Implementation of Machine Learning Algorithms to Predict Lithium-ion Battery Degradation (Residual Life Prediction)

Chinmay Pathare

Project Summary:

The project aims to address the challenge of managing battery health and replacement in industrial settings, particularly focusing on touch computers with replaceable batteries. These devices offer significant advantages in terms of flexibility and uninterrupted operation but pose challenges in tracking battery health accurately. The project proposes the development of machine learning algorithms to forecast battery capacity using data collected from touch computers. By leveraging battery parameters such as voltage, current, and temperature, the algorithms aim to predict the remaining useful life (RUL) capacity of lithium-ion batteries, thereby enabling businesses to optimize inventory management and reduce costs associated with battery replacement.

The project adopts a data-driven approach, utilizing machine learning techniques such as ARIMAX and CNN-LSTM models. These models are trained and tested using battery datasets, and their performance is evaluated based on metrics such as Mean Absolute Error (MAE) and Mean Squared Error (MSE). The results indicate that while the ARIMAX model shows promising accuracy in the short term, the CNN-LSTM model demonstrates better reliability for real-life applications, particularly in forecasting battery capacity to reach critical thresholds.

Project Motivation:

In today's industrial landscape, the integration of technology has become important for optimizing workflows and enhancing productivity. Touch computers with replaceable batteries represent a significant advancement in this regard. Combining the user-friendly interface of touchscreens with the practicality of replaceable batteries, these devices revolutionize how tasks are executed on the factory floor, in warehouses, at airports and beyond. The ability of these devices to swap batteries on the go ensures uninterrupted usage, eliminating costly downtimes associated with traditional devices that rely on fixed internal batteries. It is particularly crucial in industries where continuous operation is essential, such as manufacturing and distribution.

Lithium-ion batteries have emerged as the go-to choice for batteries for this use case, owing to their exceptional rechargeability, and longevity. These batteries provide a good power source that is ideal for sustaining the demands of modern industrial applications. The rechargeability of the lithium-ion battery eliminates the need for disposable batteries, reducing waste and environmental impact while also providing cost savings over time. Workers can simply swap out depleted batteries for fully charged ones, maintaining continuous operation without lengthy downtimes for recharging.

However, while the rechargeability of lithium-ion batteries offers undeniable benefits, it also presents a challenge in terms of tracking battery health and determining when replacements are necessary. Unlike disposable batteries, which typically exhibit a gradual decline in performance before reaching the end of their lifespan, lithium-ion batteries can degrade more subtly over time.

As time passes, repeated cycles of charging and discharging result in deterioration of the battery's components. The circumstances under which the battery is recharged, such as frequent rapid charging, can expedite this deterioration process. Additionally, factors such as temperature, duty cycle, and the construction of the battery pack significantly influence the rate at which the battery ages [1].

This uncertainty surrounding lithium-ion battery health poses a significant challenge in procurement planning for replacements. Without clear visibility,it becomes difficult to accurately predict when batteries will reach the end of their lifespan. This uncertainty can lead to inefficiencies in inventory management, as organizations may either overstock on replacement batteries, tying up capital unnecessarily, or risk running out of batteries when needed, disrupting operations. Thus, the lack of insight into battery health necessitates a careful balance between maintaining adequate replacement stock levels and minimizing excess inventory costs in industrial settings.

To sustain a healthy inventory level, many industries use stock acquisition strategies devised by utilizing standard regression models that rely on historical consumption rates to make projections. The motivation of this project is to make a shift in that approach and use a machine learning algorithm capable of utilizing battery data collected by the touch computers to forecast the remaining capacity of batteries. Such a tool would empower businesses to fine-tune their inventory management, thereby influencing inventory costs.

Literature Review:

As the core component of electronic products, in long-term use, due to various factors, the performance and life of lithium batteries will be affected, which may cause some troubles for users. That is why battery life has become a growing concern in recent years [2]. The basic composition of the battery pack is a single lithium battery. Therefore, the research of battery pack management system technology usually includes the monitoring of temperature, current and voltage of a single battery, capacity prediction, and charge state prediction of a single battery [3]. The battery's capacity serves as a crucial gauge of its performance, while its longevity is a pivotal measure of its condition. Once the battery's capacity dwindles to a specific critical threshold (often around 80% of its rated capacity, though this varies depending on the battery type), it is deemed unusable [4].

At present, although there are many kinds of lithium battery life prediction algorithms, according to the principle of modeling, it is divided into two types: physical model and data-driven. Physical model-based methods refer to the prediction of remaining life (RUL) based on the physical properties of lithium batteries (e.g., material properties and loading conditions) and degradation mechanisms. (2) The data-driven RUL prediction method is to extract the characteristic parameters that can reflect the battery health status from the monitored variables such as voltage and current and build a statistical model of the system to extrapolate the prediction of RUL [5]. Because the data-driven method does not require the establishment of a complex system physical model, it is suitable for RUL prediction of complex and changeable internal lithium battery systems [5].

Numerous data-based methods have been proposed for this use case in the last decade. These methods utilize machine learning techniques to estimate the current capacity and remaining capacity using variables like voltage, charging current, discharge current etc. Some of the common models used for this approach are vector regression, support vector regression (SVR), recurrent neural network (RNN), long short-term memory (LSTM) neural network and other techniques [6]. These methods establish the relationship between external characteristics of battery aging and SOH without a precise mechanism model, thus making them more practical [7].

Research suggests that advanced deep learning neural networks (NN) perform remarkably well in non-linear and high dimensional data modelling [8]. LSTM NN is useful for this use case which is designed to control the flow of information for long-term dependencies and eliminating vanishing or exploding gradients [8]. Researchers highly recommend using LSTM for long-term degradation state and the RUL of lithium-ion batteries. [8]. Generally, before the LSTM NN performs a prediction, a learning process is required to determine the network parameters using the historical data of LIBs. Because the degradation dynamics of LIBs are unknown and uncertain, parameter-fixed LSTM trained only with historical data demonstrates unsatisfactory adaptability and generalization in the prediction stage [8].

