
A Web Application Framework for Battery Health Prediction in Industrial IoT Networks

Seongseop Kim, Seungwoo Lee*, Minsu Kim
and Youngmin Kwon

Korea Electronics Technology Institute, South Korea
E-mail: sskim@keti.re.kr; seungwoo.lee@keti.re.kr; mskim92@keti.re.kr;
ymkwon@keti.re.kr
**Corresponding Author*

Received 25 May 2025; Accepted 15 July 2025

Abstract

This study presents a web engineering architecture for predictive battery health management in industrial IoT environments. The proposed framework leverages a scalable web-based platform that integrates data streams, web services, and machine learning modules to estimate the state of charge (SOC) of primary lithium batteries. These batteries are critical for long-term device reliability in applications such as gas advanced metering infrastructure (AMI) networks.

To overcome challenges associated with flat discharge profiles and data sparsity, the framework incorporates web-enabled data processing, online augmentation techniques (e.g., CutMix), and adaptive learning models. A key contribution of this work is the design of a modular web application layer compliant with oneM2M standards and RESTful interfaces. It includes components for real-time monitoring, automated model updates, and secure service orchestration using technologies such as HTTP bindings.

This architecture not only enables accurate SOC estimation without additional hardware but also demonstrates the critical role of web engineering in ensuring system scalability, security, and integration across heterogeneous IoT devices. Experimental validation in AMI systems confirms the effectiveness of the approach, which is extensible to broader domains such as smart utilities, environmental sensing, and industrial automation.

Keywords: Web engineering, Internet-of-Things, web-based application, oneM2M, RESTful API, online data augmentation, SOC estimation, primary lithium battery, machine learning, industrial IoT device.

1 Introduction

Primary lithium batteries are extensively utilized in industrial IoT devices due to their high energy density, low self-discharge rates, and extended operational lifetime. These attributes render them indispensable for applications requiring consistent power, including industrial sensors, monitoring systems, and advanced metering infrastructure. However, predicting the state of charge (SOC) of primary lithium batteries presents a considerable challenge, as their voltage remains stable for the majority of their lifecycle, followed by a rapid drop near the end of their usable life. This abrupt voltage decline complicates battery management and increases the likelihood of unexpected device downtime in critical IoT applications.

In industrial IoT environments, ensuring the reliable and efficient operation of distributed devices necessitates robust battery management systems alongside scalable and adaptive frameworks. Traditional SOC estimation methods, such as full discharge tests or feature-based models, are unsuitable for primary lithium batteries due to their unique discharge characteristics and the constraints inherent in IoT applications, including limited data collection capabilities and stringent energy budgets.

To address these limitations, this study proposes a web-based framework for SOC estimation tailored specifically to industrial IoT systems. The framework utilizes server-enabled data streams to collect and analyze voltage data generated during routine high-power communication events, such as those occurring in narrowband IoT (NB-IoT) devices. By integrating these voltage fluctuations into a centralized web-based platform, the framework facilitates real-time monitoring, predictive modeling, and adaptive SOC estimation without requiring additional hardware or complex testing procedures. The centralized web infrastructure further enables dynamic updates to machine

learning models, ensuring that the system remains effective across varying environmental conditions and operational scenarios.

As a practical application of the proposed framework, gas advanced metering infrastructure (AMI) systems are employed as a case study. Gas AMI networks, powered by primary lithium batteries, are designed to support operations such as remote metering and data transmission. By applying the framework, these systems can estimate SOC based on voltage changes observed during standard operations, including NB-IoT communication events. The integration of machine learning models with web-based platforms supports real-time data analysis, thereby enhancing the scalability and reliability of AMI networks.

To mitigate the challenges associated with data sparsity in resource-constrained environments, the framework incorporates advanced data augmentation techniques such as CutMix. These techniques synthetically expand and diversify training datasets, enabling machine learning models—such as Random Forest, Extra Trees, and Bagging Regressors—to achieve accurate SOC predictions even with limited input data. By processing and managing these augmented data streams through a centralized web-based infrastructure, the framework ensures scalability, adaptability, and operational efficiency across a broad spectrum of industrial IoT applications.

This study underscores the transformative potential of web-based technologies in SOC estimation and battery management for industrial IoT systems. By providing a scalable and practical solution, the proposed framework not only addresses the challenges associated with managing primary lithium batteries but also establishes a foundation for more efficient, reliable, and cost-effective operation of IoT devices in diverse application environments.

2 Related Works

Primary lithium batteries are widely used in IoT applications like military devices, smart meters, and environmental sensors because of their high energy density, long lifetime, and reliable performance under low discharge rates [1, 2]. Unlike rechargeable batteries, these batteries are designed for single use and maintain a stable voltage profile until a sudden drop near the end of their life. This unique characteristic makes predicting their remaining capacity particularly challenging [3].

Most research on battery lifetime prediction has focused on secondary (rechargeable) batteries, which exhibit gradual voltage declines and capacity degradation over time. Full discharge tests are commonly used to measure

the remaining capacity of rechargeable batteries, but these tests are not applicable to primary batteries because they permanently deplete the battery [4]. Similarly, feature-based models, which rely on trends like voltage decay or internal resistance, are difficult to apply to primary batteries due to their flat voltage profiles [5, 6].

Some studies have proposed methods tailored specifically for primary lithium batteries. Pulse load tests, for example, measure voltage responses to short, high-current pulses and use the results to estimate battery capacity. Machine learning models like backpropagation neural networks have been applied to analyze this data and improve prediction accuracy [7, 8]. However, these methods often require extensive testing under controlled conditions, which limits their applicability in real-world IoT environments where devices operate under dynamic loads and varying temperatures [9, 10].

Gas advanced metering infrastructure (AMI) systems are a notable application of primary lithium batteries. These systems rely on wireless communication technologies such as narrowband IoT (NB-IoT), which cause significant voltage changes during high-power operations. Researchers have identified that analyzing these voltage changes, such as minimum voltage during transmission and recovery voltage after the operation, can provide useful insights into battery health [11, 12]. Machine learning models, including Random Forest and support vector machines (SVMs), have been used to model these relationships and improve the accuracy of SOC estimation [13, 14]. Despite these advancements, many studies require large datasets for training, making them less suitable for resource-constrained environments like gas AMI systems [15–17].

To address the issue of limited datasets, researchers have explored data augmentation techniques such as CutMix, which combines portions of different samples to create new training data. These methods have been shown to improve the generalization of machine learning models in SOC estimation tasks [18, 19]. While data augmentation has proven effective, existing research often overlooks simpler yet practical approaches that use readily available data, such as voltage changes during normal operations, as a primary indicator for SOC [20, 21].

This study builds on prior work by emphasizing the value of using natural voltage fluctuations observed during high-power communication events in NB-IoT devices for SOC estimation. Unlike methods that rely on external tests or large datasets, this approach leverages data already generated during regular device operations, making it practical and cost-effective. Combined with machine learning models and data augmentation techniques,

this framework provides a scalable and reliable solution for battery management in gas AMI systems, ensuring long-term operational stability even in data-limited conditions.

3 Background

3.1 Industrial IoT Environments

Industrial IoT systems are integral to efficient utility management, supporting a wide range of applications in smart cities and advanced industrial operations. These systems rely on interconnected devices powered by efficient energy sources and robust communication technologies to enable real-time monitoring, data analysis, and predictive management. A notable example of Industrial IoT implementation is the advanced metering infrastructure (AMI), which facilitates the automation of metering, billing, and maintenance processes for utility providers. While AMI systems are employed as an illustrative use case in this study, the proposed framework is broadly applicable to various Industrial IoT scenarios requiring scalable, web-based solutions.

