State of Health Estimation of Batteries Using a Time-Informed Dynamic Sequence-Inverted Transformer

Janak M. Patel[1], Milad Ramezankhani[2], Anirudh Deodhar[3], and Dagnachew Birru[4]

[1,2,3,4] *Phi Labs, Quantiphi,*
Marlborough, MA 01752, USA.
janak.mpatel@quantiphi.com
milad.ramezankhani@quantiphi.com
anirudh.deodhar@quantiphi.com
dagnachew.birru@quantiphi.com

ABSTRACT

The rapid adoption of battery-powered vehicles and energy storage systems over the past decade has made battery health monitoring increasingly critical. Batteries play a central role in the efficiency and safety of these systems, yet they inevitably degrade over time due to repeated charge discharge cycles. This degradation leads to reduced energy efficiency and potential overheating, posing significant safety concerns. Accurate estimation of a battery's State of Health (SoH) is therefore essential for ensuring operational reliability and safety. Several machine learning architectures, such as LSTMs, transformers, and encoder-based models, have been proposed to estimate SoH from discharge cycle data. However, these models struggle with the irregularities inherent in real world measurements: discharge readings are often recorded at non uniform intervals, and the lengths of discharge cycles vary significantly. To address this, most existing approaches extract features from the sequences rather than processing them in full, which introduces information loss and compromises accuracy. To overcome these challenges, we propose a novel architecture: Time Informed Dynamic Sequence Inverted Transformer (TIDSIT). TIDSIT incorporates continuous time embeddings to effectively represent irregularly sampled data and utilizes padded sequences with temporal attention mechanisms to manage variable-length inputs without discarding sequence information. Experimental results on the NASA battery degradation dataset demonstrate that TIDSIT significantly outperforms existing models, achieving over a 43% reduction in prediction error compared to the strongest baseline. The model attains an $\mathrm{RMSE}\% = 0.58\%$, indicating highly accurate SoH estimation. Furthermore, the architecture exhibits strong cross-battery generalization and shows promise for broader health monitoring tasks involving irregular time-series data.

https://doi.org/10.36001/IJPHM.2026.v17i1.4729

1. INTRODUCTION

The global transition toward clean energy and decarbonization is accelerating the adoption of electric vehicles (EVs) and large-scale battery energy storage systems. In response, automotive manufacturers are rapidly advancing electrification efforts, with several regions mandating that all new vehicles be battery-electric by 2035 to meet stringent emissions targets (Swarnkar, Ramachandran, Ali, & Jabbar, 2023). At the core of these technologies lie lithium-ion batteries, praised for their high energy density and efficiency (Lu, Han, Li, Hua, & Ouyang, 2013). Despite their advantages, lithium-ion batteries naturally degrade over time as a result of repeated charge–discharge cycles. This degradation leads to diminished capacity, reduced efficiency, and heightened risks of thermal instability or failure (Belt, Utgikar, & Bloom, 2011). The consequences can be significant, ranging from safety hazards and unexpected system downtime to expensive battery replacements (Cabrera-Castillo, Niedermeier, & Jossen, 2016). Consequently, accurate estimation of a battery's *State of Health* (SoH) is critical for ensuring operational reliability, safety, and cost-effective lifecycle management (Venugopal, 2019).

In recent years, data-driven and machine learning (ML) based approaches have gained significant traction for SoH estimation due to their ability to capture complex, nonlinear degradation patterns from historical data. Traditional SoH prediction methods typically rely on trends in discharge capacity to estimate the current life of a battery (Richardson, Osborne, & Howey, 2017; D. Liu et al., 2014). However, discharge capacity alone may not sufficiently reflect internal degradation, particularly under diverse operational conditions. Consequently, more recent methods leverage multivariate time-series signals, such as voltage, current, and temperature profiles during discharge cycles to enhance predictive accuracy and model robustness

(Saha & Goebel, 2007, 2008). These richer input features enable the development of more adaptive and generalizable models across a wide range of usage scenarios. Deep learning architectures such as Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and Support Vector Regression (SVR) have been widely adopted for SoH estimation due to their ability to capture temporal dependencies and nonlinear degradation trends in battery data (Li & Chen, 2025). Additionally, deep autoencoders have been employed to extract low-dimensional representations from high-dimensional cycle data while preserving degradation related information (Xu, Guo, & Saleh, 2022). Recently, transformer-based models, initially developed for natural language processing have gained popularity in time-series applications, owing to expressiveness and scalability of their attention mechanisms (Vaswani et al., 2017). In the context of battery SoH prediction, a hybrid architecture combining convolutional neural networks (CNNs) with transformers has been proposed to enhance predictive performance (Gu et al., 2023). A transformer-based model incorporating a dedicated feature extraction mechanism for SoH prediction has also been developed (Luo, Zheng, & Shi, 2023). Furthermore, inverted Transformer (iTransformer) has been applied to battery SoH prediction, where variable-specific embeddings are used instead of conventional temporal encodings, enabling richer representation of multivariate data (Guirguis, Abdulmaksoud, Ismail, Kollmeyer, & Ahmed, 2024).

Despite the advancements in data-driven SoH estimation, existing models rely heavily on extensive feature engineering or sequence preprocessing to manage the practical challenges inherent in real-world battery data, such as *irregular sampling intervals* and *variable discharge cycle lengths* resulting from aging and inconsistent usage patterns as illustrated in Figure 1. To standardize these inputs, conventional methods typically employ interpolation or extract handcrafted statistical features from raw discharge sequences (Xu et al., 2022; Guirguis et al., 2024; Li & Chen, 2025). While such preprocessing simplifies the modeling pipeline, it often compromises temporal fidelity and may obscure subtle yet informative degradation patterns. Recent efforts have begun to move away from handcrafted features, leveraging deep learning models that can directly capture temporal dependencies. For instance, recurrent architectures like LSTMs have been used to process segments of discharge cycles while retaining sequential structure (Van & Quang, 2023). However, these models still depend on fixed-length input windows, typically derived from truncated or sampled portions of the full cycle and therefore fail to exploit the complete discharge sequence in its raw, variable-length form. This limitation can result in the loss of long-range temporal dependencies and non-uniform degradation signals critical for accurate SoH estimation. These challenges underscore the need for a modeling framework capable of directly processing raw, irregularly sampled, and variable-length time series, while preserving the full temporal structure and improving both the robustness and generalizability of SoH prediction.

