Engineering Journal
issn: 0125-8281

Article

Predictive State of Charge (SoC) Modeling Using Machine Learning Algorithms in Lithium-Ion NMC Batteries

Farah Apit Tantri[1], Dewanto Harjunowibowo[1], Endah Retno Dyartanti[2,3], Muhammad Nizam[2,4], Mufti Reza Aulia Putra[2,4], Rekyan Regasari MP[5], Tiong Hoo Lim[6], and Anif Jamaluddin[1,2,*]

1 ESMART, Department of Physics Education, Faculty of Teacher Training and Education, Universitas Sebelas Maret, Indonesia
2 Center of Excellence for Electrical Energy Storage Technology, Universitas Sebelas Maret, Surakarta, Indonesia
3 Chemical Engineering Department, Faculty of Engineering, Universitas Sebelas Maret, Surakarta, Indonesia
4 Electrial Engineering Department, Faculty of Engineering ,Universitas Sebelas Maret, Surakarta, Indonesia
5 Department of Informatics, Faculty of Computer Science, Brawijaya University, Malang, Indonesia
6 Department of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Brunei, Gadong, Brunei Darussalam
*E-mail: elhanif@staff.uns.ac.id (Corresponding author)

Abstract. The State of Charge (SoC) estimation is crucial in lithium-ion batteries to prevent excessive charging and discharging, impacting the battery's safety, stability, and efficiency. Conventional techniques are the most frequently employed method for estimating SoC. However, they are less accurate in predicting SOC due to their computational sensitivity and difficulty adapting to complex environments. This study proposed four machine learning models: Linear Regression, Multilayer Perceptron, Decision Tree, and Random Forest that were applied for SoC prediction on lithiom-ion NMC batteries. The models' performance was evaluated based on Correlation Coefficient and error values (Mean absolute error or MAE and Root mean square error or MRSE). Based on the result, the Random Forest model exhibited the best performance with a Correlation Coefficient of 1, and MAE and MRSE values of 0.2052 and 0.2712, respectively. Conversely, the Linear Regression model demonstrated the worst performance, with a Correlation Coefficient of 0.9534 and MAE and MRSE values of 5.9064 and 8.2602, respectively.

Keywords: State of charge (SoC), NMC batteries, machine learning.

ENGINEERING JOURNAL Volume 28 Issue 9
Received 12 November 2023
Accepted 13 September 2024
Published 30 September 2024
Online at https://engj.org/
DOI:10.4186/ej.2024.28.9.1

1. Introduction

The use of renewable energy has been increasing recently. Based on previous research, the use of renewable energy from 2000 to 2020 has increased 260 times from 0.03 EJ to 7.79 EJ [1]. As a result of the higher utilization of renewable energy, more energy storage technology is urgently needed to optimize renewable energy storage. There are various energy storage technologies, one of which is the battery. Batteries are widely used as energy storage because they have high endurance, easy maintenance, scalability, and low social-environmental impact. There are two types of batteries that can be used for energy storage: primary and secondary. Primary batteries are disposable batteries, while secondary batteries are rechargeable batteries [2]. The lithium-ion battery is a type of secondary battery with the advantages of high efficiency, high energy density, and environmental friendliness. Therefore, lithium-ion batteries are promising for use in renewable energy storage [3].

Lithium-ion batteries mainly comprise cathode, anode, and electrolyte solution parts. Based on the cathode type, lithium-ion batteries are divided into six, namely Lithium Cobalt Oxide (LCO), Lithium Manganese Oxide (LMO), Lithium Nickel Manganese Cobalt Oxide (NMC), Lithium Iron Phosphate (LFP), Lithium Nickel Cobalt Aluminum Oxide (NCA), Lithium Titanate Oxide (LTO) [4]. Each type of lithium-ion battery has its advantages and disadvantages. LMO-type batteries have the advantage of being low-cost but have a limited life. LCO-type batteries have a moderate cost and high energy, but low safety level. LFP-type batteries have a long service life but lower energy density compared to nickel-based batteries (NMC and NCA) [5]. The NCA-type battery offer high energy density but have unstable thermal stability, making them less safe to use. NMC-type batteries, on the other hand, have high energy density and longer lifetime [6]. Additionally, NMC batteries have high volumetric and gravimetric capacity, high nominal voltage, and low self-discharge [7]. Given these considerations, this research uses NMC batteries due to their significant advantages in renewable energy storage.

Renewable energy storage is more secure, efficient, and stable if the battery's SoC value can be known [8]. SoC estimation is used to determine the ratio of the remaining capacity of the battery to the full capacity of the battery [9]. An accurate SoC estimation can protect the battery by preventing overcharging [10]. The most commonly used SoC estimation methods are the conventional methods. The methods can be classified as Feed-Forward Backpropagation Neural Network (FF-BPNN), Coulomb Counting, Piecewise Open Circuit Voltage (OCV), and Kalman Filter [11]. However, conventional methods are less accurate in predicting SoC due to computational sensitivity and are difficult to adapt to complex environments [9]. Therefore, adopting a machine learning model is required to facilitate and provide accurate results in predicting SoC lithium-ion NMC batteries.

Machine learning is a science applied to teach computers to learn without being programmed explicitly and is part of Artificial Intelligence [12]. Machine learning uses data as input, which is then analyzed to find patterns. Based on the learning method, machine learning is divided into three types: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is predictive hence it can estimate SoC on lithium-ion batteries. Further, based on the target, supervised learning is divided into two: Classification and Regression [12]. SoC prediction on lithium-ion batteries can be applied to supervised learning regression because the prediction target is the SoC value, which includes continuous data.

