batteries

MDPI

Article

Hybrid Neural Networks for Enhanced Predictions of Remaining Useful Life in Lithium-Ion Batteries

Alireza Rastegarpanah [1,2,*,†] , Mohammed Eesa Asif [1,†] and Rustam Stolkin [1,2]

1   Department of Metallurgy & Materials Science, University of Birmingham, Birmingham B15 2TT, UK; mxa1586@student.bham.ac.uk (M.E.A.); r.stolkin@bham.ac.uk (R.S.)
2   The Faraday Institution, Quad One, Harwell Science and Innovation Campus, Didcot OX11 0RA, UK
*   Correspondence: a.rastegarpanah@bham.ac.uk
†   These authors contributed equally to this work.

Abstract: With the proliferation of electric vehicles (EVs) and the consequential increase in EV battery circulation, the need for accurate assessments of battery health and remaining useful life (RUL) is paramount, driven by environmentally friendly and sustainable goals. This study addresses this pressing concern by employing data-driven methods, specifically harnessing deep learning techniques to enhance RUL estimation for lithium-ion batteries (LIB). Leveraging the Toyota Research Institute Dataset, consisting of 124 lithium-ion batteries cycled to failure and encompassing key metrics such as capacity, temperature, resistance, and discharge time, our analysis substantially improves RUL prediction accuracy. Notably, the convolutional long short-term memory deep neural network (CLDNN) model and the transformer LSTM (temporal transformer) model have emerged as standout remaining useful life (RUL) predictors. The CLDNN model, in particular, achieved a remarkable mean absolute error (MAE) of 84.012 and a mean absolute percentage error (MAPE) of 25.676. Similarly, the temporal transformer model exhibited a notable performance, with an MAE of 85.134 and a MAPE of 28.7932. These impressive results were achieved by applying Bayesian hyperparameter optimization, further enhancing the accuracy of predictive methods. These models were bench-marked against existing approaches, demonstrating superior results with an improvement in MAPE ranging from 4.01% to 7.12%.

Keywords: deep learning; lithium-ion batteries; battery management systems; EV battery recycling; battery degradation

1. Introduction

The prediction of remaining useful life (RUL) for lithium-ion batteries is a critical task for ensuring the safe and optimal operation of battery packs, especially in applications like electric vehicles (EVs) [1]. This literature delves into two primary approaches for RUL prediction: physics-based models (PBMs) and data-driven models (DDMs). PBMs leverage the fundamental principles of electrochemistry and battery physics to simulate battery behaviour over time [2]. These models consider factors such as ion diffusion, electrode reactions [3], discharge capacity [4,5], cycles, capacity fade [6], and thermal effects [7]. While PBMs provide valuable insights into degradation mechanisms, they are computationally intensive, require detailed knowledge of electrochemical processes [2,8], and may struggle to capture real-world complexities. DDMs, on the other hand, use machine learning algorithms to learn patterns and relationships directly from available data [9,10]. They have gained prominence due to their ability to capture complex and nonlinear relationships that exist in the data, making them more adaptable and flexible than PBMs.

DDMs in machine learning are broadly categorized into two distinct groups: statistical machine learning and deep learning, each offering unique strengths and applications in various domains. Statistical machine learning, often referred to as shallow learning, typically

encompasses models that are less complex and computationally intensive. This category includes a range of methods such as support vector machines (SVMs) [10], Gaussian process regression (GPR) [11–13], random forest [14,15], and Bayesian approaches [16–18]. These techniques are particularly effective for smaller datasets where prior knowledge about the generative process of the data is available. However, they often encounter challenges when addressing complex scenarios, such as accurately capturing intricate battery characteristics and long-term dependencies within datasets [2,19].

In contrast, deep learning methods, which utilize neural networks with multiple hidden layers, are adept at handling large, complex datasets, often where there is limited a priori knowledge about the underlying processes or the most suitable features. This category includes advanced techniques like recurrent neural networks (RNNs) [20,21], convolutional neural networks (CNNs) [22,23], and various hybrid models [24,25]. Deep learning approaches are renowned for their ability to discern and learn from intricate patterns present in raw data, making them particularly effective in the analysis of multivariate time-series information. This capability makes them especially valuable in fields where understanding and predicting complex behaviours over time is crucial.

Recent advancements in sequence-to-sequence deep learning in the domain of natural language processing further contribute to the discourse on multi-horizon time-series forecasting (MTSF). Sutskever et al. introduced a powerful end-to-end approach employing multilayered long short-term memory (LSTM) networks for sequence learning, showcasing impressive results in translation tasks [26]. Similarly, Yang et al. explored incorporating a cross-entity attention mechanism in MTSF in [27]. This method minimizes assumptions on sequence structures, particularly in tasks where large labelled training sets are available. The use of time-series forecasting has also been adopted in RUL prediction, as shown by [28], which introduces a Capsule Neural Network–LSTM, a hybrid model for precise RUL prediction in mechanical systems, enhancing sensitivity to spatial features in time-series sensor data.

However, in the context of RUL prediction, improvements in robustness, generalizability, and addressing challenges like variable sampling rates and incomplete data remain areas of focus [29]. Hence, despite the progress in RUL prediction, current research has several knowledge gaps and limitations. These include (1) the need for more robust and accurate models, for enhanced generalizability, (2) the comprehensive exploration of various deep learning architectures, (3) the utilisation of complete datasets with varying sampling rates, and the consideration of factors like battery ageing and non-stationary signals.

Rationale for Advanced Model Architectures over Regression

The progression in methodologies for predicting the remaining useful life (RUL) of lithium-ion batteries underscores a pivotal shift toward enhancing the operational efficiency and reliability of battery-powered systems. Building upon the groundwork laid by Severson et al. (2019) [30], who utilized regression-based methods for early-cycle data prediction, our study uses the same dataset for advanced machine learning exploration. Diverging significantly, we employ sophisticated hybrid deep learning architectures, such as convolutional long short-term memory deep neural networks (CLDNN) and temporal transformers, which are adept at intricately modelling the spatial and temporal dynamics of battery degradation. This strategic departure from linear regression to deep learning is instrumental in capturing the multifaceted and nonlinear aspects of battery ageing, thereby elevating the precision and dependability of our predictions within a dynamic 100-cycle window. The rationale behind exploring temporal deep learning approaches based on neural networks over vanilla regression lies in their superior ability to process and learn from sequential data, providing a nuanced understanding of degradation patterns that enhance predictive accuracy and reliability, and broadening the scope for practical applications in battery life-cycle management.

Further, we refine the dataset through a rigorous feature engineering and selection process, including techniques such as linear interpolation and principal component analysis

(PCA), which enhance model training and provide a broader application range. Coupled with precise data preprocessing and sampling techniques, such as stratified random sampling and outlier removal, our methodology assures robust and dependable predictions across varied battery use cases. The incorporation of Bayesian optimization for hyperparameter tuning is a testament to our dedication to achieving optimal model performance, tailored specifically to the dataset's nuances (an overview can be seen in Figure 1). It encompasses crucial parameters such as discharge capacity, temperature, internal resistance, and discharge time.

To effectively address the heterogeneity within our dataset, we adopted a hybrid approach known as CLDNN proposed by Sainath et al. [31], which stands for convolutional, long short-term memory, and dense neural networks. CLDNN harnesses the collective power of these neural network architectures, providing a solution to the multifaceted nature of LIB RUL prediction. In addition, we repurposed the hybrid model called the temporal transformer (TT) proposed by Chadha et al. [32] to enhance prediction accuracy and robustness in LIB RUL. The TT model combines the strengths of transformer self-attention layers and LSTM architectures, presenting a unique approach to sequential modelling that effectively addresses the challenge of capturing long-term dependencies [33]. While the temporal transformer shares a name with the temporal fusion transformer (TFT) introduced by Lim et al. [34], it is important to note their architectural differences. The TFT is designed for handling multi-modal data, allowing it to incorporate various relevant features for forecasting tasks. In contrast, our dataset did not require such multi-modal capabilities, leading to divergent architectural choices in our models.
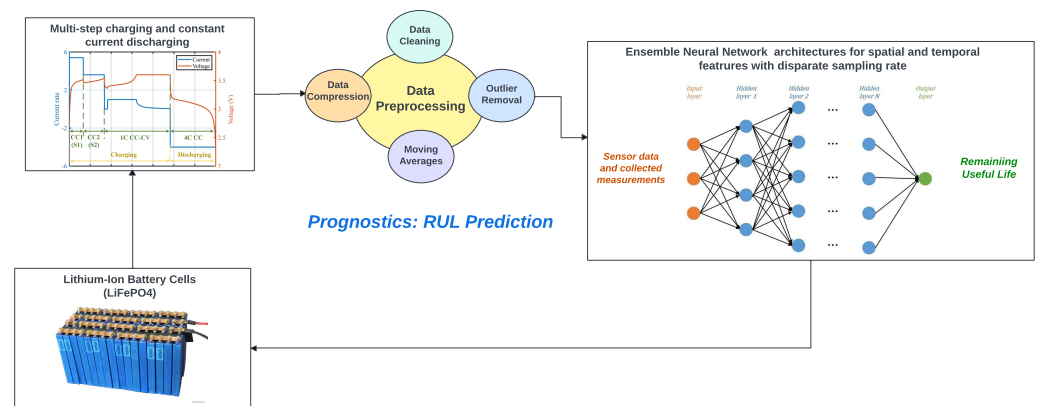

Figure 1. An overview of the proposed advanced data refinement and model optimization.

The remainder of this article is structured as follows. In Section 2, we delve into the related works within the field of lithium-ion battery RUL prediction, focusing on deep-learning approaches. Section 3 provides a comprehensive overview of our proposed methodology, followed by a detailed account of the experimental procedure. Section 4 is dedicated to discussions regarding the outcomes of our experiments. Lastly, Section 5 closes this article with our conclusion and future work.

