OPEN ACCESS
IOP Publishing | ISECAEP
Journal of Reliability Science and Engineering
J. Reliab. Sci. Eng. **2** (2026) 035301 (16pp)
https://doi.org/10.1088/3050-2454/ae89bb

# Battery voltage reconstruction via a generative model for state of health estimation

**Wei Wu, Zhen Chen*, Weijie Liu and Ershun Pan**

Department of Industrial Engineering and Management and the State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, 200240 Shanghai, People's Republic of China

E-mail: chenzhendr@sjtu.edu.cn

Received 1 April 2026, revised 19 June 2026
Accepted for publication 11 July 2026
Published 29 July 2026

**Abstract**

The quality and quantity of data influence the development and performance of advanced battery management algorithms. Existing methods that ignore the issue of incomplete data (especially in terms of voltage data) in real-world operational scenarios will limit their practical deployment and commercialization. In this study, a generative framework based on a Conditional Variational Autoencoder and a transformer structure is proposed. This framework leverages condition variables collected during battery degradation and reconstructs complete standard charging voltage curves from limited voltage segments. The battery's future state of health (SOH) can then be estimated with reconstructed charging-voltage data. By analyzing field data from an energy storage system, a simulation condition was designed to collect the degradation data. Experimental validation shows that the proposed framework can reconstruct the charging voltage profile and achieve accurate SOH estimation. Furthermore, testing under incomplete-data conditions shows that the reconstructed voltage profiles can provide useful information for online SOH estimation in battery management applications.

Keywords: lithium batteries, generative AI, SOH estimation, incomplete data, voltage reconstruction

(Some figures may appear in colour only in the online journal)

## 1. Introduction

The development of new energy storage technologies has exerted significant transformative impacts on the modern energy landscape. In the process of rapid growth in the new energy market, lithium-ion batteries hold a large market share. This has led to an urgent need for lithium-ion health management technology, with the aim of enabling high-safety and stable energy storage systems to achieve greater economic benefits. The explosive growth in artificial intelligence technology has driven continuous advancements in algorithms for battery health management [1]. Digital operation technology holds great potential in enhancing the operational efficiency and reliability of lithium-ion battery energy storage systems (BESSs) [2, 3].

A comprehensive battery dataset plays a crucial role in driving relevant research in both the academic and industrial fields. Among the most widely used publicly available datasets for battery research are those from NASA experiments. These key battery operation data, such as voltage, current, and temperature, are helpful for studies of battery degradation processes. The NASA dataset has been utilized for the development and validation of methods such as the Wavelet Packet Decomposition Convolutional Neural Network [4].

* Author to whom any correspondence should be addressed.

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.


© 2026 The Author(s). Published by IOP Publishing Ltd on behalf of the Institute of Systems Engineering, China Academy of Engineering Physics

The CALCE battery group has conducted cycling tests on a diverse set of LCO/graphite cells. Their degradation data was collected until the batteries reached at least their end-of-life threshold, defined as 80% of the rated capacity. A hybrid machine learning method was verified using this data and the NASA dataset, varying degradation profiles and testing conditions [5]. An improved vision transformer network has also been well validated on these two datasets [6]. In recent years, due to the growing interest and significant investment from numerous institutions, the development process of open-access battery datasets has accelerated. Notable examples include datasets from MIT [7], the University of Oxford [8] and Huazhong University of Science and Technology [9]. The availability of publicly accessible battery datasets addresses a major challenge in the field: obtaining high-quality degradation data through battery tests is costly and time-consuming. It can be seen that high-quality data has greatly facilitated the development of advanced battery management algorithms [10].

Many methods based on experimental data are still far from reaching the level of commercial application because they ignore various limitations of the actual operating conditions [11]. The battery management system is responsible for controlling the operation of the battery. The design of this system and its internal algorithms are highly dependent on accurate battery operation data [12]. These models must overcome numerous challenges arising from the fact that the real-world scenarios are different from the controlled conditions on which most experimental data are collected. At present, most of the methods have been developed in strictly controlled environments using relatively small but well-designed laboratory datasets [13]. In order to ensure the practical applicability of these methods, it is necessary to incorporate real-world data that can more comprehensively reflect the aging of the battery under actual operating conditions for the development of relevant algorithms [14]. So it can be seen that current research is increasingly focusing on actual vehicle operation data [15]. A comparison between laboratory data and real-world data reveals that the most significant difference lies in the absence of full discharge conditions in practical applications–understandably, batteries in real-life scenarios are rarely subjected to frequent full discharges due to safety, longevity, and operational constraints [16]. This discrepancy results in incomplete voltage data in real-world applications compared to laboratory datasets [17]. Consequently, many existing methods that have high requirements for data quality may fail when used in actual operating conditions [18]. Those methods that rely on complete charge–discharge curves for feature extraction or as direct model inputs are prone to be affected by this issue. In practical scenarios, the incompleteness of the voltage curve makes these methods inapplicable without appropriate modifications or data supplementation strategies [19].

Various approaches have been explored to develop battery health management algorithms to address the problem of insufficient or incomplete data [20]. Considering that it may not be possible to obtain data on the entire life cycle of batteries in practice, there are methods that attempt to establish prediction models using limited degradation data. A mechanism–data fusion model that needs only the ultra-early stage battery and the planned cycling protocol can forecast a battery's full-life-cycle capacity and energy curves [21]. DENG *et al* built a digital twin model of the battery pack to simulate its dynamic behavior under different degrees of aging and inconsistency and proposed a rapid health assessment method for in-use battery packs based on short-term charging data and limited labels [22]. There are also methods that focus on extracting key information from collected data [23]. The most well-known of these is health feature extraction, which establishes relationships between features and health status [24]. Li *et al* [25] combined electrochemical model aging features with multi-stage voltage and temperature-derived indicators to construct a machine learning model to improve the accuracy of battery's state of health (SOH) estimation. A novel health indicator called voltage variance during equal time interval is extracted by Liu and then applied to a hierarchical ensemble model for accurate SOH estimation [26]. Meanwhile, various data augmentation strategies have been proposed to enhance model robustness under data-scarce conditions [27–29]. The generative transfer-belief rule base method proposed by Kang, integrated data enhancement and expert knowledge reasoning to address the few-shot health condition estimation problem [30]. Approaches that directly develop models on sparse data often suffer from biased degradation estimation, since missing voltage segments typically correspond to nonlinear electrochemical phases that carry critical diagnostic information.

Although current methods have partially addressed the limitations of incomplete battery degradation data [31], the explicit reconstruction of missing voltage profiles under practical incomplete-data conditions remains insufficiently investigated [32]. Existing studies mainly focus on direct SOH estimation, feature extraction, or sample-level data augmentation, whereas sparse voltage segments may miss critical electrochemical information needed for reliable health assessment. To address the aforementioned issues, this study proposes a framework called Conditional Variational Autoencoder based on Transformer (CVAE-TRANS). The design concept of CVAE-TRANS is to conduct offline training using the benchmark data obtained under laboratory conditions, to learn the degraded features, and then use deep learning techniques to reconstruct voltages. The reconstructed voltage data are further used to support battery SOH estimation under incomplete-data scenarios. The contributions of this study are as follows:

- A CVAE-TRANS framework has been developed to address the challenges of incomplete and limited data. This framework is able to reconstruct a complete standard charging voltage curve with limited voltage segments.
- The reconstruction performance of the proposed framework is systematically validated under different charging/discharging protocols. Experimental results show that the proposed model can generate high-fidelity charging-voltage profiles from sparse inputs and outperforms representative generative and sequence-learning baselines.

- The applicability of the reconstructed voltage profiles to battery health assessment is demonstrated under non-full charging/discharging and field-data-inspired scenarios. By extracting health-related features from the reconstructed profiles, accurate SOH estimation can be achieved, indicating the potential of the proposed framework for incomplete-data battery health management.

The remainder of this paper is organized as follows. Section 2 introduces the field and experimental datasets; section 3 presents the proposed CVAE-TRANS framework and training strategy; section 4 reports the reconstruction, SOH estimation, online application, and sensitivity-analysis results; and section 5 concludes the paper.

## 2. Source of the data

This study provides battery datasets obtained from two distinct sources: field data and experimental data. The field data are acquired from an operational BESS in China. The experimental data are collected under three different operating conditions in our battery testing laboratory. Detailed descriptions of these two types of datasets will be given in the next subsections. Notably, field data differ significantly from laboratory data, and this distinction will be used to highlight existing research gaps. The battery test experiments will be introduced, along with the methodology for bridging these experimental results and field data to facilitate the development of intelligent estimation algorithms for energy storage systems.

### 2.1. Field data

The studied BESS is structured into multiple hierarchical levels ranging from smallest to largest: battery cells, modules, plug-in boxes, clusters, stacks, and container units. This complex hierarchical structure generates a large amount of data, which prompts the operators to reduce the sampling frequency in order to lower the storage costs. Events such as system upgrades or scheduled maintenance downtime frequently lead to data gaps throughout the BESS's lifecycle. Additionally, batteries are usually restricted to specific state-of-charge (SOC) ranges during charge and discharge cycling process to ensure long-term operational safety. All these factors result in various forms of data incompleteness in real-world conditions.

This paper aims to explore the strategies for addressing the issue of incomplete data by utilizing the existing limited datasets in practical scenarios. This paper collects data from a cluster in BESS as one of the research data. Specifically, the battery cluster comprises 17 series-connected plug-in boxes, each containing 12 battery modules in series connection, with each module consisting of 18 battery cells connected in parallel. The smallest unit of BESS is a 13Ah cylindrical LiFePO4 cell with a nominal voltage of 3.2 V and named 'IFR32135-13Ah'. Figure 1 presents six months of data collected from the energy storage system, with each data point corresponding to a 60 s interval. Figure 1(a) displays the voltage profile of a representative cluster within the BESS, which consists of 204 battery modules. Figure 1(b) provides a partial voltage curve graph for one of the modules. This offers a more detailed view at the module level. Figure 1(c) shows the charging voltage during the cycle of this cluster. The tail part of the charging voltage represents the relaxation stage. There was a significant drop in voltage at this stage. A single cycle is defined as one charging segment followed by one discharging segment. During these six months, the cluster carried out a total of 275 cycles, as can be seen from the color bars in the figure 1(C).

### 2.2. Experimental data

The experimental data were obtained from our Battery Testing System Laboratory using Neware battery testing equipment. The degradation experiments include 12 battery degradation trajectories under three operating conditions, as shown in figure 2. The cells used in the first and second operating conditions are 13Ah cylindrical LiFePO4 cells of the same type as those used in the BESS. The cells used in the third operating condition are cylindrical 18 650 cells, which are included to further examine the proposed framework under non-full charging/discharging conditions and battery-type variation.

The first operating condition (figure 2(B)) followed the manufacturer-recommended protocol for full charge–discharge cycling. During charging, the battery undergoes constant current (CC) charging until the voltage reaches 3.65 V, followed by constant voltage (CV) charging at this upper limit referred to as the CC–CV charging process. Discharging is carried out at a CC, corresponding to the standard CC discharging protocol.

