Predicting Capacity Change in Li-ion Batteries using Regression Models
Regresyon Modelleri ile Li-ion Bataryalarında Kapasite Değişimi Tahmini

Murat Surucu[1,*]
[1]Republic of Turkey Ministry of National Education, Ankara, Turkey
ORCIDs: 0000-0002-9889-9952
E-mails: msurucu@gmail.com
*Corresponding author.

***Abstract*—The lithium-ion battery technology has led to significant changes in the usage of rechargeable batteries due to its low discharge current, high energy capacity, and long charge/discharge cycles. The easy production of portable and high-energy density batteries has not only contributed to the proliferation of smart devices like the Internet of Things (IoT) devices but has also led to an increase in the use of electric vehicles (EVs). As battery chemistry varies based on manufacturers and storage conditions, the importance of determining the charge lifespan and capacity of batteries connected to smart devices is growing progressively. Therefore, various studies are being conducted to assess capacity and lifespan calculations for Li-Ion batteries. In this study, the behavioral patterns of Li-Ion cells in end-user products are analyzed, aiming to predict capacities for similar battery groups. For this purpose, besides a fundamental linear regression analysis, regression analysis using Gaussian Process Regression (GPR) and Convolutional Neural Networks (CNN) is carried out. The regression performance is evaluated using diverse metric criteria such as R-squared (R^2), Adjusted R-squared (Adj. R^2), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Normalized Mean Squared Error (NMSE).**

***Keywords*—Lithium-ion battery; capacity; regression; GPR; CNN**

***Özetçe*—Lityum-iyon pil teknolojisi, düşük deşarj akımı, yoğun enerji kapasitesi ve uzun şarj/deşarj döngüsü sebebiyle şarj edilebilir pillerin kullanım alanlarında büyük değişimler yaşanmasına sebep olmuştur. Taşınabilir ve enerji güvenliği yüksek bataryaların kolayca üretilmesi ile IoT gibi akıllı aygıtların yaygınlaşmasının yanısıra elektrikli araçların(EV) kullanılmasında da artış meydana gelmiştir. Pil kimyasının üretici temelli ve saklama koşulları ile değişmesi sebebiyle akıllı cihazlara bağlı pillerin şarj omrünün ve kapasitesinin tespitinin onemi gitgide artmaktadır. Bu sebeple Li-Ion pillerde de kapasite ve omür hesabında çeşitli çalışmalar yapılmaktadır. Bu çalışmada ozellikle Li-Ion hücrelerin son kullanıcı ürünlerindeki davranış modelleri incelenerek benzer batarya grupları için kapasite tahmini yapılması amaçlanmıştır. Bu sebeple temel bir doğrusal regresyon analizinin yanı sıra Gaussian Process Regression (GPR) ve Convolutional Neural Networks (CNN) ile regresyon analizi gerçekleştirilmiştir. Çeşitli metrik olçütlerle, (R-squared (R^2), Adjusted R-squared (Adj. R^2), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Normalized Mean Squared Error (NMSE)), regresyon performansı incelenmiştir.**

***Anahtar Kelimeler*—Lityum-iyon batarya; kapasite; regresyon; GPR; CNN**

I. Introduction

Lithium-ion batteries, owing to their high energy density and compatibility with rapid charging methods, have revolutionized the landscape of rechargeable batteries. They find extensive applications across a wide spectrum, ranging from unmanned aerial vehicles (UAVs), cell phones, and electric vehicles to everyday consumer products like toothbrushes and electric shavers. Additionally, they are employed in smart home sensors and Internet of Things (IoT) devices. For projects with energy-sensitive requirements such as IoT devices, cell phones, and UAVs, there has been significant research on predicting the capacity degradation over time, estimating instantaneous current outputs, and predicting voltage behavior.

