

# Large lithium-ion battery model for secure shared electric bike battery in smart cities

Received: 14 August 2024

Accepted: 26 August 2025

Published online: 25 September 2025

 Check for updates

Donghui Ding<sup>1,2</sup>, Zhao Li<sup>2,3</sup>✉, Linhao Luo<sup>4</sup>, Ming Jin<sup>5</sup>, Bin Zhu<sup>6</sup>✉, Yichen Zhong<sup>2</sup>, Junhao Hu<sup>2</sup>, Peng Cai<sup>1</sup>✉ & Huiqi Hu<sup>1</sup>

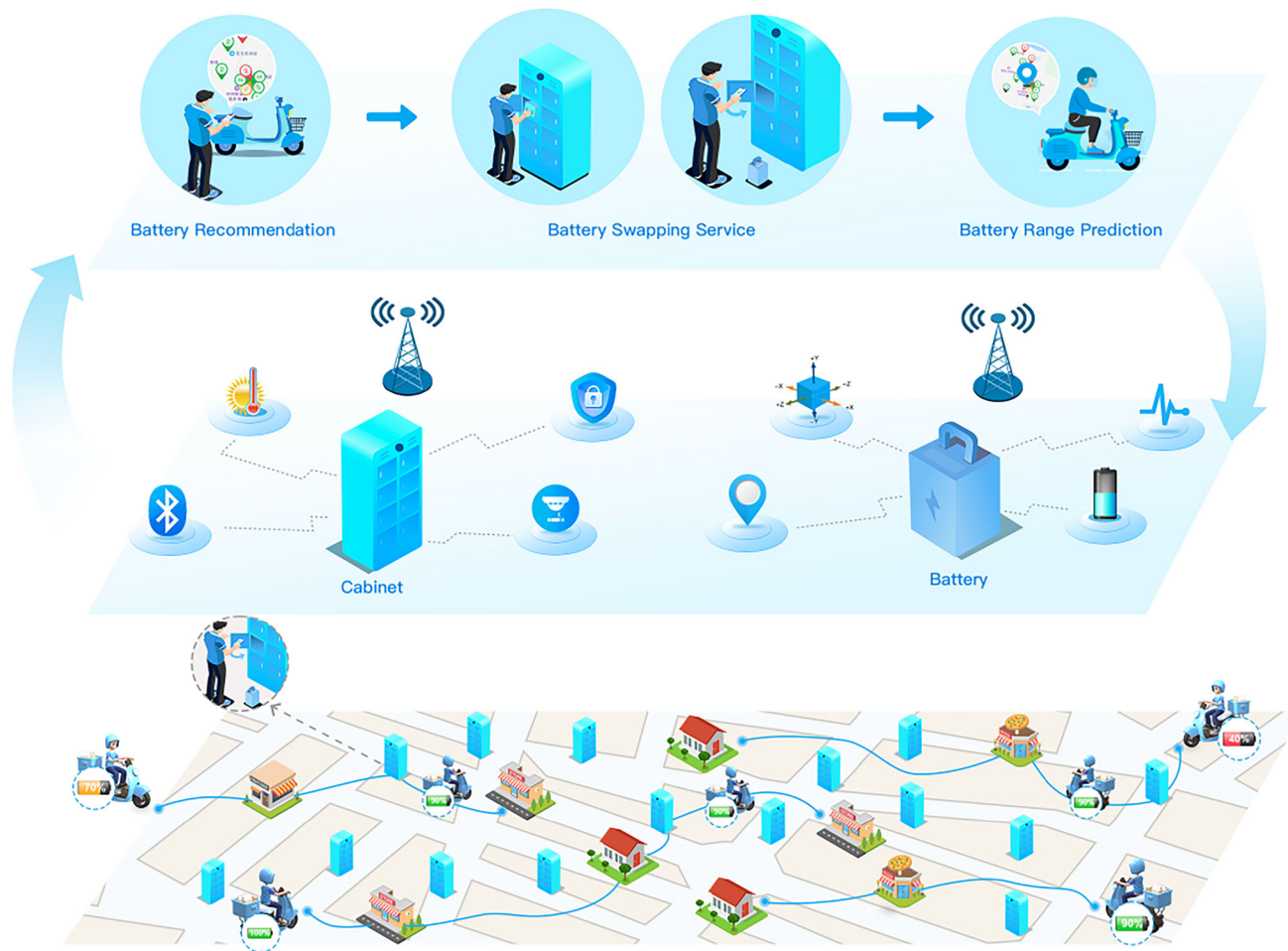
Electric bikes powered by lithium-ion batteries are increasingly used in smart cities to promote sustainable mobility and efficient delivery services. However, limited battery range and slow plug-in charging remain key challenges. Shared electric bike battery systems, facilitated by battery swapping stations, offer a promising solution by enabling quick and efficient battery replacements. However, their success hinges on accurate anomaly detection, battery health estimation and remain range prediction. These tasks remain challenging due to data scarcity, battery diversity and environmental variability. Here we show that a large-scale lithium-ion battery model trained on over ten million battery time series data enables robust and adaptable battery management across diverse real-world scenarios. The model learns complex battery behavior through unsupervised pretraining. Importantly, after efficient finetuning, the model significantly outperforms existing approaches in the three critical tasks. Deployed on cloud servers, our model enables real-time data processing, enhancing the safety, reliability and efficiency of battery swapping services. This advancement accelerates electric bike adoption, fostering sustainable urban mobility and green smart city development.

As cities worldwide strive for sustainable development, the integration of smart city technologies with zero-carbon transportation has become a key priority in combating climate change and the reduction of greenhouse gas emissions and urban air pollution<sup>1,2</sup>. Smart cities leverage cutting-edge technologies and data analytics to improve infrastructure, optimize resource management, and promote sustainable urban living<sup>3,4</sup>. Within this framework, zero-carbon solutions electric bikes (e-bikes) with lithium-ion batteries have experienced rapid growth worldwide in recent years, owing to their efficiency and eco-friendly nature<sup>5-7</sup>. E-bikes play essential roles in urban transportation, offering not only daily travel but also a reliable door-to-door service for local businesses such as on-demand delivery<sup>8</sup>. For instance, in China, social ownership of e-bikes users reached 340 million units in 2021 and the number of on-demand delivery workers is expected to exceed 10 million by 2025<sup>9</sup>.

Despite the rapid increase in the number of e-bikes, battery range limitations remain a significant concern for a large amount of e-bike

riders. A typical e-bike can only travel approximately 30–60 km on a single charge, significantly less than the daily travel distance of e-bike riders such as couriers and delivery personnel, which averages around 120 km<sup>10</sup>. Moreover, the plug-in charging is time-consuming and hampers the efficiency of on-demand delivery services<sup>11</sup>. Shared E-bike Battery (SEB) systems<sup>12</sup>, facilitated by battery-swapping stations, present a promising solution to offer quick lithium-ion battery swapping services for e-bikes<sup>13-15</sup>. As illustrated in Fig. 1, when the battery runs out, the process begins with e-bike riders receiving battery recommendations for nearby swapping stations, which consist mainly of swapping cabinets and batteries. Upon arrival at a swapping station, riders can hire a SEB and exchange their depleted batteries for fully charged ones. The swapping process only takes a few minutes, providing a convenient and efficient solution for e-bike riders. Additionally, remain range (RR) prediction technologies help riders plan their routes based on battery levels and swapping station locations. To meet

<sup>1</sup>School of Data Science and Engineering, East China Normal University, Shanghai, China. <sup>2</sup>Hangzhou Yugu Technology, Hangzhou, China. <sup>3</sup>Zhejiang Lab, Hangzhou, China. <sup>4</sup>Monash University, Melbourne, VIC, Australia. <sup>5</sup>Griffith University, Gold Coast, QLD, Australia. <sup>6</sup>Singapore Management University, Singapore, Singapore. ✉e-mail: lzjoey@gmail.com; binzhu@smu.edu.sg; pcaid@ecnu.edu.cn



**Fig. 1 | The Shared E-bike Battery (SEB) system is a cutting-edge Artificial Intelligence of Things (AIoT) solution that provides a swift lithium-ion battery swapping service for e-bike users.** At the base of the illustration, e-bike users have the convenience of swapping their depleted batteries for fully charged ones at battery swapping stations. This ensures uninterrupted riding experience as they can quickly replace their batteries and hit the road again. In the middle section of the illustration, the battery swapping stations are composed of battery swapping cabinets equipped with lithium-ion batteries. These stations act as the endpoints of the AIoT system, continuously gathering data through an array of sensors, such as

