

# De-SaTE: Denoising Self-attention Transformer Encoders for Li-ion Battery Health Prognostics

Gaurav Shinde <sup>\*</sup>, Rohan Mohapatra <sup>†</sup>, Pooja Krishan <sup>†</sup>, and Saptarshi Sengupta <sup>†</sup>

<sup>\*</sup>Department of Engineering, San José State University, San Jose, CA, USA

<sup>†</sup>Department of Computer Science, San José State University, San Jose, CA, USA

Email: <sup>\*</sup>gauravyeshwant.shinde@sjsu.edu, <sup>†</sup>rohan.mohapatra@sjsu.edu,

<sup>†</sup>pooja.krishan@sjsu.edu, <sup>†</sup>saptarshi.sengupta@sjsu.edu

**Abstract**—Lithium Ion (Li-ion) batteries have gained widespread popularity across various industries, from powering portable electronic devices to propelling electric vehicles and supporting energy storage systems. A central challenge in managing Li-ion batteries effectively is accurately predicting their Remaining Useful Life (RUL), which is a critical measure for proactive maintenance and predictive analytics. This study presents a novel approach that harnesses the power of multiple denoising modules, each trained to address specific types of noise commonly encountered in battery data. Specifically we use a denoising auto-encoder and a wavelet denoiser to generate encoded/decomposed representations, which are subsequently processed through dedicated self-attention transformer encoders. After extensive experimentation on the NASA and CALCE datasets, we are able to characterize a broad spectrum of health indicator estimations under a set of diverse noise patterns. We find that our reported error metrics on these datasets are on par or better with the best reported in recent literature.

**Index Terms**—Prognostics and Health Management, Remaining Useful Life, Denoising Auto-Encoders, Lithium-ion Batteries, Transformer, Battery Health

## I. INTRODUCTION

Lithium-ion batteries (Li-ion) have emerged as a leading energy storage solution, prized for their exceptional energy density, rapid power response, recyclability, and portability. Their unparalleled combination of energy and power density has made them the preferred choice for applications ranging from hybrid and electric vehicles to portable electronics. However, the capacity degradation of Li-ion batteries over time is a concern. Various factors, including temperature, state of charge, cycling rate, and operating conditions, can contribute to capacity loss, increased internal resistance, and eventual failure.

The capacity of Li-ion batteries serves as a direct indicator of their health, and accurately predicting their Remaining Useful Life (RUL) is essential for effective battery management. RUL is the estimated duration a battery can remain operational before reaching a critical failure point, commonly described as End Of Life (EOL), where the battery can no longer supply its rated capacity or exhibits significantly diminished performance.

Prognostics and Health Management (PHM) offers a multidisciplinary approach to address the capacity degradation and reliability concerns associated with Li-ion batteries. PHM

encompasses data acquisition, diagnostics, and the core component of prognostics, which predicts a system's RUL. In the context of Li-ion batteries, PHM provides critical insights into monitoring their health, guiding maintenance decisions, and reducing the risk of unexpected failures, particularly in safety-critical applications like electric vehicles and aerospace systems.

In the past, RUL prediction primarily relied on conventional machine learning models such as Convolutional Neural Network (CNN), Recurrent Neural network (RNN), and Long Short Term Memory (LSTM) networks. These models, while valuable, often faced challenges in capturing long-term dependencies in sequential data and required manual feature engineering. Consequently, they were less flexible in handling diverse datasets and complex real-world scenarios.

The recent shift in RUL prediction has centered around attention-based mechanisms. These approaches have gained prominence due to their innate capacity to autonomously identify and adapt to intricate temporal and spatial dependencies within data. By doing so, they eliminate the need for extensive feature engineering and enhance adaptability to the dynamic and multifaceted nature of RUL prediction tasks. This transition away from traditional architectures reflects a concerted effort to improve prediction accuracy and model flexibility in the realm of complex systems.

Chen et. al [1] explores the innovative utilization of a Denoising Auto-Encoder to enhance data representation from battery inputs, which inherently exhibit noise due to various factors. However, it's essential to acknowledge that the measurement noise in practical scenarios may not adhere to only a Gaussian distribution as assumed in the paper. To address this concern and ensure robust noise handling, our research proposes a novel approach. We employ diverse set of noise types in the denoising framework, where each noise type is associated with a dedicated auto-encoder and its corresponding transformer encoder. The decoded representations from these auto-encoders undergo transformation through the respective transformer encoders. Following this transformation process, we introduce a minimization layer, tasked with identifying the noise type that yields the minimum error value. This adaptive noise modeling approach substantially bolsters the auto encoder's capacity to capture the wide spectrum of non-Gaussian noise characteristics commonly encountered in

battery data. Consequently, due to the denoising process, this method culminates in generating higher-quality data representations for subsequent analysis, where our prediction is based on the noise type associated with the minimum error.

**Contributions:** The major contributions of this paper are:

- 1) **Multi-Faceted Noise Mitigation:** Our work introduces a comprehensive noise mitigation strategy by employing dedicated denoising auto-encoders and wavelet decomposers for various noise types present in battery operational data. Each denoising module is tailored to handle specific noise characteristics, enhancing the model's adaptability and relevance in real-world scenarios.
- 2) **Robustness to different magnitudes of noise:** The proposed denoisers are tuned to handle various levels of noise, in addition to the different types of noise, thereby making the larger prediction architecture more robust to fluctuations in input data.
- 3) **Better data representation leads to enhanced accuracy:** Our novel architecture not only effectively denoises the input data, but also significantly improves the quality of data representations. This, in turn, leads to better predictions and adds valuable contribution to the field of battery health management.
- 4) **A modular architecture for all complex processing:** The proposed architecture processes the input by passing it through the denoising modules to encode/decompose various types of noisy data, and the self-attention encoder network subsequently learns the degradation physics (from the enhanced representation) and predicts the remaining useful life.

## II. RELATED WORK

This section of this research paper entails a methodical and comprehensive evaluation of various methods for predicting RUL. The systematic breakdown of these methods facilitates a more precise understanding of their strengths, limitations, and applicability. This meticulous analysis provides an essential foundation for drawing meaningful conclusions and developing insightful recommendations. This section will provide an overview of the evolution of RUL prediction methods for Li-ion batteries, focusing on the shift from model-driven to data-driven approaches.

