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

Alireza Rastegarparnah *, Mohammed Eesa Asif , Rustam Stolkin

Posted Date: 25 January 2024

doi: 10.20944/preprints202401.1817.v1

Keywords: Deep Learning; Remaining Useful Life; Lithium-Ion Batteries; Battery Management Systems; Recycling and Reuse; Battery Degradation

Article

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

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

[1] Department of Metallurgy & Materials Science, University of Birmingham, Birmingham, B15 2TT, UK
[2] The Faraday Institution, Quad One, Harwell Science and Innovation Campus, Didcot, OX11 0RA, UK
* Correspondence: a.rastegarpanah@bham.ac.uk
† Mohammed Eesa Asif and Alireza Rastegarpanah should be considered joint first author.

Abstract: With the proliferation of electric vehicles (EVs) and the consequential increase in EV battery circulation, the need for accurate assessment of battery health and remaining useful life (RUL) is paramount, driven by environmentally friendly and sustainable options. 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-LSTM-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 notable performance with an MAE of 85.134 and a MAPE of 28.7932. These impressive results were achieved through the application of Bayesian hyperparameter optimization, further enhancing the accuracy of predictive methods. These models were benchmarked against existing approaches, demonstrating superior results with an improvement in MAPE ranging from 4.01% to 7.12%.

Keywords: deep learning; remaining useful life; lithium-ion batteries; battery management systems; recycling and reuse; 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 (PBM) and Data-driven Models (DDM). 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 compared to PBMs.

DDMs can be further categorised into statistical machine learning and deep learning methods, which provide valuable insights into their respective strengths and applications. Shallow learning, a subset of statistical machine learning, employs neural networks with a single layer, as exemplified by Support Vector Machines (SVMs). On the other hand, deep learning methods utilize neural networks with multiple hidden layers.

 doi:10.20944/preprints202401.1817.v1

Statistical learning methods, including SVMs [10], Gaussian Process Regression (GPR) [11–13], Random Forest [14,15], and Bayesian approaches [16–18], are well-suited for modeling small datasets with prior knowledge of the generative process. However, they may face challenges in capturing complex battery characteristics and long-term dependencies in the data [2,19].

In contrast, deep learning methods, represented by Recurrent Neural Networks (RNNs) [20,21], Convolutional Neural Networks (CNNs) [22,23], and hybrid models [24,25], excel in handling extensive datasets with limited knowledge about the underlying process or suitable features. They demonstrate remarkable capabilities in capturing intricate patterns from raw battery data, particularly in dealing with multi-variate time-series information.

Recent advancements in sequence-to-sequence learning in the domain of 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 cross-entity attention mechanism in MTFS in [27]. These method minimizes assumptions on sequence structures and proves effective, particularly in tasks where large labeled training sets are available.

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 [28]. Hence, despite the progress in RUL prediction, current research has several knowledge gaps and limitations. These include the 1) need for more robust and accurate models, enhanced generalizability, 2) exploration of various deep learning architectures, 3) utilisation of complete datasets with varying sampling rates, and consideration of factors like battery ageing and non-stationary signals.

Considering the limitations of previous research and methods, our study takes a comprehensive approach to address data-driven methods for RUL prediction of LIB. We recognise that the key to successful prediction lies in harnessing the potential of a diverse dataset. Our dataset comprises 124 commercial lithium-ion batteries [29], each subjected to extensive cycling until failure under specific operational conditions (an overview can be seen in Figure 1). It encompasses crucial parameters such as discharge capacity, temperature, internal resistance, and discharge time.

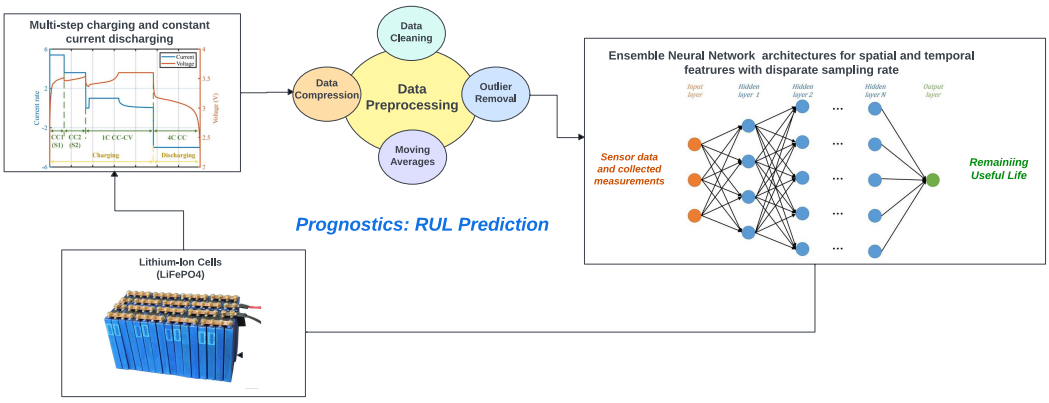


Figure 1. Overview of the project

To effectively address the heterogeneity within our dataset, we adopted a hybrid approach known as CLDNN proposed by Sainath et. al. [30], 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. [31] 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 [32]. While

 doi:10.20944/preprints202401.1817.v1

the Temporal Transformer shares a name with the Temporal Fusion Transformer (TFT) introduced by Lim et al. [33], 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.

It is worth emphasising that our experimentation phase encompassed the exploration of multiple temporal hybrid models, however, we only detail the most effective models. 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 concludes this article with our conclusion and future work.

2. Related Works

Physics-based models face challenges in predicting lithium-ion batteries RUL [2]. Performing well only in narrow domains, they require a detailed understanding of battery electrochemical processes, which is time-consuming and computationally intensive [3,8,34]. Simplified assumptions limit their ability to capture real-world complexities, resulting in lower RUL prediction accuracy. Parameterisation is challenging due to manufacturing variations and difficulty in obtaining precise model parameters [35,36]. The models exhibit limited adaptability to dynamic environments and high computational complexity, hampering real-time applicability. Calibration and validation procedures demand extensive data and resources [37]. Sensitivity to uncertainties affects accuracy, as minor errors propagate in measurements and assumptions. Consequently, data-driven methods like machine learning have gained prominence for their ability to learn complex nonlinear relationships from data, overcoming uncertainties, providing a promising alternative for RUL prediction in lithium-ion batteries [38–41].