Previous researchers have conducted research to predict SoC on lithium-ion batteries using machine learning. Machine learning models usually used to make SoC predictions include Linear Regression, Multilayer Perceptron, Decision Tree, Random Forest, and KNN. Research conducted by Zazoum employed a Linear Regression machine learning model to predict SoC lithium-ion batteries and obtained perfect results with MAE, MRSE, and R^2 values of 1.4065, 1.1120, and 1, respectively [13]. Another research has shown that a decision tree model was suitable for predicting SoC with RMSE and MSE values of 2.56 and 10.03, respectively [14]. The multilayer perceptron model could also achieve good results for predicting lithium-ion batteries with MAE values of 0.0055 – 0.0082 and MRSE values of 0.0077 [15]. In addition, the random forest model had a small error value when it was used to predict the SoC lithium-ion NMC battery with MAE values of 0.33, 0.23, 0.21, and MRSE values of 0.91, 0.65, and 0.58 [16]. However, this research has not discussed the correlation coefficient result and the comparison with other models.

Therefore, this paper evaluated four machine-learning models to predict the SoC precision of NMC lithium-ion batteries. The machine learning models used were Multilayer Perceptron, Decision Tree, and Random Forest. Then, the prediction results were compared by observing the Correlation Coefficient, MAE, and MRSE values to obtain the best machine-learning model. Besides, using the same type of battery on each model will give a fair result compared to using a different type of battery.

2. Method

To find the best model for predicting the SoC values of the NMC battery, Fig. 1 presents the stages of the study involving preparation, dataset cleaning, dataset processing, and analysis.

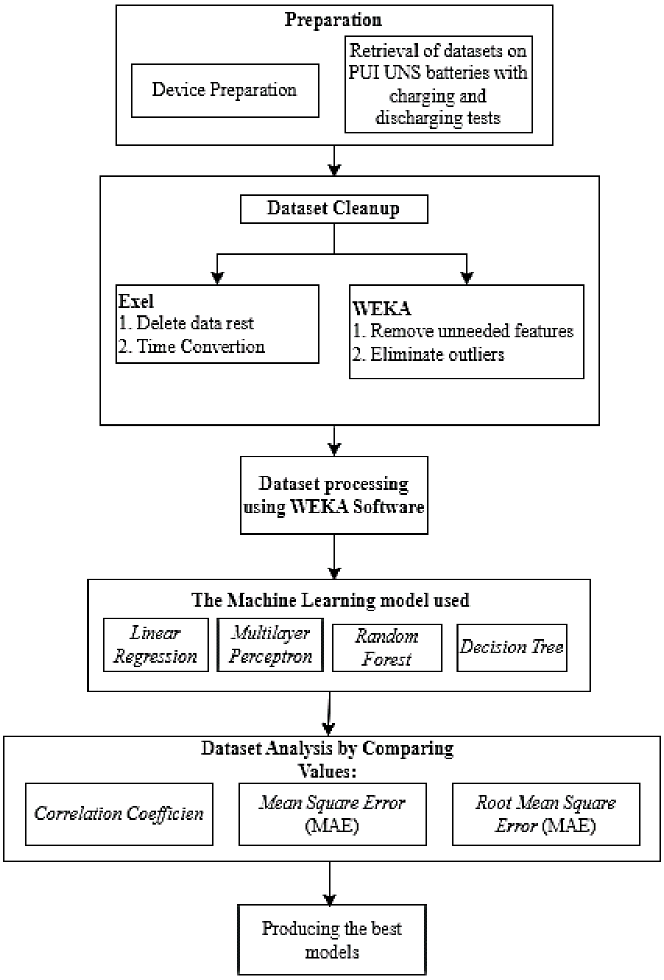

Fig. 1. Research stages.

2.1. Preparation

The dataset used in this study was secondary data taken by conducting a charging and discharging test using the Neware Battery Testing System. The charging and discharging test was done by varying the C-rate values and applying each for five cycles. Table 1 exhibits the C-rates variation on the charging and discharging test.

Table 1. C-Rate Variation Test Procedure Steps

C-Rate	Detail Steps	Cycle
2.2	Charge → Rest → Discharge → Rest	5
1.4	Charge → Rest → Discharge → Rest	5
0.7	Charge → Rest → Discharge → Rest	5
0.4	Charge → Rest → Discharge → Rest	5
0.2	Charge → Rest → Discharge → Rest	5

2.2. Data Cleaning

After obtaining the dataset, the raw data was cleaned using Excel and WEKA. Cleaning data in Excel was done by deleting the rest of the data, performing time conversions, and grouping C-rates. After that, dataset cleaning with WEKA was completed by removing unnecessary features/variables and outlier data. Outlier detection was carried out using the interquartile range (IQR) following Eq. (1) [17].

$$IQR = Q3 - Q1 \tag{1}$$
$$Outlier\ r_u > Q3 + 1.5 \times IQR$$
$$Outlier\ r_L < Q1 + 1.5 \times IQR$$

2.3. Dataset Processing Using Machine Learning

After cleaning the dataset, the data was processed using four machine learning models: Linear Regression, Multilayer Perceptron, Random Forest, and Decision Tree.

2.3.1. Linear regression

Linear Regression is a machine learning model that can be employed to estimate numerical values based on simple model parameters. It is suitable for datasets in numerical format, allowing for accurate estimation. Besides, the model can be applied for the estimation if one or several variables affect another variable [18]. The Linear Regression model works by examining the relationship between the dependent and independent variables through a straight line (regression line). The correlation value describes the strength of the relationship between two continuous variables and is expressed as Pearson product-moment correlation [19].

