This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/JBHI.2024.3392648, IEEE Transactions on Transportation Electrification

Periodic Segmentation Transformer-Based Internal Short Circuit Detection Method for Battery Packs

Zhekang Dong, *Senior Member*, *IEEE*, Shenyu Gu, Shiqi Zhou, Mengjie Yang, Chun Sing Lai, *Senior Member, IEEE*, Mingyu Gao, Xiaoyue, Ji, *Member, IEEE*

***Abstract*—With the rapid evolution of electric vehicles (EVs), assuring the security and dependability of battery packs has acquired paramount significance. Internal short circuit (ISC) within EV battery packs poses a threat to the safety and reliability of EVs. Most of existing ISC detection methods still suffer from two limitations, i.e., the dataset incompleteness and poor feature representation. To address these challenges, we develop a periodic segmentation Transformer-based ISC detection method for battery packs. Firstly, considering three different operating conditions, a comprehensive dataset encompassing three distinct ISC severity levels is constructed. Secondly, to facilitate understanding of the proposed model design, a discrete wavelet transform-based periodicity analysis module and a time-oriented segmenting module are developed. This dual-module design empowers the model to adjust the length of sliding windows adaptively, and enables the joint capture of temporal-spatial and periodic information, significantly enhancing the feature representation ability. Thirdly, experimental results show that our method outperforms the best state-of-the-art in terms of accuracy (average F1 score improved by 24.2%). Finally, robustness analysis and generalization analysis are conducted. The former one demonstrates robustness in terms of parameters within the adaptive aggregation module and input data length; the latter one demonstrates generality of feature extraction method.**

***Index Terms*—Internal short circuit, fault detection, battery packs, transformer-based neural network**

This work was supported in part by the National Natural Science Foundation of China under Grants (62206062 and 62001149), National Major Scientific Research Instrument Development Project of China under Grant 62227802, and Fundamental Research Funds for the Provincial University of Zhejiang under Grant GK229909299001-06. (Corresponding author: *Xiaoyue Ji*).

Z Dong, S Gu, S Zhou, M Yang, and M Gao are with the School of Electronics and Information, Hangzhou Dianzi University, Hangzhou, China, 310018, (e-mail: englishp@hdu.edu.cn; shenyugu@hdu.edu.cn; shiqizhou@hdu.edu.cn; 221040056@hdu.edu.cn; mackgao@hdu.edu.cn)

L C Sing is with Department of Electronic and Computer Engineering, Brunel University London, London, UB8 3PH, UK, (e-mail: chunsing.lai@brunel.ac.uk)

X. Ji is with the Center for Brain-Inspired Computing Research (CBICR), Beijing Innovation Center for Future Chip, Optical Memory National Engineering Research Center, Department of Precision Instrument, Tsinghua University, Beijing 100084, China. (e-mail: jixiaoyue@mail.tsinghua.edu.cn).

ABBREVIATIONS AND NOMENCLATURE

EVs	Electric Vehicles
ISC	Internal Shor Circuit
V_{ocv}	Open Circuit Voltage
V_t	Terminal Voltage
R_{isc}	Internal Short Circuit Resistant
R_i	Internal Resistant
I_{isc}	Internal Short Circuit Current
FUDS	Federal Urban Driving Service
UDDS	Urban Dynamometer Driving Scheme
US06	US06 Supplemental Federal Test Procedure
DWT	Discrete Wavelet Transform
POT	Peak Over Threshold
GPD	Generalized Pareto Distribution

I. INTRODUCTION

A. Background

The electric vehicle (EV) industry is flourishing on a global scale, driven by the imperative to reduce environmental pollution as fossil fuel consumption escalates and carbon neutralization targets loom [1]. EVs have emerged as an important solution for consumers seeking to align with these objectives. To meet the voltage and capacity requirements of EVs, numerous individual batteries are configured through series and parallel connections, forming battery packs [2]. Thermal runaway caused by internal short circuit (ISC) in batteries always lead to serious damage to the entire systems and even threaten people's lives [3, 4]. Accurate and timely ISC detection can be achieved through the monitoring of abnormal voltage signals, which is of utmost significance to prevent thermal runaway [5].

B. Literature review

Numerous ISC detection methods have been introduced in existing literature, falling into three primary categories [3]: parameter inconsistency-based method, model-based method, and data-driven-based method.

Parameter inconsistency-based methods mainly considering a natural phenomenon that parameters (e.g., voltage and capacity) will deviate from consistency once fault occurs [6]. Based on this, many scholars [6-9] proposed parameter inconsistency-based ISC detection methods. However, these methods need a robust cell balancing scheme within battery packs and tend to exhibit low sensitivity to minor ISC

TABLE I
PARAMETERS OF EXPERIMENT PLATFORM

Parameters	Value	Parameters	Value
Number of batteries	60	Number of battery groups	20
Nominal voltage	3.7V	Weight	48g
Normal capacity	2600mAh	Internal resistance	≤60mΩ
Diameter	18.40mm	Height	65mm

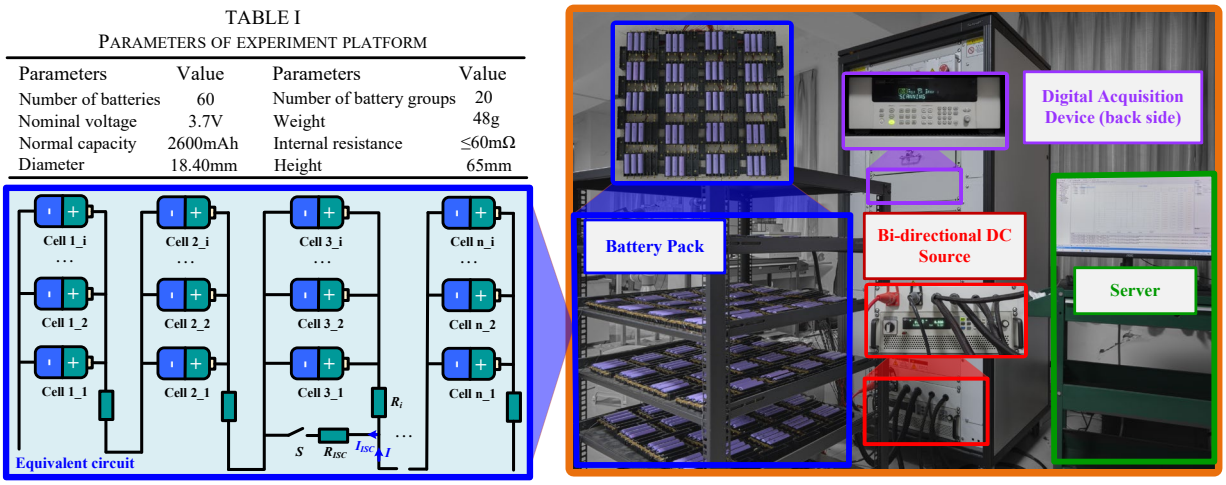

Fig. 1. Experimental platform and battery parameters

occurrences, thereby being susceptible to the influences of sensor accuracy and signal noise. Model-based method is to transform the ISC detection into model parameter/state estimation [10-12]. These kinds of methods encounter difficulties in maintaining high accuracy across varying temperatures. Meanwhile, they exhibit heightened sensitivity to model errors, leading to substantial noise and diminished precision [13]. In addition, model parameters need to be continuously updated to ensure accuracy in the presence of intricate operating conditions and battery aging.

With the rapid development of machining learning and deep learning technologies, data-driven-based methods have shown tremendous potential in the field of ISC detection. Compared to conventional machine learning technologies [14-17], deep learning technologies have the capacity to capture more profound relationships within data and exhibits enhanced generalization capabilities. A long-sequence voltage series forecasting method for ISC detection was proposed in [18], achieving early detection and warning. Cao et al. proposed an ISC diagnosis algorithm for battery packs by combining mean-difference model and bi-directional long short-term memory (Bi-LSTM) neural network [19]. Cui et al. developed a deep neural network using the electrochemical impedance spectroscopy spectrum as the input feature to predict the occurrence probability of ISC [20]. Wang et al. designed a LSTM hybrid neural network that can generate residual signals to detect ISC [21]. However, these models introduce internal memory states to maintain and update information, which may struggle to capture distant feature relationships in long time sequences. Transformer with self-attention mechanism is a potential remedy to address this problem [22-26]. Applying the Transformer into ISC detection may face two following challenges:

1) *Incomplete datasets*: Transformer-based methods need substantial data, and the use of incomplete datasets may lead to disparities between experimental outcomes and real-world scenarios.