The ability to forecast these parameters enables the determination of network connection timing for service and charging intervals of mobile devices. Battery management is notably crucial in unmanned aerial vehicles and IoT devices, as mismanaging batteries can lead to severe consequences, including complete device failure. Accordingly, the literature is replete with studies on this matter. For instance, Li et al. examined battery depletion patterns in UAVs engaged in rescue missions under extreme weather conditions [1]. Depcik et al. highlighted the advantages of Li-ion batteries in UAV flight duration compared to alternative energy storage methods [2]. Ma et al. investigated optimizations necessary for using Li-ion batteries in UAVs [3]. Shahjalal et al. delved into heat management issues during charge and discharge in Li-ion batteries [4]. Kumar et al. worked on battery life prediction for IoT devices [5], while Hemavathi focused on predicting battery life using internal impedance [6]. Likewise, Shah et al. concentrated on estimating the lifespan of EV batteries [7]. Li et al. employed GPR and later Support Vector Regression for capacity analysis in Li-ion batteries [8], [9].

In predicting the State of Health (SOH) of batteries, several electrically measurable parameters are considered. Battery internal resistance measurements have been used to estimate battery capacity directly [10]–[13]. Similarly, IC/DV capacity growth-voltage change analysis is also employed for battery life prediction [14]–[17]. Another method, the Differential Temperature Voltage (DTV) analysis, utilizes potential difference and surface temperature change for SOH estimation [18]–[20].

In this study, the NASA Prognostics Center of Excellence (PCoE) battery discharge dataset was employed [21].

Looking at similar studies conducted with comparable datasets, analyses of SOH have been carried out using methods like the Autoregressive process [22], a combination of Gaussian Process Regression (GPR) and Fuzzy Logic [23], Capacity estimation through Lumped Parameter Modeling [24], Capacity prediction using Polynomial Fitting [25], Capacity estimation employing Nonlinear AR model [26], Prediction through Kalman filtering [27], Battery life prediction using Support Vector Regression (SVR) [28], [29], and also Nonlinear Regression techniques [30].

For SOH analysis from battery discharge data, linear regression, GPR, and CNN-based regression analyses were utilized to predict the battery output voltage in terms of measured load current, load potential, and battery voltage's previous recorded value.

II. Materials & Methods

A. Dataset & Preprocess

In this study, the battery charge/discharge dataset provided by the NASA Prognostics Center of Excellence (PCoE) has been utilized [21]. The presented dataset encompasses temperature, terminal current, voltage, as well as circuit current and voltage (load or charge circuit) during the charge/discharge events of Li-ion batteries. Additionally, the dataset includes measurements necessary for internal resistance calculations during impedance measurements. A detailed representation is presented in the image below, and for comprehensive examination, the entire dataset can be accessed at https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/.

The data structure for battery B0005 is provided as an illustrative example in the image. Beneath the top-level field named "cycle," there are fields labeled "type," "ambient_temperature," "time," and "data." The "type" section contains subfields for "charge," "discharge," or "impedance." The "time" field accommodates the starting time of the "cycle" operation. The "data" section holds the recorded data. In this study, discharge patterns from batteries B0005, B0006, B0007, and B0018 have been employed.

For the utilized discharge data, only MinMax normalization has been applied as a preprocessing step. Through MinMax normalization, the measured quantities have been scaled to the range of 0 to 1, as indicated by the equation 1.

$$x_{\text{normalized}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{1}$$

Where:
- $x_{\text{normalized}}$ represents the normalized value of the measurement x,
- x_{min} is the minimum value of measurement x,
- x_{max} is the maximum value of measurement x.

This normalization process ensures that the measured values are transformed to a standardized range suitable for subsequent analysis and modeling. Three attributes have been chosen as features for the regression. These include the load current, voltage values, and the terminal potential value recorded in the previous measurement of the battery. The goal is to predict the instantaneous voltage value of the battery as the output.

B. Regression

Supervised learning can be fundamentally categorized into regression and classification problems. If we are building a learning model to predict discontinuous or labeled intervals, we are dealing with a classification problem [31], [32]. Similarly, when the values we are aiming to predict are continuous, this process is referred to as a regression problem. In the context of this study, focusing on battery capacity, the prediction task at hand is a regression process that needs to be addressed [33].

