



Degree Project in Energy Technology
Second cycle, 30 credits

Development of AI-Based Data-Driven Aging Model for Li-Ion Batteries

ZHIFU SHANG

Master of Science Thesis
Department of Energy Technology
KTH 2024

Development of AI-Based Data-Driven Aging Model for Li-Ion Batteries

Zhifu Shang

Approved 2026.01.23	Examiner Farzin Golzar	Supervisor Farzin Golzar
TRITA: TRITA-ITM-EX 2026:11	Industrial Supervisor -	Contact person -



Development of AI-Based Data-Driven Aging Model for Li-Ion Batteries

Zhifu Shang

MSc Programme Environmental Pathways for Sustainable Energy Systems (120 credits)
Supervisor: Farzin Golzar
Examiner: Farzin Golzar
Department of Energy Technology
Swedish title: Utveckling av AI-baserad datadriven åldringsmodell för litiumjonbatterier

Abstract

With the increasing application of lithium batteries, especially battery storage systems (BESS), accurately predicting their remaining lifespan has become an important research topic in the energy field. Compared to traditional batteries, lithium batteries face more complex usage scenarios, especially with varying conditions during charging and discharging, requiring prediction models with greater generality and accuracy. In recent years, some studies have introduced neural networks into lithium battery prediction to more accurately simulate their charging and discharging behavior and characteristics. Machine learning methods, including MLPs and LSTM, have made initial progress in predicting two key parameters of lithium battery lifespan: SOH and RUL.

However, due to inherent limitations, existing models still have shortcomings in generalization ability and decision accuracy.

To address these issues, this study proposes a neural network-based Transformer model. Model variants were constructed by adjusting different hyperparameter combinations, and the performance of selected models was evaluated based on MAE, RMSE on a validation set. The Transformer model is compared with MLP and LSTM models, and the performance of different models is discussed.

The results show that the Transformer-based model significantly outperforms MLPs and LSTM models in both the overall and subgroup tests, validating the effectiveness of the proposed method. In summary, this study provides a new approach to Transformer modeling for lithium-ion battery BESS applications and compares the application potential of different modeling methods.

Keywords

Li-ion Battery Aging Model, Battery Energy Storage System, MLPs, LSTM, Transformer, State of Health, Remaining Useful Life

Sammanfattning

Med den ökande användningen av litiumbatterier, särskilt batterilagringssystem (BESS), har det blivit ett viktigt forskningsämne inom energiområdet att noggrant förutsäga deras återstående livslängd. Jämfört med traditionella batterier står litiumbatterier inför mer komplexa användningsscenarier, särskilt med varierande förhållanden under laddning och urladdning, vilket kräver prediktionsmodeller med större generalitet och noggrannhet. Under senare år har vissa studier introducerat neurala nätverk i litiumbatteriprediktion för att mer exakt simulera deras laddnings- och urladdningsbeteende och egenskaper. Maskininlärningsmetoder, inklusive MLP och LSTM, har gjort inledande framsteg när det gäller att förutsäga två viktiga parametrar för litiumbatteriers livslängd: SOH och RUL.

På grund av inneboende begränsningar har dock befintliga modeller fortfarande brister i generaliseringsförmåga och beslutsnoggrannhet.

För att ta itu med dessa problem föreslår denna studie en transformatormodell baserad på neurala nätverk. Modellvarianter konstruerades genom att justera olika hyperparameterkombinationer, och prestandan för utvalda modeller utvärderades baserat på MAE, RMSE på en valideringsuppsättning. Transformatormodellen jämförs med MLP- och LSTM-modeller, och prestandan för olika modeller diskuteras.

Resultaten visar att den transformatorbaserade modellen avsevärt överträffar MLP- och LSTM-modellerna i både övergripande och subgruppstester, vilket validerar effektiviteten hos den föreslagna metoden. Sammanfattningsvis tillhandahåller denna studie en ny metod för transformatormodellering för BESS-tillämpningar med litiumjonbatterier och jämför tillämpningspotentialen hos olika modelleringsmetoder.

Nyckelord

Li-jonbatteriets Åldringsmodell, Batterilagringssystem, MLP, LSTM, Transformator, Hälsotillstånd, Återstående Livslängd

Acknowledgments

Here, I would like to express my deepest gratitude to my supervisor, Dr. Farzin Golzar, for his invaluable support and guidance throughout the entire thesis writing process. He provided me with the initial idea for this thesis and guidance on research methods, which greatly facilitated my research. Dr. Golzar is very responsible towards his students; he organized bi-weekly meetings to help me solve problems and ensure the smooth progress of the research. He also allowed me to adjust the research pace at my own rhythm, for which I am truly grateful. From choosing the research topic and developing the research plan to revising the thesis draft, his insightful feedback and suggestions were crucial in completing this work. I sincerely thank him for his kindness and unwavering support.

Stockholm, November 2025
Zhifu Shang

Contents

Abstract	i
Sammanfattning	ii
Acknowledgments	iii
List of Acronyms and Abbreviations	vii
List of Acronyms and Abbreviations	vii
List of Tables	viii
Introduction	1
Background	4
1 Preliminaries	4
1.1 State of health of Li-ion batteries	4
1.2 End of life of Li-ion batteries	4
1.3 Remaining Useful lifetime of Li-ion batteries	5
2 Problem statement	5
3 Objective	6
4 Previous work	7
4.1 Comparison between physics-based methods and machine learning-based methods	7
4.2 DNN	9
4.3 RNN	11
Methodology	13

5	Dataset	13
5.1	Discription of Dataset.....	14
5.2	Pre-process	19
6	Model description	20
6.1	MLP.....	21
6.2	LSTM	25
6.3	Transformer	28
7	Model Performance Evaluation.....	33
Results		35
8	Model Performance	35
8.1	Prediction and Result Analysis of MLP Model.....	35
8.2	Prediction and Result Analysis of LSTM Model	35
8.3	Prediction and Result Analysis of Transformer Model	37
8.4	Comparative Analysis of Prediction Performance.....	40
9	Sensitivity Analysis.....	41
9.1	Method of Sensitivity Analysis	42
9.2	Result of Sensitivity Analysis.....	42
9.3	Summary of Sensitivity Analysis	46
10	Leave-One-Out Cross-Validation Analysis	46
Discussion		49
Conclusion		51
10.1	Limitations.....	51
Future Work		52
Bibliography		57
Appendix		58

List of Acronyms and Abbreviations

CNN Convolutional Neural Network. 8, 12

DL Deep learning. 12

DNN Deep neural networks. 9, 10, 12

EOL End of Life. 1, 4, 5, 20

FFNNs Feedforward Neural Networks. 11

LIBs Low-voltage batteries LIBs. 9, 12

LSTM Long Short-Term Memory. 2, 3, 11–13, 25–27

MAE Mean absolute error. 11, 33

ML Machine Learning. 9, 19

MLP Multilayer Perceptron. 3, 13

RMSE Root mean square error. 11, 33

RNN Recurrent Neural Network. 25

RUL Remaining Useful Life. ii, iii, 1–3, 5, 6, 9, 10, 12, 13, 19, 20, 23, 27, 33

SOC State of Charge. 6, 7, 9, 10

SOH State of Health. ii, 2–5, 9–12, 19, 20

List of Tables

1	Publication trends of ML-based methods in BMS estimations (past 10 years, Web of Science)	8
3	Prediction performance (MAE and RMSE) of the MLP model across four datasets.....	36
4	Prediction performance (MAE and RMSE) of the LSTM model across four datasets.....	37
5	Prediction performance (MAE and RMSE) of the Transformer model across four datasets.	39
6	Comparison of MAE and RMSE for SOH/RUL Prediction across Four Datasets	40
7	Performance Metrics of Leave-One-Out Cross-Validation on NASA Datasets	47
8	Comparison with State-of-the-Art Cross-Battery Models.....	48

Introduction

In Battery Energy Storage Systems (BESS), lithium-ion batteries are widely chosen as power supply modules. Accurate and continuous monitoring of the remaining useful life (RUL) of lithium-ion batteries is crucial during electric vehicle operation. Battery performance gradually deteriorates with capacity degradation, making RUL a key indicator for evaluating battery aging. According to the IEEE 1188-1996 standard, it is recommended to replace a lithium-ion battery module once its capacity drops to 80%. Real-time monitoring of the health status of lithium-ion batteries and predicting their remaining useful life has always been a significant challenge in addressing availability and ensuring battery reliability and safety. The capacity degradation of commercial lithium-ion batteries exhibits a strong linear relationship in the initial stages, but shows a highly non-linear trend as it approaches EOL. In addition, various factors such as charging and discharging strategies, ambient temperature, positive and negative electrode materials, and battery capacity regeneration also lead to different trends in capacity degradation. [32, 37]. This makes predicting capacity declines complex and challenging. Accurate lifespan prediction is crucial to avoiding overcharging and over-discharging, as miscalculations of remaining usable capacity can lead to safety issues such as explosions and fires in cell phone batteries and electric vehicles. In summary, for lithium-ion batteries in energy storage systems, accurate monitoring and prediction of their remaining useful life is a crucial aspect of ensuring system controllability and safety, and is also an urgent need for promoting the development of battery technology. In this paper, battery capacity is chosen as the standard for measuring the lifespan of lithium-ion batteries; when the remaining capacity of the lithium-ion battery drops to 80% of its rated capacity, the measured lithium-ion battery is considered to have reached the failure threshold. Currently, there are two main methods for predicting the lifespan of lithium-ion batteries. The first method is the model-based method, which involves establishing a physical or mathematical model, and then finding an equivalent physical circuit model based on experience. An RUL model is then established by adjusting parameters such as resistance and current in the circuit [53]. Model-based lithium-ion battery RUL prediction methods require accurate predictions under relatively stable conditions, as changes in external conditions can easily affect the accuracy of the model, making it difficult to obtain an accurate mechanistic model. Although model-based methods provide a theoretically more rigorous way to predict RUL, they face some challenges in practical applications. For example, changes in the external environment can lead to inaccuracies in model parameters, and the complex battery behavior and chemical reactions also make modeling quite difficult. Therefore, in practical applications, it is necessary to continuously improve and optimize the model to improve its accuracy and robustness. This study will explore and optimize model-based RUL prediction methods to improve their applicability and reliability. In addition to

model-based methods, data-driven methods are also one of the main methods in the field of lithium-ion battery RUL prediction. Unlike model-based methods, data-driven methods do not focus on the internal failure mechanisms of lithium-ion batteries, but instead utilize adaptive models, which create approximate models using existing data to map the relationship between input and output data. This method does not require accurately learning the physical or mathematical model of the object under test, but rather relies on the analysis and pattern recognition of a large amount of data. Data-driven methods include various techniques, such as machine learning and deep learning. Among them, methods such as neural networks and random forests have been widely used in lithium-ion battery RUL prediction [16, 26]. These methods can automatically capture complex nonlinear relationships through a large amount of experimental data, thus achieving relatively accurate RUL prediction in the absence of an accurate physical model. Another method is the fusion model method, which combines model-based and data-driven methods. Complex methods such as neural networks and random forests are introduced to improve the applicability and robustness of the model. However, the network structure of these methods may be relatively redundant, leading to certain limitations in practical applications. In summary, data-driven methods have become the main methods for nonlinear system performance degradation analysis and lithium-ion battery RUL prediction problems, improving the accuracy of battery life prediction by utilizing a large amount of experimental data. This research aims to predict the lifespan of lithium-ion batteries based on machine learning, by thoroughly discussing the relationship between the aging mechanism of the battery charging process and directly measurable variables, and using models including deep learning and neural networks to identify the aging patterns of lithium-ion batteries as they cycle and establish an aging model to achieve accurate prediction of the remaining useful life of lithium-ion batteries. The goal of this research is to achieve an accurate prediction of the remaining useful life of lithium-ion batteries, thereby enhancing the reliability of the Battery Management System (BMS) and contributing to a comprehensive battery health management and failure warning system. Through this research, it is expected to provide a new and effective method for lithium-ion battery life prediction, providing strong support for the further development and application of battery technology.

In this study, we explored the performance of different models based on existing databases. Since a single model often struggles to generalize to battery datasets with different charging patterns and is prone to underfitting or insufficient generalization capabilities, we aimed to compare various models and employ ensemble learning methods to enhance the model's generalization ability and adaptability. Specifically, we used three algorithms for model fusion: an MLP for initial data processing; an LSTM, suitable for long-term memory; and finally, the latest transformer for final fitting and modeling. Due to time and resource limitations, this study focuses only on exploring the accuracy of predicting battery SOH and RUL in specific charging cycles. Other aspects, such as predicting battery life based on DOD (depth of discharge), are beyond the scope of this current study and will be left for future research. In the evaluation phase, we compared the model results of different fusion methods and compared them with the original baseline model. The results show that the ensemble model using stacked fusion outperforms the baseline model in terms of prediction accuracy.

To summarize, the main contributions of the study are as follows:

1. We designed and validated a transformer-based prediction method, which effectively improved the prediction accuracy of battery SOH and RUL in specific charging cycles;
2. We compared the performance of the transformer model with baseline method(LSTM and MLP), which proves the former effectively improved the prediction accuracy of battery SOH and RUL in specific charging cycles;
3. We expanded the application potential of data-driven methods in battery prediction, providing new ideas for enhancing the generalization ability of battery prediction under different charging behaviors.
4. We expanded the application potential of data-driven methods in battery prediction, providing new ideas for enhancing the generalization ability of battery prediction under different charging behaviors.
5. We developed a web-based platform for building battery SOH and RUL prediction models. Besides testing with the databases described in this paper, it also allows uploading other databases. This facilitates further research by other interested individuals building upon this work.

Background

1 Preliminaries

This section illustrates a basic introduction to the aging-related concepts of Li-ion batteries and structured models.

1.1 State of health of Li-ion batteries

Lithium-ion batteries are used in a variety of applications due to their high-performance energy density and cycle life. These batteries are particularly useful in portable electronics, electric vehicles, and renewable energy storage systems. The degradation of the battery's performance is often indicated by its SOH, which decreases over time as a result of various factors related to the batteries themselves or the ambient conditions. The health of a battery is determined by its capacity for delivering the rated power output, as compared to its original state. The SOH of a Li-ion battery, SOH_k , is represented as a percentage of the battery's health and is defined as follows:

$$SOH_k = \frac{Q_k}{Q_{initial}} \cdot 100 \quad (1)$$

Where

Q_k represents the current capacity of the battery at cycle k , $Q_{initial}$ represents the initial capacity.

As the batteries are consumed and rapidly charged and discharged, its SOH decreases by each cycle. And it can be observed from the measured voltage, current, and temperature profiles with time series.

1.2 End of life of Li-ion batteries

The EOL of a battery is defined as the point at which the battery can no longer deliver the rated capacity and power, marking the end of its useful life. EOL of a battery is

typically determined by the SOH dropping below a certain edge, generally defined as 80% of the rated capacity. It is crucial to recognize that, owing to recuperative effects, a battery's SOH may rise once more, exceeding the EOL threshold on numerous occasions.

Within the present study, the EOL indicator is defined as the initial cycle following the final instance when the SOH falls beneath the threshold. As noted earlier, the aging process of lithium-ion batteries is affected by both internal and external factors. Internal factors pertain to the chemical properties of the battery, whereas external factors encompass manufacturing procedures, environmental circumstances, and battery utilization, to name a few.

1.3 Remaining Useful lifetime of Li-ion batteries

The remaining useful lifetime (RUL) is defined as the amount of time in a certain operation before the battery reaches the predefined EOL capacity; it represents the period from the observed condition to EOL.

Since the RUL is the most direct and concerned

2 Problem statement

Despite significant progress in battery SOH and RUL prediction, the continuous advancement of the energy revolution will lead to broader and more rapid development of batteries as a crucial portable energy source under sustainable development strategies. Although data-driven methods for online prediction, anomaly detection, and fault warning of battery health status are considered one of the most promising approaches to achieving these goals, related research is still in its early stages. Several major research challenges are presented below:

(1) Real-time application of prediction models. The online application of battery state prediction models faces several challenges. Current physical models typically require high-precision measuring instruments and the solution of a large number of differential equations, limiting their application in rapid battery diagnosis and anomaly detection. In contrast, data-driven methods significantly shorten prediction time while providing sufficient accuracy, making them the most promising method for online prediction. However, current data-driven models are mainly built based on constant current and constant voltage charging and discharging strategies in the laboratory, which differ from the dynamic operating environment and equipment condition changes in actual battery use. These conditions are key factors affecting the nonlinear degradation of batteries. In addition, actual battery data collection suffers from missing values and noise, and may even be affected in extreme environments (such as extremely high/low temperatures). Therefore, further research is needed to successfully apply laboratory-built models to actual online estimation. Establishing a big data platform is crucial for online model application. By training on actual battery operation data, it is possible to build more practical and ef-

fective application models and explore optimal charging and discharging strategies to maximize battery performance and achieve energy savings.

(2) Combining machine learning prediction models with physical models. Combining physical models and machine learning models in the field of battery state prediction is a current research focus. Although single physical models based on battery operating mechanisms and machine learning models based on cycle data each have their advantages, the former suffers from shortcomings in accuracy and online application, while the latter performs poorly in interpretability, making it difficult to apply to scenarios such as battery performance maintenance [4]. Therefore, one of the current research directions is to construct a prediction framework that combines physical models and machine learning [30, 2]. This combination can be achieved in various ways, including post-fusion of the prediction results of physical and machine learning models, separate predictions under different weights, using physical models to process electrochemical information and outputting prediction results through machine learning models, and selecting machine learning model parameters based on physical models. Overall, improving the prediction accuracy and interpretability of models by fully leveraging the advantages of both types of models is one of the pressing problems that needs to be solved.

(3) Early prediction of battery degradation and analysis of its uncertainty. Early prediction of batteries is of great significance in the production, performance maintenance, recycling, and secondary utilization processes of batteries. On the one hand, this can effectively reduce costs by quickly obtaining the future health status and degradation curve of the battery without conducting large-scale cycle experiments. On the other hand, early prediction of batteries helps to quickly analyze the safety of batteries in the early stages, prevent potential safety accidents, and further improve battery performance and effectively control product quality. However, due to the nonlinear degradation and capacity regeneration phenomena of batteries, the accuracy and reliability of early prediction need to be further quantified, that is, in-depth research on uncertainty is required. Currently, early prediction of batteries and related uncertainty research are still in the initial stage, and the methods used to construct early prediction models have significant limitations in terms of accuracy and generalization performance [34].

3 Objective

The goal of this thesis is to create an AI-based aging model for Li-ion batteries for the energy management of PV system integrating on KTH-Lab building. The objective is to estimate the RUL and EOL on Li-ion batteries cell level, using reliable SOC as well as ambient factors. The model will provide insights into battery degradation, helping optimize operation strategies.