A core requirement of Industrial IoT systems is the ability to maintain reliable long-distance communication while minimizing power consumption. To meet this requirement, low power wide area network (LPWAN) technologies, such as LoRa and narrowband IoT (NB-IoT), are extensively adopted. Among these, NB-IoT offers distinct advantages due to its compatibility with existing LTE infrastructure, reducing installation costs while ensuring stable connectivity in both urban and rural environments. This makes NB-IoT particularly well-suited for battery-powered devices, such as smart meters in AMI systems, which must operate reliably for extended periods, often exceeding five years. By leveraging NB-IoT, IoT devices are able to transmit critical data efficiently while conserving energy, an essential consideration for resource-constrained environments.

To enhance interoperability and scalability across heterogeneous devices, Industrial IoT systems often adopt global standards such as oneM2M. This standard defines a common services layer, which facilitates seamless integration between devices and platforms. By utilizing web-based interfaces, oneM2M enables real-time data exchange, device management, and system integration, significantly simplifying the scaling of IoT networks. For instance, in the context of AMI systems, oneM2M allows gas meters to connect seamlessly with other IoT systems via centralized web platforms. These platforms provide operators with a unified network view, supporting remote monitoring, predictive maintenance, and efficient data analysis.

The architecture of systems like AMI generally includes IoT devices, communication networks, web-based platforms, and operational management systems. IoT devices, such as smart gas meters, are equipped with wireless communication modules that transmit periodic data, including gas usage and system status, to central servers. This data is transmitted through LPWAN protocols, such as NB-IoT, and subsequently processed on centralized platforms. Web-based platforms, in particular, play a pivotal role by storing and analyzing incoming data streams and integrating them with operational systems for tasks such as installation management, billing, and safety monitoring. The adoption of web technologies enhances scalability, enabling operators to effectively manage expansive IoT device networks with reduced effort and cost.

Data security is a critical concern in Industrial IoT systems. Web-based platforms implement end-to-end encryption protocols to protect data during transmission and storage, ensuring the confidentiality and integrity of sensitive information. This robust security framework is vital for the reliable operation of IoT networks, particularly in high-stakes applications, such as gas distribution systems or industrial monitoring, where data corruption or loss could lead to significant operational risks.

One of the primary challenges in Industrial IoT environments is the infrequent collection of data, driven by constrained communication schedules and stringent power consumption requirements as shown in Figure 1. For instance, in AMI systems, smart gas meters often transmit data only once every 24 hours to extend battery life, resulting in sparse datasets. This limitation complicates critical tasks such as estimating the state of charge (SOC) of the batteries powering these devices.

To address these challenges, the proposed framework incorporates web-based data augmentation techniques to increase the size and diversity of available datasets. Methods like CutMix generate synthetic training samples

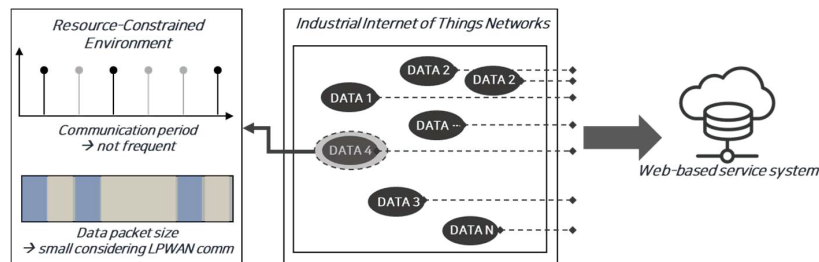


Figure 1 Resource-constrained industrial IoT network.

by combining portions of existing data, and these augmented datasets are processed and managed on web-enabled servers. By integrating these enriched datasets with ensemble machine learning models, such as Random Forest and Extra Trees, the framework enables accurate SOC predictions even under sparse data conditions.

The web-based architecture ensures that the proposed framework remains scalable and adaptable to a wide range of Industrial IoT applications beyond AMI systems. By centralizing data collection, analysis, and machine learning model updates through web platforms, the framework effectively addresses the inherent constraints of resource-limited IoT devices. This approach provides a cost-effective and efficient solution for battery management, ensuring the long-term reliability and operational efficiency of Industrial IoT systems.

3.2 Discharge Characteristics and Feature Extraction for SOC Estimation

In this study, the primary Li/SOCI2 battery used shows a specific discharge pattern under a constant discharge condition, as illustrated in Figure 2. The discharge curve displays a distinct two-stage voltage profile. At first, the voltage remains mostly stable as the battery discharges, maintaining a flat voltage profile typical of primary lithium batteries. However, when about 85% of the battery's total capacity is used, a significant voltage transition occurs, which is the start of the final discharge stage. At this point, approximately 15% of the battery's capacity remains, which is a critical marker for capacity estimation.

After this transition, predicting the remaining capacity becomes more difficult because the voltage stabilizes and stays nearly constant despite further energy depletion. This flat voltage region poses a major challenge for SOC estimation since voltage readings alone are no longer reliable indicators of the remaining capacity. As a result, accurate capacity prediction requires additional data beyond simple voltage measurements.

To improve SOC estimation accuracy, it is necessary to include other factors, such as voltage changes during high-power usage events. These additional parameters are critical for capturing more detailed patterns in the data. This highlights the need for advanced methods such as data augmentation and machine learning to improve prediction accuracy by identifying more subtle relationships within the data.

In this study, we extract several key features from voltage data. Baseline voltage (V_0) refers to the voltage level prior to the high-current event, while

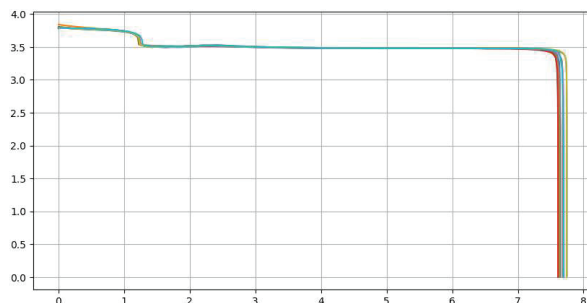


Figure 2 Primary lithium battery discharge characteristic.

minimum voltage ($V1$) represents the lowest voltage recorded during the operation. Recovery voltage ($V2$) is the stabilized voltage level following the operation, and the voltage change rate (CR) quantifies the rate of voltage drop during the event. Recovery time measures the duration required for the voltage to stabilize after the operation. These features enable the framework to identify patterns indicative of the battery's remaining capacity with high precision.

3.3 Web Engineering Considerations in IoT Battery Management Frameworks

Web engineering plays a pivotal role in the design, implementation, and scalability of the proposed SOC estimation framework. As IoT networks grow increasingly complex, the role of web-based systems transitions from mere data visualization tools to integral components that manage data flow, machine learning model updates, and system interoperability. This section outlines the key web engineering principles embedded within the proposed framework.

The framework leverages a service-oriented architecture (SOA) that adheres to web standards, including RESTful APIs and the oneM2M protocol. This enables seamless communication between edge devices, middleware (e.g., oneM2M adapters), and centralized cloud-based platforms. By decoupling components through web interfaces, system modularity is enhanced, simplifying updates and scalability across heterogeneous IoT environments.

The data augmentation process, including CutMix, is implemented as a web microservice capable of ingesting real-time data streams, applying transformations, and returning augmented datasets for training. This web-based

pipeline ensures low-latency processing and supports asynchronous task scheduling, which is critical in sparse-data environments typical of gas AMI networks.

