Comparison among Machine Learning Models Applied in Lithium-ion Battery Internal Short Circuit Detection

ZiHong Zhang[1], Mikel Arrinda[2], and Jon Perez[3]*

[1,2,3] *CIDETEC, Basque Research and Technology Alliance (BRTA), Po. Miramón 196, 20014 Donostia-San Sebastián, Spain*

zzhang@cidetec.es
marrinda@cidetec.es
jonperez@cidetec.es

ABSTRACT

The world is experimenting a decarbonization process, mainly through lithium-ion-based solutions. Nonetheless, catastrophic events have negatively affected the social acceptance of lithium-ion-based solutions. One of the most interesting projects regarding catastrophic event prevention is the internal short-circuit detection. This paper proposes to detect it using different machine-learning algorithms such as random forest and combination of random forest with neural network-based algorithms through time-instant classification and historical feature classification. The hyper-parameters have been optimized through grid-search. The selected algorithms have been trained thanks to synthetically generated data using a first-order electrical equivalent circuit model. The performance of the generated models has been verified and compared thanks to testing and validation data sets taken from the synthetically generated data. Afterward, the most accurate internal short circuit detection algorithm was selected and validated through laboratory-level data. The selected cell in this study is SLPB526495HE, a pouch cell of 3.7Ah. The generated data are time series of voltage and current, which are the variables that will be available in a real application. The results demonstrate an accuracy above 90% in detecting an internal short circuit in the most interesting cases. The validation with laboratory data has shown that an accuracy of 90% can be achieved. This paper provides learned lessons on the process of developing the internal short circuit detection machine-learning model, highlighting the potential they possess to detect accurately internal short circuits.

1. INTRODUCTION

Lithium-ion batteries have been acclaimed for their high energy density, low self-discharge rates, and environmental compatibility since a decade ago (Diouf & Pode, 2015). This battery technology has emerged as a key component in global decarbonization strategies, finding extensive application in diverse energy storage systems such as electric vehicles and smart grids (Zubi et al., 2018). Despite their notable advantages, safety concerns, particularly those stemming from internal short circuits (ISC) leading to thermal runaway, remain a primary impediment to their broader adoption(Zhan et al., 2023). Such incidents can result in battery fires or even explosions, leading to grave consequences. Thermal runaway is often initiated by ISC events (Ren et al., 2021), and the detection of such events poses significant challenges, especially during their incipient stages.

ISC faults often begin with mild severity, starting with high resistance values, which decrease as the fault progresses. In this process, the voltage, current, and State of Charge (SoC) of normal batteries and those with varying ISC resistance values exhibit similar characteristics during charging and discharging processes. This similarity significantly complicates the diagnosis of early-stage ISC faults, leading to researchers to apply data-driven algorithms (Zhang et al., 2021). The application of data-driven or machine learning (ML) algorithms in ISC fault detection can be categorized further into two main types: unsupervised and supervised learning.

Unsupervised learning methods train fault detection models using data generated during the charging and discharging processes of normal batteries. These methods identify potential anomalies by defining deviation measures between normal and abnormal data, deciphering the standard patterns within the data, and employing specific decision rules. Typical algorithms include Support Vector Machine (SVM) (Chatterjee et al., 2023), Relevance Vector Machine (RVM) (Xie et al., 2020), Kernel Principal Component Analysis (KPCA) (Schmid & Endisch, 2022), and Isolated Forest (Cheng et al., 2023), which have demonstrated effective anomaly detection capabilities in various scenarios. Nonetheless, unsupervised learning algorithms have their limitations. Particularly when the abnormal data closely resemble the normal data with no significant distributional

differences, as it is our case. These methods may struggle to distinguish between them accurately. This is why, it is challenging for unsupervised learning algorithms to detect these subtle anomalies.

Supervised learning methods utilize pre-labeled datasets for training, differentiating like this battery performance data under normal operations and abnormal conditions in the training process. Through training, these models are adept at distinguishing between normal battery behavior and potential fault signals. Typical algorithms include Random Forest (RF) (naha et al., 2020), Convolutional Neural Network (CNN) (Yang et al., 2022) and Long Short-Term Memory (LSTM) (Wang et al., 2023). Nonetheless, it is not clear which should be the one to be applied. In light of this, our research aims to explore and apply various supervised learning algorithms to help researchers find the most suitable algorithms for ISC detection.

This paper proposes the development and comparison of ISC detection supervised learning algorithms, enhancing the detection capabilities at early stages of ISC. This study seeks to provide an ISC detection algorithm selection background to fellow researchers and boost the reliability and safety of lithium-ion batteries in operational contexts.

This paper is structured as follows. The data generation is detailed in section 2. The ISC detection methods are described in section 3. The hyperparameter tuning process undergone in this paper is placed in section 4. The results are shown in section 5. The discussion is done in section 6 and the conclusions are drawn in section 7.

2. DATA GENERATION

The selected battery is SLPB526495HE. The synthetically generated data has been generated with a first-order equivalent electric model. The experimental data has been generated in laboratory testing facilities. During the training and testing phases, only virtual datasets were utilized to develop the models. In the validation phase, experimental datasets were additionally incorporated. This approach was adopted to evaluate the performance of the models trained on virtual datasets in real-world scenarios.

This study focuses on charging data. In practical scenarios, battery discharging conditions are highly complex, whereas the charging scenarios are relatively monotonous. Therefore, we chose charging data to train the model for detecting ISC anomalies during the charging phase.

