Journal of Informatics and Web Engineering
Vol. 5 No. 2 (June 2026) eISSN: 2821-370X

Analysis and Predictive Modelling of EV Charging Patterns and User Behaviour

Yi Xuan Law[1], Kah Ong Michael Goh[2*], Check Yee Law[3], Connie Tee[4], Yong Wee Sek[5], S M Hasan Mahmud[6]**

[1,2,3,4]Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450 Melaka, Malaysia
[5]Faculty of Artificial Intelligence and Cyber Security, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
[6]Department of Software Engineering, Daffodil International University, Savar, Dhaka, 1216, Bangladesh
**corresponding author: (michael.goh@mmu.edu.my; ORCiD: 0000-0002-9217-6390)*
***corresponding author: (drhasan.swe@diu.edu.bd; ORCiD: 0000-0002-6828-3559)*

Abstract - As more Electric Vehicles (EVs) are released, the ability to predict energy consumption and charging duration becomes crucial in building optimal infrastructure and a proper system of managing energy. This paper proposes a machine learning model that predicts these two important parameters using a real-world dataset. The dataset consists of 1320 EV charging sessions made between January and February 2024 on Kaggle. The data set includes vehicle specifications, time stamps of sessions, environmental conditions, and user behaviour. Feature engineering tasks followed a thorough preprocessing procedure where missing values were imputed, outliers were removed, and the type of data was converted, and included time-based transformations, interaction terms, station popularity measures. Three regression models were developed: Light Gradient Boosting Machine (LGBM), Random Forest (RF), and Support Vector Regression (SVR) to evaluate different modelling approaches and test the predictive efficacy of ensemble against kernel-based methods. These models were trained and tuned using GridSearchCV combined with TimeSeriesSplit cross-validation to maintain temporal integrity. Evaluation metrics included Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R^2). Results showed that RF achieved the highest accuracy in predicting energy consumption with an R^2 of 0.6620, while LGBM performed best in predicting charging duration with an R^2 of 0.9152. Final testing on unseen data validated the generalization capabilities of these models. The findings support practical infrastructure recommendations and demonstrate the potential of machine learning in enhancing EV charging operations.

Keywords— Electric Vehicles, Energy Consumption, Charging Duration, Machine Learning, Prediction.

Received: 15 July 2025; Accepted: 23 November 2025; Published: 16 June 2026

1. INTRODUCTION

In 2023, the United States registered around 1.4 million new Electric Vehicles (EVs), representing a more than 40% increase over 2022 [1]. The analysis results from Trends in Electric Cars [1] reveals that although the relative yearly

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2026.5.2.6

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

growth rate has begun to level down compared to prior years, overall demand for EVs and the absolute number of new registrations remain high. As the EV industry matures, there is an increasing need to transition to more sustainable and ecologically aware transportation solutions. EVs have significant environmental advantages, including a zero-tailpipe emission capable of reducing air pollution and climate change. Even their better power-efficiency compared to the vehicles that employ internal combustion engines, along with the fact that they could be made to run on renewable energy sources, lends further to their sustainability. The wider use of EVs also contributes to the health of the population, energy security and economic growth as they drive innovation and create job opportunities in the manufacturing sector and development of infrastructures [2]. The construction of efficient and accessible EV charging infrastructure is a vital component in facilitating this shift.

Globally, EV use is increasing. In 2023, approximately 14 million new EV were registered worldwide, representing for 18% of total new automobile sales [3]. China topped the market with more than 8 million new EV registrations, followed by Europe and the United States. Together, these areas accounted for 95% of global EV sales [3] . Countries such as Norway, where EVs account for more than 90% of new car sales, as well as Canada, Japan, South Korea, and the United Kingdom, are making substantial progress toward electrification [4]. The worldwide EV stock reached 40 million vehicles in 2023, a sixfold increase from 2018 [3].

The fast uptake of EVs has increased the need for efficient and dependable charging infrastructure [5]. However, the development of such infrastructure is hindered by a limited understanding of user charging behaviour. Unlike typical automobiles, EVs require longer charging times, therefore user behaviour is an important aspect in optimizing charging station rollout and operation. The lack of precise prediction models for energy consumption and charging time makes it difficult to estimate peak demand and distribute resources efficiently. The charging behaviour is very temporal and spatially variable, and planning of such systems becomes intractable with EVs. This sort of unpredictability has the potential to lead to congestion, under-utilization of efficient-stations, and low levels of customer satisfaction [6]. Without an integrated representation of the variables that affect the EV-charging dynamics, there is a possibility that strategic infrastructure deployment will be out of line with the user requirements, which will subsequently lead to poor performance of the system, and network pressure.

This research tackles these issues by introducing a Machine Learning (ML) based method for forecasting electric car energy consumption and charging duration. The suggested approach employs a systematic technique that uses historical charging data to train and improve sophisticated regression models such as Light Gradient Boosting Machine (LGBM), Random Forest (RF), and Support Vector Regression (SVR). The goal of methodically analysing and designing characteristics from charging sessions is to precisely forecast these essential indicators, giving critical information for the strategic development and efficient operation of EV charging infrastructure. This research aims to answer the following questions:

- What are the key patterns and trends in EV charging behaviour shown by the dataset?
- How to forecast the energy consumption and charging duration?
- How to improve EV charging infrastructure based on analysis and predictive models?

The research objectives are as follows:

- To identify key patterns and trends in EV charging behaviour from the dataset.
- To develop predictive models for energy consumption and charging duration.
- To make recommendations for improving EV charging infrastructure based on the analysis.

While existing studies have explored various aspects of EV charging, this work offers a distinct contribution by providing a comprehensive predictive framework that addresses both energy consumption and charging duration across diverse real-world datasets. A large EV charging dataset extracted in Omaha, Nebraska, was used to analyse household EV charging behaviour in [7] resulting in estimated connection duration, charging duration, charging demand, and next-session arrival time. They limited their analysis to user-agnostic models where the user ID was deliberately left out to better generalize. Based on these preliminary insights, the present study further extends the scope of the predictive nature and practical applicability of previous research in a few note-worthy ways.

This article presents an introduction that emphasizes the rising demand for efficient EV charging infrastructure, as well as the issues faced by unexpected charging behaviour that leads to the need for further research in this field. Section 2 includes Literature Review, which summarizes existing studies on EV charging trends, forecasting methodologies, and user behaviour. Research Methodology that consists of the systematic approach used for data preprocessing, feature engineering, data splitting, model training, hyperparameter tuning, and assessment criteria for both the EVCP and Colorado datasets is described in Section 3. Section 4 presents the results and discussions, which examine the performance of LGBM, RF, and SVR models in predicting energy consumption and charging duration for the EVCP dataset, as well as monthly energy usage for the Colorado dataset, with a comparison to traditional time series models and contributions of this research work. Finally, Section 5 concludes the article by summarizing the key findings, acknowledging limitations, and recommending future research directions.

2. LITERATURE REVIEW

Understanding EV charging patterns and anticipating important metrics such as energy consumption and charging duration is critical for optimal EV charging infrastructure development and management. Beyond these, recent research has further been focused on the EV health status and designing more accurate range estimation using both machine and deep learning algorithms [8]. The definite of EV charging and the respective energy consumption patterns have drawn attention of many scholars who utilize tools of data processing and analysis, clustering, regression, and ML [9], [10], [11]. Various investigations have been performed on temporal, environmental and behavioural aspect. For instance, low temperatures have been linked to increased charging frequency, while holidays tend to reduce it, as shown through K-means and Apriori analysis of big data from Shanxi Province [12]. Similarly, studies at university campuses have identified peak charging periods during weekday mornings and afternoons, with session durations typically under three hours, and have developed probabilistic models to estimate demand [13]. Another line of research focuses on behavioural modelling and user responsiveness. A multi-stage EV charging recommendation system capable of forecasting peak load times (5:00 AM to 8:00 PM) was designed in [14], noting user reluctance to respond to early morning invitations but responsiveness to evening pricing mechanisms. Multinomial logistic regression applied to over 16,000 charging sessions in Nebraska highlighted variations in charging duration based on time and location [15], while Functional Principal Component Analysis (FPCA) techniques have uncovered dominant charging trends tied to land-use characteristics, distinguishing between early morning/evening and daytime patterns [16]. Driving conditions were shown to significantly affect energy demand in [17], where cold weather increased consumption due to battery and cabin heating, while warm conditions had a minor positive effect; aggressive driving, elevation changes, and accessory use also raised consumption.

