Fuzzy Systems and Soft Computing
ISSN : 1819-4362

BATTERY AGING ANALYSIS: MAXIMUM CAPACITY PREDICTION USING RANDOM FOREST REGRESSION

Devendra Potnuru, Department of Electrical and Electronics Engineering, GVPCOEW, Visakhapatnam-530048, India. devendra.p.07@gmail.com;
K. Venkata Rao, Department of CSSE, Andhra University, Visakhapatnam-530003, India.

Abstract:
Accurate estimation of battery health is crucial for the reliable operation energy storage systems. This paper explores the application of Random Forest Regression (RFR) to estimate battery health, focusing on metrics such as Maximum Capacity. The study leverages a comprehensive simulation dataset comprising various operational parameters and degradation profiles from lithium-ion battery. By training the Random Forest model on this data, we aim to capture the complex, non-linear relationships inherent in battery aging processes. The proposed RFR approach is benchmarked against traditional methods, demonstrating good prediction accuracy and robustness. This research underscores the potential of machine learning techniques, particularly Random Forest Regression, in enhancing the predictive capabilities and operational efficiency of battery management systems.
Keywords: Battery, Maximum Capacity, Aging Analysis, Random Forest Regression

1. Introduction

The demand in consumer electronic systems, electric vehicles, and energy storage applications with renewable energy has significantly increased the demand for reliable and efficient battery technologies. Lithium-ion batteries, in particular, gained popular in Electric Mobility systems due to their high energy density and long charge-discharge cycle life. However, the batteries useful life is subject to reduce over time; there issues of low reliability and efficiency will occur. Accurate estimation of battery health, often quantified as the State of Health (SoH), is essential for ensuring optimal performance, safety, and longevity of battery systems.
Traditional methods for battery health estimation often rely on empirical models, electrochemical impedance spectroscopy, or equivalent circuit models. While these approaches provide valuable insights, they are often limited by their inability to account for the complex, non-linear degradation processes that occur in real-world operating conditions. Consequently, there is a risingimportance in leveraging advanced data-driven techniques to enhance the accuracy and robustness of battery health predictions.
Machine learning, with its capability to model complex patterns and relationships within large datasets, offers a promising alternative for battery health estimation. Among various machine learning techniques, Random Forest Regression (RFR) popularly used due to its advantages like robustness, multi-dimensional data handling, and avoidingover fitting. There are many applications in the literature using Random Forest algorithm across different sectors [1]-[10]. Assessment of the aging parameters for Lithium iron phosphate based battery is developed in[1], similarly acoustic classification and event detection is developed in [2], state of charge estimation (SOC) is proposed in [3], geometry and proximity estimation in [4] , energy consumption of an electric bus using the real-time data is presented in [5]. Agriculture yield prediction is proposed in [6], identification and retrieval of soil salinity from the data is developed using deep learning in [7] , space vector PWM technique is implemented for induction motor using random forest Regressor in [8], age estimation is proposed in [9],and short term load forecasting using deep random forest Regressor in [10].
This paper investigates the application of RFR for battery health prediction. We employ a comprehensive dataset form the simulation model of Lithium-ion battery aging with different charging-discharging cycles and encompassing various operational parameters and degradation profiles from lithium-ion batteries to train the RFR model. The primary objectives of this study are to assess the accuracy of RFR in predicting battery Maximum Capacity in Ampere-hour (Ah) rating and compare its performance with various operational conditions.
In the following sections, description for Lithium-Ion Battery Aging and Maximum Capacity data generation form the simulation environment, description data statistics in section-1 of the paper , implementation of random forest Regressor is described in section-3, results and discussions are

elaborated in section-4 of this paper and finally conclusion is given in section-5. The present work aims to contribute to the growing body of knowledge on advanced battery health estimation techniques and highlight the potential of Random Forest Regression in enhancing the predictive capabilities and operational efficiency of energy storage systems.

