

Article

Transfer Learning-CNN-LSTM-Based Insulation State Prediction for Energy Storage Systems

Yong Qiu ¹, Hanlin Liu ^{2,*}, Baohong Lu ¹, Yan Chen ^{2,*}, Aiwei Guan ¹ and Tianyan Jiang ²

¹ XYZ Storage Technology Corp. Ltd., Beijing 102402, China; qiuyong1980@163.com (Y.Q.); jaslu@163.com (B.L.); guanaiwei2026@163.com (A.G.)

² School of Electrical and Electronic Engineering, Chongqing University of Technology, Chongqing 400054, China

* Correspondence: 15983070297@163.com (H.L.); chenyan2012@cqut.edu.cn (Y.C.)

Abstract

Accurate estimation of insulation resistance, denoted as R_{iso} , in high-voltage direct current energy storage systems plays a pivotal role in leakage protection and thermal runaway suppression. Conventional physical measurement techniques are inherently susceptible to distortions under dynamic operating conditions due to interference from parasitic capacitance. Meanwhile, emerging data-driven approaches are often bottlenecked by cross-domain distribution shifts and the scarcity of annotated full-lifecycle data. This study proposes a hybrid framework that integrates transfer learning and residual correction within a CNN-LSTM architecture, referred to as TL-CNN-LSTM + Corr. Utilizing seven-dimensional operational features as inputs, the framework employs a one-dimensional convolutional neural network to extract high-frequency transient response patterns. Simultaneously, a long short-term memory network models the long-term, non-stationary temporal evolution of insulation degradation. To circumvent systemic biases across varying scenarios, a three-stage domain adaptation strategy consisting of pre-training, freezing, and fine-tuning was developed, which is complemented by a lightweight linear residual compensator designed to rectify amplitude drifts during abrupt operational transitions. Independent evaluations using 500 sets of real-world operational data demonstrate that the proposed model achieves high-precision predictions, yielding a root mean square error of 40.595, a mean absolute error of 32.919, and an R^2 value of 0.941. Furthermore, the model exhibits remarkable robustness against sensor noise and data loss. By ensuring cross-domain predictive consistency with minimal computational overhead, this framework provides a highly reliable and deployable solution for online insulation state monitoring in edge-side battery management systems.

Keywords: energy storage systems; insulation resistance; transfer learning; CNN-LSTM; residual correction

Academic Editors: Friedhelm Schwenker and Fabio Corti

Received: 19 April 2026

Revised: 22 May 2026

Accepted: 4 June 2026

Published: 1 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Driven by the rapid deployment of large-scale energy storage systems for renewable energy integration and grid services, system-level operational safety has emerged as a paramount concern [13]. Within high-voltage direct current energy storage systems, insulation performance—specifically the real-time insulation resistance, denoted as R_{iso} —serves as the cornerstone for mitigating short-circuit risks, preventing thermal runaway,

and maintaining vital safety margins 47. Although physical measurement techniques utilizing balanced bridges or alternating current injection are widely adopted, they are inherently vulnerable to resistor–capacitor delay effects induced by large parasitic capacitances in high-voltage environments. To overcome these delays, foundational designs of insulation detection systems 8,9 have increasingly incorporated advanced algorithms, such as extended Kalman filtering 1012 and model-based diagnostic frameworks 1313. However, despite these algorithmic improvements, these methods frequently suffer from delayed responses and severe measurement distortions under dynamic operating conditions. For instance, Zoeller et al. 15 specifically demonstrated how large parasitic capacitances critically distort transient current signals, underscoring the physical limitations of conventional hardware-based diagnostics.

To better contextualize the severity of these delayed responses, it is essential to outline the specific temporal resolution constraints of this problem. From a physical perspective, capturing the rapid dynamics of transient leakage currents and preventing localized thermal runaway precursors typically dictates a required temporal resolution of less than 100 milliseconds. However, constrained by the substantial RC delays in large parasitic capacitances, the temporal resolution achieved today in practical field deployments (e.g., standard BMS balanced bridge methods) typically lags significantly, often ranging from 2 to 10 s. Furthermore, while the best published laboratory results utilizing advanced AC injection or sophisticated filtering techniques can achieve a resolution of approximately 0.5 to 1.0 s, these are mostly realized under idealized, noise-free conditions, leaving a significant gap for real-world dynamic applications.

Data-driven approaches offer a compelling “soft-sensing” alternative to physical measurements, shifting the paradigm from internal sensing to big data utilization 16. Researchers have actively investigated data-driven methods for estimating internal ohmic resistance 1718 and overall battery states 19. Building upon these advancements, machine learning techniques have subsequently been applied to insulation monitoring to optimize testing 20, estimating R_{iso} by mapping multi-source telemetric features, such as positive and negative pole-to-ground resistances (R_p and R_n) and temperature. However, the engineering deployment of this paradigm is hampered by stringent boundary constraints. Advanced frameworks leveraging cloud-based AI 21, digital twins 22, and comprehensive laboratory-to-field datasets 2324 require immense computational resources. Furthermore, while comparative studies on deep learning models 2527 show high accuracy, the highly variable operating conditions of energy storage systems induce significant distribution shifts. To address this, recent studies have explored domain generalization 28 and deep domain adaptation 2930 for battery health predictions. Unfortunately, these complex monolithic neural frameworks typically incur high computational costs and memory overheads, precluding their direct deployment on resource-constrained edge-side microcontrollers. Consequently, they struggle to strike a viable balance between high-fidelity feature extraction and the correction of cross-domain biases in real-time edge environments.

To address the coexistence of temporal non-stationarity and distribution shifts, this study proposes a predictive framework integrating transfer learning with a CNN-LSTM architecture, augmented by a residual correction mechanism. Specifically, a one-dimensional convolutional neural network leverages its local receptive fields to act as a natural filter for high-frequency transient noise. Simultaneously, a long short-term memory layer performs gated feature compression to capture the long-term physical degradation trends of the insulation. Furthermore, to overcome the fitting bottleneck associated with small-sample target domains, a three-stage transfer strategy is employed to align deep feature distributions. This is coupled with a lightweight residual compensator cascaded at the

inference stage, designed to algebraically isolate and eliminate the systemic, deterministic biases triggered by shifts in operating conditions.

The novel contributions of this work center on achieving robust, edge-deployable insulation resistance prediction under dynamic conditions through three interconnected innovations:

1. Feature Representation and Decoupled Modeling: A dual-scale CNN-LSTM architecture is proposed to physically decouple high-frequency transient noise from slow cumulative degradation trends.
2. Model Optimization and Cross-Domain Alignment: A tailored three-stage transfer learning strategy overcomes the small-sample bottleneck, efficiently aligning deep features to mitigate performance degradation caused by operational shifts.
3. Practical Deployment and Dynamic Bias Elimination: A lightweight residual compensator is integrated to algebraically eliminate systematic biases during inference, ensuring high predictive consistency with minimal computational overhead.

2. Materials and Methods

2.1. Dataset Description and Preprocessing

2.1.1. Data Collection and Operating Conditions

Data were acquired from the online monitoring platform of a high-voltage direct current energy storage system. The dataset encompasses typical operational states—such as charging, discharging, and resting—alongside dynamic events including power perturbations, bus voltage fluctuations, and load transfers. The monitoring system continuously outputs the positive and negative pole-to-ground insulation resistances (R_p and R_n) and the synthesized insulation resistance (R_{iso}), while synchronously recording the temperature (T). To better capture asymmetric leakage currents and transient variations, derivative features were engineered, including the resistance ratio (R_p/R_n), the degree of asymmetry (asym), and the first-order differences of the resistances (dR_p and dR_n).