To support real-time monitoring and diagnostics, the framework includes web dashboards that provide insights into SOC trends, device health, and predictive failure alerts. These dashboards are built using reactive web technologies (e.g., Vue.js, React) and are integrated with backend analytics services for real-time feedback loops.

From a web engineering perspective, the proposed architecture extends existing industrial IoT frameworks by incorporating modular, real-time web microservices that support asynchronous data processing and remote model management. Unlike typical web dashboards that provide passive visualization, our implementation enables real-time interaction, dynamic service orchestration, and web-based deployment of retrained models without requiring firmware updates on edge devices.

Furthermore, the system's compliance with oneM2M allows semantic interoperability across vendor-agnostic devices, enabling plug-and-play scalability across industrial networks. The RESTful API design also adopts stateless interactions and microservice separation, which improves fault tolerance and simplifies distributed deployment. These features collectively go beyond conventional web integration in industrial IoT by enabling an adaptive, resilient, and cloud-native framework for battery health estimation.

4 Proposed Framework

4.1 Web-based IoT Network Architecture

Figure 3 illustrates the overall system architecture of a resource-constrained Industrial IoT framework, such as smart utilities. The IoT devices measure data and have wireless communication capabilities for transmitting the measured data. These resource-constrained IoT devices must operate solely on internal batteries for several years. To minimize the power usage of communication, low-power wide-area (LPWA) networks, such as NB-IoT, are mainly used for smart utility networks. Communication between IoT devices and oneM2M adapters system utilizes the NB-IoT network. However, due to NB-IoT's packet size limitation of approximately 500 bytes, a data format that minimizes packet length is employed through a UDP-based data transmission protocol.

Once the device sends sensing data to oneM2M adapter, the sensing data is converted into the oneM2M standard resource format and transmitted to

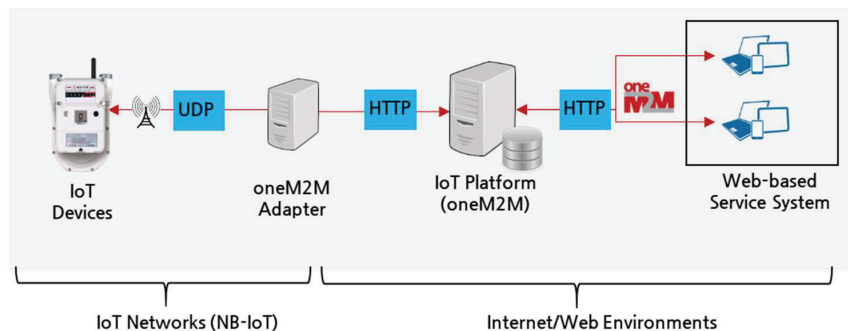


Figure 3 IoT network architecture.

the IoT platform. The IoT platform adheres to the oneM2M standard, one of the Internet of Things standards, and communication between the adapter and platform uses an HTTP binding protocol. The IoT platform manages resources and can provide web-based services by interfacing with various IoT applications.

The adoption of web technologies such as oneM2M and RESTful APIs plays a critical role in the scalability and interoperability of the proposed SOC estimation framework. The oneM2M standard provides a unified communication protocol and resource structure, which ensures seamless data integration from heterogeneous IoT devices regardless of vendor or communication protocols. This allows the system to easily scale across various industrial settings, such as smart metering, environmental monitoring, and automation.

RESTful APIs enable efficient and standardized data exchange between microservices, allowing for modular deployment, remote model updates, and integration with web-based dashboards. They also support stateless communication, which simplifies error handling and improves robustness in intermittent-connectivity environments. These characteristics are essential for battery-powered industrial IoT systems that require reliable and lightweight communication mechanisms. Therefore, oneM2M and RESTful APIs help ensure that the battery health estimation services can be deployed and maintained efficiently across large-scale IoT infrastructures with minimal manual intervention.

4.2 Voltage Response Modeling for SOC Prediction

To address the challenges associated with voltage-based capacity estimation, this study emphasizes the analysis of voltage variations occurring during

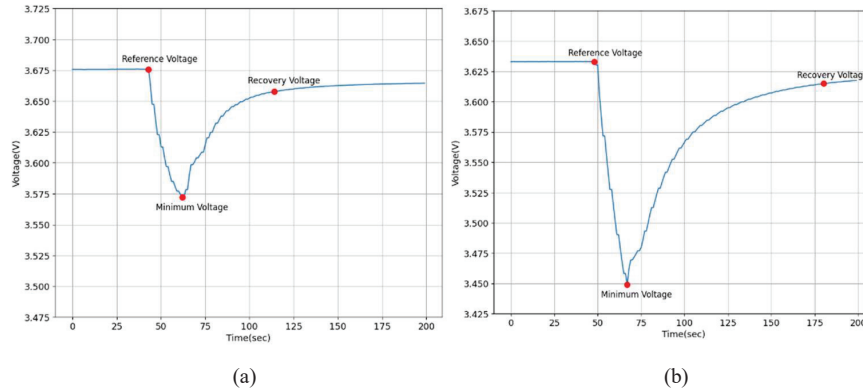


Figure 4 (a) Voltage recovery characteristic at 20% used and (b) voltage recovery characteristic at 40% used.

high-current operations, such as communication activities in industrial IoT devices. These voltage fluctuations serve as key indicators of the remaining capacity of primary lithium batteries.

As described in Section 3.2, the framework leverages key voltage response features, such as baseline, minimum, and recovery voltages, extracted during high-power NB-IoT communication events to estimate SOC under real-world conditions, induce noticeable voltage dips and subsequent recoveries. These variations offer valuable insights into the battery's performance under practical environmental conditions and are essential for accurately estimating its remaining capacity, as shown in Figure 4.

Furthermore, this study investigates the influence of environmental factors, particularly temperature, on battery performance. Controlled experiments are conducted to evaluate how changes in temperature affect voltage responses. Since temperature variations significantly alter the internal resistance and voltage characteristics of the battery, accounting for these effects enhances the precision of SOC predictions.

These voltage features were selected based on their sensitivity to battery degradation phenomena observable during real-world communication events. In particular, the minimum and recovery voltages (e.g., V_1 and V_2) are strongly affected by internal resistance, transient response characteristics, and the battery's ability to stabilize after load removal, all of which degrade over time as capacity diminishes. By capturing such transient behaviors, the selected features provide a more responsive and accurate basis for SOC estimation compared to traditional steady-state metrics. This approach is

especially suitable for primary lithium batteries, which maintain flat discharge curves for most of their lifespan. Across various temperature and load conditions, these features have shown consistent correlation with residual capacity, validating their effectiveness in practical deployments.

The integration of voltage responsiveness during high-power operations with environmental factors, such as temperature, enables a more robust and reliable estimation of the remaining capacity. By incorporating these parameters into the predictive model, the proposed method improves the robustness and accuracy of SOC estimations. This approach is particularly well-suited for deployment in dynamic and challenging IoT environments where variability in operating conditions is prevalent.

4.3 Data Driven SOC Estimation

The proposed framework is designed to estimate the remaining capacity of primary lithium batteries in industrial IoT environments, as illustrated in Figure 5. Leveraging the capabilities of web-based platforms, this approach integrates feature-based analysis with machine learning techniques to provide an effective and scalable solution for battery capacity estimation in resource-constrained scenarios. While the framework is broadly applicable across various industrial IoT systems, a gas advanced metering infrastructure (AMI) is used as an example to demonstrate its practicality and effectiveness.

