Advancing sustainable mobility

Dynamic predictive modeling of charging cycles in electric vehicles using machine learning techniques and predictive application development

Paneru, Biplov; Mainali, Durga Prasad; Paneru, Bishwash; Sapkota, Sanjog Chhetri

DOI
10.1016/j.sasc.2024.200157

Publication date
2024

Document Version
Final published version

Published in
Systems and Soft Computing

Systems and Soft Computing 6 (2024) 200157

Contents lists available at ScienceDirect

Systems and Soft Computing

journal homepage: www.journals.elsevier.com/soft-computing-letters

Advancing sustainable mobility: Dynamic predictive modeling of charging cycles in electric vehicles using machine learning techniques and predictive application development

Biplov Paneru [a,*], Durga Prasad Mainali [b,c], Bishwash Paneru [d], Sanjog Chhetri Sapkota [e]

[a] Department of Electronics and Communication Engineering, Nepal Engineering College, Pokhara University, Bhaktapur, Nepal
[b] Department of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, Netherlands
[c] Department of Technology, Innovation and Society, The Hague University of Applied Science, The Hague, Netherlands
[d] Department of Applied Sciences and Chemical Engineering, Institute of Engineering, Pulchowk Campus, Tribhuvan University, Lalitpur, Nepal
[e] Nepal Research and Collaboration Center, Nepal

ARTICLE INFO

Keywords:
Electric vehicles (EVs)
Charging cycles
Battery management system (BMS)
Regression algorithm
Tkinter application

ABSTRACT

The main goal in this research is to train various machine learning models to predict charging cycles in EV Electric Vehicles) battery systems. The considered models are gradient boosting, random forests, decision trees, and linear regression. Each of these was assessed based on its R-squared score, which is an important statistical measure in indicating the variance proportion yielded by the model. In contrast, the Random Forest model significantly improved, with an R-squared value of 0.83, thereby doing an excellent job in capturing nuances of the data. Only surpassed by the Gradient Boosting model at an astonishing R-squared score of 0.87, it is this excellent score that underlines its capability to predict the outcome quite accurately by modeling complex interrelations. In other words, gradient boosting outran the rest and provided the most robust results concerning drivers of students' performance. It also underlines how important choosing a good model is in educational analytics in order to increase the accuracy of the predictions. The use of these models in the proposed EV Battery Charging Cycle Predictor App results in accurate predictions to aid schedule maintenance and energy-related decisions. This research brings light to the future of advanced machine learning methods in enhancing the battery efficiencies of EVs and the development of electric mobility technologies. It is possible that the future work will imply the additional inclusion of real data and the integration of the application to general energy systems.

1. Introduction

The use of electric vehicles (EVs) as a technologically advanced and economically feasible means of lowering greenhouse gas emissions is growing. EVs are silent, simple to run, and do not require the fuel that conventional cars do [1]. The world's transition to more environmentally friendly methods of living has sparked a revolution in transportation, and one of the main principles of this new paradigm is the electrification of automobiles. Since they solve the problem of mobility in everyday routines, automobiles have made a significant contribution to the development of modern civilization. They are essential for personal expression as well as for transportation, and they serve as a link that unites people in a dispersed globe [2]. To minimize greenhouse gas emissions and air pollution, maximize the use of natural energy resources, and safeguard the environment, electric vehicles (EVs) were developed as a substitute for gasoline and diesel vehicles. For electric vehicles (EVs), using power produced from renewable energy sources like wind, water, and sunlight can be one of the most effective ways to lower emissions and combat climate change [3]. The need for infrastructure in rural areas grows. Given that they typically travel greater distances than urban drivers, rural drivers encounter particular difficulties with regard to the location and accessibility of charging stations. There is an increasing need for the few public fast charging facilities that are available in rural infrastructure [4]. With their environmentally friendly propulsion systems, electric vehicles (EVs) have become incredibly popular as important contributors to a more sustainable and

* Corresponding author.
E-mail addresses: biplovp019402@nec.edu.np (B. Paneru), D.P.Mainali@tudelft.nl (D.P. Mainali), rampaneru420@gmail.com (B. Paneru), kshetrisanjog@gmail.com (S.C. Sapkota).

https://doi.org/10.1016/j.sasc.2024.200157
Received 24 July 2024; Received in revised form 6 September 2024; Accepted 1 October 2024
Available online 13 October 2024

greener future. The optimization of battery technologies is at the heart of this shift, and efficient charging cycle management stands out as a crucial component for guaranteeing the longevity and performance of EVs [5,6]. When compared to traditional diesel, hybrid, electric, and alternative fuels, they can frequently lower the energy and emissions from transit bus operations [7]. For battery technology, estimating metrics like state-of-charge (SOC) and charging cycles is an essential job for the battery management system (BMS) [8]. This work explores the relationship between advanced technology and sustainable mobility by using machine learning to predict electric vehicle charging cycles in real-time. This study aims to improve EVs' efficiency and significantly contribute to the larger conversation on sustainable transportation by utilizing predictive modeling. The research's conclusions provide valuable information for developing useful tools, such as an intuitive TKinter GUI application that can be used to give stakeholders in the electric mobility space valuable data. By investigating cutting-edge machine learning techniques, this research seeks to open the door to a future in which sustainable mobility is a concrete reality rather than merely an ideal. Fig. 1 illustrates EV technology.

1.1. Literature review

In order to improve battery electric vehicle (BEV) performance and preserve market competitiveness, research and development (R&D) has become increasingly important. For numerical simulations, MATLAB is frequently used as a tool to analyze and improve overall BEV performance. An overview of future technology developments and market consequences for impending small BEVs, internal combustion engine cars (ICEVs), fuel cell electric vehicles (FCEVs), and hybrid electric vehicles (HEVs) is given in this paper. The analysis includes cost analysis until 2050, market projections, techno-economics of BEVs, and a comparison of important BEV attributes with those of other car classes. In addition, a well-to-wheel comparison of BEVs, HEVs, FCEVs, and ICE cars is investigated [1]. The authors of this book chapter thoroughly examine the most recent methods and algorithms that aim to effectively manage the energy used for EV charging and discharging based on Machine-Learning (ML) techniques. They also note the main differences and similarities among the suggested approaches. This made it possible for them to draw some insightful conclusions and to pinpoint a number of benefits, drawbacks, and issues that come with using these ML-based EV energy management techniques. These issues should be resolved in order to encourage the development of a more effective and affordable power grid [2]. The introduction of electric vehicles (EVs) as a replacement for gasoline and diesel vehicles has been implemented to minimize greenhouse gas emissions, maximize the utilization of fossil fuels, and safeguard the environment. Forecasting EV sales is critical for all parties involved, including automakers, legislators, and fuel suppliers. The quality of the prediction model is highly dependent on the data used in the modeling process. To obtain the necessary data, multiple web crawlers were employed in addition to this data. Long short-term memory (LSTM) and convolutional LSTM (ConvLSTM) models were used to forecast car sales. The hybrid model known as "Hybrid LSTM with two-dimensional Attention and Residual network" has been proposed as a way to improve LSTM performance [3]. But even with the remarkable advancements in nanosatellite technology, there is still a research gap that has to be filled. Even though the body of current literature is rich in accomplishments and possible uses, more in-depth research is still needed to fully address some problems and unrealized promise in this rapidly evolving field. This research gap highlights the ongoing need for comprehensive studies that could advance nanosatellite technology and its revolutionary effects on space exploration and technology. Future trends in the development of the space sector as well as the impacts of artificial intelligence are also included in this study. The study as a whole address the functions of nanosatellites in space development and connects them to the function of AI in this field. An On-Board Diagnostics (OBD) device attached to the driver's vehicle provided the real-world driving cycle (RDC) data used in this study. Our study's findings demonstrated that the suggested machine learning strategy could predict the state of charge (SOC) of lithium-ion batteries with an average accuracy of 95 %. Notwithstanding the significant fluctuations in the RDC data and the existence of noise in the measurements, this high degree of accuracy was attained. With an average error of less than 2 %, the suggested machine learning technique was also able to predict the EVs' Remaining Driving Range (RDR) using the anticipated SOC values [4]. In order to extract the most dependable features—namely, the EIS impedances that have a strong correlation with SOC—from EIS measurements, a feature sensitivity analysis of the data is first performed. The selected features are then fed into the models. The models are intended to create a mapping relationship between the chosen features and the SOC and to train the input features. The findings show that the GPR model's error was determined to be less than 3.8 %. With the use of onboard EIS measurements, this technique can be practically integrated into the battery management system to provide precise SOC measurements for Li-ion batteries and guarantee the correct and effective operation of electric vehicles that run on batteries [5]. This assessment covers LIB's operation, effectiveness, and shortcomings. It gives a general overview of LIB with a focus on the variables that influence their performance and the variables that lead to failures. Lastly, possible lithium battery substitutes for electric vehicles are assessed. The difficulties with these upcoming batteries are also covered. In this study, authors examine research on general issues, future battery technologies, and batteries utilized in electric vehicles. Approaches pertaining to these subjects are contrasted in terms of their benefits, drawbacks, and qualitative elements [9]. The current work

