LiBrain: LLM-Powered Li-ion Battery Diagnostics with Time-Series-Aware Retrieval-Augmented Framework for E-bikes

Zhao Li[1, 2, 3], **Zixin Lin**[2], **Donghui Ding**[4, 1 *], **Yichen Zhong**[1], **Biao Wang**[1], **Haitao Xu**[2 *], **Peng Cai**[4]

[1]Hangzhou Yugu Technology Co., Ltd.
[2]Zhejiang University
[3]Hangzhou Institute of Advanced Study, UCAS
[4]East China Normal University

lzjoey@gmail.com, linzixin@zju.edu.cn, 72275900022@stu.ecnu.edu.cn, zhongyichen@yugu.net.cn, wangbiao@yugu.net.cn, haitaoxu@zju.edu.cn, pcai@dase.ecnu.edu.cn

Abstract

The rapid proliferation of smart-city ecosystems has significantly amplified the demand for Li-ion batteries, which now serve as the primary energy source for sustainable transportation systems such as e-bikes. Ensuring battery safety and optimal performance is crucial, yet challenging due to complex intrinsic dynamics and extrinsic operating conditions. This paper presents LiBrain, an innovative LLM-powered, time-series-aware retrieval-augmented framework designed to simultaneously address both safety and performance challenges through three synergistic components: (1) a distributed IoT-enabled edge network for continuous real-time battery monitoring and data acquisition, (2) a pretrained deep multi-task diagnostic engine capable of comprehensive battery performance forecasting, and (3) a knowledge-base augmentation module that transforms technical diagnostics into clear, actionable guidance tailored for e-bike users. Functioning as an intelligent battery management assistant, LiBrain effectively bridges the gap between expert-level real-time analytics and practical, user-friendly instructions. Extensive validation across a real-world operational e-bike battery-swap network demonstrates LiBrain's exceptional capabilities, achieving a 95% adoption rate in hazardous alarm detection and 92% in battery-status prediction. In real application, LiBrain has processed over 500 million battery events, managed almost 10 million inquiries and 1 million alarms annually, and identified 10% of on-site batteries daily for proactive replacement, thereby maintaining operational safety and reliability.

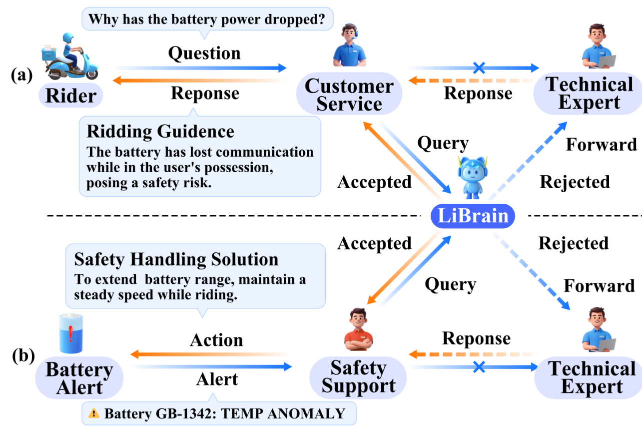

Figure 1. A comparison between conventional technical experts-dependent workflow and the proposed LiBrain-enhanced workflow in two typical real-world scenarios. In scenario (a), when having a question about battery usage such as power dropping, a rider typically contacts the customer service who rely on technical experts for professional guidance. Similarly, in scenario (b), when a battery alert such as temperature anomaly is triggered, the safety support also needs to consult technical experts for safety handling solutions. In practice, the proposed LiBrain-powered system autonomously resolves most user inquiries and battery alerts, significantly reducing human resource costs.

Introduction

The rapid development of environment-friendly smart cities has driven a substantial increase in the adoption of electric bicycles (e-bikes) as a sustainable, zero-emission mode of urban transportation. In China, the e-bike population has exceeded 350 million, establishing them as a dominant mode for daily commuting and short-distance travel. Among various power solutions, lithium-ion (Li-ion) batteries have become a primary energy source for e-bikes due to their high energy density, long cycle life, and robust power output. This preference is especially prominent in the on-demand delivery industry, where riders require extended operational ranges to support long delivery distances. To meet such requirements, battery-swap services have expanded rapidly in recent years, further solidifying the central role of Li-ion batteries in e-bike energy supply (Ding et al. 2025b).

Maintaining optimal user experience and operational safety is paramount for battery-powered e-bikes. From the user experience perspective, this necessitates accurate forecasts of battery performance and intelligent operational guidance tailored to riders. For instance, precise remaining range prediction facilitates efficient route planning (Li et al.

*Corresponding authors.

2024b,a), while actionable guidance such as avoiding overcharging and deep discharging can substantially extend battery lifespan, enhancing both reliability and riding experience. From the safety perspective, failures in Li-ion batteries can lead to hazardous events including fires and explosions, posing serious threats to users and property. Consequently, implementing timely and accurate fault diagnosis, coupled with a hierarchical fault management framework, is crucial for early anomaly detection and risk mitigation.

Traditional operational workflows rely heavily on the availability and expertise of technical experts to address user inquiries and Li-ion battery fault alerts, as illustrated by two scenarios in Fig. 1. In Fig. 1(a), when a rider raises questions regarding Li-ion battery usage—such as handling unexpected power shutdowns or assessing remaining riding range—they typically contact customer service. The representative forwards the query to a technical expert (annotated by solid blue arrow with "×"), obtains the response, and returns it to the rider. Similarly, in Fig. 1(b), when a battery fault alert is triggered, the message is sent to safety support staff, who consult a technical expert to obtain validated safety solutions and then act. Therefore, this traditional mode incurs significant human resource costs, particularly on technical personnel who must possess multi-dimensional battery expertise for accurate, instant responses.

