MDPI

Article

Attention Mechanism-Based Neural Network for Prediction of Battery Cycle Life in the Presence of Missing Data

Yixing Wang [ID] and Benben Jiang *

Department of Automation, Tsinghua University, Beijing 100084, China; yx-wang21@mails.tsinghua.edu.cn
* Correspondence: bbjiang@tsinghua.edu.cn

Abstract: As batteries become widespread applications across various domains, the prediction of battery cycle life has attracted increasing attention. However, the intricate internal mechanisms of batteries pose challenges to achieving accurate battery lifetime prediction, and the inherent patterns within temporal data from battery experiments are often elusive. Meanwhile, the commonality of missing data in real-world battery usage further complicates accurate lifetime prediction. To address these issues, this article develops a self-attention-based neural network (NN) to precisely forecast battery cycle life, leveraging an attention mechanism that proficiently manages time-series data without the need for recurrent frameworks and adeptly handles the data-missing scenarios. Furthermore, a two-stage training approach is adopted, where certain network hyperparameters are fine-tuned in a sequential manner to enhance training efficacy. The results show that the proposed self-attention-based NN approach not only achieves superior predictive precision compared with the benchmarks including Elastic Net and CNN-LSTM but also maintains resilience against missing-data scenarios, ensuring reliable battery lifetime predictions. This work highlights the superior performance of utilizing attention mechanism for battery cycle life prognostics.

Keywords: battery cycle life prediction; self-attention mechanism; neural network; missing data

1. Introduction

Recent years have seen a significant focus on battery degradation and its impact on the operational lifespan of energy storage systems. As the demand for reliable and efficient energy storage solutions grows, accurate prediction of battery cycle life has become paramount. Batteries, being crucial for a myriad of applications from portable electronics [1,2] to electric vehicles [3–5], energy storages [6,7], and renewable energy systems [8–10], suffer from degradation influenced by a complex interplay of physical, chemical, and environmental factors. These complexities pose a substantial challenge for accurately modeling and predicting battery performance over time.

Researchers have been dedicated to extending battery lifespan and enhancing system performance by exploring various methodologies for predicting battery cycle life. Traditional methods typically rely on empirical models that may oversimplify and thus fail to contain the complex dynamics of battery behavior. More sophisticated approaches such as model-based approaches can capture more intricate details but often at the cost of computational intensity and the need for battery's physical and chemical property information due to the complex electrochemical nature of the modeling [11]. The development of battery models has been critical in this endeavor, with a pioneer work on the electrochemical model (ECM) by Newman and Tiedemann [12]. The ECM, evolving through time, provides a comprehensive description of battery properties due to its modelling capabilities [13]. This enables researchers to accurately estimate the state of health (SOH) and predict the remaining useful life (RUL) of batteries [14]. Additionally, the ECM has been enhanced with thermal models to account for phenomena such as solid electrolyte interface (SEI)

layer growth—a key indicator of battery degradation [15]—furthering the precision of lifetime prediction methods [16]. The pseudo-two-dimensional (P2D) model [17], a prominent ECM variant, applies porous electrode theory principles, focusing on the order reduction and simplification of kinetic equations [18], and also serves as a competitive model for SOH estimations [19]. This model has been utilized by researchers such as Ashwin et al. [20] to simulate battery aging processes, incorporating temperature effects and SEI layer thickness and thereby modeling battery capacity fade under cyclic loading conditions. Innovations have continued with variants such as the single-particle model (SPM) [21], the simplified P2D model [17], the improved SPM model [22], and the reduced-order model (ROM) [23–26]. For instance, Parhizi et al. [27] investigated the SEI layer development using the SPM model to understand capacity degradation under various operational conditions, aiming to predict capacity fade.

Concurrently, machine learning methods have surfaced as potent alternatives, harnessing data-driven insights to capture hidden patterns within battery performance data [28]. These techniques range from regression-based algorithms and ensemble methods to recurrent neural networks (RNNs), which are particularly adept at capturing temporal dependencies. Severson et al. [29] leveraged the elastic net (EN) regression to predict battery cycle life based on early cycle charging and discharging characteristics, using features derived from voltage, capacity, temperature, and internal resistance as machine learning inputs. Zhang et al. [30] also achieved battery lifetime predictions based on feature construction, where the physical features were namely ECM parameters. In addition, Hong et al.'s approach [31] to predict a battery's RUL using voltage curve data from only a few early cycles included the use of dilated convolutional neural networks (CNNs) to handle the time-sequential voltage data and implicit sampling to gauge the neural network's output uncertainty. Thelen et al. [32] also focused on RUL prediction while using various data-driven methods, including EN or random forest, to enhance the predictive performance of an empirical capacity fade model implemented by a Gaussian process (GP) model. Nuhic et al. [33] utilized the support vector machine (SVM) for the diagnosis and prognostics of battery SOH and RUL, taking into account the influence of environmental and load conditions. Considering the usage of fully discharged voltage and internal resistance, Tseng et al. [34] built a regression model for SOH estimation and RUL prediction that incorporated particle swarm optimization for optimal parameter determination. Mansouri et al. [35] and Guo et al. [36] expanded the suite of regression techniques, implementing algorithms like LASSO, multi-layer perceptron (MLP), support vector regression (SVR), and gradient boosted trees (GBT) for RUL prediction, with some even incorporating Bayesian methods for managing uncertainty. Neural networks have also been pivotal, with studies by Khumprom et al. [37] and Ren et al. [38] employing deep neural network (DNN) structures including autoencoders for feature extraction, while Zhang et al. [39] utilized RNNs to capture capacity degradation patterns. Meanwhile, the model-based and data-driven methods can also be combined together to achieve SOH and RUL prediction, where these types of methods are commonly addressed as grey box modeling [40]. For example, Thelen et al. [32] proposed a framework where various machine learning methods, including EN, random forest (RF), GP, and neural networks (NNs), were employed to correct and enhance the predictive performance of an empirical model in the context of RUL prediction. Liao et al. [41] raised another method fusion framework where a particle filter model was assisted by two data-driven methods to measure the internal state of the model and predict future measurements, eventually predicting the RUL of batteries.

Model-based approaches are recognized for their detailed simulation of battery behaviors, providing explicit models of internal processes that accurately represent battery dynamics. On the other hand, data-driven methods have garnered acclaim for their data processing capabilities, yielding impressive predictive accuracy even in the absence of intricate domain-specific knowledge. Nevertheless, the sequential nature of data hidden in battery experiments poses a significant challenge; the above-mentioned methodologies have yet to surmount the complexity and non-linearity embedded within the data

sequences. Notably, RNNs have been employed to handle temporal sequences in many researches [39], although this advantage often entails high computational demand or a compromise in prediction precision.

In the domain of natural language processing (NLP), a groundbreaking concept known as the self-attention mechanism has been introduced with noteworthy success. Initiated by A. Vaswani et al. [42] to address translational intricacies, the attention mechanism shows a great capability for dealing with serial data without recurrent structure, reducing the complexity of networks and calculation burden. The self-attention mechanism outperforms conventional RNNs with two merits: its capability for parallel computation of temporal sequences, markedly accelerating operational throughput, and its intrinsic proficiency in maintaining focus across long-time series, overcoming the fundamental limitation encountered in traditional RNN architectures.

