Forecasting Battery Capacity and Feasibility Using the Gaussian Process Regression Method

Lardiansyah Adhwa Nugraha[1*], Tri Wrahatnolo[2]
[1,2] Electrical Engineering Department, Univerisitas Negeri Surabaya
A5 Building Ketintang Campus, Surabaya 60231, Indonesia
[1*]lardiansyah.20005@mhs.unesa.ac.id
[2] triwrahatnolo@unesa.ac.id

Abstract – The 110VDC batteries at the 150kV South Surabaya Substation have a shortage in the number of units. Therefore, they require extra supervision to ensure that protection and control equipment relying on DC power sources can operate normally during rectifier system outages, preventing potentially severe disruptions at the substation. The objective of this study is to use Matlab's forecasting degradation method for battery performance using Regression Learner, aimed at facilitating operators at the 150kV South Surabaya Substation. The research focuses on forecasting battery performance degradation using Gaussian Process Regression (GPR) with datasets obtained from observed discharging and charging tests compiled in Excel format. Data analysis techniques involve building a GPR model using Matlab software and comparing forecasted results with discharging test data over two years from PT. PLN (Persero). The study concludes that a 71% battery efficiency qualifies as sufficiently reliable backup power during AC or rectifier disruptions. This ensures continuous operation of protection and control equipment during blackouts, thereby preventing operational disruptions and serious safety issues.

Keywords: Gaussian Procces Regression, Battery, Forecasting

I. INTRODUCTION

The 150kV Surabaya Selatan Substation, also known as the 150kV Wonorejo Substation, is a conventional substation playing a crucial role in meeting the daily electricity needs of customers in the eastern part of Surabaya. Within the substation, there are two types of power sources: alternating current (AC) and direct current (DC). The DC source is used to operate protection and control equipment, both under normal and emergency conditions. This DC system is vital as it supplies power to protection relays, control systems, and other equipment requiring a DC source, such as batteries. Reliable batteries that can supply DC power to equipment in emergencies must adhere to the standards set by PT. PLN (Persero) [1], which stipulate that a battery is considered efficient if it has an efficiency of >80% and is deemed poor if the efficiency is <60%. If the efficiency ranges between 60% and 80%, the battery is considered adequate.

However, the use of batteries as a DC system in substations often encounters several issues. If the DC supply system in the substation fails, it can lead to a failure of the protection system within the substation. One of the impacts of the decreased capacity and efficiency of the batteries in the substation is the failure of the battery system function, or even serious safety issues, which can reduce the battery's lifetime [2]. Furthermore, if the battery's energy efficiency is compromised due to a decrease caused by increased internal resistance, and when the efficiency drops to an unacceptable level, a warning signal should be sent to the system to replace the battery [3]. As in the issues faced by the 150kV Surabaya Selatan Substation, where to serve as a DC source, it must first go through a rectifier with the source coming from the Auxiliary Transformer (PS). The PS Transformer at the 150kV Surabaya Selatan Substation has two units without a DC distribution panel. Consequently, the coupling that functions as a direct connection to the load can become a serious problem. If there is a power outage in the Auxiliary Transformer (PS) system, the Rectifier, which depends on the PS transformer, will also go down. In this situation, the battery will protect the equipment that uses the DC source. Additionally, the battery at the 150kV Surabaya Selatan Substation only has 1 unit, whereas the standard set by PT. PLN (Persero) is 2 units according to the SK DIR 0520 manual [1].

Maintenance of the batteries has been carried out by PT. PLN (Persero) by conducting routine testing every two years using discharging and charging test models to assess the battery's capacity and efficiency. The maintenance method applied is less effective because the battery system is affected by rapid response time when absorbing and releasing power, but all batteries age over time and with the number of cycles they have undergone. Batteries will lose power over time, thus limiting their use for long-term storage [4].

Periodic testing as a maintenance method is indeed important, but more sophisticated and effective solutions are still needed. Advances in computer technology can leverage machine learning methods to help predict battery capacity degradation without the need for periodic testing. The use of machine learning methods to predict battery data has been implemented previously, with methods such as Support Vector Machine (SVM), Long-Short Term Memory (LSTM), and Artificial Neural Network (ANN) [5]-[6].

