Fast screening of lithium-ion batteries for second use with pack-level testing and machine learning

Sijia Yang [a,b], Caiping Zhang [a,**], Jiuchun Jiang [c,d], Weige Zhang [a], Haoze Chen [a,b], Yan Jiang [c], Dirk Uwe Sauer [b,e,f], Weihan Li [b,e*]

[a] *National Active Distribution Network Technology Research Center (NANTEC), Beijing Jiaotong University, Beijing 100044, China*

[b] *Chair for Electrochemical Energy Conversion and Storage Systems, Institute for Power Electronics and Electrical Drives (ISEA), RWTH Aachen University, Campus-Boulevard 89, Aachen, 52074, Germany*

[c] *Shenzhen Precise Testing Technology Co. Ltd, Shenzhen 518100, China*

[d] *School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, China*

[e] *Jülich Aachen Research Alliance, JARA-Energy, Templergraben 55, 52056, Aachen, Germany*

[f] *Helmholtz Institute Münster (HI MS), IEK-12, Forschungszentrum Jülich, Germany*

[*, **] **Corresponding author.**

E-mail addresses: sjyang@bjtu.edu.cn (S. Yang), zhangcaiping@bjtu.edu.cn (C. Zhang), weihan.li@isea.rwth-aachen.de (W. Li).

Abstract

Fast and accurate screening of retired lithium-ion batteries is critical to an efficient and reliable second use with improved performance consistency, contributing to the sustainability of renewable energy sources. However, time-consuming testing, representative criteria extraction, and large module-to-module inconsistencies at the end of first life all pose great challenges for fast screening. This paper proposes a fast screening approach with pack-level testing and machine learning to evaluate and classify module-level aging, where disassembly of the battery pack and individual testing of modules are not required. Dynamic characteristic-based criteria are designed to extract the comprehensive performance of the retired modules, making the approach applicable for battery packs with module state-of-charge inconsistencies up to 30%. Adaptive affinity propagation clustering is utilized to classify the modules and further accelerate the screening progress. The proposed approach is implemented and validated by conducting pack-level and module-level experiments with a retired battery pack consisting of 95 modules connected in series. The screening time is reduced by at least 50% compared with approaches that require module-level testing. Reasonable static performance consistency and better dynamic performance consistency, as well as higher screening stability, are achieved, with average overall performance improvements of 18.94%, 4.83% and 34.41% compared with the three benchmarks, respectively. Its adaptability to a larger current rate shows promise for large-scale applications in second-use screening.

Keywords: lithium-ion batteries, fast screening, pack-level testing, second use, machine learning

1. Introduction

Lithium-ion batteries (LIBs), the main pillar of energy storage technology for electric vehicles (EVs), suffer from performance degradation during usage and storage in terms of capacity and power [1]. Typically, they reach their end-of-life when their remaining capacity reaches 80% of the nominal capacity [2] or their internal resistance reaches 200% of that of pristine batteries [3]. The retired EV batteries reached 14 GWh/year in 2020 globally, and this figure is expected to rise 10-fold by 2025 and surpass 900 GWh/year by 2030, which will exceed the current annual global battery production [4–6]. As the EV holdings continue to grow worldwide - potentially hitting 245 million by 2030 [7] - there is no doubt that a tidal wave of retired LIBs is on the horizon.

The second use of EV batteries plays an important role in the sustainability of new energy vehicles. It is a promising path to increase the usage time of the batteries, thereby decreasing the total lifetime costs and increasing resource utilization [8]. Instead of recycling these retired EV batteries directly at the material level, it is more economical and ecological to reuse them [9], as most of them are still adequate in their second life for other less-demanding applications in the second life, such as electric motorcycles, mobility e-scooters and stationary battery storage systems [10–12]. In fact, over 70% of the batteries may still reserve more than 80% of their nominal capacity [13,14]. However, the internal aging states of these batteries may be varied due to different aging mechanisms in their first life [1]. Such underlying inconsistencies can induce different aging speeds of the LIBs in the second life and further exacerbate the degradation of the battery packs and even cause safety issues [15,16]. Therefore, it is necessary and critical to screen the retired batteries with good consistency for the reliable operation of EV batteries in second use and for the sustainable development of the EV industry.

1.1. Literature review

Unfortunately, there are currently no uniform screening criteria or industrial standards for retired LIB packs [17]. Generally, the battery packs are disassembled into modules or cells to evaluate the condition of the batteries [8], which can be conducted at different scales with corresponding screening criteria [18]. Inner criteria, like morphological changes in electrodes [19] or electrochemical parameters [20], while straightforwardly capturing the aging, requires special equipment and domain expertise for the measurement, which is costly to apply in large-scale screening. On the other hand,

outer criteria are much easier to implement, including observational check [14,21,22], weight [23,24], self-discharge rate [25,26], thermal behavior [27], capacity [27–29], internal resistance [27,28,30], and terminal voltage at a certain state-of-charge (SOC) [22,27,31], etc.

These outer criteria can be further divided into two categories: static and dynamic characteristics-based [23,32]. Static characteristic-based criteria are state points that only reflect the performance in one SOC at a specific aging degree, whereas the dynamic characteristic-based criteria portray the performance during battery operation over a time period or a SOC interval. Nominal capacity is an example of static criteria. The voltage-capacity curve, in contrast, is a dynamic criterion. Inarguably, dynamic characteristics-based criteria provide more insights into the internal states of the batteries than static characteristics-based criteria. After all, despite having similar nominal capacities, the internal aging states of two aged cells can be varied due to different aging paths in their first life [1,33].

To improve performance consistency for the second use, outer criteria are often used in combination [18,34]. For instance, capacity, pulse discharge voltage, charge transfer resistance and lithium-ion diffusion coefficient are jointly utilized in Ref. [14]. More criteria generally mean better performance consistency and higher screening accuracy, but the problem is quite obvious – low efficiency, as most of these criteria are measured at the module/cell level, which implies that more labor is demanded to disassemble the battery pack into modules/cells for individual testing, inevitably slowing down the second use progress of LIBs [8]. As a result, many researchers have attempted to shorten the screening time while balancing the accuracy. Basically, there are four options.

The first way is to find some highly relevant features to characterize these module-/cell-level criteria based on a partial testing process instead of the complete process. Incremental capacity (IC) curves, whose features are assumed to be accessible in partial voltage windows, are commonly employed to represent capacity [35,36]. For example, the peak height of the IC curve [37] and the coordinates of the three IC peaks [38]. The latter further reduced the feature extraction time by increasing the charging current rate (C-rate) from C/10 to C/2. Considering that only part of the measurement process is involved, this is theoretically possible for the pack-level testing application. However, these methods normally extract the representative features based on partial voltage segments directly intercepted from the complete voltage curve starting from 0% SOC, while neglecting the reality that the initial charging SOC is usually non-zero due to the existence of SOC inconsistency. This may be different from the voltage curve of 0% SOC charging [39], which may affect their feature

extraction and challenge the applications of these methods. Moreover, large module-to-module inconsistencies at the end of the first life make the extraction of the representative features for each module more difficult [40], and in some cases, may not even be extractable.

The second alternative is to make greater use of dynamic characteristic-based criteria, which are likely to reduce the number of criteria to be measured while maintaining accuracy, as they can capture the comprehensive performance of the batteries mentioned earlier. Jiang et al. [28] improved screening performance consistency by adding a criterion depicting the loss of active material. Luo et al. [34] employed four criteria extracted from the short-time pulse voltage curve and electrochemical impedance spectroscopy (EIS) model, but it still took around 20 minutes to measure the criteria for one cell. Zhang et al. [41,42] used two dynamic characteristic-based criteria extracted from partial charging voltage curves of the cells. Whilst these methods have produced reasonably good results, most of them still require individual testing of the batteries and rarely discuss the voltage consistency of the batteries during operation after the screening.

Another idea to improve the screening efficiency is provided by Lai et al. [43], in which they charged the entire battery pack to avoid testing the modules individually, saving a great deal of testing time and resources. They pre-trained a support vector machine model to estimate the battery capacity based on the supplemental charging voltage and used an improved k-means algorithm for screening. The results are encouraging, but they only verified it on a simulation model, which may differ from the practical situation where large module-to-module inconsistencies exist.

The last choice is to adopt unsupervised learning rather than supervised learning, where the latter requires plenty of samples for model training, which is time-consuming when acquiring the screening criteria. The screening time for 176 batteries using fuzzy c-means is only about 13% of that of the supervised learning method [42]. Although unsupervised learning algorithms are faster for screening, the initialization of the centroids is rather random for the algorithms currently adopted in the literature [28,31,42], which is likely to result in less stable clustering outputs.

