Preprints.org

Article

Not peer-reviewed version

An Online Fault Cell Screening Method for Lithium-Ion Battery Formation Based on the Data-Driven Model with Incomplete Time-Series Data

Jianjun He , Aibin Deng , Xiang Wang * , Rihui Long , Fuxin Huang

Posted Date: 26 March 2026

doi: 10.20944/preprints202603.2033.v1

Keywords: lithium-ion battery; battery formation; time-series classification; data-driven model

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2026 doi:10.20944/preprints202603.2033.v1

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

Article

An Online Fault Cell Screening Method for Lithium-Ion Battery Formation Based on the Data-Driven Model with Incomplete Time-Series Data

Jianjun He, Aibin Deng, Xiang Wang *, Rihui Long and Fuxin Huang

The School of Automation, Central South University, Changsha, China
* Correspondence: awoshi37@163.com; Tel.: +86-155-7670-4717

Abstract

Battery formation is important for ensuring the quality and service life of cells in lithium-ion battery (LIB) production. During the formation process, fault cells, such as low open-circuit voltage cells, are offline screened after the charging stage since, in most formation protocols, the online screening process is absent. This can lead to energy waste and extend the rework cycle of the fault cells in the LIB formation process. To address this problem, this paper considers the online fault cell screening problem, the formation pre-screening, in the LIB formation process as a classification task and proposes a data-driven model based on incomplete time-series data for formation pre-screening. First, the proposed model transforms segments of the incomplete charging voltage curve (ICVC) of the LIB as tokens, which is a more compact and less redundant data representation of the ICVC. Then, the attention-based feature encoder, transformer encoder (TE), captures the dependency between tokens to extract features for the formation pre-screening. Finally, a task-specified decoder, feature enhance decoder (FED), is used to screen out fault cells online. The effectiveness of the proposed model is verified by experiments on real-world production data. The results show that the proposed model can achieve an accuracy of 98.73% and a lower miss rate of only 1.92% during formation pre-screening, which is a 2.49% improvement in accuracy and an 8.98% decrease in miss rate compared with the deep residual network. Our method can more accurately screen fault cells online than existing models, thus reducing the energy consumption and rework time of fault cells in LIB production.

Keywords: lithium-ion battery; battery formation; time-series classification; data-driven model

---

1. Introduction

Due to the advantages of high energy density and efficiency and long service life, lithium-ion batteries have been widely used in energy storage systems, electric vehicles, and portable devices [1–4]. In lithium-ion battery manufacturing, the formation process is crucial for achieving high quality and long service life of a LIB, but this process is time- and energy-consuming [5–7]. During the initial charging of cells in the formation process, the solid electrolyte interphase (SEI) [8,9] is formed, which is vital for the life and cycle performances of the LIBs. SEI can protect the electrolyte from reductive decomposition and improve the charging and discharging performances of the LIBs [10,11]. Moreover, stable and durable SEI is essential for a high-quality LIB and can be matured in the following service life [12,13]. The formation process differs among manufacturers and battery types, but the majority of formation processes include steps shown in Figure 1.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2026 doi:10.20944/preprints202603.2033.v1

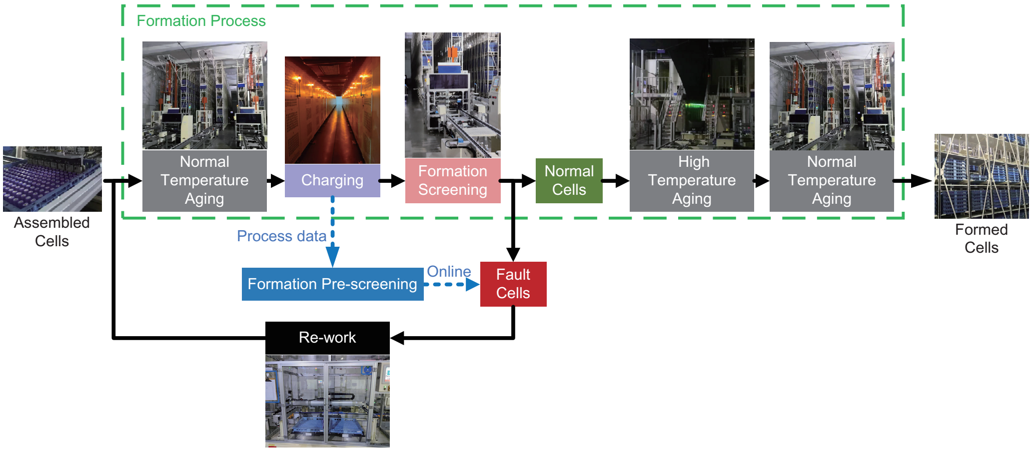

Figure 1. The formation process (green dotted box) and the relationship with formation pre-screening

In the formation process, the most important step is the charging stage, where a cell obtains the SEI. A large part of the time and energy consumed in the formation process is used by the charging process. Therefore, several studies have aimed at reducing the battery formation time and optimizing the values of the formation process parameters without affecting the battery performance. Andrew et al. [14] studied the influence of several formation processes on the discharging capacity and capacity retention to make the formation process faster. Drees et al. [15] proposed a fast formation process that could reduce the formation time while retaining the cycle performance of cells by Real-Time Negative Electrode Voltage Control and bypassing the intercalation step. An et al. [16] designed an efficient formation process and reduced the formation time by six times while guaranteeing the quality of cells and improving capacity retention. Drees et al. [17] developed an electrode equivalent circuit model for an NMC622 battery, which predicted the electrode voltages to optimize the fast charging profile, thus significantly shortening the formation time. Eras et al. [18] investigated the electricity consumption in the formation process of lead-acid batteries and proposed an energy management method to reduce energy costs and improve the energetic performance of the battery factory. Mao et al. [19] studied several different formation protocols and proposed an optimized formation protocol to decrease the high manufacturing cost caused by the long formation time. The above-mentioned studies have improved the efficiency of the formation process to a certain extent without affecting the quality of the battery. However, the fault cells cannot be screened timely in the charging process because the online screening process is not involved in most formation protocols. The fault cells can be screened only offline after charging according to certain indexes, such as the open-circuit voltage (OCV), as shown in Figure 1. This can cause energy waste and extend the rework cycle, thus increasing the cost and time of LIB production.

Recently, data-driven methods have been widely used in various industrial fields and processes. In the field of LIB manufacturing, many data-driven approaches based on machine learning and deep learning technique have been proposed for different scenarios. Yang et al. [20] proposed a density-based clustering algorithm to cluster cells using the discharge voltage curve (DVC) to improve the consistency within the battery pack. Duquesnoy et al. [21] combined the experimental results and machine learning-based algorithms to investigate the manufacturing parameters–electrode properties interdependencies and determine the effect of the calendaring process on the battery electrochemical performance. Turetskyy et al. [22] proposed a multi-output method based on data-driven models, which can assist the cell design and decision-making process in LIB production. Wang et al. [23] proposed a fast voltage-abnormal cell detection method for LIB mass production, suitabling for time-critical industrial screening. a pre-trained data-driven cell screening method for LIB grouping was introduced by Wang et al. [24]. pre-training can effectively improve feature extraction while reducing the dependence on labeled samples. Zhang et al. [25] proposed a coupled mechanism model for lithium-ion batteries , and introduced a data-driven parameter identification framework to address the large number of unknown model parameters. Huang et al. [26] proposed a hybrid

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2026 doi:10.20944/preprints202603.2033.v1

data-driven network for LIB fault prediction, aiming to address the strong nonlinearity and complex electrochemical dynamics in battery state sequence data.