Furthermore, based on research by Piyus Tagede et al. [7], it is stated that the GPR method exhibits a far superior representation ability and forecasting accuracy using Gaussian Process Regression (GPR) compared to parametric algorithms like Artificial Neural Network (ANN) and Support Vector Machine (SVM). Additionally, previous research by Zhengyu Liu et al. [8] utilized the Support Vector Regression (SVR) method for forecasting self-discharge in lithium-ion batteries. However, the SVR method requires more in-depth hyperparameter tuning, such as kernel selection and regularization parameter settings, which can be complicated and time-consuming [9]. On the other hand, GPR employs adjustable kernel functions to handle various types of data and patterns, providing greater flexibility in modeling [10]. While SVR also uses kernel functions, GPR's ability to handle complex patterns and nonlinear relationships more effectively often makes it superior in the context of forecasting [11]. Therefore, this study proposes the use of the Gaussian Process Regression (GPR) method for forecasting battery capacity degradation. Accurate forecasting of capacity degradation is a crucial function of battery management systems [12]. Several parameters used for forecasting capacity degradation include voltage, current, and temperature. The Gaussian Process Regression (GPR) method is chosen because it does not require a specific functional form to map inputs to outputs. The performance evaluation of the GPR model in this study will be analyzed using RMSE and MAPE calculations. For evaluating the feasibility of the battery at the 150kV Surabaya Selatan substation, battery efficiency will be calculated based on the predicted (forecasted) test data.

II. LITERATURE

150kV South Surabaya Substation

Substations are critical components of power generation that play a crucial role in distributing electrical energy. This substation serves as a vital asset for the power transmission system, utilities, and operators, ultimately ensuring a reliable electricity supply to consumers [13]. There are two main types of substations: conventional substations and Gas Insulated Switchgear (GIS). Conventional substations use hardwired systems for connecting and disconnecting electrical networks, with most components located outside the building [14]-[15]. GIS (Gas Insulated Switchgear), on the other hand, utilizes pressurized sulfur hexafluoride gas as both the electrical insulation and arc extinguishing medium [16].

110 VDC Battery

The battery used in the substation is of Nickel Cadmium (NiCd) type. Nickel Cadmium batteries belong to the nickel battery family, which also includes nickel-metal hydride, nickel-iron, and nickel-zinc batteries [17]. One commonly used electrochemical energy storage technology is rechargeable batteries, which convert electrochemical energy into electrical energy and are frequently used as portable power sources in various applications [18]. When the battery is used, a chemical reaction occurs that produces electricity. Conversely, during the charging process, electricity is converted into chemical energy [19]. The 110VDC battery used in the 150kV South Surabaya Substation can be seen in Figure 1, and the diagram of the 110 VDC GI South Surabaya is presented in Figure 2.

Figure 1. 110VDC Battery at South Surabaya Substation

Figure 2. Diagram of the 110 VDC GI South Surabaya

III. METHODS

Dataset

The research dataset comprises discharging and charging data from the years 2019 and 2021, encompassing voltage, current, temperature, and capacity measurements. These data were collected biennially following the Standard Operating Procedure (SOP) of the 150kV South Surabaya Substation. The training data consists of discharging and charging tests from 2019 and 2021 stored in Excel files, detailing voltage, current, temperature, and capacity. These datasets were used to train a Gaussian Process Regression (GPR) model within the Regression Learner, assessing the Root Mean Square Error (RMSE) of the training data. Meanwhile, testing data from the year 2023, including voltage, current, and temperature, were used for forecasting purposes. Table 1 presents the battery specifications used in the biennial discharging tests, and Figure 3 illustrates the data collection at the South Surabaya Substation.

Table 1. Battery Specification

Spesifikasi baterai	
Merk	SAFT NITE
Tipe	SCM 211
Jenis baterai	Alkali Nickel Cadmium
Kapasitas	211
Tegangan baterai	110VDC
Tegangan setiap sel	1,2 V
Jumlah sel baterai	86 sel
Tahun operasi	2009

Figure 3. Data Collection at the South Surabaya Substation.