To authentically simulate the cross-domain distribution shifts induced by variable operating conditions in real-world engineering applications, this study eschews the conventional random splitting method. Instead, a non-independent and identically distributed (non-I.I.D.) dataset was constructed based on distinct operational profiles. Specifically, the 500 valid sample sets were partitioned into a source domain and a target domain:

- Source Domain (D_s): Comprises 320 datasets collected under conventional, relatively stable conditions, such as standard constant-current charging, discharging, and prolonged resting. This subset is utilized for the large-scale pre-training of the model to map the fundamental physical degradation of the insulation.
- Target Domain (D_t): Contains 180 datasets characterized by complex dynamic evolutions, including severe power perturbations, transient bus voltage fluctuations, and load transfers. Because the resistor–capacitor delay effects—induced by parasitic capacitances in high-voltage systems—are particularly pronounced during dynamic operations, the target domain inherently exhibits systemic biases and feature drifts relative to the source domain.

To align with a few-shot transfer learning paradigm, the overall dataset was divided into a training set of 350 samples and an independent testing set of 150 samples. The training set consists of the entire 320-sample source domain merged with 30 randomly selected target domain samples. These 30 target samples serve exclusively as a fine-tuning set for transfer adaptation and the fitting of the residual compensator. The remaining 150 target domain samples are strictly reserved as an independent testing set to evaluate the final generalization performance of the model.

2.1.2. Data Structure and Variable Definitions

The dataset is structured for a multi-variate regression task, mapping multi-source operational features to the target insulation resistance. At any given time step t , the input feature vector is defined as:

$$X_t = [R_p, R_n, R_p / R_n, asym_t, T_c, dR_p, dR_n]^T \in R^7, \quad (1)$$

The corresponding output is the synthesized insulation resistance, denoted as $y_t = R_{iso,t}$. The detailed definitions of these features and variables are summarized in Table 1.

Table 1. Definitions of input features and output label.

Name	Symbol	Unit	Note
Positive-to-ground resistance	R_p	k Ω	measured/estimated
Negative-to-ground resistance	R_n	k Ω	measured/estimated
Resistance ratio	R_p/R_n	–	imbalance indicator
Asymmetry	asym	–	$(R_p - R_n)/(R_p + R_n)$
Temperature	T	°C	monitoring point
ΔR_p (1st diff.)	dR_p	k Ω /step	transient change
ΔR_n (1st diff.)	dR_n	k Ω /step	transient change
Output: Insulation resistance	R_{iso}	k Ω	synthesized target

2.1.3. Data Preprocessing and Input Construction

To enhance the convergence stability and generalization capability of the CNN-LSTM model across diverse operating conditions and cross-domain scenarios, a three-step preprocessing pipeline was implemented: anomaly filtering, dual-tier normalization, and input formatting.

1. Anomaly filtering: Non-physical values (such as NaN or Inf) induced by transient shocks, measurement noise, or communication packet loss were either removed or imputed. This step ensures the validity and numerical stability of the training dataset.
2. Dual-tier normalization: A sequential normalization strategy consisting of Z-score and Min-Max scaling was applied. Initially, Z-score standardization was utilized to eliminate dimensional disparities across features and mitigate the impact of extreme outliers. Subsequently, the features were scaled to a range of [0, 1] using Min-Max normalization, thereby facilitating efficient network training and cross-domain feature alignment.
3. Input construction via sliding windows: To capture the temporal dependencies inherent in the evolution of insulation resistance, a sliding window mechanism was employed to construct the sequential inputs. The window length and stride were set to $T = 10$ and $s = 1$, respectively. Consequently, each sample is formulated as a $T \times 7$ tensor, which maps to the target insulation resistance value, R_{iso} , at the subsequent time step. The detailed configurations of the model architecture and training parameters are summarized in Table 2.

Table 2. Model and training settings.

Module	Setting (Condensed)	Module	Setting (Condensed)
Input/Output	$d = 7$ (Rp, Rn, Rp/Rn, asym, T, dRp, dRn) $\rightarrow$ 1 (Riso)	Input/Output	$d = 7$ (Rp, Rn, Rp/Rn, asym, T, dRp, dRn) $\rightarrow$ 1 (Riso)
CNN + GAP	Conv1D $\times$ 2: Filters 16/32, Kernel = 3, Padding = same; GAP	CNN + GAP	Conv1D $\times$ 2: Filters 16/32, Kernel = 3, Padding = same; GAP
LSTM	12 units, OutputMode = last	LSTM	12 units, OutputMode = last
Regressor	FC24 + ReLU + Dropout 0.3 + FC1	Regressor	FC24 + ReLU + Dropout 0.3 + FC1
Optimizer	Adam; Batch = 20; GradThresh = 1	Optimizer	Adam; Batch = 20; GradThresh = 1
Schedule	stage-wise: pre-train $\rightarrow$ freeze&adapt $\rightarrow$ fine-tune; Epoch 200–500; LR 1e-4–1e-2 (piecewise)	Schedule	stage-wise: pre-train $\rightarrow$ freeze&adapt $\rightarrow$ fine-tune; Epoch 200–500; LR 1e-4–1e-2 (piecewise)
Validation	No early stopping (fixed MaxEpochs)	Validation	No early stopping (fixed MaxEpochs)
Normalization	Z-score $\rightarrow$ Min-Max (fit on train only; apply to test)	Normalization	Z-score $\rightarrow$ Min-Max (fit on train only; apply to test)

2.2. CNN-LSTM-Based Temporal Modeling

It is worth emphasizing that the integration of the CNN-LSTM architecture in this study deviates significantly from a textbook, plug-and-play application of generic mathematical tools. Standard textbook applications typically treat time-series inputs as purely homogeneous numerical sequences, directly feeding them into vanilla networks without considering underlying physical heterogeneities. In contrast, our customized structural design is strictly governed by the specific dual-scale physical characteristics inherent to insulation resistance (R_{iso}) evolution: high-frequency transient operational noise and slow, cumulative degradation trends. Instead of arbitrary hyperparameter tuning, the receptive fields of the 1D-CNN are mathematically tailored to act as local non-parametric feature filters for transient power perturbations, while the LSTM gating mechanisms are explicitly aligned to preserve long-term degradation history. This tailored coupling achieves decoupled modeling of distinct error sources, creating a specialized soft-sensing paradigm capable of countering the severe measurement distortions induced by high-voltage resistor-capacitor delay effects.

2.2.1. 1-D CNN with Global Average Pooling

Specifically, based on the configurations detailed in Table 2, the 1D-CNN module consists of two hidden convolutional layers with 16 and 32 filters, respectively, to effectively extract local high-frequency features.

Recognizing that the evolution of insulation resistance encompasses both transient operational noise and long-term degradation trends, this study employs a one-dimensional convolutional neural network (1D-CNN) to capture local perturbation features from the time-domain signals. This is coupled with global average pooling (GAP) to achieve effective feature dimensionality reduction and non-parametric denoising. Let the input sequence be denoted as $\mathbf{X} = \{x_t\}_{t=1}^T$. Assuming a convolutional kernel length of K ,

C output channels, a stride of $s = 1$, and the application of zero-padding to maintain the “same” temporal length, the convolution output $Z_t^{(c)}$ for the c -th channel at time step t is formulated as

$$Z_t^{(c)} = \phi \left(\sum_{k=1}^K w_k^{(c)} \cdot x_{t-k} + b_c \right), \quad (2)$$

where $w_k^{(c)}$ and b_c represent the convolution kernel weights and bias for the c -th channel, respectively, while ϕ denotes the nonlinear activation function. Subsequently, GAP is performed along the temporal dimension for each channel to yield a compact feature representation:

$$g_c = \frac{1}{T} \sum_{t=1}^T z_t^{(c)}, \quad g \in R^C, \quad (3)$$

The resulting compact vector g serves as the input to the subsequent LSTM regression head for predicting Riso, effectively forming a seamless pipeline of “local pattern extraction, global statistical compression, and temporal modeling.