2.1. Statistical Machine Learning

As previously noted, statistical machine learning methods, including Support Vector Machines (SVM) [10], Support vector regression with Kalman filters [42], Gaussian Process Regression (GPR) [11–13], Random Forest [14,15], and Bayesian approaches [16–18], have been employed for predicting battery remaining useful life (RUL). Tong et al. [43] introduced ADLSTM-MC, a deep-learning-based algorithm combining adaptive dropout long short-term memory (ADLSTM) and Monte Carlo (MC) simulation, achieving high precision with only 25% of degradation data. Similarly, Kwon et al. [44] 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. CNNs are used to extract salient features from sparse time series data. Lipu et al. [20] introduced, a Nonlinear Auto-regressive with Exogenous Input (NARX)-based Neural Network (NARXNN) is introduced 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 [45] to predict capacity trajectories. In [46] Zheng et al. employed a deep-LSTM network followed by multiple fully connected (FC) layers. Whereas Wang et al. in [47] introduced a bidirectional-LSTM architecture with additional FC layers. Chemali et al. [21] proposes an RNN architecture with LSTM for RUL estimation, achieving low Mean Absolute Error (MAE) values and demonstrating generalisation across different conditions. [26,27]

 doi:10.20944/preprints202401.1817.v1

2.3. Convolutional Models

Shen et al. [22] presented a Deep Convolutional Neural Network (DCNN) that achieved higher accuracy than traditional methods, such as filter-based models [48–51]. However, limitations include the fixed-size input matrix and a limited understanding of the method's interpretability. Following a similar approach, Li et al. [52] proposed another DCNN employing a time window approach for sample preparation and feature extraction. In this work [53], 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 is then integrated with ensemble learning to form DCNN-ETL. The effectiveness of the DCNN-ETL model is then bench-marked 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 limitations include applicability to other battery types and real-time implementation challenges [24].

2.4. Hybrid Models

An Auto-CNN-LSTM model for RUL estimation is proposed by Ren et al. [25], incorporating an auto-encoder to enhance feature vectors by dimensionality reduction and feature learning, achieving accurate predictions. Knowledge gaps include the need for robust generalizability, exploration of various temporal networks, and better utilisation of datasets with varying feature sampling rates. Additionally, consideration should be given to ageing effects, non-stationary signals, relevant feature incorporation, and spatial information capture. The work by Li et al. [54] introduced a directed acyclic graph network which coupled the LSTM with the CNN. In [55], Tan et al. introduced a Multi-variate Time-series focused approach for a light-weight CNN with attention mechanisms. Hybrid models like the one in [56] combine GRU and CNN for SoH estimation, while [57] uses an Attention-Assisted Temporal Convolutional Memory-Augmented Network (ATCMN) for RUL prediction from limited data. These approaches demonstrate promise but need further exploration and rigorous testing on more diverse datasets.

In the following section, we aim to address critical knowledge gaps identified in the literature regarding predicting the RUL of LIBs. These gaps encompass issues of robustness, accuracy, and generalizability in data-driven models for Battery Management Systems (BMS). To overcome these limitations, we developed end-to-end hybrid models capable of utilizing complete datasets [58], 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) [59], Long Short term Memory cells (LSTMs) [32], the Transformer network [60], Autoencoder [61], Neural Turing machines [62], Differentiable neural computer [63] 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 optimisation of model hyperparameters using Bayesian optimization techniques aimed at maximising the models' predictive accuracy. The primary objective was to develop a robust predictive model for estimating the RULs of LIB, accounting for intricate temporal dynamics and feature variations across different battery batches.

 doi:10.20944/preprints202401.1817.v1

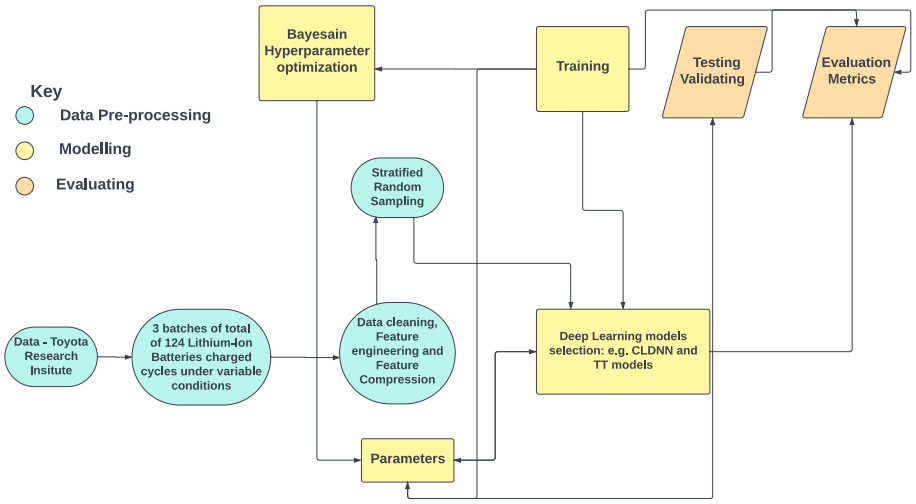


Figure 2. The study pipeline begins with a dataset comprising 124 commercial LIBs that have undergone extensive cycling until failure. Next, we implement feature selection to extract crucial features that enhance the accuracy of RUL predictions. Following this, a series of data preprocessing techniques 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

This project employed a dataset comprising 124 lithium-ion batteries of the lithium iron phosphate (LFP)/graphite type, each with a nominal capacity of 1.1 Ah and a nominal voltage of 3.3 V. These batteries underwent cycling until failure under fast-charging conditions within a convection temperature chamber set to 30°C. The dataset, sourced from the Toyota Research Institution [64], included information on various parameters, such as voltage, capacity, and current, continuously measured during cycling, spanning from a single cycle to the End Of Life. Charging involved an initial phase with current C1 until a state-of-charge (SOC) of S1 was reached, followed by charging with current C2 until reaching SOC S2, consistently set at 80% for all cells. Subsequently, the cells underwent an 80% to 100% SOC transition using a 1 C-rate constant current-constant voltage (CC-CV) charging approach, up to a 3.6 V cut-off voltage [65]. The batteries' lifetimes were determined based on the cycle number at which their capacity declined to 80% of the original value, with observed lifetimes ranging from 1350 to 2300 cycles.