Forecasting

1. Start
   Open MATLAB Software, then utilize the Regression Learner feature.
2. Set Data
   The use of data for forecasting is divided into two parts: training data and testing data. The training data includes parameters such as voltage, current, temperature, and capacity, while the testing data consists only of voltage, current, and temperature. The training data is derived from charging and discharging tests conducted in 2019 and 2021, stored in Excel file format. These datasets were employed to train a Gaussian Process Regression (GPR) model using the Regression Learner tool. Meanwhile, the testing data is obtained from the results of charging and discharging tests conducted in 2023. This data also includes voltage, current, and temperature, and is used to generate forecasted values. The GPR model trained with the training data is then utilized to predict outcomes based on the testing data.
3. Training Model GPR
   The GPR model will be trained using training data derived from charging and discharging tests conducted in 2019 and 2021, with parameters such as voltage, current, and temperature as predictors. The training results will produce a visualization graph displaying actual and predicted points, along with RMSE, MAE, and MSE values.
4. Model Evaluation
   Evaluation of the GPR model is conducted by selecting the kernel function that yields the smallest RMSE value. If the evaluation shows a significantly high RMSE value, it could indicate inaccuracies in the input data. Therefore, a thorough recheck of both the training and testing datasets is necessary.
5. Forecasting
   Forecasting is performed by exporting the model to the MATLAB command window after training the Gaussian Process Regression (GPR) model using the training data. Subsequently, input the following code into the MATLAB command window:
   >> T=readtable('datatesting2023.xlsx');
   >> yfit = trainedModel.predictFcn(T)
   yfit = (capacity result)

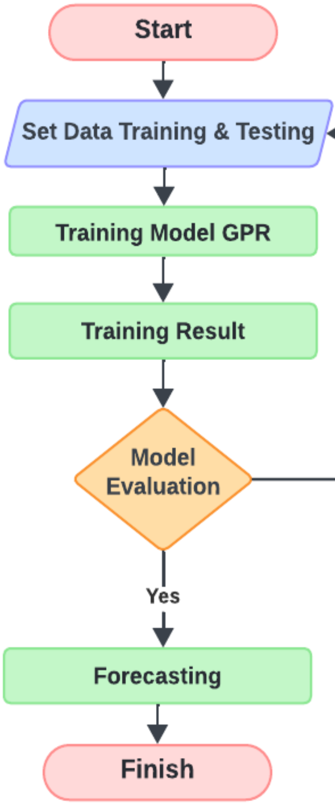


Figure 4. Forecasting Steps

Data Analysis Techniques

Data analysis involves building a Gaussian Process Regression (GPR) model using MATLAB software. Subsequently, the forecasted results obtained from the GPR model are compared with the biennial discharging test data from PT. PLN (Persero). Next, the accuracy performance of the model is analyzed using the Mean Absolute Percentage Error (MAPE) based on the forecasted results. The formula for MAPE is presented in the equation below.

$$\boldsymbol{MAPE = \frac{\sum \frac{|A-F|}{A} \times 100}{N}} \quad (1)$$

Desc:
N: Number of data points
A: Actual value (Ah)
F: Forecasted or predicted value (Ah)

Regression Learner MATLAB Features

The GPR model training is performed using the training data imported upon entering the Regression Learner feature.

The data import process involves selecting parameters or predictors for the response variable, which is capacity. Afterward, click "Start Session" to choose the model and train it using the specified training data parameters. This process is illustrated in Figure 5 for parameter selection and Figure 6 for kernel function selection.