Linear Regression is divided into Simple Linear Regression and Multiple Linear Regression. The simple Linear Regression Modelling technique is used to investigate the relationship between one dependent variable with one independent variable [20]. Meanwhile, multiple Linear Regression is employed to analyze the relationship between one or more independent parameters with one dependent parameter using several equations [18]. The formula for multiple linear Regression can be written as Eq. (2) [21]. In which y_i is a dependent variable, x_i is independent variable, β_0 is intercept (constant term), β_p is coefficient of slope of each independent variable, and ε is the model's error.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \cdots \beta_{pn} x_{ip} + \varepsilon \tag{2}$$

2.3.2. Multilayer perceptron

The Multilayer Perceptron is a feed-forward structure (information that flows in one direction) usually employed in Artificial Neural Network (ANN) architectures [22]. It consists of three layers of the network, namely the input layer (I), the output layer (L), and one hidden layer (J). The workings of the Multilayer Perceptron are depicted in Fig. 2. The number of nodes in the input layer shows the number of independent variables in the data. The number of nodes in the output layer reveals the number of dependent variables, and the hidden layer demonstrates the constraints that can be modified. There are weighted connections between the input nodes to the hidden layer nodes and the output nodes to the hidden layer. The Multilayer Perceptron model has bias nodes with magnitude +1 in both the input, output, and hidden layers. The bias node will be connected to the next node, which functions to set the output of the weights inserted into the neurons [22].

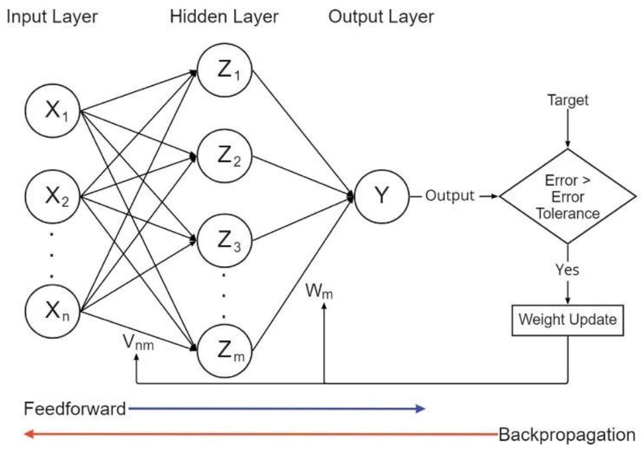

Fig. 2. Multilayer Perceptron Work.

2.3.3. Decision Tree

A decision tree is another machine learning model that can be employed for classification (classification tree) or Regression (regression tree) tasks. The model works by recursively splitting the dataset into smaller classes, and the results are represented in leaf nodes. Through this process, the training data forms a tree structure that enables predictions to be made, as illustrated in Fig. 3.

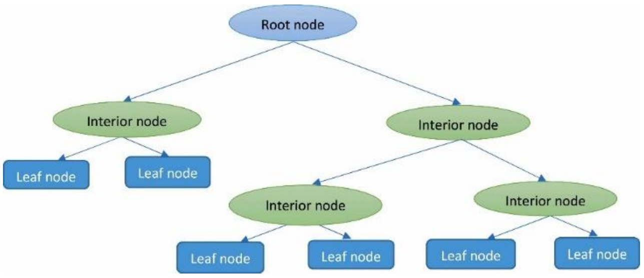

Fig. 3. Decision Tree Work.

The Decision Tree has three nodes: the root, interior, and leaf. The root node is the first broken node to form more nodes. Interior nodes represent data features and decision rules. In contrast, the leaf node is the final result of the decision [23].

The Decision Tree machine learning model operates on breaking down a complex problem into smaller, more manageable subproblems to obtain easily interpretable solutions. In this model, each leaf node denotes a simple linear regression. The Decision Tree generally divides data by data entropy and performs local linear Regression on the new data subset. The number of subsets will double as the data is split so that the tree will get deeper. The success of the Decision Tree machine learning model can be seen from the mean square error (MSE) predicted results with actual data. The output of the Decision Tree model is parameter weights and correlations. Parameters highly influencing the target will have a heavyweight [23].

2.3.4. Random Forest

Random Forest is a machine-learning model with working principles based on the Decision Tree [24]. The Random Forest randomly selects features to create several uncorrelated decision trees. The resulting decision trees are combined to produce a prediction [25].

In the Random Forest model, a decision tree is constructed using a Classification and Regression Tree (CART) by selecting a random variable and a random sample from the training data in the Random Forest regression. The steps for the Random Forest machine learning model include:

1) Sampling of data and random samples (bagging)
   Dataset D contains various feature variables p. Then the training data n is taken randomly from dataset D using the bootstrap method (sample taking with put-back). Testing data or Out-Of-Bag (OOB) data in dataset D is data that does not include training data, and it plays a crucial role in the Random Forest model as it is used to validate the model during the training process.
2) Random selection of feature variables
   The selection of the specific feature variable k (k<p) is taken from training data randomly from the variable fitting a branch in the classification tree to construct a regression. The Gini coefficient is employed to measure the impurity of the variable. Regression tree branches are generated for training data; thus, regression trees will form the Random Forest model [27]. Random Forest works as a team with multiple decision trees. The dataset is divided into various subsets, and predictions are made for each subset. After that, each subset of predictions is combined to provide an overall prediction [26].

2.4. Dataset Analysis

Furthermore, the results from applying four machine learning models were analyzed by comparing their Correlation Coefficient, MAE, and MRSE values to get the best model.

The Mean Absolute Error (MAE) is a crucial metric in our analysis, as it measures the average error of the data. A MAE value of 1 signifies a significant difference between the predicted and actual class, while 0 indicates a precise prediction. Equation (3) presents the formula for determining the MAE value, where f_i is predicted values, y_i is actual values, and n is the data number.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| \quad (3)$$

Root mean square error (RMSE) is a method applied to measure the error of a quantitative prediction model by knowing the distribution of data point deviations from the data around a linear line. RMSE can be calculated by Eq.