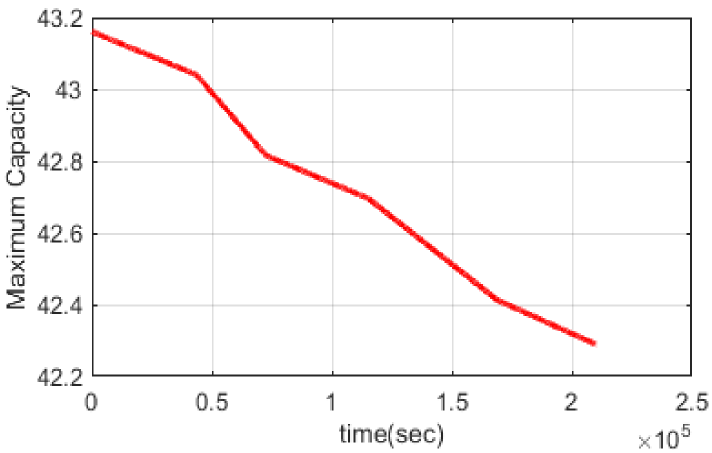
2. Lithium-Ion Battery Aging and Maximum CapacityData

In this study, data is derived from 40 Ah battery. The battery model is tested for 1000 hours of multiple discharge-charge cycles at room temperature of 25^{o}C, for various discharge rates. MATLAB/Simulink simulation data is used to observe the effects of Depth of Discharge (DOD) and discharge rate on battery life. Initially, the battery cycles to a 20% DOD at a 0.5 C discharge rate. This involves 20 Adischarge current at starting with 0.5 C-rate from an initial state of charge (SOC=100%) until the SOC reaches 80% (that means DOD of 100-80=20%), later the batter recharges to 100% SOC with the same current. Later this cycle are repeated for more number of cycle means the increased battery aging, during this process the temperature is varied from 25°C to 33°C.

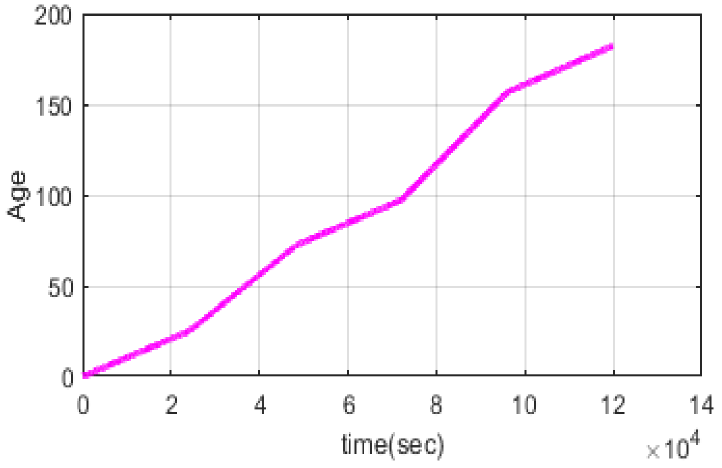
At 200 hours, the DOD is increased to 80%, and the battery cycles between 20% and 100% SOC for another 200 hours, leading to accelerated aging. At 400 hours, DOD is reduced again to 20% in cyclic manner during 200 hours, slowing the aging process. Similarly at 600 hours, 2C-rate is considered discharge current is 80 A. causing the internal cell temperature to rise to 43°C. This rapid cycling accelerates battery aging and reduces capacity. At 800 hours, the battery dischargeswith 0.5 C-rate for the remaining 200 hours. This process of charging and discharging leads to aging of the battery and decrease in its Maximum capacity. Overall description of the data is shown in Table.1.

