

**How to Cite:**

Sreedevi, C., Shrimali, M., Leela, R., Murthy, M. Y. B., & Selvaraj, D. (2022). Battery life time prediction of electric vehicle using artificial intelligence. *International Journal of Health Sciences*, 6(S1). <https://doi.org/10.53730/ijhs.v6nS1.8612>

# Battery life time prediction of electric vehicle using artificial intelligence

**Ms. Ch. Sreedevi**

Assistant Professor, Department of CSE, B V Raju Institute of Technology, Narsapur, Telangana, India,  
[sreedevi.chikkudu@bvr.it.ac.in](mailto:sreedevi.chikkudu@bvr.it.ac.in)

**Dr. Manish Shrimali**

Associate Professor, Department of Computer Science and Information Technology, Janardan Rai Nagar Rajasthan Vidyapeeth (Deemed to be University), Udaipur, Rajasthan, India,  
[manishshrimali2009@gmail.com](mailto:manishshrimali2009@gmail.com)

**Ms. Leela R**

Lecture, Department of Electronics and Communication Engineering, Government polytechnic, Nagamangala,  
[leela.ramanna@gmail.com](mailto:leela.ramanna@gmail.com)

**Mr. Mantripragada Yaswanth Bhanu Murthy**

Professor, Department of Electronics and Communication Engineering, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India,  
[mybmurthy@gmail.com](mailto:mybmurthy@gmail.com)

**Dr. D. Selvaraj**

Professor, Department of Electronics and Communication Engineering, Panimalar Engineering College, Chennai, Tamilnadu, India,  
[mails2selvaraj@yahoo.com](mailto:mails2selvaraj@yahoo.com)

**Abstract**--The widespread use of electric vehicles (EVs) is viewed as a turning point for lower emissions of co2 and much more advanced driver assistance systems. It is time-consuming to predict the battery capacity of the electric vehicle using the endurance and dependability of battery systems. Because battery deterioration is typically non-linear, predicting the capacity of the battery of charge estimation with significantly less deterioration is incredibly time-consuming. As a result of this complexity, real-time control of battery storage has proven difficult. Nevertheless, using the latest advancements in battery deterioration comprehension, simulation techniques, and testing, there is a potential to combine this expertise with new machine learning (ML) approaches to find a potential solution for this

complexity. In this study, the battery life is predicted using three different regression models Linear Regression, Ridge Regression, and Ensemble Regression. The model's error analysis is conducted to improve the battery's performance characteristics. Finally, the three ML algorithms are compared using performance parameters. Out of the three regression models, the Ensemble regression model is found to be the best model which has an accuracy of 94.8%.

**Keywords**---Life Prediction, Regression, Battery Management, Error, Accuracy.

## 1. Introduction

Reduced carbon dioxide emissions and the advent of global warming are worldwide priorities. The transportation industry contributes to the majority of carbon dioxide emissions and negative impacts on the environment. The electrification of public transit is being heralded as a strategy for lowering worldwide carbon pollution and reliance on unsustainable power sources. Yearly electric vehicle sales are up steadily from 2011, reaching 2.1 million electric vehicles being sold globally in 2019. This growth in sales is mostly due to the continuous financial and quality improvements in professional electric vehicles, additional electric vehicle alternatives offered to consumers, and growing protection of the environment. The lithium-ion battery (LIB) is a significant technology in the low-carbon era, enabling uses ranging from new energy vehicles to energy storage for regulating renewable power from solar and wind power. However, there are still challenges to wider implementation of the technology, such as inadequate energy density, poor longevity, and relatively high price. Because of the high correlation between LIB longevity and operational circumstances such as heat, charge, and operational state-of-charge range optimizing battery capacity is indeed a unique difficulty. Moreover, the susceptibility of such factors to battery capacity is determined by the cells and package design, which includes essential issues like chemical, electrodes content, compactness, package structure, and temperature control mechanism. In recent decades, the advancement of data-driven methodologies such as ML models has made significant progress in enhancing the precision of state of charge metrics with enhanced generalization capability, effective learning functionality for high precision, and connectivity. The overall efficiency of the batteries is determined based on their health condition and their life span. Artificial intelligence, such as computer vision and the deep learning (DL) approach, which are smart and adaptable, may forecast the health condition and life span of the batteries. Nevertheless, estimate and forecast results are dependent on the availability of the training dataset.

