

Article

Remaining Useful Life Prediction of Lithium-Ion Batteries by Using a Denoising Transformer-Based Neural Network

Yunlong Han ¹, Conghui Li ², Linfeng Zheng ³ , Gang Lei ¹  and Li Li ^{1,*} 

¹ Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia; yunlong.han-1@student.uts.edu.au (Y.H.); gang.lei@uts.edu.au (G.L.)

² Zibo Vocational Institute, Zibo 255314, China; 12601@zibvc.edu.cn

³ Institute of Rail Transportation, Jinan University, Zhuhai 510632, China; lfzheng@jnu.edu.cn

* Correspondence: li.li@uts.edu.au

Abstract: In this study, we introduce a novel denoising transformer-based neural network (DTNN) model for predicting the remaining useful life (RUL) of lithium-ion batteries. The proposed DTNN model significantly outperforms traditional machine learning models and other deep learning architectures in terms of accuracy and reliability. Specifically, the DTNN achieved an R^2 value of 0.991, a mean absolute percentage error (MAPE) of 0.632%, and an absolute RUL error of 3.2, which are superior to other models such as Random Forest (RF), Decision Trees (DT), Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Dual-LSTM, and DeTransformer. These results highlight the efficacy of the DTNN model in providing precise and reliable predictions for battery RUL, making it a promising tool for battery management systems in various applications.

Keywords: Li-ion battery; remaining useful life; transformer; residual learning



Citation: Han, Y.; Li, C.; Zheng, L.; Lei, G.; Li, L. Remaining Useful Life Prediction of Lithium-Ion Batteries by Using a Denoising Transformer-Based Neural Network. *Energies* **2023**, *16*, 6328. <https://doi.org/10.3390/en16176328>

Academic Editor: Carlos Miguel Costa

Received: 24 July 2023

Revised: 29 August 2023

Accepted: 30 August 2023

Published: 31 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Lithium-ion batteries power applications such as electric vehicles, power grid systems, and consumer electronics, which have become indispensable to contemporary society [1–3]. As the number of charging and discharging cycles increases, the efficiency of these batteries steadily declines, impacting their overall performance and reliability. A reliable prediction system identifying a battery's health status is essential for the safety and effective functioning of electronic devices and energy storage systems.

A battery's remaining useful life (RUL) is a crucial indicator of its health state. Accurate RUL estimation allows users to schedule maintenance, replacement, and make informed management decisions [4]. With the advent of machine learning and artificial intelligence, numerous data-driven methods for predicting battery life and assessing battery health have emerged [2].

Various models have been employed to forecast the RUL of batteries, including Decision Trees (DT) [5], Random Forest (RF) [6], Multilayer Perceptron (MLP) [7], Long Short-Term Memory (LSTM) [8], Recurrent Neural Networks (RNNs) [9], and Dual-LSTM [10]. While RF, an ensemble method, offers robust predictions by combining multiple decision trees, DT provides simpler, interpretable decision-making structures. MLP, though simple, effectively captures complex non-linear relationships. LSTM and RNNs are tailored for sequential data, with LSTM excelling in capturing long-term dependencies. Dual-LSTM enhances this by processing both past and future information [7–10]. However, these methods often face challenges like overfitting, sensitivity to noise, and computational intensity.

Our research introduces the Denoising Transformer-based Neural Network (DTNN), a novel approach aiming to enhance prediction accuracy and reliability in battery life. Unlike traditional models, the DTNN leverages the Transformer's capability to capture long-term correlations in input sequences. Additionally, it employs a one-dimensional

CNN for denoising, ensuring reduced noise and enhanced prediction accuracy [11,12]. This dual approach addresses the limitations of conventional models, offering potential advancements in accuracy and reliability.

Furthermore, our research identifies challenges in battery life prediction, including data quality issues like noise, outliers, and missing data. There is also a pressing need for models that generalise across various battery types and conditions [2,13,14]. Recognising these challenges, our DTNN method offers enhanced precision and reliability in battery RUL predictions, ensuring the safety and longevity of electronic devices and energy storage systems. This research not only contributes to the exploration of data-driven methodologies, but also lays the groundwork for future advancements in this domain.

2. Literature Review

The State of Health (SOH) serves as a vital metric in estimating the RUL of a battery, as it represents the present condition of the battery's health concerning its original capacity (C_0) when new. The SOH is quantified by expressing the battery's current capacity (C_t) as a ratio of its original capacity:

$$\text{SOH} = \frac{C_t}{C_0} \times 100(\%) \quad (1)$$

A battery's capacity gradually declines throughout the charge and discharge cycles, which causes a drop in SOH. Temperature, charging and discharging rates, and depth of discharge (DoD) [15] are a few variables that affect how quickly capacity degrades [16]. RUL, which stands for a battery's anticipated life before the battery reaches the End of Life (EOL), may be calculated using the current SOH and the rate of capacity decline. The battery is deemed to have come to its EOL and needs to be changed once the SOH falls below a predetermined level [17]. Accurate and fast RUL prediction is crucial for batteries to operate safely and reliably, especially in demanding applications like grid-scale energy storage, aerospace systems, and electric vehicles [18,19].

Scientists and engineers worldwide are putting significant effort into developing various methodologies to predict the RUL and determine the SOH of batteries. These techniques range from traditional empirical models to advanced data-driven algorithms anchored in artificial intelligence and machine learning methods [20,21].

One strategy for predicting lithium-ion batteries' RUL is through model-based methods. These techniques are founded on creating mathematical models that comprehensively depict the battery's physical and chemical properties alongside the degradation processes that take place over time [22–24]. The models subsequently predict the battery's RUL based on its current state and operating conditions. For instance, the University of California researchers in Berkeley have developed a sophisticated model that amalgamates electrochemical, thermal, and mechanical processes to anticipate RUL [22]. At the Massachusetts Institute of Technology, another group proposed a physics-based model for Lithium-ion battery capacity degradation and RUL prediction under various operating conditions [23]. Tsinghua University researchers devised a model-based method that integrates the advantages of electrochemical and equivalent circuit models, enhancing the RUL prediction's accuracy [24]. These model-based techniques have paved the way for advancements in batteries' SOH estimate and RUL prediction [25–27].

With the rise of artificial intelligence, machine learning-based techniques have gained momentum in SOH estimation and RUL prediction. These data-driven methodologies focus on processing historical data and do not require a deep understanding of battery characteristics, making them highly adaptable and user-friendly [14,28,29]. Examples of cutting-edge machine learning approaches are deep, reinforcement, and transfer learning, integrated into SOH estimation and RUL prediction models, potentially enhancing their accuracy and reliability [30].

In the domain of battery RUL prediction, the significance of shallow machine learning techniques, such as DT and RF, cannot be understated. With their inherent simplicity and interpretability, these methods have been foundational in early predictive modelling

efforts. With its clear decision-making structure, DT offers insights into the factors influencing battery degradation, making it invaluable for applications where understanding the model's reasoning is crucial [31]. RF, an ensemble method, leverages the power of multiple decision trees to enhance prediction accuracy, reducing the risk of overfitting and providing more robust predictions [32]. These shallow techniques have been pivotal in early battery RUL prediction models, setting the stage for the subsequent rise of deep learning models. As technology advanced, deep learning models, with their superior generalisation and feature extraction capabilities, began to gain traction in the field of battery RUL prediction [9,33,34].

Deep learning models have gained significant attention in battery RUL prediction due to their superior generalisation capabilities and powerful feature extraction capabilities [9,33,34]. Models like MLP, RNNs, LSTM, GRU, and Dual-LSTM have been suggested. MLP networks, for instance, can learn from the operational history data of the battery to predict its life [9]. However, while the MLP is often considered a shallow neural network, it serves as the baseline deep learning model in our experiments. It captures basic patterns in the data, but may need to improve at modelling more intricate relationships or temporal sequences as effectively as deeper architectures [35].