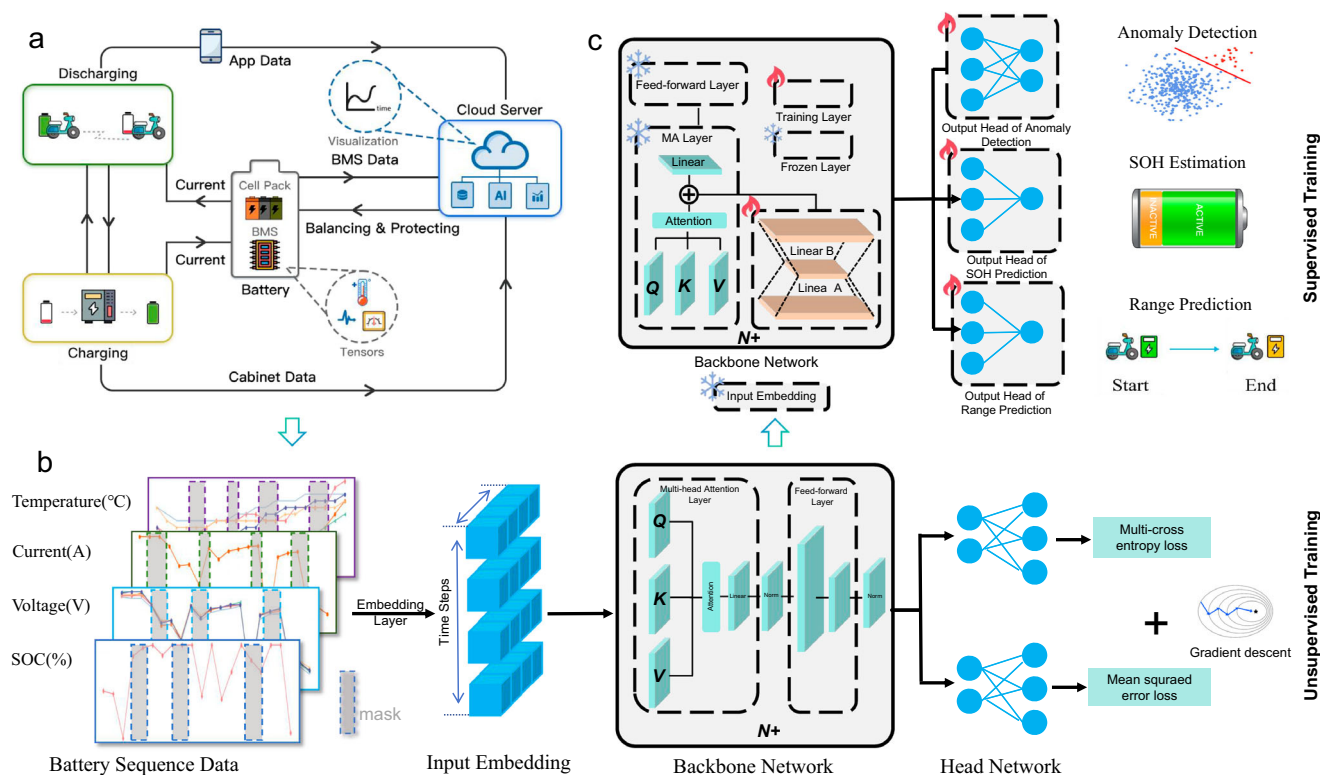
smoke and motion detectors, and maintaining seamless communication with the cloud infrastructure. At the top of the illustration, the cloud platform plays a pivotal role by offering a suite of services to users. These include battery recommendations, facilitating the battery swapping process, and providing predictions for the remain range of the e-bike, ensuring users have the information they need to plan their rides effectively. This system not only enhances the usability of e-bikes but also integrates smart technology to create a more connected and efficient e-mobility ecosystem.

the demands of e-bike riders, the SEB systems need to provide *safe* and *reliable* battery-swapping services, incorporating advanced technologies for real-time monitoring and management of battery status and station operations.

Existing works on lithium-ion battery safety and reliability have primarily focused on three tasks: anomaly detection<sup>16–20</sup>, state-of-health (SoH) estimation<sup>21–24</sup>, and range prediction<sup>25–27</sup>. Specifically, anomaly detection is significant for identifying and mitigating potential battery failures that could pose safety risks to e-bike riders. However, previous works<sup>16–20</sup> often rely on a substantial amount of anomaly data for training models, which may not always be feasible in practice. SoH estimation provides insights into the current condition and lifespan of batteries, enabling proactive maintenance and replacement to prevent unexpected failures and save costs. Yet, most SoH estimation models<sup>21,22</sup> cannot adapt to various types and states of lithium-ion batteries with real-time capacity estimation. RR prediction assists in route planning by forecasting how far a battery can travel before needing a swap. Existing methods<sup>23,24,26,27</sup>, however, struggle to generalize across different battery status time series data due to the complexity of the usage environment. In addition, the correlations

among these three tasks are still underexplored. For instance, anomalies in the battery can affect its state of health, which in turn impacts the RR, and vice versa. Understanding these interdependencies is crucial for developing comprehensive solutions that enhance both the safety and reliability of battery-swapping services.

To address these issues, we propose the **Large Lithium-ion Battery Model (LLiM)**, for the shared e-bike lithium-ion batteries management in battery-swapping stations. Building on foundation model architectures previously applied in natural language processing<sup>28–30</sup> and computer vision<sup>31,32</sup>, LLiM is pretrained with more than 10 million time series data collected during the battery usage in an unsupervised manner with mask modeling. The pretraining enables our method to capture complex patterns in the time series data of the battery. As a result, after pretraining, LLiM is finetuned to various downstream tasks, showing improved performance in anomaly detection, SoH estimation, and RR prediction across different types of lithium-ion batteries, compared to existing methods. Furthermore, experimental results demonstrate that prolonged battery usage not only degrades SoH but also elevates anomaly probability, with lithium-ion battery capacity exhibiting a positive correlation to remain riding range.



**Fig. 2 | Illustration of LLiM framework.** **a** Data processing: the battery includes a cell pack and a battery management system (BMS). The cell pack consists of multiple cells connected in series or parallel to store electricity and discharge during usage. BMS is responsible for managing the battery such as controlling the charging and discharging of battery to prevent battery from overcharging and over-discharging. It collects a variety of battery status with sensors, such as current, temperature and voltage. These data would be sent to the cloud server for further analysis. The cloud server would process the data and send instructions to the battery to protect battery usage. For example, the cloud server would send a signal to stop the battery charging and lock it in the cabinet if an anomaly is detected, and instantly notify the user through a mobile application (App) if necessary. **b** Pre-training: the pre-training of the LLiM is unsupervisedly conducted based on the

large-scale battery time series data. The battery sequence data would be first randomly masked out and represented as embeddings via embedding layers. The input embeddings are then fed into the transformer encoder which consists of multiple layers of self-attention and feed-forward neural networks. The encoder captures the complex patterns of the battery data by predicting the masked out values and optimizing with gradient descent. **c** Fine-tuning: after pre-training, LLiM is fine-tuned on the downstream tasks in a supervised manner, such as anomaly detection, SOH estimation, and remain range prediction. The fine-tuning process involves optimizing a classification or regression head that predicts the target values based on the output embeddings of the transformer encoder. The model is optimized with the supervised loss function to minimize the difference between the predicted values and the ground truth.

Importantly, we have scaled up LLiM to 1 billion parameters and provide public access via an Application Programming Interface (API). Deployed on cloud servers, LLiM integrates seamlessly with a diverse array of terminals, which operates as an Artificial Intelligence of Things (AIoT) framework. This integration ensures not only the operational efficiency but also the robust reliability of SEB services. By leveraging the cloud's scalability and the AIoT system's pervasive connectivity, LLiM stands as a pivotal asset in the realm of battery management, offering advanced analytics and real-time insights that are essential for maintaining optimal battery performance and safety. This would greatly facilitate the management of the SEB ecosystem and provide secure battery-swapping service for users, ultimately promoting the growth and popularity of e-bikes for sustainable urban transportation in smart cities. LLiM can be effortlessly tailored to accommodate a variety of lithium-ion battery types and sizes, capitalizing on the intrinsic characteristics of time series data. Our work may pave the way for further exploration into foundational models for lithium-ion battery management across diverse applications, such as electric vehicles, where data series exhibit varying lengths and cycles.

## Results

### Overview of LLiM for lithium-ion battery management

As shown in Fig. 2a, the proposed LLiM is a cloud-based foundation model designed for lithium-ion battery management in battery-

swapping stations. It integrates with edge-based battery management systems (BMS) to collect real-time battery data, enabling advanced analysis for tasks such as anomaly detection, SOH estimation, and RR prediction. By processing and visualizing this data, LLiM ensures the safety, reliability, and optimization of SEB systems, contributing to their overall operational efficiency.

The statistical data gathered from the battery usage is a multi-variable time series. It enables the profiling of the battery status and analysis of the battery behavior. To capture the complex patterns inherent in the battery data, we introduce a foundation model called LLiM that can be applied to various battery management tasks. The LLiM follows the architecture of Transformer, containing stacks of self-attention layers<sup>33</sup>. The time series data is encoded by the model to extract representations of the battery status. The extracted representations are then used to perform the specific tasks of battery management, such as anomaly detection, SOH estimation, and RR prediction. LLiM is trained in two stages: large-scale time-series pre-training and task-specific fine-tuning.

During the pre-training stage (Fig. 2b), we introduce a customized masking strategy and an unsupervised training objective. These techniques assist the model in handling the irregular and noisy nature of time series, enabling it to learn the underlying patterns within battery data. In the fine-tuning stage (Fig. 2c), we introduce a supervised training objective to adapt the model to various tasks. Specifically, we

**Table 1 | Experiment results on battery anomaly detection**

Methods	Accuracy	Precision	Recall	F1
TimesNet	0.500 ± 0.174	0.587 ± 0.250	0.525 ± 0.025	0.485 ± 0.085
DLinear	0.942 ± 0.004	0.941 ± 0.005	0.924 ± 0.008	0.933 ± 0.004
Pyraformer	0.969 ± 0.008	0.960 ± 0.017	0.934 ± 0.021	0.946 ± 0.019
FEDformer	0.758 ± 0.112	0.716 ± 0.150	0.668 ± 0.164	0.695 ± 0.152
Autoformer	0.661 ± 0.013	0.690 ± 0.147	0.511 ± 0.011	0.578 ± 0.047
PatchTST	0.952 ± 0.007	0.941 ± 0.010	0.929 ± 0.008	0.935 ± 0.003
LLiM(100M)	0.979 ± 0.001	0.981 ± 0.006	0.949 ± 0.006	0.966 ± 0.002
LLiM(1B)	<b>0.984 ± 0.002</b>	<b>0.988 ± 0.009</b>	<b>0.967 ± 0.004</b>	<b>0.975 ± 0.004</b>

The mean and standard deviation of accuracy, precision, recall and F1 score are reported. LLiM(100M) and LLiM(1B) indicate our model with 100 million and 1 billion parameters respectively. Bold values denote the maximum value within each column.