Besides that, understanding the diversity of user behaviour is critical for adapting EV charging services and infrastructure to satisfy a wide range of needs, increasing user happiness and optimizing resource allocation. Clustering and segmentation of users are two of the traditional approaches to explaining the diversity of behaviour. Various studies have exploited the unsupervised learning and hybrid structures to divide specific user probes. For example, a hybrid RF–GMM approach was used to analyse over 240,000 charging sessions in Suzhou, revealing 14 behavioural clusters grouped into four overarching categories [18]. Similarly, Gaussian Mixture Models (GMM) were employed in [19] to classify users by session duration, power consumption, and waiting time, and further RF and LSBoost algorithms were used to develop predictive models for load flexibility, concluding that smart charging outperforms First-Come, First-Served strategies. Beyond clustering, researchers have explored behavioural traits and psychological factors. In the study of [20], the BIRCH algorithm was used to survey four levels of power, five State of Charge (SOC) behaviours, and three user profiles among 220,000 records, and it also noted such psychological phenomena as mileage anxiety. At the same time, Factor and cluster analyses were conducted in [21], to categorize Battery Electric Vehicle (BEV) and Plug-in Hybrid Electric Vehicle (PHEV) users into four groups, finding that home ownership and income were key determinants of home and public charging usage.

It was found that regression-based approaches have been instrumental in uncovering how infrastructure and spatial context influence EV charging behaviour. Studies have shown that charging frequency and idle time are closely tied to infrastructure configuration and urban density. For instance, users in low-density areas tend to charge more frequently, and the presence of fast chargers significantly reduces idle time—highlighting the need for policies that promote off-peak charging and improved grid integration [22]. Expanding on this, a large-scale analysis of 148,136 sessions in Boulder, Colorado employed five regression models to capture spatial-temporal dynamics, ultimately identifying the K-Nearest Neighbours Regressor as the most effective ($R^2 = 0.99996$), demonstrating the potential of non-linear models in accurately modelling charging behaviour [23].

To this end, some researchers have taken a systematic assessment of ML based strategies towards energy and duration forecasting. A Root Rean Square Error (RMSE) of 0.159 kWh with a Mean Absolute Percentage Error (MAPE) of 12.68% was achieved using Extreme Gradient Boosting (XGBoost) on trip-segmented real-world driving data [24]. A survey of ML approaches concluded that supervised methods, particularly RF, were better suited for predicting energy and journey length, while unsupervised methods such as K-means and GMM were effective for classifying user behaviour [25]. In addition, to improve predictive capabilities, some studies have used synthetic data generation and hybrid modelling. Conditional Tabular Generative Adversarial Network (CTGAN) and kernel density estimation were used to model daily energy use and vehicle counts. Gaussian Process Regression and multilayer neural networks achieved R^2 values over 0.96 [26]. Likewise, quantile-based pipelines have been developed to assist real-time forecasting, with Quantile Regression Neural Networks (QRNN) and their online variations lowering prediction errors to as low as 5.04% PMAE [27]. In parallel, physics-informed ML models have evolved to capture sophisticated driving behaviours. A hybrid model incorporating synthetic driving data effectively simulated acceleration and regenerative braking patterns with exponential functions, attaining R^2 values over 0.99 [28]. These approaches demonstrate the power of integrating domain expertise with data-driven methodologies to increase forecasting accuracy and operational relevance.

Furthermore, a growing corpus of research has examined the relative efficacy of ML models in predicting EV charging behaviour, energy consumption, and session time. RF consistently outperforms XGBoost and Artificial Neural Network (ANN) in predicting home charging habits, with moderate accuracy (R^2 between 0.40 and 0.48). However, forecasting the time until the next session remains challenging [7]. Comparative studies have also demonstrated that model performance can vary greatly depending on input characteristics. For example, Multilayer Perceptron (MLP) performed best without weather data, but SVR increased accuracy with weather inputs [29]. Several studies have demonstrated that ensemble and gradient boosting methods such as XGBoost, CatBoost, and LGBM can achieve high predictive accuracy under specific conditions. For instance, RF achieved the lowest RMSE and MAPE on Alphonso-Charging Network (ACN) data [30], while XGBoost outperformed SVR and RF in predicting charging time at UK stations, with SVR showing inconsistent results [31].

A combination of XGBoost, RF, CatBoost, and LGBM was used in [32] to predict charging duration by vehicle type, achieving an R^2 of 0.92 with XGBoost; SHAP analysis identified SOC, season, and lighting as key contributors. These findings were extended in [33] which introduced a Metaheuristic-optimized Model (GWO-ELM) that achieved R^2 = 0.972 and RMSE = 20.427, with SHAP analysis highlighting end/start SOC and air-conditioning use as significant inputs. Beyond these, advanced time-series analysis and deep learning models have also shown strong performance in EV-related predictions. The framework proposed in [34] used Convolutional Neural Networks (CNNs) to accomplish the movie recommendation task and converted the input sequences with 2D-DCT from emobpy-collected driving information. They did improve the accuracies of single-output CNN settings with their multi-output CNN and MAE loss combined with cumulative summation improved training speed, but straight modelling is computationally more cost-effective.

Apart from these, the user preferences in EV and performance of smart basting programs have been studied empirically. Interviews and choice experiments conducted among Danish EV owners revealed a strong preference for flat-fee tariffs, home charging, and the Tesla integrated network [35]. Cost, ease of use and convenience were factors most valued by the respondents which signalled that they were ready to go out of their way to charge faster and offer prices they found desirable. In the case of the Renault ZOE fleet, the dynamic smart-charging proposal involved using Renault fleet data, which saved 71.5 percent of nonproductive time and boosted the station throughput by facilitating up to 37 charges per day, which is one more than conventional strategies that allow only 31 charges per day [36].

3. RESEARCH METHODOLOGY

3.1 EVCP Dataset

The current work used a systematic pipeline to handle data gathered through Kaggle [37] and investigate EV charging behaviour, allowing for the estimate of two major outcomes: energy consumption and charging duration. This methodical pipeline, as shown in Figure 1, began with Data Preprocessing, which involved preparing raw data for analysis. This was followed by Exploratory Data Analysis (EDA) to discover underlying trends and feature engineering to generate new, useful variables. The prepared data was then separated into distinct sets for training and evaluation. Model training was then performed using specified methods, with their performance enhanced via hyperparameter tuning. Finally, model evaluation determined the accuracy of predictions, resulting in the derivation of insights and recommendations.