2. Related Works

Physics-based models (PBMs) provide valuable insights into the underlying mechanisms of lithium-ion batteries but often face challenges when predicting their remaining useful life (RUL) [2]. These models perform optimally within narrow domains, requiring a detailed understanding of battery electrochemical processes, which can be time-consuming and computationally intensive [3,8,35]. The simplified assumptions often made in these models limit their ability to generalize, affecting RUL prediction accuracy. Furthermore, parametrisation in PBMs is challenging, due to manufacturing variations and the difficulty of obtaining precise model parameters [36,37]. These models also exhibit limited adaptability to dynamic environments and high computational complexity, which hampers their real-time applicability. Calibration and validation procedures for these models demand

extensive data and resources [38]. Consequently, data-driven methods like machine learning have gained prominence for their ability to learn complex nonlinear relationships from data, overcoming uncertainties, and providing a promising alternative for RUL prediction in LIB [39–42].

2.1. Statistical Machine Learning

As previously noted, statistical machine learning methods, including support vector machines (SVMs) [10], support vector regression with Kalman filters [43], Gaussian process regression (GPR) [11–13], random forest [14,15], and Bayesian approaches [16–18], have been employed for predicting battery RUL. Tong et al. [44] introduced a deep-learning-based algorithm called adaptive dropout long short-term memory (ADLSTM) and Monte Carlo (MC) simulations, achieving high precision with only 25% of degradation data. Similarly, Kwon et al. [45] integrated a multi-linear regression approach with a recurrent neural network to model LIB degradation.

2.2. Temporal Models

Temporal models, like recurrent neural networks (RNNs) and more advanced variants such as LSTM cells and gated recurrent units (GRUs), are employed to capture the sequential nature of battery data. Lipu et al. [20] introduced a nonlinear auto-regressive with exogenous input (NARX)-based neural network (NARXNN) for state of charge (SOC) estimation, utilizing a lighting search algorithm (LSA) to optimise hyperparameters. Signal decomposition techniques like discrete wavelet transform (DWT), empirical mode decomposition (EMD), and variational mode decomposition (VMD) are combined with the NARX model in [46] to predict capacity degradation trajectories. In [47], Zheng et al. employed a deep LSTM network followed by multiple fully connected (FC) layers, whereas Wang et al. in [48] introduced a bidirectional LSTM architecture with additional FC layers. Chemali et al. [21] proposed an RNN architecture with LSTM for RUL estimation, achieving low mean absolute error (MAE) values and demonstrating generalization across different conditions.

2.3. Convolutional Models

Shen et al. [23] presented a deep convolutional neural network (DCNN) that achieved higher accuracy than traditional methods, such as filter-based models [49–52]. However, limitations include the fixed-size input matrix and a limited understanding of the method's interpretability. Following a similar approach, Li et al. [53] proposed another DCNN, employing a time window approach for sample preparation and feature extraction. In another work, [54], Babu et al. proposed a novel deep CNN-based regression method for RUL estimation. Similarly, Shen et al. in [14] proposed a DCNN with transfer learning (DCNN–TL) model, which was then integrated with ensemble learning to form DCNN–ETL. The effectiveness of the DCNN–ETL model was then benchmarked against five other data-driven methods including random forest regression, Gaussian process regression, and DCNN. In another study, a temporal convolutional network (TCN) model was used along with causal and dilated convolutions to capture local capacity degradation, but its limitations include its applicability to other battery types and real-time implementation challenges [24].

2.4. Hybrid Models

Ren et al. proposed an auto-CNN–LSTM model for RUL estimation, incorporating an auto-encoder to enhance feature vectors through dimensionality reduction and feature learning, achieving accurate predictions [25]. Identified knowledge gaps include the need for enhanced generalizability, further exploration of various temporal networks, and improved utilization of datasets with varying feature sampling rates. Additionally, it is suggested that future research should address the ageing effects, the impact of non-stationary signals, the incorporation of relevant features, and the capture of spatial information. Li et al.

introduced a directed acyclic graph network that combined the LSTM with the CNN, offering a novel approach to RUL prediction [55]. Tan et al. focused on a multi-variate time-series approach using a lightweight CNN with attention mechanisms to enhance prediction accuracy [56]. Hybrid models, such as the one combining a GRU and a CNN for state of health (SOH) estimations, presented by Fan et al., and the attention-assisted temporal convolutional memory-augmented network (ATCMN) for RUL prediction from limited data, proposed by Fei et al., demonstrate promise but require further exploration and rigorous testing on more diverse datasets [57,58]. In previous work, the effectiveness of CNN–LSTM and CNN–LSTM–Differential-Neural-Computer models was demonstrated, showcasing not only superior predictive accuracy but also efficiency in learning temporal dependencies with fewer epochs [59,60].

In the following section, we aim to address critical knowledge gaps identified in the literature regarding the prediction of the RUL of LIBs. These gaps encompass issues of robustness, accuracy, and generalizability in DDMs for battery management systems (BMSs). To overcome these limitations, we developed end-to-end hybrid models capable of utilizing complete datasets [61], including features with varying sampling rates. By doing so, we intend to enhance the performance metrics for RUL estimation in LIBs and move closer to enabling real-time implementation of these models within BMS for practical applications. Additionally, we explored various deep learning approaches, such as convolutional neural networks (CNNs) [62], long short-term memory cells (LSTMs) [33], the Transformer network [63], autoencoder [64], neural Turing machines [65], differentiable neural computers [66], and hybrid models, while considering the impact of ageing, non-stationary signals, relevant feature incorporation, spatial information capture, and variable ambient temperature conditions.

3. Methodology

This section presents the framework (Figure 2) of the proposed LIB RUL prediction method. An important part of this research was the optimization of model hyperparameters using Bayesian optimization techniques aimed at maximizing the models' predictive accuracy. The primary objective was to develop a robust predictive model for estimating the RULs of LIBs, accounting for intricate temporal battery degradation dynamics and feature variations in sensor measurements across different battery batches.
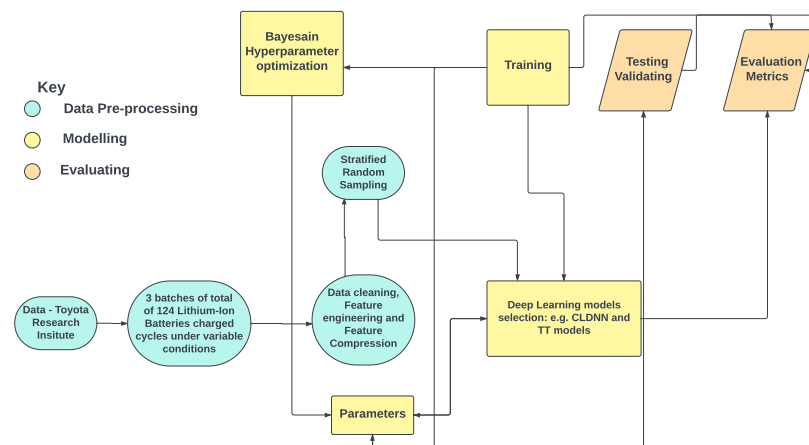

Figure 2. The study pipeline begins with a dataset comprising 124 commercial LIBs that have undergone extensive cycling until failure. Next, we implement feature selection to extract crucial features that enhance the accuracy of RUL predictions. Following this, a series of data preprocessing techniques are employed to clean and compress the dataset. Subsequently, stratified random sampling is utilized to create a representative sample. Finally, we develop a variety of hybrid deep network architectures, apply hyperparameter optimization, and evaluate these models using a testing set to determine the best-performing model.

3.1. Dataset Description

The dataset employed in this study consists of 124 commercial lithium iron phosphate (LFP)/graphite lithium-ion batteries, each with a nominal capacity of 1.1 Ah and a nominal voltage of 3.3 V. These batteries were subjected to cycling tests until failure under conditions aimed at fast charging within a temperature-controlled environment set at 30 °C. Our dataset is derived from extensive data collection efforts, capturing critical parameters such as voltage, current, temperature, and internal resistance, which were continuously measured throughout the cycling process. The term 'continuously measured' refers to the collection of data at frequent, regular intervals throughout each battery's charging and discharging cycles. This high-resolution data capture allows for an in-depth analysis of battery performance over time, enabling the precise modelling of degradation patterns. The charging protocol involved an initial phase utilizing a current of C1 until reaching a state-of-charge (SOC) of S1, followed by a transition to a current of C2 until achieving a SOC of S2, consistently maintained at 80% for all cells. Subsequently, a constant current–constant voltage (CC–CV) charging technique was applied for the transition from 80% to 100% SOC at a rate of 1 C, culminating at a cut-off voltage of 3.6 V [67]. The batteries' lifetimes, defined by the cycle count at which the capacity declined to 80% of its initial value, varied from 150 to 2300 cycles, showcasing a diverse range of degradation behaviors.

3.2. Stratified Random Sampling

Stratified random sampling was employed to create a representative dataset for the model's training and testing. This approach ensured that cells from different battery batches, characterised by varying quality control protocols, were proportionally included in the dataset. By preventing models from overfitting to specific batch attributes, this method improved model robustness and generalisation across different LIB batches.
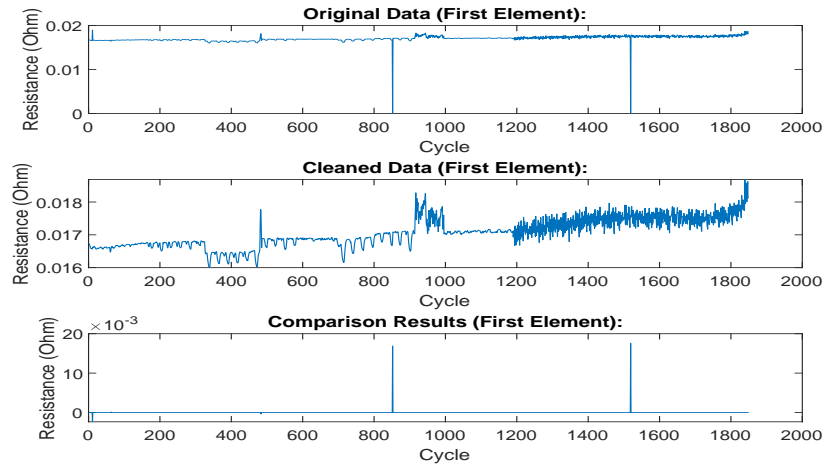
3.3. Feature Selection