**Table 2 | Statistics of three task datasets**

Dataset Type	Cell Type	#Training	#Validation	#Testing
Anomaly Detection	FE32	1800	645	645
	FE30	5493	1800	1800
	FE29	6657	2262	2262
	Total	13,950	4650	4650
SoH Estimation	FE32	38,262	12,755	12,755
	FE30	6439	2147	2147
	FE29	2275	760	760
	TG29	1917	639	639
	DT29	636	213	213
	AMP29	1293	432	432
Remain Range Prediction	FE32	6783	1162	1162
	FE30	1111	370	370
	FE29	357	119	119
	TG29	1063	354	354
	DT29	1254	418	418
	AMP29	732	244	244

optimize a classification or regression head to predict the target values based on the output embeddings of the transformer encoder. The encoder is also updated with the LoRA technique<sup>34</sup> to transfer the knowledge learned from the pre-training stage to downstream tasks. This enables the model to learn the specific patterns of each task and make accurate predictions.

To collect diverse and extensive training data, we obtain the massive time series from real-world battery usage. The training data includes more than 10 million battery usage records, which cover different types of batteries and 71 different features, such as voltage, current, and temperature. This data provides a comprehensive view of the battery status and behaviors in real-world scenarios. We assess the effectiveness of our approach for detecting battery anomalies by examining key performance indicators such as accuracy, precision, recall, and the F1 score. Furthermore, for estimating the SoH of batteries and predicting the RR, we utilize Mean Absolute Error (MAE) and Mean Squared Error (MSE) as our metrics for evaluation. Compared to existing approaches, the results demonstrate that our method exhibits improved performance and generalization ability to different types of batteries.

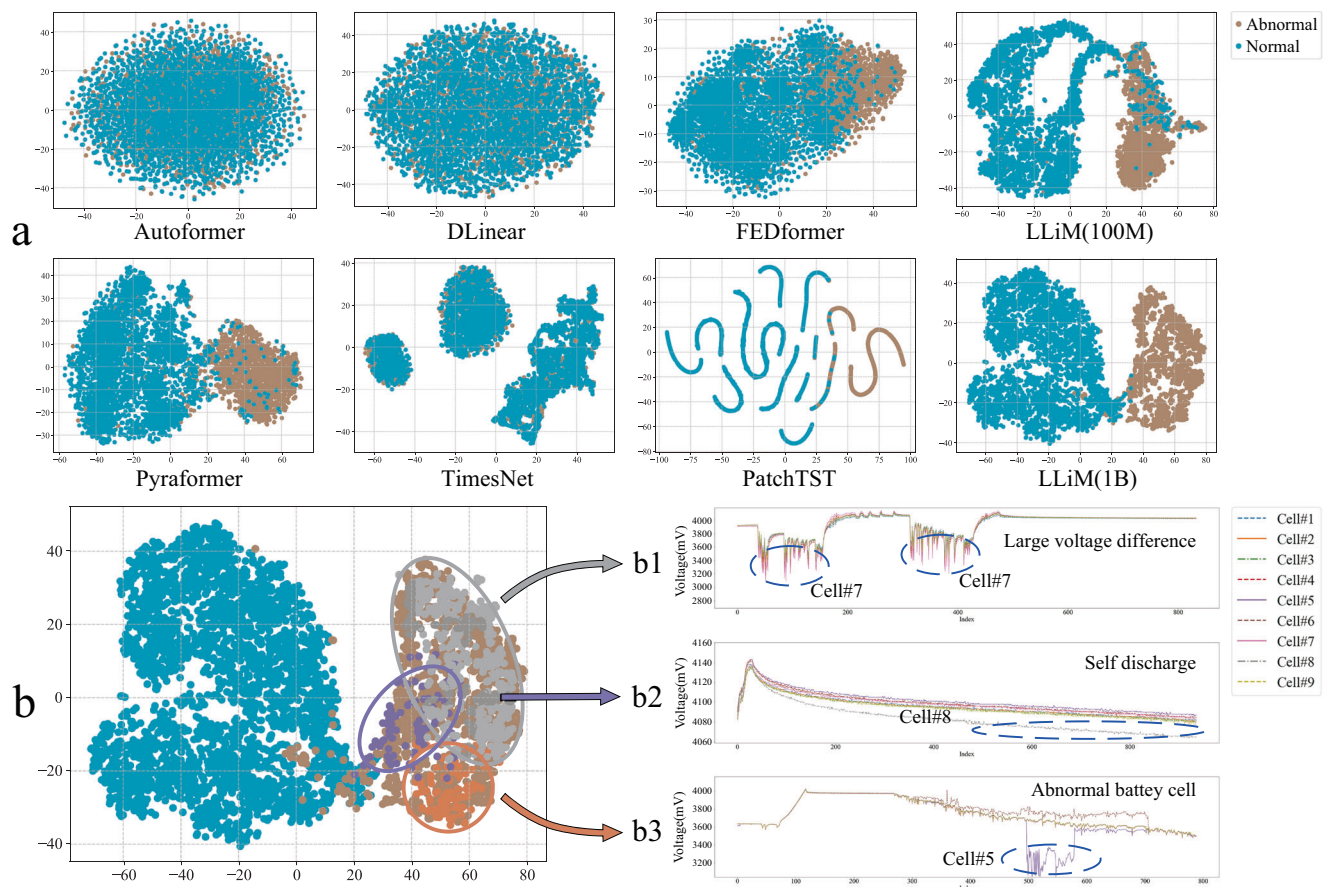
### LLiM for battery anomaly detection

In the process of using lithium-ion batteries, detecting abnormal lithium-ion batteries in advance and dealing with them in time is crucial to protect the lifespan of lithium-ion batteries and the property of users. To evaluate the performance of our model in battery anomaly

detection, we adopt a multi-layer perception as a classifier, which takes the output time series representations of the LLiM as input. The classifier is trained on a labeled battery anomaly detection dataset. In the experiment, we use two versions of LLiM with different model sizes: LLiM(100M) and LLiM(1B) with 100 million and 1 billion parameters respectively. We compare the effectiveness of our models with state-of-the-art supervised battery anomaly detection methods and present the accuracy, precision, recall and F1 score results for comparison in Table 1, including Transformer-based methods: Pyraformer<sup>35</sup>, FEDformer<sup>36</sup>, Autoformer<sup>37</sup>, PatchTST<sup>38</sup>; TCN-based methods: TimesNet<sup>39</sup>; and Linear-based methods: DLinear<sup>40</sup>. We analyzed the complete dataset of anomalous batteries (Total, seen in Table 2) due to limited data per cell type, with detailed results of different battery cells provided in Supplementary Table 11. From the results, LLiM shows statistically significant improvements in all key metrics compared to baseline approaches.

Specifically, LLiM(1B) model achieves 0.984, 0.988, 0.967 and 0.975 for accuracy, precision, recall and F1 score, respectively, and the average error of the LLiM model is much lower than that of the other supervised models, suggesting that our models not only have better performance but also good stability. The results can be attributed to LLiM's ability to integrate information from adjacent segments in the temporal data of lithium batteries. Since the information at each time point represents only the instantaneous state of the battery and is prone to noise, leveraging consecutive segment masking enables LLiM to capture continuous patterns and mitigate the effects of perturbations, enhancing the robustness of our model. Additionally, as the model size increases, the performance of the LLiM also improves. Compared to LLiM(100M), LLiM(1B) achieves an improvement of 0.5%, 0.7%, 1.9%, and 0.9% in accuracy, precision, recall and F1 score respectively. This demonstrates the effectiveness of the LLiM's large-scale pre-training and generalization capabilities.

To further demonstrate the performance of our model, we visualize the representation of the battery in latent space. We project the output embeddings of each method into a 2D space using the t-SNE algorithm<sup>41</sup>. The results of the visualization are shown in Fig. 3a. It is observed that the normal and abnormal battery representations encoded by LLiM are more separable and clustered than other methods. This indicates that LLiM can effectively capture the complex patterns of battery status data and learn the effective features for anomaly detection. To analyse the ability of LLiM to capture different anomalies, we visualize the representations of different anomalies captured by LLiM in Fig. 3b. From the visualization, we can see that the representations of different anomalies are also separated and grouped, demonstrating the effectiveness of our LLiM in capturing the difference among different anomalies. As shown in Fig. 3b1, when the battery is discharging, the voltage of cell#7 is much lower than the voltage of the other cells, which leads to a large voltage difference across the cell, resulting in poor overall battery performance. Figure 3b2 indicates that the voltage drop rate of cell#8 is much higher



**Fig. 3 | Illustration of battery data in latent space (brown color denotes abnormal samples and blue color normal ones). a** The representation visualization of the battery status data encoded by different methods. **b** Visualization of

different anomalies representations captured by LLiM(1B). Among different anomalies, b1 represents the large voltage difference, b2 indicates the self-discharge behavior, and b3 corresponds the abnormal battery cell.