Meanwhile, the issue of missing data is a prevalent concern in real-world sequential data, and its occurrence can introduce unpredictable disruptions to model training and evaluative processes. To tackle the challenge of data loss, various methods have been proposed in the literature. For example, Severson et al. [43] introduced a method employing expectation maximization and sparse discriminant analysis for high-dimensional datasets, training classifiers across varying patterns of missing data. Complementary to this, PCA-based techniques have also shown promise in addressing data insufficiencies [44]. Jeong et al. [45] tackled fairness in model training amid absent values, devising a fair mixed-integer programming (MIP) forest algorithm based on decision trees with the missing incorporated in attribute (MIA) technique [46], while Chen et al. [47] integrated a multi-head self-attention mechanism within a variational auto-encoder (VAE) network for the industrial soft-sensing problem with missing data, and the best method could satisfy the requirements of high precision in complex industrial processes. Additionally, Wu et al. [48] applied an attention-based AimNet structure for data imputation and data cleaning issues, and Yu et al. [49] proposed a novel network structure consisting of convolution modules and an attention module, along with a specially designed loss function, successfully improving the performance for the consecutively missing seismic data reconstruction problem.

In the context of battery life prediction, where the intricacies of temporal correlations and the varying importance of attributes are decisive, the ability of the attention mechanism to spotlight pertinent information holds substantial promise. However, there are still few studies focusing on the application of the self-attention mechanism on battery life prediction [50]. By effectively capturing the interplay between attributes and their temporal progression, the attention mechanism presents an opportunity to elevate prediction accuracy. Building on the progress made in attention-based models and the growing recognition of their potential in capturing temporal dependencies, this study introduces a novel application: the integration of the attention mechanism into neural networks for battery life prediction, where we combined the calculation of attention value and typical fully connected networks together to enhance battery life forecasting. To enhance the predictive performance of the attention network during training, a two-stage training approach was employed, where some hyperparameters of the network need to be tuned between the training stages. Moreover, this study delves into the attention mechanism's innate capacity to handle missing data, examining its efficacy under different data-missing possibilities across three distinct data-loss patterns: random missing, cycle missing, and time-step missing. The results demonstrate the attention mechanism's exceptional predictive performance under varied data-missing scenarios and various degree of data missingness.

The remaining sections of this paper are outlined as follows. Section 2 elaborates on the principles of the self-attention mechanism and its deployment in contexts of data-loss scenarios. Section 3 details the used dataset and the categorization of data missingness, and then Section 4 presents the predictive results for battery cycle life under different data conditions. Finally, Section 5 provides the conclusion of the work.

2. Methods

2.1. The Self-Attention Mechanism

Denote $X = [x_1^T, x_2^T, \ldots, x_L^T]^T$ as the input sequence, and $x_i^T = [x_{i1}, x_{i2}, \ldots, x_{im}]$, where L is the sequence length, and $m = d_{\text{model}}$ is the dimension of the input sequence. The calculation of attention mainly involves three aspects, i.e., query Q, key K, and value V, derived by $Q = XW^Q$, $K = XW^K$, $V = XW^V$, where W^Q, W^K, W^V are the weight matrices with the dimension of $d_{\text{model}} \times d_Q, d_{\text{model}} \times d_K$, and $d_{\text{model}} \times d_V$, respectively. As the requirement of the calculation, $d_Q = d_K$ should hold, and $d_K = d_V$ is a common setup in practical applications similar to our work.

Given the Q, K, V, the calculation of self-attention can be obtained by (1). As the $\left(\frac{QK^T}{\sqrt{d_K}}\right)$ part is a square matrix, the SoftMax operator is applied on each of its rows.

$$\text{Attention}(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d_K}}\right)V \tag{1}$$

Inspired by the idea of attention, multi-head self-attention is more popular and more capable in practical applications, where different heads can focus on different parts of the input sequence, covering important portions as much as possible. The calculation of multi-head self-attention mainly consists of a concatenation of attention heads and a linear transformation, as shown in (2):

$$\begin{aligned} \text{MultiHead}(Q, K, V) &= [\text{head}_1, \text{head}_2, \ldots, \text{head}_h]W^O \\ \text{where head}_i &= \text{Attention}\left(XW_i^Q, XW_i^K, XW_i^V\right) \end{aligned} \tag{2}$$

In addition, to emphasize location information, positional encoding is adopted before calculating the attention value. As a common choice in practice, we use sine and cosine functions for the positional encoding in the work, as shown in (3):

$$\begin{aligned} PE_{(pos,\ 2i)} &= \sin\left(pos/10,000^{2i/d_{\text{model}}}\right) \\ PE_{(pos,\ 2i+1)} &= \cos\left(pos/10,000^{2i/d_{\text{model}}}\right) \end{aligned} \tag{3}$$

where *pos* represents the position, and *i* indicates the dimension of input.

After the multi-head self-attention has been calculated, a fully connected network is employed to reduce dimension and obtain the target output. The overall structure of the entire attention-based network is described in Figure 1.
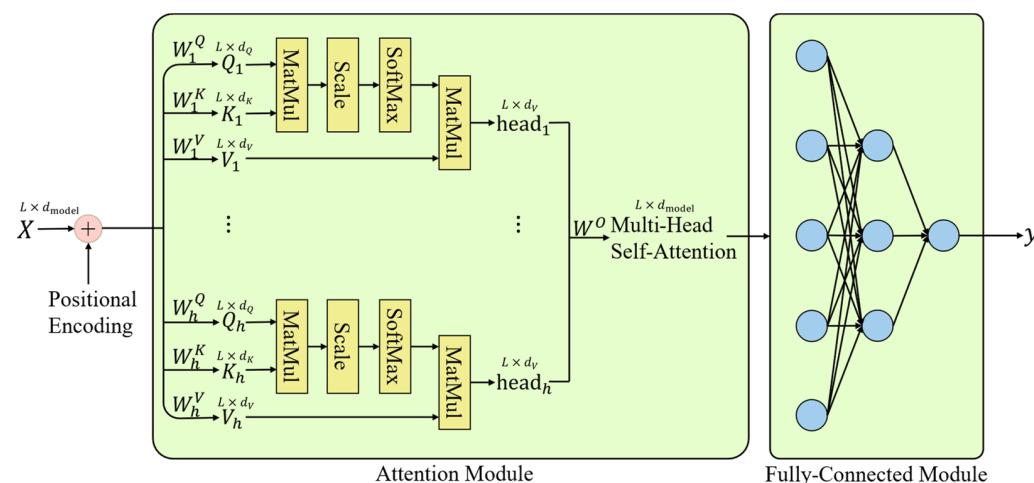

Figure 1. The structure of the attention-based network. The arrows indicate the data flow and the blue circles represent the general neurons in the network.

2.2. Handling Missing Data with the Self-Attention Mechanism

Within the domain of sequential data analysis, the self-attention mechanism has emerged as a cornerstone for capturing contextual relationships. Its innovative deployment in the face of missing data further underscores its versatility. In real-world datasets, where incomplete data are commonplace, the self-attention mechanism showcases its inherent adaptability to these challenges. Under the data-missing condition, the affected positions in the data input X will be set to zero, yet their positional encoding preserves the location's informational context. Particularly in the context of self-attention, where queries, keys, and values are derived from the same input sequence, the mechanism's intrinsic similarity computation and attention weight allocation process exhibit a degree of adaptability that can accommodate partial data availability.