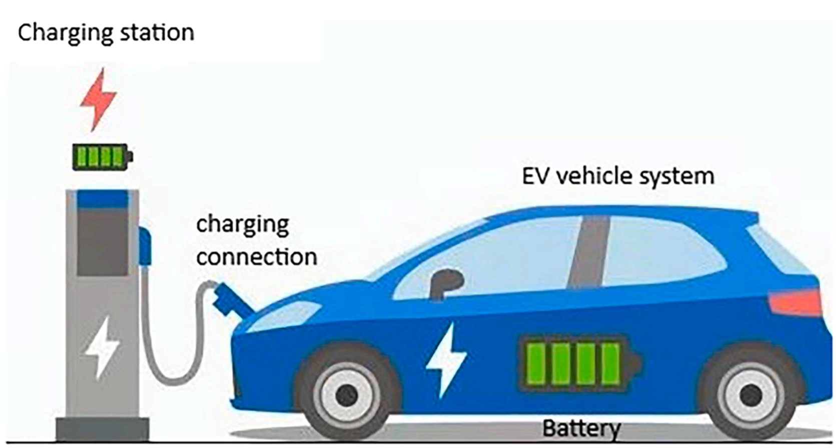

Fig. 1. EV and recharging system.

compares three machine learning algorithms in an effort to identify the best method for estimating the SOC necessary for electric vehicles. The ML model is trained using Agartala city's current car traffic. The findings demonstrate that CG-SVM performs better than the other two ML approaches in terms of performance metrics, including MSE, MAE, R-squared value, and RMSE also seen, when compared to bagged tree regressor and boosted tree regressor. CG-SVM predicts more quickly and requires less training time. thus proving that CG-SVM is a suitable machine learning technique for predicting the necessary SOC of EVs [10]. The authors of this research suggest employing widely used machine learning methods, such as random forest, SVM, XGBoost, and deep neural networks, to forecast the length of an EV session and its energy use by combining past charging data with weather, traffic, and event data. With SMAPE scores of 9.9 % and 11.6 % for session duration and energy usage, respectively, an ensemble learning model has the strongest predictive performance, outperforming previous studies in the literature. The authors show a notable improvement in both predictions when compared to earlier research on the same dataset, and we emphasize the significance of traffic and weather data for charging behavior predictions [11]. The electricity required in EV charging stations (EVCS) can be predicted by the machine learning algorithms. Machine learning (ML) has been used because of its capacity to use past data to learn and spot trends for future decision-making with little to no user participation. Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), K-Nearest Neighbors (KNN), Deep Neural Networks (DNN), and Long Short-Term Memory (LSTM) are the machine learning models employed in this article. Compared to traditional optimization methods like quadratic programming, the usage of ML offers a simpler way to coordinate EVs, and it can be applied more quickly because it uses less processing resources. With a 95 % accuracy rate, LSTM outperformed other models in forecasting the ideal power rating (PR) for EVCS, with RF, DT, DNN, SVM, KNN, and NB following closely behind. Furthermore, with an error rate of $\pm$0.7 %, LSTM was the model with the lowest error rate, ahead of RF, DT, KNN, SVM, DNN, and NB. With ML's increased speed and simplicity, the results from the LSTM model were comparable to those from earlier research that used quadratic programming [12]. The authors provide a sensitivity analysis for a mechanical model used to evaluate the energy requirement of battery-electric cars. This model is widely used in the literature, however its parameters are typically selected carelessly, resulting in erroneous energy demand predictions. They provide a unique parameter priority strategy and quantify its influence on the accuracy of energy demand estimates, allowing for better decision making throughout the model parameter selection phase. They also identify a subset of parameters that must be established in order to attain the necessary estimation accuracy [13]. This study makes use of a dataset obtained from the Hawaii Natural Energy Institute, which includes 14 unique batteries that were exposed to over 1000 cycles under controlled settings. A multi-step technique is used, beginning with data collection and preprocessing and continuing with feature selection and outlier reduction. The RUL prediction model uses machine learning models such as XGBoost, BaggingRegressor, LightGBM, CatBoost, and ExtraTreesRegressor. Feature importance analysis helps to find essential characteristics that affect battery health and longevity. Statistical assessments demonstrate no missing or duplicate data, and removing outliers increases model accuracy. Notably, XGBoost emerged as the most successful algorithm, producing almost flawless predictions. This study emphasizes the importance of RUL prediction for improving battery lifetime management, particularly in applications like as electric cars, by assuring optimal resource usage, cost efficiency, and environmental sustainability [14]. Machine learning processes are distinguished by temporal change and are crucial for channel modeling in vehicle network situations. This study intends to give theoretical underpinnings for machine learning, as well as leading models and methods for addressing the problems of IoV applications. This research conducted a comprehensive evaluation with analytical modeling for offloading mobile edge-computing choices using machine learning and Deep Reinforcement Learning (DRL) techniques for the Internet of Vehicles (IoV). The article assumes a Secure IoV edge-computing offloading architecture with diverse data processing and traffic flow. The proposed analytical model uses the Markov decision process (MDP) and machine learning (ML) to offload the decision process of various task flows in the IoV network control cycle [15]. The discharge voltage curves' changes at different temperatures and C rates are shown by this analysis. Analysis is done on changes in voltage, internal resistance, and SOC at different temperatures and C rates. This opens the door for data-driven approaches to get a SOC estimate that is more accurate. The results of this study indicate that when the C rate of charging is lower, the energy storage capacity increases and when the C rate of discharging is lower, more energy can be delivered. The results of this investigation show that for SoC values less than 90 %, the battery drain is linear. By employing this method, the SoC estimation error is reduced to 0.835 % [8]. This research offers a thorough analysis of the commercial lithium ion battery for electric vehicle's thermal runaway process. The abusive circumstances that can result in thermal runaway have been compiled in order to learn from common incidents. The three types of abuse conditions are thermal, electrical, and mechanical. Across all misuse circumstances, internal short circuit is the most prevalent characteristic. The thermal runaway is caused by a chain reaction process that causes the battery's component materials to decompose one after the other. To understand the mechanics of the chain reactions during thermal runaway, a unique energy release diagram is suggested, which is capable of quantifying the reaction kinetics for all the elements that make up the battery [16]. According to this study, the most important factor in assessing whether or not WV vehicles are feasible is the battery life of EVs. One technique for determining the state of the EV battery is incremental capacity analysis (ICA). The SOH test using ICA demonstrates that cell capacity can be increased with an RMSE of 2 %, meaning that commercial EVs can make use of it [17]. In order to improve the efficiency of the modeling process, this study makes use of multi-objective metaheuristic models such as the Ant Lion Optimizer, Keshtel Algorithm, and Keshtel and Social Engineering Optimizer, which have been applied to non-permutation flow-shop scheduling problems [18]. The model's complexity is addressed by implementing an accelerated Benders decomposition (ABD) algorithm to determine the best course of action. In addition, the bi-objective nature of the model is addressed by the implementation of three multi-objective optimization solution approaches: fuzzy multi-objective programming (FMOP), augmented ε-constraint (AEC), and weighted sum method (WSM). Subsequently, an actual case study in Iran is examined to determine whether the suggested methodology is applicable. Based on numerical results, CPU time is reduced by approximately 10.8 % using the ABD approach [19]. The change in equity and the blockchain network's pleasure are maximized in order to optimize this supply chain. To optimize the suggested mathematical model, GAMS software is utilized in an exact solution method. Based on the study and implementation of this proposed framework, BCSC was able to attain the optimal value for its financial indices, specifically the change in equity, by building a product portfolio based on a robust model. This newly proposed framework shows a 1.5 % reduction in lost revenue and an average inaccuracy of 0.33 % when compared to the integrated portfolio-supply chain model [20]. This article proposes a three-objective Harris Hawks Optimizer (HHO) scheduling algorithm that improve reliability, decrease makespan time, and use less energy. Additionally, energy consumption has been decreased by the use of dynamic voltage frequency scaling (DVFS), which lowers CPU frequency. After that, HHO is contrasted with other algorithms, including Particle Swarm Optimization (PSO), Firefly method (FA), and Whale Optimization Algorithm (WOA), and the suggested method performs better on experimental data. The suggested approach has produced results with an average makespan of 272.5 s, energy usage of 14.95 KJ, and reliability of 83 % [21]. The suggested mathematical model takes into account the fuzzy uncertainty of organ demands and transit time to maximize the overall cost. Furthermore, in order to address the uncertainty in the optimization of this mathematical model, a unique simulation-based optimization is implemented using the credibility theory. Also, a variety of test issues are used to assess the suggested model and solution technique. The numerical results show that in all studied scenarios, the ideal credibility level lies in the range of 0.2–0.6. Furthermore, in the planned organ supply chain, the patient satisfaction rate is higher than the viability rate [22].