Yapeng Zhou, Miaohua Huang proposed a novel approach by combining empirical mode decomposition (EMD) and autoregressive integrated moving average (ARIMA) to predict battery RUL [9]. They tested both ARIMA and EMD-ARIMA models to predict battery RUL. The results of their research identified that the absolute errors of RUL with ARIMA are bigger than that of EMD-ARIMA. Another approach taken by Xifeng Guo et al. was based on Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), One-dimensional Convolutional Neural Network (1D CNN) and Bi-directional Long Short-Term Memory (BiLSTM) neural network [10]. They compared the BiLSTM model with a CNN-BiLSTM model. The results show that prediction using the CEEMDAN-CNN BiLSTM model had higher accuracy [10].

Method for Analysis:

In combination with domain expertise of industry application and the research conducted on existing models for predicting battery health dataset, the following approach was selected for an end-to-end analysis:

i) Data Collection: Collect battery dataset to perform analysis.
ii) Data Cleaning: Clean data for any inconsistencies or missing values.
iii) Data Visualization: Visualize data and identify required approaches.
iv) Data Pre-Processing and Analysis.
v) Model Training and Testing.
vi) Model Validation.

Implementation:

1) Data Collection:

Currently, touch computers extensively employed in various industries predominantly run on Android-based platforms. Android.os facilitates battery monitoring through its BatteryManager class, enabling access to crucial charging parameters such as voltage, current, temperature, and datetime [11]. Additional device-specific APIs can be used to gather insights, including total cumulative charge since the battery's inception and duration for the last charge [12]. These comprehensive data streams empower users to collect battery parameters required for model training and prediction.

An effort was undertaken to locate data sources for a particular 4200mAh lithium-ion battery type used in touch computers, but the search yielded no results. Another effort was made to identify boundary parameters of a particular 4200mAh lithium-ion battery type used in touch computers. The goal was to use the boundary parameters to generate synthetic data for analysis. However, generating randomized synthetic data is not an effective approach because it fails to replicate real-life battery health information accurately. Randomized data lacks the inherent patterns, complexities, and nuances present in actual battery health data, which are crucial for developing reliable and robust predictive models. As a result, relying on randomized synthetic data may lead to inaccurate predictions and unreliable insights, hindering the effectiveness and applicability of the developed algorithms in real-world scenarios. This approach was hence not pursued for the purpose of this analysis. A useable dataset with similar battery capacity was found from NASA's public repository [13].

This dataset consists of data for three batteries (RW9, RW10 and RW11) that is collected from batteries that were continuously cycled with randomly generated current profiles. 18650 Li-ion batteries were continuously operated using a sequence of charging and discharging currents between -4.5A and 4.5A. A random walk mode of battery cycling was used for this data collection [13]:

a) Selecting a charging or discharging current at random from the set {-4.5A, -3.75A, -3A, -2.25A, -1.5A, -0.75A, 0.75A, 1.5A, 2.25A, 3A, 3.75A, 4.5A}. Negative currents are associated with charging and positive currents indicate discharging.
b) The selected current setpoint is applied until either the battery voltage goes outside the range (3.2V - 4.2V) or 5 minutes has passed.
c) These steps are identified with the comment field = "discharge (random walk)" and comment field = "charge (random walk)"
d) After each charging or discharging period, there will be a <1s period of rest while a new charging or discharging current setpoint is selected.
e) This step is identified with the comment field = "rest (random walk)"
f) Steps 2 and 3 are repeated 1500 times, then characterization cycles are preformed to benchmark battery state of health.

The variables collected in the data set are as follows [13]:

- comment (string description of step)
- type (one character identifier of step: 'C' = Charging, 'D' = Discharging, 'R' = Resting (current = 0))
- relativeTime (vector of time in seconds, referenced to the beginning of current step)
- time (vector of sample time in seconds, referenced to the beginning of the experiment)
- voltage (vector of sample voltage in units of Volts)
- current (vector of sample current in units of Amps)
- temperature (vector of sample temperature in units of degrees C)
- date and time at which the current step was started in dd-Mon-yyyy HH:MM:SS format

2) Data Cleaning:

The selected dataset has multiple variables collected. Based on the domain knowledge and research conducted, the total variables for analysis were reduced to the following:

- Type
- Relative Time (secs)
- Voltage (Volts)
- Current (Amps)
- Temperature (degree Celsius)
- Date (dd-Mon-yyyy HH:MM:SS)

First the data was cleaned by identifying any empty rows and removing them from the dataset. To enable analysis based on battery state of health (SOH), capacity was chosen to be as the response variable for the dataset. Capacity is the amount of charge a battery can store or deliver at a given current rate. It is typically measured by performing a constant-current discharge and a constant-current charge between the minimum and maximum voltage thresholds; once the maximum voltage is reached, the voltage is maintained at this upper limit by reducing the current down to a threshold. This procedure is often called "CC-CV charging" [14]. This was calculated using the method of Coulomb Counting which relies on the integration of the current drawn and supplied to a battery over time [15]. This was achieved using the *relativeTime* and *current* parameters collected from the battery dataset.

The method of coulomb counting uses the discharge states of the battery data collection to identify the SOH (capacity). The available dataset contained both charging and discharging steps which would limit the use of the data for training. Removing charging steps completely would have resulted in loss of critical information that impacts battery health [1]. As a result, a new variable called *cumulative_charge* was created using the *relativeTime* of charge and the *current* applied. This would capture the total charge applied to the battery from the first usage. This method enabled the capturing of critical information on charging and supplemented in the removal of all charging steps. Furthermore, a new variable called *cycles* was added which would count the total charge-discharge cycles the battery has gone through.

The original data also consists of reference discharge steps that capture data where the battery voltage goes outside the range (3.2V - 4.2V) or 5 minutes has passed. This acts as a full charge/discharge cycle for the given battery. However, using only these steps will not be useful in real life applications. The usage of these batteries in industrial application will most likely be randomized and not go through full cycles of charging / discharging. To replicate real life battery datasets, these reference discharge steps were removed from the dataset. This completed the data cleaning process.