The authors of [1] presented the prediction of the State -Of-charge of LIB through 6 ML techniques for electric - vehicle applications. The ML model's error analysis is performed to optimize the battery's output characteristics. Finally, performance indexes are used to evaluate all 6 methods. Depending on MSE and RMSE values, ANN and GPR are deemed to be the rational approach. The authors of [2] used a large aging database of nickel-manganese-cobalt oxide cells to

develop multiple potential fading algorithms and evaluate output results. The semi-empirical modeling method is evaluated against an ML algorithm as well as an ANN model. The nonlinear autoregressive system (NARXnet) was the most accurate at predicting capacity decline while also reducing computational resources. The author [3] proposes a hybrid ML model that combines relevance vector machines (RVMs) and convolutional neural networks (CNN). RVM is employed to create synthetic cells with extended life spans by augmenting the actual training data. The CNN trained using input-output data from the augmented training sample. A LIB cell's cycle life can be forecasted using CNN, which extracts key local invariant attributes from raw data. The implementation of the developed hybrid ML (RVM+CNN) model should then be tested on a test sample containing 83 cells having cycle lifetimes varying from 150 to 2300. In respect of cycle life predictive performance, the RVM+CNN significantly achieves good results than other ML and DL methods. A model-based approach that includes the ML method into the standard ampere-hour integral process can be used to analyze lithium batteries in NEVs (New Energy Vehicles). The author [4] develops a framework for determining SOH in live time by combining ML algorithms and the ampere-hour technique, enabling car manufacturers to precisely understand the current state of the battery in NEVs. Utilizing a long short-term memory (LSTM) network, a deep neural network could be utilized to foresee the future degradation curve of SOH, enabling the lithium battery's life cycle better foreseeable. The recommended method has an error of 0.009 for SOH forecasting and 0.021 for long-term decay curve forecasting.

Article [5] provides an in-depth examination of various ML algorithms for evaluating the SOC and RUL of battery management systems. The emphasis is on the advantages and disadvantages of each strategy. The author starts with the battery management systems' strengths and weaknesses have been identified and briefly discussed in terms of understanding how they affect performance. The use of ML methods in BMS was extensively examined in this study to generate an appropriate battery model. SOH and RUL for Li-Ion 18650 cells could be estimated using the approach outlined in the publication [6], which considers various variables, such as the cell's current SOC, internal resistance, discharge voltage, and capacity. Various battery mathematical techniques are developed and deployed on a stand-alone hardware device to determine an optimal SOH and RUL ML-based prediction model. In a real-time application, DNN methodology accurately estimates SOH with a 5% error rate, whereas LSTM efficiently handles RUL with a 10 cycle. The efficacy of this strategy to conserve battery life is demonstrated using a realistic hardware design and a range of ML models. An ML-based SOH estimate is developed using new health markers derived from localized constant-voltage (CV) charging [7]. To begin, a thorough review of many CV-based health indicators is performed to find the most instructive elements with a significant relationship to the SOH. Second, a DL model is used to create a nonlinear mapping between the specified health metric to the SOH. The recommended SOH estimate was tested over time. The proposed method seems to have high predicting accuracy, minimal charging integrity demands, and a high level of cell inconsistency resistance.

The article [8] proposes an ML technique for estimating LIB's RUL using a CNN architecture with an extra LSTM layer. As a training sample for the CNN-