2.2.2. Time Series Modeling of Insulation Resistance Based on LSTM

Compared to traditional Recurrent Neural Networks (RNNs) that often suffer from the vanishing gradient problem when processing long sequences, the LSTM model was specifically chosen because its unique gating mechanisms (forget, input, and output gates) are highly capable of capturing long-term temporal dependencies. This makes it structurally superior for modeling the slow, cumulative physical degradation trends of insulation materials alongside transient state changes.

The insulation resistance Riso is simultaneously influenced by transient operational perturbations and long-term cumulative degradation, thereby exhibiting highly dynamic characteristics defined by the coexistence of non-stationarity and short-/long-term dependencies. To effectively model these temporal dependencies, this study employs a Long Short-Term Memory (LSTM) network to perform gated updates on the feature sequences. At time step t , the input is denoted as z_t (derived from the CNN + GAP module; if the CNN is omitted, $z_t = x_t$). This input, combined with the hidden state from the previous time step h_{t-1} , is utilized to update the cell state c_t and the current hidden state h_t as follows:

$$f_t = \sigma(W_f[h_{t-1}, z_t] + b_f), \quad (4)$$

$$i_t = \sigma(W_i[h_{t-1}, z_t] + b_i), \quad (5)$$

$$\tilde{c}_t = \tanh(W_c[h_{t-1}, z_t] + b_c), \quad (6)$$

$$C_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \quad (7)$$

$$o_t = \sigma(W_o[h_{t-1}, z_t] + b_o), \quad (8)$$

$$h_t = o_t \odot \tanh(C_t), \quad (9)$$

where $\sigma(\cdot)$ denotes the Sigmoid activation function, and $\odot$ represents the Hadamard product (element-wise multiplication). To achieve explicit regression prediction, a linear regression head is adopted to map the hidden state h_t to the target variable:

$$\hat{R}_{iso}(t) = w_r^T h_t + b_r, \quad (10)$$

The model is optimized using the Mean Squared Error (MSE) as the objective function:

$$L = \frac{1}{N} \sum_{t=1}^N \left(\hat{R}_{iso}(t) - R_{iso}(t) \right)^2 \quad (11)$$

Through the aforementioned modeling architecture, the network concurrently characterizes both the long-term degradation trends and the short-term perturbation responses, thereby substantially enhancing the predictive stability and generalization capability across diverse operational conditions.

2.3. Three-Stage Transfer Learning Strategy

In practical engineering applications, target platforms (e.g., newly deployed battery clusters, novel operating conditions, or different sampling environments) typically suffer from a scarcity of annotated samples. Training a deep model from scratch under such data-limited conditions is highly susceptible to overfitting. To mitigate this, we transfer the universal representations learned from the source domain D_s to the target domain D_t . A three-stage “pre-training, feature transfer, and fine-tuning” strategy is adopted to substantially enhance the stability and generalization capability of Riso prediction under few-shot conditions.

(1) Pre-training: The CNN-LSTM model is initially trained on the source domain dataset $D_s = \{(x_i^{(s)}, R_{iso,i}^{(s)})\}_{i=1}^{N_s}$ to obtain the optimal initial parameters, denoted as $\theta^{(0)}$.

(2) Feature Transfer: The parameters of the CNN module (serving as the feature extractor) are frozen, and only the LSTM layer and the regression head are updated. This mechanism enables the model to rapidly adapt to the target domain distribution using a small subset of target samples $D_t = \{(x_j^{(t)}, R_{iso,j}^{(t)})\}_{j=1}^{N_t}$.

(3) Fine-tuning: The higher layers of the CNN are gradually unfrozen, and the entire network is optimized end-to-end utilizing a considerably smaller learning rate to further fuse the source and target features. During target domain training, a supervised loss is employed alongside a parameter constraint term to suppress catastrophic forgetting:

$$\min_{\theta} L_t(\theta) + \lambda \|\theta - \theta^{(0)}\|_2^2 \quad (12)$$

$$L_t(\theta) = \frac{1}{N_t} \sum_{j=1}^{N_t} \ell(\hat{R}_{iso,j}^{(t)}, R_{iso,j}^{(t)}) \quad \ell(\cdot) \in \{MSE, Huber\} \quad (13)$$

where λ represents the trade-off coefficient, which is designed to restrict the updated parameters from deviating excessively from the pre-trained solution. This regularization mechanism stabilizes the transfer process and subsequently bolsters the predictive performance on the target domain.

To intuitively illustrate the engineering implementation of the aforementioned theoretical model and transfer learning strategy, Figure 1. presents the complete topological schematic of the TL-CNN-LSTM hybrid prediction architecture. As depicted in the figure, the network utilizes a 1D-CNN and a Global Average Pooling (GAP) layer as front-end feature filters, employs an LSTM layer as the core temporal evolution model, and incorporates a fully connected layer at the end to execute state mapping. To address the critical challenge of scarce annotated data on target platforms (e.g., novel operating conditions or newly deployed battery clusters), this architecture deeply integrates the previously proposed three-stage domain adaptation mechanism within its optimization pipeline. Specifically, it anchors prior knowledge from the source domain via “pre-training,” circumvents small-sample overfitting through “feature freezing and transfer,” and ultimately achieves

end-to-end feature alignment in the target domain via “joint fine-tuning.” The tight coupling of this structural design and training strategy enables the model to rapidly and robustly generalize to real-world engineering scenarios characterized by complex distribution shifts, requiring only a minimal sample cost.