1) Linear Regression: Linear Regression analysis is the approach of investigating the relationship between multiple variables using linear methods. In its simplest form, the linear regression model can be defined as [34]:

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{2}$$

Here, y represents the dependent variable, while x is the independent variable. β_0 denotes the intercept, β_1 signifies the slope of the line, and ε indicates the error term.

2) Gaussian Process Regression: Gaussian Process Regression (GPR) is a non-parametric Bayesian regression method. Broadly, it is a modeling process performed to find a Gaussian-distributed kernel that maps our input feature vector to the output values. If we express it in the form of a standard linear model,

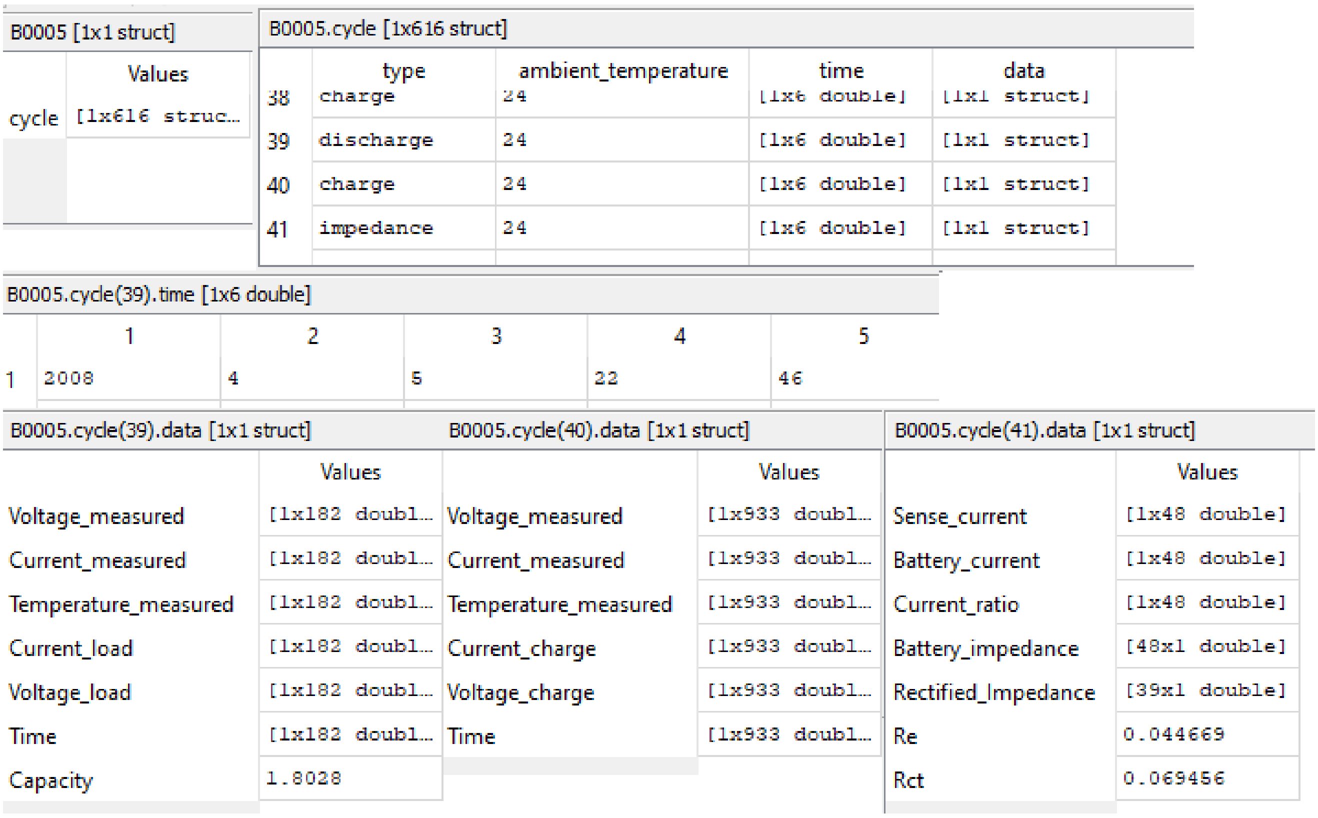
$$y = f(X) + \varepsilon \tag{3}$$

$$f(X) = X^T w \tag{4}$$

$$\varepsilon \sim \mathcal{N}(0, \sigma_n^2) \tag{5}$$

Where noise (ε) follows a Gaussian distribution with a mean of zero and a variance of σ_n^2 [33].

