UNITED ARAB EMIRATES
MINISTRY OF ECONOMY

الإمارات العربية المتحدة
وزارة الاقتصاد

شهادة تسجيل مصنف

رقم (576-2020)

تشهد وزارة الاقتصاد بأن المصنف موضوع هذه الشهادة قد تم تسجيله بإدارة المصنفات الفكرية بتاريخ2020/05/18 وذلك وفقا للنسخة المسجلة لدى الإدارة

1) **إسم المصنف** : Lithium-Ion Battery Life Prediction Based on Initial Stage-Cycles Using Machine Learning

2) **نوع المصنف** : المصنفات المكتوبة

3) **وصف المصنف** : عبارة عن مادة مكتوبة تتضمن بحث باللغة الإنجليزية ، وذلك حسب النسخة المسجلة لدينا.

4) **إسم المؤلف** : 1- اديتيا محيط ساكسينا الجنسية: الهند
2- فيلاس هاريداس جايدانى الجنسية: الهند

5) **إسم صاحب حقوق المصنف** : نفس البيانات اعلاه

وقد سلمت الشهادة الى:

بصفته: وكيل عن مؤلف

ملاحظة :

- ☐ **لا يجوز تداول المصنف بناء على هذه الشهادة، حيث يتعين الحصول على إجازة تداول مصنف من الجهة المختصة بالدولة**
- ☐ **أي كشط أو تعديل في بنود هذه الشهادة أو فصل في مرفقاتها المختومة يلغيها**
- ☐ **لاتشمل الحماية الأفكار والإجراءات وأساليب العمل**

حررت في : 18/05/2020

مدير إدارة المصنفات الفكرية

Lithium-Ion Battery Life Prediction Based on Initial Stage-Cycles Using Machine Learning

Aditya Saxena*, *Birla Institute of Technology and Science, Pilani, Dubai Campus*
E-mail: f20190089@dubai.bits-pilani.ac.in

Abstract

Accurate prediction of remaining useful life (RUL) of lithium-ion battery plays an increasingly crucial role in the intelligent battery health management systems (BMS). The advances in machine learning introduce new data-driven approaches to this problem. In this paper, the various integrated machine learning models for RUL prediction of lithium-ion battery during its first 100 life cycles are proposed. Moreover, optimal performance of these models is compared and the major factors affecting the RUL of the battery are obtained. The proposed RUL prediction models is trained for remaining cycle life estimation. The various experimentations are carried out to test the effectuality of the presented models. It is observed that the proposed approach which is applied to the real dataset of lithium-ion battery cycle life performs better than the existing approaches. The experimental results show that the accuracy of RUL prediction (Cycle number) can be improved using machine-learning approaches.

Keywords – Lithium-ion battery, Remaining useful life, Machine learning, Prediction models, Feature extraction, Data driven approach, Battery management system (BMS).

1. INTRODUCTION

Lithium-ion batteries are deployed in a wide range of applications due to their low and falling costs, high energy densities and long lifetimes. However, as in the case with many chemical, mechanical and electronic systems, long battery lifetime entails delayed feedback of performance, often many months to years [1]. Accurate prediction of lifetime using cycle-data would unlock new opportunities in battery production, use and optimization [2]. The task of predicting lithium-ion battery lifetime is critically important given its broad utility and wide variability, even when controlling for operating conditions.

Recently, with the development of battery-related technologies, more and more data of these Lithium-ion batteries have been produced [3], [4]. A database using Lithium-ion phosphate (LFP)/graphite cells has been established by Saha and Goebel [5]. Based on these datasets, the data processing methods, such as machine learning approach, have been applied gradually to analyse the properties of the batteries. These machine learning approaches are especially attractive for high-rate operating conditions, where first-principles models of degradation are often unavailable. In short, opportunities for improving upon state-of-the-art prediction models include higher accuracy, earlier prediction, greater interpretability and broader application to a wide range of cycling conditions [6],[7]. The different approaches using statistical and machine-learning techniques to predict cycle life are attractive and mechanism-agnostic alternatives. Recently, advances in computational power and data generation have enabled these techniques to accelerate progress for a variety of tasks, including prediction of material properties, identification of chemical synthesis routes and material discovery for energy storage and catalysis [8],[9],[10].