The dataset consists of three groups of approximately 48 lithium-ion battery cells each, to identify features critical for predicting the remaining useful life (RUL) of the batteries. Our analysis focuses on key parameters, which include linearly interpolated discharge capacity (Qdlin) and linearly interpolated temperature (Tdlin). These features are standardized through linear interpolation to maintain uniform analysis conditions across all cells, ensuring accurate predictions by aligning data on a consistent timeline. The interpolation process applied maintains a consistent sampling rate for all cells, a critical factor for the accurate prediction of RUL and ensuring the reliability of time-series forecasting models. The selection also encompasses internal resistance (IR) and discharge characteristics, which are vital for evaluating battery health and degradation. This approach ensures the reliability of our forecasting models by providing a detailed view of battery performance over time. These features provide descriptive, summary, and cycle data, offering a holistic view of the batteries' operational conditions, performance metrics, and cycle-by-cycle behaviour. This supports a nuanced analysis, leveraging Qdlin and Tdlin alongside other selected parameters to assess battery health and predict its lifespan effectively. The output metric, remaining cycles, is calculated to estimate each battery's potential duration until its capacity falls to 80% of its original value, serving as a key indicator of its useful life.
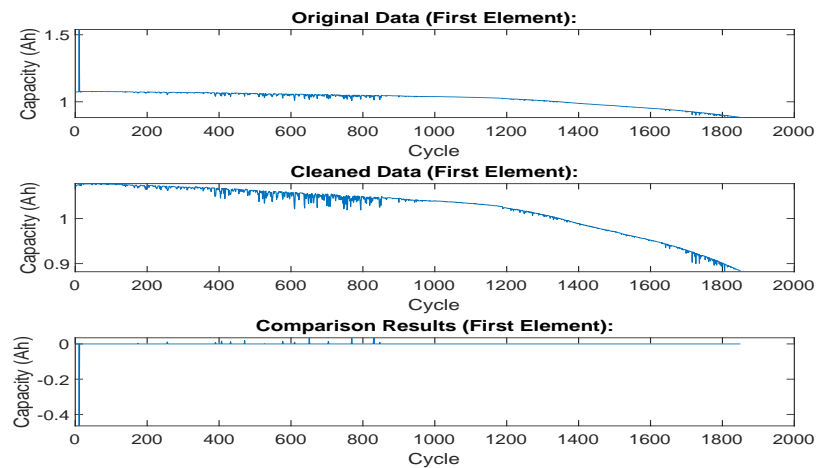
3.4. Data Preprocessing

To prepare the data for deep neural networks, we implemented a comprehensive preprocessing pipeline. Outliers in internal resistance, discharge time, and discharge quantity were removed using the fill outliers function with the cubic spline method and a moving average window of 100 (Figure 3a, Figure 3b and Figure 3c, respectively). These figures show the features before and after preprocessing (note: these plots depict the first cell in the first batch). Smoothing techniques, such as moving average filters with a window size of 15, were applied to discharge data. To capture temporal dynamics in Qdlin and
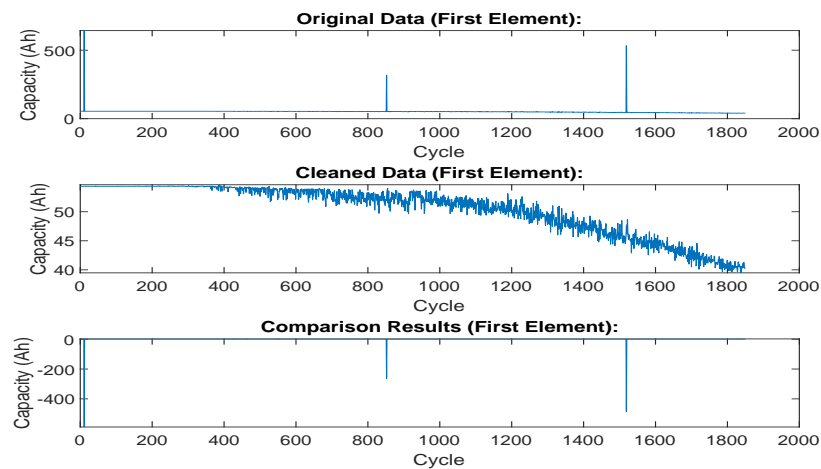
Tdlin, the sampling rate was standardized to 1000 entries per cycle. Finally, principal component analysis (PCA) was employed to reduce the dimensionality of Qdlin and Tdlin while preserving their salient features.

(**a**)

(**b**)

(**c**)

Figure 3. Comparison of data features before and after preprocessing. (**a**) Internal resistance, (**b**) discharge time, (**c**) discharge Quantity.

3.5. RUL Estimation

Traditional supervised learning approaches rely on labelled training data, where each data point is associated with a known target value. However, in the context of prognostics, this assumption often does not hold. The remaining useful life (RUL) of individual components is typically not known a priori, making it challenging to train predictive models using standard regression techniques. To overcome this hurdle, researchers have explored alternative approaches, such as employing physics-based models or utilizing machine learning algorithms to estimate the RUL of training data. While assigning a constant RUL value to all training points may seem like a straightforward solution, this can lead to inaccurate representations of the actual degradation patterns and hinder the model's ability to generalize effectively. A more sophisticated approach involves estimating the RUL based on a suitable model, as demonstrated by the use of a deep convolution neural network [23]. This approach offers a more realistic representation of the degradation process and can potentially enhance the model's predictive performance.

3.6. Metrics

A variety of metrics were used in this study. The success of the model was calculated based on the deviation (e_i) of the predicted number (model prediction $\hat{y}_i : RUL_{predicted}$) of cycles from the actual number (ground truth $y_i : RUL_{truth}$) of cycles remaining after every 100 cycle count, as shown in Equation (1). The choice of loss function significantly impacts the outcome of RUL prediction. Along with the mean absolute error (*MAE*), shown in Equation (3), and the mean absolute percentage error (*MAPE*), shown in Equation (4), which simply average the absolute errors and calculate the percentage, respectively, we also tracked the root mean squared error (*RMSE*), shown in Equation (2), which squares the errors before averaging, placing greater emphasis on larger deviations. This sensitivity to larger errors makes *RMSE* a more suitable metric for prognostics, where accurately predicting RUL is crucial, and substantial errors can lead to poor performance. Additionally, the rationale behind tracking these specific metrics is that they allow for meaningful comparisons against state-of-the-art approaches. Success in our study is quantitatively measured by the accuracy of our RUL prediction models. The *MAE* provides a direct measure of prediction accuracy in the same units as the dataset (number of cycles), *MAPE* offers a percentage-based error that is independent of the scale, and *RMSE* gives a sense of the error distribution and penalizes larger errors more severely. Our success criterion was to minimize these error metrics, thereby improving the predictive accuracy of our models.

$$e_i = \hat{y}_i - y_i \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \tag{2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |e_i| \tag{3}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|y_i - \hat{y}_i|}{|y_i|}\right) \times 100 \tag{4}$$
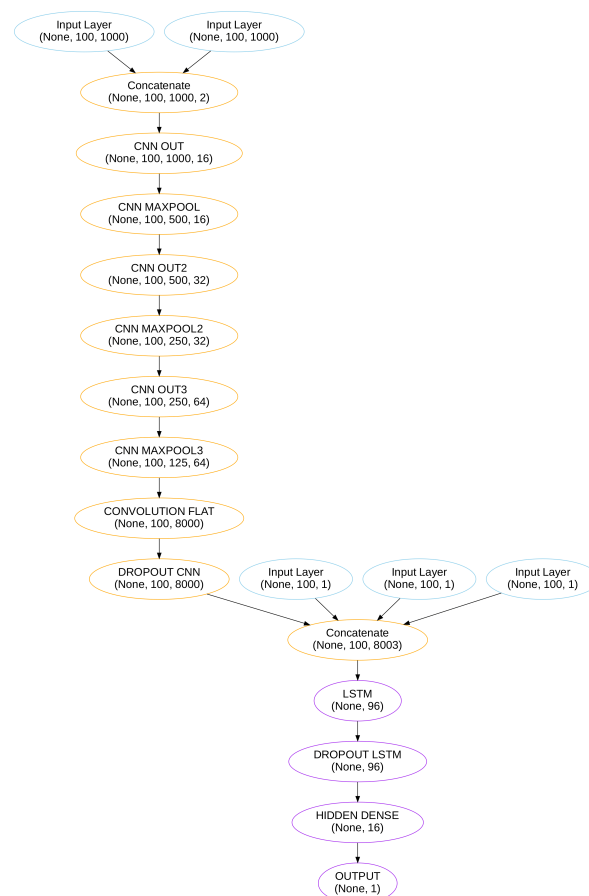
3.7. Proposed Architectures

It is worth noting that our experimentation phase encompassed the exploration of multiple temporal hybrid models; however, we only detail the most effective models. The most successful architectures in our study were the convolutional long short-term memory deep neural network (CLDNN), originally proposed by Sainath et al. [31] for natural language processing and specifically used for large vocabulary continuous speech recognition (LVCSR) tasks [31]. In their work, CLDNN outperformed the Gaussian mixture

model (GMM) and hidden Markov model (HMM) systems [68]. We repurposed the CLDNN architecture for the specific task of RUL estimation.

Another noteworthy model in our investigation was the temporal transformer (TT), initially introduced by Chadha et al. [32] for RUL estimation using the commercial modular aero-propulsion system simulation (C-MAPSS) dataset. The TT model demonstrated effectiveness in predicting the remaining useful life of aircraft engines. Ma et al. [69] presented a similar use case where their model utilized multi-head attention to capture global features from various representation sub-spaces. Although originally designed for predicting the remaining useful life of aerospace engines, we adapted these models for estimating the remaining useful life of lithium-ion batteries (LIB). This adaptation involved specific architectural deviations and adjustments to hyperparameters.