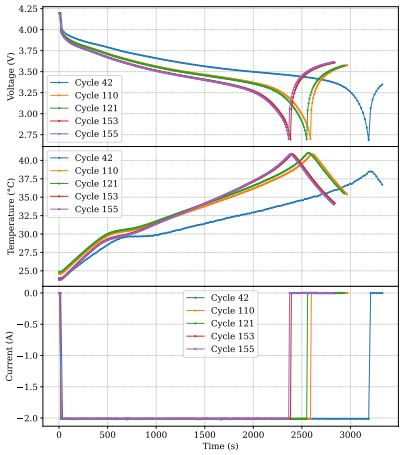


Figure 1. Visualization of five randomly selected discharge cycles showing raw input features : voltage (V), temperature (°C), and current (A) plotted against time. The figure highlights two key characteristics of real-world battery data: (1) *irregular sampling intervals* within each cycle, and (2) *variable sequence lengths* across cycles. Notably, later cycles become progressively shorter, illustrating the effect of battery degradation on discharge duration over time.

To overcome these critical limitations, we propose a novel deep learning architecture: the **Time-Informed Dynamic Sequence Inverted Transformer (TIDSIT)**. TIDSIT is designed from the ground up to directly address the challenges posed by *irregularly sampled* and *variable-length* time-series data in battery health monitoring, eliminating the need for lossy feature extraction or sequence truncation. Unlike conventional approaches that operate on preprocessed or fixed-length segments, TIDSIT processes entire discharge sequences in their native form, preserving complete temporal dynamics and enabling more faithful modeling of the battery's degradation trajectory. To effectively handle irregular sampling, TIDSIT incorporates *continuous-time embeddings*, inspired by recent work in time-series forecasting (Kim & Lee, 2024). This embedding mechanism explicitly models the non-uniform intervals between observations, allowing the model to leverage temporal information. Additionally, TIDSIT incorporates data variate embeddings to explicitly model multivariate sensor inputs such as voltage, current, and temperature, allowing the network to capture variable-specific temporal dynamics. This structured representation facilitates a more expressive encoding of battery behavior over time and is inspired by recent advances in multivariate time-series forecasting (Y. Liu et al., 2023). To accommodate variable-length discharge cycles, TIDSIT utilizes padded sequences along with a temporal attention mechanism that dynamically attends over all valid timestamps. By integrating these components within a transformer-based architecture, TIDSIT enables direct, end-to-end processing of

raw discharge data, achieving accurate and generalizable SoH estimation without reliance on hand-engineered features. **The key contributions of this work are as follows:**

- We present **TIDSIT**, a novel transformer-based architecture specifically designed for battery SoH estimation from raw discharge cycle data with irregular sampling intervals and variable sequence lengths.
- We address key limitations in existing SoH estimation approaches by introducing a suite of architectural innovations that enable robust learning from irregular, multivariate time series with variable sequence lengths. These include: (i) a **continuous-time embedding mechanism** to encode non-uniform temporal structures; (ii) **data variate embeddings** to capture feature-wise relationships among voltage, current, and temperature; and (iii) a **temporal attention mechanism** that operates on padded sequences to preserve full temporal resolution across variable-length discharge cycles.

2. METHODOLOGY

2.1. Problem Specification

The primary objective of this work is to develop a robust and generalizable deep learning model for estimating the *SoH* of lithium-ion batteries, based on multivariate time-series data collected during individual discharge cycles. SoH is a key indicator of battery condition, commonly defined as the ratio between the current discharge capacity and the nominal (rated) capacity of the battery (Guo, Qiu, Hou, Liaw, & Zhang, 2014):

$$\text{SoH}(t) = \frac{C_{\text{current}}(t)}{C_{\text{rated}}} \tag{1}$$

where $C_{\text{current}}(t)$ represents the discharge capacity measured at cycle t, and C_{rated} denotes the capacity of a new, fully functional battery. The goal is to predict the SoH at each cycle using only the sensor signals recorded during that specific discharge cycle without relying on information from future degradation cycles. Each discharge cycle is represented as a multivariate, irregularly sampled time series:

$$\mathcal{X}^{(i)} = \{(x_1, \tau_1), (x_2, \tau_2), \ldots, (x_{T_i}, \tau_{T_i})\} \tag{2}$$

where $\mathcal{X}^{(i)}$ is the i-th discharge cycle, $x_j \in \mathbb{R}^d$ is a d-dimensional vector comprising sensor readings (e.g., voltage, current, temperature) recorded at time τ_j, and T_i is the number of observations in cycle i. Due to practical constraints in battery usage and asynchronous data logging, both the sequence length T_i and the time intervals $\Delta\tau_j = \tau_j - \tau_{j-1}$ vary significantly across different cycles. The corresponding label for each cycle is the scalar SoH value:

$$y^{(i)} = \text{SoH}^{(i)} \in (0, 1] \tag{3}$$

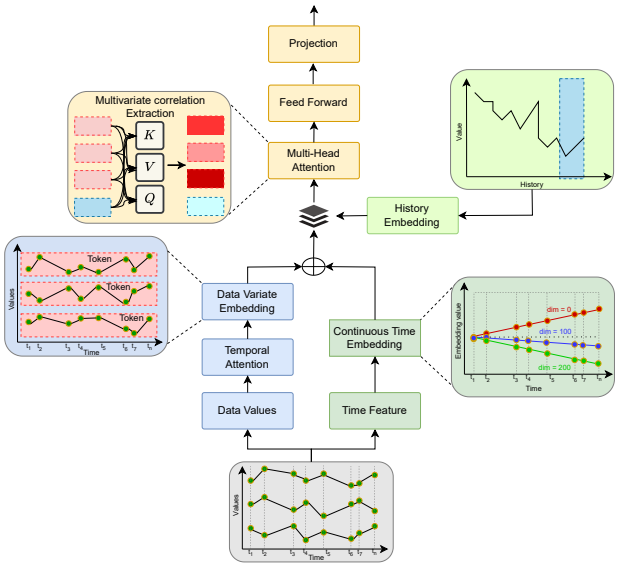


Figure 2. Schematic of the proposed **Time-Informed Dynamic Sequence Inverted Transformer (TIDSIT)** architecture for battery SoH estimation. Each input consists of a multivariate, irregularly sampled discharge cycle, represented as a sequence of sensor readings: voltage, current, and temperature along with corresponding timestamps. To accommodate varying sequence lengths, raw features are first passed through a *temporal attention block* with padding masks, allowing the model to handle variable-length inputs without information loss. The output is processed by a *data variate embedding* layer that learns a distinct representation for each sensor channel. In parallel, *continuous-time embeddings* are computed based on the non-uniform timestamps and added to the variable-specific representations, preserving fine-grained temporal structure. A learned *SoH history embedding* derived from the SoH values of the past few discharge cycles is then concatenated with the fused output. The resulting enriched sequence is passed through a transformer encoder where multi-head self-attention captures both intra-step interactions and long-range dependencies. Finally, the output is aggregated and passed through a feed-forward network and a projection head to yield the predicted scalar SoH for the current cycle.