The second operating condition (figure 2(C)) uses CC protocol for both charging and discharging. Unlike the first operating condition, it omits the CV phase during charging, and the charging voltage is allowed to reach a theoretical upper limit of 3.9 V. The second operating condition is designed to obtain the degradation conditions of a wide SOC range under real operating conditions without CV charging. As shown in figure 1(C), there is almost no CV charging segment in the charging process under real operating conditions, and the voltage drops sharply at the next time point after the end of charging.

The third operating condition (figure 2(D)) simulates a partial charge–discharge scenario representative of real-world usage, with CC charging and discharging within a voltage window of 2.97–3.54 V. A standard full charge–discharge cycle is periodically incorporated into the third scenario as a calibration procedure, performed every 30 cycles. This calibration procedure can obtain battery's actual usable discharge capacity which can serve as a label for SOH estimation. These calibration cycles constitute 1/10 the number of cycles in the partial charge–discharge scenario and their influence on the main scenario is neglected in this study. The measured capacities from these calibration cycles are interpolated to generate virtual capacity labels for the intervening non-full charge–discharge cycles.

In figure 2(A), data1-*i*–*j* denotes the *j*th cell under the *i*th operating condition. Thus, data1–1–1 represents the first 13Ah LiFePO4 cell under the standard CC–CV charging and

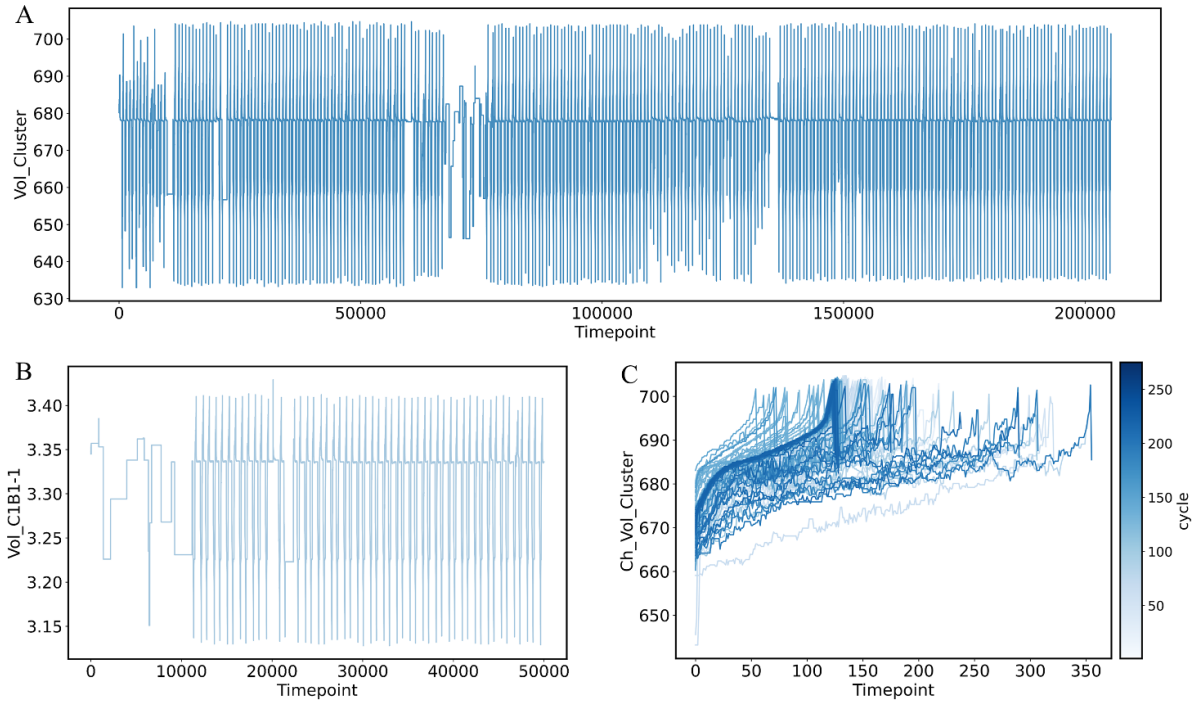



**Figure 1.** Battery energy storage system data. Voltage data, measured in volts (V). The subfigure A: cluster voltage. The subfigure B: partial battery module voltage. The subfigure C: charging segment in cluster voltage.

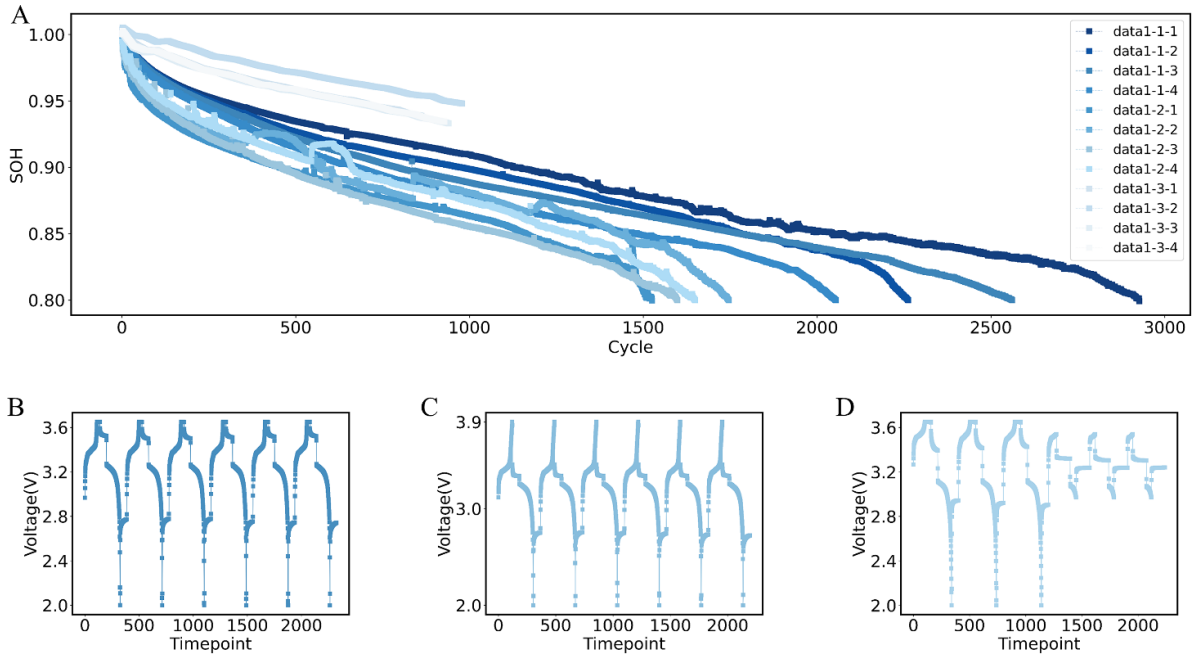



**Figure 2.** Experimental data. The subfigure A: 12 battery degradation data. The subfigure B: standard charging and discharging protocol. The subfigure C: constant current charging and discharging protocol. The subfigure D: standard charging and discharging calibration and non-full charging and discharging protocol. The horizontal axis Timepoint in figures B-D is measured in seconds(s).

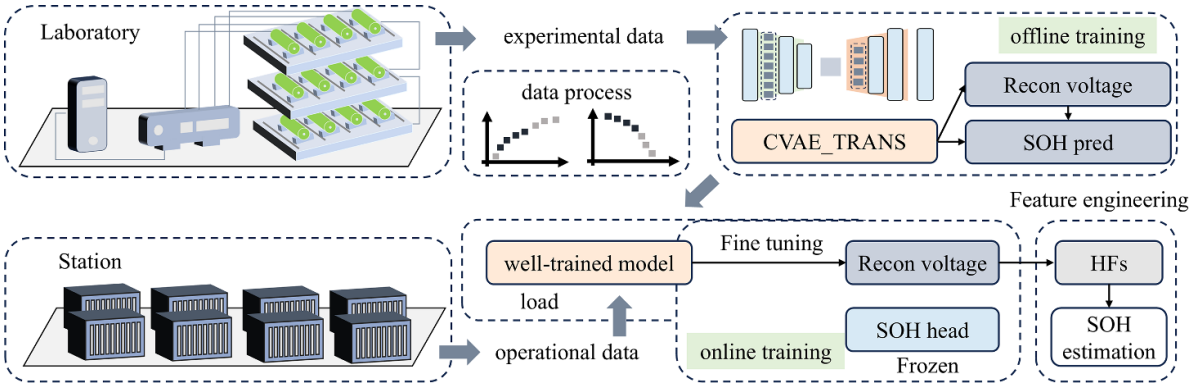



**Figure 3.** The whole framework of the voltage reconstruction and SOH estimation.

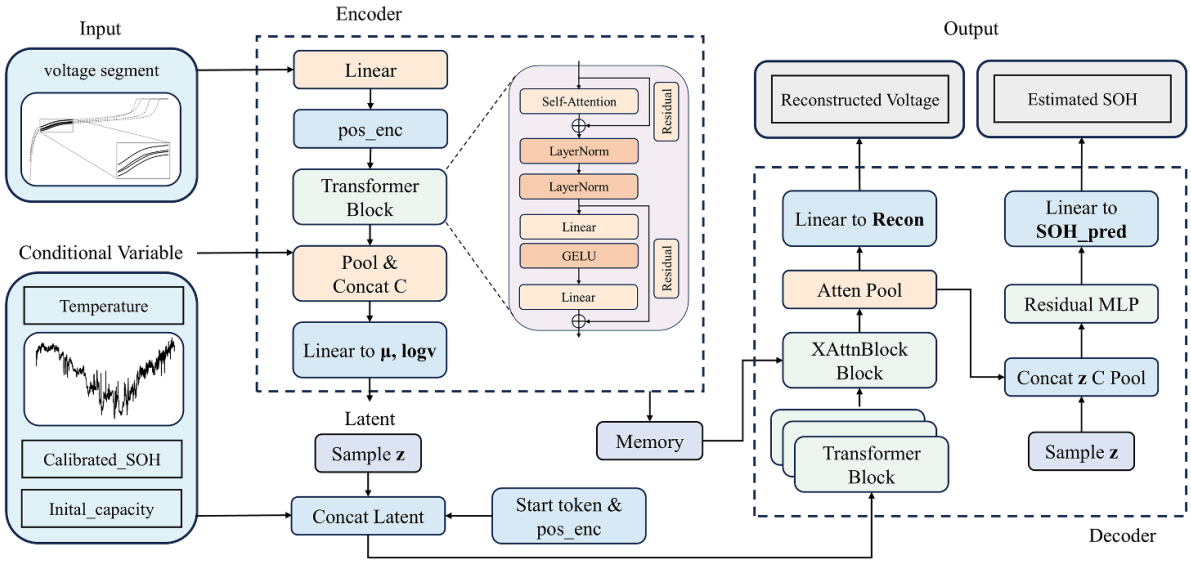



**Figure 4.** The proposed model.

CC discharging protocol. And it should be noted that data1-3 corresponds to cylindrical 18 650 cells rather than the 13Ah cells used in data1–1 and data1–2.