Our approach follows Occam's razor principle, which emphasizes simplicity and efficiency in model selection without sacrificing the predictive accuracy needed for effective RUL estimation [70]. This method allows for scalable solutions in battery degradation prediction by using the advantages of neural network architectures while avoiding unnecessary complexity. Regarding dataset size concerns, our hybrid deep learning models aim to balance complexity with dataset limitations. We enhanced data utilization through preprocessing strategies and incorporated dropout, regularization, and early stopping in our training processes to prevent overfitting, ensuring consistent performance (please refer to Figure 4a,b).
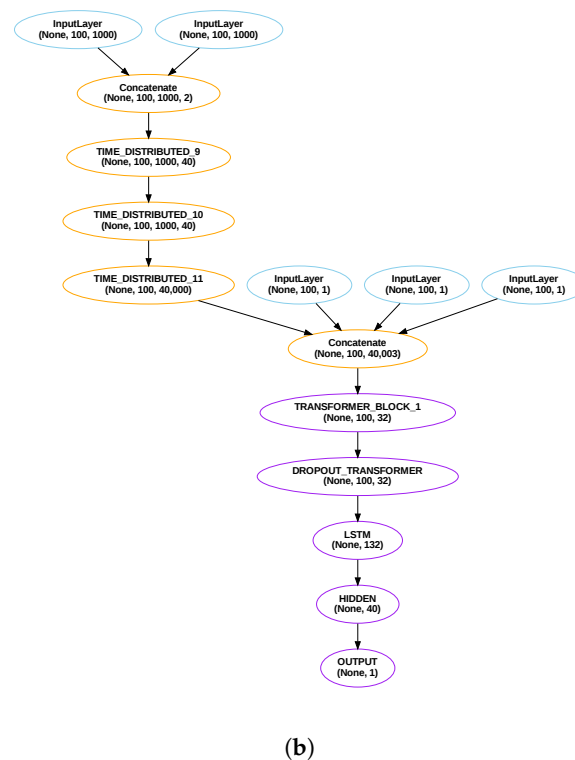

(**a**)

Figure 4. *Cont.*

(**b**)

Figure 4. Comparison of the CLDNN and temporal transformer architecture. (**a**) The CLDNN model. (**b**) The temporal transformer model.

3.7.1. The Convolutional Long Short-term Memory Deep Neural Network (CLDNN)

Adapting a convolutional long short-term memory deep neural network (CLDNN), initially developed for natural language processing (NLP) tasks with tokenized sentences, to predict remaining useful life (RUL) in lithium-ion batteries (LIB) necessitates several architectural and algorithmic modifications, as follows:

- Input Representation: In NLP tasks, CLDNN takes tokenized sentences as input. For RUL prediction, the input representation needs to be tailored to the characteristics of battery data. Time-series data from sensors measuring various parameters (voltage, current, temperature, etc.) were used as input. The input data were reshaped into a format suitable for time-series analysis.
- Sequence Length and Padding: LIB data have variable lengths of sequences as the cycle count for each battery differs, unlike fixed-length sentences in NLP. Padding or trimming sequences to a uniform length was not necessary. The network architecture was able to handle variable-length input sequences.
- Temporal Features: LIB data are inherently temporal, reflecting the degradation of the battery over time. The CLDNN architecture incorporated mechanisms to capture temporal dependencies effectively. Long short-term memory (LSTM) layers allowed us to model temporal patterns.
- Feature Extraction: The features relevant to RUL prediction in LIB differ from those important for NLP tasks. Modifications to the convolutional layers had to be made to extract features that are indicative of the battery's health and degradation.
- Hyperparameter Tuning: The hyperparameters, such as learning rate (l_r), filter size (n_f), kernel parameters (k_s), activation functions (a), and dropout rates (d_r) needed adjustment for the new task. Bayesian hyperparameter tuning was used to optimize the model for RUL prediction.
- Fine-tuning: The model was fine-tuned with different optimizer choices to minimize loss functions.

The final optimized CLDNN architecture excels at predicting the remaining useful life (RUL) of lithium-ion batteries. It comprises a total of 1,518,665 trainable parameters and combines convolutional neural networks (CNN), long short-term memory (LSTM), and dense neural networks (DNN). Each component of the network serves a specific function, as follows:

1. **CNN:** The CNN layers are used for feature extraction. They capture intricate spatial patterns within the input data, which, in this case, are time-series data from sensors measuring various parameters like voltage, current, and temperature.
2. **LSTM:** The LSTM layers are used to model temporal patterns in the data. They capture the temporal dependencies in the data, reflecting the degradation of the battery over time.
3. **DNN:** The dense neural networks are used for prediction. They contribute to the model's regularization and refined prediction capabilities.

The CLDNN architecture for our task is shown in Algorithm 1 and Figure 4a. This model demonstrates a strong efficiency in handling the heterogeneity of the dataset.

Algorithm 1 CLDNN

1: **procedure** DEFINESPATIALLAYER($Q_{dlin}, T_{dlin}, IR, DT, QD, hp$)
  Input: Q_{dlin}, T_{dlin}: Input data tensors, IR, DT, QD: Additional inputs, hp: Hyperparameters
  Concatenate: (Q_{dlin} and T_{dlin})
  Apply Convolutional Layers and Dropout:
  $h_1 \leftarrow \text{Conv1D}([Q_{dlin}, T_{dlin}], n_f, k_s, s, a, \text{padding} = \texttt{same})$
  $h_2 \leftarrow \text{MaxPooling1D}(h_1)$
  $h_3 \leftarrow \text{Conv1D}(h_2, 2n_f, k_s, s, a, \text{padding} = same)$
  $h_4 \leftarrow \text{MaxPooling1D}(h_3)$
  $h_5 \leftarrow \text{Conv1D}(h_4, 4n_f, k_s, s, a, \text{padding} = same)$
  $h_6 \leftarrow \text{MaxPooling1D}(h_5)$
  $h_7 \leftarrow \text{Flatten}(h_6)$
  $h_8 \leftarrow \text{Dropout}(h_7, d_r)$
  Features Concatenation:
  $h_9 \leftarrow \text{Concatenate}(h_8, IR, DT, QD)$
  Return h_9
2: **end procedure**
3: **procedure** DEFINELSTMLAYER(h_9, n_u, a_u)
  Input: h_9, n_u, a_u
  $h_{10} \leftarrow \text{LSTM}(h_9, n_u, a_u)$
  Return h_{10}
4: **end procedure**
5: **procedure** DEFINEDENSELAYERS(h_{10}, n_d, a_d)
  $h_{11} \leftarrow \text{Dense}(h_{10}, n_d, a_d)$
  $RUL \leftarrow \text{Dense}(h_{11}, 1, relu\_cut)$
  Return: RUL
6: **end procedure**

3.7.2. Temporal Transformer

Adapting the temporal transformer (TT) model, which was originally designed for the estimation of the remaining useful life (RUL) of aircraft engines, for the estimation of RUL in lithium-ion batteries (LIB), involves several architectural differences and adjustments. The following are some modifications that might be considered.

- Input Representation: Adjustments to the input representation to accommodate the characteristics of LIB data were required. The original model took input sequences related to engine parameters. LIB data consists of time-series measurements of capacity, temperature, resistance, and discharge time.

- Attention Mechanisms: Multi-head attention mechanisms, which were used in the original model by Ma et al. [69], needed adjustments for LIB data. Attention mechanisms were tailored to focus on features relevant to battery degradation patterns about the linearly interpolated feature.
- Model Size and Complexity: The overall size and complexity of the LIB dataset required an increase in the size and complexity of the TT model. This involved adding layers, adjusting attention mechanisms, and increasing the model depth, which led to 3,936,281 trainable parameters.
- Hyperparameter Tuning: Fine-tuning hyperparameters using Bayesian optimization specific to LIB data was required. This included learning rates, the number of attention heads, embedding dimensions, layer sizes, and dropout rates.

The temporal transformer (TT) model we used is a hybrid network that combines the transformer and long short-term memory (LSTM). Each component of the network serves a specific function, as follows:

1. **Transformer:** The transformer's multi-head self-attention mechanism allows the model to decipher complex temporal dependencies within the dataset. By parallel processing different parts of the input sequence, it extracts a rich contextual understanding of each data point.
2. **LSTM:** The LSTM units capture long-term relationships between features, enhancing the model's predictive capabilities.
3. **Feed-forward neural networks (FFNs):** The architecture employs FFNs for further refinement, facilitating the modelling of non-linear data relationships.

The TT algorithm is outlined in the provided Algorithm 2 and Figure 4b (note: $x \in \mathbb{R}^{B \times T \times D}$: input data with dimensions $B \times T \times D$, where B is the batch size, T is the time dimension, and D is the feature dimension).

The SelfAttention function in Algorithm 2 computes weighted representations of input data by considering inter-dependencies across multiple dimensions, employing a scaled dot-product attention mechanism. The TransformerBlock further refines these representations through layer normalization and feed-forward networks, enhancing their expressiveness while retaining sequential relationships. This modular and hierarchical structure allows the LSTM–transformer to capture patterns in sequential data, making it versatile for offering a robust solution for accurate RUL predictions in lithium-ion batteries.

3.8. Hyperparameter Optimization

The hyperparameter tuning for the proposed LIB RUL prediction models was achieved using Bayesian Optimization. Hyperparameter tuning is a critical step in the development of machine learning models, involving the search for optimal configurations to enhance predictive accuracy [71]. In this study, Bayesian optimization was chosen due to its effectiveness in handling non-linear and complex search spaces. Unlike traditional grid search or cross-validation methods [72], Bayesian optimization uses probabilistic models to predict the performance of different hyperparameter configurations, guiding the search toward promising regions [73]. This is particularly beneficial in high-dimensional spaces, where an exhaustive search becomes computationally expensive.

The specific implementation of this tuning process utilized the Keras Tuner library [74], a choice motivated by its compatibility and ease of integration with deep learning models (which can be seen in Algorithm 3). The process began with the definition of a hypermodel class, referred to as "MyHyperModel", derived from Keras Tuner's "HyperModel" class. This class encapsulated not only the architecture of the LIB RUL prediction models, including CLDNN or TT but also the hyperparameters' search space. Furthermore, a specialized Bayesian optimization tuner class, named "MyBayesianOptimizationTuner", was created by extending the "BayesianOptimization" class from Keras Tuner. This custom tuner was configured to minimize the validation mean absolute error (MAE), thereby optimizing the models' performance.