3) Data Visualization

One of the crucial steps in choosing machine learning models is to visualize the dataset to identify visible data patterns that could impact model performance or require data pre-processing. The following plots were graphed for the RW9 battery dataset:

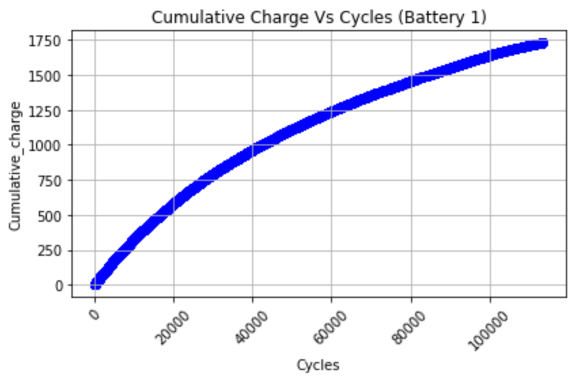

Fig (1) : Cumulative Charge (RW9)

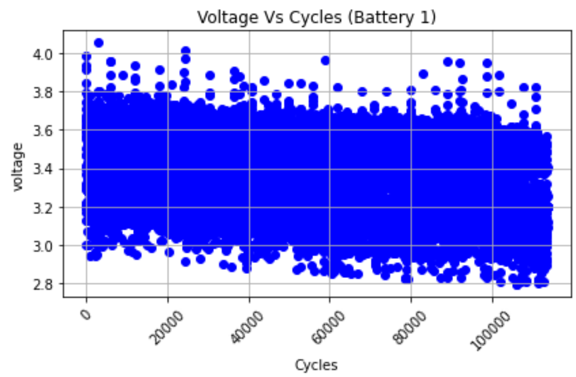

Fig (2) : Voltage (RW9)

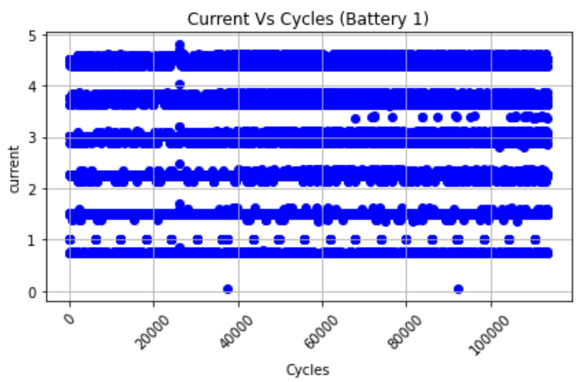

Fig (3) : Current (RW9)

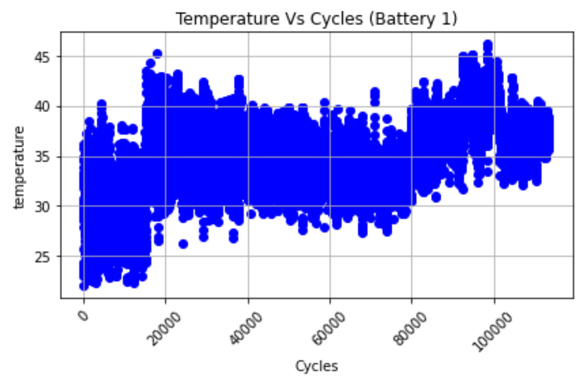

Fig (4) : Temprature (RW9)

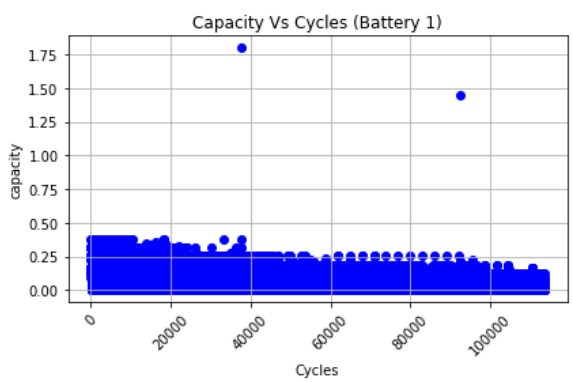

Fig (5) : Capacity (RW9)

The visual analysis of cumulative charge chart indicates that the battery was potentially charged for longer periods of time during the initial cycles compared to the other cycles or it could be indicative of degradation in the charging capacity of the battery itself. The voltage measured across the cycles has a broad range distribution. However, we do notice a gradual downward trend in the data. The current chart indicates randomness based on design parameter as anticipated. It can be inferred from the temperature chart that the average battery temperature sees an upward trend as it is cycled. The capacity data shows a range of capacity that is lower than the actual capacity of the battery. This was anticipated as the full discharge cycles were removed from the dataset. The assumption made for this analysis is that all batteries were performing at their 100% capacity at the beginning of the data collection. The graphs also indicate that the data is very noisy and should be pre-processed before implementing a machine learning model to reduce variability.