The operational conditions of the generated data are the same for the synthetically generated one and the one generated through laboratory tests: an ambient temperature of 25°C and a charge process at constant charge mode from 1% SoC to the maximum voltage value.

The extracted feature data during this process included the battery voltage and the battery's current voltage increment relative to its voltage before charging an amount equivalent to 1% of its nominal capacity, denominated as voltage difference, see Eq. (1).

$$V_{diff\_k} = V_k - V_{k-1\%c} \quad (1)$$

2.1. Virtual dataset

The virtual data set used for model training and testing was generated by a first-order electric equivalent circuit model (Arrinda, Oyarbide, Macicior, Muxika, et al., 2021) for the SLPB526495HE battery, as shown in Figure 1.

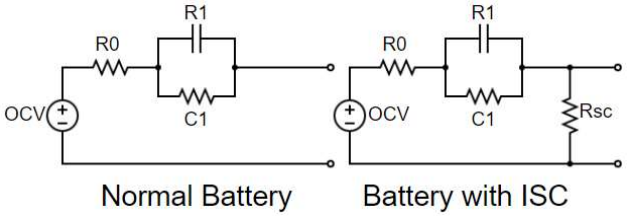

Figure 1: Equivalent electric circuit models.

The parameters of the model were obtained by conducting specific modeling tests on the SLPB526495HE battery, see Figure 2.

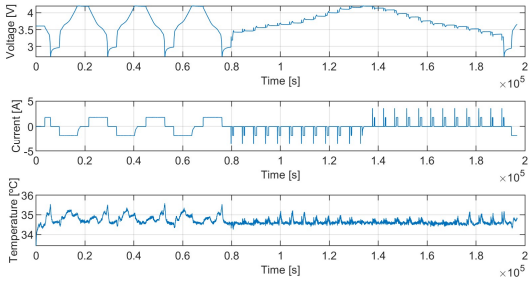

Figure 2: Modeling tests. A capacity and pulse based impedance characterization and OCV characterization test is performed at controlled ambient temperature.

The built model was run to obtain data from a normal battery and a battery with different level of ISC faults (a total of 21 stages): 5Ω, 50Ω, 100Ω, 150Ω, 200Ω, 250Ω, 300Ω, 350Ω, 400Ω, 450Ω, 500Ω, 550Ω, 600Ω, 650Ω, 700Ω, 750Ω, 800Ω, 850Ω, 900Ω, 950Ω and 1kΩ.

2.2. Experimental dataset

The experimental Data Set used for the model validation was generated in the laboratory by cycling the cell with constant current charging tests from 1% SOC to maximum voltage value, see Figure 3. The reference test has been tested only with the cell. The ISC has been emulated by connecting an external bleed resistor of 10 Ω and performing the charge.

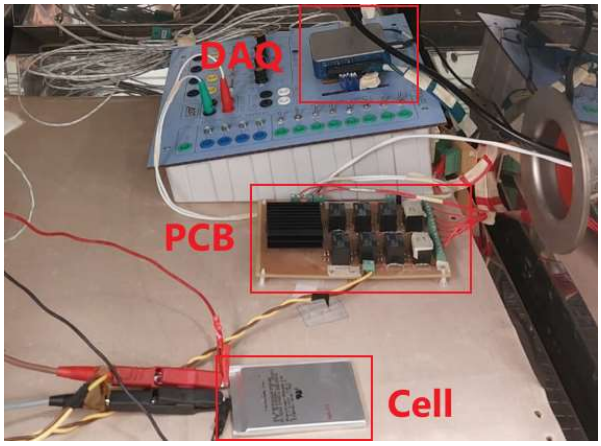

Figure 3: Experimental setup. Within a controlled temperature chamber, a battery cell is interfaced with a Printed Circuit Board (PCB), which incorporates various resistors to simulate an Internal Short Circuit (ISC) phenomenon. This setup is further connected to a Data Acquisition System (DAQ) for comprehensive data collection and monitoring.

3. DETECTION METHODS

This paper presents and compares various ISC detection solutions based on different ML models. These solutions can be divided into two main categories according to the data types they utilize: the Instantaneous Feature-based Method and the Historical Feature-based Method.

3.1. Instantaneous Feature-based Method

As illustrated in Figure 4, the ISC detection solutions under the Instantaneous Feature-based Method determine whether the battery is in a normal state or experiencing an ISC anomaly by analyzing the feature data at every single moment and treating it as a binary classification task at every moment. The ISC detection solutions proposed in this paper that are under the Instantaneous Feature-based Method category are the RF solution and the RF combined with Multilayer Perceptron (RF+MLP) solution.

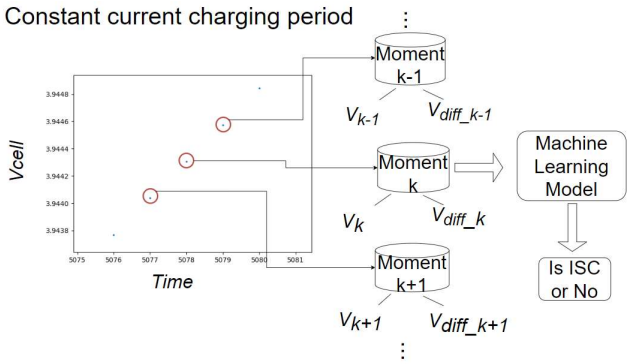

Figure 4: Instant-based methods' main concept diagram. Instant-based methods treat each moment k as an independent data point, characterized by two features: the cell voltage at time k (Vk) and the voltage difference (Vdiff_k) at time k. A machine learning model is then employed to classify each time point as either a normal or an ISC label.