A clear research gap exists at the intersection of battery electric vehicle (BEV) technology, energy management applications of machine learning (ML), and electric vehicle (EV) sales forecasting, despite the fact that these topics are well covered in the literature. The current research offers insightful information about the market ramifications, energy management with machine learning, and the techno-economic features of BEVs. But there hasn't been much research done on a comprehensive strategy that combines machine learning (ML)-based energy management with BEV forecasting models to solve practical issues like improving overall sustainability, forecasting sales, and maximizing BEV performance. Additionally, the research gap in the use of ML techniques for charging cycles estimation is addressed by our suggested methodology we hope to improve SOC predictions and incorporate this data into a workable battery management system. This methodology guarantees the accurate and efficient functioning of BEVs, thereby advancing the establishment of a more dependable and sustainable electric mobility environment.

To close this research gap, we propose an approach that focuses on the synergistic integration of machine learning techniques for energy management with advanced forecasting models for predicting EV battery charging phenomenon. Our approach leverages machine learning (ML) insights to provide a complete framework that effectively manages the energy consumed for EV charging and discharging, building on the foundation laid by previous research. Our approach also goes beyond discrete components by offering a unified framework that combines forecasting models, machine learning-based energy management, and market dynamics. A number of machine learning (ML)-based energy management strategies were found in the literature research; these will be incorporated into our conceptual framework to produce a comprehensive forecasting model that can anticipate charging cycles.

2. Methodology

2.1. Data description

Forecasting battery charging cycles has important implications for electric cars' long-term viability, affordability, and driving experience. Our utilization of machine learning techniques to accurately forecast charging cycles makes our contribution to the continuous advancement of electric mobility toward a future of transportation that is more environmentally conscious and resilient possible. Fig. 2 shows the EV vehicles advancement and needs. Each of the 100,000 rows in the artificial dataset inspired by Mendely publication [23] used for this study represents a distinct electric vehicle charging cycle scenario and is developed with a relationship maintaining with the various feature parameters. The artificial dataset portion data is tabulated in Table 1.

The dataset involved in modelling is crucial to understanding the models' performance. The features involved in the modelling process are explained below. Fig. 3 shows the histogram distribution of the applied dataset.

2.1.1. Temperature

A continuous variable that shows the temperature of the surrounding air while the battery is charging. Temperature ranges of 15 to 35 degrees Celsius are used to simulate various climate conditions.

2.1.2. Initial state of health (Initial SoH)

The battery's initial condition at the start of the charging cycle is known as the Initial State of Health, or Initial SoH. The continuous variable has a range of 0.8 to 1.0, with 1.0 denoting a fully functional battery.

2.1.3. Final state of health (Final SoH)

The estimated health at the conclusion of the charging cycle, taking

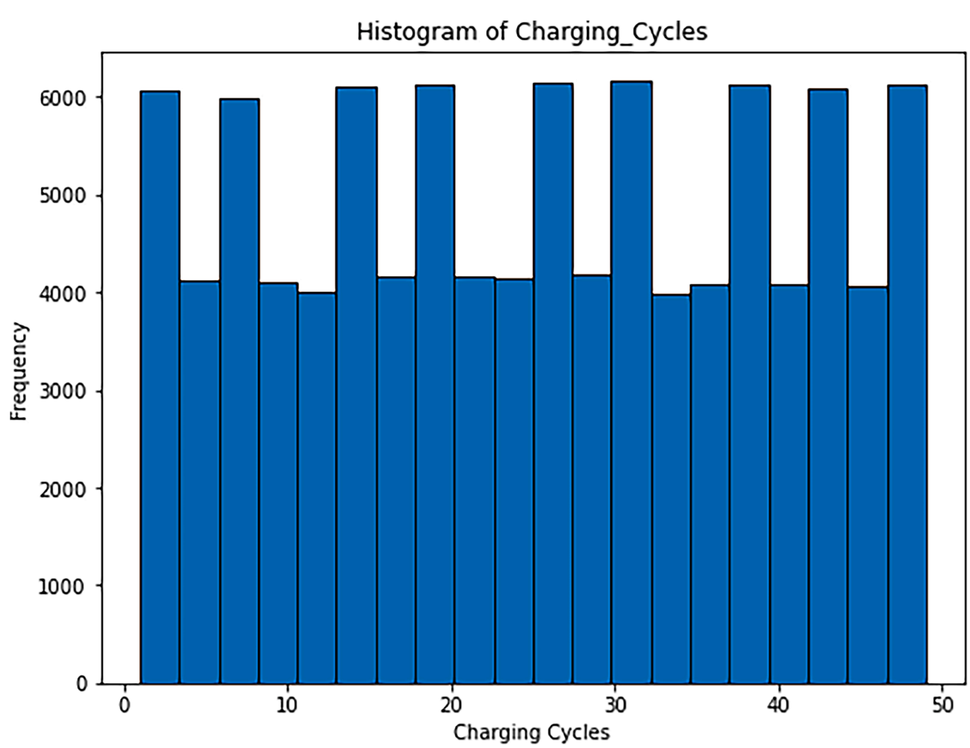

Fig. 2. Histogram of charging cycles.

Table 1
Dataset overview.

Charging_Cycles	Temperature	Initial_SoH	Final_SoH	Initial_SoC	Final_SoC
39	31.471	0.8555	0.873	0.576	0.413
29	29.692	0.894	0.901	0.436	0.294
15	20.594	0.855	0.856	0.644	0.552
43	34.824	0.893	0.875	0.674	0.546
8	24.810	0.874	0.891	0.575	0.432

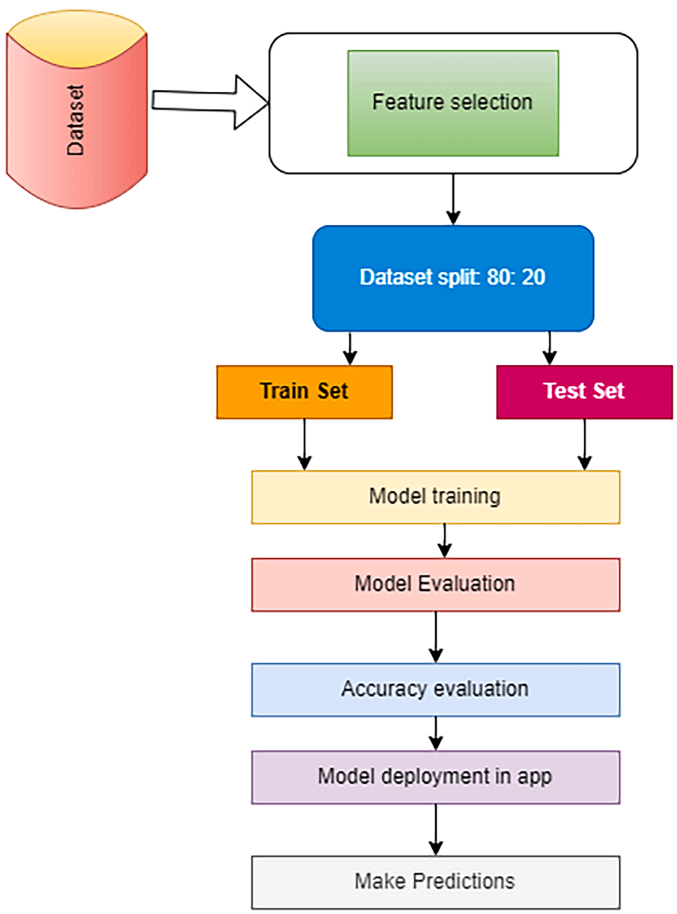

Fig. 3. Proposed Method architecture.

into account the effects of temperature and charging habits.