and the learning task is to identify a function f_θ parameterized by θ, that maps each variable-length, irregularly sampled sequence to a scalar estimate of SoH:

$$\hat{y}^{(i)} = f_\theta(\mathcal{X}^{(i)}) \tag{4}$$

such that the prediction $\hat{y}^{(i)}$ closely approximates the ground truth label $y^{(i)}$ for all discharge cycles in the dataset.

2.2. Time-Informed Dynamic Sequence Inverted Transformer (TIDSIT) Architecture

To address the challenges posed by irregular sampling and variable-length inputs, we propose the **TIDSIT**. As illustrated

in Figure 2, TIDSIT is a transformer-based architecture that processes raw multivariate discharge sequences, comprising asynchronous voltage, current, and temperature readings without the need for truncation, interpolation, or handcrafted features. TIDSIT is built upon three core components: (1) a *continuous-time embedding* that encodes irregular timestamp information; (2) a *data variate embedding module* that learns sensor-wise latent representations; and (3) a *temporal attention block* that mitigates the effects of sequence padding while capturing temporal dependencies within features.

In addition, we introduce an optional *SoH history embedding* module that incorporates contextual degradation information from prior cycles when such estimates are available. These representations are fused and passed through a transformer encoder, which models intra-cycle sensor dynamics while optionally leveraging inter-cycle degradation context to generate accurate SoH estimates across diverse batteries and usage profiles.

Temporal Attention. To effectively process sequences of varying lengths, we pad each input to a fixed length T, where T denotes the maximum sequence length observed across the dataset. Let each original discharge cycle be represented as $X_o^{(i)} = [x_1, x_2, \ldots, x_{T_i}] \in \mathbb{R}^{T_i \times d}$, where T_i is the number of the number of observations of i-th cycle, and d is the number of sensor features (e.g., voltage, current, temperature). Sequences shorter than T are end-padded with a sentinel value of -1, resulting in inputs of shape $\mathbb{R}^{T \times d}$. Since all sensor readings are normalized to the range $[0, 1]$, this padding value is clearly distinguishable and does not conflict with valid measurements. To prevent these artificial values from influencing the learning process, we introduce a *temporal attention block* that applies a binary padding mask. This mask ensures that attention scores involving padded positions are excluded from computation, allowing subsequent layers to ignore non-informative inputs. Additionally, this block captures temporal dependencies across the valid (unpadded) portions of each sequence before any data variate embedding is applied, enabling early reasoning about discharge dynamics. Given the padded input sequence $X^{(i)} = [x_1, x_2, \ldots, x_T] \in \mathbb{R}^{T \times d}$, the scaled dot-product attention mechanism is computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}} + M\right)V \quad (5)$$

where $Q = XW_Q$, $K = XW_K$, $V = XW_V$, and $W_Q, W_K, W_V \in \mathbb{R}^{d \times d_k}$ are learnable projection matrices. The mask matrix $M \in \mathbb{R}^{T \times T}$ assigns large negative values to padded positions to eliminate their influence during attention computation (Devlin, Chang, Lee, & Toutanova, 2019). Multiple attention heads are combined as:

$$\text{MultiHead}(X) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W_O \quad (6)$$

where $W_O \in \mathbb{R}^{(h \cdot d_k) \times d}$ is a learnable output projection matrix that maps the concatenated output of all attention heads back to the original model dimension d. This is followed by a residual connection and layer normalization to stabilize training and improve gradient flow (Vaswani et al., 2017):

$$Z = \text{LayerNorm}(X + \text{Dropout}(\text{MultiHead}(X))) \quad (7)$$

The final output of this block, denoted $Z^{(i)} \in \mathbb{R}^{T \times d}$, is a temporally attended sequence retaining the original time step structure and feature dimensionality. This serves as the input to the data variate embedding module.

Data Variate Embedding. To capture inter-variable correlations across the entire discharge cycle, we implement a data variate embedding strategy that operates on the output of the temporal attention block (Y. Liu et al., 2023). Let the attended sequence output from the previous block be denoted as $Z^{(i)} \in \mathbb{R}^{T \times d}$. We transpose this to obtain $Z_{\text{var}}^{(i)} \in \mathbb{R}^{d \times T}$, where each row $z_{1:T}^{(v)} \in \mathbb{R}^T$ corresponds to the full time series of the v-th variable. Each variable sequence is then independently projected into a shared latent space:

$$z^{(v)} = z_{1:T}^{(v)} W^{(v)} + b^{(v)}, \quad \text{for } v = 1, \ldots, d. \quad (8)$$

where $W^{(v)} \in \mathbb{R}^{T \times h}$ and $b^{(v)} \in \mathbb{R}^h$ are learnable parameters. The set of embeddings $\{z^{(1)}, \ldots, z^{(d)}\}$ is stacked as:

$$Z^{\text{var}} = [z^{(1)}; z^{(2)}; \ldots; z^{(d)}] \in \mathbb{R}^{d \times h}. \quad (9)$$

enabling the model to learn high-level representations for each sensor signal that preserve its unique temporal behavior. This transposition fundamentally changes how the attention mechanism operates. In conventional Transformers, tokens correspond to time steps, and attention models temporal relationships between observations. In contrast, the inverted representation treats each sensor variable as a token summarizing the full temporal evolution of that variable across the discharge cycle. Consequently, the self-attention mechanism learns relationships between sensor channels rather than between individual timestamps. This design provides two advantages. First, it allows the model to capture cross-sensor dependencies (e.g., interactions between voltage decay patterns and temperature fluctuations) that are indicative of battery degradation. Second, it reduces the effective sequence length from T time steps to d variables, which significantly lowers the computational complexity of the attention operation from $\mathcal{O}(T^2)$ to $\mathcal{O}(d^2)$. Since d is typically small (three sensor channels in this work), this representation improves computational efficiency while still preserving the complete temporal profile of each variable.