In industrial IoT applications, devices such as environmental sensors, industrial controllers, and metering systems are required to operate reliably over extended periods with minimal maintenance. Accurate monitoring of battery capacity is essential to prevent unexpected device downtime. The proposed framework addresses this need by analyzing voltage variations that naturally occur during high-current operations, such as those involved in data transmissions over narrowband IoT (NB-IoT). These voltage fluctuations are captured and processed using the voltage responsiveness measurement module, which extracts and analyzes critical battery behavior metrics.

The proposed framework uses the voltage response features outlined in Section 3.2 as input variables for machine learning-based SOC estimation, enabling accurate capacity estimation without relying on direct SOC measurements, which are often impractical in real-world IoT environments.

The integration of web-based platforms significantly enhances the framework's scalability and adaptability. Centralized data collection and processing on a web-enabled server allows the framework to manage data streams from numerous IoT devices in real-time. This capability is particularly valuable

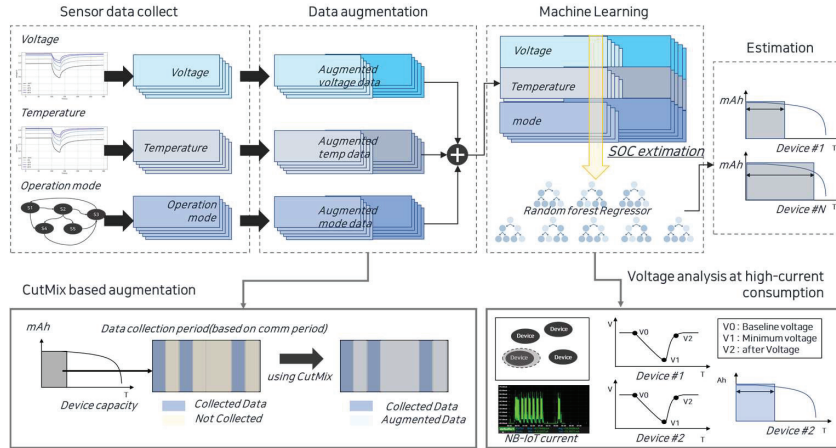


Figure 5 Proposed framework for web-based battery SOC estimation.

in large-scale deployments, such as gas AMI networks, where thousands of devices generate data intermittently. For instance, edge devices in AMI systems transmit data once every 24 hours to conserve energy, resulting in sparse datasets. The centralized web-based platform bridges this gap by aggregating, analyzing, and visualizing the collected data, ensuring continuous system monitoring and efficient battery management.

Moreover, the framework employs advanced machine learning techniques to dynamically adapt to variations in operating conditions. Ensemble models, such as Random Forest and Extra Trees, analyze the extracted voltage features to identify patterns indicative of the battery’s remaining capacity. These models are trained and updated on the web platform, enabling rapid deployment of improved predictive models across all connected devices.

Although gas AMI systems serve as a case study, the modular and web-based architecture of the framework makes it highly adaptable to other industrial IoT applications. It can be applied to diverse resource-constrained environments, such as environmental monitoring systems, industrial automation, and smart agriculture. By centralizing battery data management and model updates, the framework supports long-term device operation and mitigates the risk of unexpected battery failures.

In summary, the proposed framework combines voltage responsiveness analysis, machine learning, and centralized web-based processing to deliver a robust solution for SOC estimation in industrial IoT environments (Figure 6). The integration of web-based platforms not only ensures scalability and

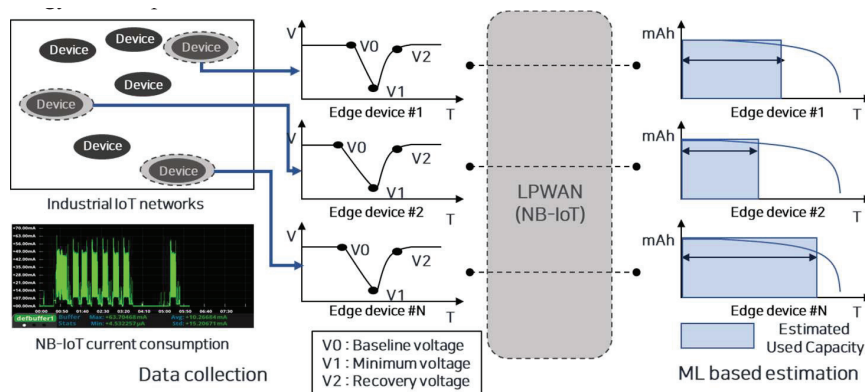


Figure 6 Machine learning based SOC estimation framework in smart metering networks.

reliability but also simplifies the deployment and maintenance of predictive models across diverse IoT systems. This study highlights the potential of web-based technologies to transform battery management, demonstrating their effectiveness through the example of gas AMI systems.

By employing advanced machine learning algorithms, such as Random Forest Regressor, Extra Trees, Bagging Regressor, and XGBoost, the framework develops predictive models capable of analyzing complex relationships in the data. These ensemble models were selected due to their robustness in handling high-dimensional, nonlinear feature interactions, which are common in voltage-based SOC estimation. In addition to their predictive performance, ensemble methods like Random Forest and Extra Trees offer strong model interpretability by enabling feature importance analysis. We analyzed the contribution of each voltage feature using built-in importance scores in tree-based models. The results indicated that minimum voltage (V1) and recovery voltage (V2) were the most influential predictors of residual battery capacity, aligning with prior observations on voltage responsiveness during high-load events. Such interpretability is essential for deploying the models in industrial IoT systems, where transparent decision-making and diagnostic capabilities are critical. This architecture is particularly well-suited to resource-constrained IoT environments, where frequent data collection and communication are restricted by energy consumption limitations.

This framework not only enhances the accuracy of SOC estimation but also supports proactive maintenance strategies in industrial IoT environments. By providing reliable predictions of battery capacity, the framework enables operators to identify potential issues before they lead to device

failures, ensuring uninterrupted operation of connected systems. As demonstrated in the example of gas advanced metering infrastructure (AMI) systems, the proposed approach leverages web-based platforms to centralize data processing and monitoring, enhancing scalability and operational efficiency.

Through the integration of web technologies, the framework facilitates real-time updates to machine learning models and ensures seamless communication across distributed IoT devices. This capability not only improves the reliability of SOC predictions but also supports dynamic adaptation to changing operational conditions. By addressing battery management challenges with a scalable and efficient solution, the framework strengthens the overall efficiency and reliability of industrial IoT applications while reducing maintenance costs and minimizing downtime.

Compared to traditional SOC estimation methods such as Coulomb counting, full discharge tests, or open-circuit voltage (OCV) profiling, the proposed framework offers several practical advantages. Coulomb counting requires accurate current sensing and integration over time, which is not feasible in low-power IoT devices due to hardware limitations and energy constraints. Full discharge tests permanently deplete the battery and are unsuitable for primary lithium cells. OCV profiling assumes a strong correlation between voltage and SOC, which is not applicable to primary batteries due to their flat discharge profiles.

In contrast, our approach leverages voltage transients during normal high-current communication events—data that is already generated by the device in routine operation. This allows for non-invasive, real-time SOC estimation without requiring additional sensors, invasive testing, or large datasets. The integration with web-based platforms further enhances its practicality by enabling remote model updates and centralized analytics.