DOI:10.4186/ej.2024.28.9.1

(4) [27]. In which, $\hat{y}_1, \hat{y}_2, \dots \dots, \hat{y}_n$ is the predicted values, $y_1, y_2, \dots \dots, y_n$ denotes observed values, and n is data number. The smaller RMSE value indicates a higher accuracy value [28].

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(y_1 - \hat{y}_1)^2}{n}} \tag{4}$$

3. Results and Discussion

The cleaned NMC 18650 battery dataset was used to generate plots that are presented in Fig. 4. This graph was utilized as a parameter to compare the accuracy of the SoC prediction applying the machine learning models. Figure 4 displays the charging curve (Fig. 4(a)) and discharging curve (Fig. 4(b)) of the NMC 18650 battery with 5 different C-rate values, which were the actual data. Each C-Rate value was represented by a distinct point shape.

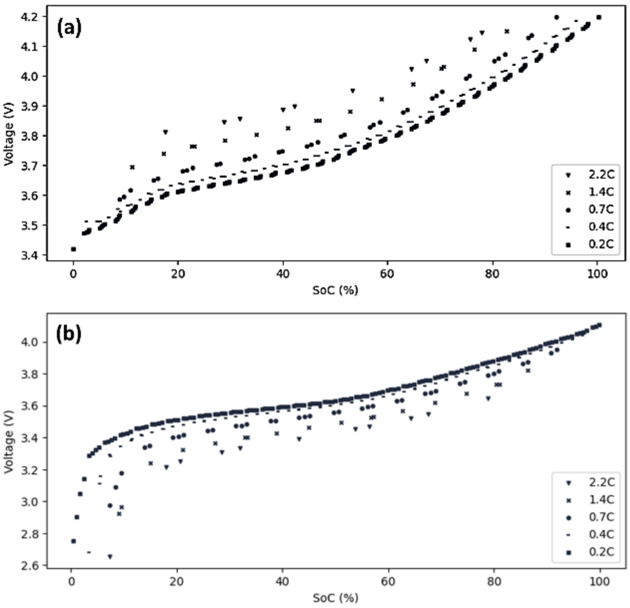

Fig. 4. Charge and discharge voltages at various C-rates in dependence on SoC for the NMC battery.

The charging curve (Fig. 4(a)) constantly increased, starting from 0% SoC to 100% or from the voltage of 3.4 V to 4.2 V, exhibiting the maximum voltage of the NMC battery. Meanwhile, in the discharging process (Fig. 4(b)), it improved dramatically at the cut-off voltage of 2.6 V until it approached the value of 3.4 V, then the curve rose constantly.

3.1. Linear Regression

The SoC prediction in the charging and discharging conditions obtained with the Linear Regression is presented in Fig. 5. In the charging process, the curve starts forming at SoC points of more than 7.2% to a SOC value of 108%, while the discharging graph begins developing at the SoC value of 1.77% to more than 104.5%. Figure 6 exhibits the plot of the actual and predicted SoC in the charging and discharging process using the Linear Regression model, drawing specific curves. It denotes that the prediction results differ from the actual data. Besides, the charge and discharge voltages vs. the SoC curves also demonstrate distinct shapes, with the curves plotted in Fig. 4 as the actual data. This difference might be due to how linear regression works, drawing a straight line to find the relationship between the dependent and independent variables [19]. The Correlation Coefficient, MAE, and MRSE values of the SoC prediction achieved using this model were 0.9534, 5.9064, and 8.2602, respectively.

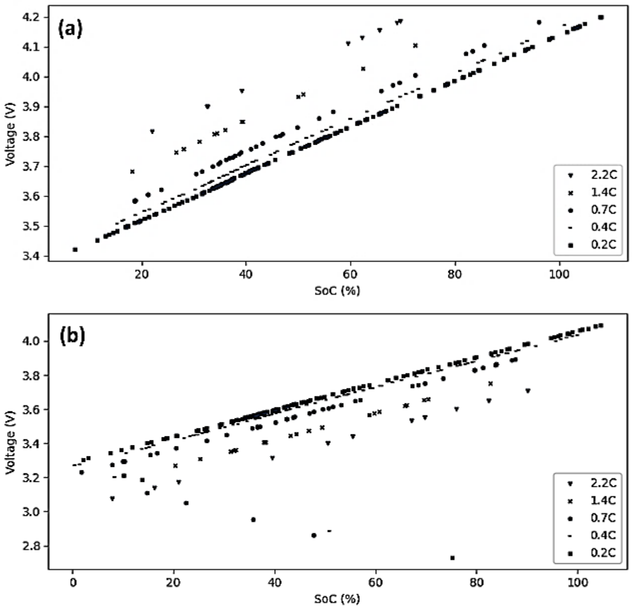

Fig. 5. Charge and discharge voltages at various C-rates vs. predicted SoC using the Linear Regression model.

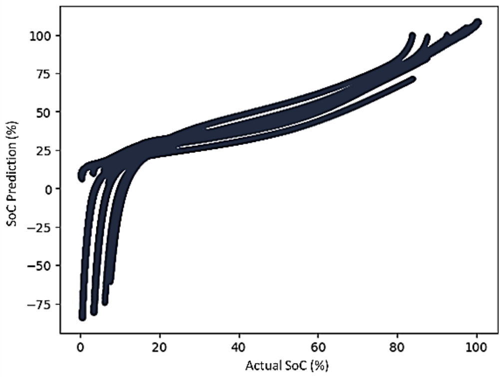

Fig. 6. Actual SoC Data with Predicted SoC on Linear Regression.