3) Convolutional Neural Network: The regression problem, as mentioned before, is the process of approximating a continuous output in terms of attributes through specific parameters and functions. On the other hand, CNN is a multi-layered perceptual system that includes at least one layer with a convolution operation and can be utilized for both classification and regression tasks [35], [36]. The cornerstone of CNN, the convolutional layer, is where the convolution operation takes place. These layers are typically responsible for determining data attributes. The convolutional layer allows focusing on crucial points rather than processing the entire dataset, enabling the creation of a more profound learning algorithm [37]. By leveraging the learned weight coefficients specific to the



B0005 [1x1 struct]

	Values
cycle	[1x616 struc...

B0005.cycle [1x616 struct]

	type	ambient_temperature	time	data
38	charge	24	[1x6 double]	[1x1 struct]
39	discharge	24	[1x6 double]	[1x1 struct]
40	charge	24	[1x6 double]	[1x1 struct]
41	impedance	24	[1x6 double]	[1x1 struct]

B0005.cycle(39).time [1x6 double]

	1	2	3	4	5
1	2008	4	5	22	46

B0005.cycle(39).data [1x1 struct]

	Values
Voltage_measured	[1x182 doubl...
Current_measured	[1x182 doubl...
Temperature_measured	[1x182 doubl...
Current_load	[1x182 doubl...
Voltage_load	[1x182 doubl...
Time	[1x182 doubl...
Capacity	1.8028

B0005.cycle(40).data [1x1 struct]

	Values
Voltage_measured	[1x933 doubl...
Current_measured	[1x933 doubl...
Temperature_measured	[1x933 doubl...
Current_charge	[1x933 doubl...
Voltage_charge	[1x933 doubl...
Time	[1x933 doubl...

B0005.cycle(41).data [1x1 struct]

	Values
Sense_current	[1x48 double]
Battery_current	[1x48 double]
Current_ratio	[1x48 double]
Battery_impedance	[48x1 double]
Rectified_Impedance	[39x1 double]
Re	0.044669
Rct	0.069456

Figure 1: Data structure for battery B0005

convolutional network, the input data undergoes convolution to compute the output. For instance, the convolution operation $Y = X * W$ is illustrated in Figure 2, and the calculation of the y_0 value from the outputs is presented in Equation 6. Similarly, other values within the Y matrix are computed by sliding the W matrix over the X matrix. Ultimately, Equation 7 is derived for the convolution operation.

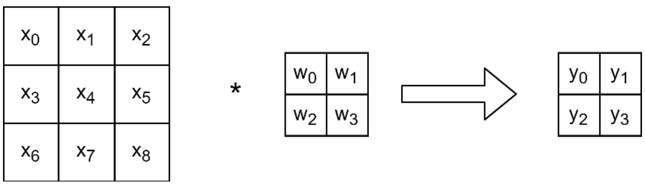


Figure 2: A convolution process example

$$y_0 = x_0 w_0 + x_1 w_1 + x_3 w_2 + x_4 w_3 \tag{6}$$

$$y(n) = x(n) \cdot w(n) = \sum_{k=0}^{n} w(k) x(n-k) \tag{7}$$

The convolutional layer can be utilized for classification or regression tasks based on the type of activation layer used in its output. In this study, a linear function has been employed.