Researchers have introduced RNN-based frameworks, such as LSTM, to tackle these limitations. These networks can automatically capture long-term dependencies in sequence data and handle variable-length sequence data [8]. Dual-LSTM, an improved version of LSTM, simultaneously learns two different sequence data to provide more accurate RUL predictions [10]. This model has shown superior prediction performance compared to MLP networks, showcasing the power of deep learning techniques in the realm of battery life prediction [36].

Despite the progress made in battery SOH estimation and RUL prediction, several challenges remain, including the improvement of data quality, issues related to noise, outliers, and missing data, and enhancing the prediction models' generalisability and scalability for different battery types and use cases [11,30,37]. Future research directions focusing on these challenges, coupled with the incorporation of advanced machine learning techniques, promise to further increase the accuracy and reliability of SOH estimation and RUL prediction models.

A SucMulti-Head Self-Attention sublayer aimstation for specific applications will significantly enhance battery-powered systems' overall safety, effectiveness, and durability. These advantages will facilitate better-informed decisions about battery management, replacement, and maintenance, improving the performance and safety of electronic devices, electric vehicles, aerospace systems, and grid-scale energy storage systems [11].

3. Methodology

The proposed method seeks to leverage historical data to forecast the capacity of lithium-ion batteries better, employing the DTNN model as its foundation. The DTNN model, in its fundamental structure, comprises four primary components: input normalisation, denoising, transformer layers, and prediction. In our study, we have extended this original architecture by integrating pertinent findings from a range of published literature to afford a comprehensive understanding of the model, an outline of which can be observed in Figure 1.

We have incorporated an enhanced design for the initial input stage that introduces one-dimensional convolutions (1-D Conv) into every model layer. The intention behind this integration is to filter and capture meaningful local data features that provide insight into the behaviour of lithium-ion batteries. These locally extracted features are combined to generate a global characteristic representation through an addition operation. This global representation carries a broader view of the dataset, providing a comprehensive understanding of the battery capacity trends.

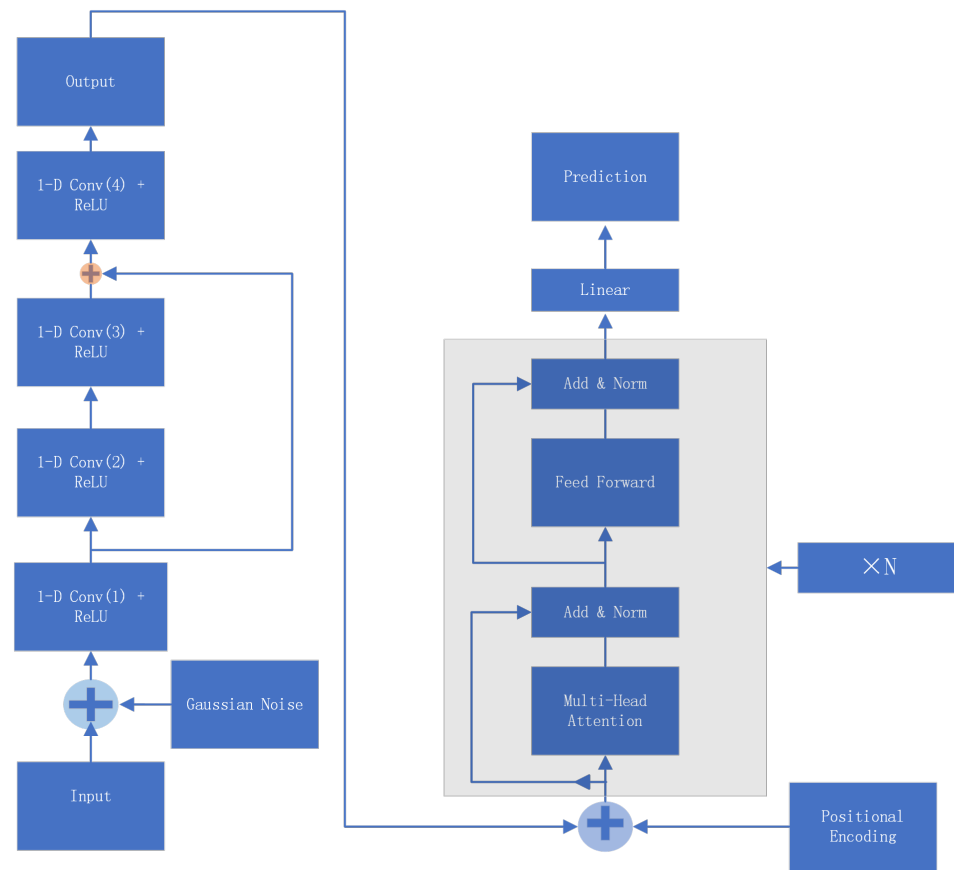


Figure 1. Flowchart for capacity prediction by using DTNN.

In addition, our modified method aims to enhance the overall accuracy and reliability of the predictions by reducing potential noise within the data. To this end, we have incorporated residual learning into the model. This technique facilitates the reduction of noise points in the images, improving the clarity of the data and, subsequently, the precision of the results.

Regarding data encoding, we have elected to use absolute positional encoding. This decision is based on the fact that our data adheres strictly to temporal sequencing rather than relative positioning. This encoding methodology allows for the temporal nature of our data to be accurately represented within the model, thereby enhancing its predictive capabilities.

Furthermore, we have made alterations to the original transformer layer. Specifically, we have eliminated the masked multi-head attention that characterised the original model, primarily because it was deemed unsuitable for applications to time series data. This modification not only simplifies the model but also enhances its robustness, contributing to the overall reliability of the prediction results.

Our modified transformer model effectively harmonises critical feature extraction processes, noise reduction, and temporal recognition. This balanced and integrated approach results in a model capable of producing accurate and robust predictions of lithium-ion battery capacities, leveraging historical data to anticipate future trends. This approach holds significant promise for the continued study and development of efficient lithium-ion battery management systems.

3.1. Input Denoising

In our approach, the first step involves input data normalisation, an essential pre-processing operation that standardises the sequence of battery capacities into a range typically between 0 and 1. This normalisation plays a crucial role in ensuring the stability

and robustness of our neural network model, effectively safeguarding it from potential disruptions due to variations in data distribution.

Following this, we embark on the denoising process for the battery data. In our pre-processing strategy, we apply one-dimensional convolution coupled with residual learning to reduce image noise, an often overlooked but accuracy-enhancing approach. By appropriately increasing the data width, we enhance the distinctiveness of data features, thereby boosting the effectiveness of denoising.

Simultaneously, we introduce Gaussian noise and use a denoising encoder to minimise interference, a critical part of the denoising process. After the third layer of image processing convolution, we incorporate residual learning, further reducing image noise and enhancing the precision of denoising. This strategy compensates for the denoising encoder's shortcomings in handling image noise, thereby increasing our model's predictive accuracy.

Therefore, our pre-processing steps include input data normalisation, one-dimensional convolution, residual learning, and the use of a denoising encoder. They effectively diminish noise interference and pave the way for subsequent processing stages.

3.2. Transformer

The conventional architecture of the Transformer consists of a sequence-to-sequence framework, which includes an encoder and a decoder. The encoder is responsible for taking an input sequence and converting it into a vector with a high number of dimensions. Subsequently, this vector is inputted into the decoder to generate a sequence of outputs [11]. In this study, we employ a transformer-based encoder to capture long-term dependencies related to capacity degradation in battery operation records.

