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

Gaurav Shinde *, Rohan Mohapatra †, Pooja Krishan †, and Saptarshi Sengupta †
*Department of Engineering, San José State University, San Jose, CA, USA
†Department of Computer Science, San José State University, San Jose, CA, USA
Email: *gauravyeshwant.shinde@sjsu.edu, †rohan.mohapatra@sjsu.edu,
†pooja.krishan@sjsu.edu, †saptarshi.sengupta@sjsu.edu

***Abstract*—The usage of Lithium-ion (Li-ion) batteries has gained widespread popularity across various industries, from powering portable electronic devices to propelling electric vehicles and supporting energy storage systems. A central challenge in Li-ion battery reliability lies in 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, a denoising auto-encoder and a wavelet denoiser are used to generate encoded/decomposed representations, which are subsequently processed through dedicated self-attention transformer encoders. After extensive experimentation on NASA and CALCE data, a broad spectrum of health indicator values are estimated under a set of diverse noise patterns. The reported error metrics on these data are on par with or better than the state-of-the-art 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) are the 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. But, Li-ion battery capacity can deteriorate with time, influenced by factors such as temperature, state of charge, cycling rate, and operating conditions, leading to reduced performance and potential failure. The battery capacity is a key health indicator, crucial for accurately forecasting the Remaining Useful Life (RUL) before reaching End of Life (EOL) when performance significantly degrades or rated capacity can no longer be sustained.

Prognostics and Health Management (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) [1], Recurrent Neural network (RNN) [2], and Long Short Term Memory (LSTM) [3] networks. These models, while valuable, often faced challenges in capturing long-term dependencies in sequential data and required manual feature engineering. Consequently, they are less flexible in handling diverse datasets and complex real-world scenarios.

Recent RUL prediction trends focus on attention-based mechanisms, known for autonomously capturing intricate temporal and spatial data dependencies, reducing the requirement for extensive feature engineering. This shift away from traditional approaches aims to boost prediction accuracy and adaptability in complex systems [4]. Chen et. al [5] explore the 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 measurement noise in practical scenarios may not adhere to only a Gaussian distribution as assumed here. To address this concern and ensure robust noise handling, this research proposes a novel approach. Diverse set of noise types are implemented 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. Then, a minimization layer which identifies the noise type that yields the minimum error value is introduced. This adaptive noise modeling approach bolsters the auto-encoder's capacity to capture the spectrum of non-Gaussian noise characteristics commonly encountered in battery data. Due to the denoising process, this method culminates in generating higher-quality data representations for subsequent analysis, where the 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:** This 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 noise levels and noise distributions, thus making the architec-

979-8-3503-2445-7/23/$31.00 ©2023 IEEE

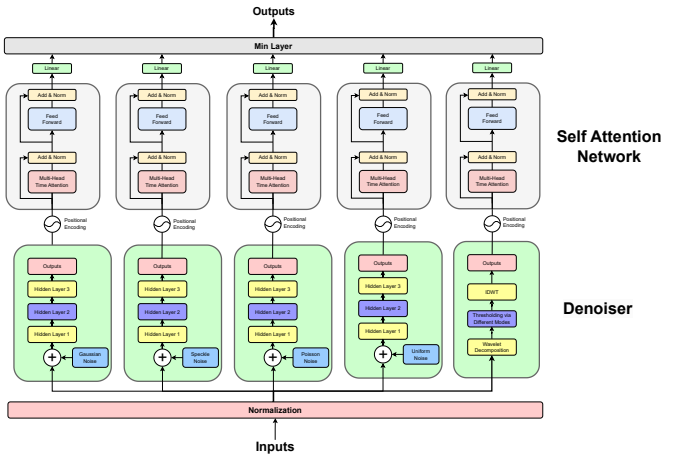


Fig. 1: The proposed **De-SaTE**: Denoising Self-attention Transformer Encoder architecture

ture more robust to fluctuations in the input.

3) **Better data representation leading to enhanced accuracy:** This architecture effectively denoises the input data and improves the quality of data representations leading to better predictions.
4) **A modular architecture for all complex processing:** The proposed architecture processes the input by passing it through the denoising modules to encode various types of noisy data, and the self-attention encoder network subsequently learns the degradation physics and predicts the remaining useful life.