4 Previous work

4.1 Comparison between physics-based methods and machine learning-based methods

As mentioned earlier, models for predicting the health status of lithium-ion batteries can be broadly categorized into two types: physics-based methods and machine learning-based methods. Machine learning-based methods utilize various models trained on a set of input-output data from lithium-ion battery cells. These models predict the output state based on specific input conditions. These models require different laboratory conditions, and therefore the input and output data will vary for different specific BMS estimation tasks. Typical input data for these methods includes basic parameters such as voltage, current, temperature, time, and charge/discharge cycle counts during charging/discharging, as well as indirect measurements such as cell internal resistance and state of charge (SOC). Data collection frameworks are used to record this information, and additional data can be generated based on battery or battery system models. Physics-based methods, on the other hand, rely on the derivation and empirical formulas of battery models of electrochemical kinetics, where a significant portion of the required parameters are theoretical assumptions. Given the complexity of real-world applications, the predictions of these models in practical applications are highly dependent on the robustness of the cell's chemical properties and uncontrollable factors such as the battery pack's interconnection methods. In contrast, machine learning-based methods do not rely on any specific assumptions about the battery model. Conversely, they establish black-box relationships between input and output data through training, without explicitly considering battery physics models.

From a development time perspective, model-based methods are more time-consuming than data-driven methods. This is because model-based methods require a comprehensive understanding of the underlying physics and chemistry of the battery system and are related to factors relevant to battery production, making this both difficult and time-consuming. In contrast, data-driven methods, based on a large amount of existing data, can more efficiently shorten development cycles using machine learning algorithms, making them more suitable for large companies with substantial data accumulation. Therefore, the choice of method depends on various factors, including data quality, required accuracy, and the specificity of the application environment.

The advantages and disadvantages of the two methods are shown in Table 1. In recent years, research based on machine learning has become increasingly popular, especially since machine learning algorithms have higher usability than model-based methods, leading to their greater favor among researchers. Figure 1 shows the classification system of research related to battery life prediction in the literature, illustrating commonly used methods for cell life prediction from both machine learning and model-based perspectives.

The current review of cell state prediction covers a variety of technical paths, including data-driven methods, model-driven strategies, or a combination of both. Reference [5] provides a comprehensive review of research based on models and machine learning, aiming to ensure the safe operation and optimized utilization of battery systems. Its core

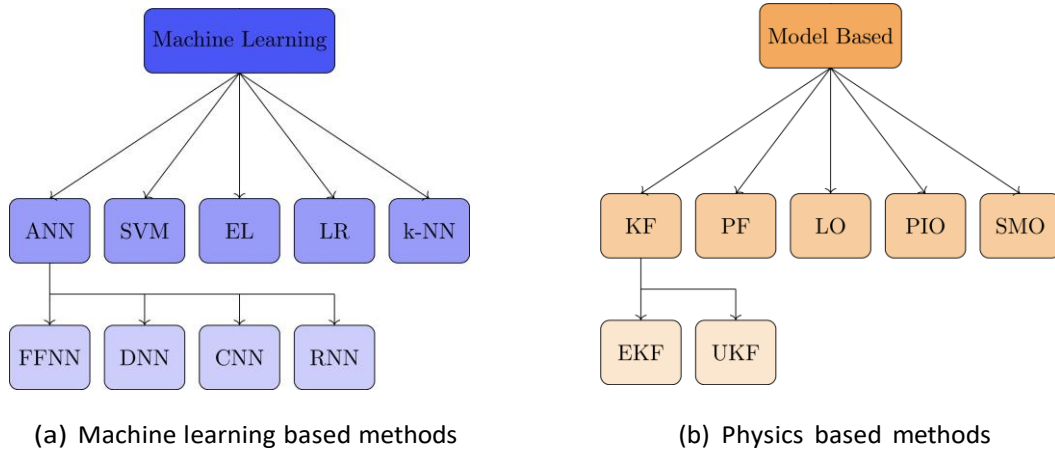


Figure 1: ANN: Artificial Neural Network, FFNN: Feed Forward Neural Network, DNN: Deep Neural Network, CNN: Convolutional Neural Network, RNN: Recurrent Neural Network, SVM: Support Vector Machine, EL: Ensemble learning, LR: Linear Regression, k-NN: k-Nearest Neighbors; KF: Kalman Filter, EKF: Extended Kalman Filter, UKF: Uncented Kalman Filter, PF: Particle Filter, LO: Luenberger Observer, PIO: Proportional Integral Observer, SMO: Sliding Mode Observer

Method	Advantages	Disadvantages
Model-based	<ul style="list-style-type: none"> Reliable and accurate Universal validity 	<ul style="list-style-type: none"> Extensive domain knowledge Since it requires extensive domain knowledge, which can be time-consuming and challenging to achieve
Data-driven	<ul style="list-style-type: none"> Does not require extensive domain knowledge Rapidly learns from large datasets, more efficient in development time 	<ul style="list-style-type: none"> Unpredictable black box model Requires large high-quality datasets (time-consuming to obtain)

Table 1: Publication trends of ML-based methods in BMS estimations (past 10 years, Web of Science)

objective is to highlight the increasing attention given to data-driven methods in the field of battery health management. The research of [52] focuses on the prediction and health management (PHM) direction of LIBs, incorporates deep learning algorithms into the ML model system, and elaborates on the three core links of deep learning-based LIBs PHM technology: data acquisition, deep learning method selection, and performance verification. The review of [24] specifically analyzes the advantages and disadvantages of model-based and machine learning-based SOH estimation and RUL prediction technologies. This study clearly points out that data-driven battery health prediction methods are gradually becoming a hot topic in the industry. In contrast, the work of [7] only compares and analyzes the two types of research based on models and machine learning from four dimensions: technical methods, advantages and disadvantages, limitations and errors. [14] focuses on machine learning-driven prediction models, summarizing relevant research results into two categories: one focuses on exploring new technical methods, and the other focuses on quantitative comparison of prediction accuracy and model complexity. In the review on specific state estimation, [1] systematically sorted out the research on SOC estimation, divided the prediction techniques used into four categories: interpretable, indicative, restrictive, and error-related, and gave an overall evaluation. The study of [41] specifically integrated the relevant results on SOH estimation. Under the above research background, this literature made a horizontal comparison of 5 different machine learning methods. In general, the existing review studies either focus on single state estimation techniques or cover multi-dimensional prediction methods, all of which reflect the development trend of multi-technology integration in the field of battery health management.

4.2 DNN

Deep learning (DL), as a core branch of machine learning (ML), is based on artificial neural networks (ANNs). The core definition of "deep" is that the network architecture contains multiple hidden layers [13]. In the age prediction of Li-ion battery, deep neural networks (DNNs), convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are three mainstream DL algorithms. They adapt to different assessment needs with their different network topologies. Among them, DNN are essentially fully connected neural networks (FFNNs) with multiple hidden layers. One or more activation functions can be embedded in the process of information transmission to the hidden layers, making them an efficient tool for mining complex features in data. [20] A dual DNN model architecture is used to estimate the remaining useful life (RUL) of LIBs. One DNN model performs statistical analysis on impedance decay related to capacity decay caused by increased distortion, while the other DNN model focuses on predicting the capacity decay trend. The accuracy of RUL estimation is improved by the collaboration of the two models. [83] It is pointed out that accurate feature indicators are the key to accurate evaluation of battery performance. This study abandons the discharge voltage parameter commonly used in early SOH calculations and instead uses the first derivative of the discharge curve, dQ/dV , as the core evaluation parameter, and uses DNN as the parameter optimization algorithm. [46] In view of the SOH evaluation needs of secondary utilization batteries, a new estimation strategy based on DNN is proposed, and a Markov chain error correction mechanism is introduced to effectively improve the prediction accuracy of the model.

[12] A new DNN architecture with a fusion convolution training strategy is designed, which is specifically applied to battery RUL estimation. Its single-cycle error rate is only 6.46%. [11] An end-to-end deep learning system based on DNN is developed. This system can complete the RUL estimation of LIBs with only short-term measured data. [50] By mining the complex mapping relationship between SOH feature indicators and actual SOH values through DNN, several DNN-based SOH estimators were successfully developed.

[33] Although the proposed technical solution focuses on capacity estimation under SOC and 2C rates as the core application scenario, the kernel function of the DNN was optimized and modified, so that the battery SOH evaluation can maintain high accuracy under different initial SOC and C rate current conditions. [28] A DNN-based transfer learning-based RUL estimation technique was proposed. This method first selects the reference battery with the closest performance to the target battery from the historical test dataset by means of the average Euclidean distance, and then uses the selected dataset to train an estimation model based on stacked noise reduction autoencoder (SDA), and finally achieves accurate prediction of the target battery's RUL. [48] An innovative DNN network structure was designed. By introducing a specific polynomial function, the computational complexity was simplified while ensuring the evaluation performance of the algorithm.

[21] A DNN state classifier was used to diagnose and analyze the fault state of individual battery cells, thereby indirectly achieving accurate prediction of the battery's health state. [18] Prediction and health management (PHM) is emphasized as the core technology to ensure the reliable and safe operation of the battery in SOH and RUL prediction. This study introduces the preliminary development results of the PHM system based on the DNN algorithm. [38] A hybrid estimation method was proposed, which combines the extended Kalman filter technique based on the circuit model with the data-driven fully connected DNN, specifically for the accurate estimation of the battery's SOH. [40] Based on one year of real-world data from 700 unique vehicles with different driving modes, a DNN model was constructed to achieve intelligent prediction of battery SOH.

[17] Using the public database of the NASA Ames Center of Excellence for Prediction (PCoE), the performance correlation between DNN and other ML methods such as Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), ANNs, and Logistic Regression (LR) in the SOH estimation task was explored.

[44] Unlike traditional research paths, the battery aging process was evaluated by combining constant-condition simulation with hardware platform testing. A DNN based on an ML framework was used to predict SOH, and a multi-dimensional battery statistical data set was constructed.

[19] A DNN algorithm for SOH estimation was designed, dividing the SOH level into five gradient categories within the range of 100% to 80% to adapt to the evaluation needs of different aging stages. The aforementioned research covers multiple applications of DNN in battery SOH/RUL estimation, fault diagnosis, and PHM system development. It includes both optimization and innovation of single models and integration of advanced technologies such as hybrid architecture and transfer learning, fully demonstrating the flexibility and efficiency of DNN in the field of battery state assessment.

4.3 RNN

Recurrent Neural Networks (RNNs) are a type of neural network derived from FFNNs. Their core advantage lies in their memory characteristics—the ability to retain the computational data of previous hidden layers and use it as the input for the current hidden layer [41]. In layman’s terms, the model’s estimation of the current state needs to combine real-time input information with historical state data. It is this innovative design of memory characteristics that makes RNNs an efficient model for processing sequential inputs [36]. However, when dealing with long-term time series, RNNs face the problem of long-term dependency, which may lead to the vanishing gradient phenomenon. To solve this problem, the academic community has proposed the LSTM scheme with a controlled storage mechanism [36].

[43] proposed a SOH estimation scheme based on historical operating data of lithium-ion batteries. This scheme uses Independent Recurrent Neural Networks (IndRNNs) to capture the complex nonlinear characteristics of batteries and solves the gradient dependency problem, enabling the network to learn the long-term correlation law of capacity decay. Experimental results show that this IndRNN-based model performs reliably in SOH prediction, with a root mean square error (RMSE) as low as 1.14%, a mean absolute error (MAE) of 1.33%, and a maximum error (MAX) of only 2.5943% during testing. These values are far below the $\pm 5\%$ SOH error threshold allowed in the electric vehicle field, providing a precise and reliable SOH estimation technique for lithium-ion batteries (LIBs).

[23] To meet the SOH prediction requirements, the network structure of LSTM was optimized by modifying the connection paths between gates: the enhanced LSTM filters fresh information and determines old information in the input and forget gates, respectively, while feeding back past cell states to the input and output gates. This preserves key effective information while reducing redundant error signals. Test results on the NASA public dataset show an average RMSE of only 2.16

[6] A multi-channel SOH estimation scheme based on voltage (V), current (I), and temperature (T) curves was designed. Compared to the conventional method based solely on voltage signals, this scheme reduced the estimation error by 25%. In this study, the LSTM network is used to efficiently combine degradation data from multiple battery cells, which significantly improved the accuracy and stability of the SOH estimation of battery packs. This conclusion was verified in battery pack tests with series, parallel, and hybrid series-parallel architectures [45].

[49]’s research optimized the input features of the LSTM model by combining grey relational analysis (GRA) with entropy weighting, and then trained the LSTM, achieving an accurate mapping between the improved features and the SOH estimation results. [47] The k-means clustering algorithm is used to divide the voltage/current (V/I) data points into 30 sub-regions, which can effectively characterize the aging state of the battery. Based on this clustering result, the LSTM model only needs a small number of V/I samples collected in a short time to estimate the SOH through the density distribution of the sub-regions. This technique shows high accuracy, strong flexibility and noise resistance, and the average error rate is lower than the set threshold.

[51] A new path for predicting the lifespan of lithium-ion batteries based on deep learning is proposed. Its core is to use a high-efficiency LSTM network variant—this variant can adapt to data with different input dimensions and can be trained using more labeled samples. In addition, the study also designed an online verification mechanism, which can dynamically adjust the model parameters according to the real-time available data of the system to achieve continuous optimization of model performance. This scheme has the potential for online deployment and is suitable for real-time application scenarios.

[15] It is pointed out that in addition to expanding the diversity of input signals, sparse sampling technology is also an effective means to improve model performance. Sparse data can concisely characterize the core features of a battery, while optimization algorithms can leverage the sparsity of signals to achieve signal reconstruction with less data without losing key information; by rationally designing the network topology, the technical advantages of the LSTM algorithm can be fully utilized, simplifying its online implementation process.

[42] By using the Adaptive Moment Estimation (Adam) algorithm, it was confirmed that this method can significantly optimize the operating efficiency of neural networks, providing algorithmic support for improving the performance of LSTM models.

[25] SOH was defined as the time period corresponding to the equal discharge voltage difference, which was extracted by the constant current (CC) voltage curve; to monitor the nonlinear degradation mode of the battery, the study adopted the Monotonic Echo State Network (MONESNs) technology, and combined with feature extraction and screening techniques, carried out relevant research on the optimization of SOH features of the LSTM model, and improved the stability and accuracy of the evaluation through model fusion, and the entire process was deduced based on probability distribution.

The aforementioned research focuses on the application of LSTM and its variants in battery SOH/RUL estimation, covering multiple dimensions such as feature optimization, network structure improvement, data sampling strategies, and algorithm fusion. This fully demonstrates the technical advantages and application flexibility of recurrent neural networks in processing battery depletion data over time.

The recent rise of DL can be attributed to the availability of cloud platforms and the accessibility of big data. Processing massive volumes of data is a well-known challenge, and DL algorithm has proven advantageous in scenarios involving large-scale datasets. Among DL algorithms, DNNs are particularly well-suited for handling large amounts of data because they are able to extract high-level features from raw, unprocessed data. CNN excel at automatically extracting features through convolution techniques when analyzing multidimensional data. For time-series data, LSTM models are considered the best choice. While DL offers many advantages, it is worth noting that it also suffers from drawbacks such as high computational cost and long processing time. Because it processes a large number of input parameters, it requires large datasets, thus acknowledging both its strengths and weaknesses. With the increasing use of LIBs in complex devices and high-voltage applications, accurate estimation of SOH and RUL of LIBs is of great significance.

Methodology

This chapter investigates how to construct an accurate RUL prediction model for lithium-ion batteries. Addressing the problem of low accuracy in prediction models built under specific charging and discharging conditions, this paper proposes using MLP, LSTM, and transformer to extract key influencing factors, and then training a high-precision model, ultimately comparing these three models for prediction, thereby building a more accurate RUL prediction model for lithium-ion batteries.

This paper mainly consists of four research contents, and the implementation steps are as follows:

(1) Extraction of indirect influencing factors of lithium batteries. In this research content, the first step is to select a suitable public dataset for lithium batteries. The dataset selected in this paper is the NASA lithium-ion battery aging dataset. The raw data is obtained, processed, and analyzed. Referring to authoritative literature in this field and combining with personal knowledge, battery aging features are extracted and finally validated.

(2) Establishment RUL prediction model for lithium batteries. Addressing the problems of low accuracy and overfitting of RUL models due to the small amount of data under small sample conditions, this paper uses the extracted influencing factors and lithium-ion battery capacity as input data for prediction after integration.

5 Dataset

Currently, the main research focus in academia is on improving the accuracy of machine learning prediction models, while relatively little research has been done on the data itself used as input for these models. Given that raw datasets often face problems of high dimensionality and data redundancy, these factors can negatively impact the accuracy of lithium-ion battery life prediction. Therefore, thorough preprocessing of lithium-ion battery datasets becomes crucial. This involves in-depth analysis of the internal data structure and effective extraction of key feature parameters, aiming to simplify and optimize the parameters. This process not only improves the accuracy of battery life prediction but also enhances computational efficiency, laying a solid foundation for future prediction work.

5.1 Discription of Dataset

The dataset released by NASA contains experimental data on 18650-type lithium-ion batteries. To get these data, the batteries were repeated charging and discharging to collect data. The batteries are standard 18650-type lithium-ion batteries with a graphite anode and a nickel-cobalt-manganese oxide (NMC-LiNiCoMnO₂) cathode [8]. This data is stored in MATLAB's ".mat" format and records the results of charging, discharging, and resistance measurement experiments conducted under different temperature conditions and charge/discharge modes. The collected data includes parameters such as battery capacity, voltage, resistance, and current, which are key indicators for evaluating battery aging. For example, the B0005 battery data clearly shows the structure of this experimental data. This study uses the NASA publicly available dataset of 18650-type lithium-ion batteries with a rated capacity of 2Ah, specifically the first group of batteries in the dataset (B0005, B0006, B0007, B008). This data was obtained at room temperature (24°C) through cyclic charge and discharge and resistance measurement experiments. The third and sixth groups measured the battery's cycling performance under more extreme conditions (43°C and 4°C, respectively).