In our research, we utilise a configuration of transformer encoders in a stacked architecture to extract salient features from regenerated data indicative of battery degradation. Each encoder is bifurcated into two integral sublayers: multi-head self-attention and a feed-forward network.

We introduce positional encoding (PE) to account for the sequence's temporal dimension, an essential aspect overlooked by the inherent design of Transformer models. For this purpose, sine and cosine functions are utilised at different frequencies to represent relative positional encoding within the sequence [38]:

$$PE(t, 2i) = \sin\left(t/10000^{2i/m}\right) \quad (2)$$

$$PE(t, 2i + 1) = \cos\left(t/10000^{2i/m}\right) \quad (3)$$

where t is the time step, i is the dimension of the feature, and m is the length of the input sequence.

The Multi-Head Self-Attention sublayer aims to identify the relationships among features while disregarding their relative positions within the sequence [39–41]. The y -th attention ($y \in [1, h]$) is defined based on the representation of the $(l - 1)$ -th layer, denoted as \mathbf{H}^{l-1} , and h parallel attention functions:

$$\text{head}_y = \text{Attention}\left(\mathbf{H}^{l-1}\mathbf{W}_Q^l, \mathbf{H}^{l-1}\mathbf{W}_K^l, \mathbf{H}^{l-1}\mathbf{W}_V^l\right) \quad (4)$$

The matrices $\{W_Q^l, W_K^l, W_V^l\} \in R^{d \times d_h}$ are the project weights. The concept of projection weights refers to the assignment of numerical values to different variables or factors to determine their relative importance or contribution. The variable l denotes the layer within a transformer model, \mathbf{H} represents the hidden states within the transformer model, and h represents the quantity of 'heads' in a multi-head self-attention mechanism. Let Q_l , K_l , and V_l represent the query, key, and value, respectively. In practical implementation, the attention function is computed simultaneously on a set of queries packed together

into a matrix Q_l . Similarly, the keys and values are also packed into matrices K_l and V_l . The output matrix is computed as follows [38]:

$$\text{Attention}(Q_l, K_l, V_l) = \text{softmax}\left(\frac{Q_l K_l^T}{\sqrt{d_h}}\right) V_l \quad (5)$$

where $d_h = d/h$.

This methodology mitigates the problem of vanishingly small gradients and concurrently facilitates a more uniform attention distribution. Consequently, the multi-head attention mechanism can be characterised as follows:

$$\text{multi-head}(H^{l-1}) = [\text{head}_1; \text{head}_2; \dots; \text{head}_h] W^O \quad (6)$$

where the weight W^O is subject to training.

The Feed-Forward Network is utilised to apply two distinct mappings, namely linear and ReLU non-linear, to each time step identically and independently. Next, we obtain the value of H^l from the previous multi-head layer (H^{l-1}) using the following procedure [11]:

$$H^l = \text{FFN}\left(\text{multi-head}(H^{l-1})\right) \quad (7)$$

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2 \quad (8)$$

where $W_1 \in \mathbb{R}^{d_{\text{model}} \times d_{\text{interm}}}$ and $W_2 \in \mathbb{R}^{d_{\text{interm}} \times d_{\text{model}}}$ are the weight matrix, $b_1 \in \mathbb{R}^{d_{\text{model}}}$ and $b_2 \in \mathbb{R}^{d_{\text{interm}}}$ are the bias, $\text{ReLU}(x) = \max(0, x)$, d_{model} represents the vector dimension of input and output sequence elements, and d_{interm} is the dimension of the hidden layer mapping before ReLU activation.

3.3. Prediction

In predicting battery capacity, we employed the attention-based DTNN model. With its self-attention mechanism, this model can handle dependencies at any position within the input sequence. It offers significant advantages in capturing battery usage patterns' complexity and time dependency.

In practical applications, we leverage all the DTNN model's connected layers to map the last unit's information for future battery capacity prediction. The optimisation of the model is achieved by minimising the discrepancy between the predicted values and the actual battery capacities, providing high accuracy and robustness for the battery capacity prediction task.

3.4. Learning

In the learning process of our battery capacity prediction model, we utilised an objective function to optimise the tasks of denoising and prediction simultaneously [11]. This objective function is defined as:

$$\mathcal{L} = \sum_{t=T+1}^n (x_t - \hat{x}_t)^2 + \alpha \sum_{i=1}^n \ell(\tilde{x}_i - \hat{x}_i) + \lambda \Omega(\Theta),$$

Here, x_t is the t -th capacity of x , \hat{x}_t is the predicted value of x_i ; letting $x_i = \{x_{i+1}, x_{i+2}, \dots, x_{i+m}\}$ be the slice of input with m samples of a sequence, then \hat{x}_i is the predicted value of x_i , \tilde{x}_i is the vector after Gaussian noise is added to x_i , α and λ are parameters that control the relative contribution of each task and the regularisation level, respectively, $\ell(\cdot)$ is a loss function, and Θ denotes the learning parameters of our model. Through this approach, the learning process of our model not only focuses on the accuracy of battery capacity prediction, but also minimises the impact of noise on the prediction results.

3.5. Complexity Analysis of the DTNN Method

The DTNN method, as presented in this study, leverages the power of transformers, which inherently have a computational complexity of $O(n^2)$ for sequence length n . This quadratic complexity arises from the self-attention mechanism, where each element in the sequence attends to every other element. However, it is worth noting that the benefits of this mechanism, such as the ability to capture long-range dependencies in the data, often outweigh the computational costs, especially for shorter sequences. In battery life prediction, where sequences might not be exceedingly long, the DTNN method remains computationally feasible. Moreover, while adding to the algorithmic intricacy, the denoising aspect of the model does not significantly increase the computational complexity, but provides robustness against noisy data. Such denoising capabilities are invaluable in real-world scenarios, where data might be corrupted or incomplete, while the DTNN method is more complex than traditional methods, its accuracy and robustness justify the increased computational costs.

4. Experiment Setup

In this study, we utilised a publicly accessible dataset, specifically the one provided by NASA Ames Research Center. The dataset encapsulates the properties of four distinct lithium batteries, each undergoing three cyclical processes: charging, discharging, and impedance measurement. This data acquisition from NASA's resources allowed us to explore nuanced patterns within these battery cycles.

To assess the RUL prediction performance, we utilised six evaluation metrics: Relative Error (RE), Mean Absolute Error (MAE), Root Mean Square Error ($RMSE$), R^2 , Mean Absolute Percentage Error ($MAPE$), and RUL Error (RUL_{er}). These metrics provide comprehensive insight into the predictive model's performance. The four evaluation indicators are set as follows [11]:

$$RE = \frac{|RUL_{true} - RUL_{pred}|}{|RUL_{true}|} \quad (9)$$

$$RMSE = \sqrt{\frac{1}{n-T} \sum_{t=T+1}^n (x_t - \hat{x}_t)^2} \quad (10)$$

$$MAE = \frac{1}{n-T} \sum_{t=T+1}^n \|x_t - \hat{x}_t\| \quad (11)$$

$$R^2 = 1 - \frac{\sum_{t=1}^n (x_t - \hat{x}_t)^2}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad (12)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{x_t - \hat{x}_t}{x_t} \right| \times 100 \quad (13)$$

$$RUL_{er} = |RUL_{pred} - RUL_{true}| \quad (14)$$

In this context, n represents the length of a sequence, and T represents the length of samples generated from a series specifically for training purposes.

We used a leave-one-out methodology in the evaluation stage; a random battery sample was chosen, and the remaining batteries were used to train our model. A new battery sample was used for validation throughout each of the five iterations of this method. The average of the results from all batteries throughout these iterations was used to calculate the final indicator of the model's performance.