4.4 Challenges in Resource-constrained IoT Environments

IoT devices in industrial environments, including wireless sensors, controllers, and edge devices, operate under strict resource constraints, particularly in terms of power consumption and communication frequency. These devices are designed to function reliably over extended periods, often exceeding five years, and are typically powered by a single primary lithium battery. To achieve such longevity, communication activities are minimized, and data transmission is restricted to essential operations, such as daily usage reports or event-triggered updates through networked communication systems. While this design is effective in conserving power, it significantly limits

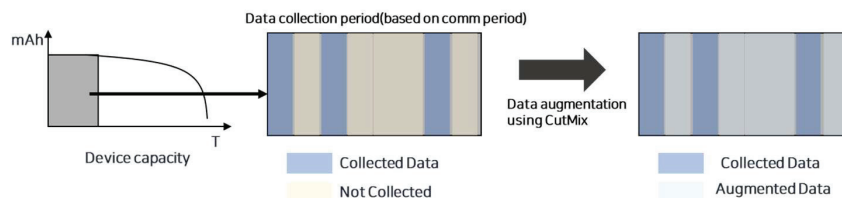


Figure 7 Data augmentation for ML based SOC estimation.

the frequency and volume of data collection, posing substantial challenges for tasks like state of charge (SOC) prediction.

The primary challenge in SOC estimation for resource-constrained IoT devices lies in the insufficient and intermittent nature of available data. Traditional machine learning models generally require large, diverse datasets to identify patterns and generate reliable predictions. However, the sparse datasets resulting from limited communication schedules in IoT devices, such as those in gas advanced metering infrastructure (AMI) systems, can impair the performance of SOC prediction models. Furthermore, these devices often operate under highly dynamic conditions. Factors such as temperature fluctuations, variable load profiles, and transitions between low-power and high-power modes create complex voltage and current behaviors that are difficult to model with limited data.

To overcome these challenges, the proposed framework employs web-based online data augmentation techniques to enhance the quantity and diversity of training data, which is shown in Figure 7. These methods dynamically generate synthetic data points in real time using centralized, network-enabled processing. For instance, techniques like CutMix combine portions of existing data samples to create augmented datasets, allowing machine learning models to capture complex patterns and improve prediction accuracy. By leveraging web-enabled servers for the processing and management of these augmented datasets, the framework ensures scalability and seamless integration across distributed IoT networks.

In addition to augmenting data, the web-based approach enables centralized coordination of data flows and updates to machine learning models. This networked infrastructure supports the continuous refinement of SOC prediction models, allowing them to adapt dynamically to changing operational conditions. For example, in gas AMI systems, where smart meters typically transmit data only once per day to conserve energy, the centralized platform aggregates data from multiple devices, applies augmentation techniques, and updates the predictive models. This approach enhances the robustness

of SOC estimation despite the inherent limitations posed by sparse data availability.

Accurate SOC predictions are critical to ensuring the uninterrupted operation of industrial IoT devices. In systems like gas AMI networks, inaccurate SOC estimation can result in premature battery replacements, leading to increased maintenance costs, or unexpected battery failures, causing service interruptions. By addressing these limitations through web-based solutions, the framework provides a cost-effective and scalable approach to battery management. For instance, centralized web platforms facilitate real-time monitoring and predictive analysis, mitigating the risks associated with battery depletion and minimizing device downtime.

While gas AMI systems serve as an example application, the proposed framework is broadly applicable to other industrial IoT environments, such as environmental monitoring, smart agriculture, and industrial automation. The web-based architecture ensures that the solution can scale across diverse IoT networks, enabling dynamic data integration, real-time analytics, and proactive maintenance strategies. By combining advanced data augmentation methods with centralized web processing, the framework provides reliable SOC predictions and improves the overall efficiency and reliability of industrial IoT applications.

4.5 Data Augmentation Techniques

In this study, CutMix was applied as the primary data augmentation technique to address the issue of limited data availability in gas AMI systems. These techniques are particularly effective in expanding small datasets, which often result from the limited communication frequency and power constraints of gas AMI devices.

CutMix introduces structure into data augmentation by cutting a region from one sample and pasting it onto another. The new sample is created as:

$$\hat{x} = M \odot x_1 + (1 - M) \odot x_2, \quad \hat{y} = \lambda y_1 + (1 - \lambda)y_2 \quad (1)$$

Here, \odot represents the mask for the cut region, and λ is the proportion of the cut area relative to the total area. CutMix is particularly effective for time-series data, such as voltage and current profiles, as it preserves local structures while introducing diversity by swapping data features between samples.

By applying CutMix, this study expanded the dataset's size and variability, allowing models to generalize better across different operating conditions. These techniques were critical in scenarios where the dataset was reduced

(e.g., reduced original data) or augmented further. This augmentation helped improve the predictive performance of machine learning models, including Random Forest Regressor, Extra Trees, Bagging Regressor, and XGBoost.

The effectiveness of CutMix was evaluated using various error metrics, including mean absolute error (MAE), root mean squared error (RMSE), and relative error. These metrics were calculated across three datasets: original data, reduced data, reduced data with CutMix augmentation. The results demonstrated significant improvements in model robustness and prediction accuracy when CutMix was applied.

While several time-series augmentation methods such as Time Warping, Jittering, and Window Slicing were considered, CutMix was selected due to its effectiveness in preserving localized structural patterns while introducing diversity at the segment level. These characteristics are particularly suitable for voltage–time sequences, where partial shape retention is important for SOC prediction. Moreover, CutMix enables controlled mixing ratios and preserves temporal alignment, making it compatible with our voltage feature extraction approach.

5 Experimental Results

The performance of the proposed residual capacity prediction framework was evaluated through experiments that measured voltage variations across different battery capacities. Voltage data was collected during communication operations using IoT devices powered by primary lithium batteries. Specifically, current consumption patterns during NB-IoT communication events were utilized, with seven distinct patterns varying in intensity and duration. These patterns simulated diverse operational conditions, enabling a detailed analysis of voltage changes under varying battery capacities.

The collected data was used to develop a predictive module for estimating the remaining capacity of primary lithium batteries. These batteries exhibit a characteristic voltage drop when approximately 15% of their capacity remains, and this behavior was leveraged for residual capacity prediction. As previously defined, voltage response features were used to train SOC estimation models using the experimental dataset. These features were then used to train machine learning models to accurately estimate residual capacity under diverse operational scenarios, as shown in Figure 8.

The current consumption pattern utilized in the experiment consists of a total of seven variations, each characterized by different current patterns and durations. These diverse current consumption patterns were employed

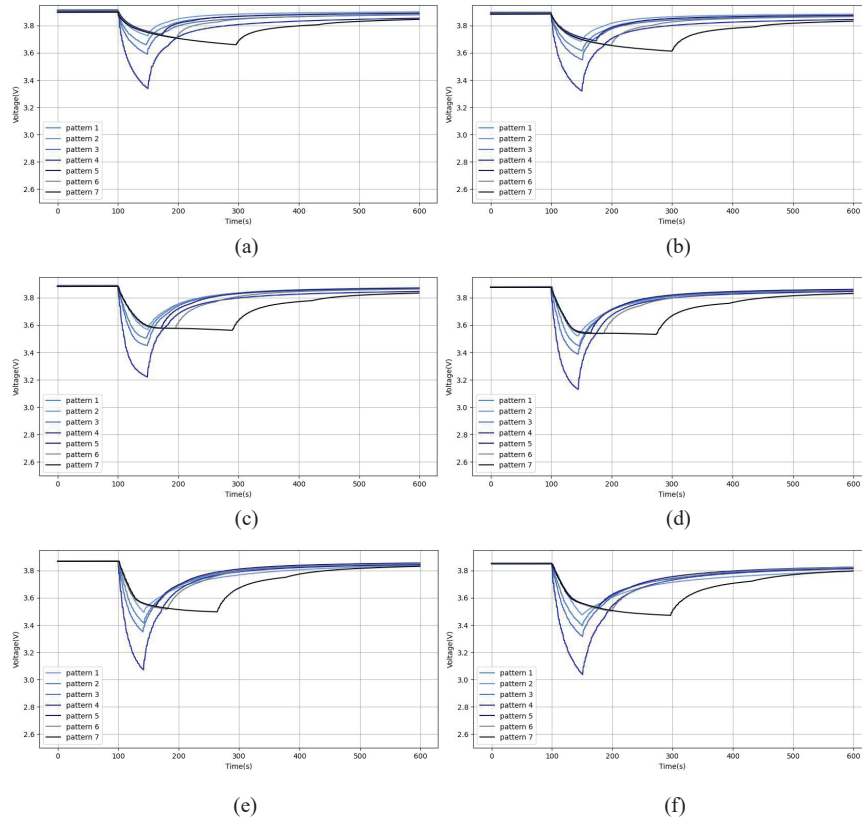


Figure 8 Voltage change characteristics according to current consumption pattern at capacity of (a) 100%, (b) 80%, (c) 60%, (d) 40%, (e) 30%, (f) 20%.

to measure voltage variation data resulting from current consumption at different capacity levels.

