Predicting the Battery Health of Lithium-ion Batteries using Machine Learning

Asvika A P[1], Nandhakishore M[2], Pradeep Kumar M[3]
[1,2,3]*Department of Electronics and Communication Engineering, Kumaraguru College of Technology, Coimbatore - 641049*

***Abstract*— Lithium-ion batteries commonly applied in contemporary technologies are a vital source of green energy solutions to a variety of real-time issues owing to their energy density, long-life period and efficiency benefits. Playing its role, it's also required to pay attention to its maintenance, safety and cost of storage to make it useful for different applications. For its attainment, this piece of work emphasizes presenting an ML-based model to make predictions for battery health. This piece of work utilizes ML models including Decision Tree (DT) and Long Short-Term Memory (LSTM) for the purpose of analyzing data and predicting outcomes. Data captured in real-time from Lithium Iron Phosphate (LFP) battery and Lithium Nickel Manganese Cobalt Oxide (NMC) cells with the aid of external support, were pre-processed and divided into training and test datasets. The model design was created to make the best predictions from the data and lead to precise predictions. On testing and training using LFP battery data, the model predictions were measured with results demonstrating that DT registered a higher Mean Absolute Error (MAE) of 0.9375 whereas LSTM indicated a reduced MAE of 0.01575 and therefore is more suitable for battery health prediction. From this performance, data of NMC cells was treated using LSTM model since it highly comprehends sequential data, identifying the long-term dependencies. These outcomes support efficient and effective monitoring, maintenance, safety and restricting storage cost of the Lithium-ion batteries. This work supports comprehension of the prediction power of the ML methods which has a key role in molding the Lithium-ion batteries as sustainable energy source and solution.**

***Index Terms*—Decision Tree (DT), Lithium-ion Battery, Lithium Nickel Manganese Cobalt Oxide (NMC), Long Short-Term Memory (LSTM), Machine Learning, Remaining Useful Life (RUL), State of Health (SoH).**

I. INTRODUCTION

The world's use of lithium-ion batteries (LiBs) has revolutionized contemporary energy storage technology. Their benefits and users enhance the demand that makes them inevitable for green energy use. Nevertheless, regardless of their benefits, LiBs age with time, causing increase in internal resistance and safety risks [1]. Precise prediction of State of Health (SoH) and Remaining Useful Life (RUL) is required to attain highest performance, reliability, and safety in battery applications.

Historically, battery degradation has been monitored using electrochemical modeling and experimentation techniques, which while precise, are computation and time intensive [2]. Machine Learning (ML)-based techniques are a real alternative in the sense that they use data-driven algorithms to analyze degradation patterns and provide real-time predictions. Such models can analyze large amounts of historical battery data, learn about degradation patterns, and predict future battery performance accurately. ML methods have been found to hold potential in improving battery diagnostics, saving cost, and making predictive maintenance policy possible [3].

Different ML algorithms have been used for battery health prediction. Our proposal is aimed at utilizing DT and LSTM owing to their high-precision prediction. DT models offer an understandable and a computationally effective means of predicting SoH based on major battery parameters like voltage, current [4]. In case of capturing difficult, long-range dependencies in the degradation of a battery, it has been seen that LSTM networks are more efficient, as it is established to deal with sequence data and non-linear degradation behavior [5].

Recent research shows that LSTM models are more accurate than traditional statistics in predicting SoH and RUL with higher accuracy. Researchers in [6] employed LSTM-based SoH prediction and achieved significantly reduced error rates compared to conventional methods. With data collected from batteries and cells, we move forward employing DT and LSTM to significantly reduce the error rates by providing a more functional and accurate model in real-life.

The results are designed to improve battery management systems (BMS) with safety, reliability, and extended use of Li-ion batteries in mission-critical applications. With the growing demand for energy storage and electrification, integrating AI-powered battery diagnostics will be the key to creating sustainable energy solutions. With advancements in battery technology, smart ML-based prediction systems will be the game-changer for improving performance and avoiding failures, ultimately defining the future of energy storage and electric mobility.