In summary, existing studies generally use the data obtained from module/cell-level testing to classify the modules/cells, suggesting the necessity of complete disassembly of the battery pack for individual testing, while methods to classify the module-level aging using data obtained from practical pack-level testing have not been reported. In addition, little research for fast screening has been conducted on pack-level testing with large module-to-module inconsistencies. There is a possibility

that representative features used for classification in the state-of-the-art literature may not be extractable for each module in LIB packs with large module-to-module inconsistencies, causing the existing methods to be inapplicable. Finally, as most unsupervised learning algorithms have fairly random centroid initialization, it is imperative to investigate the incorporation of unsupervised learning with stability.

1.2. Contributions of the work

To address the above research gaps and accelerate the second-use progress of EV batteries, a fast and accurate screening performed by pack-level testing is proposed for the evaluation and classification of module-level aging. The main contributions of this work are as follows:

1) Only pack-level testing is involved for screening, which means that this approach is time- and energy-efficient, as there is no need to test the batteries individually.

2) Dynamic characteristic-based criteria are designed to cover the comprehensive performance of retired modules, making the approach applicable for battery packs with large module-to-module SOC inconsistencies of up to 30%.

3) Adaptive affinity propagation clustering is applied to classify the modules and further save time from an algorithmic perspective, achieving stable and accurate screening without pre-training the model.

4) The verification is conducted on the real testing data of a 5 parallel-95 series (5P95S) battery pack. The static and dynamic performance consistency is evaluated and compared with three benchmarks after the screening.

2. Methodology

2.1. Dynamic characteristic-based screening principle

As mentioned previously, screening based on static-characteristic criteria may be incomprehensive to reflect the battery performance. Therefore, we propose a machine learning screening approach that enables fast, accurate and stable screening using dynamic characteristic-based criteria extracted from pack-level testing. The basic idea is to extract the distinguishing criteria that the batteries exhibit during operation. The framework of our proposed approach is shown in Fig.1. Here we are dealing with the LIB pack with a large number of modules in series, in our case, 95 modules in series. The modules were considered to be the minimum screening unit in this study since

screening retired batteries at the module level balances the economics of battery disassembly with the flexibility of regrouping [18]. The screening data are generated from the modules' voltage data obtained from the pack-level testing. Both charge and discharge processes are included because the terminal voltage of the batteries may behave differently due to the direction of polarization. Dynamic characteristic-based criteria, consisting of four criteria of voltage proportion and two criteria of phase transformation reactions number, are extracted from the module terminal voltage and the corresponding IC curves from the pack-level testing. We employ the adaptive affinity propagation algorithm to cluster the data, as it does not require model pre-training, further improving the screening efficiency from the algorithmic perspective. The consistency performance of the screened modules is evaluated and validated with the module-level and pack-level experiments.

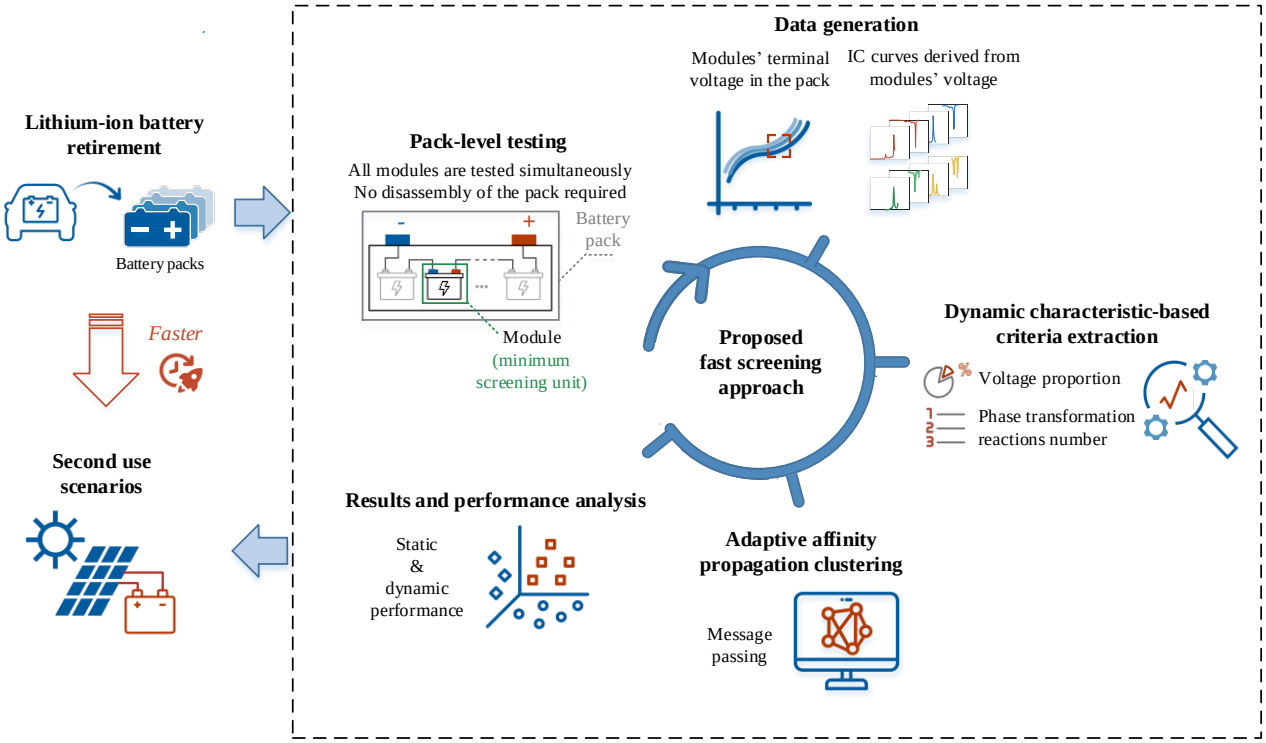

Fig.1. The framework of the proposed screening approach for retired LIBs.

The cyclic aging of the LIB packs is affected by many extrinsic factors, including temperature, C-rate, depth-of-discharge, mechanical stress, storage time and operating time [44]. These extrinsic factors, combined with the intrinsic factors, exert varied effects on the batteries and will result in inconsistencies among them. The SOC interval is a dominant factor [37], and it is also a consequence of degradation. Therefore, we extract the dynamic criteria characterizing the SOC intervals during battery operation based on the modules' terminal voltage. As the SOC is related to the terminal voltage, we calculate the voltage proportion that the low and high voltage of the module accounts for in the

charging and discharging operations of the battery pack to denote the low and high SOC intervals. The peaks observed on the IC curve of the $LiFePO_4$ battery, corresponding to the staging in the graphite of the negative electrode, take place mainly in the full-cell voltage plateau of 3.20V-3.40V. So we chose voltage proportions below 3.20V and above 3.40V of the module during the pack charge-discharge process as four criteria, defined as the Vl_{chg}, Vh_{chg}, Vl_{dch} and Vh_{dch} for $LiFePO_4$ battery, which can be computed by Eq.(1). The criteria are extracted from the modules' voltages in the pack-level testing, whose curves are shown in Fig. 2(a) and (b). The subscripts *chg* and *dch* indicate whether the criteria are extracted from charge or discharge.

$$\begin{aligned} Vl &= \Delta V_{<3.20V} / \Delta V_{interval} \\ Vh &= \Delta V_{>3.40V} / \Delta V_{interval} \end{aligned} \tag{1}$$

Where $\triangle V_{<3.20V}$ and $\triangle V_{>3.40V}$ denote the voltage changes of module index *im* when the terminal voltage is below 3.20V and above 3.40V, respectively. The $\triangle V_{interval}$ is the total voltage change measured at the end of the charging or discharging process. These variables are illustrated in Fig.2 (c). When a battery is operated in the low SOC interval, the value of Vl_{chg} and Vl_{dch} will be larger compared with those operating in the middle SOC interval, as the voltage plate of the $LiFePO_4$ battery results in little voltage change. Vl_{chg}, Vh_{chg}, Vl_{dch} and Vh_{dch} are, therefore, four indicators that provide a good description of the SOC interval.

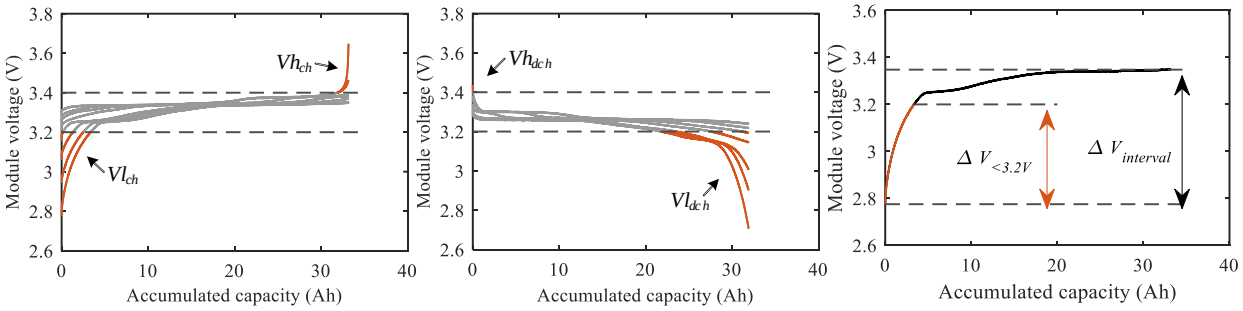

Fig. 2. Voltage proportion criteria extracted from the modules' voltage obtained from pack-level testing during (a) charging, (b) discharging, and (c) an illustrative example of the variables used to calculate Vl_{chg}. (Note: only a small amount of module voltage has been plotted for clarity of presentation.)