4) Data Pre-Processing and Analysis

In efforts to avoid overfitting the data and reduce the complexity of the dataset, a correlation matrix was generated to identify variables that were highly correlated:

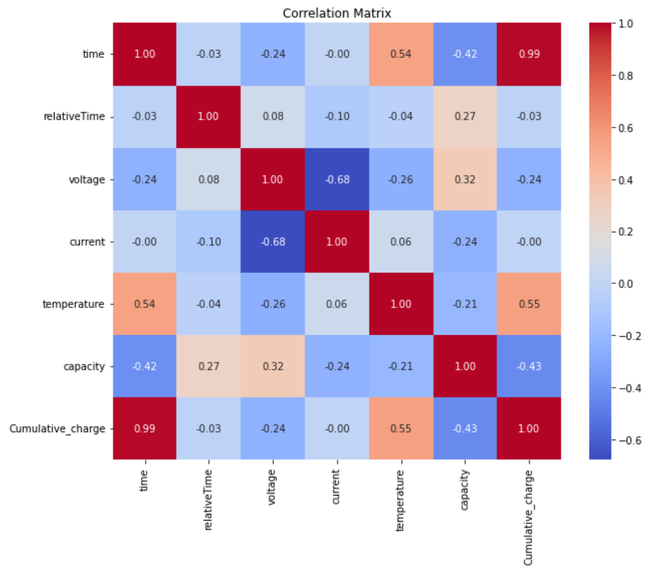

Fig (6) : Correlation Matrix for RW9 Dataset

The plot indicates there is heavy correlation between variables *time* and *Cumulative_charge*. Since time variable only provides insight into time elapsed from beginning of the experiment, this variable was removed from further analysis. Since capacity variable is a calculated field based on collected dataset, the correlation needed to be carefully examined. We notice from the two fields used to calculate the capacity parameter, *relativeTime* has a higher correlation with it. As a result, the decision was made to drop the variable *relativetime* from further analysis to avoid overfitting the data.

The visual analysis showed that the data had a lot of noise and variance. To reduce this noise and variability, EMD technique was used on the dataset. Empirical mode decomposition (EMD) stands as a mathematical technique for time domain decomposition, introduced by scientists N. E. Huang et al. in 1998 [16]. This method transforms a set of time series into distinct, locally narrow band components termed intrinsic mode functions (IMFs) [9]. This technique was applied to variables voltage, capacity and temperature. Since the current discharged was chosen as a design parameter and does not variate based on battery SOH, this was not smoothed using EMD analysis. The resulting residues from the EMD analysis are plotted the graphs below:

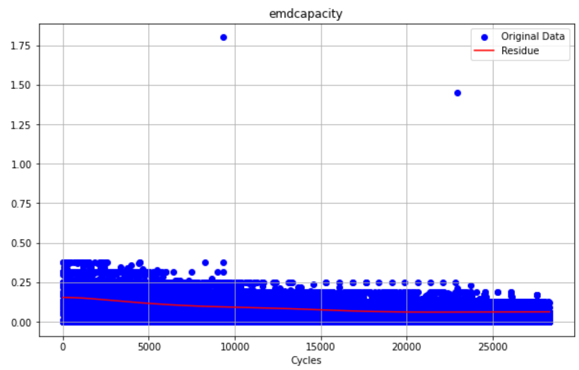

Fig (7): Capacity (RW9 EMD Implementation)

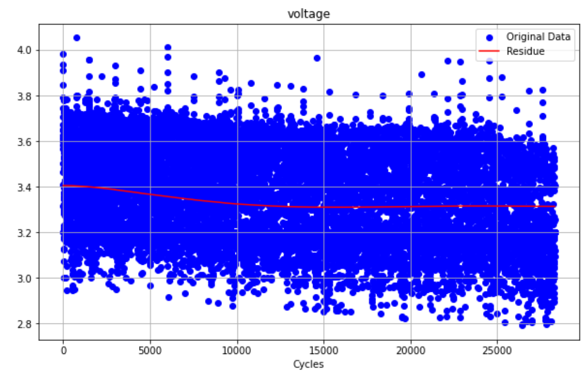

Fig (8): Voltage (RW9 EMD Implementation)

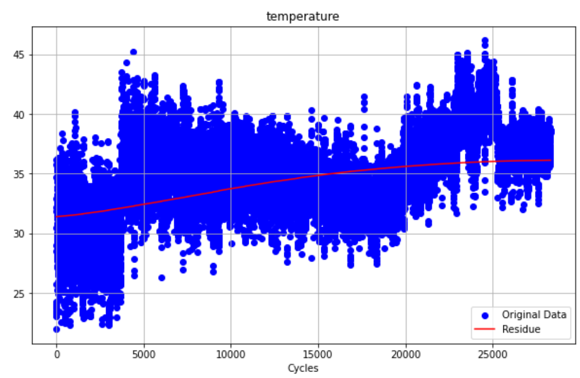

Fig (9): Temperature (RW9 EMD Implementation)

The visual representation of the residuals plotted against the original data shows that the EMD analysis was able to identify the data trends and reduce the variability in them. These residuals were used for further analysis.

Since our data is a time series collection, time series analysis was chosen to be one of the models. To identify data patterns that will enable in selecting the right time series model, the autocorrelation (ACF) and partial autocorrelation (PACF) graphs were plotted for the response variable (capacity).

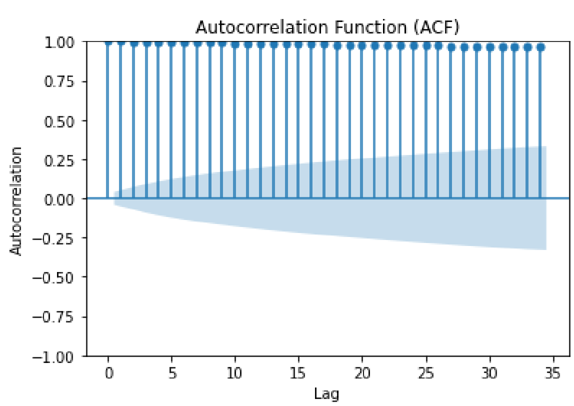

Fig (10) : ACF Plot (RW9 Capacity)

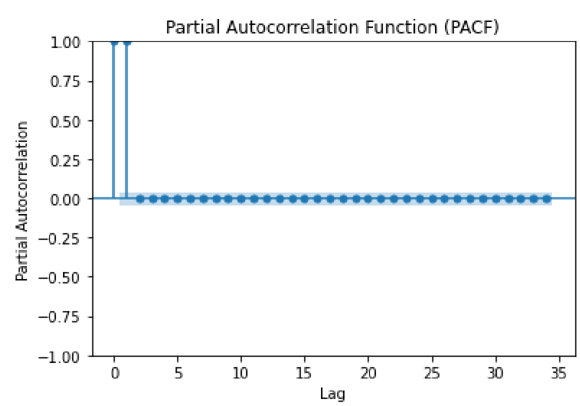

Fig (11) : PACF Plot (RW9 Capacity)

The Autocorrelation Function (ACF) plot suggests a persistent correlation between observations over time in the time series data, indicating a long-term dependency. Conversely, the Partial Autocorrelation Function (PACF) plot illustrates a rapid decline in correlation for higher lags, implying that direct correlation diminishes beyond the first lag. These observations align with the selection of the Autoregressive Integrated Moving Average (ARIMA) model, as it captures both the long-term dependencies highlighted by the ACF plot and the immediate correlations depicted by the PACF plot, making it suitable for modeling the time series data. The battery dataset is a multivariate dataset with univariate analysis required on capacity, as a result ARIMA with exogenous variables (ARIMAX) was selected.