An average Lithium-ion battery is capable of performing 700-800 cycles before dying out and discarded away completely. Degradation of batteries usually starts with the completion of 450-500 life cycles and continues to degrade before the chemicals present inside the cell are completely exhausted and of no use for any application [11]. Therefore, in this paper, data-driven models has been developed that accurately predict the cycle life of commercial lithium-ion phosphate (LFP) or graphite cells using early-cycle data and classify the given batteries, with no prior knowledge of degradation mechanisms and correlating the parameters which are responsible for determining the state of the Lithium-ion battery in its initial early stage life cycle (1-100).

2. DATA SET DETAILS

From the dataset [12] six features of the Lithium-ion battery cell is observed and taken into account during each cycle of the battery from beginning till the battery has completely degraded and of no use, which is around 760 life cycle. The six features used for early-stage prediction are; Charge Capacity, Discharge Capacity, Internal Resistance, Temperature, Current, and Voltage. For early-stage prediction, machine learning models are trained on the first hundred life cycles (1-100 life cycle), considering the above six factors on which the Lithium-ion battery cycle life depends.

3. MACHINE LEARNING APPROACHES AND MODELS

Based on the given dataset, six Machine Learning models has been created, namely: Deep Neural Network, Gradient Boosting Trees (GBT), Naïve Bayes, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Kernel PCA.

3.1 Deep Neural Network (DNN)

DNN was established using Python and open-source library Keras from Tensorflow. Our DNN consist of three hidden layers and one output layers. The activation function used in the hidden layers is "relu" function and the activation used in the output layer is "softmax" function. The optimizer used for the DNN is "adam" optimizer for its better performance on big data. The data was divided into the batch size of '10' and number of epochs used- '100'. Number of hidden layers used are '4' followed by 'binary cross-entropy' loss function.

3.2 Gradient Boosting Trees (GBT)

GBT is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision tree [13]. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function. The model had a maximum depth(max_depth) of '2', number of trees(n_estimators) of '20' with a learning rate of '0.1'.(Among other various learning rates used)

3.3 Naïve Bayes

Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of conditional independence between every pair of features given the value of the class variable. Naïve Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed form expression which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers. To implement- 'from sklearn.naive_bayes import GaussianNB', done on the basis of [14].

3.4 Principal Component Analysis (PCA)

PCA is a statistical procedure that uses an orthogonal transformation which converts a set of correlated variables to a set of uncorrelated variables. PCA is a most widely used tool in exploratory data analysis and in machine learning for predictive models. Moreover, PCA is an unsupervised statistical technique used to examine the interrelations among a set of variables. The PCA model consist of '2' number of components(n_components) followed by a Logistic Regression model with parameters – solver = 'lbfgs' , multi_class = 'auto'.

3.5 Linear Discriminant Analysis (LDA)

LDA is a dimensionality reduction technique which is commonly used for the supervised classification problems. It is used for modelling differences in groups i.e. separating two or more classes. It is used to project the features in higher dimension space into a lower dimension space. LDA explicitly attempts to model the difference between the classes of data. The PCA model consist of '2' number of components (n-components) followed by a Logistic Regression model.

3.6 Kernel Principal Component Analysis (kernel PCA)

Kernel PCA is an extension of principal component analysis using techniques of kernel methods. Using a kernel, the originally linear operations of PCA are performed in a reproducing kernel Hilbert space. Kernel PCA uses a kernel function to project dataset into a higher dimensional feature space, where it is linearly separable. It is similar to the idea of Support Vector Machines. The model consist of '2' components followed by kernel function 'rbf' (radial basis function).

4. PROPOSED MODEL

In literature, the most efficient machine learning models available; Deep Neural Network (DNN) [15], Gradient Boosting Trees (GBT) [16], Naïve Bayes, Principal Component Analysis (PCA) [17], Linear Discriminant Analysis (LDA) [18] and Kernel PCA [19]. The DNN model is run using Keras library provided by Tensor-flow. Rest of the machine learning models are run using Scikit-learn library. All these machine

learning models are run on the Anaconda Navigator software build with the Spyder integrated-development environment. The Pearson's coefficient of correlation is found using the SPSS software.

