trained on the same data set to compare the pure performances of the algorithms. PC and ANOVA methods, which are feature selection methods, were applied prior to the training of algorithms. In the PC method, features above 0.5 were selected based on correlation coefficient values as in Scenario 1. As a result, six features were determined for the training phase of the algorithm. In the ANOVA method, the F-scores for each feature is calculated by the selection method. In order to determine the features for the algorithm, a limit of 10 was determined for the F-scores, and the features above this value were selected in Scenario 2. Therefore, four feature were determined for the training phase of the algorithm.

Prior to attaining predictions, the values contained in the input variables of the data set are normalized by scaling them between 0 and 1. This normalization process eliminates the units of the various data types, reduces computational time, requires less memory to store the data, and ensures that multiple data columns can be benchmarked in a consistent manner. In SARIMAX, GRU networks and XGBoost, the input matrix is $N \times M$ in which N is number of the features and M is the length of the data set. In SARIMAX, auto-ARIMA is used to obtain the values of model (p,d,q) and the model values is $(2,2,1)$. To predict the EV charging demand, the batch of GRU networks size set to 48. The optimizer of this model was Adam and the back-propagation of the error function was implemented by using L1 loss function [28]. The predictions obtained from neural networks are often unreliable on the basis of unseen data because neural networks are prone to overfitting. A dropout layer was implemented with a dropout probability of 0.2 in order to prevent the network from overfitting. In XGBoost, the booster was set to gradient boosting tree, the loss function was defined squared error, and maximum depth of the tree was set to 6. Finally, the learning rate of both models were set to 0.03 and number of the epoch of both models were defined as 100.

In order to verify the validity and robustness of both algorithms, the data was randomly divided into two sets, namely, training sets and test sets. As much as 80% of the entire data set was made up of the training set, only 20% was made up of the test set. This study employs R^2, MAE, and $MAPE$ to assess the performance of SARIMAX, GRU, and XGBoost. The formulas for these performance metrics are as follows:

$$R^2(\%) = \left(1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right) \times 100 \tag{12}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{13}$$

$$MAPE(\%) = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{|y_i|} \tag{14}$$

where y_i is actual output, $\hat{y}_i$ is predicted output, $\bar{y}$ is mean of y_i and n indicates the number of observations [29,30].

In accordance with the results presented in Table 1, two different feature selection methods were used to assign the inputs to be included in the algorithms and these methods were evaluated under two scenarios. Scenario 1 consists of inputs selected using Pearson correlation method and processed into SARIMAX, GRU and XGBoost algorithms, while Scenario 2 is composed of inputs selected utilising ANOVA method and processed into SARIMAX*, GRU*

and XGBoost* algorithms. The R^2 values for all models are generally above 90%, and the model with the highest R^2 score is XGBoost*. Moreover, the best performances of the MAPE are 14.33 % for XGBoost*, 15.66% for XGBoost, 18.77% for GRU* networks, and 19.08 for GRU networks. Both statistical models have MAPE values over 20%. The best elapsed time is also XGBoost and XGBoost*. The MAE rankings are also the same as the MAPE rankings, with the exception of XGBoost and XGBoost*. The results indicate that the most accurate approach for prediction of EV charging demand was obtained by using XGBoost* model having a MAPE of 14.33%. In other aspects, while the MAPE of the GRU* network model is comparable to that of the XGBoost* model, it is worth noting that the GRU* model takes longer to execute than the XGBoost* model.

Table 1. Performance comparison of the applied methods.

