Machine Learning Based Battery Anomaly Detection using Empirical Data

Md Shahriar Nazim
Dept. of Electronics Engineering
Kookmin University
Seoul, South Korea
shahriarnazim45@gmail.com

Yeong Min Jang
Dept. of Electronics Engineering
Kookmin University
Seoul, South Korea
yjang@kookmin.ac.kr

ByungDeok Chung
CEO, ENS. Co. Ltd
Ansan, South Korea
bdchung@ens-km.co.kr

***Abstract*—In the context of energy storage systems (ESS), this work investigates the use of machine learning approaches for anomaly identification utilizing empirical site data. Making advantage of the empirical data gathered from the operational environment, the study concentrates on using precise anomaly detection techniques—mainly the Isolation Forest method. The Isolation forest approach is utilized to detect abnormalities in the empirical data obtained by ESS operations. It is well-known for its effectiveness in locating outliers in datasets. In order to improve the operational dependability and safety of Energy Storage Systems (ESS), this study explores the application of the Isolation Forest technique as a powerful tool for identifying anomalies in the site data. The results of the study show that, Isolation forest can detect anomalies with the accuracy of 99.43%.**

***Index Terms*—Anomaly detection, BMS, Isolation Forest, Local outlier factor, Machine learning.**

I. Introduction

The extensive use of renewable energy sources has completely changed the energy landscape in the modern era of technological growth. The need for energy storage solutions has increased due to the increasing integration of solar, wind, and other renewable resources into the power grid; this has highlighted the importance of battery systems [1]. Simultaneously, the increased dependence on batteries in consumer devices and grid-scale applications highlights how crucial it is to guarantee their dependability and safety. In the early stages of this transformative transition to sustainable energy, battery system performance and safety have become top priorities [2]. Since batteries are essential for storing irregular renewable energy for later use and powering numerous applications, it is necessary that strict safety measures and effective performance optimization be implemented. The identification and management of abnormal circumstances in battery systems has attracted considerable interest as a strategy to enhance both efficiency and security [3]. It is not only safer to avoid possible risks but also more effective and long-lasting when irregularities in battery performance can be quickly detected and addressed. Within this framework, the new approach for abnormal condition detection in battery systems is made up of machine learning-based methodologies. [4] Taking advantage of machine learning algorithms' capabilities presents a viable way to detect, anticipate, and handle anomalies that might compromise battery safety or performance.

A. Literature Review

Recent research has achieved substantial advances in the detection of battery anomalies. Both Zhao et. al. and Bhaskar et.al. [5], [6] suggest data-driven methodologies for detecting anomalies in lead-acid and lithium-ion battery packs. Zhao's study focuses on the use of unsupervised anomaly detection techniques, whereas Bhaskar's approach generates mean-based residuals and evaluates them using principal component analysis utilizing real-time voltage and temperature data. Both investigations show that their strategies are effective at finding anomalies. Li et. al. [7] add to this body of work by presenting a data-driven technique for identifying thermal abnormalities in EV and ES batteries that is resistant to data loss and requires little reference data. These studies show that data-driven approaches have the potential to improve the safety and dependability of battery systems. Haider et. al. [8] proposed a new battery anomaly detection method based on time series clustering. Lee et. al. [9] present an efficient single-model strategy for offline detection of faulty batteries in UPS systems utilizing isolation forest and hyper-parameter adjustment. A multi-model solution also covers online anomaly detection, displaying significant performance in detecting faulty batteries during operation. Diao et. al. [10] explore five data-driven methods to detect early signs of degradation in ongoing reliability tests of lithium-ion batteries. It examines regression models, support vector machines, outlier detection strategies, and sequential probability ratio tests before settling on an ensemble approach because no single method consistently provides the earliest warnings. Based on continuing reliability testing, the suggested technique supports device manufacturers in making warranty, recall, and technical decisions.