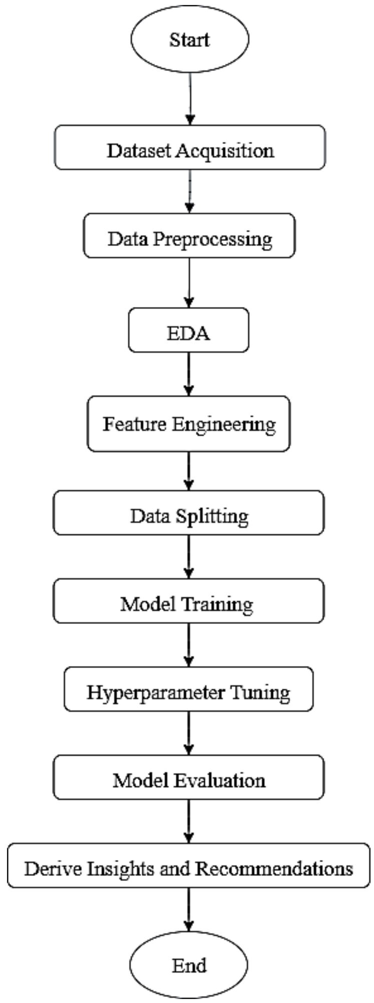

Figure 1. Proposed Framework Diagram for EVCP Dataset

The dataset utilized in this study was obtained from Kaggle [37] and included 1,320 unique electric car charging sessions collected between January and February 2024. This dataset's key elements were the start and end dates/times, port type, vehicle model, SOC at start and end, charging duration, temperature, and charger type. Missing values were resolved during the Data Preprocessing step using approaches such as median imputation, and outliers were discovered and dealt with to guarantee data quality and appropriateness for modelling. The ensuing EDA aimed to uncover patterns at the individual session level, such as the correlations between charging duration, energy spent, and factors such as SOC levels, charger types, and temperature fluctuations. This thorough study aided the feature engineering process, which enriched the dataset with useful characteristics for analysis and predictive modelling. This entailed numerous significant transformations, including the extraction of time-based characteristics from start date/time, such as Start Hour and its sinusoidal modifications [38], as well as a binary Is_Weekend variable. In addition, categorical variables were transformed to one-hot encoding to prepare them for ML methods. To train and analyse the model, the dataset was selectively partitioned into time segments. The training set contained data from January 1 to February 10, 2024 (41 days). The validation set contained data from the next week, February 11, 2024, through February 17, 2024 (7 days). Finally, the most recent week of data, February 18, 2024 to February 24, 2024 (7 days), was chosen as the test set. This temporal split ensured that the model was trained on historical data and then tested on new, previously unknown data, simulating real-world deployment scenarios.

The findings from these studies aided the selection of three ML techniques for model development: LGBM, RF, and SVR. LGBM was chosen because of its computational efficiency and ability to recognize complicated feature relationships. RF was chosen because of the architecture's ensemble-based nature, which reduces overfitting and improves generalization. SVR was chosen because of its capacity to match high-dimensional, nonlinear connections. Recognizing that SVR is sensitive to the values of the features used as input, all numeric features were normalized using the StandardScaler procedure, ensuring that no bias distance was calculated during training.

To optimize the models, methodical hyperparameter optimization provided its optimizations through the GridSearchCV algorithm performance technique. In the like vein, the TimeSeriesSplit cross-validation mechanism was specifically tailored to maintain chronological sequence and order of EV charging sessions, as well as allowing future-looking estimations to be made. Every split exercised the model in data until a certain timestamp and evaluated it on subsequent observations thus simulating the actual deployment conditions. The approach has the strength of structural time series of chronological organization, and as [39] argued, it facilitated a more life-assured and believable assessment of time-series practices. GridSearchCV provided an organised search over a grid of hyperparameter values and used cross-validation to choose the optimised configuration that has the best average performance over folds. It was established by Ahmad et al. [40] that such extensive approach always enhances model accuracy when comparing it with untuned models.

In this paper, the performance was measured using the common regression performance metrics, such as the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and R^2. After the hyperparameter optimization process, the trained models were used to predict the energy consumption and charging duration until the final week of the dataset February 18, 2024 - February 24, 2024. The forecasts are a practical estimate of the ability of the models to extrapolate to unseen data and are informative on the design and formulation of EV charging infrastructure policies.

3.2 Colorado Dataset

The effectiveness of the EV charging stations was estimated in this study by use of Colorado Public Charge Dataset [41]. An analytic-modelling framework was selected based on a comparison and implemented to predict the total energy usage each month. Data Preparation was the first stage of analysis and highly detailed session records were reduced to obtain the total monthly energy consumption. This important stage of data compilation placed the dataset in a condition to be converted to a variable of monthly forecast. The further treatment of the data set included a set of EDA techniques, such as Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, which provided an explanation of overall dynamics, demonstrated a strong annual trend, and identified autoregressive correlations affecting the relationship between the energy consumed on a monthly basis.

The results of the empirical analysis conducted in the current work became the input to the next feature engineering stage, as they provided a range of time matter variables. These characteristics included month, year, sine/cosine features, lagged features, rolling window features, and seasonal differences to capture time dynamics. These qualities were aimed at reflecting the animistic aspect of the temporal dimension. Data splitting was done in a temporal scheme which guaranteed a spatially coherent separation. Model training followed with a series of ML processes done with hyperparameters optimized with hyperparameter tuning. Finally, model evaluation judged performance against established metrics and compared results with past studies, leading to the conclusion of the analysis.

To construct and analyse the models, the temporal splitting technique was adopted, and data from January 2018 to May 2021 were assigned to a training set, with further observations from June 2021 to May 2022 kept for testing. The setup will ensure that the models are evaluated on unobserved future data hence simulating real life prediction problems. To adhere to the EVCP dataset, ML algorithms of LGBM, RF and SVR were chosen. The hyperparameters were tuned using GridSearchCV along with cross-validation that was conducted using TimeSeriesSplit, which kept in order the temporal pattern of data and levelled out the process of model optimization. MAE, Root Mean Squared Error (RMSE) and MAPE assessment measures were deployed. The above results were then standardized with the results of past studies, based on conventional time series models such as the SARIMA models [42], therefore, giving a point of reference to the current analysis.

4. RESULTS AND DISCUSSIONS

4.1 EVCP Dataset: Energy Consumption Prediction

In the frame of predicting the energy consumption, the RF model performed better with a MAE of 10.0437, RMSE of 13.0424 and R^2 of 0.6620. LGBM was the next best performer, whose MAE of 10.0483, RMSE of 13.0991 and R^2

of 0.6591, represented the best possible result. The visual performance of the LGBM is depicted in Figure 2. This figure depicts the LGBM model's performance on the test set for predicting energy consumption. The model accurately predicts energy usage, as evidenced by its MAE of 10.0483, RMSE of 13.0991, and R^2 of 0.6591. SVR on the other hand had significantly worse performance judging by MAE of 14.0537 17.8198 of RMSE and 0.3690 of R^2.

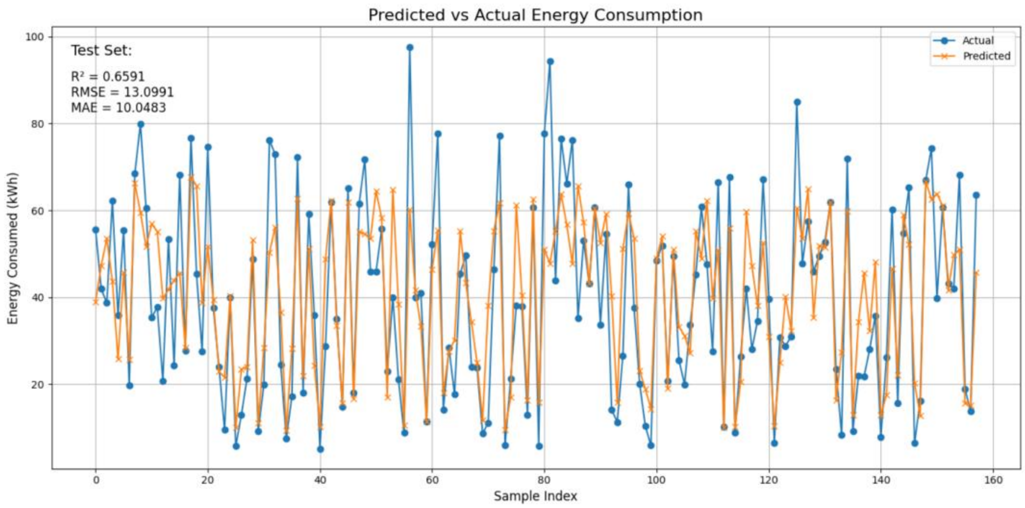

Figure 2. Comparison Between Predicted and Actual Energy Consumption Using LGBM Model

Figures 3 and 4 show the visualization of RF and SVR as the means of energy consumption prediction. Figure 3 compares the RF model's predictions to the actual energy consumption on the test set. The model accurately predicts energy consumption trends, with an MAE of 10.0437, RMSE of 13.0424, and R^2 of 0.6620. Figure 4 depicts the SVR model's performance in predicting energy usage on the test set. Compared to tree-based models, the visual alignment between predicted and actual values is less exact, with an MAE of 14.0537, RMSE of 17.8198, and R^2 of 0.3690. Aggregately, these results indicate that ensemble tree-based methods usually perform better than the kernel-based method when dealing with nonlinear, interactive nature of the dynamic line with the energy consumption of EVs.