To overcome these limitations, we present LiBrain, a Large Language Model (LLM)-powered intelligent agent for Li-ion battery detection. LiBrain is designed to comprehensively analyze multi-dimensional battery performances and generate intelligent actionable guidelines or safety management solutions. As shown in Fig. 1, in the new LiBrain-augmented operational mode, customer service and safety support staff can first query LiBrain to obtain immediate responses. If the responses are deemed valid, they will be accepted to answer user inquiries or take emergency safety measures; otherwise they will be sent to the technical experts for further consultation. The proposed approach reduces the workload of the technical experts, thereby enabling a more efficient maintenance of both user experience and battery safety. Apparently, the acceptance rate of LiBrain's responses by front-line operators (customer service or safety support staff) is directly determined by its capabilities, underscoring the importance of constructing a robust and high-performing LiBrain framework.

The core of LiBrain lies in real-time accurate analysis of the performance and safety status of Li-ion batteries, relying on real-time data acquisition pipeline and algorithmic modeling for accurate battery performance prediction. On the data acquisition side, LiBrain leverages an AIoT (Artificial Intelligence of Things)-based sensing and communication infrastructure, in which distributed IoT-enabled sensing units continuously monitor battery telemetry (e.g., voltage, current, temperature, state of charge (SOC)) and transmit them to edge or cloud processing nodes. On the modeling side, attributed to the development of deep learning, a series of works have been proposed to improve the performance of battery analysis. The majority of predictive algorithms predominantly focused on individual technical metrics such as state of health (SOH) or remaining useful life (RUL) (Wu et al. 2025; Guo et al. 2023; Tomar et al. 2025). Graph neural network-based approach has also been explored to estimate SOH by capturing spatio-temporal degradation dynamics in partial discharging curves of Li-ion batteries (Peng, Liu, and Li 2025). In addition, anomaly detection has also been investigated for battery safety assurance (Cao et al. 2025). A novel deep sequence model approach was proposed by (Zheng et al. 2023) to detect anomalies in charging sequences within the Industrial Internet of Things (IIoT) framework. Nevertheless, these methods are limited in their capacity to capture multi-dimensional performances of Li-ion batteries simultaneously.

The emergence of general-purpose LLMs such as ChatGPT (OpenAI 2024) and DeepSeek (DeepSeek-AI et al. 2025) has created new opportunities, yet their performance in specialized domains such as battery diagnostics remains inadequate. Three primary enhancement approaches have emerged: domain-specific pretraining, task-specific fine-tuning, and retrieval-augmented generation (RAG) coupled with curated knowledge bases. For example, recent works on pretrained foundational models for Li-ion battery, LLiM (Ding et al. 2025a) and LiPM (Li et al. 2025), have demonstrated that pretraining on large-scale, domain-specific Li-ion battery datasets significantly improves downstream task performance, especially when integrated with task-specific fine-tuning modules. Furthermore, the RAG framework enhances the capabilities of domain-specific LLMs by employing a retrieval–augmentation–generation pipeline that retrieves relevant domain knowledge from curated knowledge bases and integrate it into input prompts (Gao et al. 2024; Knollmeyer, Caymazer, and Grossmann 2025). Building on this paradigm, LiBrain utilizes LLiM as a backbone to effectively address six core Li-ion battery analytics tasks spanning safety and performance prediction. In addition, LiBrain leverages RAG technique to retrieve relevant information from a curated knowledge base on Li-ion battery, augment the predictive analytics generated by LLiM and generate user-tailored actionable guidelines.

LiBrain is an LLM-powered intelligent agent for Li-ion battery diagnostics, integrating real-time battery telemetry, high-accuracy performance prediction, and RAG over a curated domain-specific knowledge base. Leveraging the reasoning and decision-making capabilities of LLMs, LiBrain is able to conduct comprehensive battery assessments, generate user-friendly operation suggestions, and perform intelligent safety management, thereby ensuring both the operational performance and safety of Li-ion batteries. To summarize, our work makes the following contributions:

- We significantly extend the predictive tasks of a pretrained Li-ion battery foundation model to six downstream tasks including battery safety and performance.
- We develop the first domain-specific knowledge base for Li-ion battery diagnostics and utilize RAG pipeline to effectively fuse real-time battery data, predictive performance analytics and retrieved battery knowledge to generate accurate actionable guidelines tailored for users.
- We validate the efficacy of LiBrain in two representative real-world scenarios in a commercial battery swapping

network for e-bikes where LiBrain achieves over 92% success rate in battery-status prediction task and 95% in hazardous alarm detection task.

The remainder of this paper is organized as follows. In next section, we elaborate on the architecture of the LiBrain framework. Following that, we present a comprehensive experimental evaluation, including the performance of the core diagnostic engine and an ablation study of the framework's components. We then detail the system's real-world deployment, covering its application scenarios, technical configuration, user interface, and illustrative case studies. Finally, we conclude the paper and outline directions for future work.

LiBrain Framework

In this section, we present the framework of LiBrain, as illustrated in Fig. 2. LiBrain synergistically integrates three core components to ensure it functions as an intelligent battery management assistant. The following subsections elaborate on specific functionality of each component.

AIoT-enbaled Battery Data Acquisition

In order to monitor the highly variable external operational conditions and internal status of batteries, LiBrain constructs an AIoT-based sensory network to collect real-time battery data. Specifically, the network consists of specialized charging cabinets and Li-ion batteries. Each battery consists of a battery pack, comprising multiple cells, and a Battery Management System (BMS), which is crucial for ensuring both safety and operational efficiency, as it continuously monitors key parameters, including current, temperature, and cell-level voltage, through an array of embedded sensors.