The complete proposed approach is demonstrated as shown in Fig. 1. Moreover, the Table 1 explain the major steps and their description.

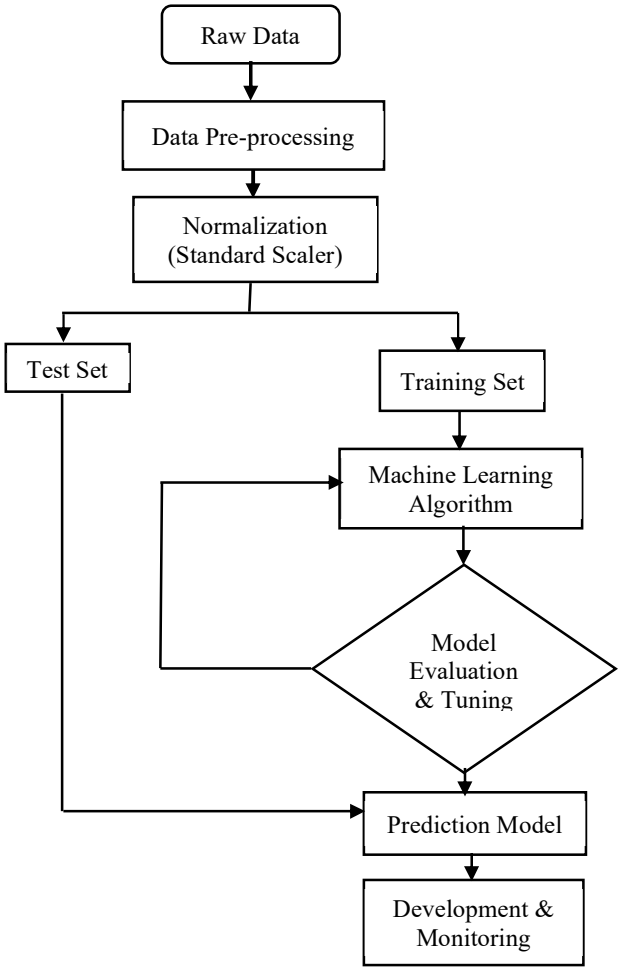

Fig. 1: Flow chart of proposed approach based on machine learning

Table 1: The most important features before machine learning model implementation on dataset

Feature	**Description**
Raw Data	Raw data of lithium-ion phosphate (LFP)/ graphite cells APR18650M1A is generated using a intelligent Battery Management System (BMS) and converted into categorical dataset.
Data Pre-Processing	The dataset is then cleaned, processed and unknown values are being averaged out and filled.
Normalization	Standard Scaler function is implemented on the dataset to reduce and eliminate data reluctance and improve data integrity.
Test/Train Set	The dataset is split into two halves – Testing set and Training set in the ratio of 0.25

5. EXPERIMENTATIONS AND RESULTS

5.1 Analysis of Dataset

The dataset [12] consists of 124 commercial lithium-ion batteries cycled to failure under fast-changing conditions. These lithium-ion phosphate (LFP)/ graphite cells APR18650M1A, are cycled in horizontal cylindrical fixtures on a 48-channel Arbin-LBT potentiostat in a forced convection temperature chamber set to 30˚C. These cells have a nominal capacity of 1.1 Ah and a nominal voltage of 3.3 V (Irregularity in dataset are not taken into account for ideal condition). Some of the important feature are as follows:

- All cells are cycled with one-step or two-step charging policies. The charging time varies from 8 to 13.3 minutes (0-80% SOC). There are generally two cells tested per policy, with the exception of 3.6C (80%).
- 1 minute and 1 second rests are placed after reaching 80% SOC during charging and after discharging, respectively.
- Cycle to 80% of nominal capacity (0.88 Ah).
- An initial 0.1C (10-hour discharge) cycle was performed in the beginning of each test.
- The cut-off currents for the constant-voltage steps are 0.02C for both charge and discharge.
- The pulse width of the current-resistance test is 30 ms.