In the case of missing data within self-attention, the absence of values at specific positions—manifested as random zeros within the input X—directly affects their influence on the similarity calculations between queries and keys. Denote the data input with missing values as X'; then, the query, key, and value can be articulated as $Q' = (X' + PE)W^Q$, $K' = (X' + PE)W^K$, and $V' = (X' + PE)W^V$, respectively. While missing values do not disrupt the attention values' computation, the resultant attention scores inherently reflect the data's absence and maintain positional context, as depicted in Figure 2. Consequently, the mechanism intuitively assigns greater attention weights to the present data, ensuring that the calculated scores naturally diminish the impact of any missing elements, thereby implicitly moderating the undue emphasis on missing values. This intrinsic modulation of attention allows the model to concentrate on the segments of data that are available, resonating with the fundamental principle of the self-attention mechanism to prioritize segments enriched with information.
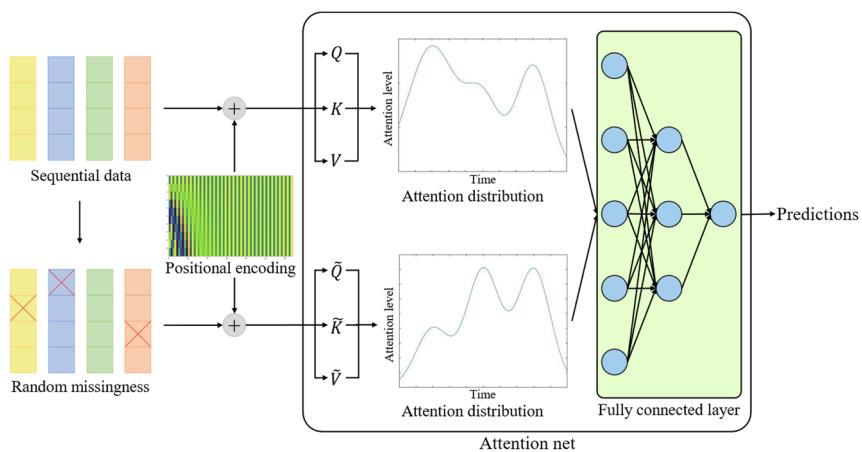

Figure 2. The self-attention mechanism for handling data missingness. When sequences with missing data (the pattern with red crosses) are fed into the network along with positional encodings, the attention-based network retains its original computational structure but adapts its self-attention values computation in response to the missing data, consequently altering attention distribution across different positions, thereby autonomously handling missing data scenarios.

Leveraging the attention mechanism to handle time-series data suffering from random losses presents several advantages. Its inherent adaptability to missing data scenarios is particularly notable, as it dynamically recalibrates significance to the intact portions of data, adeptly bridging the voids left by absent values. This flexibility ensures the model's predictive performance remains robust, even when confronted with incomplete sequences. Additionally, the attention mechanism's ability to selectively emphasize informative segments while minimizing the impact of missing or noisy data facilitates the extraction of pivotal patterns and dependencies, ultimately enhancing prediction accuracy.

2.3. Training Process

Overfitting poses a significant challenge in network training. To avoid the overfitting issue, our approach incorporates a two-stage training methodology, as illustrated in Figure 3. This strategy involves distinct settings for the regularization coefficient and batch size in each stage. Initially, the training commences with a smaller regularization coefficient and a reduced batch size for a duration of 50 epochs. The network parameters are then preserved at the point where the training error reaches its minimum during this initial phase. Before the second stage, the division of the dataset into training and validation sets is re-randomized. The subsequent stage of training proceeds from the previously saved network parameters, now employing an increased regularization coefficient and a larger batch size. This two-stage technique for training enables the network to learn with enhanced precision while effectively preventing overfitting.
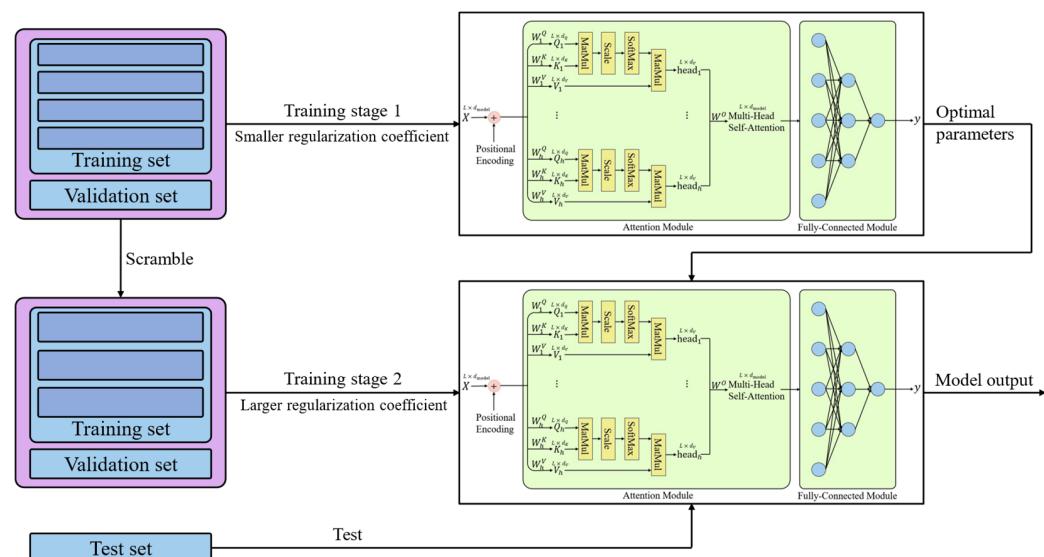
Figure 3. Flowchart of the two-stage training process.

3. Description of Dataset and Data-Missing Patterns

3.1. Dataset and Input Data Construction

This work utilizes the dataset referenced in [29], which comprises 124 commercial high-power LFP/graphite battery cells, each with a nominal capacity of 1.1 Ah and a nominal voltage of 3.3 V. These cells underwent cycling tests under varying constant current–constant voltage (CC-CV) charging protocols with several different charging rates while maintaining a constant environment temperature of 30 °C within experimental chambers. The CC stage of the charging process featured C rates ranging from 2 C to 6 C across different SOC intervals. Upon reaching 80% SOC, a CV stage ensued until the completion of the charging process. For discharging, all cells underwent another CC-CV discharge process at a rate of 4 C down to 2.0 V, with a current cutoff of C/50. Under such cycling tests, the lifespan of all batteries ranged from 148 to 2233 cycles.