The model under consideration encompasses six crucial parameters: the sample size, the learning rate, the depth, the hidden size, the transformer regularisation, and the task ratio. The sample size value can be assigned to 5 to 10% of the sequence length.

In the process of hyperparameter tuning, we employed a grid search methodology to optimise six key parameters: the sample size, learning rate, depth, hidden size, transformer

regularisation, and task ratio. The grid search was conducted over multiple iterations, each assessing a unique combination of hyperparameters. We utilised a five-fold cross-validation scheme to ensure a robust evaluation, with the Mean Squared Error (MSE) serving as the performance metric. The learning rate was selected from a pre-defined set $\{10^{-4}, 5 \times 10^{-4}, 10^{-3}, 5 \times 10^{-3}, 10^{-2}\}$. The depth value was restricted to $\{1, 2, 3, 4\}$, and the transformer regularisation was chosen from the set $\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}\}$. The task ratio was selected from the interval (0,1) [11]. A constant sample size of 16 was chosen across all experiments, which was influenced by specific parameters from NASA's battery dataset. A summary of the grid search results indicated that a combination of specific hyperparameters yielded the best performance in terms of MSE, thereby guiding the final model configuration.

In this work, to further illustrate the efficiency and robustness of our model, we conducted comparison experiments with various other popular machine learning architectures: DT, RF, MLP, RNNs, LSTM, GRU, Dual-LSTM and DeTransformer [11]. Each model was trained and tested under similar conditions to maintain a fair comparison. The sample size was consistently set at 16 for all models.

The learning rate for MLP, RNNs, LSTM, GRU, Dual-LSTM and DeTransformer models was set at 10^{-2} , 10^{-3} , 10^{-3} , 10^{-3} , 10^{-3} and 5×10^{-3} , respectively. Similarly, the depth for DT, RF, MLP, RNNs, LSTM, GRU, and Dual-LSTM models was set at 2 and DeTransformer at 1. The hidden size, another significant factor influencing the models' performance, was set at 8 for MLP and RF, 32 for DeTransformer, and 64 for DT, RNNs, LSTM, GRU, and Dual-LSTM. Furthermore, to regularise the models and prevent overfitting, the transformer regularisation was set at 10^{-6} , as shown in Table 1.

Each model's performance was then evaluated based on the accuracy of its battery capacity predictions. This comparative analysis offered critical insights into the performance variations among different deep learning models. It demonstrated how specific models are more effective than others in handling complex, time-dependent sequences and noisy data—characteristics intrinsic to battery capacity prediction. Moreover, it shows how our proposed DTNN model outperforms these standard architectures in accurately predicting battery capacity, demonstrating its significant potential for real-world applications.

RE has the highest correlation with battery RUL among the four evaluation metrics of battery RUL, so we use RE as our main evaluation index.

Table 1. Optimal parameter of RE score for NASA dataset.

Dataset Models	Sample Size	Learning Rate	Depth	Hidden Size	Trans Reg
DT	16	0.001	2	64	10^{-6}
RF	16	0.01	2	8	10^{-6}
MLP	16	0.01	2	8	10^{-6}
RNN	16	0.001	2	64	10^{-6}
LSTM	16	0.001	2	64	10^{-6}
GRU	16	0.001	2	64	10^{-6}
Dual-LSTM	16	0.001	2	64	10^{-6}
DeTransformer	16	0.005	1	32	10^{-6}
DTNN	16	0.005	1	32	10^{-6}

5. Experiment Result

5.1. Comparative Analysis and Evaluation

The performance of our method has been validated through experiments conducted on various datasets. Table 2 presents the results of the R^2 , MAPE, RE, MAE, RMSE, and RUL_{er} scores achieved by different methods.

Table 2. Comparison of performance metrics for different models.

	RF	DT	MLP	RNN	LSTM	GRU	Dual-LSTM	DeTransformer	DTNN
RE	0.2969	0.3997	0.3871	0.2924	0.2716	0.3342	0.2641	0.2312	0.0351
RMSE	0.0962	0.1522	0.1402	0.0827	0.0952	0.0916	0.0831	0.0792	0.005
MAE	0.0838	0.163	0.1564	0.0744	0.0866	0.0912	0.0883	0.0852	0.0272
R^2	0.977	0.971	0.972	0.965	0.968	0.967	0.969	0.975	0.991
MAPE	1.431	1.672	1.215	1.542	1.479	1.452	1.353	1.120	0.632
RUL_{er}	26.1	35.3	34.6	25.6	23.8	27.4	23	20.8	3.2

Our in-depth experimental investigation was designed to test and compare a multitude of different models on their proficiency in making accurate predictions, with a particular focus on the NASA dataset. These experiments were pivotal in helping us understand the relative strengths and weaknesses of these models when applied to real-world data.

The proposed DTNN model demonstrated the most superior results among the various models tested. Table 2 compares the models based on six crucial metrics: RE, R^2 , MAPE, MAE, RMSE and RUL_{er} . Notably, the DTNN stood out by consistently scoring the highest across these metrics, thus underlining its superior prediction capabilities.

When examining our DTNN compared with the DeTransformer as our primary control group, a significant difference in performance is apparent across all evaluation metrics. Our DTNN model excels in denoising and showcases superior predictive accuracy, embodying an integrative and synergistic approach to the problem. This superiority is evident with a mere RE of 0.0351 for DTNN, a substantial improvement over the DeTransformer's RE of 0.2252. Furthermore, our DTNN outperforms the DeTransformer in other key metrics, such as RMSE, MAE, and RUL_{er} , signifying a substantial leap in predictive accuracy and reliability.

Beyond the DeTransformer, the DTNN model stands out compared to other baseline methods, including MLP, RNN, LSTM, GRU, and Dual-LSTM. Despite these models having different optimal parameters, the DTNN consistently shows lower error rates across all evaluation metrics. This is a testament to DTNN's robustness and versatility in handling complex time-series prediction tasks, even compared to other advanced deep learning architectures.

A distinct advantage of the DTNN model, evident through the experiments, is its robustness and stability. It efficiently provides reliable predictions irrespective of whether the capacity sequence is long or short. This commendable performance is primarily attributed to the model's adeptness at extracting critical temporal information from the capacity sequences, which plays a vital role in accurate prediction.

As we delved deeper into the baseline methods, it was observed that the MLP did not meet the standards set by the other models. Its primary drawback lies in its inability to adequately account for temporal information, which is crucial for the accurate prediction of RUL. Contrarily, our model and the other RNN-based models performed significantly better, predicting trends more accurately than the MLP, reinforcing the necessity of integrating sequential information into these models for proficient RUL prediction.

The attention networks inherent in the DTNN model are designed to adeptly capture broad patterns by effectively modelling relationships among historical capacity attributes. This ability allows our model to proficiently simulate the impacts of historical capacities on sequence states, significantly boosting its overall performance. Particularly in the case of the NASA dataset, the DTNN model proved superior to the others in terms of RE metrics, which are directly tied to predicting a battery's RUL. Our models utilise a denoising encoder to improve further representation and reduce raw sequence noise.