Furthermore, to facilitate data-driven residual capacity prediction for primary lithium batteries, experiments were conducted to measure voltage variations by altering temperature conditions using a constant temperature and humidity chamber. These experiments involved varying the temperature conditions while observing the voltage changes resulting from different current consumption patterns, as shown in Figure 9.

The obtained voltage variation data during communication operations were utilized to develop a module for predicting the remaining capacity of the primary battery commonly used in the research. These batteries exhibit a characteristic voltage change at the point where only 15% of capacity

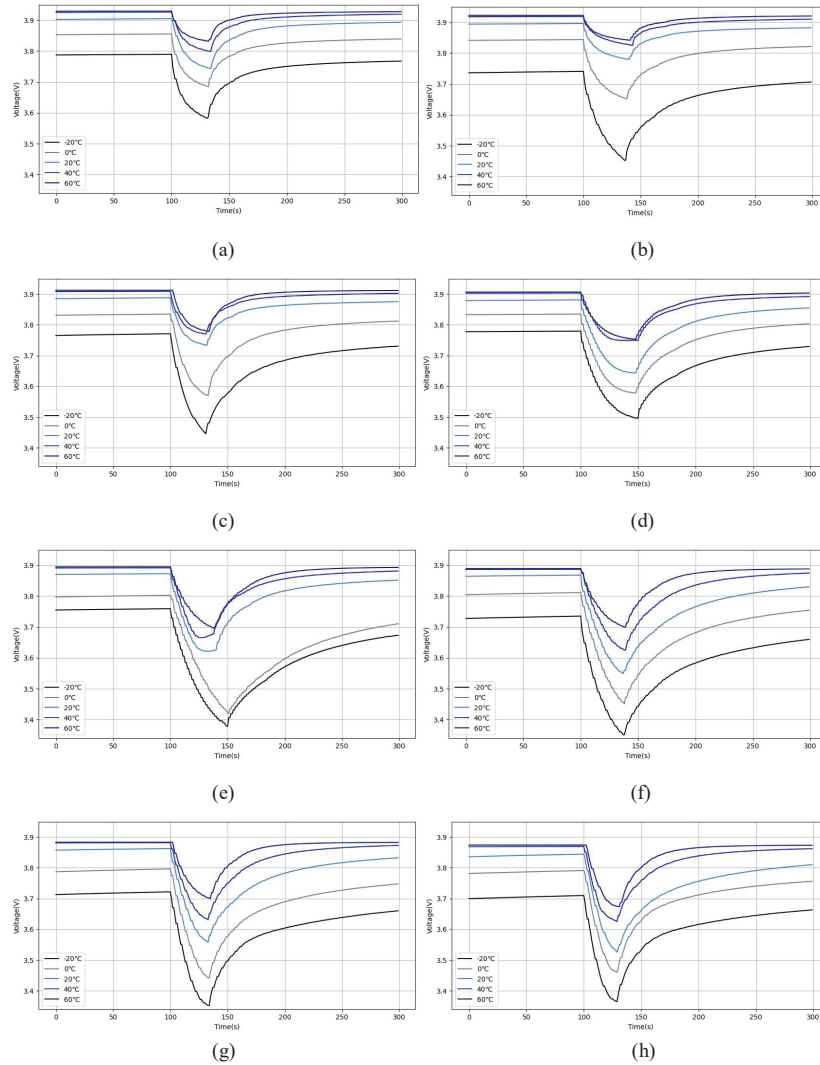
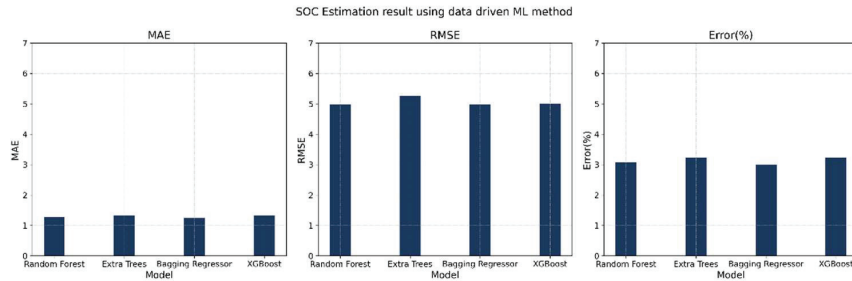


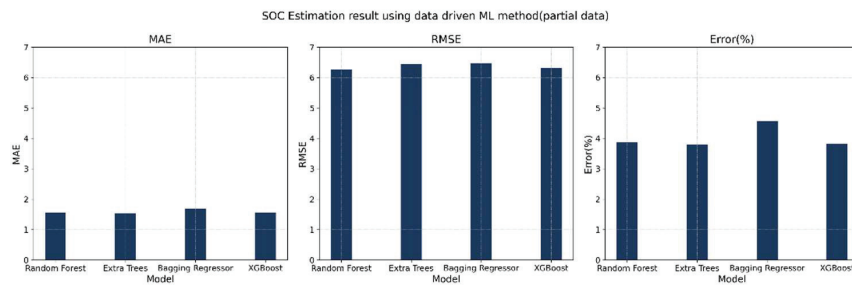
Figure 9 Voltage change characteristics according to temperature condition of (a) 100%, (b) 80%, (c) 70%, (d) 60%, (e) 50%, (f) 40%, (g) 30%, (h) 20%.

remains. Leveraging this behavior, the module measures voltage variations to predict the current remaining capacity of the battery.

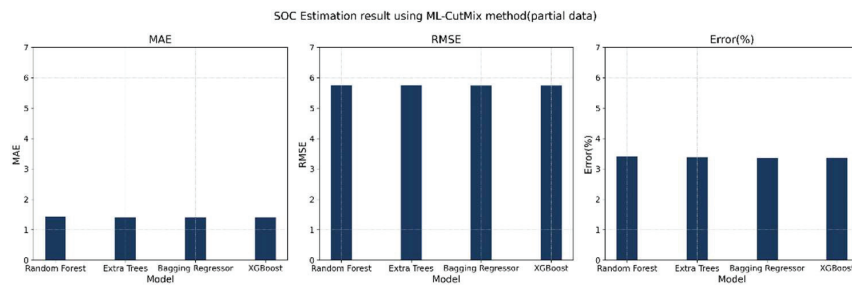
The performance of the models was evaluated under three different scenarios: the original dataset, a reduced dataset with only 30% of the original data, and the reduced dataset augmented using CutMix. Figure 10 and Table 1



(a)



(b)



(c)

Figure 10 (a) SOC estimation result using the data driven ML method, (b) SOC estimation result using the data driven ML method (partial data), and (c) SOC estimation result the using the ML-Cutmix method (partial data).

summarize the results, illustrating the impact of data augmentation on model accuracy and generalization.