LSTM model, the author uses data from 124 commercialized LIB cycled under fast-changing conditions. In this study, just 100 cycles are used to establish recommendations for the remaining cycles. According to a model developed by experts, the mean absolute error on prediction of the current cycle was 0.1099, whereas the remaining cycles had an error of 0.0741. The article [9] describes how to forecast RUL in batteries using data-driven methodologies. The discrepancy between a battery's capacity and voltage charts can be used to calculate important parameters. These curves have been identified as indicators of battery deterioration. Using the information obtained from such a chart, the author provides a practical method for measuring the RUL of batteries. The differential curves provide 19 features, which are subsequently assessed and retrieved. SBL is a popular ML technique in the RUL estimation industry, and it is employed to identify battery RUL in advance. Finally, the method's effectiveness is assessed by examining the training and testing flaws. The researcher's [10] aim is to find what the RUL of LIB will be. They present a novel statistical technique for detecting critical moments in the discharge cycle by evaluating the LIB dataset. The LIB analysis assisted in identifying a relationship between the number of cycles and the pattern. This research suggests two feature sets based on the most essential criteria. On the feature set under examination, the LIB dataset is used to evaluate a range of ML regression models. A comparison analysis is performed to guarantee that the proposed merger criteria operate properly.

The paper is organized in the following way: The data required to predict the battery life span using artificial intelligence for this study and the features extracted from the data are discussed in Section 2. The classification algorithms are addressed in section 3 which include Linear regression, Ridge regression, and Ensemble regression. Section 4 discusses the results obtained from this study and the comparison of Machine Learning models based on error parameters. Finally, in section 5 the conclusion of this study is given.

## **2. Methodology**

In this study, the required dataset to predict the battery life of electric vehicles is collected from the MIT battery database. The collected database is a raw unprocessed dataset that is not compatible with the Machine Learning algorithms. Thus the raw dataset is pre-processed, making it compatible to be evaluated by the machine learning algorithms. As the dataset collected contains several features incorporated in them, it makes the prediction process complex, Thus only the required features are extracted from the dataset.

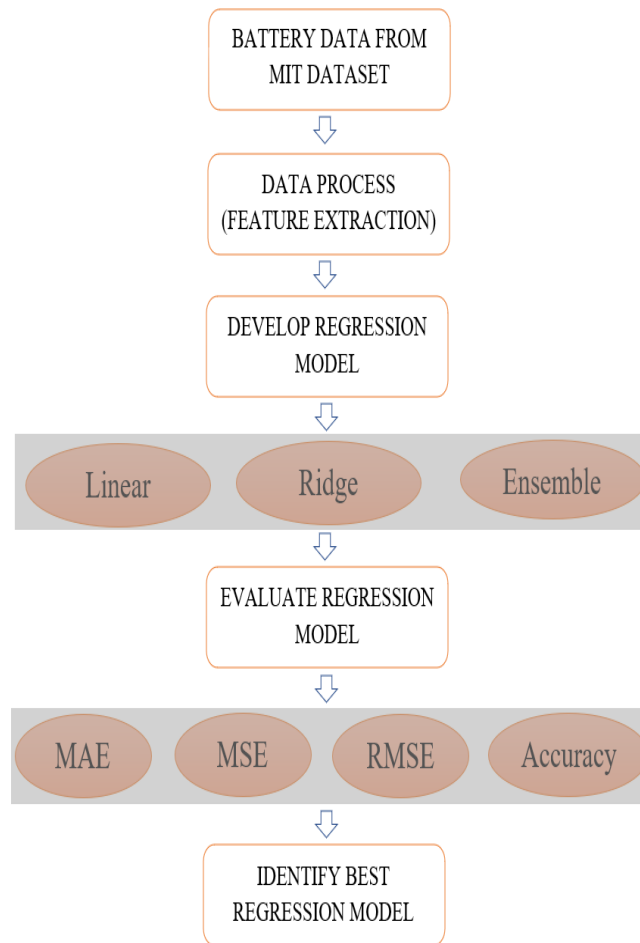


Fig.1 Block diagram

In this study, only 8 features that are significant and relevant for the prediction process are extracted from the dataset. The dataset is then divided into two parts, namely, the training set and the testing set. It is divided by an 8:2 ratio. Once the features are extracted and the data is divided, the classification regression models are developed. In this study, three regression models are used: Linear Regression, Ridge Regression, and Ensemble regression. Initially, the models are trained using the training set and then tested using the testing set. To find the most efficient model among the three models, the measuring parameters such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Square Error (MSE) values are calculated. The calculated values are evaluated and then compared to find the best model among the three models. The accuracy of each model is also calculated. Figure 1 given shows the block diagram of the prediction of battery life of electric vehicle time using artificial intelligence.