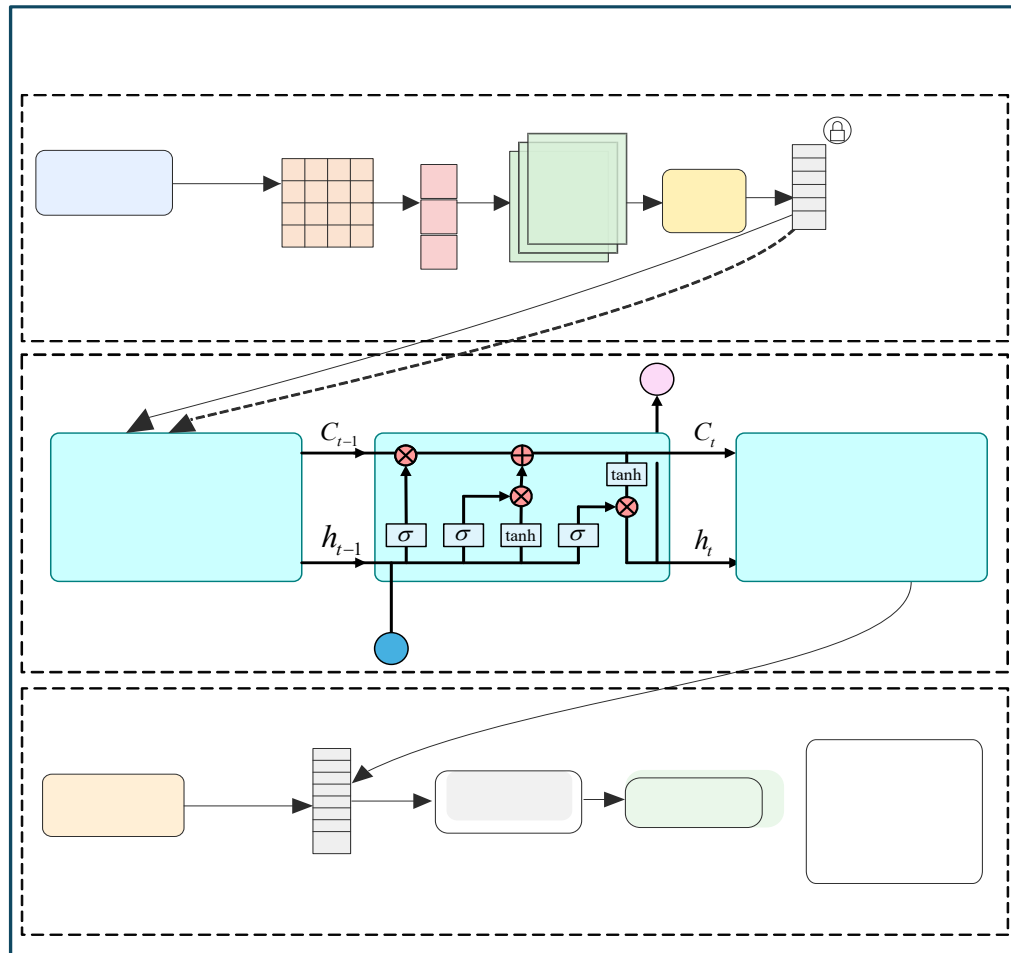


Figure 1. Schematic of the transfer learning-driven CNN-LSTM framework for Riso prediction. (Padlocks denote frozen CNN weights; red crosses/pluses indicate element-wise multiplication/addition in LSTM; solid arrows standard flows, dashed transfer paths.)

2.4. Residual Correction Module

Although the three-stage transfer learning strategy effectively aligns the source and target feature distributions, the end-to-end fine-tuned model may still exhibit residual low-order systemic biases in its amplitude response, primarily constrained by the scarcity of annotated target samples. To compensate for these linearly separable bias components, a lightweight residual correction module is integrated during the inference phase. This module precisely rectifies the systematic mean drift of the output without necessitating the retraining of the deep neural network.

Given the input feature X_t and the preliminary prediction $\hat{y}_t$, the residual is defined as $r_t = y_t - \hat{y}_t$. Assuming that the residual is composed of a deterministic linear bias and random noise:

$$r_t = b(z_t) + \varepsilon_t, \quad z_t = [1, X_t^T, \hat{y}_t]^T, \quad \hat{r}_t = \phi^T z_t, \quad (14)$$

where the optimal parameter vector ϕ^* is obtained by minimizing the regularized residual loss over the training set:

$$\phi^* = \arg \min_{\phi} \sum_{t \in D_{tr}} (r_t - \phi^T z_t)^2 + \lambda \|\phi\|_2^2, \quad (15)$$

The final corrected output is formulated as:

$$\tilde{y}_t = \hat{y}_t + \hat{r}_t, \quad (16)$$

This module effectively enhances cross-domain predictive consistency while maintaining a remarkably low computational overhead, rendering it highly conducive to practical edge-side deployment.

3. Results

- This section systematically validates the proposed transfer learning-enhanced CNN-LSTM and its associated residual correction framework. For clarity and conciseness, the models are denoted as follows throughout the subsequent text: TL refers to the transfer learning CNN-LSTM model (i.e., TL-CNN-LSTM), and TL + Corr denotes the complete framework incorporating the residual correction module (i.e., TL-CNN-LSTM + Corr). The baseline models retain their standard acronyms: CNN-LSTM, TCN, GRU, and SVR. Beyond overarching error metrics (RMSE, MAE, and R^2), this section comprehensively evaluates the generalization capability and practical deployability of each model on real-world energy storage operational data. The assessment spans multiple dimensions, including trend consistency, amplitude fidelity, robustness against perturbations, and the statistical structure of prediction errors. To ensure a fair and rigorous comparison, all models share an identical input feature architecture (7 inputs/1 output), standardized data cleaning and normalization pipelines, and uniform evaluation criteria. The testing set is kept strictly independent and is exclusively utilized for the final performance reporting post-training.

3.1. Model Training and Optimization

- All deep learning models in our study, including the proposed CNN-LSTM architecture and baseline models, were programmed and implemented using MATLAB R2022b.
 - During the model training and evaluation phases, this study strictly adheres to the aforementioned condition-based non-independent and identically distributed (non-I.I.D.) partitioning strategy. The model utilizes 350 sample sequences as the training set and 150 target domain sample sequences as the testing set. The training process is driven by the Adam optimizer (batch size = 20), employing a predefined maximum number of epochs (MaxEpochs) coupled with a piecewise learning rate decay schedule. Notably, neither a validation set nor early stopping mechanisms are utilized, explicitly preventing the testing data from inadvertently influencing any training decisions. The specific training pipeline proceeds as follows:
4. Pre-training: The baseline CNN-LSTM model is initially pre-trained using the 320 source domain samples collected under stable operating conditions.
 5. Transfer Adaptation: The feature distribution undergoes transfer fine-tuning (TL) via the proposed three-stage strategy, exclusively utilizing the 30 dynamic target domain samples embedded within the training set.
 6. Residual Correction: The residual correction module is linearly fitted relying solely on the aforementioned 30-sample fine-tuning set. Subsequently, during the inference phase, it executes output compensation across the 150 independent testing samples.

Figure 2 illustrates the training loss trajectories, demonstrating that the transfer learning-enhanced TL model achieves faster convergence and a smoother loss reduction compared to the baseline CNN-LSTM. This observation directly corroborates the efficacy of the proposed transfer learning strategy.

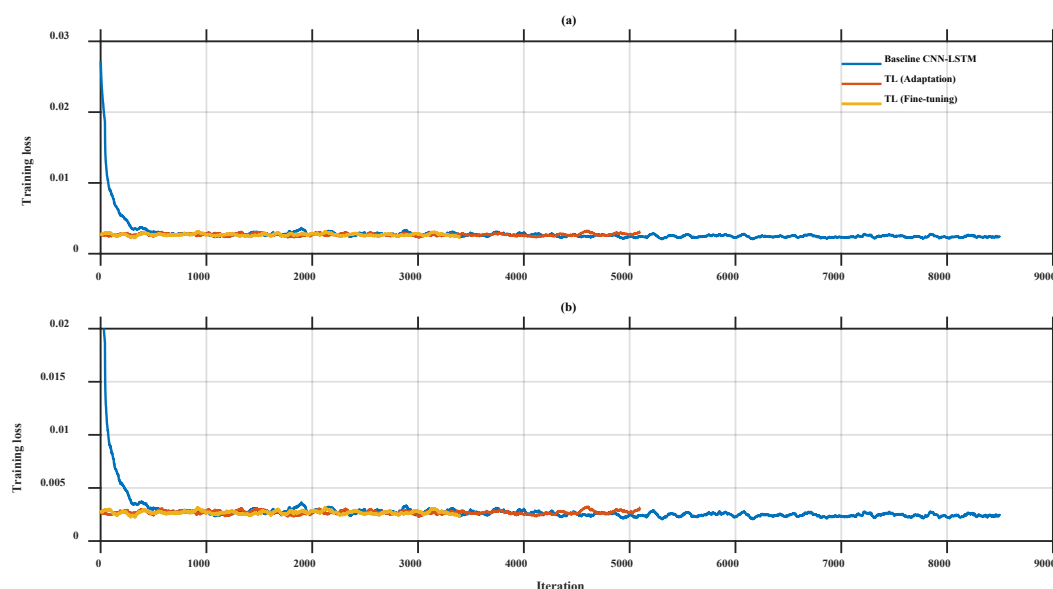


Figure 2. Training loss curves: (a) global; (b) zoomed-in.

The overarching experimental framework is anchored by the progressive evaluation of the CNN-LSTM, TL, and TL + Corr architectures, which are concurrently benchmarked against representative baseline models, including TCN, GRU, and SVR. Furthermore, to ensure an equitable and rigorous assessment, all models are evaluated on the identical independent testing set utilizing uniform performance metrics.