Algorithm 2 Temporal Transformer

1: **procedure** DEFINETIMEDISTRIBUTEDLAYER(Q_{dlin}, T_{dlin}, hp)
    Input: $Q_{dlin}, T_{dlin} \in \mathbb{R}^{B \times T \times D}$ and Hyperparameters hp ▷ *Qdlin, Tdlin*: Input data tensors, hp: Hyperparameters
    Concatenate: (Q_{dlin} and T_{dlin})   ▷ Merge input features
    Apply TimeDistributed Dense Layer and Dropout: ▷ Feature transformation and regularization
    $h \leftarrow \text{TimeDistributedDense}([Q_{dlin}, T_{dlin}], hp)$
    Flatten h ▷ Prepare for Transformer input
    Return $[Q_{dlin}, T_{dlin}]$ ▷ Linearly interpolated features compressed
2: **end procedure**
3: **procedure** DEFINESTRUCTURING(hp, IR, DT, QD)
    Extract Hyperparameters: embed_dim, num_heads, ff_dim, lstm_units, dense_units ▷ Model configuration parameters
    Define Inputs: IR, DT, QD ▷ Additional model inputs
    Features Concatenation: $[Q_{dlin}, T_{dlin}, IR, DT, QD]$ ▷ Combine all input features
    Apply TimeDistributed Dense Layer and Dropout: ▷ Further feature transformation
4:     $h \leftarrow \text{TimeDistributedDense}([Q_{dlin}, T_{dlin}, IR, DT, QD], hp)$
    Flatten h ▷ Prepare for Transformer block
    Return $[Q_{dlin}, T_{dlin}, IR, DT, QD]$ ▷ We now proceed to apply the attention mechanism across each data point
5: **end procedure**
6: **procedure** DEFINEMULTIHEADSELFATTENTION(x, D_h)   ▷ Uses three sets of weight matrices W_q, W_k, W_v to transform the input data into query (Q), key (K), and value (V)
    Input: $x \in \mathbb{R}^{B \times T \times D}, D_h$ ▷ Obtained from DefineStructuring()
    Parameters: $W_q, W_k, W_v \in \mathbb{R}^{D \times D_h}, W_o \in \mathbb{R}^{D_h \times D}$ ▷ D_h is number of attention heads, W is weight matrices
    $Q, K, V \leftarrow xW_q, xW_k, xW_v$
    $H \leftarrow \text{Attention}(Q, K, V)$ ▷ Computing attention scores Q and K representations, obtained by linear transformations using W_q and W_k.
    $h \leftarrow HW_o$
    Return h
7: **end procedure**
8: **procedure** DEFINETRANSFORMERENCODERBLOCK(x, D_h, F)
    Input: $x \in \mathbb{R}^{B \times T \times D}, D_h, F$
    Parameters: $W_q, W_k, W_v \in \mathbb{R}^{D \times D_h}, W_o \in \mathbb{R}^{D_h \times D}, W_1, W_2 \in \mathbb{R}^{D \times F}, b_1, b_2 \in \mathbb{R}^F$ ▷ F is the feedforward dimension
    $h \leftarrow \text{MultiHeadSelfAttention}(x, D_h)$
    $h \leftarrow \text{LayerNorm}(h + x)$ ▷ Apply layer normalization to $h + x$
    $u \leftarrow W_1 h + b_1$ ▷ Perform a linear transformation on h and add bias b_1
    $u \leftarrow \text{ReLU}(u)$ ▷ Apply the ReLU activation function for non-linearity
    $v \leftarrow W_2 u + b_2$
    $z \leftarrow \text{LayerNorm}(v + h)$ ▷ Perform another linear transformation
9:     **Return** $z \in \mathbb{R}^{B \times T \times D}$
10: **end procedure**
11: **Dropout Layer**
12: **LSTM Layers:**   ▷ LSTM for sequential data processing
    $h_1 \leftarrow \text{LSTM}(z, lstm\_units)$
13: **Dense Layers:**   ▷ Dense layer for feature extraction
    $h_2 \leftarrow \text{Dense}(h_1, dense\_units, activation)$ ▷ Output layer for remaining useful life prediction
    $RUL \leftarrow \text{Dense}(h_2, 1, relu\_cut)$
14: **Return:** $RUL \in \mathbb{R}^{B \times 1}$ ▷ Final model output

---

Algorithm 3 Hyperparameter Optimization with Keras Tuner

1: **procedure** DEFINEHYPERMODELCLASS
   Class MyHyperModel(HyperModel):
   Inputs: ▷ Define hyperparameters and model architecture
   Define build: ▷ Build and compile models with varying hyperparameters (e.g.,
   CLDNN, TRANSFORMER–LSTM)
   Return model
2: **end procedure**
3: **procedure** DEFINECUSTOMBAYESIANOPTIMIZATIONTUNER
   Class MyBayesianOptimizationTuner(BayesianOptimization):
   Define initialization: ▷ Define custom Bayesian Optimization tuner
   Define on_error: ▷ Handle errors during the optimization for increased modularity
   and robustness
4: **end procedure**
5: **procedure** HYPERMODELINSTANCEANDTUNER
   Instantiate Hypermodel and Tuner:
   hypermodel ← MyHyperModel(...)
   tuner ← MyBayesianOptimizationTuner(hypermodel, objective=val_mae) ▷ Optimize
   for the lowest validation MAE
   tuner.search(max_trials=100, epochs=10, dataset_train, batch_size=512)
6: **end procedure**
7: **procedure** BESTHYPERPARAMETERSANDMODEL   ▷ Extract optimized
   hyperparameters and associated weights
   best_hp ← tuner.oracle.get_best_trials(1)
   best_model.set_weights(tuner.get_best_trial().get_weights())
8: **end procedure**

The hyperparameter search was conducted over a substantial span of 100 trials, encompassing 10 epochs for each trial on the training dataset. This comprehensive approach ensured a thorough exploration of the hyperparameter space, thereby identifying configurations that yield the most accurate predictions. Finally, the performance of the optimized models was rigorously evaluated on the validation dataset, ensuring the reliability and effectiveness of the hyperparameter tuning process in enhancing the predictive capabilities of the LIB RUL prediction models.

After completing the search, the optimal hyperparameter set was retrieved, and a new model was built using these parameters. This process ensured that the final CLDNN and TT model was fine-tuned for optimal performance, leveraging the power of Bayesian optimization to navigate the hyperparameter space efficiently.

4. Results and Discussion

This section presents and compares our findings and results to other algorithms, leveraging the validation data. During model development, it became evident that the most effective models necessitated two cardinal attributes. Firstly, they were required to be capable of managing sparse data by proficiently extracting significant features. Secondly, they were expected to possess the capability to learn both temporal and spatial relationships between the features, and the remaining cycles of each lithium-ion battery.

After these findings, a comparative analysis was undertaken among an assortment of hybrid models, which embraced these two critical characteristics. The performance of these models is visually presented in Figure 5a–c where key performance indicators such as loss, mean absolute error (MAE), mean absolute percentage error (MAPE) and mean squared error (MSE) are prominently depicted. It is noteworthy that each figure includes a comparison of all the hybrid neural network architectures developed, encompassing the baseline CLDNN, the optimized CLDNN, the transformer–LSTM, and its optimized counterpart.
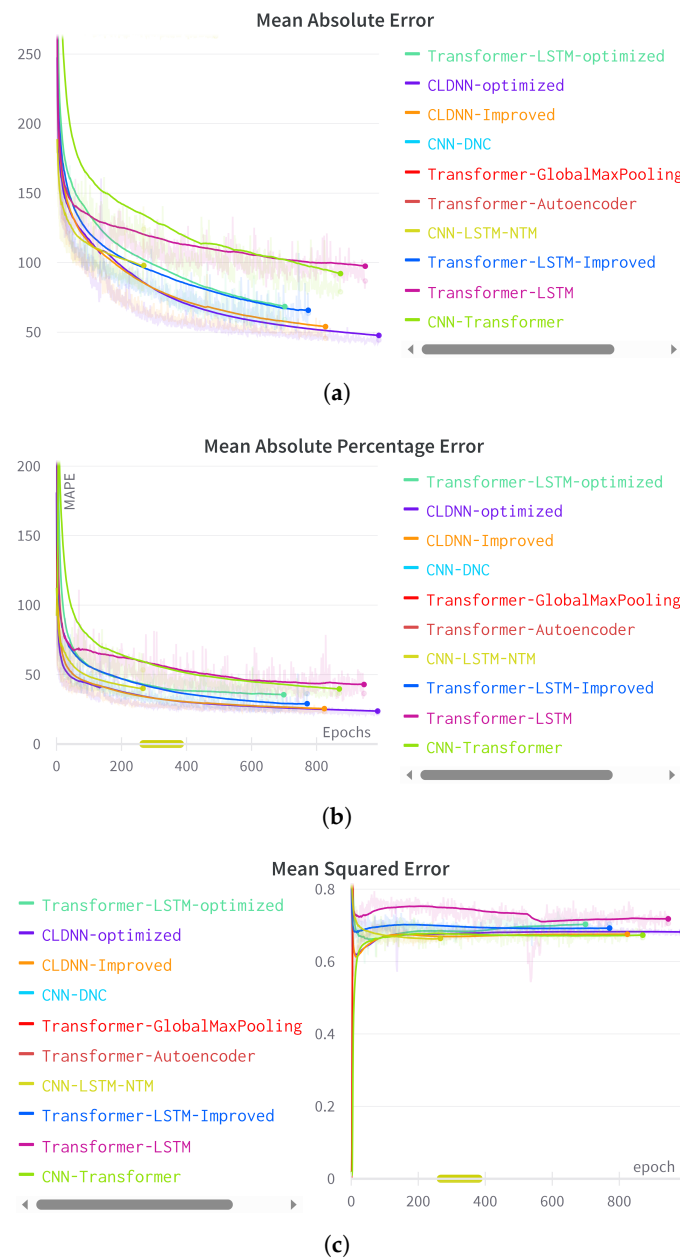
Figure 5. Comparison of error metrics between the top-performing models and other tested models, as detailed in Table 1. (**a**) MAE—mean absolute error for the best performing models. (**b**) MAPE—mean absolute percentage error for the best performing models. (**c**) MSE—mean squared error for the best performing models.
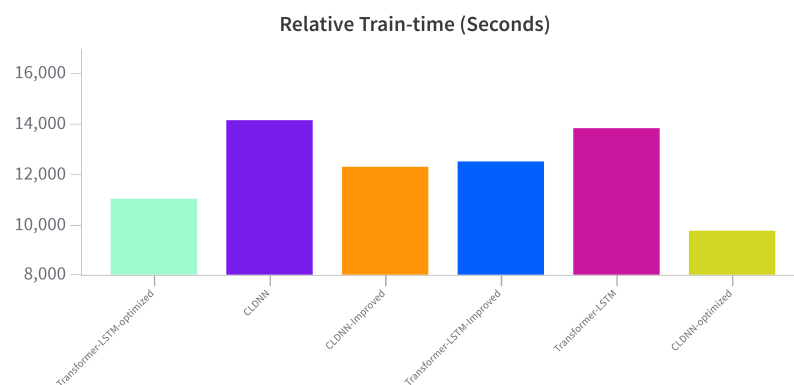
4.1. Comparing All the Tested Temporal Models

When placed in contrast with other models that embody the desired characteristics, such as convolution neural network–differential neural computer (CNN–DCN), CNN–LSTM–neural Turing machine (CNN–LSTM–NTM), CNN–transformers, and transformer–autoencoder, it becomes evident that these models exhibit higher error metrics, thereby underscoring their relative inefficacy in accurately predicting the RUL for lithium-ion batteries, as can be seen in Table 1.