3.2. Multilayer Perceptron

The profile of SoC prediction in the charging and discharging processes using the Multilayer Perceptron is presented in Fig. 7. In the charging process (Fig. 7(a)), the curve begins at the SoC point of 2.5% and ends at 96.1%, starting at a voltage of 3.4 V up to 4.2 V. Meanwhile, the discharging curve (Fig. 7(b)) rises drastically from a voltage of 2.6V to a value nearly 3.4 V then increases steadily until it reaches 4.2V. The discharging curve forms at the SoC value of 0% to less than 94.98%. It shows that the minimum and maximum voltages of the charge and discharge did not match the actual data. Besides, the Correlation Coefficient, MAE, and MRSE values obtained using this model were 0.9884, 3.3571, and 4.172, respectively. Thus, this model's prediction results are considered to have high enough errors.

Another study conducted by Hossain in 2017 predicting SoC lithium-ion NMC batteries with the Multilayer Perceptron model using Backpropagation Neural Network (BPNN) demonstrated promising results with MAE values around 0.0055 – 0.0082 and MRSE values in about 0.0077 – 0.0108 [15]. The difference in the error values in the two studies was very adrift because Hossain employed the BPNN, where the iterations in the training process would update the weights and bias so that the error value could be smaller.

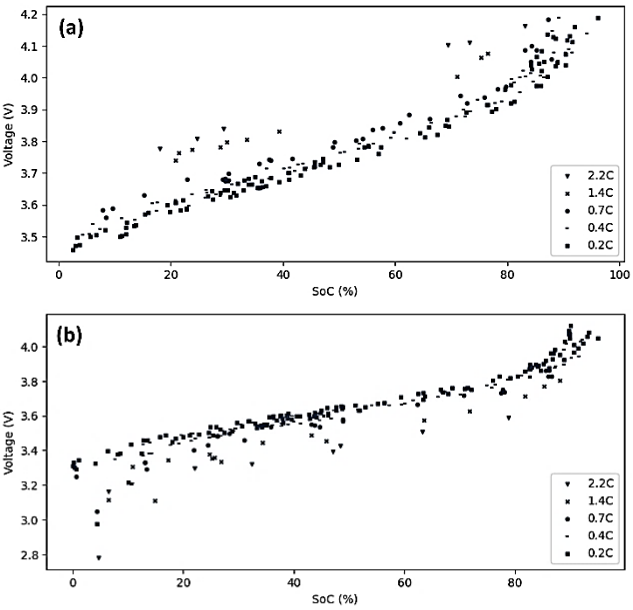

Fig. 7. Charge and discharge voltages at various C-rates vs. predicted SoC using the Multilayer Perceptron model.

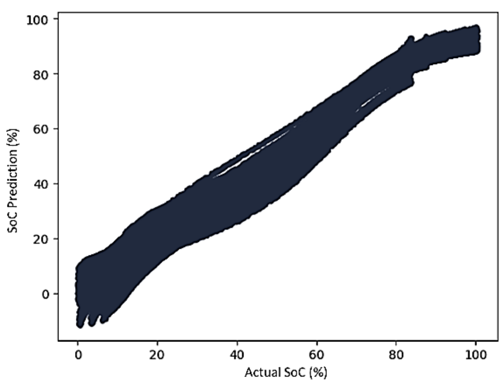

Fig. 8. Actual SoC Data with Predicted SoC on Multilayer Perceptron.

Furthermore, Fig. 8 illustrates the relationship between SoC values of the actual data and predicted SOC values using the Multilayer Perceptron model. The graph depicts a line with a wide shape, indicating that the prediction results had spreading values. It suggests that the model was less accurate in predicting the SoC value of NMC lithium-ion batteries.

3.3. Decision Tree

The curves of SoC prediction vs. voltages in the charging and discharging process through the Decision Tree are presented in Fig. 9. The charging curve (Fig. 9(a)) depicts a linear increase between the voltages and the SoC values. The curve starts at a voltage of 3.4V up to 4.2V, with SoC prediction beginning from 0.9% and ending at less than 99% value. The discharging curve (Fig. 9(b)) appears to increase drastically from the voltage of 2.6 V to close to 3.4 V, and then it constantly grows up to 4.2 V. Besides, it begins at the SoC value greater than 1.39% to a value less than 99.2%. After processing the dataset, the Correlation Coefficient, MAE, and MRSE values attained were 0.9997, 0.537, and 0.6615, respectively. The charge and discharge voltage range almost correspond to the actual data. Besides, the curve shapes are similar to the actual data in Fig. 4. This model's results also produce relatively small error values.

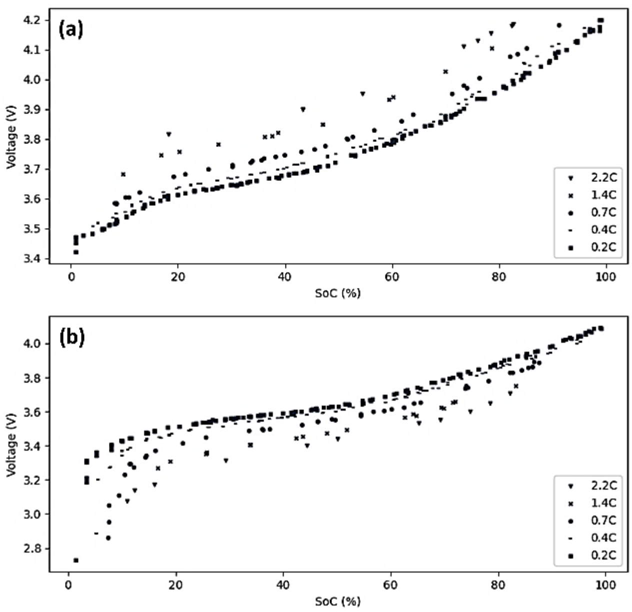

Fig. 9. Charge and discharge voltages at various C-rates vs. predicted SoC using the Decision Tree model.