### A. Model Based Methods

Model-based methods attempt to set up mathematical or physical models to describe degradation processes, and update model parameters using measured data [2] [3]. Model-based methods could incorporate both expert knowledge and real-time information. Consequently, they may work well in RUL prediction of batteries [4]. However, model-based approaches require accurate knowledge of the battery's internal structure and operating conditions, which can be challenging to obtain in practice. Additionally, these models may not generalize well to different battery chemistry or operating conditions.

#### 1) Physical Model

Physical model quantifies the factors that influence a battery's performance, and obtains the description of performance evolution. However, considering that several factors may interact with each other to impact the degradation of performance, it is not easy to make a reliable and precise model to simulate battery system. Thus, this approach usually focuses on the specific physical and chemical phenomena occurring during utilization [5]. Physics-based models rely on mathematical equations to describe the battery's physical attributes and controlling principles.

#### 2) Electrochemical Model

Electrochemical models are based on precise mathematical models of electrochemical processes that occur within the battery, such as chemical reactions, lithium ion and electron movement, and heat impacts. The electrochemical model can represent the internal variables of the battery well, so it can achieve high accuracy [6]. However, due to the complexity and nonlinearity of battery behavior, as well as the challenge of precisely describing the electrochemical processes within the battery, establishing accurate electrochemical-based models can be difficult.

#### 3) Adaptive Filter Method

Adaptive filter is a digital filter whose coefficient varies with the target in order for the filter to converge to the optimal state [7]. The cost function is the optimization criteria, and the most common is the root mean square of the error signal between the adaptive filter output and the intended signal [8].

#### 4) Stochastic Process Methods

Stochastic process methods are based on statistical theory and combined with other mathematical principles [9]. These are based on the notion that battery degradation is a stochastic process that can be modeled using probabilistic methods. The advantage of stochastic process approaches is that they can represent the unpredictability and uncertainty inherent in the battery deterioration process, but they may need more complicated modeling and computational techniques.

### B. Data Driven Methods

A data-driven method for predicting the RUL of lithium-ion batteries involves analyzing the battery's operating data to estimate its degradation level and predict when it will reach the end of its useful life. This method relies on the analysis of battery performance metrics such as voltage, current, and temperature, as well as historical data. The data-driven method can directly mine the degradation information of lithium-ion battery through historical data, and there is no need to establish a specific mathematical model [10]. The data-driven RUL prediction method is a widely used method at this stage [11]. Overall, data-driven RUL prediction methods have the potential to improve battery performance and reduce costs associated with premature battery failure, making them an important area of research and development for the energy storage industry.

### 1) Traditional Machine Learning

This section provides an overview of traditional machine learning methods for RUL estimation of Li-ion batteries. Zhou et al. [12] proposed an incremental optimized Relevance Vector Machine (RVM) that achieves efficient online training for model updating using battery data from the National Aeronautics and Space Administration (NASA) Prognostics Center of Excellence (PCoE). While this method can provide accurate predictions, it requires a large amount of training data and the selection of hyperparameters can be challenging. Refer [12] for mathematical computations of the RVM output. Qin et al. [13] proposed a Feature Vector Selection (FVS) and RVM approach that addresses the memory consumption issue. The method involves using RVM to estimate the RUL and FVS to remove redundant data points from the input data. Additionally, Chen et al. [14] uses Cuckoo Search (CS) based RVM with hybrid kernel for increased accuracy. This approach uses the CS algorithm to intelligently determine kernel parameters and their weights. While this method shows promising results, the complexity of the kernel functions and their parameters can be a concern. To further improve prediction accuracy, multiple ensemble models have been proposed. Li et al. [15] proposed RUL estimation based on Ensemble Learning with Least Square Support Vector Machine (LS-SVM), while Patil et al. [16] proposed a multistage Support Vector Machine (SVM) based approach that integrates the classification and regression attributes of Support Vector (SV) based machine learning technique. Wang et al. [17] proposed RUL estimation using Support Vector Regression (SVR) optimized by Artificial Bee Colony (ABC). The ABC algorithm optimizes SVR core parameters. Over the years, traditional machine learning methods have been widely used for RUL estimation of Li-ion batteries. While these methods have shown promising results, they also have their limitations. Overall, these methods offer a range of approaches to address the challenge of RUL estimation and can be useful in various applications, but careful consideration is needed when selecting the most appropriate method for a particular scenario.

### 2) The Advent of Deep Learning

Deep learning models automatically learn relevant features from raw data, capture complex relationships between input features and output targets, and generalize well to new data. Various deep learning techniques have been proposed for RUL estimation, including the LSTM-RNN method proposed by Zhang et al. [18], which captures long-term temporal dependencies of capacity degradation by using a sequence of memory cells to store past information. Ren et al. [19] proposed an integrated deep learning approach that combines the Auto-encoder with Deep Neural Network (DNN) to extract multi-dimensional features and estimate RUL. Liu et al. [20] proposed a Bayesian Model Averaging (BMA) and LSTM ensemble method that accounts for uncertainty and provides better prediction accuracy. Zhang et al. [21] proposed an online estimation method that combines partial incremental capacity with an Artificial Neural Network (ANN) for estimating battery State of Health (SOH) and RUL. Zhou et al. [22] proposed

a Temporal Convolution Network (TCN) based framework that uses causal and dilated convolution techniques to capture local capacity regeneration and improve prediction accuracy. Ren et al. [23] proposed an Auto-encoder-Convolutional Neural Network-Long Short Term Memory Model that mines deeper information in finite data by using an Auto-encoder to extract features from raw data and a combination of convolutional and LSTM layers to learn temporal patterns in the data. Additionally, Li et al. [24] designed an AST-LSTM NN for multiple battery sharing predictions, which uses a variant of the LSTM model to predict the RUL of multiple batteries. Overall, the shift from traditional machine learning to deep learning has led to significant improvements in the accuracy and robustness of RUL estimation methods for Li-ion batteries. However, deep learning models can be computationally expensive to train and require large amounts of data, which may limit their applicability in some domains.

In recent years, the landscape of RUL estimation for Lithium Ion Batteries (LIB) has undergone a remarkable transformation with the emergence of transformer-based models. This paradigm shift can be attributed to the extraordinary capabilities of transformers in processing sequential data efficiently, rendering them exceptionally well-suited for intricate time-series forecasting tasks such as RUL estimation. Transformers have risen to prominence due to their innate prowess in capturing not only short-term but also long-term dependencies within data, endowing them with the capacity to model the intricate and evolving degradation patterns of Li-ion batteries with precision.