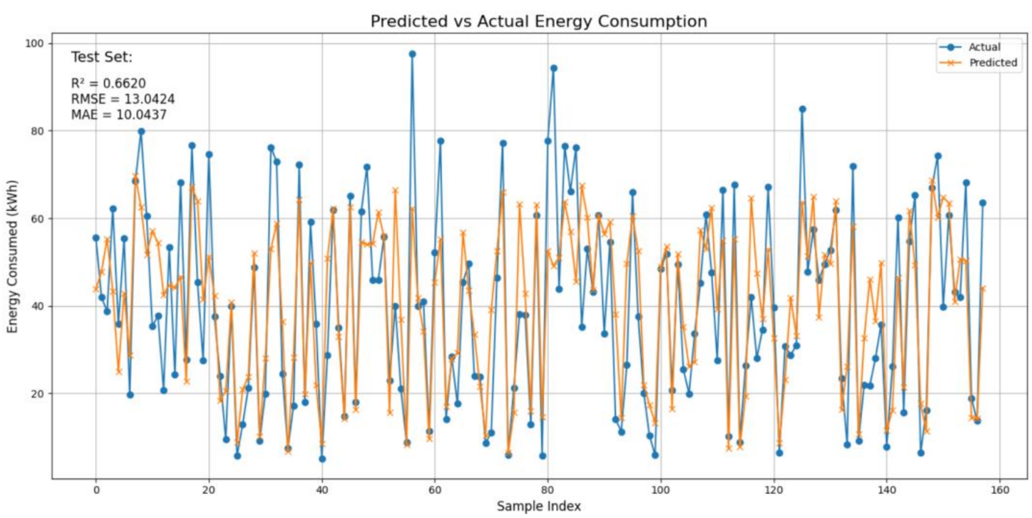

Figure 3. Comparison Between Predicted and Actual Energy Consumption Using RF Model

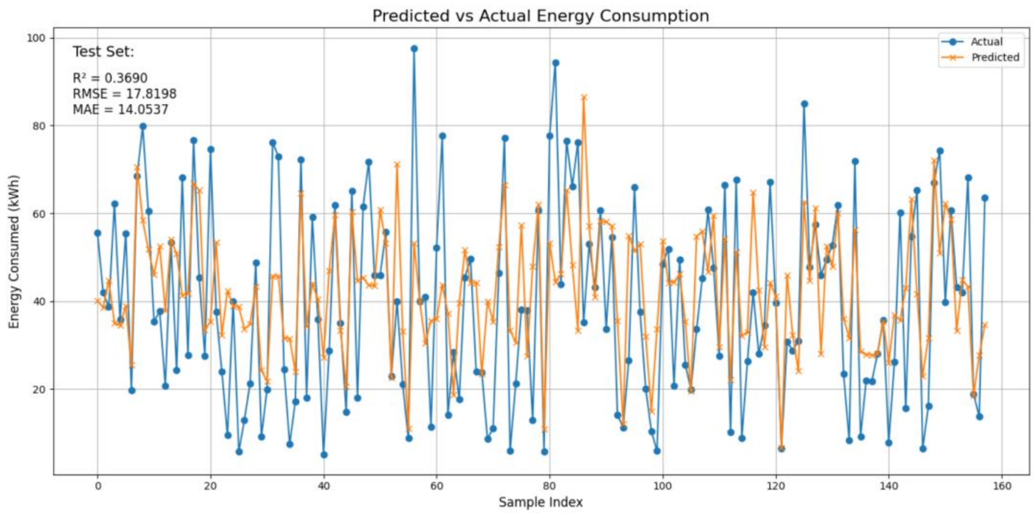

Figure 4. Comparison Between Predicted and Actual Energy Consumption Using SVR Model

4.2 EVCP Dataset: Charging Duration Prediction

A similar form of modelling was applied to predict the time of charging periods. Figure 5 shows the LGBM model's excellent performance in forecasting charging duration on the test set. Predictions closely match actual values, demonstrating great accuracy with an MAE of 0.1974, RMSE of 0.2987, and R^2 of 0.9152. The models have been finetuned as before through GridSearchCV with TimeSeriesSplit. Its performance on this task showed that LGBM performed better as compared to the other techniques. The desired values of the measures used in this model were MAE of 0.1974, RMSE of 0.2987, and a significantly high R^2 of 0.9152, thus exhibiting a higher level of accuracy and off-measured generalization in the case of using in new data. The LGBM model performance results in terms of charging duration are shown in Figure 5. Compared to [7], the RF model yielded a R^2 value of 0.47 in predicting charging time. However, our RF model recorded a substantially higher value of $R^2 = 0.7801$, with an MAE of 0.3727 and RMSE of 0.4811, suggesting an improved goodness of fit in forecasting. This shows that we have expanded the predictive nature and practical relevance of earlier studies to acquire a more complete understanding of EV charging dynamics, that can better support infrastructure planning and user engagement tactics.

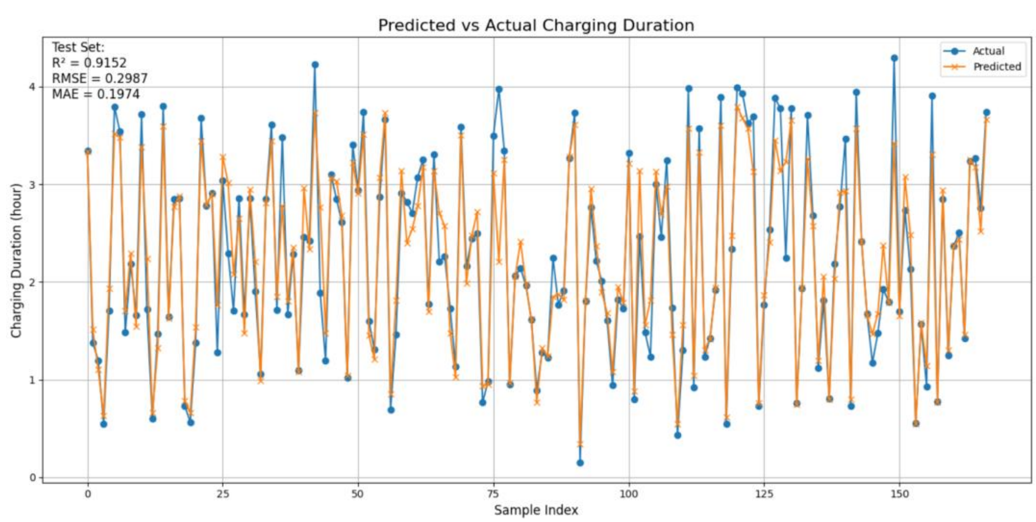

Figure 5. Comparison Between Predicted and Actual Charging Duration Using LGBM Model

Figure 6 shows the corresponding RF model performance of charging duration. This figure compares the RF model's predictions for charging duration to the actual values on the test set. Despite strong overall alignment, the model's performance is characterized by an MAE of 0.3727, RMSE of 0.4811, and R^2 of 0.7801. The SVR was not too good with the MAE of 0.4376, RMSE of 0.5713 and the $R^2$of 0.6900. Figure 7 illustrates the graphical output of charging period by using SVR model. All these findings help to confirm the fact that complexities of EV charging datasets are best represented by a tree-based ensemble model. On the test set, this figure shows how well the SVR model predicted charging duration. The forecasts are less consistent with actual values, with an MAE of 0.4376, RMSE of 0.5713, and R^2 of 0.6900.