Figure 1 shows the steps involved in the prediction of the battery's life of electric vehicle time using artificial intelligence. From figure 1, the very first step involved in this study is the collection of data. In the next step, the data is processed, where the required and relevant features are extracted. The next step involves the

development of regression models. Once the models are developed and trained using the training dataset, the models are tested using the testing dataset. The regression models are then evaluated using various parameters such as MAE, MSE, RMSE, and Accuracy to find the best model. Finally, the best and most efficient model is identified using these parameter values.

### 3. Data Acquisition and Feature Extraction

The dataset required for this study is collected from the MIT Battery Database which is considered to be a trusted repository of battery datasets. The dataset gathered is deemed raw and unprocessed. The unprocessed raw database is incompatible with machine learning methods. As a result, the initially obtained raw dataset is pre-processed using a variety of approaches to make it suitable for machine learning algorithms. The pre-processed data is then split into two parts: a training set and a testing set. The data in the training set is used to train the machine learning models, whereas the data in the testing set is used to evaluate the machine learning technique once it has been trained using the training dataset. As the collected dataset from the internet consists of several features, that makes it difficult to analyze and classify the output. It also results in increased computational time. To remove all this complexity, only the essential features are extracted from the dataset. For this study, the dataset collected contains several features but only 8 specifically required features were extracted for the classification purpose. Figure 2 given below tabulates the 8 extracted features from the dataset.

DeltaQ_var	DeltaQ_min	CapFadeCycle2Slope	CapFadeCycle2Intercept	Qd2	AvgChargeTime	MinIR	IRDiff2And100
-5.0839	-1.9638	6.4708e-06	1.0809	1.0753	13.409	0.016764	-3.3898e-05
-4.3754	-1.6928	1.6313e-05	1.0841	1.0797	12.025	0.016098	4.4186e-05
-4.1464	-1.5889	8.1708e-06	1.08	1.0761	10.968	0.015923	-0.00012443
-3.8068	-1.4216	-8.491e-06	1.0974	1.0939	10.025	0.016083	-3.7309e-05
-4.1181	-1.6089	2.2859e-05	1.0589	1.0538	11.669	0.015963	-0.00030445

Fig.2 Features extracted from the dataset.

### 4. Classification Methods

In this study, Machine-learning (ML) methods are particularly applied to compute the battery's life span. Here three different types of regression machine learning models are used to predict the life of the battery. The regression models used in the study are Linear Regression, Ridge Regression, and Ensemble Regression. These ML algorithms are briefed below.

#### A. Linear Regression

Linear regression is a popular technique for prognostic analysis and prediction. It is also considered to be a reasonable method for predicting a numerical outcome[11]. It is a supervised machine learning model that uses regression to

estimate the targeted estimation value based on subjective variables. It is mostly used for determining the connection between variables and prediction. Linear regression predicts parameters that are sustained/real or numerical. The linear regression algorithm demonstrates a linear correlation between a dependent (y) factor and one or more self-reliant (y) factors, thus the name. Because linear regression demonstrates a linear correlation, it determines how well the values of the dependent variable change concerning the effect of the independent variable. In this study, the training set is given as the input to the Linear regression model to train the model first. Figure 3 given below show the scatterplot of the linear regression model when trained with the training set. The plot shows the relation between the actual battery cycle life and the predicted battery life cycle. And from the plot, it is clear that the plot is linear, and hence linear regression model is considered one of the potential models to predict the battery life span.