## 3. Methods

This study aims to address the challenges posed by incomplete data (particularly voltage data) in SOH estimation for real-world energy storage applications. By reconstructing voltage profiles, the proposed framework not only enables online battery SOH estimation but also facilitates the practical deployment of existing data-driven approaches (as highlighted in the literature review, many of which demand high-quality voltage data) in commercial energy storage systems.

Following a detailed introduction of the two battery datasets, this section presents the complete workflow of the CVAE-TRANS model for generating full charge voltage profiles and estimating SOH in practical application scenarios as shown in figure 3. To address the complexity and incompleteness of operational data from energy storage systems, experimental data collected under controlled laboratory conditions are preprocessed and used to train and validate the proposed CVAE-TRANS model.

It is noteworthy that, in practical scenarios, degradation data corresponding to standard charge/discharge protocols may be sourced directly from battery manufacturers' technical manuals, thereby circumventing the need for time-consuming experimental procedures. The CVAE-TRANS model reconstructs voltage profiles and learns to obtain battery's SOH by modeling charge/discharge protocols under standard operating conditions. The well-trained model can be deployed online and further fine-tuned using voltage data spanning a broad SOC range from the energy storage systems. Feature engineering is then applied to the reconstructed voltage data to extract health features. These health features can be applied to assist in monitoring and obtaining health status, thereby enabling online estimation of battery's SOH.

### *3.1. Model inputs*

The model proposed in this paper can estimate the SOH of the battery by history degradation data and the reconstructed voltage data, as shown in figure 4. The model takes processed

voltage segments and conditional variables as inputs. The conditional variables include the temperature, calibrated SOH, and initial capacity. These variables include the operating conditions that affect the battery aging process (such as temperature), the differences between batteries, and the calibration information during the operation. The batch ($B$) of voltage segments selected from the raw voltage data is represented as:

$$\mathbf{x}_{\text{seq}} = [x_1, x_2, \ldots, x_{x_{\text{dim}}}] \in \mathbb{R}^{B \times x_{\text{dim}}}$$

where $B$ denotes the batch size of the inputs and $x_{\text{dim}}$ refers to the selected segment length from the raw voltage sequence. Here in this study, $x_{\text{dim}}$ is set to 30, corresponding to 30 time points of voltage data.

The design of this model is to generate a simulated complete voltage segment using only a portion of the actual scene's voltage data. During the offline training stage, this study segmented and expanded the voltage data obtained from experimental data. Segmentation ensures consistency with future application scenarios, while expansion maintains the length of the output voltage data points unchanged. The expansion operation solves the problem of inconsistent voltage curve length caused by the decay of battery aging during charging time. An appropriate delay is applied during the CV charging stage, and setting the output voltage length to 150 ensures that the number of output data points remains consistent.

Battery degradation is also affected by various other factors. So this study takes three conditional variables to assist in model training. The conditional variables selected are temperature, initial battery capacity, and calibrated SOH values. Temperature is an important factor affecting battery degradation rates, and is available in most battery health management systems equipped with temperature sensors. Furthermore, although batteries made of the same material have the same rated capacity in the technical manual, their initial capacity often varies due to various factors [33]. Therefore, during the model training process, the initial capacity is used as an important reference variable. Previous studies have shown that as the battery ages, its voltage curve undergoes characteristic changes, which can also serve as a reliable indicator of the battery's SOH [34]. For experimental data, the ratio of the most recently measured discharge capacity to the initial discharge capacity from standard charge–discharge cycles is used as the calibrated SOH value. In practical applications or experiments involving partial charge–discharge conditions, SOH is calculated as the ratio of the maximum available discharge capacity from the most recent calibration cycle to the rated capacity. The corresponding conditional variables are represented as:

$$\mathbf{c} = [c_1, c_2, c_3] \in \mathbb{R}^{B \times \text{cond_dim}}$$

where, in this study, cond_dim = 3, corresponding to the temperature, initial capacity, and latest calibrated SOH value for each sample in the batch.

### 3.2. CVAE-TRANS model

The CVAE-TRANS framework is adopted because CVAE and Transformer play complementary roles in this task. Reconstructing a complete charging-voltage profile from a sparse voltage segment is a conditional generation problem, where temperature, initial capacity and calibrated SOH can help reduce the uncertainty caused by incomplete inputs. Therefore, the CVAE structure is used to introduce these degradation-related variables into the latent space. Meanwhile, the Transformer is used to capture temporal dependencies and nonlocal relationships among voltage points, which are difficult to model sufficiently with purely recurrent structures such as LSTM. Thus, CVAE-TRANS combines condition-guided generation with Transformer-based voltage-sequence modeling, making it suitable for incomplete voltage-profile reconstruction and subsequent SOH estimation.

For the voltage input to the model, an initial input projection is performed. This process transforms each scalar measurement into a high-dimensional representation, yielding a feature space suitable for subsequent Transformer-based processing. Unlike directly feeding raw scalars into the model, the linear projection allows the model to adaptively learn effective feature representations from the input domain during training. For each scalar input at time step $t$, a shared linear projection is applied to map it into a high-dimensional token space:

$$h_{b,t}^{(0)} = W_{\text{in}} x_{b,t} + \mathbf{b}_{\text{in}}, \quad \forall b \in [1, B], t \in [1, x_{\text{dim}}] \tag{1}$$

where $W_{\text{in}}$ and $\mathbf{b}_{\text{in}}$ are learnable parameters. And $h^{(0)}$ is the projected feature sequence for all samples and time steps.

Then, to preserve temporal order information, a learnable positional embedding vector is added to the projected feature at each time step $t$:

$$h_{b,t}^{(1)} = h_{b,t}^{(0)} + p_{\text{enc},t} \tag{2}$$

Alternatively, in batch form:

$$\mathbf{h}^{(1)} = \mathbf{h}^{(0)} + \mathbf{P}_{\text{enc}} \tag{3}$$

where $\mathbf{P}_{\text{enc}}$ is broadcasted across the batch dimension.

These operations can appropriately process the sequence data for use in the subsequent self-attention mechanism module. This enables the model to capture rich feature information as well as the inherent temporal structure of the battery data signal.

A standard Transformer encoder block with multi-head self-attention, residual connections, layer normalization, and a feedforward network is then used to extract contextual voltage representations:

$$\mathbf{h}^{(2)} = \text{TransformerEnc}\left(\mathbf{h}^{(1)}\right) \tag{4}$$

where $\mathbf{h}^{(2)}$ denotes the encoder output. Since the Transformer encoder follows the standard architecture, its unchanged self-attention and feedforward details are omitted here.

To produce a fixed-dimensional summary of each sample's sequence representation, a pooling operation is then applied. Average pooling is defined as:

$$\bar{\mathbf{h}} = \frac{1}{x_{\text{dim}}} \sum_{t=1}^{x_{\text{dim}}} h_t^{(2)} \in \mathbb{R}^{B \times d_{\text{model}}} \tag{5}$$

If attention pooling is employed, a learnable query vector $\mathbf{q}_{\text{pool}}$ attends over the sequence, generating attention weights and a weighted summary. This pooling step compresses the sequence representation into a fixed-dimensional representation $\bar{\mathbf{h}}$, effectively summarizing salient temporal features for subsequent parameterization:

$$\alpha_{b,t} = \frac{\exp\left(\mathbf{q}_{\text{pool}}^{\top}\mathbf{h}_{b,t}^{(2)}\right)}{\sum_{s=1}^{x_{\text{dim}}}\exp\left(\mathbf{q}_{\text{pool}}^{\top}\mathbf{h}_{b,s}^{(2)}\right)}, \tag{6}$$

$$\bar{\mathbf{h}}_b = \sum_{t=1}^{x_{\text{dim}}}\alpha_{b,t}\mathbf{h}_{b,t}^{(2)}, \tag{7}$$

where $\mathbf{q}_{\text{pool}} \in \mathbb{R}^{d_{\text{model}}}$ is a learnable query vector for encoder-side attention pooling.

The pooled sequence representation is concatenated with the conditional vector for each sample:

$$h_{\text{enc}} = \left[\bar{\mathbf{h}} \,\|\, \mathbf{c}\right] \in \mathbb{R}^{B\times(d_{\text{model}}+\text{cond_dim})} \tag{8}$$

where $\mathbf{c}$ is the conditional variable (temperature, initial capacity, SOH).

Subsequently, a linear transformation projects this concatenated representation into parameters of a Gaussian latent variable. This step defines the mean and (log-)variance of the approximate posterior distribution $q_\phi(\mathbf{z} \mid \mathbf{x}_{\text{seq}}, \mathbf{c})$, which plays a central role in the variational inference framework:

$$\left[\mu, \log\sigma^2\right] = h_{\text{enc}}W_{\text{enc}} + \mathbf{b}_{\text{enc}}. \tag{9}$$

After obtaining the mean and log-variance of the latent variable, we apply the reparameterization trick to enable backpropagation through the stochastic latent code:

$$\mathbf{z} = \boldsymbol{\mu} + \exp\left(0.5\log\sigma^2\right)\odot\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon}\sim\mathcal{N}(0,I) \tag{10}$$

where $\odot$ denotes element-wise multiplication, and $\boldsymbol{\epsilon}$ is sampled from a standard normal distribution.

The decoder reconstructs the output sequence from the latent representation $\mathbf{z}$ and the associated conditional variables $\mathbf{c}$. This process begins by preparing an initial sequence of decoder tokens that embed both temporal position and sample-specific latent information, serving as the starting point for sequential reconstruction. First, an initial decoder token tensor is created, where all time steps are initialized identically. Then, positional information is injected through a learnable positional embedding. Finally, the latent and conditional information-after being projected-is broadcast and added to each time step in the sequence:

$$\mathbf{l}_{\text{dec}} = \text{LatProj}\left([\mathbf{z}\|\mathbf{c}]\right) \tag{11}$$

$$\mathbf{T}^{(0)} = \mathbf{1}_B \otimes \mathbf{s}_{\text{start}} \tag{12}$$

$$\mathbf{T}^{(1)} = \mathbf{T}^{(0)} + \mathbf{P}_{\text{dec}} \tag{13}$$

$$\mathbf{T}^{(2)} = \mathbf{T}^{(1)} + \mathbf{l}_{\text{dec}}. \tag{14}$$

Here, LatProj is a linear projection layer that maps the concatenated latent and conditional vector to the model dimension, $\mathbf{s}_{\text{start}}$ is a learnable start token, and $\mathbf{P}_{\text{dec}}$ denotes the positional embedding for the decoder. The resulting tensor $\mathbf{T}^{(2)}$ serves as the input to the decoder block, enabling it to condition the reconstruction process on both sequence position and latent context.