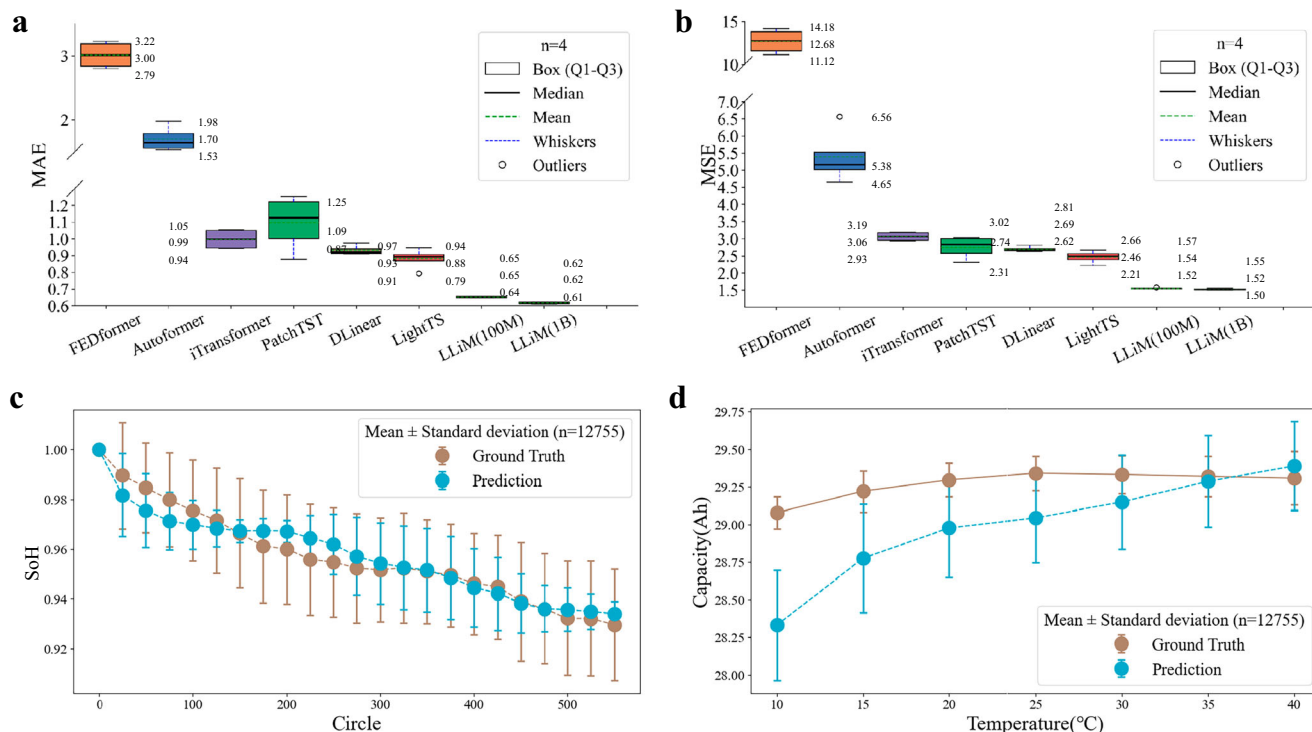
than that of other cells, which indicates that cell#8 has serious self-discharge abnormality, affecting the safety and usability of the battery. Figure 3b3 shows that cell#5 suddenly has a cell abnormality where the voltage suddenly drops by several hundred millivolts and fluctuates erratically, indicating that the battery must be removed from service and the condition of cell#5 checked to ensure battery safety. Different abnormalities need different ways to be dealt with, by identifying different abnormalities in advance, targeted action can be taken in advance to better protect the safety of the battery. This allows our model to play a more significant role in practical applications.

### LLiM for battery state of health estimation

SoH of a battery is a measure of its current capacity compared to its nominal capacity. When a new battery is manufactured, the capacity is same as the nominal capacity as per the battery specification, so the battery is in optimal health,  $\text{SoH} = 100\%$ . Because the nominal capacity of the lithium-ion battery is rated when manufacturing, estimating SoH of the battery is converted to estimating the current actual capacity, and the nominal capacity of experimental batteries is 30Ah. As the battery is used, its capacity decays, making it less efficient and difficult to estimate the actual capacity, which affects the reliability of the battery. In addition, the temperature of the environment seriously affects the actual capacity and lithium-ion batteries suffer performance degradation at low temperatures. We use a regression model that takes the output time series representations of the LLiM as input and predicts the capacity of the battery. The regression model is fine-tuned on a labeled dataset by minimizing the MSE between the predicted capacity and the ground truth capacity, which is obtained in the experiments. The details of the dataset

construction can be found in section “Datasets”. We performed the analysis using the FE32 dataset as it contains the largest number of battery units and exhibits a broad distribution of cycle counts, with detailed results for each individual cell provided in Supplementary Table 15.

In experiments, we also compare the two versions of our model, LLiM(100M) and LLiM(1B), with other existing methods including Transformer-based methods: iTransformer<sup>40</sup>, FEDformer<sup>36</sup>, Autoformer<sup>37</sup>, PatchTST<sup>38</sup>; TCN-based methods: LightTS<sup>42</sup>; and Linear-based methods: DLinear<sup>39</sup>. The experimental results are shown in Fig. 4a, b. We compare the MSE and MAE of the actual capacity estimation results. As shown, LLiM(1B) has the smallest MAE (0.62) and MSE (1.52) and has the smallest mean error. In addition, LLiM(100M) model ranks second after the LLiM(1B) model, demonstrating that LLiM excels at learning key information in time-series data. This also indicates that LLiM’s learning ability strengthens as the number of model parameters increases. It is worth noting that the actual capacity MAE=0.62 means the averaged SoH error is 2.06%, indicating that our model also has consistent performance in battery SoH estimation. In other words, when the user checks the state of charge (SoC) of the battery, which is defined the remain capacity over total actual capacity, then the error of SoC between the displayed value and the actual value is at most about 2% if the remain capacity is calculated accurately. This minimal discrepancy ensures that the user does not notice any difference in SoC, and therefore have no impact on the user’s experience or usage. Compared to these existing methods, LLiM effectively learns to capture charging and discharging patterns by reconstructing major battery features such as cell voltage and battery temperature



**Fig. 4 | Experiment results on battery SoH estimation.**  $n = 4$  in **a**, **b** represents 4 repeated experiments conducted for each algorithm. In **c**,  $n = 12775$  refers to the number of samples in the test set. Each Box annotation displays three values in top-to-bottom order: maximum, average, and minimum. **a** The comparison of MAE of

different methods on battery actual capacity estimation. **b** The comparison of MSE of different methods on battery actual capacity estimation. **c** The changes of SoH under different cycles. **d** The changes of actual capacity under different temperatures.

during pretraining, which manages to model changes in battery voltage and distance, leading to significantly better performance in SoH estimation.

In addition, we analyze the estimation performance of LLiM across varying battery cycle counts (defined as one full charge-discharge cycle), as presented in Fig. 4c. The prediction errors remain relatively consistent across different cycle numbers, suggesting that LLiM captures general degradation patterns of lithium-ion batteries regardless of cycle count. Notably, predictions align closely with ground truth values after 300 cycles, likely due to the dataset's majority of samples falling within the 300–500 cycle range. As expected, battery SoH decreases with increasing cycles, consistent with degradation behavior. However, variation in SoH at identical cycle counts indicates that additional factors (e.g., temperature during charging/discharging) influence degradation rates<sup>43</sup>.

Furthermore, we conducted experiments to evaluate the capacity prediction performance of LLiM across different temperatures and its impact on actual capacity under identical cycling conditions (Fig. 4d). Results indicated that capacity increased marginally between 10 and 25 °C but remained stable at 25–40 °C. At lower temperatures, LLiM exhibited a higher MAE, suggesting reduced prediction accuracy compared to higher temperature conditions, though the overall error remained below 2.6%. Additionally, we identified a statistically significant correlation between SoH degradation and amplified anomaly probabilities during cycling, with detailed results provided in the Supplementary Note C.4.5.

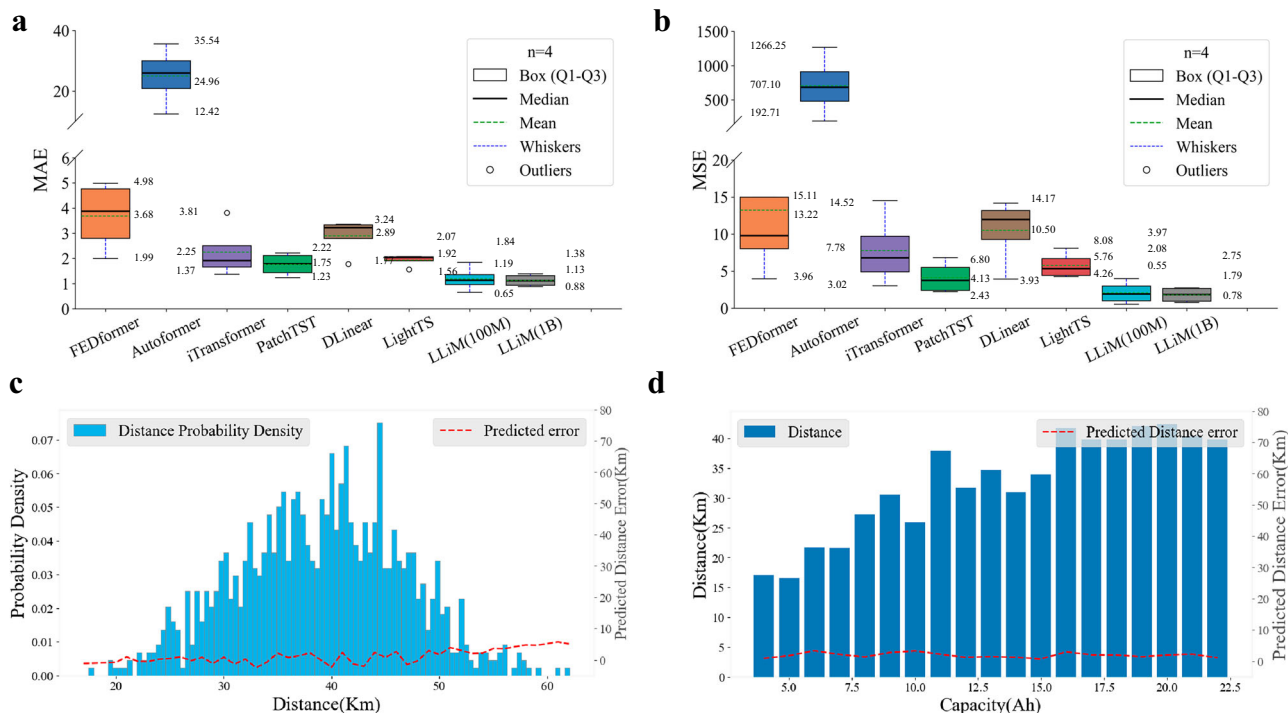
### LLiM for remain range prediction

Providing reliable battery swapping services necessitates accurate prediction of the RR of the battery. The RR signifies the distance that the battery is capable of traversing under prevailing conditions. By fine-tuning the pre-trained LLiM, an effective regression model is developed to predict RR under current battery conditions.