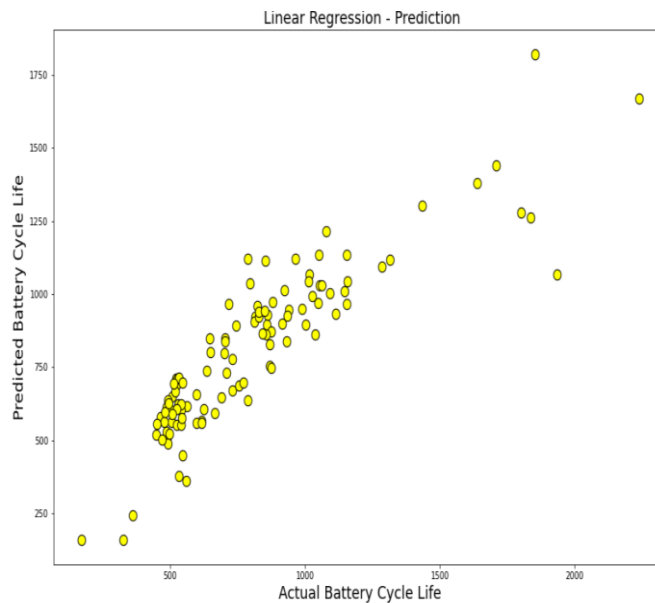


Fig. 3 Scatterplot of Linear Regression model.

### B. Ridge Regression

Ridge regression is a network optimization technique for analyzing datasets with multi-collinearity. When there is a problem with multicollinearity, the smallest squares are impartial, and deviations are enormous, resulting in projected values that are far from the true numbers. Ridge regression is a compensated smallest-square technique that tackles the least-square theory's generalization error[12]. Ridge regression performs well whenever there is a subgroup of genuine coefficients that are minimal or even null. In this study, the training set is fed into the Ridge regression model to train the model first. Figure 4 depicts the scatterplot of the Ridge regression model after training with the training dataset. The plot depicts the relationship between the actual battery cycle life and the expected battery life cycle. And it is evident from the plot that the plot is linear, hence the Ridge regression model is considered one of the probable models to estimate the battery life span.

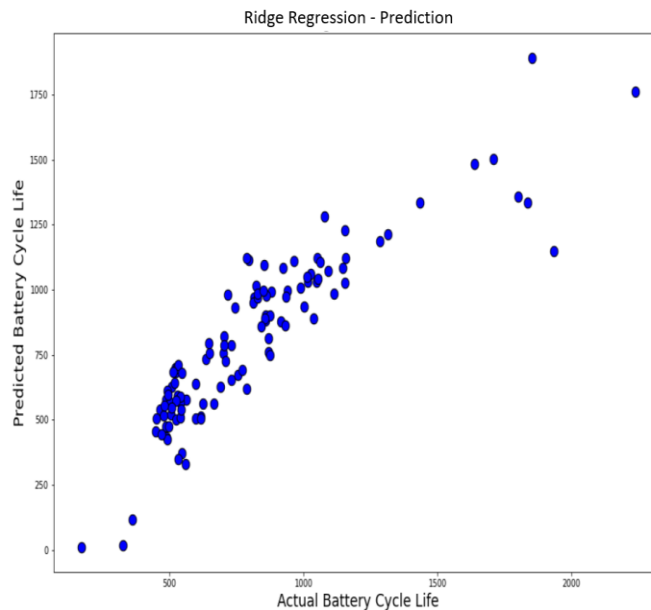


Fig. 4 Scatterplot of Ridge Regression model.

### C. Ensemble Regression

An ensemble is a collection of elements that are perceived as a whole instead of separate. An Ensemble technique generates numerous models and then integrates those to address the problem. Ensemble approaches to aid in improving the model's structural rigidity. Ensemble regression aims to increase predictive performance in learning situations with a quantitative target attribute by combining multiple models. By leveraging individual ML techniques, ensemble ML frameworks are predicted to increase prediction performance [13]. In this study, the Linear regression model and the Ridge regression model is combined to introduce a hybrid Ensemble regression model. Figure 5 given below depicts the scatterplot of the ensemble regression model. Here, the training set is given as the input to the Ensemble regression model to train the model first. The plot shows the relation between the actual battery cycle life and the predicted battery life cycle. And from the plot, it is clear that the plot is linear and hence ensemble regression model is considered one of the potential models to predict the battery life span.