Table.1 Statistical Description of Data related to Lithium-Ion Battery Aging and Maximum Capacity

	Voltage	**current**	**SOC**	**Temp**	**Age**	**Maximum Capacity**
count	209930	209930	209930.	209930	209930	209930
mean	13.500621	5.670682	85.644884	33.118008	92.069137	42.720527
std	0.422658	31.159280	15.994282	3.811884	56.562089	0.269990
min	12.373688	-20.000000	19.166021	25.000000	0.000000	42.289759
25%	13.159353	-20.000000	79.818505	30.800539	39.228940	42.473138
50%	13.524340	-20.000000	88.559745	31.122280	91.228943	42.724534
75%	13.878877	20.000000	97.887968	38.379905	143.895665	42.972747
max	14.379213	80.000000	100.000000	41.015490	182.312999	43.160000

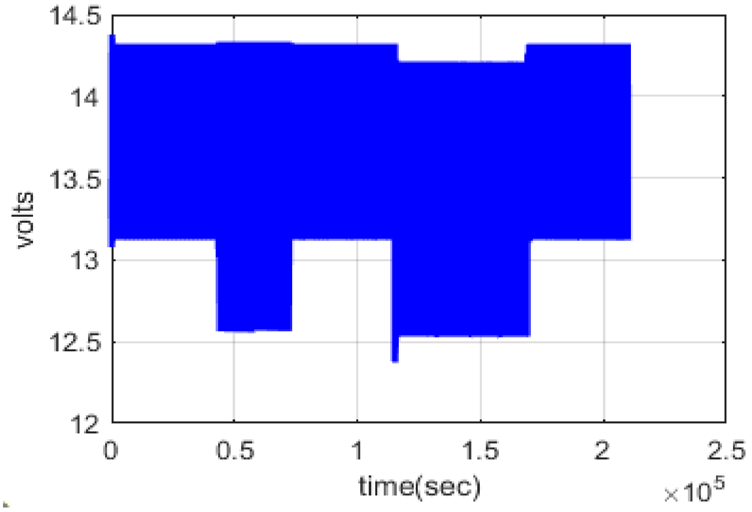


(a)

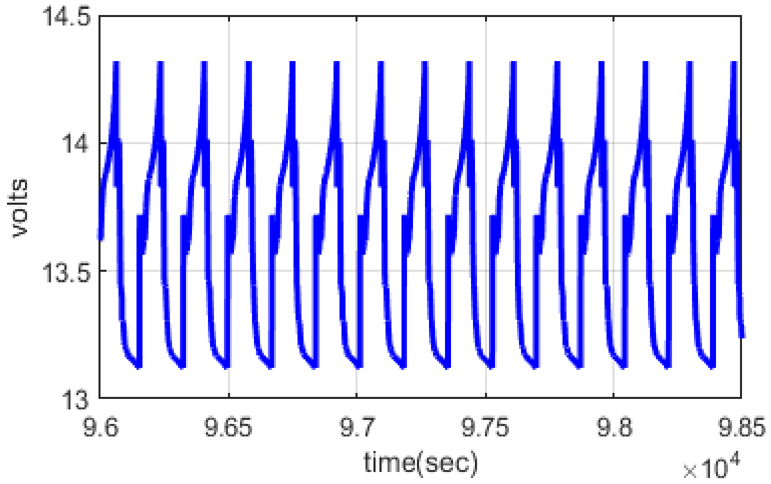


(b)

Fig.1(a) Maximum Capacity of Battery with respect to Aging (b) Aging of Battery involving many charging-discharging cycles

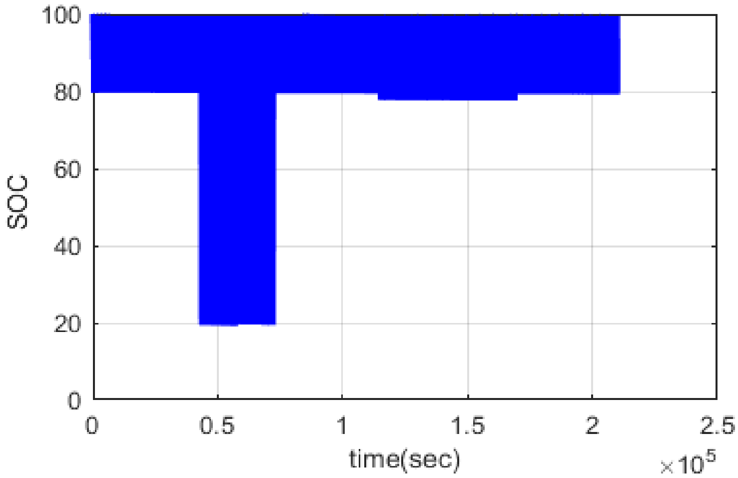


(a)