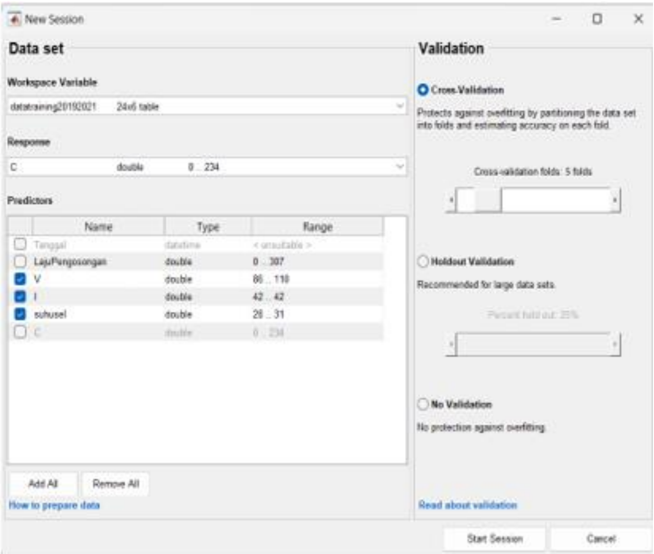


Figure 5. Parameter Selection

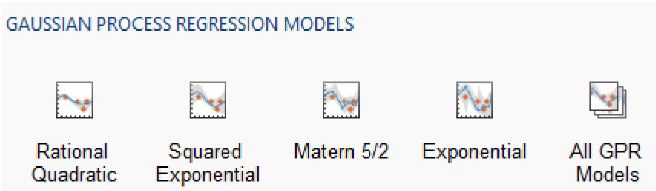


Figure 6. Kernel Function

IV.RESULTS AND DISCUSSION

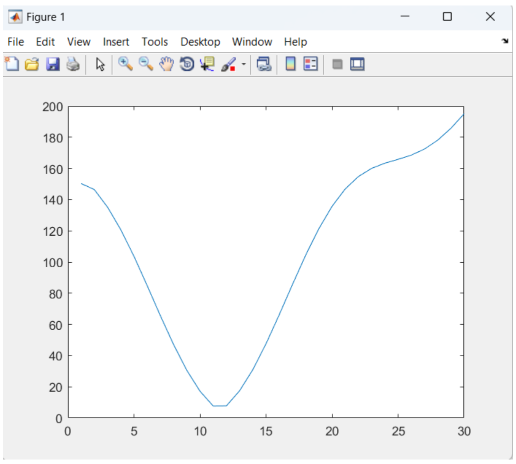
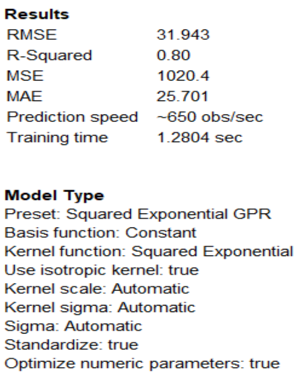


Figure 7. Testing Data Graph



Results

RMSE	31.943
R-Squared	0.80
MSE	1020.4
MAE	25.701
Prediction speed	~650 obs/sec
Training time	1.2804 sec

Model Type
Preset: Squared Exponential GPR
Basis function: Constant
Kernel function: Squared Exponential
Use isotropic kernel: true
Kernel scale: Automatic
Kernel sigma: Automatic
Sigma: Automatic
Standardize: true
Optimize numeric parameters: true

Figure 8. Error rate Training Data

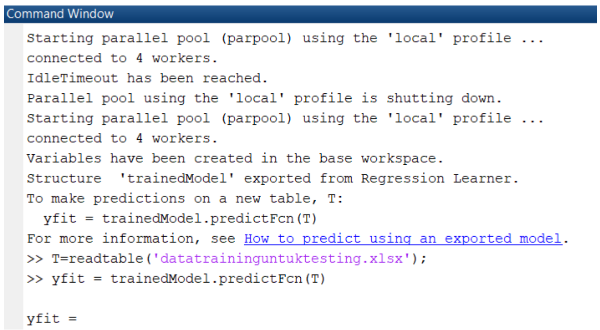
Based on the training results above, the smallest RMSE value was obtained using the Squared Exponential GPR kernel function, which was 31.943, with an MSE of 1020.4 and an MAE of 25.701. According to [20], accuracy indicators such as MAPE, RMSE, and MAE are used to evaluate the accuracy of prediction results, where yi and y^i represent the actual data and the estimated battery capacity, respectively. The closer these indicator values are to zero, the more accurate the prediction results. From the obtained RMSE and MAE values, it is evident that the model's accuracy performance is poor. RMSE measures the average prediction error made by the model on the training data [21].

If the model demonstrates good performance on the training data, as indicated by a low RMSE value, it is expected to provide accurate predictions on the testing data, and vice versa. Therefore, the next step is to perform forecasting to determine the battery capacity on the testing data for the year 2023.

The initial step for performing forecasting is to export the model and then enter the Matlab Command window by writing the following source code:

```
>> T=readtable('datatesting2023.xlsx');
>> yfit = trainedModel.predictFcn(T)
```

yfit = (capacity result)



Command Window

```
Starting parallel pool (parpool) using the 'local' profile ...
connected to 4 workers.
IdleTimeout has been reached.
Parallel pool using the 'local' profile is shutting down.
Starting parallel pool (parpool) using the 'local' profile ...
connected to 4 workers.
Variables have been created in the base workspace.
Structure  'trainedModel' exported from Regression Learner.
To make predictions on a new table, T:
  yfit = trainedModel.predictFcn(T)
For more information, see How to predict using an exported model.
>> T=readtable('datatraininguntuktesting.xlsx');
>> yfit = trainedModel.predictFcn(T)

yfit =
```