	Feature Selection Method	Models	Performance Metrics			Elapsed Time (s)
			R^2 (%)	MAPE (%)	MAE	
Scenario 1	Pearson Correlation	**XGBoost**	95.73	15.66	0.193	Best
		GRU	94.86	19.08	1.666	Good
		SARIMAX	91.13	20.92	3.000	Medium
Scenario 2	ANOVA	**XGBoost***	96.58	14.33	1.833	Best
		GRU*	96.36	18.77	2.166	Good
		SARIMAX*	90.74	21.68	2.800	Medium

Furthermore. if the primary objective is to achieve minimal error, the XGBoost* model, which combines ANOVA feature selection, is the most suitable choice for the prediction of the EV charging demand when considering all the results together. The GRU* model is a great option when both minimum error and elapsed time considered together. Unseemly, it is not recommended to use SARIMAX models for predicting EV charging demand in this study, as they do not appear to be a suitable fit for the data set.

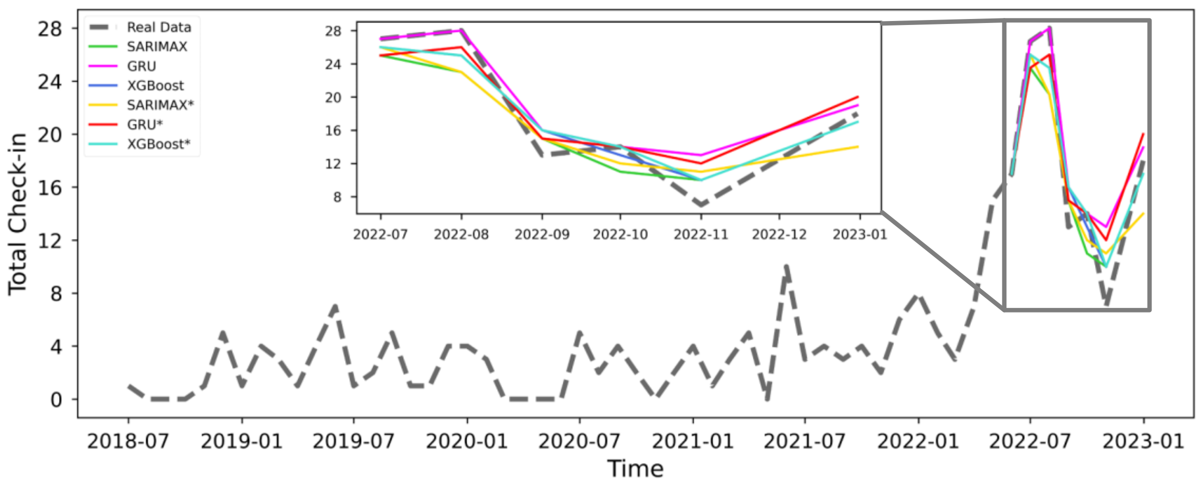

Figure 8. Comparison of the prediction EV charging demand.

Moreover, the results of the prediction of EV charging demand are compared as shown in Figure 8. In recent years, it can be seen from the monthly check-in numbers in the figure that the interest in EVs has increased gradually in Türkiye. It is also shown from the figure that XGBoost* and GRU* models are most reliable option for the prediction of the EV charging demand.

CONCLUSION

Nowadays, there has been a shift in vehicle fuel types from fossil fuels to electricity with global warming becoming an increasingly pressing concern. This has led to a rise in popularity of electric vehicles (EVs), which offer several advantages over conventional vehicles such as being environmentally friendly, having lower maintenance requirements, and providing better mileage performance in city driving. However, the integration of EVs into the existing infrastructure and their high initial investment cost remain a few of the challenges associated with their adoption.

In order to address the gap in the literature, this study applied various statistical- and ML-based techniques such as SARIMAX, GRU networks, and XGBoost to real-time data obtained from PlugShare and EMRA for the Eastern Mediterranean Region of Türkiye to predict EV charging demand. The models considered various input parameters, including historical monthly check-in, electricity, gasoline, diesel, LPG, number of years, and number of months, to predict the charging demand for the upcoming month.