The first group of data mainly studied in this paper was obtained at room temperature of 24°C, with a charging current of 1.5A and a discharging current of 2A. This data was provided by NASA, and the specific experimental procedure was as follows: First, during the lithium battery charge and discharge experiments, a constant current (CC) followed by a constant voltage (CV) mode was adopted. This experimental procedure first involves continuously charging the battery, with the charging current set at 1.5A. When the battery voltage reaches and stabilizes at the maximum value of 4.2V, charging continues until the current drops to 0.2A. This decrease in current marks the completion of the lithium-ion battery charging process. During this process, the battery first undergoes a constant current charging phase until the voltage reaches the preset maximum point. To optimize charging efficiency and avoid overcharging, the charging mode will then switch to constant voltage mode. At this point, the current will gradually decrease to the set threshold, which means the battery is fully charged. This effectively ensures the charging efficiency and safety of the battery. In the lithium-ion battery discharge experiment, the constant current discharge method was used. This method discharges the battery with a current of 2A. In the initial stage of discharge, the battery discharges at a higher current until the voltage drops to a preset stable value. This stable voltage value is a key indicator during the discharge process, marking the end of the discharge process. The constant current discharge method can help accurately evaluate the discharge characteristics of a battery, such as its discharge capacity and battery performance under different discharge conditions. By monitoring voltage changes during continuous discharge, a better understanding of battery performance and stability can be achieved. The frequency range for measuring the electrochemical impedance of the battery is from 0.1 Hz to 5 kHz.

As shown in Figure 2, the maximum effective capacity of a battery generally tends to decrease as it undergoes more charge-discharge cycles. However, this capacity reduction is not a monotonically decreasing process, but is accompanied by repeated capacity recovery, i.e., battery capacity fluctuation. This nonlinear and variability is a key challenge to the accuracy of battery state estimation [31]. Depending on different discharge strategies and ambient temperatures, the capacity decay curve of a battery will be different, and the

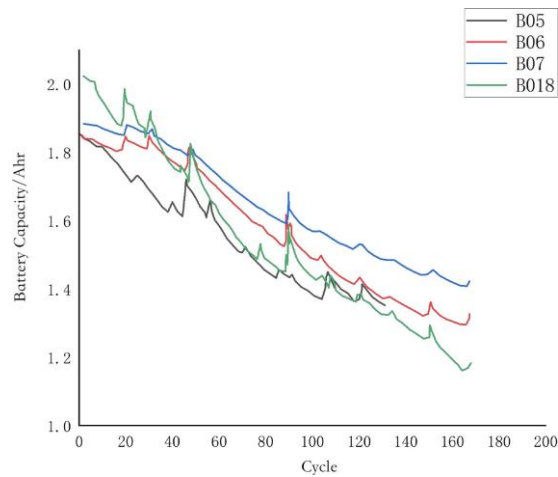


Figure 2: Battery Capacity of B005-B018 Cells at Different Cycle Numbers

fluctuation range of its capacity recovery will also vary.

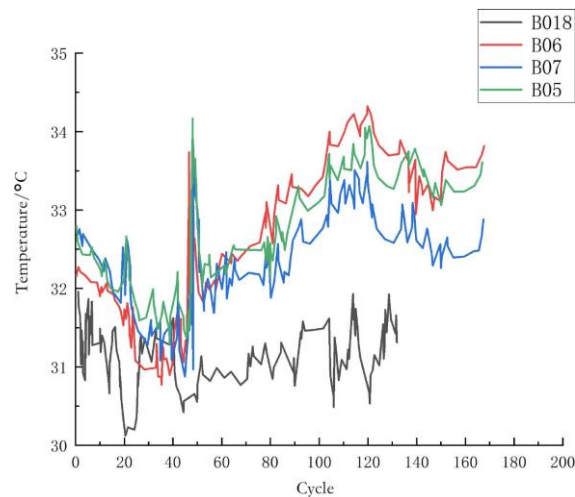
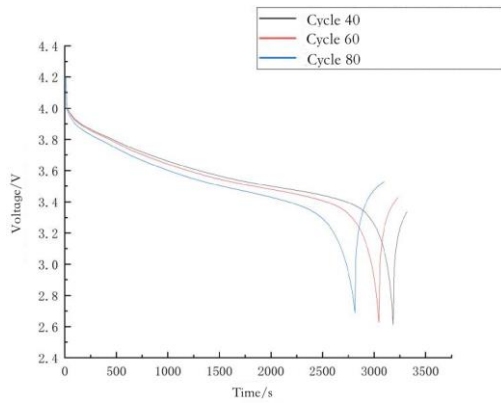


Figure 3: Influence of Cycle Numbers on Temperature in Lithium-ion Batteries

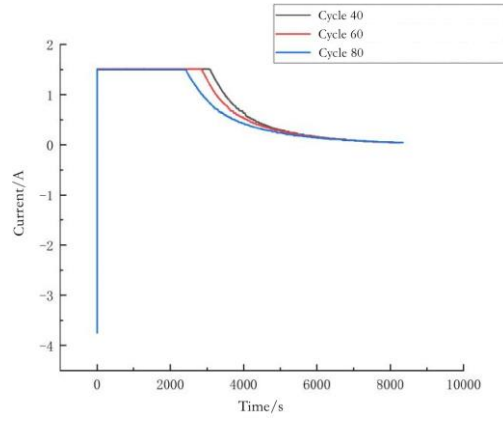
During a single charge-discharge cycle, the battery temperature typically rises gradually and slowly. This phenomenon is primarily due to the heat generated by the electrochemical reactions during charging and discharging (Joule heating), causing the battery's average temperature to generally exceed the ambient temperature. In most cases, except for batteries used in extremely low-temperature environments, the average battery temperature exhibits a curvilinear upward trend with increasing cycle count. However, it is important to emphasize that the average battery temperature does not simply increase or decrease with increasing cycle count; temperature fluctuations are particularly pronounced in batteries cycled at room temperature.

Next, taking the B005 battery as an example, we will study the battery discharge under relevant influencing factors based on the dataset provided by NASA laboratories.

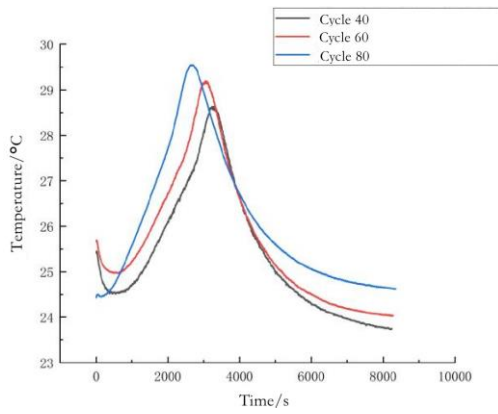
Figures 4 show the data on the changes of key parameters of battery B0005 over time during the charging process. It was observed that the battery terminal voltage gradually



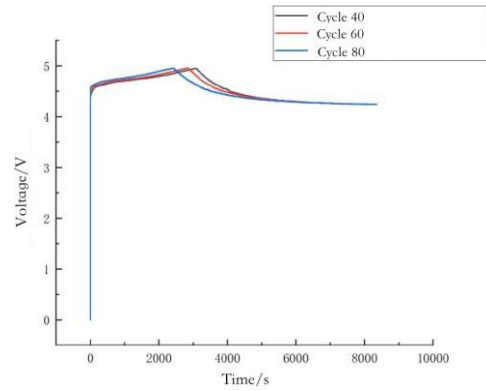
(a) Charging Battery Terminal Voltage



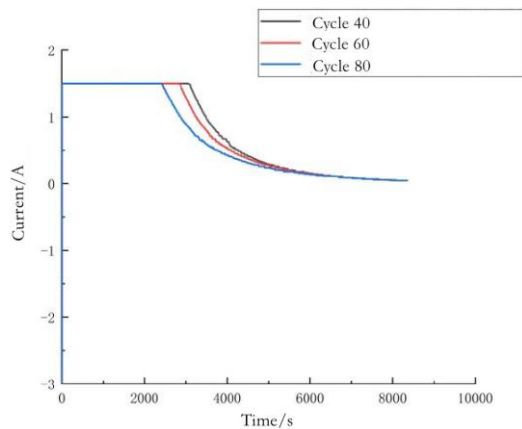
(b) Charging Current



(c) Battery Temperature during Charging



(d) Measured Voltage of the Charging Battery



(e) Charging Measured Current

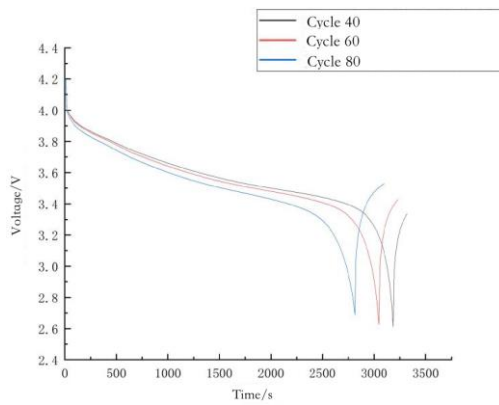
Figure 4: Charging Battery Data

increased during charging, eventually reaching a peak of 4.2V and stabilizing.

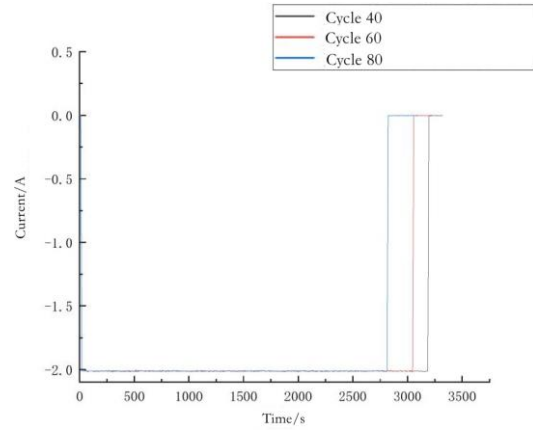
With the increase of charging cycles, the time it took for the battery to reach its peak voltage gradually shortened. Initially, the output current remained at 1.5A, gradually decreasing until it reached zero as charging progressed. Notably, the time at which the current began to decrease advanced with the increase of charging cycles. The battery temperature showed a trend of first increasing and then decreasing, and the peak temperature increased accordingly with the increase of charge-discharge cycles. The measured battery voltage increased in the early stages of charging, reached its peak, and then gradually decreased, with the time to reach the peak advancing with the increase of charge-discharge cycles. Finally, the trend of the current measured during charging was basically consistent with that of the battery output current. As the number of charging cycles increases, the time required for the battery to reach its peak voltage gradually decreases. Initially, the battery output current is maintained at 1.5A, and then the current will gradually decrease until it reaches zero as charging progresses. Notably, the time when the current drops will be earlier as the number of charging cycles increases. The battery temperature shows a trend of first increasing and then decreasing, with the peak temperature increasing with the increase of the number of charge-discharge cycles. The measured battery voltage increases during the initial charging phase, reaches its peak, and then gradually decreases. Similar to current, the time when the voltage reaches its peak value is earlier as the number of charging cycles increases. In summary, the trend of current change measured during charging is basically consistent with the trend of battery output current change.

Shown in Figures 5, during the discharge, the battery's terminal voltage decreases and experiences a brief but rapid rebound after reaching a certain threshold. With more charge-discharge cycles, the time before the minimum voltage point gradually shortens. Furthermore, with more charge-discharge cycles, the minimum voltage value gradually increases. During discharge, the battery outputs a constant current of 2A, and the time required to complete the discharge decreases with increasing number of cycles. During discharge, the battery temperature continuously rises until it begins to decrease near the end of the discharge phase. In particular, with more cycles, the rate of increase in temperature accelerates and the peak temperature also gradually increases. The load voltage of the discharging battery continuously decreases throughout the discharge cycle until the end of the discharge. With more battery cycles, the time before the load voltage drops to 0V becomes shorter. In the end, the curve of the discharge load current is the same as the curve of the discharge battery's output current, indicating that they exhibit a consistent trend.

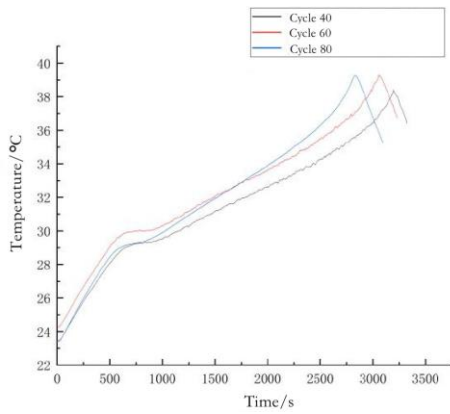
The raw data is stored in MATLAB structs provided by the NASA Li-ion Battery Aging Datasets, hosted by the Prognostics Center of Excellence (PCoE) [35]. Based on the extraction logic in backend/data_utils.py, the file structure consists of a top-level dictionary (keyed by the battery ID, e.g., B0005) containing a cycle array. Each element in this array represents a specific operation (Charge, Discharge, or Impedance), shown in the figure 6.



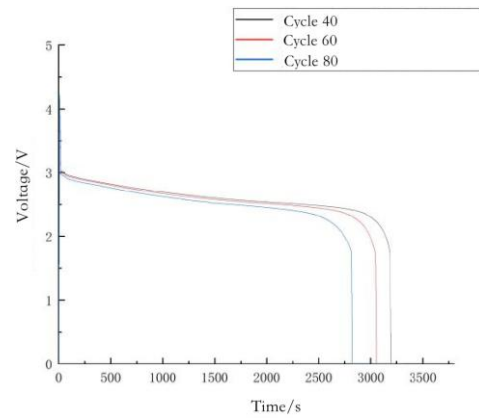
(a) Discharge Battery Terminal Voltage



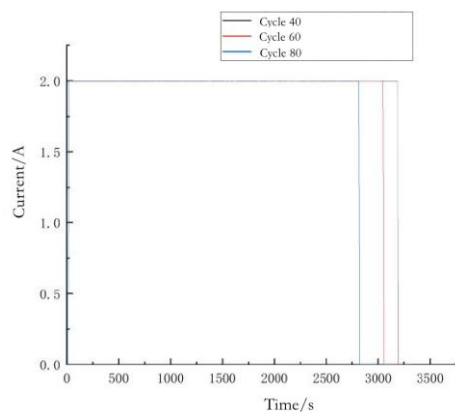
(b) Discharge Battery Output Current



(c) Discharge Battery Temperature



(d) Discharge Battery Load Voltage



(e) Discharge Load Current

Figure 5: Discharging Battery Data

```

B0005.mat
  B0005 (struct)
    cycle (1xN array of structs)
      type (string: 'charge', 'discharge', 'impedance')
      ambient_temperature (double)
      time (date vector)
      data (struct)
        Voltage_measured (1xM array)
        Current_measured (1xM array)
        Temperature_measured (1xM array)
        Current_charge (Used for capacity integration)
        Voltage_charge (1xM array)
        Time (Used for dt calculation)

```

Figure 6: MATLAB file structure

5.2 Pre-process

Preprocessing is a critical step in developing ML for lithium-ion battery aging prediction. This process determines the quality of input features and the model's accuracy. The objective of this phase is to transform raw battery operation data from the NASA dataset into more meaningful and clean features.

First, we extract the valid charge cycles from the raw dataset. This involves filtering out incomplete cycles and abnormal data points. Next, we calculate the capacity (Ah) of the battery at each valid charge cycle by summing the charging current over time. Integrating the current (in amperes, A) during a charging cycle yields the total charge in ampere-hours (Ah), which is the battery capacity. Battery capacity is raw data that measures a battery's available energy storage capacity, as well as its related state and lifespan.

After obtaining the capacity for each cycle, the calculation of SOH follows the formula 1.

In our algorithm, the initial capacity used to calculate SOH is not the nominal value from the battery datasheet. Because the nominal value may deviate from the actual performance, we define the denominator as the maximum capacity in the first 5 charging cycles. This is because the capacity of a new battery typically fluctuates slightly during the first few cycles. We speculate that this may be due to the electrode material not yet being stable. Taking the maximum value from the early cycles more reasonably reflects the true initial performance of the battery, which can avoid underestimating or overestimating the initial capacity due to short-term fluctuations.

For each cycle, divide the current capacity (Q_k) by the calibrated initial capacity ($Q_{initial}$) to obtain SOH. For example, if the battery's initial maximum capacity is 2.0 Ah, and its current capacity at the 100th cycle is 1.8 Ah, then SOH is $(1.8/2.0) \times 100 = 90\%$, indicating that the capacity has decayed to 90% of its initial state. The figure 7 shows the actual SOH values obtained from the formula for each dataset, which uses B0005, B0007, and B0018.mat as examples.

After calculating SOH, we still need to quantize RUL. This is to understand how many

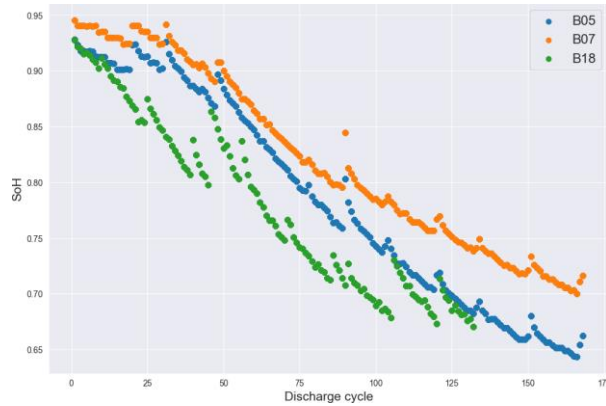


Figure 7: SOH of B0005, B0006, B0018

more cycles a battery can run before reaching the end of its lifespan (EOL). In the algorithm, the threshold EOL is set to 80.0% SOH. This threshold is a widely accepted industry standard because the industry generally believes that batteries below this threshold typically have significant energy loss and higher safety risks. The calculation logic for RUL is as follows: 1. For each predicted SOH value, the code counts the number of consecutive loops in which SOH remains above the 80% threshold. 2. Counting stops when SOH falls below the threshold. Store the counted number of loops as RUL for a specific cycle i . For example, if the predicted SOH remains above 80% for the next 35 cycles at the 200th cycle (until the 235th cycle, when the SOH drops to 79%), then the RUL at the 200th cycle is 35 cycles.

The above preprocessing steps include calibration using the maximum capacity of early cycles, effective cycle extraction, and capacity integration into SOH calculation; and RUL calculation based on the industry-standard EOL 80% threshold.

6 Model description

Currently, various data-driven methods exist for predicting battery health status. These methods treat changes in battery capacity as a time-series data sequence. During the battery charging and discharging cycle, each time point corresponds to a specific set of data, including voltage, current, and temperature. As time progresses and the number of charge-discharge cycles increases, these data form a time series reflecting the changes in battery capacity. By analyzing this time series, we can find that battery capacity is not constant but gradually decays with increasing usage time and charge-discharge cycles, exhibiting certain patterns. The core of data-driven methods lies in using historical degradation data to reveal these patterns. Through learning from a large amount of historical data, the algorithm can identify the characteristic patterns of data under different battery health states. When faced with new data, the algorithm can predict future changes in battery performance based on the learned patterns. However, each model has its advantages and disadvantages. This paper will subsequently use MLP, LSTM (with RNN), and Transformer, which are currently relatively effective in the field of battery life prediction, to build models and compare their performance.