3.2. Performance Metrics and Accuracy Evaluation

- The proposed methodology is fundamentally underpinned by the physical correlations between the input features (e.g., R_p , R_n , temperature, and their derivatives) and the underlying insulation degradation. To rigorously validate the efficacy of the transfer learning and residual correction mechanisms, this study contrasts the predictive performance of the original CNN-LSTM, the TL model, and the complete TL + Corr framework. Furthermore, a comprehensive horizontal comparison is conducted against typical baseline models, including TCN, GRU, and SVR. Table 3 presents the quantitative results, specifically the RMSE, MAE, and R^2 metrics, evaluated on the independent testing set.

Table 3. Performance comparison of original CNN-LSTM and transfer learning CNN-LSTM models.

Dataset	Model	RMSE	MAE
Test	CNN-LSTM	47.432	38.848
Test	TL	42.765	35.205
Test	TL + Corr	40.595	32.919
Test	TCN	43.533	34.808
Test	GRU	42.052	33.980
Test	SVR	58.904	40.574

Quantitative Results: Compared to the baseline CNN-LSTM, the TL model exhibits superior performance across all evaluated metrics. This substantiates that transfer learning effectively mitigates the degradation in generalization capabilities induced by cross-condition distribution shifts. Furthermore, the integration of the residual correction module (TL + Corr) yields the optimal predictive performance (RMSE = 40.595, MAE = 32.919, $R^2 = 0.941$). This demonstrates that the residual compensation successfully curtails systemic biases, thereby significantly enhancing both absolute accuracy and cross-scenario predictive consistency.

Trend and Amplitude Consistency: Beyond overarching statistical metrics, the dynamic behavior of the models is evaluated via two visualization paradigms: (1) Temporal tracking profiles (Figure 3), which assess the model's capacity to track abrupt transient changes and peak-to-valley fluctuations (i.e., mitigating temporal lag and amplitude offsets); and (2) Scatter consistency plots, coupled with detailed error statistics (refer to Section 3.4), utilized to detect systemic biases and the degree of error concentration. Figure 3 compares the predicted R_{iso} values against the ground truth on the testing set, demonstrating that the TL + Corr framework possesses superior tracking fidelity, particularly during abrupt transient events and extreme peak-to-valley regions.

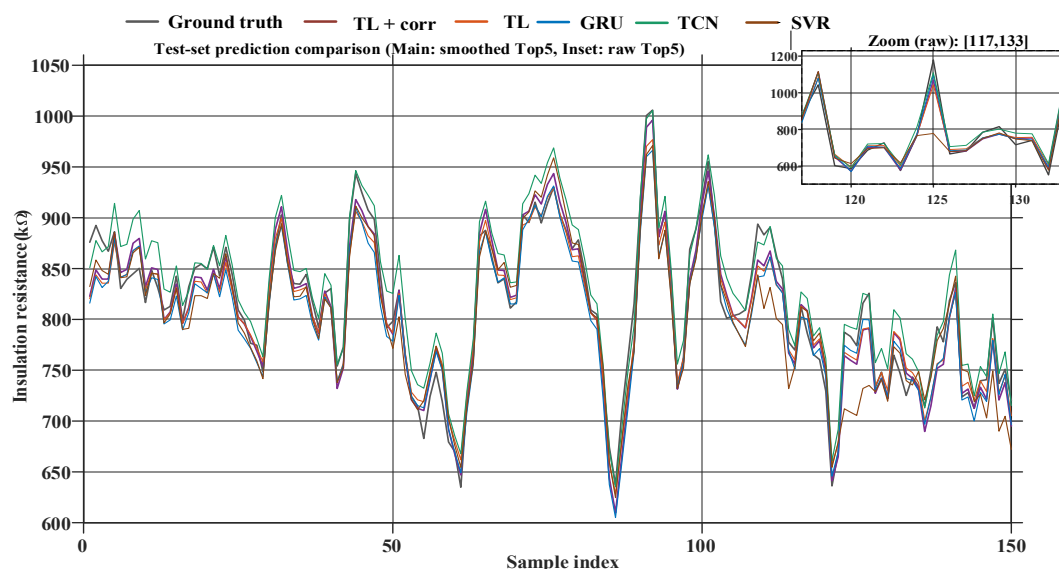


Figure 3. Comparison of R_{iso} predictions by different models.

3.3. Robustness Evaluation Under Perturbations

Achieving low prediction errors exclusively on pristine data does not inherently guarantee the practical deployability of a model. To rigorously evaluate the model's robustness against measurement noise, data loss, and sensor drift, we froze the learned model parameters, injected controlled perturbations exclusively into the testing inputs, and quantified the resultant performance variations. Three specific perturbation scenarios were configured:

7. Noise ($\sigma = 0.03$): Zero-mean Gaussian noise was superimposed onto each input channel within the normalized domain.
8. Missing (10%): Ten percent of the feature values were randomly nullified and subsequently imputed using mechanisms strictly consistent with online operational protocols.
9. Drift (+0.03): A fixed bias was injected into critical channels to emulate the slow drift characteristic of aging sensors.

The relative performance degradation metrics—specifically $\Delta RMSE$, ΔMAE , and ΔR^2 —were calculated for comprehensive assessment. As illustrated in Figure 4., both the TL and TL + Corr models exhibit highly controlled performance degradation across all perturbation scenarios. The TL model inherently bolsters feature robustness through cross-domain alignment, while the TL + Corr framework further curtails residual systemic biases. Consequently, the TL + Corr model achieves consistently smaller increments in $\Delta RMSE$ and ΔMAE , alongside a more contained reduction in R^2 . This synergistic advantage renders the proposed framework exceptionally well-suited for robust online deployment in real-world scenarios.

Furthermore, while traditional model-based measurements rely on analytical sensitivity equations to determine which physical circuit changes necessitate system re-analysis, the proposed framework achieves equivalent operational insight through its hierarchical domain adaptation mechanisms. Specifically, minor physical variations—such as gradual sensor aging or slight ambient thermal drifts—manifest as linear systemic biases in the multi-variate feature space. Our framework inherently absorbs these minor variations via the lightweight Residual Correction Module without requiring a full network retrain, indicating that such minor changes can be computationally “ignored.” Conversely, fundamental physical circuit alterations—such as battery pack replacements, significant high-voltage bus topology modifications, or severe sensor faults—induce drastic, non-linear cross-domain distribution shifts. In our deployment paradigm, when the real-time compensated residual errors persistently exceed a predefined safety threshold, it serves as a deterministic trigger for re-analysis. This dynamically prompts the Transfer Adaptation phase, utilizing a small subset of newly acquired target data to fine-tune the frozen network architecture. Thus, the framework provides clear, data-driven guidance on managing hardware variations, mirroring the sensitivity insights of physical models.

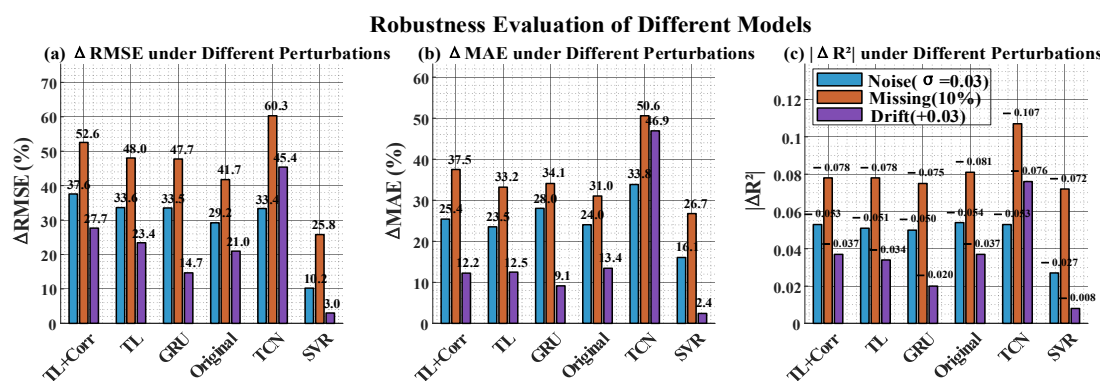


Figure 4. Robustness evaluation of different models under perturbations.