2.1.4. Initial state of charge (Initial SOC)

This represents the battery's initial charge level at the start of the charging cycle, which can be anywhere between 0.2 and 0.8.

2.1.5. Final state of charge (Final SOC)

The estimated state of charge after the charging cycle, taking patterns of charging and discharging into consideration.

2.1.6. Charging cycles

The target variable representing the number of charging cycles experienced by the electric vehicle's battery. This is the parameter the machine learning model aims to predict.

2.1.7. Splitting the dataset

In this section, the data was prepared for training and testing. It separated the synthetic dataset into training and testing sets after first dividing it into features (X) and the target variable (y) using the train_test_split function.

The histogram is shows the charging cycles variation through a histogram representations. Predicting battery charging cycles is the main goal of this study. A significant factor in determining the lifespan and general operational efficiency of an electric vehicle is its charging cycle, which is commonly defined as the number of times a battery can be charged and discharged before reaching the end of its useful life. For the management of electric vehicle batteries to be optimized, it is essential to comprehend and predict these charging cycles with accuracy. The block diagram representation is shown in Fig. 3.

2.2. Proposed models

Regression [24] is a statistical approach for modeling the connection between a dependent variable and one or more independent variables. It is often used in predictive modeling assignments when the aim is to forecast continuous numerical data. In regression analysis, the goal is to determine the best-fitting line or curve that explains the connection between the independent and dependent variables. This relationship is frequently represented by a mathematical function, such as a linear equation, which may be used to forecast the dependent variable, given new values for the independent variables [25].

2.2.1. Random forest regression

Random Forest Regression is a machine learning approach that applies the ideas of regression to ensemble learning. Ensemble learning entails integrating many models to enhance prediction accuracy. Random Forest Regression creates a collection of decision trees during training. Each decision tree is built from a portion of the training data and a random collection of features. The randomness introduced in the selection of both data and features contributes to the decorrelation of individual trees and fosters ensemble diversity [24]. This is also the case with the random forest classifier, which performs best with a larger number of trees [26]. The architecture of random forest model is shown in Fig. 4.

2.2.2. Support vector machine (SVM)

SVM for regression tasks, the supervised learning algorithm regression is employed. The method finds the hyperplane that maximizes the margin between the nearest data point (support vectors) and the hyperplane that best fits the data points. The goal of SVM regression is to reduce prediction error while preserving strong generalization capacity. It works well in situations when there are more dimensions than samples and in high-dimensional settings.

2.2.3. Categorical bossting (CatBoost)

A gradient-boosting technique called CatBoost Regression was created especially to deal with categorical characteristics in datasets. The acronym for this is "categorical boosting." CatBoost sequentially constructs an ensemble of decision trees and adds more trees to gradually minimize the target function. It eliminates the requirement for preprocessing by automatically handling categorical features. When comparing CatBoost to other gradient boosting techniques, less hyperparameter tweaking is frequently needed.

2.2.4. Gated recurrent unit (GRU)

Though they use gating techniques to regulate information flow

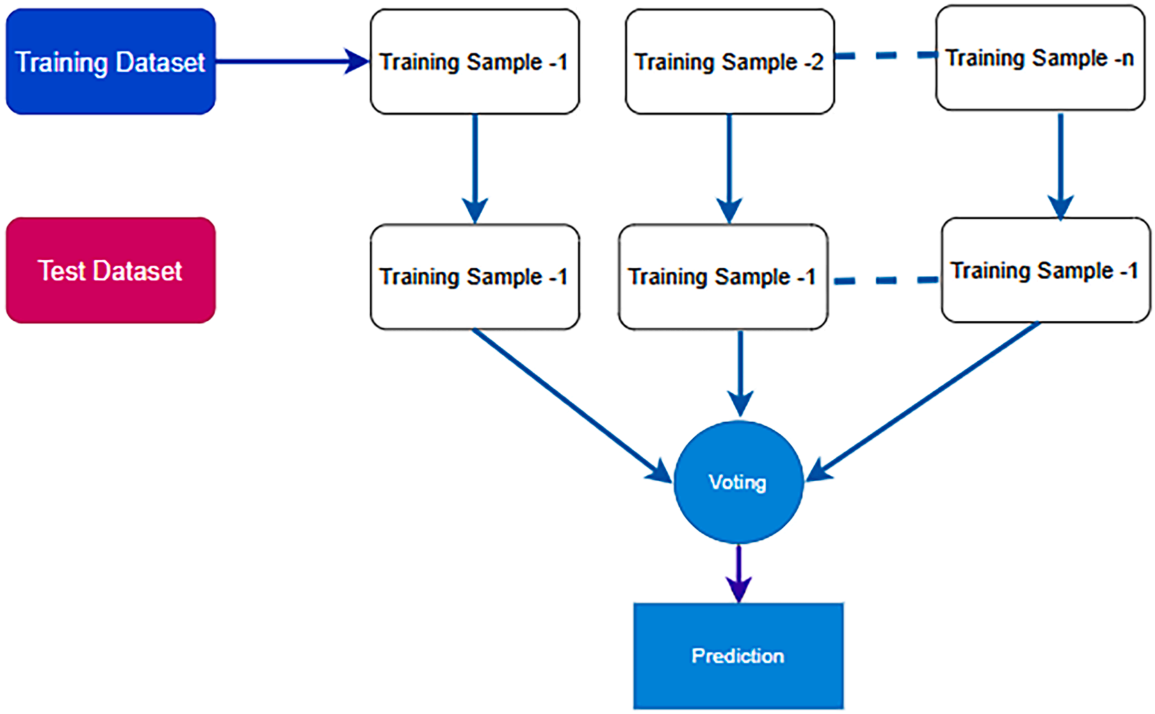

Fig. 4. Random Forest Architecture.

within the network, GRUs resemble conventional RNNs in other ways. These gating mechanisms may help GRUs identify long-term dependencies in sequential data more successfully by selectively updating and resetting their hidden state in response to the input data. The hyperparameter details of all the algorithms are presented in Table 1.

During prediction, the output of each decision tree is combined, usually by average, to give the final forecast. This ensemble strategy produces more robust and accurate predictions than a single decision tree because it decreases overfitting and variation in the model. Random Forest Regression is widely utilized in a variety of fields, including banking, healthcare, and ecology, because to its adaptability, scalability, and capacity to handle complicated datasets with high-dimensional feature spaces.

2.3. Performance metrics

The script evaluated the model's performance, generated predictions on the test set, and reported evaluation results using metrics such as R-squared (R^2) and root mean squared error (RMSE). In the end, the script reported the input attributes and the expected mileage after estimating the mileage for a new input using the trained model. The mean absolute error (MAE) is the average absolute difference between actual and anticipated values, and it provides a clear indicator of the model's average prediction error in the same units as the target variable. The mean squared error (MSE) is calculated by averaging the squared disparities between actual and anticipated values, offering a measure that penalizes greater mistakes more strongly, making it sensitive to outliers. Finally, the root mean squared error (RMSE) is the square root of the MSE, making it a more understandable measure of model performance in the same units as the target variable. The RMSE is important for determining the average size of mistakes, and the indicated component aids in comprehending the metric's underlying computation.

$$R^2 = \frac{\sum_{a=1}^{n}(x_a - x_m)^2 - \sum_{a=1}^{n}(x_a - y_a)^2}{\sum_{a=1}^{n}(x_a - x_m)^2} \tag{1}$$

$$MSE = \frac{1}{n}\sum\nolimits_{a=1}^{n}(x_a - y_a)^2 \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum\nolimits_{a=1}^{n}(x_a - y_a)^2} \tag{3}$$

Where, x_a, y_a, and x_m refer to the observed, measured, and average values for each n observation respectively.