3.1.1. Data treatment

A dataset for the charging process on a battery without an ISC was gathered with no ISC labels. A dataset of battery charging data that reflects the 21 stages of ISC fault conditions has been generated with ISC anomaly labels. The dataset with ISC anomaly labels has a significantly higher volume of data compared to the one with no ISC labels. To address this issue, a down-sampling method was adopted to balance the label distribution in the training dataset.

The specific down-sampling process involves randomly selecting data from the dataset with ISC anomaly labels. The total number of data points is the same in both data sets, 4000. The data with ISC anomaly labels is evenly taken from all the simulated cases. This method ensures the consistency of the total volume of ISC anomaly data with normal data and guarantees a balanced sampling quantity of different stages of ISC anomaly data generated from different short-circuit resistance values.

The final step in data handling involves splitting the dataset obtained through down-sampling into a training set and a test set based on an 80% to 20% ratio. This approach allows the model to train on a substantial portion of the data while retaining a separate subset for evaluation, ensuring that the model's performance can be accurately assessed.

3.1.2. Random Forest (RF)

The RF classifier, as a widely used ML model for classification tasks, leverages ensemble learning techniques to enhance the accuracy and stability of predictions. This model employs bootstrapping to draw multiple subsets of samples with replacements from the original training dataset and randomly selects subsets of features during the construction of each decision tree. In classification tasks, RF makes the final decision by aggregating the predictions from all its decision trees, adopting the class supported by the majority of the trees as the prediction outcome.

3.1.3. Randon Forest with Multilayer Perceptron (RF+MLP)

The RF+MLP combines the RF and MLP to detect the presence of ISC phenomena in batteries. The main workflow consists of training first a RF classifier to be used to predict the data. Then, the prediction results from each decision tree within the RF classifier are used as new input features of the MLP. This approach aims to leverage the MLP to learn the relationships between decision trees, thereby enhancing the

model's ability to distinguish between data with ISC and data without ISC.

3.2. Historical Feature-based Method

The historical feature-based method for detecting ISC events utilizes a sliding window technique. Starting from the (n+1)th time point, it combines the feature data of that moment with the feature data from the preceding n moments to construct a time series window. The window then slides forward, step by step, continuing this process to generate a series of time series window data, see Figure 5. Subsequently, deep learning models specifically designed for time series classification are applied to distinguish between ISC event data and no ISC data.

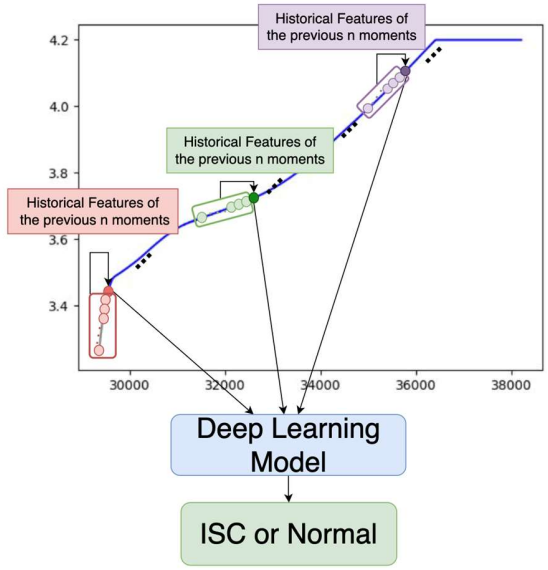

Figure 5: Historical feature-based methods' main concept diagram.

3.2.1. Data treatment

A series of data preprocessing steps are necessary. Initially, the sliding window technique is applied to transform the charging data of batteries without ISC and the charging data of batteries with ISC anomaly using different short-circuit resistance values. Starting from the (n+1)th moment, the feature data of each moment and its preceding n moments are combined to form individual 3x(n+1) dimensional time series windows. Here, 3 represents the number of features:

- The battery voltage at each moment.
- The voltage difference or the voltage increment of the battery at each moment relative to its voltage before charging an amount equivalent to 1% of its nominal capacity.
- The probability that the current moment might correspond to ISC data as determined by the Random Forest classifier analyzing the current battery voltage at each moment and the voltage difference.

Subsequently, down-sampling of ISC time series window data is performed as in the data treatment performed for the instant-based methods to balance the data label distribution and prevent data bias issues during the training process.

Unlike the RF algorithm, neural network models typically require data normalization prior to training. This normalization accelerates model convergence, prevents issues with vanishing or exploding gradients, and enhances the model's generalization capability to new data. Common data normalization methods include the min-max normalization and the Z-score normalization (Patro & sahu, 2015).

The min-max normalization method adjusts the scale of the data so that all features have values ranging between 0 and 1. Specifically, for each feature, this is achieved by subtracting the minimum value of that feature from each value, then dividing by the difference between the maximum and minimum values of that feature, Eq. (2).

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

The Z-score normalization, also known as standard score normalization, normalizes the data by subtracting the mean of each feature from its values and then dividing by its standard deviation, resulting in a dataset with a mean of 0 and a standard deviation of 1, Eq. (3).

$$x_{norm} = \frac{x - \mu}{\sigma} \tag{3}$$

Beyond the normalization methods, the choice of normalization strategy is crucial and can be based on one of the considered normalization processes: normalization-by-moment, normalization-by-feature, and normalization-by-window.