Based on the prior knowledge of Lithium-ion phosphate (LFP)/graphite cells, we initialized six most important parameters in order to the predict early-life of the battery: 1. Charge Capacity, 2. Discharge Capacity, 3. Internal Resistance, 4. Temperature, 5. Current, 6. Voltage. Charge capacity and Discharge capacity are the basic attributes of lithium-ion batteries. Internal resistance characterizes the battery assembly process and internal state. Temperature represents the external environment of the battery. Current and Voltage are both intensive properties of the Lithium-ion Battery representing the current state. Initial 100 cycles are considered.

The average charging rates ranging from 3.6 C, the manufacturer's recommended fast-charging rate, to 6C to probe the performance of current-generation power cells under extreme fast-charging conditions (~10min charging), an area of significant commercial interest [20]. By deliberately varying the charging conditions, a dataset generated that captures a wide range of cycle lives, from approximately 150 to 2,300 cycles (average cycle life of 806 with a standard deviation of 377).

The observed conditions of Lithium-ion battery during first 100 cycles under nominal conditions is shown in Fig. 2. The graph is plotted between Voltage (V) and Cycle number is shown in Fig. 2(a). The Pearson's coefficient of correlation of this slope is 0.984. The graph is plotted between Current (mA) and Cycle number is shown in Fig. 2(b) and Pearson's coefficient of correlation of this slope is 0.174. Fig. 2(c) shows the relation between Charge Capacity (Ah) and Cycle number. The Pearson's coefficient of correlation of this slope is 0.891. Similarly, the graph is plotted between Discharge Capacity (Ah) and Cycle number shown in in Fig. 2(d), and the Pearson's coefficient of correlation of this slope is 0.922. Moreover, Fig. 2(e) depict the relation between internal Resistance (m-Ohm) and Cycle number. The Pearson's coefficient of correlation of this slope is 0.992. Finally, Fig. 2(f) shows the relation between Temperature (˚C) and Cycle number. The Pearson's coefficient of correlation of this slope is -0.221.

5.2 Preprocessing of Dataset

Standardization of datasets is a common requirement for many machine learning estimators. In order to eliminate the negative effects caused by different ranges of values, normalization of dataset is done to reduce and even eliminate data redundancy and improve data integrity. In addition, the dataset is transformed into the same interval, which facilitates the overall training of the depth model. Machine-Learning algorithms benefit from standardization of the dataset. It is done by standardizing the features by removing the mean and scaling to unit variance. Mathematically, normalization can be represented as follows:

$$Normalization\ z = \frac{x-\mu}{\sigma} \tag{1}$$

where μ is the mean defined as $$\mu = \frac{1}{N}\sum_{i=1}^{N}(x_i) \tag{2}$$

and σ is the standard deviation defined as $$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2} \tag{3}$$

