Prognostics of internal short circuit in lithium-ion batteries using a hybrid deep learning approach

Yongtao Yao [a],*, Xinyu Du [b], Shengbing Jiang [b], Rasoul Salehi [b], Weisong Shi [a]

[a] University of Delaware, Newark, DE 19702, USA
[b] General Motors Warren, MI 48092, USA

HIGHLIGHTS

- HyCARLT pinpoints internal shorts during real-world driving.
- Two-stage design beats extreme healthy–faulty data imbalance.
- OCV + balanced-Ah features embed electrochemistry into the DL model.
- Validated on 10 000 EVs over two years; ISC error ≈ 15 %.

ARTICLE INFO

ABSTRACT

Short resistance estimation is vital for diagnosing internal short circuit (ISC) faults in batteries, impacting battery lifespan and EV performance. Current estimation methods are often limited by error sources introduced during the estimation of charge difference based on the state-of-charge (SOC), such as inaccuracies in capacity and SOC/OCV (open circuit voltage) mapping coefficients, as seen in methods like Cell Voltage Droop, the Battery SOC Difference Relative to Module Median (dSOC). While some approaches may mitigate these interference factors, they rely on steady-state constant-current conditions, which restricts their applicability under dynamic, real-world operational scenarios. To address these limitations, it is essential to develop methods that can handle diverse operating conditions while mitigating data imbalance issues. This work introduces HyCARLT (Hybrid Classification and Regression with LSTM-Augmented Transformer), a hybrid model that combines a deep residual network classification module with a regression module incorporating LSTM and Transformer architectures to handle data imbalance issue and improve resistance estimation accuracy. The method is validated using data from over 10,000 vehicles and demonstrates better accuracy in short resistance estimation compared to the state-of-the-art approaches. The validation results demonstrate HyCARLT is a promising approach for ISC fault detection and short resistance estimation under diverse and dynamic conditions.

1. Introduction

Electric vehicles (EVs) are rapidly becoming a cornerstone of the modern transportation landscape, driven by the global shift toward more sustainable and environmentally friendly technologies. Central to the operation and efficiency of EVs is the battery, which serves not only as the power source but also as a critical component that directly affects the vehicle's performance, driving range and experience. Internal short circuits (ISCs) occur when a conductive pathway forms between the positive and negative electrodes inside a battery cell, bypassing the separator and causing abnormal current flow [1]. ISC, can drastically degrade battery performance by increasing energy consumption, reducing efficiency, diminishing capacity, and impairing

overall functionality [2]. This necessitates their timely detection and diagnosis to maintain the health and safety of EVs.

There are some straightforward methods to identify potential internal short circuits in EV batteries. For instance, voltage droop rate and cell voltage performance detect ISCs by directly observing how fast the cell voltage droops in response to a short circuit [3]. While these methods are effective under conditions with minimal external interference, such as stable temperatures and limited cell aging, their reliability decreases in the presence of external factors. These factors, such as temperature fluctuations or significant aging, can either mimic the voltage signatures of an ISC or obscure them, leading to

---

potential inaccuracies. Furthermore, such methods do not provide a direct estimation of short resistance, which is a key parameter in battery diagnostics to quantify the severity of internal short circuits [4]. Accurate estimation of short resistance is vital for the early detection of potential issues, enabling timely interventions that can prolong battery life and ensure the reliability of the vehicle's operation [5].

Estimating short resistance in EV batteries is a challenging task due to the dynamic and complex nature of battery operations. Traditional methods, such as electrochemical and equivalent circuit models, laid the theoretical foundation for understanding battery behavior. These methods, from the pioneering electrochemical frameworks by Newman and Balsara [6] to the simplified circuit models by Rand and Woods [7], rely on static assumptions and approximations that do not adequately reflect real-world variability caused by factors like temperature, state of charge (SOC), and aging. Even with advancements introduced by Doyle et al. [8], and refinements by He et al. [9], these methods struggle to model the nonlinear and time-variant characteristics essential for precise short resistance estimations in real-time applications. Building upon these traditional principles, physical methods for estimating short resistance, such as the Cell Droop Rate [10] and dSOC [11] approaches, attempt to make these theoretical models more practically applicable. However, indirect D&P methods, such as voltage-based ISC detection techniques, cannot quantify fault severity and are prone to false positives and negatives under operational variability, such as temperature fluctuations and cell aging. Direct D&P methods, while focus more on short resistance estimation, still exhibit significant errors in complicated real-world conditions, e.g. the capacity variation, uncertainty in SOC/OCV (open circuit voltage) mapping or other noises and disturbances. To make it more robust under diverse and challenging environments, data-driven solutions may help.

Machine learning (ML) has emerged as a powerful alternative for battery diagnostics, offering data-driven insights that can adapt more flexibly to complex, changing conditions. Early ML applications focused on state-of-health (SOH) and SOC estimation [12–14], while more recent efforts incorporate neural networks for predicting aging and capacity fade [13]. Although such techniques show potential in handling nonlinear battery behaviors, the direct application of ML or deep learning (DL) to short resistance estimation has remained limited. A substantial challenge lies in the inherent imbalance of the dataset, particularly regarding the distribution of short resistance values among healthy and faulty cells. Faulty cells constitute a much smaller proportion compared to healthy cells, and even within the faulty population, data are unevenly distributed across varying levels of fault severity. This skewed distribution complicates model training, as the deep learning algorithm may become biased toward predicting healthy cells, ultimately degrading performance when identifying and estimating short-circuit faults. The scarcity of short resistance dataset has led most DNN-based research to focus on classification tasks, such as categorizing fault severity, rather than directly estimating short resistance. Even when deep learning techniques have been applied to short resistance estimation, these studies often rely on small-scale, specially curated datasets that lack diversity and fail to represent the unbalanced distribution of short resistance values typically observed in real-world scenarios [14]. In most cases, these datasets only contain several discrete resistance values, making it difficult for the resulting models to generalize beyond the limited conditions they are trained on.

To address these challenges, we propose HyCARLT, a novel hybrid classification and regression approach specifically tailored for short resistance estimation in EV batteries. The key methodological innovations of HyCARLT include: (i) integrating a deep residual network (ResNet)-based classifier to accurately isolate faulty cells and significantly reduce data imbalance; (ii) employing an LSTM-Augmented Transformer regression module designed to effectively capture both short-term sequential dynamics and long-term dependencies in battery operational data; (iii) implementing a specialized weighted loss function within the regression stage to ensure the model prioritizes estimation accuracy for rare and critical low-resistance ISC faults. Additionally, HyCARLT incorporates carefully designed data augmentation strategies – combining oversampling of faulty cases and undersampling of healthy cases – to further improve model robustness. By combining these innovations, HyCARLT offers a comprehensive, data-driven diagnostic framework that addresses the limitations of existing methods, particularly under realistic, diverse, and imbalanced operational conditions. Our contributions are highlighted as follows:

- Our proposed deep learning method achieves better performance in estimating the short resistance of batteries using real-world vehicle data during normal operation. Validated on data from over 10,000 vehicles, it achieves a relative error of 15% on faulty cells and no healthy cells are false classified as unhealthy, representing a significant improvement over existing state-of-the-art physical methods such as Cell Droop Rate and dSOC, which exhibit estimation errors ranging from 50% to 80%.
- The method proposed in this report provides reliable estimations across diverse operating conditions without requiring steady-state assumptions or post-hoc maturation, underscoring its robustness and practical utility.
- Furthermore, the integration of a hybrid model—combining a deep residual network for classification with an LSTM-Transformer regression module—effectively addresses data imbalance, reduces false positives, and enhances fault detection capabilities. This design leverages tailored data augmentation techniques, including under-sampling of healthy cells and over-sampling of faulty cells, to further improve diagnostic performance, ensuring accurate and scalable short resistance estimation under real-world conditions.

Despite its strong performance, HyCARLT has several limitations. First, its accuracy depends on the quality and diversity of training data, and rare failure cases remain underrepresented. Second, while data balancing techniques improve fault detection, estimation errors for low short resistance values persist. Third, the classification threshold at 2000 Ω is empirically set and may require adjustment for different battery chemistries. Finally, as a deep learning model, HyCARLT lacks physical interpretability, which may limit its adoption in safety-critical applications. Future work can address these issues by integrating physics-informed models and expanding training datasets.

The structure of this report is organized as follows: In Section 2, an overview of the foundational techniques used for estimating short resistance is discussed. The methodology is provided in Section 3, including data preprocessing, model architecture, and training process. In Section 4, the experimental evaluation is discussed, including a comparative analysis with existing models and the proposed model.

2. Related work

This section explores how advanced deep learning architectures can be adapted to improve short resistance estimation in electric vehicle batteries. By leveraging classification and regression paradigms, models such as convolutional networks, recurrent networks, transformers, and hybrid designs can effectively capture temporal patterns and manage long-range dependencies. The discussion emphasizes the potential of hybrid frameworks to address the limitations of traditional diagnostic tools, particularly in handling imbalanced data and adapting to dynamic battery states.