The findings of the study suggest that the statistical- and ML-based techniques used provide reasonably accurate predictions of EV charging demand. However, upon closer examination of the results, the XGBoost* model stands out for its successful convergence. Furthermore, while the performance of GRU* networks is comparable to that of the XGBoost* model, GRU* networks elapsed time is longer than elapsed time of the XGBoost* model. In future studies, it is planned to expand the data set with supervised learning-based methods to estimate EV charging demand based on cities. In addition to the data set to be expanded, feature variation will be added to the data set, and different methods will be included in the future studies in addition to the methods used in the current study.

ACKNOWLEDGMENT

On February 6, 2023, the series of earthquakes that directly affected eleven provinces in Türkiye caused thousands of casualties and enormous destruction. Most of the provinces included in this study are within the scope of the earthquake-affected region. Since the data set collected in this study took place before the earthquake, we regret to state that some charging stations where the data were collected are currently damaged and cannot function properly.

NOMENCLATURE

EV	Electric Vehicle
ML	Machine Learning
PC	Pearson Correlation
ANOVA	Analysis of Variance
XGBoost	eXtreme Gradient Boosting
MAPE	Mean Absolute Percentage Error
MAE	Mean Absolute Error
ICE	Internal Combustion Engine
Seq2Seq	Sequence to Sequence
LSTM	Long Short-Term Memory
ARIMA	Auto-Regressive Integrated Moving Average

EMRA	Energy Market Regulatory Authority
SARIMAX	Seasonal Regressive Integrated Moving Average eXogenous
P_MC	Previous Monthly Check-in
E	Electricity
G	Gasoline
D	Diesel
LPG	Liquefied Petroleum Gas
Y	Years
M	Month
RNN	Recurrent Neural Network

REFERENCES

1. Miškić, J., Pukšec, T., and Duić, N., Sustainability of energy, water, and environmental systems: a view of recent advances, *Clean Technologies and Environmental Policy*, 1-8, 2022.
2. Carvalho, E. F., Sousa, J. A., and Lagarto, J. H., Assessing electric vehicle CO2 emissions in the Portuguese power system using a marginal generation approach, *International Journal of Sustainable Energy Planning and Management*, *26*, 47-66, 2020.
3. Schill, W. P., and Gerbaulet, C., Power system impacts of electric vehicles in Germany: Charging with coal or renewables?, Applied Energy, 156, 185-196, 2015.
4. Ardeshiri, R. R., Balagopal, B., Alsabbagh, A., Ma, C., and Chow, M. Y., Machine learning approaches in battery management systems: State of the art: Remaining useful life and fault detection, In *2020 2nd IEEE International Conference on Industrial Electronics for Sustainable Energy Systems (IESES)* (Vol. 1, pp. 61-66), IEEE, 2020.
5. Bachute, M. R., and Subhedar, J. M., Autonomous driving architectures: insights of machine learning and deep learning algorithms, *Machine Learning with Applications*, *6*, 100164, 2021.
6. Mosavi, A., and Bahmani, A., Energy consumption prediction using machine learning; a review, 2019.
7. Ahmed, M., Zheng, Y., Amine, A., Fathiannasab, H., and Chen, Z., The role of artificial intelligence in the mass adoption of electric vehicles, *Joule*, *5*(9), 2296-2322, 2021.
8. Yi, Z., Liu, X. C., Wei, R., Chen, X., and Dai, J., Electric vehicle charging demand forecasting using deep learning model, *Journal of Intelligent Transportation Systems*, *26*(6), 690-703, 2022.
9. Sumanasena, V., Gunasekara, L., Kahawala, S., Mills, N., De Silva, D., Jalili, M., ... and Jennings, A., Artificial Intelligence for Electric Vehicle Infrastructure: Demand Profiling, Data Augmentation, Demand Forecasting, Demand Explainability and Charge Optimisation, *Energies*, *16*(5), 2245, 2023.
10. Amini, M. H., Karabasoglu, O., Ilić, M. D., Boroojeni, K. G., and Iyengar, S. S., ARIMA-based demand forecasting method considering probabilistic model of electric vehicles' parking lots, In *2015 IEEE Power & Energy Society General Meeting* (pp. 1-5). IEEE, 2015.
11. Kim, Y., and Kim, S., Forecasting charging demand of electric vehicles using time-series models, *Energies*, *14*(5), 1487, 2021.
12. Hu, T., Liu, K., and Ma, H., Probabilistic electric vehicle charging demand forecast based on deep learning and machine theory of mind, In *2021 IEEE Transportation Electrification Conference & Expo (ITEC)* (pp. 795-799), IEEE, 2021.
13. Arias, M. B., and Bae, S., Electric vehicle charging demand forecasting model based on big data technologies, *Applied energy*, *183*, 327-339, 2016.
14. Xydas, E. S., Marmaras, C. E., Cipcigan, L. M., Hassan, A. S., and Jenkins, N., Forecasting electric vehicle charging demand using support vector machines, In *2013 48th International Universities' Power Engineering Conference (UPEC)* (pp. 1-6), IEEE, 2013.