6.1 MLP

The Multi-Layer Perception (MLP)[29] model can also be regarded as a DNN model. It is a network model based on the basic principles of traditional neural networks, which simulates the way the deep nervous system of the human brain solves complex problems and combines the network topology. As a multi-layer feedforward neural network with deep learning capabilities, the multi-layer perceptron plays a key role in deep learning models. It is characterized by autonomous learning, high intelligence, strong fault tolerance, and multi-layer parallel processing capabilities. It can combine the processing and storage of information, and can fit highly complex nonlinear functions into linear continuous functions with arbitrary precision [3] to realize the analysis and processing of logical structures. These advantages have also promoted the application and development of MLP models in speech analysis, image analysis, intelligent devices and other fields[27, 39].

The MLP model consists of an input layer, one or more hidden layers, and an output layer. Its multi-layered network contains a large number of neurons, which map a set of input vectors to a set of output vectors. The MLP model structure is shown in Figure 8. In the MLP model, each layer contains several neurons. Neurons in the same layer are not directly connected, but neurons in adjacent layers are fully connected through weighted summation. Except for the input layer, each neuron in each layer has a non-linear activation function. The multilayer perceptron receives data through the input layer, the neurons in the hidden layers analyze, process, and transmit the data, and the output layer outputs the data. The entire process achieves multi-layered optimization of the data.

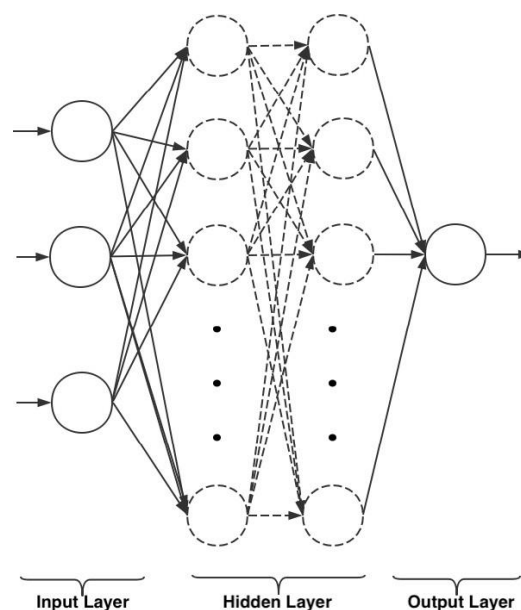


Figure 8: MLP model Structure

MLP determines the number of neurons in the input and output layers based on the target requirements; the number of hidden layers and their neurons are determined according to the set error requirements. In the MLP model, the output of the j neuron in the i layer is given by the formula:

$$y_j^{(i)} = f_j^{(i)} \left(\mathbf{W}_j^{(i)} \cdot \mathbf{y}^{(i-1)} + \mathbf{b}_j^{(i)} \right) \quad (2)$$

Where $\mathbf{W}_j^{(i)}$ is the weight vector of the j neuron in the i layer, directed from the $i - 1$ layer to the j neuron in the i layer; $\mathbf{y}^{(i-1)}$ is the output vector of the $i - 1$ layer; $\mathbf{b}_j^{(i)}$ is the bias vector of the j neuron in the i layer; and $f_j^{(i)}$ is the activation function of the j neuron in the i layer. Two commonly used activation functions are $f_j^{(i)}$:

(1) Sigmoid activation function The expression of the Sigmoid activation function is shown in formula 3.

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \quad (3)$$

All parameters of an MLP model are the weights and biases between layers. The selection of parameters affects the prediction performance of the MLP to a certain extent. Therefore, it is necessary to optimize the weights and biases of the MLP to make the output value of the MLP model closer to the true value, thereby improving the prediction accuracy of the model.

(2) Tanh function The Tanh function is a non-linear function, and its expression is shown in equation 4.

$$\tanh(n) = \frac{1 - e^{-2n}}{1 + e^{-2n}} \quad (4)$$

The Tanh function has a value range of $[-1, 1]$. When features differ significantly, the Tanh function performs well and continuously amplifies the feature effect during iteration. In practical applications, the Tanh function is superior. Therefore, this paper selects the Tanh function as the activation function $f_j^{(i)}$.

Assuming the MLP model has m ($m \geq 3$) layers, then the m layer is the output layer. The process from the hidden layer to the output layer can be viewed as a multi-class logistic regression, i.e., softmax regression. The output formula for the m layer is shown below.

$$Y = G(\mathbf{W}^{(m)} \cdot \mathbf{y}^{(m-1)} + \mathbf{b}^{(m)}) \quad (5)$$

Where Y is the output of the output layer, $\mathbf{W}^{(m)}$ is the weight vector of the m layer, $\mathbf{y}^{(m-1)}$ is the output vector of the $m - 1$ layer, $\mathbf{b}^{(m)}$ is the bias vector of the m layer, and the function G is the softmax function.

All parameters of an MLP model are the weights and biases between layers. The selection of parameters affects the prediction performance of the MLP to a certain extent. Therefore, it is necessary to optimize the weights and biases of the MLP to make the

output value of the MLP model closer to the true value, thereby improving the prediction accuracy of the model.

As a promising deep learning model, the multilayer perceptron has the following characteristics:

(1) High degree of parallelism

A multilayer perceptron is composed of multiple interconnected networks, with a large number of neurons connected in parallel within each layer. While the function of a single neuron is not complex, the overall efficiency of information analysis is superior when a large number of simple neurons are active in parallel. Even with a large amount of information, the processing power of the multilayer perceptron is not diminished.

(2) Highly nonlinear global effects

As can be seen from the architecture of a multilayer perceptron, each layer contains several neurons, and there are numerous mapping relationships between these neurons. This allows the nonlinear input to be fitted into a linear output through layer-by-layer mapping, and the output data also affects the neurons. The operation of a multilayer perceptron is based on global interaction RULes; its model function is not merely the sum of local performances, but rather should start from the whole, with close connections between the previous and next processing steps, ultimately solving the problem through layer-by-layer processing and feedback.

(3) Good fault tolerance and strong self-learning ability

MLPs utilize their structure to distribute information across the weights of individual neurons. The stored information is not apparent from a single weight, thus employing a distributed storage method. This significantly improves the model's fault tolerance, enabling MLPs to perform tasks such as feature extraction, type analysis, speech recognition, and data classification. Furthermore, MLPs refine their parameters and structure through training, demonstrating strong self-learning capabilities and adaptability to their environment.

The steps for constructing a prediction model based on a multilayer perceptron (MLP) and performing prediction using an MLP are as follows:

1. Import training data and prediction data;
2. Normalize the data;
3. Initialize the parameters of the MLP model;
4. Start iteration and calculate the error for each iteration;
5. Determine if the maximum number of iterations or the target error value has been reached. If yes, proceed to the next step; otherwise, return to step (4);
6. Perform prediction experiments using the trained model;

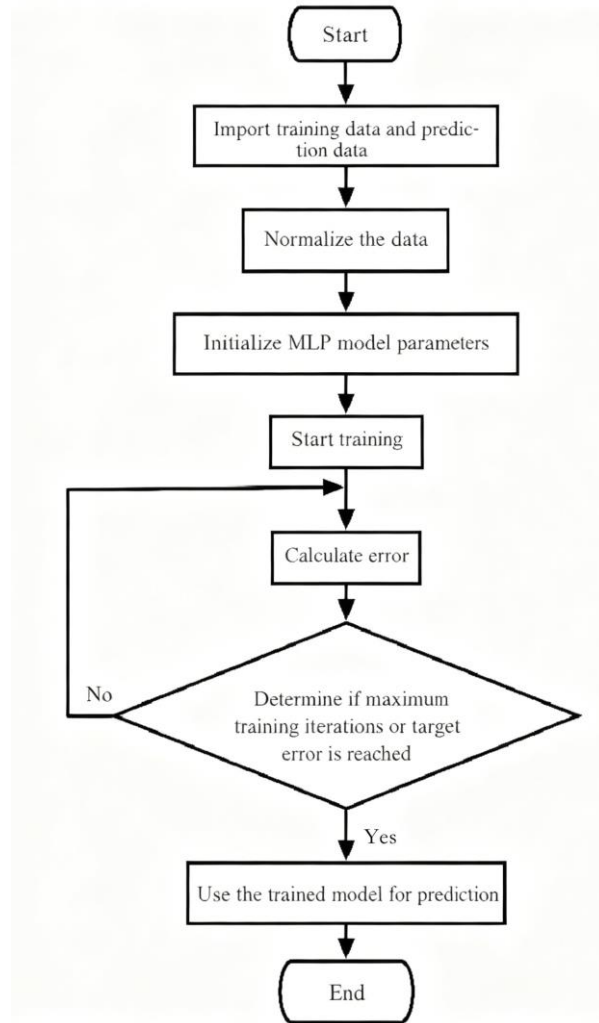


Figure 9: Modeling flowchart of MLPs

7. Output the prediction experiment results.

In the battery health status prediction task, MLP can effectively evaluate and predict the health status of a battery by learning the non-linear relationships in the battery's operating data. It can capture the complex correlation patterns between battery capacity degradation and factors such as voltage, current, and temperature. However, MLP also has certain limitations. Due to its fully connected structure, as the number of neurons in the hidden layer increases, the number of model parameters increases sharply. This not only leads to a significant increase in computation, increasing training time and computational resource consumption, but also easily causes overfitting problems, resulting in the model performing well on the training set but having poor generalization ability on the test set or in practical applications, making it unable to accurately predict new data. Furthermore, MLP has difficulty fully utilizing the time series information of data when processing data with temporal characteristics, as it does not explicitly consider the temporal dependencies between data, and its ability to capture the dynamic changes in battery capacity over time is relatively weak.

6.2 LSTM

Long Short-Term Memory (LSTM) is a variant of RNN and is widely used in deep learning tasks for processing sequential data. LSTM was introduced to overcome problems such as vanishing and exploding gradients encountered by traditional RNNs when processing long sequences, thereby improving the model's ability to model long-term dependencies. In the past few decades, LSTM has become an important tool in fields such as natural language processing, speech recognition, and time series analysis.

Figure illustrates the internal structure of a LSTM network. In RNN node, typically only a single tanh activation function is used. The input information at the current time step includes the hidden information $h_{t-1}^{(i)}$ output by the previous node and the new input information X_t of the current node. Compared to RNN, the LSTM structure, in addition to containing an m-dimensional activation function, also introduces a sigmoid function, forming a gating mechanism to regulate the information transmission path and update the transmitted information. This forms the internal structure of the network, enabling LSTM to filter information states, selectively remembering or forgetting information.

The LSTM neural network proposed by Hochreiter and Schmidhuber can be regarded as a special variant of the RNN [10]. The original intention of this network was to address the fact that traditional neural networks rarely have memory functions, and traditional neural networks often face the problem of gradient vanishing or gradient exploding, which is unavoidable in the network training process. The root cause of this problem is that in the learning process of traditional neural networks, the number of network layers gradually increases over time [22].

LSTM neural networks have successfully overcome the limitations of traditional neural networks by introducing hidden cell states into their structure and carefully designing three "gate" structures. The ingenious use of these gate structures allows for increasing

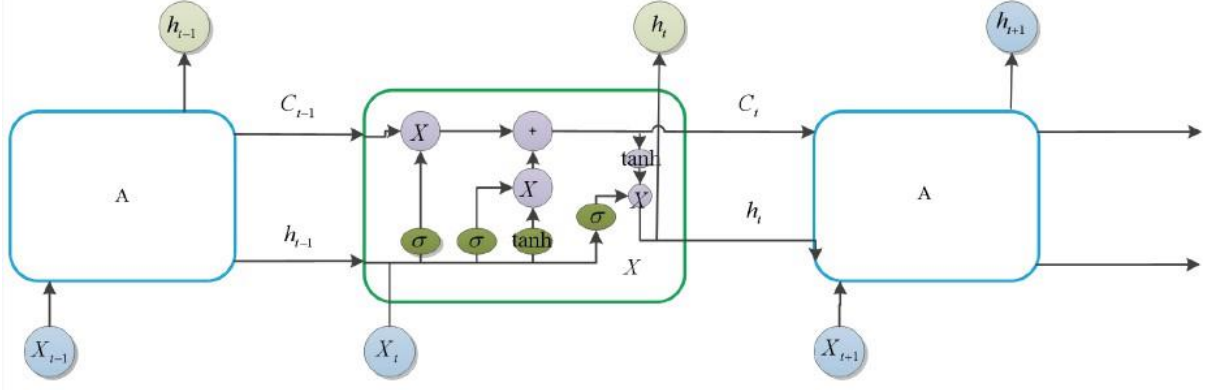


Figure 10: LSTM Cell Structure

or decreasing the flow of information to the cell state, enabling LSTM networks to effectively handle long-term dependencies. When inputting data, researchers can adjust the input data to these "gates" as needed, making the neural network's self-recurrent weights adjustable, thus achieving flexible information retention. These three gate structures are the input gate, forget gate, and output gate. The input gate determines whether to retain some information, the forget gate mainly determines whether to discard some information, and the output gate calculates the cell state results from the input and forget gates. The LSTM neural network achieves the storage of long-term data information by updating its internal state, where the letter A represents the same isomorphic cell.

In Figure 10, is the Sigmoid activation function, whose output range is (0,1), where 0 indicates that the state is discarded and 1 indicates that the state is successfully passed; while the output range of tanh is (-1,1).

Forget Gate: Responsible for determining which information from the previous time step should be forgotten and which information should be retained in the cell state.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (6)$$

In the above equation, W_f represents the weight matrix, h_{t-1} represents the output state at the previous time step, x_t represents the input state at the current time step, b_f represents the bias value, and f_t represents the output value.

Input Gate: Based on the current input and the output of the previous time step, it determines whether to add new information to the cell state.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (7)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (8)$$

In the above equation, W_i represents the weight matrix, b_i represents the bias value, W_C represents the weight matrix of the cell state, b_C is the bias value of the cell state, and \tilde{C}_t represents the candidate cell state. Cell state: Responsible for storing and transmitting information, and its state is updated under the regulation of the forget gate and the input gate. The process of updating the cell state is as follows:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (9)$$

In the above equation, f_t represents the output of the forget gate, i_t represents the output of the input gate, and \tilde{C}_t represents the cell state at the previous time step.

$y^{(t)}$ represents the output of the output layer at time t , while the output of the "output layer" at various intermediate times in the sequence can be represented by the following mathematical expression.

$$y^{(t)} = \sigma(Vh^{(t)} + c) \quad (10)$$

V represents a fully connected matrix.

Based on the principles of the LSTM algorithm described above, an LSTM battery prediction algorithm was designed experimentally, with particular attention paid to the close relationship between and . Using the current usable capacity of the battery as an indicator of its aging state, the RUL of battery B05 in the NASA battery dataset was predicted. The algorithm model was further adjusted in the experiment, and the prediction accuracy under different prediction step sizes was studied, providing a reference for multi-step forward prediction of batteries. The performance of the constructed LSTM battery prediction model was evaluated using the metrics MAE and RMSE.

Input Gate: The input gate in the LSTM network is another key component in its internal memory unit, playing the role of controlling the input of new information. The design of the input gate allows the LSTM network to better adapt to changes in sequential data and capture key input information. It controls the influence of new input on memory. It outputs weights for updating the memory through sigmoid and tanh activation functions. In the LSTM network, the input gate is iterated at each time step, determining the impact of the new input information at the current time step on the internal memory unit. This mechanism allows the LSTM to adapt more flexibly to changes in sequential data. The input gate works in conjunction with the forget gate and output gate, jointly regulating the content of the internal memory unit, achieving accurate modeling of sequential data by the LSTM network.

Output Gate: The output gate in the LSTM network is a crucial component of the network, responsible for controlling the information output from the internal memory unit. The design of the output gate allows the LSTM network to better adapt to inputs at different time steps, improving its performance on sequential data. It determines how much memory to output at a specific time. The final memory is output through sigmoid and tanh activation functions. In the LSTM network, the output gate is iterated at each time step, regulating the output information based on the current input and the hidden state of the previous time step. This mechanism allows the LSTM network to better adapt to different patterns in sequential data. The output gate works in conjunction with other components such as the input gate and the forget gate, jointly regulating the content of the internal memory unit to achieve accurate modeling of sequential data by the LSTM network.

The LSTM model uses the 18650 type lithium-ion battery test dataset publicly available from NASA. The feature data extracted from the B0005 battery is preprocessed to obtain training and testing data. This data includes multimodal information such as discharge voltage, discharge current, and temperature of the lithium-ion battery during discharge, which is integrated into four-dimensional data, with battery capacity as the output. A corresponding program is written, and the constructed LSTM model is trained and tested

on the experimental platform. By evaluating the training and testing results, parameters are adjusted to obtain the prediction results of the remaining useful life of the lithium-ion battery.

6.3 Transformer

As an innovative deep learning model, the Transformer has achieved remarkable results in sequence transformation fields such as machine translation by utilizing self-attention mechanisms. Its core idea lies in abandoning the use of RNNs and CNNs in sequence processing, relying entirely on self-attention mechanisms to handle long-range dependencies in the input sequence. As shown in Figure 11, the core architecture of the Transformer model mainly consists of two key components: the encoder component responsible for processing the input sequence and the decoder component responsible for generating the output sequence. These two components interact through a self-attention mechanism to achieve the transformation from the input sequence to the output sequence. The overall flow of the Transformer mechanism is as follows:

(1) Input: The input sequence is first converted into word embeddings and then positional encoding is added to inject word order information.

(2) Encoder: The encoder is mainly responsible for transforming the raw data into a sequence feature representation with rich contextual information. The encoder consists of a series of repeatedly stacked identical layers, each layer containing two key sub-components: a self-attention mechanism sub-layer and a feed-forward network (FFN) sub-layer. The self-attention mechanism enhances the model's sensitivity to global information by quantifying the correlation between elements within the sequence, while the FFN applies a non-linear transformation to mine deeper feature representations. Furthermore, the encoder introduces residual connections and layer normalization techniques between each pair of sub-layers to address common problems such as vanishing or exploding gradients.

(3) Decoder: The decoder is mainly responsible for using the contextual information provided by the encoder and its own generated historical information to gradually generate the target sequence. The decoder also consists of multiple layers, each containing a self-attention sublayer (masked to prevent future information leakage), a cross-attention sublayer (mainly used to integrate contextual information from the encoder during the decoding stage), and an FFN sublayer. Residual connections and normalization are also performed between these sublayers.

(4) Output: By performing a linear transformation on the decoder output and then processing it through the softmax function, the decoder's internal representation can be effectively converted into the probability distribution of the target sequence.