These studies demonstrate the feasibility and effectiveness of data-driven models in LIB manufacturing. Inspired by the above advances, this study adopts a data-driven online screening strategy using process data during formation charging to further reduce energy consumption and improve production efficiency in LIB formation. By leveraging the rich temporal information embedded in formation-stage process signals, the proposed method aims to achieve accurate and efficient cell screening without introducing additional waiting time.

During the formation charging, the process data of each cell are collected as time series, such as a charge voltage curve (CVC). These raw time series data contain information on the magnitude and variation of charging, which can reflect the quality of cells. For online fault cell screening, an incomplete version of the CVC, which is shown in Figure 2, is used to classify the fault cells before the formation charging terminates. Nevertheless, the raw ICVC data of cells are long sequences and have high redundancy. The characteristics of the ICVC data of the normal cells and part of the fault cells are similar and overlap, as shown in Figure 3, which makes it challenging to directly screen fault cells online based on the raw ICVC data.

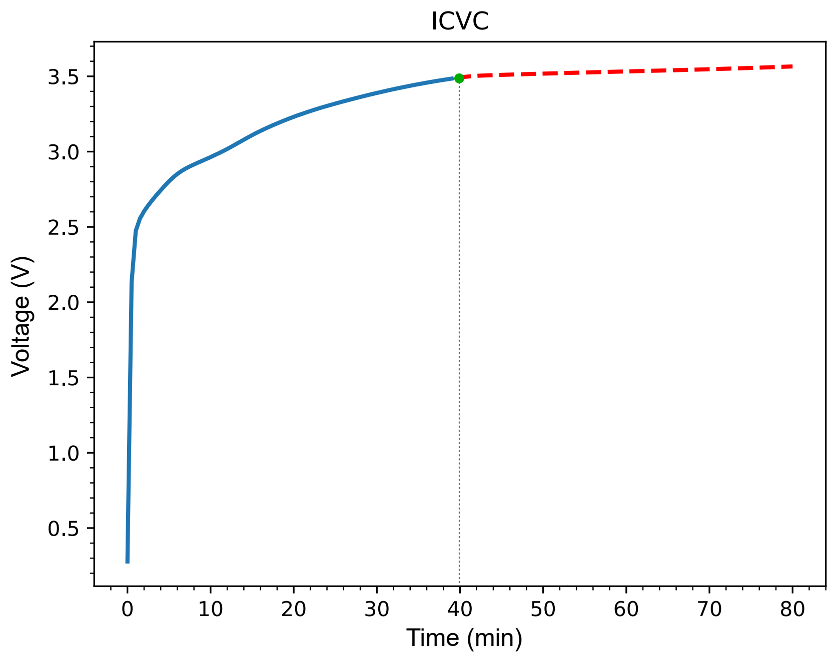

Figure 2. The ICVC data (blue solid line) with the cut-off time of 40 min (green point) and the corresponding CVC data (blue solid line and red dash line).

The attention mechanism (AM) is an effective neural network structure in the fields of natural language processing [27], speech recognition [28], computer vision [29], and time series forecasting [30]. The AM is inspired by the human visual system, and it attempts to focus on the saliency part of a scene or an image while neglecting irrelevant information [31]. Recently, there have been several modifications and improvements in the AM, such as multi-dimensional attention, hierarchical attention, self-attention, and memory-based attention. These variants have been designed for the purposes of better performance and more flexible application, but the operational principle of the AMs is the same [32]. Among different types of AMs, self-attention represents a special and effective AM. Self-attention was first proposed by Wang et al. [33] which calculate attention weight only based on the input data. Self-attention aims to capture long-range, multi-level intricacy relations and dependency within input data [34]. For the ICVC data in the formation process, self-attention can reduce data redundancy and

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2026 doi:10.20944/preprints202603.2033.v1

extract distinctive features by modeling the dependency of substages in the ICVC data for better online screening performance.

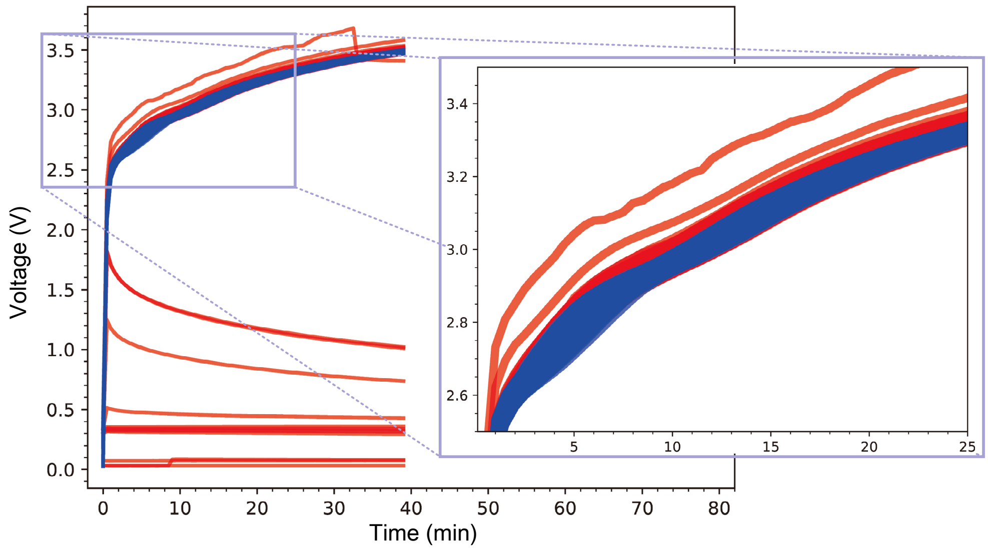

Figure 3. Raw ICVC data consisting of both normal cells (blue line) and fault cells (red line); the inset represents the zoomed view of the ICVC data.

This study proposes an online cell screening method for LIB formation, which screen the quality of cells during the formation charging according to a part of the process data, such as ICVC. In this study, the formation pre-screening is considered a classification problem, and a data-driven model is proposed for solving the formation pre-screening problem. The proposed model first transforms the input ICVC data into a collection of tokens, which represent a more compact and less redundant data representation of the ICVC. Then, the feature encoder, which is based on the TE [35] , extracts the features of tokens for the formation pre-screening task. Finally, a task-specified decoder, namely the FED, is designed to enhance the features to improve online cell screening performance. The proposed data-driven model is named TE-FED. To the best of the authors' knowledge, there has been no study that introduces an online fault cell screening method into the formation protocol of LIB production. The results of the experiments conducted on real-world industrial data obtained from the LIB production demonstrate that the proposed model can screen out fault cells 40 min earlier than offline screening methods can, with an accuracy of 98.73% and a miss rate of 1.92%, which is a 2.49% improvement in accuracy and an 8.98% decrease in miss rate compared with deep residual network (ResNet). Our method can accurately screen fault cells online, thus reducing the energy consumption and rework time of fault cells in LIB production.

- An online fault cell screening method is proposed to compensate for the lack of online screening techniques in formation protocol, which can reduce energy consumption and improve the production efficiency of the LIB formation.
- The TE based feature encoder is adopted to tokenize the ICVC data and convert it into more compact and less redundant token-based feature for more conducive to feature extraction.
- The FED, a task-specified decoder, is proposed to enhances the quality of the token-based feature with a dual branch feature decoder to capture different types of features and relationships between tokens.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2026 doi:10.20944/preprints202603.2033.v1

The rest of this paper is organized as follows. The overall structure and framework of the proposed method are described in Section 2. In Section 3, the implementation detail of the experiments and experimental results are presented. The conclusions are drawn in 4.