Since long term dependency is critical, the other model selected for analysis based on the research completed in literature review was a CNN-BiLSTM model. CNN models are usually used for image processing, but they can also be used for numerical analysis [10]. However, since the battery's previous usage is more critical in its degradation curve, adding a weightage to the future values was deemed to be unnecessary. As a result, a CNN-LSTM model was selected where the bi-directional LSTM was removed.

5) Model Training and Testing

a) ARIMAX Model:

The first model selected for analysis was ARIMAX. The data was converted into a time series format using the *date* variables as the index with an hourly frequency. The data was then split into training and testing data using a 70/30 split. ARIMAX model requires the selection of three input parameters, viz. autoregressive parameter (p), degree of differencing (d) and moving average parameter (q). To select the best model inputs, parameter tuning logic was implemented for p in range(1,3), d in range(0,2) and q in range (1,3). The Akaike Information Criterion (AIC) was used to evaluate goodness of fit to balance model accuracy and simplicity. The resulting parameters were used to train a final ARIMAX model using the training data. The fitted model was then used to predict the battery capacity for three total battery datasets (including RW9).

Two approaches have been used in forecasting. In the first forecasting approach, the input variables (*voltage, current, temperature, Cumulative_charge, cycles)* were used from the readily available dataset. This is a direct approach that helps validate the model accuracy with a pre-existing dataset referenced as method (1). The second approach, referenced as method (2), was uniquely developed to identify the model fit for real life use cases. In industrial settings, the future battery parameters will not be available for inputs to predict future capacity. This data will only be available once the battery has reached that stage in its lifecycle, at which point the collected data itself can be used for capacity identification. This would render the prediction model unusable. To mitigate this risk, a logic was developed to use the most recently collected variables as the model input for next steps. This would enable businesses to use the battery's most recent state of health to forecast its remaining useful life. Both approaches were implemented for testing and validation.

b) CNN-LSTM Model:

The second model selected for implementation was CNN-LSTM. This neural network architecture leverages the strengths of both Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks for time series forecasting tasks. The features were first standardized using the standard scaler function to ensure each feature has a mean of 0 and a standard deviation of 1. They were then reshaped to be compatible with the CNN input requirements that include batch size, time steps and features. The data was then split into training and testing data using a 70/30 split. The model architecture begins with two sets of Conv1D layers followed by max pooling and dropout regularization that facilitate feature extraction from sequential data while mitigating overfitting. Subsequently, an LSTM layer is used to capture dependencies in forward direction. The LSTM layer is augmented with Rectified Linear Unit (ReLU) activation functions to introduce non-linearity and enhance the network's learning capacity. Finally, the architecture concludes with fully connected Dense layers, facilitating regression output. The hyperparameters for this architecture were optimized using a nested loop for filters, kernel size and LSTM values. The AIC was used to evaluate the goodness of fit to balance model accuracy and simplicity. The resulting parameters were used to train a final CNN-LSTM model using the training data. The fitted model was then used to predict the battery capacity for three total battery datasets.

Like the ARIMAX model, two approaches were used for model fit validation. One with input variables from the collected data for all battery sets and one with input variables chosen from available data from recent data collections. Both approaches were implemented with the assumption that 20% of the data will be collected at the time of forecasting.

6) Results:

The two approaches used in evaluation of the model fit were labelled as forecasting method (1) and forecasting method (2). Forecasting method (1) uses the collected variables from available dataset for forecasting and Forecasting method (2) uses the proposed approach of using the most recently collected variables as the model inputs for forecasting. The resulting model performances were as follows:

		Forecasting Method (1)			Forecasting Method (2)		
		MAE	MSE	RMSE	MAE	MSE	RMSE
RW9 (Training Data)	ARIMAX	0.5880%	0.0049%	0.6980%	0.6510%	0.0046%	0.6838%
	CNN-LSTM	0.1820%	0.0004%	0.2110%	2.0100%	0.0610%	2.4610%
RW10	ARIMAX	1.3270%	0.0287%	1.6930%	0.3050%	0.0011%	0.3344%
	CNN-LSTM	0.5186%	0.0494%	0.7030%	1.8620%	0.0604%	2.4580%
RW11	ARIMAX	5.7030%	0.3250%	5.7030%	6.0450%	0.3660%	6.0510%
	CNN-LSTM	0.4680%	0.0026%	0.5080%	2.3880%	0.0841%	2.9010%

Table (1) : Model Performance Comparison Table

The CNN-LSTM averages with a score of MAE: 0.3895%, MSE: 0.0175% and 0.4740% for method (1) and MAE: 2.0867%, MSE: 0.0685% and 2.6067% for model (2). In comparison, the ARIMAX averages with a score of MAE: 2.5393%, MSE: 0.01195% and 2.6980% for method (1) and MAE: 2.3337%, MSE: 0.1239% and 2.3564% for model (2). Across the examined series of batteries, the CNN-LSTM model demonstrates better accuracy over the ARIMAX model, as evidenced by lower Mean Absolute Error (MAE) and Mean Squared Error (MSE) in both Forecasting (1) and Forecasting (2) methods. The CNN-LSTM model does show high variability in forecasting method (2) as shown by the RMSE values. However, selecting the better model cannot be achieved by looking at the model performance metrics as they alone do not provide any insight into the useability of these models.