In the Transformer model, the Transformer architecture itself does not contain recurrent or convolutional layers, and therefore cannot capture the sequential information of the sequence. Therefore, to address the issue that the Self-Attention mechanism itself does not include positional information of elements in the input sequence, a positional encoding module is introduced. Initially, the Transformer used absolute positional encoding. This

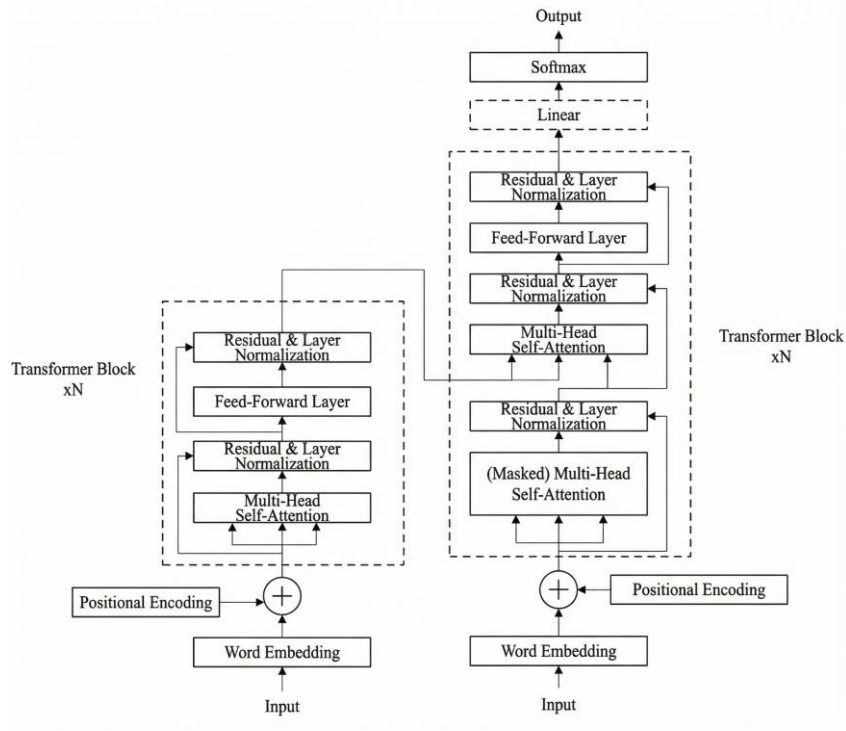


Figure 11: Transformer Overall Architecture

encoding method generates a fixed vector representation for each position based on sine waves and cosine functions, and its calculation process is as follows:

$$PE_{(pos\ 2i)} = \sin\left(pos/10000^{\frac{2i}{d_{model}}}\right) \quad (11)$$

$$PE_{(pos\ 2i)} = \cos\left(pos/10000^{\frac{2i}{d_{model}}}\right) \quad (12)$$

Here, pos represents the position index in the sequence, i represents the dimension index, and d_{model} is the method by which position encoding can encode sine and cosine waves of different frequencies in different dimensions, enabling the model to distinguish positions at different distances, and this periodic encoding ensures that the model can generalize to long sequences that have never been seen before.

It is worth noting that, with the development of research, in addition to absolute position encoding, researchers have also explored the method of relative position encoding. This encoding directly reflects the relative positional relationship between elements in the sequence, rather than a single absolute positional information, in order to improve the robustness of the model to changes in sequence structure. For example, the Transformer-XL model uses relative position encoding to further optimize the performance of the Transformer on long sequence tasks [9].

The self-attention mechanism is a core component of the Transformer model. It allows the model to focus on different parts of the sequence when processing the input sequence,

thereby better understanding and representing the contextual relationships in the sequence. It plays a key role in modeling long-distance dependencies. The overall process is shown in Figure 12.

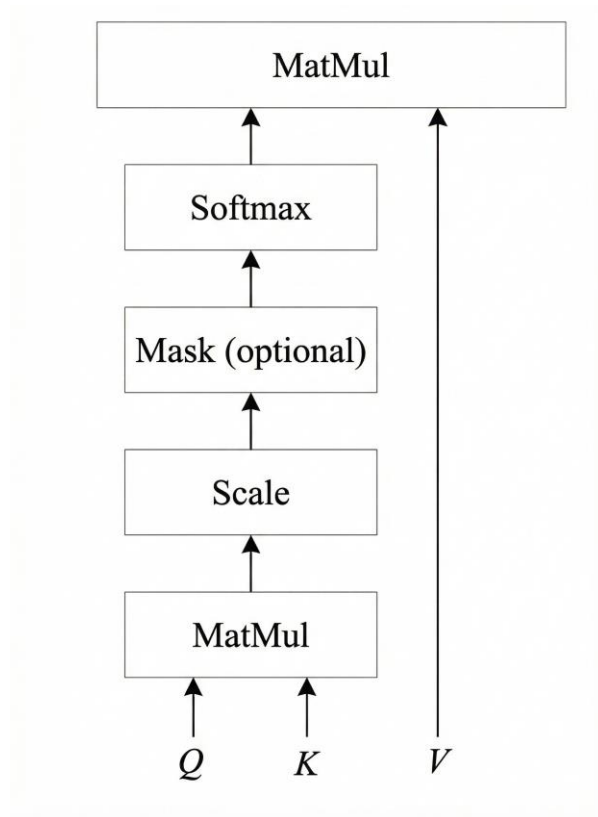


Figure 12: Self-attention mechanism

Specifically, the self-attention mechanism utilizes three core concepts: a query matrix, a key matrix, and a value matrix to dynamically adjust the representation of each element in the sequence. These three matrices are generated from the input sequence through three different linear transformations, and their calculation process is as follows:

$$Q = XW^Q \quad (13)$$

$$K = XW^K \quad (14)$$

$$V = XW^V \quad (15)$$

Here, K , W^K , and W^V are trainable weight matrices, representing the projections of the query, key, and value, respectively, while X is the input matrix after word embedding and position encoding of the input sequence. Attention weights are calculated using the query matrix Q and the key K , and then the context representation of each position is obtained by weighting the value matrix V with these attention weights. The calculation process is as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (16)$$

Here, d_k is the dimension of the key vector. The attention score matrix is obtained by performing a dot product between the lookup matrix Q and the transpose of the key matrix K . This is then normalized using a softmax function to obtain the attention weight matrix. To ensure gradient stability, the result is typically divided by a scaling factor $\sqrt{d_k}$ before calculating the *softmax*. Finally, the normalized attention weights are used to perform a weighted sum of the value vectors, yielding the output of the self-attention mechanism, which contains weighted information for all positions in the input sequence.

In practice, self-attention mechanisms are often extended to multi-head self-attention mechanisms to capture more fine-grained features. As shown in Figure 13, by dividing the input query, key, and value matrix into multiple parts, multiple self-attention heads are computed in parallel, capturing features in the input sequence from different perspectives, thereby enhancing its expressive power and ability to process complex data.

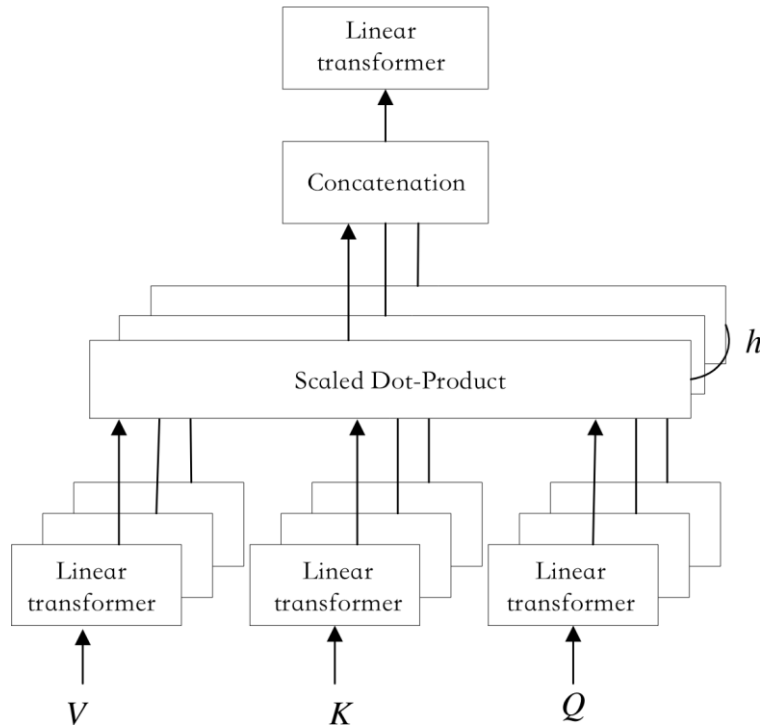


Figure 13: Multi-Head Attention Mechanism

Multi-head attention allows the model to learn in different subspaces, enabling it to better understand complex dependencies. Its computational process is as follows:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (17)$$

$$\text{where head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (18)$$

Where h is the number of parallel self-attention layers, $W_i^Q, W_i^K, W_i^V \in R^{d_{\text{model}} \times d_k}$ are the weight vectors of the query, key, and value matrices of the i -th time series, respectively. $W_i^O \in R^{d_{\text{model}} \times d_k}$, and their dimensional relationships are: $d_v = d_k = d_{\text{model}}/h$. W_i^Q, W_i^K, W_i^V and W_i^O are parameter matrices obtained through learning and training.

The feedforward layer in the Transformer model is a crucial component of the encoder and decoder blocks, located after the self-attention layer. Its design aims to enhance the model's expressive power and learn more complex feature interactions, compensating for the limitations of the self-attention layer in handling complex language structures or patterns.

The feedforward layer cleverly combines linear transformations with non-linear activation functions to transform the input data from the original feature space to a deeper feature representation. Specifically, a linear transformation layer first maps the input data from the original feature space to a higher-dimensional hidden space, enhancing the data's expressive power and enabling it to capture more complex and abstract feature patterns. Then, these high-dimensional features are passed through a non-linear activation function, allowing the network to learn and simulate complex non-linear relationships. Finally, a linear transformation layer maps the non-linearly activated features back from the high-dimensional space to a lower-dimensional space, reducing feature redundancy and extracting more critical and effective feature representations for downstream tasks. This strategy of first increasing dimensionality and then decreasing it not only helps the model capture deeper features of the input data, but also helps prevent overfitting to some extent and improves the model's generalization ability. The calculation formula is as follows:

$$FFN(x) = \sigma(xW_1 + b_1)W_2 + b_2 \quad (19)$$

Where $x \in R^{n \times d_{\text{model}}}$, $W_1 \in R^{d_{\text{model}} \times d_{\text{ff}}}$, $W_2 \in R^{d_{\text{ff}} \times d_{\text{model}}}$, $\sigma(\cot)$ is the activation function, usually the simple and efficient ReLU function is chosen as the activation function, n is the length of the sequence, d_{model} is the dimension of the input, d_{ff} is the dimension of the intermediate layer of the feedforward network, which is usually larger than d_{model} . W_1 and W_2 are the weight matrices of the first and second linear transformations, respectively, b_1 and b_2 are the corresponding bias vectors, and x is the output of this self-attention layer, which serves as the input of the FFN layer.

Residual connections are a widely used technique in deep learning, playing a crucial role, especially in CNN and Transformer models. In transformer, residual connections are applied after each sub-layer (self-attention layer and feedforward layer). Specifically, the operation involves adding the output of the current sub-layer to its input, followed by normalization. The calculation steps are as follows:

$$\text{Output} = \text{LayerNorm}(x + \text{Sublayer}(x)) \quad (20)$$

Where x is the input of the sub-layer, $\text{Sublayer}(x)$ represents the transformation before the sub-layer (self-attention layer or feedforward layer), and LayerNorm is the layer normalization operation.

In battery health prediction, the Transformer model has shown immense application potential. It can effectively process long sequences of data generated during battery operation, accurately capturing complex patterns and long-term dependencies of battery performance over time. By learning from a large amount of historical data, the Transformer model can model the health status of a battery from a global perspective, thereby achieving accurate prediction of future changes in battery performance. When predicting the remaining lifespan of a battery, the Transformer model can comprehensively consider various data throughout the battery's entire lifecycle, including early charge and discharge data as well as recent operating status, thus providing more reliable prediction results. Compared with traditional models, the Transformer model's parallel computing capabilities and ability to capture long-range dependencies when processing long sequence data give it a significant advantage in battery health prediction tasks, and it is expected to provide more powerful decision support for battery management systems.

7 Model Performance Evaluation

Using the current usable capacity of the battery as an indicator of its aging state, the RUL of the NASA battery dataset B005 was predicted. The first part of the experiment predicted battery states using different proportions of the training set data at the same prediction step size. The accuracy and stability of the model were evaluated using performance metrics to provide a reference and comparison for the construction of a new model in the fourth part. The second part of the experiment further adjusted the algorithm model and studied the prediction accuracy under different prediction step size inputs, providing a reference for multi-step forward prediction of batteries. The MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error) metrics were used to evaluate the performance of the constructed LSTM battery prediction model. These metrics quantify the accuracy and performance of the model in predicting remaining useful life.

This study uses a method with 100 iterations (epochs=100) and uses MAE as the loss function of the model. Adam is chosen as the optimizer for the LSTM model to ensure that the scaling transformation of the gradient does not affect the parameter update. The Adam optimizer is computationally simple and efficient, and it is very suitable for use in case of a lot of noise in the actual operation of the battery [69].

This paper uses two error MAE, and RMSE—as criteria for evaluating the predictive ability of the model. MSE represents the magnitude of error variation, and RMSE evaluates the dispersion of the data.

The methods for calculating MSE and RMSE are shown in equations 21, and 22.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (21)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{g_N} \right)^2} \times 100\% \quad (22)$$

Where n represents the number of samples, y_i represents the true value of battery cycles, and \hat{y}_i represents the predicted value of the model. The model is trained to improve prediction accuracy by continuously minimizing the MSE error.

Results

During the training process of the model, in order to prevent underfitting or overfitting caused by a single training dataset, and to ensure the general applicability and accuracy of the model prediction. This chapter presents the results of this study, mainly analyzing the performance of MLP, LSTM and Transformer models for 'B0005.mat', 'B0006.mat', 'B0007.mat', and 'B0008.mat' from NASA's Prognostics Data Repository. One part involves comparing the predicted SOH values with the actual data values. The other part involves predicting the RUL of the aforementioned battery sample group. This chapter also makes a comparative analysis of three SOH prediction models. Additionally, a sensitivity analysis was added, mainly focusing on the impact of sequence length and data noise on model performance. Finally, Leave-One-Out Cross-Validation tests the model's generalization capability and performance on a completely new battery.

8 Model Performance

8.1 Prediction and Result Analysis of MLP Model

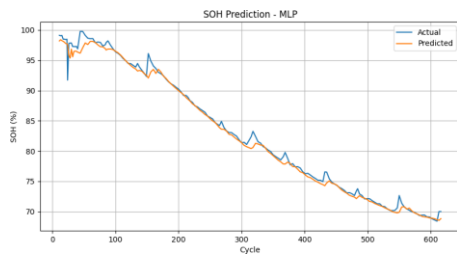
The figure 14 below shows a visualization of the comparison between the predictions and the actual values of SOH based on the MLP model. As can be seen from the figure, the SOH prediction value of the MLP is close to 99% at the beginning of the cycle, and then the value begins to decrease as the number of cycles increases. The actual values, however, generally show a decreasing trend as the number of cycles increases. Observation of the prediction results for B0005 batteries revealed that when using MLP for prediction, the prediction results for some cycles showed a significant deviation from the actual values. Furthermore, as the number of iterations increases, the deviation between the predicted and actual values also shows a gradual increasing trend. For each dataset, the performance of the MLP model is presented in Table 3 as MAE and RMSE. Figure 15 shows the RUL prediction results of the MLP model.

8.2 Prediction and Result Analysis of LSTM Model

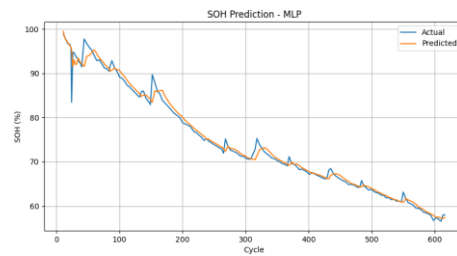
The figure 16 below shows a visualization of the comparison between the predictions and the actual values of the SOH model based on LSTM. As can be seen from the figure, the SOH prediction value of LSTM also begins to decrease as the number of iterations

Table 3: Prediction performance (MAE and RMSE) of the MLP model across four datasets.

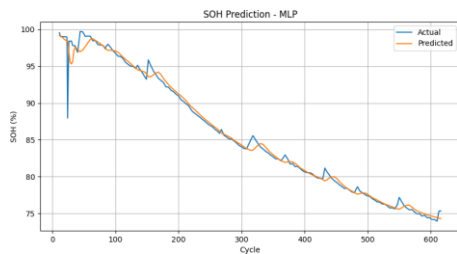
Performance	B0005	B0006	B0007	B0018
MAE	0.45683	0.8359	0.45373	0.9160
RMSE	0.9734	1.6549	1.0896	1.6769



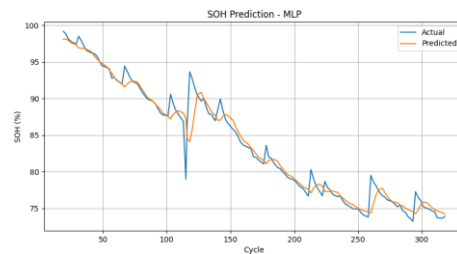
(a) B0005



(b) B0006



(c) B0007



(d) B0018

Figure 14: SOH Prediction of MLP

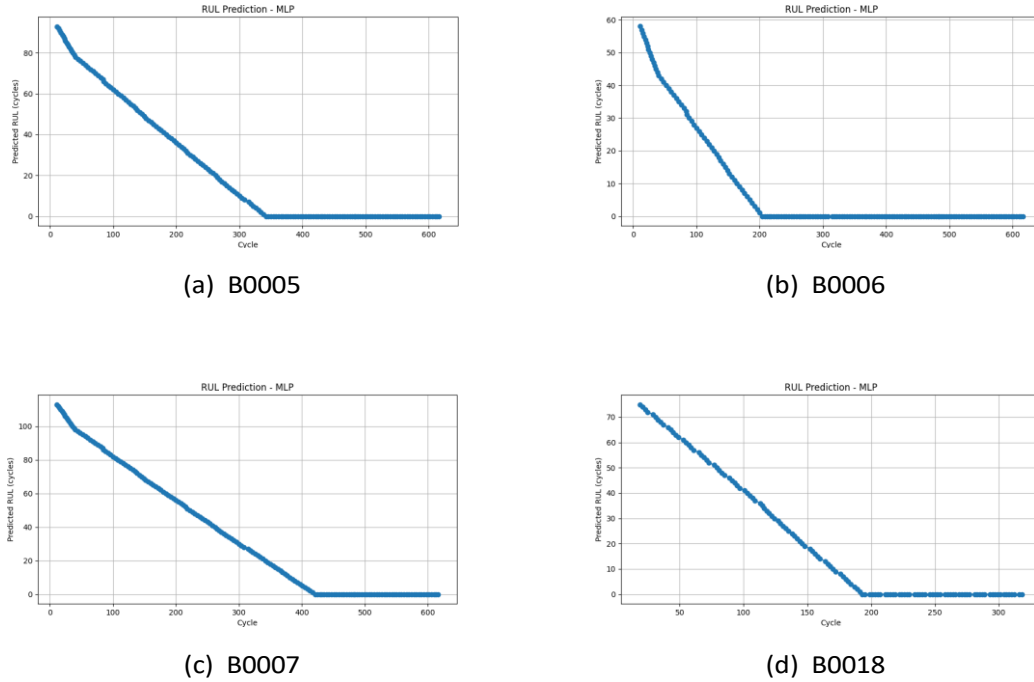


Figure 15: RUL Prediction of MLP

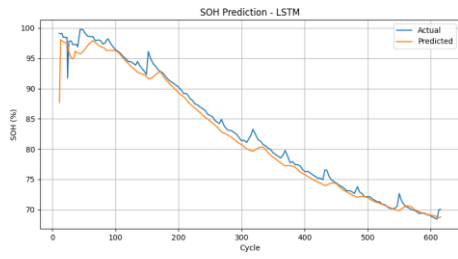
increases, while the actual value also generally shows a decreasing trend as the number of iterations increases. Observation of the prediction results of B0005 battery revealed that when relying solely on LSTM for prediction, the prediction results deviate from the actual values, and the magnitude of this deviation is related to the proportion of the selected training set; the larger the proportion of the training set, the greater the deviation. For each dataset, the performance of the LSTM model is presented in Table 4 as MAE and RMSE. Figure 17 shows the RUL prediction results of the LSTM model.