However, the terminal voltage proportion alone is far from sufficient. For the main voltage plateau, the proposed *ICreact* criterion, which represents the number of phase transformation reactions exhibited at the pack-level testing, i.e., the number of IC peaks, is used to evaluate the inconsistency and classify the batteries. The *ICreact* values for modules with different numbers of IC peaks are

suggested in Table 1. Fig.3 (a) and (b) show the IC curves obtained from the module-level and pack-level testing. During the pack-level testing of the 5P95S battery pack, all modules were tested in the pack simultaneously, while the module-level testing was conducted on the 95 modules individually (the detailed procedure is described in Section 3), which is why the IC curves displayed in Fig.3 are complete at the module level, whereas the IC curves at the pack-level are only partial (inconsistencies among the modules cause other modules to reach cut-off condition early).

The peak ⑤ during discharge is not as distinguishable as it is during charge (Fig.3(a)). Therefore, we only identified peak ② and peak ①. During charge and discharge, we noticed that the height of peak ① of some modules varied considerably. As shown by the red line in Fig.3(b), the height of peak ① is almost equal to the height of peak ② during discharge, which poses a challenge for screening methods based on the peak height of the IC curves. We believe that the dramatic change in the height of peak ① may be related to the differences in the polarization process induced by different initial SOCs. That is also the reason why we incorporate both charge and discharge directions in criteria extraction: to capture comprehensive inconsistency performance. As illustrated in Fig.3 (b), the shape and height of peak ① are similar for the two batteries during pack-level charge, but distinct differences can be observed during discharge. We found that this was not an exception, and thus, we used the relative height with respect to peak ② to identify these modules. If the height of peak ② is not twice the height of peak ① during discharge (both peaks are of similar height), then the *ICreact* value for the module exhibiting this performance will be 2. Otherwise, the *ICreact* value will be 1.5, as these modules are likely to display only half of peak ① during the charge at the pack level. The *ICreact* criterion, although based on the IC curve, is not concerned with the exact height of the peaks and is, therefore, insensitive to the noise errors introduced by the IC calculation. It is worth mentioning that the parameters presented may need to be slightly modified to suit the different battery electrode materials, as they have varied voltage characteristics. Yet the essence is the same, that is, to characterize the differences reflected by SOC during pack-level operation.

Table 1 The *ICreact* values for modules with different numbers of IC peaks.

Item		*ICreact* values
Charge	Discharge	
half-②	half-②	0.5
complete-②	complete-②	1
complete-② and half-①	half-② and complete-① or complete-② and complete-① (② is twice as high as ①)	1.5
complete-② and complete-①	complete-② and complete-①	2

After all the dynamic characteristics-based criteria have been obtained (Eq.(2)), each criterion is normalized by centering the data to have a mean zero to aid comparison. The subscripts *chg* and *dch* indicate whether the criteria are extracted from charge or discharge.

$$\theta_{dp} = [Vl_{chg}, Vh_{chg}, ICreact_{chg}, Vl_{dch}, Vh_{dch}, ICreact_{dch}] \quad (2)$$

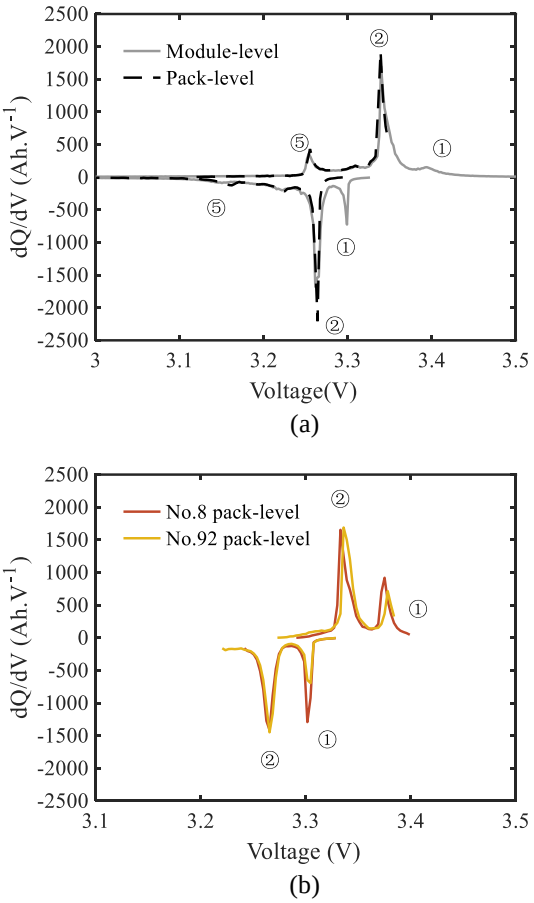

Fig. 3. The IC curves obtained from the module-level and pack-level testing of (a) No.39 battery; and IC curves obtained from the pack-level operation of (b) No.8 and No.92 batteries.

2.2. Adaptive affinity propagation clustering algorithm

Developed based on the concept of "message passing" between data points, unlike clustering algorithms such as k-means or k-medoids, a predetermined or estimated number of clusters is not a necessity for affinity propagation (AP) [45]. Adaptive affinity propagation (adAP) was developed by Wang et al. [46] to improve preference tuning, the convergence and to eliminate oscillations during the iterations in the original AP algorithm. Instead of randomly choosing the centers as an initialization, adAP treats all the data points as potential clustering centers (called exemplars), contributing to a more

stable result. The main equations are as follows and a detailed description is given in Ref.[46,47].

Suppose the dataset is $D_{n\times m} = \{d_1, d_2, d_3, \ldots, d_n\}$, where d_i contains m features and n samples. The similarity matrix S (Eq.(3)) is an $n \times n$ matrix that describes the similarity between d_i and d_j. The diagonal of S represents the likelihood of a particular input being an exemplar, known as "preferences", and is set as half of the median of the input similarities at the initial.

$$S(i,j) = -\left\| d_i - d_j \right\|^2 \quad (i \neq j) \tag{3}$$

The responsibility $R(i,k)$ conveys the attraction messages from data point i to candidate exemplar point k, indicating how well-suited point k is to serve as the exemplar for point i when taking other potential exemplars for point i into account. In turn, the availability $A(i,k)$, sent from candidate exemplar point k to point i, reflects how appropriate it would be for point i to choose point k as its exemplar, considering the support from other points that point k should be an exemplar. Eq.(4) and (5) show the formulas for responsibility and availability, respectively.

$$\begin{aligned} R(i,k) &= S(i,k) - \max_{k' \neq k} \left\{ A(i,k') + S(i,k') \right\} \\ R(k,k) &= S(k,k) - \max_{k' \neq k} \left\{ S(k,k') \right\} \end{aligned} \tag{4}$$

$$\begin{aligned} A(i,k) &= \min \left\{ 0, R(k,k) + \sum_{i' \notin \{i,k\}} \max \left\{ 0, R(i',k) \right\} \right\} \\ A(k,k) &= \sum_{i' \neq k} \max \left\{ 0, R(i',k) \right\} \end{aligned} \tag{5}$$

During the iterative process, R_t and A_t are updated with R_{t-1} and A_{t-1} from the previous iteration using the damping factor λ, which is demonstrated in Eq.(6). The preference p is also updated adaptively with the current number of exemplars K, i.e., the decrease of p will accelerate until the K exemplary decline. This is realized through Eq.(7)-(8), where the p will decrease $b \times p_{step}$ every certain iteration if the number of exemplars converges to K. The λ will be increased in steps of 0.05 if oscillations occur within the monitoring window. If the λ is larger than 0.85, the p will gradually decrease by step p_{step} until oscillations disappear.

$$\begin{aligned} R_t &= (1-\lambda) \times R_t + \lambda \times R_{t-1} \\ A_t &= (1-\lambda) \times A_t + \lambda \times A_{t-1} \end{aligned} \tag{6}$$

$$p = p + b \times p_{step} \tag{7}$$

$$p_{step} = 0.01 p_m / q$$
$$q = 0.1\sqrt{K + 50} \tag{8}$$

Where b is the counter for each K exemplary converges, p_m is the median of the similarity.