Standard practice in battery cycling experiments involves the recording of several parameters in each cycle, including time, current, voltage, charging capacity, discharging capacity, and temperature, as depicted in Figure 4a–e. Recognizing the significance of the interplay among capacity, temperature, and voltage in battery analysis, we further computed the scaled discharging capacity and discharging temperature through fitting and interpolation against the voltage curve, as shown in Figure 4f,g. In addition, to obtain deeper insights from the cycling data, we derived the gradient of discharging capacity in relation to voltage (Figure 4h), i.e., the cycle-wise incremental capacity of the cells. In total, nine attributes of battery cycling data are employed in our work.
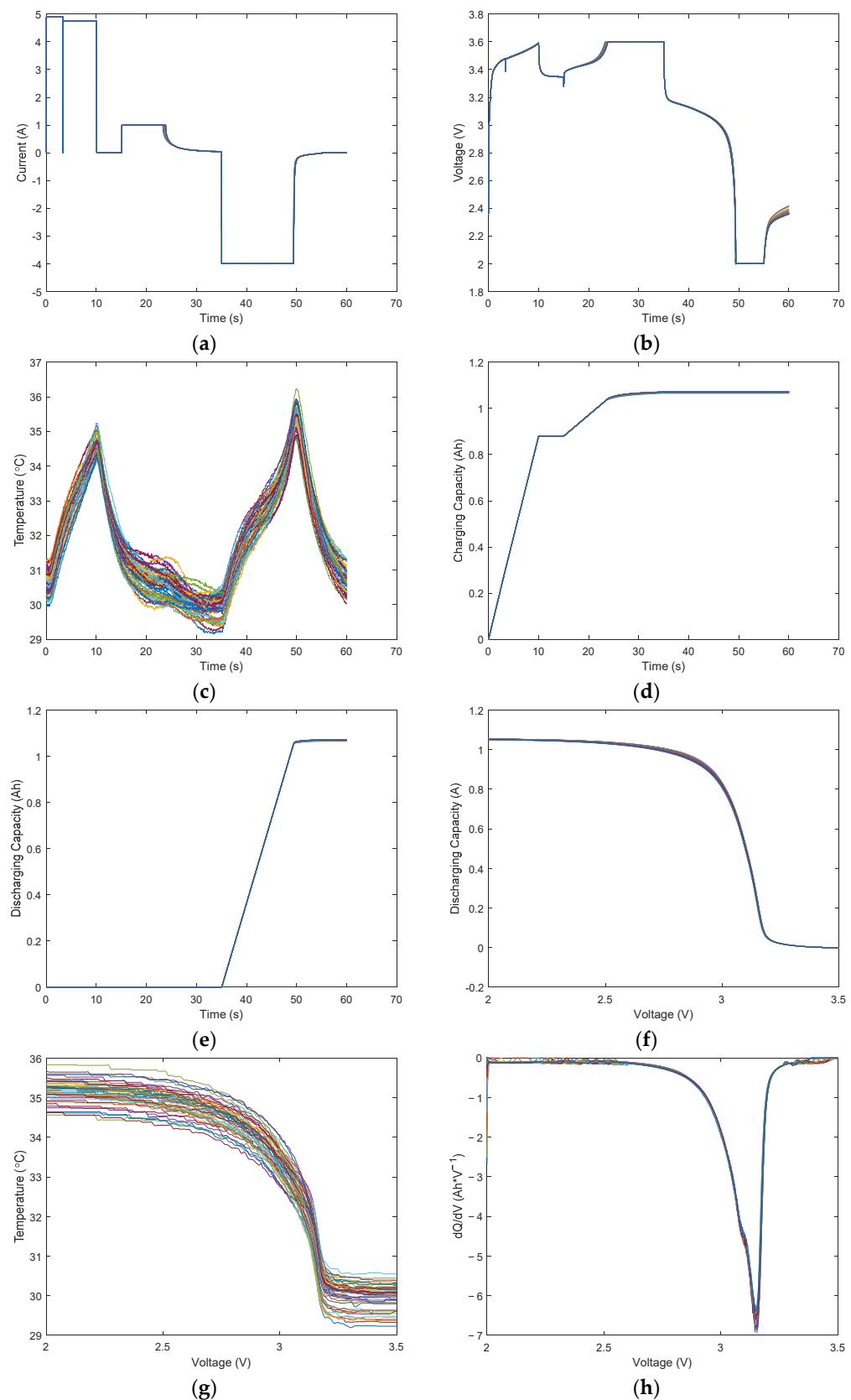
Figure 4. The contents of input data. (**a**) Current. (**b**) Voltage. (**c**) Temperature. (**d**) Charging capacity. (**e**) Discharging Capacity. (**f**) Discharging capacity variation against voltage. (**g**) Temperature variation against voltage. (**h**) Incremental capacity. The different line colors indicate different cycles of the displayed battery cell.

To optimize the utilization of sequential data from the cycling experiments, we adopted a unique structure for input data, diverging from conventional intuitive input formats, as illustrated in Figure 5. Our input construction (as depicted in Figure 5a) involves vertically concatenating the nine attributes to form an extended array for each cycle, while the horizontal dimension represents the different cycles. The principal benefit of this structured approach is the strategic reconstruction of the matrix's dimensions, coupled with a reduction in its length. This effectively mitigates the computational demands, particularly in the subsequent phases of attention calculation, ensuring a more efficient processing workflow.
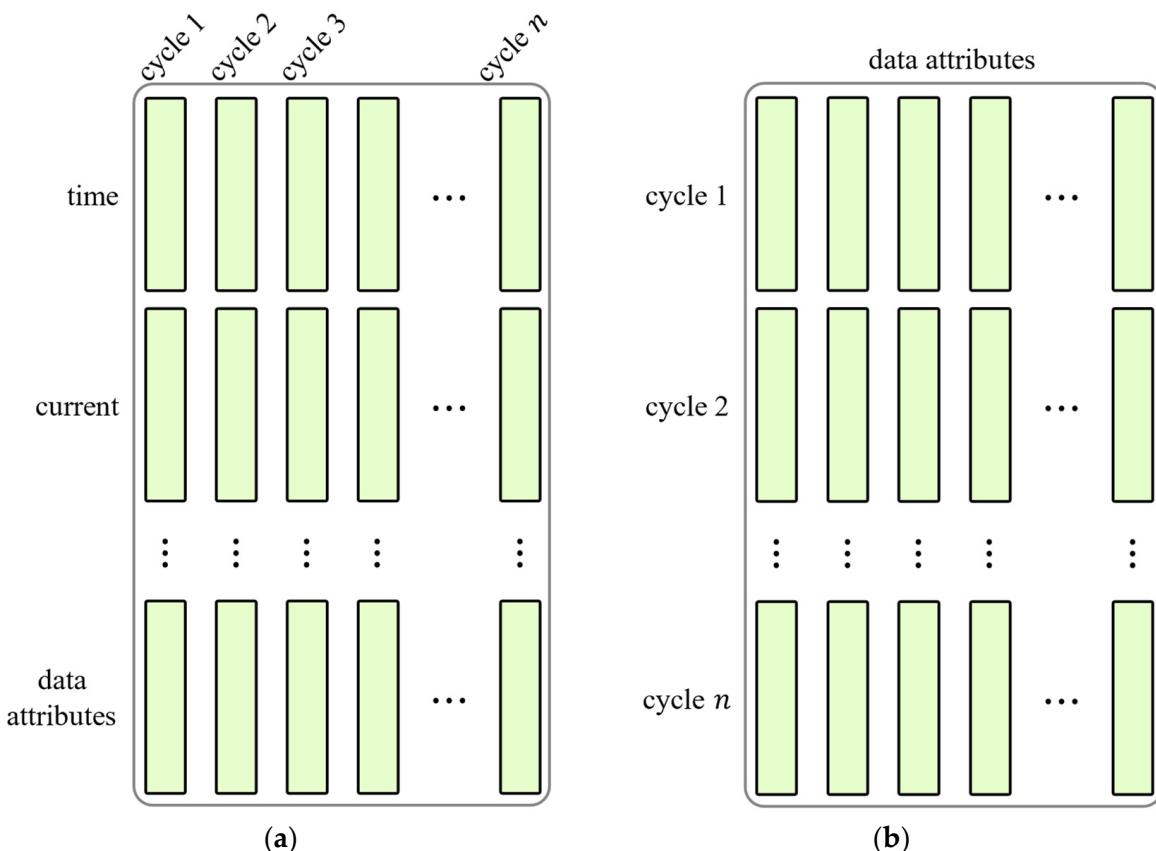
Figure 5. The structure of input data. (**a**) The input structure used in this work. (**b**) The intuitive input structure.

3.2. Addition of Missing Data

This work primarily investigates scenarios involving data loss [43]. We consider three different patterns of missing data: random missing, cycle missing, and time-step missing, which is illustrated in Figure 6. Random missing drops data points in a completely random manner, while cycle missing and time-step missing drop data points by entire columns or rows, reflecting real-world instances of data loss in specific cycles or at particular time steps.