Table 1. Results of six different ensemble NN models.

Model	MSE	MAE	MAPE	RMSE %
CLDNN	0.6754	84.012	25.676	0.8218
CNN–DCN	1.402	95.6365	46.408	1.1841
CNN–LSTM–NTM	1.333	284.887	-	1.1546
Transformer–LSTM	0.7136	85.134	28.7932	0.8444
CNN–transformers	0.6783	92.127	36.981	0.8236
Transformer–autoencoder	1.524	288.951	-	1.2345

4.2. Best Performing Models

In the course of assessing these results, two models, namely the convolutional, LSTM, densely connected (CLDNN) and the transformer–LSTM (temporal transformer), emerged as the most proficient in predicting the RUL, as outlined in Table 1 and Figure 5. The CLDNN model exhibited an MAE of 84.012, a MAPE of 25.676, and a MSE of 0.6754. Conversely, the temporal transformer model recorded a MAE of 85.134, a MAPE of 28.7932, and a MSE of 0.7136. To view the model's training duration, please refer to Figure 6.

Figure 6. Training times of the baseline and optimized models in seconds.

4.3. Observations

Several key observations were drawn from the hyperparameter setups (Tables 2 and 3) of the best-performing hybrid models.

- The optimized temporal transformer had fewer embedding dimensions and a lower number of attention heads compared to the original model. This meant that the original model was over-parametrised.
- Both transformer–LSTM and CLDNN models have 'optimized' versions with distinguishable hyperparameter configurations. For instance, the dense layer in the optimized models contains an increased number of units, with the optimized transformer–LSTM model having 64 units, compared to its original 40. Additionally, the optimized configurations have a reduced dropout rate and learning rate.
- With a learning rate of 0.001, the original transformer–LSTM model is ten times more robust than its optimized counterpart, which has a learning rate of 0.0001, preventing gradient explosion and overshooting the minimum in the optimized model.

Table 2. Hyperparameter comparison for CLDNN.

Parameter	Original	Optimized
Convolution filter	56	44
Convolution kernel	27	12
Dense layer Activation	tanh	tanh
Dense layer units	40	64
LSTM layer activation	tanh	tanh
LSTM layer units	132	108
Dropout rate–CNN	0.45	0.3
Dropout rate–LSTM	0.4	0.3
Output activation	relu	relu_cut
Learning rate	0.001	0.0001

Table 3. Hyperparameter comparison for transformer–LSTM.

Hyperparameter	Original	Optimized
Embedding Dimension (embed_dim)	64	32
Number of Attention Heads (num_heads)	8	2
Hidden layer size (ff_dim)	32	32
Dropout after attention	0.3	0.2
Dense layer activation	tanh	tanh
Dense layer units	40	64
LSTM layer activation	relu	tanh
LSTM layer units	132	108
Dropout rate–LSTM	0.4	0.3
Output activation	relu_cut	relu
Learning rate	0.001	0.0001

4.4. Comparing Convolutional–Long Short-Term Memory–Deep Neural Network (CLDNN) and Temporal Transformer (TT) Models to Existing Approaches

In comparing our results to other approaches in the literature, several standout models, namely autoencoder–DNN, auto–CNN–LSTM, and ATCMN, exhibit relevance and importance to our research. These models, as presented in Table 4, share a hybrid deep learning architecture similar to ours, combining different network architectures to enhance prediction. Notably, they distinguish themselves by predicting the remaining useful life (RUL) of batteries in terms of remaining cycle counts, aligning with our research objectives and providing a more actionable metric for battery health management compared to other models that predominantly focus on binary classification tasks (predicting whether the battery has remaining cycles or has surpassed the end of life threshold). While these models bear significance due to their hybrid architecture and cycle count prediction approach, our CLDNN and transformer–LSTM models demonstrate superior performance in RUL prediction for lithium-ion batteries (LIBs), especially when compared to other temporal models like LSTM–RNN, Deep-CNN, and Temporal-CNN [75]. Table 4 also highlights the average inference time during the validation of each of the approaches. The average inference time of the CLDNN and TT is also reasonable, making them efficient for real-time applications.

The ATCMN model, as described by Fei et al. [58], utilizes discharging time, voltage, and capacity for RUL estimation. In contrast, our models demonstrate superior performance in predicting remaining cycles for LIBs within a 100-cycle moving window, employing a CC–CV (constant current–constant voltage) charge policy. The CC–CV charging approach involves initially charging the battery at a constant current until a set voltage threshold is reached, followed by maintaining this voltage while the current gradually decreases as the battery becomes fully charged. This method is widely used due to its effectiveness in extending the battery's lifespan and optimizing its performance. Typical parameters of the CC–CV charge policy include the constant charge current, the voltage

threshold at which the transition to constant voltage occurs, and the termination current at which the charge cycle is considered complete. These parameters are not dynamic but are predetermined based on the battery's characteristics and the desired charging performance. The Auto–CNN–LSTM model by Ren et al. [25] operates on a distinct dataset with less variation and limited diversity in input parameters, which the authors acknowledge. This disparity in datasets, compared to our more extensive and diverse dataset, may lead to variations in outcomes between the two models. Our comprehensive dataset provides a robust foundation for modelling and prediction, ultimately enhancing the reliability and applicability of our findings, especially when considering real-time variables such as voltage, capacity, battery body temperature, and internal resistance within the 100-cycle moving window.

Table 4. Caparison of CLDNN and TT against other deep learning algorithms.

Model	MAE	RMSE	MAPE	Avg. Inference Time
Auto–CNN–LSTM [25]	-	5.03	-	-
ATCMN [58]	84	-	32.8%	18 ms
Deep CNN [23]	-	1.986	-	12.3 ms
Deep CNN transfer-learning [14]	-	1.361	-	133.9 s
CLDNN (ours)	84.012	0.8218	25.676 %	15.5 ms
Transformer–LSTM (ours)	85.134	0.8444	28.7923%	16.7 ms

5. Conclusions and Future Work

This study addresses the critical challenge of accurately estimating the RUL of LIBs within the context of electric vehicles. By leveraging deep learning techniques and utilizing a rich dataset from the Toyota Research Institute, we have developed and evaluated two hybrid models: CLDNN and TT. Our contributions encompass the creation of a pre-processed high-quality dataset through stratified random sampling, by implementation of a comprehensive data preprocessing pipeline, and the development of two hybrid models. This pipeline ensures feature consistency and captures temporal dynamics, thereby laying the foundation for precise RUL predictions.

Both the CLDNN and TT models exhibited commendable performance, surpassing existing approaches with mean absolute errors (MAEs) of 84.012 and 85.134, respectively. Furthermore, they demonstrated improvements in mean absolute percentage error (MAPE), which ranged from 4.01% to 7.12%. These models prove to be well-suited for LIB RUL prediction, making substantial contributions to battery recycling and sustainability within the electric vehicle industry. Additionally, these models demonstrated superior inference time, highlighting their efficiency and applicability in real-world scenarios where rapid decision-making is crucial. This enhancement in inference speed, combined with their predictive accuracy, positions the CLDNN and TT models as highly effective tools for advancing the reliability and efficiency of battery systems in the electric vehicle sector.

While the current achievements in battery health estimation models are commendable, there are several avenues for improvement. Expanding the validation of these models to a wider array of real-world datasets is essential to strengthening confidence in their practical applicability. Investigating advanced techniques such as complex data augmentation, and alternative ensemble methods, and exploring the potential of liquid neural networks (LNNs) could lead to improvements in model performance, offering more adaptable and robust solutions for battery health estimation. Additionally, the application of graph neural

networks (GNNs) in explaining the lithium-ion battery's (LIB) remaining useful life (RUL) has been suggested as a means to reduce the number of parameters while potentially outperforming traditional physics-based models. Incorporating diverse battery chemistries into future research is also critical. The current dataset is limited to a specific type of battery (lithium ferrous phosphate/Gr, 1.1 Ah, 3.3 V), which may not fully represent the performance across different battery types. By including a variety of battery chemistries, such as LFP batteries with a capacity of 3 Ah or nickel cobalt manganese (NCM) batteries with a capacity of 3 Ah and a nominal voltage of 3.8 V, models can be trained to generalize better across different battery systems. The benefit of including diverse battery chemistries lies in the ability to create more universal and robust predictive models that can adapt to various battery behaviours and degradation patterns, ultimately leading to improved safety, reliability, and efficiency in battery usage. Additionally, a more exhaustive comparison with existing RUL prediction methods, encompassing both traditional statistical models and contemporary machine learning approaches, could be conducted.

Despite these achievements, several areas for future improvement have been identified. Real-world implementation and validation on a broader dataset are crucial for bolstering confidence in the models' applicability. Exploring complex augmentation methods, alternative ensemble solutions, and liquid neural networks (LNNs) [76–78] could further refine model performance and introduce more efficient, adaptable, and robust approaches to battery health estimation. There is also potential for the use of explaining the LIB RUL using graph neural networks, which have also been shown to significantly reduce parameter count and perform better than their traditional physics-based model counterparts [79,80]. Future research may leverage LNNs or GNNs, known for their dynamic adaptability, to potentially enhance RUL prediction for LIBs. With reduced computational intensity, these networks may offer superior generalization and efficiency for large-scale applications like electric vehicle battery management systems.