2.4. Stratified cross validation

It is crucial to understand the classifier's performance in a real-world application. One of the most popular techniques in the literature for validating performance is the so-called k-fold cross-validation. This method's primary idea is to randomly divide the items into k groups. However, the validation may be weakened by the data shift issue . We can obtain a more robust validation, if we randomly select from the majority and minority classes in accordance with the original distribution since the partitions' (folds') distributions will resemble the original distribution. This technique is known as stratified cross-validation [27]. The script used in this program model was used in Google Collab's compiler which generated the following results: the datasets overview is shown in Table 2 .

2.5. Tkinter-based application GUI development

The default Python GUI (Graphical User Interface) package is called

Table 2
Details of all hyperparameters.

Models	Hyperparameters settings
SVM	Kernel ='rbf',c = 100, gamma=0.1,epsilon =0.1
RF	N_estimator =100, depth=8
CatBoost	Iterations =1000,Learning rate =0.1, depth=6,loss function=MSE, Verbose =100
GRU	Gru units=50,optimizer='adam',loss function =MSE, epoch= 100, batch size = 32

Tkinter. It offers a quick and simple method for developing window-based apps, enabling programmers to design windows, buttons, text fields, labels, and other elements required for user interaction. This interface application allows users to make real time prediction on charging cycles, and the integrated model is deployed with the application to give charging cycles values using other parameters.

3. Results and discussions

3.1. Results of the developed model

Fig. 5 shows the model score comparison plot. The MSE of 0.0587 shows little difference between the actual and anticipated charging cycles. This implies that the model is effective because its predictions agree reasonably well with the observed values. The R-squared (R2) as the model's outstanding capacity to explain the variance in the charging cycle data is demonstrated by its R-squared value of 0.9997. The model captures nearly every variation in the dataset, proving its accuracy. Fig. 6 shows the result scores for various utilized models.

The Random Forest Regression [24,25] model's outstanding performance has encouraging ramifications for sustainable mobility. Precise forecasts of charging cycles provide the necessary foresight for proactive maintenance, energy conservation, and improving the user experience for all parties involved in the ecosystem of electric vehicles.

Because it can prolong battery life, the model's accuracy in predicting charging cycles is especially noteworthy. Proactive maintenance strategies that are based on precise forecasts can help reduce the environmental impact of battery replacement, which is in line with sustainability's main objectives.

Additionally, the robustness of the model in capturing the complex relationships between different features and charging cycles is indicated by its high R-squared value, which also demonstrates the model's explanatory power. This lays the groundwork for additional study and practical uses of the model, which could be used to improve electric vehicle charging plans.

A graph of actual charging cycles versus anticipated charging cycles is displayed in the Fig. 6. The expected number of charging cycles is displayed on the x-axis, while the actual number is displayed on the y-axis. The ideal prediction is represented by the blue line, where the number of charging cycles that actually occur and the number that is anticipated coincide. The red line indicates the real number of charging cycles. Table 3 shows the random forest results in both training and testing sets

The model-depicted results are shown in Table 3. And the best results were obtained with the same model we developed. Since the data points are grouped around the blue line, the predictions are most of the time correct. Some data points, though, fall outside or above the line. This indicates that, compared to expectations, some batteries have cycled more or less. It also depicts two vertical lines in the figure. The maximum anticipated number of charging cycles is displayed on the right line, while the minimum predicted number is displayed on the left. Since every data point falls inside these lines, every battery cycle has occurred within the expected range. The battery life cycle and remaining useful life can be calculated in the future, as done in [14,28,29] with regression algorithms [24,25], and our work on a new dataset has been successful in making predictions on charging cycles of batteries.

The figure indicates that, overall, the estimations of the number of charging cycles are accurate. Some batteries have cycled more or fewer times than anticipated, and there is some variation from the predictions. The details of all other models are tabulated in Table 4.

3.2. Models results comparison and application development

The Table 4. simplifies comparing the performance of each regressor by offering a succinct summary of their measurements. Table 5 shows the details of cross validated scores across 10 folds.

Our results, which were carefully examined using stratified cross-validation [27], highlight the created model's resilience and

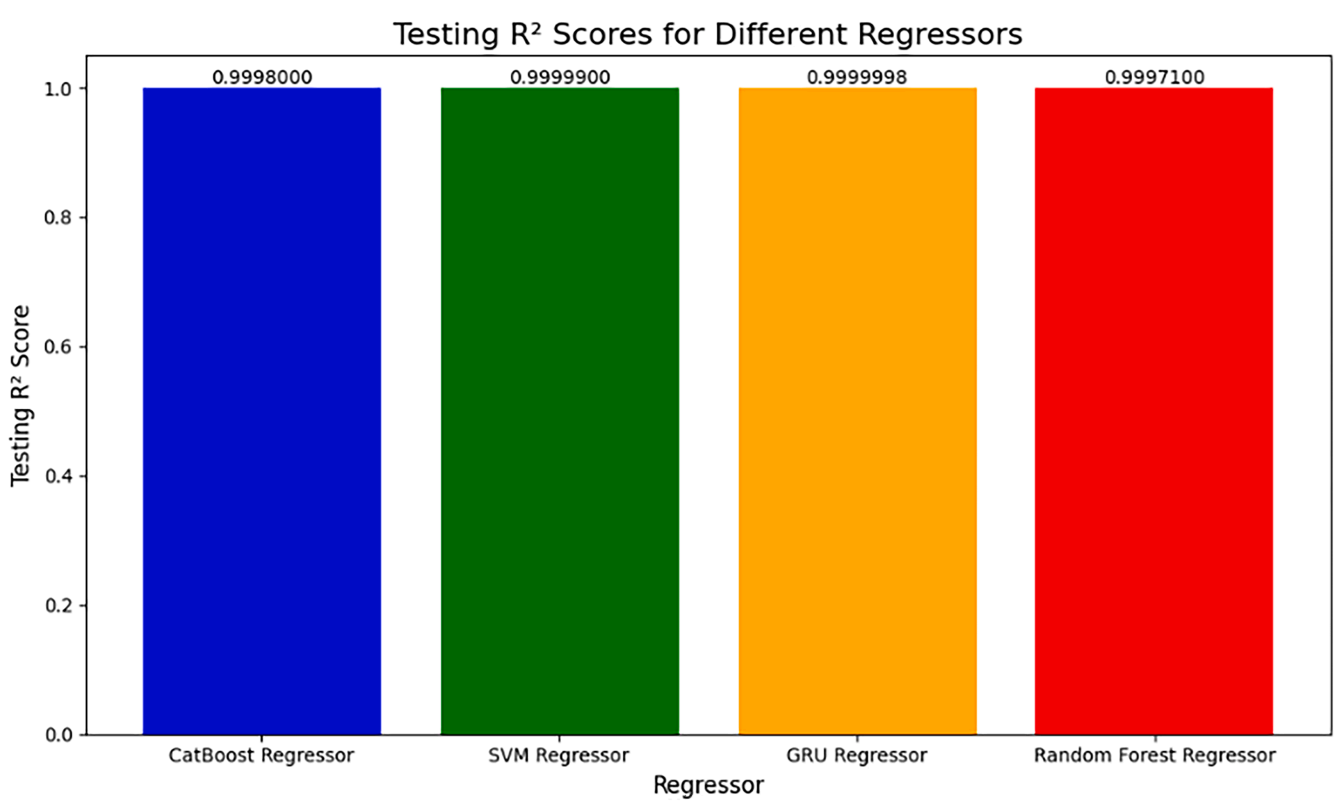

Fig. 5. Accuracy comparison bar diagram for models.