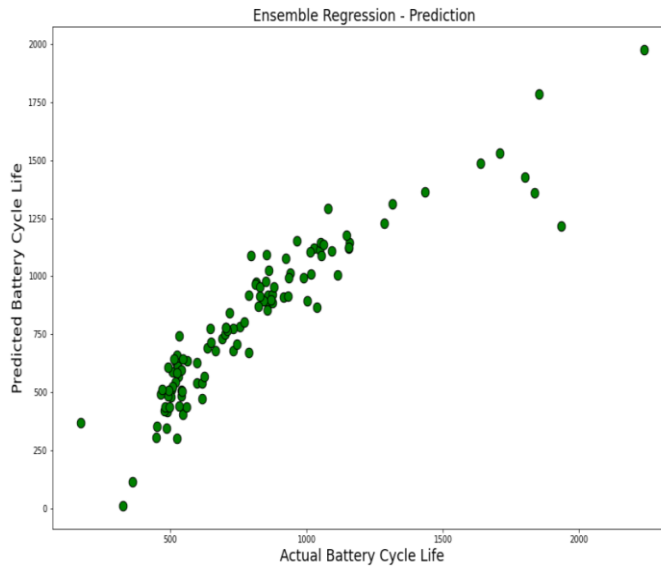


Fig.5 Scatterplot of Ensemble Regression model.

## 5. Result and Discussion

The dataset required for this study is collected from the MIT battery database, which is considered to be a trusted website for the database. The raw dataset is subsequently pre-processed to ensure compatibility with machine learning methods. The pre-processed dataset is then separated into 8:2 training and testing sets. As a result, 80 percent of the pre-processed data is designated the training dataset, while the remaining 20 percent is considered the testing dataset. After splitting the dataset the essential features are extracted from the dataset. Since the dataset contains several features, to reduce the complexity of the classification process only the relevant features are extracted from the dataset. In this study, the crucial 8 features are extracted from the dataset. Once the features are extracted the dataset is subjected to the classification process. In this study, machine learning models such as Linear regression, Ridge regression, and Ensemble regression are used to predict the battery life of electric vehicles. To select the efficient model and to improve the accuracy, the model's error parameters such as Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE) values are evaluated.

MAE determines the average degree of errors in a group of projections without taking into account their orientation. It is the arithmetic average of differences among predictions and actual observation over the test sample, with all unique deviations given equal value. The equation to calculate the MAE value is given below.

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x| \quad [1]$$

The RMSE is the square root of the average of the squares error among actual and projected values for the target attribute. The equation to calculate the RMSE value is given below.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2} \quad [2]$$

The degree of error in predictive methods is measured by MSE. The MSE values are calculated by the average of the square difference between both the predicted and actual values.

$$MSE = \frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2 \quad [3]$$

Figure 6 given below shows the comparison of machine learning models based on error parameters.

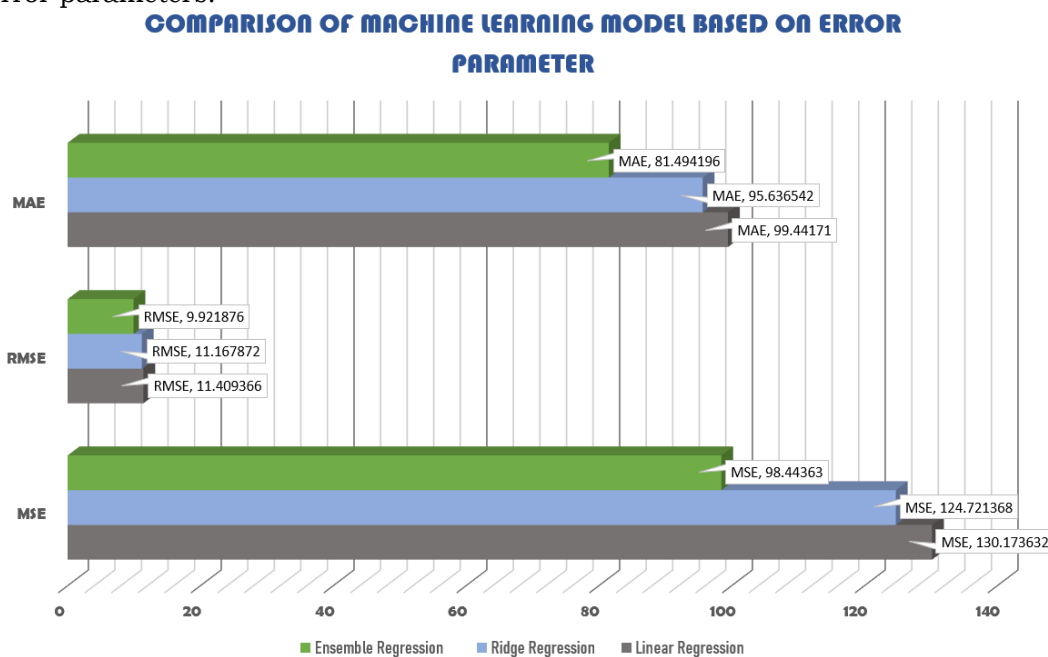


Fig.6 Comparison of Machine Learning models based on error parameters.