The following complex description of the dataset and statistical analysis of the dataset followed the method in [42]. The NASA dataset focuses on the 18,650 Li-ion battery, utilising accelerated ageing experiments for data collection. These batteries are categorised into nine groups, each containing 3–4 lithium batteries with a rated capacity of 2 Ah. We selected B05, B06, B07, and B18 as our experimental and prediction subjects. The experimental

environment was maintained at a constant temperature of 24 °C. The discharge cut-off voltages were set at 2.7 V, 2.5 V, 2.2 V, and 2.5 V, respectively, with a continuous discharge current of 2 A and a charging current of 1.5 A. The Electrochemical Impedance Spectroscopy (EIS) frequency ranged between 0.1 and 5 kHz. The charging process involved maintaining a temperature of approximately 24 °C and charging with a steady current of 1.5 A until the working voltage reached the maximum cut-off voltage of 4.2 V. The charging then shifted to constant voltage until the current dropped to 20 mA. The discharge process involved a continuous current discharge at 1 C until the working voltage of the four batteries dropped to their respective minimum cut-off voltages. Impedance measurements were intermittently conducted between charging and discharging cycles to record battery resistance.

Figure 2 shows that the battery's releasable capacity gradually decreases as the charge–discharge cycle progresses. Interestingly, there is a phased increase in the capacity decay process, termed “capacity self-recovery”. This phenomenon occurs when the battery's charge–discharge ends, and a short-term placement results in a temporary localised increase in capacity. This is attributed to the battery's internal reaction formula reactants accumulating on the electrode, weakening the internal reaction. When placed aside, these inductors have a chance to dissipate, thereby increasing the capacity for the next charge–discharge cycle. This is a manufacturer setting to ensure that after battery ageing, the usable capacity remains as consistent as possible with a new battery. Figure 2 shows that the number of capacity data of B18 is much smaller than that of the other three batteries in the same battery pack. This is due to NASA's consideration that the EIS test frequency will somewhat affect the battery's health, where batteries B05, B06, and B07 underwent 278 EIS tests, while B18 only underwent 53 EIS tests.

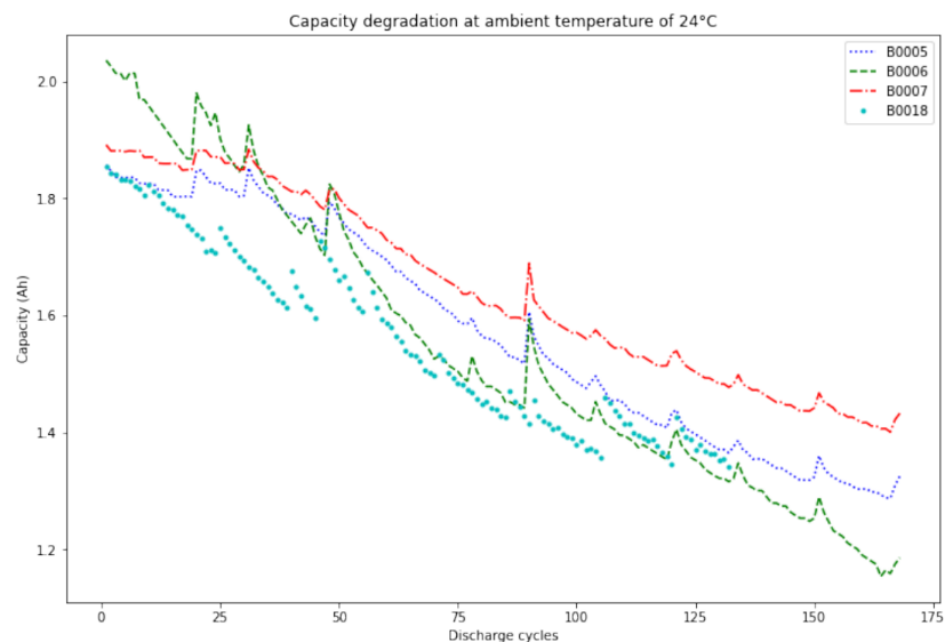


Figure 2. NASA battery data for capacity degradation.

Figures 3–6 demonstrate the prediction performance of the proposed DTNN method for batteries B05, B06, B07 and B18, where the y-axis is the State of Charge (SOC). In these tests, NASA employed the BatteryAgingARC-FY08Q4 model as the test group in this controlled environment. The charging procedure was consistent with the aforementioned method, and a steady 2 A current was sustained during the discharge phase until the batteries reached their respective voltages. Our methodology was applied to predict the NASA dataset, utilising the initial 60% of the data for training and the remainder for prediction. The results showcased a minimal discrepancy between forecasted values and actual experimental outcomes. We established the standard battery usage time as the duration before its capacity fell below 70% of its initial value, represented by a dashed

line in the figure. This comprehensive approach provides a holistic view of the battery’s performance and degradation over time, ensuring accurate predictions and insights.

Figure 7 presents the boxplots of prediction errors for different batteries using the DTNN method. The y-axis represents the difference between the predicted values and the actual values, which indicates the prediction error.

The boxplot shows that the median prediction error for all batteries is deficient, between 0 and 0.0023. This suggests that the DTNN method is generally accurate in its predictions. However, some outliers are represented by the discrete points outside the main body of the boxplot. These outliers, especially those significantly distant from the main plot, indicate instances where the prediction was notably off from the actual value.

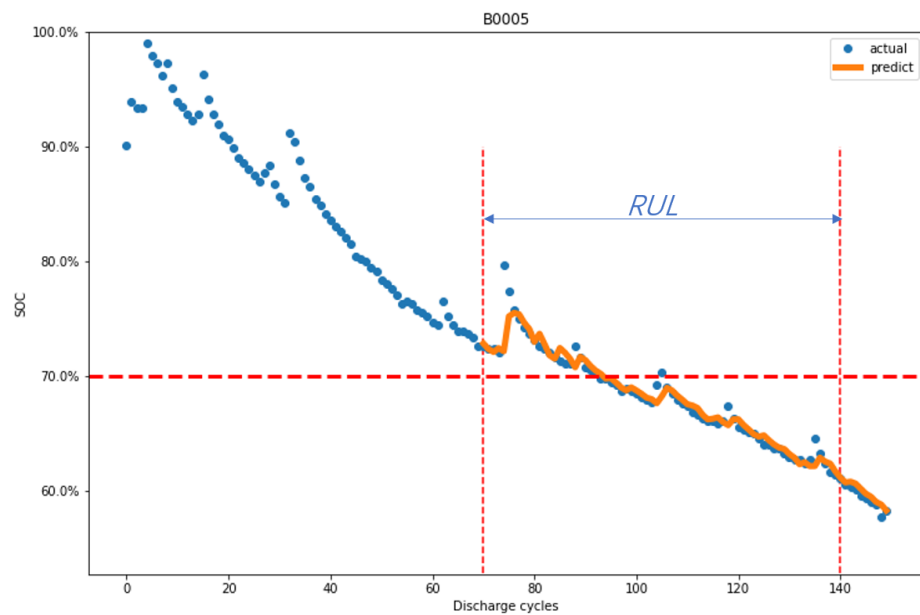


Figure 3. DTNN prediction performance for battery B0005.

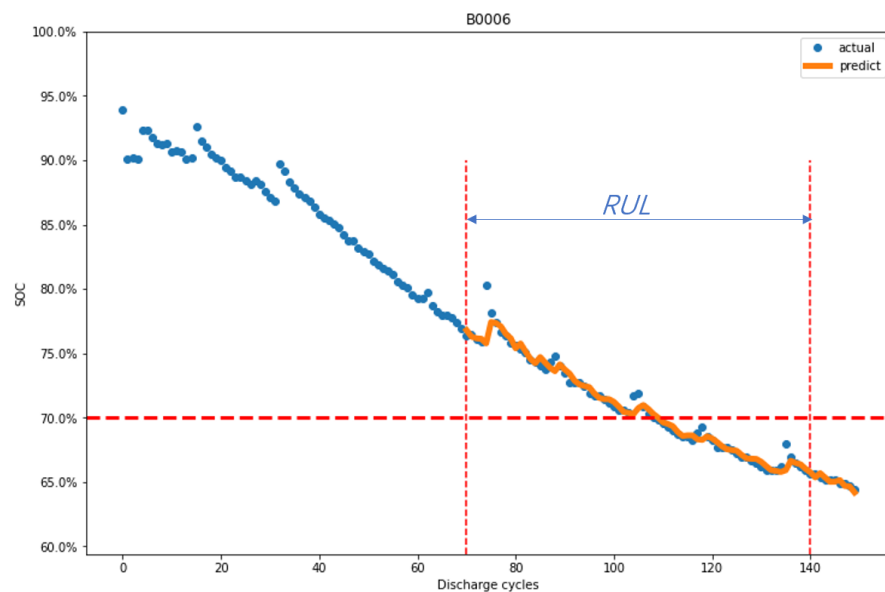


Figure 4. DTNN prediction performance for battery B0006.