We evaluate the performance of the proposed framework in predicting the RR using actual RR values as labels. Similar to SoH estimation task, we conducted our analysis using the FE32 dataset, as seen in Table 2, with detailed results for each individual cell provided in Supplementary Table 19. Figure 5a, b illustrate MSE and MAE of the predicted RR for each model. The results indicate that the MAE and MSE of LLiM are significantly lower compared to other baseline models same as the SoH task. Specifically, the MAE and MSE for LLiM(100M) are approximately 1.19 and 2.08, while those for LLiM(1B) are approximately 1.13 and 1.79, respectively. This enables the user to plan their journey and determine when to swap the battery, which greatly improves the reliability of the battery-swapping services. These findings demonstrate that LLiM manages to accurately predict RR based on utilized battery records, thereby enhancing real-time estimation of the reachable range. Similar to SoH, the effectiveness of LLiM can be attributed to its pretraining masking strategy to reconstruct the major battery features, which greatly enhances the relationship between the used capacity and the distance traveled. This capability provides crucial information to support users' journey decisions and significantly improves the reliability of the battery swapping service.

As shown in Fig. 5c, the prediction error of the model is relatively small when the number of samples is large, although the error is relatively large when the distance is over 50 km due to the small number of samples, but the prediction accuracy will continue to improve as the riders continue to use and the data is enriched.

We also analyzed the relationship between remain battery capacity and RR mileage, as illustrated in Fig. 5d. Our observations indicate a generally positive association between these two variables, though exceptions were noted where higher remain capacity corresponded to reduced mileage. To explore these discrepancies, we examined specific scenarios involving short trips with high remain capacity and long trips with low remain capacity. The data suggest that under conditions



**Fig. 5 | Experiment results on battery remain range prediction.**  $n = 4$  in **a**, **b** represents 4 repeated experiments conducted for each algorithm. Each Box annotation displays three values in top-to-bottom order: maximum, average, and minimum. **a** The comparison of MAE of different methods on remain range

prediction. **b** The comparison of MSE of different methods on remain range prediction. **c** The prediction errors w.r.t different distances. **d** The prediction errors w.r.t different distances.

of high remain capacity, riders tend to utilize higher discharge currents, which may lead to reduced mileage due to increased energy consumption. Conversely, when remain capacity is lower, riders appear to adopt more conservative power usage patterns, resulting in comparatively extended mileage under these conditions.

## Discussion

In this work, we introduce LLiM, an unsupervised pre-trained large model specifically designed for the domain of e-bike lithium-ion batteries. LLiM is based on a Transformer encoder network pre-trained on more than 10 million real time-series data points collected from lithium-ion battery BMS. Notably, LLiM is scaled up to a parameter count of 1 billion. Our downstream experiments demonstrated that LLiM effectively addresses core technical challenges in the lithium-ion battery domain, such as anomaly detection, SoH estimation and range prediction. The results highlight the potential of large models for lithium-ion battery time-series data.

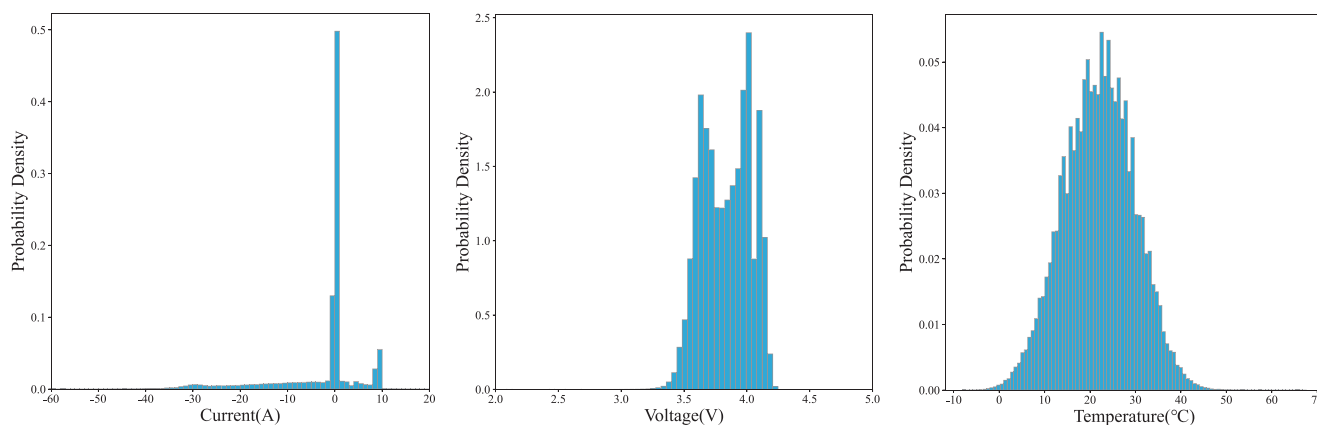
The LLiM model effectively provides *safe* and *reliable* battery-swapping services, which are crucial for the success of the SEB business model. By significantly improving the accuracy of lithium-ion battery anomaly detection, LLiM helps identify batteries with potential safety risks early, allowing for proactive measures to ensure battery safety and protect users' personal and property safety. Meanwhile, LLiM can accurately predict the actual capacity of lithium-ion batteries and the remaining riding distance for users, with a SoH estimation error of less than 3% and a remaining riding distance error of less than 1.2 km. This precision enables users to plan their trips in advance, avoiding issues such as power depletion during rides, thereby enhancing the riding experience. Accurate battery levels and remaining riding distances make SEB increasingly popular among e-bike vehicle users, reducing charging wait times and enabling virtually unlimited range through the battery-swapping station network.

We recognize that the development of large artificial intelligence models leads to increased energy consumption, for example<sup>44</sup>,

highlights that the widespread adoption of AI, particularly generative AI, has triggered a new crisis due to its substantial energy demands. Training and operating these models require significant energy resources. However, we address these concerns in several ways. First, as demonstrated in the Supplementary Note C.1.5 and C.6, the development, and operation of LLiM involve relatively modest energy consumption. Second, we have conducted additional experiments to demonstrate that LLiM(1B) exhibits significantly better robustness and performance in handling challenging tasks, such as battery bulging anomalies detection. Detailed experimental results are also provided in the Supplementary Note C.3.2. Third, advancements in hardware and algorithms offer potential solutions for mitigating the energy demands of large AI models. For example, GPUs demonstrate consistent annual progress in performance and energy efficiency metrics, while algorithmic approaches such as Multi-head Latent Attention<sup>45</sup> can reduce computational requirements. These technical advancements, along with improved transparency and industry collaboration, support the sustainable development of artificial intelligence technologies.

Furthermore, AI technologies like LLiM can play a significant role in addressing environmental challenges. As highlighted in ref. 46, AI can reduce ecological footprints and carbon emissions while promoting energy transitions, with the most substantial impact observed in energy transitions. Similarly, LLiM ensures the success of the SEB business model, which facilitates the transition of the timely delivery industry from traditional energy sources to new energy solutions. For instance, delivery personnel often require a 120-kilometer range, making fossil fuels a preferred option to avoid the inconvenience of frequent charging. By providing safe and reliable battery-swapping services, SEB enables zero-carbon emissions for two-wheeled electric vehicles, further optimizing environmental impact and contributing to sustainable urban transportation solutions.

Building upon the success of LLiM's pretraining with an extensive dataset and its subsequent finetuning for various battery management



**Fig. 6 | Distribution of values of BMS data including current, battery cell voltage and temperature.** The x-axis represents different values, and the y-axis denotes the probability density (integrated to unity for each plot). The current

values are predominantly concentrated between  $-10$  A and  $10$  A. The battery cell voltage ranges from  $3.0$  to  $4.2$  V, while the temperature values are centered around  $20$  °C with a distribution closely approximating a normal distribution.

tasks, future work will focus on expanding the model's capabilities and applicability. Our immediate goals include enhancing the model's adaptability to handle an even broader spectrum of lithium-ion batteries used in electric vehicles and other emerging technologies. We aim to refine the model further to accommodate the technical challenges posed by varying data lengths and operational cycles, ensuring robust performance across a wider range of applications. Additionally, we will explore the integration of LLiM with more sophisticated AIoT systems to improve predictive maintenance, optimize energy consumption, and enhance the overall user experience. By continuously updating the pretraining process with new data and feedback mechanisms, we will strive to maintain LLiM at the forefront of battery management technology, thereby contributing to the advancement of sustainable and smart urban transportation solutions. Furthermore, we will investigate the potential of transferring the knowledge encapsulated within LLiM to other domains, such as predictive modeling in renewable energy systems, thereby broadening the impact of our foundational model.