For fault diagnostics and prognostics, classification and regression are two primary predictive paradigms, each is suited for specific types of outputs. Classification involves predicting categorical labels from input data. The output is discrete, such as "faulty" or "healthy" in binary classification, or one of several categories in multiclass classification. Models for classification often output probabilities over classes,

with the predicted label being the one with the highest probability. Regression, in contrast, deals with predicting continuous values. The output can take any numerical value within a range, making regression ideal for tasks such as estimating numerical quantities or forecasting trends. For instance, predicting the short resistance of a battery cell is a regression task, where the model aims to provide precise numerical estimates. Loss functions like Mean Squared Error (MSE) are commonly employed to minimize the prediction error.

Short resistance estimation is fundamentally a time-series regression task. DNNs have become central to time-series analysis to tackle temporal modeling. Fully Convolutional Networks (FCNs) [15] are designed to process time series data by utilizing convolutional layers to extract features, with enhancements like FCNPlus introducing modifications to improve performance. Inception-based models, such as InceptionTime, adapt the Inception architecture for time series classification, employing multiple filters of varying lengths to capture diverse patterns; further developments include InceptionTimePlus variants like InceptionTimePlus and InceptionTimeXLPlus [16], which offer expanded architectures for handling more complex data. Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) networks [17], are adept at modeling temporal dependencies in sequential data, with advanced versions like LSTMPlus and GRUPlus providing enhanced capabilities [16]. Hybrid models that combine RNNs with FCNs, such as LSTM-FCN [17], leverage the strengths of both architectures to improve classification accuracy. Residual Networks (ResNets) [15] incorporate residual connections to facilitate the training of deeper networks, with adaptations like ResNetPlus [18] and one-dimensional versions such as xresnet1d18plus tailored for time series tasks. Transformer-based models, including TransformerModel [19] and MultiTSTPlus, have been applied to time series analysis to capture long-range dependencies, with models like XCM and XCMPlus offering explainable convolutional neural network approaches. Other notable models include Multi-Layer Perceptrons (MLPs) and their gated variants (gMLP) [20], Temporal Convolutional Networks (TCNs) [21], and architectures that integrate attention mechanisms, such as LSTMAttention and its enhanced version LSTMAttentionPlus [16]. Informer [22] pioneers the use of probabilistic sparse self-attention to reduce the quadratic cost of vanilla Transformers, making it much better at handling long time-series forecasts. Building on this idea, Autoformer [23] introduces an auto-correlation decomposition that explicitly separates seasonal and trend components, further improving long-range accuracy. FEDformer [24] extends the decomposition view by combining Fourier and wavelet blocks inside the attention pipeline, yielding better efficiency-accuracy trade-offs. More recently, PatchTST [25] reimagines time-series inputs as non-overlapping "patches", enabling pure Transformer encoders (without convolutions or RNNs) to capture local context with fewer parameters. Beyond these mainstream variants, several 2024 releases have pushed the frontier: PSFormer [26] shares parameters across segment-wise windows to lower memory, Ister [27] employs interpretable seasonal-trend attention that exposes internal recurrence patterns, and TEAFormers [28] augment standard attention with tensor decomposition to model multi-dimensional battery signals compactly. These architectures provide strong baselines for time-series regression tasks.

In this work, we revisit three fundamental yet versatile models – ResNet, LSTM, and Transformer – as essential building blocks to construct a hybrid framework tailored for short resistance estimation with better performance. Both local temporal patterns and long-range dependencies are captured while maintaining computational efficiency, the complementary strengths of these foundational models are leveraged. ResNet overcomes the vanishing gradient issue in very deep networks by introducing "skip connections." These skip connections allow gradients to bypass certain layers, thereby stabilizing training in deep models. The use of ResNet could exploit its strong feature extraction capabilities, allowing it to detect subtle yet critical indicators of ISC faults over time by leveraging residual connections to

model complex dependencies within battery data, which has shown 96.67% prediction accuracy in identifying ISC faults for classification tasks [29]. LSTM networks are a specialized type of recurrent neural network (RNN) designed to address the limitations of traditional RNNs, such as vanishing and exploding gradients. These problems often hinder RNNs from learning long-term dependencies. LSTMs, introduced by Hochreiter and Schmidhuber [30], overcome these challenges by using memory cells and gating mechanisms – specifically input, forget, and output gates – that regulate the flow of information. This structure allows LSTMs to retain or discard information selectively, making them particularly effective for time series tasks where both short-term and long-term patterns need to be understood. In battery diagnostics, LSTMs have proven valuable for estimating the SOH and SOC, predicting degradation trends, and detecting early faults [31]. Their ability to capture temporal dependencies is crucial for identifying patterns in voltage and current data that indicate potential faults or degradation in battery cells. A self-attention mechanism is employed by transformers that allows the model to focus on different parts of the input data, emphasizing critical elements that influence predictions [32]. Unlike LSTMs, Transformers do not rely on sequential processing, enabling them to capture long-range dependencies more efficiently and in parallel. This capability significantly improves training and inference speed. Transformers have demonstrated strong performance in domains ranging from natural language processing to image recognition, and recently in battery diagnostics for tasks like SOC estimation and fault detection [31]. Their ability to model complex interactions without the constraints of sequential order makes them ideal for capturing the dynamic behaviors of batteries.

3. Methodology

To address the severe data imbalance inherent in real-world battery data – where healthy cells significantly outnumber ISC faulty cells – our proposed HyCARLT model integrates both classification and regression modules. Specifically, a deep residual network-based classifier first determines the presence of ISC faults, allowing the subsequent LSTM-Transformer regression module to focus exclusively on accurately estimating the severity of confirmed faults. This two-stage design contrasts with purely regression-based methods, which typically do not adequately compensate for data imbalance, often resulting in less reliable fault estimations.

The proposed diagnostic model, illustrated in Fig. 1, is designed to estimate the short circuit resistance (Rs) of battery cells through a DNN architecture. The input to the model consists of a time series dataset that includes multiple features, such as cell group (CG) OCV, ampere-hour (Amphr), module median OCV (mmOCV), module median Amphr (mmAmphr), and timestamp data arranged in a Nx5 format. The input data undergoes preprocessing steps, including normalization and transformation, ensuring that it is suitable for HyCARLT. The final output of the HyCARLT is a regression prediction of the short resistance.

3.1. Physical explanation of feature selection and model inputs

To avoid confusion between open-circuit voltage (OCV) and instantaneous terminal voltage, we define OCV here as the "relaxed" terminal voltage measured after the module current has been zero for a dwell long enough that the voltage drift falls below $1\ \mathrm{mV\ h^{-1}}$. Throughout this work voltage is sampled only in such rest windows, so the recorded value is a practical surrogate for the true thermodynamic OCV.

The selected features – cell-group relaxed voltage (denoted OCV_r), module-median relaxed voltage ($mmOCV_r$), cell balanced ampere-hour (Amphr), and module-median balanced ampere-hour (mmAmphr) – are closely related to ISC faults through measurable physical phenomena occurring within battery modules. Each cell group, which consists of three parallel-connected cells, forms the basic analysis unit (hereafter referred to simply as "cell"). Under ideal conditions, all eight healthy
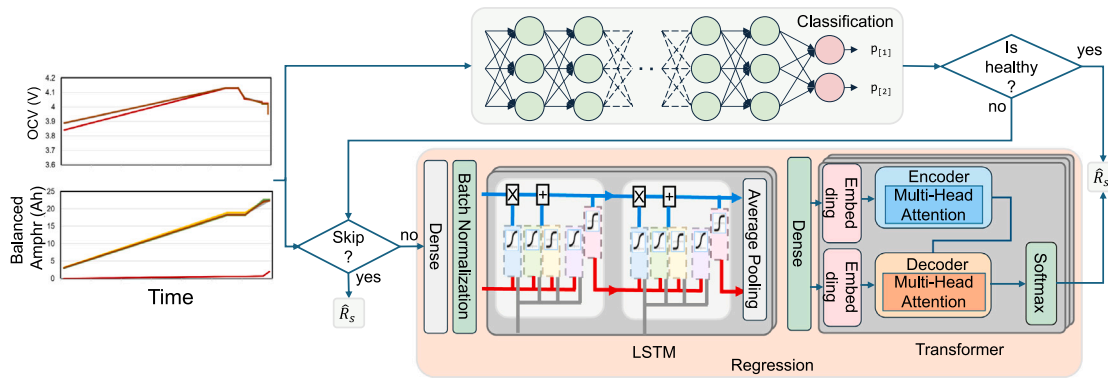
Fig. 1. The proposed HyCARLT model integrates a classification module and a regression module to accurately estimate short-circuit resistance in electric vehicle batteries. First, a classifier determines whether the cell is healthy or faulty. Healthy samples are quickly assigned a nominal short resistance, while only the faulty samples pass to the LSTM-Augmented Transformer regression stage. This selective routing ensures the model focuses computational effort where it matters most, delivering robust and reliable resistance predictions even under highly imbalanced conditions.

cells within a module exhibit identical OCV_r profiles and the balancing circuit remains inactive. When an ISC arises in a particular cell its OCV_r drifts downward; the lower the short resistance, the faster this drift.