Reducing parameter counts to enhance model efficiency, measuring processing times in online scenarios, and investigating the alignment between hyperparameter optimization and a comparison between physics-based models are promising avenues for future research. In conclusion, while this study represents a significant step forward in battery health estimation, ongoing research should focus on diversifying data sources, simplifying model complexities, and exploring emerging technologies, such as LNNs, to further advance this field.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial neural network
AE	Autoencoder
BMS	Battery management system
CLDNN	CNN–LSTM–deep neural networks
CNN	Convolution neural network
DDA	Data-driven approaches
DL	Deep learning
EOL	End of life
EM	Electrochemical model
GPR	Gaussian process regression
IR	Internal resistance
LSTM	Long short-term memory
LIB	Lithium-ion batteries
MAE	Mean absolute error
MAPE	Mean absolute percentage error
QD	Quantity of discharge
Qdlin	Linearly interpolated discharge capacity
DDM	Data-driven model
RUL	remaining useful life
RMSE	Root mean square error
RNN	Recurrent neural network
SOC	State of charge
SOH	State of health
SVM	Support vector machine
Tdlin	Linearly interpolated temperature
TT	Temporal transformer

References

1. Plett, G.L. *Battery Management Systems. Volume I, Battery Modeling*; Artech House: Boston, MA, USA, 2015.
2. Yang, B.; Wang, J.; Cao, P.; Zhu, T.; Shu, H.; Chen, J.; Zhang, J.; Zhu, J. Classification, summarization and perspectives on state-of-charge estimation of lithium-ion batteries used in electric vehicles: A critical comprehensive survey. *J. Energy Storage* **2021**, *39*, 102572. [CrossRef]
3. Dickinson, E.J.; Wain, A.J. The Butler-Volmer equation in electrochemical theory: Origins, value, and practical application. *J. Electroanal. Chem.* **2020**, *872*, 114145. [CrossRef]
4. Tian, Y.; Chen, C.; Xia, B.; Sun, W.; Xu, Z.; Zheng, W. An Adaptive Gain Nonlinear Observer for State of Charge Estimation of Lithium-Ion Batteries in Electric Vehicles. *Energies* **2014**, *7*, 5995–6012. [CrossRef]
5. Ouyang, M.; Liu, G.; Lu, L.; Li, J.; Han, X. Enhancing the estimation accuracy in low state-of-charge area: A novel onboard battery model through surface state of charge determination. *J. Power Sources* **2014**, *270*, 221–237. [CrossRef]
6. Wang, S.; Zhang, J.; Gharbi, O.; Vivier, V.; Gao, M.; Orazem, M.E. Electrochemical impedance spectroscopy. *Nat. Rev. Methods Prim.* **2021**, *1*, 41. [CrossRef]
7. Ramadesigan, V.; Boovaragavan, V.; Pirkle, J.C.; Subramanian, V.R. Efficient reformulation of solid-phase diffusion in physics-based lithium-ion battery models. *J. Electrochem. Soc.* **2010**, *157*, A854. [CrossRef]
8. 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]
9. Li, Y.; Liu, K.; Foley, A.M.; Zülke, A.; Berecibar, M.; Nanini-Maury, E.; Van Mierlo, J.; Hoster, H.E. Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review. *Renew. Sustain. Energy Rev.* **2019**, *113*, 109254. [CrossRef]
10. Anton, J.C.A.; Nieto, P.J.G.; Viejo, C.B.; Vilán, J.A.V. Support vector machines used to estimate the battery state of charge. *IEEE Trans. Power Electron.* **2013**, *28*, 5919–5926. [CrossRef]
11. Deng, Z.; Hu, X.; Lin, X.; Che, Y.; Xu, L.; Guo, W. Data-driven state of charge estimation for lithium-ion battery packs based on Gaussian process regression. *Energy* **2020**, *205*, 118000. [CrossRef]
12. Si, X.S.; Wang, W.; Hu, C.H.; Zhou, D.H. Remaining useful life estimation—A review on the statistical data driven approaches. *Eur. J. Oper. Res.* **2011**, *213*, 1–14. [CrossRef]
13. Liu, J.; Chen, Z. Remaining useful life prediction of lithium-ion batteries based on health indicator and Gaussian process regression model. *IEEE Access* **2019**, *7*, 39474–39484. [CrossRef]

14. Shen, S.; Sadoughi, M.; Li, M.; Wang, Z.; Hu, C. Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries. *Appl. Energy* **2020**, *260*, 114296. [CrossRef]

15. Alfarizi, M.G.; Tajiani, B.; Vatn, J.; Yin, S. Optimized random forest model for remaining useful life prediction of experimental bearings. *IEEE Trans. Ind. Inform.* **2022**, *19*, 7771. [CrossRef]

16. Hu, X.; Jiang, J.; Cao, D.; Egardt, B. Battery Health Prognosis for Electric Vehicles Using Sample Entropy and Sparse Bayesian Predictive Modeling. *IEEE Trans. Ind. Electron.* **2016**, *63*, 2645–2656. [CrossRef]

17. Kraus, M.; Feuerriegel, S. Forecasting remaining useful life: Interpretable deep learning approach via variational Bayesian inferences. *Decis. Support Syst.* **2019**, *125*, 113100. [CrossRef]

18. Mosallam, A.; Medjaher, K.; Zerhouni, N. Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. *J. Intell. Manuf.* **2016**, *27*, 1037–1048. [CrossRef]

19. Guan, Q.; Wei, X. The Statistical Data-driven Remaining Useful Life Prediction—A Review on the Wiener Process-based Method. In Proceedings of the 2023 Prognostics and Health Management Conference (PHM), Paris, France, 31 May–2 June 2023; pp. 64–68.

20. Lipu, M.S.H.; Hannan, M.A.; Hussain, A.; Saad, M.H.M.; Ayob, A.; Blaabjerg, F. State of Charge Estimation for Lithium-Ion Battery Using Recurrent NARX Neural Network Model Based Lighting Search Algorithm. *IEEE Access* **2018**, *6*, 28150–28161. [CrossRef]

21. Chemali, E.; Kollmeyer, P.J.; Preindl, M.; Ahmed, R.; Emadi, A. Long Short-Term Memory Networks for Accurate State-of-Charge Estimation of Li-ion Batteries. *IEEE Trans. Ind. Electron.* **2018**, *65*, 6730–6739. [CrossRef]

22. How, D.N.; Hannan, M.A.; Lipu, M.S.H.; Sahari, K.S.; Ker, P.J.; Muttaqi, K.M. State-of-charge estimation of li-ion battery in electric vehicles: A deep neural network approach. *IEEE Trans. Ind. Appl.* **2020**, *56*, 5565–5574. [CrossRef]

23. Shen, S.; Sadoughi, M.; Chen, X.; Hong, M.; Hu, C. A deep learning method for online capacity estimation of lithium-ion batteries. *J. Energy Storage* **2019**, *25*, 100817. [CrossRef]

24. Zhou, D.; Li, Z.; Zhu, J.; Zhang, H.; Hou, L. State of health monitoring and remaining useful life prediction of lithium-ion batteries based on temporal convolutional network. *IEEE Access* **2020**, *8*, 53307–53320. [CrossRef]

25. Ren, L.; Dong, J.; Wang, X.; Meng, Z.; Zhao, L.; Deen, M.J. A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life. *IEEE Trans. Ind. Inform.* **2020**, *17*, 3478–3487. [CrossRef]

26. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. *Neural Inf. Process. Syst.* **2014** , *27*. [CrossRef]

27. Yang, S.; Eisenach, C.; Madeka, D. MQRetNN: Multi-Horizon Time Series Forecasting with Retrieval Augmentation. *arXiv* **2022**, arXiv:2207.10517. [CrossRef].

28. Zhao, C.; Huang, X.; Li, Y.; Li, S. A Novel Cap-LSTM Model for Remaining Useful Life Prediction. *IEEE Sens. J.* **2021**, *21*, 23498–23509. [CrossRef]

29. Xiong, R.; Li, L.; Tian, J. Towards a smarter battery management system: A critical review on battery state of health monitoring methods. *J. Power Sources* **2018**, *405*, 18–29. [CrossRef]

30. 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]

31. Sainath, T.N.; Vinyals, O.; Senior, A.; Sak, H. Convolutional, long short-term memory, fully connected deep neural networks. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, 19–24 April 2015; pp. 4580–4584.

32. Chadha, G.S.; Shah, S.R.B.; Schwung, A.; Ding, S.X. Shared temporal attention transformer for remaining useful lifetime estimation. *IEEE Access* **2022**, *10*, 74244–74258. [CrossRef]

33. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

34. Lim, B.; Arık, S.Ö.; Loeff, N.; Pfister, T. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [CrossRef]

35. Lyu, P.; Liu, X.; Qu, J.; Zhao, J.; Huo, Y.; Qu, Z.; Rao, Z. Recent advances of thermal safety of lithium ion battery for energy storage. *Energy Storage Mater.* **2020**, *31*, 195–220. [CrossRef]

36. Tremblay, O.; Dessaint, L.A.; Dekkiche, A. A Generic Battery Model for the Dynamic Simulation of Hybrid Electric Vehicles. In Proceedings of the 2007 IEEE Vehicle Power and Propulsion Conference, Arlington, TX, USA, 9–12 September 2007; pp. 284–289.