C. Metrics & CV

In the performance measurement of regression analysis, the correlation coefficient R^2 is commonly employed [38]. Additionally, in the literature, metrics such as adjusted R^2 , Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Normalized Mean Squared Error (NMSE) are often selected as supplementary indicators alongside the correlation coefficient [39]–[41]. The formulas for the metrics used here are provided below:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{8}$$

$$\text{Adj. } R^2 = 1 - \frac{(1 - R^2) \cdot (n-1)}{n - p - 1} \tag{9}$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \tag{10}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{11}$$

$$\text{NMSE} = \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y})^2} \tag{12}$$

Where:

- n is the number of data points,
- p is the number of predictors,
- y_i represents the observed value,
- $\hat{y}_i$ represents the predicted value,
- $\bar{y}$ is the mean of observed values.

The preprocessed dataset, after undergoing normalization, has had 30% of its data set aside as test data before applying regression analysis. A 10-fold Cross-Validation (CV) process has been conducted on the remaining 70% of the data. Performance values have been calculated using the remaining 30% test data for metric measurements.

III. Results & Discussion

A. Results

In this study, three different regression analysis methods have been employed. These are linear regression, GP regression, and CNN regression methods. The generated models were trained and validated using 70% of the data reserved for training, employing the 10-fold cross-validation technique, repeated 10 times. Subsequently, these models were tested using the previously set aside 30% test data, and their performance was measured using five different metrics. As shown in Table 1, battery voltage exhibits a strong dependence on the selected features.

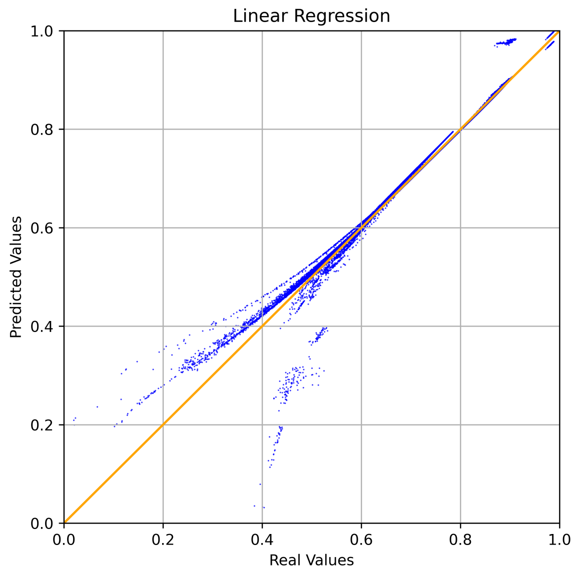


Figure 3: Linear regression

The response of the first model, the Linear Regression model, for the test data can be observed in Figure 3. Remarkably accurate predictions are obtained for battery voltage values of 50% or more of the maximum value. Similarly, for Figure 4, GPR yields results similar to the Linear Regression model, yet with superior performance. Finally, in the investigation involving the CNN model, the distribution depicted in

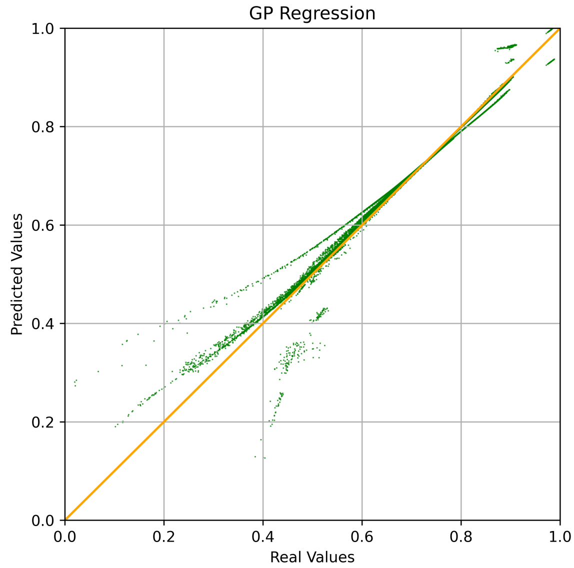


Figure 4: Gaussian process regression

Figure 5 is attained. Slightly more successful outcomes are achieved compared to the previous models. In Figure 6, the predicted battery voltages for the test values by all models are collectively plotted. It can be observed that the CNN model's performance is marginally better.