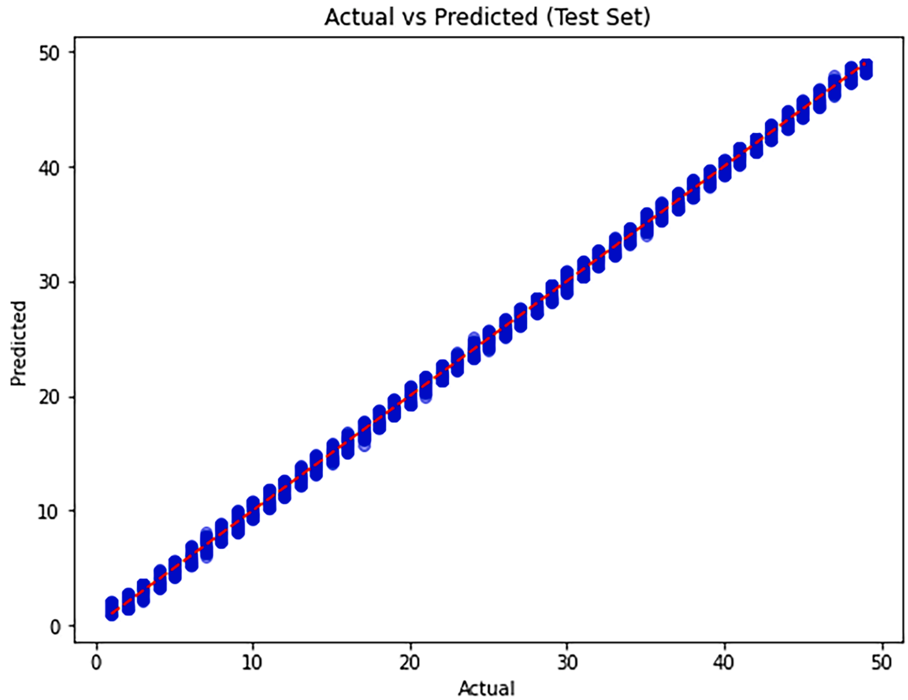

Fig. 6. Actual vs predicted charging cycles plot.

Table 3
Random forests model prediction results.

Metric	Value
Training Metrics	
RMSE	0.0963
MSE	0.00927
R^2 Score	0.9999
Testing Metrics	
RMSE	0.2412
MSE	0.0581
R^2 Score	0.9997

Table 4
Results obtained from other Regressor models.

Regressor	Training MSE	Testing MSE	Training R^2 Score	Testing R^2 Score
CatBoost Regressor	0.0313	0.0377	0.9998	0.9998
SVM Regressor	0.0020	0.0020	0.99999	0.99999
GRU Regressor	0.0000327	0.0000329	0.999999	0.9999998

Table 5
Different fold accuracies after stratified cross-validation for random forest.

Testing folds	R^2 error
Testing Score 1	**0.99971**
Testing Score 2	**0.9997**
Testing Score 3	**0.99972**
Testing Score 4	**0.99972**
Testing Score 5	**0.99971**
Testing Score 6	**0.999720**
Testing Score 7	**0.99973**
Testing Score 8	**0.99971**
Testing Score 9	**0.99973**
Testing Score 10	**0.99972**
Mean of Testing Scores	**0.99972**
Standard Deviation of Testing Scores	**9.5970 e-06**

dependability. The model was remarkably consistent across folds, as evidenced by its remarkable mean R-squared value of 0.99997 and small standard deviation of ±0.0001. The impressive standard deviation of ±9.59706e-06 using this dataset further supports the model's accuracy in forecasting charging cycles. These striking outcomes demonstrate the potential of incorporating state-of-the-art machine learning methods into real-world applications and validating our method's effectiveness.

The ratings above demonstrate the ability of the models to predict charging cycles for electric vehicle (EV) charging cycles, indicating their effectiveness in predicting charging behavior. Concerning both training and testing datasets, the Random Forest Regression model demonstrated remarkable predictive accuracy as demonstrated by its low mean squared error (MSE) and high R2 values. This shows that the model generalizes well to new instances and captures the underlying patterns in the data accurately. Similar to this, though with different performance criteria, the CatBoost, SVM, and GRU regressions all showed good prediction ability. These results highlight how well machine learning models can anticipate EV charging cycles, providing insightful information that can be used to improve energy efficiency, optimize battery management tactics, and guarantee the dependability of electric cars.

During testing, the EV Battery Charging Cycle Predictor App showed strong predictive capabilities. Based on user-provided input parameters, the application consistently predicted the number of charging cycles an electric vehicle battery would undergo. The mean squared error (MSE) and R-squared values acquired during the model training phase were used to assess the prediction accuracy.

The GUI User Interface (UI) is shown in Fig. 7 which is integrated with random forest trained model with '.pkl', extension. The application's capacity to accurately predict charging cycles is in line with the goals of maximizing battery efficiency, minimizing environmental effects, and improving the sustainability of electric vehicles as a whole. Fig. 7 shows the GUI interface for the visualization of charging cycles of EVs. By interacting with the application, users can obtain important knowledge about the anticipated life of their EV batteries, which helps them make well-informed decisions about long-term planning, maintenance, and charging techniques [2].

The findings highlight the usefulness of machine learning models

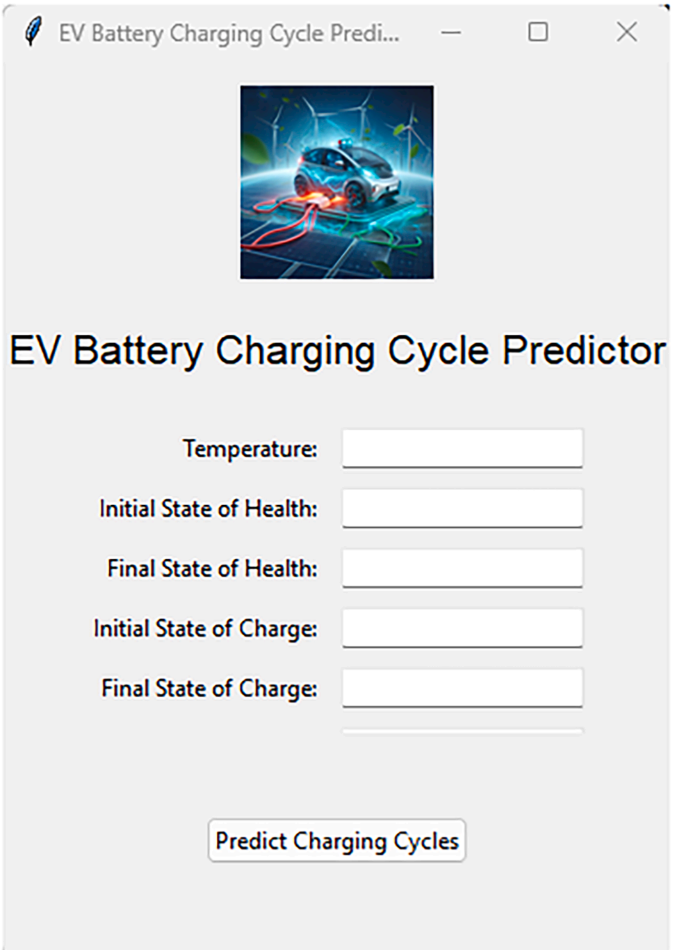

Fig. 7. EV charging cycles predictor application GUI.

incorporated into intuitive applications, which makes predictive analytics more accessible in the context of electric car technology [30,31]. The promising results motivate additional research into analogous applications for various electric vehicle platforms as well as the incorporation of real-world data to improve prediction precision. The random forest regression algorithm's feasibility for such a model's development can be verified with the help of such regression tasks [24].

The EV Battery Charging Cycle Predictor App is an example of how research findings are translated into useful tools that connect cutting-edge machine learning methods with electric vehicle end users [15]. The effective execution of this application creates opportunities for further advancements, such as the improvement of predictive models, the enlargement of input features, and the integration into more comprehensive electric vehicle management systems [2]. The comparison of this research work with previous work is tabulated in Table 6.