## Methods

In this section, we will introduce the details of the proposed foundation model LLiM, which is based on the widely used transformer-encoder architecture<sup>33</sup>. The overall framework of LLiM is shown in Fig. 2b, c. LLiM is an unsupervised large model pre-trained on a large-scale dataset of battery time series data. After pre-training, our method can be fine-tuned on various downstream tasks, such as anomaly detection, SoH estimation and RR prediction. Finally, we will introduce the details of pre-training, anomaly detection, SoH estimation and RR prediction datasets.

### Large-scale model pre-training

The LLiM is a large model pre-trained on large-scale battery time series data in an unsupervised manner. The pre-training process aims to learn the intrinsic patterns of the battery data. The pre-training objective is to minimize the reconstruction error between the input data and the reconstructed data. We will introduce the pre-training process in detail in the following sections.

**Data preprocessing.** The sensor inside the batteries would collect various types of data, such as voltage, current and temperature during the charging and discharging process. This information would be sent to the cloud server at a variable frequency to improve the efficiency of data transmission. In practice, during lithium battery usage, data is collected every 30 s in the active state (e.g., charging or discharging),

every 120 s in the stationary state (idle but unused), and every 900 s in the storage state (prolonged inactivity). Thus, the raw data is an irregularly discrete time series, denoted as

$$\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_t, \dots, \mathbf{X}_T\}, \quad (1)$$

where  $\mathbf{X} \in \mathbb{R}^{T \times F}$ ,  $T$  is the length of the time series and  $F$  is the number of features.  $\mathbf{X}_t$  is the data point at timestamp  $t$ , which contains the status of the battery. The features of the data point  $\mathbf{X}_t$  can be roughly grouped into numerical features (e.g., voltage and current) and categorical features (e.g., motion status). There are total 28 numerical features and 43 categorical features in the raw data.

The raw data is usually noisy and contains irrelevant information. Therefore, it is essential to preprocess the raw data before feeding it into the model.

For categorical features, we encode each category into a  $d$ -dimension embedding  $\mathbf{Z}_t^{\text{cat}} \in \mathbb{R}^d$ . These embeddings are learned during the pre-training process. For numerical features, most of them conform to a similar normal distribution. For instance, a battery pack consists of many cells, and most lithium-ion cells have a nominal voltage range of  $3.0$ – $4.2$  volts. Typically, riders opt to swap the battery when the SOC dips below 30% of the remaining power. As depicted in Fig. 6, the collected data predominantly falls within the  $3.6$ – $4.0$  volts range. The values of the temperatures collected follow a near-standard normal distribution. About 50% of the current values are 0, and the rest of the current values (current  $< 0$ ) show an approximate uniform distribution under discharging, with most of the charging current values being around 9A. Consequently, we proceed to standardize the numerical features, defined as

$$\mathbf{X}_t^{\text{num}'} = \frac{\mathbf{X}_t^{\text{num}} - \boldsymbol{\mu}}{\boldsymbol{\sigma}}. \quad (2)$$

The standardization would make the numerical features have a mean of 0 and a standard deviation of 1, which is more suitable for the MSE loss used for model training. After the standardization, we adopt a fully-connected layer to project the numerical features into a  $d$ -dimensional space, which is formulated as

$$\mathbf{Z}_t^{\text{num}} = \mathbf{W}_{\text{num}} \mathbf{X}_t^{\text{num}'} + \mathbf{B}_{\text{num}}. \quad (3)$$

where  $\mathbf{W}_{\text{num}}$  is the weights of fully-connected layer. The numerical features  $\mathbf{Z}_t^{\text{num}}$  and categorical features  $\mathbf{Z}_t^{\text{cat}}$  are then concatenated to

form the input data for the model as

$$\mathbf{Z}_t = [\mathbf{Z}_t^{\text{num}}, \mathbf{Z}_t^{\text{cat}_1}, \mathbf{Z}_t^{\text{cat}_2}, \dots, \mathbf{Z}_t^{\text{cat}_n}], \mathbf{Z}_t \in \mathbb{R}^d, \quad (4)$$

where  $n$  is the total number of categorical features and  $d$  is the dimension of embedding, it means the total embedding size of numerical features and categorical features is  $d$ .

**Mask-based model pre-training.** The pre-training objective is to minimize the reconstruction error between the input data and the reconstructed data, following the widely adopted mask-based training strategy in ref. 28. Specifically, given a sequence of battery data  $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T\}$ , we will randomly mask a portion of the data and feed the masked data into the model. This is formulated as

$$\tilde{\mathbf{X}} = \{\mathbf{X}_1, [\text{MASK}], \dots, \mathbf{X}_T\}, \quad (5)$$

The model is then trained to predict the masked features based on the unmasked data.

However, unlike the training of natural language processing models, the battery data is a time series. The current of the battery is more irregular and contains more noise due to the usage environment. For instance, the user could accelerate or decelerate the vehicle at any time, leading to a sudden change in the current, which makes it difficult to predict the current at the next time step based on previous data. Moreover, the battery data is discretely sampled leading the situation between two data points unknown. To address this problem, we propose a mask-based training strategy to better capture the intrinsic patterns of the battery time series for the pre-training of the LLiM.

**Current-preserved masking.** We keep the current value unmasked during the training process. As the current is one of the most important features of the battery status, both the discharge power and voltage are directly related to the current. In electrical circuits, Ohm's law<sup>47</sup> states that the voltage  $V$  across an electrical conductor is directly proportional to the current  $I$  passing through the conductor, the mathematical expression is  $V = IR$ . For DC, power equals current multiplied by voltage,  $P = VI$ . The discharge current during the discharging process reacts to the amount of power of the motor of the e-bike and also represents the riding behavior of the user. Therefore, we believe that preserving the current value would help the model to learn the intrinsic patterns of the battery data.

**Capacity-preserved masking.** During the masking, we also calculate the capacity difference between the two consecutive time steps. The capacity-voltage curve of lithium-ion battery reflects the performance of lithium-ion battery very well, in which the horizontal axis reflects the battery's charging and discharging capacity, charge state and other information, and the vertical axis contains the battery's voltage plateau, inflection point, polarization and other information. As the battery is used for a longer time, the battery capacity degrades, and the charging and discharging capacities corresponding to the same voltage range become smaller. Since there is a certain time interval between the two consecutive points in the reported sequence, and the intermediate charging and discharging status is unknown, it is infeasible to predict and restore the discharge capacity and voltage of a certain point in the sequence by using data from adjacent points. In this way, we preserve the capacity difference between the two consecutive time steps to help the model learn the intrinsic patterns of the battery data.

**Time-preserved masking.** The battery data is a time series with irregular intervals between consecutive data points. To retain temporal information, we calculate the time gap between each pair of consecutive data points and maintain it during masking. This approach

enables the model to better capture the influence of time on battery status. For example, if a lithium-ion battery stops discharging after continuous discharge, the battery's voltage will rebound and rise over time. Similarly, after stopping charging after continuous charging it will rebound down with time.

In the training process, we would mask out the rest of the features except for the aforementioned current, capacity difference, and time gap. For the numerical features, we would replace them with 0 following the previous masked-based methods<sup>48,49</sup>. For the categorical features, we would replace them with a special category called [MASK], whose embedding is also learned during the training process. Additionally, to increase the difficulty of the training, we would mask out consecutive data points<sup>50</sup>, which is set to 3 in experiments. In this way, the model is forced to learn the long-term dependencies between the data points.

After the masking, we first embed the masked data  $\tilde{\mathbf{X}}$  into embedding  $\tilde{\mathbf{Z}}$ . Then, we feed it into the model and learn to predict the masked features, which is formulated as

$$\tilde{\mathbf{Z}} = \{\mathbf{Z}_1, [\text{MASK}], \dots, \mathbf{Z}_T\}, \quad (6)$$

$$\hat{\mathbf{Z}} = \text{TransformerEncoder}(\tilde{\mathbf{Z}}), \quad (7)$$

where  $\hat{\mathbf{Z}}$  is the output of the Transformer encoder model. In the model, we replace the layer normalization with the RMS Normalization, which contributed to a measurable acceleration in training<sup>51</sup>.

**Loss function.** The pre-training objective is to minimize the reconstruction error between the input data and the reconstructed data. To better recover the numerical and categorical features, we adopt a mixed reconstruction loss (Fig. 2b). Specifically, after the backbone network, we get the output of the Transformer encoder model,  $\hat{\mathbf{Z}}$ . For numerical and categorical features, we design different head network to calculate the corresponding loss. We feed the output  $\hat{\mathbf{Z}}$  into a fully connected layer (FC) to reconstruct the original numerical features,  $\hat{\mathbf{Z}}_t^{\text{num}}$  means the predicting result of the numerical features and  $t$  is the masked time step. However, for every categorical feature, we feed the output  $\hat{\mathbf{Z}}$  into a FC to get the logits of the categorical features,  $\hat{\mathbf{Z}}_t^{\text{cat}_i}$ , and then adopt a cross-entropy loss to predict the original categories. The overall loss function is formulated as

$$\hat{\mathbf{Z}}_t^{\text{cat}_i} = \text{FC}_{\text{cat}_i}(\hat{\mathbf{Z}}), \quad (8)$$

$$\hat{\mathbf{Z}}_i^{\text{num}} = \text{FC}_{\text{num}}(\hat{\mathbf{Z}}), \quad (9)$$

$$\mathcal{L} = \sum_{t \in \text{masked}} \left( \sum_i \mathbf{X}_t^{\text{cat}_i} \log P(\hat{\mathbf{X}}_t^{\text{cat}_i} | \hat{\mathbf{Z}}_t^{\text{cat}_i}) + \lambda (\mathbf{X}_t^{\text{num}} - \hat{\mathbf{Z}}_t^{\text{num}})^2 \right), \quad (10)$$

where FC denotes the fully connected layer and  $P(\cdot)$  also denotes a softmax function, and  $\lambda$  is a hyperparameter to balance the numerical and categorical loss.