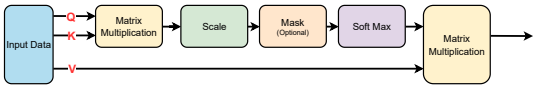
II. Self-Attention with Variable Denoising

In the input data $x = \{x_1, x_2, \ldots, x_n\}$, where $x \in (0, 1]$. A normalized input is produced, $x' = \frac{x}{C_0}$ 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. The data passes through the self-attention layers, each one intricately connected to its respective noise reducer, which then yields individual metric values such as Relative Error (RE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). A minimization strategy is employed, which selects the minimum value among these metrics, thereby obtaining the optimal error estimate. This architecture ensures robust performance and is 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.

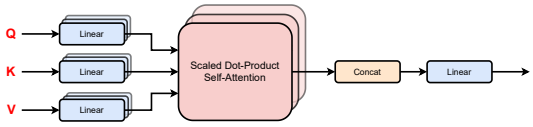
A. Positional Encoding

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

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



(a) Scaled Dot-Product Self-Attention mechanism



(b) Multi-Head Self-Attention mechanism

Fig. 2: Multi-Head Self-Attention architecture

B. Self Attention

The encoder's self-attention mechanism [6] 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 queries, keys and values as shown Fig. 2a.

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK'}{\sqrt{d_k}})V \tag{3}$$

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 and d_k is a scaling factor to stabilize the gradients.

C. Multi-head Attention

In Fig. 2b, the multi-head attention process is outlined. The self-attention mechanism is often used in multiple heads [6] to capture different types of dependencies:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, ..., \text{head}_h) \cdot W^O \tag{4}$$

Here, head_i represents the output of the i-th attention head and W^O is a learnable weight matrix.

D. Feed forward network

After computing attentions, the output passes through a Feed-Forward Neural network:

$$\text{FFN}(x) = \text{ReLU}(W_1 \cdot x + b_1) \cdot W_2 + b_2 \tag{5}$$

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

E. Learning

The proposed architecture can be divided into two tasks: *denoising* and *metric 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 L [5] is defined as follows:.

$$L = \sum_{t=T+1}^{n} (x_t - \hat{x_t})^2 + \delta \sum_{i=1}^{n} \ell(x_{corr} - \hat{x_i}) + \alpha\psi(L_{rate}) \tag{6}$$

where, n is the number of samples, δ controls relative contribution of each task, $\ell(\cdot)$ is the loss function, α is a

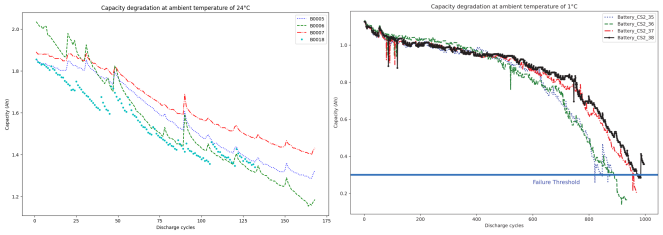
regularization parameter, $\psi(\cdot)$ denotes regularization and L_{rate} denotes learning parameters.

The denoising effect along with penalized loss acts like a regularizer. Regularization techniques [7] add a penalty term to the loss function, boosting model performance to tend towards smaller weights or simpler representations.

III. Experimental Setup

A. Dataset Description

Two datasets from National Aeronautics and Space Administration (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 of records from four different Li-ion batteries (B0005, B0006, B0007 and B0018), each subjected to three distinct operations: charging, discharging, and impedance measurements [8], [9]. The CALCE dataset (CS2_35, CS2_36, CS2_37 and CS2_38) is sourced from the Center for Advanced Life Cycle Engineering (CALCE) at the University of Maryland [10]. Figures 3a and 3b illustrate the capacity degradation trends observed across various batteries in these datasets.



(a) Degradation trend on the NASA dataset

(b) Degradation trend on the CALCE dataset

Fig. 3: Capacity vs. degradation cycles

B. Noise distributions

1) Gaussian Noise

Gaussian noise, characterized by mean (μ) and standard deviation (σ) has the Probability Density Function (PDF):

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{7}$$

2) Speckle Noise

Speckle noise is often multiplicative, where input values are multiplied by random values. Its PDF is:

$$f(x;\gamma) = \frac{1}{\gamma^2} e^{-\frac{x}{\gamma}} \tag{8}$$

where γ is a parameter controlling the noise intensity.