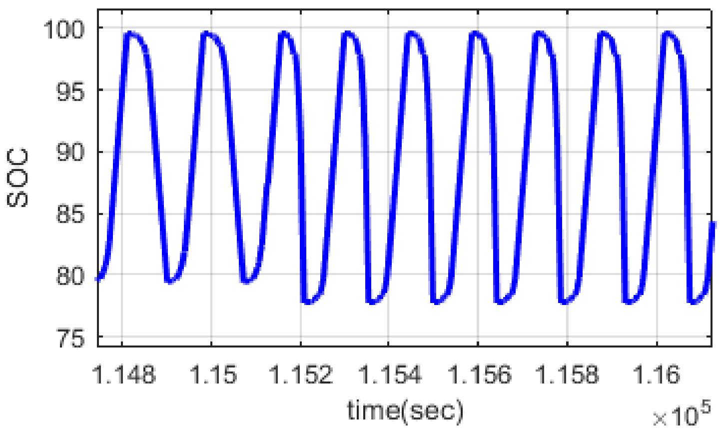


(b)

Fig.2(a) Voltage of Battery during charging –discharging cycles(b) Zoomed view of voltage during charging –discharging cycle

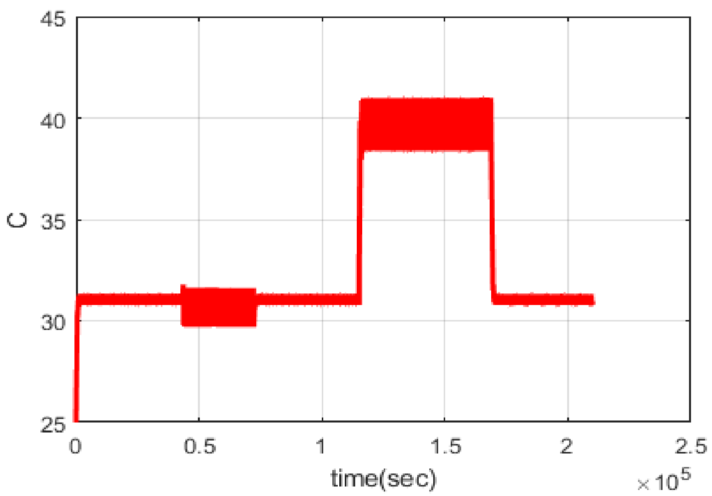


(a)

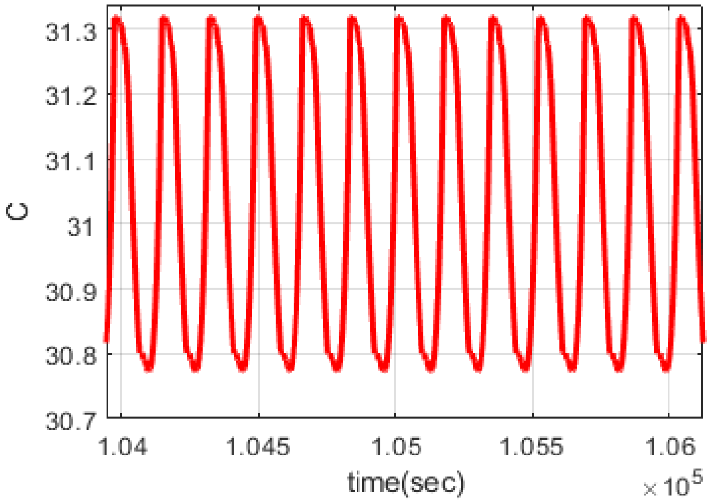


(b)