Continuous-Time Embedding. To incorporate fine-grained temporal structure from irregularly sampled sequences, we apply a continuous-time embedding scheme inspired by (Kim & Lee, 2024). Given the timestamp vector $\boldsymbol{\tau}^{(i)} = [\tau_1, \tau_2, \ldots, \tau_T]$, we normalize it to $[0, 1]$. The normalized vector is projected:

$$E^{\text{time}} = \boldsymbol{\tau} W_{\text{time}} + b_{\text{time}}, \quad E^{\text{time}} \in \mathbb{R}^{1\times h}. \tag{10}$$

with $W_{\text{time}} \in \mathbb{R}^{T\times h}$ and $b_{\text{time}} \in \mathbb{R}^h$. To retain information about the temporal ordering and relative timing of events within each irregularly sampled sequence analogous to positional encoding in large language models, we add the continuous-time embedding to the data variate embedding:

$$Z^{\text{fused}} = Z^{\text{var}} + E^{\text{time}} \tag{11}$$

The continuous-time embedding provides the model with explicit information about the relative timing of measurements. Unlike standard positional encoding that assume uniform sampling, battery discharge measurements are recorded at irregular intervals due to asynchronous sensing and logging processes. By projecting the normalized timestamp vector into the latent space, the model learns a representation that reflects both the ordering of observations and the temporal spacing between them. This enables the transformer to distinguish between sequences that share similar sensor values but occur over different temporal dynamics, which is particularly important for capturing degradation patterns that unfold at varying rates across discharge cycles.

SoH History Embedding. To optionally incorporate degradation information from previous discharge cycles, we maintain a fixed-length vector of past SoH values $[y^{(i-1)}, \ldots, y^{(i-p)}] \in \mathbb{R}^p$, where p denotes the history window size. These values may correspond either to previously estimated SoH values available from the battery management system or to predictions generated by the model in an online inference setting. The history vector is mapped via a learned projection:

$$E^{\text{hist}} = [y^{(i-1)}, \ldots, y^{(i-p)}] W_{\text{hist}} + b_{\text{hist}}, \quad E^{\text{hist}} \in \mathbb{R}^{1\times h} \tag{12}$$

where $W_{\text{hist}} \in \mathbb{R}^{p\times h}$ and $b_{\text{hist}} \in \mathbb{R}^h$ are learnable parameters. The final sequence representation passed to the transformer encoder is formed by concatenating the fused embedding Z^{fused} with the history embedding E^{hist} along the temporal dimension:

$$Z^{\text{input}} = [Z^{\text{fused}}; E^{\text{hist}}] \in \mathbb{R}^{(d+1)\times h} \tag{13}$$

When historical SoH estimates are unavailable, this module can be omitted without modifying the remaining architecture. In this case, the encoder operates solely on the fused sensor representations Z^{fused}, allowing the model to estimate SoH directly from the current discharge cycle.

Transformer Encoder and Output. The final input Z^{input} is processed by a Transformer encoder consisting of a multi-head self-attention layer followed by a position-wise feed-forward network. The encoded representation is then passed through a projection layer to yield the final SoH prediction $\hat{y}^{(i)}$.

3. RESULTS AND DISCUSSION

3.1. Datasets and Benchmark

3.1.1. Dataset Description

We evaluate the proposed TIDSIT architecture using the battery degradation dataset provided by the NASA Ames Prognostics Center of Excellence (PCoE) (Saha & Goebel, 2007). This dataset consists of sensor measurements collected from lithium-ion 18650 battery cells undergoing repeated charge-discharge cycles under controlled aging protocols and varying ambient temperatures. For each discharge cycle, the dataset provides multivariate time-series data including voltage, current, and temperature readings. These three sensor channels: voltage (V), current (I), and temperature (T), serve as the input features to our model. The target output is the battery's remaining capacity, measured at the end of each discharge cycle, which is used to compute the ground-truth SoH. During the aging process, periodic capacity tests were conducted to track the progressive degradation of each battery. The end-of-life (EOL) criterion is defined as a 30% reduction in capacity from the nominal rating (i.e., from 2 Ah to 1.4 Ah). Consequently, each cycle is represented as a variable-length multivariate sequence of [V, I, T] signals along with a single scalar SoH label.

To assess the model's generalization capability, we follow a cross-battery evaluation setup. Specifically, data from batteries B0005 and B0006 are used for training, while battery B0007 is reserved for testing. This ensures that the model is evaluated on entirely unseen degradation patterns, simulating a realistic deployment scenario. This dataset presents a challenging setting for data-driven SoH estimation, characterized by irregular sampling intervals, varying sequence lengths, and heterogeneous degradation trajectories, making it well-suited for benchmarking the robustness of the proposed TIDSIT architecture.

Data Preprocessing and Normalization. The raw battery discharge data consist of multiple cycles with variable-length sensor measurements recorded during each discharge process.

Table 1. Summary of benchmark model input representations and preprocessing pipelines.

Model	Input Type	Preprocessing	Sequence Handling
FNN	Time-series sequence	Subsequence extraction	Fixed length
LSTM (Fixed-range)	Time-series sequence	Window truncation	Fixed length
LSTM (Feature-based)	Feature vector	Handcrafted feature extraction	Fixed vector
i-Transformer (Feature Extraction)	Feature vector	Handcrafted feature extraction	Fixed vector
TIDSIT (Ours)	Time-series sequence	Padding + masking	Variable length