2) *Hyperparameter sensitivity*: Transformer-based methods are always sensitive to hyperparameters, especially to the length of sliding windows, which can influence the efficiency of feature extraction and impact results.

Based on these, developing a comprehensive and publicly accessible dataset, enhancing feature extraction capabilities, and adaptively adjusting hyperparameters are crucial for improving accuracy and reliability of ISC detection model in real-world applications.

C. Contribution of our study

To fully exploit the potential of Transformer in ISC detection, the main contributions of this paper can be summarized as follows:

- To address the incomplete dataset issue, a comprehensive dataset based on series-parallel batteries is constructed. Dataset includes three distinct operating conditions and three varying severity levels of ISC.
- In order to reduce the impact of hyperparameter sensitivity on model performance, an adaptive adjustment of the length of sliding windows based on periodicity analysis module and time-oriented segmentation module is proposed.
- A novel periodic segmentation Transformer model capable of extracting temporal-spatial and periodic information simultaneously for ISC detection within battery packs is proposed. The model is demonstrated to have generalized feature extraction and robustness to input data length. Compare to the state-of-the-art (SOTA) models, this method demonstrates an average F1 score improvement of 24.2%.

D. Organization of this paper

Section II describes a complete dataset construction process including data acquisition and the preprocessing steps. Section III describes the ISC detection process, model construction, training, and testing procedure. To verify the effectiveness and robustness of the proposed method, a series of comparative experiments and analysis are conducted in Section IV. Finally, conclusions are drawn in Section V.

II. DATASET

Considering the incompleteness and non-public nature of the datasets utilized in almost all existing works, a comprehensive ISC dataset is constructed. This dataset encompasses voltage time-series data across three distinct ISC

severity levels under three distinct operating conditions. The specific data acquisition and preprocessing are described below.

A. Data acquisition

The specific experimental platform and battery parameters are provided in Fig. 1. Specifically, the ISC experimental platform (as shown in the right of Fig. 1) includes two battery packs, each pack consists of 60 series-parallel Li-ion batteries, a bi-directional DC source (ITECH IT6012-500-80), a high-precision digital acquisition device (KEYSIGHT 34980A), and a server (Advantech IPC-610). The voltage and current measurements exhibit a precision of 0.1%, with each voltage sensor possessing a maximum range of 5 V. The sampling frequency is set as 10 Hz. This configuration comprises 60 lithium-ion batteries (corresponding battery parameters are collected in Table I) divided into 20 groups, with parallel connections within each group and series connections between the groups, forming a battery pack. The initial charging entails reaching 4.2 V through a constant current of 1 C. All batteries underwent pre-experiment testing and exhibited normal performance. Prior to the discharge test, a one-hour resting period was observed before conducting the failure experiments, to minimize voltage variations between battery cells. In practical scenarios, the scope of battery system troubleshooting is typically confined to time-series signals such as voltage, temperature, and current [27]. In ISC detection, the time-series signal is always selected as voltage.

The connection diagram of the battery pack and ISC generator is shown in the left of Fig. 1. $Cell_{n\_i}$ is the number of the battery in battery pack, V_{ocv} is the open circuit voltage, R_{isc} and R_i are the ISC resistant and internal resistant, respectively. I_{isc} is the ISC current, and I is the total current of one cell. When the switch is turned off, the battery is in the normal state, and the equivalent equation can be written by:

$$V_t = V_{OCV} + R_i I \tag{1}$$

When the switch is turned on, the battery is in a faulty state, the equivalent equation can be rewritten by:

$$V_t = V_{OCV} + R_i (I - I_{ISC}) \tag{2}$$

To ensure the completeness of the dataset and validate the generalization of the model, a series of tests are conducted under three operating conditions at room temperature, Federal Urban Driving Service (FUDS), Urban Dynamometer Driving Scheme (UDDS), and US06 Supplemental Federal Test Procedure (US06). For each condition, three power resistors are employed with values of 1 Ω, 3 Ω, and 5 Ω to simulate high, medium and low ISC severity respectively, these values are demonstrated [6, 7] to be closer to severity in real-world scenarios. The training set comprises voltage data from 60 cells in their normal state, while the testing set consisting of voltage data from another set of 60 cells are randomly chosen under the same operating condition to mimic ISC. For each operating condition, 120000 sample points are collected as training sets and 120000 sample points are collected as testing sets for each severity level of ISC. Faulty battery groups are randomly generated, with the total number of selected groups ranging from 3 to 5. The specific faulty battery groups are listed in the Table II. The model is trained separately for each operating condition.

TABLE II
FAULTY BATTERY GROUPS

Operation condition	Severity level	Faulty battery groups
FUDS	Low	4, 7, 9, 16
	Medium	1, 9, 15, 19
	High	7, 9, 10, 14, 15
UDDS	Low	3, 6, 9, 18, 10
	Medium	3, 10, 16, 19
	High	3, 7, 16
US06	Low	5, 6, 11, 14
	Medium	8, 9, 19, 20
	High	1, 6, 17

A comparison of different datasets is collected in Table III.

TABLE III
COMPARISON OF DIFFERENT DATASET

Dataset	Number of ISC Severity	Number of Operating Conditions	Battery connection method
Ours	3	3	Series-parallel
Moeini dataset [6]	3	×	Single battery
Ma dataset [7]	2	1	Single battery
Lai dataset [8]	3	1	Series
Meng dataset [9]	3	1	Single battery
Feng dataset [10]	1	1	Single battery
Hu dataset [11]	2	1	Single battery
Kong dataset [12]	3	×	Single battery
Naha dataset [14]	2	×	Single battery
Kriston dataset [15]	2	×	Single battery
Chen dataset [16]	1	1	Single battery
Cui dataset [18]	4	1	Single battery
Cui dataset [20]	3	×	Single battery
Wang dataset [21]	5	×	Series

Note: "×" represents the operating condition is not considered.

To better simulate the real-world scenarios, two initiatives are implemented: 1) Compared with the battery connection method (single battery or series connection) in existing datasets [6-12, 14-16, 18, 20, 21], the series-parallel connection of batteries are used in our dataset to meet voltage and capacity requirements in EVs. 2) Our dataset considers three operating conditions with three severity levels of ISC fault. Both of these, our dataset is a more comprehensive one compared to existing datasets.

B. Data preprocess

The voltage data from each individual group within the battery pack is selected as the input. The acquired battery voltage data can be represented using a matrix format: $\mathbf{X} \in \mathbb{R}^{T \times M}$, T represents the length of the battery voltage data, and M represents the number of groups within the battery pack. In the collected dataset, labels of the same dimensions (0 and 1) are provided, where '0' denoting the battery is at normal condition, and '1' denoting the presence of ISC. To enhance the robustness of the entire system, the data normalization is applied across the training, validation, and testing sets. The normalization process can be mathematically expressed by:

$$x_t = \frac{x_t - \min(\Gamma)}{\max(\Gamma) - \min(\Gamma) + \varepsilon} \tag{3}$$