B. Research Gap

Even with the large amount of research on battery anomaly detection, there is still a significant research gap regarding improving model accuracy. Even though previous research has made significant improvements to the field, there is still an apparent desire to further refine and increase the accuracy of these models. In order to close the current accuracy gap and improve the dependability and effectiveness of anomaly detection frameworks, this research investigates a new tuned Isolation forest model for anomaly detection within battery systems.

C. Our Approach

This study improves on earlier attempts to identify anomalies in batteries by utilizing Isolation Forest and Local Outlier Factor models to target accuracy issues. In order to increase safety and efficiency in a variety of applications, the study compares these models in order to determine how well they can improve anomaly detection in battery systems.

II. Methodology

The main goal of this research is to use the Local Outlier Factor and Isolation Forest algorithms to improve battery anomaly detection accuracy. For training, actual data from the Battery Management System (BMS) of a 100KW photovoltaic (PV) facility is used. After close inspection, the voltage, current, and temperature data are the most important metrics to identify abnormal situations in the battery system out of all the ones that were gathered. As a result, due to their critical importance in the identification of battery irregularities, these three characteristics are the only ones selected for model training. In the input dataset contains different features such as battery temperatures (T_{batt}), battery terminal voltage (V_{batt}) and battery current (I_{batt}).

Algorithm 1 Machine Learning-based Battery Anomaly Detection

Require: $V_{\text{batt}}, I_{\text{batt}}, T_{\text{batt}}$
Ensure: Anomaly conditions for voltage, current, and temperature
function main():
$df \leftarrow$ Data from PV plant $\{df \leftarrow dataframe$ Selecting only voltage, current, temperature data$\}$
$new_{\text{df}} \leftarrow$ New dataframe {Split the data for train and test purposes}
$X_{\text{train}}, X_{\text{test}} \leftarrow$ Split(new_df) {Train model}
$Definemodels \leftarrow Isolationforest, LocalOutlierFactor$
for all model **in** models **do**
  trainModel(model, X_{train})
  Save train model
  test the model using X_{test}
  $comparison\_data \leftarrow$ Accuracy, recall, F1 score
end for

Algorithm 1 starts by gathering information from the Battery Management System (BMS) of the PV plant and naming it df. Then, a new dataset called new_df is created with the sole purpose of including the crucial variables of temperature, voltage, and current that were taken from the original dataset. To make the process of evaluating the model easier, the dataset is then separated into two subsets: one for testing and another for training. The program progressively applies two different anomaly detection algorithms, Isolation Forest and Local Outlier Factor, using a for loop. This loop uses the training dataset to train both models, and then saves the trained models for later use. After training, the algorithm uses the assigned test set to evaluate each model's performance, calculating evaluation metrics including accuracy, recall, and F1 score. Based on the critical current, voltage, and temperature parameters taken from the BMS dataset, this iterative process enables a systematic comparison of the Isolation Forest and Local Outlier Factor models in terms of their capacity to identify anomalies within battery systems, offering insights into their efficacy.

Three primary metrics—accuracy, recall, and F1 score—are used in the evaluation phase to evaluate the effectiveness of the anomaly detection models. These metrics are important measures of how well the models detect abnormalities in the battery system. By comparing the percentage of accurately discovered anomalies to the total number of forecasts produced, accuracy evaluates how accurate the model's predictions are overall. Recall, sometimes referred to as sensitivity or true positive rate, measures how well the model can detect every real anomaly in the dataset. It calculates the percentage of accurately detected anomalies out of all the actual abnormalities that are there. The F1 score provides a fair assessment by taking into account both false positives and false negatives. It is a harmonic mean of precision and recall. This measure takes into consideration how well the model detects anomalies with the least number of misclassifications. The combined assessment of accuracy, recall, and F1 score allows for a more complete perspective of the performance of the anomaly detection models and their capacity to detect anomalies inside battery systems. Equations 1, 2, 3, and 4 are utilized to calculate such performance matrices.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total Sample Size}} \tag{1}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True positive} + \text{False negative}} \tag{2}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True positive} + \text{False Negative}} \tag{3}$$