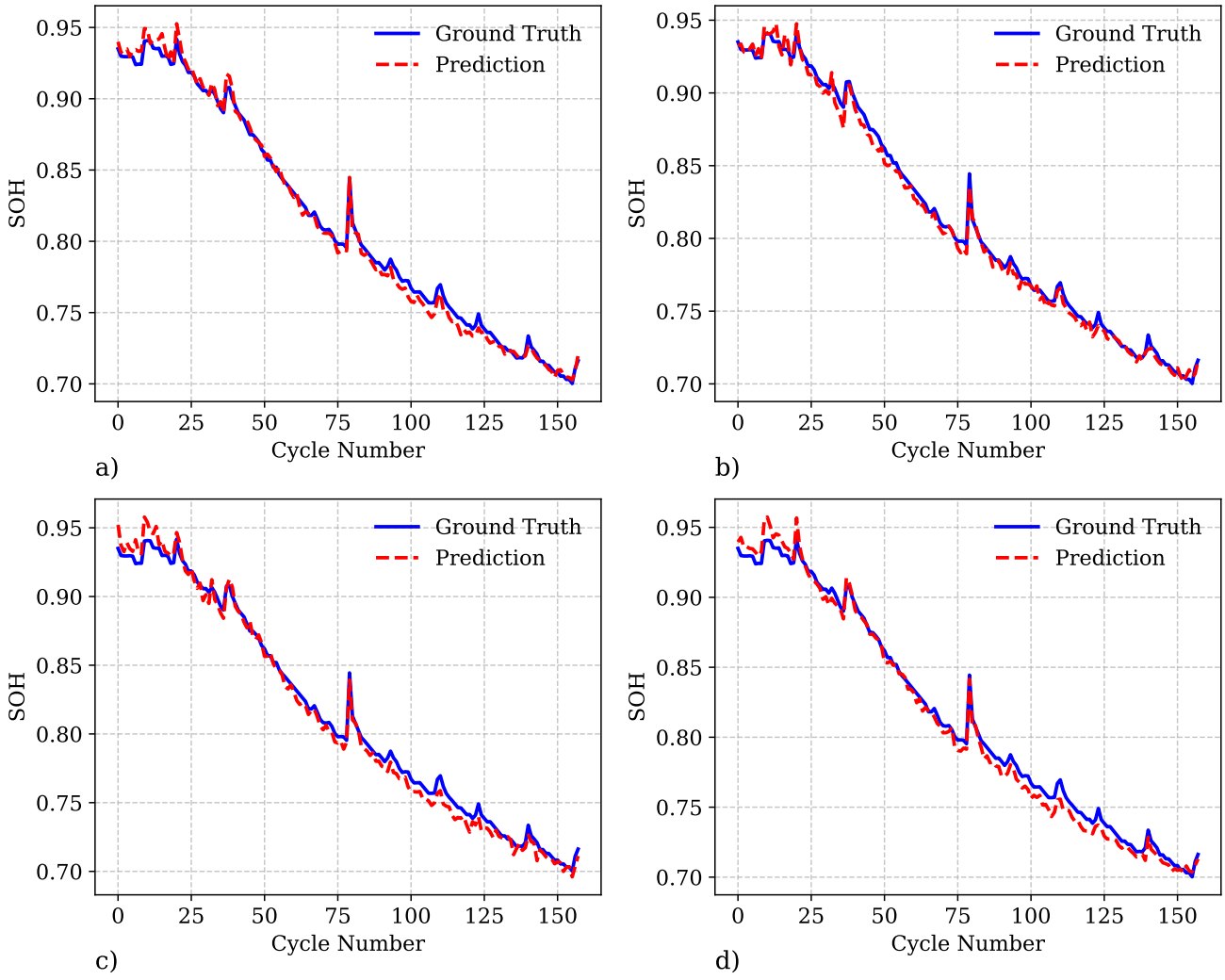


Figure 3. Predicted versus ground truth SoH values for the B0007 battery (unseen during training). a)–d) compare the SoH prediction performance of different transformer variants: a) TIDSIT, b) Reformer, c) Flowformer, and d) Informer. The figure demonstrates the model's generalization capability by accurately tracking the degradation trend on unseen test data.TIDSIT demonstrates the best alignment with the true degradation trajectory, indicating superior generalization to real-world battery behavior with irregular sampling and variable cycle lengths.

Table 2. Comparison of SoH estimation performance on the B0007 test battery. Reported metrics are taken from the corresponding referenced works unless otherwise stated. Entries marked "N/A" indicate that the metric was not reported in the original publication.

Model	RMSE	RMSE%
FNN	0.032	N/A
LSTM (Fixed-range)	0.0082	N/A
LSTM (Feature-based)	0.0094	N/A
i-Transformer (Feature-based)	N/A	1.47
TIDSIT (Ours)	0.0047	0.58

Table 3. Comparison of SoH estimation performance using different Transformer encoder variants. Only the encoder block is modified, while all other components of TIDSIT are retained.

Transformer Encoder	RMSE	RMSE %
Reformer	0.0050	0.61
Informer	0.0071	0.89
Flowformer	0.0065	0.80
TIDSIT	0.0047	0.58

For each cycle i, we construct a multivariate sequence

$$X^{(i)} = \{x_1^{(i)}, x_2^{(i)}, ..., x_{T_i}^{(i)}\}, \quad x_t^{(i)} \in \mathbb{R}^C \tag{14}$$

where T_i denotes the number of measurements in cycle i and $C = 4$ corresponds to the sensor channels: voltage, temperature, current, and timestamp. Measurements are grouped by cycle index, and each discharge cycle is treated as one input sequence with an associated cycle-level capacity value used to compute the SOH target. Since discharge cycles have variable lengths, all sequences are padded to a fixed maximum length T_{max} using a sentinel value ($-\infty$). Padding is applied at the end of the sequence (post-padding). The initial sentinel value is chosen so that padded elements are excluded from the normalization statistics. Sensor measurements (voltage, temperature, and current) are normalized using min–max scaling:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min} + \epsilon} \tag{15}$$

where x_{min} and x_{max} are computed globally from the training data across all training batteries to prevent information leakage. The same normalization statistics are then applied to the test batteries. After normalization, padded values are converted from $-\infty$ to a fixed sentinel value of -1. This allows the temporal attention module to explicitly identify padded positions while ensuring they were not included in the normalization process.

3.1.2. Benchmark Models

To assess the performance of the proposed TIDSIT architecture, we compare it against a range of baseline models that represent both conventional and recent approaches to SoH estimation:

- **Feedforward Neural Network (FNN):** A fully connected neural network trained on fixed-length subsequences extracted from the discharge data, rather than processing the entire variable-length sequence (Van & Quang, 2023). Each sensor signal (voltage, current, and temperature) is divided into $M = 35$ equal temporal ranges per cycle. The values within each range are averaged and normalized to $[0, 1]$ using min-max scaling. The resulting features from all three sensors are concatenated to form a $3M$-dimensional input vector per cycle. This representation removes temporal ordering and summarizes the discharge profile as a fixed-length vector (Van & Quang, 2023).
- **LSTM (Fixed-range):** A Long Short-Term Memory network trained on fixed-length subsequences extracted from the discharge data, rather than processing the entire variable-length sequence (Van & Quang, 2023). This model uses the same range-based representation as the FNN, where each cycle is converted into a sequence of M averaged segments per sensor. The concatenated sequence of size $3M$ is treated as a fixed-length time-series input to the LSTM. Temporal dynamics are modeled only across these coarse-grained segments rather than the original high-resolution sequence (Van & Quang, 2023).
- **LSTM (Feature Extraction):** A feature-driven approach where handcrafted health indicators (HIs) are extracted from voltage, current, and temperature profiles and used as inputs to an LSTM model for SoH prediction. These features capture key degradation characteristics, including peak temperature, time-to-peak, average charge/discharge temperature, voltage differences, current slopes, and energy-related measures, resulting in an initial set of 16 candidate features. Pearson correlation analysis is applied to select the most informative features. Detailed definitions and extraction procedures for these features can be found in (Li & Chen, 2025).
- **i-Transformer (Feature Extraction):** A transformer-based model employing an inverted embedding scheme that operates on extracted statistical features rather than raw time-series data. Specifically, two features are computed from the first 100 time steps of each discharge cycle: (i) voltage range ($\Delta V = V_{\max} - V_{\min}$) and (ii) variance of temperature. These features are used as input tokens to the model, following the design described in (Guirguis et al., 2024).

Table 1 summarizes the input representation, preprocessing steps, and sequence handling strategy for each benchmark model, highlighting the differences between feature-based approaches and sequence-based methods. For all baselines, the input representations and preprocessing pipelines strictly

Table 4. Ablation study showing the impact of removing each architectural component from the full TIDSIT model on SoH estimation performance. "w/o" indicates that the corresponding component is removed from the full model configuration.

Configuration	History Embedding	Temporal Attention	Continuous-Time Embedding	Variate Embedding	RMSE	RMSE%
Full Model (TIDSIT)	✓	✓	✓	✓	**0.0047**	**0.58**
w/o Variate Embedding	✓	✓	✓	×	0.0058	0.71
w/o History Embedding	×	✓	✓	✓	0.0065	0.81
w/o Temporal Attention	✓	×	✓	✓	0.0190	2.52
w/o Continuous-Time Embedding	✓	✓	×	✓	0.0369	4.79

follow the methodologies described in their respective referenced works, with no modifications to feature design or data processing.

3.1.3. Evaluation Metrics

Following prior work in battery SoH estimation, we evaluate model performance using the Root Mean Squared Error (RMSE) and the Root Mean Squared Percentage Error (RMSE%). The RMSE measures the absolute prediction error and is defined as

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y^{(i)} - \hat{y}^{(i)}\right)^2} \tag{16}$$

where $y^{(i)}$ denotes the ground-truth SoH for cycle i, $\hat{y}^{(i)}$ denotes the predicted SoH, and N is the total number of evaluation samples. To enable scale-independent comparison with existing studies, we also report the Root Mean Squared Percentage Error (RMSE%), defined as

$$RMSE\% = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y^{(i)} - \hat{y}^{(i)}}{y^{(i)}}\right)^2} \times 100 \tag{17}$$

This metric expresses the prediction error as a percentage relative to the true SoH value, allowing consistent comparison across datasets and prior literature. Together, RMSE and RMSE% provide complementary insights into model performance. RMSE captures the absolute prediction error in capacity units, which is directly relevant for practical battery management applications. In contrast, RMSE% provides a normalized, scale-independent measure that facilitates comparison across batteries with different degradation levels. As a result, these metrics provide a balanced and sufficient evaluation of both absolute accuracy and relative performance.

3.2. Model Hyperparameters and Training Protocol

The hyperparameters of TIDSIT were selected based on a combination of preliminary experiments and established design practices for transformer-based time-series models, with the objective of balancing model capacity and computational efficiency. Due to the limited number of batteries in the dataset, we do not employ a dedicated validation split. Instead, a small number of exploratory runs were conducted to identify stable hyperparameter settings, which were then fixed and used consistently across all experiments. We set the model's hidden dimension to 42, with a single encoder layer and 8 attention heads to balance representational capacity and computational efficiency. A dropout rate of 0.1 is used for regularization, and the feed-forward network dimension is set to 168 (four times the hidden dimension), following standard transformer design principles. ReLU is used as the activation function. To accommodate variable-length discharge cycles within a batch, we pad all sequences to a maximum length of 371, corresponding to the longest sequence in the dataset, and use masking to ignore padded positions during attention computation. The model is trained on these padded sequences with a past SoH history window of size 10 provided as additional input context. In the reported experiments, the SoH history embedding uses previous ground-truth SoH values from the dataset. Training is performed using a batch size of 16.

For training, we use the Adam optimizer with an initial learning rate of 1×10^{-4}. The model is trained for 1200 epochs with a step learning rate scheduler that reduces the learning rate every 200 epochs with a decay factor $\gamma = 0.05$. All experiments were conducted on a standard laptop equipped with an 11th Gen Intel(R) Core(TM) i5-1135G7 CPU running at 2.40GHz and 8 GB RAM. No GPU acceleration was used during training or evaluation. Despite the limited computational resources, the proposed TIDSIT model achieved efficient training times, underscoring its practicality for lightweight deployment.

3.3. Results

To evaluate the effectiveness of the proposed TIDSIT architecture, we assess its performance on the B0007 battery dataset, which was entirely held out during training. Table **??** reports the SoH estimation accuracy of TIDSIT compared to several benchmark models, including FNN (Van & Quang, 2023), LSTM (fixed-range (Van & Quang, 2023) and feature-based variants (Li & Chen, 2025)), and the i-Transformer(Feature Extraction) (Guirguis et al., 2024). For several baseline methods, only one evaluation metric was reported in the original reference, hence the missing metric reported as N/A. TID-

SIT achieves an RMSE of 0.0047, which marks a significant improvement over all baselines. Notably, it outperforms the best baseline LSTM (Fixed-range), which records an RMSE of 0.0082. This corresponds to more than a **50% reduction in prediction error**, highlighting TIDSIT's superior ability to learn directly from raw, irregularly sampled time series. The FNN baseline performs worst, with an RMSE of 0.032, underscoring the limitations of static input modeling. In terms of RMSE percentage, TIDSIT achieves RMSE% = 0.58%, outperforming the i-Transformer(Feature Extraction) (RMSE%=1.47%), despite the latter relying on extracted statistical features. This highlights TIDSIT's strong generalization capability without requiring handcrafted features or preprocessing. Figure 3 illustrates the predicted versus ground truth SoH values for the B0007 battery. The predictions closely follow the true degradation trajectory, confirming TIDSIT's capacity to generalize to unseen battery instances and accurately capture both long-term trends and short-term variations.