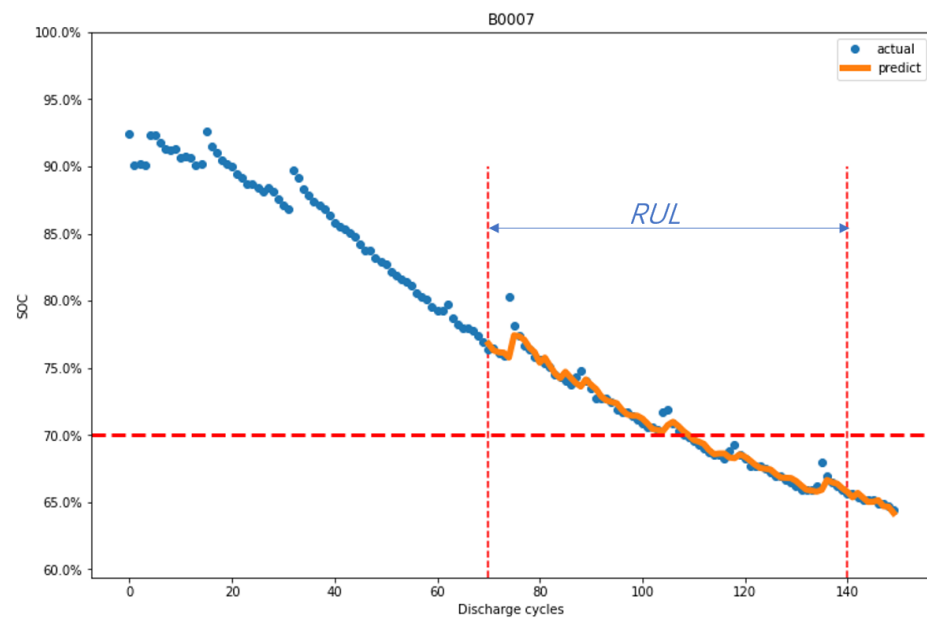


Figure 5. DTNN prediction performance for battery B0007.

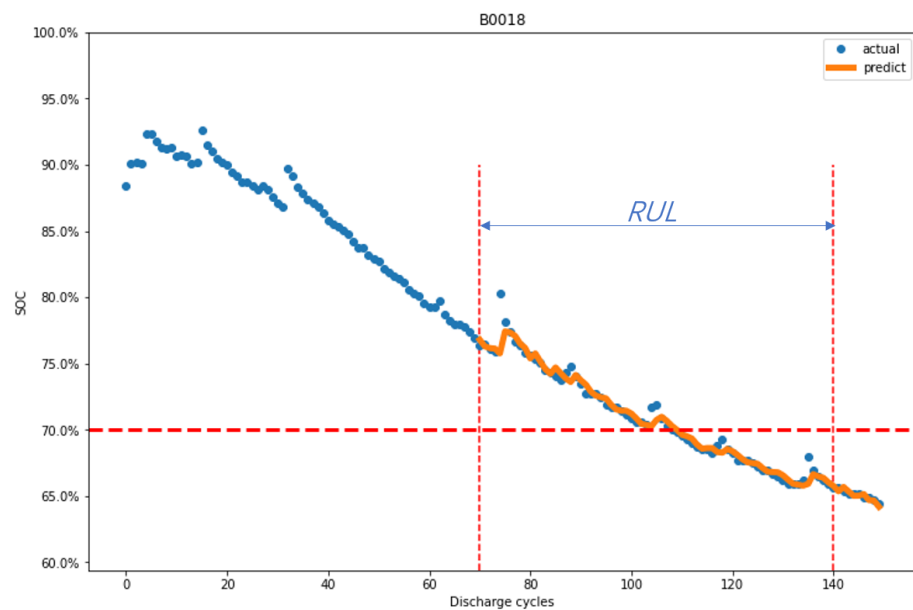


Figure 6. DTNN prediction performance for battery B0018.

The presence of these outliers can be attributed to the peak values in the prediction graphs. While our model aims to match the actual values closely, it also prioritises robustness to ensure compatibility with most batteries. As a result, the feature boundaries defined by the model are smoothed out, which might not capture sharp peaks or sudden changes in the actual data effectively.

From a battery's physical perspective, these peak values or sudden changes can be caused by various factors, including battery usage patterns, external environmental factors, or internal battery conditions. It is beneficial to delve deeper into the specific physical properties of batteries that lead to these peak values. By understanding these, we can refine our model to handle such scenarios better and reduce the prediction error.

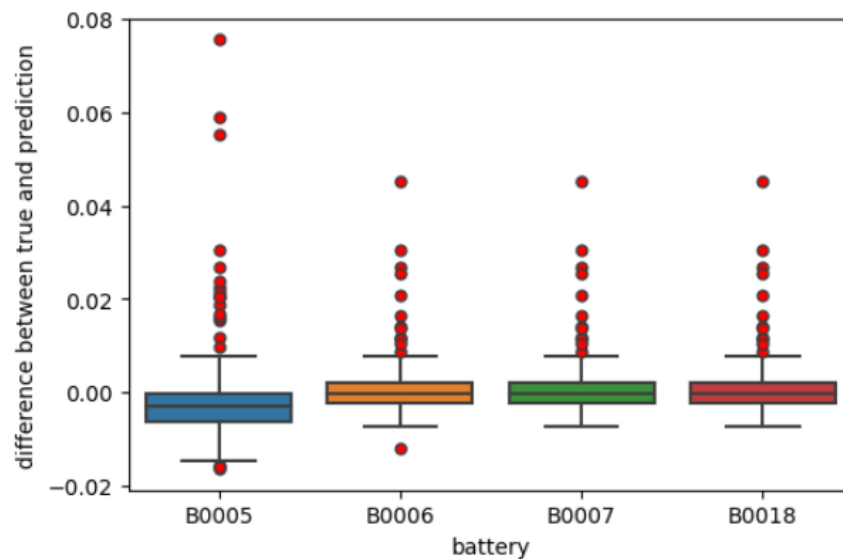


Figure 7. Boxplots of prediction errors.

In summary, our DTNN model, enhanced with a multi-head attention network, successfully learns features concurrently, making it exceptionally proficient at predicting the RUL of batteries with high accuracy. Including a denoising encoder further bolsters the performance of our model, making it an effective and highly efficient tool for accurate RUL prediction, mainly when applied to the NASA dataset.

5.2. Encoder Optimisation and Effects

In the preliminary phase of utilising NASA's dataset, images were transformed into current capacity data by integrating CNN and residual learning into the encoder. This enhancement facilitated significant improvements in image noise reduction and pre-processing, eliminating anomalous battery data noise in the primary sequence. Consequently, compared to baseline methods, our decoder processes highly accurate data.