The normalization-by-moment strategy involves normalizing the values of all features at each specific moment. It treats each point in time independently, adjusting the features across all samples at that particular moment to conform to the chosen normalization scale. This approach is useful when the relative magnitudes of features at each moment are important for the model to recognize patterns over time.

The normalization-by-feature strategy operates on each feature across all moments. It normalizes the values of a single feature over the entire dataset, ensuring that the feature's values are on the same scale across all time points. This is particularly beneficial when you want the model to understand the behavior of each feature independently across time, emphasizing the feature's overall distribution without the influence of varying scales.

The normalization-by-window approach treats all the data within a sliding window as a whole for normalization purposes. Each window is normalized independently, meaning that the scale of the features is adjusted within the context of that window. This strategy is useful when the relationship between features within each window is critical to identifying patterns, and it aims to preserve the internal dynamics of each time window.

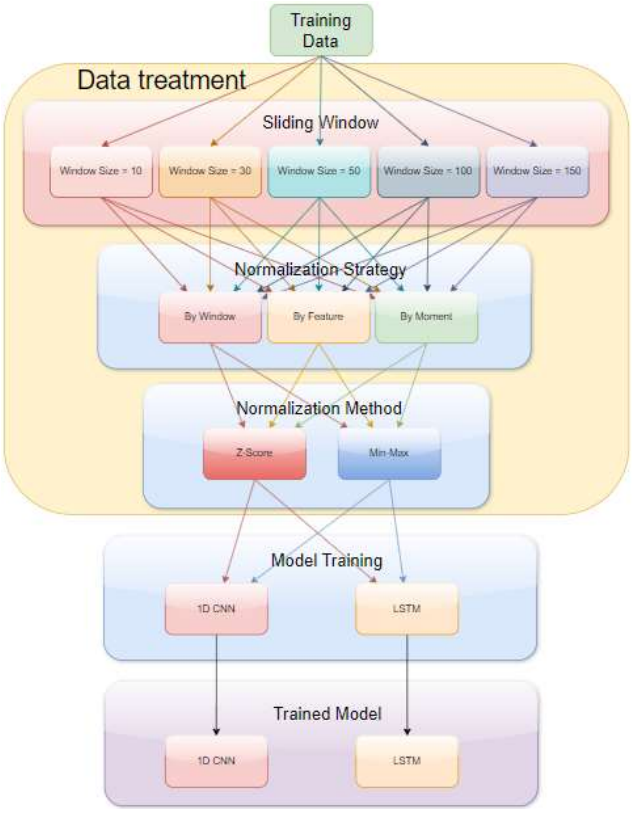

Figure 6: Data treatment processing selection diagram.

As a result, six distinct data preprocessing schemes have been developed based on the aforementioned. To identify the most suitable data processing approach for various models and to find the optimal sliding window size, an experimental workflow was designed as follows (Figure 6):

- Experiment with window sizes equals to 10, 30, 50, 100, 150 are evaluated.
- The six different data preprocessing schemes are applied to each window size experiment, resulting in 30 different data processing configurations.
- Train 1D CNN and LSTM networks using the processed data.
- Evaluate the performance of these models through cross-validation to determine the most suitable data processing method and window size for each model type.

3.2.2. Random Forest with Convolutional Neural Network (RF+CNN)

1D CNN algorithms are frequently employed for processing sequential data, such as time series data. A 1D CNN processes input data through a series of specific layers to extract useful features for classification or other tasks. The fundamental architecture of a simple 1D CNN consists of an input layer, convolutional layer, activation function, pooling layer, fully connected layer, and output layer. In the context of time series classification tasks, the input layer initially receives the raw data. This is followed by the convolutional layer, where multiple kernels slide across all features of the data to perform convolution operations and extract features, which are subsequently subjected to an activation function. The pooling layer then reduces the dimensionality of the feature maps, decreasing the volume of data that needs to be processed. Finally, the fully connected layer and the output layer classify the previously extracted features, producing the final outcome.

3.2.3. Random Forest with Long short-term memory (RF+LSTM)

LSTM networks are a specialized type of Recurrent Neural Networks (RNNs) particularly suited for classifying, processing, and predicting based on time series data. LSTMs are adept at addressing issues of vanishing or exploding gradients, which are common with traditional RNNs. The basic structure of an LSTM includes an input layer, LSTM layer, hidden layers, and an output layer. Within the LSTM layer, each LSTM unit contains several key components: Cell State, Input Gate, Forget Gate, Output Gate, and Hidden State. When LSTMs are employed for time series data classification tasks, data is initially decomposed into individual time steps through the Input Layer and then fed into the LSTM layer. This layer captures long-term and short-term relationships within the time series data by maintaining, ignoring, or updating information through an internal state and three gate structures. The output from the LSTM layer is then passed to one or more Hidden Layers for further feature extraction, with the final classification result being produced by the output layer.

4. HYPERPARAMETER TUNING

Hyperparameter tuning plays a crucial role in the training of ML and deep learning models, as the choice of hyperparameters directly affects the performance, learning capability, and generalization ability of the model. During the training of various models mentioned before, such as RF, MLP, 1D CNNs, and LSTMs, experimenting with multiple combinations of hyperparameters is an effective method to find the relatively optimal hyperparameter settings.