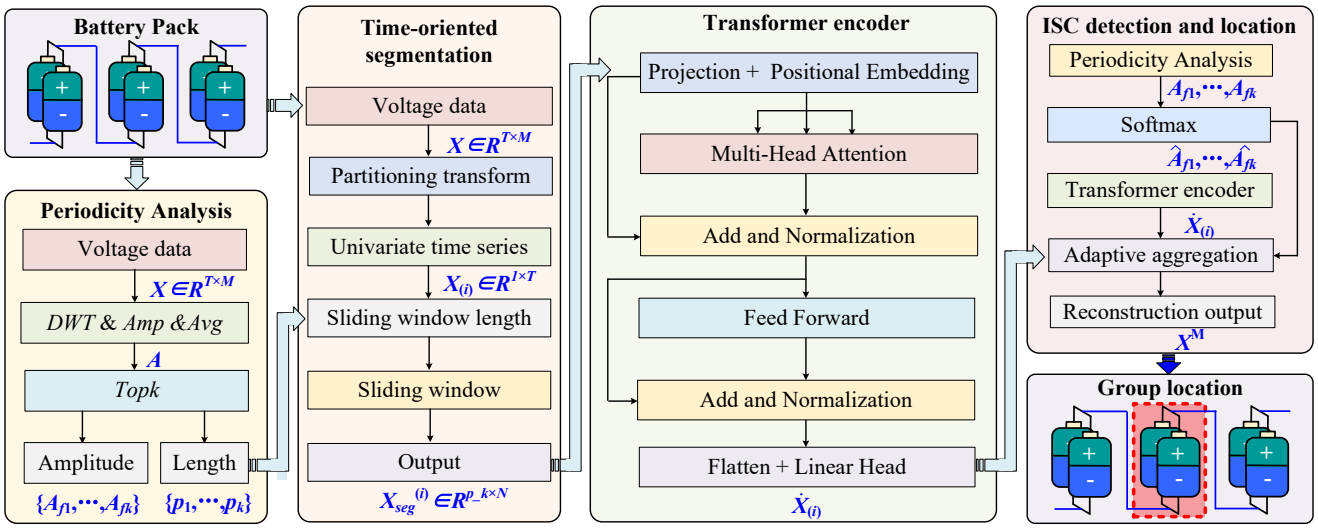

Fig. 2. Structure of periodic segmentation Transformer model.

where $\min(\Gamma)$ and $\max(\Gamma)$ represent the maximum and minimum vectors for each cell within the training set respectively. To prevent division by zero, a small constant ε is added. Through normalization, the data is confined within the interval of [0,1).

III. Methodology

The process for ISC detection of EV battery packs is outlined as follows:

1) *Feature extraction*: The original data from the training set is injected to the periodic segmentation Transformer model. The model extracts latent periodic information from the time-series data.

2) *Training phase*: During the training phase, preprocessed training and validation sets are input into the model separately. The model with the lowest validation set error is retained.

3) *Testing phase*: During the testing phase, the preprocessed test set is fed into the model, generating a reconstructed output.

4) *ISC detection*: Threshold values are determined using the reconstruction results and the actual values. These threshold values are obtained through peak over threshold (POT) theory [3]. This procedure facilitates the localization of faults, subsequently enabling the computation of precision, recall, and $F1$ scores.

A. Periodic segmentation Transformer

Transformer is a neural network model designed for sequence-to-sequence tasks, and it has achieved remarkable performance in both natural language processing and computer vision [22]. The Transformer model relies entirely on a self-attention mechanism for sequence modeling. This self-attention mechanism enables the model to dynamically calculate weights based on relationships between different positions within the input sequence, facilitating the capture of long-range information and the effective processing of lengthy sequences. Furthermore, the Transformer introduces positional encoding, which injects positional information into the input sequence feature vectors, allowing the model to differentiate vectors from different positions. In contrast to other sequence-to-sequence neural network models (like RNNs and LSTMs), the Transformer boasts parallel computation capabilities that enhance computational efficiency and reduce runtime [28]. In this paper, the periodic segmentation Transformer model is proposed. The specific structure (as shown in Fig. 2), the corresponding pseudocode (as illustrated in Table IV), and the detailed description are provided below:

TABLE IV
THE ALGORITHMIC PSEUDOCODE OF THE PROPOSED ISC DETECTION METHOD

Algorithm 1: Description of periodic segmentation Transformer model

Input: Raw battery voltage data $\mathbf{X}\in\mathbb{R}^{T\times M}$, *Topk* value K, Transformer encoder E, Iteration limit P
Output: Reconstructed result X^M
1: Analyze $\mathbf{X}$ in the frequency domain, average the amplitude across M dimensions: $\mathbf{A} = Avg\{Amp[DWT(\mathbf{X})]\}$
2: Select top K amplitude values and their corresponding frequencies: $\{f_1,\ldots,f_k\} = Topk(\mathbf{A})$
3: Calculate the sliding window lengths using the selected frequencies: $p\_i = T/f_i$, $i\in\{1,\ldots,K\}$
4: Partition $\mathbf{X}$ into M univariate time series: $\mathbf{X}_{(i)}\in\mathbb{R}^{1\times T}$, $i\in\{1,\ldots,M\}$
5: For each $\mathbf{X}_{(i)}$, create K distinct segmented time windows along the temporal dimension:
  for i from 1 to M **do**
    for k from 1 to K **do**
      Extract sliding windows of length p_k from $\mathbf{X}_{(i)}$: $\mathbf{X}_{seg\_k}^{(i)}\in\mathbb{R}^{p\_k\times N}$
6: Training the model:
  for k from 1 to K **do**
    $n = 0$
    do
      for i from 1 to M **do**
        $\dot{\mathbf{X}}_{(i)}^{k} = E(\mathbf{X}_{seg\_k}^{(i)})$
        $L = \frac{1}{M}\sum_{i=1}^{M}\left\|\dot{\mathbf{X}}_{(i)}^{k} - \mathbf{X}_{seg\_k}^{(i)}\right\|_2$
        Update weights of E using L
        $n = n+1$
      while $n < P$
      Reconstructed result of each K value: $\dot{\mathbf{X}}^k = \left(\dot{\mathbf{X}}_{(1)}^{k},\ldots,\dot{\mathbf{X}}_{(M)}^{k}\right)$
7: Calculate weights to each period:
  $\hat{A}_{f\_1},\ldots,\hat{A}_{f\_k} = Soft\max(A_{f\_1},\ldots,A_{f\_k})$
8: Output reconstructed result: $X^M = \sum_{k=1}^{K}\hat{A}_{f\_k}\times\dot{\mathbf{X}}^k$

Step 1: The raw battery voltage data $\mathbf{X}\in\mathbb{R}^{T\times M}$ is used as input of the periodicity analysis module to extract the periodic

information. The output is then utilized as prior knowledge in the time-oriented segmentation module.

Step 2: The raw battery voltage data $\mathbf{X} \in \mathbb{R}^{T \times M}$ is then split into $\mathbf{X}_{(i)} \in \mathbb{R}^{I \times T}$ (i = 1, …, M) that needs to be further cut according to the periodic characteristics. The split periods server as tokens of the Transformer encoder.

Step 3: When battery voltage data have more than one potential period, the reconstruction process will begin. The reconstructed results are compared with the target results in terms of residuals. The compared results are used to guide the weight updating.

Step 4: Repeat **Steps 2 ~ 3** to obtain the best reconstruction result.

Step 5: The multiple reconstructed results are aggregated into a single reconstruction result through the adaptive aggregation module.

For clarity, the specific description of different modules in the periodic segmentation Transformer are provided below.

1. Periodicity analysis module

The Fourier transform is a widely used signal processing method that can convert time-domain signals into frequency-domain signals [29]. Nevertheless, the Fourier transform is deficient analyzing non-stationary signals, as it does not take into account the amplitude and temporal localization of the signal, thus lacking temporal information during time-frequency conversions [30]. To accurately capture characteristic features of non-stationary signals, the discrete wavelet transform (DWT) in introduced in this work. The discrete wavelet transform is defined as follows:

$$DWT(j,h) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} \mathbf{X}(t) \psi\left(\frac{t - h2^j}{2^j}\right) dt \tag{4}$$

where ψ(t) represents the mother wavelet, while $\mathbf{X}$(t) represents the original signal. j and h represent scale parameters denoting dilation and translation, respectively. j captures the oscillation frequency and wavelength, and h reflects the translation distance of the wavelet. The mathematical expression of time series in the frequency domain via the periodicity analysis module is provided below:

$$\begin{aligned} \mathbf{A} &= Avg\{Amp[DWT(\mathbf{X})]\}, \\ \{f_1, \cdots, f_k\} &= Topk(\mathbf{A}), \\ p_i &= \left\lceil \frac{T}{f_i} \right\rceil, i \in \{1, \cdots, k\}. \end{aligned} \tag{5}$$

where $DWT(\cdot)$ and $Amp(\cdot)$ represent the DWT and the computation of frequency-domain amplitudes, respectively. $\mathbf{A} \in \mathbb{R}^T$ represents the amplitude of each frequency, and can be obtained by calculating the arithmetic mean of M feature parameters using $Avg(\cdot)$. $\{f_1, \ldots, f_k\}$ represent frequencies corresponding to the $Topk(\cdot)$ amplitude in $\mathbf{A}$. p_i denotes period of corresponding frequencies. The approximate coefficients of the wavelet transform depict the low-frequency component of the signal, while the detail coefficients portray the high-frequency portion. The absolute value of the approximate coefficients reflects the amplitude of the low-frequency components of the signal, encompassing both the overall trend and information about longer periods.