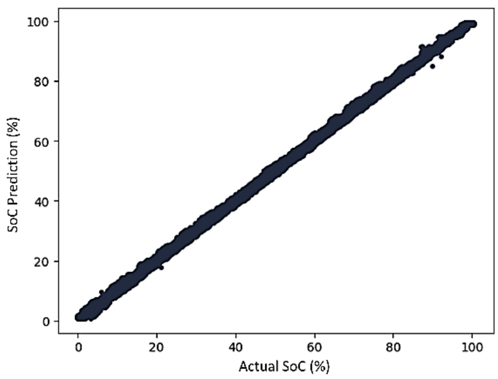

Fig. 10. Graph of actual and predicted SoC with the Decision Tree model.

Figure 7 displays the curve between actual and predicted SoC with the Decision Tree model. The curve draws a relatively heavy straight line representing small dots outside the straight line. It denotes that some predicted SoC data did not match the actual SoC value. Nevertheless, the error values obtained were relatively small because the model principle breaks down a big problem into smaller problems to get an easy-to-understand solution [23]. From the results, it can be seen that the Decision Tree model was quite good at predicting the SoC value of the NMC lithium-ion battery.

3.4. Random Forest

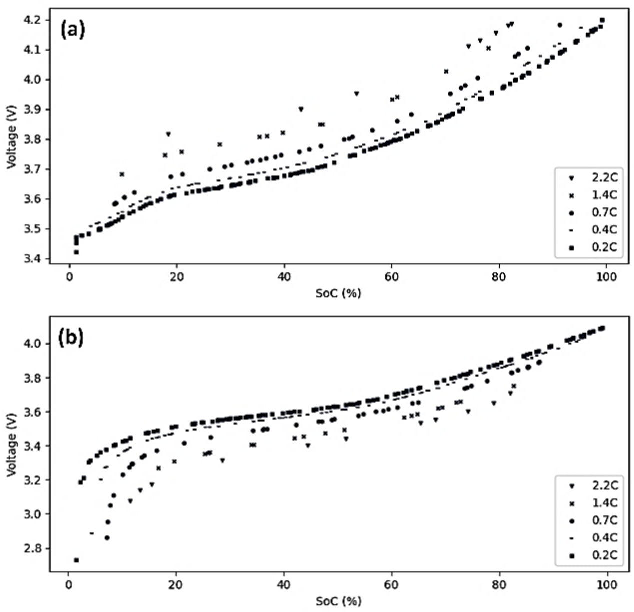

Fig. 11. Charge and discharge voltages at various C-rates vs. predicted SoC using the Random Forest model.

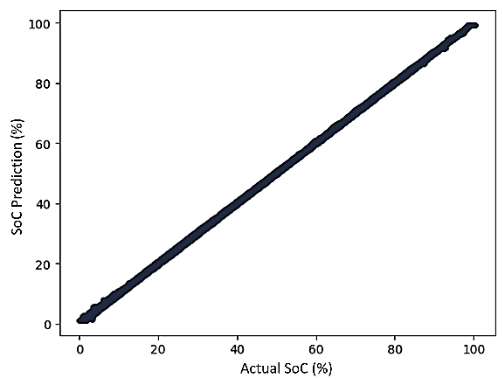

Fig. 12. Actual SoC Data with Predicted SoC on Random Forest.

The relationship between the voltages and predicted SoC in the charging and discharging process is displayed in Fig. 11. The charging curve (Fig. 11(a)) increases at a voltage of 3.4V up to 4.2V, with SoC values creating at 1.3% and ending at 99%. Meanwhile, the discharging curve (Fig. 11(b)) improves drastically from 2.6 V to nearly 3.4 V, then rises constantly to 4.2 V. Besides, it begins at SoC values greater than 1.5% to less than 99%. The dataset processed using the Random Forest machine learning model results in Correlation Coefficient, MAE, and MRSE values of 1, 0.2052, and 0.2712, respectively. The charge and discharge voltage range using this model was almost the same as the actual data. Figure 12 reveals the graph between the actual and predicted SoC values with the Random Forest model. It draws a thin linear line indicating a good prediction result. In fact, the SoC prediction with the Random Forest model revealed Correlation Coefficient, MAE, and MRSE values of 1, 0.2052, and 0.2712, respectively. The Correlation Coefficient of 1 denotes the perfect prediction results.

Table 2 exhibits the Correlation Coefficient, MAE, and MRSE values of the SoC prediction using the four machine learning models. It can be seen that the Random Fores was the most accurate model compared to others due to its lowest MAE and MRSE values, while the Linear Regression demonstrated the contrary results with the highest errors. Other research by Lipu in 2018 applying the Random Forest model to predict the NMC Battery SoC also resulted in relatively small error values of MAE (0.33, 0.23, and 0.21) and MRSE (0.91, 0.65, and 0.58) [16]. The values were close to those obtained in this research, denoting that the Random Forest model is suitable for predicting SoC on NMC batteries. Nevertheless, there was a difference between the two studies where the Lipu study used three temperatures to produce MAE and MRSE, resulting in 3 each. In contrast, this study only applied one temperature, creating one MAE and MRSE value.

Table 2. Correlation Coefficient, MAE, and MRSE values of the SoC prediction using the four machine learning models.