## III. SELF-ATTENTION WITH VARIABLE DENOISING

In this paper, a novel architecture designed to handle diverse noise distributions in battery capacity data is presented. A data pre-processing step involving normalization is introduced to enhance model robustness.

$x = \{x_1, x_2, \dots, x_n\}$ , denotes input capacity mapped to  $(0, 1]$   
 $x_0 = \frac{x}{C_0}$ , where  $C_0$  denotes rated capacity.

Subsequently, multiple denoising schemes, each trained to mitigate a specific noise type are leveraged. The encoded representations from these denoising modules are then subjected to individual self-attention layers. Once the data has passed through the self-attention layers, each one intricately connected to its respective noise reducer, then yield individual metric values such as Relative Error (RE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). In this context, a minimization strategy is employed, which selects the minimum value among these metrics, thereby obtaining the most optimal error estimate. This architecture not only ensures robust performance but also proves highly effective in estimating failures, even when confronted with the presence of diverse noise types inherent in battery data.

The system architecture of the entire process is shown in Fig. 1.

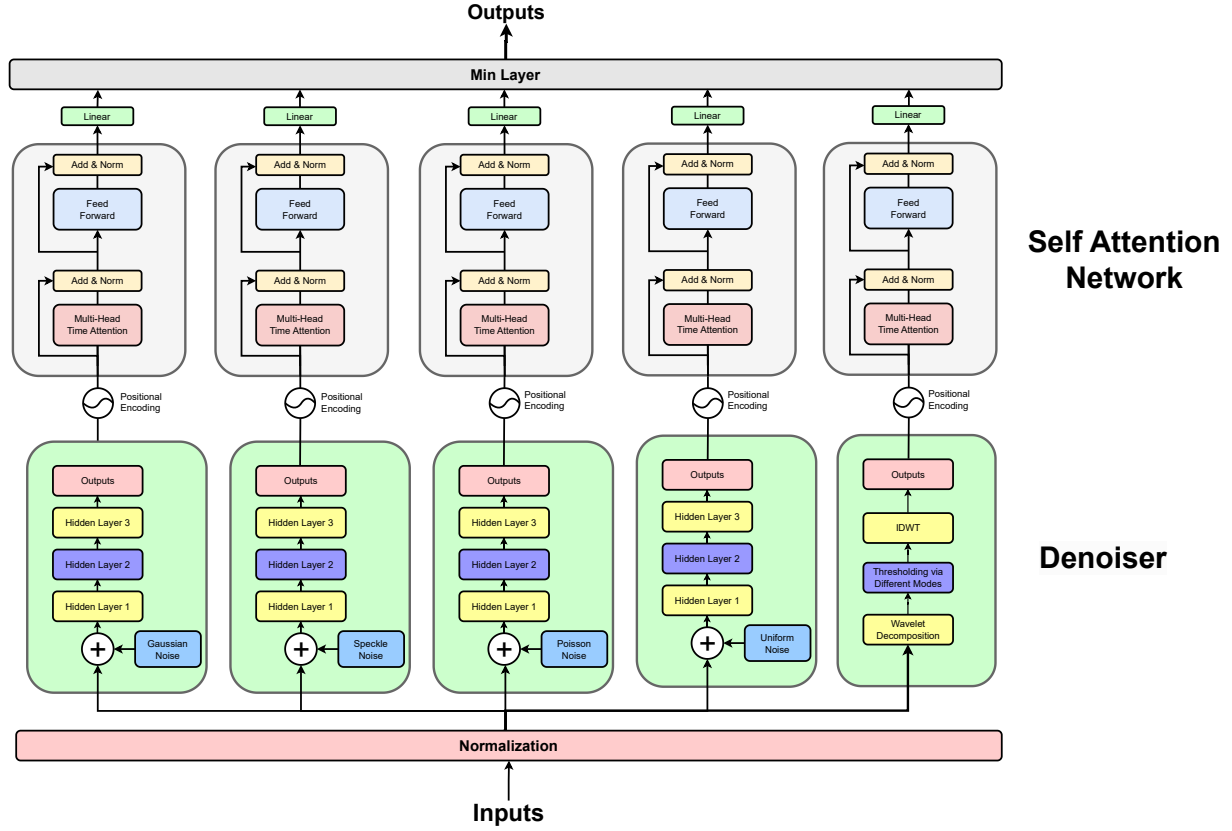


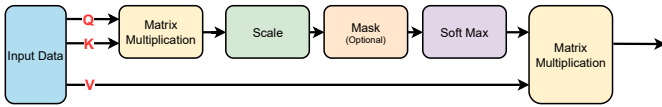
Fig. 1: The proposed De-SaT: Denoising Self-attention Transformer Encoder architecture

#### A. Positional Encoding

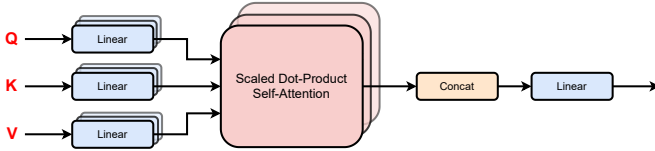
$$PE_{(pos,2k)} = \sin\left(\frac{pos}{10000^{2k/d_k}}\right) \quad (1)$$

$$PE_{(pos,2k+1)} = \cos\left(\frac{pos}{10000^{2k/d_k}}\right) \quad (2)$$

#### B. Self Attention



(a) Scaled Dot-Product Self-Attention mechanism



(b) Multi-Head Self-Attention mechanism

Fig. 2: Multi-Head Self-Attention architecture

The encoder's self-attention mechanism computes attention scores for each position in the input sequence, allowing the model to weigh the importance of different elements in the sequence when encoding a particular position. This is typically

computed using a weighted sum of the queries, keys and values as shown Fig. 6a

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

Here,

$Q$  represents the query matrix for the input sequence  
 $K$  represents the key matrix for the input sequence  
 $V$  represents the value matrix for the input sequence  
 $d_k$  is a scaling factor to stabilize the gradients

#### C. Multi-head Attention

In Fig. 6b, we outline the multi-head attention process. The self-attention mechanism is often used in multiple heads to capture different types of dependencies:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^O$$

Here,

$\text{head}_i$  represents the output of the  $i$ -th attention head.

$W^O$  is a learnable weight matrix.

#### D. Feed forward network

After attention, the output passes through a feed-forward neural network:

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2 \quad (3)$$

Here,

$W_1, b_1, W_2, b_2$  are learnable weights and biases.