To get insight into the useability of these approaches, a visual representation of forecasted RUL capacity against the original RUL capacity was plotted for each battery. An example of the forecasted plot is provided below, and the other battery plots have been added to the Appendix for reference:

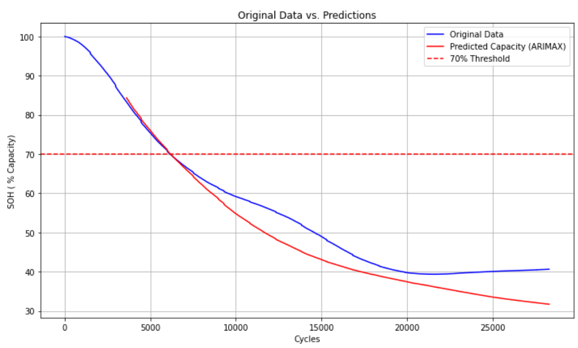

Fig (12) : ARIMAX Method 1 (RW9)

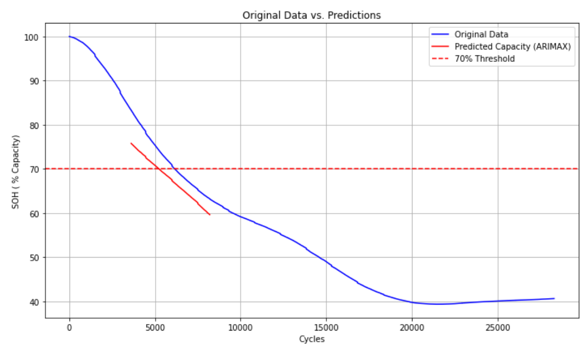

Fig (13) : ARIMAX Method 2 (RW9)

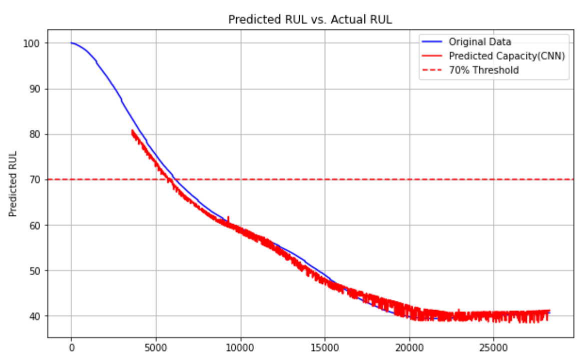

Fig (14) : CNN-LSTM Method 1 (RW9)

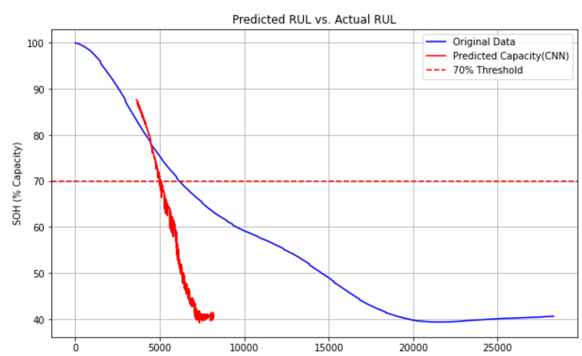

Fig (15) : CNN-LSTM Method 2 (RW9)

The visual representation shows the limitations of method (2) in comparison to method (1) for both ARIMAX and CNN-LSTM forecasting. Method (2) is limited to forecasting equal number of cycles to the available previous data. Therefore, method (2) is unlikely to predict when the battery will reach 70% threshold in the early stages of the battery life cycle. The model used in method (2) is also more accurate in the short-term forecasting compared to long term. Method (1) in comparison with method (2) provides better accuracy but less practicality as the upcoming variable information will not be available in a real-life situation.

For the forecasting on RW9 battery, it appears that the ARIMAX model provides better accuracy than CNN-LSTM model for method (2) based on visual representation and the performance values from Table (1). However, when the trained models from both approaches are used for forecasting on other battery datasets, the ARIMA model failed to converge to 70% through the entire life cycle of the battery RW11 [Appendix]. In comparison, the CNN-LSTM model offers better reliability between the two. The comparison table for RUL capacity on each battery dataset is below:

		Actual RUL	RUL Cycle (Method 1)	% Error (1)	RUL Cycle (Method 2)	% Error (2)
RW9 (Training Data)	**ARIMAX**	6163	5236	-15%	6190	0.44%
	CNN-LSTM	6163	5726	-7%	4988	-19.07%
RW10	**ARIMAX**	7609	8708	14%	8175	7.44%
	CNN-LSTM	7609	7100	-7%	5399	-29.04%
RW11	**ARIMAX**	5972	Failed to Forecast 70%	-	Failed to Forecast 70%	-
	CNN-LSTM	5972	4922	-18%	4626	-22.54%

Table (2) : RUL Capacity Forecast Error Table

The CNN-LSTM model shows higher accuracy than the ARIMAX model for method (1), however a lower accuracy is seen in method (2) from the CNN-LSTM model when ARIMAX model was able to converge to 70% capacity forecasting. In general, the ARIMAX model predicts an RUL capacity to be later than the recorded actual RUL of the batteries. The CNN-LSTM model in contrast predicts the RUL capacity earlier than the recorded RUL of the batteries.

Based on the results data, although ARIMAX shows more promising results in terms of accuracy than CNN-LSTM model in the short term, the reliability of the CNN-LSTM model would be preferred for a real-life use case. Moreover, the risk of overusing the batteries in the production environment which could potentially result in battery explosions due to high temperatures cannot be ignored. A more conservative model with ~70% accuracy might be better suited for the motivation of this project. However, this decision could change from a case-to-case basis based on business requirements.