In addition, it is important to note that while the attention mechanism inherently possesses the ability to adjust to missing data, this capability is inversely related to the extent of the data gaps. To maintain the balance during training and ensure the effectiveness of the mechanism, we set the dropout probability at a default of 0.01, with an upper limit of 0.1 under any circumstances. This approach enables us to systematically evaluate the robustness of the attention mechanism across different levels of data missing while preventing excessive data loss that could compromise the model's performance.
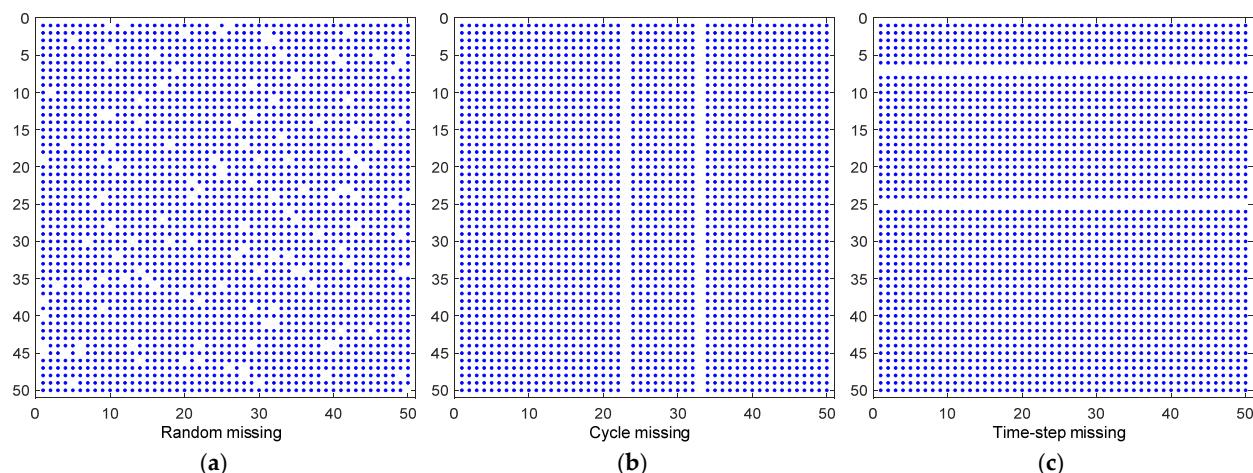
Figure 6. Examples of the various types of data missing [43]: (**a**) random missing, (**b**) cycle missing, and (**c**) time-step missing.

4. Results

To measure the predictive performance of the attention-based network approach, the metrics of average percentage error (APE) and root mean square error (RMSE) are used in this work, which is defined as follows:

$$\text{APE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \tag{4}$$

and

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

where y_i is the observed cycle life, and $\hat{y}_i$ is the predicted cycle life.

The prediction results of the proposed method, in scenarios without data missing, are displayed in Figure 7, wherein the conducted battery lifetime prediction using the initial 20, 50, and 100 cycles, respectively, is shown. As can be seen from the results in Figure 7, the cycle life predictions in all three conditions are relatively accurate (i.e., close to the ideal prediction line, represented by the black solid line in the figures), with the test APE of 13.02%, 10.59%, and 15.16%, respectively. In addition, compared to the results presented in [29], it is evident that our approach effectively addresses the challenge of outliers, showcasing enhanced predictive accuracy and reliability.
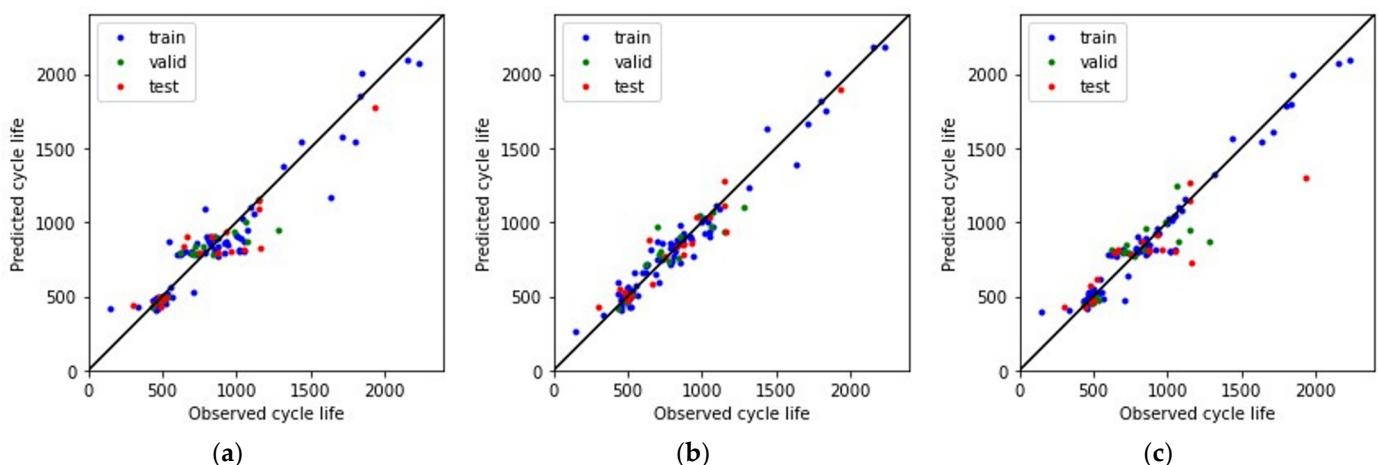
Figure 7. The prediction result of the proposed attention-based network approach by using different cycles of battery data without data missing: (**a**) 20 cycles, (**b**) 50 cycles, and (**c**) 100 cycles.

Additionally, we conducted an analysis to examine the impact of altering the number of cycles utilized for lifetime prediction, with the findings showcased in Figure 8. As demonstrated in Figure 8, utilizing the first 50 cycles for prediction obtains the optimal performance, where the average test APE of ten different dataset splits reaches 9.34%. In contrast, employing either fewer or more cycles results in increased APEs. The reason for this phenomenon is that a lower number of cycles may lead to underutilization of the network's capacity to process available information, and the lack of information may also deteriorate the performance. Conversely, an excessively high cycle number can introduce redundant data, potentially impeding the network's training efficiency and overall predictive performance. This balance highlights the importance of selecting appropriate information to achieve optimal prediction accuracy when solving battery lifetime prediction problems.
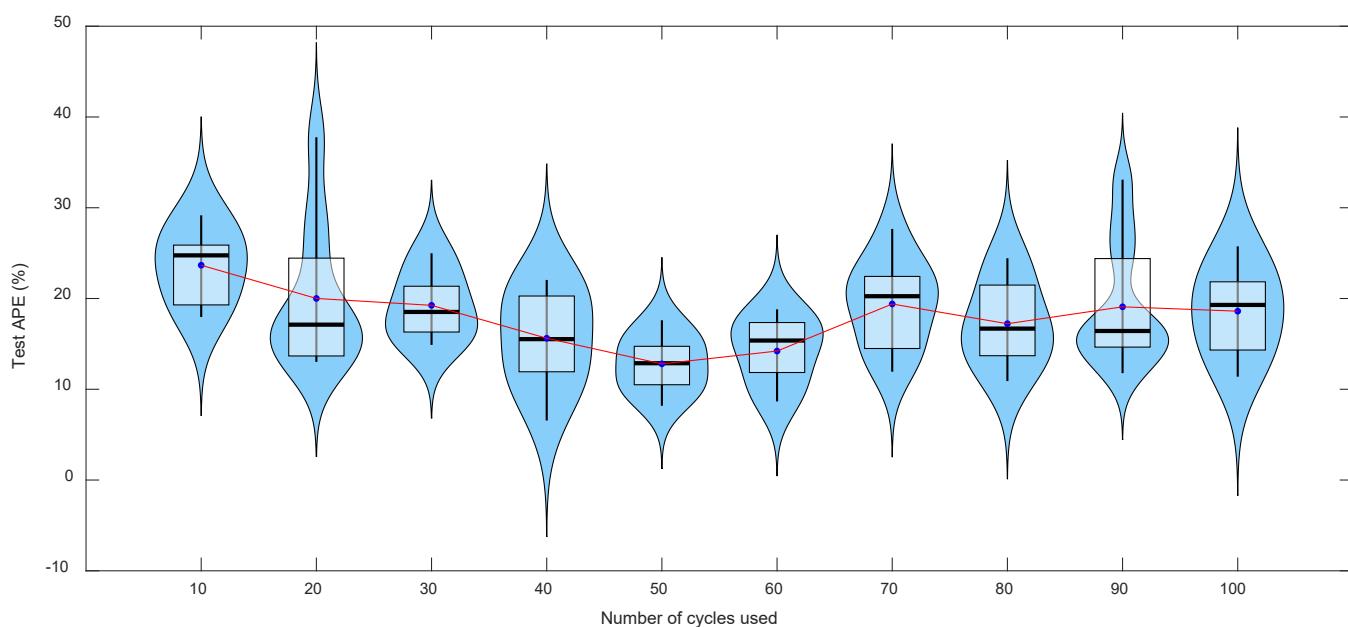
Figure 8. The violin plot for the influence of using different number of cycles. Under each number, 10 different splits of train/valid/test datasets were conducted. In the figure, the bold lines represent median APE values, the boxes represent quartiles of the APE values, and blue dots represent mean APE values with the red line connecting them, and the blue areas represent the probability distribution of the APE values.