4.1. RF Classifier Hyperparameter Tuning

The use of the grid search through "GridSearchCV" tool from the Scikit-learn python's library is proposed to systematically explore and optimize the hyperparameter settings of the RF classifier (Arrinda, Oyarbide, Macicior, & Muxika, 2021). Initially, we defined a search space containing various combinations of hyperparameters, including the number of decision trees, the maximum depth of the trees, the minimum number of samples required to split an internal node, the minimum number of samples required at a leaf node, and whether bootstrap sampling is used.

"GridSearchCV" tested each of the 2,400 different hyperparameter combinations defined in our search space and employed cross-validation to comprehensively assess the performance of each combination. The training dataset was divided into five subsets, with one subset being used as the validation set to evaluate the model and the remaining four subsets for training. The performance of each combination was assessed based on the average results of these five validations.

The optimal combination of hyperparameters for the model was finally identified by analyzing and comparing. After determining the best hyperparameters, these parameters were used with the full training dataset to conduct the final training of the random forest classifier, ensuring the model achieved optimal predictive performance.

4.2. Neural Networks Hyperparameter Tuning

The grid search method used for identifying the optimal hyperparameters of the RF Classifier was considered unsuitable for neural networks due to time cost concerns. Hence, the "Keras Tuner" Python's library is proposed to perform hyperparameter optimization through random search of neural network based models. Similar to "GridSearchCV", before starting the random search, a search space for each model is defined. However, the distinct feature of the random search method provided by "Keras Tuner" is that it does not attempt every possible combination of hyperparameters. Instead, it randomly selects n combinations of hyperparameters from the defined search space to experiment with. A key advantage of this approach is its ability to significantly reduce the search time while still maintaining the possibility of discovering well-performing hyperparameter sets.

5. RESULTS

The most suitable data processing approach and the most optimal hyperparameters for each model are shown in Figure 7. Each trained model has been validated both by virtual datasets and experimental datasets.

RandomForest
Data processing

- None

Hyperparameter

- Num of estimators: 500
- Bootstrap: True
- Max_depth: None
- Min_samples_leaf: 1
- Min_samples_split: 2

CNN
Data processing

- Sliding Window Size: 100
- Normalization Strategy: By Window
- Normalization Method: Z-Score

Hyperparameter

- Num of Module: 1
- Num of Convolutional Layer: 1
- Num of kernels: 96
- Kernel size: 5
- Activation function: relu
- L1 regularizer: No
- L2 regularizer: 0.00011731100894294
- MaxPooling: Yes
- Pool size: 3
- Dropout: No
- Num of Full Connected layer: 1
- Num of units: 256
- Activation function: relu
- L1 regularizer: 0.00020918054059229
- L2 regularizer: 0.00102808799175713
- Dropout: Yes
- Dropout rate: 0.3
- Learning rate: 0.0003967960974673

MLP
Data processing

- None

Hyperparameter

- Num of Hidden Layer: 1
- Dropout: Yes
- Dropout rate : 0.5
- Num of hidden units: 64
- Activation function: relu
- L1 Regularizer: 2.5099081541393e-05
- L2 Regularizer: 0.0004041274348191
- Learning rate: 0.000494137845717255

LSTM
Data processing

- Sliding Window Size: 30
- Normalization Strategy: By Feature
- Normalization Method: Min-Max

Hyperparameter

- Num of LSTM units: 256
- Dropout: No
- Recurrent dropout: Yes
- Recurrent dropout rate: 0.3
- L1 regularizer: No
- L2 regularizer: 2.5732638229762e-05
- Num of Hidden Layer: 1
- Num of Hidden units: 160
- Activation function: tanh
- L1 regularizer: 0.000423110271510091
- L2 regularizer: 8.380137024491e-05
- Dropout: No

Figure 7 The most suitable data processing approach and the most optimal hyperparameters for each model.

5.1. Validation with virtual dataset

The virtual dataset comprises one charging data set generated by a normal battery electric equivalent circuit model and other two charging data sets generated respectively by ISC battery electric equivalent circuit models, characterized by short-circuit resistances of 10Ω and 510Ω, respectively.

The validation results of models of the instantaneous feature-based method are shown in Figure 8, whereas the results for models utilizing historical feature-based methods are illustrated in Figure 9. These figures illustrate the temporal variation of cell voltage (Vcell) during the constant current charging process of normal batteries and batteries experiencing ISC faults with short circuit resistances of 10 ohms and 510 ohms. Furthermore, the figures depict the fault detection outcomes at each time point during the charging process, as predicted by the RF model and the RF model integrated with MLP. The prediction outcomes are marked in blue and red, indicating correct predictions and incorrect predictions, respectively.

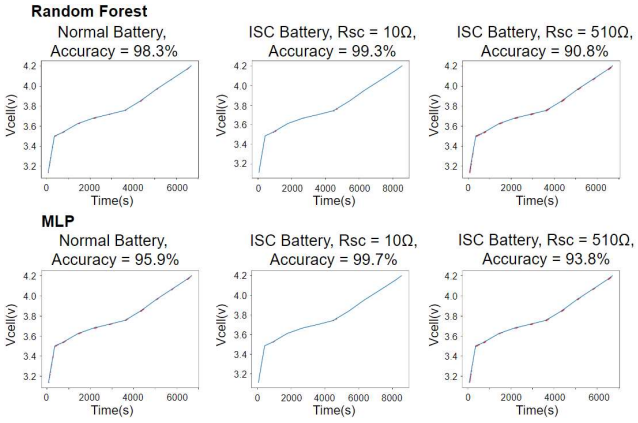

Figure 8: Validation Results of models of Instantaneous Feature-based method with virtual dataset. The blue line is the Vcell vs time, and the red points represent the moments where the prediction is wrong.