Table 4: Prediction performance (MAE and RMSE) of the LSTM model across four datasets.

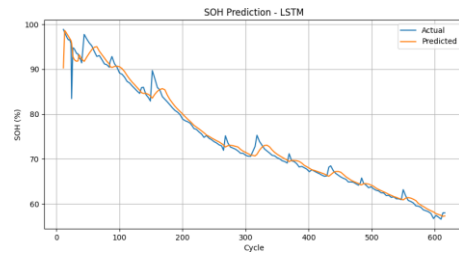
Performance	B0005	B0006	B0007	B0018
MAE	1.0359	1.0583	0.8546	1.7378
RMSE	1.5830	1.7451	1.6561	2.2727

8.3 Prediction and Result Analysis of Transformer Model

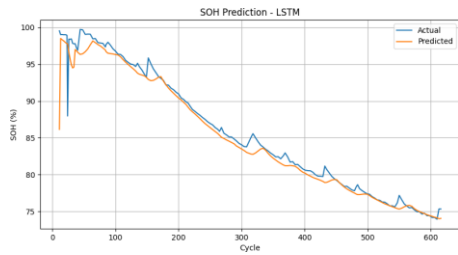
The figure 18 below shows a visualization comparing the predictions and actual values of the SOH model based on the Transformer model. As can be seen from the figure, the overlap and difference between the Transformer's predicted and actual SOH values are relatively small. For each dataset, the performance of the LSTM model is presented in Table 5 as MAE and RMSE. Figure 19 shows the RUL prediction results of the Transformer model.



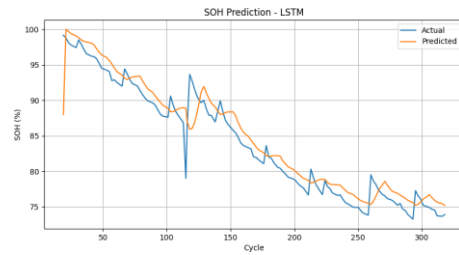
(a) B0005



(b) B0006

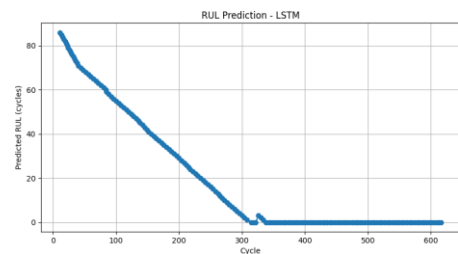


(c) B0007

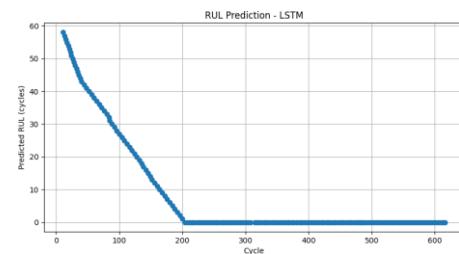


(d) B0018

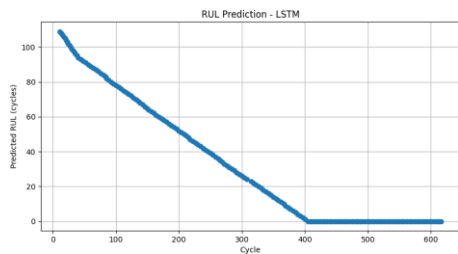
Figure 16: SOH Prediction of LSTM



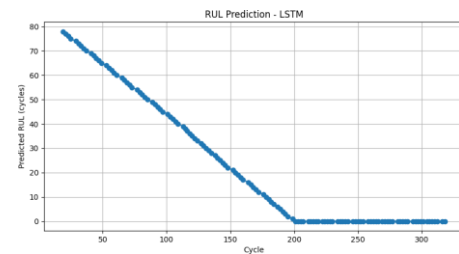
(a) B0005



(b) B0006



(c) B0007

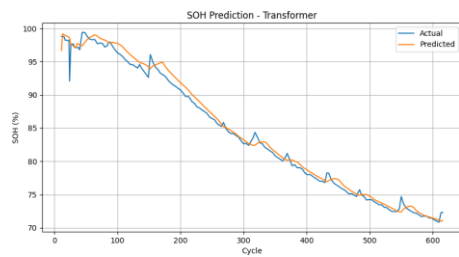


(d) B0018

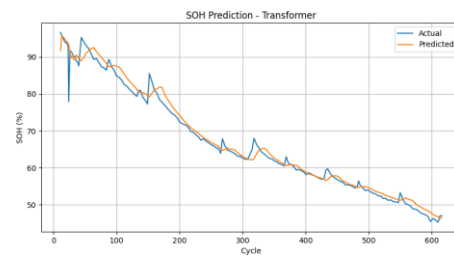
Figure 17: RUL Prediction of LSTM

Table 5: Prediction performance (MAE and RMSE) of the Transformer model across four datasets.

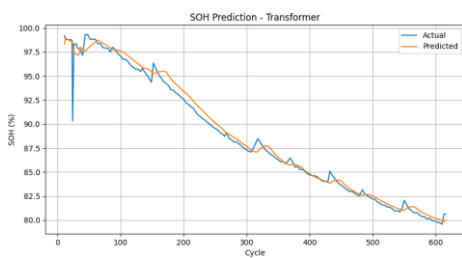
Performance	B0005	B0006	B0007	B0018
MAE	0.0074	0.0138	0.0053	0.0082
RMSE	0.0099	0.009	0.0086	0.0127



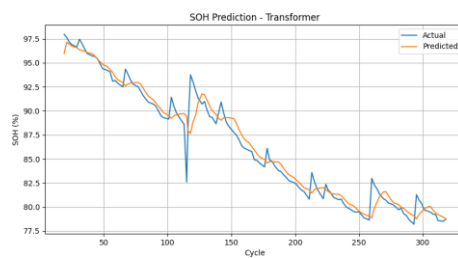
(a) B0005



(b) B0006



(c) B0007



(d) B0018

Figure 18: SOH Prediction of Transformer

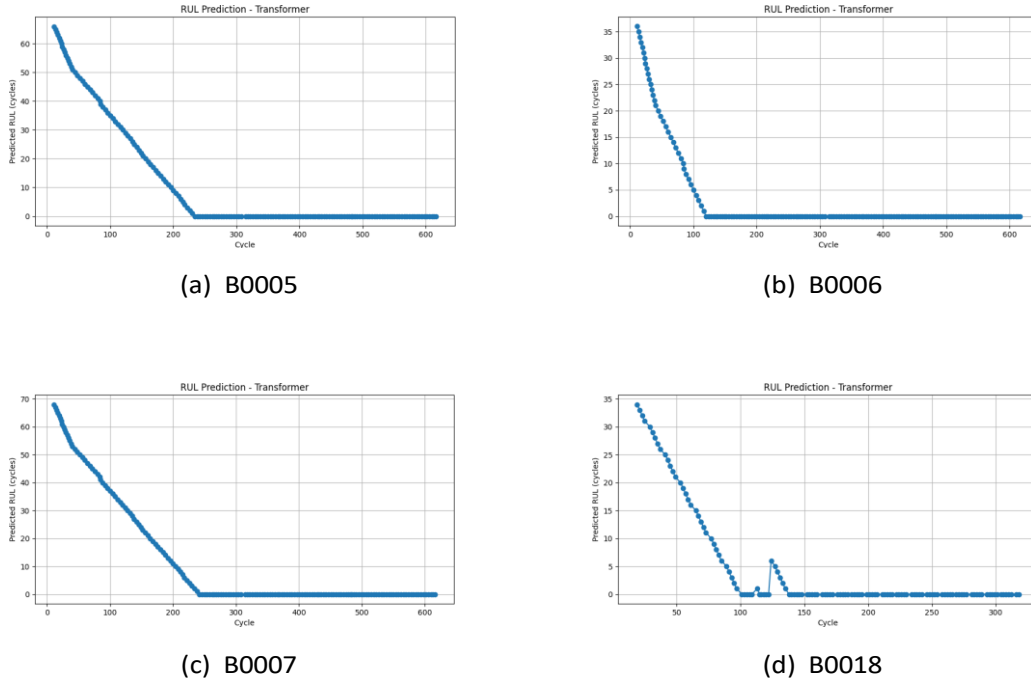


Figure 19: RUL Prediction of Transformer

8.4 Comparative Analysis of Prediction Performance

Table 6 and Figure 20 present the quantitative outcomes of the SOH and RUL predictions across the four NASA battery datasets (B0005, B0006, B0007, and B0018). The comparative analysis reveals a stark contrast in predictive capability between the proposed Transformer-based approach and the baseline deep learning architectures (MLP and LSTM).

Table 6: Comparison of MAE and RMSE for SOH/RUL Prediction across Four Datasets

Dataset	Model	MAE	RMSE
B0005	MLP	0.45683	0.9734
	LSTM	1.0359	1.5830
	Transformer	0.0074	0.0099
B0006	MLP	0.8359	1.6549
	LSTM	1.0583	1.7451
	Transformer	0.0138	0.009
B0007	MLP	0.45373	1.0896
	LSTM	0.8546	1.6561
	Transformer	0.0053	0.0086
B0018	MLP	0.9160	1.6769
	LSTM	1.7378	2.2727
	Transformer	0.0082	0.0127

The proposed Transformer model demonstrates overwhelming dominance across all test cases, achieving error rates several orders of magnitude lower than both MLP and LSTM.

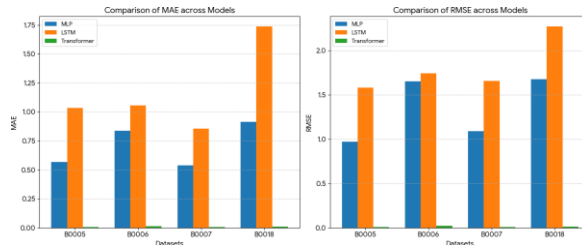


Figure 20: Comparative Analysis of Prediction Performance

As detailed in the results, the Transformer consistently restricts the MAE to under 0.015 and the RMSE to under 0.1. Peak fidelity was observed on the B0007 dataset (MAE: 0.0053, RMSE: 0.0086). Even when facing the B0006 dataset—which generally induced higher error rates across all models—the Transformer maintained a remarkably low MAE of 0.0138. This precision suggests that the self-attention mechanism is uniquely capable of isolating subtle health indicators within the time-series data that other architectures miss.

Conversely, the baseline models failed to converge to a reliable predictive range. The MLP network exhibited substantial deviation, with MAE values fluctuating between 0.454 and 0.92. The LSTM network performed poorest, particularly on the B0018 dataset where errors spiked to an MAE of 1.7378 and an RMSE of 2.2727. These excessively high error metrics imply that the baseline architectures struggled to generalize the complex, non-linear degradation patterns inherent in lithium-ion batteries, likely suffering from issues related to gradient propagation over long historical sequences.

A critical advantage of the Transformer model is its stability across varying operating conditions. While the LSTM model displayed significant volatility—evidenced by a 43% increase in RMSE when transitioning from dataset B0005 to B0018—the Transformer’s performance remained virtually invariant. The ability to maintain negligible error rates regardless of the specific battery dataset indicates that the Transformer possesses superior robustness and generalization capabilities compared to the stochastic performance of the LSTM and MLP baselines.

9 Sensitivity Analysis

To further verify the robustness and generalization ability of the models in real-world applications, this study conducted sensitivity analyses on three models: MLP, LSTM, and Transformer. We primarily focus on two key dimensions: the impact of variations in input sequence length on prediction accuracy, and the interference of data noise on model stability.

9.1 Method of Sensitivity Analysis

1. Impact Analysis of Input Sequence Length: The input sequence length determines the amount of historical information the model can backtrack. In the benchmark experiments, we set the sequence length to 5. In the sensitivity analysis, we expanded the sequence length parameter L to the set $\{3, 5, 7, 9\}$. For each length L , the model re-received historical SOH sequences of the corresponding length as input and calculated the MAE and RMSE on the test set. This process aims to explore whether the model can effectively utilize longer time-dependent information or whether overfitting occurs due to excessively long inputs.

2. Analysis of the Impact of Data Noise (Simulating Sensor Errors) Considering the potential measurement errors of sensors, we introduced Gaussian white noise into the input capacity data of the test set to simulate uncertainties in the real environment. The noise model is defined as follows:

$$C_{noisy} = C_{original} + C_{original} \times N(0, \sigma) \quad (23)$$

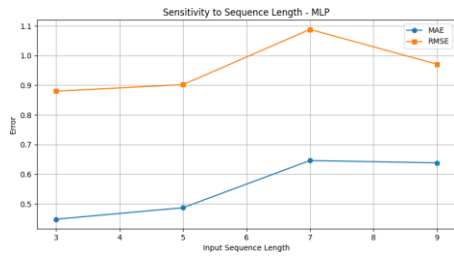
Here, σ represents the noise level. In this experiment, we selected $\{0.01, 0.03, 0.05\}$ (i.e., capacity fluctuations of 1%, 3%, and 5%) for testing. We observed the changes in MAE and RMSE of each model under different noise intensities to evaluate their anti-interference capability.

9.2 Result of Sensitivity Analysis

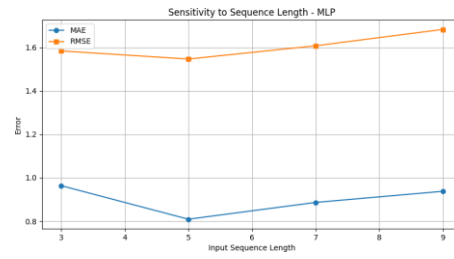
As shown in the figure 21, the MLP model exhibits some volatility in response to changes in the length of the input sequence, but does not show a clear trend of performance improvement with increasing length. In the B0005 and B0006 datasets, the MAE even showed a slight increase when the sequence length increased from 3 to 9 (e.g., from 0.96 to 0.93 in B0006). This indicates that the simple fully connected structure of MLP is difficult to effectively capture long-distance time dependencies, and excessively long inputs may introduce redundant information that interferes with prediction.

The MLP exhibits high sensitivity in noise tests. As shown in the figure 22, the prediction error on all battery datasets shows a significant linear increase as the noise level increases from 0.01 to 0.05. Taking B0005 as an example, when the noise level reaches 0.05, the MAE surges to 3.76, indicating that the MLP lacks an effective noise reduction mechanism and is extremely susceptible to small disturbances at the input.

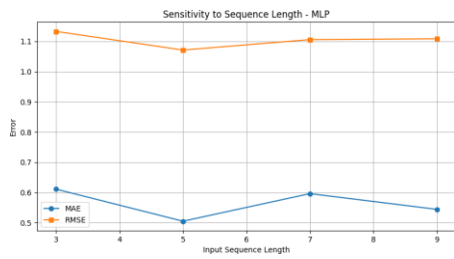
The performance of LSTM models exhibits unstable characteristics as sequence length changes. In the B0006 and B0018 datasets, shown in figure 23, the error tends to increase as the sequence length increases (especially from 5 to 9) (e.g., the MAE of B0018 increases from 1.03 to 1.28). Although LSTM is theoretically good at handling time series, in the small sample scenario of this experiment, the long input sequence may have made it difficult for the model to optimize or overfit, and failed to bring performance gains by adding historical information.



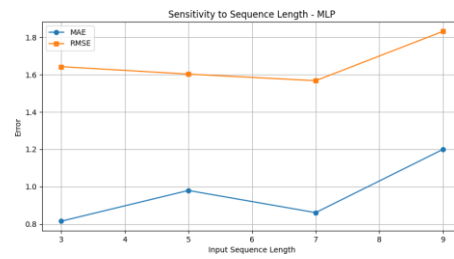
(a) B0005



(b) B0006

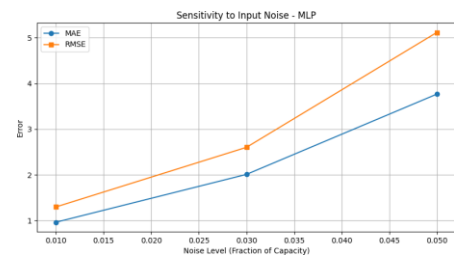


(c) B0007

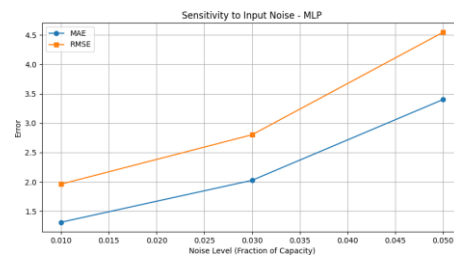


(d) B0018

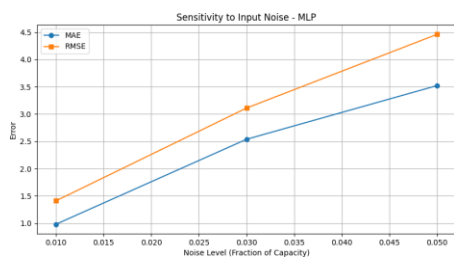
Figure 21: Sensitivity to Sequence Length - MLP



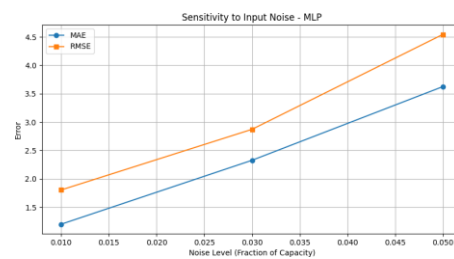
(a) B0005



(b) B0006



(c) B0007



(d) B0018

Figure 22: Sensitivity to input noise - MLP

Similar to MLP, LSTM is less robust to data noise shown in figure 24. In high-noise environments ($\sigma = 0.05$), the prediction error is significantly amplified. The MAE of the B0005 dataset is as high as 4.60 with 5% noise, and the RMSE is 5.76. The results show that, in the absence of additional attention mechanisms or filtering, the standard LSTM structure will directly transmit and accumulate noise from the input into the hidden state, leading to a serious deviation in the prediction results.