2. Proposed Method

To implement formation pre-screening, the ICVC data of a charging cell, which are denoted by $\mathbf{x} = (x_0, x_1, x_2, ..., x_{d_{ICVC}})$, are divided into several segments and concatenated to form segmented ICVC data, as shown in Figure 4, where $\mathbf{x} \in \mathbf{R}^{d_{ICVC}}$, $x_i \in \mathbf{R}$, $0 \le i \le d_{ICVC}$, $d_{ICVC} \le d_{CVC}$. d_{ICVC} is the length of the input ICVC. d_{CVC} is the length of the CVC data. Then, each row of the segmented ICVC data is embedded as a token through segment embedding, and information on the segment order is added to tokens by position encoding. Finally, tokens are processed by the TE to capture their correlations and produce an attentive feature for the FED to screen fault cells before formation charging is conducted. The overall architecture of the TE-FED is shown in Figure 4, where it can be seen that the TE-FED includes four parts, segment embedding, position encoding, TE, and FED. More details about the four parts are provided in the following sections.

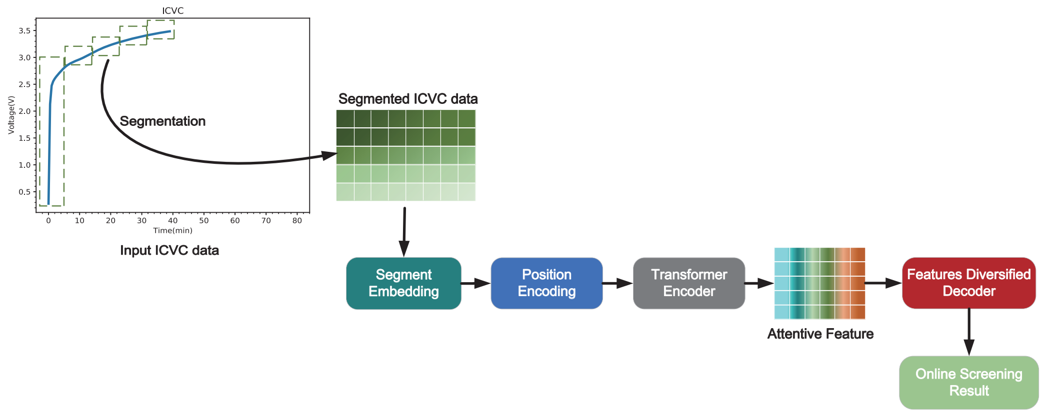

Figure 4. The overall structure of the TE-FED.

2.1. Segment Embedding

During the segment embedding, the input ICVC data are converted through several charging substages by dividing $\mathbf{x}$ into n non-overlapping segments, $\mathbf{f}_j \in \mathbf{R}^{d_{seg}}$, $0 \le j \le n$, where d_{seg} is the length of a segment. Each segment represents a charging phase and contains local charging information. A one-dimension convolution layer is used to transform a segment $\mathbf{f}_j$ into a d_{token}-dimension vector, which represents a token $\mathbf{t}_j \in \mathbf{R}^{d_{token}}$ for the corresponding charging substage. for the corresponding charging substage. This vectorial representation of the data is more compact and less redundant than the segment $\mathbf{f}_j$. The segment embedding is performed as follows:

$$\mathbf{t}_j = Conv1d(\mathbf{f}_j) \tag{1}$$

$$\mathbf{T} = Concat(\mathbf{t}_0, \mathbf{t}_1, ..., \mathbf{t}_n) \tag{2}$$

where $Conv1d(.)$ denotes a one-dimension convolution. Next, tokens are concatenated to form a token matrix $\mathbf{T} = (\mathbf{t}_0, \mathbf{t}_1, ..., \mathbf{t}_n)^T \in \mathbf{R}^{n \times d_{token}}$, where $d_{token} < d_{seg}$ and $\mathbf{T}$ is a token-based representation of $\mathbf{x}$, and where $(.)^T$ is the transpose operator and $Concat(.)$ is the concatenation operator. The block diagram of segment embedding is shown in Figure 5.

2.2. Position Encoding

The ICVC data of a charging cell obtained after segment embedding, denoted by $\mathbf{x}$, are decomposed into several substages to conduct formation pre-screening, and a token-based data representation of the ICVC is obtained. However, the ICVC structure is destroyed during segmentation and segment embedding. Moreover, the chronological order of the substages is lost, but this information is important

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2026
doi:10.20944/preprints202603.2033.v1

to the TE to obtain token dependency. Therefore, to emphasize the position order of different segments in $\mathbf{x}$, this study inserts the position information into each token in $\mathbf{T}$ through position encoding to help the TE apprehend the sequential order of tokens. Generally, there are mainly two types of position encoding, learnable and fixed position encodings [36]. Fixed position encodings represent a sequential order with a pre-defined function, such as a cosine function, and are effective and have a low cost. However, they are calculated in advance and cannot be updated with model training. Learnable position encodings use a set of learnable parameters to represent a sequential order. These learnable parameters can be updated with model training to adapt different applications. Therefore, learnable position encoding is used herein, and can be expressed as follows:

$$\mathbf{T}_{PE} = \alpha \mathbf{T} + \mathbf{P} \tag{3}$$

where $\mathbf{P} \in \mathbf{R}^{n \times d_{token}}$ is the position encoding matrix, whose each element is a learnable parameter; $\mathbf{T}_{PE} \in \mathbf{R}^{n \times d_{token}}$ is the token matrix after position encoding; α is a scale factor that prevents the input data from being overwhelmed by position information, which could lead to a degeneration in the model performance.

Specifically, $\alpha = \sqrt{d_{token}}$ is used in the proposed model. The details of the position encoding in the proposed model are shown in Figure 5.

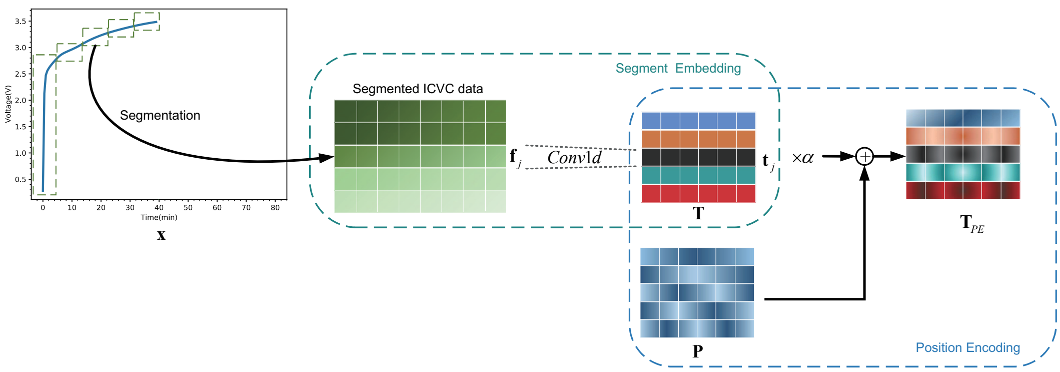

Figure 5. The block diagram of segment embedding and position encoding.