Our DTNN employs residual learning to expedite training and augment denoising performance. The quickened training procedure boosts the efficiency of CNN, mitigating time consumption without compromising the algorithm's performance. Our approach can address Gaussian denoising even at unidentified noise levels, unlike most extant models, which are predominantly trained to handle specific Gaussian white noise models at known noise levels.

5.3. Model Comparison Using the Diebold-Mariano Test

To rigorously compare the predictive accuracy of our proposed DTNN method with other comparative methods, we employ the Diebold-Mariano (DM) test. The DM test statistic is given by:

$$DM = \frac{\bar{d}}{\sqrt{\frac{1}{T} \sum_{t=1}^T (d_t - \bar{d})^2}} \quad (15)$$

where \bar{d} represents the mean of the forecast error differences d_t , and T denotes the number of forecasts. Under the null hypothesis, which posits that both models possess equivalent predictive accuracy, the DM statistic follows an asymptotic normal distribution.

For our comparative analysis, the computed DM statistic and p -value is:

In the given DM test results, we compared the benchmark model DTNN with several other prediction methods. The results show that DTNN outperforms all other methods in all cases, as shown in Table 3. All p -values are less than 0.05, proving the superiority

of DTNN. Among them, the difference between DTNN and DT is the largest, and the difference with DeTransformer is the smallest.

Table 3. DM test results for comparing various methods with DTNN.

Method	<i>p</i> -Value	DM Value
RF	0.019	−4.17
DT	0.015	−5.26
MLP	0.018	−4.35
RNNs	0.024	−4.12
LSTM	0.026	−3.24
GRU	0.031	−4.01
Dual-LSTM	0.024	−2.95
DeTransformer	0.04	−1.98

6. Conclusions

In conclusion, the experimental results reveal the significant potential of our proposed method, the DTNN model, in predicting the RUL of lithium-ion batteries, mainly when applied to the NASA dataset. By leveraging a denoising encoder for feature extraction and noise reduction, our method improves upon existing models in handling noisy and complex battery life cycle data.

One key strength of our model is its ability to extract and utilise temporal information from the capacity sequences effectively. This capability was demonstrated in how our model consistently outperformed other models, including MLP and RNN-based models, in predicting battery RUL. Our findings confirmed the necessity of incorporating sequential information for robust and accurate RUL prediction.

While the DT and RF provided reasonable accuracy, it was outperformed by deeper architectures, especially our proposed DTNN. This underscores the limitations of shallower networks like DT and RF in modelling complex relationships and the advantages of employing more sophisticated deep learning models for tasks like RUL prediction of lithium-ion batteries.

A distinct feature of our analysis was setting a 70% capacity threshold to define the standard battery usage time. By focusing on the period before the battery capacity falls below this level, we were able to hone our predictions and focus on the most critical part of the battery's life cycle.

However, despite the promising results, our model has certain limitations. The DTNN, while effective, may require more computational resources compared to simpler models. Additionally, while our model showed excellent performance on the NASA dataset, its performance on other datasets or real-world scenarios remains to be validated. Future work should also explore integrating other features or external factors that affect battery degradation, such as environmental conditions or usage patterns.

Furthermore, it is essential to continue refining the model and testing its performance across different types of batteries, use cases, and operating conditions. This will help ensure its adaptability and scalability while addressing ongoing data quality and generalisability challenges. With further research and refinement, the DTNN model has the potential to significantly improve battery management practices by providing reliable and timely predictions of battery RUL, thereby contributing to the safety, efficiency, and sustainability of battery-powered systems.

Author Contributions: Conceptualisation, Y.H.; methodology, Y.H.; software, C.L.; validation, Y.H.; formal analysis, Y.H.; investigation, Y.H., G.L. and L.L.; resources, Y.H., G.L. and L.L.; data curation, Y.H. and L.Z.; writing—original draft preparation, Y.H.; writing—review and editing, Y.H., G.L. and L.L.; visualisation, Y.H. and C.L.; supervision, G.L., L.Z. and L.L.; project administration, G.L. and L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Armand, M.; Tarascon, J. Building Better Batteries. *Nature* **2008**, *451*, 652–657. [[CrossRef](#)] [[PubMed](#)]
2. Severson, K.; Attia, P.; Jin, N.; Perkins, N.; Jiang, B.; Yang, Z.; Chen, M.; Aykol, M.; Herring, P.; Fraggedakis, D.; et al. Data-Driven Prediction of Battery Cycle Life Before Capacity Degradation. *Nat. Energy* **2019**, *4*, 383–391. [[CrossRef](#)]
3. Singh, B.; Dubey, P.K. Distributed power generation planning for distribution networks using electric vehicles: Systematic attention to challenges and opportunities. *J. Energy Storage* **2022**, *48*, 104030. [[CrossRef](#)]
4. Tran, M.K.; Panchal, S.; Khang, T.D.; Panchal, K.; Fraser, R.; Fowler, M. Concept Review of a Cloud-Based Smart Battery Management System for Lithium-Ion Batteries: Feasibility, Logistics, and Functionality. *Batteries* **2022**, *8*, 19. [[CrossRef](#)] [[PubMed](#)]
5. Zheng, Z.; Peng, J.; Deng, K.; Gao, K.; Li, H.; Chen, B.; Yang, Y.; Huang, Z. A Novel Method for Lithium-Ion Battery Remaining Useful Life Prediction Using Time Window and Gradient Boosting Decision Trees. In Proceedings of the 2019 10th International Conference on Power Electronics and ECCE Asia (ICPE 2019-ECCE Asia), Busan, Republic of Korea, 27–30 May 2019; pp. 3297–3302.
6. Chen, Z.; Sun, M.; Shu, X.; Shen, J.; Xiao, R. On-board state of health estimation for lithium-ion batteries based on random forest. In Proceedings of the 2018 IEEE International Conference on Industrial Technology (ICIT), Lyon, France, 20–22 February 2018; pp. 1754–1759.
7. Sun, F.; Liu, J.; Wu, J.; Pei, C.; Lin, X.; Ou, W.; Jiang, P. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019; pp. 1441–1450.
8. Ma, M.; Mao, Z. Deep-Convolution-Based LSTM Network for Remaining Useful Life Prediction. *IEEE Trans. Ind. Inform.* **2021**, *17*, 1658–1667. [[CrossRef](#)]
9. Khalid, A.; Sundararajan, A.; Acharya, I.; Sarwat, A.I. Prediction of Li-Ion Battery State of Charge Using Multilayer Perceptron and Long Short-Term Memory Models. In Proceedings of the 2019 IEEE Transportation Electrification Conference and Expo (ITEC), Detroit, MI, USA, 19–21 June 2019; pp. 1–6.
10. Shi, Z.; Chehade, A. A dual-LSTM framework combining change point detection and remaining useful life prediction. *Reliab. Eng. Syst. Saf.* **2021**, *205*, 107257. [[CrossRef](#)]
11. Chen, D.; Hong, W.; Zhou, X. Transformer Network for Remaining Useful Life Prediction of Lithium-Ion Batteries. *IEEE Access* **2022**, *10*, 19621–19628. [[CrossRef](#)]
12. Hu, W.; Zhao, S. Remaining useful life prediction of lithium-ion batteries based on wavelet denoising and transformer neural network. *Front. Energy Res.* **2022**, *10*, 969168. [[CrossRef](#)]
13. Barre, A.; Deguilhem, B.; Grolleau, S.; Gerard, M.; Suard, F.; Riu, D. A Review on Lithium-Ion Battery Ageing Mechanisms and Estimations for Automotive Applications. *J. Power Sources* **2013**, *241*, 680–689. [[CrossRef](#)]
14. Xing, Y.; Ma, W.; Tsui, K.L.; Pecht, M. An ensemble model for predicting the remaining useful performance of lithium-ion batteries. *Microelectron. Reliab.* **2011**, *51*, 1084–1091. [[CrossRef](#)]
15. Hlal, M.I.; Ramachandaramurthy, V.K.; Sarhan, A.; Pouryekta, A.; Subramaniam, U. Optimum battery depth of discharge for off-grid solar PV/battery system. *J. Energy Storage* **2019**, *26*, 100999. [[CrossRef](#)]
16. Wang, J.; Liu, P.; Hicks-Garner, J.; Sherman, E.; Soukiazian, S.; Verbrugge, M.; Tataria, H.; Musser, J.; Finamore, P. Cycle-life model for graphite-LiFePO₄ cells. *J. Power Sources* **2011**, *196*, 3942–3948. [[CrossRef](#)]
17. Kwon, S.; Han, D.; Park, J.; Lee, P.Y.; Kim, J. Joint state-of-health and remaining-useful-life prediction based on multi-level long short-term memory model prognostic framework considering cell voltage inconsistency reflected health indicators. *J. Energy Storage* **2022**, *55*, 105731. [[CrossRef](#)]
18. Hesse, H.C.; Schimpe, M.; Kucevic, D.; Jossen, A. Lithium-Ion Battery Storage for the Grid—A Review of Stationary Battery Storage System Design Tailored for Applications in Modern Power Grids. *Energies* **2017**, *10*, 2107. [[CrossRef](#)]
19. Karoń, G. Energy in Smart Urban Transportation with Systemic Use of Electric Vehicles. *Energies* **2022**, *15*, 5751. [[CrossRef](#)]
20. Cakiroglu, C.; Islam, K.; Bekdaş, G.; Nehdi, M.L. Data-driven ensemble learning approach for optimal design of cantilever soldier pile retaining walls. *Structures* **2023**, *51*, 1268–1280. [[CrossRef](#)]