#### E. Learning

Our architecture can be divided into two tasks: *denoising* and (*Relative Error (RE)*, *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*) *evaluation*. The learning procedure optimizes both tasks simultaneously within a unified framework. Mean Square Error (MSE) is used to evaluate loss, and the objective function [1] is defined as follows:

$$L = \sum_{t=T+1}^n (x_t - bx_t)^2 + \alpha \sum_{i=1}^n \phi(ex_i - bx_i) + \lambda \psi(L_r P) \quad (4)$$

where,

$\alpha$  : denotes a parameter to control the relative contribution of each task;

$\lambda$  : denotes a regularization parameter;

$\phi(\cdot)$  : denotes the regularization;

$L_r P$  : denotes the learning parameters of our model.

The denoising effect along with penalized loss acts like a regularizer. Regularization techniques [25] add a penalty term to the loss function that the model is trying to minimize, boosting the model performance to tend towards smaller weights or simpler representations.

### IV. EXPERIMENTAL SETUP

#### A. Dataset Description

Two rich datasets from NASA and Center for Advanced Life Cycle Engineering (CALCE) were used to conduct the experiments. The NASA dataset, acquired from the NASA Ames Research Center, comprises records from four different Li-ion batteries, each subjected to three distinct operations: charging, discharging, and impedance measurements [26], [27]. The CALCE dataset is sourced from the Center for Advanced Life Cycle Engineering (CALCE) at the University of Maryland [28].

#### B. Data Visualization

Figures 3a and 3b illustrate the capacity degradation trends observed across various batteries in the NASA and CALCE datasets, respectively.

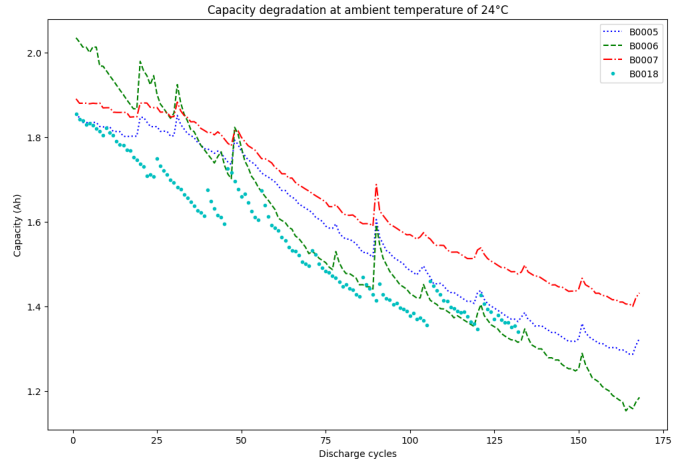
#### C. Noise distributions

##### 1) Gaussian Noise

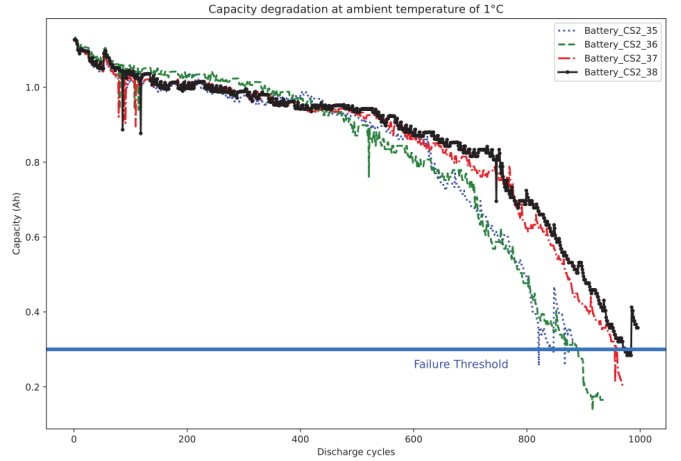
Gaussian noise is typically characterized by a mean ( $\mu$ ) and standard deviation ( $\sigma$ ).

The Probability Density Function (PDF) of Gaussian noise is given by

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (5)$$



(a) Degradation Trend on the NASA Dataset



(b) Degradation Trend on the CALCE Dataset

Fig. 3: Capacity vs. degradation cycles across NASA and CALCE datasets

##### 2) Speckle Noise

Speckle noise is often modeled as a multiplicative noise, where input values are multiplied by random values.

The PDF of Speckle noise is given by:

$$f(x; \gamma) = \frac{1}{\gamma^2} e^{-\frac{x}{\gamma}} \quad (6)$$

where  $\gamma$  is a parameter controlling the noise intensity.

##### 3) Poisson Noise

Poisson noise is characterized by a single parameter, usually the mean ( $\lambda$ ). Poisson noise often occurs when events occur at a consistent average rate but with randomness in the exact timing or occurrence of these events. This noise can be explained as instrumentation noise in battery health prognosis.

The PDF of Poisson noise is given by:

$$f(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (7)$$

#### 4) Uniform Noise

Uniform noise is characterized by a minimum value ( $a$ ) and a maximum value ( $b$ ), defining the range of possible noise values.

The PDF of Uniform noise is a constant within the range  $[a, b]$  and zero outside that range:

$$f(x; a, b) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

#### D. Denoising Autoencoder

A denoising autoencoder is a type of artificial neural network used for unsupervised learning and feature extraction. It is particularly effective in removing noise from data and learning robust representations. The basic idea behind a denoising autoencoder is to train the network to reconstruct clean data from noisy input. It consists of an encoder network that maps the input data to a latent representation and a decoder network that reconstructs the data from this representation. During training, the encoder learns to capture essential features while the decoder learns to remove noise. The loss function used for training typically measures the reconstruction error, encouraging the network to minimize the difference between the clean input and the reconstructed output.

Let  $x_{0t} = x_{0(t+1)}, x_{0(t+2)}, \dots, x_{0(t+m)} \in x_0$  denote the slice of input with  $m$  samples of a sequence. Gaussian noise is added to the normalized input to obtain the corrupted vector,  $\text{ext}$ .

DAE serves two purposes: denoising the raw input and learning a nonlinear representation [1]:

$$z = a(W\text{ext} + b) \quad (9)$$

where  $W$ ,  $b$ ,  $a(\cdot)$ , and  $z$  denote weight, bias, activation function, and the output of the DAE encoder, respectively.

Then, to reconstruct the input vector, the latent representation is mapped back to the input space, defined as follows:

$$\hat{x}_t = f_0(W_0 z + b_0) \quad (10)$$

where  $W_0$ ,  $b_0$ ,  $z$ , and  $f_0(\cdot)$  denote weight, bias, output, and map function of the output layer of the DAE encoder, respectively.