II. RELATED WORK

Precise SoH and RUL prediction of LiBs is essential in EV and renewable energy system applications. Many methods have been evolved over the years, classified into broad categories of physics-based models, machine learning (ML) methods, and hybrid techniques.

In the early period, most of the research efforts were directed toward applying electrochemical methods and equivalent circuit models (ECMs) for battery health estimation. He et al. [3] compared various ECM-based models of State of Charge (SoC) prediction based on their accuracy and their inability to respond to actual battery aging under real-time conditions. Plett [2] also used Extended Kalman Filtering (EKF) to estimate battery states, but the approach necessitated a painstaking process of parameter tuning and suffered from nonlinear issues. Zhang and Lee [7] discussed some model-based prognostic techniques with a focus on the computational intensiveness of electrochemical impedance spectroscopy (EIS) and system identification complexity.

ML and deep learning (DL) have been used widely to surpass the shortcomings of physics-based models. Ecker et al. [8] showed that SoH could be predicted with low error rates using Support Vector Machines (SVM) and Random Forest (RF) from historical cycling data. Severson et al. [9] stated a Gaussian Process Regression (GPR)-based method with enhanced SoH prediction. DL models like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have also been used widely due to their ability to learn complex temporal interactions. Wu et al. [10] explained an LSTM-based approach which aims to enhance accuracy for SoH estimation than traditional ML algorithms.

Hybrid approaches contain a combination of physics-based constraints with deep learning frameworks have also been proposed. Chen et al. [11] came up with a Physics-Informed Neural Networks (PINN), where they integrated electrochemical knowledge into neural networks to increase SoH predictions' interpretability and accuracy. Li et al. [12] also designed a hybrid battery state-of-health estimation framework using partial differential equations (PDEs) and neural networks (RNNs) to improve sensor noise robustness.

Furthermore, transfer learning and domain adaptation techniques have been utilized to improve model generalizability across different battery chemistries. Zhang et al. [13] reported that small-sized datasets of new battery types were adequate for fine-tuning pre-trained LSTM models, without the need for vast amounts of labelled data. Yang et al. [14] attempted unsupervised domain, which was successful in applying SoH estimation models to real field datasets captured from diverse operating conditions.

Relying on such studies, our work utilizes Decision Tree and LSTM models to provide precise real-time SoH and RUL prediction with real-time battery and cell datasets for training and testing. Using time-series learning and data-driven knowledge, our approach focuses on enhancement of battery diagnosis, predictive maintenance, and efficiency of overall Battery Management System (BMS) of EVs and energy storage systems.

III. METHODOLOGY

SoH and RUL prediction of LiBs needs to be done with the use of advanced ML methods, which are more accurate compared to the conventional model-based methods. DT and LSTM are utilized in this work for Soh and RUL prediction.

A. Machine Learning

In battery diagnosis, ML is a key component used for the identification of degradation trends. In contrast to physics-based models, ML algorithms can be implemented in real-time in BMS since they don't need profound knowledge of electrochemical mechanisms. LSTMs possess excellent prediction abilities and hence are well suited for tracking battery conditions over time. DTs are fast and interpretable.

B. Decision Tree (DT) Model

DT algorithm is also a popular ML battery health predictor because of the hierarchical nature of DT and relatively low computational expense. DT models capture nonlinear input variables and aging patterns but lacks temporal awareness and is therefore restricted from examining long-term aging patterns.

C. Long Short-Term Memory (LSTM) Model

LSTMs, or the subset RNN, are particularly suitable for sequential data processing. LSTMs leverage memory cells and gate control structures to store critical information across charge-discharge cycles so that they can be very efficient in predicting long-term battery health trends. Previous work [5] has shown that models based on LSTM are superior to the traditional ML methods in learning lithium-ion battery long-term degradation patterns.

D. Implementation Overview

A dataset containing voltage, current, capacity, and various other features of a LFP battery and NMC cells at 1C rating were collected and used for training and testing. The DT model provided a quick estimation, while the LSTM model captured complex degradation patterns over time. Preprocessing comprises normalization and train-test split and Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) was used to estimate prediction accuracy. With DT and LSTM models, this paper aims to improve SoH and RUL prediction accuracy to enable more efficient battery management and predictive maintenance.