Through the periodicity analysis module, the periodic information of the time series is extracted, along with the corresponding frequencies and amplitudes. The entire module can be rewritten by:

$$\mathbf{A}, \{f_1, \cdots, f_k\}, \{p_1, \cdots, p_k\} = Period(\mathbf{X}) \tag{6}$$

In prior works, the majority of research efforts have predominantly focused either on temporal characteristics between adjacent timestamps [23-25] or on temporal features across various dimensions [31]. Given the pervasive existence of periodicity in time-series data, the extraction of periodic features assumes particular significance. Even for time series with less evident periodic features, this module can still provide latent periodicities for the network. For extracting periodic information, this module also provides the basis for time-oriented segmentation module.

2. Time-oriented segmentation module

After data preprocessing, the raw data $\mathbf{X} \in \mathbb{R}^{T \times M}$ is partitioned. This partitioning transforms the multivariate time series into M individual univariate time series. The i-th univariate time series can be represented as $\mathbf{X}^{(i)} \in \mathbb{R}^{I \times T}$. The resulting univariate time series, are individually input into separate instances of the model, as shown in Fig. 2. These instances share weights within a common backbone module, but their forward computations are independent of each other. For each individual univariate time series $\mathbf{X}_{1:T}^{(i)} = (\mathbf{X}_1^{(i)}, \ldots, \mathbf{X}_T^{(i)})$, sliding windows are extracted along the temporal dimension. These sliding windows can be either overlapping or non-overlapping. The length of each sliding window is defined as the length of the top K period p_k. The non-overlapping length between two consecutive windows is defined as S. The number of time windows is denoted as N. The univariate time series can be represented by $\mathbf{X}_{seg}^{(i)} \in \mathbb{R}^{p\_k \times N}$ after segmentation along the temporal dimension, where $N = (T - p_k) / S + 2$ denoting the number of tokens fed into the network. Conventional practice makes each timestamp input individually as a token, which means the token count in previous approaches is approximately S times greater than that used in this study.

This approach enables an adaptive adjustment of sliding window lengths. This effectively addresses the fluctuation in performance caused by manually setting time window parameters in prior works. Also, by significantly reducing the number of input tokens, both the computational complexity and processing time experience substantial reduction. This allows the model to capture extended historical sequences. Moreover, with the sliding window length aligned to the latent periodicity of the time series, the model becomes adapt at capturing the underlying relationships within and between periods. This proficiency results in enhanced fitting capabilities and improved performance in ISC detection. Notably, the combination of periodicity analysis module and time-oriented segmentation module makes the joint extraction of temporal-spatial and periodic information available.

3. Transformer encoder

In the periodic segmentation Transformer architecture, a foundational Transformer encoder is employed to map the

original input signals to a latent representation space. After segmenting the input along the temporal dimension, the matrix $\mathbf{X}_{seg}^{(i)} \in \mathbb{R}^{p\_k \times N}$ is endowed with positional encoding to preserve the temporal information of the input tokens. Through a trainable linear projection $\mathbf{W}_{seg}^{(i)} \in \mathbb{R}^{D \times p\_k}$, a learnable positional encoding $\mathbf{W}_{pos}^{(i)} \in \mathbb{R}^{D \times N}$, $\mathbf{X}_d^{(i)} = \mathbf{W}_{seg}\mathbf{X}_{seg}^{(i)} + \mathbf{W}_{pos}$ maps the input data to the Transformer latent space of dimension D. The resulting positional encoding $\mathbf{X}_d^{(i)} \in \mathbb{R}^{D \times N}$ will serve as input to the fundamental Transformer encoder. The Transformer encoder introduces multi-head attention, where each head h = 1, …, H transforms the input into three matrices: Q (query), K (key), and V (value). Their computation are as follows:

$$\begin{aligned} Q_h^{(i)} &= (\mathbf{X}_d^{(i)})^T \mathbf{W}_h^Q \\ K_h^{(i)} &= (\mathbf{X}_d^{i})^T \mathbf{W}_h^K \\ V_h^{(i)} &= (\mathbf{X}_d^{(i)})^T \mathbf{W}_h^V \end{aligned} \tag{7}$$

where $\mathbf{W}_h^Q$, $\mathbf{W}_h^k \in \mathbb{R}^{D \times d\_k}$, $\mathbf{W}_h^V \in \mathbb{R}^{d\_k \times D}$. The scaled dot-product operation is employed to obtain the attention outputs.

$$\begin{aligned} (O_h^{(i)})^T &= Attention(Q_h^{(i)}, K_h^{(i)}, V_h^{(i)}) \\ &= Soft\max\left(\frac{Q_h^{(i)}(K_h^{(i)})^T}{\sqrt{d_k}}\right)V_h^{(i)} \end{aligned} \tag{8}$$

where $\mathrm{O}_h^{(i)} \in \mathbb{R}^{D \times N}$, *Soft*max function allocates a learnable convex combination weight to each element in matrix $\mathrm{V}_h^{(i)}$, compressing matrix $\mathrm{V}_h^{(i)}$ into a more compact representative embedding. This simplifies inference during downstream neural network operations. Unlike conventional attention operation, the scaled dot-product operation scales weights by the square root of d_k, reducing their variance and promoting stable training.

The multi-head attention module encompasses a batch normalization layer and a feedforward network with residual connections. Finally, a flattening layer with a linear output layer is applied to obtain the reconstruction result $\dot{\mathrm{X}}^{(i)} = \dot{\mathrm{X}}_1^{(i)}, \ldots, \dot{\mathrm{X}}_{\mathrm{T}}^{(i)}$.

Then, the loss function that can calculate the error between the reconstruction result and the ground truth is mathematically expressed by:

$$\mathcal{L} = \frac{1}{M}\sum_{i=1}^{M}\left\| \dot{\mathbf{X}}_{1:T}^{(i)} - \mathbf{X}_{1:T}^{(i)} \right\|_2 \tag{9}$$

4. Adaptive aggregation

Within the periodicity analysis module, the amplitude values of frequencies are sorted in descending order. The top K amplitude values, denoted as $\{\mathrm{A}_{f\_1}, \ldots, \mathrm{A}_{f\_k}\}$, correspond to the top K most significant frequencies $\{f_1, \ldots, f_k\}$ and capture the latent periodic features of the time series. Due to the presence of K distinct segmented time windows for different univariate time series, the model produces K distinct reconstruction outcomes for each univariate time series. Consequently, these diverse reconstruction outcomes require aggregation into a final unified reconstruction result.

Given that each amplitude value A reflects the importance of the corresponding frequency and period, it also determines the significance of the associated reconstruction outcome. Hence, an amplitude-based adaptive aggregation method is proposed:

$$\begin{aligned} \hat{A}_{f\_1}, \ldots, \hat{A}_{f\_k} &= Soft\max(A_{f\_1}, \ldots, A_{f\_k}) \\ X^M &= \sum_{i=1}^{k} \hat{A}_{f_i} \times \dot{X}^{M,i} \end{aligned} \tag{10}$$

where $\hat{A}_{f\_k}$ denotes the weights obtained from the softmax equation. These weights are multiplied with the corresponding reconstructed results $\dot{X}^{M,i}$ from different periods within a specific dimension. The summation of these weighted results yields the final reconstruction output.

This approach allocates weights to the corresponding reconstruction outcomes based on the magnitude of their amplitude values. Based on the experiment conducted in Section IV, the impact of variations in the added hyperparameter K on experimental results can be safely disregarded. This technique not only aggregates different latent periodic features but also mitigates the impact of less pronounced latent periodic features, thus preventing substantial deviations between the final reconstruction result and the ground truth.

B. Fault detection and evaluation metrics

During the testing phase, the test set $\dot{X}$ is the input of the trained model, $\dot{X}'$ is the output. The criterion can be written by:

$$s = \left\| X' - \dot{X}' \right\|_2 \tag{11}$$

As the model segments the test set into M individual univariate time series, s_i represents the criterion of each dimension. For a fair comparison, same approach is utilized by employing the POT method for threshold selection. Notably, the POT method is a statistical approach that employs extreme value theory to fit data distributions using the generalized pareto distribution (GPD) [32]. It dynamically determines thresholds by identifying appropriate values at risk. When the criterion surpasses the threshold, the corresponding timestamp is labeled as a faulty state (label 1), otherwise, the timestamps are assigned the label 0.