In this network, identity and ReLU functions are used as the decoding and encoding activation, respectively. Finally, the objective function is defined as follows:

$$L_d = \frac{1}{n} \sum_{t=1}^n \left( (\text{ext} - \hat{x}_t)^2 \right) + \lambda (\|W\|_F^2 + \|W_0\|_F^2) \quad (11)$$

#### E. Wavelet Transformation and Denoiser

Wavelet denoising [29] typically involves thresholding coefficients obtained from wavelet transforms. Wavelet denoising is a signal processing technique for removing noise from a voltage signal coming from a battery or another source. It is especially beneficial for dealing with signals that have been tainted by noise of various types, such as random electrical noise or interference.

#### 1) Discrete Wavelet Transform (DWT)

The DWT decomposes a signal or image into wavelet coefficients at different scales and positions:

$$\mathbf{W} = \text{DWT}(\mathbf{X}) \quad (12)$$

Here,  $\mathbf{X}$  is the original signal or image, and  $\mathbf{W}$  contains the wavelet coefficients.

#### 2) Thresholding

Thresholding is applied to the wavelet coefficients to remove or reduce noise. A common method is *soft* thresholding:

$$\hat{W}_{i,j} = \text{sign}(W_{i,j}) \cdot \max(|W_{i,j}| - \lambda, 0) \quad (13)$$

Another approach to thresholding is *hard* thresholding. Hard thresholding sets coefficients below a certain threshold value to zero and retains those above the threshold. It is defined as follows:

$$\hat{W}_{i,j} = \begin{cases} W_{i,j}, & \text{if } |W_{i,j}| \geq \lambda \\ 0, & \text{if } |W_{i,j}| < \lambda \end{cases}$$

Another thresholding type is *Garrote*. Garrote thresholding is a variation that penalizes large coefficients more than smaller ones. It is defined as follows:

$$\hat{W}_{i,j} = \frac{\text{sign}(W_{i,j}) \cdot \max(|W_{i,j}| - \lambda, 0)}{1 + \frac{\lambda}{|W_{i,j}|}} \quad (14)$$

Where:

- 1)  $\hat{W}_{i,j}$  is the denoised coefficient.
- 2)  $W_{i,j}$  is the original coefficient.
- 3)  $\lambda$  is the threshold value.

#### 3) Inverse Discrete Wavelet Transform (IDWT)

Finally, the denoised signal is reconstructed using the inverse DWT:

$$\hat{\mathbf{X}} = \text{IDWT}(\hat{\mathbf{W}})$$

Here,  $\hat{\mathbf{X}}$  is the denoised signal or image, and  $\hat{\mathbf{W}}$  contains the denoised wavelet coefficients.

#### F. Training and Evaluation

Four types of noise were introduced: Gaussian, Speckle, Poisson, and Uniform with varying noise levels (small, medium, and relatively high). We evaluated the optimal hyperparameters by a grid search.

- *Learning Rate*: 1e-3 and 1e-2
- *Number of Layers*: 1 and 2
- *Hidden Dimension*: 16 and 32
- *Noise Levels*: 0.001, 0.01, and 0.05
- *Epochs*: 2000

Evaluation Metrics:

For evaluation, we used three key metrics:

- 1) **Relative Error (RE)**: Relative Error measures the relative difference between predicted and actual values and is represented as follows:

$$RE = \frac{|Y - \hat{Y}|}{|Y|} \quad (15)$$

where,  $\hat{Y}$  represents the predicted value, and  $Y$  represents the actual value.

- 2) Root Mean Square Error (RMSE): RMSE calculates the square root of the mean of the squared differences between predicted and actual values:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{Y}_i - Y_i)^2} \quad (16)$$

where,  $\hat{Y}_i$  represents the predicted value for the  $i$ -th sample,  $Y_i$  represents the actual value for the  $i$ -th sample, and  $N$  is the total number of samples.

- 3) Mean Absolute Error (MAE): MAE calculates the mean of the absolute differences between predicted and actual values:

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{Y}_i - Y_i| \quad (17)$$

where,  $\hat{Y}_i$  represents the predicted value for the  $i$ -th sample,  $Y_i$  represents the actual value for the  $i$ -th sample, and  $N$  is the total number of samples.

This systematic exploration aimed to identify the best hyper-parameter combination for accurate predictions amidst diverse noise distributions, ensuring model robustness and scalability.

## V. RESULTS

The results of the RE, MAE, and RMSE values under various noise distributions for the NASA and CALCE datasets are shown in Figures 4 and 5. We also so comparative analysis as shown in Fig. 6 of how each noise affects the key metrics used in this paper. The values are tabulated in Tables I and II.

TABLE I: Results for the NASA dataset

Noise and metrics		LR	NoL	HD	$\alpha$	NL	Result
Gaussian	RE	0.01	1	16	1e-05	0.05	0.1674
	MAE	0.01	1	16	1e-05	0.05	0.0806
	RMSE	0.01	2	16	1e-05	0.01	0.0957
Speckle	RE	0.01	1	16	0.0001	0.05	0.1869
	MAE	0.01	1	16	1e-05	0.01	0.0807
	RMSE	0.01	1	16	1e-05	0.01	0.0935
Poisson	RE	0.01	2	16	0.0001	0.01	0.1876
	MAE	0.01	2	16	1e-5	0.05	0.0860
	RMSE	0.01	2	32	1e-5	0.001	0.1013
Uniform	RE	0.001	2	32	1e-5	0.05	0.2285
	MAE	0.001	1	32	0.0001	0.001	0.0891
	RMSE	0.001	1	32	0.0001	0.001	0.0781

The provided parameter configurations were determined as optimal for the NASA and CALCE datasets. For the NASA dataset, the optimal settings encompass a learning rate (Learning rate (LR)) of 0.01, Number of layers (NoL) of 1, 16 hidden dimensions (Hidden Dimensions (HD)), and an  $\alpha$  of 1e-05. In contrast, the optimal parameters for the CALCE dataset involve a LR of 0.001, a NoL of 1, 32 hidden dimensions (HD), and an  $\alpha$  of 0.01. These configurations were fine-tuned for the auto-encoder and transformer encoder layers working in parallel.