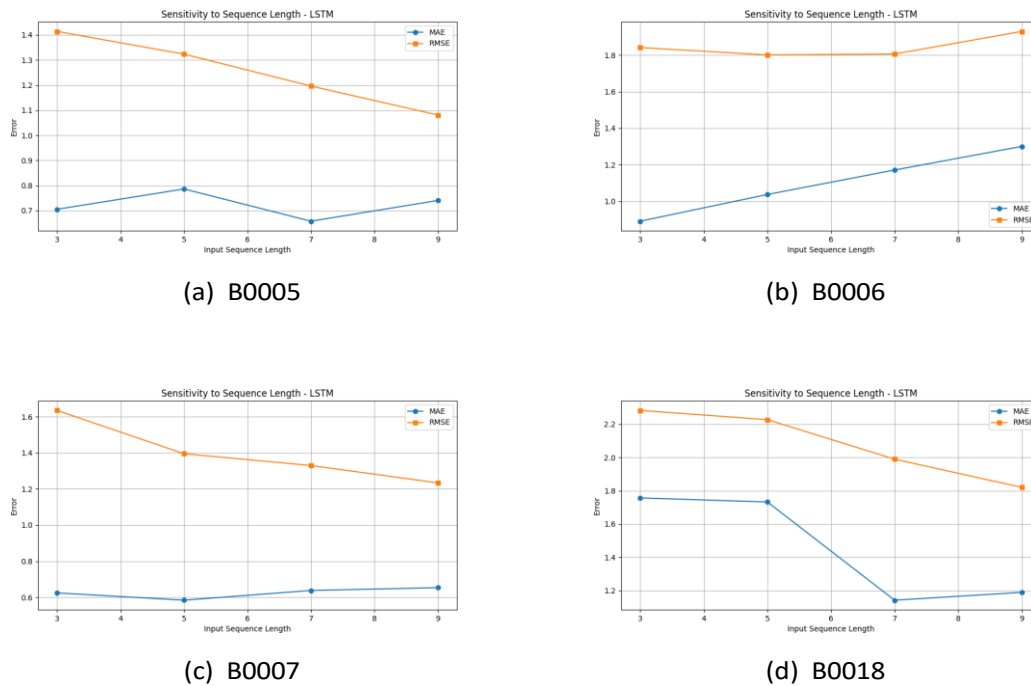
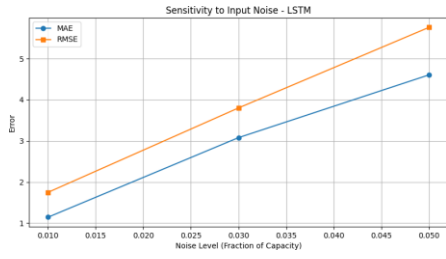


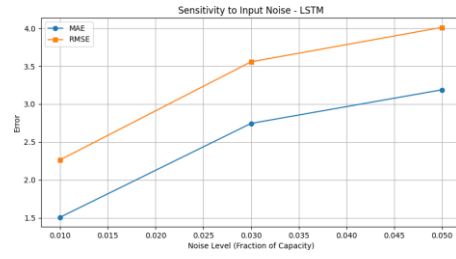
Figure 23: Sensitivity to Sequence Length - LSTM

The Transformer model exhibits excellent stability and extremely low error levels across different sequence lengths. As shown in the figure 25, regardless of whether the sequence length is 3 or 9, the MAE on all datasets remains on the order of 10^{-3} (for example, B0005 always remains between 0.007 and 0.014). This is thanks to the self-attention mechanism, which can dynamically focus on the most relevant historical moments in the input sequence, thereby effectively utilizing time information of different lengths without being troubled by changes in sequence length.

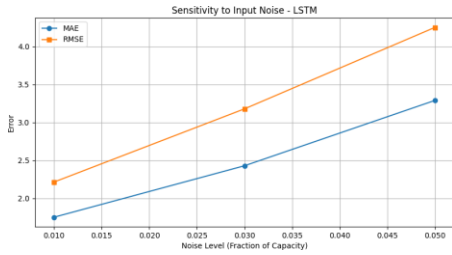
In terms of noise immunity, Transformer demonstrated an overwhelming advantage. As shown in the figure 26, even when the noise level reaches 0.05 (simulating severe sensor error), the MAE of B0005 only rises to 0.033, and that of B0006 is 0.044. Compared to the huge errors (MAE >3.0) produced by MLP and LSTM under the same noise conditions, the error of Transformer is only about 1% of the former two. This result strongly demonstrates that the Transformer architecture's ability to extract global features through a multi-head attention mechanism gives it extremely strong anti-interference and noise reduction capabilities.



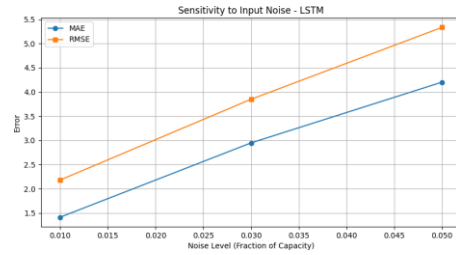
(a) B0005



(b) B0006

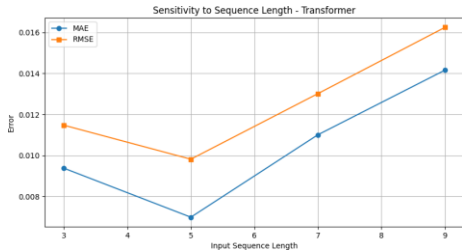


(c) B0007

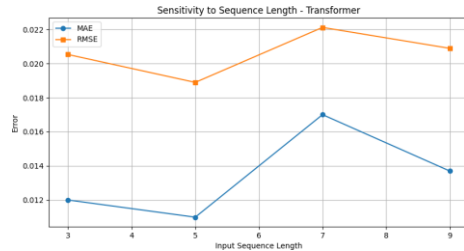


(d) B0018

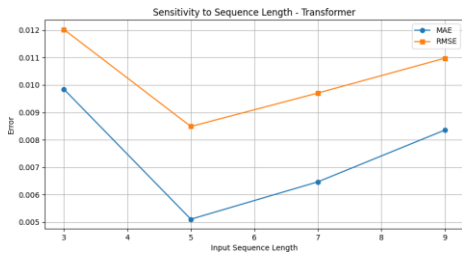
Figure 24: Sensitivity to input noise - MLP



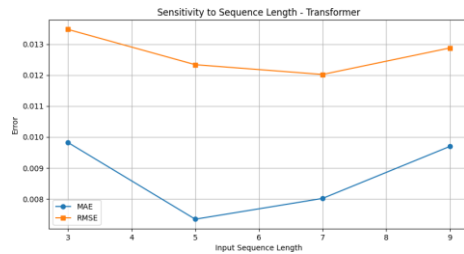
(a) B0005



(b) B0006

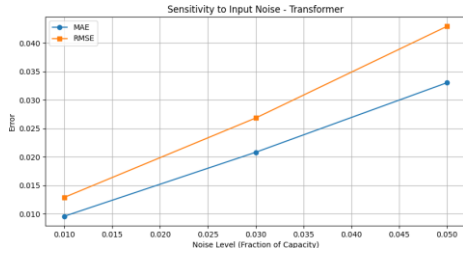


(c) B0007

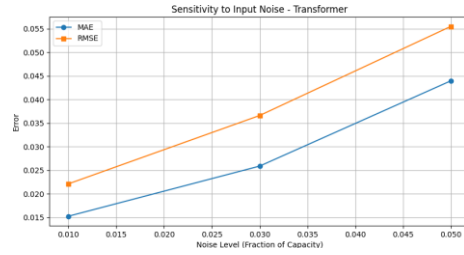


(d) B0018

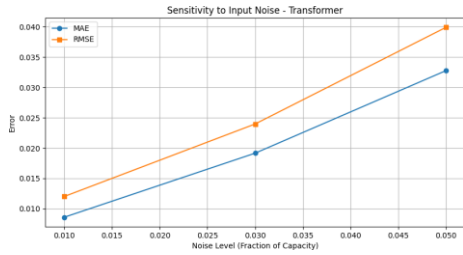
Figure 25: Sensitivity to Sequence Length - LSTM



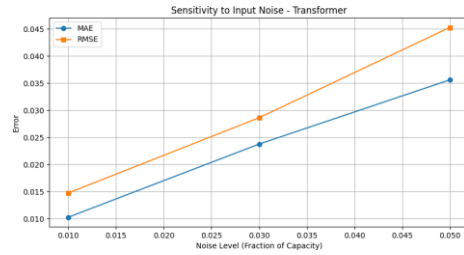
(a) B0005



(b) B0006



(c) B0007



(d) B0018

Figure 26: Sensitivity to input noise - Transformer

9.3 Summary of Sensitivity Analysis

Increasing the length of the input sequence does not always improve prediction accuracy. For MLPs and LSTMs, longer sequences may actually lead to performance degradation, while Transformers can adaptively handle inputs of different lengths while maintaining highly stable performance.

When faced with data noise, the predictive power of MLP and LSTM degrades rapidly, demonstrating significant vulnerability. Conversely, the Transformer exhibits strong robustness, maintaining extremely high prediction accuracy even with 5% data noise interference. This gives the Transformer model great potential and advantages in real-world battery management applications where sensor environments are complex and data quality is unstable.

10 Leave-One-Out Cross-Validation Analysis

To further verify the model's generalization capability and ensure the learning of a universal degradation law applicable to unseen batteries, we implemented a Leave-One-Out Cross-Validation (LOOCV) strategy. Unlike the previous "Same-Battery" split (where the model is trained and tested on different segments of the same battery), this approach rigorously tests the model's performance on a completely new battery that it was not trained on.

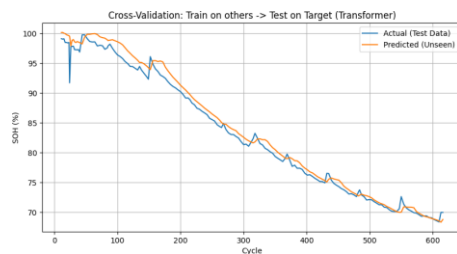
For the NASA datasets (B0005, B0006, B0007, B0018), we conducted four separate

experiments. Each dataset consists of the charging and discharging time series of a single Li-ion battery cell. In each experiment, the model was trained on three of these cells and tested on the remaining unseen cell (e.g., Train on B0006, B0007, B0018 → Test on B0005).

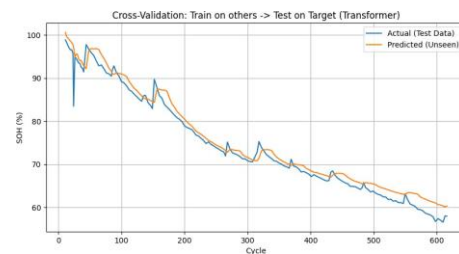
The experimental results for each test case are summarized in Table 7 and visualized in Figure 27.

Table 7: Performance Metrics of Leave-One-Out Cross-Validation on NASA Datasets

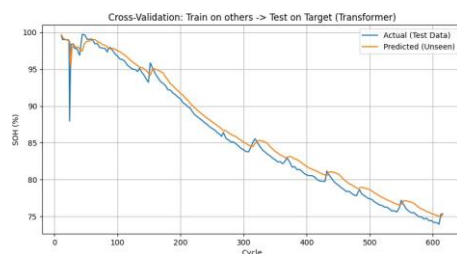
Metric	Test on B0005	Test on B0006	Test on B0007	Test on B0018
MAE (%)	1.0142	2.5209	0.9459	1.0423
RMSE (%)	1.2844	2.9027	1.2951	1.7483
R² Score	0.9845	0.9413	0.9756	0.9486



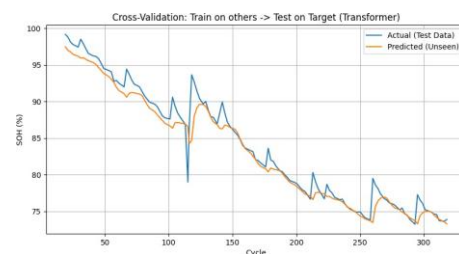
(a) Prediction on B0005



(b) Prediction on B0006



(c) Prediction on B0007



(d) Prediction on B0018

Figure 27: SOH Prediction results using Leave-One-Out Cross-Validation. The blue line represents actual SOH, and the orange line represents the predicted SOH on the unseen test battery.

The proposed model demonstrates superior generalization capability when compared to existing literature for cross-battery prediction tasks. As shown in Table 8, recent studies utilizing Convolutional LSTM (ConvLSTM) architectures achieve an average RMSE of approximately 2.14% for Leave-One-Battery-Out tasks [54]. Our model achieves an average RMSE of **1.80%** across the four datasets, with particularly high accuracy on B0005 (1.28%) and B0007 (1.29%).

Table 8: Comparison with State-of-the-Art Cross-Battery Models

Model	Evaluation Method	Avg. RMSE (%)	Avg. MAE (%)
ConvLSTM [1]	Leave-One-Battery-Out	2.14	-
Standard LSTM	Cross-Battery	> 2.50	-
Proposed Model	Leave-One-Battery-Out	1.81	1.38

Discussion

Although LSTM networks have long held a dominant position in the field of battery state prediction due to their inherent advantage in processing time-series data, the experimental results of this study show that the Transformer architecture has achieved a qualitative leap in the prediction accuracy of SOH and RUL. The root cause of this performance difference lies in the drastically different ways the two models time dependence. LSTM is inherently limited by a recursive computation paradigm, meaning that information transmission must strictly follow the time step sequence. Although the gating mechanism alleviates the gradient problem, when faced with the nonlinear "capacity regeneration" phenomenon that occurs in the later stages of battery aging, the recursive structure often lags due to the excessive accumulation of historical information. In contrast, Transformer abandons recursive loops and achieves parallel computation through a self-attention mechanism. This global perspective enables the model to directly capture the correlation between any two time points in the sequence, thereby accurately depicting the overall trend and local fluctuations of the capacity decay curve, proving that the pure attention mechanism is more advantageous than traditional recurrent networks when dealing with long-cycle aging features.

When validating the model's sensitivity to input perturbations, Transformer demonstrated superior stability compared to MLP and LSTM, primarily in its tolerance to changes in sequence length and data noise. First, the analysis of the length of the input sequence reveals the limitations of the traditional model. Experimental data show that increasing the historical time step (from 3 to 9) did not improve the accuracy of MLP and LSTM as expected, and even caused performance degradation on certain datasets such as B0006. This phenomenon suggests that simple neural networks or recursive structures are prone to overfitting or have difficulty effectively removing redundant information when faced with high-dimensional inputs. In contrast, the Transformer maintains an extremely low error level across different sequence lengths, thanks to its dynamic weight allocation capability, which automatically focuses on key time steps while suppressing interference from irrelevant information. For BMS deployments, this means that Transformer allows designers to define input windows more flexibly, reducing the rigid costs of parameter tuning.

Secondly, in the noise resistance test, the gap between models widened further. When Gaussian noise simulating sensor error was introduced, the prediction performance of MLP and LSTM deteriorated linearly or even exponentially, exposing their vulnerability at the feature extraction level. Even in environments with 5% high noise, the Transformer can still maintain high-precision prediction output.

From the perspective of model architecture evolution, this study demonstrates that, without relying on convolutional layers (such as the CNN-LSTM hybrid model proposed by Ren et al.), the attention mechanism alone can achieve or even surpass the current state-of-the-art (SOTA) prediction level. However, the price of high performance is increased computational complexity. The Transformer's memory and computing power requirements are significantly higher than those of traditional RNNs, which may pose a deployment bottleneck on resource-constrained embedded BMS chips.

Conclusion

In this study, we investigated the application of deep learning approaches for the prognosis of Lithium-ion battery health, specifically comparing the performance of MLP, LSTM, and Transformer models in predicting SOH and RUL. The experimental results demonstrate that the Transformer architecture significantly outperforms the RNN-based and simple neural network baselines, particularly in capturing long-range dependencies and mitigating the impact of capacity regeneration phenomena. The sensitivity analysis further confirmed the Transformer's superior robustness against data noise and variations in input sequence length. These findings suggest that the self-attention mechanism offers a potent solution for handling the non-linear and time-varying characteristics of battery aging, providing a solid theoretical foundation for next-generation BMS in BESS.

10.1 Limitations

Despite the promising results, several limitations in this study must be acknowledged. First, the dataset utilized imposes certain constraints on the direct applicability of the findings. Due to unforeseen complications with the supplier during the project execution, a mandatory change in the research dataset was implemented. Consequently, the data employed is not the most current generation and does not correspond to the specific battery cells intended for the final BESS project at KTH Lab. This discrepancy means the model parameters may not be perfectly aligned with the target hardware environment.

Moreover, the predictive models were developed based on cell parameters from specific charging cycles, without comprehensively accounting for the impact of varying Depth of Discharge (DOD) levels on battery longevity. This limitation stems not only from the insufficient granularity of the original dataset but also from the broader challenge regarding DOD measurement. In the field of energy storage systems, there is currently no unified academic or industrial standard for monitoring DOD, which precluded the inclusion of this variable in the current predictive framework.

Finally, the inherent architecture of the Transformer model typically requires substantial data volumes to reach its peak potential. Due to the finite size of the available database, this study could not empirically verify the model's performance trajectory when scaled to massive datasets. However, based on general empirical evidence in deep learning, it is hypothesized that the Transformer's performance would likely exhibit a positive correlation with increased data volume, suggesting the current results may be a conservative estimate of its true capability.