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.

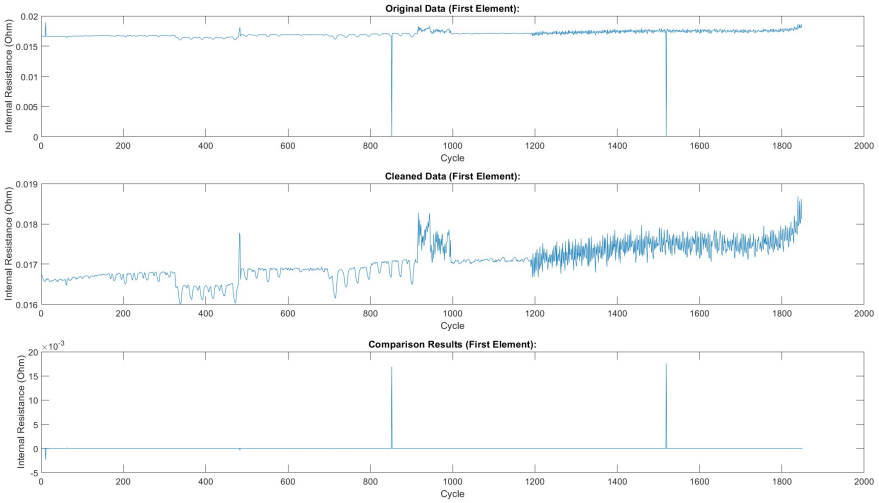
3.3. Feature Selection

The dataset, divided into three batches of approximately 48 cells each, was prepared for analysis. Essential features for RUL prediction were identified, including linearly interpolated Discharge Capacity (Qdlin), Linearly Interpolated Temperature (Tdlin), Internal Resistance (IR), discharge data, discharge time, and remaining cycles. Qdlin and Tdlin were interpolated to maintain a consistent sampling rate for all cells.

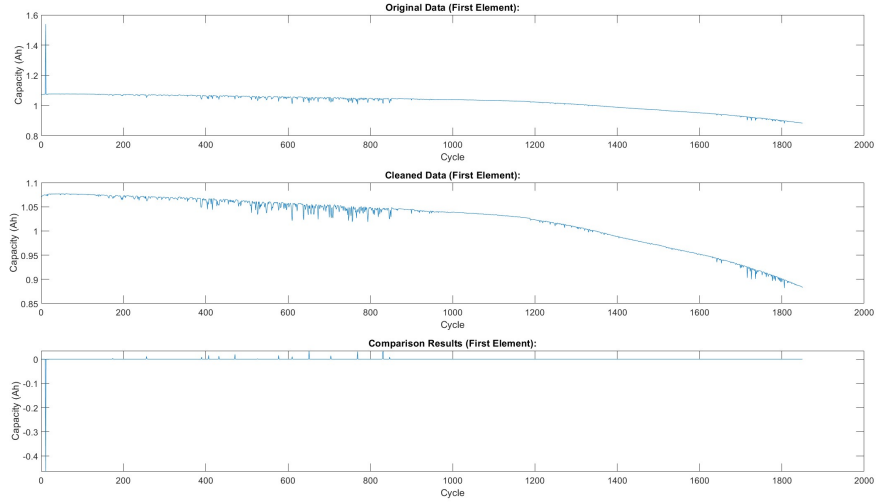
 doi:10.20944/preprints202401.1817.v1

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–c, 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, PCA was employed to reduce the dimensionality of Qdlin and Tdlin while preserving their essential features.



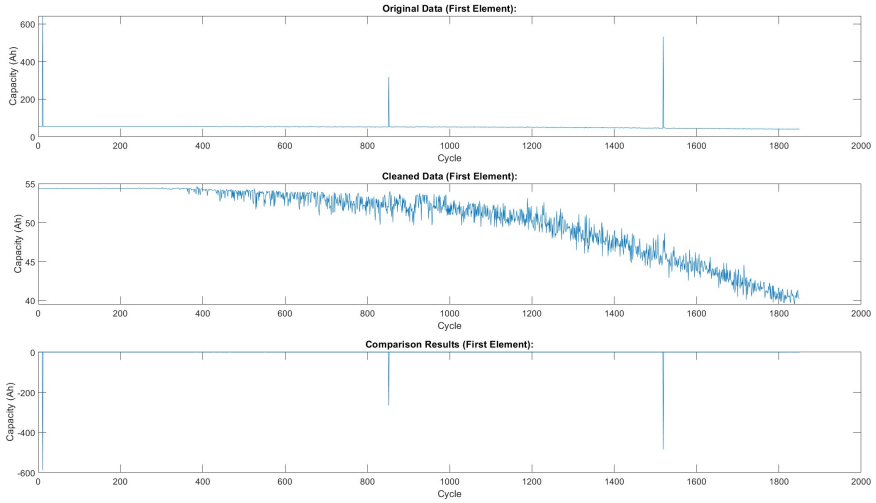
(**a**)Internal resistance



(**b**)Discharge time

Figure 3. *Cont.*

 doi:10.20944/preprints202401.1817.v1



(**c**)Discharge Quantity

Figure 3. Comparison of Data Features Before and After Preprocessing.

3.5. RUL estimation

Traditional supervised learning approaches rely on the existence of 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 [66]. 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 Mean Absolute Error (MAE) shown in Equation (3) and 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.

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

 doi:10.20944/preprints202401.1817.v1

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

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

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

3.7. Proposed Architectures

The most successful architectures in our study were the Convolutional Long Short-Term Memory Deep Neural Network (CLDNN), originally proposed by Sainath et al. [30] for Natural Language Processing specifically used for Large Vocabulary Continuous Speech Recognition (LVCSR) tasks [30]. In their work, CLDNN outperformed Gaussian Mixture Model (GMM) and Hidden Markov Model (HMM) systems [67]. We repurposed the CLDNN architecture for the specific task of Remaining Useful Life (RUL) estimation.

Another noteworthy model in our investigation was the Temporal-Transformer (TT), initially introduced by Chadha et al.[31] 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. [68] 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.