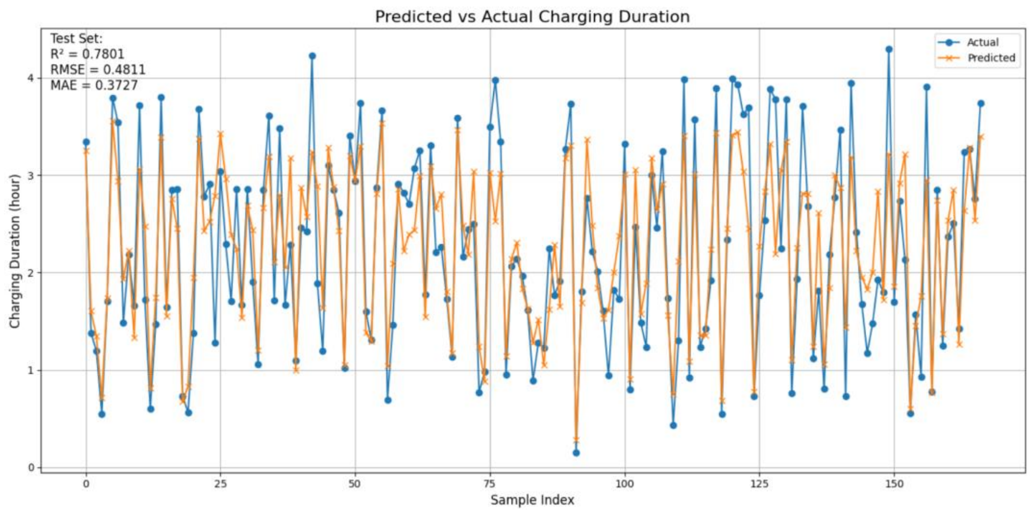

Figure 6. Comparison Between Predicted and Actual Charging Duration Using RF Model

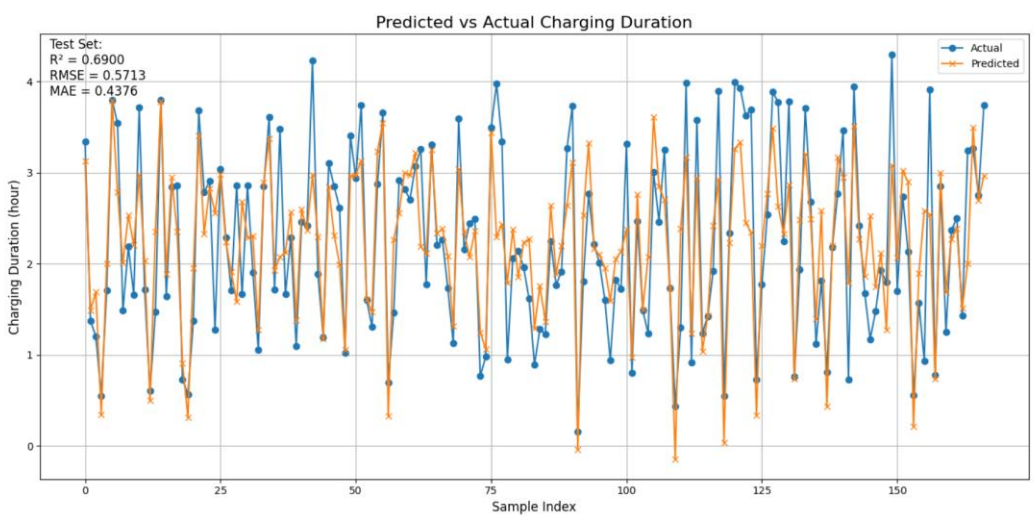

Figure 7. Comparison Between Predicted and Actual Charging Duration Using SVR Model

4.3 EVCP Dataset: Recommendations for Improving EV Charging Infrastructure

Several infrastructural suggestions were developed. The analysis of usage trends across charger types revealed that DC Fast Chargers are required for brief, high-speed sessions, particularly in high-traffic regions. Besides that, Level 1 and Level 2 chargers are better suited for longer-duration charging in residential or low-demand zones. In addition, the strategic deployment of charger types should be adapted per city. For example, Houston and Chicago would benefit from greater Level 2 charger deployments, but New York would require an urgent expansion in public fast-charging choices due to its high EV density.

Furthermore, infrastructure should be consistent with user behaviour patterns. Commuters often charge for short periods of time on weekday evenings, indicating the necessity for Level 2 chargers in commercial districts and residential regions. Long-distance travellers have a high weekend activity rate and prefer late-night charging. This indicates the need of fast-charging stations near roads and trip destinations. Casual drivers, with more diverse habits, would benefit from the ubiquitous availability of Level 1 and Level 2 charges in public places such as malls and parks.

Finally, spatial and temporal heatmaps of charging activity highlight the importance of dynamic capacity planning in metropolitan hotspots such as Los Angeles and San Francisco. While many cities already have extensive infrastructure, increasing EV adoption would demand a commensurate increase in charging station availability to avoid congestion and downtime.

4.4 Colorado Dataset: Energy Usage Prediction

LGBM produced the most precise forecast as compared to the other two models being explored with the lowest RMSE, MAE, and MAPE on the test set. LGBM generated 6051.90 in MAE, 6790.83 in RMSE, and 27.66% in MAPE. The model successfully caught the intricate temporal connections in the data, resulting in higher predicted accuracy. Figure 8 illustrates the LGBM forecast alongside the historical energy consumption data. This figure displays the LGBM model's monthly energy usage forecast for the Colorado dataset, showing its alignment with historical data. The model effectively captures temporal trends, yielding an MAE of 6051.90, RMSE of 6790.83, and MAPE of 27.66%, indicating precise predictions for the test period. Though RF is a decent ensemble algorithm, it performed noticeably worse than LGBM in this forecasting task: 8924.97, 9556.52, and 41.47% were the RMSE, MAE, and MAPE of the model, respectively.

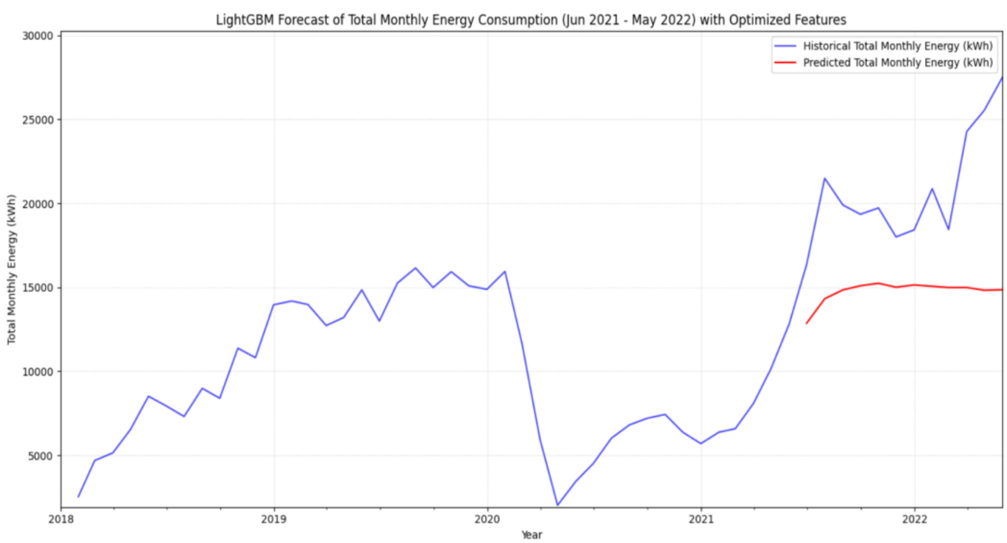

Figure 8. Energy Usage Forecasted Using LGBM for 1 Year

Figure 9 presents the RF forecast compared with the historical energy usage. This image depicts the RF model's monthly energy usage projection for the Colorado dataset. Compared to LGBM, its predictions are less accurate, with an MAE of 8924.97, RMSE of 9556.52, and MAPE of 41.47%, indicating a worse capacity to capture complicated temporal dynamics in this dataset. The performance was the worst when using SVR, as it had the largest RMSE, MAE, and MAPE values of 9156.20, 9698.64, and 42.73%, respectively. Figure 10 shows the SVR forecast against the historical energy consumption data. Such disparities help to emphasize the fact that SVR is not highly effective in terms of capturing the complex temporal dependencies and the feature distributions that the energy consumption data set contains, even when the feature scaling of standards scaler has been applied. This graph depicts the SVR model's monthly energy consumption projection for the Colorado dataset. SVR has the least exact performance among the models, with an MAE of 9156.20, RMSE of 9698.64, and MAPE of 42.73%, highlighting its shortcomings in capturing the intricate temporal correlations of the energy consumption data.