Precision, recall, and $F1$ score are employed to assess the ISC detection performance of all models as:

$$\begin{aligned} precision &= \frac{TP}{TP + FP} \\ recall &= \frac{TP}{TP + FN} \\ F1 &= 2 \cdot \frac{precision \cdot recall}{precision + recall} \end{aligned} \tag{12}$$

where TP denotes true positives, FP stands for false positives, and FN represents false negatives. Notably, precision, recall, and $F1$ are all used as the evaluation indexes for model performance, higher indexes indicate better performance on ISC detection.

IV. Experiment

This section presents a comprehensive evaluation of the proposed model through a series of experiments. Specifically, the experimental environment and parameter settings are described, followed by a detailed analysis of the detection results. Then, an ablation study is conducted to highlight the importance of key modules in the proposed method.

TABLE VI
RESULTS OF EACH MODEL ON THE FUDS, UDDS, US06, SMD DATASETS

Evaluation indicators		FUDS			UDDS			US06			SMD
		Low	Medium	High	Low	Medium	High	Low	Medium	High	
Ours	Prec.	0.7712	0.8994	0.8514	0.6201	0.9027	0.8967	0.4953	0.9115	0.8540	0.8984
	Rec.	0.7089	0.7506	0.8077	0.6110	0.7971	1.0	0.7874	0.7649	1.0	0.8799
	F1	**0.7387**	**0.8183**	**0.8292**	**0.6155**	**0.8466**	**0.9455**	**0.6081**	**0.8318**	**0.9212**	**0.8890**
PatchTST [26]	Prec.	0.6989	0.7245	0.7114	0.6873	0.8978	0.8934	0.4934	0.8153	0.8319	0.8752
	Rec.	0.7089	0.7506	0.7871	0.4073	0.7971	1.0	0.7874	0.7649	1.0	0.8092
	F1	0.7039	0.7373	0.7473	0.5115	0.8445	0.9437	0.6067	0.7893	0.9082	0.8409
TimesNet [25]	Prec.	0.7664	0.7924	0.8328	0.3808	0.8456	0.8804	0.4647	0.8489	0.8502	0.8790
	Rec.	0.7089	0.7506	0.7871	0.4073	0.7971	1.0	0.7874	0.7649	1.0	0.8145
	F1	0.7365	0.7709	0.8093	0.3936	0.8206	0.9364	0.5844	0.8047	0.9190	0.8455
Non-Stationary [24]	Prec.	0.5616	0.5511	0.5520	0.3208	0.8267	0.8617	0.4647	0.7482	0.7814	0.8848
	Rec.	0.7089	0.7506	0.7871	0.4073	0.7971	1.0	0.7874	0.7649	1.0	0.8063
	F1	0.6268	0.6355	0.6489	0.3589	0.8117	0.9257	0.5844	0.7654	0.8773	0.8437
LSTM Hybrid [21]	Prec.	0.6406	0.6849	0.7262	0.4280	0.8186	0.8741	0.4431	0.5874	0.5455	0.8875
	Rec.	0.7089	0.7506	0.7871	0.4257	0.7971	1.0	0.4551	0.5332	0.6839	0.8105
	F1	0.6731	0.7162	0.7554	0.4269	0.8077	0.9328	0.4490	0.5584	0.6069	0.8473
DNN [20]	Prec.	0.4795	0.5065	0.5169	0.4265	0.6704	0.8073	0.4638	0.5701	0.5039	0.8141
	Rec.	0.7089	0.7506	0.7871	0.4257	0.5680	1.0	0.7874	0.7649	1.0	0.7183
	F1	0.5721	0.6048	0.6240	0.4261	0.6150	0.8934	0.5838	0.6533	0.6701	0.7632
Bi-LSTM [19]	Prec.	0.3306	0.4106	0.7135	0.3077	0.7604	0.8294	0.3304	0.3427	0.3308	0.8672
	Rec.	0.7089	0.5396	0.3935	0.4257	0.5680	1.0	0.4551	0.5050	0.6839	0.8030
	F1	0.4510	0.4663	0.5073	0.3572	0.6503	0.9067	0.3829	0.4083	0.4459	0.8339
LSVSF [18]	Prec.	0.4564	0.4976	0.4909	0.3171	0.6648	0.8408	0.4431	0.5787	0.5376	0.8745
	Rec.	0.7089	0.7506	0.7871	0.6110	0.5680	1.0	0.4551	0.7649	1.0	0.7749
	F1	0.5553	0.5985	0.6047	0.4176	0.6126	0.9135	0.4490	0.6588	0.6993	0.8217

Furthermore, the parameter sensitivity analysis, robustness analysis, and effectiveness analysis are conducted, respectively.

A. *Experimental environment and parameter settings*

The series of experiments carried out by the proposed model are performed on a desktop computer (I7-10700K, RTX 3080, Windows 10 OS, Python 3.8, PyTorch 1.12). Inspired by the parameter setting in [23-27], the specific parameter settings are shown in Table V. In order to balance training time and model performance, a sliding window length of 100 and 10 epochs are set, consistent with [23-27]. In this paper, the *K* value has minimal impact on the detection results and is set to 3. To accommodate computer resources, a batch size of 128 is chosen. Theoretically, variations in batch size have little effect on model performance. Based on the default parameters in traditional Transformer model, d_model, d_feedforword, and n_layers are set to 512, 1024, and 6, respectively.

TABLE V
PARAMETER SETTING

parameter	Value	
	Our model	Other models
sliding window length	adaptive	100 (fixed)
K	3	-
batch size	128	128
epoch	10	10
d_model	512	512
d_feedforward	1024	1024
n_layers	6	6

B. *Detection result*

To verify the generalization capability of ISC detection model, all models were tested on a simulated ISC dataset encompassing fault severities under different operating conditions (FUDS, UDDS, and US06), as well as a publicly available anomaly detection dataset, i.e., the server machine dataset (SMD) [33]. SMD is a five-week long multivariate dataset collected from a large Internet company. The proposed model is compared with LSVSF [18], Bi-LSTM [19], DNN [20], LSTM hybrid network [21], Nonstationary Transformer [24], TimesNet [25], and PatchTST [26]. Notably, the selection criterion is that all these competitors are SOTA methods published in TOP journals or conferences within three years.

To simulate real-world scenarios, all models were trained using the same parameter set and subsequently employed for ISC detection across varying severity levels. The test results are depicted in Fig. 3 and Table VI.

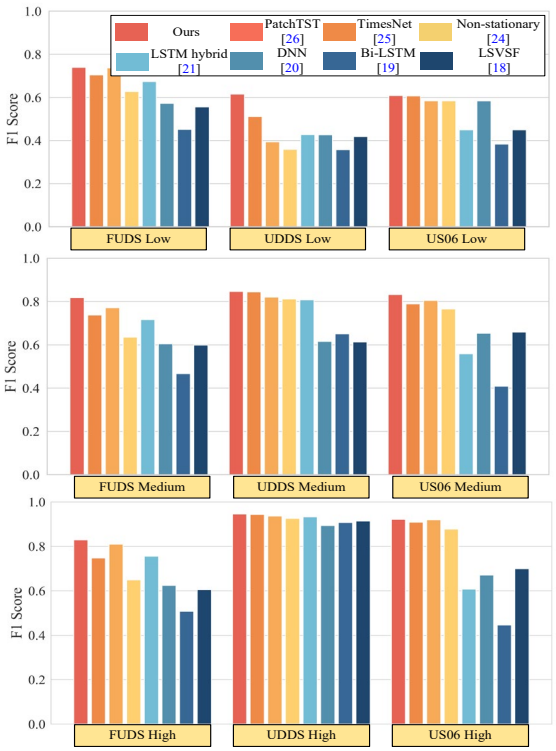

Fig. 3. F1 score of each model on three different operating conditions.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/JBHI.2024.3392648, IEEE Transactions on Transportation Electrification

Specifically, experimental results indicate that the proposed model achieves the best performance across all datasets, with its F1 score surpassing other models by approximately 24.2%. When detecting high severity ISC, all models exhibited high recall and precision, suggesting their capacity for terminal ISC diagnosis solely based on battery pack voltage data. However, as impedance increases and severity diminishes, the impact on the battery lessens, leading to a corresponding decrease in model detection performance. For medium and low severity, other methods witness varying degrees of decline, around 10.5% and 29.8% lower compared to high severity, respectively. Notably, the proposed method demonstrates relatively stable performance despite variations in ISC severity, with a decline of around 7.1% and 26.6%, outperforming the other competitors. For different ISC severities, the lower the severity, the greater the improvement in F1 scores. For high, medium and low severity, the proposed model's F1 scores are enhanced by 20.8%, 26.2%, and 31.3%, respectively.