15. Moon, H., Park, S. Y., Jeong, C., and Lee, J., Forecasting electricity demand of electric vehicles by analyzing consumers' charging patterns, *Transportation Research Part D: Transport and Environment*, *62*, 64-79, 2018.
16. Tolun, O. C., and Tutsoy, O., A Non-Parametric Driving Range Estimation Algorithm for Electric Autonomous Vehicles with Multi-Layer Neural Networks, *4th International Congress on Engineering Sciences and Multidisciplinary Approaches at: Istanbul, Türkiye,* 2022.
17. OpenStreetMap contributors, Türkiye [Map], OpenStreetMap Foundation, Retrieved March 15, 2023, from https://www.openstreetmap.org/relation/174737, 2023.
18. PlugShare EV Charging Map, https://www.plugshare.com/, 2022.
19. Energy Market Regulatory Authority, https://www.epdk.gov.tr/Home/En, 2022.
20. Liu, X. S., A probabilistic explanation of Pearson's correlation, *Teaching Statistics, 41*(3), 115-117, 2019.
21. Liu, Y., Mu, Y., Chen, K., Li, Y., and Guo, J., Daily activity feature selection in smart homes based on pearson correlation coefficient. *Neural Processing Letters, 51,* 1771-1787, 2020.
22. Effrosynidis, D., and Arampatzis, A., An evaluation of feature selection methods for environmental data, *Ecological Informatics, 61,* 101224, 2021.
23. Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., and Lang, M., Benchmark for filter methods for feature selection in high-dimensional classification data. *Computational Statistics & Data Analysis, 143,* 106839, 2020.
24. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y., Learning phrase representations using RNN encoder-decoder for statistical machine translation, *arXiv preprint arXiv:1406.1078,* 2014.
25. Zor, K., and Buluş, K., A benchmark of GRU and LSTM networks for short-term electric load forecasting, *In 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)* (pp. 598-602). IEEE, 2021.
26. Ozdemir, A. C., Buluş, K., and Zor, K., Medium-to long-term nickel price forecasting using LSTM and GRU networks, *Resources Policy, 78,* 102906, 2022.
27. Yan, H., Yan, K., and Ji, G., Optimization and prediction in the early design stage of office buildings using genetic and XGBoost algorithms, *Building and Environment, 218,* 109081, 2022.
28. Kingma, D. P., and Ba, J., Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980,* 2014.
29. Zor, K., Timur, O., Çelik, Ö., Yıldırım, H., and Teke, A., Interpretation of error calculation methods in the context of energy forecasting, In *12th Conference on Sustainable Development of Energy, Water and Environment Systems, Dubrovnik, Croatia* (Vol. 722, pp. 1-9), 2017.
30. Timur, O., Zor, K., Çelik, Ö., Teke, A., and İbrikçi, T., Application of statistical and artificial intelligence techniques for medium-term electrical energy forecasting: a case study for a regional hospital, *Journal of Sustainable Development of Energy, Water and Environment Systems*, *8*(3), 520-536, 2020.