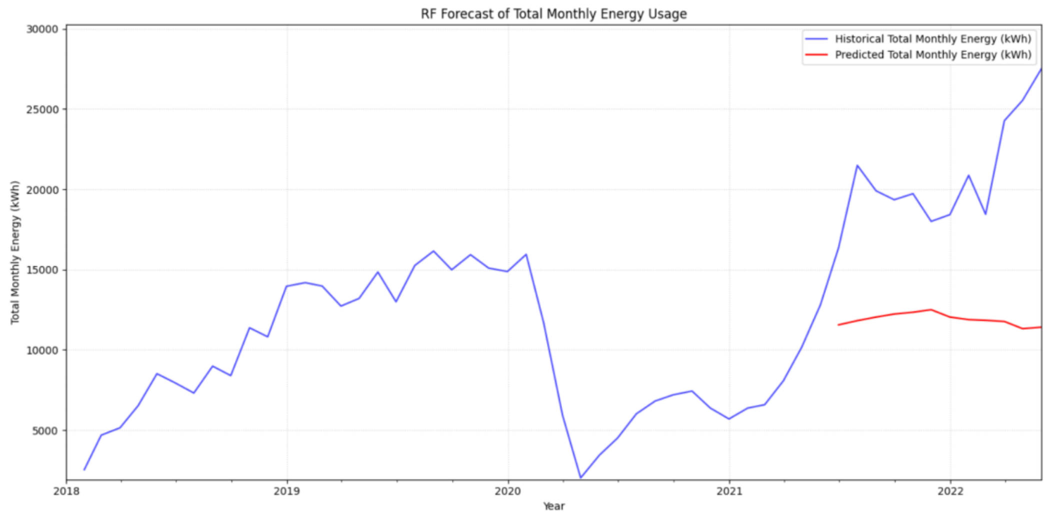

Figure 9. Energy Usage Forecasted Using RF for 1 Year

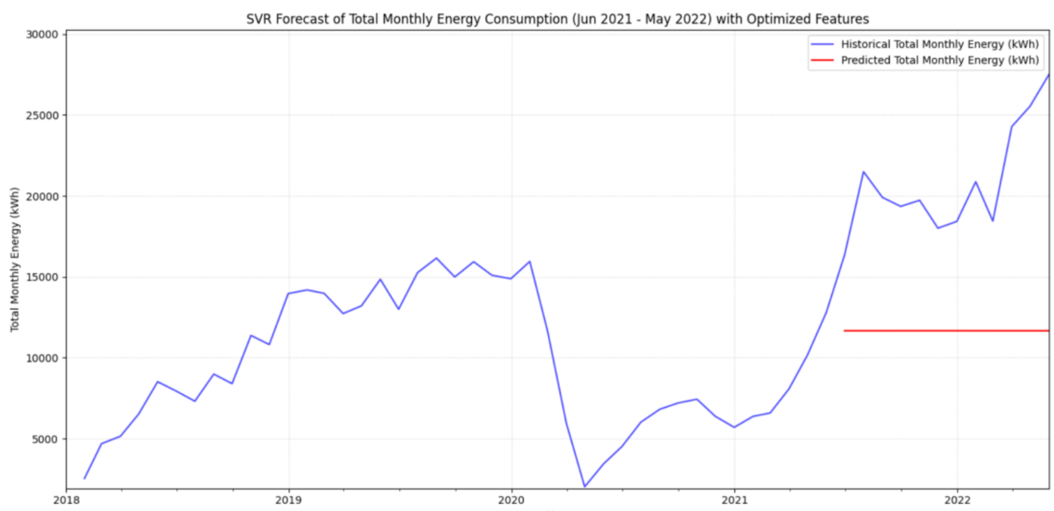

Figure 10. Energy Usage Forecasted Using SVR For 1 Year

To offer a larger perspective for the performance of the ML models, a comparison was made with the results from [42] who used classical time series models on the same dataset. Table 1 shows the findings of the results, along with the best-performing model from this study (LGBM). The SARIMA model from [42] had much lower RMSE, MAE, and MAPE values than all the ML models evaluated in this study, including the best-performing LGBM model. This suggests that for this specific type of EV charging station energy data, traditional time series models such as SARIMA, which are specifically designed to capture autoregressive, integrated, and moving average components as well as seasonality, may be better suited or tuned for this dataset than the ML models examined here. The performance gap might be attributed mostly to the COVID-19 pandemic's unusual impact on energy consumption patterns, which caused severe non-stationarity and sudden fluctuations that traditional ML models cannot capture without specific intervention modelling. The pandemic generated significant changes in human behaviour and economic activity, resulting in concept drift and making past patterns less valid for future predictions [43]. ML methods, which learn from previous patterns, have a difficult time generalizing from such a disturbed and non-linear historical background to anticipate future trends. Traditional time series models may handle this unrepresented disruption differently or be more resilient to it. Despite SARIMA's superior performance in the referenced paper, the ML models in this study continue to provide valuable insights into feature importance and serve as a reliable alternative option, particularly in scenarios where data is less strictly time dependent.

Table 1. Comparative Performance

Model	MAE	RMSE	MAPE
ARMA	2885.26	3227.79	25%
ARIMA	2484.33	3022.34	22%
SARIMA	1869.13	1572	16%
LGBM	6051.90	6790.83	27.66%

Based on the results and findings, our contributions include leveraging two distinct real-world datasets: the EVCP dataset (session-level) and the Colorado dataset (monthly aggregated). This ensures the generalizability and robustness of the predictive models across varying granularities and contexts, providing a broader scope than many single-dataset studies. Besides that, we create and test models for both energy consumption and charging duration, giving a fuller knowledge of EV charging dynamics required for comprehensive infrastructure development. This approach not only goes beyond forecasting a single parameter but provides a more integrated solution. In addition, the findings are translated into practical and insightful suggestions for optimizing EV charging infrastructure rollout, improving user engagement techniques, and influencing policy development. The findings not only directly impact choices on charger type allocation, provide strategic placement depending on user behaviour (e.g., catering to commuters or long-distance travellers), but enable dynamic capacity planning in high-traffic regions. This practical relevance demonstrates how our findings may immediately help to better support urban transportation and sustainable energy governance.

5. CONCLUSION

This study successfully achieved its primary goals of analysing EV charging behaviour, developing predictive models for energy consumption and charging duration, and providing data-driven infrastructure recommendations. For the EVCP dataset, RF was proven to be the best for energy consumption prediction whereas LGBM excelled in charging duration prediction. Tree-based models consistently outperformed SVR for both tasks. For the Colorado dataset, LGBM was the best-performing ML model for monthly energy forecasting. However, it is important to note that a traditional time series model from prior research showed superior accuracy as compared to the tested ML models for this dataset. Overall, the research findings provide valuable insights for strategic charger deployment, capacity optimization, and tailoring infrastructure to diverse user needs, contributing to smarter mobility planning.

In spite of its strengths, there are some limitations in this study. First, while the dataset is extensive, it only covers 55 days of activity and may not reflect seasonal trends or larger regional variety. Second, certain attributes were imputed or derived, which may result in bias or estimation inaccuracy. Third, the data does not contain real-time factors such as the cost of power, queue of chargers and traffic information, which may have an impact on user behaviour and subsequent model accuracy. Future research could enhance the current project with the addition of live-streaming information, the use of other sources, such as weather APIs, energy tariffs, or GPS-based travel history, as well as applying more elaborate deep learning architectures.