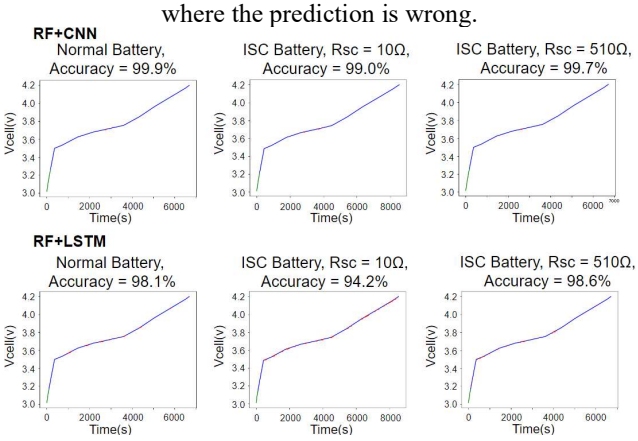

Figure 9: Validation Results of models of Historical Feature-based Methods with virtual dataset. The green part represents the moments when the charging amount is still less than 1% nominal capacity.

5.2. Validation with experimental dataset

The experimental dataset consists of two datasets. One of them is the charging dataset of a real battery without any fault. The other one is the charging dataset from the same battery connected with a 10Ω short-circuit resistance emulating an ISC condition.

The battery model used to generate the virtual data assumed a state of health (SoH) of 100%, whereas the battery utilized for the experimental dataset did not have the exact same SoH. Hence, Eq. (4) should be employed for calculating the voltage difference of the experimental dataset.

$$V_{diff\_\exp\_k} = (V_k - V_{k-1\%c}) \cdot SoH_{exp} \qquad (4)$$

Figure 10 and Figure 11 illustrate the validation results of Instantaneous Feature-based method models and Historical Feature-based method models respectively. These figures also use red and blue markers to denote the accuracy of predictions in relation to the actual conditions, where red indicates a mismatch between predicted and actual label, and blue signifies correct predictions.

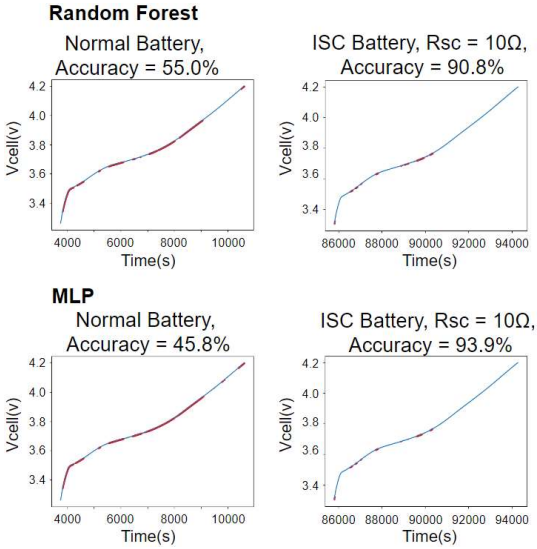

Figure 10: Validation Results of models of Instantaneous Feature-based method with experimental dataset.

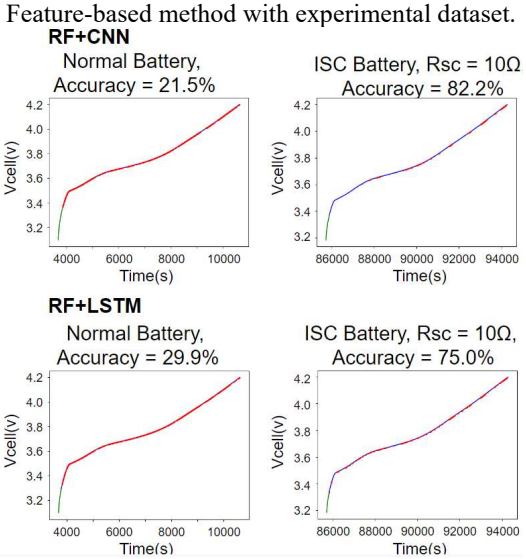

Figure 11: Validation Results of models of Historical Feature-based Methods with virtual dataset.

6. DISCUSSION

This study introduced four innovative methods for detecting ISC faults, all of which demonstrated a prediction accuracy above 90% (and 8 of 12 close to 99%) during the validation process with virtual data, including the identification of normal data and ISC data with 10Ω and 510Ω short-circuit resistances. The validation results are presented in Table1.

Table 1. Validation Accuracy of models on Different Validation Dataset.

Dataset	RF	RF+MLP	RF+CNN	RF+LSTM
Normal (Virtual)	98.3%	95.9%	99.9%	98.1%
Normal (Experimental)	55.0%	45.8%	21.5%	29.9%
10 Ω (Virtual)	99.3%	99.7%	99.0%	94.2%
10 Ω (Experimental)	90.8%	93.9%	82.2%	75.0%
510 Ω (Virtual)	90.8%	93.8%	99.7%	98.6%

The RF method, being the earliest and simplest instantaneous feature-based method, shows excellent performance in distinguishing between ISC and normal data within the virtual dataset. It achieved an accuracy of 98.3% for normal data and 99.3% for severe ISC detection (lower short-circuit resistance values), though its accuracy slightly decreased to 90.8% for early-stage ISC detection (higher short-circuit resistance values). To enhance the model's accuracy in predicting high resistance value ISC conditions, three other detection methods based on model stacking were proposed, using RF as the base model whose output serves as additional input features for other models.

