Real-Time Hazard Detection in Battery Systems Using Random Forest Classifiers and SMOTE

Mohamed A. Dini, Sium Bin Noor, Dong-Seong Kim, Jae Min Lee
Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea
(mohamedabubakar407)@gmail.com (siumbinmoor, dskim, ljmpaul)@kumoh.ac.kr

***Abstract*—For battery-powered systems to be more dependable and long-lasting, effective battery management is essential. This study uses the NASA battery dataset to develop a machine learning-based system for forecasting dangerous battery conditions. In order to anticipate possible failures before they happen, the system analyzes charge and discharge cycle data, paying special attention to anomalies related to temperature. Battery states are classified as safe or dangerous using a Random Forest Classifier, and class imbalance is addressed using the Synthetic Minority Over-sampling Technique (SMOTE). The suggested system provides precise, real-time forecasts, enhancing battery longevity and safety across a range of industrialapplications.**

***Index Terms*—Battery Health, Class Imbalance, Dangerous Conditions Prediction, Machine Learning, NASA Battery Dataset, Random Forest Classifier, SMOTE, Battery Management System**

I. Introduction

For battery-powered systems to be dependable and long-lasting, battery management is essential. To avoid failures and improve safety, it is essential to accurately predict dangerous events, such as temperature anomalies. By detecting possible hazards early on, machine learning techniques—especially when used on large datasets like the NASA battery dataset [1], [2]—can greatly enhance battery health monitoring.

Important variables like voltage, current, and temperature are included in the NASA battery dataset and can be used to inform predictive maintenance. Despite advancements in battery health monitoring, there is still a lack of research on using machine learning models to identify temperature-related risks [3], [4]. With an emphasis on temperature-related anomalies, this study suggests a machine learning-based system to analyze charge and discharge cycle data and forecast hazardous battery conditions [5], [6].

In this paper, a Random Forest Classifier-based predictive system for identifying safe and dangerous battery states is presented. By using SMOTE to address class imbalance, the system offers precise predictions in real time. By enabling early detection of hazardous events, the goal is to improve battery safety by prolonging battery lifespan and preventing failures [7], [8].

II. System Design

The proposed system predicts dangerous events in battery charging and discharging cycles using the NASA battery dataset. It involves data preprocessing, feature engineering, and model training to build a predictive model for battery health monitoring.

A. 1. Data Preprocessing

The battery data includes voltage, current, and temperature readings. Key preprocessing steps include:

1) 1.1 Temperature Handling: Missing or non-numeric temperature data T_{measured} is converted to numeric values and filled with the mean temperature T_{mean}:

$$T_{\text{mean}} = \frac{1}{N}\sum_{i=1}^{N} T_i \tag{1}$$

2) 1.2 Temperature Rate Calculation: The rate of temperature change $\frac{dT}{dt}$ is calculated as:

$$\frac{dT}{dt} = T_{i+1} - T_i \tag{2}$$

This feature helps detect rapid temperature changes indicative of hazardous conditions.

3) 1.3 Danger Label Creation: A binary danger label is assigned based on a 45°C threshold. If $T_i > 45°C$, the label is 1 (dangerous), otherwise 0 (safe):

$$\text{danger}_i = \begin{cases} 1 & \text{if } T_i > 45°C \\ 0 & \text{if } T_i \leq 45°C \end{cases} \tag{3}$$

B. 2. Feature Engineering and Data Combination

The features for training include voltage, current, mean temperature, and the temperature change rate:

$$X = \{V_{\text{converted}}, I_{\text{converted}}, T_{\text{mean}}, \frac{dT}{dt}\}$$

These features are extracted from both charge and discharge events and combined into a single dataset.

C. 3. Model Training

A Random Forest Classifier is trained to predict dangerous events. The model minimizes the binary cross-entropy loss:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)\right] \tag{4}$$

SMOTE is applied to handle class imbalance by generating synthetic samples for the minority class.