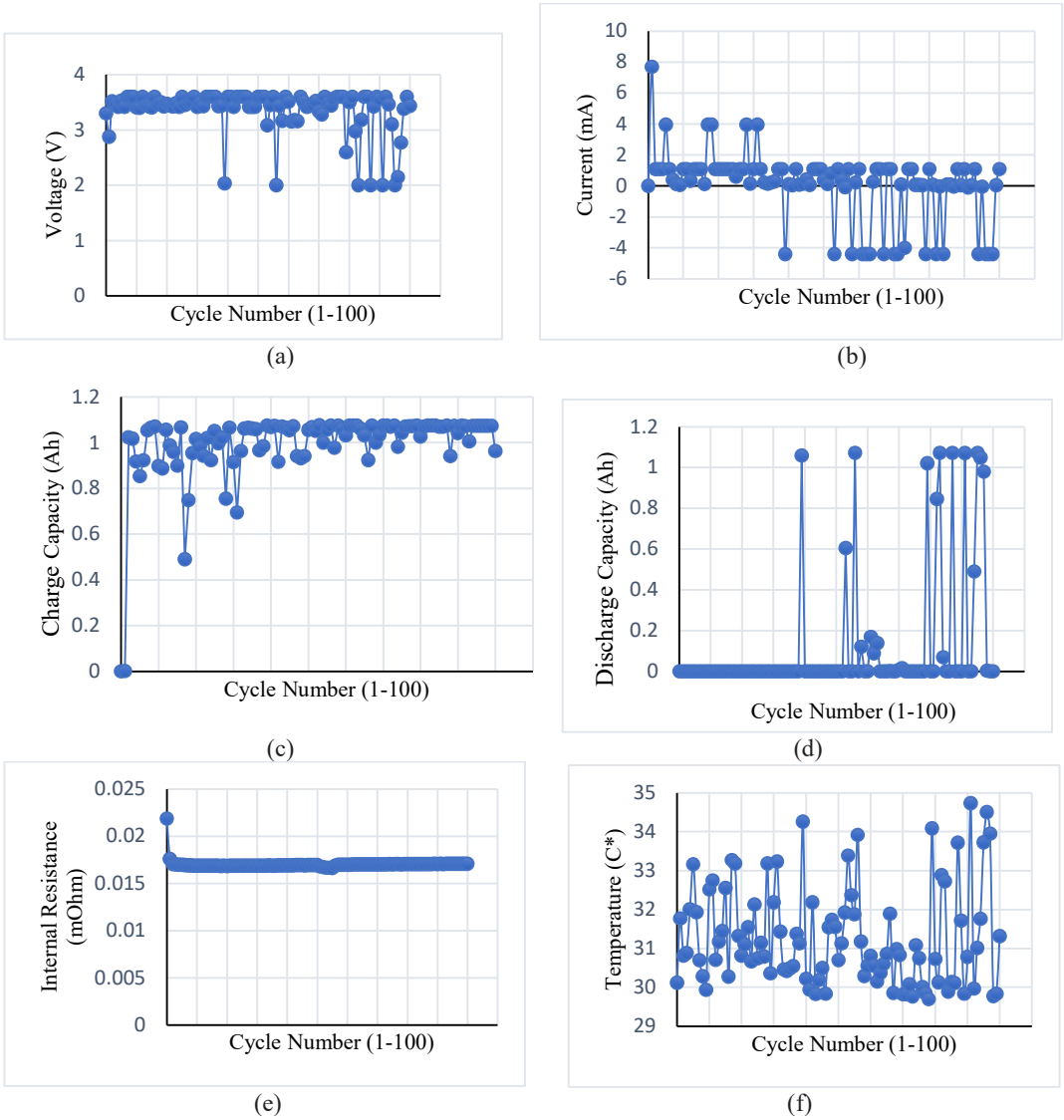

Fig. 2: Observed conditions of Lithium-ion battery during first 100 cycles under nominal conditions. a) Voltage and Cycle number, (b) Current (mA) and Cycle number, c) Charge Capacity (Ah) and Cycle number, d) Discharge Capacity (Ah) and Cycle number, e) Internal Resistance (m-Ohm) and Cycle number, f) Temperature and Cycle number.

The proposed algorithm is defined stepwise as given Algorithm 1.

Algorithm 1

```
1. From sklearn.preprocessing import StandardScaler
2. sc = StandardScaler()
3. X_train = sc.fit_transform(X_train)
4. X_test = sc.transform(X_test)
5.classifier=Sequential();classifier.add(Dense(activation="relu",input_dim=10,units=6,kernel_initializer="uniform"))
6. classifier.add(Dense(activation="relu", units=6, kernel_initializer="uniform"))
7. classifier.add(Dense(activation="relu", units=6, kernel_initializer = "uniform"))
8. classifier.add(Dense(activation="softmax",units=1,kernel_initializer="uniform"))
9. classifier.compile(optimizer='adam',loss='binary_crossentropy',metrics=['accuracy'])
10. classifier.fit(X_train, y_train, epochs = 100,batch_size = 10)
11. y_pred = classifier.predict(X_test)
```

Based on the above proposed algorithm the various experimentations are carried out using six different models and the accuracy has been compared among themselves to find out the efficient model. Fig. 3 shows the prediction accuracies of the six machine learning models on lithium-ion phosphate (LFP)/ graphite cells APR18650M1A dataset.