This stream of high-fidelity sensor data is transmitted to a cloud-based data processing pipeline. The cloud backend performs real-time analytics to detect operational anomalies and dispatches control commands back to the BMS when necessary (e.g., initiating a charging cut-off). Although of high value, the collected raw time-series data is mostly structured tabular data lacking semantic interpretability. For instance, the collected data record "SOC: 70%" is likely abstruse to general users due to the domain-specific terminology SOC and the value 70%. Even general-purpose LLMs may fail to completely capture the insights of these simplified terms if they are not pretrained on domain knowledge. Hence, it is necessary to transform these domain-specific technical data into natural language phrases comprehensible to regular users. This semantically grounded real-time data is a key output of this module and is utilized in the final LLM-powered reasoning step. Concretely, we programmatically convert the original "SOC: 70%" into the textual description "The battery's current state of charge is 70%." Similarly, all historical and real-time battery data are transformed into textually formatted battery status descriptions with higher interpretability for both general users and LLMs. Subsequently, the raw battery data and transformed textual descriptions are delivered to the following LLM-powered forecasting module for further prediction of battery performances.

LLiM-Powered Battery Performance Forecasting

While the AIoT system provides instantaneous, real-time data snapshots, the core predictive and analysis task is performed by LLiM module (Ding et al. 2025a), a foundation model specifically engineered for Li-ion battery analytics. By processing extensive historical battery time-series data, LLiM learns the universal principles of battery degradation, including characteristic voltage curves and performance decay in operation. This deep temporal understanding allows it to identify latent patterns and emerging risks that are impossible to detect from real-time data alone, providing a comprehensive assessment of a battery's health and future state. With a one-billion-parameter, transformer-based architecture, LLiM is pre-trained on an extensive corpus of multivariate time-series data, capturing the operational dynamics of batteries under diverse real-world conditions. This pretraining phase endows the model with a profound understanding of the intricate electrochemical and thermal patterns inherent to battery degradation and performance.

To extend LLiM into a multi-task diagnostic engine, we append and fine-tune specialized prediction heads onto the pre-trained backbone for six tasks spanning both safety (anomaly detection, internal short circuit (ISC), and bulging detection) and performance (SOH, RUL, and remaining range estimation). The input to the model is a preprocessed feature vector, $\boldsymbol{x}$, comprising 71 distinct operational features, including but not limited to cell voltage, current, and temperature. First, the input feature vector is passed through the LLiM backbone to generate a shared high-level representation, $\boldsymbol{h}$:

$$\boldsymbol{h} = \mathrm{LLiM}_{\mathrm{backbone}}(\boldsymbol{x}; \boldsymbol{\theta_b}) \tag{1}$$

This representation is then passed to the task-specific heads. For instance, the safety tasks are formulated as binary classifications:

$$\boldsymbol{p(y = 1|x)} = \mathrm{LLiM}_{\mathrm{head}}(\boldsymbol{x}; \boldsymbol{\theta_h}) \tag{2}$$

and the performance tasks are formulated as regressions:

$$\boldsymbol{\hat{y}} = \mathrm{LLiM}_{\mathrm{head}}(\boldsymbol{h}; \boldsymbol{\theta_h}) \tag{3}$$

The LLiM backbone network is shared, while the LLiM head networks are task-specific, with $\boldsymbol{\theta_b}$ and $\boldsymbol{\theta_h}$ denoting their corresponding network parameters. For instance, in anomaly detection, this feature vector is processed by a fine-tuned LLiM variant to generate a specific risk classification. A parallel methodology is employed for other predictive targets. The model's outputs constitute a comprehensive diagnostic profile, providing binary classifications for critical safety events and quantitative estimations for key performance indicators.

Collectively, the real-time battery state data from the AIoT module and the prognostic assessments from LLiM provide the comprehensive technical information for the subsequent reasoning stage, where LLiM's output serves as a key intermediate representation encapsulating predictions of the battery's future state.