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

III. Result and Discussion

Figure 1 shows a comparison of the performance of Isolation Forest and Local Outlier Factor in detecting anomaly circumstances in the dataset. Isolation Forest has a notable ability to detect a significant fraction of true anomaly situations, correlating closely with ground truth anomalies. In contrast, the Local Outlier Factor, represented in the same image, has a larger potential to discover numerous false anomaly conditions, which is considered undesirable in anomaly detection frameworks. This pattern is further emphasized in Figure 2, which focuses on temperature data. Isolation Forest continues to demonstrate adept anomaly detection, collecting major real-world anomalies. However, in this time, isolation forest detect more false anomalies compare to previous. The Local Outlier factor, on the other hand, indicates a significant number of erroneous anomaly circumstances within the temperature data, demonstrating its sensitivity to false positives.

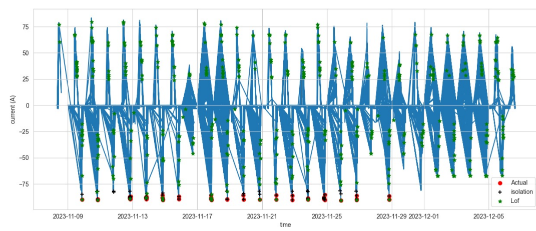

Fig. 1. Detected anomalies in current

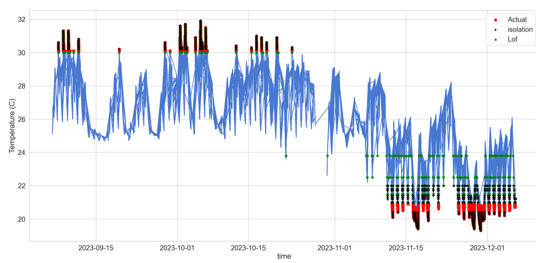

Fig. 2. Detected anomalies in temperature

Isolation Forest performed well in current data analysis, with an accuracy of 99.43%, precision of 99.69%, recall of 90.22%, and an F1 score of 94.72%. This demonstrates Isolation Forest's outstanding accuracy and precision in detecting anomalies while balancing recall and F1 score. In contrast, the Local Outlier Factor, when applied to the identical current data, showed somewhat lower accuracy 97%, but commendable precision 98.82%, recall 97.70%, and F1 score 98.26%. This demonstrates the ability of Local Outlier Factor to maintain high precision and balanced performance metrics in anomaly detection for current data.

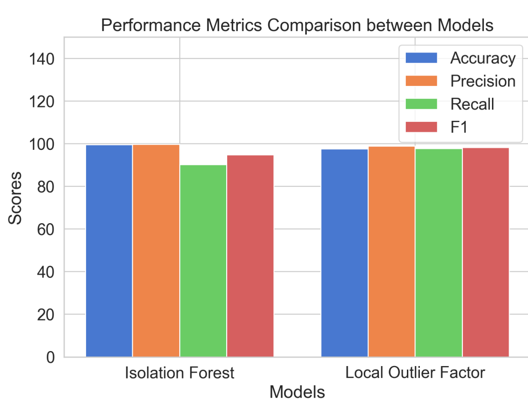

Fig. 3. Performance analysis for current data

When it came to temperature data, Isolation Forest performed admirably, with an accuracy of 97.55%, precision of 96.70%, recall of 93.82%, and F1 score of 95.23%. Despite having a slightly lower recall than current data, Isolation Forest maintained great accuracy and precision, demonstrating its dependability in spotting anomalies within temperature measures. In contrast, the accuracy of the Local Outlier Factor, when applied to the identical temperature data, was 46.08%. It did, however, show substantial precision 92.48%, recall 90.22%, and an F1 score of 91.34%, suggesting its capacity to attain balanced performance in anomaly detection but with lower accuracy than Isolation Forest.