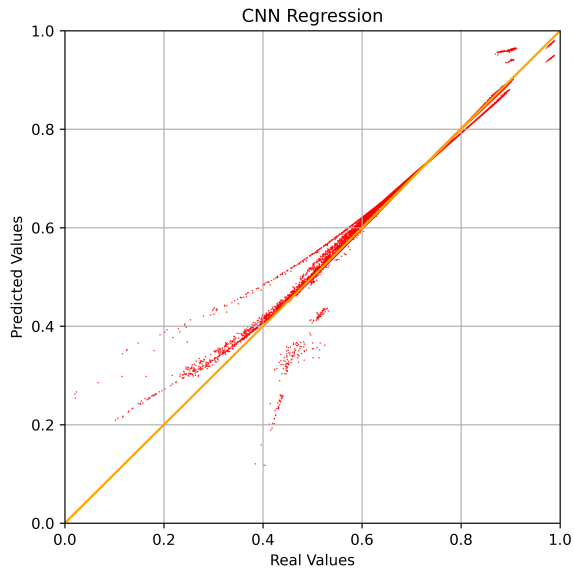


Figure 5: CNN regression

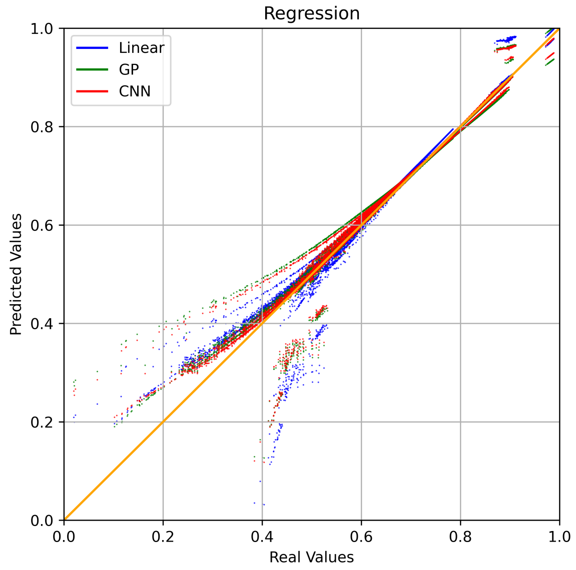


Figure 6: All regression results

Model	R^2	Adj.R^2	RMSE	NMSE	MAPE
Linear Reg.	0.980569	0.980568	0.013996	0.074926	0.008070
Gaussian Proc. Reg.	0.984854	0.984853	0.012357	0.097037	0.007583
CNN Reg.	0.986838	0.986837	0.011519	0.090538	0.006785

Table I: Model Performances

B. *Discussion*

An R^2 value of over 98% has been achieved for all three selected regression models. To attain even higher performance values, adjustments can be made to the models' hyperparameters, such as increasing the number of neurons in the CNN model. For future studies, it is considered beneficial to explore adjustments of hyperparameters specific to the models, as well as experimenting with other machine learning methods.

References

[1] Li N, Liu X, Yu B, Li L, Xu J, Tan Q. Study on the environmental adaptability of lithium-ion battery powered UAV under extreme temperature conditions. Energy 2021; 219: 119481.

[2] Depcik C, Cassady T, Collicott B, Burugupally SP, Li X, Alam SS, Arandia JR, Hobeck J. Comparison of lithium ion batteries, hydrogen fueled combustion engines, and a hydrogen fuel cell in powering a small unmanned aerial vehicle. Energy Conversion and Management 2020; 207: 112514.

[3] Ma Y, Chiang SW, Chu X, Li J, Gan L, Xu C, Yao Y, He Y, Li B, Kang F, Du H. Thermal design and optimization of lithium ion batteries for unmanned aerial vehicles. Energy Storage 2019; 1(1): e48.