3.4. Error Structure and Residual Characteristics

To comprehensively assess the predictive reliability of the proposed model under cross-domain scenarios, this section performs an in-depth diagnostic analysis across three dimensions: temporal error correlation, statistical distribution, and prediction consistency. Figure 5. illustrates the absolute error sequences alongside the scatter consistency plots for the evaluated models. The experimental results demonstrate that the TL + Corr framework achieves the highest alignment with the $y = x$ reference line across the entire measurement range. Notably, in high-resistance regions, conventional models such as SVR and TCN exhibit significant heteroscedasticity due to the impact of cross-domain distribution shifts, with scatter points becoming sparse and deviating markedly from the ideal line. In contrast, the TL + Corr model maintains a tight prediction envelope, effectively suppressing the variance expansion triggered by operational transitions. Furthermore, the error

sequences in Figure 5a confirm that the residual correction mechanism successfully eliminates the dynamic peak errors prevalent in the baseline CNN-LSTM architecture.

The residual statistical characteristics presented in Figure 6, further substantiate the model's reliability. The cumulative distribution function (CDF) curves in Figure 6a reveal that over 90% of the relative absolute errors for the TL + Corr model are concentrated within a narrow interval of less than 0.1, significantly outperforming the original CNN-LSTM. As shown in the boxplots in Figure 6b, TL + Corr not only achieves the lowest median error but also remarkably compresses the interquartile range (IQR) while drastically reducing the number of extreme outliers.

Such a near-Gaussian, zero-centered residual structure is of paramount importance for industrial-grade Battery Management Systems (BMS). This suggests that the residual correction module has successfully isolated systemic biases, leaving the remaining errors to consist primarily of irreducible stochastic noise. In practical applications, this error distribution facilitates more precise settings for insulation alarm thresholds, thereby minimizing the risk of false alarms induced by model drift.

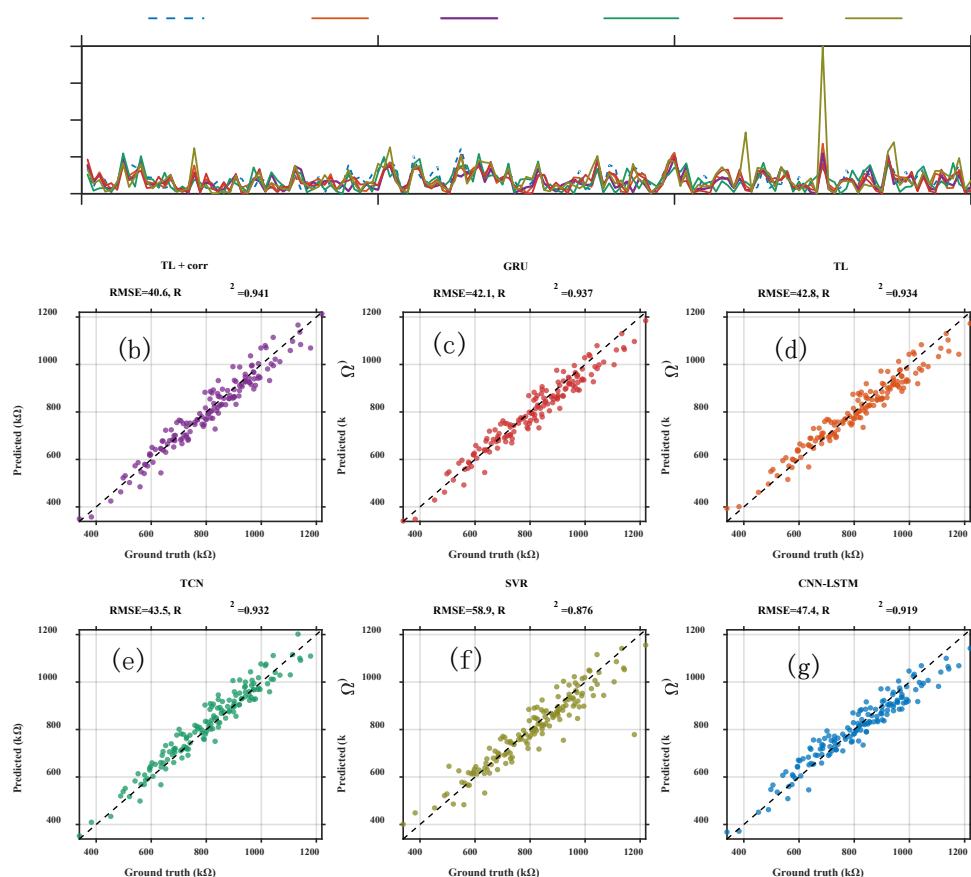


Figure 5. Error analysis on test set: (a) absolute errors; (b–g) scatter plots for the six models.

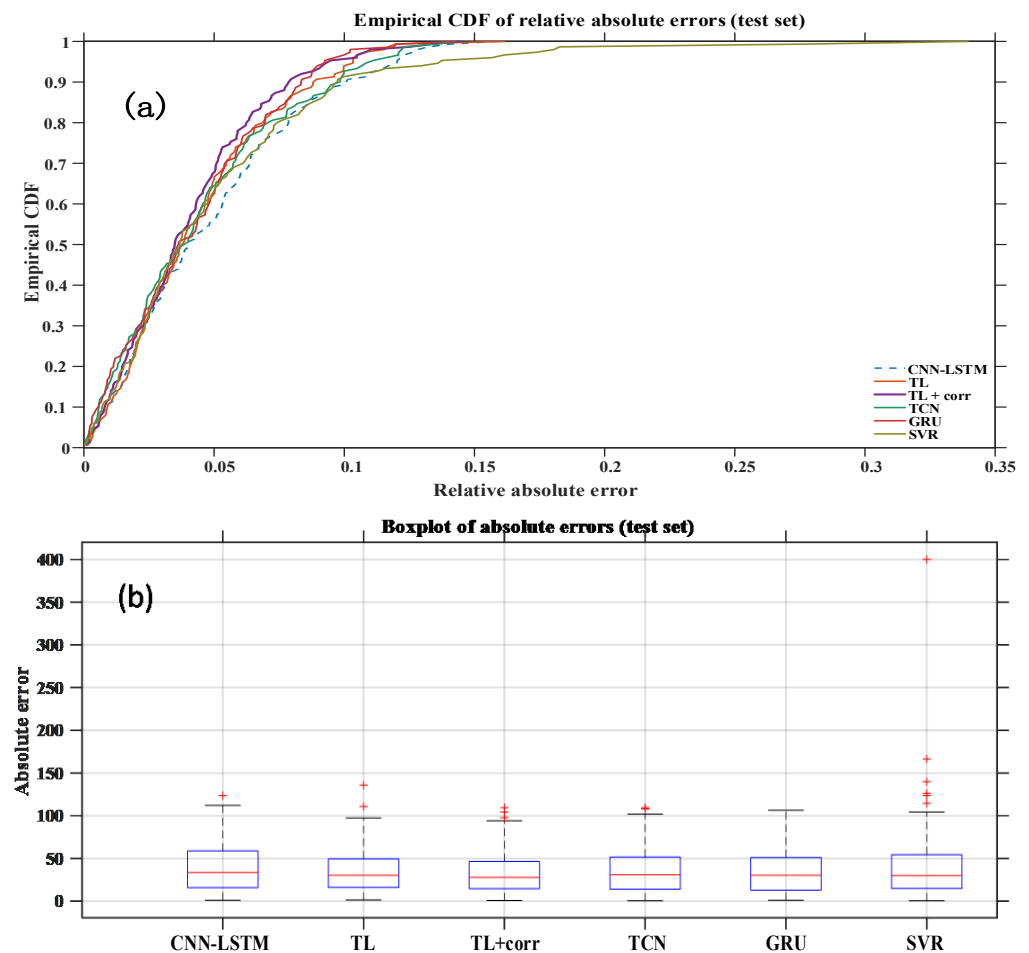


Figure 6. (a) empirical CDF; (b) boxplot of errors/residuals across mode. (In subfigure (b), the blue boxes represent the interquartile range, and the red '+' symbols indicate outliers.)