The sequence of decoder tokens is further processed by a stack of Transformer decoder blocks-specifically, three stacked Transformer blocks structurally identical to the encoder blocks but without pooling. This enables the decoder to construct a full-sequence representation in an autoregressive-like manner (though the entire sequence is available in this case) before applying cross-attention.

After the decoder self-attention layers, the decoder tokens attend to the encoder output (referred to as the "memory") via a cross-attention mechanism, which allows the decoder to reference the encoded input sequence at each step of reconstruction:

$$\mathbf{T}_{\text{cross}} = \text{CrossAttn}\left(\mathbf{T}_{\text{dec}},\, \mathbf{H}_{\text{enc}},\, \mathbf{H}_{\text{enc}}\right) \tag{15}$$

where $\mathbf{T}_{\text{dec}}$ is the output from the last decoder self-attention block, $\mathbf{H}_{\text{enc}}$ is the encoder output sequence.

The cross-attended output is then refined through a feedforward layer and residual connection with layer normalization:

$$\tilde{\mathbf{T}}_{\text{cross}} = \text{LayerNorm}\left(\mathbf{T}_{\text{dec}} + \mathbf{T}_{\text{cross}}\right) \tag{16}$$

$$\mathbf{T}_{\text{out}} = \tilde{\mathbf{T}}_{\text{cross}} + \text{FFN}\left(\tilde{\mathbf{T}}_{\text{cross}}\right). \tag{17}$$

This enables the decoder to access all encoded input information at every time step, improving the fidelity and expressiveness of sequence reconstruction.

The final decoder output is projected back into the original signal domain (battery voltage) using a linear transformation:

$$\hat{x} = W_{\text{o}}\,\mathbf{T}_{\text{out}} + \mathbf{b}_{\text{o}}. \tag{18}$$

Another decoder output is the estimated SOH value, computed by a multilayer perceptron:

$$\widehat{\text{SOH}} = f_{\text{soh}}\left(\mathbf{u}\right) \in \mathbb{R}^B \tag{19}$$

where $f_{\text{soh}}(\cdot)$ denotes a multilayer perceptron. The unified feature representation $\mathbf{u} = [\mathbf{z}\,\|\,\mathbf{c}\,\|\,\mathbf{q}]$ is formed by combining three components: the latent vector $\mathbf{z}$ from the CVAE encoder, the conditional vector $\mathbf{c}$ (including operating temperature and calibrated SOH), and the global context vector $\mathbf{q}$ obtained via attention pooling,

$$\mathbf{q} = \text{AttnPool}(\mathbf{q}_0,\, \mathbf{T}_{\text{out}},\, \mathbf{T}_{\text{out}}),$$

where $\mathbf{q}_0$ is the initial learnable query embedding that serves as the query input for the attention pooling operation. It aggregates global contextual information from the decoder output $\mathbf{T}_{\text{out}}$ to produce a holistic representation of the reconstructed voltage dynamics. By integrating both the global context and the latent–conditional information, this SOH estimation head provides a robust and context-aware prediction of battery health for each input sample.

### 3.3. Model training

The dataset division was performed at the cycle level rather than at the voltage-point level. Each sample corresponds to one charge/discharge cycle. For each valid cycle, the middle $x_{\text{dim}} = 30$ charging-voltage points were used as the sparse input, and the full charging-voltage sequence was padded or truncated to 150 points as the reconstruction target. The condition vector consisted of temperature, initial capacity and calibrated SOH. The SOH label was defined as the next-cycle SOH, so the last cycle of each cell was excluded because no next-cycle label was available. Dataset 1 and Dataset 2 were used for offline voltage-reconstruction evaluation over eight independent cell-level subsets. Dataset 3 was used to simulate online adaptation: the complete calibration cycles were used for fine-tuning, while overlapping cycle indices were removed from the non-full charge/discharge records before validation/testing. The operational BESS data were used only as a field application demonstration and were not used to train the offline benchmark model.

*3.3.1. Three-phase training.* To maximize predictive accuracy while maintaining a well-behaved latent space, a three-stage training strategy is adopted, outlined in algorithm 1. Each stage targets a specific sub-objective and gradually releases optimization degrees of freedom, allowing the model to converge to a high-quality solution that jointly serves voltage generation and battery SOH diagnosis.

The weights of these losses are selected to maintain voltage reconstruction as the primary objective while gradually introducing SOH-related supervision. Specifically, the KL weights $\beta_1$ and $\beta_3$ prevent an unstructured latent space but are kept small to avoid over-smoothing the reconstructed voltage profiles. The SOH weight $\gamma$ is introduced after reconstruction pretraining to reduce interference between waveform generation and health regression. The ranking weight $\lambda$rank is set to a small value because it is used only as a latent-space regularizer. The final values are selected according to validation performance and are further examined in the sensitivity analysis.

The model is first exposed only to the reconstruction objective, minimizing:

$$\mathcal{L}_{\text{p1}} = \mathcal{L}_{\text{rec}} + \beta_1 \mathcal{L}_{\text{KL}} \tag{20}$$

where the multiscale loss $\mathcal{L}_{\text{p1}}$ enforces global-and-local waveform fidelity while $\mathcal{L}_{\text{KL}}$ regularities the latent posterior, so that the encoder–decoder pair learns an information-rich code $\mathbf{z}$ before any SOH supervision is introduced.

In the second phase the encoder and decoder are frozen and only the regression branch is trained with $\mathcal{L}_{\text{p2}} = \gamma_2 \mathcal{L}_{\text{soh}}$. $\mathcal{L}_{\text{soh}}$ is the mean-squared error between the predicted SOH and the reference value; the regressor receives $[\mathbf{z}\|\mathbf{c}\|\mathbf{q}]$, thereby exploiting frozen generative features without destructive interference from reconstruction gradients. Finally, all parameters are unfrozen and fine-tuned end-to-end by minimizing:

$$\mathcal{L}_{\text{p3}} = \mathcal{L}_{\text{rec}} + \beta_3 \mathcal{L}_{\text{KL}} + \gamma_3 \mathcal{L}_{\text{soh}} + \lambda_{\text{rank}} \mathcal{L}_{\text{rank}} \tag{21}$$

**Algorithm 1.** Three-Phase Training for CVAE-TRANS.

**Input:** Training set $\mathcal{D}_{\text{tr}}$, Validation set $\mathcal{D}_{\text{val}}$, Test set $\mathcal{D}_{\text{test}}$
**Output:** Trained model $\theta^\star$, Test losses $\mathcal{L}_{\text{rec}}^{\text{test}}, \mathcal{L}_{\text{soh}}^{\text{test}}$
Initialise model parameters $\theta$ and learning rates $\{lr_{\text{recon}}, lr_{\text{soh}}, lr_{\text{finetune}}\}$
Define phase epochs $(E_1, E_2, E_3)$ and weights $(\beta_1, \beta_3, \gamma_2, \gamma_3, \lambda_{\text{rank}})$
**Phase 1:** *Reconstruction pre-training*
**for** epoch $= 1$ to $E_1$ **do**
  **for all** mini-batch $(x_{\text{seg}}, c, v_{\text{full}}) \in \mathcal{D}_{\text{tr}}$ **do**
    $(\hat{v}, \mu, \log\sigma^2) \leftarrow \text{model}(x_{\text{seg}}, c)$
    $\mathcal{L}_{\text{p1}} \leftarrow \mathcal{L}_{\text{rec}}(\hat{v}, v_{\text{full}}) + \beta_1 \mathcal{L}_{\text{KL}}(\mu, \log\sigma^2)$
    Update encoder_decoder parameters by $\nabla_\theta \mathcal{L}_{\text{p1}}$
  **end for**
**end for**
**Phase 2:** *SOH-head regression*
**Freeze** encoder_decoder parameters
**Unfreeze** SOH-head parameters
**for** epoch $= 1$ to $E_2$ **do**
  **for all** mini-batch $(x_{\text{seg}}, c, s_{\text{true}}) \in \mathcal{D}_{\text{tr}}$ **do**
    $(\mathbf{z}, \text{pool}) \leftarrow \text{encode_decode}(x_{\text{seg}}, c)$
    no gradients
    $\hat{s} \leftarrow \text{SOH-head}([\mathbf{z}\|\mathbf{c}\|\mathbf{q}])$
    $\mathcal{L}_{\text{p2}} \leftarrow \gamma_2 \mathcal{L}_{\text{soh}}(\hat{s}, s_{\text{true}})$
    Update SOH-head parameters by $\nabla_\theta \mathcal{L}_{\text{p2}}$
  **end for**
**end for**
**Phase 3:** *Joint fine-tuning*
**Unfreeze** entire model
**for** epoch $= 1$ to $E_3$ **do**
  **for all** mini-batch $(x_{\text{seg}}, c, v_{\text{full}}, s_{\text{true}}) \in \mathcal{D}_{\text{tr}}$ **do**
    $(\hat{v}, \hat{s}, \mu, \log\sigma^2) \leftarrow \text{model}(x_{\text{seg}}, c)$
    $\mathcal{L}_{\text{p3}} \leftarrow \mathcal{L}_{\text{rec}}(\hat{v}, v_{\text{full}}) + \beta_3 \mathcal{L}_{\text{KL}}(\mu, \log\sigma^2) + \gamma_3 \mathcal{L}_{\text{soh}}(\hat{s}, s_{\text{true}}) + \lambda_{\text{rank}} \mathcal{L}_{\text{rank}}(\mu, s_{\text{true}})$
    Update all parameters by $\nabla_\theta \mathcal{L}_{\text{p3}}$
  **end for**
**end for**
Select best checkpoint $\theta^\star$ w.r.t. validation loss
Evaluate $\theta^\star$ on $\mathcal{D}_{\text{test}}$ to obtain $\mathcal{L}_{\text{rec}}^{\text{test}}, \mathcal{L}_{\text{soh}}^{\text{test}}$
**return** $\theta^\star$, $\mathcal{L}_{\text{rec}}^{\text{test}}$, $\mathcal{L}_{\text{soh}}^{\text{test}}$

where $\mathcal{L}_{\text{rank}}$ aligns the pairwise distances of latent means with the corresponding SOH differences, thereby encouraging degradation-related ordering in the latent space. This joint objective lets improvements in reconstruction sharpen the latent manifold, which in turn enhances SOH prediction. This staged strategy first stabilizes voltage reconstruction and then introduces SOH-related supervision, which reduces the interference between the generative and regression objectives.