[4] Shahjalal M, Shams T, Islam MdE, Alam W, Modak M, Hossain SB, Ramadesigan V, Ahmed MdR, Ahmed H, Iqbal A. A review of thermal management for Li-ion batteries: Prospects, challenges, and issues [Homepage on the Internet]. Journal of Energy Storage 2021; 39: 102518

[5] Kumar A, Hoque MA, Nurmi P, Pecht MG, Tarkoma S, Song J. Battery health estimation for IoT devices using V-Edge dynamics. Proceedings of the 21st International Workshop on Mobile Computing Systems and Applications 2020; 1: 56-61.

[6] Hemavathi S. Li-ion battery health estimation based on battery internal impedance measurement. Book chapter in Innovations in Sustainable Energy and Technology 2021; 183-193.

[7] Shah FA, Shahzad Sheikh S, Mir UI, Owais Athar S. Battery health monitoring for commercialized electric vehicle batteries: Lithium-ion. 2019 International Conference on Power Generation Systems and Renewable Energy Technologies (PGSRET) 2019; 1: 1-6.

[8] Li X, Yuan C, Li X, Wang Z. State of health estimation for Li-Ion battery using incremental capacity analysis and Gaussian process regression. Energy 2020; 190: 116467.

[9] Li X, Yuan C, Wang Z. State of health estimation for Li-ion battery via partial incremental capacity analysis based on support vector regression. Energy 2020; 203: 117852.

[10] Farmann A, Waag W, Marongiu A, Sauer DU. Critical review of on-board capacity estimation techniques for lithium-ion batteries in electric and hybrid electric vehicles. Journal of Power Sources 2015; 281: 114-130.

[11] Min Zhu, Wensong Hu, Kar NC. The SOH estimation of LiFePO4 battery based on internal resistance with Grey Markov Chain. IEEE Transportation Electrification Conference and Expo (ITEC) 2016; 1: 1-6.

[12] Balagopal B, Chow M-Y. The state of the art approaches to estimate the state of health (SOH) and state of function (SOF) of lithium Ion batteries. 13th International Conference on Industrial Informatics (INDIN) 2015; 1: 1302-1307.

[13] Chen ZP, Wang QT. The application of UKF algorithm for 18650-type lithium battery SOH estimation. Applied Mechanics and Materials 2014; 519-520: 1077-1082.

[14] Pastor-Fernandez C, Widanage WD, Chouchelamane GH, Marco J. A SoH diagnosis and prognosis method to identify and quantify degradation modes in li-ion batteries using the IC/DV technique. 6th Hybrid and Electric Vehicles Conference 2016; 1: 1-6.

[15] Yao H, Jia X, Zhao Q, Cheng Z-J, Guo B. Novel lithium-ion battery state-of-health estimation method using a genetic programming model. IEEE Access 2020; 8: 95333-95344.

[16] Pastor-Fernandez C, Uddin K, Chouchelamane GH, Widanage WD, Marco J. A comparison between electrochemical impedance spectroscopy and incremental capacity-differential voltage as li-ion diagnostic techniques to identify and quantify the effects of degradation modes within battery management systems. Journal of Power Sources 2017; 360: 301-318.

[17] Harting N, Wolff N, Roder F, Krewer U. State-of-health diagnosis of lithium-ion batteries using nonlinear frequency response analysis. Journal of The Electrochemical Society 2019; 166(2): A277-A285.

[18] Wang Z, Yuan C, Li X. Lithium battery state-of-health estimation via differential thermal voltammetry with Gaussian process regression. IEEE Transactions on Transportation Electrification 2021; 7(1): 16-25.

[19] Shibagaki T, Merla Y, Offer GJ. Tracking degradation in lithium iron phosphate batteries using differential thermal voltammetry. Journal of Power Sources 2018; 374: 188-195.

[20] Yang J, Cai Y, Mi C. A battery capacity estimation method using surface temperature change under constant-current charge scenario. IEEE Energy Conversion Congress and Exposition (ECCE) 2021; 1: 1687-1691.

[21] Saha B, Goebel K. Battery Data Set: NASA Prognostics Data Repository. NASA Ames Research Center, Moffett Field, CA, 2007.