Fig.3(a) Battery SOC (State-of –Charge) during charging –discharging cycles(b) Zoomed view of SOC during charging –discharging cycle

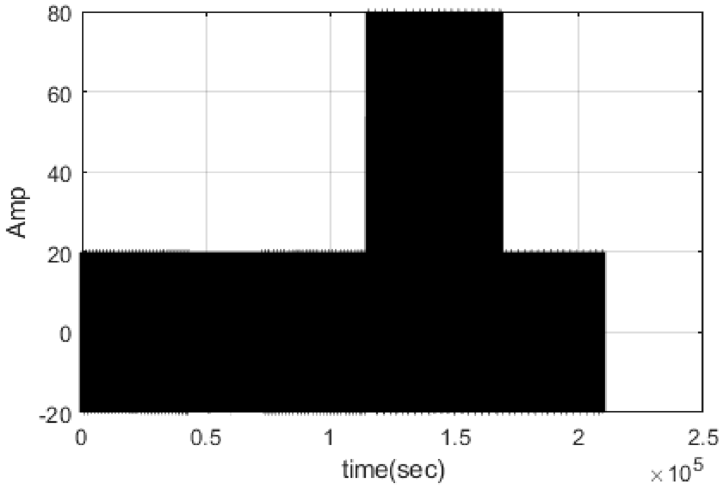


(a)



(b)

Fig.4(a) Various Temperature conditions considered(b) Zoomed view



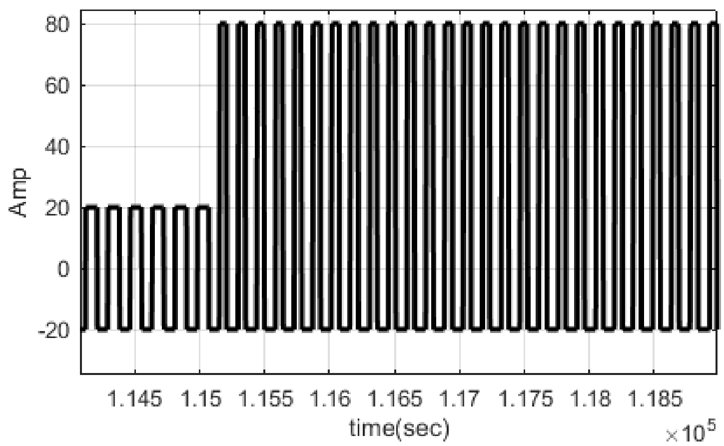


Fig.5(a) Battery Current during charging –discharging cycles(b) Zoomed view of Current

3. Implementation of Random Forest Regression

Random Forest is popular algorithm prediction problems. It constructs numerous decision trees and combines their outputs to achieve more accurate and reliable predictions. This method was created by Leo Breiman and Adele Cutler.