These datasets are the foundation for training and validating the DT and LSTM models with real-world validation of battery health prediction methods.

E. Dataset Description

Two datasets mentioned in Table I were collected for this work using controlled battery testing experiments under external aid. The datasets have real time data measurements of voltage, current, capacity, and other features over a single cycle.

TABLE I. DATASET DESCRIPTION

Dataset	Description			
	Battery Chemistry	*Nominal Capacity*	*Testing Condition*	*Features recorded*
Dataset 1 (LFP Battery)	LiFePO$_4$	48 Ah	Constant current discharging (CC-DC) at room temperature.	Voltage, Current, Discharge capacity, Power, dQ/dV, Energy, Capacity
Dataset 2 (NMC Cell)	LiNiMnCoO$_2$	4.85 Ah	CC-DC at room temperature at 1C rating.	Voltage, Current, Capacity, Energy

These datasets are the foundation for training and validating the DT and LSTM models with real-world validation of battery health prediction methods.

F. Data Preprocessing

To provide uniformity and enhance model performance, all the features were normalized through Min-Max Scaling, converting values within the range 0 to 1. For Dataset 1, both DT and LSTM models were employed for SOH prediction. The DT model requires feature scaling, whereas the LSTM model implicated scaling both input features and target labels to ensure uniformity in sequential learning.

SoH was computed as:

$$\text{SoH}=\frac{Ct}{Cnom}.100 \qquad (1)$$

Where Ct refers to measured Capacity, and Cnom refers to Nominal capacity.

As the error rate in LSTM models is significantly low, we proceeded to utilize them for Soh and RUL prediction in the event of Dataset 2.

RUL was estimated by:

$$\text{RUL}=\frac{SoHactual-\text{SoHeol}}{SoHactual-SoHeol}.Total\ Lifespan \qquad (2)$$

Where SoHactual is the current state of health, SoHeol represents the End-of-Life threshold, and SoHinitial is the initial health of the battery.

The structured preprocessing strategy was utilized, which involved missing value handling and normalizing features. This preprocessing ensures standard data representation, effective training of the model, and precise battery health prediction.

G. Model Training

The models were trained using pre-processed datasets to predict SoH and RUL of lithium-ion battery and cells. Two approaches were used:

- The DT model was trained as a baseline estimator using key parameters such as voltage, current, capacity and power to predict Soh for Dataset 1. The model was constructed with a maximum depth of 5 and 42 random states.
- The LSTM model was trained on sequential battery and cell data of both Datasets. The model is constructed using multiple LSTM layers with a dense layer as the output layer to capture long-term trends of degradation. It was trained with the Adam optimizer and MSE as the loss function.

The hyperparameters used are summarized in Table II.

Models are compared and evaluated following training on Mean Absolute Error (MAE) and MSE to evaluate their performance and precision.

IV. RESULTS AND DISCUSSIONS

A. State of Health (SoH) Prediction

The accuracy of SoH prediction models was checked against both DT and LSTM networks. Fig. 1 illustrates actual and predicted values from the DT model. Results have a clear likeness between actual and predicted, with the DT model successfully capturing patterns of battery degradation.

Similarly, LSTM was trained and tested for SoH prediction on both data sets.

TABLE II. TRAINING HYPERPARAMETERS

Model	Model Training				
	Layers	*Units per Layer*	*Batch Size*	*Epoch*	*Optimizer*
DT Model (SoH)-Dataset 1	-	-	-	-	-
LSTM (SoH)-Dataset 1	2LSTM +1 Dense	128,64,1	32	50	Adam
LSTM (SoH)-Dataset 2	2LSTM +1 Dense	100,50,1	32	100	Adam

LSTM performed better than DT in identifying temporal relationships in battery degradation as shown in Fig. 2 and Fig. 3. MinMax scaling used in preprocessing facilitated stable and effective model convergence.