[22] Long B, Xian W, Jiang L, Liu Z. An improved autoregressive model by particle swarm optimization for prognostics of lithium-ion batteries. Microelectronics Reliability 2013; 53(6): 821-831.

[23] Kang W, Xiao J, Xiao M, Hu Y, Zhu H, Li J. Research on remaining useful life prognostics based on fuzzy evaluation-Gaussian process regression method. IEEE Access 2020; 8: 71965-71973.

[24] Saha B, Goebel K. Modeling Li-ion battery capacity depletion in a particle filtering framework. In Proceedings of the Annual Conference of the Prognostics And Health Management Society 2021; 1: 2909-2924.

[25] Penna JAM, Nascimento CL, Rodrigues LR. Health monitoring and remaining useful life estimation of lithium-ion aeronautical batteries. IEEE Aerospace Conference 2012; 1: 1-12.

[26] Liu D, Luo Y, Liu J, Peng Y, Guo L, Pecht M. Lithium-ion battery remaining useful life estimation based on fusion nonlinear degradation AR model and RPF algorithm. Neural Computing and Applications 2013; 25(3-4): 557-572.

[27] Huang M, Zhang Q. Prediction of remaining useful life of lithium-ion battery based on UKF. 2020 Chinese Automation Congress (CAC) 2020; 1: 4502-4506.

[28] Shi Y, Yang Y, Wen J, Cui F, Wang J. Remaining useful life Prediction for lithium-ion battery based on CEEMDAN and SVR. IEEE 18th International Conference on Industrial Informatics (INDIN) 2020; 1: 888-893.

[29] Wang Z, Zeng S, Guo J, Qin T. Remaining capacity estimation of lithium-ion batteries based on the constant voltage charging profile. PLoS ONE 2018; 13(7): e0200169.

[30] Meng J, Ricco M, Luo G, Swierczynski M, Stroe DI, Stroe AI, Teodorescu R. An overview and comparison of online implementable soc estimation methods for lithium-ion battery. IEEE Transactions on Industry Applications 2018; 54(2): 1583-1591.

[31] Egeli SS, Isler Y. Determining the relation between the count number and x-ray energy level in pyroelectric materials using linear regression analysis. Journal of Intelligent Systems with Applications 2021; 4(1): 58-60.

[32] Dogdu B, Ertugrul O. Statistical relationship between strontium content and cooling rate on a356 alloy by using regression analysis. Journal of Intelligent Systems with Applications 2021; 4(1): 31-37.

[33] Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. Cambridge, Mass. Mit Press. 2006.

[34] Montgomery DC. Introduction To Linear Regression Analysis. S.L.: John Wiley, 2021.

[35] Ozdemir V, Caliskan A, Yigit A. Machine learning based electric energy consumption prediction of a large-scaled production plant with small-scaled data. Journal of Intelligent Systems with Applications 2020; 3(2): 84-89.

[36] Kızıltas Koc S, Yeniad M. Diabetes prediction using machine learning techniques. Journal of Intelligent Systems with Applications 2021; 4(2): 150-152.

[37] Habibi Aghdam H, Jahani Heravi E. Guide to Convolutional Neural Networks. Springer International Publishing, 2017.

[38] Spiess A-N, Neumeyer N. An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: A Monte Carlo approach. BMC Pharmacology and Toxicology 2010; 10(1): 6.

[39] Leach FJ, Henson RK. The use and impact of adjusted r 2 effects in published regression research. Multiple Linear Regression Viewpoints 2007; 33(1): 1-11.

[40] Botchkarev A. A new typology design of performance metrics to measure errors in machine learning regression algorithms. Interdisciplinary Journal of Information, Knowledge, and Management (IJIKM) 2019; 14: 045-076.

[41] Spuler M, Sarasola-Sanz A, Birbaumer N, Rosenstiel W, Ramos-Murguialday A. Comparing metrics to evaluate performance of regression methods for decoding of neural signals. 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2015: 1083-1086.