4. Discussion

The results demonstrate that the proposed TL + Corr (Transfer Learning with Residual Correction) framework significantly outperforms baseline models in both accuracy and robustness. From an algorithmic and physical perspective, the superior performance of the TL + Corr model under noise, missing data, and sensor drift can be attributed to its decoupled error-handling mechanism. The 1-D CNN acts as a temporal feature extractor that inherently filters high-frequency transient noise through its convolutional kernels and Global Average Pooling (GAP), while the LSTM captures the underlying physical degradation trend of the insulation materials. When faced with cross-domain distribution shifts—such as slow sensor drift—the frozen feature transfer prevents the model from overfitting to the target domain's noise. More importantly, the residual correction module explicitly isolates and models the systematic, deterministic bias caused by operating condition switches or drift. By subtracting this linear bias during inference, the framework successfully recenters the prediction errors to a near-zero-mean Gaussian distribution, ensuring strict amplitude consistency even during abrupt voltage transitions. Furthermore, to determine the optimal network depth, we evaluated the model's performance by varying the number of hidden convolutional layers. Our observations revealed that reducing the CNN to a single hidden layer resulted in underfitting, failing to adequately capture the complex transient perturbations. Conversely, increasing the architecture to three or more hidden layers not only increased the computational burden—making it less suitable

for edge-side BMS deployment—but also exacerbated the risk of overfitting on the small-sample target domain. Consequently, a two-layer configuration was adopted to strike the optimal balance between feature extraction capability and parameter efficiency.

Compared to conventional physical measurement techniques widely used in current energy storage systems, such as the unbalanced bridge method or AC signal injection, this data-driven “soft sensing” approach offers substantial deployment advantages for Battery Management Systems (BMS). Traditional hardware-based methods are fundamentally constrained by the RC delay caused by large parasitic Y-capacitors in high-voltage DC buses. This delay leads to slow response times and severe measurement distortion during dynamic load switching or transient power disturbances. In contrast, the proposed framework leverages existing multi-source telemetry (e.g., R_p , R_n , temperature, and their derivatives) to predict the insulation state instantly. It completely eliminates the need for additional injection circuits or physical relays, reducing hardware costs. Furthermore, the lightweight architecture of the 1-D CNN combined with a small-scale LSTM and a linear corrector ensures low computational overhead, enabling millisecond-level online inference on standard BMS edge microcontrollers.

Despite its promising performance and high deployment potential, the proposed framework exhibits two primary limitations. First, while the three-stage transfer learning strategy successfully reduces the reliance on massive datasets, the optimization pipeline still fundamentally requires a small subset of labeled target-domain samples to fine-tune the network and fit the residual corrector. In practical field applications, obtaining accurately labeled R_{iso} data for a newly commissioned battery cluster can be challenging. Second, the residual correction module currently assumes a linear relationship for systematic bias. While effective for gradual sensor drift, its small-sample learning capability may be restricted under extreme, out-of-distribution catastrophic scenarios (e.g., sudden mechanical puncture or severe thermal runaway precursors), where the insulation degradation exhibits highly non-linear and chaotic behaviors. Future research will focus on unsupervised domain adaptation to eliminate the dependency on target-domain labels and explore dynamic, non-linear residual modeling to further enhance the system’s resilience under extreme boundary conditions.

5. Conclusions

To address the critical challenges of cross-domain distribution shifts and dynamic perturbations in predicting the insulation resistance of high-voltage direct current energy storage systems, this study proposes the TL-CNN-LSTM + Corr framework. Experimental results derived from 500 sets of real-world operational data demonstrate the following: the three-stage transfer learning strategy effectively mitigates the risk of overfitting on small-sample target domains; the hybrid CNN-LSTM architecture achieves the decoupled modeling of localized transient noise and long-term degradation trends; and the residual compensation module successfully eliminates the linear amplitude biases that persist post-fine-tuning. In quantitative evaluations, the framework achieved an RMSE of 40.595, an MAE of 32.919, and an R^2 of 0.941 on an independent testing set, while maintaining a highly stable error distribution under severe perturbations, including noise injection, random data missing, and sensor drift. Crucially, directly addressing the primary motivation of overcoming stray capacitance-induced RC delays, this data-driven soft-sensing approach advances the temporal resolution of insulation measurement from the traditional 2–10 s—typical of hardware-based balanced bridges—to a sub-100 millisecond algorithmic inference scale. This represents a critical quantitative step toward the timely, real-time detection of rapid transient leakage currents and thermal runaway precursors.

Furthermore, while the proposed CNN-LSTM with residual correction serves as a highly effective paradigm for decoupled spatio-temporal modeling, we acknowledge that

it is not the exclusively viable mathematical approach. Other classes of mathematical frameworks, such as Physics-Informed Neural Networks (PINNs) or advanced continuous-time State-Space Models (e.g., Mamba), represent highly promising alternatives that may perform equally well or better in bounding non-linear uncertainties. Consequently, future research will focus on evaluating these broader mathematical classes, alongside exploring unsupervised domain adaptation methodologies to further optimize deployment under extreme degradation conditions.

Author Contributions: H.L. performed the methodology and wrote the manuscript; Y.C. provided supervision and validation; Y.Q. was responsible for conceptualization and funding; B.L. handled project administration and formal analysis; A.G. and T.J. contributed to the investigation. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Key Laboratory of Electrochemical Energy Safety, Ministry of Emergency Management, No. EES2025KF14.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy or commercial restrictions.

Acknowledgments: The authors would like to thank the technical staff at XYZ Storage Technology Corp. Ltd. for their assistance during data collection.