In conclusion, the paper outlines the great potential of ML to enable data-based decision-making in the framework of the EV charging infrastructure. By scrutinizing systematically constructed models as well as testing their predictions in empirical conditions, the study provides a solid foundation to smart-mobility design as well as sustainable-energy governance.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for the suggestions to improve the paper.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Yi Xuan Law: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation;
Kah Ong Michael Goh: Project Administration, Supervision, Writing – Review & Editing;
Check Yee Law: Project Administration, Writing – Review & Editing;
Connie Tee: Review & Editing;
Wee Sek Yong: Review & Editing;
S M Hasan Mahmud: Co-Supervision, Review & Editing.

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/. This research did not involve any human participants, animal experiments, or data collected from social media platforms. Therefore, ethical approval and informed consent were not required.

DATA AVAILABILITY

The data that support the findings of this study are openly available in EVCP at https://www.kaggle.com/datasets/valakhorasani/electric-vehicle-charging-patterns/data.

REFERENCES

[1] International Energy Agency (IEA), "Global EV outlook 2024," IEA, Paris, 2024. [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2024. [Accessed: Jun 12, 2025]

[2] M. Bassey, "Adoption of Electric Vehicles (EVs) as a sustainable transportation solution: Reshaping the trends in conventional automobile applications Aniekan Essienubong Ikpe Akwa Ibom State Polytechnic," IEA, Paris, 2024. [Online]. Available: www.iksadyayinevi.com.

[3] International Energy Agency (IEA), "Global EV outlook 2024 moving towards increased affordability," IEA, Paris, 2024. [Online]. Available: www.iea.org.

[4] International Energy Agency (IEA), "Global EV outlook 2023: Catching up with climate ambitions," IEA, Paris, 2023. [Online]. Available: www.iea.org.

[5] S. S. G. Acharige, M. E. Haque, M. T. Arif, N. Hosseinzadeh, K. N. Hasan, and A. M. T. Oo, "Review of electric vehicle charging technologies, standards, architectures, and converter configurations," *IEEE Access*, vol. 11, pp. 41218–41255, 2023, doi: 10.1109/ACCESS.2023.3267164.

[6] A. Alsabbagh, and C. Ma, "Distributed charging management of electric vehicles considering different customer behaviors," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5119–5127, Aug. 2020, doi: 10.1109/TII.2019.2952254.

[7] A. Almaghrebi, K. James, F. Al Juheshi, and M. Alahmad, “Insights into household electric vehicle charging behavior: Analysis and predictive modeling,” *Energies*, vol. 17, no. 4, pp. 925, Feb. 2024, doi: 10.3390/en17040925.

[8] J. Jayaram, J. Chetan, and B. Nayak, “Electric vehicle health monitoring with electric vehicle range prediction and route planning,” *Journal of Informatics and Web Engineering*, vol. 3, no. 1, pp. 265–276, Feb. 2024, doi: 10.33093/jiwe.2024.3.1.18.

[9] N. Wang, Y. Lyu, H. Tian, and Y. Guo, “Research on charging patterns of electric taxis based on high-dimensional cluster analysis: A case study of Hangzhou, China,” *Transportation*, 2025, doi: 10.1007/s11116-024-10574-6.

[10] Y. Xiong, B. Wang, C.-C. Chu, and R. Gadh, “Electric vehicle driver clustering using statistical model and machine learning,” in *2018 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, 2019, doi: 10.1109/PESGM.2018.8586132.

[11] S. Shahriar, A. R. Al-Ali, A. H. Osman, S. Dhou, and M. Nijim, “Prediction of EV charging behavior using machine learning,” *IEEE Access*, vol. 9, pp. 111576–111586, 2021, doi: 10.1109/ACCESS.2021.3103119.

[12] Y. Wang *et al.*, “The analysis of electrical vehicles charging behavior based on charging big data,” *2019 IEEE 4th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2019*, pp. 63–67, Apr. 2019, doi: 10.1109/ICCCBDA.2019.8725776.

[13] J. Yusuf, S. Ula, and A. S. M. J. Hasan, “Analyses and applications of plug-in electric vehicle charging stations’ user behavior in a large university campus community,” *Proceedings - 2020 International Conference on Smart Grids and Energy Systems, SGES 2020*, pp. 928–933, Nov. 2020, doi: 10.1109/SGES51519.2020.00170.

[14] X. Li, Y. Sun, M. Shi, and Y. Jia, “Multi-stage charging recommendation of charging station considering user’s charging behavior,” *2023 5th Asia Energy and Electrical Engineering Symposium, AEEES 2023*, pp. 1247–1251, 2023, doi: 10.1109/AEEES56888.2023.10114356.

[15] A. Almaghrebi, S. Shom, F. Al Juheshi, K. James, and M. Alahmad, “Analysis of user charging behavior at public charging stations,” *ITEC 2019 - 2019 IEEE Transportation Electrification Conference and Expo*, Jun. 2019, doi: 10.1109/ITEC.2019.8790534.

[16] C. Chen, Y. Song, X. Hu, & I. Guardiola, “Analysis of electric vehicle charging behavior patterns with function principal component analysis approach”, *Journal of Advanced Transportation*, vol. 2020, pp. 1-12, 2020, doi: 10.1155/2020/8850654.

[17] A. Skuza and R. Jurecki, “Analysis of factors affecting the energy consumption of an EV vehicle - A literature study”, *Iop Conference Series Materials Science and Engineering,* vol. 1247, no. 1, pp. 012001, 2022, doi: 10.1088/1757-899x/1247/1/012001.

[18] W. Yao, Z. Yu, and S. Gao, “Research on data-driven cluster analysis of electric vehicle charging sessions,” *2023 IEEE Sustainable Power and Energy Conference, iSPEC 2023*, 2023, doi: 10.1109/ISPEC58282.2023.10402864.

[19] S. Singh, B. Vaidya, & H. Mouftah, “Smart EV charging strategies based on charging behavior”, *Frontiers in Energy Research*, vol. 10, 2022, doi: 10.3389/fenrg.2022.773440.

[20] D. Yan, C. Y. Luo, Y. Li, B. Zhu, M. L. Yan, and S. L. Yao, “Charging behavior analysis based on BIRCH clustering,” *2022 12th International Conference on Power and Energy Systems, ICPES 2022*, pp. 450–454, 2022, doi: 10.1109/ICPES56491.2022.10072610.

[21] J. H. Lee, D. Chakraborty, S. J. Hardman, and G. Tal, “Exploring electric vehicle charging patterns: Mixed usage of charging infrastructure,” *Transportation Research Part D Transport and Environment*, vol. 79, pp. 102249, Feb. 2020, doi: 10.1016/J.TRD.2020.102249.

[22] M. van der Kam, W. van Sark, and F. Alkemade, "Multiple roads ahead: How charging behavior can guide charging infrastructure roll-out policy," *Transportation Research Part D Transport and Environment*, vol. 85, pp. 102452, Aug. 2020, doi: 10.1016/J.TRD.2020.102452.

[23] A. Alsarhan, A. Alnatsheh, M. Aljaidi, T. Al Makkawi, M. Aljamal, and T. Alsarhan, "Optimizing electric vehicle charging infrastructure through machine learning: A study of charging patterns and energy consumption," *International Journal of Interactive Mobile Technologies* , vol. 18, no. 21, pp. 149–170, Nov. 2024, doi: 10.3991/ijim.v18i21.50843.

[24] J. Zhang, Z. Wang, P. Liu, and Z. Zhang, "Energy consumption analysis and prediction of electric vehicles based on real-world driving data," *Applied Energy*, vol. 275, Oct. 2020, doi: 10.1016/j.apenergy.2020.115408.

[25] S. Shahriar, A. R. Al-Ali, A. H. Osman, S. Dhou, and M. Nijim, "Machine learning approaches for EV charging behavior: A review," *IEEE Access*, vol. 8, pp. 168980–168993, 2020, doi: 10.1109/ACCESS.2020.3023388.