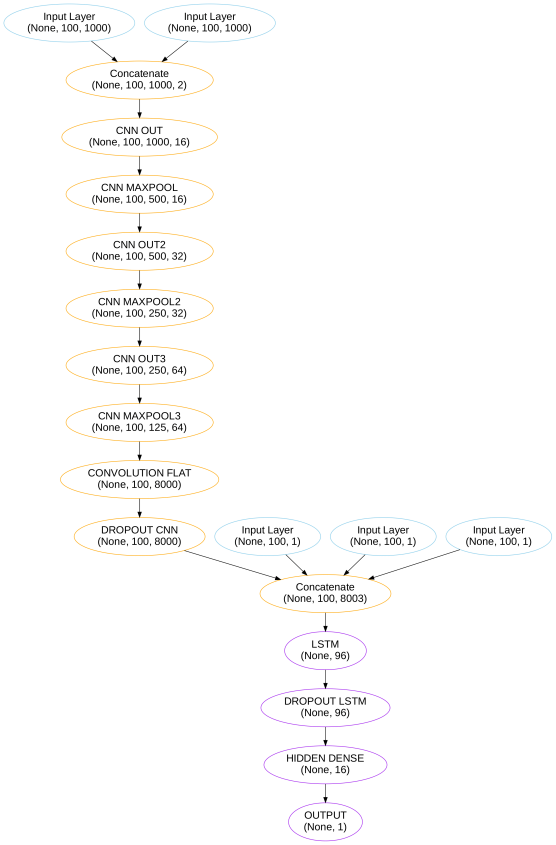
3.7.1. *CLDNN*

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.

- **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 was reshaped into a format suitable for time series analysis.
- **Sequence Length and Padding:** LIB data has 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 is inherently temporal, reflecting the degradation of the battery over time. The CLDNN architecture incorporates mechanisms to capture temporal dependencies effectively. Long-Short-Term Memory (LSTM) layers allow 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 is fine-tuned with different optimizer choices to minimize loss functions.

 doi:10.20944/preprints202401.1817.v1

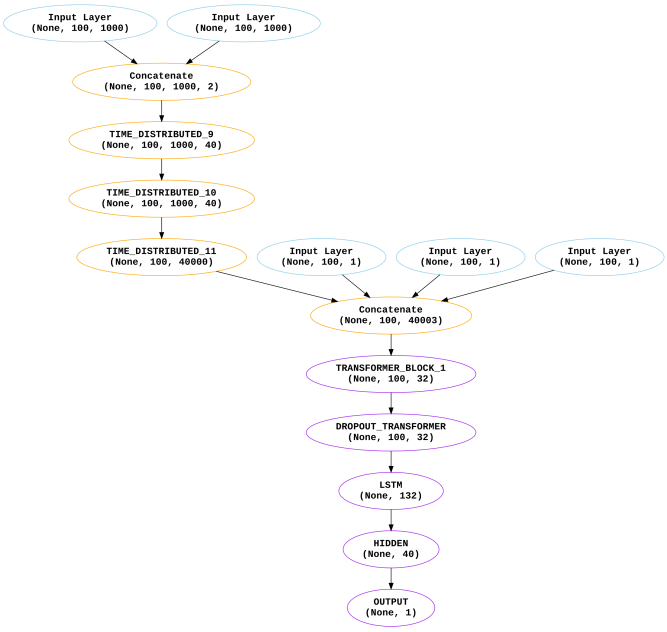
The final optimized CLDNN architecture excelled at predicting the remaining useful cycles (RUL) for lithium-ion batteries. It comprises a total of `1,518,665` trainable parameters and integrates Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Dense Neural Networks (DNN) to leverage the unique strengths of each component. The CNN layers capture intricate spatial patterns within features, while the LSTM layers facilitate the capture of temporal dependencies. Dropout layers and dense layers 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.



(a) The CLDNN model.

Figure 4. *Cont*.

 doi:10.20944/preprints202401.1817.v1



(**b**)The Temporal Transformer model.

Figure 4. Comparison of the CLDNN and Temporal Transformer architecture

Algorithm 1 CLDNN

procedure DEFINESPATIALLAYER($Q_{dlin}, T_{dlin}, IR, DT, QD, hp$)
Input: $Q_{dlin}, T_{dlin} \in \mathbb{R}^{B\times T\times D}$, IR, DT, QD and Hyperparameters hp ▷ $Qdlin, Tdlin$: Input data tensors, hp: Hyperparameters
Concatenate: (Q_{dlin} and T_{dlin}) ▷ Merge input features
Apply Convolutional Layers and Dropout:
$h_1 \leftarrow \text{Conv1D}([Q_{dlin}, T_{dlin}], n_f, k_s, s, a, \text{padding} = \texttt{same})$
$h_2 \leftarrow \text{MaxPooling1D}(h_1)$
$h_3 \leftarrow \text{Conv1D}(h_2, 2n_f, k_s, s, a, \text{padding} = \texttt{same})$
$h_4 \leftarrow \text{MaxPooling1D}(h_3)$
$h_5 \leftarrow \text{Conv1D}(h_4, 4n_f, k_s, s, a, \text{padding} = \texttt{same})$
$h_6 \leftarrow \text{MaxPooling1D}(h_5)$
$h_7 \leftarrow \text{Flatten}(h_6)$
$h_8 \leftarrow \text{Dropout}(h_7, d_r)$ ▷ Output of the CNN is a set of high-level features, h_8
Features Concatenation: ▷ Concatenation of detailed features with additional inputs
$h_9 \leftarrow \text{Concatenate}(h_8, IR, DT, QD)$ ▷ Feature transformation and regularization
Return h_9 ▷ Prepare for the LSTM layers
end procedure
procedure DEFINELSTMLAYER(h_9, n_u, a_u)
Input: h_9, n_u, a_u
$h_{10} \leftarrow \text{LSTM}(h_9, n_u, a_u)$ ▷ Producing a set of hidden states, h_{10}, capturing the temporal evolution of the system
Return h_{10} ▷ Prepare for the Dense layers
end procedure
procedure DEFINEDENSELAYERS(h_{10}, n_d, a_d) ▷ Learn complex non-linear relationships between the hidden states and the RU
$h_{11} \leftarrow \text{Dense}(h_{10}, n_d, a_d)$
$RUL \leftarrow \text{Dense}(h_{11}, 1, \texttt{relu\_cut})$
Return: $RUL \in \mathbb{R}^{B\times 1}$ ▷ Final model output
end procedure

 doi:10.20944/preprints202401.1817.v1

3.7.2. *Temporal-Transformer*

Adapting the Temporal-Transformer (TT) model, originally designed for the Remaining Useful Life (RUL) estimation of aircraft engines, for the estimation of RUL in Lithium-Ion Batteries (LIB), involves several architectural differences and adjustments. Here are some modifications that might be considered.