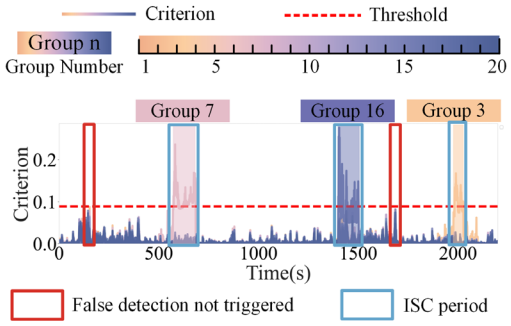

Fig. 4. Detection results under UDDS operational condition using our method.

Fig. 4 shows the detection results under UDDS operational condition using our method. Notably, the 'Criterion' line (colored solid line) shows the detection output of different battery groups, the 'Threshold' line (red dash line) indicates a boundary above which the detection is considered positive, the color shaded areas represent the detected ISC, different colors

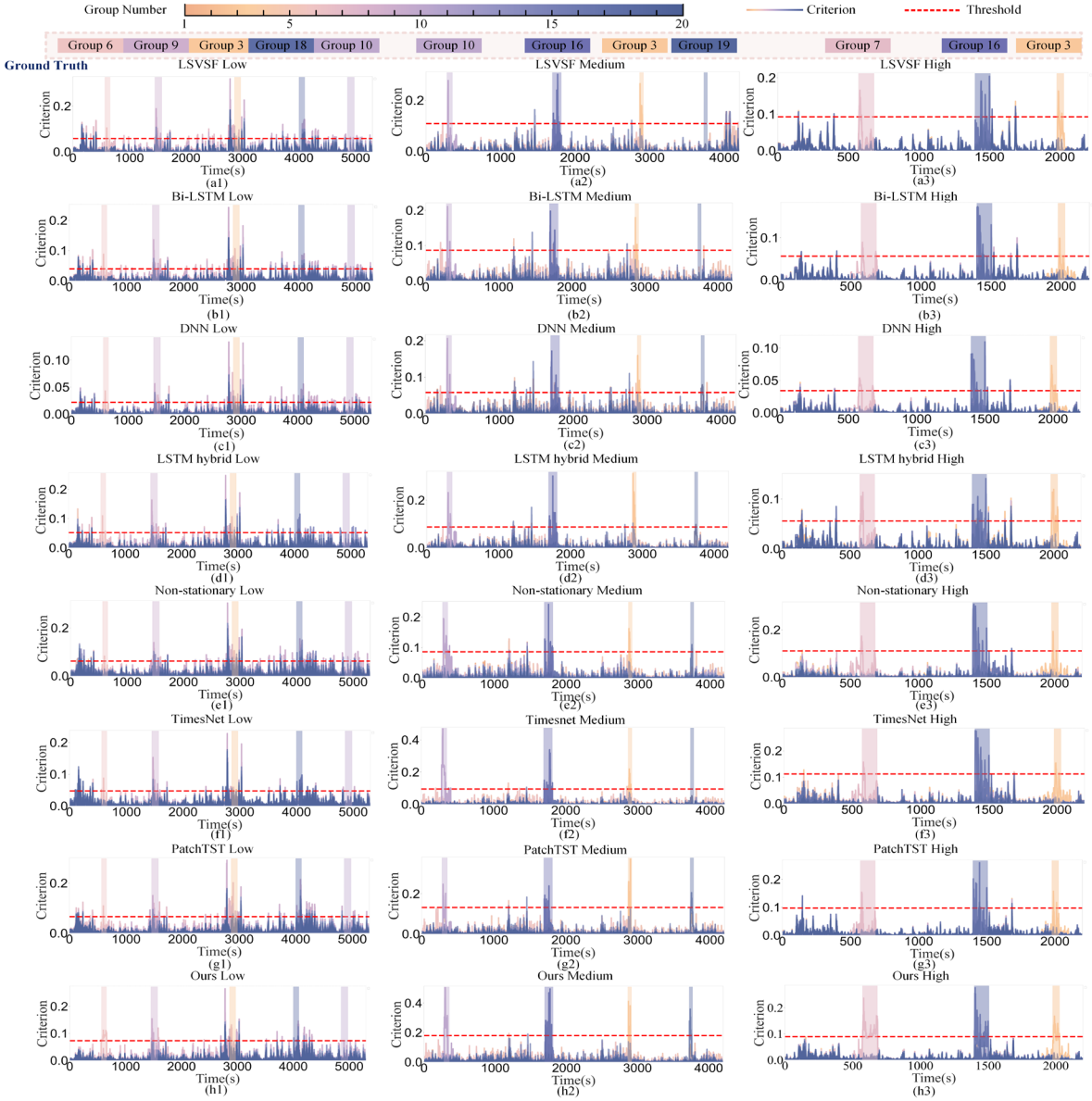

Fig. 5. Detection results under UDDS operational condition for various models: (a1) ~ (a3) LSVSF, (b1) ~ (b3) Bi-LSTM, (c1) ~ (c3) DNN, (d1) ~ (d3) LSTM Hybrid, (e1) ~ (e3) Non-stationary Transformer, (f1) ~ (f3) TimesNet, (g1) ~ (g3) PatchTST, (h1) ~ (h3) Our method.

represent different battery groups. Under normal conditions, the criterion is predominantly below the threshold, while in the case of a ISC fault, criterion lies above the threshold.

Fig. 5 shows the detection results under UDDS operational condition for various models. The visualized results demonstrate that our model has fewer instances of false positives and captures more true positives. The most consistent alignment with the ISC periods of our model indicates higher detection accuracy for all severity levels. Namely, our model exhibits the best performance, and the main reasons may be that our method significantly improves the extraction of features from time-series data, the temporal-spatial and periodic information can be captured simultaneously. Similarly, TimesNet is the closest competitor. It utilizes 2D convolution to extract spatial and periodic information from time-series data. But its feature extraction ability is inferior to Transformer-based models, resulting in notable deviations between reconstruction results and ground truth. Among these competitors, Bi-LSTM and LSVSF show worse performance, the main reasons can be concluded that Bi-LSTM network has difficulty in preserving sufficient long-term memory for distant historical data, which may lead to inadequate feature extraction, particularly in capturing periodic information. LSVSF extracts temporal-spatial information with non-adaptive time window, which captures periodic information insufficiently, resulting in poor performance. Overall, our approach is more sensitive to detecting ISC fault for all severity levels compared to other models. These results underscore the strong generalization ability and robustness of the proposed model, making it suitable for real-world engineering applications.

C. Ablation study

To assess the significance of each module, a series of ablation experiments are conducted by progressively excluding key modules and observing their impact on the F1 score.

Specifically, the periodicity analysis module is excluded, setting the sliding time window for segmenting time series to 100 (following prior work). Then, the time-oriented segmentation module is excluded, alternatively using the latent period computed within the periodicity analysis module as the sliding window length. The time-series data is input into the model, following the previous methodology. The experimental results collected in Table VII can be summarized below:

- Removing the periodic feature extraction module will lead to a 15.4% decrease of average F1 score. This is attributed to the fixed time window length of 100, which fails to capture latent periodic patterns, thereby hindering the model's ability to learn inter-period features. Based on this, extracting feature information between time-series periods is proved crucial.
- Omitting the time dimension segmentation module results in a 10.7% decrease of average F1 score. After removing this module, the model receives singular timestamps as tokens instead of the required time-series segments. This restricts the model to learn features within the periods, neglecting inter-period features. The time dimension segmentation module enables long time-series input lengths, thus facilitating more comprehensive feature extraction from the time-series data.