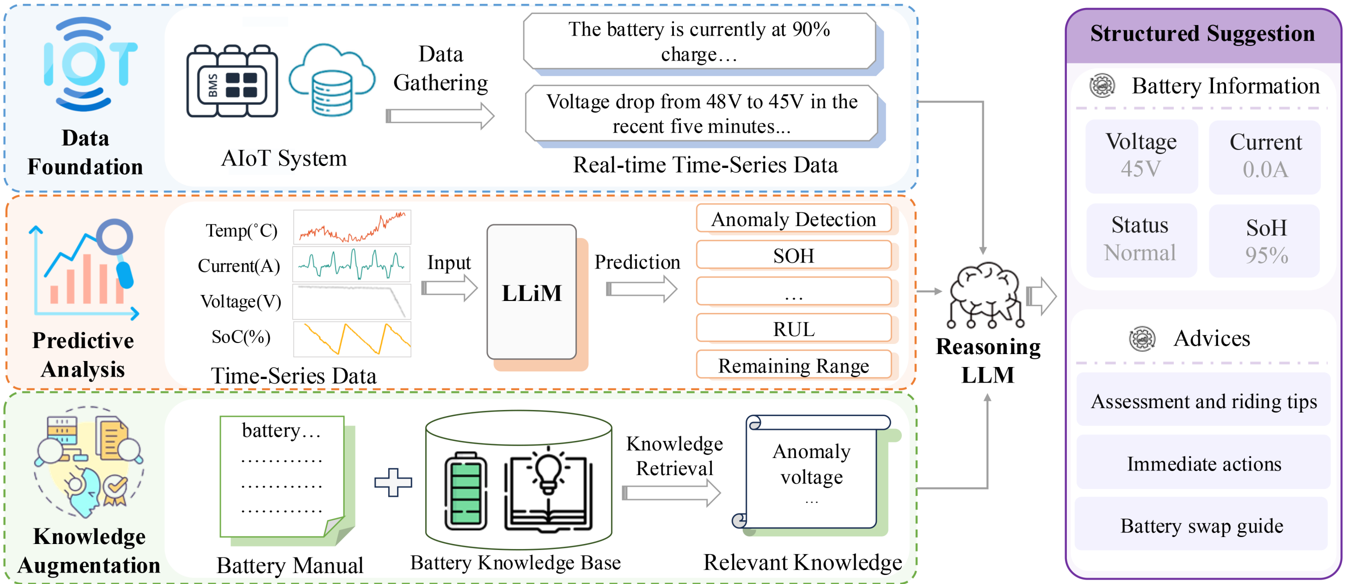

Figure 2. The overall architecture of the LiBrain framework. The workflow initiates with an AIoT System (top) to collect real-time battery time-series data, which is then processed and fed into the LLiM prediction module (middle) to generate multi-dimensional forecasts spanning both battery safety and performance. The forecasting results are utilized to retrieve relevant information from the curated domain-specific knowledge base (bottom) with an RAG pipeline. Finally, a reasoning LLM synthesizes all the information to produce a comprehensive structured suggestion for users.

User-tailored Knowledge Base Augmentation

The real-time battery data and performance predictions from LLiM usually contain domain-specific technical terminology. This presents a significant barrier for riders to comprehend and take appropriate actions, potentially leading to battery degradation and safety risks. To bridge this knowledge gap, LiBrain incorporates a meticulously curated Li-ion battery knowledge base and leverages an RAG module to translate complex technical diagnostics into actionable, user-friendly guidance.

We construct a high-quality knowledge base including both a fundamental document and key feature dictionary of Li-ion batteries. The fundamental document is built from technical files such as manufacturer specification sheets, introducing basic information of Li-ion batteries. The key feature dictionary is composed of 188 curated domain-specific knowledge entries on Li-ion batteries. Each entry consists of a specific battery metric, its definition and operational instructions. For example, the entry on SOH specifies: (1) SOH denotes the ratio of the battery's current capacity to its rated (nominal) capacity; it is 100% for a new cell and declines progressively with aging during use; (2) when SOH falls below 70%, the battery is considered to have reached end of life and should be retired, as both its energy and power performance have degraded substantially.

Based on the knowledge base, we employ a multi-step RAG strategy to iteratively integrate multi-dimensional battery information for reasoning. A single-step RAG pipeline is insufficient due to the complexity of Li-ion battery diagnostics, which involves synthesizing multi-dimensional information from static attributes to dynamic real-time data from the AIoT network, and forecasting analytics from LLiM. Prior work has explored the insufficiency of one-step retrieve-and-read approach and proposed various solutions. For example, (Shao et al. 2023) proposed that structuring the LLM interaction into a series of iterative calls markedly enhanced complex reasoning performance. (Trivedi et al. 2023) integrated chain-of-thought (CoT) paradigm with RAG and proposed an interleaving retrieval with multiple CoT steps to implement reasoning RAG. (Xiong et al. 2024) proposed an iterative RAG model that uses follow-up queries to improve the accuracy of medical QA in complex medical scenarios. Inspired by these methods, we propose a multi-step RAG approach that is essential to accurately process and integrate diverse battery information, yielding a more comprehensive diagnosis.

Specifically, we categorize the acquired battery information into three classes: basic information, anomaly information, and performance information. During the retrieval phase, we employ BM25 method (Robertson and Zaragoza 2009) to select the most relevant knowledge entries to rider inquiries or safety alert messages, producing an initial output. Subsequently, we feed the initial output together with the input queries and anomaly information into a second RAG step to obtain an intermediate result. Finally, we perform a third RAG step using the intermediate result, the input queries and the performance information to generate the final diagnostics and user guidelines. The mathematical expression of the process is formulated as below. Given a user query q, retrieve the most relevant document(s) d^* or subset

$\mathcal{D}^* \subseteq \mathcal{D}$ from a knowledge base $\mathcal{D} = \{d_1, d_2, \dots, d_N\}$:

$$\mathcal{D}^* = \text{Retrieve}(q, \mathcal{D}) = \underset{d \in \mathcal{D}}{\text{arg top-k}}\, \text{BM25}(q, d) \tag{4}$$

where $\text{BM25}(q, d)$ is the BM25 score of document d given query q and arg top-k returns the top k most related documents with highest BM25 scores. At step t, the retrieved documents $\mathcal{D}_t^*$, the output a_{t-1} at previous step, and the original query q are fed into a generative model (e.g., an LLM) to produce the final answer a_t as:

$$a_t = \text{LLM}(q, a_{t-1}, \mathcal{D}_t^*) \tag{5}$$

We also designed scenario-specific prompts in each step of RAG to improve reasoning ability of LLM on different types of battery information. This multi-stage procedure ensures that each RAG step focuses on a single feature type, avoiding mutual interference while iteratively aggregating evidence, producing comprehensive and accurate outcomes.

Experiment

To evaluate the comprehensive diagnostic capability of LiBrain in both battery performance and safety domains, we conducted a series of experiments on real-world datasets. This section details the dataset, the evaluation metrics used, and a comprehensive analysis of the results.

Dataset

We constructed datasets encompassing both of the application scenarios described in Fig. 1: rider-initiated inquiries and system-initiated alerts. For every case in the dataset, the input sample includes the trigger event and the associated battery's operational data. To establish a ground truth, each case was analyzed and labeled by technical experts with a standard diagnostic report and recommended action.

- Scenario A: rider-initiated inquiries. This dataset contains 104 historical cases, where the trigger event is the rider's textual inquiry.
- Scenario B: system-initiated alerts. This dataset consists of 108 historical cases, where the trigger event is the system-generated alert data.