In reference to previous studies, our study achieved higher accuracies and error metrics rating when predicting battery charging cycles across multiple popular models. For example, the research by [32], compared to the maximum R^2 of 0.97 and 1.27 MAE of Kernelized SVR for predicting the driving range and state of charge, we reached a R^2 of 0.9999 with much lower MSE of 0.00003. Upon reviewing prior research, we found that our work had greater accuracies and error metrics assessment when predicting battery charging cycles across numerous frequently used models. The work by [32] at increasing prediction of performance and wellbeing, reached a maximum R^2 of 0.97 and 1.27 MAE of Kernelized SVR for predicting driving range and state of charge, while we achieved an encouraging R^2 of 0.9999 with much lower MSE of 0.00003. Besides, in [35], the lowest values of RMSE for

Table 6
Comparison with previous works.

S. N	References	Models employed	Predictor variables	Results
1	[32]	Kernelized SVR	State of charge and driving range	The maximum R^2 and lowest MAE exhibited are 0.97 and 1.27 respectively
2	[33]	Random forest and Extreme learning machine (ELM)	State of charge (%)	RMSE =5.90 (RF) and 6.31 (ELM), MAE = 4.43 (RF),5.11(ELM)
3	[34]	Bidirectional GRU with attention layers	State of charge (%)	Accuracy of 77.15 %
4	[35]	ANN,LSTM	State of health	Best results by Federate learning and ANN. ANN (RMSE =0.733, MAE=0.49284) FL-ANN (RMSE =0.7073, MAE=0.6243)
5	[36]	Linear regression (LR), Extra trees (ET), Boosting method (BR)	State of health	R^2 of 0.99 and MSE of 0.03
6	[37]	Support Vector Machine-Recursive Feature Elimination (SVM-RFE), Diffusion Convolutional Recurrent Neural Network dDCRNN) (dDCRNN+SVM-RFE	State of health	Best model RMSE = 0.014, MAE =0.011, MAPE =0.32
7	Current study	SVM, CatBoost, GRU	Charging cycles	Testing results of MSE = 0.00003 and R^2 = 0.9999

the models developed by using Federated Learning and ANN were equal to 0. 7073 and MAE of 0. 6243 to predict the state of health, although our model demonstrated a considerably higher performance than the former. Models applied by [36] including Linear Regression and Extra Trees yielded an R^2 of 0. 99 and MSE of 0.The forecasting accuracy of MG machines is quite promising and it signifies that the strength of this technique is matching extremely well with the prediction strength of the algorithm which is still lower in comparison to the finding of the present study but the business is been reported a high performer. Finally, the paper [37] that integrated the identification of the most important features using SVM-RFE and the prediction model based on the dDCRNN led to the RMSE of 0. 014 and MAE of 0. 011 for state of health prediction, our proposed model was able to yield a considerably lower error and better accuracy of the predictions. These comparisons emphasise on the benefits of the research approach involving SVM, CatBoost and GRU for the prediction of battery charging cycles.

3.3. Future scopes

Numerous new directions for future study are opened up by the creation of machine learning models for predicting the charging cycles of electric vehicle (EV) batteries. Integrating real-world data is one important area that would improve the generalizability and accuracy of the model. Though the use of an artificial dataset inspired by a publication doesn't ensure reliability, this paves a path for real-time integration and performance checking of battery management and charging systems. The real-time data acquisition is needed for creating a more reliable dataset. Also, predicting only a single parameter isn't the best approach for solving numerous issues related to battery management and charging systems [38]. Through integration of data from many electric vehicle platforms and varied environmental circumstances, the models may be adjusted to better accurately represent battery behaviors. Adding new features, such as charging routines, car usage patterns, and battery chemistry, could be a viable next step in order to gain a better knowledge of how batteries degrade. Moreover, integrating these predictive models into more extensive energy management systems may make it easier to optimize energy allocation and charging infrastructures, which would ultimately increase the sustainability of electric car operations. The model can be expanded with more rigorous validation techniques, and an application on a mobile app version can be developed with IoT (Internet of Things)-based data acquisition for predicting parameters related to vehicle charging.

Furthermore, the approach can be applied not only to electric vehicles but also to other areas where battery life prediction is equally important, including grid energy storage. Energy ecosystems could be completely transformed by combining these models with other cutting-edge machine learning methods, such as deep reinforcement learning or blockchain technology [20] for safe, decentralized battery data management. The creation of more user-friendly programs with integrated predictive models, which assist users in better managing their battery usage and maintenance schedules, represents another possible research field. Last but not least, working with EV manufacturers may spur additional advancements in model development and battery technology, expanding the availability of sustainable transportation options across the globe.

4. Conclusion

Finally, this study successfully illustrates how machine learning models—in particular, Random Forest Regression—can be used to predict electric vehicle battery charging cycles with high accuracy. The study emphasizes the significance of stratified cross-validation, feature selection, and model evaluation in creating reliable models that can capture intricate interactions between variables. The model demonstrates high predictive capabilities with an impressive R-squared value of 0.9997 and a low mean squared error (MSE), which is in line with the objective of maximizing energy efficiency and extending battery life. The use of such applications can allow integration of real-time data monitoring systems for predicting various parameters. The artificial data can be utilized for examining various real-time battery conditions and performance evaluations.

The EV Battery Charging Cycle Predictor App's successful creation highlights the usefulness of this research even more. By empowering EV owners to make knowledgeable decisions on long-term use and battery maintenance, this tool will help the ecosystem supporting electric vehicles become more efficient and sustainable. The study's encouraging findings open up new avenues for the investigation and application of machine learning methods in battery management systems, which will advance efforts to promote electric mobility more broadly. Future research can continue to develop and accelerate the shift toward greener and more sustainable transportation solutions by building on these results.

Declaration of LLM use

The authors would like to declare that they utilized LLM tools for minimizing grammar errors and writing.

CRediT authorship contribution statement

Biplov Paneru: Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Durga Prasad Mainali:** Software, Supervision, Validation, Writing – review & editing. **Bishwash Paneru:** Software, Resources, Methodology, Investigation, Formal analysis. **Sanjog Chhetri Sapkota:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data is would be made available on request to authors.

References

[1] M.S. Bin Ahmad, A. Pesyridis, P. Sphicas, A. Mahmoudzadeh Andwari, A. Gharehghani, B.M. Vaglieco, Electric vehicle modelling for future technology and market penetration analysis, Front. Mech. Eng 8 (2022) 896547, https://doi.org/10.3389/fmech.2022.896547.

[2] Vardakas, J. & Zenginis, I. & Verikoukis, C. (2020). Machine learning methodologies for electric-vehicle energy management strategies. 10.1201/9780429329401-4.

[3] S. Afandizadeh, D. Sharifi, N. Kalantari, et al., Using machine learning methods to predict electric vehicles penetration in the automotive market, Sci. Rep. 13 (2023) 8345, https://doi.org/10.1038/s41598-023-35366-3.

[4] M.A. Eissa, P. Chen, Machine learning-based electric vehicle battery state of charge prediction and driving range estimation for rural applications, IFAC-PapersOnLine 56 (3) (2023) 355–360, https://doi.org/10.1016/j.ifacol.2023.12.050. ISSN 2405-8963,.

[5] I. Babaeiyazdi, A. Rezaei-Zare, S. Shokrzadeh, State of charge prediction of EV Li-ion batteries using EIS: a machine learning approach, Energy 223 (2021), https://doi.org/10.1016/j.energy.2021.120116 article 120116, ISSN 0360-5442.

[6] FCEVs | hydrogen mobility Europe, Hydrogen Mob. Europe | (2015). Nov. 19,, https://h2me.eu/about/fcevs/.

[7] Y. Xu, F.E. Gbologah, D.-Y. Lee, H. Liu, M.O. Rodgers, R.L. Guensler, Assessment of alternative fuel and powertrain transit bus options using real-world operations data: life-cycle fuel and emissions modeling, Appl. Energy 154 (2015) 143–159, https://doi.org/10.1016/j.apenergy.2015.04.112. Sep.

[8] P. Kumari, A.K. Singh, N. Kumar, Electric vehicle battery state-of-charge estimation based on optimized deep learning strategy with varying temperature at different C Rate, J. Eng. Res. 11 (3) (2023) 158–163, https://doi.org/10.1016/j.jer.2023.100113.