Figure 9. Source Code Forecasting

Next, a comparison between the actual data and the predicted (forecasted) data will be conducted by writing the following source code

```
>> actual_data = ...; % Your actual data
predicted_data = ...; % Your predicted data

rmse = sqrt(mean((actual_data - predicted_data).^2));

disp(['RMSE: ', num2str(rmse)]);
```

Figure 10. Comparison of Actual Data and Forecasting Data Source Code

```
>> actual_data = [199, 189, 169, 149, 128, 105, 84, 61, 43, 21, 0, 0, 11, 23, 35, 46, 54, 67, 75, 90, 102, 113, 120, 136, 145, 155, 168, 175, 190, 199] ;
>> predicted_data = [150.3151, 146.4439, 135.1778, 120.7555, 103.6763, 84.8847, 65.5617, 47.0720, 30.5689, 17.1710, 7.6568, 7.7625, 17.4513, 30.9676, 47.5791, 66.1822, 85.5518, 104.3311, 121.3463, 135.7005, 146.8820, 154.8395, 160.0631, 163.3662, 165.8310, 168.5466, 172.4417, 178.1235, 185.7442, 195.0334] ;
>>rmse = sqrt(mean((actual_data - predicted_data).^2));
        disp(['RMSE: ', num2str(rmse)]);
RMSE: 27.103
```

Figure 11.Comparison of Actual Data and Forecasting Data Source Code

From the RMSE results above, it is shown that the RMSE value from the comparison between the actual data and the forecasting results using the 2023 testing data is 27.103. According to the study by Galal Uddin et al. [22], an RMSE value greater than 20% falls into the poor category.

In addition to using RMSE, the model's performance can also be evaluated based on the MAPE analysis results [23]. Therefore, to validate the accuracy performance of the GPR model on the 2023 testing data, MAPE calculations were performed using the formula from Equation 1. Based on the MAPE calculations, the accuracy performance of the Gaussian model on the 2023 discharging and charging test data is 70.8%. According to research [20], an MAPE category ≥ 50% indicates poor performance.

Next, to assess battery feasibility, battery efficiency will be calculated using the following equation:

$$\eta baterai = \frac{Cd}{Cc} \times 100\% \qquad (2)$$

$$\eta baterai = \frac{150{,}3}{211} \times 100\%$$

$$\eta baterai = 71\%$$

Explantion :
Cd : Predicted battery capacity (Ah)
Cc : Original factory battery capacity (Ah)

According to the standards outlined in SKDIR 0520-2.K/DIR/2014, a guideline for AC DC supply maintenance by PT. PLN (Persero) [1], batteries are classified based on their efficiency. Batteries with efficiencies above 80% are deemed good, those between 60% and 80% are considered adequate, and those below 60% are categorized as poor. Hence, with a calculated battery efficiency above 71%, it indicates that the battery remains sufficiently operational. In the event of anomalies such as blackouts, the battery would still be in a condition considered adequate

V. CONCLUSION

The performance of the Gaussian Process Regression (GPR) model on both training data and the comparison between actual and forecasted data, based on RMSE values of 31.943 and 37.103 respectively, indicates inadequate performance. Similarly, the MAPE calculation between actual and forecasted data shows poor performance at 70.8%. However, the battery efficiency of 71% is considered adequate for operation, falling within the 60%-80% range defined by the standards in SKDIR 0520-2.K/DIR/2014 for AC DC supply maintenance by PT. PLN (Persero) [1].
The implementation of the GPR model for forecasting in this study has been deemed insufficient due to limitations in the available data. Further experimentation with alternative models better suited to the dataset is necessary.

Acknowledgment

Thank you to Prof. Dr. Tri Wrahatnolo, M.Pd., M.T., as the supervisor who has provided valuable guidance and input throughout this research.

REFERENCES

[1] P. PLN, "Buku Sistem Suplai AC DC Final," *19.Buku Sist. Suplai AC DC Final*, vol. 44, no. 8, pp. 1689–1699, 2011.