37. Zhao, X.; Wang, Y.; Sahinoglu, Z.; Wada, T.; Hara, S.; Callafon, R. State-of-charge estimation for batteries: A multi-model approach. In Proceedings of the American Control Conference, Portland, OR, USA, 4–6 June 2014; pp. 2779–2785. [CrossRef]

38. Song, W.; Wu, D.; Shen, W.; Boulet, B. A Remaining Useful Life Prediction Method for Lithium-ion Battery Based on Temporal Transformer Network. *Procedia Comput. Sci.* **2023**, *217*, 1830. [CrossRef]

39. Yang, Y. A machine-learning prediction method of lithium-ion battery life based on charge process for different applications. *Appl. Energy* **2021**, *292*, 116897. [CrossRef]

40. Yang, D.; Wang, Y.; Pan, R.; Chen, R.; Chen, Z. A neural network based state-of-health estimation of lithium-ion battery in electric vehicles. *Energy Procedia* **2017**, *105*, 2059–2064. [CrossRef]

41. Lotfi, N.; Li, J.; Landers, R.G.; Park, J. Li-ion Battery State of Health Estimation based on an improved Single Particle model. In Proceedings of the 2017 American Control Conference (ACC), Seattle, WA, USA, 24–26 May 2017; pp. 86–91. [CrossRef]

42. El-Dalahmeh, M.; Al-Greer, M.; El-Dalahmeh, M.; Bashir, I. Physics-based model informed smooth particle filter for remaining useful life prediction of lithium-ion battery. *Measurement* **2023**, *214*, 112838. [CrossRef]

43. Xue, Z.; Zhang, Y.; Cheng, C.; Ma, G. Remaining useful life prediction of lithium-ion batteries with adaptive unscented kalman filter and optimized support vector regression. *Neurocomputing* **2020**, *376*, 95–102. [CrossRef]

44. Tong, Z.; Miao, J.; Tong, S.; Lu, Y. Early prediction of remaining useful life for Lithium-ion batteries based on a hybrid machine learning method. *J. Clean. Prod.* **2021**, *317*, 128265. [CrossRef]

45. Kwon, S.J.; Han, D.; Choi, J.H.; Lim, J.H.; Lee, S.E.; Kim, J. Remaining-useful-life prediction via multiple linear regression and recurrent neural network reflecting degradation information of 20 Ah $LiNi_xMnyCo_{1-x-y}O_2$ pouch cell. *J. Electroanal. Chem.* **2020**, *858*, 113729. [CrossRef]

46. Bashir, I.; Al-Greer, M.; et al. Lithium-ion Batteries Capacity Degradation Trajectory Prediction Based on Decomposition Techniques and NARX Algorithm. In Proceedings of the 2022 57th International Universities Power Engineering Conference (UPEC), Istanbul, Turkey, 30 August–2 September 2022; pp. 1–6.

47. Zheng, S.; Ristovski, K.; Farahat, A.; Gupta, C. Long Short-Term Memory Network for Remaining Useful Life estimation. In Proceedings of the 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), Dallas, TX, USA, 19–21 June 2017; pp. 88–95. [CrossRef]

48. Wang, J.; Wen, G.; Yang, S.; Liu, Y. Remaining Useful Life Estimation in Prognostics Using Deep Bidirectional LSTM Neural Network. In Proceedings of the 2018 Prognostics and System Health Management Conference (PHM-Chongqing), Chongqing, China, 26–28 October 2018; pp. 1037–1042. [CrossRef]

49. Spagnol, P.; Rossi, S.; Savaresi, S.M. Kalman filter SoC estimation for Li-ion bateries. In Proceedings of the 2011 IEEE International Conference on Control Applications (CCA), Denver, CO, USA, 28–30 September 2011; pp. 587–592.

50. Hu, C.; Youn, B.D.; Chung, J. A multiscale framework with extended Kalman filter for lithium-ion battery SOC and capacity estimation. *Appl. Energy* **2012**, *92*, 694–704. [CrossRef]

51. Dai, H.; Wei, X.; Sun, Z.; Wang, J.; Gu, W. Online cell SOC estimation of Li-ion battery packs using a dual time-scale Kalman filtering for EV applications. *Appl. Energy* **2012**, *95*, 227–237. [CrossRef]

52. Lambert, B. *A Student's Guide to Bayesian Statistics*; SAGE: New York, NY, USA, 2018.

53. Li, X.; Ding, Q.; Sun, J.Q. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliab. Eng. Syst. Saf.* **2018**, *172*, 1–11. [CrossRef]

54. Sateesh Babu, G.; Zhao, P.; Li, X.L. Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life. In *Database Systems for Advanced Applications*; Navathe, S.B., Wu, W., Shekhar, S., Du, X., Wang, X.S., Xiong, H., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 214–228.

55. Li, J.; Li, X.; He, D. A Directed Acyclic Graph Network Combined With CNN and LSTM for Remaining Useful Life Prediction. *IEEE Access* **2019**, *7*, 75464–75475. [CrossRef]

56. Tan, W.M.; Teo, T.H. Remaining Useful Life Prediction Using Temporal Convolution with Attention. *AI* **2021**, *2*, 48–70. [CrossRef]

57. Fan, Y.; Xiao, F.; Li, C.; Yang, G.; Tang, X. A novel deep learning framework for state of health estimation of lithium-ion battery. *J. Energy Storage* **2020**, *32*, 101741. [CrossRef]

58. Fei, Z.; Zhang, Z.; Yang, F.; Tsui, K.L. A deep attention-assisted and memory-augmented temporal convolutional network based model for rapid lithium-ion battery remaining useful life predictions with limited data. *J. Energy Storage* **2023**, *62*, 106903. [CrossRef]

59. Rastegarpanah, A.; Contreras, C.A.; Stolkin, R. Hyperparameter-optimized CNN and CNN-LSTM for Predicting the Remaining Useful Life of Lithium-Ion Batteries. In Proceedings of the 2023 Eleventh International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, Egypt, 21–23 November 2023; pp. 110–115. [CrossRef]

60. Rastegarpanah, A.; Contreras, C.A.; Stolkin, R. Harnessing CNN-DNC and CNN-LSTM-DNC Architectures for Enhanced Lithium-Ion Remaining Useful Life Prediction. In Proceedings of the 2023 Eleventh International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, Egypt, 21–23 November 2023; pp. 116–121. [CrossRef]

61. Li, P.; Zhang, Z.; Grosu, R.; Deng, Z.; Hou, J.; Rong, Y.; Wu, R. An end-to-end neural network framework for state-of-health estimation and remaining useful life prediction of electric vehicle lithium batteries. *Renew. Sustain. Energy Rev.* **2022**, *156*, 111843. [CrossRef]

62. LeCun, Y.; Bengio, Y.; et al. Convolutional networks for images, speech, and time series. In *The Handbook of Brain Theory and Neural Networks*; MIT Press: Cambridge, MA, USA, 1995; Volume 3361, p. 1995.

63. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2017; Voume 30.

64. Zhai, J.; Zhang, S.; Chen, J.; He, Q. Autoencoder and its various variants. In Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 7–10 October 2018; pp. 415–419.

65. Collier, M.; Beel, J. Implementing neural turing machines. In Proceedings of the Artificial Neural Networks and Machine Learning—ICANN 2018: 27th International Conference on Artificial Neural Networks, 2018. Available online: https://www.nature.com/articles/nature20101 (accessed on 10 January 2023).

66. Graves, A.; Wayne, G.; Reynolds, M.; Harley, T.; Danihelka, I.; Grabska-Barwińska, A.; Colmenarejo, S.G.; Grefenstette, E.; Ramalho, T.; Agapiou, J.; et al. Hybrid computing using a neural network with dynamic external memory. *Nature* **2016**, *538*, 471–476. [CrossRef]

67. 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. Electrif.* **2023**, *9*, 2628–2644. [CrossRef]

68. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.R.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N.; et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97. [CrossRef]

69. Ma, Q.; Zhang, M.; Xu, Y.; Song, J.; Zhang, T. Remaining Useful Life Estimation for Turbofan Engine with Transformer-based Deep Architecture. In Proceedings of the 2021 26th International Conference on Automation and Computing (ICAC), Portsmouth, UK, 2–4 September 2021; pp. 1–6. [CrossRef]

70. Blumer, A.; Ehrenfeucht, A.; Haussler, D.; Warmuth, M.K. Occam's razor. *Inf. Process. Lett.* **1987**, *24*, 377–380. [CrossRef]

71. Probst, P.; Boulesteix, A.L.; Bischl, B. Tunability: Importance of hyperparameters of machine learning algorithms. *J. Mach. Learn. Res.* **2019**, *20*, 1934–1965.

72. Malakouti, S.M.; Menhaj, M.B.; Suratgar, A.A. The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction. *Clean. Eng. Technol.* **2023**, *15*, 100664. [CrossRef]

73. Wu, J.; Chen, X.Y.; Zhang, H.; Xiong, L.D.; Lei, H.; Deng, S.H. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.* **2019**, *17*, 26–40.

74. Gulli, A.; Pal, S. *Deep Learning with Keras*; Packt Publishing Ltd.: Birmingham, UK, 2017.

75. Wang, S.; Jin, S.; Bai, D.; Fan, Y.; Shi, H.; Fernandez, C. A critical review of improved deep learning methods for the remaining useful life prediction of lithium-ion batteries. *Energy Rep.* **2021**, *7*, 5562–5574. [CrossRef]

76. Hasani, R.; Lechner, M.; Amini, A.; Rus, D.; Grosu, R. Liquid time-constant networks. In Proceedings of the AAAI Conference on Artificial Intelligence, 2021; Volume 35, pp. 7657–7666. Available online: https://ojs.aaai.org/index.php/AAAI/article/view/16936 (accessed on 10 January 2023).

77. Hasani, R. Interpretable Recurrent Neural Networks in Continuous-Time Control Environments. Ph.D. Thesis, Technische Universität Wien, Vienna, Austria, 2020.

78. Hashemi, S.R.; Bahadoran Baghbadorani, A.; Esmaeeli, R.; Mahajan, A.; Farhad, S. Machine learning-based model for lithium-ion batteries in BMS of electric/hybrid electric aircraft. *Int. J. Energy Res.* **2021**, *45*, 5747–5765. [CrossRef]

79. Wang, Z.; Yang, F.; Xu, Q.; Wang, Y.; Yan, H.; Xie, M. Capacity estimation of lithium-ion batteries based on data aggregation and feature fusion via graph neural network. *Appl. Energy* **2023**, *336*, 120808. [CrossRef]

80. Lam, R.; Sanchez-Gonzalez, A.; Willson, M.; Wirnsberger, P.; Fortunato, M.; Alet, F.; Ravuri, S.; Ewalds, T.; Eaton-Rosen, Z.; Hu, W.; et al. Learning skillful medium-range global weather forecasting. *Science* **2023**, *382*, 1416–1421. [CrossRef]