3.4. Comparison with Transformer Variants

To isolate the contribution of the encoder architecture in SoH estimation, we perform a comparative analysis by replacing the Transformer encoder block in TIDSIT with several well-known Transformer variants designed for long-sequence modeling: **Reformer** (Kitaev, Kaiser, & Levskaya, 2020), **Informer** (Zhou et al., 2021), and **Flowformer** (Huang et al., 2022). These variants were selected because they represent different approaches for improving Transformer efficiency and scalability for long time-series sequences, including sparse attention (Reformer), probabilistic sparse attention (Informer), and flow-based attention mechanisms (Flowformer). This comparison allows us to examine potential trade-offs between computational efficiency and modeling accuracy in the context of battery discharge sequences. Importantly, all other components of the TIDSIT pipeline such as the temporal attention block, data variate embedding, continuous-time embedding, and SoH history embedding are retained unchanged across all variants. This design choice ensures that the comparison focuses solely on the encoder's ability to capture intra-cycle and inter-cycle dependencies.

Table 3 summarizes the SoH prediction performance of these encoder variants on the B0007 test battery. While Reformer, Informer, and Flowformer show competitive results, TIDSIT outperforms them all, achieving the lowest RMSE of 0.0047 and the lowest RMSE%=0.58%. These results suggest that, for the moderately sized discharge sequences considered in this study (maximum length of 371), standard attention remains sufficient for capturing temporal dependencies. While Reformer, Informer, and Flowformer are designed to improve computational efficiency for very long sequences, the benefits of sparse or linear attention mechanisms are less pronounced at this sequence scale.

3.5. Ablation Study

To evaluate the contribution of each architectural component in TIDSIT, we conduct an ablation study by systematically removing one module at a time and measuring the resulting SoH prediction performance. The modules evaluated include: (1) the *Temporal Attention Block*, which handles variable-length sequences and captures time-step dependencies; (2) the *Continuous-Time Embedding*, which encodes irregular sampling intervals; (3) the *Data Variate Embedding*, which enables variable-wise representation learning; and (4) the *SoH History Embedding*, which incorporates contextual information from previous cycles.

Table 4 reports the RMSE and RMSE percentage on the unseen B0007 test battery for each variant. The full TIDSIT model achieves the best performance, with an RMSE of 0.0047 and RMSE%=0.58%. Removing any single component leads to a noticeable drop in performance, highlighting the importance of each module. Among all components, removing the *Continuous-Time Embedding* causes the most significant degradation (RMSE = 0.0369), underscoring its critical role in modeling temporal irregularity inherent in real-world battery data. The absence of the *Temporal Attention Block* also results in a substantial error increase (RMSE = 0.0190), reflecting its importance in handling padded inputs and capturing intra-cycle dependencies early in the pipeline. Excluding the *Data Variate Embedding* leads to a moderate increase in error (RMSE = 0.0058), but also introduces a large computational overhead training time increases from 40.4 minutes (with variate embedding) to 154.4 minutes (without it). This is because, without variate embedding, the model must process each time step as an independent token, significantly increasing the sequence length and computational cost. This highlights the dual benefit of the variate embedding: improved performance and substantially faster training. Interestingly, the removal of the *SoH History Embedding* causes only a minor performance drop (RMSE = 0.0065), suggesting that while historical degradation trends provide complementary context, the model predominantly learns to infer SoH from the current discharge sequence itself. This indicates that TIDSIT is not simply correlating past SoH values but is effectively learning degradation dynamics directly from raw sensor inputs.

3.6. Availability of SoH History

The SoH history embedding incorporates the previous p cycle-level SoH values $[y^{(i-1)}, \ldots, y^{(i-p)}] \in \mathbb{R}^p$ to provide the model with additional degradation context. In practical battery management systems, SoH is typically estimated periodically through capacity tests or model-based estimators rather than being directly measured at every cycle. Therefore, during deployment, these historical SoH values may correspond either to previously estimated SoH values available from the system or to the model's own predictions from earlier cycles in an

online inference setting. Notably, our ablation results indicate that removing the SoH History Embedding results in only a minor performance drop (RMSE = 0.0065), suggesting that TIDSIT predominantly learns to infer SoH from the current discharge sequence itself, while the historical SoH information provides only complementary context. This indicates that the proposed model does not rely on ground-truth SoH labels at inference time.

3.7. Additional Cross-Battery Evaluation

To further evaluate the robustness of the proposed architecture, we conducted additional cross-battery experiments by varying the training and test battery combinations. In the primary setup used throughout this study, the model is trained on batteries B0005 and B0006 and evaluated on the unseen battery B0007. To assess generalization under different training configurations, we performed two additional experiments. When training on batteries B0006 and B0007 and testing on B0005, TIDSIT achieved an RMSE%=1.02%. In another configuration, where the model was trained on B0005 and B0007 and evaluated on B0006, TIDSIT obtained an RMSE%=2.35%.

For comparison, we refer to the reported results of the i-Transformer baseline from prior literature for the B0006 test configuration, where an RMSE%=3.8% is reported. Results for other train–test battery combinations were not available in the referenced study; therefore, the comparison is restricted to this shared evaluation setting. Within this configuration, the proposed TIDSIT architecture achieves lower error, indicating improved cross-battery generalization.

4. CONCLUSION

In this work, we introduced the **Time-Informed Dynamic Sequence Inverted Transformer (TIDSIT)**, a novel transformer-based architecture for accurate and generalizable estimation of battery State of Health (SoH) using raw multivariate discharge sequences. By effectively handling irregular sampling, variable-length sequences, and limited degradation history, TIDSIT enables end-to-end learning without the need for handcrafted feature extraction or sequence truncation. Our experiments demonstrate that TIDSIT significantly outperforms traditional models such as feedforward neural networks and LSTMs, as well as recent transformer variants, achieving more than 50% reduction in prediction error on unseen test data. Furthermore, ablation studies highlight the critical importance of each architectural component including temporal attention, continuous-time embedding, data variate embedding, and SoH history context in achieving robust SoH predictions. Overall, TIDSIT establishes a strong foundation for reliable battery health monitoring and paves the way for future research into scalable, interpretable, and real-time prognostics using deep sequence models.