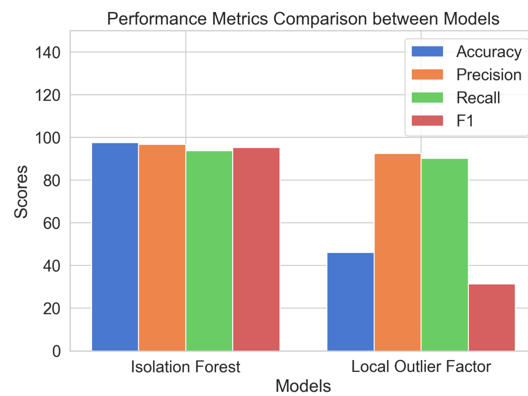

Fig. 4. Performance analysis for temperature data

Based on the presented performance measurements, it appears that Isolation Forest outperforms Local Outlier Factor in the context of battery anomaly detection. Isolation Forest regularly outperforms current and temperature data analysis in terms of accuracy, precision, and F1 score. Its consistent performance in correctly recognizing anomalies while maintaining high precision shows that it is effective in distinguishing anomalies inside battery systems. Local Outlier Factor, on the other hand, despite displaying balanced metrics, has worse accuracy in both current and temperature data evaluations.

IV. Conclusion

This study thoroughly investigated the effectiveness of Isolation Forest and Local Outlier Factor in detecting battery anomalies using current and temperature data from a PV plant's battery management system. In terms of accuracy, precision, recall, and F1 score, Isolation Forest consistently outperformed Local Outlier Factor in both datasets. Its superiority in recognizing actual anomalies inside battery systems is seen in its robust performance in precisely identifying anomalies with increased accuracy. While Local Outlier Factor had balanced performance measures, it had lesser accuracy across evaluations. The findings significantly support the use of Isolation Forest as a more powerful and dependable model for successful battery anomaly detection, providing significant insights into improving anomaly detection frameworks inside battery systems.

Acknowledgment

This work was supported by the Technology development Program (S3098815) funded by the Ministry of SMEs and Startups (MSS, Korea).

References

[1] Wali, Safat Bin, et al. "Battery storage systems integrated renewable energy sources: A biblio metric analysis towards future directions." Journal of Energy Storage 35 (2021): 102296.

[2] Wali, Safat Bin, et al. "Battery storage systems integrated renewable energy sources: A biblio metric analysis towards future directions." Journal of Energy Storage 35 (2021): 102296.

[3] Chen, Yuqing, et al. "A review of lithium-ion battery safety concerns: The issues, strategies, and testing standards." Journal of Energy Chemistry 59 (2021): 83-99.

[4] Finegan, Donal P., et al. "The application of data-driven methods and physics-based learning for improving battery safety." Joule 5.2 (2021): 316-329.

[5] Zhao, Wenjie, et al. "Anomaly detection of aircraft lead-acid battery." Quality and Reliability Engineering International 37.3 (2021): 1186-1197.

[6] Bhaskar, Kiran, et al. "Data-driven anomaly detection in large battery pa4ks." arXiv preprint arXiv:2210.15773 (2022).

[7] Li, Xiaojun, et al. "Data-driven thermal anomaly detection for batteries using unsupervised shape clustering." 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE). IEEE, 2021.

[8] Haider, Syed Naeem, Qianchuan Zhao, and Xueliang Li. "Data driven battery anomaly detection based on shape based clustering for the data centers class." Journal of Energy Storage 29 (2020): 101479.

[9] Lee, Chun-Hsiang, et al. "Anomaly detection of storage battery based on isolation forest and hyperparameter tuning." Proceedings of the 2020 5th International Conference on Mathematics and Artificial Intelligence. 2020.

[10] Diao, Weiping, Ijaz Haider Naqvi, and Michael Pecht. "Early detection of anomalous degradation behavior in lithium-ion batteries." Journal of Energy Storage 32 (2020): 101710.