Once the OCV_r of a faulty cell departs by roughly 5 mV from its neighbors, the module's active-balancing circuit is triggered, forcing the healthy cells to dissipate charge so that module voltages realign. Hence, the severity of ISC governs both the rate of voltage divergence in the faulty cell and the duty-cycle of balancing in its healthy peers.

Mechanistically, the discharge quantity ΔQ_j of the jth cell over a window Δt decomposes into three terms:

$$\Delta Q_j = \Delta Q_j(\text{Usage}) + \Delta Q_j(\text{ISC}) + \Delta Q_j(\text{Balance}), \quad (1)$$

while the module median (assumed healthy) contains only

$$\Delta Q_{mm} = \Delta Q_{mm}(\text{Usage}) + \Delta Q_{mm}(\text{Balance}). \quad (2)$$

Because all cells share the same load current, the usage terms cancel: $\Delta Q_j(\text{Usage}) = \Delta Q_{mm}(\text{Usage})$. The remaining quantities are mapped to our measured features:

- ΔQ_j and ΔQ_{mm} are inferred from the respective changes in OCV_r and $mmOCV_r$ via the local dQ/dV relationship.
- $\Delta Q_j(\text{Balance})$ and $\Delta Q_{mm}(\text{Balance})$ are provided directly by the logged Amphr and mmAmphr counters maintained by the BMS.
- $\Delta Q_j(\text{ISC})$ equals $\int (V/R_s)\,dt \approx \int (OCV_r/R_s)\,dt$, linking short-circuit resistance R_s to the observable voltage depression in the faulty cell.

Eqs. (1)–(2) make "no a-priori assumption" that imbalance or ISC is present; they simply express charge conservation. The deep network learns to distinguish healthy from faulty behavior by correlating the observed divergences in OCV_r and balancing current with the hidden ISC term, thereby enabling detection even when the true battery state is unknown beforehand.

3.2. Data preprocessing

The raw data collected from the vehicle fleet is inherently simple. Each cell record in the raw dataset contains only two attributes – OCV and balanced Ampere-hour (Amphr) – with timestamps. Sampling is sparse and non-uniform, averaging about 10 points per day, but in many cases fewer. If a vehicle remains idle without charging or driving, no data are uploaded, and hence no records appear in the dataset. Although this relatively sparse sampling and minimalistic attribute set restrict detailed temporal resolution, the strength of this dataset lies in its vast scale – over 10,000 vehicles monitored over up to two years – providing extensive coverage of diverse operational conditions. Despite

these limitations, the recorded OCV and Amphr values are sensitive indicators of internal battery dynamics and ISC events, enabling effective training of our proposed model, as illustrated in Fig. 2.

Fig. 2 presents one month of data monitoring for 8 cells from a vehicle. In Fig. 2 (a), the initial OCV of Cell 108 is lower than that of the other cells. However, over time, the OCV difference gradually decreases. In Fig. 2 (b), it can be observed that the balanced Amphr of the other cells continues to increase, while the balanced Amphr of Cell 108 remains near zero. It is the balancing mechanism that leads to the gradual reduction in the OCV difference.

It is important to clarify that the real-world data employed in this study are acquired directly from a central database of a large-scale EV fleet via direct database queries. Specific details regarding the internal data collection methodologies, sensor configurations, and operational strategies used by the fleet management system are not available to us. The data reflects genuine user behavior and real-world driving conditions, which result in irregular and sometimes sparse data uploads, dependent on vehicle usage patterns. Explicit charging or discharging statuses are not provided; instead, changes in OCV and balanced Amphr inherently reflect cumulative charging and discharging activities. Due to the irregular and sparse sampling frequency (approximately ten samples per day on average, with frequent days having fewer or no samples), continuous trends observed in data visualization – such as the sustained OCV increase from September 27, 2023, to October 20, 2023 – are attributable to intermittent and short partial charging events combined with limited vehicle usage, rather than indicating continuous uninterrupted charging.

The primary features extracted for this work include the cell OCV, module median OCV, cell balanced ampere-hour, and module median balanced Amphr, alongside the corresponding timestamp. To prepare the data for model training, the time series is segmented into windows of 128 consecutive timestamps, resulting in a tensor with the shape of 128×5 for each input sample. This window size is chosen to balance the trade-off between capturing sufficient temporal information and maintaining computational efficiency.

The data labeling is performed using General Motors' proprietary internal algorithm specifically developed to estimate short-circuit resistance. Due to confidentiality and company policy constraints, the specific details of this method cannot be publicly disclosed at this time. All data used in this study are collected from normal vehicle driving without artificial intervention to induce battery faults. Faulty battery samples are identified retrospectively through rigorous statistical consistency checks across multiple independent diagnostic methods. These include publicly documented approaches such as the dSOC Method [11] and the Cell Droop Rate Method [10], which effectively detect ISC faults but have limited accuracy in short-resistance estimation. To ensure the most accurate labeling, an internal proprietary algorithm,
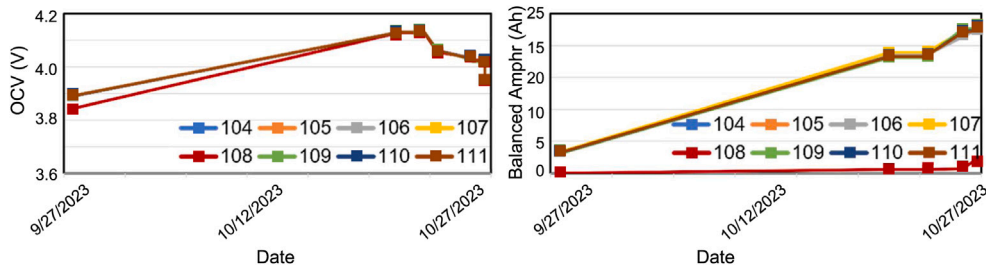
Fig. 2. Driving Vehicle Data. 10 samples/day for OCV and Amphr. (a) OCV signals of 8 cells from one module (Cell 108 shows a short-circuit fault). (b) Balanced Amphr signals of 8 cells from one module. *Markers indicate actual raw database samples; gaps between markers reflect idle periods with no uploaded data.*

which significantly improves upon the published methods in terms of resistance estimation accuracy, is ultimately utilized. Only samples consistently verified as faulty by all independent evaluations are labeled as faulty, ensuring the robustness, accuracy, and reliability of the training and testing datasets.

Once these estimates are obtained, we apply an additional processing step by setting a threshold of 2000 Ω. Any estimated values exceeding this threshold indicate that cells are considered operationally normal. The classification threshold of 2000 Ω utilized in our study is determined based on battery-specific parameters and operational characteristics. Specifically, for the battery cells analyzed (capacity ∼300 Ah), this threshold corresponds to a minor internal discharge rate, where it takes approximately one month (35.94 days) in a resting state to cause a voltage drop of about 5 mV, resulting in a negligible SOC reduction (∼0.57%). For different battery chemistries or cell capacities, we recommend recalculating this threshold based on actual measured OCV-SOC relationships, cell capacity, and empirical discharge parameters to maintain model generality and applicability. By applying this threshold, the classification task is simplified, focusing the model's learning process on regions where faults are most critical. This approach prevents excessive computational and representational resources from being spent on healthy samples, where variations in short resistance are operationally insignificant. Furthermore, it helps to streamline data preprocessing and ensures consistent labeling criteria.

In addition, to address the challenges posed by the inherent imbalance in the dataset, healthy cells with short resistance value labeled at 2000 Ω are under sampled, and faulty cells with low short resistance values are oversampled, which helps to mitigate the bias toward predicting the dominant class and enhances the model's ability to learn from the rarer ISC fault cases. Oversampling faulty cells could potentially introduce bias or lead to model overfitting due to duplicated samples. To address this, our hybrid architecture employs a classification module that robustly filters out healthy cells first. Given that faulty cells show distinct OCV behaviors compared to healthy cells due to module-level balancing mechanisms, our classification module achieves high accuracy despite oversampling. Furthermore, the regression module, focused only on the confirmed faulty cells, effectively leverages oversampling as an increased training epoch strategy, thus minimizing bias or overfitting risks. The data is collected from proprietary vehicle database, consisting of autonomously reported records from over 10,000 vehicles, each submitting roughly 10 data records daily. Both sampling times and intervals varied and were not fixed. Initially, missing and anomalous values are identified and filtered out or interpolated based on their impact on data integrity. Subsequently, the data is segmented into sequences of 128 temporally continuous data points from the same battery module, creating time windows used as model inputs. Timestamps are also included as a feature to incorporate temporal information crucial for short-circuit resistance estimation. All extracted features are normalized to mitigate numerical biases.