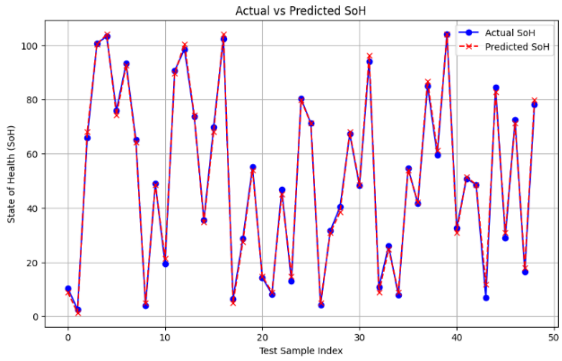

Fig. 1. Decision Tree prediction for Dataset 1

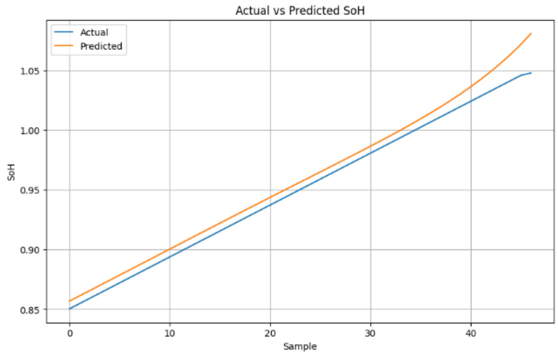

Fig. 2. LSTM prediction for Dataset 1

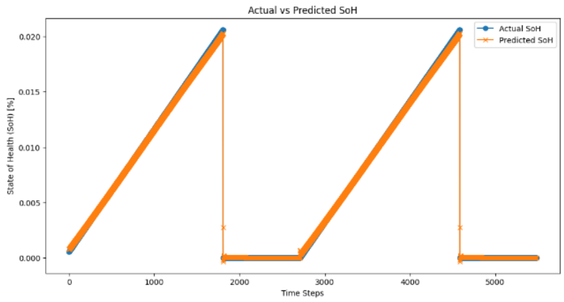

Fig. 3. LSTM prediction for Dataset 2

B. *Remaining Useful Life (RUL) Prediction*

For RUL estimation, the LSTM model was used due to its capability to process sequential time-series data. The estimated RUL values were extremely close to the actual degradation patterns, which validates the model's effectiveness in estimating battery life. Fig. 4 shows the predicted vs. actual RUL comparison, where the LSTM model generalizes across different data.

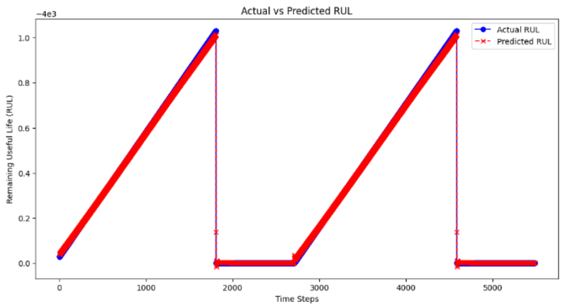

Fig. 4. LSTM results for Dataset 2 – RUL

C. *Performance evaluation*

To measure the prediction accuracy in numerical terms, MAE, Root Mean Square Error (RMSE), and R^2 scores were estimated for both the models. LSTM surpassed DT by exhibiting lower RMSE and greater R^2 scores, validating its efficacy for forecasting long-term degradation. A concise comparison of performance measures is outlined in Table III.

TABLE III. EVALUATION RESULTS

Dataset	*Model*	*MAE*	*RMSE*	*R^2 Score*
Dataset 1	DT (SoH)	0.9375	1.4725	0.9978
Dataset 1	LSTM (SoH)	0.01575	0.01598	0.86166
Dataset 2	LSTM (SoH)	0.00004	0.00039	0.99666
Dataset 2	LSTM (RUL)	0.00251	0.01985	0.99666

D. *Discussions*

The results show that while Decision Trees provide a fast and interpretable approach for SoH prediction, LSTMs deliver better performance due to their ability to capture sequential dependencies. The trained models provide promising potential for real-time battery health monitoring applications. However, deviations occur in these predictions which requires furthermore model fine-tuning and dataset expansion to improve accuracy of predictions.