As shown in Figure 10(a), the models trained on the original dataset achieved the best performance metrics overall. Random Forest recorded the lowest MAE (1.267), RMSE (4.976), and error (%) (3.08%). Bagging

Table 1 Results of ML based SOC estimation

	Model	MAE	RMSE	Error (%)
Original data	Random Forest	1.2674	4.9764	3.0771
	Extra Trees	1.3252	5.2670	3.2204
	Bagging Regressor	1.2424	4.9889	2.9915
	XGBoost	1.3225	4.9994	3.2225
Partial data	Random Forest	1.5647	6.2713	3.8625
	Extra Trees	1.5351	6.4467	3.7951
	Bagging Regressor	1.6887	6.4652	4.5618
	XGBoost	1.5629	6.3040	3.8171
Partial data with CutMix	Random Forest	1.4321	5.7568	3.4045
	Extra Trees	1.4080	5.7542	3.3860
	Bagging Regressor	1.4175	5.7394	3.3535
	XGBoost	1.4105	5.7398	3.3711

Regressor also performed well, achieving a similar RMSE of 4.989 and slightly lower error (%) 2.99%. Extra Trees and XGBoost, while slightly less accurate, demonstrated consistent results, confirming their robustness across varying operational states. These findings underscore the efficacy of ensemble methods, particularly Random Forest and Bagging Regressor, in capturing complex feature interactions within the dataset.

Figure 10(b) illustrates the impact of using the reduced dataset, which contained only 50% of the original data. The performance of all models declined significantly, emphasizing the challenges posed by limited data diversity. For instance, Random Forest's MAE increased from 1.267 to 1.5647, and its RMSE rose from 4.976 to 6.2713. Similarly, Bagging Regressor and XGBoost exhibited higher error rates, reflecting their reduced ability to generalize under data-constrained conditions.

To address these limitations, CutMix augmentation was applied to the reduced dataset. The results, shown in Figure 10(c), demonstrate a clear improvement in model performance compared to the reduced dataset alone. Random Forest achieved an MAE of 1.4321 and an RMSE of 5.7568, which marked a substantial recovery in accuracy. Extra Trees also showed significant gains, with an MAE of 1.4080 and an RMSE of 5.7542. These improvements highlight the effectiveness of CutMix in synthetically increasing data diversity by mixing data patches between samples. However, Bagging Regressor exhibited less consistent improvements, with an error (%) increasing slightly compared to the original dataset, suggesting that its sensitivity to augmented data may vary depending on the augmentation.

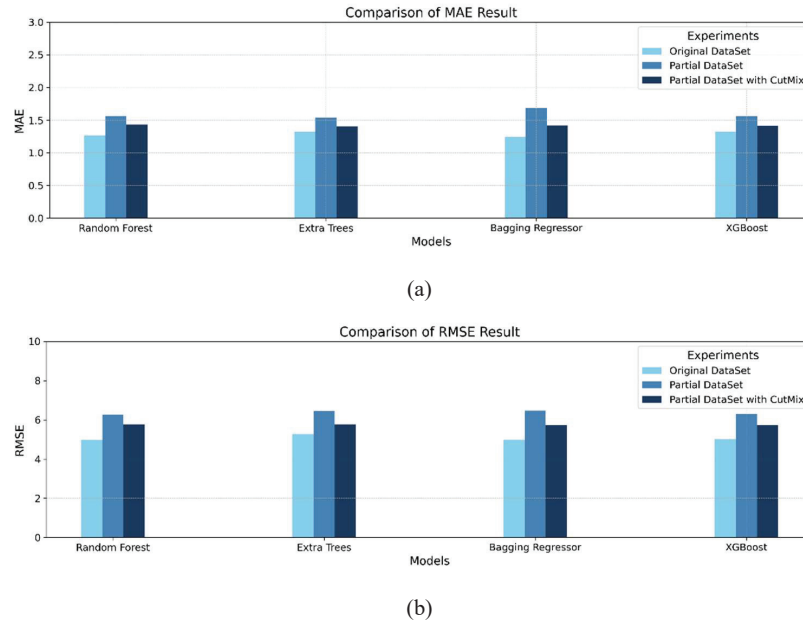


Figure 11 (a) Comparison of the MAE result and (b) comparison of the RMSE result.

Figure 11(a) provides a detailed comparison of MAE results between the partial dataset and the CutMix-augmented partial dataset. The models demonstrated consistent improvements when trained on the augmented dataset. For instance, Random Forest’s MAE decreased from 1.5647 in the partial dataset to 1.4321 in the CutMix-augmented dataset, reflecting a 9% improvement. Similarly, Extra Trees and XGBoost exhibited reductions in MAE, with Extra Trees achieving the most significant enhancement. These results confirm that CutMix effectively compensates for data sparsity by generating diverse and representative training samples, thereby reducing prediction errors in resource-constrained scenarios.

Figure 11(b) illustrates the RMSE results under the same conditions as Figure 10(a), reinforcing the findings related to error reduction through CutMix augmentation. Random Forest showed a marked reduction in RMSE from 6.2713 to 5.7568, while Extra Trees exhibited a comparable improvement, decreasing its RMSE from 6.4467 to 5.7542. These improvements underline the capability of CutMix to enhance model performance by augmenting the variability in the training data and mitigating overfitting, particularly in ensemble-based models like Random Forest and Extra Trees.

In summary, Figure 11 highlights the critical impact of CutMix on reducing both MAE and RMSE in machine learning models trained on limited data. The results validate the use of data augmentation techniques as an effective strategy for improving SOC estimation accuracy and generalization in resource-constrained environments. These findings further emphasize the versatility of ensemble models in leveraging augmented data to achieve robust performance across diverse operational states.

6 Conclusion

This study presents a framework for estimating the state of charge (SOC) of primary lithium batteries deployed in industrial IoT applications, such as gas advanced metering infrastructure (AMI) networks. The proposed approach leverages voltage variations observed during high-power communication events, such as those in NB-IoT devices, to estimate the remaining battery capacity. By focusing on natural voltage changes occurring during typical device operations, the framework offers a practical method for SOC estimation that does not require additional hardware or complex testing procedures.

To address the challenge of limited data availability in resource-constrained environments, the framework incorporates web-based online data augmentation techniques, such as CutMix, which enhance model robustness by synthetically expanding the size and diversity of the dataset in real time. Moreover, the use of centralized, web-enabled data processing facilitates the seamless integration of data streams from IoT devices, thereby improving the scalability and efficiency of the framework. This combination of voltage-based feature analysis and network-integrated data augmentation enables accurate and reliable SOC predictions, even in scenarios where data is sparse.

The proposed framework provides a scalable and effective solution for battery management in industrial IoT settings, ensuring long-term device operation while minimizing maintenance requirements. Applications such as gas AMI networks illustrate the practicality of the framework; however, it is designed to adapt to a wide variety of industrial IoT environments, including environmental monitoring and industrial automation, where reliable and efficient battery management is equally essential.

Future research will focus on expanding the framework to incorporate additional battery chemistries and to account for more complex environmental factors, such as humidity variations and dynamic load conditions. These enhancements aim to further improve the accuracy and versatility of the

framework across a broader range of IoT applications. Additionally, ongoing exploration of web-based and network-enabled data augmentation methods will help refine the framework, ensuring its continued effectiveness in diverse and evolving operational environments.