3) Poisson Noise

Poisson noise is characterized by its mean (λ). Poisson noise manifests when events occur at a consistent average rate but with randomness in the exact timing or occurrence of these events. This noise, explained as instrumentation noise in battery health prognosis has the PDF:

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \tag{9}$$

4) Uniform Noise

Uniform noise characterized by a minimum value (a) and a maximum value (b), has the PDF:

$$f(x;a,b) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

C. Denoising Autoencoder

A denoising autoencoder is effective in removing noise from data and learning robust representations. The network is trained to reconstruct clean data from noisy input. It consists of an encoder that maps the input data to a latent representation and a decoder 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 defined in Eqn. 6 typically measures reconstruction error, encouraging the network to minimize differences between clean inputs and reconstructed outputs.

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

DAE serves two purposes: denoising the raw input and learning a nonlinear representation [5]:

$$z = a\left(W \cdot x_{corr} + b\right) \tag{11}$$

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

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

$$\hat{x}_t = f_0\left(W_0 \cdot z + b_0\right) \tag{12}$$

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.

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

$$L_d = \frac{1}{n}\sum_{t=1}^{n}\left(x_{corr} - \hat{x}_t\right)^2 + \alpha\left(\|W\|_F^2 + \|W_0\|_F^2\right) \tag{13}$$

where $\|\cdot\|_F$ is the Fibonacci-norm, α is the regularization parameter, and n is the number of samples.

D. Wavelet Transformation and Denoiser

Wavelet denoising is a signal processing technique for removing noise from a voltage signal coming from a battery or another source. Wavelet denoising [11] typically involves thresholding coefficients obtained from wavelet transforms. A typical transformation involves passing the signal through Discrete Wavelet Transform (DWT), thresholding the wavelet coefficients. The denoised signal is reconstructed using the inverse DWT. The process is outlined below.

1) DWT

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

$$\theta = \text{DWT}(I) \tag{14}$$

I is the original signal, θ contains wavelet coefficients.

2) Thresholding

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

$$\hat{\theta}_{i,j} = \text{sign}(\theta_{i,j}) \cdot \max\left(|\theta_{i,j}| - \epsilon, 0\right) \tag{15}$$

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

$$\hat{\theta}_{i,j} = \begin{cases} \theta_{i,j}, & \text{if } |\theta_{i,j}| \geq \epsilon \\ 0, & \text{if } |\theta_{i,j}| < \epsilon \end{cases} \tag{16}$$

An additional thresholding method, *Garrote*, is a variation penalizing large coefficients than smaller ones and is given as:

$$\hat{\theta}_{i,j} = \frac{\text{sign}(\theta_{i,j}) \cdot \max\left(|\theta_{i,j}| - \epsilon, 0\right)}{1 + \frac{\epsilon}{|\theta_{i,j}|}} \tag{17}$$

where, $\hat{\theta}_{i,j}$ is the denoised coefficient, $\theta_{i,j}$ is the original coefficient, and ϵ is the threshold value.

3) Inverse Discrete Wavelet Transform (IDWT)

IDWT reconstructs signals using this transformation:

$$\hat{I} = \text{IDWT}(\hat{\theta}) \tag{18}$$

$\hat{I}$ is the denoised signal, $\hat{\theta}$ contains the denoised wavelet coefficients.

E. Training and Evaluation

Four types of noise were introduced: Gaussian, Speckle, Poisson, and Uniform with varying noise levels (small, medium, and relatively high). The optimal hyperparameters are evaluated 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

The experiments were run on a system using Python 3.10, Tensorflow 2.0 and Keras with Nvidia A100 and Nvidia T4 GPUs.

The models are evaluated using three key metrics outlined in Appendix B. This systematic exploration is aimed to identify the best hyperparameters for accurate predictions amidst diverse noise distributions, ensuring robustness and scalability.

IV. Results

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

The network performed optimally on the NASA dataset with a Learning rate (LR) of 0.01, Number of layers (NoL) of 1, 16 Hidden Dimensions (HD), and an α of 1e-05. In contrast, the optimal parameters for the CALCE dataset involve a LR of 0.001, a NoL of 1, 32 HD, and an α of 0.01.

TABLE I: Results for the NASA dataset

Noise and metrics		LR	NoL	HD	α	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**

TABLE II: Results for the CALCE dataset

Noise and metrics		LR	NoL	HD	α	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

These configurations were fine-tuned for the auto-encoder and transformer encoder layers via a grid search.