The ensuing results, encompassing Relative Error (RE), Mean Absolute Error (MAE), and Root Mean Square Error

TABLE II: Results for the CALCE dataset

Noise and metrics		LR	NoL	HD	$\alpha$	NL	Result
Gaussian	RE	0.001	1	32	0.01	0.001	0.052
	MAE	0.001	1	32	0.01	0.01	0.008
	RMSE	0.001	1	32	0.01	0.01	0.09
Speckle	RE	0.001	1	32	0.01	0.001	0.052
	MAE	0.001	1	32	0.01	0.01	0.008
	RMSE	0.001	1	32	0.01	0.01	0.091
Poisson	RE	0.001	1	32	0.01	0.01	0.033
	MAE	0.001	1	32	0.01	0.01	0.024
	RMSE	0.001	1	32	0.01	0.01	0.152
Uniform	RE	0.001	1	32	0.01	0.001	0.052
	MAE	0.001	1	32	0.01	0.001	0.009
	RMSE	0.001	1	32	0.01	0.001	0.093

(RMSE) values, under various noise distributions for both datasets, offer valuable insights into the model's performance and its capability to handle different noise types frequently encountered in battery data.

Continuing with our evaluation, we conducted additional experiments using wavelet denoising in conjunction with the optimal parameter settings obtained previously for both datasets, focusing on the major metric RE. Since RE is highly related to the Remaining Useful Life (RUL) of a battery, it serves as our primary evaluation metric. After rigorous testing, we observed that RE achieved superior results when paired with a denoising autoencoder followed by a transformer encoder. We explored three wavelet denoising modes - soft, hard, and garrote - each with three distinct thresholds (0.001, 0.01, and 0.05) to comprehensively assess their impact on the overall performance. The results of the study is compiled in Table III.

TABLE III: Mean RE distribution for different wavelet modes and thresholds

Dataset	Wavelet Mode	Threshold		
		<b>0.001</b>	<b>0.01</b>	<b>0.05</b>
<b>NASA</b>	Soft	0.29	0.31	0.52
	Hard	0.213	0.12	0.17
	Garotte	0.24	0.27	0.22
<b>CALCE</b>	Soft	0.76	0.78	0.81
	Hard	0.56	0.61	0.62
	Garotte	0.65	0.67	0.71

## VI. CONCLUSION AND FUTURE WORK

Our work uses a denoising framework to filter out noise from the NASA and CALCE lithium-ion battery data to estimate the RE, MAE, and RMSE metrics. In addition to the usual modeling of Gaussian noise, this study extends the recent literature to model multiple types of noise distributions. The findings show that Poisson noise produces a lower RE of 0.033 over the other noises for the CALCE dataset. However, for the NASA dataset, the Gaussian noise yields enhanced performance across the metrics. The architecture with the ensemble of denoising autoencoders has produced lower RE, MAE, and RMSE and better performance compared to the past works in the literature.



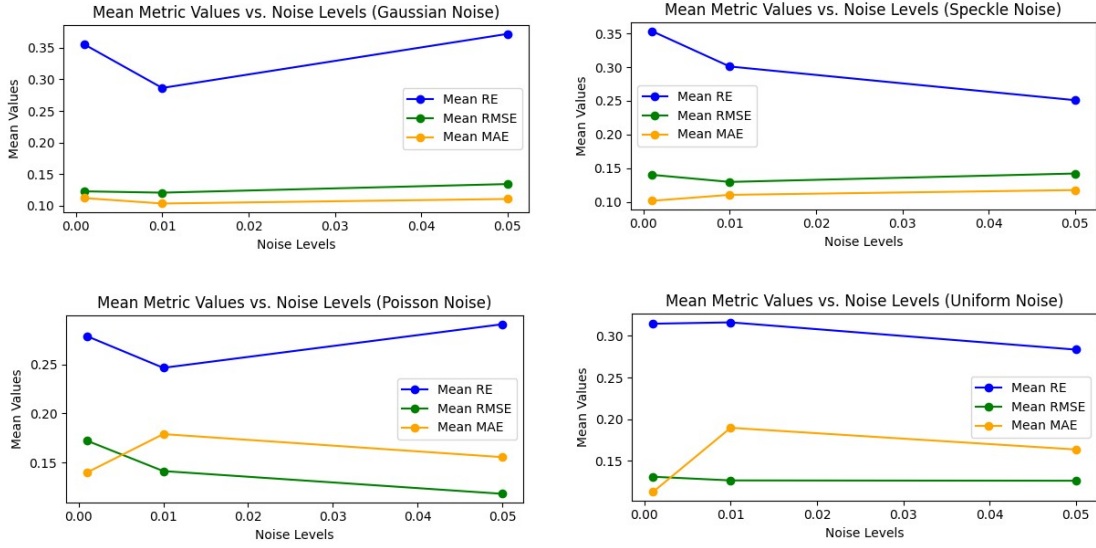


Fig. 4: Mean RE, RMSE, MAE values on the NASA data under different types of noise with varying levels

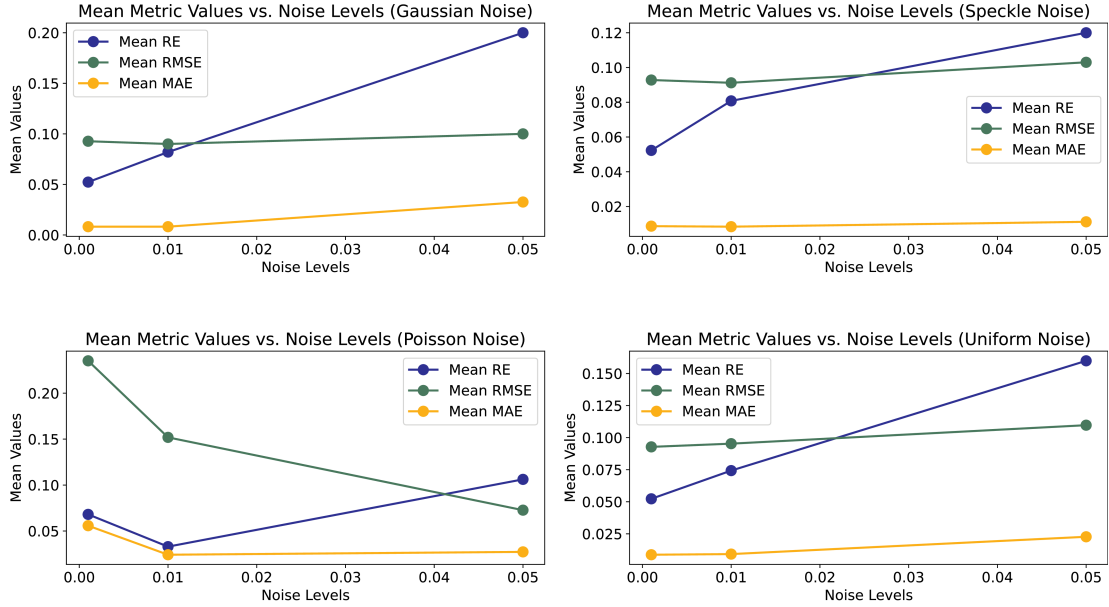


Fig. 5: Mean RE, RMSE, MAE values on the CALCE data under different types of noise with varying levels

In future, we may study the response of our proposed architecture to adversarial attacks and devise defense strategies accordingly thereby adding to its robustness.