Metrics

We define our primary evaluation metric as the Adoption Rate, which holistically measures the framework's real-world effectiveness. We classify an event as a successful adoption when a customer service representative, using the LiBrain-generated guidance, successfully resolves a case—whether a rider-initiated inquiry (Scenario A) or a system-initiated alert (Scenario B). Conversely, any case that must be escalated to a technical expert for final judgment is classified as a non-adoption. Formally, the customer service adoption rate is calculated as:

$$\textit{Adoption Rate} = 1 - \frac{\textit{non-adoption}}{\textit{total}} \tag{6}$$

where *total* denotes the total number of diagnostic queries processed by customer service representatives and *non-adoption* denotes the number of cases subsequently escalated to a technical expert.

LLiM Forecasting Performance Evaluation

To validate the LLiM's forecasting capabilities, we benchmarked it against three state-of-the-art time-series models: DLinear (Zeng et al. 2023), a powerful linear model; iTransformer (Liu et al. 2023), which inverts the Transformer's attention mechanism to focus on multivariate correlations; and PatchTST (Nie et al. 2023), which uses a patching technique to learn local semantic information. We assess performance across all six diagnostic tasks using standard metrics. For the classification tasks of anomaly, bulging, and ISC detection, we use precision and recall to measure predictive accuracy and completeness. For the regression tasks of remaining range, SOH, and RUL estimation, we use Mean Absolute Error (MAE) and Mean Squared Error (MSE) to evaluate the magnitude of the prediction errors, where lower values are better.

The results presented in Table 1 demonstrate the superiority of LLiM across all six diagnostic tasks. In the safety-critical classification tasks including anomaly, bulging and ISC detection, LLiM achieves the best performance in both precison and recall among all methods. Particularly in anomaly detection, LLiM achieves precision of 99.2% and recall of 95.9%, exceeding all other methods by 6.7% to 35.6%. The performance gap is even more pronounced in the performance-related regression tasks, where LLiM achieves a significant reduction in both MAE and MSE, especially in RUL estimation where LLiM outperforms the second best model PatchTST by 55.8% in MAE and 78.2% in MSE. This superior performance of LLiM can be attributed to the pre-training on a massive corpus of Li-ion battery time-series data, which endows it with a deep, foundational understanding of the underlying electrochemical dynamics that simpler, general-purpose forecasting models fail to capture.

Ablation Study

We evaluate the performance of the LiBrain framework through a series of comparative and ablation studies designed to validate its effectiveness and the contribution of its core components.

- Full LiBrain Framework: To select the optimal backbone for our framework and investigate its general effectiveness when integrated with various models, we conduct a comparative study of LiBrain's performance with several state-of-the-art LLMs, including models from the LLaMa, DeepSeek, GPT, and Qwen.
- Ablation Study: To isolate and quantify the contribution of each component, we test three ablated versions of our primary model: one without the LLiM module (w/o LLiM), one without the Knowledge Base (w/o KB), and another with LLM backbone only (LLM-Only).

Our analysis of the experimental results, presented in Table 2 and Table 3, yields two primary conclusions regarding the framework's effectiveness and design. Our comparative study demonstrates the effectiveness of the LiBrain framework and provides clear insights into the performance of different LLM backbones. The Deepseek-R1-Distill-LLaMA3.3-70B (DRDL-70B) emerged as the top-performing model, achieving the highest adoption rate of

Model	Anomalies		Bulging		ISC		Remaining Range		SOH		RUL	
	P	R	P	R	P	R	MAE	MSE	MAE	MSE	MAE	MSE
DLinear	0.929	0.916	0.380	0.575	0.416	0.410	7.925	106.828	2.491	11.623	55.369	4435.100
iTransformer	0.732	0.574	0.585	0.511	0.699	**0.606**	9.255	158.993	1.853	7.342	48.103	3292.861
PatchTST	0.929	0.868	0.480	0.766	0.660	0.573	7.930	100.292	2.158	8.817	24.943	824.654
LLiM	**0.992**	**0.959**	**0.800**	**0.800**	**0.750**	**0.606**	**2.015**	**11.386**	**1.026**	**3.550**	**11.031**	**179.927**

Table 1: Performance comparison of LLiM against baseline models across various battery diagnostic tasks. The best performance for each metric is highlighted in bold. P and R denote precision and recall, respectively.

LLM Backbone	Adoption Rate	
	Scenario A	Scenario B
LiBrain (Qwen3-32B)	77.45%	87.04%
LiBrain (LLaMA3.3-70B)	**92.31%**	94.44%
LiBrain (GPT-4o-Mini)	89.42%	91.67%
LiBrain (Deepseek-V3-671B)	91.35%	87.96%
LiBrain (DRDL-70B)	**92.31%**	**95.37%**

Table 2: Comparative study of the full LiBrain framework powered by different LLM backbones. This experiment benchmarks our chosen model against other state-of-the-art alternatives, with the top-performing result in each column highlighted in bold.

Configuration	Adoption Rate	
	Scenario A	Scenario B
LiBrain (DRDL-70B, Full)	92.31%	95.37%
LiBrain (w/o LLiM)	75.00%	78.00%
LiBrain (w/o KB)	71.15%	74.01%
LLM-Only (DRDL-70B)	47.12%	56.00%

Table 3: Ablation study on the selected DRDL-70B model. This experiment demonstrates the performance contribution of the LLiM and Knowledge Base (KB) components.

95.37% in the safety-critical Scenario B and 92.31% in Scenario A. Given its superior performance in handling both rider-initiated inquiries and system-initiated alerts, we selected it for our final deployment to maximize safety and reliability.

The standard LLaMA-3.3-70B model also performed well, matching the same highest adoption rate in Scenario A (92.31%) and achieving a strong 94.44% in Scenario B. Similarly, the GPT-4o-Mini model achieved robust adoption rates exceeding 89% in both scenarios. This validates that the LiBrain framework can effectively leverage various modern LLMs without necessitating a large-scale model, making it a flexible and practical solution for real-world deployment. Despite being the smallest model tested, Qwen-32B still achieved a respectable 87.04% adoption rate in the complex Scenario B, which further highlights the efficacy of LiBrain framework.