The availabilities and responsibilities are combined together to identify exemplars, displayed as Eq.(9). The element with the maximum criterion value $C(i,k)$ will be designated to be an exemplar. Elements corresponding to the rows which share the same exemplar are clustered together. The iterations are terminated by reaching K=2, so that the lower limit of the preference is obtained. We set the maximum number of repetitions to be as large as 50000 to ensure that it will not affect whether the algorithm reaches K=2. The minimum number of samples in one cluster is set to 4, suggesting that a cluster will be omitted if a cluster contains fewer than 5% of the total sample (95*5%).

$$C(i,k) = R(i,k) + A(i,k) \tag{9}$$

If the dataset $D_{n\times m}$ is divided into K clusters C_i (i = 1,2,…,k), the optimal number of clusters can be determined by the silhouette coefficient, which can be calculated as Eq.(10). With $Sil(t)$ for each sample, the overall average silhouette Sil for n samples of the dataset can be obtained. The largest overall average silhouette indicates the best clustering quality and the optimal cluster number. A series of Sil values corresponding to clustering solutions under different cluster numbers are calculated, and the optimal clustering is the case where the number of clusters yields the largest Sil.

$$Sil(t) = \frac{b(t) - a(t)}{\max\{a(t), b(t)\}} \tag{10}$$

where $a(t)$ is the average Euclidean distance of sample t of C_j to all other samples in C_j, and $b(t)$ is the minimum value of the average Euclidean distance of sample t of C_j to all samples in another cluster C_i (i = 1,2,…,k and $i \neq j$).

3. Experimental

3.1. Configuration of the retired battery pack

The LIB pack retired from an electric vehicle with a mileage of 32,500 kilometers that had been operating in a southern Chinese city for over three years was used for validation. The pack has not been equalized throughout its lifetime. The configuration of the battery pack is 5P95S, with 5 cells connected in parallel to form a 5P1S module and 95 modules bolted together in series to form a battery pack. The parameters of the 5P1S module are shown in Table 2.

Table 2 Specifications of the 5P1S module.

Item	Parameters
Electrode material	$LiFePO_4$-graphite
Nominal capacity	60Ah
Nominal energy	18.5Wh
Weight	1.738kg
Nominal voltage	3.2V
End-of-charge voltage	3.65V
End-of-discharge voltage	2.5V

3.2. Experimental design

A series of tests were carried out on the battery pack and the modules to determine the inconsistency of the modules and to verify the effectiveness of the proposed approach. The experiments designed for implementation and validation of the classification as well as the corresponding specific testing procedures, are outlined in Fig.4. First, we conducted the pack capacity test to obtain the present aging state of the battery pack. The pack capacity test takes much less time to perform than the module capacity test that follows, and the testing data is used for classification implementation. Then, we designed the capability test at 0.3C, 0.7C, and 1.2C at the pack level to not only acquire a picture of the inconsistencies among the modules but also to set the stage for verifying the adaptability of the proposed approach under diverse scenarios. Prior to the module-level experiments, the battery pack was fully charged at 0.1C until any modules reached the end-of-charge voltage. To evaluate and validate the performance consistency of the classification, the battery pack was disassembled into 95 modules for module-level testing, which continued to be charged individually to their end-of-charge voltage at 0.1C to acquire the SOC inconsistency. Modules comprising 5 cells connected in parallel were not further disassembled. Finally, we performed the module capacity test on 95 modules individually under 0.1C to obtain the voltage-capacity reference values for classification validation. The open circuit voltage (OCV)-SOC and charging-based IC of the modules are also derived from this test. All data were recorded at 1s time intervals.

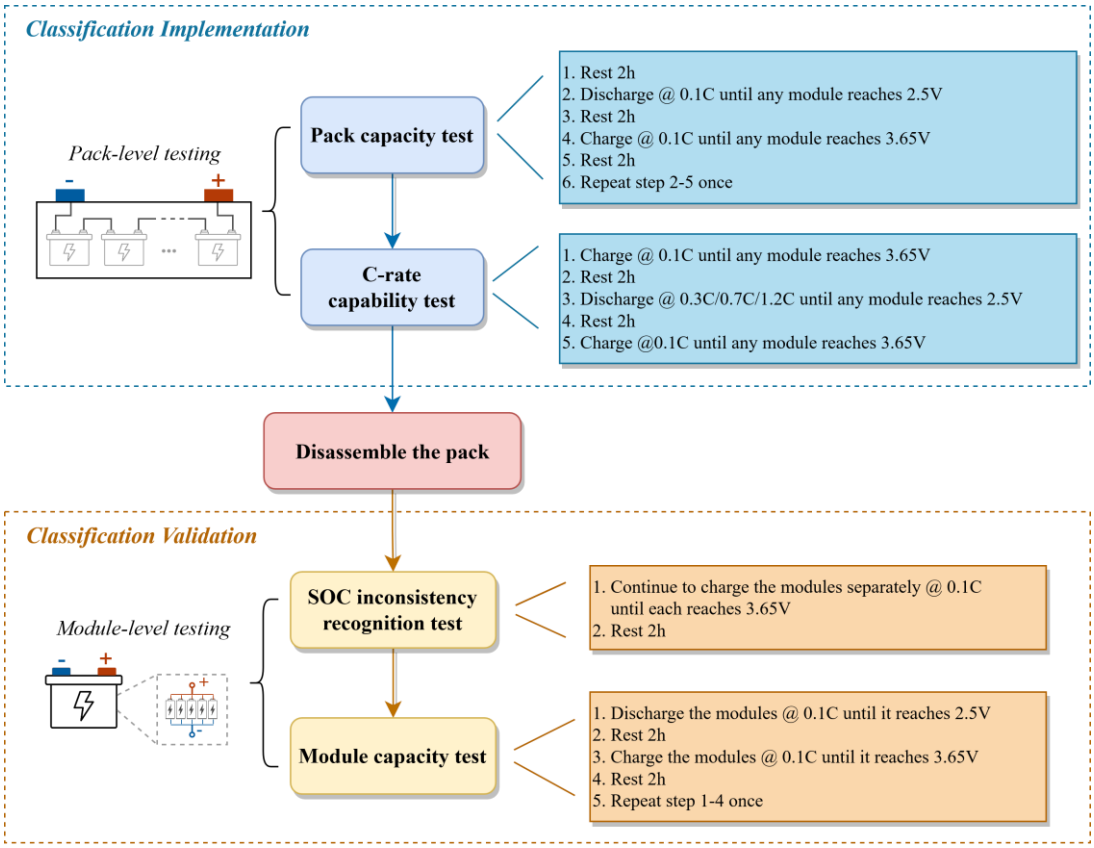

Fig. 4. The experiment procedures for implementation and validation of the proposed screening approach.

Fig.S1 demonstrates the experimental platform composition and the schematic principle diagram. For pack-level experiments, a test bench comprised of an Arbin BT2000 testing system, a BMS, and a host computer was used, with the battery pack's ambient temperature at roughly 23°C. For both current and battery pack voltage, the measurement error was ±0.05% of the full-scale range. The voltage precision for module voltage acquisition was ±0.02% of the full-scale range. As for module-level experiments, an Arbin BT2000 and a Maccor Series 4000 were employed. The Maccor Series 4000 had a current and voltage accuracy of ±0.05% and ±0.02% of the full-scale range, respectively. The module experiments were carried out in a thermal chamber set at 25°C.

3.3.Performance inconsistency analysis of the retired battery pack

The pack and module capacity is defined as the last charge capacity measured in the pack capacity test and module capacity test, respectively. As a result of module degradation and SOC inconsistency, the measured capacity of the battery pack is around 32.90 Ah. Namely, more than 45% of the nominal capacity is lost. The maximum and minimum value of module capacity retention is 85.22% and 73.72%, respectively. The average capacity of the modules is 49.46 Ah (capacity retention: 82.4%), with a standard deviation of 1.50 Ah (3%). Despite the small variation in the capacity of these modules,

the difference in their dynamic performance is noticeable, which is shown in supplementary Fig.S2 labeled as without (w/o) screening. The results of the pack C-rate capability test are plotted in Fig.5. The capacity or energy efficiency are defined as the capacity or energy of the battery pack discharged at a given C-rate divided by the maximum charge capacity or energy of the battery pack. The 0.1C charging data is considered to be the maximum value that can be achieved by the battery pack and is taken as the denominator when calculating capacity or energy efficiency. As expected, as the C-rate increases, both capacity and energy efficiency decrease, indicating large module-to-module inconsistencies. The terminal voltages of the modules in the pack capacity test confirm the severe inconsistencies in SOC interval. Fig.5 (b) reveals the IC curve of the maximum and minimum voltage in the pack under 0.1C charging. The overlap between these two voltages is only around 50mV, which may pose a challenge for some methods [42,43] where the prerequisite of small module-to-module consistencies have to be met.