Therefore, we opted to use 50 cycles of battery data to examine how the attention network performs under different scenarios of missing data, including random missing, cycle missing, and time-step missing. The results of this analysis are detailed in Table 1. Notably, the attention network exhibited exceptional stability in scenarios characterized by cycle missing. In such cases where data points were consecutively absent in a cyclic pattern, the attention mechanism demonstrated its resilience, effectively preserving its capability to discern relationships within the data sequence, notwithstanding these gaps. The result highlights the mechanism's ability to adapt to predictable patterns of data gaps.

Table 1. Model metrics for different data-missing types.

	APE (%)			RMSE (Cycles)		
	Train	**Validation**	**Test**	**Train**	**Validation**	**Test**
Random missing	9.4	10.8	11.3	82	119	104
Cycle missing	8.8	9.3	10.3	77	102	100
Time-step missing	9.8	9.8	11.1	80	108	99

On the other hand, in scenarios characterized by random missing data and time-step missing data, the attention network's performance closely approached its full potential. Despite the unpredictability and challenges associated with the two types of data absence, the network effectively leveraged the available information, maintaining a considerable level of predictive accuracy.

These findings shed light on the adaptability and robustness of the attention network across various scenarios of data absence. Its consistent performance in cycle-missing scenarios and substantial resilience in situations of random missing data and time-step missing data underscore its capability to handle patterns of missing data effectively. The overall capability of the attention mechanism in lifetime prediction demonstrates its versatility and potential in diverse data conditions.

To benchmark the performance of our method under conditions with data missing, we also conducted validations against several other methods from existing literature, specifically focusing on the cycle-missing scenario. The results of these comparative tests are presented in Table 2. As the method described in [29] solely utilizes the 2nd and 100th cycles of battery data for feature construction, the replication of its result did not incorporate scenarios of data missing. Instead, the basic performance of this method was used as a benchmark for comparison. The findings indicate that, under identical conditions of missing data, our proposed attention-based method demonstrates superior adaptability and enhanced predictive performance compared to the alternatives. This further validates the efficacy of the proposed attention-based network approach in predicting battery lifetime, particularly in situations with missing data.

Table 2. Model metrics for different prediction methods with data missing.

	APE (%)			RMSE (Cycles)		
	Train	**Validation**	**Test**	**Train**	**Validation**	**Test**
Attention net	8.8	9.3	10.3	77	102	100
Elastic Net [29]	11.2	-	14.4	112	-	165
CNN	17.8	14.1	18.7	189	135	239
CNN-LSTM [51]	12.7	12.3	15.8	98	96	140

Lastly, to further explore the adaptability and capabilities of the attention-based network in the scenarios of extreme data missing, we assessed the predictive performance of our proposed method under varying data-missing degrees. The results of these tests are illustrated in Figure 9. In the experimental settings, all the cases are randomly repeated five times, and the circle in the figure represents the mean value of test APEs, while the shaded areas represent the APE values within one standard deviation. The results demonstrate that the attention-based network consistently exhibits robust predictive performance in scenarios of cycle missing and random missing, even as the degree of data missingness escalates to 10%. However, it encounters slight limitations in the context of time-step missing, preserving its effectiveness up to a maximum of 5% degree of data missingness. Overall, the superior ability of the attention-based network for battery lifetime prediction with missing data underscores its significance in real-world settings where data loss can be a prevailing issue.
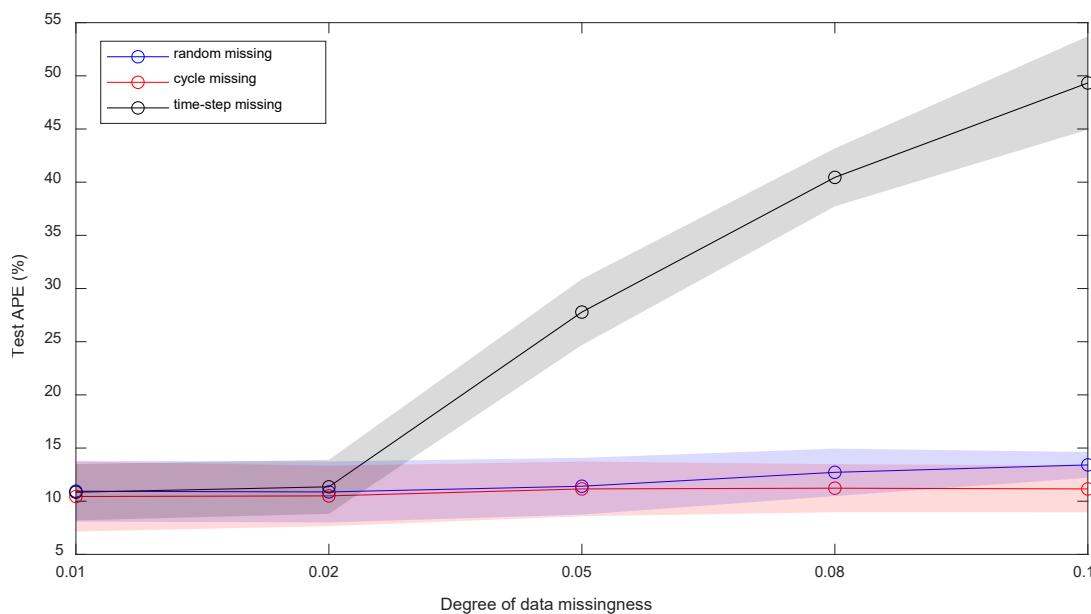
Figure 9. The impact of data-missing degree on battery lifetime prediction in different data-missing scenarios. The shaded areas represent the values within one standard deviation across the results of five different train/valid/test splits.

5. Conclusions

In this article, we proposed an innovative neural network that incorporates the self-attention mechanism, specifically designed to tackle the challenges of extracting meaningful information from sequential data in battery cycling experiments for accurate battery cycle life prediction. The results from our study indicate that our proposed self-attention mechanism-based network approach performs higher accuracy compared to other benchmarks in this field under the conditions without data missingness, where the average test APE reaches 9.34%, improving the performance by approximately 35.1% relative to the original article of the dataset [29]. Furthermore, our proposed method exhibits robust adaptability across three distinct data-missing scenarios, i.e., random missing, cycle missing, and time-step missing, achieving test APEs of 11.3%, 10.9%, and 11.1%, respectively, and maintaining good performance under all data-missing scenarios. This adaptability plays a pivotal role in preserving the accuracy of the predictions under various data-loss conditions. Moreover, even in the face of these data-missing scenarios, our proposed method consistently outperforms the benchmarking methods within the field, with an average test APE of 10.3%, signifying its superior predictive power. Additionally, the attention-based network maintains good predictive performance even as the extent of data missing intensifies. Notably, the test APEs under cycle-missing and random-missing scenarios remained under 13.5% even as the data-missing degree escalated to 10%, showcasing its resilience in increasingly challenging circumstances.