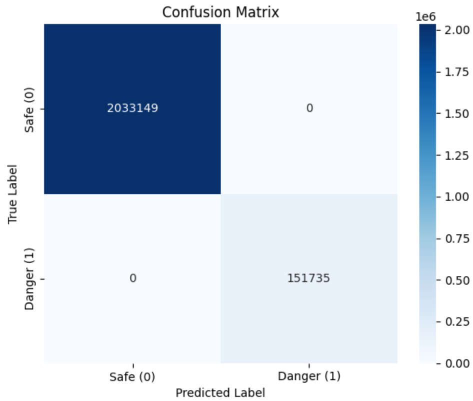

Fig. 1. System flowchart

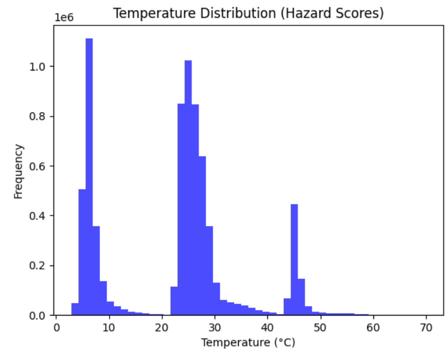

Fig. 2. System flowchart

D. System Flowchart

The system flowchart can be summarized as:

- **Load Data**: Load metadata and charge/discharge data files.
- **Preprocess Data**: Handle missing values, convert units, and calculate temperature change.
- **Feature Engineering**: Extract features like voltage, current, and temperature.
- **Train Model**: Train the Random Forest model with the processed dataset.
- **Predict Danger**: Predict potential dangerous events with the trained model.

III. Performance Evaluation

The performance of the proposed system is evaluated using the confusion matrix and temperature distribution. These metrics offer insights into the model's accuracy and potential overfitting.

A. Confusion Matrix

The confusion matrix, shown in 1 indicates the following:

- True Positives (Dangerous events correctly predicted): 151,735 - True Negatives (Safe events correctly predicted): 2,033,149 - No False Positives or False Negatives

This suggests high accuracy, but the absence of false positives and false negatives raises concerns about possible overfitting, especially given the class imbalance, where safe events far outnumber dangerous ones. The model might have over-learned the safe class, potentially neglecting dangerous events.

B. Temperature Distribution

The temperature distribution, shown in 2, reveals:

- Most temperatures are below 30°C, with few events exceeding 40°C.

- This skewed distribution suggests that the model is mostly trained on safe events, reinforcing the overfitting risk.

IV. Conclusion

This paper presents a machine learning-based system for predicting dangerous battery conditions using the NASA battery dataset. By applying a Random Forest Classifier and addressing class imbalance with SMOTE, the system accurately classifies battery states as safe or dangerous. The proposed system enhances battery health monitoring, providing real-time, reliable predictions to improve safety and longevity. Future work will focus on validating the model on new datasets to ensure generalization and robustness.

V. Acknowledgement

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the IITP grant funded by the Korea government(MSIT) (IITP-2025-RS-2020-II201612, 33%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 33%) and by the MSIT, Korea, under the ITRC support program(IITP-2025-RS-2024-00438430, 34%)

References

[1] J. Nkoju, Y. Moyo, and S. Ngobeni, "Battery management systems: Current status and future directions," *Journal of Battery Technology*, vol. 14, pp. 56–72, 2023.

[2] A. Ahmed, T. Zhang, and Z. Liu, "Predictive maintenance for lithium-ion batteries: A machine learning approach," *IEEE Transactions on Industrial Informatics*, vol. 20, pp. 1054–1062, 2024.

[3] J. Nkoju and Y. Moyo, "Predicting battery life using data from nasa battery dataset," *Journal of Energy Storage*, vol. 32, pp. 85–94, 2024.

[4] R. Zhao, L. Zhang, and J. Wang, "Advanced battery management systems for electric vehicles," *IEEE Transactions on Transportation Electrification*, vol. 12, pp. 224–235, 2023.

[5] Q. Li, L. Xie, and Y. Wu, "Machine learning for battery fault diagnosis: A survey," *Energy Reports*, vol. 9, pp. 1452–1465, 2023.

[6] L. A. C. Ahakonye, C. I. Nwakanma, J. M. Lee, and D.-S. Kim, "Time-efficient deep learning-based energy consumption prediction for smart factory," in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2022, pp. 879–882.

[7] M. Tayeb, M. A. Dini, and J. M. Lee, "Predictive maintenance of batteries in industrial applications," *Journal of Battery Systems*, vol. 28, pp. 65–75, 2024.

[8] M. A. Dini, M. J. A. Shanto, G.-H. Kwon, D.-S. Kim, and T. Jun, "Emotion type recognition scheme using eeg based signals," , pp. 489–490, 2023.