*3.3.2. Online fine-tuning strategy.* Models trained on laboratory or manufacturer data may show distribution shifts when directly applied to field operating conditions. Therefore, an online fine-tuning strategy is adopted, as shown in algorithm 2. The offline-trained parameters are first loaded as the initialization, and voltage segments collected from the target operating scenario are then used for adaptation. When calibration cycles

**Algorithm 2.** Online fine-tuning strategy.

**Input:** Training loader $\mathcal{D}_{\text{tr-new}}$, epochs $E$, weights $(\beta, \gamma, \lambda)$
**Output:** Fine-tuned parameters $\theta^{\star\star}$
**for** $e \leftarrow 1$ **to** $E$ **do**
  loss_sum $\leftarrow 0$
  **for all** $(x_{\text{seg}}, c, v_{\text{full}}, s_{\text{true}}) \in \mathcal{D}_{\text{tr-new}}$ **do**
    $(\hat{v}, \hat{s}, \mu, \log\sigma^2) \leftarrow \text{model}(x_{\text{seg}}, c)$
    $\mathcal{L}_{\text{rec}} \leftarrow \text{MSE}(\hat{v}, v_{\text{full}}) + 0.2 \sum_{w} \text{MSE}(\hat{v}_w, v_w)$
    $\mathcal{L}_{\text{KL}} \leftarrow -\frac{1}{2} \sum_{i} (1 + \log\sigma_i^2 - \mu_i^2 - \sigma_i^2)$
    $\mathcal{L}_{\text{soh}} \leftarrow \text{MSE}(\hat{s}, s_{\text{true}})$
    $\mathcal{L}_{\text{rank}} \leftarrow \text{MSE}(\|\mu_i - \mu_j\|_2, |s_i - s_j|)$
    $\mathcal{L}_{\text{joint}} \leftarrow \mathcal{L}_{\text{rec}} + \beta\mathcal{L}_{\text{KL}} + \gamma\mathcal{L}_{\text{soh}} + \lambda\mathcal{L}_{\text{rank}}$
    Back-propagate $\mathcal{L}_{\text{joint}}$ and update parameters
    loss_sum $\leftarrow$ loss_sum $+ \mathcal{L}_{\text{joint}}$
  **end for**
**end for**
$\theta^{\star\star} \leftarrow$ parameters with lowest validation loss
**return** $\theta^{\star\star}$

are available, the measured capacity from the most recent calibration cycle is used to construct the SOH label; otherwise, pseudo-SOH labels obtained from auxiliary calibration information are used for illustrative field-data validation. During fine-tuning, the same joint loss is adopted, but a smaller learning rate is used to avoid destroying the reconstruction ability learned from the offline data. In this way, the model preserves the general voltage-reconstruction capability learned from standard protocols while adapting to the voltage distribution and condition variables of the target application scenario.

## 4. Results

The results discussed in the following subsections are based on four datasets. These datasets contain complete charge and discharge cycles as well as measured capacity values. These calibrated capacity values can provide SOH labels for the validation of the proposed method and serve as the basis for offline training. Meanwhile, in order to demonstrate the applicability and effectiveness of the method in real-world scenarios, datasets with more complex conditions were also utilized.

### 4.1. Generative results

*4.1.1. Latent space analysis.* PCA and t-SNE dimensionality reduction methods were employed to evaluate the learned latent representations $\boldsymbol{\mu}(\mathbf{z})$. Principal component analysis is used to represent the main directions of variance within the latent space, while t-SNE focuses on the clustering behavior between cycles. The results can be seen from figure 5.

The estimated SOH results in Dataset 1 show a continuous and monotonous trend. These results indicate that this potential representation is highly consistent with the battery aging situation. Taking the results of Dataset 1–1 and Dataset 1–3 as examples, the principal component analysis of $\boldsymbol{\mu}(\mathbf{z})$ clearly shows the principal components directly related to SOH. This correlation is also confirmed by the smooth trajectory graph in t-SNE. This indicates that for these subsets, the aging condition of the batteries can actually be captured and encoded by the proposed model. Additionally, some of the patterns in the figure illustrate more complex potential geometric structures. In the PCA and t-SNE results, they manifest as the presence of multiple different clusters as well as branched or circular patterns. Taking the results of datasets 1–2 and 1–4 examples, it may suggest that the latent variables also encode other factors such as operating conditions or periodic information. Datasets 2–1 to 2–4 further illustrate a variety of latent structures. PCA results reveal varying degrees of linearity and clustering across these subsets. In summary, these latent-space analyses provide valuable insights into the encoding behavior of the CVAE-TRANS model. These insights will guide subsequent detailed analyses of voltage reconstruction quality and SOH prediction accuracy.

These results indicate that the latent space learned by the proposed model contains both degradation-related and condition-related information. The relatively smooth trajectories suggest that the model can encode the gradual aging evolution of batteries, while the clustered or branched structures imply that the latent variables may also reflect differences in operating protocols, calibration behavior, or cell-to-cell variation. This observation supports the introduction of conditional variables in the CVAE structure, because sparse voltage segments alone may not be sufficient to distinguish different degradation states under heterogeneous operating conditions.

*4.1.2. Reconstructed voltage analysis.* This subsection will provide the voltage reconstruction performance of the well-trained model. Here, the reconstruction target is set as the full charging voltage curve (which is usually not affected by dynamic operating conditions during the charging process). Two datasets are employed: Dataset 1 acquired under the manufacturer's standard CC–CV charge/discharge protocol and Dataset 2 containing only the constant-current charge phase. After training, a single representative voltage curve will be randomly selected from each dataset, as shown in the figure 6. These curves were selected to visually demonstrate the ability of the trained model to reconstruct the standard charging voltage under different conditions.

In this study, two evaluation indicators-Mean absolute error and Root-mean-square error (MAE and RMSE)-are used to display the voltage-reconstruction performance:

$$\text{MAE} = \frac{1}{B}\sum_{b=1}^{B}\left(\frac{1}{T}\sum_{t=1}^{T}|\hat{v}_{b,t} - v_{b,t}|\right) \tag{22}$$

$$\text{RMSE} = \sqrt{\frac{1}{BT}\sum_{b=1}^{B}\sum_{t=1}^{T}(\hat{v}_{b,t} - v_{b,t})^2}. \tag{23}$$

And two similarity metrics-Pearson correlation coefficient ($\rho$) and the Discrete Fréchet distance (DFD)-are used to assess shape congruence and offer complementary perspectives on

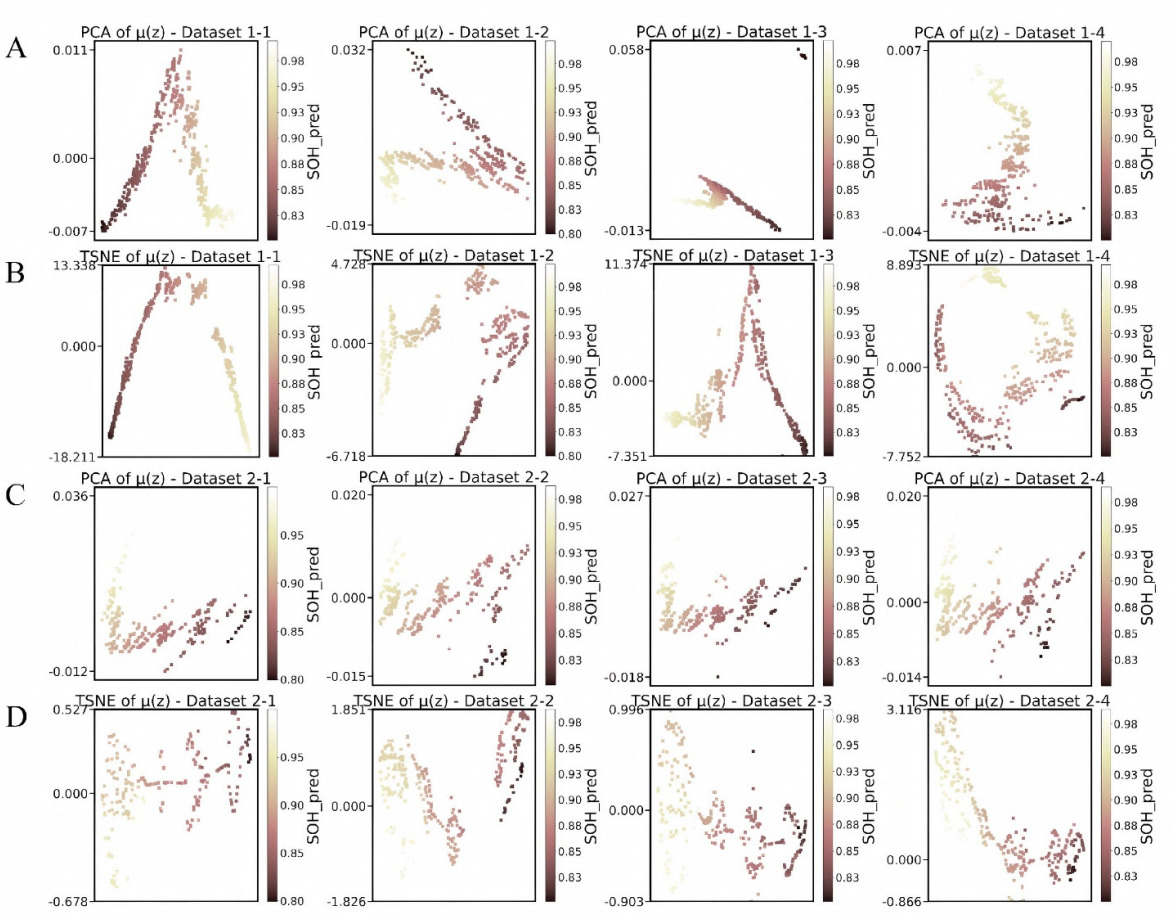



**Figure 5.** The results of PCA and t-SNE dimensionality reduction. Row A in the figure shows the PCA results for Dataset 1 and row B shows the TSNE results for Dataset 1. Row C shows the PCA results for Dataset 2 and row D shows the TSNE results for Dataset 2.

the fidelity of the reconstructed waveforms. The formulas are:

$$\rho = \frac{\sum_{b=1}^{B}\sum_{t=1}^{T}\left(\hat{v}_{b,t}-\bar{\hat{v}}\right)\left(v_{b,t}-\bar{v}\right)}{\sqrt{\sum_{b=1}^{B}\sum_{t=1}^{T}\left(\hat{v}_{b,t}-\bar{\hat{v}}\right)^2}\sqrt{\sum_{b=1}^{B}\sum_{t=1}^{T}\left(v_{b,t}-\bar{v}\right)^2}} \tag{24}$$

$$d_{\mathrm{F}}(P,Q) = \min_{\sigma}\max_{(i,j)\in\sigma}\left|P_i - Q_j\right| \tag{25}$$

where $B$ represents the batch size used during the evaluation; $T$ represents number of time steps per voltage sequence; $\hat{v}_{b,t}$ is the reconstructed voltage of batch element $b$ at time step $t$; $\bar{\hat{v}}$ and $\bar{v}$ are global means of $\hat{v}_{b,t}$ and $v_{b,t}$; $P_i$, $Q_j$ represent $i$th and $j$th points of the discrete curves $P$ and $Q$; $\sigma$ represents any monotone path from $(1,1)$ to $(|P|,|Q|)$ in the Fréchet grid.

The aforementioned evaluation metrics are computed for voltage curves in the validation sets, not merely for the curves presented visually. The results are summarized in table 1. The relatively low MAE (with a mean value lower than 0.02) and RMSE score, high Pearson correlation coefficient, and consistently small DFD all indicate the superior performance of the proposed model.