Machine learning model	Correlation Coefficient	MAE	MRSE
Linear Regression	0.9534	5.9064	8.2602
Multilayer Perceptron	0.9884	3.3571	4.172
Decision Tree	0.9997	0.537	0.6615
Random Forest	1	0.2052	0.2712

4. Conclusion

The prediction of the SoC values of lithium-ion NMC batteries with the machine learning approach has been conducted using four machine models, including Linear Regression, Multilayer Perceptron, Decision Tree, and Random Forest. The Correlation Coefficient, MAE, and MRSE values obtained from the four models were compared to reveal the best model to predict NMC battery SoC. The results showed that the Random Forest model had the best performance, with Correlation Coefficient, MAE, and MRSE values of 1, 0.2052, and 0.2712, respectively.

DOI:10.4186/ej.2024.28.9.1

Acknowledgment

The author thanks UNS Sebelas Maret, which has financially supported this Project with contract number 228/UN27.22/PT.01.03/2023.

References

[1] C. C. Lee, F. Wang, and Y. F. Chang, "Does green finance promote renewable energy? Evidence from China," *Resour. Policy*, vol. 82, no. December 2022, pp. 103439, 2023. doi: 10.1016/j.resourpol.2023.103439.

[2] M. Nasution, "Karakteristik Baterai Sebagai Penyimpan Energi Listrik Secara Spesifik," *JET (Journal Electr. Technol.*, vol. 6, no. 1, pp. 35–40, 2021. [Online]. Available: https://jurnal.uisu.ac.id/index.php/jet/article/view/3797.

[3] L. da Silva Lima *et al.*, "Life cycle assessment of lithium-ion batteries and vanadium redox flow batteries-based renewable energy storage systems," *Sustain. Energy Technol. Assessments*, vol. 46, pp. 101286, 2021. doi: 10.1016/j.seta.2021.101286.

[4] F. A. Perdana, "Baterai Lithium," *INKUIRI J. Pendidik. IPA*, vol. 9, no. 2, p. 113 - 109, 2021. doi: 10.20961/inkuiri.v9i2.50082.

[5] I. Viantama and B. M. Suyitno, "Analisis Perbandingan Sistem Kinerja Motor Penggerak Pada Mobil Listrik Kapasitas 75 kWh," *J. Asiimetrik J. Ilm. Rekayasa Inov.*, vol. 3, pp. 157–164, 2021. doi: 10.35814/asiimetrik.v3i2.2083.

[6] G. Berckmans, M. Messagie, J. Smekens, N. Omar, L. Vanhaverbeke, and J. V. Mierlo, "Cost projection of state of the art lithium-ion batteries for electric vehicles up to 2030," *Energies*, vol. 10, no, 9, pp. 134, 2017. https://doi.org/10.3390/en10091314.

[7] M. Malik, K. H. Chan, and G. Azimi, "Review on the synthesis of LiNixMnyCo1-x-yO2 (NMC) cathodes for lithium-ion batteries," *Mater. Today Energy*, vol. 28, p. 101066, 2022. doi: 10.1016/j.mtener.2022.101066.

[8] Z. Ren and C. Du, "A review of machine learning state-of-charge and state-of-health estimation algorithms for lithium-ion batteries," *Energy Reports*, vol. 9, pp. 2993–3021, 2023. doi: 10.1016/j.egyr.2023.01.108.

[9] J. Hong, Z. Wang, W. Chen, L. Y. Wang, and C. Qu, "Online joint-prediction of multi-forward-step battery SOC using LSTM neural networks and multiple linear regression for real-world electric vehicles," *J. Energy Storage*, vol. 30, no. February, p. 101459, 2020. doi: 10.1016/j.est.2020.101459.

[10] W.Y. Chang, "The State of Charge Estimating Methods for Battery: A Review," *ISRN Appl. Math.*, vol. 2013, no. 1, pp. 1–7, 2013. doi: 10.1155/2013/953792.

[11] M. N. Habibi, M. Imron, D. Prasetyo, and N. A. Windarko, "Estimasi State of Charge (SOC) Pada Baterai Lithium – Ion Menggunakan Feed-Forward Backpropagation Neural Network Dua Tingkat," vol. 8, no. 2, pp. 82 - 91, 2020. DOI: https://doi.org/10.32487/jtt.v8i2.846

[12] W. K. Bickel, D. C. Tomlinson, W. H. Craft, M. Ma, C. L. Dwyer, Y. H. Yeh, A. N. Tegge, R. F. Lemos, L. N. Athamneh*e*, "Predictors of smoking cessation outcomes identified by machine learning: A systematic review," vol. 6, no. January, 2023. doi: 10.1016/j.addicn.2023.100068

[13] B. Zazoum, "Lithium-ion battery state of charge prediction based on machine learning approach," *Energy Reports*, vol. 9, pp. 1152–1158, 2023. doi: 10.1016/j.egyr.2023.03.091

[14] S. Jafari, Z. Shahbazi, Y. C. Byun, and S. J. Lee, "Lithium-ion battery estimation in online framework using extreme gradient boosting machine learning approach," *Mathematics*, vol. 10, no. 6, p. 888, 2022. doi: 10.3390/math10060888

[15] M. S. Hossain Lipu, M. A. Hannan, and A. Hussain, "Feature selection and optimal neural network algorithm for the state of charge estimation of lithium-ion battery for electric vehicle application," *Int. J. Renew. Energy Res.*, vol. 7, no. 4, pp. 1701–1708, 2017. doi: 10.20508/ijrer.v7i4.6237.g7211