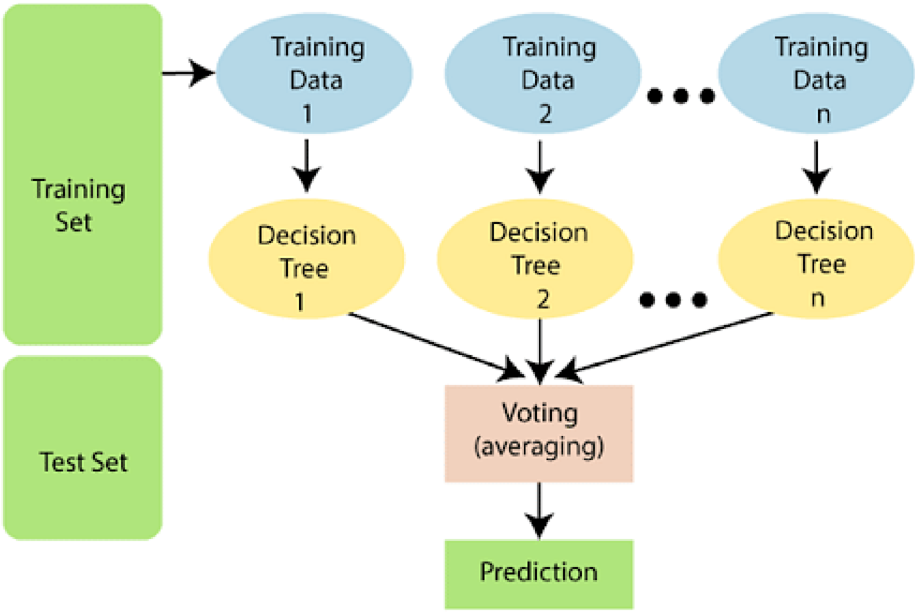


Fig.6 Random forest algorithm schematic diagram

Random Forest Regression is an excellent algorithm for regressionand working of this algorithm is presented in Fig.6. It consists of following features

(i) Bootstrap Sampling:

It is selecting different data sets and considered for training with data replacement randomly. In the fields of statistics and machine learning, this resampling method involves repeatedly selecting samples from the original dataset with replacement, typically to estimate a population parameter.

(ii) Training Multiple Trees:

Each subset of data is used to train a decision tree. In the forest, each tree learns from a randomly selected sample of the data.

(iii) Aggregation of Results:

The final prediction is derived by averaging the predictions of all individual decision trees.

Python Sample Code for Regression

```
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error

# Generating some sample data
X = np.random.rand(100, 5)
y = np.random.rand(100)

# Splitting data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initializing the Random Forest Regressor
rf = RandomForestRegressor(n_estimators=100, random_state=42)

# Training the model
rf.fit(X_train, y_train)

# Making predictions
y_pred = rf.predict(X_test)

# Evaluating the model
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")
```

4. Results and Discussions

The Random Forest Regressor is applied to the data frame described in Table.1 with the default parameters viz. number of estimators=100, error criteria is squared error, boost strap sampling, minimum sample split=2 , minimum sample leaf=1 ,minimum weight fraction leaf=0 etc.

In this work, total data frame consists of 209930 rows × 6 columns, in which 80% of the data is used for training and testing. The size of the data frame for training is 167944 rows x 5 columns and for testing is 41986 rows x 5 columns.It is observed that the RFR algorithm is estimated the battery health successfully. The Fig.7 shows the predicted Maximum capacity (Ampere-hour (Ah)) of the battery with respect to the Actual Maximum Capacity (Ah) of the Battery. It is observed that Actual and estimated values nearly same. The absolute error between these two values is shown in Fig.8. The absolute error is negligibly less and is near to zero value. It is a known fact that the capacity of the battery will reduce with the aging of a battery due to more number of charging & discharging cycles. The Fig. 9 shows the actual and predicted maximum capacity of the battery with respect to the battery aging. It is also observed that the predicted value is as expected with actual value with respect to the battery aging. The mean square error (MSE) of prediction is 0.0056448.