Recommendations:

Other approaches should be explored before selecting a final model for implementation. This project evaluates the approach of using recently collected battery data to forecast RUL capacity. This was a pivotal decision that greatly impacted the results. As the results in Table (2) indicate, the CNN-LSTM model provides better accuracy in general to the ARIMAX model using method (1). This is also reflected in the model performance metrics from Table (1), that shows the CNN-LSTM to be more accurate on average for all datasets using method (1). This indicates the variable input selection in model (2) influenced the model performances. This performance can be improved by using alternative methods to first forecast the individual input variables using forecasting models like ARIMA for univariate data or exponential smoothing for univariate data on the variables voltage, temperature, and cumulative capacity. The correlation matrix [Fig .6] shows a moderately high negative correlation between voltage and current. This relationship can be used to predict the current values using predicted voltage values.

Another approach that can be taken is to collect more battery data and find the average input variable values at each cycle step. This will help provide more reliable historic data for the battery type and can be used as an input for future steps. A reinforcement learning (RL) model can then be used to make input variable changes in real time, based on future collected variable data. Both approaches have the potential to increase the model's accuracy but might come at a higher cost of implementation that may not be required for the business use case.

The finalized model can then be deployed to analyze battery health and RUL capacity for on-hand inventory. Using the forecast, the number of batteries that will reach the RUL capacity can be obtained. This will enable businesses to accurately predict the battery consumption for the upcoming cycles and place orders accordingly. The implementation of this logic can provide businesses with an opportunity to reduce their inventory costs and supply chain waste.

References:

[1] Loveridge, M., & Dowson, M. (2021). Why Batteries Fail and How to Improve Them: Understanding Degradation to Advance Lithium-Ion Battery Performance [Review of Why Batteries Fail and How to Improve Them: Understanding Degradation to Advance Lithium-Ion Battery Performance]. Faraday Insights, 10.

[2] Dunstan R. A. Smart Battery Providing Battery Life and Recharge Time Prediction. 1996. https://patents.google.com/patent/US5565759A/en.

[3] Rakhmatov D., Vrudhula S., Wallach D. A. Battery lifetime prediction for energy-aware computing. Proceedings of the International Symposium on Low Power Electronics and Design; August 2002; Monterey, CA, USA. IEEE

[4] Perone S. P., Spindler W. C. Battery lifetime prediction by pattern recognition. Application to lead-acid battery life-cycling test data. Journal of Power Sources. 1984;13(1):23–38. doi: 10.1016/0378-7753(84)80051-8.

[5] Ma D, Qin X. Residual Life Prediction of Lithium Batteries Based on Data Mining. Comput Intell Neurosci. 2022 Jun 13;2022:4520160. doi: 10.1155/2022/4520160. PMID: 35958783; PMCID: PMC9359824.

[6] Guangfeng Wang, Naxin Cui, Changlong Li, Zhongrui Cui, Haitao Yuan. A state-of-health estimation method based on incremental capacity analysis for Li-ion battery considering charging/discharging rate, Journal of Energy Storage, Volume 73, Part B, 2023,109010, SSN 2352-152X, https://doi.org/10.1016/j.est.2023.109010.

[7] Ma Yan, Shan Ce, Gao Jinwu, Chen Hong. A novel method for state of health estimation of lithium-ion batteries based on improved lstm and health indicators extraction. Energy, 251 (2022), Article 123973

[8] Zhuqing Wang, Ning Liu, Chilian Chen, Yangming Guo, Adaptive self-attention LSTM for RUL prediction of lithium-ion batteries, Information Sciences, Volume 635,2023,Pages 398-413,ISSN 0020-0255,https://doi.org/10.1016/j.ins.2023.01.100.

[9] Yapeng Zhou, Miaohua Huang, Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and ARIMA model, Microelectronics Reliability, Volume 65,2016,Pages 265-273,ISSN 0026-2714,https://doi.org/10.1016/j.microrel.2016.07.151.

[10] Xifeng Guo, Kaize Wang, Shu Yao, Guojiang Fu, Yi Ning, RUL prediction of lithium ion battery based on CEEMDAN-CNN BiLSTM model,Energy Reports, Volume 9, Supplement 10, 2023, Pages 1299-1306, ISSN 2352-4847, https://doi.org/10.1016/j.egyr.2023.05.121.

[11] Android Developers, BatteryManager API. https://developer.android.com/reference/android/os/BatteryManager

[12] Technologies, Z. (n.d.). Battery Intent API reference – TechDocs https://techdocs.zebra.com/emdk-for-android/7-0/guide/reference/refbatteryintent/

[13] Prognostics Center of Excellence Data Set Repository - NASA. (n.d.). NASA. https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/v

[14] Hassini, M., Redondo-Iglesias, E., & Venet, P. (2023). Lithium–Ion battery data: From production to prediction. Batteries, 9(7), 385. https://doi.org/10.3390/batteries9070385

[15] Stefanopoulou, A. G., & Kim, Y. (2015). System-level management of rechargeable lithium-ion batteries. In Elsevier eBooks (pp. 281–302). https://doi.org/10.1016/b978-1-78242-090-3.00010-9

[16] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society A Mathematical, Physical and Engineering Science, 454 (1998), pp. 903-995

[17] Li, X., Yu, D., Vilsen, S. B., & Store, D. I. (2023). The development of machine learning-based remaining useful life prediction for lithium-ion batteries. Journal of Energy Chemistry, 82, 103–121. https://doi.org/10.1016/j.jechem.