### Fine-tuning on downstream tasks

After the pre-training, LLiM is fine-tuned on the downstream tasks of anomaly detection, SoH estimation and RR prediction. The fine-tuning process aims to learn the specific patterns of the battery data for the downstream tasks. During the finetuning, we simply add an additional layer on top of the pre-trained model to predict the anomaly score, SoH and RR. For the pre-trained model, we adopt the low-rank adaptation<sup>34</sup> to efficiently update the parameters of the attention layers and improve the performance on down-stream tasks. Specifically, for the pre-trained weight matrix  $\mathbf{W} \in \mathbb{R}^{d \times k}$ , we update it by

**Table 3 | Statistics of the pre-training dataset**

Cell Type	Training dataset		Testing dataset	
	Quantity	Ratio	Quantity	Ratio
FE32	5,728,218	56.6%	86,835	43.4%
FE30	1,901,082	18.8%	62,722	31.4%
FE29	634,526	6.3%	12,443	6.2%
TG29	1,111,386	11.0%	22,000	11.0%
DT29	240,084	2.4%	6000	3.0%
AMP29	512,832	5.1%	10,000	5.0%
Total	10,128,128	100%	200,000	100%

representing it with a low-rank decomposition, formulated as

$$\mathbf{W}' = \mathbf{W} + \Delta\mathbf{W}, \quad (11)$$

$$= \mathbf{W} + \mathbf{B}\mathbf{A}, \quad (12)$$

where  $\mathbf{B} \in \mathbb{R}^{d \times r}$ ,  $\mathbf{A} \in \mathbb{R}^{r \times k}$ , and the rank  $r \ll \min(d, k)$ . During fine-tuning,  $\mathbf{W}$  is frozen, and  $\mathbf{A}$ ,  $\mathbf{B}$  are optimized using downstream task loss functions. The inference process can be formulated as

$$\mathbf{H} = \mathbf{W}\mathbf{X} + \Delta\mathbf{W}\mathbf{X} = \mathbf{W}\mathbf{x} + \mathbf{B}\mathbf{A}\mathbf{X}, \quad (13)$$

where  $\mathbf{X}$  is the input feature and  $\mathbf{H}$  is the output hidden representation.

The anomaly detection is a binary classification task, where the model is trained to predict whether the battery is normal or abnormal. The loss function is formulated as follows

$$\hat{y} = \text{Softmax}(\text{FC}(\hat{\mathbf{Z}})), \quad (14)$$

$$\mathcal{L}_{\text{anomaly}} = - \sum_i y_i \log(\hat{y}_i) - (1 - y_i) \log(1 - \hat{y}_i), \quad (15)$$

where  $\hat{\mathbf{Z}}$  denotes the output of the pre-trained model. We adopt a FC followed by a softmax function to predict the anomaly score.

The SoH estimation and RR prediction are regression tasks, where we adopt a new FC to predict the target value separately. The loss function is formulated as follows

$$\hat{r} = \text{FC}(\hat{\mathbf{Z}}), \quad (16)$$

$$\mathcal{L}_{\text{SoH/RR}} = \sum_i (r_i - \hat{r}_i)^2. \quad (17)$$

where  $r_i$  denotes the predicted value and  $\hat{r}_i$  is the target value of the  $i$  battery in the two regression tasks.

## Datasets

In the work, we use data gathered from real-world lithium-ion batteries charging and discharging process for SEB, covering battery BMS data from January to December 2023. Geographically, the battery BMS data were distributed across more than 50 cities in China, encompassing major first- and second-tier cities spanning from southern to northern regions, including Shenzhen, Hangzhou, and Nanjing. Due to the high energy density and good low-temperature performance, the batteries currently deployed in SEB scenarios are all ternary lithium batteries, specifically lithium nickel cobalt manganese oxide and lithium nickel cobalt aluminum oxide. The dataset comprises six types of battery cells: FE32, FE30, FE29, DT29, TG29, and AMP29. Because of the differences in the material, some batteries exhibit notable divergence in their charge and discharge curves, and the open-circuit voltage to

state-of-charge (OCV-SOC) curves for each type of battery cell can be found in Supplementary Fig. 1. The nominal capacity of each type of battery cell is approximately 30 Ah, with variations in capacity for different cells.

The data is divided into four groups: pre-training, battery anomaly detection, SoH estimation, and RR prediction. For pre-training, we only utilized the BMS data reported by the lithium batteries themselves. For the other three downstream tasks, we also need to obtain labels for each sample, which will be discussed in detail next.

- **Pre-training Dataset** Due to the continuous BMS data reporting and storage of lithium-ion batteries on cloud servers, each lithium battery had accumulated a long time series of data. We employed a sliding window approach to segment the battery time series data into non-overlapping sequences of 600 time points each. This process yielded a final pre-training dataset of 10.1 million sequences, collectively containing 6 billion time data points. Each time point contains 71 features, including 43 numerical features and 28 categorical features. These features are primarily collected and reported by the BMS through various sensors, a detailed description of the features can be found in the Supplementary Note A.2. Additionally, we reserved an extra 200,000 sequences to validate the performance of the pretrained model. The statistics of the dataset are shown in Table 3.
- **Battery Anomaly Detection Dataset** We randomly selected 23,250 lithium-ion batteries that were used during 2023 year. These lithium-ion batteries have been used in the market and are recalled to detect anomalies using instruments, such as X-ray imaging, battery self-discharge tester, battery capacity analyzer, etc. Based on the abnormal batteries, We collect the battery status of these batteries 500 time steps before they are recalled to create a time-series record and mark them separately for abnormality. These abnormal batteries will serve as positive samples in the anomaly detection dataset, labeled as 1, while the batteries that pass all the inspections will be included as negative samples, labeled as 0. The number of features contained at each time step is the same as that of the pre-training data, containing 71 different features. The three types of batteries, FE32, FE30, and FE29, have been used for a longer period, and the majority of the abnormal batteries accumulated in the system are primarily composed of these three types. In the end, we collect 23,250 records for anomaly detection, including 7750 positive samples and 15,500 negative samples, as shown in Table 2.
- **State of Health Estimation Dataset** SoH of a battery is a measure of the battery's capacity compared to its original capacity. It is an important metric for battery management and maintenance. As the battery is used, the capacity of the battery will decrease, making it less efficient and hard to estimate the actual capacity. The capacity estimation process constitutes a regression analysis task, requiring precise measurement of actual battery capacity. To obtain reliable capacity labels for our dataset, we conducted full charge-discharge cycles using precision capacity analyzers within temperature-controlled environmental chambers on batteries returned. In this dataset, we provided 63,772 FE32 batteries and measured the actual capacity of each battery. The FE32 batteries are the primary batteries on the market, with a wide distribution of cycle counts. Similarly, we collected sequential data on each battery during the charge and discharge process before it was returned, with a fixed sequence length of 500. Additional cell types are also included in the study, as shown in Table 2, with complete results for individual cells available in Supplementary Table 15.
- **Remain Range Prediction Dataset** Predicting the battery's RR is crucial to offering reliable battery swapping services. The RR indicates how far a vehicle can travel with its current battery capacity. Similar to the capacity estimation dataset, the RR dataset

comprises 9107 FE32 records, each record contains a rider's complete discharge data. To determine the actual RR at every timestamp, we track the e-bikes' geographical position reported by the BMS and integrate this data from the current timestamp to the end of the battery discharging process. We select the preceding 450 steps of battery status prior to the measurement for the model input and predict the RR at each timestamp. Statistics for the remaining range prediction dataset are presented in Table 2, with results per individual cell in Supplementary Table 19.

### Data availability

The pre-training dataset is restricted due to commercial confidentiality and user privacy laws. Access requires institutional agreements and non-disclosure agreements. Requests should be directed to the corresponding author Z.L. The datasets for the three downstream tasks are available with <https://doi.org/10.6084/m9.figshare.28260095> at the link [https://figshare.com/articles/dataset/Datasets\\_of\\_three\\_downstream\\_tasks\\_in\\_LLiM/28260095](https://figshare.com/articles/dataset/Datasets_of_three_downstream_tasks_in_LLiM/28260095). Source data are provided with this paper.

### Code availability

The pre-training code is not publicly available due to its commercial nature. The code for the three downstream tasks is available for download at GitHub (<https://github.com/ddhdzt/LLiM>) and has been archived in Zenodo with the <https://doi.org/10.5281/zenodo.16421724>.