The combination of RF with MLP slightly improved the accuracy for high resistance value ISC conditions to 93.8%, at the cost of reduced accuracy for normal battery data predictions, which fell to 95.9%. This is a consequence of the weight distribution of RF's estimators provided by the MLP.

The combination of RF with CNN performed best in virtual data validation. It achieved a 1.6% increase in accuracy for normal data predictions, reaching 99.9%. Moreover, this method improved the detection accuracy for 510Ω short-circuit resistance ISC from 90.8% to 99.7%. Nonetheless, its accuracy for detecting lower resistance ISC slightly fell by 0.3% to 99.0% compared to using RF alone.

The combination of RF with LSTM networks increased the prediction accuracy for high resistance ISC detection to 98.6% while maintaining the accuracy for normal data predictions at 98.1%. Nevertheless, lower resistance ISC detection slightly decreased to 94.2%.

The validation process with experimental data has shown a decrease in accuracy. All models experienced a significant performance decline in the validation with the experimental dataset, especially in predicting normal data. The instantaneous feature-based methods outperformed the historical feature-based methods overall in the experimental dataset. RF and RF+MLP maintained over 90% accuracy in predicting low resistance ISC conditions, but their accuracy in predicting normal data dropped to 55% and 45.8%, respectively. The HFM-based methods had less than 30% accuracy in predicting normal data, and their detection accuracy for ISC faults did not reach 80%.

This performance drop could be attributed to overfitting on virtual normal data, as the virtual normal battery could generate only a single set of normal battery charging data, and to discrepancies between data generated by equivalent circuit models and real data, preventing the RF model from accurately learning subtle changes in real conditions. The performance of model stacking methods was impacted by the base model RF's performance; if RF could not provide accurate predictions, the overall performance of the stacked models was negatively affected.

In real applications, as the performance gap between electric equivalent circuit model and actual battery increases, and therefore, the negative effect of stacked models will be amplified, making the RF model a preferred choice for ISC detection algorithms in the absence of real data. Nonetheless, if the performance gap could be resolved, or if real-life data could be used to further train the stacked models, models like RF+CNN could achieve significantly higher accuracy levels in early-stage ISC detection compared to the RF model alone.

7. CONCLUSION

This paper developed four methods for detecting ISC faults. Data is obtained from simulations and experiments at lab level. The ISC fault detection methods are trained using the virtual data. The normalization method, normalization strategy, are performed by using an exhaustive method and hyperparameter tuning is done by grid-search and random-search. After the training and hyperparameter tuning, these methods have been evaluated respectively by conducting validations and performance comparisons with both the virtual and experimental datasets.

The validation with virtual data shows that the historical feature-based method combining RF and CNN demonstrated superior performance. However, the limitations of virtual data became apparent during the validation with experimental data. The base model, the RF model, fails to achieve satisfactory prediction results on the experimental dataset. It suffers a drop of accuracy from 98.3% to 55% in describing data without ISC. This limitation further impacted the overall performance of the RF+CNN methods in the experimental data validation, having higher accuracy drops than the ones observed in RF.

To address this issue in future research, one of the potential approaches could be integrating digital twin and cloud computing technologies. Digital twins can facilitate the collection of extensive real-world data to refine the model, while cloud computing, combined with the gathered real-world data, can enable continuous learning for the model.

This method ensures the model's adaptability to real-world data. Moreover, as a battery's SoH gradually declines, affecting its charging data throughout its use, continuous learning can also allow the model to adjust to these changes. This synergy of multiple technologies considerably augments the flexibility and universality of ISC detection systems, equipping them to accommodate a wider array of scenarios and conditions.

ACKNOWLEDGEMENT

This project has received funding from the European Union's Horizon Europeresearch and innovation programme under grant agreement No 101103821. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

NOMENCLATURE

SoH_{exp}	*battery's SoH in the experimental dataset*
V_k	voltage at the kth moment
$V_{k\text{-}1\%c}$	voltage at the moment before charging 1% of the nominal capacity prior to the k^{th} moment.
V_{diff_k}	voltage difference of virtual dataset between k^{th} and k-1%c^{th} moment
$V_{diff\_exp\_k}$	voltage difference of experimental dataset between k^{th} and k-1%c^{th} moment
x	unnormalized original value
x_{min}	minimum value among dataset
x_{max}	maximum value among dataset
x_{norm}	normalized value
μ	mean value of dataset
σ	standard deviation of dataset

REFERENCES

Arrinda, M., Oyarbide, M., Macicior, H., & Muxika, E. (2021). Unified Evaluation Framework for Stochastic Algorithms Applied to Remaining Useful Life Prognosis Problems. *Batteries*, *7*(35), 1–27. https://doi.org/https://doi.org/10.3390/batteries7020035

Arrinda, M., Oyarbide, M., Macicior, H., Muxika, E., Popp, H., Jahn, M., Ganev, B., & Cendoya, I. (2021). Application Dependent End-of-Life Threshold Definition Methodology for Batteries in Electric Vehicles. *Batteries*, *7*(1), 12. https://doi.org/10.3390/batteries7010012