2.3. Transformer Encoder

Due to the superior performance of the transformer for sequential data, this study introduces the TE into the formation pre-screening task to improve the feature extraction performance. The structure details of the TE are illustrated in Figure 6, where it can be seen that it is stacked by L encoder layers, and $\mathbf{T}_{PE}$ and $\mathbf{Q}_o^L$ denote the input and output of the TE, respectively. The encoder layer is based on multi-head attention with p heads and a point-wise feed-forward network (FNN).

2.3.1. Multi-Head Attention

Instead of performing self-attention in a certain feature space, multi-head attention is performed, which mapping input queries to different subspaces and performing self-attention in turn. The self-attention operation can calculate the similarity between tokens in input data and express it as a weight to indicate the relationship between different tokens and their contribution to the output, as follows:

$$\mathbf{Q}_t^l = \mathbf{Q}^l \mathbf{W}_t^{l,Q} \tag{4}$$

$$\mathbf{K}_t^l = \mathbf{Q}^l \mathbf{W}_t^{l,K} \tag{5}$$

$$\mathbf{V}_t^l = \mathbf{Q}^l \mathbf{W}_t^{l,V} \tag{6}$$

$$\mathbf{H}_t^l = SelfAttention(\mathbf{Q}_t^l, \mathbf{K}_t^l, \mathbf{V}_t^l) = Softmax(\frac{\mathbf{Q}_t^l (\mathbf{K}_t^l)^T}{\sqrt{d_k}})\mathbf{V}_t^l \tag{7}$$

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2026 doi:10.20944/preprints202603.2033.v1

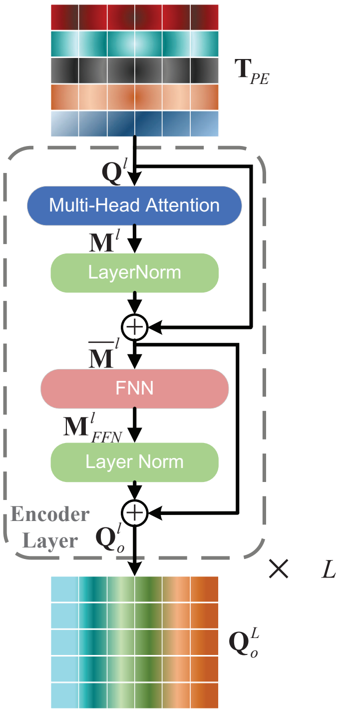

Figure 6. The structure of the TE.

where $\mathbf{Q}^l \in \mathbf{R}^{L_Q \times d_{in}}$ is the input query matrix of l-th encoder layer, whose each row is a query vector, L_Q and d_{in} are the number of the query vector in $\mathbf{Q}^l$ and the dimension of the query vector,respectively; $\mathbf{W}_t^{l,Q} \in \mathbf{R}^{d_{in} \times d_q}$, $\mathbf{W}_t^{l,K} \in \mathbf{R}^{d_{in} \times d_k}$, and $\mathbf{W}_t^{l,V} \in \mathbf{R}^{d_{in} \times d_v}$ are the linear projection operators for generating the t-th head's queries $\mathbf{Q}_t^l \in \mathbf{R}^{L_Q \times d_q}$, keys $\mathbf{K}_t^l \in \mathbf{R}^{L_Q \times d_k}$, values $\mathbf{V}_t^l \in \mathbf{R}^{L_Q \times d_v}$, respectively; d_q, d_k, and d_v are the dimension of query vector, key vector, and value vector in $\mathbf{Q}_t^l$, $\mathbf{K}_t^l$, and $\mathbf{V}_t^l$. $Softmax(.)$ is the soft-max function; $\mathbf{H}_t^l \in \mathbf{R}^{L_Q \times d_v}(0 < t < p)$ is the attention output of the t-th head.

Further, the variant subspace features are concatenated and synthesized by a linear transformation to form the encoded features denoted by $\mathbf{M}^l$; each feature vector in $\mathbf{M}^l$ contains correlation information on queries in input $\mathbf{Q}^l$. The multi-head attention with heads in the th encoder layer is expressed as follows:

$$\mathbf{H}_o^l = Concat(\mathbf{H}_0^l, \mathbf{H}_1^l, ..., \mathbf{H}_p^l) \tag{8}$$

$$\mathbf{M}^l = MultiHeadAttention(\mathbf{Q}^l) = \mathbf{H}_o^l \mathbf{W}^{l,O} \tag{9}$$

where $\mathbf{H}_o^l \in \mathbf{R}^{L_Q \times p d_v}$ is obtained by concatenating the results of all heads; $\mathbf{H}_t^l$ is transformed by $\mathbf{W}_O^l \in \mathbf{R}^{p d_v \times d_{out}}$ to obtain the encoded features $\mathbf{M}^l \in \mathbf{R}^{L_Q \times d_{out}}$; $\mathbf{M}^l$ is normalized by layer normalization [37] and added with a residual connection of $\mathbf{Q}^l$ to obtain $\overline{\mathbf{M}}^l \in \mathbf{R}^{L_Q \times d_{out}}$.

2.3.2. Point-Wise Feed-Forward Network

A point-wise FFN is a fully-connected network that processes the elements of $\overline{\mathbf{M}}^l$ separately and identically as:

$$\mathbf{M}_{FFN}^l = FFN(\overline{\mathbf{M}}^l) = Dropout(Relu(\overline{\mathbf{M}}^l \mathbf{W}_1^l + \mathbf{b}_1^l))\mathbf{W}_2^l + \mathbf{b}_2^l \tag{10}$$

where $\mathbf{W}_1^l$ and $\mathbf{W}_2^l$ are the network weights of the FFN; $\mathbf{b}_1^l$ and $\mathbf{b}_2^l$ are network biases of the FFN; $Relu(.)$ is the linear rectified function [38]. $Dropout(.)$ represents the dropout layer, which suppresses the input with a probability of q to avoid model overfitting.

The output of the encoder layer $\mathbf{Q}_o^l$ is obtained by normalizing $\mathbf{M}_{FFN}^l$ and adding it with a residual connection of $\overline{\mathbf{M}}^l$.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2026 doi:10.20944/preprints202603.2033.v1

2.4. Feature Enhance Decoder

To decode the output features of the TE and screen out the fault cells online, this study designs a task-specified decoder called the feature enhance decoder, which enhances feature diversity by transforming the input features into various feature spaces with convolution and pooling operations to capture different types of features and relationships between tokens. Then, these features are combined to classify fault cells online.

The FED structure is shown in Figure 7. The input features $\mathbf{Q}_o^l$ are processed by a convolution layer with a kernel size of $1 \times 1 \times 32$ first and then normalized by layer normalization and activated by Gelu [39] to obtain unenhanced features $\mathbf{F}$. Next, $\mathbf{F}$ is divided into two branches to obtain different types of features, as shown in Figure 7. The right branch (refer to Figure 7) synthesizes the features between adjacent heads and adjacent tokens using a convolution filter to obtain the convolutional features $\mathbf{F}_{Conv}$ that contain the local correlation between adjacent tokens and heads. Meanwhile, the left branch (refer to Figure 7) produces the pooling features $\mathbf{F}_{Pool}$ by using a max-pool and convolution filter, which contain global features of the input data. The dropout layers are added after each convolution layer to avoid model overfitting. Afterward, the $\mathbf{F}_{Conv}$ and $\mathbf{F}_{Pool}$ are channel-wise concatenated and pooled by global average pooling [40,41] that pool each channel to produce a comprehensive feature vector. Finally, a linear layer and a soft-max layer are used to transform the feature vector into classification probability for the formation pre-screening task.