A notable result concerns DeepSeek-V3: despite its 671B-parameter capacity, its performance was solid yet below the leading models. DeepSeek-V3's limited reasoning ability appears to reduce its ability to exploit our structured Knowledge Base, while the Mixture-of-Experts (MoE) topology activates only 37B parameters per token, constraining its effective per-token compute to a level potentially lower than other smaller dense models. Considered jointly, these observations indicate that for our task, reasoning capability and per-token compute matter more than parameter scale alone.

Our ablation study (Table 3) provides a quantitative validation of each component's indispensable contribution to our synergistic architecture. The full LiBrain configuration significantly outperforms all ablated versions. The removal of the LLiM module results in a notable performance drop to 75.00% (Scenario A) and 78.00% (Scenario B), confirming that LLiM's time-series-aware forecasting is essential for a comprehensive diagnosis. The degradation in both scenarios is expected, as LLiM is responsible for predicting six critical indicators spanning both performance (SOH, RUL, remaining range), which are often central to rider inquiries, and safety (anomalies, ISC, bulging detection), which are the primary triggers for system alerts. Without these deep, prognostic inputs, the framework's ability to accurately assess the battery's condition is fundamentally compromised. The impact is also pronounced when the knowledge base is removed, with adoption rates dropping sharply to 71.15% and 74.01%. This demonstrates the critical role of the domain-specific knowledge base. Without it, the framework loses its ability to translate complex technical data into accessible, actionable guidance. This gap between raw diagnostics and user-friendly recommendations likely leads to confusion and a lack of trust, causing the observed decline in adoption rates. The LLM-only model achieves the lowest adoption rates, which decrease markedly to only 47.12% in Scenario A and 56.00% in Scenario B. This provides strong evidence that a naive application of a general-purpose LLM is insufficient for this complex, specialized task and confirms that our integrated, multi-component approach is essential for a comprehensive battery diagnostic.

Deployment

Application Scenarios

This section presents details of the real-world application scenarios of the LiBrain framework and its graphical user interfaces (GUIs) in a commercial battery swap network for online battery service and onsite battery detection. In the battery swap network, increasing usage by riders leads to faster battery performance degradation, leading to more safety alerts and rider inquiries on battery usage. We have fully deployed the LiBrain framework into our online battery service, where it is actively used by customer service

Figure 3. The GUI of the workflows of LiBrain applied in online battery service and onsite battery detection scenarios. (a) For online battery service in a battery swap network, riders inquire customer services via the mobile application, where the representative consults LiBrain for battery diagnostics and user-tailored advice; (b) For onsite battery detection, LiBrain collects real-time battery data from the detection device and generates a comprehensive diagnostic report.

representatives via a dedicated GUI, as shown in Fig. 3(a). Concretely, in this example, a rider contacts customer service reporting an issue of shorter range Fig. 3(a1) and (a2). The representative organizes and forwards the inquiry to LiBrain and obtains a diagnostic report within the representative's interface Fig. 3(a3). The report displays both the basic information (battery ID, voltage, current, temperature) and core technical indicators (SOH, SOC, remaining range), confirming the battery's status is normal. It also provides clear and helpful advice that the rider should replace the current battery with a fully charged one within 13km of remaining range. After reviewing the report, the representative sends the complete, validated guidelines to the rider, efficiently resolving the inquiry Fig. 3(a3).

In the onsite battery detection scenario, e-bike users ride their vehicles to designated inspection stations, where technicians perform Li-ion battery diagnostics, as shown in Fig. 3(b1). The technicians connect detection equipment to the vehicle, conducting charge-discharge cycling while the device collects real-time battery data. Upon acquiring comprehensive metrics, LiBrain automatically generates a diagnostic report containing three sections: Battery Information, Inspection Indicators, and Recommendations, as shown in Fig. 3(b2). Key indicators like SOH and battery status are predicted precisely using LiBrain's LLiM. The Recommendations section, derived from LiBrain's battery knowledge base, provides usage guidance to optimize Li-ion battery performance and extend service life.

Deployed within the e-bike battery swap system, LiBrain has facilitated the intelligent inspection of over 500,000 batteries for hundreds of millions of times. Annually, the system handles more than 10 million guidance to rider inquiries and processes over 1 million battery fault alarms, effectively ensuring both user experience and battery safety. More importantly, it can process over 100 Queries Per Second (QPS) in real time simultaneously, significantly enhancing customer service efficiency. In onsite battery detection applications, LiBrain has evaluated over 50 Li-ion batteries at inspection stations per day, identifying approximately 10% as non-compliant with safety standards. The diagnostic core metrics and actionable recommendations empower users to preemptively identify battery hazards, enabling timely replacement of compromised units—a critical safeguard for community safety and critical infrastructure protection.

Deployment Configuration

The LiBrain framework is developed using Python 3.12 and is deployed on a local enterprise server cluster. This cluster is equipped with multiple NVIDIA A100 GPUs, providing computational power for both model training and inference. The LLiM prediction and LLM reasoning services are co-located in this environment for low-latency communication, which is essential for providing real-time diagnostic responses. We deployed the open source LLM with vLLM framework (Kwon et al. 2023) and leveraged advanced techniques such as tensor parallelism, PagedAtten-