[18] Zhou, Y., & Huang, M. (2016). Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and ARIMA model. Microelectronics Reliability, 65, 265–273. https://doi.org/10.1016/j.microrel.2016.07.151

[19] https://chat.openai.com/

Appendix:

1) Battery RW11 Prediction Charts:

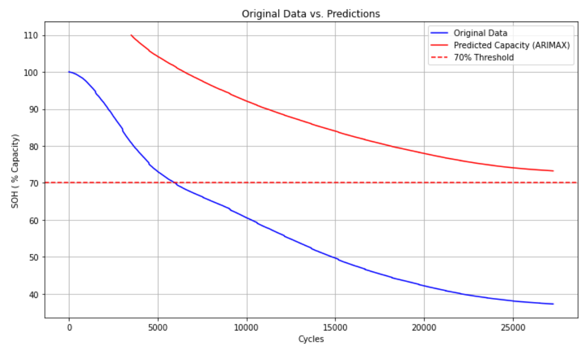

Fig (16) : ARIMAX Method 1 (RW11)

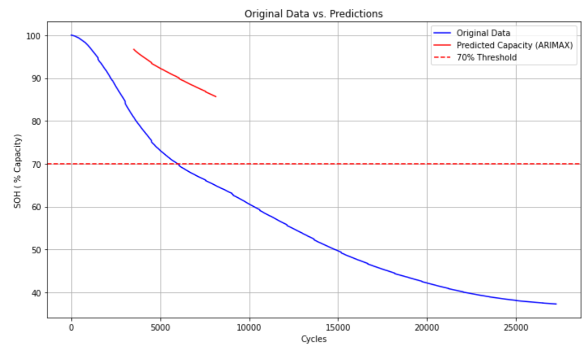

Fig (17) : ARIMAX Method 2 (RW11)

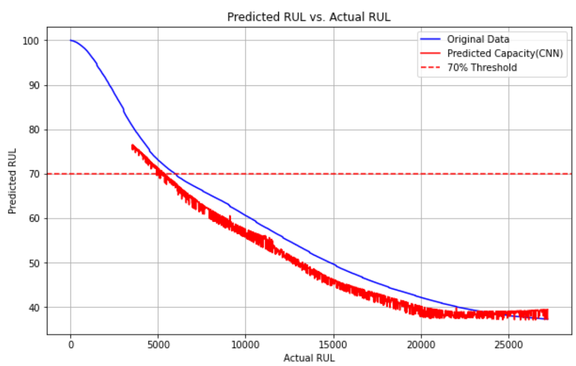

Fig (18) : CNN-SLTM Method 1 (RW11)

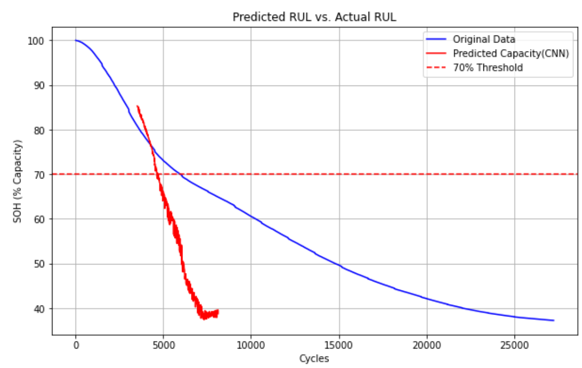

Fig (19) : CNN-LSTM Method 2 (RW11)

2) Battery RW10 Prediction Charts:

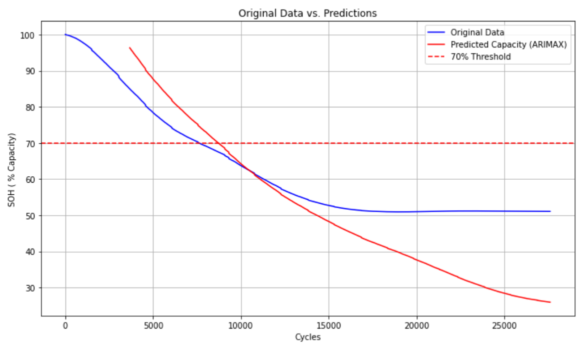

Fig (20) : ARIMAX Method 1 (RW10)

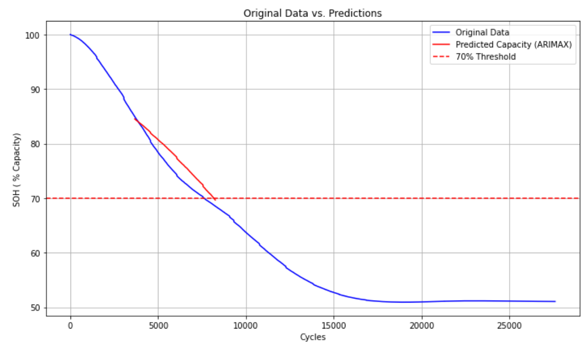

Fig (21) : ARIMAX Method 2 (RW10)

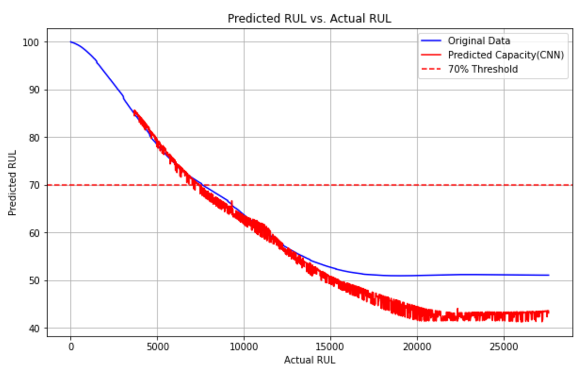

Fig (22) : CNN-SLTM Method 1 (RW11)

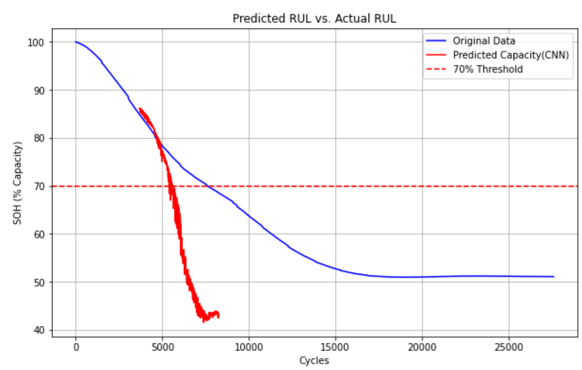

Fig (23) : CNN-LSTM Method 2 (RW11)