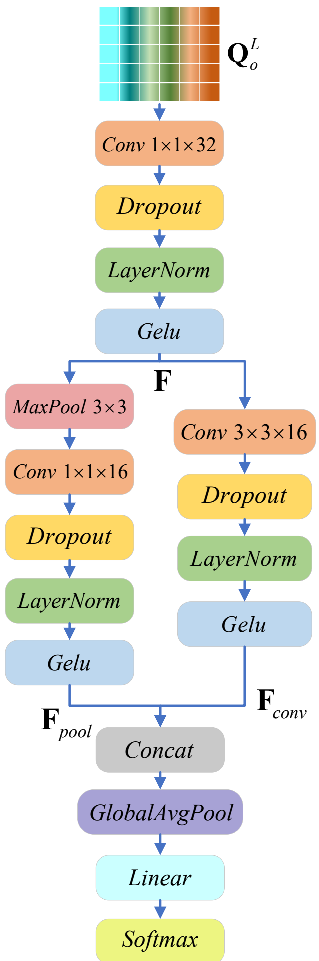

Figure 7. The structure of FED.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2026 doi:10.20944/preprints202603.2033.v1

3. Experiments

3.1. Data Preparation

The experiments conducted in this study were conducted using the practical production data of a 18650 LIB from an automatic formation system. The CVC data of each cell were collected as raw data, and the corresponding OCV result was used as a label to form a dataset for the formation pre-screening task. Particularly, the CVC and OCV data on 47,959 cells were collected. The CVC data were aligned by linear interpolation into the reference time step $\mathbf{T}_{ref} = \{0, 0.5, 1, ..., 80\}$, which contained 161 data points with a time interval of 30 s. For each cell, if the corresponding OCV result was in the range of 3.54 V–3.55 V, then the cell was considered well-formed, which denoted a positive sample; otherwise, the cell was regarded as a fault cell, that was, a negative sample.

Due to the advance in production technology and quality management technique in the battery industry, the number of fault cells has been far less than the number of normal cells in the real-world formation process. To avoid the imbalance in the number of positive and negative samples affected model performance, the hybrid resampling technique [42] was employed to balance the raw dataset by reducing the number of majorities (i.e., a well-formed cell) and increasing the number of minorities.

Finally, the CVC data were trimmed to a certain length as input ICVC data and paired with the corresponding OCV labels to construct a formation pre-screening dataset used for the model's training, validation, and test. The proportion of positive and negative samples in the formation pre-screening dataset was the same as in the resampled dataset. Detailed information on the dataset is given in Table 1.

Table 1. The detail of the formation pre-screening dataset.

Data Set		Positive	Negative	Total
Raw Data		93.23%	6.77%	47959
Resampled Data		66.65%	33.35%	46928
Formation Pre-Screening Data	Train	66.65%	33.35%	28160
	Validation	66.65%	33.35%	9384
	Test	66.65%	33.35%	9384

3.2. Experimental Settings

Currently, there is no existing formation pre-screening method that could be used for comparison with the proposed method. Therefore, several deep learning-based models were adopted for the formation pre-screening task and compared with the proposed mode. The baseline models used in the comparison include the multi-layer perceptron (MLP), fully-convolutional network (FCN), and ResNet. The detailed structures of these models have been discussed in detail by Taherkhani et al. [43].

The performances of the above models in the formation pre-screening task were evaluated using several evaluation metrics: accuracy, F1-value, G-mean, and miss rate. The average running time of each model was recorded to obtain its computation complexity. The details about the evaluation indexes are shown in Table 2, where TP denotes the number of true positive predictions; TN is the number of true negative predictions; FP is the number of false positive predictions, and FN is the number of false negative predictions.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

 doi:10.20944/preprints202603.2033.v1

Table 2. The details of the evaluation indexes.

Evaluation index	Definition	Value range	Best value
Accuracy	$\frac{TP+TN}{P+N}$	$[0,1]$	1
F1-value	$2 \times \frac{Precision \times Recall}{Precision + Recall}$	$[0,1]$	1
G-mean	$\sqrt{Recall \times Specificity}$	$[0,1]$	1
Miss Rate	$\frac{N-TN}{N}$	$[0,1]$	1

Accuracy is a naive evaluation metric in binary classification problems, and it represents a ratio of the number of correct predictions to the total number of test samples [44]. The accuracy index is a basic metric of model performance, but it is susceptible to variations in sample ratio and a large bias for extremely imbalanced datasets.

The F1-value represents a harmonic average of the precision and recall for a binary classification problem. It can alleviate the evaluation bias caused by an imbalanced dataset and obtain more objective evaluation results of model capacity. The precision is a ratio of the number of true positive predictions to the number of samples that are predicted as positive samples [45]. The recall is a ratio of the number of true positive predictions to the number of all positive samples.

The G-mean denotes a combination of the recall index and specificity index, which can reflect the model performance on an imbalanced dataset. The specificity represents a ratio of the number of true negative predictions to the number of samples that are predicted as negative samples.

The miss rate is a vital index for the formation pre-screening task because screening out all the negative samples is important for quality control in battery manufacturing.

The precision, recall, and specificity values are respectively calculated by:

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$$Specificity = \frac{TN}{TN + FP} \tag{13}$$

The default hyperparameter of the TE-FED was set to be the same as the number of heads $p = 8$; the number of hidden units in the FFN was $d_{hid} = 256$; the number of encoder layers was $L = 6$; the segment length was $d_{seg} = 16$; the token length was set to $d_{token} = 12$; dropout probability was $q = 0.9$; incomplete ICVC data had $d_{ICVC} = 80$. Further, the k-fold cross-validation [46] was used to train and evaluate models, and $k = 5$, and the ratio of training and validation data was the same as that on the formation pre-screening dataset in Table 1. All experiments were performed on a desktop platform with an Intel Core i9-10980XE, 128-GB RAMs, and RTX 3090 24GB. All of these models were implemented with PyTorch [47].

3.3. Model Training

The proposed model was trained using Adam optimizer [48] and the learning rate scheduler called stepLR. The parameters of the Adam optimizer were: $\beta_1 = 0.9$, $\beta_2 = 0.9$, and $\varepsilon = 10^{-9}$. The stepLR was formulated as follows:

$$lr_{epoch} = lr_{init} \times \gamma^{\left\lceil \frac{epoch}{v} \right\rceil} \tag{14}$$

where lr_{init} is the initial learning rate; $\gamma \in [0,1]$ is the decay factor.; v is the step size; $epoch$ is the index of current epoch; lr_{epoch} is the learning rate of the current epoch; $\lceil . \rceil$ is the integer ceiling function, and in this study, $lr_{init} = 0.0001$, $\gamma = 0.95$, $v = 10$, and the maximum number of training epochs was 512.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2026 doi:10.20944/preprints202603.2033.v1

The loss function in the formation pre-screening task was the cross-entropy loss function, and the loss between the classification result and the ground true label was propagated from the decoder output to the input of the TE-FED. As shown in Figure 8, the training loss reduced and became stable after hundreds of epochs. Meanwhile, the validation loss declined in fluctuations and stabilized eventually, which indicated that the model converged.