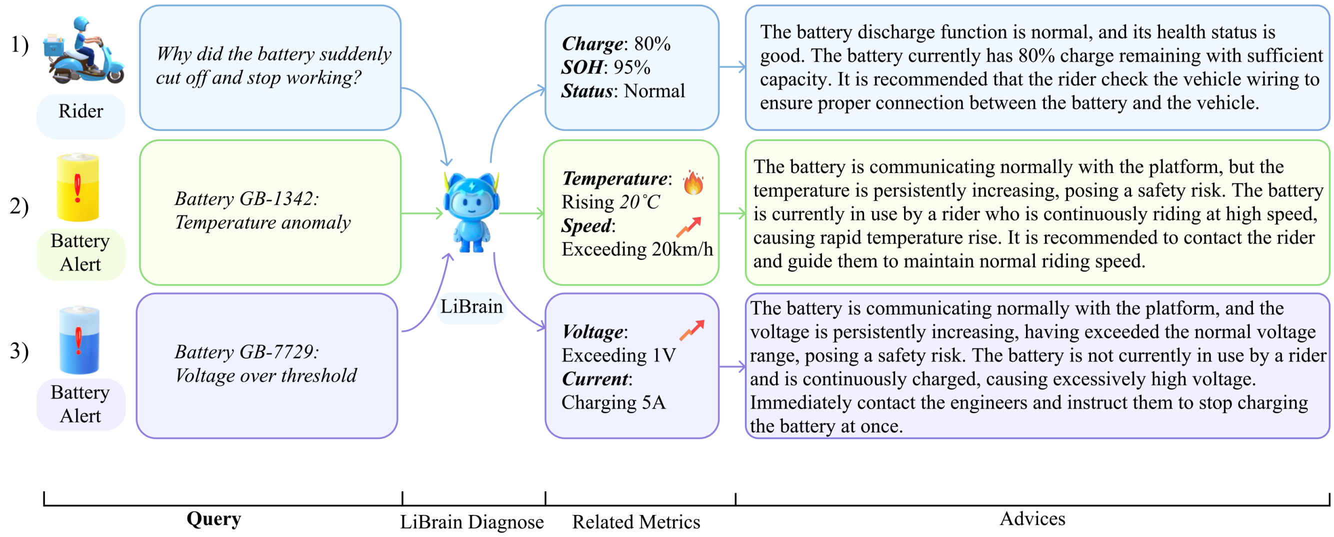

Figure 4. Three representative case studies of LiBrain implementation in e-bike battery swap network. Case 1 represents a typical performance-related scenario where the rider inquires LiBrain for battery diagnostics and actionable guidance. Cases 2 and 3 represent a safety-related scenario where LiBrain analyzes the causes and proposes solutions for system-triggered alerts.

tion and dynamic batching to maximize GPU utilization and handle concurrent requests effectively. Other techniques such as quantization options like GPTQ (Frantar et al. 2023) can further reduce memory usage without significant performance loss. The operational flow commences when either a rider-initiated inquiry or a system-initiated alert triggers the LiBrain backend, which subsequently orchestrates the core modules to generate and deliver the comprehensive diagnostic report.

Case Study

Furthermore, we present LiBrain's specific issues and recommendations in two application scenarios, as shown in Fig. 4. In Scenario A, approximately 80% of rider inquiries concern battery performance during cycling, including power loss, insufficient range, and failure to discharge. For example, in Case 1 of Fig. 4, when receiving an inquiry about battery power-off, LiBrain first assesses the battery's SOH, the primary metric for battery performance. Based on the overall battery condition, it further analyzes specific dimensional metrics-such as capacity, voltage consistency, and potential abnormalities-impacting performance. It then determines whether the issue stems from the battery itself. If battery-related anomalies are confirmed, LiBrain provides a final diagnosis and recommendations. If all battery data appears normal, it guides the rider to inspect connection issues between the battery and the vehicle.

In Scenario B, the alarm system generates various alerts, with temperature and voltage anomalies being the most prevalent, accounting for 40% and 12% of alerts, respectively. Notably, during periods of high ambient temperatures, temperature-related anomalies can exceed 50%. The alarm system specifies the battery ID and root cause for each alert, exemplified by case 2 and 3. Upon receiving an alarm alert, LiBrain immediately queries the current status information of the corresponding battery, including real-time temperature, voltage, and communication status, then provides safety handling solutions. Given the critical safety implications, it heavily relies on key indicators calculated by the LLiM model - anomaly, bulging, and internal short circuit. If the alarmed battery is assessed as risk-free, LiBrain outputs the analysis results and dismisses the alarm. For batteries identified with potential risks, LiBrain generates corresponding safety protocols based on risk levels.

Conclusion

We present LiBrain, an LLM-powered, time-series-aware, retrieval-augmented framework that unifies battery safety assurance and performance optimization through three synergistic components: (1) a distributed IoT-enabled edge network for real-time monitoring, (2) a pretrained deep multi-task diagnostic engine for performance forecasting, and (3) a knowledge-base augmentation module that turns diagnostics into actionable guidance for e-bike users. Acting as an intelligent battery management assistant, LiBrain bridges expert analytics with practical instructions. Our validation on two typical scenarios in a large-scale, real-world e-bike battery swap network confirms its exceptional capabilities. LiBrain achieves adoption rate of 95% in hazardous alerts detection and 92% in battery-status prediction. In practice, LiBrain has already inspected 500 million-plus battery events, fielded 10 million annual rider inquiries and 1 million fault alarms, all while sustaining real-time throughput above 100 QPS. Daily, it flags 10 % of on-site batteries as unsafe, issuing actionable diagnostics that enable prompt replacement and safeguard both users and infrastructure.

References

Cao, R.; Zhang, Z.; Shi, R.; Lu, J.; Zheng, Y.; Sun, Y.; Liu, X.; and Yang, S. 2025. Model-constrained Deep Learning for Online Fault Diagnosis in Li-ion Batteries over Stochastic Conditions. *Nature Communications*, 16(1): 1651.

DeepSeek-AI; Liu, A.; Feng, B.; Xue, B.; et al. 2025. DeepSeek-V3 Technical Report. arXiv:2412.19437.