- **Input Representation:** Adjustment to the input representation to accommodate the characteristics of LIB data. 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 were used in the original model by Ma et al. [68] needed adjustments for LIB data. Tailoring attention mechanisms 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-tune hyperparameters using Bayesian optimization specific to LIB data. This includes learning rates, the number of attention heads, embedding dimensions, layer sizes, and dropout rates.

The temporal Transformer (TT) architecture, combines two potent neural network paradigms: the Transformer and Long Short-Term Memory (LSTM). The Transformer's multi-head self-attention mechanism empowers 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. Meanwhile, LSTM units adeptly capture long-term relationships between features, enhancing the model's predictive capabilities. The architecture employs feed-forward neural networks (FFN) for further refinement, facilitating the modelling of non-linear data relationships. Regularisation techniques namely dropout and layer normalisation are employed to ensure training stability and generalisation. 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.

 doi:10.20944/preprints202401.1817.v1

Algorithm 2 Temporal-Transformer

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
$h \leftarrow$ TimeDistributedDense($[Q_{dlin}, T_{dlin}], hp$)
Flatten h ▷ Prepare for Transformer input
Return $[Q_{dlin}, T_{dlin}]$ ▷ Linearly interpolated features will have the same dimension as the other features
end procedure
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
    $h \leftarrow$ 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 the each data point
end procedure
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
$Q, K, V \leftarrow xW_q, xW_k, xW_v$
$H \leftarrow$ 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
end procedure
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
$h \leftarrow$ MultiHeadSelfAttention(x, D_h)
$h \leftarrow$ 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
$u \leftarrow$ ReLU(u) ▷ Apply the ReLU activation function for non-linearity
$v \leftarrow W_2 u + b_2$
$z \leftarrow$ LayerNorm$(v + h)$ ▷ Perform another linear transformation
    Return $z \in \mathbb{R}^{B\times T\times D}$
end procedure
Dropout Layer
LSTM Layers: ▷ LSTM for sequential data processing
$h_1 \leftarrow$ LSTM$(z,$ `lstm_units`$)$
Dense Layers: ▷ Dense layer for feature extraction
$h_2 \leftarrow$ Dense$(h_1,$ `dense_units, activation`$)$ ▷ Output layer for Remaining Useful Life prediction
$RUL \leftarrow$ Dense$(h_2, 1,$ `relu_cut`$)$
Return: $RUL \in \mathbb{R}^{B\times 1}$ ▷ Final model output

 doi:10.20944/preprints202401.1817.v1

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 [69]. 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, Bayesian Optimization uses probabilistic models to predict the performance of different hyperparameter configurations, guiding the search toward promising regions [70]. This is particularly beneficial in high-dimensional spaces, where an exhaustive search becomes computationally expensive. The implementation of Bayesian Optimization can be seen in Algorithm 3, which was developed using the Keras Tuner library [71]. The tuning process involved defining a hypermodel class (`MyHyperModel`) that inherits from the Keras Tuner `HyperModel` class. This class encapsulates the structure of the LIB RUL prediction models (CLDNN or TT), as well as the search space of hyperparameters. Additionally, a custom Bayesian Optimization tuner class (`MyBayesianOptimizationTuner`) was defined, extending the `BayesianOptimization` class from Keras Tuner. The tuner was configured to minimize, which was set as the validation mean absolute error (MAE) of the LIB RUL prediction models. The search was conducted over a specified number of trials (100 in this case) on the training dataset for a given number of epochs (10 in this case), and the models' performance was evaluated on the validation dataset.

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**

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.

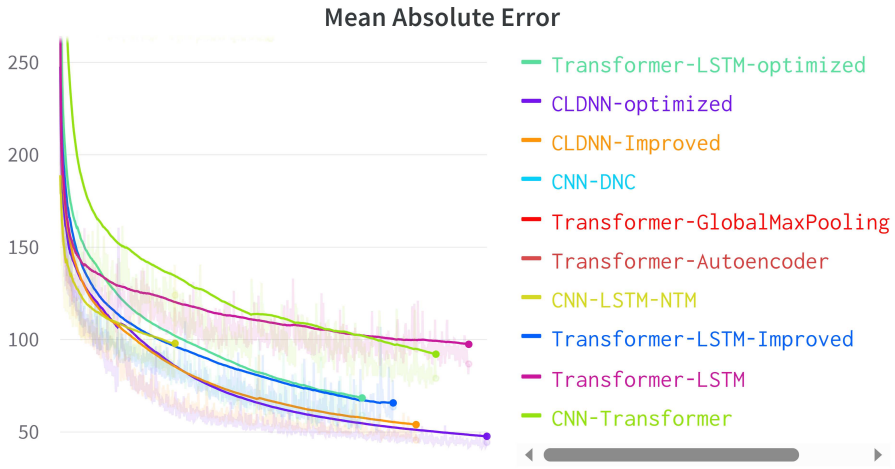
4. Results and Discussion

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

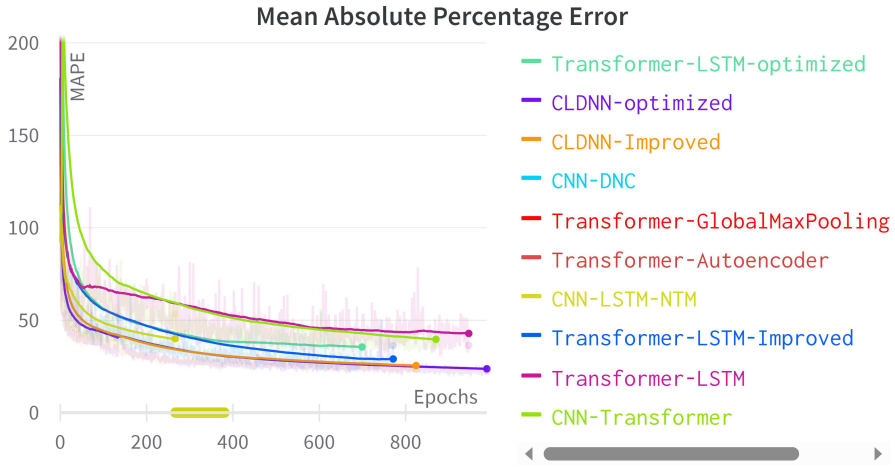
 doi:10.20944/preprints202401.1817.v1

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 the following 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 hybrid neural network architectures developed, encompassing the baseline CLDNN, the optimized CLDNN, the Transformers-LSTM, and its optimized counterpart.