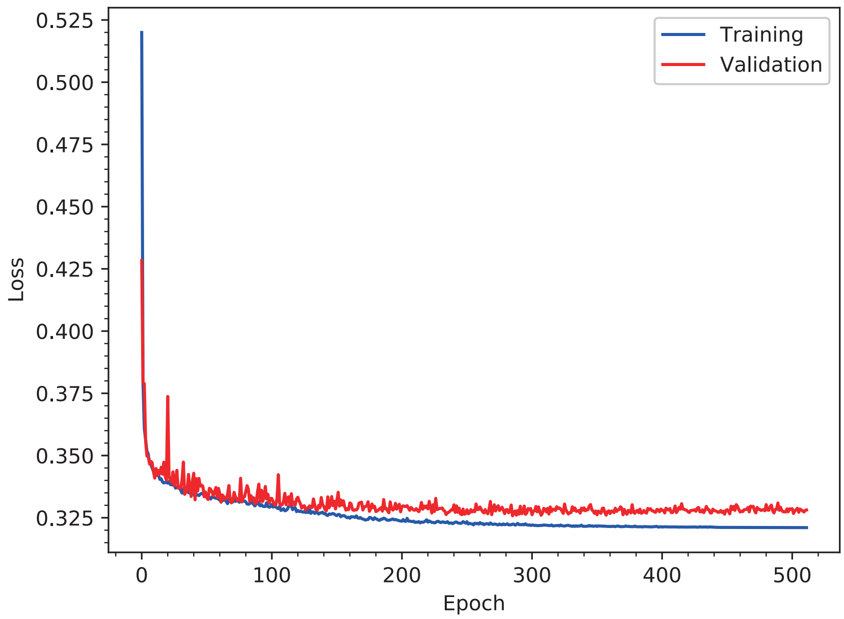

Figure 8. The training loss and validation loss of the TE-FED.

3.4. Results and Discussions

The overall evaluation of different data-driven models for formation pre-screening is shown in Table 3, where the best result of each index is boldfaced. As shown in Table 3, the proposed TE-FED achieved the best accuracy result of 98.73% among all the models, which was 4.46% higher than that of the MLP, and 2.49% higher than that of the ResNet. Furthermore, the lowest miss rate among all models, i.e., 1.92%, was obtained by the TE-FED, which was a decrease of 8.98% compared with ResNet. This was because the attention-based encoder could capture the relationship within input tokens and provide better features for the formation pre-screening task. The result of the TE-FED regarding the F1-value and G-mean was also the best among all models. Regarding the running time, the MLP achieved the best result of 0.3517 ms per sample. However, the cells were screened simultaneously, and the running time performance of the proposed model could satisfy the requirement of the formation pre-screening task in real-world production. Overall, the proposed model could pre-screen the cells that failed during the formation charging at high accuracy and low miss rate.

Table 3. The experimental result of models comparison.

Model	Evaluation Index				
	Accuracy	F1-value	G-mean	Miss Rate	Running Time(ms)
MLP	0.9427	0.9585	0.9148	0.1566	**0.3517**
FCN	0.9569	0.9685	0.9366	0.1179	0.7940
ResNet	0.9624	0.9725	0.9430	0.1090	2.2079
TE-FED	**0.9873**	**0.9905**	**0.9857**	**0.0192**	3.9116

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2026 doi:10.20944/preprints202603.2033.v1

To illustrate the capability of the TE-FED better, the TE-FED was used for different d_{ICVC} to investigate the performance of the model in the formation pre-screening task. For the TE-FED, the shorter the length of ICVC data used in formation pre-screening was, the earlier the fault cells could be screened out. However, a short input length could lead to higher uncertainty and damage the online screening performance of the model. In the experiment, the TE-FED was trained and tested on the ICVC data with different lengths to reveal the online screening performance of the model on different input lengths. Exclusively, the model was trained and tested on the ICVC data with $d_{ICVC} = 40, 80$, and 161 in turn for comparison. The experimental results are shown in Table 4.

Table 4. The experimental results of the proposed model for different input lengths.

Input length d_{ICVC}	Evaluation Index		
	Accuracy	F1-value	G-mean
40	0.9613	0.9732	0.9507
80	0.9873	0.9905	0.9857
161	**0.9907**	**0.9933**	**0.9936**

As shown in Table 4, the best result was achieved when the model trained on the full-sequence input was $d_{ICVC} = 161$, but such input data made the formation pre-screening meaningless. The proposed model could also achieve a significant result by using only half or a quarter of the CVC data to pre-screening the cells. At the input length $d_{ICVC} = 40$, the proposed model could achieve an accuracy of 96.13%, and the fault cells could be screened out 60 min early. At the input length of $d_{ICVC} = 80$, the proposed model achieved an accuracy of 98.73%, which was only 0.34% less than the full-sequence length prediction, and the fault cells could be screened out 40 min earlier.

To further investigate the sensitivity of the TE-FED on the hyperparameter, the online screening performance of the proposed model was examined for different input segment lengths d_{seg}. Before passing into the encoder, the ICVC data were divided into several fixed-length segments, each of which represented a substage of charging and was embedded as tokens. The relationship between the tokens could indicate the quality of formation charging. However, different segment lengths could cause online screening performance variations because the scales of the substage were charged. Thus, the segment length was changed, and the corresponding variations in the model performance were observed to investigate the sensitivity of the proposed model on d_{seg}. We performed sensitivity experiments on segment length using the proposed model structure with default settings. The experimental results are shown in Table 5.

Table 5. The experimental results of the proposed model for different segment lengths.

Segment Length d_{seg}	Evaluation Index		
	Accuracy	F1-value	G-mean
8	0.9742	0.9809	0.9644
16	0.9873	0.9905	0.9857
40	**0.9876**	**0.9907**	**0.9863**
80	0.9617	0.9712	0.9590

As shown in Table 5, the optimal result was obtained at $d_{seg} = 40$. The performance of the proposed model at $d_{seg} = 16$ was almost the same as that at $d_{seg} = 40$, indicating that the quality information of the tokens at $d_{seg} = 40$ and $d_{seg} = 16$ showed no significant difference. Thus, $d_{seg} = 16$ did not affect the proposed model's performance. However, there were certain degradations in model performances when $d_{seg} = 8$ and $d_{seg} = 80$. Namely, at $d_{seg} = 8$, the short length of the segment

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2026 doi:10.20944/preprints202603.2033.v1

could not guarantee the validity of quality information; meanwhile, at $d_{seg} = 80$, the long length of the segment could produce defective tokens due to the charging substages with different quality information were overlapped.

3.5. Ablation Study

The proposed TE-FED is an encoder-decoder model that contains four main parts, segment embedding (SE), position encoding (PE), TE, and FED. To analyze the effect of these components on the formation pre-screening task, four variant models were designed. The first variant was the w/o SE, which represented the TE-FED that did not use the SE to transform the segments of the ICVC data into tokens as input but the original ICVC segments. The second variant was the w/o PE that did not use the PE to inject the segment order information into the ICVC tokens. The third variant was the w/o TE, which used a simple MLP encoder as a feature encoder to replace the TE. The MLP encoder was a two-layers neural network with a sigmoid activation function [49] and layer normalization. The last variant was the w/o FED that did not use the FED but a simple MLP decoder as a feature decoder. The MLP decoder was a two-layer neural network with sigmoid and softmax activation functions. The experimental result of the ablation study is shown in Table 6.

Table 6. The experimental result of ablation study.