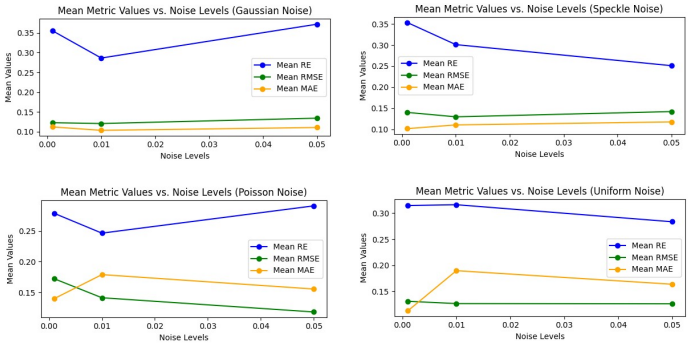
RE is highly related to the RUL of a battery, and serves as the primary evaluation metric. After thorough experimentation, it is observed that RE achieved superior results when paired with a denoising autoencoder followed by a transformer encoder. The three wavelet denoising modes - Soft, Hard, and Garrote - each with three distinct thresholds (0.001, 0.01, and 0.05) are explored to comprehensively assess their impact on the overall performance. Results are tabulated in Table III.

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

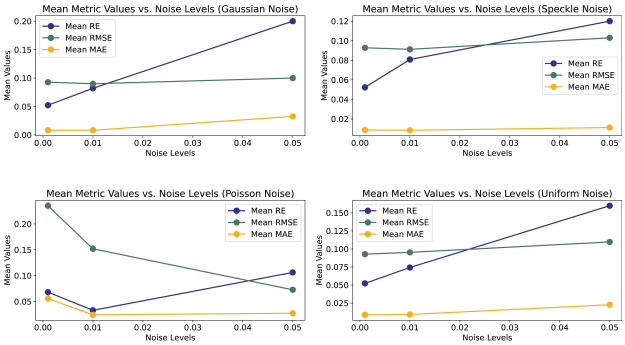
Dataset	Wavelet Denoising Mode	Threshold		
		0.001	0.01	0.05
NASA	Soft	0.29	0.31	0.52
	Hard	0.213	0.12	0.17
	Garotte	0.24	0.27	0.22
CALCE	Soft	0.76	0.78	0.81
	Hard	0.56	0.61	0.62
	Garotte	0.65	0.67	0.71

V. Conclusion and Future Work

This 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 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

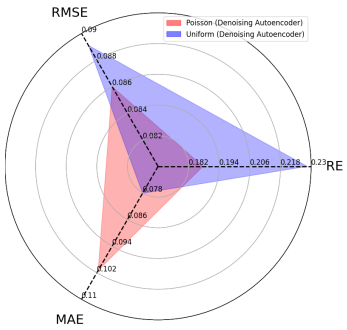


(a) Mean Relative Error (RE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) values on the NASA data under different types of noise with varying levels

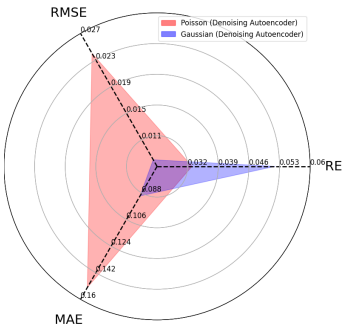


(b) Mean RE, RMSE, MAE values on the CALCE data under different types of noise with varying levels

Fig. 4: Comparison of metrics under diverse types and levels of noise



(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. 5: Effects of different noise on metrics (RE, RMSE, MAE)

TABLE IV: Model Evaluation on Li-ion Battery Datasets

Dataset	Model	Metrics		
		RE	**MAE**	**RMSE**
NASA	MLP [12]	0.3851	0.1379	0.1541
	RNN [2]	0.2851	0.0749	0.0848
	LSTM [3]	0.2648	0.0829	0.0905
	GRU [13]	0.3044	0.0806	0.0921
	Dual-LSTM [14]	0.2557	0.0815	0.0879
	DeTransformer [5]	0.2252	**0.0713**	0.0802
	De-SaTE	**0.1674**	0.0806	**0.0781**
CALCE	MLP [12]	0.4018	0.1557	0.2038
	RNN [2]	0.1614	0.0938	0.1099
	LSTM [3]	0.0902	0.0582	0.0736
	GRU [13]	0.1319	0.0671	0.0946
	Dual-LSTM [14]	0.0885	0.0636	0.0874
	DeTransformer [5]	0.0764	0.0613	**0.0705**
	De-SaTE	**0.0330**	**0.0080**	0.090

over the other noises for the CALCE dataset. However, Gaussian noise yields enhanced performance across RE and MAE for the NASA dataset. The proposed architecture produces lower RE, MAE, and RMSE compared to past work.