In addition, as a next step, we intend to introduce a more traditional encoder-decoder model to extend the present capabilities and further advance our understanding of battery health prognostics.

## VII. ABBREVIATIONS

<b>Li-ion</b>	Lithium Ion
<b>RUL</b>	Remaining Useful Life
<b>PHM</b>	Prognostics and Health Management
<b>LIB</b>	Lithium Ion Batteries
<b>EOL</b>	End Of Life
<b>RVM</b>	Relevance Vector Machine

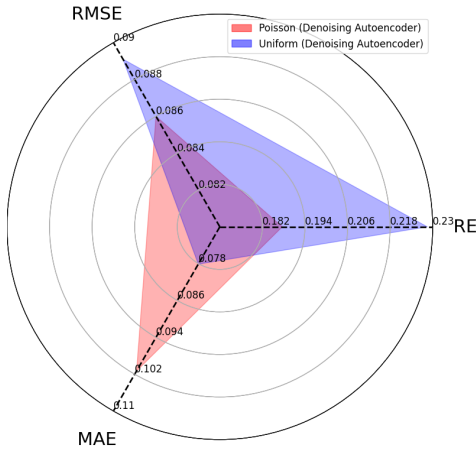
<b>BMA</b>	Bayesian Model Averaging
<b>SOH</b>	State of Health
<b>NASA</b>	National Aeronautics and Space Administration
<b>CALCE</b>	Center for Advanced Life Cycle Engineering
<b>PCoE</b>	Prognostics Center of Excellence
<b>FVS</b>	Feature Vector Selection
<b>CS</b>	Cuckoo Search
<b>SVM</b>	Support Vector Machine
<b>SV</b>	Support Vector
<b>SVR</b>	Support Vector Regression
<b>ABC</b>	Artificial Bee Colony
<b>DNN</b>	Deep Neural Network
<b>LSTM</b>	Long Short Term Memory
<b>ANN</b>	Artificial Neural Network
<b>TCN</b>	Temporal Convolution Network



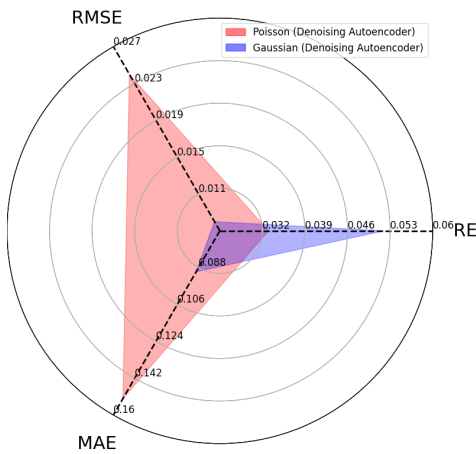
content is solely the responsibility of the author(s) and does not necessarily represent the official views of San José State University.

## REFERENCES

- [1] D. Chen, W. Hong, and X. Zhou, "Transformer network for remaining useful life prediction of lithium-ion batteries," *IEEE Access*, vol. 10, pp. 19 621–19 628, 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9714323>
- [2] Z. Huang, Z. Xu, W. Wang, and Y. Sun, "Remaining useful life prediction for a nonlinear heterogeneous wiener process model with an adaptive drift," *IEEE Transactions on Reliability*, vol. 64, no. 2, pp. 687–700, 2015. [Online]. Available: <https://ieeexplore.ieee.org/document/7051292>
- [3] H. Hanachi, J. Liu, A. Banerjee, Y. Chen, and A. Koul, "A physics-based modeling approach for performance monitoring in gas turbine engines," *IEEE Transactions on Reliability*, vol. 64, no. 1, pp. 197–205, 2015. [Online]. Available: <https://ieeexplore.ieee.org/document/6963507>
- [4] Y. Lei, N. Li, S. Gontarz, J. Lin, S. Radkowski, and J. Dybala, "A model-based method for remaining useful life prediction of machinery," *IEEE Transactions on Reliability*, vol. 65, no. 3, pp. 1314–1326, 2016. [Online]. Available: [https://www.researchgate.net/publication/304611910\\_A\\_Model-Based\\_Method\\_for\\_Remaining\\_Useful\\_Life\\_Prediction\\_of\\_Machinery](https://www.researchgate.net/publication/304611910_A_Model-Based_Method_for_Remaining_Useful_Life_Prediction_of_Machinery)
- [5] A. Barré, B. Deguilhem, S. Grolleau, M. Gérard, F. Suard, and D. Riu, "A review on lithium-ion battery ageing mechanisms and estimations for automotive applications," *Journal of Power Sources*, vol. 241, pp. 680–689, 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378775313008185>
- [6] "A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems," *Renewable and Sustainable Energy Reviews*, vol. 131, p. 110015, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1364032120303063>
- [7] M. Ahwiadi and W. Wang, "An enhanced mutated particle filter technique for system state estimation and battery life prediction," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 3, pp. 923–935, 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8421633>
- [8] L. Zhang, Z. Mu, and C. Sun, "Remaining useful life prediction for lithium-ion batteries based on exponential model and particle filter," *IEEE Access*, vol. 6, pp. 17 729–17 740, 2018. [Online]. Available: <https://ieeexplore.ieee.org/document/8318570>
- [9] "Battery health management for small-size rotary-wing electric unmanned aerial vehicles: An efficient approach for constrained computing platforms," *Reliability Engineering & System Safety*, vol. 182, pp. 166–178, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0951832018301406>
- [10] J. Zhao, L. Tian, L. Cheng, Y. Zhang, and C. Zhu, "Review on rul prediction methods for lithium-ion battery," in *2022 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)*, 2022, pp. 1501–1505. [Online]. Available: <https://ieeexplore.ieee.org/document/9949753>
- [11] Y. Zhu, Y. Shang, B. Duan, X. Gu, S. Li, and G. Chen, "A data-driven method for lithium-ion batteries remaining useful life prediction based on optimal hyperparameters," in *2022 41st Chinese Control Conference (CCC)*, 2022, pp. 7388–7392. [Online]. Available: <https://ieeexplore.ieee.org/document/9902792>
- [12] J. Zhou, D. Liu, Y. Peng, and X. Peng, "An optimized relevance vector machine with incremental learning strategy for lithium-ion battery remaining useful life estimation," in *2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2013, pp. 561–565. [Online]. Available: <https://ieeexplore.ieee.org/document/6555479>
- [13] X. Qin, Q. Zhao, H. Zhao, W. Feng, and X. Guan, "Prognostics of remaining useful life for lithium-ion batteries based on a feature vector selection and relevance vector machine approach," in *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 2017, pp. 1–6. [Online]. Available: <https://ieeexplore.ieee.org/document/7998297>
- [14] Y. Chen, C. Zhang, N. Zhang, X. Guo, H. Wang, and Y. Chen, "Cuckoo search based relevance vector machine with hybrid kernel for battery remaining useful life prediction," in *2019 Prognostics and*