From figure 6 it is given that the MAE, RMSE, and MSE value for the Ensemble regression model is 81.49%, 9.9%, and 98.4%, which is considerably lesser than the other two regression models making it the most efficient model among the three models. The MAE value for Ridge and Linear regression model is about 95.63% and 99.44%, the RMSE value for Ridge and Linear regression model is about 11.16% and 11.40% and the MSE value for Ridge and Linear regression model is about 124.72% and 130.17% respectively, making Ridge regression model second-best model and the Linear regression model the least accurate model. Table 1 given below gives the accuracy percentage of the three regression models used in this study.

Table.1 Accuracy percentage of the Machine Learning models

Model	Accuracy Rate
Linear Regression	0.9107284795803038
Ridge Regression	0.928050063821243
Ensemble Regression	0.9489445485983534

From table 1 it is clear that the Ensemble Regression model has the highest output accuracy with 94.89% making it the most efficient and ideal model. The Ridge Regression model has an accuracy of about 92.80% and the Linear Regression model has the least accuracy percentage among the three models with 91.07%. Figure 7 given below shows the graphical representation of the accuracy comparison of regression models

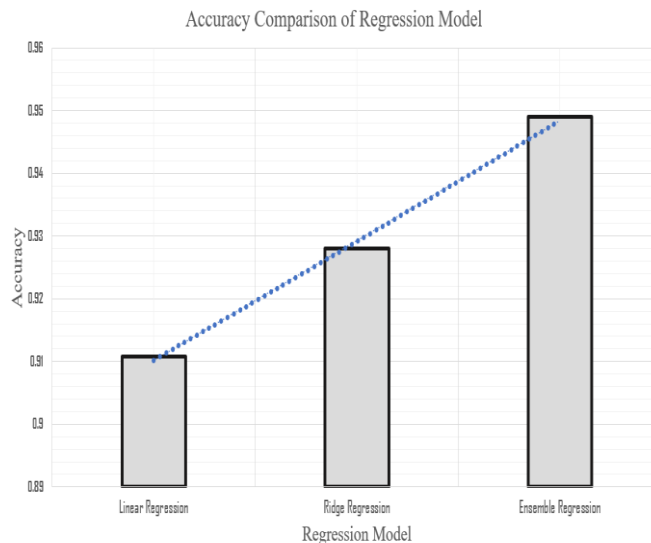


Fig.7 Graphical representation of accuracy comparison of regression models.

From figure 7, it is shown that the Linear regression model has the least accuracy value when compared to the other two regression models. The Ridge regression model, on the other hand, has greater accuracy than the linear regression model but the model's accuracy level is not up to the mark. But the Ensemble regression model which is a hybrid model, a combination of both linear regression model and ridge regression model yields the highest accuracy value among the three regression models. Hence, the Ensemble regression model is considered to be the best and most efficient regression model to predict the battery life in electric vehicles among the other two regression models.

## 6. Conclusion

Batteries will undoubtedly play an important role in a reduced carbon environment, with significant uses including electric vehicles. Maximizing the lifespan and effectiveness of such technologies remains a serious challenge. Nevertheless, with recent breakthroughs in comprehending battery

performance/lifetime, a wide range of diagnostic procedures, and the emergence of machine learning methodologies, there is room for further optimal control of battery systems. Estimation of battery life span is critical in the functioning of the electrical vehicle's battery system. The life of a battery is estimated in this paper using three machine learning techniques: linear regression, ridge regression, and ensemble regression. Because of their enhanced processing of non-linear data, machine learning techniques are chosen for evaluating the battery's life span. With less MAE, RMSE, and MSE error values and high accuracy of 94.89% the Ensemble regression model shows efficient performance while outperforming other models. The optimized attributes fed into the machine learning algorithms forecast the battery's life duration, assisting investors and researchers in determining the optimal batteries for particular purposes. Based on life span projections, ensemble regression will aid in the creation of the best battery system for electric vehicles.