Ding, D.; Li, Z.; Luo, L.; Jin, M.; Zhu, B.; Zhong, Y.; Hu, J.; Cai, P.; and Hu, H. 2025a. Large Lithium-ion Battery Model for Secure Shared Electric Bike Battery in Smart Cities. *Nature Communications*, 16(1): 8415.

Ding, D.; Li, Z.; Zhang, J.; Liu, X.; Zhang, J.; Li, Y.; Cai, P.; Liu, J.; and Long, G. 2025b. eBaaS: AIoT-Enabled eBike Battery-Swap as a Service for Last-Mile Delivery. In *Proceedings of the ACM on Web Conference 2025*, WWW '25, 5045–5053. New York, NY, USA: Association for Computing Machinery. ISBN 9798400712746.

Frantar, E.; Ashkboos, S.; Hoefler, T.; and Alistarh, D. 2023. GPTQ: Accurate Post-Training Quantization for Generative Pre-trained Transformers. arXiv:2210.17323.

Gao, Y.; Xiong, Y.; Gao, X.; Jia, K.; Pan, J.; Bi, Y.; Dai, Y.; Sun, J.; Wang, M.; and Wang, H. 2024. Retrieval-Augmented Generation for Large Language Models: A Survey. arXiv:2312.10997.

Guo, F.; Huang, G.; Zhang, W.; Wen, A.; Li, T.; He, H.; Huang, H.; and Zhu, S. 2023. Lithium Battery State-of-Health Estimation Based on Sample Data Generation and Temporal Convolutional Neural Network. *Energies*, 16(24): 8010.

Knollmeyer, S.; Caymazer, O.; and Grossmann, D. 2025. Document GraphRAG: Knowledge Graph Enhanced Retrieval Augmented Generation for Document Question Answering Within the Manufacturing Domain. *Electronics*, 14(11): 2102.

Kwon, W.; Li, Z.; Zhuang, S.; Sheng, Y.; Zheng, L.; Yu, C. H.; Gonzalez, J. E.; Zhang, H.; and Stoica, I. 2023. Efficient Memory Management for Large Language Model Serving with PagedAttention. In *Proceedings of the ACM SIGOPS 29th Symposium on Operating Systems Principles*, 611–626.

Li, J.; Yang, Y.; Su, H.; Liu, J.; Chen, Y.; Zhang, J.; and Pan, L. 2025. LiPM: Foundation Model for Lithium-Ion Battery Analysis. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, KDD '25, 1412–1423. New York, NY, USA: Association for Computing Machinery.

Li, Z.; Liu, Y.; Zhou, C.; Liu, X.; Pan, X.; Cao, B.; and Wu, X. 2024a. Transformer-based Graph Neural Networks for Battery Range Prediction in AIoT Battery-Swap Services. In *2024 IEEE International Conference on Web Services*, 1168–1176.

Li, Z.; Ren, G.; Gu, Y.; Zhou, S.; Liu, X.; Huang, J.; and Li, M. 2024b. Real-time E-bike Route Planning with Battery Range Prediction. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, WSDM '24, 1070–1073. New York, NY, USA: Association for Computing Machinery.

Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; and Long, M. 2023. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. *arXiv:2310.06625*.

Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2023. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In *The Eleventh International Conference on Learning Representations*. OpenReview.net.

OpenAI. 2024. GPT-4o mini. Model documentation. Accessed 2025-08.

Peng, H.; Liu, C.; and Li, H. 2025. Large-Language-Model-Enabled Health Management for Internet of Batteries in Electric Vehicles. *IEEE Internet of Things Journal*, 12(6): 6082–6094.

Robertson, S.; and Zaragoza, H. 2009. The Probabilistic Relevance Framework: BM25 and Beyond. *Foundations and Trends in Information Retrieval*, 3(4): 333–389.

Shao, Z.; Gong, Y.; Shen, Y.; Huang, M.; Duan, N.; and Chen, W. 2023. Enhancing Retrieval-Augmented Large Language Models with Iterative Retrieval-Generation Synergy. In Bouamor, H.; Pino, J.; and Bali, K., eds., *Findings of the Association for Computational Linguistics: EMNLP 2023*, 9248–9274. Singapore: Association for Computational Linguistics.

Tomar, A.; Gupta, M.; Mittal, J.; Arya, A.; and Varshney, U. 2025. Prediction of SOH and RUL for Li-Ion Batteries in EV Based on AttentiveLSTM Multi-Task Model. *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, 6(4): 1733–1743.

Trivedi, H.; Balasubramanian, N.; Khot, T.; and Sabharwal, A. 2023. Interleaving Retrieval with Chain-of-Thought Reasoning for Knowledge-Intensive Multi-Step Questions. In Rogers, A.; Boyd-Graber, J.; and Okazaki, N., eds., *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 10014–10037. Toronto, Canada: Association for Computational Linguistics.

Wu, J.; Sun, Z.; Li, D.; He, W.; Yang, D.; Wu, Z.; Geng, X.; Yang, H.; Wang, H.; Hu, L.; Tu, H.; and He, X. 2025. Efficient Estimating and Clustering Lithium-ion Batteries with a Deep-Learning Approach. *Communications Engineering*, 4(1): 151.

Xiong, G.; Jin, Q.; Wang, X.; Zhang, M.; Lu, Z.; and Zhang, A. 2024. Improving Retrieval-augmented Generation in Medicine with Iterative Follow-up Questions. In *Biocomputing 2025: Proceedings of the Pacific Symposium*, 199–214. World Scientific.

Zeng, A.; Chen, M.; Zhang, L.; and Xu, Q. 2023. Are Transformers Effective for Time Series Forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, 9, 11121–11128. AAAI Press.

Zheng, L.; Ding, D.; Li, Z.; Gao, J.; Xiao, J.; Chen, H.; Dustdar, S.; and Zhang, J. 2023. Anomaly Detection in Battery Charging Systems: A Deep Sequence Model Approach. In *2023 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking*, 587–594.