Variant Model		w/o SE	w/o PE	w/o TE	w/o FED	TE-FED
Component	SE	×	✓	✓	✓	✓
	PE	✓	×	✓	✓	✓
	TE	✓	✓	×	✓	✓
	FED	✓	✓	✓	×	✓
Evaluation Index	Accuracy	0.9512	0.9302	0.8845	0.9638	0.9873
	F1-value	0.9635	0.9475	0.9189	0.9732	0.9905
	G-mean	0.9434	0.9224	0.8227	0.9644	0.9857

The performance degeneration of the w/o SE and w/o PE demonstrated the effectiveness of token transformation and position information in the formation screening task. The result of the w/o TE indicated that a simple encoder structure could not achieve sufficient data features for formation screening, and the TE could obtain better online screening performance due to the capacity of dependency modeling. Finally, the result of the w/o FED demonstrated that the FED could help to improve the online screening performance further.

4. Conclusions

This paper proposes an online fault cell screening method based on the data-driven model for LIB formation. The proposed screening model is an attention-based model named TE-FED, which can capture the relation between segments of ICVC data to screen fault cells online. The production data of a 18650 LIB are used in experiments to verify the feasibility of the TE-FED. The experimental result demonstrates that in the online screening task on the ICVC data with , the proposed model has an accuracy of 98.73% and a miss rate of 1.92%. The proposed model can screen out fault cells 40 min earlier than the other offline screening models; the screened fault cells are proceeded to the rework process or discarded in advance, which can reduce the energy consumption in battery formation and shorten the screening cycle of fault cells. In addition, the proposed model can screen out fault cells 60 min earlier, achieving an accuracy of 96.13%. However, the running time of the proposed model is not optimal compared to the comparison models, which can affect the production speed when cells are pre-screened sequentially.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2026 doi:10.20944/preprints202603.2033.v1

In future work, we will simplify the model structure and reduce the running time to meet the requirements of various production environments. Moreover, since the aging temperature also affects the quality of battery formation, we will use the temperature curve during aging and research online fault cell screening methods based on multi-source data to obtain better screening performance and efficiency in the process of LIB production.

Author Contributions: Conceptualization, X.W.; methodology, A.D. and X.W.; software, A.D. and X.W.; writing—original draft preparation, A.D. and X.W.; writing—review and editing, R.L. and F.H.; visualization, X.W.; supervision, J.H.; funding acquisition, J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key and Plan under Grant 2020YFB1710600.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Coester, A.; Hofkes, M.W.; Papyrakis, E. Economic analysis of batteries: Impact on security of electricity supply and renewable energy expansion in Germany. *Applied Energy* **2020**, *275*, 115364. https://doi.org/https://doi.org/10.1016/j.apenergy.2020.115364.
2. Zhang, J.; Mou, Y.; Yang, H.; Guo, K.; Wang, C.; Song, J.; Tian, W.; Qiu, S.; Su, G. Experimental feasibility study on the application of high-temperature heat pipes in the cooling process of graphitization furnaces. *Energy* **2025**, *334*, 137650–137650. https://doi.org/10.1016/J.ENERGY.2025.137650.
3. Shen, N.; Wang, Y.; Peng, H.; Hou, Z. Renewable Energy Green Innovation, Fossil Energy Consumption, and Air Pollution—Spatial Empirical Analysis Based on China. *Sustainability* **2020**, *12*. https://doi.org/10.3390/su12166397.
4. Yang, C. Running battery electric vehicles with extended range: Coupling cost and energy analysis. *Applied Energy* **2022**, *306*, 118116.
5. Zhu, Z.W.; Qiu, J.Y.; Wang, L.; Cao, G.P.; He, X.M.; Wang, J.; Zhang, H. Application of Artificial Intelligence to Lithium-Ion Battery Research and Development. *Journal of Electrochemistry* **2022**, *28*, 2219003. https://doi.org/10.13208/j.electrochem.2219003.
6. Attia, P.M.; Harris, S.J.; Chueh, W.C. Benefits of Fast Battery Formation in a Model System. *Journal of The Electrochemical Society* **2021**, *168*, 050543. https://doi.org/10.1149/1945-7111/abff35.
7. Schomburg, F.; Heidrich, B.; Wennemar, S.; Drees, R.; Roth, T.; Kurrat, M.; Heimes, H.; Jossen, A.; Winter, M.; Cheong, J.Y.; et al. Lithium-ion battery cell formation: status and future directions towards a knowledge-based process design. *Energy & Environmental Science* **2024**, *17*, 2686–2733.
8. Su, C.; He, M.; Shi, J.; Amine, R.; Zhang, J.; Amine, K. Solvation Rule for Solid-Electrolyte Interphase Enabler in Lithium-Metal Batteries. *Angewandte Chemie International Edition* **2020**, *59*, 18229–18233. https://doi.org/10.1002/anie.202008081.
9. Li, B.; Chao, Y.; Li, M.; Xiao, Y.; Li, R.; Yang, K.; Cui, X.; Xu, G.; Li, L.; Yang, C.; et al. A Review of Solid Electrolyte Interphase (SEI) and Dendrite Formation in Lithium Batteries. *Electrochemical Energy Reviews* **2023**, *6*, 7. https://doi.org/10.1007/s41918-022-00147-5.
10. Li, B.; Chao, Y.; Li, M.; Xiao, Y.; Li, R.; Yang, K.; Cui, X.; Xu, G.; Li, L.; Yang, C.; et al. A review of solid electrolyte interphase (SEI) and dendrite formation in lithium batteries. *Electrochemical Energy Reviews* **2023**, *6*, 7.
11. Li, R.; Bao, L.; Chen, L.; Zha, C.; Dong, J.; Qi, N.; Tang, R.; Lu, Y.; Wang, M.; Huang, R.; et al. Accelerated aging of lithium-ion batteries: bridging battery aging analysis and operational lifetime prediction. *Science Bulletin* **2023**, *68*, 3055–3079. https://doi.org/https://doi.org/10.1016/j.scib.2023.10.029.
12. Guo, C.; Du, K.; Tao, R.; Guo, Y.; Yao, S.; Wang, J.; Wang, D.; Liang, J.; Lu, S.Y. Inorganic filler enhanced formation of stable inorganic-rich solid electrolyte interphase for high performance lithium metal batteries. *Advanced Functional Materials* **2023**, *33*, 2301111.
13. Yan, C.; Yao, Y.X.; Cai, W.L.; Xu, L.; Kaskel, S.; Park, H.S.; Huang, J.Q. The influence of formation temperature on the solid electrolyte interphase of graphite in lithium ion batteries. *Journal of Energy Chemistry* **2020**, *49*, 335–338. https://doi.org/https://doi.org/10.1016/j.jechem.2020.02.052.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2026 doi:10.20944/preprints202603.2033.v1