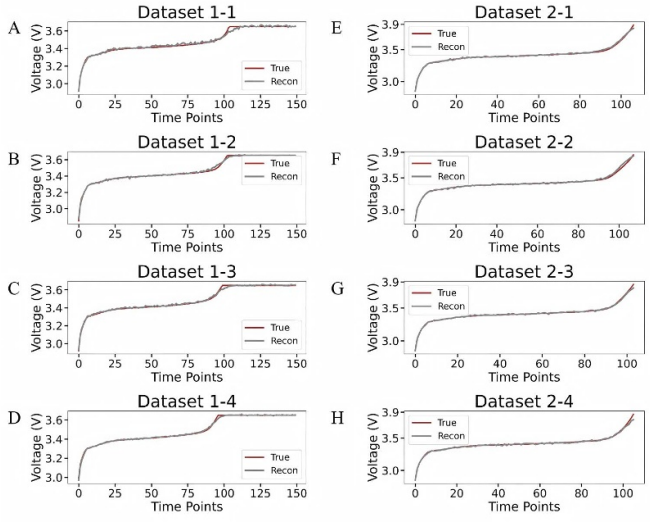



**Figure 6.** The performance of voltage reconstruction.

Three other methods-single CVAE, LSTM, and conditional generative adversarial network (CGAN)-were employed for comparative voltage reconstruction experiments, and the results are presented in table 1, where the best values are

**Table 1.** Comparison of the proposed CVAE-TRANS framework with other models (voltage–reconstruction performance).

| | Proposed (CVAE-TRANS) | | | | CVAE | | | | LSTM | | | | CGAN | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Data | MAE | RMSE | $\rho$ | F | MAE | RMSE | $\rho$ | F | MAE | RMSE | $\rho$ | F | MAE | RMSE | $\rho$ | F |
| 1–1 | 0.018 | 0.031 | 0.977 | **0.020** | 0.036 | 0.043 | 0.978 | 0.112 | **0.013** | **0.022** | **0.989** | 0.100 | 0.023 | 0.034 | 0.974 | 0.105 |
| 1–2 | **0.012** | **0.020** | **0.991** | **0.015** | 0.017 | 0.024 | 0.990 | 0.064 | 0.013 | 0.023 | **0.991** | 0.050 | 0.033 | 0.041 | 0.978 | 0.094 |
| 1–3 | **0.013** | **0.022** | **0.989** | **0.014** | 0.023 | 0.051 | 0.943 | 0.063 | 0.033 | 0.058 | 0.924 | 0.063 | 0.031 | 0.058 | 0.935 | 0.092 |
| 1–4 | 0.010 | 0.017 | 0.993 | **0.028** | 0.011 | 0.019 | 0.991 | 0.065 | **0.009** | **0.015** | **0.995** | 0.048 | 0.037 | 0.043 | 0.981 | 0.121 |
| 2–1 | **0.017** | **0.029** | **0.993** | **0.029** | 0.024 | 0.033 | **0.993** | 0.062 | 0.028 | 0.056 | 0.979 | 0.067 | 0.038 | 0.049 | 0.988 | 0.096 |
| 2–2 | **0.015** | **0.025** | **0.995** | **0.023** | 0.033 | 0.041 | 0.993 | 0.084 | 0.026 | 0.048 | 0.985 | 0.062 | 0.029 | 0.040 | 0.991 | 0.110 |
| 2–3 | **0.014** | **0.025** | **0.995** | **0.022** | 0.017 | 0.030 | 0.993 | 0.068 | 0.017 | 0.026 | **0.995** | 0.063 | 0.036 | 0.048 | 0.988 | 0.105 |
| 2–4 | **0.017** | 0.030 | 0.993 | **0.023** | 0.022 | 0.032 | **0.994** | 0.062 | **0.017** | **0.028** | **0.994** | 0.061 | 0.049 | 0.062 | 0.979 | 0.146 |
| **Avg.** | **0.014** | **0.025** | **0.991** | **0.022** | 0.024 | 0.034 | 0.985 | 0.073 | 0.020 | 0.035 | 0.982 | 0.069 | 0.034 | 0.047 | 0.978 | 0.109 |

*Notes:* Metrics are rounded to three decimals. Lower MAE, RMSE, and *F* (Fréchet distance) indicate better reconstruction performance, whereas a higher $\rho$ indicates better reconstruction performance. The best result for each metric in each row is highlighted in bold; ties are also highlighted in bold.

**Table 2.** Additionally comparison with other models (average results).

| Method | MAE | RMSE | $\rho$ | *F* |
|---|---|---|---|---|
| CVAE-TRANS | 0.014 | 0.025 | 0.991 | 0.022 |
| BiLSTM | 0.014 | 0.028 | 0.987 | 0.062 |
| CSDI | 0.020 | 0.032 | 0.985 | 0.072 |
| TimesNet | 0.028 | 0.042 | 0.974 | 0.101 |

highlighted in bold. Except for the LSTM model in the first dataset 1–1 which performed slightly better, in all the other datasets, the proposed CVAE-TRANS method outperformed all the other methods. In eight representative cases, the proposed CVAE-TRANS method achieved the lowest average MAE (0.014) and RMSE (0.025), and also had the highest Pearson correlation (0.991), indicating its excellent performance in voltage reconstruction. To expand the comparison scope, TimesNet and a CSDI-style conditional diffusion baseline were additionally evaluated under the same eight batteries and $x_{\mathrm{dim}} = 30$ input setting. In the expanded benchmark, CVAE-TRANS achieved the lowest average RMSE and Fréchet distance, with a similarly low MAE, indicating stronger overall waveform fidelity as shown in table 2.

Overall, these results indicate that the proposed model is capable of reconstructing a complete and physically consistent charging curve even with only 30 sparse sampled voltage data and their associated conditioning variables. The superior reconstruction performance can be attributed to the complementary roles of the Transformer and CVAE components. The Transformer encoder-decoder captures temporal dependencies among sparse voltage points and helps recover nonlocal voltage-evolution patterns, while the conditional latent representation introduces temperature, initial capacity and calibrated SOH into the generation process. Compared with models that rely only on sequential mapping or unconditional generation, this design reduces the uncertainty caused by incomplete voltage inputs. In addition, the improvement in Pearson correlation and Fréchet distance shows that the proposed model preserves not only pointwise voltage accuracy

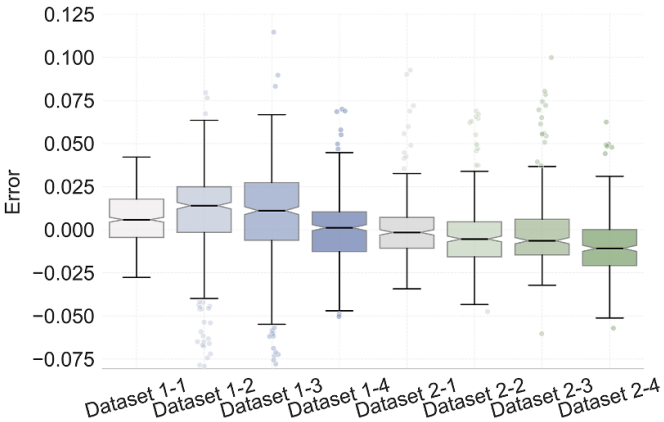



**Figure 7.** SOH estimation error for eight batteries.

but also the overall waveform shape, which is important for subsequent feature extraction and health estimation.

After the reconstructed voltage profile is fused with auxiliary battery information in the latent space, the resulting feature vector is passed to the SOH-estimation head to obtain the battery's future SOH. Figure 7 presents the prediction errors for all eight validation subsets as notched box-and-whisker plots. The median error remains within 0.01 SOH units range that indicates negligible systematic bias in all the battery datasets. The inter-quartile ranges are uniformly tight (less than 0.03) and more than 90% of errors fall inside $\pm$ 0.04. The low MAE and RMSE values reported in table 1 confirm that a latent representation derived from only 30 voltage points is sufficiently informative for reliable regression tasks. Although Datasets 1–2 and 1–3 exhibit slightly broader tails and Datasets 2–2 to 2–4 show a mild negative skew, no subset evidences catastrophic outliers. These results demonstrate that the CVAE-TRANS model can reconstruct charging curves and translate them into accurate SOH forecasts with heterogeneous operating conditions. Some deviations of the specific dataset highlight avenues for future refinement (e.g. incorporating additional conditional variables such as precise temperature or current-rate embeddings). The SOH-estimation performance further confirms that the reconstructed voltage profiles

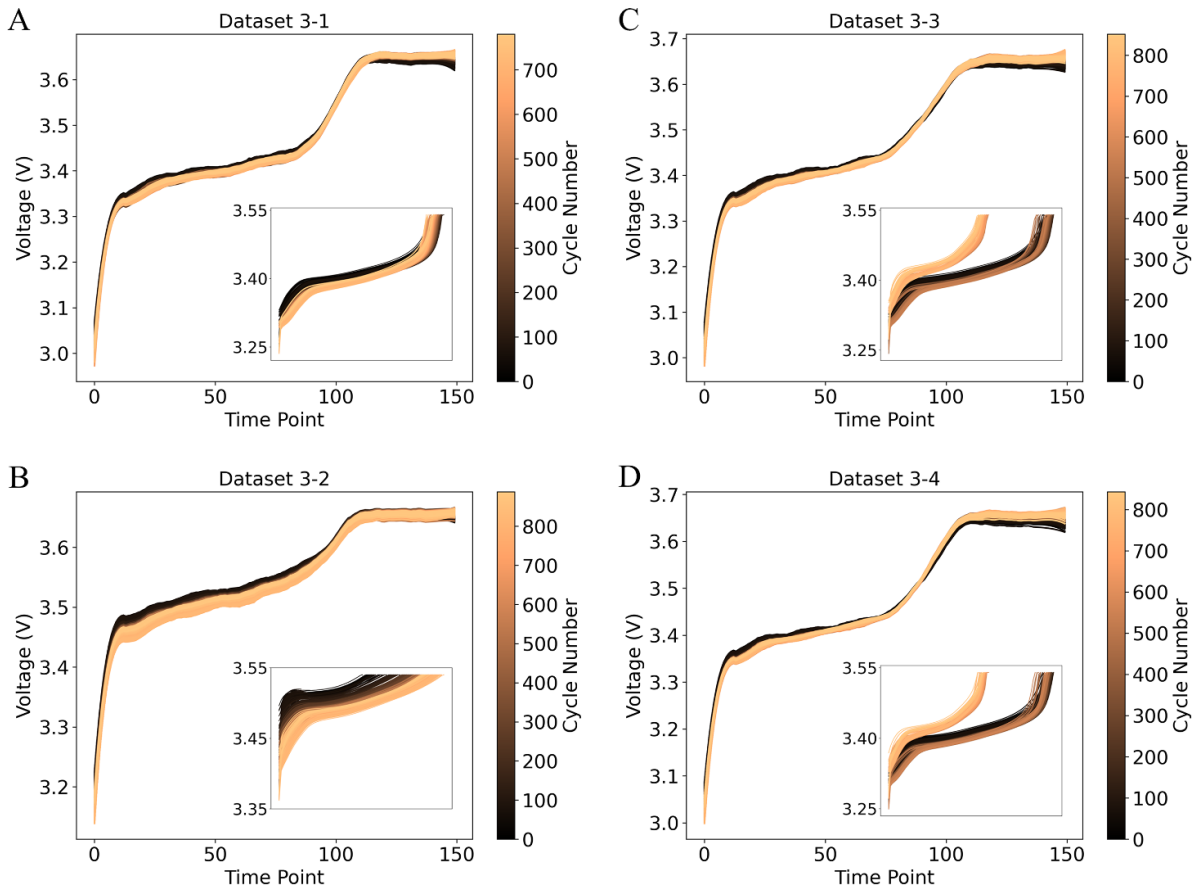



**Figure 8.** Reconstructed voltage and real voltage Dataset 3. Figures A-D show the reconstructed charging voltages for batteries 1–4 in Dataset 3. The inset gives the actual voltage.

contain degradation-sensitive information. Although only a short voltage segment is used as the model input, the reconstructed full-profile representation provides a broader basis for extracting aging-related features than the original sparse segment. The slightly larger error tails observed in some subsets may be related to differences in aging trajectories, operating protocols, or insufficiently described condition variables. This suggests that the reconstruction-based SOH estimation framework is effective, but its performance can be further improved by incorporating richer condition information such as current rate, SOC window, and more detailed temperature statistics.