Due to confidentiality constraints, detailed battery cell data collection setups and internal label generation procedures cannot be disclosed. Despite this, our described preprocessing and experimental

design remains transparent enough for methodological evaluation by the broader research community.

Additionally, we evaluate our method using an open-source dataset from the study by Jia et al. [21]. The dataset, publicly accessible at GitHub,[1] includes features such as voltage, current, capacity, energy, total_capacity, total_energy, and labels representing R_s values. The original authors provided preprocessing scripts for segmenting raw data into fixed-length time windows suitable for training.

3.3. Model design

The proposed hybrid framework, HyCARLT (Hybrid Classification and Regression with LSTM-Augmented Transformer), aims to accurately estimate the short resistance in electric vehicle batteries under real-world conditions. Our approach is grounded in a two-stage process: an initial classification step identifies whether a sample is faulty or healthy, and then a regression step refines the short resistance estimation for the faulty subset. This integrated pipeline addresses the challenges of extreme data imbalance and operational variability in battery diagnostics.

At the input level, the model takes a multivariate time series $X = [x_1, x_2, \ldots, x_N]^T \in \mathbb{R}^{N \times d}$, where N is the sequence length and $d = 5$ represents the number of features per time step. Each data segment corresponds to a window of battery operational conditions, and the goal is to estimate the short resistance for that interval. Since only a relatively small number of samples are faulty, dealing directly with regression over the entire dataset would waste computational resources on the abundant healthy samples and risk biasing the model. Instead, a classification module $f_c(\bullet; \Theta_c)$ is integrated to determine if X represents a faulty case:

$$f_c : \mathbb{R}^{N \times d} \to [0, 1], \hat{p} = f_c(X; \Theta_c), \tag{3}$$

where $\hat{p}$ represents the probability that the sample is faulty, defined as having a short resistance below a threshold (2000 Ω in this work). The classification network maps the input from $\mathbb{R}^{N \times d}$ to a probability space [0,1], producing a fault probability $\hat{p} = f_c(X; \Theta_c)$. As shown in Eq. (2), setting a decision threshold τ, the sample is considered faulty if $\hat{p} \geq \tau$, and healthy otherwise,

$$\hat{y}_c = \begin{cases} 1, & \text{if } f_c(X; \Theta_c) \geq \tau; \\ 0, & \text{if } f_c(X; \Theta_c) < \tau. \end{cases} \tag{4}$$

By introducing this classification step, healthy samples can be quickly discarded by assigning them a default short resistance of 2000 Ω, effectively bypassing unnecessary regression computations. This design significantly reduces the impact of data imbalance, as the regression model then exclusively focuses on the critical minority class of faulty samples.

To implement the classification module, a deep residual network architecture (ResNet) is employed, using the xresnet1d34 backbone

---

[1] https://github.com/orDrink/Short-circuit-resistance-estimator

from the tsai library [16]. This backbone comprises custom residual blocks with varying kernel sizes (7, 5, and 3), enabling the extraction of multi-scale temporal features. Each residual block includes batch normalization and ReLU activation to maintain stable training. The network's layered structure captures hierarchical patterns, culminating in a global average pooling layer and a fully connected output layer that delivers a probability estimate for the fault condition. Once a sample is classified as healthy, the inference ends and the estimated short resistance is set to 2000 Ω.

When the classification output $\hat{y}_c = 1$ indicates a faulty sample, the model proceeds to the regression stage. Here, the input X is passed to a regression network $f_r(\bullet; \Theta_r)$ that produces a continuous-valued estimate of short resistance $\hat{R}_s = f_r(X; \Theta_r)$. This regression model integrates an LSTM layer and a Transformer module. The LSTM layer first processes the input sequence X of shape $N \times d$, capturing immediate temporal dependencies and short-term fluctuations in battery behavior. Formally, it is denoted as:

$$Z_L = L(X) \in \mathbb{R}^{N \times h}, \tag{5}$$

where h is the latent dimension. Next, the Transformer module $T(\cdot)$ applies multi-head self-attention to Z_L, identifying complex, long-range interactions and patterns over the input time window. This step yields a transformed representation as:

$$Z_T = T(Z_L) \in \mathbb{R}^h. \tag{6}$$

The final regression output is computed as:

$$\hat{R}_s = W_r Z_T + b_r, \tag{7}$$

where W_r and b_r are trainable parameters of the output layer. The combination of LSTM and Transformer architectures empowers the regression model to capture both short-term dynamics and long-range dependencies in the battery's temporal data. This synergy ensures that the model remains sensitive to subtle but critical shifts in the battery's behavior, improving the accuracy and robustness of the short resistance estimation.

HyCARLT's hybrid inference process can be succinctly described as:

$$\hat{R}_s(X) = \begin{cases} 2000\Omega, & \text{if } f_c(X; \Theta_c) < \tau; \\ f_r(X; \Theta_r), & \text{if } f_c(X; \Theta_c) \geq \tau. \end{cases} \tag{8}$$

To further enhance the model's performance on rare ISC events, we employ weighted loss functions and data augmentation during training, emphasizing faulty samples and guiding the model to prioritize critical cases. Through this architecture and training strategy, HyCARLT establishes itself as a scalable, accurate, and efficient solution for short resistance estimation in electric vehicle battery diagnostics.

3.4. Loss functions

The choice of loss functions in both classification and regression stages is critical for effectively handling the highly imbalanced nature of the dataset and for ensuring robust fault detection and accurate resistance estimation. Standard loss functions, such as unweighted cross-entropy for classification or plain mean squared error (MSE) for regression, often fail to properly emphasize minority classes or crucial intervals of the output variable, especially when the distribution of fault conditions is severely skewed. To overcome these challenges, this section details the use of weighted loss functions. During training, higher importance (or weight) is assigned to samples with short resistance values lower than 2000 Ω, particularly those below 1000 Ω, as this range is most critical for detecting faults. More specifically, a weight of 0.1 is assigned to healthy cells and 2.0 to faulty cells, thereby emphasizing the importance of accurately predicting faulty samples. By applying this weighted approach, the model could prioritize learning from data where faulty patterns are most likely to emerge, ensuring

that it could accurately capture the complex relationship between input features and short resistance, particularly in the lower short resistance ranges.

For the classification module, a weighted binary cross-entropy loss is employed to balance the contribution of healthy and faulty samples to the gradient updates. Consider a binary label $y_i \in \{0, 1\}$ for the ith training sample, where $y_i = 1$ indicates a fault and $y_i = 0$ corresponds to a healthy sample. Let $\hat{p}_i = f_c(X_i; \Theta_c)$ be the predicted probability that sample i is faulty, with f_c denoting the classification network and Θ_c its parameters. In an imbalanced setting with far fewer faulty samples, a standard binary cross-entropy loss would cause the model to favor predicting the majority class, thus overlooking rare but critical faults. To mitigate this issue, the weighted binary cross-entropy loss is introduced as:

$$\mathcal{L}_{\mathcal{C}} = -\frac{1}{M} \sum_{i=1}^{M} \left[w_c^{(1)} y_i \log(\hat{p}_i) + w_c^{(0)} (1 - y_i) \log(1 - \hat{p}_i) \right], \tag{9}$$

where M is the number of samples, and $w_c^{(1)}$, $w_c^{(0)}$ are weights assigned to the faulty and healthy classes respectively. By increasing $w_c^{(1)}$ relative to $w_c^{(0)}$, the model devotes greater representational capacity to identifying faults, thereby reducing the likelihood of failing to detect rare but significant anomalies.

For the regression module, which estimates the short resistance $\hat{R}_s = f_r(X; \Theta_r)$ for samples identified as faulty by the classification model, a weighted mean squared error (WMSE) loss is employed. Let $R_{s,j}$ and $\hat{R}_{s,j} = f_r(X_j; \Theta_r)$ be the true and predicted short resistance values for the jth faulty sample. Although the classification stage already filters out healthy samples, imbalances persist within the distribution of fault severity levels. Severe faults with very low short resistance values tend to be rarer, but they are of heightened interest in practical settings. Treating all fault instances equally with a standard MSE loss risks underrepresenting these critical low-resistance cases. To address this, WMSE assigns higher weights to critical intervals of the short resistance range, ensuring the model focuses on accurately predicting severe faults:

$$\mathcal{L}_r = \frac{1}{M'} \sum_{j=1}^{M'} w_r^{(j)} \left(R_{s,j} - f_r(X_j; \Theta_r) \right)^2, \tag{10}$$

where M' is the number of faulty samples used for training and $w_r^{(j)}$ is a sample-dependent weight that can be set higher for more critical fault regions. This weighting scheme directs the model's learning process to improve precision where it is most necessary, aligning model performance with practical diagnostic priorities.

The combination of weighted binary cross-entropy for classification and weighted MSE for regression offers a coherent solution to the inherent data imbalance problem. Rather than naively resampling data or relying solely on standard loss functions, these weighted approaches incorporate domain insights directly into the training objective. By manipulating weights, the model can concentrate its learning capacity on fault conditions that matter most, reducing missed detections and enhancing the precision of short resistance estimation in severe cases.