Acknowledgement

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. RS-2021-KP002425, Development of a household gas energy usage (AMI) data trading platform and associated services).

References

- [1] Q. Sun, X. Ye, H. Li, W. Li, R. Yuan and G. Zhai, 'Life Prediction of Lithium Thionyl Chloride Battery Based on Pulse Load Test and Accelerated Degradation Test', 2021 3rd International Conference on System Reliability and Safety Engineering (SRSE), pp. 174–180, Harbin, 2021.
- [2] Q. Sun, X. Ye, H. Li, W. Li, R. Yuan and G. Zhai, 'Estimation of Lithium Primary Battery Capacity Based on Pulse Load Test', 2021 3rd International Conference on System Reliability and Safety Engineering (SRSE), pp. 139–144, Harbin, China, 2021
- [3] M. Coleman, W. G. Hurley and C. K. Lee, 'An Improved Battery Characterization Method Using a Two-Pulse Load Test'. in *IEEE Transactions on Energy Conversion*, vol. 23, no. 2, pp. 708–713, June 2008.
- [4] S. Li, Z. Chen, Q. Liu, W. Shi and K. Li, 'Modeling and Analysis of Performance Degradation Data for Reliability Assessment: A Review', in *IEEE Access*, vol. 8, pp. 74648–74678, 2020.
- [5] L. Cailian and Z. chun, 'Life prediction of battery based on random forest optimized by genetic algorithm', 2020 IEEE International Conference on Prognostics and Health Management (ICPHM), Detroit, MI, USA, pp. 1–6, 2020.
- [6] C. Li, Z. Chen, J. Cui, Y. Wang and F. Zou, 'The lithium-ion battery state-of-charge estimation using random forest regression', 2014 Prognostics and System Health Management Conference (PHM-2014 Hunan), Zhangjiajie, China, 2014.
- [7] Z. Xiao, H. Fang, Z. Li and Y. Chang, 'Remaining Useful Life Prediction of Lithium-ion Battery Based on Unscented Kalman Filter and Back

- propagation Neural Network’, 2019 IEEE 8th Data Driven Control and Learning Systems Conference (DDCLS), Dali, China, pp. 47–52, 2019.
- [8] T. Liu, W. Huang, R. Pan, Y. Wang, M. Tan and J. Chen, ‘State of Health Estimation for Lithium-Ion Batteries by Using Partial Battery Data with a Hybrid Neural Network Model’, 2023 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia), Chongqing, China, pp. 1783–1787, 2023.
- [9] Peiyu Chen, Zhaoyong Mao, Chengyi Lu, Bo Li, Wenjun Ding, Junqiu Li, ‘Effective capacity early estimation of lithium thionyl chloride batteries for autonomous underwater vehicles’, 2024 Journal of Power Sources, Volume 595, 2024.
- [10] Yassine Manane, Rachid Yazami, ‘Accurate state of charge assessment of lithium-manganese dioxide primary batteries,’ 2017 Journal of Power Sources, Volume 359, pp. 422–426, 2017.
- [11] Z. Wei, K. Liu, X. Liu, Y. Li, L. Du and F. Gao, ‘Multilevel Data-Driven Battery Management: From Internal Sensing to Big Data Utilization’, in IEEE Transactions on Transportation Electrification, vol. 9, no. 4, pp. 4805–4823, Dec. 2023.
- [12] X. Xu, C. -S. Huang, M.-Y. Chow, H. Luo and S. Yin, ‘Data-driven SOC Estimation with Adaptive Residual Generator for Li-ion Battery’, IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society, Singapore, pp. 2612–2616, 2020.
- [13] S. V. Vamsi, K. M. Nagabushanam, K. V. Kumar, S. V. Tewari and T. Mahto, ‘State of Health of Lithium-ion Batteries by Data-Driven Technique with Optimized Gaussian Process Regression’, 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), Bangalore, India, pp. 1–6, 2023.
- [14] F. Cui, Z. Li, C. Liu and Y. Shi, ‘A Data-Driven Hybrid Approach for Capacity Estimation on Lithium-ion Battery’, 2022 China Automation Congress (CAC), Xiamen, China, pp. 5895–5898, 2022.
- [15] J. Zhou, D. Liu, Y. Peng and X. Peng, ‘Dynamic battery remaining useful life estimation: An on-line data-driven approach’, 2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings, Graz, Austria, pp. 2196–2199, 2012.
- [16] J. Zhao, L. Tian, L. Cheng, Y. Zhang and C. Zhu, ‘Review on RUL Prediction Methods for Lithium-ion Battery’, 2022 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia), Shanghai, China, pp. 1501–1505, 2022.

- [17] M. Abbas, I. Cho and J. Kim, ‘Mathematical analysis of battery data for development of data-driven degradation model’, 2021 24th International Conference on Electrical Machines and Systems (ICEMS), Gyeongju, Korea, Republic of, pp. 992–997, 2021.
- [18] X. -S. Si, W. Wang, C. -H. Hu, D. -H. Zhou and M. G. Pecht, ‘Remaining Useful Life Estimation Based on a Nonlinear Diffusion Degradation Process’, in *IEEE Transactions on Reliability*, vol. 61, no. 1, pp. 50–67, March 2012.
- [19] Y. Li, H. Sheng and Y. Cheng, ‘Battery State of Health Estimation with Incremental Capacity Analysis Technique’, 2020 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD), Xi’an, China, pp. 384–389, 2020.
- [20] M. Lee, D. Lee and J. Kim, “Lithium-ion battery application time series data augmentation based on generative adversarial network for training deep learning algorithm,” 2024 IEEE 10th International Power Electronics and Motion Control Conference (IPEMC2024-ECCE Asia), Chengdu, China, 2024, pp. 4403–4408.
- [21] V. Maiya, V. R. Urs, J. Channegowda and C. Lingaraj, “Robust Battery Data Augmentation Techniques for Reliable State-of-Charge Estimation using Contrastive Learning,” *2023 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, Bangalore, India, 2023, pp. 1–6.

Biographies



Seongseop Kim received his bachelor’s degree in Electronics Engineering from Kyungpook National University in 2017 and his master’s degree in Electronics Engineering from Kyungpook National University in 2019. He worked as a Researcher at the Agency for Defense Development (ADD)

from 2019 to 2020 and has been working as a Senior Researcher at the Korea Electronics Technology Institute (KETI) since 2020. His research areas include on-device AI, embedded systems, parallel processing, and AI-based signal analysis.



Seungwoo Lee received his bachelor's degree in Computer Science from Yonsei University in 2008, his master's degree in Computer Science from Yonsei University in 2010, and his Doctor of Philosophy degree in Computer Science from Yonsei University in 2015. He has been working as a Principal Researcher at the Korea Electronics Technology Institute (KETI) since 2014. His research areas include IoT platforms, AI applications, and computer systems.



Minsu Kim received his bachelor's degree in Electronic and Telecommunication Engineering from Kwangwoon University in 2017 and his Doctor of Philosophy degree in Electronic and Telecommunication Engineering from Kwangwoon University in 2022. He has been working as a Senior Researcher at the Korea Electronics Technology Institute (KETI) since 2023.

His research areas include IoT platforms, industrial domain applications, and AI-based data analysis.



Youngmin Kwon received his bachelor's degree in Electronics Engineering from Yeungnam University in 2002, his master's degree in Electronics Engineering from Yeungnam University in 2004, and completed doctoral coursework in Convergence Systems Engineering at Hanyang University in 2019. He has been working as a Principal Researcher at the Korea Electronics Technology Institute (KETI) since 2004. His research areas include data fusion signal processing, high-speed communication networks, and SoC.