D. Sensitivity to K value

The choice of *K* can affect model performance. Specifically, when the *K* value is higher, more frequencies are included and

TABLE VII
RESULTS OF ABLATION STUDY

Method	Adaptive sliding window	FUDS			UDDS			US06			SMD
		Low	Medium	High	Low	Medium	High	Low	Medium	High	
Periodic segmentation	Yes	**0.7387**	**0.8183**	**0.8292**	**0.6155**	**0.8466**	**0.9455**	**0.6081**	**0.8318**	**0.9212**	**0.8890**
w/o time-oriented segmentation	No	0.7009	0.7364	0.7485	0.4563	0.7657	0.9382	0.4588	0.6038	0.7437	0.8642
w/o periodicity analysis	No	0.7029	0.7103	0.7237	0.4327	0.7816	0.9380	0.4026	0.5913	0.7564	0.8463

TABLE VIII
THE EXPERIMENTAL RESULTS WITH DIFFERENT *K* VALUES

Evaluation indicators		*K* Value									
		1		2		3		4		5	
		F1	Times (s)	F1	Times (s)	F1	Times (s)	F1	Times (s)	F1	Times (s)
FUDS	Low	0.7354	21.84	0.7343	34.18	0.7365	48.52	0.7387	57.95	0.7100	70.32
	Medium	0.7934	21.80	0.8046	34.03	0.8183	48.25	0.8147	57.62	0.8169	69.87
	High	0.8133	21.73	0.7462	34.37	0.8292	48.79	0.7450	57.77	0.7709	70.33
UDDS	Low	0.6155	21.60	0.5828	34.01	0.6150	48.46	0.5879	57.68	0.5893	69.98
	Medium	0.8257	21.62	0.8369	34.02	0.8466	48.48	0.8457	57.60	0.8493	70.18
	High	0.9187	21.56	0.9419	34.22	0.9455	48.34	0.9382	57.96	0.9437	69.71
US06	Low	0.5590	21.83	0.5362	34.25	0.6081	48.30	0.5819	58.28	0.5356	69.90
	Medium	0.8106	21.67	0.8215	33.79	0.8318	48.58	0.8314	57.79	0.8367	69.93
	High	0.9147	21.60	0.9212	34.28	0.9190	48.43	0.9169	58.11	0.9061	70.32
SMD		0.8589	35.16	0.8890	63.20	0.8696	91.24	0.8629	125.26	0.8655	160.35

Note: F1 and Time denote the F1 score and the total time consumption (including training time and testing time), respectively.

more comprehensive periodic information can be obtained. However, including too many frequencies might also introduce noise and irrelevant information, potentially leading to overfitting. To explore the sensitivity of model performance to the value of K, experiments with different values of K ranging from 1 to 5 for each dataset are conducted, the corresponding results are shown in Fig. 6 and Table VIII.

The experimental results indicate that the K value has limited impact on the performance of ISC detection. It is noted that the model performs optimally when the K value is set to 3. At this setting, the model not only achieves better performance but also exerts a minimal impact on time consumption. Consequently, the K value is set as 3 to achieve the tradeoff between efficiency and accuracy.

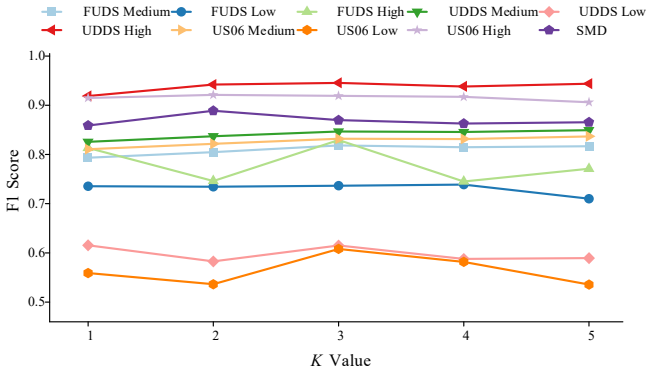

Fig. 6. F1 score with different K values.

E. Data length robustness test

In real-world scenarios, it is challenging to obtain complete voltage curves to meet the requirement of high detection efficiency. Test data of varying lengths exhibit different maximum and minimum voltage values, and the shape of the curve can change after normalization operation. 40% of the training data is used and this incomplete discharge curve is employed to test the model. The experimental results are summarized in Table IX. From Table IX, the performance of the model varies in response to different severity levels of ISC. When the severity level is low, the average F1 score decreases by approximately 21.0%; For a medium-level severity of ISC, the average F1 score decreases by around 6.7%; For a high-level severity of ISC, the average F1 score decreases by about 5.8%. It can be observed that as the severity of ISC increases, the robustness becomes stronger. Namely, even in scenarios with limited data, the detection performance can maintain a high level. This indicates that the proposed method is capable of effectively conducting ISC detection under conditions of incomplete data, displaying good applicability and resilience.

TABLE IX
TEST RESULTS USING 40% OF THE DATA

Evaluation indicators		F1score	
		Whole dataset	40% dataset
FUDS	Low	0.7387	0.6572
	Medium	0.8183	0.7738
	High	0.8066	0.7462
UDDS	Low	0.6150	0.4563
	Medium	0.8466	0.7643
	High	0.9455	0.9437
US06	Low	0.6081	0.4490
	Medium	0.8318	0.7903
	High	0.9212	0.8301
SMD		0.8890	0.8566

F. Effectiveness of feature extraction

In this subsection, we mainly focus on the effectiveness analysis of the extracting periodic features from time-series data by performing time dimension segmentation.

To validate the effectiveness of the feature extraction, the proposed periodic feature extraction method is integrated into two existing models: PatchTST and Non-stationary Transformer. The results are depicted in Table X. The proposed feature extraction method enhances the performance on both two datasets. Specifically, it can improve the performance of PatchTST (F1 score increased by 4.7%) and Non-stationary Transformer (F1 score increased by 10.2%). Meanwhile, for high-, medium- and low-severity levels of ISC, F1 scores are improved by 5.3%, 6.3%, and 12.9%, respectively. This demonstrates that the proposed feature extraction method effectively enhances performance of the proposed model, enabling better identification of anomalies caused by subtle ISC issues.

V. CONCLUSION

This paper proposes a periodic segmentation Transformer to perform ISC detection. Specifically, a comprehensive dataset encompassing three ISC severity levels under varying operating conditions is collected. Then, a dual-module design

TABLE X
RESULTS OF FEATURE EXTRACTION EFFECTIVENESS

Dataset	FUDS			UDDS			US06			SMD
model	High	Medium	Low	High	Medium	Low	High	Medium	Low	
PatchTST [26]	0.7473	0.7373	0.7039	0.9437	0.8445	0.5115	0.9082	0.7893	0.6067	0.8409
+ours	**0.8133**	**0.7934**	**0.7354**	**0.9455**	**0.8457**	**0.6150**	**0.9147**	**0.8106**	**0.6081**	**0.8589**
promotion	8.12%	7.61%	4.28%	0.19%	0.14%	16.90%	0.71%	2.70%	0.23%	2.10%
Non-stationary [24]	0.6489	0.6355	0.6268	0.9257	0.8117	0.3589	0.8773	0.7654	0.5844	0.8437
+ours	**0.7405**	**0.7373**	**0.7089**	**0.9437**	**0.8445**	**0.4944**	**0.9279**	**0.8229**	**0.5955**	**0.8458**
promotion	12.37%	16.02%	11.58%	1.91%	4.04%	27.41%	5.45%	7.51%	1.86%	0.25%
Our model	0.8292	0.8183	0.7387	0.9455	0.8466	0.6155	0.9212	0.8318	0.6081	0.8890

comprising a periodicity analysis module and a time-oriented segmentation module enables the adaptive adjustment of the length of sliding windows. Meanwhile, the temporal-spatial and periodic information can be extracted simultaneously via Transformer encoder. Furthermore, a series of experiments and analysis (i.e., parameter sensitivity analysis, robustness analysis, and effectiveness analysis) are conducted. The experimental results demonstrate that the proposed model consistently achieved the highest F1 scores, exhibiting an average improvement of 24.2% over various SOTA methods. Considering the safety implications of ISC in batteries, our method shows promise for accurate ISC detection. With the rapid development in sensing technology and artificial intelligence, how to achieve accurate and fast ISC detection under connected vehicle environment is our future work.