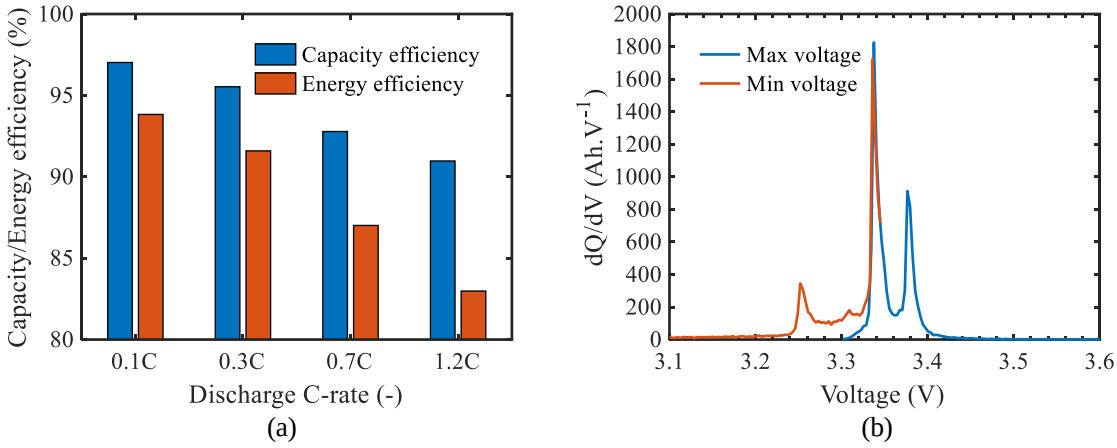

Fig. 5. Performance inconsistency analysis of the retired LIB pack: (a) capacity and energy efficiency, (b) charging-based IC curves for the maximum and minimum module voltages under 0.1C.

To further quantify the degree of SOC inconsistency, the modules' operating SOC is also calculated, whose equations are given in Supplementary Note S1 (Eq.(S1)-(S3)). The data from the pack capacity test, module capacity test and SOC inconsistency recognition test are used to obtain the operating SOC. According to Fig. 6 (a) and (b), the range and standard deviation of modules' SOC are around 30% and 8.5%. When the battery pack is at 100% SOC, up to 34.74% of the modules have more than 10% SOC remaining before they are full. When the battery pack is at 0% SOC, the percentage of modules with at least 10% discharged capacity increases to over 90%. Half of them can even release above 30% capacity. The distribution of the modules' Δ SOC is demonstrated in Fig.6 (c),

whose maximum and minimum values are 72.6% and 61.6%, confirming the large SOC inconsistency.

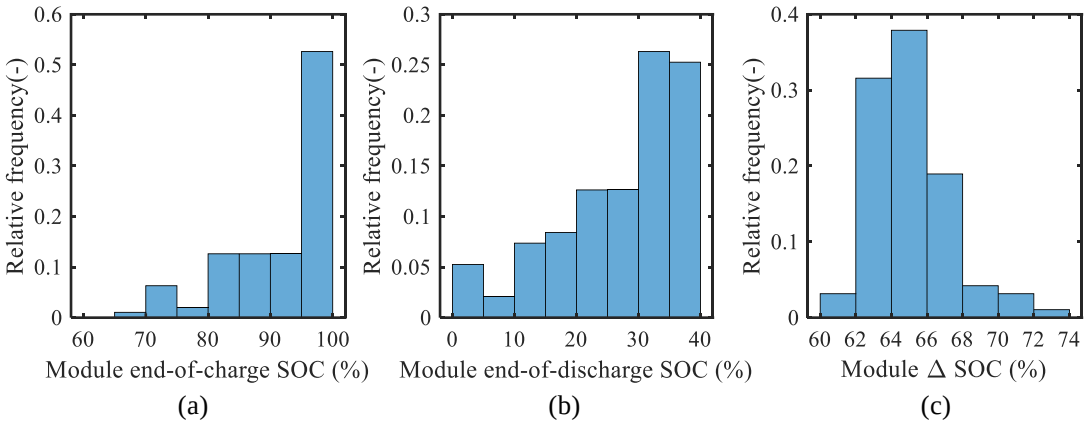

Fig. 6. Histograms of the relative frequency of module (a) end-of-charge SOC at pack 100% SOC, (b) end-of-discharge SOC at pack 0% SOC and (c) Δ SOC in the retired LIB pack.

4. Results and discussion

4.1. Comparison of the screening results

The screening process is based on pack-level testing and the performance consistency of the screened modules is evaluated and validated with the module-level and pack-level experiments. To verify the effectiveness of the proposed screening approach, dynamic characteristics-based criteria, static characteristics-based criteria and random selection approaches are selected as benchmarks for comparison. The 5P1S module is considered a minimum screening unit in this study and all benchmarks from the literature are recalculated with our 5P95S battery pack data.

Approach 1: Our proposed approach. The six dynamic characteristics-based criteria for each module obtained by Eq.(1) and Table 1 are used as input to the adAP algorithm for classification.

Approach 2: A quick screening approach based on fuzzy c-means from Ref.[42]. Given that they also used two dynamic characteristics-based criteria, namely the gradient and the volatility of a symmetrical charging voltage interval near the main phase transformation, they were included for comparison. Moreover, they did the screening based on the $LiFePO_4$ batteries, so these features should be potentially applicable to our data (also $LiFePO_4$). However, we cannot extract these criteria based on the modules' terminal voltage obtained from our pack-level testing data for all the modules due to the limited operating range of the module induced by the large module-to-module SOC inconsistencies. Hence, for Approach 2, these two criteria are still extracted based on our module-level testing data.

Approach 3: Directly capacity screening based on k-means. To verify the merits of dynamic characteristics-based criteria, we adopt the static characteristics-based criteria as an alternative comparison, in which case the measured module capacity acquired from the module capacity test is used as the screening criterion, and then k-means clustering is applied.

Approach 4: Random selection. Because of its easy-to-implement nature, the random selection approach is included, with the quantities of modules per class remaining consistent with the clustering results of Approach 1.

The optimal number of clusters is determined automatically in the adAP algorithm according to Eq.(10), which in our case, the algorithm returned a silhouette coefficient of 0.7546, corresponding to 4 classes. To make the results among different approaches comparable, the number of clusters for fuzzy c-means and k-means is equally set to four. We calculate the mean of modules' capacity in each classified group and name them in ascending order. In other words, the mean value of the module capacity is the smallest in Class 1 and the largest in Class 4. Given that the number of clusters set to four may not be the optimal classification for Approach 2 and Approach 3, performance consistency results for Approach 1-3 under the different number of cluster settings are provided in Supplementary Note S2.

The performance consistency of the screening results is evaluated from two perspectives: static (capacity) and dynamic performance (OCV-SOC, charging-based IC and terminal voltage curve during C-rate capability test) within the classified group. The standard deviation varies with the magnitude of the samples in the dataset and is not intuitive when comparing data series with different means, as is the case for different classification groups. The coefficient of variation (CV) provides a good comparison of the degree of variation from one data series to another, even if the means are highly different from one another. Therefore, statistical indicators of *range* and *CV* were employed for both static and dynamic performance to quantify the inconsistency under each classified group, and can be computed as follows,

$$range_{EvalP,C_i} = \max(EvalP_{im,C_i}) - \min(EvalP_{im,C_i}) \quad im = k_1, k_2, ..., k_n \tag{11}$$

$$mean_{EvalP,C_i} = \frac{1}{n} \sum_{im=k_1}^{k_n} EvalP_{im,C_i} \tag{12}$$

$$std_{EvalP,C_i} = \sqrt{\frac{1}{n-1}\sum_{im=k_1}^{k_n}\left|EvalP_{im,C_i} - mean_{EvalP,C_i}\right|^2} \quad (13)$$

$$CV_{EvalP,C_i} = std_{EvalP,C_i} / mean_{EvalP,C_i} \quad (14)$$

where the subscript *EvalP* indicates the evaluated performance, which in our case is *cap*, *OCV*, and *volt*, representing the module capacity, the OCV under each SOC point, and terminal voltage during C-rate capability test, respectively. The *im* represents the index of the module and *n* is the module quantities in Class C_i (i = 1,2,3,4), respectively. The $range_{EvalP,C_i}$, $mean_{EvalP,C_i}$, std_{EvalP,C_i} and CV_{EvalP,C_i} are the range, mean, standard deviation and coefficient of variation of the evaluated performance of modules index of k_1, k_2,…, k_n in Class C_i, accordingly.