V. CONCLUSIONS

In this research, we trained and tested ML models for SoH and RUL estimation of LiBs on two different datasets: one from a LFP battery and another from an NMC cells. A DT model was used for SoH estimation in Dataset 1, whereas an LSTM model was utilized for both datasets. The models were developed and tested using real-time battery data, pre-processed with MinMax scaling for feature normalization.

Experimental outcomes showed that the LSTM-based model performed better than the DT model in identifying intricate temporal relationships, presenting a great accuracy for SoH and RUL prediction. The visual representation of actual and predicted values also proved the execution of the proposed method. These outcomes provide implications for the power of deep learning in battery health estimation as a starting point for future developments in predictive maintenance and battery management systems.

Future research can investigate the incorporation of other battery parameters, real-time implementation, and optimization methods to enhance predictive accuracy and model generalization further.

VI. ACKNOWLEDGEMENT

We sincerely acknowledge the support of Harshini K of the Department of Electrical and Electronics and Sebastin Daniel J of the Department of Electronics and Instrumentation of Kumaraguru College of Technology. Their efforts under the Research and Exploration (Re) Forum were of great help in facilitating the ease of conducting the data gathering process. Their dedication and support greatly assisted in the completion of this study.

REFERENCES

[1] J. Zhang and J. Lee, “A review on prognostics and health monitoring of Li-ion battery for EVs,” *J. Power Sources*, vol. 196, no. 15, pp. 6007–6014, 2011.

[2] G. L. Plett, “Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 1. Background,” *J. Power Sources*, vol. 134, no. 2, pp. 252–261, 2004.

[3] H. He, R. Xiong, and J. Fan, “Evaluation of lithium-ion battery equivalent circuit models for state of charge estimation by an experimental approach,” *Energies*, vol. 4, no. 4, pp. 582–598, 2011.

[4] Y. Xing, E. W. M. Ma, K. L. Tsui, and M. Pecht, “Battery management systems in electric and hybrid vehicles,” *Energies*, vol. 4, no. 11, pp. 1840–1857, 2011.

[5] C. Chen, X. Hu, and Y. Li, “State of health estimation of lithium-ion batteries based on LSTM recurrent neural network,” *IEEE Trans. Ind. Electron.*, vol. 66, no. 11, pp. 8770–8780, 2019.

[6] W. Waag, C. Fleischer, and D. U. Sauer, “Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles,” *J. Power Sources*, vol. 258, pp. 321–339, 2014.

[7] C. Zhang and J. Lee, “A review on prognostics and health management of lithium-ion

batteries," *IEEE Trans. Rel.*, vol. 61, no. 2, pp. 519–531, 2012.

[8] M. Ecker, T. K. Kim, S. Schnackenburg, and D. U. Sauer, "Prediction of lithium-ion battery aging for automotive applications using support vector machines," *J. Power Sources*, vol. 248, pp. 841–851, 2014.

[9] K. A. Severson, P. M. Attia, N. Jin, and R. D. Braatz, "Data-driven prediction of battery cycle life before capacity degradation," *Nat. Energy*, vol. 4, no. 5, pp. 383–391, 2019.

[10] C. Wu, J. Cao, and Z. Zhang, "Long short-term memory neural network for lithium-ion battery state of health estimation," *IEEE Trans. Ind. Electron.*, vol. 67, no. 8, pp. 6765–6775, 2020.

[11] Y. Chen, T. Liu, and X. Hu, "Physics-informed neural networks for lithium-ion battery state estimation," *IEEE Trans. Ind. Electron.*, vol. 68, no. 8, pp. 7454–7464, 2021.

[12] Y. Li, W. Wei, and Z. Han, "Hybrid physics-informed deep learning model for lithium-ion battery health prognosis," *IEEE Trans. Transport. Electrific.*, vol. 8, no. 3, pp. 3261–3272, 2022.

[13] X. Zhang, L. Chen, and H. Wang, "Transfer learning-based state of health prediction for lithium-ion batteries across different chemistries," *IEEE Access*, vol. 9, pp. 77452–77464, 2021.

[14] Y. Yang, S. Sun, and B. Wang, "Battery remaining useful life prediction using ensemble learning," *IEEE Trans. Transport. Electrific.*, vol. 7, no. 2, pp. 254–266, 2021.