3.5. Training optimization strategy

To ensure efficient and accurate training of the classification and regression modules, tailored optimization strategies are employed. For the classification module, the Adam optimizer with an initial learning rate of 0.001 is utilized to accelerate convergence. A dynamic learning rate schedule is implemented, reducing the learning rate by a factor of 0.1 when the validation loss plateaus, enabling the model to refine its learning during later stages of training. Additionally, early stopping is applied based on validation performance to prevent overfitting, ensuring the model generalizes well to unseen data. The training process is conducted for a maximum of 300 epochs, though early stopping often results in fewer actual training iterations.

The regression module, designed to estimate short resistance values for faulty samples, is similarly optimized using the Adam optimizer with an initial learning rate of 0.001. A dynamic learning rate schedule is applied here as well, adjusting the learning rate during training to balance stability and learning capacity. Early stopping is employed to halt training once the validation loss no longer improves, reducing unnecessary computational cost and preventing overfitting. Given the greater complexity of the regression task, training is conducted for up to 500 epochs, providing sufficient iterations to capture the intricate temporal and contextual relationships in the data.

Both modules utilize mini-batch training to efficiently process the large dataset, and model checkpoints are saved at each epoch to ensure recovery in case of interruptions.

4. Experimental evaluation and discussion

The evaluation strategy and results of the proposed hybrid model for short resistance estimation in EV batteries are presented in this section. Our experimental design aims to comprehensively assess the classification and regression capabilities of HyCARLT while addressing the challenges posed by dataset imbalance and real-world driving conditions. Furthermore, we benchmark the hybrid model's performance against SOTA regression models as baseline methods, highlighting its advantages in accuracy and robustness.

To provide more comprehensive theoretical evidence for our conclusions, we explicitly clarify the rationale behind our experimental design, emphasizing the significance of separating fault classification from resistance estimation. Furthermore, we highlight the electrochemical rationale underpinning the feature set selection and clarify the representativeness and robustness of our real-world dataset.

4.1. Experiment setup

The computational infrastructure for training the DNN models relies on a high-performance computing (HPC) system specifically optimized for machine learning workloads. The system's core processing power is provided by a processor operating at a base clock frequency of 2.8 GHz, with support for 64 threads to enable substantial parallel computational throughput. This is complemented by 512 GB of system memory, essential for handling the large datasets required for deep learning tasks and ensuring efficient execution of memory-intensive operations. GPU acceleration is provided by a pair of GPUs, each equipped with 48 GB of VRAM. These GPUs, optimized for deep learning applications, deliver high memory bandwidth and computational power for matrix-intensive operations, utilizing CUDA Version 12.2 for acceleration.

The training process is conducted in two stages to optimize the performance of the classification and regression tasks. The first stage focuses on training the ResNet-based classification module, which contains approximately 3 million parameters. This module is responsible for distinguishing healthy and faulty samples, requiring about three hours of training time for 300 epochs. The inference time for the classification module is approximately 10 ms per instance, demonstrating its suitability for real-time fault detection. The second stage involves training the regression module, which integrates LSTM and Transformer architectures to estimate the short resistance values for samples classified as faulty. This regression model, comprising approximately 0.6 million parameters, completes training in about two hours for 500 epochs. The LSTM layer captures short-term temporal dependencies, while the Transformer module models complex, long-range interactions. The inference time for this module is approximately 50 ms per instance, ensuring efficient operation during deployment.

This two-stage approach explicitly addresses the severe imbalance by isolating healthy cells in the classification phase, allowing the regression phase to concentrate exclusively on accurately estimating resistance for faulty cells.

4.2. Data sets and evaluation metrics

To validate our approach, we first use an open-source dataset described in [21], consisting of 584,148 training samples and 64,905 testing samples. Each sample consists of sequences with 120 time steps, and the ground truth includes five distinct short resistance levels spanning a broad range. Our proposed regression module is adopted to demonstrate the model's effectiveness on a different data distribution.

Another data employed in this study is collected under real-world driving conditions. The dataset originates from an extensive fleet of over 10,000 electric vehicles monitored for up to two years. The data encompasses diverse operational conditions, including substantial seasonal temperature variations, capturing both summer and winter scenarios—and various realistic driving and charging/discharging profiles. Although extreme temperature conditions (such as unusually low or high ambient temperatures) and extreme charging/discharging events are not explicitly separated or emphasized in isolation, the extensive temporal and seasonal coverage of our dataset ensures a representative range of typical operational conditions encountered by EV batteries in practical use. Therefore, our evaluation provides a robust assessment of our model's performance under realistic operating conditions commonly experienced in real-world EV applications. Approximately 10 samples are gathered per day from each vehicle, resulting in a comprehensive dataset encompassing both healthy and faulty cells. Among these cells, about 380 exhibited ISC faults. To obtain the training data, we first select 380 faulty cells with short resistance less than 2000 Ω. We then randomly sample 380 vehicles from total 10,000 vehicles, without any faulty cells. Since each vehicle contains 96 or 192 cells, the total number of healthy cells is initially much larger than the number of faulty cells. To achieve a balanced training dataset, the faulty cell data is oversampled by duplication, resulting in a double number of faulty samples. Similarly, the healthy data is randomly downsampled, reducing the number of healthy samples by a ratio of approximately 4:1 ratio. After oversampling and downsampling, the ratio of healthy samples to faulty samples is set to be approximately 1:1, ensuring a balanced dataset. After balancing, the combined dataset (containing both faulty and healthy cells) is split into training set, validation set, and testing set using an 80%:10%:10% ratio. This split applies specifically to the balanced dataset, which consists of 380 faulty cells and 380 healthy cells. The training and validation sets are used during model development for parameter tuning and overfitting prevention. The testing set in the combined dataset comprises both healthy and faulty cells (i.e., the 10% portion from the split of the 380-faulty and 380-healthy sample pool). For the cells with short resistance values below 2000 Ω in the testing set, are included in Test Set 1 to evaluate the model's performance to examine the model's fault isolation capability and the model's accuracy on short resistance estimation. In addition to Test Set 1, we establish another distinct data set, Test Set 2. This set is entirely independent of the balanced dataset and contains only healthy cells that are not selected previously. Specifically, it consists of data from all remaining vehicles after excluding the 380 vehicles containing soft faulty cells and the 380 vehicles selected to represent healthy cells for the balanced dataset. Since this test set represents most of the real-world distribution, it provides insights into the model's generalization capability and its robustness in recognizing healthy cells without misclassification.

This data organization ensures comprehensive performance assessment. Test Set 1 enables direct evaluation of classification and regression performance in identifying and quantifying faults, while Test Set 2 examines the model's ability to generalize to a much larger population of unseen healthy cells, thereby ensuring reliable deployment in real-world conditions.

For classification, F1-score is used as the main metric, with recall, precision, and accuracy calculated using Eqs. (11)–(14). Here, True Positive (TP) represents correctly classified faulty samples, True Negative (TN) represents correctly classified healthy samples, False Positive

(FP) is the number of false positives, and False Negative (FN) is the number of false negatives. This approach allowed for a more nuanced evaluation of the model's classification performance.

$$\text{Precision} = TP/(TP + FP),\tag{11}$$

$$\text{Recall} = TP/(TP + FN),\tag{12}$$

$$\text{F1} - \text{score} = \frac{2 * Precision * Recall}{Precision + Recall},\tag{13}$$

$$\text{Accuracy} = (TP + TN)/(TP + FP + FN + TN).\tag{14}$$

For the regression task, we employ multiple evaluation metrics to capture different aspects of prediction accuracy, as shown in Eqs. (15)–(19). The Mean Squared Error (MSE) measures the average of squared prediction errors. The Root Mean Squared Error (RMSE) is the square root of MSE, making it more interpretable in the same units as the target variable. The Mean Absolute Error (MAE) provides a straightforward average of absolute residuals, offering a robust measure against outliers. The Mean Absolute Percentage Error (MAPE) focuses on the relative magnitude of errors, which is particularly helpful for assessing the estimation quality across a wide range of short resistance values. Finally, the coefficient of determination (R^2) quantifies how well the regression model explains the variance in the true labels, with values closer to 1 indicating better fits. Each metric highlights a distinct perspective of regression performance, ensuring a comprehensive evaluation of the proposed model.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2,\tag{15}$$

$$\text{RMSE} = \sqrt{\text{MSE}},\tag{16}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|,\tag{17}$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%,\tag{18}$$

$$R^2 = 1 - \frac{\sum\left(y_i - \hat{y}_i\right)^2}{\sum\left(y_i - \bar{y}\right)^2}.\tag{19}$$

4.3. Evaluation for proposed model

In this section, an ablation study is performed on an open-source dataset to identify the best Transformer-based backbone for ISC resistance estimation. The chosen regression module is compared to a CNN-based baseline, confirming its effectiveness on the same open-source dataset. Finally, the hybrid classification-regression design is evaluated on a real-world driving vehicle dataset and benchmark it against existing SOTA methods, demonstrating the overall performance and robustness of our approach.