To combine the static and dynamic performance, the overall performance (OP_{C_i}) of the modules (defined by Eq.(15)) is calculated for each class, in which two static performance metrics and two dynamic performance metrics are included. The reciprocal of the mean module capacity ($1/mean_{cap,C_i}$) and the CV of the module capacity (CV_{cap,C_i}) in each class represent the static performance. The mean value of CV_{OCV} under each SOC ($mean_{CV_{OCV},Ci}$) and the mean value of CV_{volt} under each accumulated ampere-hours ($mean_{CV_{volt},Ci}$) are calculated and used as metrics to illustrate the dynamic performance inconsistency. The $1/mean_{cap,C_i}$, CV_{cap,C_i}, $mean_{CV_{OCV},Ci}$, $mean_{CV_{volt},Ci}$ are normalized to make the performances comparable. The smaller the value, the better the overall performance, indicating that a good performance should have a large mean module capacity, small variations in module capacity, OCV-SOC and terminal voltage during C-rate capability test.

$$OP_{C_i} = \frac{1}{4}\left(\left(\frac{1}{mean_{cap,C_i}}\right)_{norm} + (CV_{cap,C_i})_{norm} + (mean_{CV_{OCV},C_i})_{norm} + (mean_{CV_{volt},C_i})_{norm}\right) \quad (15)$$
$$i = 1,2,3,4$$

Table 3 compares the static capacity consistency in each class using four approaches. Unsurprisingly, the $mean_{cap}$ of Approach 4 shows no differences regardless of the class (all around 49.4Ah), implying the worst consistency among the other approaches, given that it was based purely on randomness. Thus, we exclude it from the comparison of the dynamic performance consistency. Approach 1 divided the 95 modules into 4 classes whose $mean_{cap}$ are separated. The $mean_{cap}$ for

Approach 1 from Class 1 to Class 4 are 46.02Ah, 48.29Ah, 49.08Ah and 50.47Ah, respectively. In Approach 1-3, the $mean_{cap}$ of the modules for each class is separated, i.e., around 46Ah, 48Ah, 49Ah, and 50Ah, respectively, which illustrates the ability for effective screening and a degree of division accuracy of the three approaches. The $mean_{cap,C_i}$ for each class are similar across the other three approaches (Eq.(12)), so it is reasonable to compare the range and CV of the capacity ($range_{cap,C_i}$ and CV_{cap,C_i}).

At first glance, Approach 3 appears to be superior in all aspects to Approach 1 and Approach 2. Indeed, in terms of capacity, it has the least range and CV in each class. However, in the dynamic performance consistency analysis that follows, we will reveal that this result is 'deceptive'. Approach 1 and Approach 2 have the same $range_{cap}$ in Class 1, but Approach 1 is slightly better than Approach 2 in terms of the CV_{cap}. Compared with Approach 2, the average $range_{cap}$ and CV_{cap} using Approach 1 are reduced by 22.10% and 17.36%, specifically by 52.32% and 34.25% in Class 2, 25.23% and 19.67% in Class 3, and 10.83% and 12.26% in Class 4, respectively.

Table 3 Comparison of capacity consistency in each class using four approaches (units for *mean* and *range*: Ah, and the value of *CV* has no units and is multiplied by 100 for demonstration purposes).

	Class 1			Class 2			Class 3			Class 4		
Approach	*mean*	*range*	*CV*	*mean*	*range*	*CV*	*mean*	*range*	*CV*	*mean*	*range*	*CV*
1	46.02	2.972	2.254	48.29	2.515	1.509	49.08	2.809	1.659	50.47	3.239	1.422
2	46.02	2.972	2.330	48.56	5.275	2.295	49.57	3.757	2.065	50.47	3.633	1.620
3	45.77	2.566	1.971	48.09	1.686	1.043	49.67	1.353	0.773	50.91	1.806	0.962
4	49.46	3.897	2.841	49.46	5.188	2.925	49.48	5.677	2.934	49.46	6.749	2.975

The OCV-SOC curve is calculated as the average of the discharging and charging voltage obtained from the module capacity test. Fig.7 illustrates the range ($range_{OCV,C_i}$) as well as the CV (CV_{OCV,C_i}) of the OCV curves for all modules in Class C_i under each SOC point employing three approaches. After the screening, the $range_{OCV}$ is indeed reduced in each class (Supplementary Fig.S2). The results of the module quantities of Class 1-4 for three approaches are marked in white in Fig.8. In the first class, even though Approach 1 has one more module than Approach 2 and Approach 3, the $range_{OCV}$ of Approach 1 is less than or equal to that of Approach 2 in more than 88% of the SOC intervals and the CV_{OCV} is even slightly smaller in the 20-40% SOC intervals (Fig.7 (a) (e)). As expected, the variation is much larger in both Class 2 and Class 3. The CV_{OCV} of Approach 2 is greater than that of Approach

1 by exceeding 90% and 80% of the SOC intervals in Class 2 and Class 3, respectively. In fact, there are around 30% of SOC intervals where the $range_{OCV}$ of Approach 2 is twice as large as that of Approach 1. Over 80% of the SOC intervals of CV_{OCV} of Approach 1 is smaller than or equal to that of Approach 2 in Class 4 (Fig.7 (d) (h)). In comparison to Approach 2, the average $range_{OCV}$ and CV_{OCV} of four classes using Approach 1 are decreased by 14.80% and 9.87%, respectively.

Similar conclusions can be drawn from the results of Approach 3 and Approach 1. Surprisingly, the $range_{OCV}$ and CV_{OCV} of Approach 3 are noticeably larger in both Class 2 and Class 3 compared with that of Approach 1, while the $range_{cap}$ and CV_{cap} of Approach 3 are actually smaller. This is because batteries with similar capacities are not necessarily expected to manifest a similar thermodynamic behavior, confirming the importance of dynamic characteristic-based criteria in screening. Approach 3 achieves marginally better OCV-SOC consistency in Class 4 since the module quantity is 19 batteries less than in Approach 1. Although Approach 1 is slightly worse than Approach 3 in terms of $range_{OCV}$ for the last class, this difference is still within an acceptable margin of no more than 5mV maximum. Overall, the average $range_{OCV}$ and CV_{OCV} of four classes utilizing Approach 1 are 9.39% and 8.67% lower than those using Approach 3.

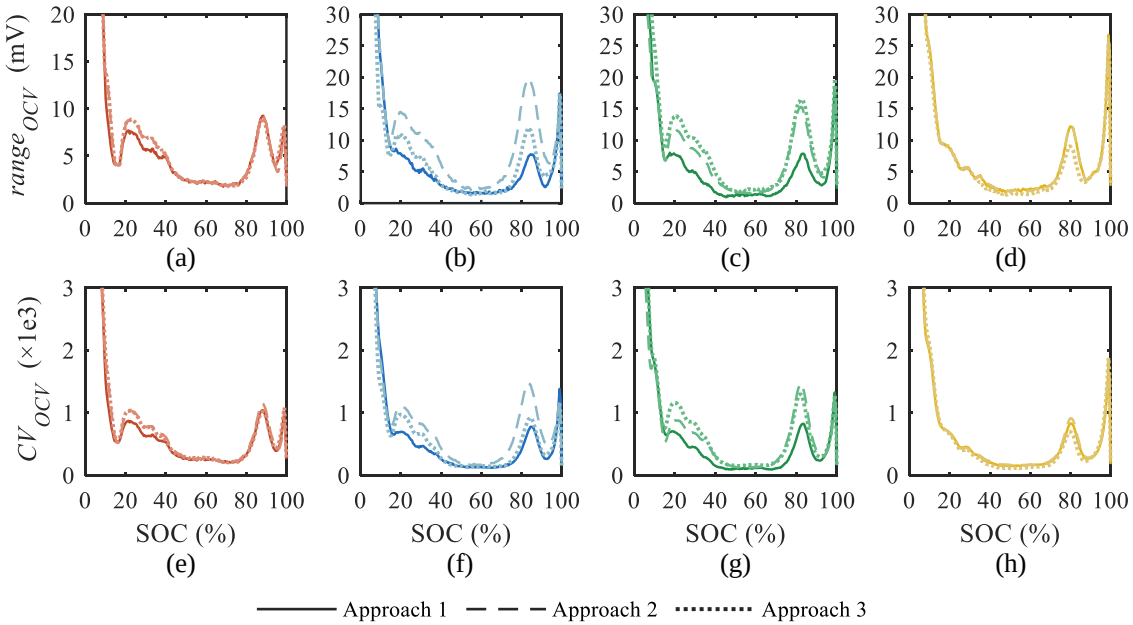

Fig. 7. The range and the coefficient of variation of OCV-SOC in (a)(e) Class 1, (b)(f) Class 2, (c)(g) Class 3, and (d)(h) Class 4 under three approaches (solid: Approach 1, dash: Approach 2, dot: Approach 3).