When considering a broader range of applications, our proposed self-attention-based neural network, as a supervised machine learning approach, has the potential to remain applicable under different conditions, such as varying depths of discharge (DOD). DOD is a critical factor that directly influences battery performance and lifetime, with shallower discharges generally leading to longer battery life compared to deeper discharges. The primary requirement for adapting our method to diverse conditions is the collection of relevant data under the respective conditions to serve as the training set. With appropriate training, our method can predict battery lifetime or other performance metrics under those specific conditions. The successful implementation of the attention mechanism in this context not only highlights its potential in battery cycle life prediction but also underscores its applicability in critical areas such as battery management systems that rely on accurate cycle life prediction.

References

1. Liang, Y.; Zhao, C.Z.; Yuan, H.; Chen, Y.; Zhang, W.; Huang, J.Q.; Yu, D.; Liu, Y.; Titirici, M.M.; Chueh, Y.L.; et al. A Review of Rechargeable Batteries for Portable Electronic Devices. *InfoMat* **2019**, *1*, 6–32. [CrossRef]
2. Olabi, A.G.; Abbas, Q.; Shinde, P.A.; Abdelkareem, M.A. Rechargeable Batteries: Technological Advancement, Challenges, Current and Emerging Applications. *Energy* **2023**, *266*, 126408. [CrossRef]
3. König, A.; Nicoletti, L.; Schröder, D.; Wolff, S.; Waclaw, A.; Lienkamp, M. An Overview of Parameter and Cost for Battery Electric Vehicles. *World Electr. Veh. J.* **2021**, *12*, 21. [CrossRef]
4. Lebrouhi, B.E.; Khattari, Y.; Lamrani, B.; Maaroufi, M.; Zeraouli, Y.; Kousksou, T. Key Challenges for a Large-Scale Development of Battery Electric Vehicles: A Comprehensive Review. *J. Energy Storage* **2021**, *44*, 103273. [CrossRef]
5. Tran, M.K.; Bhatti, A.; Vrolyk, R.; Wong, D.; Panchal, S.; Fowler, M.; Fraser, R. A Review of Range Extenders in Battery Electric Vehicles: Current Progress and Future Perspectives. *World Electr. Veh. J.* **2021**, *12*, 54. [CrossRef]
6. Cheng, X.; Liu, H.; Yuan, H.; Peng, H.; Tang, C.; Huang, J.; Zhang, Q. A Perspective on Sustainable Energy Materials for Lithium Batteries. *SusMat* **2021**, *1*, 38–50. [CrossRef]
7. Hannan, M.A.; Wali, S.B.; Ker, P.J.; Rahman, M.S.A.; Mansor, M.; Ramachandaramurthy, V.K.; Muttaqi, K.M.; Mahlia, T.M.I.; Dong, Z.Y. Battery Energy-Storage System: A Review of Technologies, Optimization Objectives, Constraints, Approaches, and Outstanding Issues. *J. Energy Storage* **2021**, *42*, 103023. [CrossRef]
8. Weiss, M.; Ruess, R.; Kasnatscheew, J.; Levartovsky, Y.; Levy, N.R.; Minnmann, P.; Stolz, L.; Waldmann, T.; Wohlfahrt-Mehrens, M.; Aurbach, D.; et al. Fast Charging of Lithium-Ion Batteries: A Review of Materials Aspects. *Adv. Energy Mater.* **2021**, *11*, 2101126. [CrossRef]
9. Kebede, A.A.; Kalogiannis, T.; Van Mierlo, J.; Berecibar, M. A Comprehensive Review of Stationary Energy Storage Devices for Large Scale Renewable Energy Sources Grid Integration. *Renew. Sustain. Energy Rev.* **2022**, *159*, 112213. [CrossRef]
10. Chen, T.; Jin, Y.; Lv, H.; Yang, A.; Liu, M.; Chen, B.; Xie, Y.; Chen, Q. Applications of Lithium-Ion Batteries in Grid-Scale Energy Storage Systems. *Trans. Tianjin Univ.* **2020**, *26*, 208–217. [CrossRef]
11. Park, S.; Pozzi, A.; Whitmeyer, M.; Perez, H.; Kandel, A.; Kim, G.; Choi, Y.; Joe, W.T.; Raimondo, D.M.; Moura, S. A Deep Reinforcement Learning Framework for Fast Charging of Li-Ion Batteries. *IEEE Trans. Transp. Electrif.* **2022**, *8*, 2770–2784. [CrossRef]
12. Newman, J.; Tiedemann, W. Porous-electrode Theory with Battery Applications. *AIChE J.* **1975**, *21*, 25–41. [CrossRef]
13. Elmahallawy, M.; Elfouly, T.; Alouani, A.; Massoud, A.M. A Comprehensive Review of Lithium-Ion Batteries Modeling, and State of Health and Remaining Useful Lifetime Prediction. *IEEE Access* **2022**, *10*, 119040–119070. [CrossRef]
14. Khodadadi Sadabadi, K.; Jin, X.; Rizzoni, G. Prediction of Remaining Useful Life for a Composite Electrode Lithium Ion Battery Cell Using an Electrochemical Model to Estimate the State of Health. *J. Power Sources* **2021**, *481*, 228861. [CrossRef]
15. Lam, F.; Allam, A.; Joe, W.T.; Choi, Y.; Onori, S. Offline Multiobjective Optimization for Fast Charging and Reduced Degradation in Lithium Ion Battery Cells. In Proceedings of the 2021 American Control Conference (ACC), New Orleans, LA, USA, 25–28 May 2021; pp. 4441–4446. [CrossRef]
16. O'Kane, S.E.J.; Ai, W.; Madabattula, G.; Alonso-Alvarez, D.; Timms, R.; Sulzer, V.; Edge, J.S.; Wu, B.; Offer, G.J.; Marinescu, M. Lithium-Ion Battery Degradation: How to Model It. *Phys. Chem. Chem. Phys.* **2022**, *24*, 7909–7922. [CrossRef] [PubMed]
17. Jokar, A.; Rajabloo, B.; Désilets, M.; Lacroix, M. Review of Simplified Pseudo-Two-Dimensional Models of Lithium-Ion Batteries. *J. Power Sources* **2016**, *327*, 44–55. [CrossRef]
18. Kong, X.; Plett, G.L.; Scott Trimboli, M.; Zhang, Z.; Qiao, D.; Zhao, T.; Zheng, Y. Pseudo-Two-Dimensional Model and Impedance Diagnosis of Micro Internal Short Circuit in Lithium-Ion Cells. *J. Energy Storage* **2020**, *27*, 101085. [CrossRef]
19. Liu, B.; Tang, X.; Gao, F. Joint Estimation of Battery State-of-Charge and State-of-Health Based on a Simplified Pseudo-Two-Dimensional Model. *Electrochim. Acta* **2020**, *344*, 136098. [CrossRef]
20. Ashwin, T.R.; Chung, Y.M.; Wang, J. Capacity Fade Modelling of Lithium-Ion Battery under Cyclic Loading Conditions. *J. Power Sources* **2016**, *328*, 586–598. [CrossRef]
21. Han, X.; Ouyang, M.; Lu, L.; Li, J. Simplification of Physics-Based Electrochemical Model for Lithium Ion Battery on Electric Vehicle. Part I: Diffusion Simplification and Single Particle Model. *J. Power Sources* **2015**, *278*, 802–813. [CrossRef]

22. Han, X.; Ouyang, M.; Lu, L.; Li, J. Simplification of Physics-Based Electrochemical Model for Lithium Ion Battery on Electric Vehicle. Part II: Pseudo-Two-Dimensional Model Simplification and State of Charge Estimation. *J. Power Sources* **2015**, *278*, 814–825. [CrossRef]