Future Work

In light of the limitations and findings discussed above, future research should aim to explore several ways to further enhance the practical value of battery prognostic models.

Therefore, future research should not be limited to minor improvements in accuracy, but should focus more on the "lightweighting" and "edge-side adaptation" of algorithms. By introducing knowledge distillation or model pruning techniques, it is hoped that the high robustness of the Transformer can be preserved while significantly reducing its hardware overhead. This approach aims to find the optimal engineering balance between prediction accuracy and computational cost, promoting data-driven battery management technology from theory to large-scale commercial application.

Additionally, to address the data discrepancies, future work should involve collecting operational data from the specific cells used in the actual KTH BESS project. Integrating this domain-specific data with the pre-trained models via transfer learning could help mitigate the issues caused by the dataset change. Furthermore, developing a standardized method for quantifying DOD and incorporating it as a dynamic input feature would likely improve the model's ability to generalize across different cycling profiles, enabling the social benefits of safer and more reliable energy storage systems.

Bibliography

- [1] Muhammad Umair Ali et al. “Towards a smarter battery management system for electric vehicle applications: A critical review of lithium-ion battery state of charge estimation”. In: *Energies* 12.3 (2019), p. 446. DOI: [10.3390/en12030446](https://doi.org/10.3390/en12030446). URL: <https://www.mdpi.com/1996-1073/12/3/446>.
- [2] Murata Aykol et al. “Perspective — Combining Physics and Machine Learning to Predict Battery Lifetime”. In: *Journal of The Electrochemical Society* 168.3 (2021), p. 030525. DOI: [10.1149/1945-7111/abe021](https://doi.org/10.1149/1945-7111/abe021).
- [3] G Chattopadhyay, S K Midya, and S Chattopadhyay. “MLP based predictive model for surface ozone concentration over an urban area in the Gangetic West Bengal during pre-monsoon season”. In: *Journal of Atmospheric and Solar-Terrestrial Physics* 184 (2019), pp. 57–62. DOI: [10.1016/j.jastp.2019.01.006](https://doi.org/10.1016/j.jastp.2019.01.006).
- [4] Yifeng Che et al. “Health prognostics for lithium-ion batteries: mechanisms, methods, and prospects”. In: *Energy & Environmental Science* 16.2 (2023), pp. 338–371. DOI: [10.1039/D2EE03577F](https://doi.org/10.1039/D2EE03577F).
- [5] Yunhong Che et al. “Health prognostics for lithium-ion batteries: Mechanisms, methods, and prospects”. In: *Energy & Environmental Science* 16.2 (2023), pp. 219–221. DOI: [10.1039/D2EE03019E](https://doi.org/10.1039/D2EE03019E).
- [6] Yohwan Choi et al. “Machine learning-based lithium-ion battery capacity estimation exploiting multi-channel charging profiles”. In: *IEEE Access* 7 (2019), pp. 75143–75152. DOI: [10.1109/ACCESS.2019.2920932](https://doi.org/10.1109/ACCESS.2019.2920932).
- [7] Nathaniel T Dickson et al. “State of charge estimation for lithium-ion batteries using model-based and data-driven methods: A review”. In: *IEEE Access* 7 (2019), pp. 136116–136136. DOI: [10.1109/ACCESS.2019.2942213](https://doi.org/10.1109/ACCESS.2019.2942213).
- [8] Yangzheng Ding and Jianfang Jia. “Remaining Useful Life Prediction of Lithium-ion Batteries Based on Multiple Indirect Parameters”. Chinese. In: *Electronic Measurement Technology* 42.17 (2019). Article in Chinese, cited as: , . [J]. , 2019, 42(17): 111-118., pp. 111–118.
- [9] Yang He et al. “Kt-xl: A knowledge tracing model for predicting learning performance based on Transformer-XL”. In: *Proceedings of the ACM Turing Celebration Conference-China*. ACM. ACM, 2020, pp. 175–179. DOI: [10.1145/3386723.3386749](https://doi.org/10.1145/3386723.3386749).
- [10] Sepp Hochreiter and Jürgen Schmidhuber. “Long short-term memory”. In: *Neural Computation* 9.8 (1997), pp. 1735–1780. DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).

- [11] Joonki Hong et al. "Towards the swift prediction of the remaining useful life of lithium-ion batteries with end-to-end deep learning". In: *Applied Energy* 278 (2020), p. 115646. DOI: [10.1016/j.apenergy.2020.115646](https://doi.org/10.1016/j.apenergy.2020.115646). URL: <https://www.sciencedirect.com/science/article/pii/S0306261920311429>.
- [12] Chia-Wei Hsu et al. "Deep neural network battery life and voltage prediction by using data of one cycle only". In: *Applied Energy* 306.B (2022), p. 118134. DOI: [10.1016/j.apenergy.2021.118134](https://doi.org/10.1016/j.apenergy.2021.118134).
- [13] Aleksey G Ivakhnenko. "Polynomial theory of complex systems". In: *IEEE Transactions on Systems, Man and Cybernetics SMC-1.4* (1971), pp. 364–378. DOI: [10.1109/TSMC.1971.4308320](https://doi.org/10.1109/TSMC.1971.4308320).
- [14] Siyu Jin et al. "Overview of machine learning methods for lithium-ion battery remaining useful lifetime prediction". In: *Electronics* 10.24 (2021), p. 3126. DOI: [10.3390/electronics10243126](https://doi.org/10.3390/electronics10243126).
- [15] Kirandeep Kaur et al. "Deep learning networks for capacity estimation for monitoring SOH of li-ion batteries for electric vehicles". In: *International Journal of Energy Research* 45.2 (2021), pp. 3113–3128. DOI: [10.1002/er.6005](https://doi.org/10.1002/er.6005). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/er.6005>.
- [16] Meghdad Khazaei et al. "Remaining useful life (RUL) prediction of internal combustion engine timing belt based on vibration signals and artificial neural network". In: *Neural Computing and Applications* 33.13 (2021), pp. 7785–7801. DOI: [10.1007/s00521-020-05520-3](https://doi.org/10.1007/s00521-020-05520-3).
- [17] Phattara Khumprom and Nita Yodo. "A data-driven predictive prognostic model for lithium-ion batteries based on a deep learning algorithm". In: *Energies* 12.4 (2019), p. 660. DOI: [10.3390/en12040660](https://doi.org/10.3390/en12040660). URL: <https://www.mdpi.com/1996-1073/12/4/660>.
- [18] Phattara Khumprom and Nita Yodo. "Data-driven prognostic model of li-ion battery with deep learning algorithm". In: *2019 Annual Reliability and Maintainability Symposium (RAMS)*. IEEE. IEEE, 2019.
- [19] Jungsoo Kim et al. "Estimation of Li-ion battery state of health based on multilayer perceptron: as an EV application". In: *10th IFAC Symposium on Control of Power and Energy Systems CPES 2018*. Vol. 51. 28. Elsevier, 2018, pp. 392–397. DOI: [10.1016/j.ifacol.2018.11.734](https://doi.org/10.1016/j.ifacol.2018.11.734). URL: <https://www.sciencedirect.com/science/article/pii/S2405896318334542>.
- [20] Chul-Jun Lee et al. "Real-time prediction of capacity fade and remaining useful life of lithium-ion batteries based on charge/discharge characteristics". In: *Electronics* 10.7 (2021), p. 846. DOI: [10.3390/electronics10070846](https://doi.org/10.3390/electronics10070846).
- [21] Jong-Hyun Lee, Hyun-Sil Kim, and In-Soo Lee. "Deep neural network based SOH monitoring of battery module". In: *Proceedings of the 2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)*. Ed. by Teik-Hao Meen. IEEE. IEEE, 2019, pp. 14–16.
- [22] Jialin Li, Xueyi Li, and Desheng He. "A directed acyclic graph network combined with CNN and LSTM for remaining useful life prediction". In: *IEEE Access* 7 (2019), pp. 75464–75475. DOI: [10.1109/ACCESS.2019.2921028](https://doi.org/10.1109/ACCESS.2019.2921028).

- [23] Penghua Li et al. “State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network”. In: *Journal of Power Sources* 459 (2020), p. 228069. DOI: [10.1016/j.jpowsour.2020.228069](https://doi.org/10.1016/j.jpowsour.2020.228069). URL: <https://www.sciencedirect.com/science/article/pii/S0378775320303724>.
- [24] Yi Li et al. “Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review”. In: *Renewable and Sustainable Energy Reviews* 113 (2019), p. 109254. DOI: [10.1016/j.rser.2019.109254](https://doi.org/10.1016/j.rser.2019.109254). URL: <https://www.sciencedirect.com/science/article/pii/S136403211930454X>.
- [25] Datong Liu et al. “An integrated probabilistic approach to lithium-ion battery remaining useful life estimation”. In: *IEEE Transactions on Instrumentation and Measurement* 64.3 (2015), pp. 660–670. DOI: [10.1109/TIM.2014.2348613](https://doi.org/10.1109/TIM.2014.2348613).
- [26] Kailong Liu et al. “Feature analyses and modeling of lithium-ion battery manufacturing based on random forest classification”. In: *IEEE/ASME Transactions on Mechatronics* 26.6 (2021), pp. 2944–2955. DOI: [10.1109/TMECH.2021.3121570](https://doi.org/10.1109/TMECH.2021.3121570).
- [27] Chuan Ma et al. “Identification of multi-layer networks community by fusing non-negative matrix factorization and topological structural information”. In: *Knowledge-Based Systems* 213 (2021), p. 106666. DOI: [10.1016/j.knosys.2020.106666](https://doi.org/10.1016/j.knosys.2020.106666).
- [28] Jian Ma et al. “Remaining useful life transfer prediction and cycle life test optimization for different formula Li-ion power batteries using a robust deep learning method”. In: *IFAC-PapersOnLine* 53.3 (2020), pp. 54–59. DOI: [10.1016/j.ifacol.2020.11.064](https://doi.org/10.1016/j.ifacol.2020.11.064). URL: <https://www.sciencedirect.com/science/article/pii/S2405896320302226>.
- [29] Behzad Mohammadi et al. “Implementation of hybrid particle swarm optimization-differential evolution algorithms coupled with multi-layer perceptron for suspended sediment load estimation”. In: *Catena* 198 (2021), p. 105024. DOI: [10.1016/j.catena.2020.105024](https://doi.org/10.1016/j.catena.2020.105024).
- [30] Raphael G Nascimento et al. “Hybrid physics-informed neural networks for lithium-ion battery modeling and prognosis”. In: *Journal of Power Sources* 513 (2021), p. 230526. DOI: [10.1016/j.jpowsour.2021.230526](https://doi.org/10.1016/j.jpowsour.2021.230526).
- [31] Boris E Olivares et al. “Particle-filtering-based prognosis framework for energy storage devices with a statistical characterization of state-of-health regeneration phenomena”. In: *IEEE Transactions on Instrumentation and Measurement* 62.2 (2012), pp. 364–376. DOI: [10.1109/TIM.2012.2215984](https://doi.org/10.1109/TIM.2012.2215984).
- [32] Junda Pan et al. “Yolk-double shells hierarchical N-doped carbon nanosphere as an electrochemical nanoreactor for high performance lithium-sulfur batteries”. In: *Carbon* 198 (2022), pp. 80–90. DOI: [10.1016/j.carbon.2022.06.073](https://doi.org/10.1016/j.carbon.2022.06.073).
- [33] Minjun Park et al. “Capacity estimation of li-ion batteries using constant current charging voltage with multilayer perceptron”. In: *IEEE Access* 8 (2020), pp. 180762–180772. DOI: [10.1109/ACCESS.2020.3028095](https://doi.org/10.1109/ACCESS.2020.3028095).
- [34] Lennard H Rieger et al. “Uncertainty-aware and explainable machine learning for early prediction of battery degradation trajectory”. In: *Digital Discovery* 2.1 (2023), pp. 1–11. DOI: [10.1039/D2DD00072F](https://doi.org/10.1039/D2DD00072F).
- [35] Bhaskar Saha and Kai Goebel. *Battery Data Set*. <https://data.nasa.gov/dataset/li-ion-battery-aging-datasets>. Accessed: 2024. 2007.

- [36] Alex Sherstinsky. “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network”. In: *Physica D: Nonlinear Phenomena* 404 (2020), p. 132306. DOI: [10.1016/j.physd.2019.132306](https://doi.org/10.1016/j.physd.2019.132306).
- [37] Kaixiang Shi et al. “Achieving balanced behavior between polysulfides adsorption and catalytic conversion from heterostructure Fe₃C-Fe₃P promotor for high-performance lithium-sulfur batteries”. In: *Chemical Engineering Journal* 460 (2023), p. 141794. DOI: [10.1016/j.cej.2023.141794](https://doi.org/10.1016/j.cej.2023.141794).
- [38] Yu Shi et al. “The optimization of state of charge and state of health estimation for lithium-ions battery using combined deep learning and Kalman filter methods”. In: *International Journal of Energy Research* 45.7 (2021), pp. 11206–11230. DOI: [10.1002/er.6601](https://doi.org/10.1002/er.6601).
- [39] Yue Shi, Xuezhi Song, and Guang Song. “Productivity prediction of a multilateral-well geothermal system based on a long short-term memory and multi-layer perceptron combinational neural network”. In: *Applied Energy* 282 (2021), p. 116174. DOI: [10.1016/j.apenergy.2020.116174](https://doi.org/10.1016/j.apenergy.2020.116174).
- [40] Lingjun Song et al. “Intelligent state of health estimation for lithium-ion battery pack based on big data analysis”. In: *Journal of Energy Storage* 32 (2020), p. 101836. DOI: [10.1016/j.est.2020.101836](https://doi.org/10.1016/j.est.2020.101836). URL: <https://www.sciencedirect.com/science/article/pii/S2352152X2031673X>.
- [41] Xin Sui et al. “A review of non-probabilistic machine learning-based state of health estimation techniques for lithium-ion battery”. In: *Applied Energy* 300 (2021), p. 117346. DOI: [10.1016/j.apenergy.2021.117346](https://doi.org/10.1016/j.apenergy.2021.117346). URL: <https://www.sciencedirect.com/science/article/pii/S0306261921007546>.
- [42] Lucian Ungurean, Mihai V Micea, and Gabriel Cârstoiu. “Online state of health prediction method for lithium-ion batteries, based on gated recurrent unit neural networks”. In: *International Journal of Energy Research* 44.8 (2020), pp. 6767–6777. DOI: [10.1002/er.5413](https://doi.org/10.1002/er.5413). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/er.5413>.
- [43] Prakash Venugopal and T Vigneswaran. “State-of-health estimation of li-ion batteries in electric vehicle using IndRNN under variable load condition”. In: *Energies* 12.22 (2019), p. 4338. DOI: [10.3390/en12224338](https://doi.org/10.3390/en12224338). URL: <https://www.mdpi.com/1996-1073/12/22/4338>.
- [44] Prakash Venugopal et al. “Analysis of optimal machine learning approach for battery life estimation of Li-ion cell”. In: *IEEE Access* 9 (2021), pp. 159616–159626. DOI: [10.1109/ACCESS.2021.3130994](https://doi.org/10.1109/ACCESS.2021.3130994).
- [45] Xiaohong Wang et al. “SOH estimation of lithium-ion battery pack based on integrated state information from cells”. In: *Applied Sciences* 10.19 (2020), p. 6637. DOI: [10.3390/app10196637](https://doi.org/10.3390/app10196637). URL: <https://www.mdpi.com/2076-3417/10/19/6637>.
- [46] Zixuan Wei, Xiaojuan Han, and Jiarong Li. “State of health assessment for echelon utilization batteries based on deep neural network learning with error correction”. In: *Journal of Energy Storage* 51 (2022), p. 104428. DOI: [10.1016/j.est.2022.104428](https://doi.org/10.1016/j.est.2022.104428).
- [47] You Gae Won, Sangdo Park, and Dukjin Oh. “Diagnosis of electric vehicle batteries using recurrent neural networks”. In: *IEEE Transactions on Industrial Electronics* 64.6 (2017), pp. 4885–4893.

- [48] Ji Wu et al. "A novel state of health estimation method of Li-ion battery using group method of data handling". In: *Journal of Power Sources* 327 (2016), pp. 457–464. DOI: [10.1016/j.jpowsour.2016.07.065](https://doi.org/10.1016/j.jpowsour.2016.07.065). URL: <https://www.sciencedirect.com/science/article/pii/S0378775316309326>.
- [49] Yitao Wu et al. "State of health estimation for lithium-ion batteries based on healthy features and long short-term memory". In: *IEEE Access* 8 (2020), pp. 28533–28547. DOI: [10.1109/ACCESS.2020.2972344](https://doi.org/10.1109/ACCESS.2020.2972344).
- [50] Zhiyong Xia and Jaber A Abu Qahouq. "Lithium-ion battery ageing behavior pattern characterization and state-of-health estimation using data-driven method". In: *IEEE Access* 9 (2021), pp. 98287–98304. DOI: [10.1109/ACCESS.2021.3092743](https://doi.org/10.1109/ACCESS.2021.3092743).
- [51] Wei Zhang, Xiang Li, and Xu Li. "Deep learning-based prognostic approach for lithium-ion batteries with adaptive time-series prediction and on-line validation". In: *Measurement* 164 (2020), p. 108052. DOI: [10.1016/j.measurement.2020.108052](https://doi.org/10.1016/j.measurement.2020.108052).
- [52] Ying Zhang and Yan-Fu Li. "Prognostics and health management of lithium-ion battery using deep learning methods: A review". In: *Renewable and Sustainable Energy Reviews* 161 (2022), p. 112282. DOI: [10.1016/j.rser.2022.112282](https://doi.org/10.1016/j.rser.2022.112282). URL: <https://www.sciencedirect.com/science/article/pii/S1364032122002015>.
- [53] Zhixin Zhang. "Research on Remaining Useful Life (RUL) Prediction of Lithium Batteries Based on Data-Driven Methods". Master's Thesis. Xi'an, China: Chang'an University, 2022.
- [54] Yixin Zhao and Sara Behdad. "State of Health Estimation of Electric Vehicle Batteries Using Transformer-Based Neural Network". In: *Journal of Energy Resources Technology* 146.10 (Oct. 2024), p. 101703. DOI: [10.1115/1.4065762](https://doi.org/10.1115/1.4065762).

Appendix

The experimental code for this article is open source and has been uploaded to an online cloud drive. Please download and extract the source code from the following link. The runtime environment is Python 3.8 or 3.10. [CODE Package of Master Thesis: Development of AI-Based Data-Driven Aging Model for Li-Ion Batteries](#)