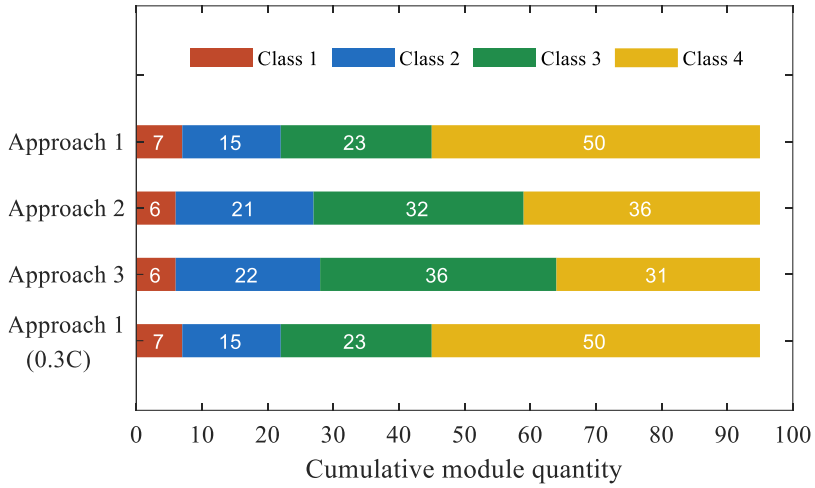

Fig. 8. Results of the module quantities for three screening approaches.

Fig.9 (a)-(d) shows the charging-based IC curves of Approach 1 in different classified groups with a zoomed-in voltage interval of 3.365-3.415V. The IC curves are calculated using the charging voltage obtained from the module capacity test. The peak height for peak ① around 3.38V, in general, is also in ascending order with the class number. The majority of the peak ① height is about 300, 450, 600 and 850 Ah·V^{-1} in Class 1, Class 2, Class 3 and Class 4, respectively. The peak shapes among different classes are varied, indicating a more consistent aging loss in each class and effective separation. In the first class, only one module is different between Approach 2 and Approach 3. Except for the additional module in Approach 1, there is also only one module different in Approach 1 and Approach 3. Therefore, the performance of charging-based IC curves of Approach 2 and Approach 3 in Class 1 would be quite close to that of Approach 1 and are not presented here. The charging-based IC curves behave similarly in Class 4 and are also not shown. This is in line with the previous analysis that the OCV-SOC inconsistency among the three approaches is relatively small. Nevertheless, the situation is different for Class 2 and Class 3. As can be seen from Fig.9 (b)(e)(f), the maximum value of peak ① in Class 2 is below 500Ah·V^{-1} in Approach 1 while it reaches around 750 Ah·V^{-1} in Approach 2 and Approach 3. Similarly, in Fig.9 (c) (h) of Class 3, the maximum value of peak ② is around 3500Ah·V^{-1} in Approach 1 while it reaches around 4000 Ah·V^{-1} in Approach 3, which is almost similar to the value of Approach 1 in Class 4. This shows that Approach 1 can better divide the modules into separate performances but with similar performances within the classes. As can be seen from Fig.9 (b)(e)(f) and (c)(g)(h), the blue for Class 2 and the green for Class 3, the peak ① around 3.38V is more concentrated in Approach 1 than in Approach 2 and Approach 3, which illustrates the good dynamic performance consistency of modules using Approach 1.

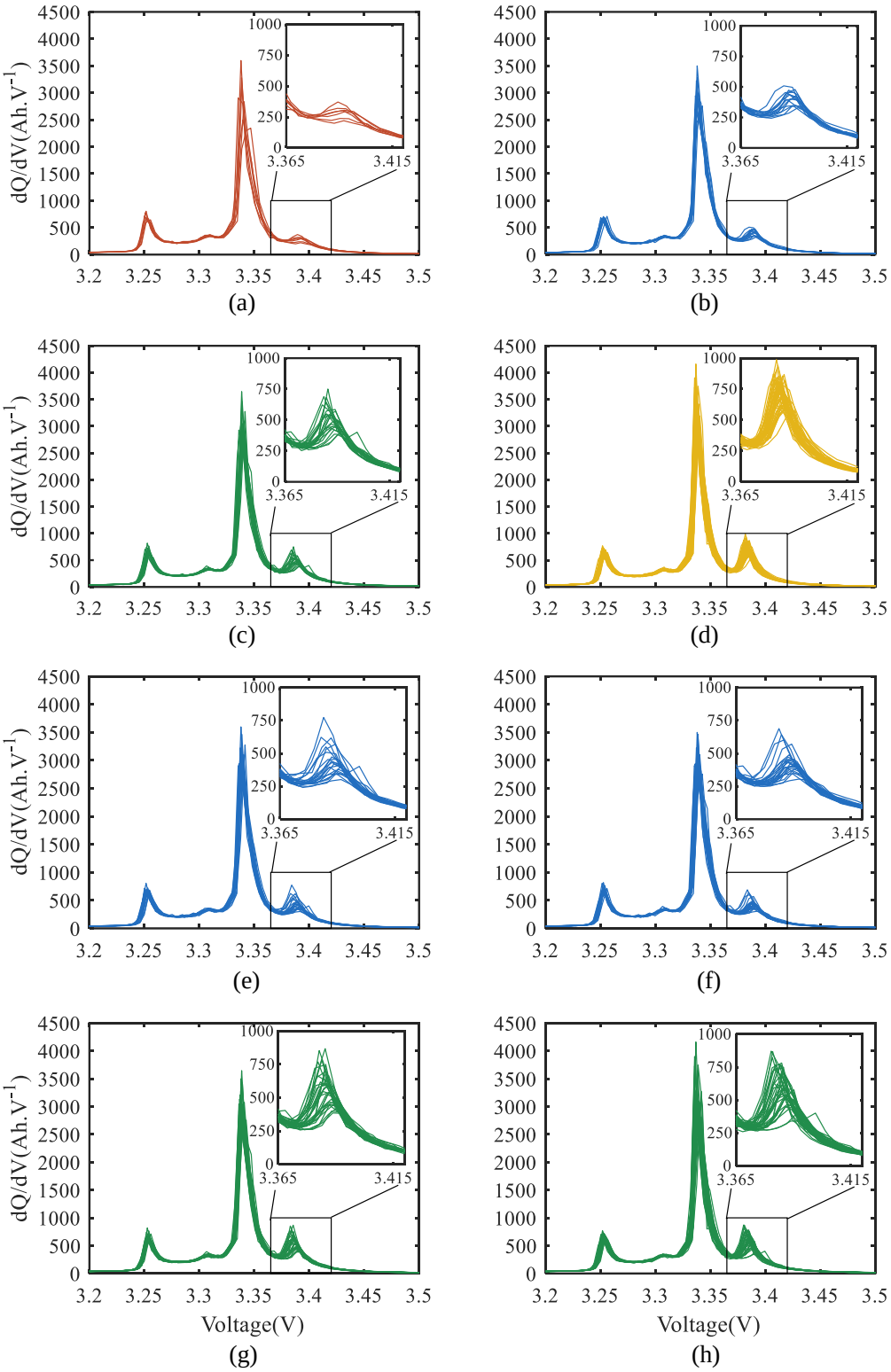

Fig. 9. Charging-based IC curves of classified modules employing three approaches: results of Approach 1 in (a) Class 1, (b) Class 2, (c) Class 3, and (d) Class 4; results of Approach 2 in (e) Class 2 and (g) Class 3; results of Approach 3 in (f) Class 2 and (h) Class 3.

Fig.10 demonstrates the initial, mean and maximum range and CV of the terminal voltage ($range_{volt}$ and CV_{volt}) during a 0.7C discharge for Approach (Ap.) 1-3. The results of the 1.2C discharge are similar to those of 0.7C and are presented in Supplementary Fig.S3. The initial $range_{volt}$ and CV_{volt} is determined using the first data point at the beginning of the testing. In general, the $range_{volt}$ and CV_{volt} display similar trends. The mean and maximum CV_{volt} are smaller overall in Approach 1 than in Approach 2 and Approach 3, meaning that the inconsistency of Approach 1 is indeed reduced compared with the others. More specifically, the mean CV_{volt} of Approach 1 is reduced by 44.3% and 38.4% on average compared with Approach 2 and Approach 3. The maximum $range_{volt}$ of Approach 1 in Class 1 is 0.285V while that of Approach 2 and Approach 3 reaches 0.448V, nearly 1.6 times larger. Compared with Approach 2 and Approach 3, the maximum $range_{volt}$ and maximum CV_{volt} of Approach 1 in Class 2 and Class 3 is reduced by more than 50%. The dynamic performance of the terminal voltage of Approach 2 and Approach 3 is fairly similar in terms of mean and maximum values. This is partly because the dynamic characteristic-based criteria used in Approach 2 are highly correlated with capacity. For Class 4, the variation in $range_{volt}$ and CV_{volt} is not obvious for the three approaches, which is in agreement with the previous analysis. From Fig.10 (d), Class 3 and Class 4 in Approach 2 have a similar initial CV_{volt}. However, the differences between these two classes in terms of mean and maximum CV_{volt} are major, with the values of Class 2 being almost twice as large as those of Class 3, again confirming that static characteristic-based criterion alone (e.g., initial voltage point or the OCV at a certain SOC) is insufficient to ensure consistent dynamic performance during operation.