21. Zhang, S.Y.; Chen, S.Z.; Jiang, X.; Han, W.S. Data-driven prediction of FRP strengthened reinforced concrete beam capacity based on interpretable ensemble learning algorithms. *Structures* **2022**, *43*, 860–877. [[CrossRef](#)]
22. Guo, W.; Sun, Z.; Vilsen, S.B.; Meng, J.; Stroe, D.I. Review of “grey box” lifetime modeling for lithium-ion battery: Combining physics and data-driven methods. *J. Energy Storage* **2022**, *56 Pt A*, 105992. [[CrossRef](#)]
23. Xu, L.; Deng, Z.; Xie, Y.; Lin, X.; Hu, X. A novel hybrid physics-based and data-driven approach for degradation trajectory prediction in Li-ion batteries. *IEEE Trans. Transp. Electrification* **2022**, *9*, 2628–2644. [[CrossRef](#)]
24. Chen, L.; Tong, Y.; Dong, Z. Li-ion battery performance degradation modeling for the optimal design and energy management of electrified propulsion systems. *Energies* **2020**, *13*, 1629. [[CrossRef](#)]
25. Li, Y.; Zhou, Z.; Wu, W.T. Three-dimensional thermal modeling of Li-ion battery cell and 50 V Li-ion battery pack cooled by mini-channel cold plate. *Appl. Therm. Eng.* **2019**, *147*, 829–840. [[CrossRef](#)]
26. Tran, M.K.; DaCosta, A.; Mevawalla, A.; Panchal, S.; Fowler, M. Comparative Study of Equivalent Circuit Models Performance in Four Common Lithium-Ion Batteries: LFP, NMC, LMO, NCA. *Batteries* **2021**, *7*, 51. [[CrossRef](#)]
27. Varini, M.; Campana, P.E.; Lindbergh, G. A semi-empirical, electrochemistry-based model for Li-ion battery performance prediction over lifetime. *J. Energy Storage* **2019**, *25*, 100819. [[CrossRef](#)]
28. Saha, B.; Goebel, K.; Poll, S.; Christophersen, J. Prognostics Methods for Battery Health Monitoring Using a Bayesian Framework. *IEEE Trans. Instrum. Meas.* **2007**, *58*, 291–296. [[CrossRef](#)]
29. Burgos-Mellado, C.; Orchard, M.E.; Kazerani, M.; Cárdenas, R.; Sáez, D. Particle-filtering-based estimation of maximum available power state in Lithium-Ion batteries. *Appl. Energy* **2016**, *161*, 349–363. [[CrossRef](#)]
30. Berecibar, M.; Gandiaga, I.; Villarreal, I.; Omar, N.; Van Mierlo, J.; Van den Bossche, P. Critical Review of State of Health Estimation Methods of Li-Ion Batteries for Real Applications. *Renew. Sustain. Energy Rev.* **2016**, *56*, 572–587. [[CrossRef](#)]
31. Nhu, V.H.; Shahabi, H.; Nohani, E.; Shirzadi, A.; Al-Ansari, N.; Bahrami, S.; Miraki, S.; Geertsema, M.; Nguyen, H. Daily Water Level Prediction of Zrebar Lake (Iran): A Comparison between M5P, Random Forest, Random Tree and Reduced Error Pruning Trees Algorithms. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 479. [[CrossRef](#)]
32. Cha, G.W.; Moon, H.; Kim, Y.m.; Hong, W.; Hwang, J.H.; Park, W.; Kim, Y.C. Development of a Prediction Model for Demolition Waste Generation Using a Random Forest Algorithm Based on Small DataSets. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6997. [[CrossRef](#)]
33. Nascimento, R.G.; Corbetta, M.; Kulkarni, C.S.; Viana, F.A. Hybrid physics-informed neural networks for lithium-ion battery modeling and prognosis. *J. Power Sources* **2021**, *513*, 230526. [[CrossRef](#)]
34. Ardeshiri, R.R.; Razavi-Far, R.; Li, T.; Wang, X.; Ma, C.; Liu, M. Gated recurrent unit least-squares generative adversarial network for battery cycle life prediction. *Measurement* **2022**, *196*, 111046. [[CrossRef](#)]
35. Yang, Y.; Wu, Z.; Yang, Y.; Lian, S.; Guo, F.; Wang, Z. A Survey of Information Extraction Based on Deep Learning. *Appl. Sci.* **2022**, *12*, 9691. [[CrossRef](#)]
36. Yin, A.; Tan, Z.; Tan, J. Life Prediction of Battery Using a Neural Gaussian Process with Early Discharge Characteristics. *Sensors* **2021**, *21*, 1087. [[CrossRef](#)] [[PubMed](#)]
37. Zhao, J.; Ling, H.; Liu, J.; Wang, J.; Burke, A.F.; Lian, Y. Machine learning for predicting battery capacity for electric vehicles. *eTransportation* **2023**, *15*, 100214. [[CrossRef](#)]
38. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.u.; Polosukhin, I. Attention is All you Need. In *Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
39. Wu, L.; Li, S.; Hsieh, C.J.; Sharpnack, J. SSE-PT: Sequential Recommendation Via Personalized Transformer. In Proceedings of the 14th ACM Conference on Recommender Systems; Association for Computing Machinery, New York, NY, USA, 18–22 September 2020; pp. 328–337.
40. Cornia, M.; Stefanini, M.; Baraldi, L.; Cucchiara, R. Meshed-Memory Transformer for Image Captioning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
41. Gu, F. Research on Residual Learning of Deep CNN for Image Denoising. In Proceedings of the 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi’an, China, 15–17 April 2022; pp. 1458–1461.
42. Alkesaiberi, A.; Harrou, F.; Sun, Y. Efficient Wind Power Prediction Using Machine Learning Methods: A Comparative Study. *Energies* **2022**, *15*, 2327. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.