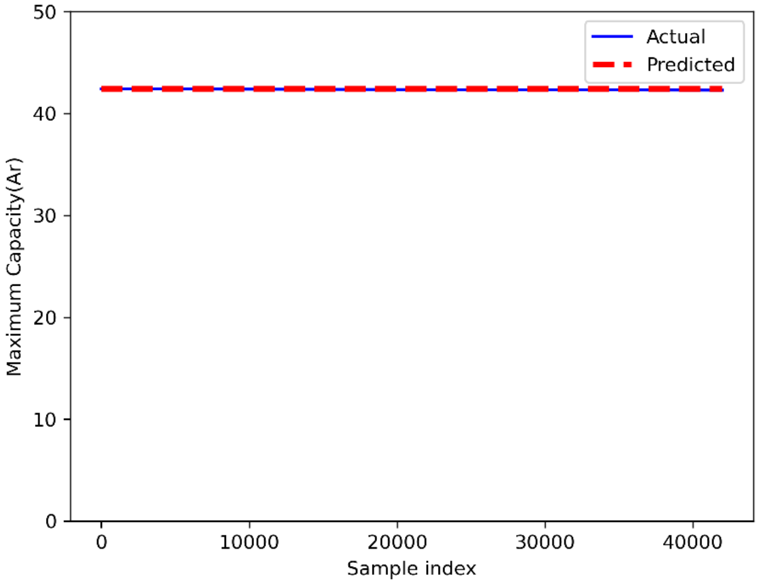


Fig.7 Maximum Capacity of a Battery Predicted with respect to Actual

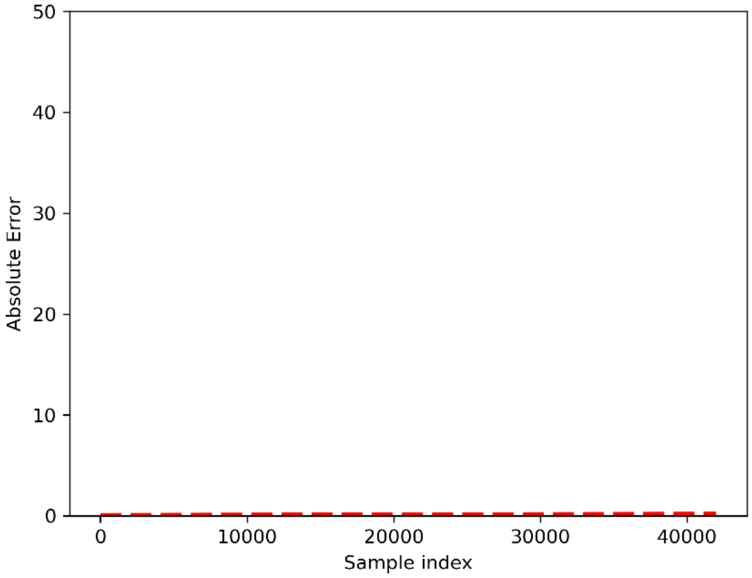


Fig.8Absolute error between predicted Maximum Capacity and Actual Value

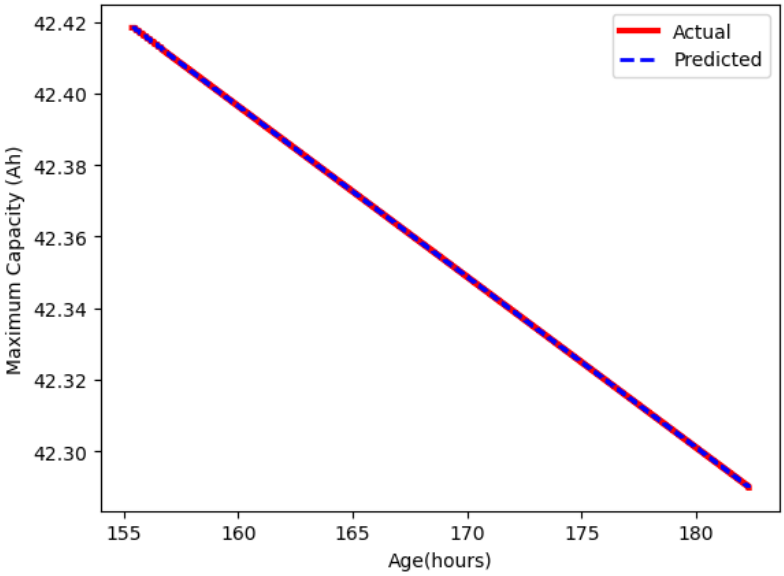


Fig.9 Maximum Capacity (Ah) of Battery withthe increase of Aging (number of charging & discharging cycles) and their actual and predicted values

5. Conclusion

In this work, the use of Random Forest Regression (RFR) to calculate battery health is investigated. The study leverages a comprehensive dataset comprising various operational parameters such as voltage, current, SOC, Temperature, Age, Maximum Capacity of lithium-ion batteries. By training the Random Forest model on this data, captured the complex, non-linear relationships inherent in battery aging processes. In this work successfully predicted the battery Maximum capacity with respect to the other operational conditions.