In future, the response of the proposed architecture to adversarial attacks may be proposed and defense strategies may be devised accordingly, thereby adding to its robustness. As a next step, a more traditional encoder-decoder model might be introduced to extend the present capabilities and advance the understanding of battery health prognostics.

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 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. Gao, X. Liu, Z. Zhu, and Q. Yang, "A hybrid cnn-bilstm approach for remaining useful life prediction of evs lithium-ion battery," *Measurement and Control*, vol. 56, no. 1-2, pp. 371–383, 2023.

[2] J. Liu, A. Saxena, K. Goebel, B. Saha, and W. Wang, "An adaptive recurrent neural network for remaining useful life prediction of lithium-ion batteries," *Annual Conference of the PHM Society*, vol. 2, no. 1, 2010. [Online]. Available: https://papers.phmsociety.org/index.php/phmconf/article/view/1896

[3] Y. Zhang, R. Xiong, H. He, and M. G. Pecht, "Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 7, pp. 5695–5705, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8289406

[4] P. Khumprom and N. Yodo, "A data-driven predictive prognostic model for lithium-ion batteries based on a deep learning algorithm," *Energies*, vol. 12, no. 4, 2019. [Online]. Available: https://www.mdpi.com/1996-1073/12/4/660

[5] 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

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[7] 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

[8] 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

[9] 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

[10] 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

[11] 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

[12] Y. Wu, W. Li, Y. Wang, and K. Zhang, "Remaining useful life prediction of lithium-ion batteries using neural network and bat-based particle filter," *IEEE Access*, vol. 7, pp. 54 843–54 854, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8698866

[13] B. Xiao, Y. Liu, and B. Xiao, "Accurate state-of-charge estimation approach for lithium-ion batteries by gated recurrent unit with ensemble optimizer," *IEEE Access*, vol. PP, pp. 1–1, 04 2019. [Online]. Available: https://www.researchgate.net/publication/332650682_Accurate_State-of-Charge_Estimation_Approach_for_Lithium-Ion_Batteries_by_Gated_Recurrent_Unit_With_Ensemble_Optimizer

[14] Z. Shi and A. Chehade, "A dual-lstm framework combining change point detection and remaining useful life prediction," *Reliability Engineering & System Safety*, vol. 205, p. 107257, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0951832020307572

[15] 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

[16] 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

[17] 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

[18] 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

[19] "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

[20] 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

[21] 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

[22] "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

[23] 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

[24] 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

[25] 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

[26] 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

[27] 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 System Health Management Conference (PHM-Qingdao)*, 2019, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/8942856

[28] *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

[29] "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

[30] "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

[31] Y. Zhang, R. Xiong, H. He, and Z. Liu, "A lstm-rnn method for the lithuim-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

[32] 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

[33] 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

[34] "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

[35] 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

[36] 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

APPENDIX A
RELATED WORK

Li-ion batteries are a ubiquitous power source in industries like electric cars, drone technology, and various other applications, offering efficient energy storage. Battery prognosis has seen rising trends with the widespread use of Li-ion batteries in the industry. Recent literature [4] has seen drift from model-based techniques to deep-learning based models.

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 [15]–[17]. However, these methods require accurate knowledge of the battery's internal structure and operating conditions, which can be challenging to obtain in practice.

1) Physical Model

Physical model quantifies the factors that influence a battery's performance. This approach usually focuses on the specific physical and chemical phenomena occurring during utilization [18]. 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 [19]. However, due to the complexity and non-linearity 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 for the filter to converge to the optimal state [20], [21]. Multiple research studies have demonstrated that adaptive filters function well in RUL estimation of Li-ion batteries.

4) Stochastic Process Methods

Stochastic process methods are based on the notion that battery degradation is a stochastic process that can be modeled using probabilistic methods. The advantage of stochastic 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 [22].