REFERENCES

[1] Z. Wei, K. Liu, X. Liu, Y. Li, L. Du, and F. Gao, "Multilevel data-driven battery management: From internal sensing to big data utilization," *IEEE Trans. Transport. Electrific.*, vol. 9, no. 4, pp. 4805–4823, Dec. 2023.

[2] Z. Wei, J. Hu, H. He, Y. Li, and B. Xiong, "Load current and state-of-charge coestimation for current sensor-free lithium-ion battery," *IEEE Trans. Power Electron.*, vol. 36, no. 10, pp. 10970–10975, Oct. 2021.

[3] G. Zhang, X. Wei, X. Tang, J. Zhu, S. Chen, and H. Dai, "Internal short circuit mechanisms, experimental approaches and detection methods of lithium-ion batteries for electric vehicles: A review," *Renew. Sustain. Energy Rev.*, vol. 141, May 2021, Art. no. 110790.

[4] X. Lai et al., "Mechanism modeling detection and prevention of the internal short circuit in lithium-ion batteries: Recent advances and perspectives", *Energy Storage Mater.*, vol. 35, pp. 470-499, Mar. 2021.

[5] R. Song, X. Liu, Z. Wei, F. Pan, Y. Wang and H. He, "Safety and longevity-enhanced energy management of fuel cell hybrid electric vehicle with machine learning approach," *IEEE Trans. Transport. Electrific.*, vol. 10, no. 2, pp. 2562-2571, Jul. 2023.

[6] A. Moeini and S. Wang, "Fast and precise detection of internal short circuit on li-ion battery," *Proc. IEEE Energy Convers. Congr. Expo.*, Sept. 2018, pp. 2759-2766.

[7] R. Ma, J. He, and Y. Deng, "Investigation and comparison of the electrochemical impedance spectroscopy and internal resistance indicators for early-stage internal short circuit detection through battery aging," *J. Energy Storage*, vol. 54, Oct. 2022.

[8] X. Lai et al., "Online detection of early-stage internal short circuits in series-connected lithium-ion battery packs based on state-of-charge correlation," *J. Energy Storage*, vol. 30, Aug. 2020.

[9] J. Meng, M. Boukhnifer, C. Delpha and D. Diallo, "Incipient short-circuit fault diagnosis of lithium-ion batteries," *J. Energy Storage*, vol. 31, Oct. 2020.

[10] X. Feng, X. He, L. Lu, and M. Ouyang, "Analysis on the fault features for internal short circuit detection using an electrochemical-thermal coupled model," *J. Electrochem. Soc.*, vol. 165, no. 2, pp. A155–A167, Jan. 2018.

[11] J. Hu, H. He, Z. Wei, and Y. Li, "Disturbance-immune and aging robust internal short circuit diagnostic for lithium-ion battery," *IEEE Trans. on Ind. Electron.*, vol. 69, no. 2, pp. 1988–1999, Feb. 2022.

[12] X. Kong, G. L. Plett, M. Scott Trimboli, Z. Zhang, D. Qiao, T. Zhao, and Y. Zheng, "Pseudo-two-dimensional model and impedance diagnosis of micro internal short circuit in lithium-ion cells," *J. Energy Storage*, vol. 27, Feb. 2020.

[13] Y. Kang, X. Yang, Z. Zhou, B. Duan, Q. Liu, Y. Shang, and C. Zhang, "A comparative study of fault diagnostic methods for lithium-ion batteries based on a standardized fault feature comparison method," *J. Cleaner Product.*, vol. 278, Jan. 2021.

[14] A. Naha, A. Khandelwal, S. Agarwal, P. Tagade, K. S. Hariharan, A. Kaushik, A. Yadu, S. M. Kolake, S. Han, and B. Oh, "Internal short circuit detection in Li-ion batteries using supervised machine learning," *Sci. Rep.*, vol. 10, no. 1, Jan. 2020.

[15] A. Kriston, A. Podias, I. Adanouj, and A. Pfrang, "Analysis of the effect of thermal runaway initiation conditions on the severity of thermal runaway-numerical simulation and machine learning study," *J. Electrochemical Soc.*, vol. 167, no. 9, Jun. 2020.

[16] A. Chen, W. Zhang, C. Zhang, Z. Wang, and X. Fan, "A novel AlCu internal short circuit detection method for lithium-ion batteries based on on-board signal processing," *J. Energy Storage*, vol. 52, p. 104748, Aug. 2022.

[17] M. Schmid, J. Kleiner, and C. Endisch, "Early detection of internal short circuits in series-connected battery packs based on nonlinear process monitoring," *J. Energy Storage*, vol. 48, Apr. 2022.

[18] B. Cui, H. Wang, R. Li, L. Xiang, J. Du, H. Zhao, S. Li, X. Zhao, G. Yin, X. Cheng, Y. Ma, H. Huo, P. Zuo, G. Han, and C. Du, "Long-sequence voltage series forecasting for internal short circuit early detection of lithium-ion batteries," *Patterns*, vol. 4, no. 6, p. 100732, Apr. 2023.

[19] R. Cao, Z. Zhang, J. Lin, J. Lu, L. Zhang, L. Xiao, X. Liu, and S. Yang, "Reliable online internal short circuit diagnosis on lithium-ion battery packs via voltage anomaly detection based on the mean-difference model and the adaptive prediction algorithm," *Batteries*, vol. 8, no. 11, p. 224, Nov. 2022.

[20] B. Cui, H. Wang, R. Li, L. Xiang, J. Du, H. Zhao, S. Li, X. Zhao, G. Yin, X. Cheng, Y. Ma, H. Huo, P. Zuo, and C. Du, "Internal short circuit early detection of lithium-ion batteries from impedance spectroscopy using deep learning," *J. Power Sources*, vol. 563, Apr. 2023.

[21] H. Wang, J. Nie, Z. He, M. Gao, W. Song, and Z. Dong, "A reconstruction-based model with transformer and long short-term memory for internal short circuit detection in battery packs," *Energy Rep.*, vol. 9, pp. 2420–2430, Dec. 2023.

[22] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," *Proc. Adv. Neural Inf. Process. Syst.*, Jun. 2017, pp. 5999–6009.

[23] S. Tuli, G. Casale, and N. R. Jennings, "TranAD: Deep transformer networks for anomaly detection in multivariate time series data," *Proc. VLDB Endow.*, vol. 15, no. 6, pp. 1201–1214, Feb. 2022.

[24] Y. Liu, H. Wu, J. Wang, and M. Long, "Non-stationary transformers: Exploring the stationarity in time series forecasting," *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2022, pp. 1–13.

[25] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "TimesNet: Temporal 2d-variation modeling for general time series analysis," *Proc. Int. Conf. Learn. Representations*, Dec. 2023, pp. 1–12.

[26] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," *Proc. Int. Conf. Learn. Representations*, Dec. 2023, pp. 1–12.

[27] R. Guo, L. Lu, M. Ouyang, and X. Feng, "Mechanism of the entire overdischarge process and overdischarge-induced internal short circuit in lithium-ion batteries," *Sci. Rep.*, vol. 6, no. 1, Jul. 2016.

[28] X. Jia, H. Wu, R. Zhang, and M. Peng, "Csformer: Enhancing deep learning efficiency for intelligent iot," *Comput. Commun.*, vol. 214, pp. 33–45, Jan. 2024.

[29] E. Rajaby and S. M. Sayedi, "A structured review of sparse fast Fourier transform algorithms," *Digit. Signal Process.*, vol. 123, p. 103403, Apr. 2022.

[30] T. Guo, T. Zhang, E. Lim, M. López-Benítez, F. Ma, and L. Yu, "A Review of Wavelet Analysis and Its Applications: Challenges and Opportunities," *IEEE Access*, vol. 10, pp. 58869-58903, Jun. 2022.

[31] C. Ding, S. Sun, and J. Zhao, "MST-GAT: A multimodal spatial–temporal graph attention network for time series anomaly detection," *Inf. Fusion*, vol. 89, pp. 527–536, Jan. 2023.

[32] A. Siffer, P.-A. Fouque, A. Termier, and C. Largouet, "Anomaly detection in streams with extreme value theory," *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 1067–1075.

[33] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust anomaly detection for multivariate time series through stochastic recurrent neural network," *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 2828–2837.