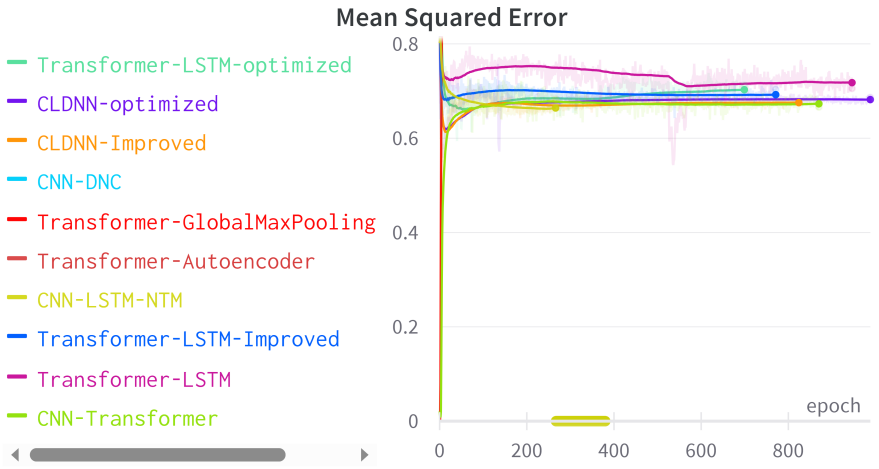
(**a**)MAE- Mean Absolute Error for the best performing models.



(**b**)MAPE- Mean Absolute Percentage Error for the best performing models.

Figure 5. *Cont*.

 doi:10.20944/preprints202401.1817.v1



(**c**)MSE- Mean Squared Error for the best performing models.

Figure 5. Comparison of different error metrics 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 6 different 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 5a–c. The CLDNN model exhibited an MAE of 84.012, MAPE of 25.676, and MSE of 0.6754. Conversely, the Temporal-Transformer model recorded a MAE of 85.134, MAPE of 28.7932, and MSE of 0.7136. To view the model's train duration refer to Figure 6.

 doi:10.20944/preprints202401.1817.v1

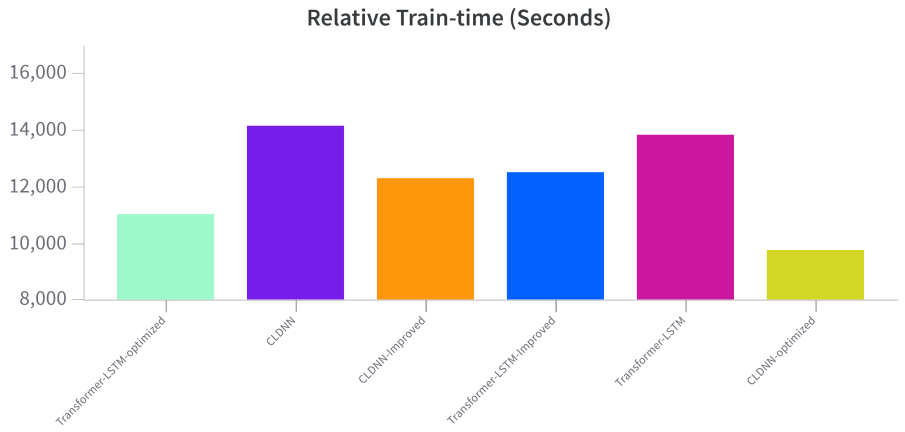


Figure 6. Training time of the Baseline and Optimized models in Seconds.

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 lower number of attention heads compared to the original model. This meant that the original model was over-parameterized.
- 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 Transformer-LSTM-optimized 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

 doi:10.20944/preprints202401.1817.v1

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 CLDNN and TT models to existing approaches

In comparing our results to other approaches in the literature, several standout models, namely Auto encoder-DNN, Auto-CNN-LSTM, and ATCMN, exhibit relevance and importance to our research. These models, as presented in Table 1, 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 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, DCNN, and TCNN (Table 4) [72]. Table 4 also highlights the Avg 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 [57], a model similar to ours, utilises discharging time, voltage, and capacity for RUL estimation; however, our models outperform it in predicting remaining cycles for LIBs within a 100-cycle moving window and using a CC-CV charge policy. Notably, the Auto-CNN-LSTM model by Ren et al. [25] operates on a distinct dataset with less variation and limited diversity in input parameters, acknowledged by the authors themselves. This dataset disparity, compared to our more extensive and diverse dataset, underscores the potential for variations in outcomes between the two models, highlighting the importance of our comprehensive dataset in providing a robust foundation for modelling and prediction, ultimately enhancing the reliability and applicability of our findings.

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 [57]	84	-	32.8%	18 ms
Deep-CNN[66]	-	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

 doi:10.20944/preprints202401.1817.v1

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) ranging 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.

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) [73–75] 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's counterparts [76,77]. 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.

Data Availability Statement: The dataset used in this study can be found at https://data.matr.io/published by the Toyota Research Institute.

Acknowledgments: This work was supported in part by the project called "Research and Development of a Highly Automated and Safe Streamlined Process for Increase Lithium-ion Battery Repurposing and Recycling" (REBELION) under Grant 101104241 and in part by the UK Research and Innovation (UKRI) project "Reuse and Recycling of Lithium-Ion Batteries" (RELIB) under RELiB2 Grant FIRG005 and RELIB3 Grant FIRG057.