[9] M. Şen, M. Özcan, Y.R. Eker, A review on the lithium-ion battery problems used in electric vehicles, Next Sustain. 3 (2024) 100036, https://doi.org/10.1016/j.nxsust.2024.100036.

[10] D. Adhya, A. Pal, A.K. Chakraborty, A. Bhattacharya, Machine learning application for prediction of EV charging demand for the scenario of Agartala, India, in: 2022 4th International Conference on Energy, Power and Environment (ICEPE), Shillong, India, 2022, pp. 1–5, https://doi.org/10.1109/ICEPE55035.2022.9798369.

[11] S. Shahriar, A. Al-Ali, A. Osman, S. Dhou, M. Nijim, Prediction of EV Charging Behavior Using Machine Learning, IEEE Access, 2021, p. 1, https://doi.org/10.1109/ACCESS.2021.3103119. -1.

[12] M. Shibl, L. Ismail, A. Massoud, Machine learning-based management of electric vehicles charging: towards highly-dispersed fast chargers, Energies 13 (20) (2020) 5429, https://doi.org/10.3390/en13205429.

[13] J. Asamer, A. Graser, B. Heilmann, M. Ruthmair, Sensitivity analysis for energy demand estimation of electric vehicles, Transp. Res. Part D 46 (2016) 182–199, https://doi.org/10.1016/j.trd.2016.03.017.

[14] K. Karthick, S. Ravivarman, R. Priyanka, Optimizing electric vehicle battery life: a machine learning approach for sustainable transportation, World Electr. Veh. J. 15 (2024) 60, https://doi.org/10.3390/wevj15020060.

[15] E.S. Ali, M.K. Hasan, R. Hassan, R.A. Saeed, M.B. Hassan, S. Islam, N. Nazmus Shaker, B. Savitri, Machine learning technologies for secure vehicular communication in internet of vehicles: recent advances and applications, *Secur. Commun. Netw.*, 2021 (2021) 23, https://doi.org/10.1155/2021/8868355. Article ID 8868355pages.

[16] X. Feng, M. Ouyang, X. Liu, L. Lu, Y. Xia, X. He, Thermal runaway mechanism of lithium ion battery for electric vehicles: a review, Energy Storage Mater. 10 (2018) 246–267, https://doi.org/10.1016/j.ensm.2017.05.013.

[17] G. Wang, N. Cui, C. Li, Z. Cui, H. Yuan, A state-of-health estimation method based on incremental capacity analysis for Li-ion battery considering charging/discharging rate, J. Energy Storage 73 (2023) 109010, https://doi.org/10.1016/j.est.2023.109010.

[18] A. Goli, A. Ala, M. Hajiaghaei-Keshteli, Efficient multi-objective meta-heuristic algorithms for energy-aware non-permutation flow-shop scheduling problem, Expert. Syst. Appl. 213 (2023) 119077, https://doi.org/10.1016/j.eswa.2022.119077.

[19] A. Goli, E.B. Tirkolaee, Designing a portfolio-based closed-loop supply chain network for dairy products with a financial approach: accelerated Benders decomposition algorithm, Comput. Oper. Res. 155 (2023) 106244, https://doi.org/10.1016/j.cor.2023.106244.

[20] A. Goli, Integration of blockchain-enabled closed-loop supply chain and robust product portfolio design, Comput. Ind. Eng. 179 (2023) 109211, https://doi.org/10.1016/j.cie.2023.109211.

[21] A. Montazerolghaem, M. Khosravi, F. Rezaee, M.R. Khayyambashi, An optimal workflow scheduling method in cloud-fog computing using three-objective Harris-Hawks algorithm, in: 2022 12th International Conference on Computer and

Knowledge Engineering (ICCKE), Mashhad, Iran, Islamic Republic of,, 2022, pp. 300–306, https://doi.org/10.1109/ICCKE57176.2022.9960123.
[22] A. Goli, A. Ala, S. Mirjalili, A robust possibilistic programming framework for designing an organ transplant supply chain under uncertainty, Ann. Oper. Res. 328 (2023) 493–530, https://doi.org/10.1007/s10479-022-04829-7.
[23] P. Kollmeyer, C. Vidal, M. Naguib, M. Skells, LG 18650HG2 Li-ion battery data and example deep neural network xEV SOC estimator script", Mendeley Data (2020) V3, https://doi.org/10.17632/cp3473x7xv.3.
[24] A. Liaw, M. Wiener, Classification and Regression by RandomForest, Forest 23 (2001).
[25] Sarstedt, M. & Mooi, E. (2014). Regression Analysis. 10.1007/978-3-642-53965-7_7.
[26] S.A.A. Balabied, H.F. Eid, Utilizing random forest algorithm for early detection of academic underperformance in open learning environments, PeerJ Comput. Sci. 9 (2023) e1708, https://doi.org/10.7717/peerj-cs.1708. Nov 22PMID: 38077552; PMCID: PMC10703007.
[27] S. Szeghalmy, A. Fazekas, A comparative study of the use of stratified cross-validation and distribution-balanced stratified cross-validation in imbalanced learning, Sensors 23 (2023) 2333, https://doi.org/10.3390/s23042333.
[28] Y. Wang, Y. Zhao, S. Addepalli, Remaining useful life prediction using deep learning approaches: a review, Procedia Manuf. 49 (2020) 81–88.
[29] J.-H. Chou, F.-K. Wang, S.-C. Lo, Predicting future capacity of lithium-ion batteries using transfer learning method, J. Energy Storage 71 (2023) 108120. Zhao, J.; Ling, H.; Liu, J.; Wang, J.; Burke, A.F.; Lian, Y. Machine learning for predicting battery capacity for electric vehicles. eTransportation 2023, 15, 100214.
[30] J. Wu, L. Kong, Z. Cheng, Y. Yang, H. Zuo, RUL prediction for lithium batteries using a novel ensemble learning method, Energy Rep. 8 (12) (2022) 313–326. Suppl.
[31] N. Omase, S. Mittal, S. Palaniraja, P. Guchhait, M. Patil, P. Mundra, A comprehensive review of electric vehicle charging infrastructure and associated challenges, Int. J. Sci. Res. Arch. 10 (2023) 834–840, https://doi.org/10.30574/ijsra.2023.10.1.0832.
[32] M.A. Eissa, P. Chen, Machine learning-based electric vehicle battery state of charge prediction and driving range estimation for rural applications, IFAC-PapersOnLine 56 (3) (2023) 355–360, https://doi.org/10.1016/j.ifacol.2023.12.050.
[33] M.H. Sulaiman, Z. Mustaffa, State of charge estimation for electric vehicles using random Forest, Green Energy Intell. Transp. (2024) 100177, https://doi.org/10.1016/j.geits.2024.100177.
[34] D. Zhao, H. Wang, Z. Liu, Charging-Related state prediction for electric vehicles using the deep learning model, J. Adv. Transp. 2022 (2022) 1–12, https://doi.org/10.1155/2022/4372168.
[35] X. Lv, Y. Cheng, S. Ma, H. Jiang, State of health estimation method based on real data of electric vehicles using federated learning, Int. J. Electrochem. Sci. 19 (8) (2024) 100591, https://doi.org/10.1016/j.ijoes.2024.100591.
[36] K. Akbar, Y. Zou, Q. Awais, M.J.A. Baig, M. Jamil, A Machine Learning-Based Robust State of Health (SOH) prediction model for electric vehicle batteries, Electronics (Basel) 11 (8) (2022) 1216, https://doi.org/10.3390/electronics11081216.
[37] R.M. Alhazmi, State of health prediction in electric vehicle batteries using a deep learning model, World Electr. Veh. J. 15 (9) (2024) 385, https://doi.org/10.3390/wevj15090385.
[38] Y. Chen et al., "Manganese oxide-based multifunctionalized mesoporous silica nanoparticles for pH-responsive MRI, ultrasonography and circumvention of MDR in cancer cells.," vol. 33, no. 29, pp. 7126–37, Oct. 2012, doi: 10.1016/j.biomaterials.2012.06.059.