### *4.2. Online application scenarios*

As verified earlier on two full-charge/discharge datasets, the proposed generative framework cannot only reconstructs charging-voltage profiles with high fidelity but also achieve accurate SOH forecasts. The above two datasets are also considered part of offline training in our overall workflow whether the data originate from a manufacturer's specification or a laboratory test bench, degradation trajectories obtained under standard protocols provide a reliable set of initial parameters for batteries fabricated from the same chemistry). During deployment, the model can be fine-tuned online with newly acquired field data, allowing it to adapt seamlessly to unseen operating regimes.

The remainder of this subsection describes the online application of the model using Dataset 3 and operational energy storage system data are used for illustration purposes.

*4.2.1. Validation on dataset 3.* Since Dataset 3 possesses the same non-full charge–discharge protocol cycles as real-world operating conditions and includes accurately measured capacity label data, it is used here to validate the effectiveness of the online steps proposed in the framework. During the online application stage, the complete charge–discharge calibration cycles included in dataset 3 (as shown in the charge–discharge protocol depicted in figure 2(D) are regarded as representative data covering a wide SOC range collected under actual operating conditions, and are used for the fine-tuning of the model. As shown in figure 3, the fine-tuned model can reconstruct complete charging voltage curves under partial charging conditions. Then, feature engineering combined with a regression model can be applied in SOH estimation. Other non-full charge–discharge cycles (illustrated by the embedded subplot in figure 8) are preserved as validation sets. As illustrated in figures 8(C) and (D), Datasets 3–3 and 3–4 were collected under modified operating conditions in which both the charge- and discharge-rates were increased to 2C. The

**Table 3.** Performance metrics on Dataset 3.

| | Feature correlation | | SOH estimation | |
|---|---|---|---|---|
| Dataset | Mean (V) | d$V$d$Q$ | MAE | RMSE |
| 3–1 | 0.6035 | –0.7805 | 0.0087 | 0.0117 |
| 3–2 | 0.7694 | –0.7821 | 0.0069 | 0.0101 |
| 3–3 | 0.5204 | –0.7512 | 0.0086 | 0.0119 |
| 3–4 | 0.4998 | –0.7725 | 0.0081 | 0.0110 |

steeper upward trend observed in the corresponding voltage curve and the changed CV plateau period both confirm that this high-rate experimental scheme is effective. By evaluating the model on these two datasets, the proposed reconstruction framework demonstrated stable performance even when the electrochemical operating conditions changed significantly. In practical applications, the current rate may vary dynamically. This further proves that this method is applicable to real-world applications.

As shown in figure 8, the shape of the reconstructed voltage curve undergoes significant changes with the increase in the number of cycles. Feature engineering can utilize this change to estimate the health status of the battery. Specifically, the voltage curve is first smoothed with a Savitzky–Golay filter; statistical voltage features and the differential-feature curves are then extracted from the smoothed curves. The mean values are mapped to the measured battery capacities. Furthermore, an XGBoost-based multivariate regression model is established to map these extracted health features to the battery's SOH. The performance results of extracted health features and the regression model are summarized in table 3. In this table, the Pearson correlation coefficients between two features (average voltage and d$V$/d$Q$) and capacity were calculated. The correlation results are listed in the "Feature Correlation" column. When employing standard end-to-end training strategies, the stability and accuracy of d$V$d$Q$'s feature correlation results fall significantly short of those achieved by our proposed method. With a learning rate of $1 \times 10^{-3}$ and extended training over 100 epochs, the results are 0.7481, 0.4168, 0.6606, and 0.5638. Compared with the results in table 3, this performance gap is quite significant, which further highlights the superiority of the three-stage training strategy we proposed.

It should be emphasized that the Savitzky–Golay smoothing algorithm, feature extraction methods, and XGBoost regressor employed in this study can be replaced with more sophisticated alternatives depending on the specific situation. They are only used here to illustrate that our voltage reconstruction method can provide more comprehensive information for battery health management systems in multiple dimensions. The smoothing method and XGboost model parameters used here are shown in table 4, which also gives some of the parameters used in online fine-tuning. The central contribution of this work is the reliable completion of otherwise missing voltage data; the reconstructed profiles not only broaden the data foundation available to battery-health-management systems, but also unlock the online deployment of many existing voltage-driven diagnostics and prognostics that were previously constrained by data incompleteness.

**Table 4.** Key hyperparameters and settings used during model training.

| Stage | Parameter | Setting |
|---|---|---|
| Smoothing | Savitzky–Golay window length | 25 |
| | Polynomial order | 3 |
| XGBoost | $n_{\text{estimators}}$ | 500 |
| | Learning rate | 0.01 |
| | Max depth | 5 |
| | Min child weight | 3 |
| | Subsample | 0.8 |
| | Colsample by tree | 0.8 |
| | $\ell_1$ regularization ($\alpha$) | 0.01 |
| | $\ell_2$ regularization ($\lambda$) | 1.0 |
| Fine-tuning | Learning rate | $1 \times 10^{-4}$ |
| | Epochs | 50 |
| | SOH loss weight $\gamma$ | 0.2 |

*4.2.2. Validation on operational energy storage system data.* The experimental data from Dataset 3 effectively illustrates the online update procedure and demonstrates favorable results. Then, data from an energy storage system will be used to validate and illustrate the methodology. Fine-tuning steps can be trained locally on a computer equipped with an Intel i7-9700 K CPU and an RTX 2070 GPU. Each batch (205 samples $\times$ 150 steps) of fine-tuning took 0.48 seconds with less than 100MB of memory usage.

Three cells (B1–1, B1–2, and B1–3) were selected from a sub-battery box within the energy storage system dataset for verification. During the processing of the original data from the energy storage station, a single charging and discharging operation is defined as one "cycle". Voltage curves spanning a wide SOC window are extracted and treated as complete charge–discharge curves inputs for fine-tuning the offline-trained model. The chosen actual operational data were then fed into the offline-trained model to generate complete charge–discharge voltage profiles. The reconstruction results were compared with the corresponding actual voltage segments. Table 5 shows that the reconstruction error was small and the performance was stable across all cycles.

A virtual capacity label was adopted in the absence of a calibration cycle capacity. The corresponding results of SOH estimation are summarized in table 5. Although the running data lacks accurate capacity measurements, the auxiliary calibration curve provides reliable pseudo-labels to facilitate the reconstruction of the voltage curve. In addition, the relationship between cycle index and curve offset patterns was analyzed. The results of Pearson correlation coefficients are shown in table 5. The correlation coefficients in the table still remain at a moderate level (below 0.7) even without using the

**Table 5.** Performance metrics on three cells from field data.

| | Vol | | SOH | | |
|---|---|---|---|---|---|
| Cell | MAE | std | MAE | R2 | Pearson |
| B1-1 | 0.0753 | 0.0325 | 0.0139 | 0.9189 | 0.6614 |
| B1-2 | 0.0932 | 0.0316 | 0.0153 | 0.9067 | 0.6120 |
| B1-3 | 0.0745 | 0.0316 | 0.0148 | 0.9141 | 0.6769 |

**Table 6.** Sensitivity analysis results.

| Parameter | Value | SOH MSE ($\times 10^{-4}$) | SOH $R^2$ |
|---|---|---|---|
| *Baseline (used in main experiments)* | | | |
| — | – | 3.380 | 0.767 |
| *Phase-1 KL weight* $\beta_1$ (baseline = 0.3) | | | |
| $\beta_1$ | 0.05 | 3.305 | 0.772 |
| $\beta_1$ | 0.15 | 3.685 | 0.746 |
| $\beta_1$ | 0.50 | 3.854 | 0.734 |
| *Phase-3 KL weight* $\beta_3$ (baseline = 0.1) | | | |
| $\beta_3$ | 0.05 | 6.011 | 0.586 |
| $\beta_3$ | 0.20 | 3.638 | 0.749 |
| $\beta_3$ | 0.30 | 3.125 | 0.785 |
| *Transformer hidden dimension* $d_{\text{model}}$ (baseline = 64) | | | |
| $d_{\text{model}}$ | 32 | 5.135 | 0.646 |
| $d_{\text{model}}$ | 96 | 8.215 | 0.434 |
| $d_{\text{model}}$ | 128 | 3.309 | 0.772 |
| *Attention heads* $n_{\text{head}}$ (baseline = 4) | | | |
| $n_{\text{head}}$ | 2 | 4.219 | 0.709 |
| $n_{\text{head}}$ | 8 | 3.395 | 0.766 |
| *Latent dimension* $z_{\text{dim}}$ (baseline = 16) | | | |
| $z_{\text{dim}}$ | 8 | 3.294 | 0.773 |
| $z_{\text{dim}}$ | 24 | 3.276 | 0.774 |
| $z_{\text{dim}}$ | 32 | 3.928 | 0.729 |

calibration curve. As more operational data are used for model training in the future, the correlation and overall estimation accuracy are expected to be further improved.

This result indicates that, under the assumed virtual SOH values, the model is still capable of reconstructing the charging voltage curve. This capability enables the establishment of a predictive maintenance mode, in which impending performance degradation can be detected and mitigated in real time. This also demonstrates that the generative CVAE-TRANS framework has practical application value for large-scale BESSs.

*4.2.3. Sensitivity analysis.* The main hyperparameters have a limited impact on the model. The mean square error (MSE) of the SOH estimation in table 6 remains within the range of $3.1 \times 10^{-4}$ to $3.9 \times 10^{-4}$ for most configurations. The $R^2$ values remain consistently between 0.73 and 0.78. The impact of KL divergence weights and latent dimensions on results is also minor. After fine-tuning these parameters, it was found that the MSE decreased only by about 2%–3%. In summary, these changes had a relatively limited overall impact. The baseline configuration has already achieved a good balance in the accuracy of reconstruction and battery SOH estimation.