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

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
AE	Auto-Encoder
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.*; Boston Artech House, 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. *Journal of Energy Storage* **2021**, *39*, 102572. doi:https://doi.org/10.1016/j.est.2021.102572.
3. Dickinson, E.J.; Wain, A.J. The Butler-Volmer equation in electrochemical theory: Origins, value, and practical application. *Journal of Electroanalytical Chemistry* **2020**, *872*, 114145.
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. doi:10.3390/en7095995.
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. *Journal of Power Sources* **2014**, *270*, 221–237.
6. Wang, S.; Zhang, J.; Gharbi, O.; others. Electrochemical impedance spectroscopy. *Nat Rev Methods Primers* **2021**, *1*, 41. doi:10.1038/s43586-021-00039-w.
7. Ramadesigan, V.; Boovaragavan, V.; Pirkle, J.C.; Subramanian, V.R. Efficient reformulation of solid-phase diffusion in physics-based lithium-ion battery models. *Journal of The Electrochemical Society* **2010**, *157*, A854.
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. *Journal of Power Sources* **2015**, *278*, 814–825. doi:https://doi.org/10.1016/j.jpowsour.2014.08.089.
9. Li, Y.; Liu, K.; Foley, A.M.; ZÃijlke, 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. *Renewable and Sustainable Energy Reviews* **2019**, *113*, 109254. doi:https://doi.org/10.1016/j.rser.2019.109254.
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 Transactions on power electronics* **2013**, *28*, 5919–5926.
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. doi:https://doi.org/10.1016/j.energy.2020.118000.
12. Si, X.S.; Wang, W.; Hu, C.H.; Zhou, D.H. Remaining useful life estimation–a review on the statistical data driven approaches. *European journal of operational research* **2011**, *213*, 1–14.
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.
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. *Applied Energy* **2020**, *260*, 114296. doi:https://doi.org/10.1016/j.apenergy.2019.114296.
15. Alfarizi, M.G.; Tajiani, B.; Vatn, J.; Yin, S. Optimized random forest model for remaining useful life prediction of experimental bearings. *IEEE Transactions on Industrial Informatics* **2022**.

16. Hu, X.; Jiang, J.; Cao, D.; Egardt, B. Battery Health Prognosis for Electric Vehicles Using Sample Entropy and Sparse Bayesian Predictive Modeling. *IEEE Transactions on Industrial Electronics* **2016**, *63*, 2645–2656. doi:10.1109/TIE.2015.2461523.
17. Kraus, M.; Feuerriegel, S. Forecasting remaining useful life: Interpretable deep learning approach via variational Bayesian inferences. *Decision Support Systems* **2019**, *125*, 113100.
18. Mosallam, A.; Medjaher, K.; Zerhouni, N. Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. *Journal of Intelligent Manufacturing* **2016**, *27*, 1037–1048.
19. Guan, Q.; Wei, X. The Statistical Data-driven Remaining Useful Life PredictionâĂŤA Review on the Wiener Process-based Method. 2023 Prognostics and Health Management Conference (PHM). IEEE, 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. doi:10.1109/ACCESS.2018.2837156.
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 Transactions on Industrial Electronics* **2018**, *65*, 6730–6739. doi:10.1109/TIE.2017.2787586.
22. Shen, S.; Sadoughi, M.; Chen, X.; Hong, M.; Hu, C. A deep learning method for online capacity estimation of lithium-ion batteries. *Journal of Energy Storage* **2019**.
23. 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 Transactions on Industry Applications* **2020**, *56*, 5565–5574.
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.
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 Transactions on Industrial Informatics* **2020**, *17*, 3478–3487.
26. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. *Neural Information Processing Systems* **2014**, *abs/1409.3215*, [1409.3215].
27. Yang, S.; Eisenach, C.; Madeka, D. MQRetNN: Multi-Horizon Time Series Forecasting with Retrieval Augmentation, 2022, [arXiv:cs.LG/2207.10517].
28. Xiong, R.; Li, L.; Tian, J. Towards a smarter battery management system: A critical review on battery state of health monitoring methods. *Journal of Power Sources* **2018**, *405*, 18–29. doi:https://doi.org/10.1016/j.jpowsour.2018.10.019.
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.; others. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy* **2019**, *4*, 383–391.
30. Sainath, T.N.; Vinyals, O.; Senior, A.; Sak, H. Convolutional, long short-term memory, fully connected deep neural networks. 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP). Ieee, 2015, pp. 4580–4584.
31. 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.
32. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Computation* **1997**, *9*, 1735âĂŞ1780. doi:10.1162/neco.1997.9.8.1735.
33. Lim, B.; ArÄśk, S.ï£¡.; Loeff, N.; Pfister, T. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting* **2021**, *37*, 1748–1764. doi:https://doi.org/10.1016/j.ijforecast.2021.03.012.
34. 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 Materials* **2020**, *31*, 195–220.
35. Tremblay, O.; Dessaint, L.A.; Dekkiche, A. A Generic Battery Model for the Dynamic Simulation of Hybrid Electric Vehicles. *2007 IEEE Vehicle Power and Propulsion Conference* **2007**, pp. 284–289.
36. Zhao, X.; Wang, Y.; Sahinoglu, Z.; Wada, T.; Hara, S.; Callafon, R. State-of-charge estimation for batteries: A multi-model approach. 2014, pp. 2779–2785. doi:10.1109/ACC.2014.6858976.
37. Song, W.; Wu, D.; Shen, W.; Boulet, B. A Remaining Useful Life Prediction Method for Lithium-ion Battery Based on Temporal Transformer Network. *Procedia Computer Science* **2023**, *217*, 1830–1838. 4th International Conference on Industry 4.0 and Smart Manufacturing, doi:https://doi.org/10.1016/j.procs.2022.12.383.