[16] M. S. H. Lipu, A. Ayob, M. H. M. Saad, A. Hussain, M. A. Hannan, and M. Faisal, "State of charge estimation for lithium-ion battery based on random forests technique with gravitational search algorithm," *Asia-Pacific Power Energy Eng. Conf. APPEEC*, vol. 2018-Octob, pp. 45–50, 2018. doi: 10.1109/APPEEC.2018.8566648

[17] I. Cho, S. Park, and J. Kim, "A fire risk assessment method for high-capacity battery packs using interquartile range filter," *J. Energy Storage*, vol. 50, no. April, p. 104663, 2022. doi: 10.1016/j.est.2022.104663

[18] J. P. Halawa, A. Hermawan, and J. Junaedi, "Implementation of linear regression algorithm to predict stock prices based on historical data," *bit-Tech*, vol. 5, no. 2, pp. 103–112, 2022. doi: 10.32877/bt.v5i2.616

[19] P. Schober and L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesth. Analg.*, vol. 126, no. 5, pp. 1763–1768, 2018. doi: 10.1213/ANE.0000000000002864.

[20] P. Surya and I. L. Aroquiaraj, "Crop yield prediction in agriculture using data mining predictive analytic techniques," *Int. J. Res. Anal. Rev.*, vol. 5, no. 4, pp. 783–787, 2018.

[21] M. Ngabire, T. Wang, X. Xue, J. Liao, G. Sahbeni, C. Huang, H. Duan, and X. Song, "Soil salinization mapping across different sandy land-cover types in the Shiyang River Basin: A remote sensing and multiple linear regression approach," *Remote Sens. Appl. Soc. Environ.*, vol. 28, no. June, p. 100847, 2022. doi: 10.1016/j.rsase.2022.100847

[22] P. Hunasigi, S. Jedhe, M. Mane, and V. Patil-Shinde, "Multilayer perceptron neural network based models for prediction of the rainfall and reference crop evapotranspiration for sub-humid climate of Dapoli,

Ratnagiri District, India," *Acta Ecol. Sin.*, vol. 43, no. 1, pp. 154–201, 2023. doi: 10.1016/j.chnaes.2022.09.004

[23] R. Latifah, E. S. Wulandari, and P. E. Kreshna, "Model Decision Tree Untuk Prediksi Jadwal Kerja Menggunakan Scikit-Learn," *J. Univ. Muhammadiyah Jakarta*, pp. 1–6, 2019. [Online]. Available: https://jurnal.umj.ac.id/index.php/semnastek/article/download/5239/3517

[24] A. Rachmat, "Survei Penerapan Model Machine Learning Dalam Bidang Keamanan Informasi," *J. Sist. Cerdas*, vol. 2, no. 1, pp. 47–60, 2019. doi: 10.37396/jsc.v2i1.20

[25] M. Parzinger, L. Hanfstaengl, F. Sigg, U. Spindler, U. Wellisch, and M. Wirnsberger, "Comparison of different training data sets from simulation and experimental measurement with artificial users for occupancy detection — Using machine learning methods Random Forest and LASSO," *Build. Environ.*, vol. 223, no. June, p. 109313, 2022. doi: 10.1016/j.buildenv.2022.109313

[26] A. Saravanan, S. Parida, M. Murugan, M. S. Reddy, P. V. Elumalai, and S. K. Dash, "Thermal performance prediction of a solar air heater with a C-shape finned absorber plate using RF, LR and KNN models of Machine learning," *Therm. Sci. Eng. Prog.*, vol. 38, no. December 2022, p. 101630, 2023. doi: 10.1016/j.tsep.2022.101630

[27] A. S. B. Karno, "Prediksi Data Time Series Saham Bank BRI Dengan Mesin Belajar LSTM (Long ShortTerm Memory)," *J. Inform. Inf. Secur.*, vol. 1, no. 1, pp. 1–8, 2020. doi: 10.31599/jiforty.v1i1.133

[28] B. Rudianto, "Analisis Pengaruh Sebaran Ground Control Point Terhadap Ketelitian Objek Pada Peta Citra Hasil Ortorektifikasi," *J. Itenas Rekayasa*, vol. 15, no. 1, p. 218798, 2011.

DOI:10.4186/ej.2024.28.9.1

Farah Apit Tantri is a student at the Department of Physics Education, Faculty of Teacher Training and Education, Universitas Sebelas Maret, Indonesia. Her current research is on applying machine learning to batteries and education. Research experience that has been carried out is making puzzle learning media using augmented reality with funding from the Ministry of Education, Culture, Research, and Technology.

Dewanto Harjunowibowo, Ph.D. is an assistant professor in the Department of Physics Education, Faculty of Teacher Training and Education, Universitas Sebelas Maret, Indonesia. He graduated from the Sustainable Energy Technology Program at the University of Nottingham, United Kingdom. Currently, he is pursuing research on heat transfer for energy conservation in buildings. He has also published works on artificial intelligence and automatic systems based on IoT.

Endah Retno Dyartanti, photograph and biography not available at the time of publication.

Muhammad Nizam, photograph and biography not available at the time of publication.

Mufti Reza Aulia Putra, photograph and biography not available at the time of publication.

Muhammad Nizam, photograph and biography not available at the time of publication.

Rekyan Regasari MP, photograph and biography not available at the time of publication.

Lim Tiong Hoo is an Assistant Professor at the Engineering Faculty of Universiti Brunei Darussalam. He is also a Director of the planning development office. His research interest is on Artificial Intelligence and Smart Systems.

Anif Jamaluddin is a lecture of Universitas Sebelas Maret. He is a researcher at the center of excellent energy storage. He holds a Ph.D. in Energy Engineering from National Central University, Taiwan, specializing in energy storage materials. His research areas include graphene, carbon-based materials for energy storage, and materials optimation and characterization. He can be contacted at email: elhanif@staff.uns.ac.id