B. Data Driven Methods

Data-driven methods rely extensively on analyzing the battery's operating data to estimate its degradation level and predict when it will reach the end of its useful life. Such a method can directly mime the degradation information of lithium-ion battery through historical data, and there is no need to establish a specific mathematical model [23], [24].

1) Traditional Machine Learning

This section provides an overview of traditional machine learning methods for RUL estimation of Li-ion batteries. The techniques that utilize Relevance Vector Machine (RVM) [25]–[27] achieve efficient online training for updating the model using battery data. However, they pose a memory consumption issue with increased model complexity. Multiple ensemble models [27]–[30] have been proposed to improve prediction accuracy. 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. LSTM models and their variants [31]–[34] can extract multi-dimensional features and estimate RUL with high precision. Zhang et al. [35] 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. Temporal Convolution Network (TCN) [36] that uses causal and dilated convolution techniques to capture local capacity regeneration has improved prediction accuracy. 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 Li-ion batteries has undergone a remarkable transformation with the emergence of transformer-based models [5]. 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 long-term dependencies within data, endowing them with the capacity to model the intricate and evolving degradation patterns of Li-ion batteries with precision.

APPENDIX B
EVALUATION METRICS

A. Relative Error (RE):

Relative Error measures the relative difference between predicted and actual values and is represented as follows:

$$RE = \frac{|Y - \hat{Y}|}{|Y|} \tag{19}$$

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

B. Root Mean Square Error (RMSE):

RMSE calculates the square root of the mean of the squared differences between predicted and actual values:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2} \tag{20}$$

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.

C. Mean Absolute Error (MAE):

MAE calculates the mean of the absolute differences between predicted and actual values:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{Y}_i - Y_i| \tag{21}$$

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.

APPENDIX C
HIGH-LEVEL OVERVIEW OF THE EXPERIMENTAL FRAMEWORK

The initial phase of the experimental framework is shown in Fig. 6 in which the input data is standardized to establish a uniform scale, ensuring coherence. Following that, the denoising modules are used, which are designed to refine the data by filtering out extraneous noise, thereby improving the dataset's integrity. The transformer-encoder component extracts intricate features from refined data and computes critical metrics like RE, MAE, and RMSE. This sequential protocol, encapsulates the series of operations that form the foundation of our experimental pipeline.

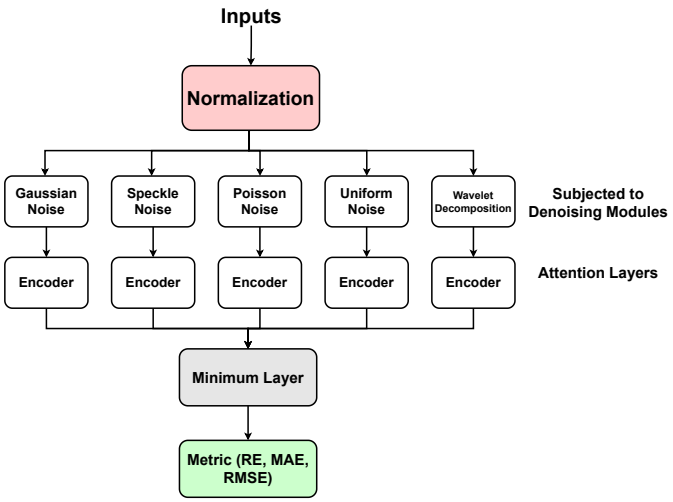


Fig. 6: High-level overview of the experimental framework

APPENDIX D
ABBREVIATIONS

Below we list all the abbreviations used in this paper.

Li-ion	Lithium-ion
RUL	Remaining Useful Life
PHM	Prognostics and Health Management
EOL	End of Life
RVM	Relevance Vector Machine
SOH	State of Health
NASA	National Aeronautics and Space Administration
CALCE	Center for Advanced Life Cycle Engineering
LSTM	Long Short Term Memory
ANN	Artificial Neural Network
TCN	Temporal Convolution Network
CNN	Convolutional Neural Network
RNN	Recurrent Neural network
PDF	Probability Density Function
RE	Relative Error
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
DWT	Discrete Wavelet Transform
IDWT	Inverse Discrete Wavelet Transform
NoL	Number of layers
HD	Hidden Dimensions
LR	Learning rate