[26] T. Etem, "Machine learning approaches for predicting electric vehicle charging demand and energy consumption," *8th International Artificial Intelligence and Data Processing Symposium, IDAP 2024*, pp. 1-7, 2024, doi: 10.1109/IDAP64064.2024.10710701.

[27] Q. Zhu, Y. Huang, C. F. Lee, P. Liu, J. Zhang, and T. Wik, "Predicting electric vehicle energy consumption from field data using machine learning," *IEEE Transactions on Transportation Electrification*, vol. 11, no. 1, p. 2120-2132, 2024, doi: 10.1109/TTE.2024.3416532.

[28] H. Mediouni, A. Ezzouhri, Z. Charouh, K. El Harouri, S. El Hani, and M. Ghogho, "Energy consumption prediction and analysis for electric vehicles: A hybrid approach," *Energies*, vol. 15, no. 17, pp. 6490, Sep. 2022, doi: 10.3390/en15176490.

[29] D. A. Renata *et al.*, "Modeling of electric vehicle charging energy consumption using machine learning," *2021 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2021*, pp. 1-6, 2021, doi: 10.1109/ICACSIS53237.2021.9631324.

[30] H. Rathore, H. K. Meena, and P. Jain, "Prediction of EV energy consumption using random forest and XGBoost," in *Proceedings - 2nd International Conference on Power Electronics and Energy, ICPEE 2023*, Institute of Electrical and Electronics Engineers Inc., pp. 1-6, 2023, doi: 10.1109/ICPEE54198.2023.10060798.

[31] Z. M. Sarkin Adar, A. Alhayd, and G. Todeschini, "Predicting EV charging duration using machine learning and charging transactions at three sites," in *Proceedings of the IEEE International Conference on Industrial Technology*, Institute of Electrical and Electronics Engineers Inc., pp. 1-6, 2024. doi: 10.1109/ICIT58233.2024.10540858.

[32] I. Ullah, K. Liu, T. Yamamoto, M. Zahid, and A. Jamal, "Prediction of electric vehicle charging duration time using ensemble machine learning algorithm and Shapley Additive Explanations," *International Journal of Energy Research*, vol. 46, no. 11, pp. 15211–15230, Sep. 2022, doi: 10.1002/er.8219.

[33] I. Ullah, K. Liu, T. Yamamoto, M. Shafiullah, and A. Jamal, "Grey Wolf Optimizer-Based machine learning algorithm to predict electric vehicle charging duration time," *Transportation Letters*, vol. 15, no. 8, pp. 889–906, 2023, doi: 10.1080/19427867.2022.2111902.

[34] X. Hu, and B. Sikdar, "Energy consumption prediction of electrical vehicles through transformation of time series data," *2023 IEEE 3rd International Conference on Sustainable Energy and Future Electric Transportation, SeFet 2023*, pp. 1-7, 2023, doi: 10.1109/SEFET57834.2023.10245102.

[35] A. A. Visaria, A. F. Jensen, M. Thorhauge, and S. E. Mabit, "User preferences for EV charging, pricing schemes, and charging infrastructure," *Transportation Research Part a Policy and Practice*, vol. 165, pp. 120–143, Nov. 2022, doi: 10.1016/J.TRA.2022.08.013.

[36] F. Lo Franco, V. Cirimele, M. Ricco, V. Monteiro, J. L. Afonso, and G. Grandi, "Smart charging for electric car-sharing fleets based on charging duration forecasting and planning," *Sustainability*, vol. 14, no. 19, pp. 12077, Oct. 2022, doi: 10.3390/su141912077.

[37] V. Khorasani, “Electric vehicle charging patterns,” Kaggle, 2025. [Online]. Available: https://www.kaggle.com/datasets/valakhorasani/electric-vehicle-charging-patterns/data. [Accessed: Jun. 11, 2025].

[38] A. Bansal, K. Balaji, and Z. Lalani, “Temporal encoding strategies for energy time series prediction,” *preprint arXiv:2503.15456*, Mar. 2025, doi: 10.48550/arXiv.2503.15456. [Online]. Available: https://arxiv.org/pdf/2503.15456.

[39] B. K. S. Bondili, S. P. Bolagani, and R. Thatikonda, “Comparative analysis of ml algorithms for forecasting time series stock returns,” *4th International Conference on Sentiment Analysis and Deep Learning, ICSADL 2025 - Proceedings*, pp. 727–732, 2025, doi: 10.1109/ICSADL65848.2025.10933249.

[40] G. N. Ahmad, H. Fatima, Shafiullah, A. Salah Saidi, and Imdadullah, “Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV,” *IEEE Access*, vol. 10, pp. 80151–80173, 2022, doi: 10.1109/ACCESS.2022.3165792.

[41] City of Boulder, “Electric vehicle charging station data,” *City of Boulder Open Data*. [Online]. Available: https://open-data.bouldercolorado.gov/datasets/95992b3938be4622b07f0b05eba95d4c/explore. [Accessed: Jun. 11, 2025].

[42] K. C. Akshay, G. H. Grace, K. Gunasekaran, and R. Samikannu, “Power consumption prediction for electric vehicle charging stations and forecasting income,” *Scientific Reports*, vol. 14, no. 1, pp. 6497, 2024, doi: 10.1038/s41598-024-56507-2.

[43] M. Muller and M. Salathe, “Addressing machine learning concept drift reveals declining vaccine sentiment during the COVID-19 pandemic,” *preprint, arXiv:2012.02197*, Dec. 2020, doi: 10.48550iv.2012.02197.

BIOGRAPHIES OF AUTHORS

	**Yi Xuan Law** is a dedicated student at Multimedia University, currently pursuing a Bachelor of Information Technology (Hons) in Business Intelligence & Analytics within the Faculty of Information Science & Technology. With a keen interest in analytical modelling and its applications, her final year project explores complex patterns in EV charging behaviour, aiming to develop predictive models for energy consumption and charging duration. Her work contributes insights for optimizing charging infrastructure and supporting sustainable energy governance. She can be contacted at email: alicialaw38@gmail.com.
	**Michael Goh Kah Ong** is an Associate Professor at Multimedia University, specializing in AI, biometrics, machine learning, bioinformatics, and cybersecurity. His research has appeared in top-tier journals such as IEEE Access, Applied Soft Computing, and Expert Systems with Applications. He has received prestigious awards including the ITEX Gold Medal, RICES Gold Medal, and iNVENTX Award. His recent work focuses on advancing intelligent systems in healthcare, security, and computer vision through high-impact, interdisciplinary research and innovation. He can be contacted at email: michael.goh@mmu.edu.my.
	**Check Yee Law** is a lecturer at Faculty of Information Science and Technology (FIST), Multimedia University (MMU), Melaka, Malaysia. Her research interests span the fields of teaching and learning to development of systems, software, and mobile applications. Topics of interest include but are not limited to educational technology, information systems, user-centred design, visual analytics, Internet of Things (IoT), and mobile computing. She can be contacted at email: cylaw@mmu.edu.my.

	**Connie Tee** received both the MSc (IT) and PhD (IT) degrees from Multimedia University in 2005 and 2015, respectively. She is a Professor in the Faculty of Information Science and Technology Multimedia University since 2021. She is currently holding the position of the Dean of the Institute for Postgraduate Studies. Her research interests include computer vision, machine learning, deep learning and image processing. She is a senior member of the IEEE. She can be contacted at email: tee.connie@mmu.edu.my.
	**Yong Wee Sek** is a researcher at the Faculty of Artificial Intelligence and Cyber Security (FAIX), Universiti Teknikal Malaysia Melaka. His research interests include technology adoption, Internet of Things (IoT), mobile computing, smart farming, and supply chain innovation. He has published in high-impact journals such as IEEE Access, IET Computer Vision, and AI Open. He actively contributes as a reviewer for journals including IEEE Access, F1000, and Brain Informatics. He can be contacted at email: ywsek@utem.edu.my.
	**S M Hasan Mahmud** is an Associate Professor in the Department of Software Engineering at Daffodil International University. His research areas are Machine Learning, Data Science, and Image Processing. He can be contacted at email: drhasan.swe@diu.edu.bd.