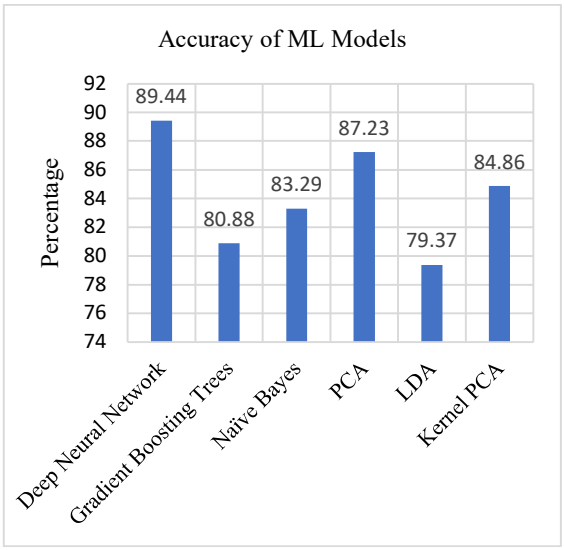

Fig. 3: Accuracy of different machine learning models

Moreover, the accuracy of the different models has been enumerated in Table 2 along with training and testing datasets.

Table 2: Accuracy Table for training and testing set

Model	Training Set	Testing Set
Deep Neural Network	92.023%	89.44%
Gradient Boosting Trees	84.01%	80.88%
Naïve Bayes	85.89%	83.29%
PCA	90.67%	87.23%
LDA	83.31%	79.37%
Kernel PCA	88.27%	84.86%

With the following results as shown in Table 2, it is found that Deep Neural Network outperforms the remaining 5 machine learning models on both training set and testing set. PCA model is the second-best model showing good results on the testing set and training set. Kernel PCA is third best predicting model followed by Naïve Bayes and Gradient Boosting Trees with LDA showing the least accuracy on the testing test. The decrease in accuracy on testing test as compared to training set signifies that the models were prevented from overfitting and therefore improves the optimization of the algorithms. Thus, Deep Neural Network and PCA are top predicting models and can be further improved by changing hyperparameters and with more data.

The experimental results show the effectiveness and better accuracy of the proposed approach as compared with the other linear regression, Bayesian regression, decision trees and other shallow models. Moreover, the Pearson's coefficient of correlation signifies that the Internal Resistance, Discharge Capacity and Charge Capacity are the major factors affecting the degradation of Lithium-Ion batteries with time followed by Discharge Capacity for the initial 100 life-cycles of the battery when in used.

CONCLUSION

Data-driven machine learning models shows a promising future for diagnostics and prognostics of lithium-ion batteries and enables emerging applications in their development, manufacturing and optimization. In this study, a cycle life prediction models are proposed using early-cycle discharge data. The presented models can achieve increasing performance in training speed, while obtaining results comparable to the traditional approaches. This work highlights the promise of combining data generation and data-driven modelling for understanding and developing complex systems such as lithium-ion batteries. However, it is observed that there is still room for improvement of top performing models. Therefore, an improve the proposed approach for lithium-ion battery RUL prediction in the contexts of different working conditions of the battery can be modified which may be the future work.

In future work, the presented model can be used in applications that can accelerate research and development of new battery designs and reduce the time and cost of production. It can shorten the time for validating new types of batteries, which is especially important given rapid advances in material. With the machine learning techniques, electric vehicle batteries determined to have short lifespans could be used instead to power street lights or back up data centres. Recyclers could find cells from used EV battery packs with enough capacity left for a second life and can help in development of improved Battery Management System (BMS)

ACKNOWLEDGEMENT

This work was supported by Birla Institute of Technology and Science, Pilani, Dubai Campus, Dubai, UAE. Special thanks to Professor Vilas H Gaidhane, Head of Department, Electrical & Electronics Engineering, BITS Pilani Dubai Campus.