4.3.1. A comparative study of CNN and transformer variants

The HyCARLT is first compared with the CNN-based baseline from [21], which relies on multiple Conv1D layers, dropout, and a final dense layer for regression. Both models are trained and tested on the same dataset (containing 584,148 training samples and 64,905 test samples). Table 1 shows that HyCARLT achieves lower MSE, RMSE, and MAPE than the CNN baseline (6.54% vs. 9.10%), while the baseline obtains a smaller MAE. This trade-off occurs because the CNN baseline's convolutional blocks can minimize overall absolute deviations on typical resistance values but are less adaptive to extreme faults, resulting in higher percentage errors for rare yet critical low-resistance conditions. In contrast, the HyCARLT architecture tracks temporal patterns more effectively, handling both small and large resistance ranges. These results highlight the effectiveness of pairing LSTM layers for short-term

sequence modeling with Transformer attention for capturing long-range dependencies, which leads to consistent performance across diverse operating conditions.

To test whether HyCARLT's two-stage design still offers an advantage against the latest Transformer forecasters, we benchmark four Transformer—variants Informer, Autoformer, FEDformer and PatchTST on the dataset. Their results, shown in the final four rows of Table 1, indicate that although several variants (e.g., PatchTST) narrow the regression gap, none can simultaneously match HyCARLT's fault-classification recall and its low relative-error estimates on faulty cells. This confirms that explicitly separating 'fault decision' from 'fault magnitude' – and training each branch on a class-balanced subset – remains crucial when the healthy/faulty ratio exceeds 100:1 in the wild.

4.3.2. Ablation study on the open-source dataset

To assess the effectiveness of different Transformer-based backbones for ISC resistance estimation, an ablation study is conducted using the same open-source dataset. Table 2 compares four model variants: a pure Transformer, CNN+Transformer, deepResNet+Transformer and HyCARLT. Note that in all tables, the best performance are highlighted in blue and bold, while the second-best results are marked in bold for clear comparison.

As shown in Table 2, although the ResNet+Transformer variant yields lower MSE, RMSE, MAE, and a slightly higher R^2, it suffers from a larger MAPE (17.77%). By contrast, HyCARLT yields a MAPE of 6.54%, reflecting more stable relative errors across imbalanced short resistance distributions. One reason for this disparity is that, while ResNet deepens feature extraction and excels in absolute-error metrics, it may overestimate resistance values for lower ranges. LSTM layers, on the other hand, excel at capturing sequential dependencies crucial for fault scenarios where small deviations at low resistances can disproportionately affect relative error. Thus, when MAPE is the primary concern—particularly in highly skewed fault distributions - HyCARLT is the better choice.

4.3.3. Evaluation on the driving vehicle dataset

In the proposed hybrid model, the classification and regression tasks are integrated for optimal performance. The ResNet-based classification model serves as the classification module, determining whether a battery cell is healthy. This classification is formulated as a binary problem, where a label of 1 is assigned to short resistance values below 2000 Ω (indicating a faulty cell), and a label of 0 is assigned to values at 2000 Ω (representing a healthy cell). Trained on this binary labeling, the classification model effectively distinguishes healthy cells from faulty ones.

On Test Set 1, the classification module achieves a recall of 0.92, a precision of 0.88, and an F1-score of 0.90, demonstrating its robustness in distinguishing healthy and faulty cells. An accuracy of 0.91 further validates the model's reliability. On Test Set 2, which includes healthy samples from over 10,000 EVs, the model does not report any false positives.

The performance of HyCARLT surpasses that of the best SOTA method [xresnet1d34] in both test sets, as illustrated in Fig. 3. On Test Set 1, HyCARLT achieves a relative error of 15.0%, significantly lower than the 26.6% achieved by the SOTA method. Figs. 3(a) and 3(c) demonstrate that the SOTA method exhibits higher errors for short resistance values below 750 Ω, while HyCARLT maintains a more uniform estimation error, almost consistently below 20%. Notably, Figs. 3(a) and 3(c) display only the distribution of estimated values below 2000 Ω, primarily focusing on the estimation results for faulty cells. In contrast, Figs. 3(b) and 3(d) illustrate the distribution of estimated values at 2000 Ω and above, mainly emphasizing the estimation results for healthy cells. According to Eq. (9), relative error becomes more sensitive for lower ground truth values; for instance, estimating 100 Ω as 150 Ω results in a 50% error, whereas estimating 1000 Ω as 1050 Ω only results in a 5% error. This highlights HyCARLT's superior accuracy

Table 1

Comparison with CNN and transformer variants using an open-source dataset.

Model	MSE (Ω^2)	RMSE (Ω)	MAE (Ω)	MAPE (%)	R^2
CNN-Based Baseline [21]	4.99×10^9	7.06×10^4	2.74×10^4	9.10	0.98
HyCARLT (Proposed)	**3.39×10^9**	**5.82×10^4**	**3.29×10^4**	**6.54**	**0.98**
Informer [22]	1.13×10^{10}	1.06×10^5	4.55×10^4	12.39	0.95
Autoformer [23]	4.05×10^9	6.36×10^4	3.80×10^4	10.81	0.97
FEDformer [24]	3.85×10^9	6.20×10^4	3.60×10^4	9.62	0.97
PatchTST [25]	**6.89×10^9**	**8.30×10^4**	**3.33×10^4**	**8.29**	**0.97**

Table 2

Ablation study on transformer-based regression modules using the open-source dataset. Each row represents a different backbone.

Model	MSE (Ω^2)	RMSE (Ω)	MAE (Ω)	MAPE (%)	R^2
Transformer Only	2.29×10^{10}	1.51×10^5	6.13×10^4	3881.05	0.89
CNN + Transformer	2.98×10^{10}	1.72×10^5	1.00×10^5	19.41	0.87
ResNet + Transformer	**2.72×10^8**	**1.65×10^4**	**8.42×10^3**	**17.77**	**0.99**
HyCARLT (Proposed)	**3.39×10^9**	**5.82×10^4**	**3.29×10^4**	**6.54**	**0.98**

Note: the error magnitudes are large because the ground-truth short resistance for healthy cells is set to $10^6\,\Omega$ in the source dataset [21]; false-positive predictions therefore inflate error metrics.
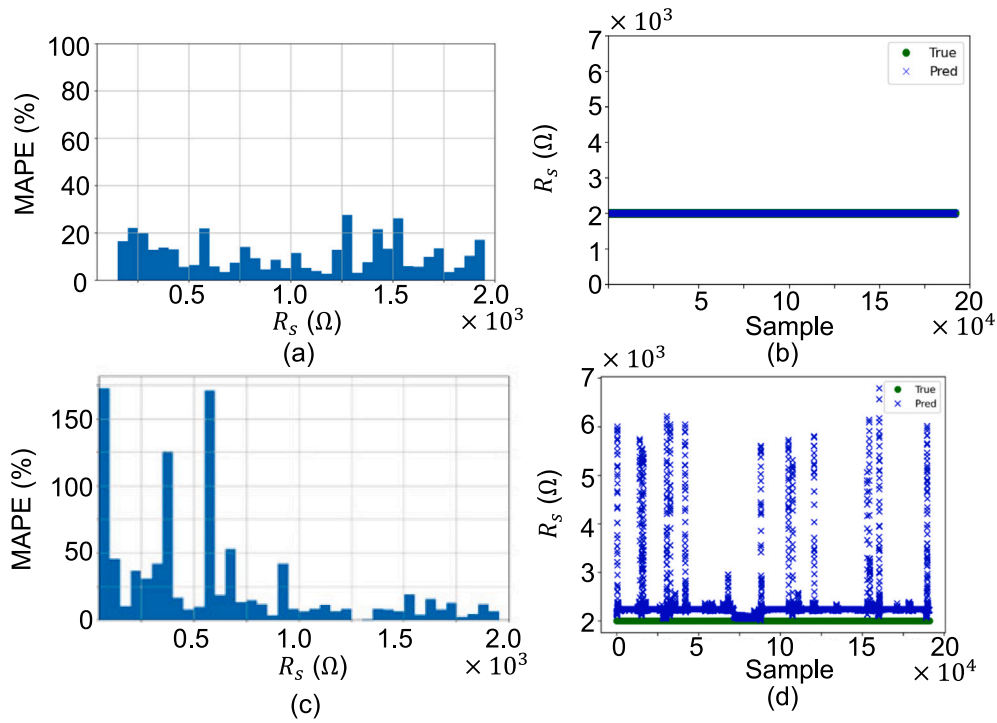
Fig. 3. Performance of HyCARLT and the best SOTA method [xresnet1d34_deeperplus] on two test sets. (a) Evaluation on Test Set 1 using HyCARLT; (b) Evaluation on Test Set 2 using HyCARLT; (c) Evaluation on Test Set 1 using the best SOTA method; (d) Evaluation on Test Set 2 using the best SOTA method.

for lower short resistance values, where it closely approximates the ground truth.