(a) Poisson noise with lower RE and RMSE but higher MAE than Uniform noise on NASA Dataset (evaluated with Denoising Autoencoder)



(b) Poisson noise with lower RE but higher RMSE and MAE than Gaussian noise on CALCE Dataset (evaluated with Denoising Autoencoder)

Fig. 6: Comparative analysis on effects of different noise on metrics (RE, RMSE, MAE)

<b>CNN</b>	Convolutional Neural Network
<b>RNN</b>	Recurrent Neural network
<b>PDF</b>	Probability Density Function
<b>RE</b>	Relative Error
<b>RMSE</b>	Root Mean Squared Error
<b>MAE</b>	Mean Absolute Error
<b>DWT</b>	Discrete Wavelet Transform
<b>IDWT</b>	Inverse Discrete Wavelet Transform
<b>NoL</b>	Number of layers
<b>HD</b>	Hidden Dimensions
<b>LR</b>	Learning rate

## ACKNOWLEDGMENT

The research reported in this publication was supported by the Division of Research and Innovation at San José State University under Award Number 23-UGA-08-044. The

- System Health Management Conference (PHM-Qingdao)*, 2019, pp. 1–6. [Online]. Available: <https://ieeexplore.ieee.org/document/8942856>
- [15] *LITHIUM-ION BATTERY REMAINING USEFUL LIFE ESTIMATION BASED ON ENSEMBLE LEARNING WITH LS-SVM ALGORITHM*, 2017, pp. 217–232. [Online]. Available: <https://ieeexplore.ieee.org/document/7656791>
  - [16] “A novel multistage support vector machine based approach for li ion battery remaining useful life estimation,” *Applied Energy*, vol. 159, pp. 285–297, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261915010557>
  - [17] “A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems,” *Renewable and Sustainable Energy Reviews*, vol. 131, p. 110015, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1364032120303063>
  - [18] Y. Zhang, R. Xiong, H. He, and Z. Liu, “A lstm-rnn method for the lithium-ion battery remaining useful life prediction,” in *2017 Prognostics and System Health Management Conference (PHM-Harbin)*, 2017, pp. 1–4. [Online]. Available: <https://ieeexplore.ieee.org/document/8079316>
  - [19] L. Ren, L. Zhao, S. Hong, S. Zhao, H. Wang, and L. Zhang, “Remaining useful life prediction for lithium-ion battery: A deep learning approach,” *IEEE Access*, vol. 6, pp. 50 587–50 598, 2018. [Online]. Available: <https://ieeexplore.ieee.org/document/8418374>
  - [20] Y. Liu, G. Zhao, and X. Peng, “Deep learning prognostics for lithium-ion battery based on ensembled long short-term memory networks,” *IEEE Access*, vol. 7, pp. 155 130–155 142, 2019. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8815721>
  - [21] S. Zhang, B. Zhai, X. Guo, K. Wang, N. Peng, and X. Zhang, “Synchronous estimation of state of health and remaining useful lifetime for lithium-ion battery using the incremental capacity and artificial neural networks,” *Journal of Energy Storage*, vol. 26, p. 100951, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352152X19307340>
  - [22] D. Zhou, Z. Li, J. Zhu, H. Zhang, and L. Hou, “State of health monitoring and remaining useful life prediction of lithium-ion batteries based on temporal convolutional network,” *IEEE Access*, vol. 8, pp. 53 307–53 320, 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9037249>
  - [23] L. Ren, J. Dong, X. Wang, Z. Meng, L. Zhao, and M. J. Deen, “A data-driven auto-cnn-lstm prediction model for lithium-ion battery remaining useful life,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3478–3487, 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9137406>
  - [24] “State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network,” *Journal of Power Sources*, vol. 459, p. 228069, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378775320303724>
  - [25] R. Kress, *Tikhonov Regularization*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1989, pp. 243–258. [Online]. Available: [https://doi.org/10.1007/978-3-642-97146-4\\_16](https://doi.org/10.1007/978-3-642-97146-4_16)
  - [26] C. Kulkarni and A. Guarneros, “Small satellite power simulation data set,” NASA Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA, 2008. [Online]. Available: <https://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository>
  - [27] B. Saha and K. Goebel, “Uncertainty management for diagnostics and prognostics of batteries using bayesian techniques,” 04 2008, pp. 1 – 8. [Online]. Available: [https://www.researchgate.net/publication/224314763\\_Uncertainty\\_Management\\_for\\_Diagnostics\\_and\\_Prognostics\\_of\\_Batteries\\_using\\_Bayesian\\_Techniques](https://www.researchgate.net/publication/224314763_Uncertainty_Management_for_Diagnostics_and_Prognostics_of_Batteries_using_Bayesian_Techniques)
  - [28] Y. Xing, E. W. Ma, K.-L. Tsui, and M. Pecht, “An ensemble model for predicting the remaining useful performance of lithium-ion batteries,” *Microelectronics Reliability*, vol. 53, no. 6, pp. 811–820, 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0026271412005227>
  - [29] D. Cho and T. D. Bui, “Multivariate statistical modeling for image denoising using wavelet transforms,” *Signal Processing: Image Communication*, vol. 20, no. 1, pp. 77–89, 2005. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0923596504000979>