Conflicts of Interest: Author Yong Qiu, Baohong Lu and Aiwei Guan were employed by the company XYZ Storage Technology Corp. Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Moa, E.H.Y.; Go, Y.I. Large-scale energy storage system: Safety and risk assessment. *Sustain. Energy Res.* **2023**, *10*, 13. <https://doi.org/10.1186/s40807-023-00082-z>.
2. Al Khdhairi, Y.M.S.; Vural, M. Large-scale battery energy storage system integration to power systems with high penetration of renewable energy sources for regulation services: A critical review. *J. Renew. Energy Environ.* **2025**, *12*, 1–28. <https://doi.org/10.30501/jree.2025.486400.2149>.
3. Chen, S.; Kong, G.; Zhang, Y.; Li, G. Fire and explosion risk analysis and prevention and control technology of lithium-ion battery energy storage systems. *Chin. Eng. Sci.* **2025**, *27*, 269–286. <https://doi.org/10.15302/J-SSCAE-2025.01.026>.
4. Guo, F.; She, D.; Zhang, X.; Han, Y. Design of a high-voltage insulation resistance detection system for commercial vehicles. *Batteries* **2025**, *11*, 143. <https://doi.org/10.3390/batteries11040143>.
5. Tian, J.; Yin, J. Insulation fault monitoring of lithium-ion battery pack: Recursive least square with adaptive forgetting factor. *J. Energy Storage* **2023**, *67*, 107624. <https://doi.org/10.1016/j.est.2023.107624>.
6. Du, J.; Gong, M. Insulation monitoring method for DC systems with ground capacitance in electric vehicles. *Appl. Sci.* **2019**, *9*, 2607. <https://doi.org/10.3390/app9132607>.
7. Li, Q.; Li, Q.; Wang, L.; Wu, Q.; Qin, G.; Zhang, H.; Wang, H. Improving the accuracy of insulation resistance measurements of a single arm unbalanced bridge using a parameter matching design and filtering algorithm for battery packs used in electric ships. *Measurement* **2022**, *195*, 111133. <https://doi.org/10.1016/j.measurement.2022.111133>.
8. Zhu, J.; Yan, J. *Insulation Resistance Detection Designs in GESS-BMS*; Texas Instruments: Dallas, TX, USA, 2025; Appl. Note SLUAB06.
9. Mao, Z. Comprehensive analysis of energy storage system for electric vehicles. *Highlights Sci. Eng. Technol.* **2023**, *71*, 184–193. <https://doi.org/10.54097/hset.v71i.12693>.
10. Song, C.; Shao, Y.; Song, S.; Peng, S.; Zhou, F.; Chang, C.; Wang, D. Insulation resistance monitoring algorithm for battery pack in electric vehicle based on extended Kalman filtering. *Energies* **2017**, *10*, 714. <https://doi.org/10.3390/en10050714>.

11. Tian, J.; Wang, Y.; Yang, D.; Zhang, X.; Chen, Z. A real-time insulation detection method for battery packs used in electric vehicles. *J. Power Sources* **2018**, *385*, 1–9. <https://doi.org/10.1016/j.jpowsour.2018.03.018>.
12. Li, Z.; Cui, X.; He, Z.; Li, E.; Wang, Y. A novel method for electric vehicle insulation detection based on the extended Kalman filter algorithm. *Measurement* **2024**, *229*, 114419. <https://doi.org/10.1016/j.measurement.2024.114419>.
13. Kraleti, R.S.; Zawodniok, M.; Jagannathan, S. Model based diagnostics and prognostics of three-phase induction motor for vapor compressor applications. In Proceedings of the 2012 IEEE Conference on Prognostics and Health Management (ICPHM), Denver, CO, USA, 18–21 June 2012; Article 6299525. <https://doi.org/10.1109/ICPHM.2012.6299525>.
14. Wang, Y.; Tian, J.; Chen, Z.; Liu, X. Model-based insulation fault diagnosis for lithium-ion battery pack in electric vehicles. *Measurement* **2019**, *131*, 443–451. <https://doi.org/10.1016/j.measurement.2018.09.007>.
15. Zoeller, C. Influence of parasitic capacitances of IGBT inverter on insulation condition monitoring of traction machines based on current signal transients analysis. In Proceedings of the 2016 18th European Conference on Power Electronics and Applications (EPE'16 ECCE Europe), Karlsruhe, Germany, 5–9 September 2016. <https://doi.org/10.1109/EPE.2016.7695580>.
16. Wei, Z.; Liu, K.; Liu, X.; Li, Y.; Du, L.; Gao, F. Multilevel Data-Driven Battery Management: From Internal Sensing to Big Data Utilization. *IEEE Trans. Transp. Electr.* **2023**, *9*, 4805–4823. <https://doi.org/10.1109/TTE.2023.3301990>.
17. Liang, K.; Zhang, Z.; Liu, P.; Wang, Z.; Jiang, S. Data-driven ohmic resistance estimation of battery packs for electric vehicles. *Energies* **2019**, *12*, 4772. <https://doi.org/10.3390/en12244772>.
18. Seidl, M.; Raji, A.; Bauer, S. Data driven analysis of lithium-ion battery internal resistance towards reliable state of health prediction. *J. Power Sources* **2021**, *513*, 230519. <https://doi.org/10.1016/j.jpowsour.2021.230519>.
19. Zhao, D.; Li, H.; Zhou, F.; Zhong, Y.; Zhang, G.; Liu, Z.; Hou, J. Research progress on data-driven methods for battery states estimation of electric buses. *World Electr. Veh. J.* **2023**, *14*, 145. <https://doi.org/10.3390/wevj14060145>.
20. Liu, S.; Zhao, G.; Zhang, H. Intelligent insulation testing and optimization based on machine learning. *Electronics* **2025**, *14*, 109. <https://doi.org/10.3390/electronics14010109>.
21. Shi, D.; Zhao, J.; Eze, C.; Wang, Z.; Wang, J.; Lian, Y.; Burke, A.F. Cloud-Based artificial intelligence framework for battery management system. *Energies* **2023**, *16*, 4403. <https://doi.org/10.3390/en16114403>.
22. Njoku, J.N.; Eneh, A.U.; Nwakanma, C.I.; Lee, J.-M.; Kim, D.-S. HyBaTwin: Web-Based Hybrid Digital Twin Platform for Electric Vehicle Battery Capacity Estimation. *J. Korean Inst. Commun. Inf. Sci.* **2025**, *50*, 549–560. <https://doi.org/10.7840/kics.2025.50.4.549>.
23. Zhao, Y.; Liu, H.; Deng, Z.; Li, T.; Jiang, H.; Ling, Z.; Wang, X.; Zhang, L.; Ouyang, X. Machine learning bridging battery field data and laboratory data. *arXiv* **2025**, arXiv:2505.05364. <https://doi.org/10.48550/arXiv.2505.05364>.
24. Tan, R.; Hong, W.; Tang, J.; Lu, X.; Ma, R.; Zheng, X.; Li, J.; Huang, J.; Zhang, T.-Y. BatteryLife: A Comprehensive Dataset and Benchmark for Battery Life Prediction. *arXiv* **2025**, arXiv:2502.18807. <https://doi.org/10.1145/3711896.3737372>.
25. Hussein, H.M.; Esoofally, M.; Donekal, A.; Rafin, S.M.S.H.; Mohammed, O. Comparative study-based data-driven models for lithium-ion battery state-of-charge estimation. *Batteries* **2024**, *10*, 89. <https://doi.org/10.3390/batteries10030089>.
26. Lu, J.; Xiong, R.; Tian, J.; Wang, C.; Sun, F. Deep learning to estimate lithium-ion battery state of health without additional degradation experiments. *Nat. Commun.* **2023**, *14*, 2760. <https://doi.org/10.1038/s41467-023-38458-w>.
27. Al-Rahamneh, A.; Izco, I.; Serrano-Hernandez, A.; Faulin, J. Machine Learning-Based State-of-Health Estimation of Battery Management Systems Using Experimental and Simulation Data. *Mathematics* **2025**, *13*, 2247. <https://doi.org/10.3390/math13142247>.
28. Chen, L.; Bao, X.; Lopes, A.M.; Li, X.; Kong, H.; Chai, Y.; Li, P. Domain generalization-based state-of-health estimation of lithium-ion batteries. *J. Power Sources* **2024**, *611*, 234696. <https://doi.org/10.1016/j.jpowsour.2024.234696>.
29. Yang, S.; Zhou, J.; Chen, B.; An, R.; Zhao, Z.; Fan, Y.; Guan, Q.; Tan, X. Deep domain adaptation for cross-chemistry battery SOH prediction with relaxation voltage features. *Energy* **2025**, *339*, 139018. <https://doi.org/10.1016/j.energy.2025.139018>.
30. Zhang, H.; Li, Y.; Zheng, S.; Lu, Z.; Gui, X.; Xu, W.; Bian, J. Accurate battery lifetime prediction across diverse aging conditions with deep learning. *arXiv* **2023**, arXiv:2310.05052. <https://doi.org/10.48550/arXiv.2310.05052>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.