On Test Set 2, HyCARLT did not misclassify any healthy cells as faulty, while the best time-series regression analysis method still misclassified some healthy cells. As shown in Fig. 3(d), directly applying regression analysis to Test Set 2 results in fluctuating short resistance estimates, with some estimated values dropping below 2000 Ω. In contrast, as illustrated in Fig. 3(b), the classification model constrains the resistance estimates of healthy cells to 2000 Ω, preventing fluctuations that could lead to false positives. HyCARLT demonstrates superior performance compared to existing time-series regression analysis methods, excelling particularly in scenarios with lower short resistance values and achieving error-free classification and regression for healthy cells.

These findings show that the hybrid classification-regression strategy of HyCARLT generalizes well to diverse datasets, effectively addresses data imbalance, and surpasses prominent single-stage regression architectures for ISC resistance estimation.

4.4. Performance comparison

Having established HyCARLT's efficacy on both our real-world dataset and the open-source dataset, we next extend our evaluations to a broader set of advanced time-series regression models. The goal is to confirm that our hybrid design consistently outperforms conventional single-stage regression approaches across multiple battery diagnostic scenarios. The performance of various time-series regression models is compared, to highlight the advantages of our approach in handling the unique challenges posed by the highly imbalanced dataset used

Table 3
Performance of models with different architectures.

Model	Relative error (%)
FCN [15]	29.5
FCNPlus [16]	29.6
InceptionTime [33]	23.4
InCoordTime [16]	26.9
XCoordTime [16]	26.1
InceptionTimePlus [16]	19.5
InceptionTimeXLPlus [16]	19.5
MultiInceptionTimePlus [16]	24.1
InceptionRocketPlus [16]	83.1
MLP [34]	88.6
gMLP [20]	**15.7**
OmniScaleCNN [35]	55.8
RNN [17]	48.5
LSTM [17]	48.8
RNNPlus [16]	45.0
LSTMPlus [16]	47.4
GRUPlus [16]	43.2
RNN_FCN [16]	29.8
LSTM_FCN [17]	30.7
GRU_FCN [36]	29.8
MLSTM_FCN [17]	29.7
MGRU_FCN [16]	27.9
RNN_FCNPlus [16]	28.8
LSTM_FCNPlus [16]	30.2
GRU_FCNPlus [16]	30.5
MRNN_FCNPlus [16]	25.9
MLSTM_FCNPlus [16]	33.4
MGRU_FCNPlus [16]	26.2
ResCNN [37]	24.2
ResNet [18]	19.8
ResNetPlus [16]	18.9
TCN [21]	30.6
MultiTSTPlus [16]	19.9
TransformerModel [19]	73.2
XCM [38]	51.9
XCMPlus [16]	48.3
xresnet1d18plus [16]	30.4
xresnet1d34plus [16]	25.1
xresnet1d50plus [16]	25.6
xresnet1d101plus [16]	33.9
xresnet1d152plus [16]	26.1
xresnet1d18_deepplus [16]	25.9
xresnet1d34_deepplus [16]	21.1
xresnet1d50_deepplus [16]	19.5
xresnet1d18_deeperplus [16]	26.8
xresnet1d34_deeperplus [16]	26.6
xresnet1d50_deeperplus [16]	26.0
XceptionTime [39]	23.4
XceptionTimePlus [16]	24.1
TSSequencer [16]	21.6
TSSequencerPlus [16]	20.4
ConvTran [40]	33.9
ConvTranPlus [16]	35.4
RNNAttention [41]	22.5
LSTMAttention [42]	22.8
GRUAttention [43]	21.6
RNNAttentionPlus [16]	20.6
LSTMAttentionPlus [16]	22.3
GRUAttentionPlus [16]	22.6
TransformerRNNPlus [16]	225.1
TransformerGRUPlus [16]	227.6
HyCARLT (Proposed)	**15.0**

To highlight the efficacy of our proposed hybrid methodology, HyCARLT, it is instructive to compare its performance against an extensive set of SOTA regression approaches that tackle the short-circuit resistance estimation task without any preliminary classification stage. These alternative models, drawn from a wide array of neural architectures, operate on the entire dataset directly and attempt to fit the severely imbalanced distribution of short-circuit resistance values in a single step. Table 3 provides a comprehensive benchmarking of numerous baseline architectures—including convolutional, recurrent, and transformer-based models. The goal is to achieve low error on faulty samples, where the estimation is challenging and critical.

As illustrated in Table 3, architectures like MLP or Transformer-Model may handle simpler distributions but yield large errors when confronting the highly skewed data, resulting in poor error metrics on Test Set 1. Even more specialized networks, such as various InceptionTime-based, RNN-based or FCN-based hybrids, often fail to overcome the intrinsic imbalance, showing inflated errors particularly in the lower resistance ranges where faults are rarer and inherently harder to model.

In contrast, HyCARLT achieves superior performance, which reports a relative short-resistance estimation error of 15%, substantially outperforming most SOTA regression models. The classification results on Test Set 2 are not reported since we mainly evaluate the regression model capability in this work. Notably, even some architectures tuned for time-series tasks – such as multi-scale convolutional networks or recurrent networks with attention – fall short of HyCARLT's balanced performance. While certain regression architectures may show intermittent strong performance on one of the test sets, none match HyCARLT's consistently strong performance on both. This comparison underscores the core advantage of the hybrid strategy: rather than forcing a single model to simultaneously master the trivial but dominant healthy regime and the complex, minority faulty regime, HyCARLT assigns complementary tasks to specialized modules. The result is a method that not only outperforms individual baselines but also sets a higher standard for addressing extreme imbalance in battery short-circuit resistance estimation.

The comparison with SOTA regression approaches reinforces the significance of our hybrid design. The ResNet front-end isolates abrupt OCV drops, the LSTM captures mid-range balancing dynamics, and the Transformer models long-range patterns, enabling the pipeline to remain sensitive to early ISC signatures. The end-to-end HyCARLT pipeline excels where traditional models fail, delivering improvements in fault detection accuracy, short resistance estimation, and overall reliability. This comprehensive performance uplift across various neural architectures illustrates that the hybrid paradigm is not a marginal enhancement, but rather a robust, scalable solution to the longstanding imbalance challenges in the EV battery diagnostic domain.

The robust theoretical foundation of our two-stage hybrid design, electrochemically grounded feature set, and the extensive representativeness of our dataset collectively provide comprehensive intrinsic support for our results. These methodological choices significantly mitigate common error sources encountered by purely regression-based approaches, theoretically justifying our improved ISC detection and resistance estimation performance.

5. Conclusion

This work proposes a deep learning model that integrates a ResNet, LSTM, and Transformer architectures for short resistance estimation in EV batteries. Validated on data from over 10,000 vehicles, the method achieves a relative estimation error of 15% on faulty cells and introduces no false positives for healthy cells, significantly outperforming traditional approaches such as Cell Droop Rate and dSOC, which can exhibit errors of up to 50%–80% under certain conditions. By augmenting the dataset – oversampling faulty cells and undersampling healthy cells – and employing a deep-residual-network-based classification module,

in this study. These models represent SOTA approaches used for time-series analysis short resistance estimation tasks. All models have been retrained using our training dataset.

As shown in Table 3, the models include a range of architectures, from simple fully connected networks (FCN) to advanced models like ResNet, Transformer, and temporal convolutional networks (TCN). Each model is tested on Test Set 1 to evaluate its ability to handle regression tasks under varying conditions.

the model effectively mitigates data imbalance and focuses the regression stage on the critical faulty subset. Although labels are derived from an existing algorithm due to the absence of ground truth, the model demonstrates adaptability by learning intrinsic data patterns rather than simply mirroring the labeling mechanism. This design ensures that it can be retrained with experimentally obtained or expert-annotated labels should they become available, preserving accuracy and effectiveness.

Despite these promising results, several limitations remain. First, the model's performance is contingent on the coverage and representativeness of the training data, which may not capture every real-world operational regime. Second, while the classification step addresses overall imbalance, extremely rare or novel fault scenarios might still require additional data collection or refined augmentation strategies. Future work will explore incorporating physics-informed constraints into the network, further expanding the diversity of training samples, and evaluating the approach across different battery chemistries and usage profiles, which would enhance the method's interpretability, broaden its applicability, and ensure more robust real-time diagnostics for EV batteries.

CRediT authorship contribution statement

Yongtao Yao: Writing – original draft, Visualization, Software, Methodology, Investigation. **Xinyu Du:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Shengbing Jiang:** Methodology, Investigation, Conceptualization. **Rasoul Salehi:** Methodology, Investigation, Conceptualization. **Weisong Shi:** Writing – review & editing, Supervision.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work we used ChatGPT in order to improve readability and language of the work. The tool was used solely for language polishing, and all figures, tables, formulas, and the ideas presented in the manuscript are entirely original or derived from relevant references. After using this tool, we reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Data availability

The data that has been used is confidential.