ACKNOWLEDGMENT

The authors would like to sincerely thank Quantiphi for their financial support of this research.

REFERENCES

Belt, J., Utgikar, V., & Bloom, I. (2011). Calendar and phev cycle life aging of high-energy, lithium-ion cells containing blended spinel and layered-oxide cathodes. *Journal of Power Sources*, *196*(23), 10213–10221.

Cabrera-Castillo, E., Niedermeier, F., & Jossen, A. (2016). Calculation of the state of safety (sos) for lithium ion batteries. *Journal of Power Sources*, *324*, 509–520.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171–4186).

Gu, X., See, K. W., Li, P., Shan, K., Wang, Y., Zhao, L., ... Zhang, N. (2023). A novel state-of-health estimation for the lithium-ion battery using a convolutional neural network and transformer model. *Energy*, *262*, 125501.

Guirguis, J., Abdulmaksoud, A., Ismail, M., Kollmeyer, P. J., & Ahmed, R. (2024). Transformer-based deep learning strategies for lithium-ion batteries sox estimation using regular and inverted embedding. *IEEE Access*.

Guo, Z., Qiu, X., Hou, G., Liaw, B. Y., & Zhang, C. (2014). State of health estimation for lithium ion batteries based on charging curves. *Journal of Power Sources*, *249*, 457–462.

Huang, Z., Shi, X., Zhang, C., Wang, Q., Cheung, K. C., Qin, H., ... Li, H. (2022). Flowformer: A transformer architecture for optical flow. In *European conference on computer vision* (pp. 668–685).

Kim, B., & Lee, J.-G. (2024). Continuous-time linear positional embedding for irregular time series forecasting. *arXiv preprint arXiv:2409.20092*.

Kitaev, N., Kaiser, Ł., & Levskaya, A. (2020). Reformer: The efficient transformer. *arXiv preprint arXiv:2001.04451*.

Li, K., & Chen, X. (2025). Machine learning-based lithium battery state of health prediction research. *Applied Sciences*, *15*(2), 516.

Liu, D., Luo, Y., Liu, J., Peng, Y., Guo, L., & Pecht, M. (2014). Lithium-ion battery remaining useful life estimation based on fusion nonlinear degradation ar model and rpf algorithm. *Neural Computing and Applications*, *25*, 557–572.

Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., & Long, M. (2023). itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*.

Lu, L., Han, X., Li, J., Hua, J., & Ouyang, M. (2013). A review on the key issues for lithium-ion battery management in electric vehicles. *Journal of power sources*, *226*, 272–288.

Luo, K., Zheng, H., & Shi, Z. (2023). A simple feature extraction method for estimating the whole life cycle state of health of lithium-ion batteries using transformer-based neural network. *Journal of Power Sources*, *576*, 233139.

Richardson, R. R., Osborne, M. A., & Howey, D. A. (2017). Gaussian process regression for forecasting battery state of health. *Journal of Power Sources*, *357*, 209–219.

Saha, B., & Goebel, K. (2007). Battery data set. *NASA AMES prognostics data repository*.

Saha, B., & Goebel, K. (2008). Uncertainty management for diagnostics and prognostics of batteries using bayesian techniques. In *2008 ieee aerospace conference* (pp. 1–8).

Swarnkar, R., Ramachandran, H., Ali, S. H. M., & Jabbar, R. (2023). A systematic literature review of state of health and state of charge estimation methods for batteries used in electric vehicle applications. *World Electric Vehicle Journal*, *14*(9), 247.

Van, C. N., & Quang, D. T. (2023). Estimation of soh and internal resistances of lithium ion battery based on lstm network. *International Journal of Electrochemical Science*, *18*(6), 100166.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Venugopal, P. (2019). State-of-health estimation of li-ion batteries in electric vehicle using indrnn under variable load condition. *Energies*, *12*(22), 4338.

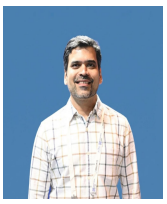
Xu, Z., Guo, Y., & Saleh, J. H. (2022). A physics-informed dynamic deep autoencoder for accurate state-of-health prediction of lithium-ion battery. *Neural Computing and Applications*, *34*(18), 15997–16017.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 35, pp. 11106–11115).

BIOGRAPHIES

Janak M. Patel completed his Master of Technology in Chemical Engineering from the Department of Chemical Engineering, IIT Bombay, India, in 2022. He is currently working as a Research Engineer at Phi Lab, Quantiphi, where he focuses on digital twin development for manufacturing and engineering applications. He has over three years of experience in applying machine learning and deep learning techniques and has filed several patents and published research papers in reputed conference proceedings. His current research interests include developing a novel graph and transformer based architectures for solving complex business problems.

Milad Ramezankhani is a Research Scientist at Quantiphi, with over seven years of experience advancing AI-driven solutions for complex scientific and engineering systems. He holds a Ph.D. in Mechanical Engineering from the University of British Columbia and has authored more than 24 peer-reviewed papers in top-tier journals and international conferences. At Phi Labs, he leads the development of next-generation scientific machine learning models and digital twins, bridging computational theory with real-world industrial applications. His work focuses on novel algorithms—including transformers, graph neural networks, diffusion models, and neural PDE solvers—to enable predictive modeling, control, and optimization of physical systems.



Anirudh Deodhar is a Principal Architect at Quantiphi, where he leverages over 15 years of expertise in process modeling, computational fluid dynamics, and ML/AI to build industrial digital twins. He has led the development and deployment of digital twin solutions across the energy, manufacturing, healthcare, and logistics sectors. Anirudh has filed 30+ patents and co-authored 25+ peer-reviewed papers. His contributions have been recognized with prestigious honors, including the Tata Innovista, the TCS Distinguished Engineer award, and the Quantiphi Annual Leadership award. His current research explores AI Physics, AI-driven Resource Scheduling, and Agentic AI. An active member of ASME, Anirudh holds a Master's in Mechanical Engineering from the University of Cincinnati. Outside of his professional life, he is a student of Indian history and an avid cricket fan.

Dagnachew Birru heads Quantiphi's R&D department and is driven by passion to improve the lives of people through data, AI and advanced digital technologies, solving practical problems, and creating business value. He holds a Ph.D. in Electronics Engineering and Signal Processing from Delft University of Technology. His background includes many years of experience as an idea-driven technology researcher, developer, and leader in systems employing AI/ML and smart algorithms, resulting in 50+ issued patents.