14. Andrew, W.; Peyman, M.; M., A.P.; Valentin, S.; Suhak, L.; Greg, L.; Anna, S. Predicting the impact of formation protocols on battery lifetime immediately after manufacturing. *Joule* **2021**, *5*, 2971–2992. https://doi.org/10.1016/J.JOULE.2021.09.015.
15. Drees, R.; Lienesch, F.; Kurrat, M. Fast Charging Formation of Lithium-Ion Batteries Based on Real-Time Negative Electrode Voltage Control. *Energy Technology* **2023**, *11*, 2200868.
16. An, S.J.; Li, J.; Du, Z.; Daniel, C.; Wood, D.L. Fast formation cycling for lithium ion batteries. *Journal of Power Sources* **2017**, *342*, 846–852. https://doi.org/https://doi.org/10.1016/j.jpowsour.2017.01.011.
17. Drees, R.; Lienesch, F.; Kurrat, M. Fast charging lithium-ion battery formation based on simulations with an electrode equivalent circuit model. *Journal of Energy Storage* **2021**, *36*, 102345. https://doi.org/https://doi.org/10.1016/j.est.2021.102345.
18. Eras, J.J.C.; Sagastume, A.; Santos, V.S.; Herrera, H.H.; Vandecasteele, C. Energy management in the formation of light, starter, and ignition lead-acid batteries. *Energy Efficiency* **2019**, *12*, 1219–1236.
19. Mao, C.; An, S.J.; Meyer, H.M.; Li, J.; Wood, M.; Ruther, R.E.; Wood, D.L. Balancing formation time and electrochemical performance of high energy lithium-ion batteries. *Journal of Power Sources* **2018**, *402*, 107–115. https://doi.org/https://doi.org/10.1016/j.jpowsour.2018.09.019.
20. Yang, J.; Ma, G.; Gao, M.; He, Z. Battery grouping with time series clustering based on features. In Proceedings of the 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), 2017, pp. 1319–1323. https://doi.org/10.1109/ISIE.2017.8001436.
21. Duquesnoy, M.; Lombardo, T.; Chouchane, M.; Primo, E.N.; Franco, A.A. Data-driven assessment of electrode calendering process by combining experimental results, in silico mesostructures generation and machine learning. *Journal of Power Sources* **2020**, *480*, 229103. https://doi.org/https://doi.org/10.1016/j.jpowsour.2020.229103.
22. Turetskyy, A.; Wessel, J.; Herrmann, C.; Thiede, S. Battery production design using multi-output machine learning models. *Energy Storage Materials* **2021**, *38*, 93–112. https://doi.org/https://doi.org/10.1016/j.ensm.2021.03.002.
23. Wang, X.; He, J.; Huang, F.; Liu, Z.; Deng, A.; Long, R. A Novel Voltage-Abnormal Cell Detection Method for Lithium-Ion Battery Mass Production Based on Data-Driven Model with Multi-Source Time Series Data. *Energies* **2024**, *17*, 3472–3472. https://doi.org/10.3390/EN17143472.
24. Wang, X.; He, J.; Huang, F.; Shen, S.; Liu, Z. Abnormal cell cause localization based on contrastive pre-training and unsupervised data-driven model for lithium-ion battery manufacturing. *Journal of Energy Storage* **2024**, *101*, 113743–113743. https://doi.org/10.1016/J.EST.2024.113743.
25. Zhang, W.; Xie, Y.; He, H.; Long, Z.; Zhuang, L.; Zhou, J. Multi-physics coupling model parameter identification of lithium-ion battery based on data driven method and genetic algorithm. *Energy* **2025**, *314*, 134120.
26. Huang, Y.; Gong, X.; Lin, Z.; Xu, L. A hybrid data-driven method for voltage state prediction and fault warning of Li-ion batteries. *Case Studies in Thermal Engineering* **2024**, *64*, 105420.
27. Min, B.; Ross, H.; Sulem, E.; Veyseh, A.P.B.; Nguyen, T.H.; Sainz, O.; Agirre, E.; Heintz, I.; Roth, D. Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys* **2023**, *56*, 1–40.
28. Zhang, Q.; Lu, H.; Sak, H.; Tripathi, A.; McDermott, E.; Koo, S.; Kumar, S. Transformer Transducer: A Streamable Speech Recognition Model with Transformer Encoders and RNN-T Loss. In Proceedings of the ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 7829–7833. https://doi.org/10.1109/ICASSP40776.2020.9053896.
29. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, 2021, [arXiv:cs.CV/2010.11929].
30. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting, 2021, [arXiv:cs.LG/2012.07436].
31. Sasibhooshan, R.; Kumaraswamy, S.; Sasidharan, S. Image caption generation using visual attention prediction and contextual spatial relation extraction. *Journal of Big Data* **2023**, *10*, 18.
32. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. https://doi.org/https://doi.org/10.1016/j.neucom.2021.03.091.
33. Wang, B.; Liu, K.; Zhao, J. Inner Attention based Recurrent Neural Networks for Answer Selection. In Proceedings of the Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, 2016; pp. 1288–1297. https://doi.org/10.18653/v1/P16-1122.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2026 doi:10.20944/preprints202603.2033.v1

34. Chaudhari, S.; Mithal, V.; Polatkan, G.; Ramanath, R. An Attentive Survey of Attention Models, 2021, [arXiv:cs.LG/1904.02874].
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *CoRR* **2017**, *abs/1706.03762*, [1706.03762].
36. Wang, B.; Shang, L.; Lioma, C.; Jiang, X.; Yang, H.; Liu, Q.; Simonsen, J.G. On Position Embeddings in BERT, 2021. International Conference on Learning Representations.
37. Huang, L.; Qin, J.; Zhou, Y.; Zhu, F.; Liu, L.; Shao, L. Normalization techniques in training dnns: Methodology, analysis and application. *IEEE transactions on pattern analysis and machine intelligence* **2023**, *45*, 10173–10196.
38. Agarap, A.F. Deep Learning using Rectified Linear Units (ReLU). *CoRR* **2018**, *abs/1803.08375*, [1803.08375].
39. Hendrycks, D.; Gimpel, K. Gaussian Error Linear Units (GELUs), 2020. Preprint at https://arxiv.org/abs/1606.08415.
40. Jianjun, L.; Yu, H.; Ming, Z.; Gang, L.; Baohua, Z. Multi-scale residual network model combined with Global Average Pooling for action recognition. *Multimedia Tools and Applications* **2021**, *81*, 1375–1393. https://doi.org/10.1007/S11042-021-11435-5.
41. Zhang, H.T.; Sun, Z.Y.; Zhou, J.; Gao, S.; Dong, J.H.; Liu, Y.; Bai, X.; Ma, J.L.; Li, M.; Li, G.; et al. Computed tomography–based COVID–19 triage through a deep neural network using mask–weighted global average pooling. *Frontiers in cellular and infection microbiology* **2023**, *13*, 1116285.
42. Ahmed, Z.; Das, S. A novel hybrid resampling approach to address class-imbalanced issues. *SN Computer Science* **2024**, *5*, 865.
43. Taherkhani, A.; Cosma, G.; McGinnity, T. A deep convolutional neural network for time series classification with intermediate targets. *SN Computer Science* **2023**, *4*, 832.
44. Carrington, A.M.; Manuel, D.G.; Fieguth, P.W.; Ramsay, T.; Osmani, V.; Wernly, B.; Bennett, C.; Hawken, S.; Magwood, O.; Sheikh, Y.; et al. Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2022**, *45*, 329–341.
45. Powers, D.M.W. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, 2020. Preprint at https://arxiv.org/abs/2010.16061.
46. Yan, T.; Shen, S.L.; Zhou, A.; Chen, X. Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm. *Journal of Rock Mechanics and Geotechnical Engineering* **2022**, *14*, 1292–1303.
47. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al., PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*; Curran Associates Inc.: Red Hook, NY, USA, 2019.
48. Salem, H.; Kabeel, A.; El-Said, E.M.; Elzeki, O.M. Predictive modelling for solar power-driven hybrid desalination system using artificial neural network regression with Adam optimization. *Desalination* **2022**, *522*, 115411.
49. Lederer, J. Activation Functions in Artificial Neural Networks: A Systematic Overview, 2021, [arXiv:cs.LG/2101.09957].

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

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.