## References

- [1] Chandran, V., K Patil, C., Karthick, A., Ganeshaperumal, D., Rahim, R., and Ghosh, A., 2021. "State of charge estimation of lithium-ion battery for electric vehicles using machine learning algorithms". *World Electric Vehicle Journal*, 12(1), p.38.
- [2] Hosen, M.S., Jaguemont, J., Van Mierlo, J. and Berecibar, M., 2021. "Battery lifetime prediction and performance assessment of different modeling approaches". *Iscience*, 24(2), p.102060.
- [3] S. Shen, V. Nemani, J. Liu, C. Hu and Z. Wang, "A Hybrid Machine Learning Model for Battery Cycle Life Prediction with Early Cycle Data," *2020 IEEE Transportation Electrification Conference & Expo (ITEC)*, 2020, pp. 181-184.
- [4] D. Chang et al., "State of Health Estimation and Remaining Useful Life Prediction of The Lithium Battery for New Energy Vehicles with Long Short-Term Memory Neural Network," *2020 5th International Conference on Universal Village (UV)*, 2020, pp. 1-6.
- [5] R. R. Ardeshiri, B. Balagopal, A. Alsabbagh, C. Ma and M. -Y. Chow, "Machine Learning Approaches in Battery Management Systems: State of the Art: Remaining useful life and fault detection," *2020 2nd IEEE International Conference on Industrial Electronics for Sustainable Energy Systems (IESES)*, 2020, pp. 61-66.
- [6] P. Venugopal et al., "Analysis of Optimal Machine Learning Approach for Battery Life Estimation of Li-Ion Cell," *In IEEE Access*, vol. 9, pp. 159616-159626, 2021.
- [7] H. Ruan, Z. Wei and H. He, "State of Health Estimation of Li-ion Battery Based on Regional Constant Voltage Charging," *2021 IEEE 12th Energy Conversion Congress & Exposition - Asia (ECCE-Asia)*, 2021, pp. 950-952.
- [8] A. Rastegarpanah, Y. Wang and R. Stolkin, "Predicting the Remaining Life of Lithium-ion Batteries Using a CNN-LSTM Model," *2022 8th International Conference on Mechatronics and Robotics Engineering (ICMRE)*, 2022, pp. 73-78.
- [9] S. S. Afshari, S. Cui, X. Xu, and X. Liang, "Remaining Useful Life Early Prediction of Batteries Based on the Differential Voltage and Differential Capacity Curves," *in IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-9, 2022, Art no. 6500709.

- [10] Y. S. Jain, A. D. Veer, G. S. Sawant, Y. R. Jain, and S. S. Udmale, "Novel Statistical Analysis Approach for Remaining Useful Life Prediction of Lithium-Ion Battery," 2021 *12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021*, pp. 1-6.
- [11] James, G., Witten, D., Hastie, T. and Tibshirani, R., 2021. "Linear regression". In *An introduction to statistical learning* (pp. 59-128). Springer, New York, NY.
- [12] Wang, X., Wang, X., Ma, B., Li, Q. and Shi, Y.Q., 2021. "High precision error prediction algorithm based on ridge regression predictor for reversible data hiding". *IEEE Signal Processing Letters*, 28, pp.1125-1129.
- [13] Ngo, N.T., Pham, A.D., Truong, T.T.H., Truong, N.S., Huynh, N.T. and Pham, T.M., 2022. "An ensemble machine learning model for enhancing the prediction accuracy of energy consumption in buildings". *Arabian Journal for Science and Engineering*, 47(4), pp.4105-4117.
- [14] Widana, I.K., Sumetri, N.W., Sutapa, I.K., Suryasa, W. (2021). Anthropometric measures for better cardiovascular and musculoskeletal health. *Computer Applications in Engineering Education*, 29(3), 550-561. <https://doi.org/10.1002/cae.22202>