[2] X. Wu, A. J. Conejo, and S. Mathew, "Optimal Siting of Batteries in Distribution Systems to Enhance Reliability," *IEEE Trans. Power Deliv.*, vol. 36, no. 5, pp. 3118–3127, 2021,

[3] M. Wei, P. Balaya, M. Ye, and Z. Song, "Remaining useful life prediction for 18650 sodium-ion batteries based on incremental capacity analysis," *Energy*, vol. 261, no. PA, p. 125151, 2022,

[4] J. Liu *et al.*, "The TWh challenge: Next generation batteries for energy storage and electric vehicles," *Next Energy*, vol. 1, no. 1, p. 100015, 2023

[5] X. Feng *et al.*, "Online State-of-Health Estimation for Li-Ion Battery Using Partial Charging Segment Based on Support Vector Machine," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8583–8592, 2019,

[6] Y. Yang, "A machine-learning prediction method of lithium-ion battery life based on charge process for different applications," *Appl. Energy*, vol. 292, no. April, p. 116897, 2021,

[7] P. Tagade *et al.*, "Deep Gaussian process regression for lithium-ion battery health prognosis and degradation mode diagnosis," *J. Power Sources*, vol. 445, no. March 2019, p. 227281, 2020,

[8] Z. Liu, H. He, J. Xie, K. Wang, and W. Huang, "Self-discharge prediction method for lithium-ion batteries based on improved support vector machine," *J. Energy Storage*, vol. 55, no. PB, p. 105571, 2022,

[9] J. Y. Hsia and C. J. Lin, "Parameter Selection for Linear Support Vector Regression," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 31, no. 12, pp. 5639–5644, 2020,

[10] H. Fathipour-Azar, "Machine learning-assisted distinct element model calibration: ANFIS, SVM, GPR, and MARS approaches," *Acta Geotech.*, vol. 17, no. 4, pp. 1207–1217, 2022,

[11] P. Das, D. A. Sachindra, and K. Chanda, "Machine Learning-Based Rainfall Forecasting with Multiple Non-Linear Feature Selection

Algorithms," *Water Resour. Manag.*, vol. 36, no. 15, pp. 6043–6071, 2022

[12] R. R. Richardson, C. R. Birkl, M. A. Osborne, and D. A. Howey, "Gaussian Process Regression for in Situ Capacity Estimation of Lithium-Ion Batteries," *IEEE Trans. Ind. Informatics*, vol. 15, no. 1, pp. 127–138, 2019,

[13] G. R. Santos, E. Zancul, G. Manassero, and M. Spinola, "From conventional to smart substations: A classification model," *Electr. Power Syst. Res.*, vol. 226, no. September 2023, p. 109887, 2024,

[14] M. Dsouza, R. S. Dhara, and R. C. Bouyer, "Modularization of High Voltage Gas Insulated Substations," *IEEE Trans. Ind. Appl.*, vol. 56, no. 5, pp. 4662–4669, 2020,

[15] S. Sabihuddin, "Implementation of a Microgrid Substation for Automatic Frequency Response Implementation of a Microgrid Substation for Automatic Frequency Response," no. February 2020, 2024,

[16] A. P. Purnomoadi, A. Rodrigo Mor, and J. J. Smit, "Spacer flashover in Gas Insulated Switchgear (GIS) with humid SF6 under different electrical stresses," *Int. J. Electr. Power Energy Syst.*, vol. 116, no. June 2019, 2020,

[17] Y. Bai *et al.*, "Reversible and irreversible heat generation of NCA/Si–C pouch cell during electrochemical energy-storage process," *J. Energy Chem.*, vol. 29, pp. 95–102, 2019,.

[18] S. A. Hasib *et al.*, "A Comprehensive Review of Available Battery Datasets, RUL Prediction Approaches, and Advanced Battery Management," *IEEE Access*, vol. 9, pp. 86166–86193, 2021,

[19] A. Turksoy, A. Teke, and A. Alkaya, "A comprehensive overview of the dc-dc converter-based battery charge balancing methods in electric vehicles," *Renew. Sustain. Energy Rev.*, vol. 133, no. October 2019, p. 110274, 2020,

[20] F. Cui, Z. Li, Y. Zhang, K. Wu, and Y. Shi, "A Hybrid Approach for Predicting Future Capacities of Lithium-ion Battery Based on N-BEATS Model," *Proc. - 2023 China Autom. Congr. CAC 2023*, pp. 5294–5297, 2023,

[21] S. Ameer *et al.*, "Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities," *IEEE Access*, vol. 7, pp. 128325–128338, 2019,

[22] M. G. Uddin, S. Nash, M. T. Mahammad Diganta, A. Rahman, and A. I. Olbert, "Robust machine learning algorithms for predicting coastal water quality index," *J. Environ. Manage.*, vol. 321, no. June, p. 115923, 2022,

[23] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, 2021,