23. Lee, J.L.; Chemistruck, A.; Plett, G.L. Discrete-Time Realization of Transcendental Impedance Models, with Application to Modeling Spherical Solid Diffusion. *J. Power Sources* **2012**, *206*, 367–377. [CrossRef]

24. Stetzel, K.D.; Aldrich, L.L.; Trimboli, M.S.; Plett, G.L. Electrochemical State and Internal Variables Estimation Using a Reduced-Order Physics-Based Model of a Lithium-Ion Cell and an Extended Kalman Filter. *J. Power Sources* **2015**, *278*, 490–505. [CrossRef]

25. Deng, Z.; Hu, X.; Lin, X.; Xu, L.; Li, J.; Guo, W. A Reduced-Order Electrochemical Model for All-Solid-State Batteries. *IEEE Trans. Transp. Electrif.* **2021**, *7*, 464–473. [CrossRef]

26. Lai, X.; He, L.; Wang, S.; Zhou, L.; Zhang, Y.; Sun, T.; Zheng, Y. Co-Estimation of State of Charge and State of Power for Lithium-Ion Batteries Based on Fractional Variable-Order Model. *J. Clean. Prod.* **2020**, *255*, 120203. [CrossRef]

27. Parhizi, M.; Pathak, M.; Ostanek, J.K.; Jain, A. An Iterative Analytical Model for Aging Analysis of Li-Ion Cells. *J. Power Sources* **2022**, *517*, 230667. [CrossRef]

28. Ng, M.F.; Zhao, J.; Yan, Q.; Conduit, G.J.; Seh, Z.W. Predicting the State of Charge and Health of Batteries Using Data-Driven Machine Learning. *Nat. Mach. Intell.* **2020**, *2*, 161–170. [CrossRef]

29. Severson, K.A.; Attia, P.M.; Jin, N.; Perkins, N.; Jiang, B.; Yang, Z.; Chen, M.H.; Aykol, M.; Herring, P.K.; Fraggedakis, D.; et al. Data-Driven Prediction of Battery Cycle Life before Capacity Degradation. *Nat. Energy* **2019**, *4*, 383–391. [CrossRef]

30. Zhang, Y.; Feng, X.; Zhao, M.; Xiong, R. In-Situ Battery Life Prognostics amid Mixed Operation Conditions Using Physics-Driven Machine Learning. *J. Power Sources* **2023**, *577*, 233246. [CrossRef]

31. Hong, J.; Lee, D.; Jeong, E.R.; Yi, Y. Towards the Swift Prediction of the Remaining Useful Life of Lithium-Ion Batteries with End-to-End Deep Learning. *Appl. Energy* **2020**, *278*, 115646. [CrossRef]

32. Thelen, A.; Li, M.; Hu, C.; Bekyarova, E.; Kalinin, S.; Sanghadasa, M. Augmented Model-Based Framework for Battery Remaining Useful Life Prediction. *Appl. Energy* **2022**, *324*, 119624. [CrossRef]

33. Nuhic, A.; Terzimehic, T.; Soczka-Guth, T.; Buchholz, M.; Dietmayer, K. Health Diagnosis and Remaining Useful Life Prognostics of Lithium-Ion Batteries Using Data-Driven Methods. *J. Power Sources* **2013**, *239*, 680–688. [CrossRef]

34. Tseng, K.H.; Liang, J.W.; Chang, W.; Huang, S.C. Regression Models Using Fully Discharged Voltage and Internal Resistance for State of Health Estimation of Lithium-Ion Batteries. *Energies* **2015**, *8*, 2889–2907. [CrossRef]

35. Mansouri, S.S.; Karvelis, P.; Georgoulas, G.; Nikolakopoulos, G. Remaining Useful Battery Life Prediction for UAVs Based on Machine Learning. *IFAC-PapersOnLine* **2017**, *50*, 4727–4732. [CrossRef]

36. Guo, J.; Li, Z.; Pecht, M. A Bayesian Approach for Li-Ion Battery Capacity Fade Modeling and Cycles to Failure Prognostics. *J. Power Sources* **2015**, *281*, 173–184. [CrossRef]

37. Khumprom, P.; Yodo, N. A Data-Driven Predictive Prognostic Model for Lithium-Ion Batteries Based on a Deep Learning Algorithm. *Energies* **2019**, *12*, 660. [CrossRef]

38. Ren, L.; Zhao, L.; Hong, S.; Zhao, S.; Wang, H.; Zhang, L. Remaining Useful Life Prediction for Lithium-Ion Battery: A Deep Learning Approach. *IEEE Access* **2018**, *6*, 50587–50598. [CrossRef]

39. Zhang, Y.; Xiong, R.; He, H.; Liu, Z. A LSTM-RNN Method for the Lithuim-Ion Battery Remaining Useful Life Prediction. In Proceedings of the 2017 Prognostics and System Health Management Conference, PHM-Harbin 2017—Proceedings, Harbin, China, 9–12 July 2017; pp. 1–4.

40. 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*, 105992. [CrossRef]

41. Liao, L.; Köttig, F. A Hybrid Framework Combining Data-Driven and Model-Based Methods for System Remaining Useful Life Prediction. *Appl. Soft Comput. J.* **2016**, *44*, 191–199. [CrossRef]

42. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008. [CrossRef]

43. Severson, K.A.; Monian, B.; Love, J.C.; Braatz, R.D. A Method for Learning a Sparse Classifier in the Presence of Missing Data for High-Dimensional Biological Datasets. *Bioinformatics* **2017**, *33*, 2897–2905. [CrossRef] [PubMed]

44. Severson, K.A.; Molaro, M.C.; Braatz, R.D. Principal Component Analysis of Process Datasets with Missing Values. *Processes* **2017**, *5*, 38. [CrossRef]

45. Jeong, H.; Wang, H.; Calmon, F.P. Fairness without Imputation: A Decision Tree Approach for Fair Prediction with Missing Values. In Proceedings of the 36th AAAI Conference on Artificial Intelligence, Virtual, 22 February–1 March 2022.

46. Twala, B.E.T.H.; Jones, M.C.; Hand, D.J. Good Methods for Coping with Missing Data in Decision Trees. *Pattern Recognit. Lett.* **2008**, *29*, 950–956. [CrossRef]

47. Chen, L.; Xu, Y.; Zhu, Q.X.; He, Y.L. Adaptive Multi-Head Self-Attention Based Supervised VAE for Industrial Soft Sensing with Missing Data. *IEEE Trans. Autom. Sci. Eng.* **2023**, 1–12. [CrossRef]

48. Wu, R.; Zhang, A.; Ilyas, I.F.; Rekatsinas, T. Attention-Based Learning for Missing Data Imputation in HoloClean. In Proceedings of the 3rd MLSys Conference, Austin, TX, USA, 15 March 2020.

49. Yu, J.; Wu, B. Attention and Hybrid Loss Guided Deep Learning for Consecutively Missing Seismic Data Reconstruction. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5902108. [CrossRef]

50. Ansari, S.; Ayob, A.; Hossain Lipu, M.S.; Hussain, A.; Saad, M.H.M. Remaining Useful Life Prediction for Lithium-Ion Battery Storage System: A Comprehensive Review of Methods, Key Factors, Issues and Future Outlook. *Energy Rep.* **2022**, *8*, 12153–12185. [CrossRef]
51. Ma, G.; Zhang, Y.; Cheng, C.; Zhou, B.; Hu, P.; Yuan, Y. Remaining Useful Life Prediction of Lithium-Ion Batteries Based on False Nearest Neighbors and a Hybrid Neural Network. *Appl. Energy* **2019**, *253*, 113626. [CrossRef]