References

[1] J. Hu, Z. Wei, H. He, An online adaptive internal short circuit detection method of lithium-ion battery, Automot. Innov. 4 (2021) 93–102.

[2] D. Qiao, X. Wei, B. Jiang, W. Fan, H. Gong, X. Lai, Y. Zheng, H. Dai, Data-driven fault diagnosis of internal short circuit for series-connected battery packs using partial voltage curves, IEEE Trans. Ind. Inform. (2024).

[3] J. Hu, Z. Wei, H. He, Improved internal short circuit detection method for lithium-ion battery with self-diagnosis characteristic, in: IECON 2020 the 46th Annual Conference of the IEEE Industrial Electronics Society, IEEE, 2020, pp. 3741–3746.

[4] G. Li, B. Li, S. Wang, P. Chen, C. Li, B. Liu, A fast model-free method for identifying internal short circuit resistance in lithium-ion battery, IEEE Trans. Power Electron. (2024).

[5] A. Chen, W. Zhang, B. Sun, H. Li, X. Fan, Online estimation of internal short circuit resistance for large-format lithium-ion batteries combining a reconstruction method of model-predicted voltage, World Electr. Veh. J. 13 (9) (2022) 170.

[6] J. Newman, N.P. Balsara, Electrochemical Systems, John Wiley & Sons, 2021.

[7] D.A.J. Rand, R. Woods, Batteries for Electric Vehicles, 1981.

[8] M. Doyle, T.F. Fuller, J. Newman, Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell, J. Electrochem. Soc. 140 (6) (1993) 1526.

[9] H. He, R. Xiong, J. Fan, Evaluation of lithium-ion battery equivalent circuit models for state of charge estimation by an experimental approach, Energies 4 (4) (2011) 582–598.

[10] Y.-Y. Wang, Y. Zeng, T.R. Garrick, A.C. Baughman, Method and system for self-discharge prognostics for vehicle battery cells with an internal short circuit, 2023. US Patent number: US 11, 733, 309 B2.

[11] Y.-Y. Wang, A.C. Baughman, Thermal runaway prognosis by detecting abnormal cell voltage and soc degeneration, 2022. US Patent Publication Number: US 20220352737A1.

[12] Z. Chen, M. Sun, X. Shu, R. Xiao, J. Shen, Online state of health estimation for lithium-ion batteries based on support vector machine, Appl. Sci. 8 (6) (2018) 925.

[13] C. Zhao, P.B. Andersen, C. Træholt, S. Hashemi, Data-driven battery health prognosis with partial-discharge information, J. Energy Storage 65 (2023) 107151.

[14] K.A. Severson, P.M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M.H. Chen, M. Aykol, P.K. Herring, D. Fraggedakis, et al., Data-driven prediction of battery cycle life before capacity degradation, Nat. Energy 4 (5) (2019) 383–391.

[15] Z. Wang, W. Yan, T. Oates, Time series classification from scratch with deep neural networks: A strong baseline, in: 2017 International Joint Conference on Neural Networks, IJCNN, IEEE, 2017, pp. 1578–1585.

[16] I. Oguiza, V. Rodriguez-Fernandez, D. Neoh, filipj8, J-M, R. Kainkaryam, D. Mistry, Z. Yang, D. Williams, R. Cho, geoHeil, C. Versek, samlloydig, D. Mistry, imilas, A. Golinski, D. Muhr, H. Jeff, MichaelDee, T. Capelle, timeseriesAI/tsai: V0.3.9 (0.3.9), Zenodo, 2024, [Online]. Available: URL https://doi.org/10.5281/zenodo.10647659.

[17] F. Karim, S. Majumdar, H. Darabi, S. Harford, Multivariate LSTM-FCNs for time series classification, Neural Netw. 116 (2019) 237–245.

[18] N.M. Dipu, S.A. Shohan, K. Salam, Ocular disease detection using advanced neural network based classification algorithms, Asian J. Converg. Technol. (ISSN: 2350-1146) 7 (2) (2021) 91–99.

[19] Y. Tatsunami, M. Taki, Sequencer: Deep lstm for image classification, Adv. Neural Inf. Process. Syst. 35 (2022) 38204–38217.

[20] H. Liu, Z. Dai, D. So, Q.V. Le, Pay attention to mlps, Adv. Neural Inf. Process. Syst. 34 (2021) 9204–9215.

[21] Y. Jia, J. Xu, Data-driven short circuit resistance estimation in battery safety issues, J. Energy Chem. 79 (2023) 37–44.

[22] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, W. Zhang, Informer: Beyond efficient transformer for long sequence time-series forecasting, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, 2021, pp. 11106–11115.

[23] H. Wu, J. Xu, J. Wang, M. Long, Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting, Adv. Neural Inf. Process. Syst. 34 (2021) 22419–22430.

[24] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, R. Jin, Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting, in: International Conference on Machine Learning, PMLR, 2022, pp. 27268–27286.

[25] Y. Nie, N.H. Nguyen, P. Sinthong, J. Kalagnanam, A time series is worth 64 words: Long-term forecasting with transformers, 2022, arXiv preprint arXiv:2211.14730.

[26] Y. Wang, J. Xu, F. Ma, S.-L. Huang, D.D. Sun, X.-P. Zhang, Psformer: Parameter-efficient transformer with segment attention for time series forecasting, 2024, arXiv preprint arXiv:2411.01419.

[27] F. Cao, S. Yang, Z. Chen, Y. Liu, L. Cui, Ister: Inverted seasonal-trend decomposition transformer for explainable multivariate time series forecasting, 2024, arXiv preprint arXiv:2412.18798.

[28] L. Kong, E. Chen, Y. Chen, Y. Han, Teaformers: Tensor-augmented transformers for multi-dimensional time series forecasting, 2024, arXiv preprint arXiv:2410.20439.

[29] G. Zhu, T. Sun, Y. Xu, Y. Zheng, L. Zhou, Identification of internal short-circuit faults in lithium-ion batteries based on a multi-machine learning fusion, Batteries 9 (3) (2023) 154.

[30] A. Graves, A. Graves, Long short-term memory, Supervised Seq. Label. Recurr. Neural Netw. (2012) 37–45.

[31] Y. Toughzaoui, S.B. Toosi, H. Chaoui, H. Louahlia, R. Petrone, S. Le Masson, H. Gualous, State of health estimation and remaining useful life assessment of lithium-ion batteries: A comparative study, J. Energy Storage 51 (2022) 104520.

[32] A. Vaswani, Attention is all you need, Adv. Neural Inf. Process. Syst. (2017).

[33] H. Ismail Fawaz, B. Lucas, G. Forestier, C. Pelletier, D.F. Schmidt, J. Weber, G.I. Webb, L. Idoumghar, P.-A. Muller, F. Petitjean, Inceptiontime: Finding alexnet for time series classification, Data Min. Knowl. Discov. 34 (6) (2020) 1936–1962.

[34] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.-A. Muller, Deep learning for time series classification: A review, Data Min. Knowl. Discov. 33 (4) (2019) 917–963.

[35] W. Tang, G. Long, L. Liu, T. Zhou, M. Blumenstein, J. Jiang, Omni-scale cnns: A simple and effective kernel size configuration for time series classification, 2020, arXiv preprint arXiv:2002.10061.

[36] N. Elsayed, A.S. Maida, M. Bayoumi, Deep gated recurrent and convolutional network hybrid model for univariate time series classification, 2018, arXiv preprint arXiv:1812.07683.

[37] S. Bose, A. Dey, Rescnn: An alternative implementation of convolutional neural networks, in: 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering, UPCON, IEEE, 2021, pp. 1–5.

[38] K. Fauvel, T. Lin, V. Masson, É. Fromont, A. Termier, Xcm: An explainable convolutional neural network for multivariate time series classification, Mathematics 9 (23) (2021) 3137.

[39] E. Rahimian, S. Zabihi, S.F. Atashzar, A. Asif, A. Mohammadi, Xceptiontime: A novel deep architecture based on depthwise separable convolutions for hand gesture classification, 2019, arXiv preprint arXiv:1911.03803.

[40] N.M. Foumani, C.W. Tan, G.I. Webb, M. Salehi, Improving position encoding of transformers for multivariate time series classification, Data Min. Knowl. Discov. 38 (1) (2024) 22–48.

[41] L. Feng, F. Tung, H. Hajimirsadeghi, M.O. Ahmed, Y. Bengio, G. Mori, Attention as an RNN, 2024, arXiv preprint arXiv:2405.13956.

[42] Q. Kang, E.J. Chen, Z.-C. Li, H.-B. Luo, Y. Liu, Attention-based LSTM predictive model for the attitude and position of shield machine in tunneling, Undergr. Space 13 (2023) 335–350.

[43] X. Li, CNN-GRU model based on attention mechanism for large-scale energy storage optimization in smart grid, Front. Energy Res. 11 (2023) 1228256.