### References

- Huang-Lachmann, J. T. Systematic review of smart cities and climate change adaptation. *Sustain. Account. Manag. Policy J.* **10**, 745–772 (2019).
- García Fernández, C. & Peek, D. Smart and sustainable? Positioning adaptation to climate change in the European smart city. *Smart Cities* **3**, 511–526 (2020).
- Belaïd, F. & Arora, A. *Smart Cities Social and Environmental Challenges and Opportunities for Local Authorities* (Springer Cham, 2023).
- Addas, A. The concept of smart cities: a sustainability aspect for future urban development. *Front. Environ. Sci.* **11**, 1241593 (2023).
- Iyer, N. V. & Badami, M. G. Two-wheeled motor vehicle technology in India: evolution, prospects and issues. *Energy Policy* **35**, 4319–4331 (2007).
- Bździuch, D. & Grzegoek, W. A two-wheeled, self-balancing electric vehicle used as an environmentally friendly individual means of transport. *IOP Conf. Ser. Mater. Sci. Eng.* **148**, 012003 (2016).
- Gu, T., Kim, I. & Currie, G. The two-wheeled renaissance in China—an empirical review of bicycle, E-bike, and motorbike development. *Int. J. Sustain. Transp.* **15**, 239–258 (2021).
- Ilin, V., Veličkovi, C. M., Garunović, N. & Simić, D. Last-mile delivery with electric vehicles, unmanned aerial vehicles, and e-scooters and e-bikes. *J. Road. Traffic Eng.* **69**, 37–42 (2023).
- Deloitte. *White Paper on the Electric Two-wheeled Vehicle Industry*. Technical report (Deloitte, Beijing, 2023). <https://www.deloitte.com/cn/zh/industries/consumer/perspectives/electric-two-wheelers-industry-whitepaper.html>.
- Wu, L., Sun, X., Lin, Y. & Lin, Y. Market multi node expansion and structural optimization in Luyuan Group Holdings-electric two wheel vehicle business expected to accelerate development. Tianfeng Securities, Shenzhen (June 6, 2024). <http://hangyan.co/reports/3384855411659113520>.
- Vallera, A., Nunes, P. & Brito, M. Why we need battery swapping technology. *Energy Policy* **157**, 112481 (2021).
- Li, Z. et al. Transformer-based graph neural networks for battery range prediction in AIoT battery-swap services. In *Proc. IEEE International Conference on Web Services* 1168–1176 (IEEE, 2024).
- Feng, Y. & Lu, X. Deployment and operation of battery swapping stations for electric two-wheelers based on machine learning. *J. Adv. Transp.* **2022**, 1 (2022).
- Huang, F. Understanding user acceptance of battery swapping service of sustainable transport: an empirical study of a battery swap station for electric scooters, Taiwan. *Int. J. Sustain. Transp.* **14**, 294–307 (2020).
- Ding, D. et al. eBaaS: AIoT-Enabled eBike Battery-Swap as a service for last-mile delivery. In *Proc. ACM on Web Conference* 5045–5053 (ACM, 2025).
- Yuan, H., Cui, N., Li, C., Cui, Z. & Chang, L. Early stage internal short circuit fault diagnosis for lithium-ion batteries based on local-outlier detection. *J. Energy Storage* **57**, 106196 (2023).
- Zhang, X., Liu, P., Lin, N., Zhang, Z. & Wang, Z. A novel battery abnormality detection method using interpretable autoencoder. *Appl. Energy* **330**, 120312 (2023).
- Zheng, L. et al. Anomaly detection in battery charging systems: a deep sequence model approach. In *Proc. IEEE International Conference on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking* 587–594 (IEEE, 2023).
- Wen, M., Ibrahim, M. S., Meda, A. H., Zhang, G. & Fan, J. In-situ early anomaly detection and remaining useful lifetime prediction for high-power white leds with distance and entropy-based long short-term memory recurrent neural networks. *Expert Syst. Appl.* **238**, 121832 (2024).
- Song, Y. et al. Detection of voltage fault in lithium-ion battery based on equivalent circuit model-informed neural network. *IEEE Trans. Instrum. Meas.* **73**, 1–10 (2024).
- Harper, G. et al. Recycling lithium-ion batteries from electric vehicles. *Nature* **575**, 75–86 (2019).
- Waag, W., Käbitz, S. & Sauer, D. U. Experimental investigation of the lithium-ion battery impedance characteristic at various conditions and aging states and its influence on the application. *Appl. Energy* **102**, 885–897 (2013).
- Zhu, J. et al. Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation. *Nat. Commun.* **13**, 2261 (2022).
- Li, Y., Li, K., Liu, X., Wang, Y. & Zhang, L. Lithium-ion battery capacity estimation—a pruned convolutional neural network approach assisted with transfer learning. *Appl. Energy* **285**, 116410 (2021).
- Li, Z. et al. Real-time e-bike route planning with battery range prediction. In *Proc. 17th ACM International Conference on Web Search and Data Mining* 1070–1073 (ACM, 2024).
- Ren, L. et al. Remaining useful life prediction for lithium-ion battery: a deep learning approach. *IEEE Access.* **6**, 50587–50598 (2018).
- Chen, Z., Chen, L., Shen, W. & Xu, K. Remaining useful life prediction of lithium-ion battery via a sequence decomposition and deep learning integrated approach. *IEEE Trans. Veh. Technol.* **71**, 1466–1479 (2021).
- Devlin, J., Chang, M., Lee, K. & Toutanova, K. Bert: pre-training of deep bidirectional transformers for language understanding. In *Proc. North American Chapter of the Association for Computational Linguistics* (ACL, 2019).
- Brown, T. et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **33**, 1877–1901 (2020).
- Paass, G. & Giesselbach, S. *Foundation Models for Natural Language Processing: Pre-trained Language Models Integrating Media*. (Springer Cham, 2023).
- Dosovitskiy, A. et al. An image is worth 16 × 16 words: transformers for image recognition at scale. In *Proc. Ninth International Conference on Learning Representations* (OpenReview, 2021).
- Zhang, Q., Xu, Y., Zhang, J. & Tao, D. Vitaev2: vision transformer advanced by exploring inductive bias for image recognition and beyond. *Int. J. Comput. Vision.* **131**, 1141–1162 (2023).

33. Vaswani, A. et al. Attention is all you need. In *Proc. Advance in Neural Information Processing Systems*. 30 (NeurIPS Foundation, 2017).
34. Hu, E. J. et al. LoRA: low-rank adaptation of large language models. In *Proc. International Conference on Learning Representations* (OpenReview, 2022).
35. Liu, S. et al. Pyraformer: low-complexity pyramidal attention for long-range time series modeling and forecasting. In *Proc. International Conference on Learning Representations* (OpenReview, 2021).
36. Zhou, T. et al. FEDformer: frequency enhanced decomposed transformer for long-term series forecasting. In *Proc. 39th International Conference on Machine Learning* (PMLR, 2022).
37. Wu, H., Xu, J., Wang, J. & Long, M. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting. In *Proc. Advances in Neural Information Processing Systems* (NeurIPS Foundation, 2021).
38. Nie, Y., Nguyen, N., Sinthong, P. & Kalagnanam, J. A time series is worth 64 words: long-term forecasting with transformers. In *Proc. International Conference on Learning Representations* (OpenReview, 2023).
39. Zeng, A., Chen, M., Zhang, L. & Xu, Q. Are transformers effective for time series forecasting? In *Proc. AAAI Conference on Artificial Intelligence* (AAAI Press, 2023).
40. Liu, Y. et al. iTransformer: inverted transformers are effective for time series forecasting. In *Proc. Twelfth International Conference on Learning Representations* (OpenReview, 2024).
41. Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
42. Campos, D. et al. LightTS: lightweight time series classification with adaptive ensemble distillation. In *Proc. ACM on Management of Data* (ACM, 2023).
43. Tian, J., Xiong, R. & Shen, W. State-of-health estimation based on differential temperature for lithium ion batteries. *IEEE Trans. Power Electron.* **35**, 10363–10373 (2020).
44. Bourzac, K. Fixing AI's energy crisis. *Nature*. <https://doi.org/10.1038/d41586-024-03408-z> (2024).
45. DeepSeek-AI. et al. DeepSeek-V3 Technical Report. *arXiv* <https://doi.org/10.48550/arXiv.2412.19437> (2024).
46. Wang, Q., Li, Y. & Li, R. Ecological footprints, carbon emissions, and energy transitions: the impact of artificial intelligence (AI). *Humanit. Soc. Sci. Commun.* **11**, 1–18 (2024).
47. Schagrin, M. Resistance to Ohm's law. *Am. J. Phys.* **31**, 536–547 (1963).
48. Dong, J. et al. SimMTM: A simple pre-training framework for masked time-series modeling. In *Proc. Advances in Neural Information Processing Systems* (NeurIPS Foundation, 2024).
49. Zha, M., Wong, S., Liu, M., Zhang, T. & Chen, K. Time series generation with masked autoencoder. *arXiv* <https://doi.org/10.48550/arXiv.2201.07006> (2022).
50. Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A. & Eickhoff, C. A transformer-based framework for multivariate time series representation learning. In *Proc. 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* 2114–2124 (ACM, 2021).
51. Zhang, B. & Sennrich, R. Root mean square layer normalization. In *Proc. Advances in Neural Information Processing Systems* (NeurIPS Foundation, 2019).

## Acknowledgements

This work was supported in part by National Key R&D Program of China (2023YFB4502400).

## Author contributions

D.D. contributed to design and implementation of the algorithm and edited the manuscript. Z.L. was responsible for designing the framework of the manuscript and supervised the work. L.L. and B.Z. contributed to the writing of the manuscript. M.J., Y.Z., and J.H. were involved in the preparation of the datasets and conducted experiments. P.C. and H.H. assisted in revising the work.

## Competing interests

The authors D.D., Z.L., Y.Z., and J.H. are employees of Hangzhou Yugu Technology Co., Ltd. and declare competing financial interests. The company derives commercial value from battery-analytics services; the performance of these services depends on the proprietary data discussed in this study. Other authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-63678-7>.

**Correspondence** and requests for materials should be addressed to Zhao Li, Bin Zhu or Peng Cai.

**Peer review information** *Nature Communications* thanks Sarang D. Supekar who co-reviewed with Miriam Stevens, Siyuan Wu, Yiquan Wu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025