**Table 7.** Influence of input voltage dimension on CVAE-TRANS reconstruction performance.

| $x_{\text{dim}}$ | MAE | RMSE | $\rho$ | $F$ |
|---|---|---|---|---|
| 10 | 0.018 | 0.030 | 0.985 | 0.081 |
| 20 | 0.017 | 0.029 | 0.985 | 0.079 |
| 30 | 0.014 | 0.025 | 0.991 | 0.022 |
| 45 | 0.015 | 0.028 | 0.986 | 0.076 |
| 60 | 0.016 | 0.028 | 0.986 | 0.077 |

An input-dimension sensitivity analysis was further conducted with $x_{\text{dim}} = 10, 20, 30, 45, 60$ to verify the choice of the observed voltage length. The selected $x_{\text{dim}} = 30$ setting achieved the best average reconstruction performance, while longer inputs required more observed voltage points without improving the overall accuracy as shown in table 7.

## 5. Conclusion

This study systematically analyzed the limitations of existing battery management methods and described the complexity of actual operating scenarios. The problem of incomplete data in real scenarios urgently needs to be addressed. We developed and validated a novel CVAE-TRANS framework to reconstruct complete charging voltage curves from sparse input data. It then provides an enriched data foundation for subsequent battery SOH estimation. The experimental results show that the MAE of the reconstructed voltage does not exceed 0.018, and the linear correlation coefficient is not less than 0.97. Furthermore, the online application scenarios have verified the practicality and universality of the proposed framework. These results suggest that the proposed framework has potential for practical battery health management under incomplete-data conditions, although further validation across more battery types and operating scenarios is still required.

Future research directions include expanding the range of conditioning variables to cover more factors that affect the aging degree of batteries. Moreover, we will explore and develop complementary algorithms tailored to this framework to enhance data utilization and further improve the accuracy of the generated voltage data. We will also develop lightweight or uncertainty-aware models suitable for online applications. These efforts will focus on achieving reliable, transparent and scalable battery health management in actual energy storage systems.

## Acknowledgments

This work is sponsored by the National Natural Science Foundation of China under Grants 72471143 and 72571178.

## ORCID iD

Zhen Chen 0000-0003-2590-0307

## References


[1] Wang L, Qiu Y, Yuan W, Tian Y and Zhou Z 2025 Next-generation battery safety management: machine learning assisted life-time prediction and performance enhancement *J. Energy Chem.* **109** 726–39

[2] Xu L, Wu F, Chen R and Li L 2023 Data-driven-aided strategies in battery lifecycle management: prediction, monitoring and optimization *Energy Storage Mater.* **59** 102785

[3] Zhao J, Feng X, Pang Q, Wang J, Lian Y, Ouyang M and Burke A F 2023 Battery prognostics and health management from a machine learning perspective *J. Power Sources* **581** 233474

[4] Ding P, Liu X, Li H, Huang Z, Zhang K, Shao L and Abedinia O 2021 Useful life prediction based on wavelet packet decomposition and two-dimensional convolutional neural network for lithium-ion batteries *Renew. Sustain. Energy Rev.* **148** 111287

[5] Zraibi B, Okar C, Chaoui H and Mansouri M 2021 Remaining useful life assessment for lithium-ion batteries using CNN-LSTM-DNN hybrid method *IEEE Trans. Veh. Technol.* **70** 4252–61

[6] Chen L, Xie S, Lopes A M, Li H, Bao X, Zhang C and Li. P 2024 A new SOH estimation method for lithium-ion batteries based on model-data-fusion *Energy* **286** 129597

[7] Severson K A *et al* 2019 Data-driven prediction of battery cycle life before capacity degradation *Nat. Energy* **4** 383–91

[8] Birkl C 2017 Diagnosis and prognosis of degradation in lithium-ion batteries *PhD Thesis* University of Oxford

[9] Ma G, Xu S, Jiang B, Cheng C, Yang X, Shen Y, Yang T, Huang Y, Ding H and Yuan Y 2022 Real-time personalized health status prediction of lithium-ion batteries using deep transfer learning *Energy Environ. Sci.* **15** 4083–94

[10] Zheng W, Wu D, Yuan C, Jiang H and Wang S 2025 Advancements in battery state prediction methods: a comprehensive review of data-driven approaches and degradation mechanisms *J. Power Sources* **659** 238192

[11] Li H, Wang S, Yang L, Zhou Z, Meng Y, Zhang W, Shuan Z, Li Y and Lv F 2025 SOH estimation method for lithium-ion battery packs under real-world operating conditions based on a new attenuated model without additional experiments *Energy* **330** 136802

[12] Ali A O, Abdelrehim O, Saafan M M, Elmarghany M R and Hamed A M 2025 Comprehensive review of battery management systems for electric vehicles: thermal management, charging strategies and emerging technologies *J. Power Sources* **658** 238269

[13] Shu X, Shen J, Guo F, Liu Y, Zhang Y, Chen Z and Zhao H 2025 Towards real-world battery health intelligence: a review of machine learning advances and challenges in soh estimation *eTransportation* **26** 100509

[14] Sulzer V *et al* 2021 The challenge and opportunity of battery lifetime prediction from field data *Joule* **5** 1934–55

[15] Liu H, Li C, Hu X, Li J, Zhang K, Xie Y, Wu R and Song Z 2025 Multi-modal framework for battery state of health evaluation using open-source electric vehicle data *Nat. Commun.* **16** 1137

[16] Liu Y, Hou B, Ahmed M, Mao Z, Feng J and Chen Z 2024 A hybrid deep learning approach for remaining useful life prediction of lithium-ion batteries based on discharging fragments *Appl. Energy* **358** 122555

[17] Wu W, Chen Z, Liu W, Zhou Di, Xia T and Pan E 2025 Battery health prognosis in data-deficient practical scenarios via reconstructed voltage-based machine learning *Cell Rep. Phys. Sci.* **6** 102442

[18] Ruan H, Wei Z, Shang W, Wang X and He. H 2023 Artificial intelligence-based health diagnostic of lithium-ion battery leveraging transient stage of constant current and constant voltage charging *Appl. Energy* **336** 120751

[19] Zhao M, Zhang Y and Wang H 2024 Battery degradation stage detection and life prediction without accessing historical operating data *Energy Storage Mater.* **69** 103441

[20] Li X, Yuan C, Wang Z, He J and Yu. S 2022 Lithium battery state-of-health estimation and remaining useful lifetime prediction based on non-parametric aging model and particle filter algorithm *eTransportation* **11** 100156

[21] Cui B *et al* 2024 Ultra-early prediction of lithium-ion battery performance using mechanism and data-driven fusion model *Appl. Energy* **353** 122080

[22] Deng Z, Xu L, Liu H, Hu X, Wang B and Zhou J 2024 Rapid health estimation of in-service battery packs based on limited labels and domain adaptation *J. Energy Chem.* **89** 345–54

[23] Zhang D, Peng C, Lai X, Liu F, Liu T and Liu J 2025 Improved state of health estimation for lithium-ion batteries via multidimensional health features and long short-term memory modeling *J. Energy Storage* **132** 117852

[24] Zeng Z, Liang W, Wang T, Hong Z, Chang Q and Yang S 2025 A method for predicting the capacity of lithium-ion batteries based on Pearson correlation coefficient-guided multi-objective particle swarm optimization *Comput. Ind. Eng.* **210** 111514

[25] Li X, Ju L, Geng G and Jiang Q 2023 Data-driven state-of-health estimation for lithium-ion battery based on aging features *Energy* **274** 127378

[26] Liu W, Xu Y and Feng X 2020 A hierarchical and flexible data-driven method for online state-of-health estimation of li-ion battery *IEEE Trans. Veh. Technol.* **69** 14739–48

[27] Channegowda J, Maiya V and Lingaraj C 2023 A deep learning approach towards generating high-fidelity diverse synthetic battery datasets *IEEE Trans. Ind. Appl.* **59** 4620–7

[28] Kim J, Moon H, Yoon K, Chun H, Lee M, Ko J and Han S 2025 A physics-driven generative model to accelerate artificial intelligence development for lithium-ion battery diagnostics *Appl. Energy* **391** 125873

[29] Yang H, Peng S, Chen C, Tang A, Yu Q, Zhang J and Li. R 2026 Data-augmented SOH estimation for lithium-ion batteries under small-sample conditions: a hybrid STL-transformer-TimesNet approach *Energy* **342** 139463

[30] Kang W, Xiao J and Xue J 2023 Generative knowledge-based transfer learning for few-shot health condition estimation *Complex Intell. Syst.* **9** 965–79

[31] Guo N, Chen S, Tao J, Liu Y, Wan J and Li. X 2024 Semi-supervised learning for explainable few-shot battery lifetime prediction *Joule* **8** 1820–36

[32] Yao J and Han T 2023 Data-driven lithium-ion batteries capacity estimation based on deep transfer learning using partial segment of charging/discharging data *Energy* **271** 127033

[33] Chen Z, Liu W, Zhou Di, Xia T and Pan E 2025 Inconsistency identification for lithium-ion battery energy storage systems using deep embedded clustering *Appl. Energy* **388** 125677

[34] Wu W, Chen Z, Liu W and Pan E 2025 Correlation based-graph neural network for health prognosis of non-fully charged and discharged lithium-ion batteries *J. Power Sources* **629** 235984



**Wei Wu** received a BS degree in Mechanical Design from Nanjing university of science and technology in 2018 and MS degrees in Mechanical Design from the University of Manchester, UK, in 2019. He then received his PhD degree in the Department of Industrial Engineering & Management, Shanghai Jiao Tong University. His research interests include battery health management, health state and remaining useful life prediction.



**Zhen Chen** received a BS degree in Quality and Reliability Engineering from Beihang University in 2014 and a PhD degree in Mechanical Engineering from Shanghai Jiao Tong University in 2019. He is now an associate professor in the Department of Industrial Engineering & Management, Shanghai Jiao Tong University. He worked as a postdoc at the Department of Energy, Politecnico di Milano, Italy. His research interests include reliability engineering, battery health management and condition monitoring of complex systems.



**Weijie Liu** received a BS degree in Statistics and an MS degree in Applied Statistics from Xiamen University in 2020 and in 2023. He is currently pursuing a PhD degree in the Department of Industrial Engineering & Management, Shanghai Jiao Tong University, China. His research interests include health management of energy storage systems and industrial big data analysis.



**Ershun Pan** received BS and MS degrees in Mechanical Design and Manufacturing from Northeastern University, Shenyang, China, in 1997 and a PhD degree in Mechanical Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2000. He is now a professor and the header in the Department of Industrial Engineering & Management, Shanghai Jiao Tong University, China. His current research interests include quality and reliability engineering, production system optimization, and lean manufacturing.