38. Yang, Y. A machine-learning prediction method of lithium-ion battery life based on charge process for different applications. *Applied Energy* **2021**, *292*, 116897.
39. 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.
40. Lotfi, N.; Li, J.; Landers, R.G.; Park, J. Li-ion Battery State of Health Estimation based on an improved Single Particle model. 2017 American Control Conference (ACC), 2017, pp. 86–91. doi:10.23919/ACC.2017.7962935.
41. 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. doi:https://doi.org/10.1016/j.measurement.2023.112838.
42. 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. doi:https://doi.org/10.1016/j.neucom.2019.09.074.
43. 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. *Journal of Cleaner Production* **2021**, *317*, 128265. doi:https://doi.org/10.1016/j.jclepro.2021.128265.
44. 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 20Ah LiNixMnyCo1- x-yO2 pouch cell. *Journal of Electroanalytical Chemistry* **2020**, *858*, 113729.
45. Bashir, I.; Al-Greer, M.; others. Lithium-ion Batteries Capacity Degradation Trajectory Prediction Based on Decomposition Techniques and NARX Algorithm. 2022 57th International Universities Power Engineering Conference (UPEC). IEEE, 2022, pp. 1–6.
46. Zheng, S.; Ristovski, K.; Farahat, A.; Gupta, C. Long Short-Term Memory Network for Remaining Useful Life estimation. 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), 2017, pp. 88–95. doi:10.1109/ICPHM.2017.7998311.
47. Wang, J.; Wen, G.; Yang, S.; Liu, Y. Remaining Useful Life Estimation in Prognostics Using Deep Bidirectional LSTM Neural Network. 2018 Prognostics and System Health Management Conference (PHM-Chongqing), 2018, pp. 1037–1042. doi:10.1109/PHM-Chongqing.2018.00184.
48. Spagnol, P.; Rossi, S.; Savaresi, S.M. Kalman filter SoC estimation for Li-ion batteries. 2011 IEEE International Conference on Control Applications (CCA). IEEE, 2011, pp. 587–592.
49. Hu, C.; Youn, B.D.; Chung, J. A multiscale framework with extended Kalman filter for lithium-ion battery SOC and capacity estimation. *Applied Energy* **2012**, *92*, 694–704.
50. 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. *Applied Energy* **2012**, *95*, 227–237. doi:https://doi.org/10.1016/j.apenergy.2012.02.044.
51. Lambert, B. *A StudentâĂŹs Guide to Bayesian Statistics*; SAGE, 2018.
52. Li, X.; Ding, Q.; Sun, J.Q. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety* **2018**, *172*, 1–11. doi:https://doi.org/10.1016/j.ress.2017.11.021.
53. Sateesh Babu, G.; Zhao, P.; Li, X.L. Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life. Database Systems for Advanced Applications; Navathe, S.B.; Wu, W.; Shekhar, S.; Du, X.; Wang, X.S.; Xiong, H., Eds.; Springer International Publishing: Cham, 2016; pp. 214–228.
54. 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. doi:10.1109/ACCESS.2019.2919566.
55. Tan, W.M.; Teo, T.H. Remaining Useful Life Prediction Using Temporal Convolution with Attention. *AI* **2021**, *2*, 48–70. doi:10.3390/ai2010005.
56. Fan, Y.; Xiao, F.; Li, C.; Yang, G.; Tang, X. A novel deep learning framework for state of health estimation of lithium-ion battery. *Journal of Energy Storage* **2020**, *32*, 101741.
57. 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. *Journal of Energy Storage* **2023**, *62*, 106903. doi:https://doi.org/10.1016/j.est.2023.106903.
58. 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. *Renewable and Sustainable Energy Reviews* **2022**, *156*, 111843.

59. LeCun, Y.; Bengio, Y.; others. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks* **1995**, *3361*, 1995.
60. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Advances in neural information processing systems* **2017**, *30*.
61. Zhai, J.; Zhang, S.; Chen, J.; He, Q. Autoencoder and its various variants. 2018 IEEE international conference on systems, man, and cybernetics (SMC). IEEE, 2018, pp. 415–419.
62. Collier, M.; Beel, J. Implementing neural turing machines. Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27. Springer, 2018, pp. 94–104.
63. Graves, A.; Wayne, G.; Reynolds, M.; Harley, T.; Danihelka, I.; Grabska-Barwińska, A.; Colmenarejo, S.G.; Grefenstette, E.; Ramalho, T.; Agapiou, J.; others. Hybrid computing using a neural network with dynamic external memory. *Nature* **2016**, *538*, 471–476.
64. Severson, K.A.; Attia, P.M.; Jin, N.; Perkins, N.; Jiang, B.; Yang, Z.; Chen, M.H.; Aykol, M.; Herring, P.K.; Fraggedakis, D.; Bazant, M.Z.; Harris, S.J.; Chueh, W.C.; Braatz, R.D. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy* **2019**, *4*, 383 – 391. doi:https://doi.org/10.1038/s41560-019-0356-8.
65. 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 Transactions on Transportation Electrification* **2023**, *9*, 2628âĂŞ2644. doi:https://doi.org/10.1109/TTE.2022.3212024.
66. Shen, S.; Sadoughi, M.; Chen, X.; Hong, M.; Hu, C. A deep learning method for online capacity estimation of lithium-ion batteries. *Journal of Energy Storage* **2019**, *25*, 100817. doi:https://doi.org/10.1016/j.est.2019.100817.
67. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.r.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N.; Kingsbury, B. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine* **2012**, *29*, 82–97. doi:10.1109/MSP.2012.2205597.
68. Ma, Q.; Zhang, M.; Xu, Y.; Song, J.; Zhang, T. Remaining Useful Life Estimation for Turbofan Engine with Transformer-based Deep Architecture. 2021 26th International Conference on Automation and Computing (ICAC), 2021, pp. 1–6. doi:10.23919/ICAC50006.2021.9594150.
69. Probst, P.; Boulesteix, A.L.; Bischl, B. Tunability: Importance of hyperparameters of machine learning algorithms. *The Journal of Machine Learning Research* **2019**, *20*, 1934–1965.
70. 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. *Journal of Electronic Science and Technology* **2019**, *17*, 26–40.
71. Gulli, A.; Pal, S. *Deep learning with Keras*; Packt Publishing Ltd, 2017.
72. 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 Reports* **2021**, *7*, 5562–5574. doi:https://doi.org/10.1016/j.egyr.2021.08.182.
73. Hasani, R.; Lechner, M.; Amini, A.; Rus, D.; Grosu, R. Liquid time-constant networks. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, Vol. 35, pp. 7657–7666.
74. Hasani, R. Interpretable recurrent neural networks in continuous-time control environments. PhD thesis, Wien, 2020.
75. 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. *International Journal of Energy Research* **2021**, *45*, 5747–5765.
76. 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. *Applied Energy* **2023**, *336*, 120808.
77. Lam, R.; Sanchez-Gonzalez, A.; Willson, M.; Wirnsberger, P.; Fortunato, M.; Alet, F.; Ravuri, S.; Ewalds, T.; Eaton-Rosen, Z.; Hu, W.; Merose, A.; Hoyer, S.; Holland, G.; Vinyals, O.; Stott, J.; Pritzel, A.; Mohamed, S.; Battaglia, P. Learning skillful medium-range global weather forecasting. *Science*, *0*, eadi2336. doi:10.1126/science.adi2336.

 doi:10.20944/preprints202401.1817.v1

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.