PROPOSED MACHINE-LEARNING MODELS

All the machine-learning models proposed in this paper are uploaded on:
https://github.com/aditya77777?tab=repositories

REFERENCES

[1] R. Schmuch, R. Wagner, G. Hörpel, T. Placke, and M. Winter, "Performance and cost of materials for lithium-based rechargeable automotive batteries", *Nat. Energy*., 3(4), pp. 267–278, 2018.
[2] B. Nykvist and M. Nilsson, "Rapidly falling costs of battery packs for electric vehicles", *Nat. Clim. Change*, 5, pp. 329–332, 2015.
[3] Y. Zhang, R. Xiong, H. He and M. Pecht, "Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries", *IEEE Trans. Veh. Technol*., 67(7), pp. 5695–5705, 2018.
[4] J. Liu, A. Saxena, K. Goebel, B. Saha and W. Wang, "An adaptive recurrent neural network for remaining useful life prediction of lithium-ion batteries", National aeronautics and space administration moffett field caames research center, 2010.
[5] B. Saha and K. Goebel, "Battery data set", NASA Ames Progn. Data Repos. 2007.
[6] K. Goebel, B. Saha, A. Saxena, J. R. Celaya,"Prognostics in battery health management", *IEEE instrumentation & measurement magazine*, 11, pp. 33-40, 2008.
[7] M. Berecibar, M. Garmendia, I. Gandiaga, J. Crego and I. Villarreal, "State of health estimation algorithm of LiFePO4 battery packs based on differential voltage curves for battery management system application", *Energy*, 103, pp. 784–796, 2016.
[8] X. Hu, J. Jiang, J. D. Cao, B. Egardt, "Battery health prognosis for electric vehicles using sample entropy and sparse Bayesian predictive modeling", *IEEE Trans. Ind. Electron*., 63(4), pp. 2645–2656, 2016.
[9] L. Ward, A. Agrawal, A. Choudhary and C. Wolverton, "A general-purpose machine learning framework for predicting properties of inorganic materials", *NPJ Comput. Mater*., 2, pp.16028, 2016.
[10] P. Raccuglia, K. C. Elbert, P. D. Adler, C. Falk, M.B. Wenny, A Mollo, M. Zeller, S. A. Friedler, J. Schrier, and A. J. Norquist, "Machine-learning-assisted materials discovery using failed experiments", Nature, 573, pp. 73–77, 2016.
[11] M. Dubarry, C. Truchot and B. Y. Liaw, "Synthesize battery degradation modes via a diagnostic and prognostic model", *J. Power Sources*, 219, pp. 204–216, 2012.
[12] The datasets used in this study are available at - https://data.matr.io/1
[13] M. A Patil, P Tagade, K. S. Hariharan, S. M. Kolake and T. Song, A novel multistage Support Vector Machine based approach for Li ion battery remaining useful life estimation", *Applied energy,* 159, pp 285-29 2015.
[14] R. R. Richardson, C. R. Birkl, M. A. Osborne and D. A. Howey, "Gaussian process regression for

capacity estimation of lithium-ion batteries", *IEEE Trans. Ind. Inform*., 15(1), pp. 127–138,2019.
[15] A survey of deep neural network architectures and their applications.
W Liu, Z Wang, X Liu, N Zeng, Y Liu, FE Alsaadi - Neurocomputing, 2017 - Elsevier
[16] A highly efficient gradient boosting decision tree.
G Ke, Q Meng, T Finley, T Wang, W Chen- Advances in neural , 2017
[17] Probabilistic principal component analysis.
ME Tipping, CM Bishop - The Royal Statistical Society: Series, 1999
[18] Fisher discriminant analysis with kernels
S Mika, G Ratsch, J Weston- Neural networks for, 1999 – IEEE Explore.
[19] Kernel principal component analysis.
B Schölkopf, A Smola, KR Müller - International conference on artificial , 1997 - Springer
[20] E. Sarasketa-Zabala et al., "Understanding lithium inventory loss and sudden performance fade in cylindrical cells during cycling with deep-discharge steps", *J. Phys. Chem. C*, 119, pp. 896–906, 2015.