References

[1] N. Omar, M. A. Monem, Y. Firouz, J. Salminen, J. Smekens, O. Hegazy, H. Gaulous, G. Mulder, P Van den Bossche, T. Coosemans, J. Van Mierlo, Lithium iron phosphate based battery - Assessment of the aging parameters and development of cycle life model, Applied Energy, Volume 113, January 2014, Pages 1575-1585.

[2] H. Phan, M. Maaß, R. Mazur and A. Mertins, "Random Regression Forests for Acoustic Event Detection and Classification," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 1, pp. 20-31, Jan. 2015, doi: 10.1109/TASLP.2014.2367814.

[3] M. S. HossainLipu et al., "Real-Time State of Charge Estimation of Lithium-Ion Batteries Using Optimized Random Forest Regression Algorithm," in IEEE Transactions on Intelligent Vehicles, vol. 8, no. 1, pp. 639-648, Jan. 2023, doi: 10.1109/TIV.2022.3161301.

[4] J. S. Rhodes, A. Cutler and K. R. Moon, "Geometry- and Accuracy-Preserving Random Forest Proximities," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 9, pp. 10947-10959, 1 Sept. 2023, doi: 10.1109/TPAMI.2023.3263774.

[5] Y. Liu and H. Liang, "A Data-Driven Approach for Electric Bus Energy Consumption Estimation," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 10, pp. 17027-17038, Oct. 2022, doi: 10.1109/TITS.2022.3152679.

[6] P. Sharma, P. Dadheech, N. Aneja and S. Aneja, "Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning," in IEEE Access, vol. 11, pp. 111255-111264, 2023, doi: 10.1109/ACCESS.2023.3321861.

[7] Q. Zhang, L. Li, R. Sun, D. Zhu, C. Zhang and Q. Chen, "Retrieval of the Soil Salinity From Sentinel-1 Dual-Polarized SAR Data Based on Deep Neural Network Regression," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 4006905, doi: 10.1109/LGRS.2020.3041059.

[8] M. A. Hannan, J. A. Ali, A. Mohamed and M. N. Uddin, "A Random Forest Regression Based Space Vector PWM Inverter Controller for the Induction Motor Drive," in IEEE Transactions on Industrial Electronics, vol. 64, no. 4, pp. 2689-2699, April 2017, doi: 10.1109/TIE.2016.2631121.

[9] W. Shen, Y. Guo, Y. Wang, K. Zhao, B. Wang and A. Yuille, "Deep Differentiable Random Forests for Age Estimation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 2, pp. 404-419, 1 Feb. 2021, doi: 10.1109/TPAMI.2019.2937294.
[10] L. Yin, Z. Sun, F. Gao and H. Liu, "Deep Forest Regression for Short-Term Load Forecasting of Power Systems," in IEEE Access, vol. 8, pp. 49090-49099, 2020, doi: 10.1109/ACCESS.2020.2979686.
[11] H. Sahour, V. Gholami, J. Torkaman, M. Vazifedan, and S. Saeedi, ‘Random forest and extreme gradient boosting algorithms for streamflow modeling using vessel features and tree-rings’, Environmental Earth Sciences, vol. 80, no. 22. Springer Science and Business Media LLC, Nov. 2021. doi: 10.1007/s12665-021-10054-5.
[12] 1. O. Tremblay, L.-A. Dessaint, A.-I. Dekkiche, A Generic Battery Model for the Dynamic Simulation of Hybrid Electric Vehicles, 2007 IEEE® Vehicle Power and Propulsion Conference, September 9-13, 2007, Arlington/Texas, USA.