Chatterjee, S., Kumar Gatla, R., Sinha, P., Jena, C., Kundu, S., Panda, B., Nanda, L., & Pradhan, A. (2023). Fault detection of a Li-ion battery using SVM based machine learning and unscented Kalman filter. *Materials Today: Proceedings*, *74*. https://doi.org/10.1016/j.matpr.2022.10.279

Cheng, X., Li, X., & Ma, X. (2023). A method for battery fault diagnosis and early warning combining isolated forest algorithm and sliding window. *Energy Science and Engineering*, *11*(12), 4493–4504. https://doi.org/10.1002/ESE3.1593

Diouf, B., & Pode, R. (2015). Potential of lithium-ion batteries in renewable energy. In *Renewable Energy* (Vol. 76). https://doi.org/10.1016/j.renene.2014.11.058

naha, A., Khandelwal, A., Agarwal, S., tagade, piyush, Hariharan, K. S., Kaushik, A., Yadu, A., Mayya Kolake, S., Han, S., & oh, B. (2020). *internal short circuit detection in Li-ion batteries using supervised machine learning*. https://doi.org/10.1038/s41598-020-58021-7

Patro, S. G. K., & sahu, K. K. (2015). Normalization: A Preprocessing Stage. *IARJSET*. https://doi.org/10.17148/iarjset.2015.2305

Ren, D., Feng, X., Liu, L., Hsu, H., Lu, L., Wang, L., He, X., & Ouyang, M. (2021). Investigating the relationship between internal short circuit and thermal runaway of lithium-ion batteries under thermal abuse condition. *Energy Storage Materials*, *34*. https://doi.org/10.1016/j.ensm.2020.10.020

Schmid, M., & Endisch, C. (2022). Online diagnosis of soft internal short circuits in series-connected battery packs using modified kernel principal component analysis. *Journal of Energy Storage*, *53*. https://doi.org/10.1016/j.est.2022.104815

Wang, H., Nie, J., He, Z., Gao, M., Song, W., & Dong, Z. (2023). A reconstruction-based model with transformer and long short-term memory for internal short circuit detection in battery packs. *Energy Reports*, *9*. https://doi.org/10.1016/j.egyr.2023.01.092

Xie, J., Zhang, L., Yao, T., & Li, Z. (2020). Quantitative diagnosis of internal short circuit for cylindrical li-ion batteries based on multiclass relevance vector machine. *Journal of Energy Storage*, *32*. https://doi.org/10.1016/j.est.2020.101957

Yang, N., Song, Z., Amini, M. R., & Hofmann, H. (2022). Internal Short Circuit Detection for Parallel-Connected Battery Cells Using Convolutional Neural Network. *Automotive Innovation*, *5*(2). https://doi.org/10.1007/s42154-022-00180-6

Zhan, J., Deng, Y., Ren, J., Gao, Y., Liu, Y., Rao, S., Li, W., & Gao, Z. (2023). Cell Design for Improving Low-Temperature Performance of Lithium-Ion Batteries for Electric Vehicles. In *Batteries* (Vol. 9, Issue 7). https://doi.org/10.3390/batteries9070373

Zhang, G., Wei, X., Tang, X., Zhu, J., Chen, S., & Dai, H. (2021). Internal short circuit mechanisms, experimental approaches and detection methods of lithium-ion batteries for electric vehicles: A review. In *Renewable and Sustainable Energy Reviews* (Vol. 141). https://doi.org/10.1016/j.rser.2021.110790

Zubi, G., Dufo-López, R., Carvalho, M., & Pasaoglu, G. (2018). The lithium-ion battery: State of the art and future perspectives. In *Renewable and Sustainable Energy Reviews* (Vol. 89). https://doi.org/10.1016/j.rser.2018.03.002

BIOGRAPHIES

Zihong Zhang received the B.S. degree in electrical engineering and automatization in 2019 at Shanghai Institute of Technology, Shanghai, China. He obtained M.S. in computational methods in sciences at University of Navarra, Pamplona, Navarra, Spain in 2021. In 2023, he completed the second M.S. in Artificial Intelligence at Tecnun, University of Navarra, Donostia, Gipuzkoa, Spain. His main responsibility in CIDETEC concentrates on development machine learning and deep learning models for applications in fault diagnosis, predictive maintenance, digital twins for lithium-ion batteries.

Mikel Arrinda received the B.S. degree in industrial electronic engineering in 2012 at MU, Mondragon, Basque Country, Spain. In 2013 completed his studies with a M.S. in integration of renewable energy sources into the electricity grid at EHU, Bilbao, Basque Country, Spain. In 2020, he obtained his Phd degree in apply engineering at CIDETEC institute for Energy Storage. His research interest is focused on lithium-ion batteries in terms of electric-thermal-aging modeling, SoX algorithms, thermal control strategies, digital twin applications, AI implementation and safety measures.

Jon Perez received the B.S. degree in electrical engineering in 2020 at Tecnun University of Navarre, Donostia – San Sebastian, Basque Country, Spain. In 2022, he received the M.Sc degree in Energy and Power Electronics from Mondragon University, Mondragon, Basque Country, Spain. From 2022 onwards, he is pursuing a PhD degree in applied engineering at CIDETEC institute for Energy Storage. His research interests are focused on lithium-ion batteries aging detection, SoX algorithms and early detection of Thermal Runaway and other failures.