Table 4 compares the overall performance of four approaches in four classes. The overall performance of Approach 1 in four classes is 0.7593, 0.6146, 0.6283 and 0.5992, meaning that modules in Class 1 have the worst overall performance and modules in Class 4 have the best overall performance. The average overall performance for Approach 1 is 0.6503, which is improved by 18.94%, 4.83% and 34.41% compared with Approach 2, Approach 3 and Approach 4, respectively.

Table 4 The overall performance for four approaches in four classes.

	Proposed	Benchmarks		
Class	Approach 1	Approach 2	Approach 3	Approach 4
1	0.7593	0.8500	0.8244	0.8950
2	0.6146	0.8713	0.6977	0.8344
3	0.6283	0.7491	0.6594	0.9374
4	0.5992	0.6236	0.5456	0.8298
average	0.6503	0.7735	0.6818	0.8741

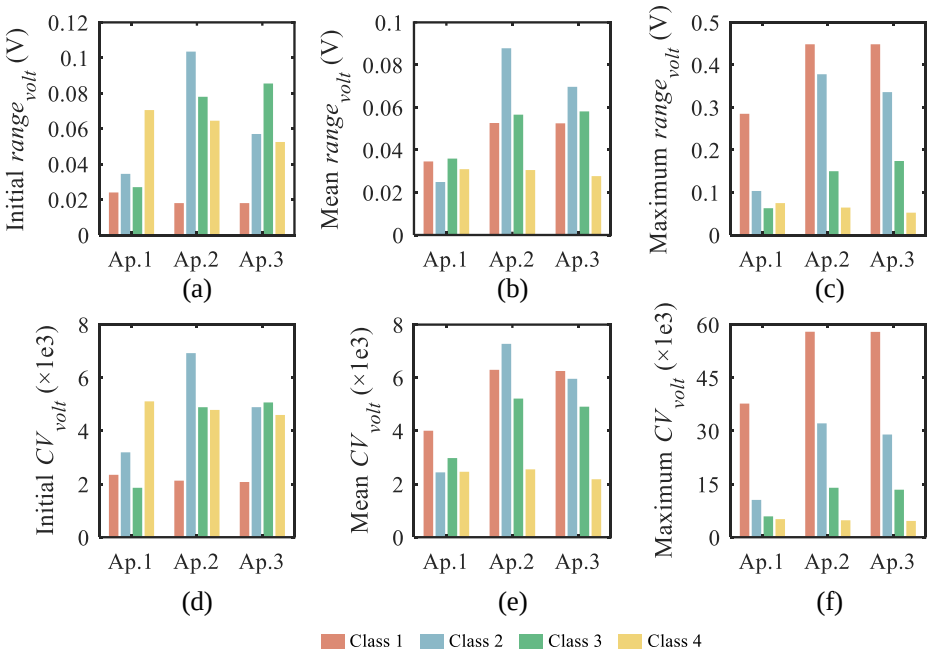

Fig. 10. The range and coefficient of variation of the terminal voltage during the 0.7C discharge for three approaches: (a)(d) initial, (b)(e) mean and (c)(f) maximum.

As for the screening time, compared with the approaches that require module-level criteria testing, the proposed approach shortens the testing time of the screening. Take the 5P95S battery pack in the paper as an example, a battery pack testing equipment containing auxiliary voltage measurements or battery management systems is enough to conduct the screening. However, it takes at least twice as long as the proposed approach if the number of channels in the module/cell testing equipment is assumed to be 50. Not to mention the time and effort involved in battery pack disassembly. The cost, energy and time saved are more considerable when the proposed approach is applied to large-scale retired LIB packs compared with approaches based on criteria that require module/cell-level testing.

To sum up, our proposed approach (Approach 1) can effectively screen the 95 modules into four classes with separated $mean_{cap}$. The modules in Class 1 have the smallest mean capacity and worst overall performance, while the modules in Class 4 have the largest mean capacity and best overall performance. The screening results show that our proposed approach has better module static and dynamic performance consistency among the four approaches. The average overall performance of the proposed approach is improved by 18.94%, 4.83%, and 34.41% compared with Approach 2, Approach 3 and Approach 4, respectively. In addition, our proposed approach only requires pack-level testing and does not rely on individual module-level testing to obtain the relevant screening criteria,

outperforming Approach 2 and Approach 3 in terms of screening time. Furthermore, although Approach 3 has the best static performance consistency, the results reveal that batteries with only similar capacities do not necessarily guarantee similar dynamic performance consistency, such as terminal voltage during operation. Therefore, it is indeed of necessity and significance to include dynamic characteristic-based criteria in screening in order to improve the performance consistency of the system.

4.2. Comparison of the algorithm stability

As the initialization of the centroid for most unsupervised learning clustering is random, in this section, we compare the screening results from multiple runs of the algorithms to show the stability of the adAP algorithm. The algorithms involved are adAP, fuzzy c-means and k-means, which are used respectively in Approach 1-3. Since the input criteria are different in Approach 1 and Approach 2, to show the advantages of the adAP algorithm, we employed our proposed dynamic characteristic-based criteria with fuzzy c-means for complementary illustration (referred to as Approach 5). Each approach was run 1000 times with the settings fixed in the algorithm (the setting is listed in Supplementary Note S3). In Approach 5, about 14.9% of the screening results in Class 3 had fewer than 4 samples, which has been subtracted from the probabilities shown in Table 5. For each run, the mean of the modules' capacity was used to identify the comparable classes of different algorithms. The standard deviation of the OCV at each SOC was calculated in each class (std_{OCV,C_i}). Then, the mean standard deviation of the OCV across the 0-100% SOC, shown as Eq.(16), was taken as the criterion to compare the stability of the algorithm, whose value would change if the screening results changed.

$$mean_{std_{OCV,C_i}} = \frac{1}{N_{soc}} \sum_{soc_j=0}^{100} std_{OCV,C_i}(soc_j)\ i = 1,2,3,4 \tag{16}$$

Where $std_{OCV,C_i}(soc_j)$ is the standard deviation of OCV of modules at soc_j in cluster C_i (i = 1,2,3,4). N_{soc} is the number of sampled SOC.

Finally, we computed the probability of different results for this criterion during the multiple runs of the algorithm, as presented in Table 5. As expected, the results of Approach 1 remain unchanged during the multiple runs, yielding the most stable results. This is because the adAP treats all the data points as exemplars and scans the "preferences". Moreover, the $mean_{std_{OCV,C_i}}$ of Approach 1 maintains the lowest in the first three classes compared with Approach 2 and Approach 3, supporting the

conclusions in Section 4.1. The value for the last class is not the smallest, which is not surprising and is consistent with our previous analysis. The probability that $mean_{std_{OCV,C_i}}$ exhibits some other value in either Approach 2 or Approach 3 implies that the batteries in the clusters have changed, resulting in more than one classification result. The reason for the poor stability is the random assignment of centroids in both approaches during initialization. For Approach 5, the same mean standard deviation as Approach 1 is achieved in each class (Table 5), which validates the effectiveness of the proposed dynamic characteristic-based criteria to some extent. But even with the same input criteria, Approach 5 shows some other values of probability, confirming the advantage of the adAP over fuzzy c-means in terms of stability and the necessity of adopting it.

Table 5 Probability of the mean standard deviation of the OCV over 0-100% SOC interval in each class during 1000 runs of the four approaches.

Class	Approach	$mean_{std_{OCV,C_i}}$ (mV)	Probability (%)
1	1	3.072	100.0
	2	3.274; 3.759; 3.831; 4.028	60.6; 27.6; 11.6; 0.2
	3	3.332	100.0
	5	3.072; 3.1839	53.4; 31.7
2	1	2.711	100.0
	2	3.367; 2.922; 2.951; 2.867	60.6; 27.6; 11.6; 0.2
	3	2.980; 2.884	59.0; 41.0
	5	2.711; 2.439	53.4; 31.7
3	1	2.5378	100.0
	2	2.848; 3.077; 3.043; 3.205	60.6; 31.9; 7.3; 0.2
	3	3.285; 3.265	59.0; 41.0
	5	2.538; 2.652	53.4; 31.7
4	1	2.746	100.0
	2	2.856; 2.449; 2.400; 2.452	60.6; 31.9; 7.3; 0.2
	3	2.667	100.0
	5	2.746; 2.803	53.4; 31.7

4.3. Discussion

4.3.1. Explanation of the approaches' results

Previously, we demonstrated the importance of dynamic characteristic-based criteria in screening. However, Approach 2, which also utilizes two dynamic characteristic-based criteria, failed to deliver

good screening results. This is probably because these dynamic criteria are still chosen by their correlation with capacity, which is essentially a "static" criterion. This also explains the similar performance of Approach 2 and Approach 3 in terms of dynamic consistency performance. Although the dynamic criteria of Approach 2 were capacity-dependent, it did not achieve similar screening results as Approach 3 regarding capacity consistency. Upon further investigation, we find that the Pearson correlation coefficient ρ between criteria and module capacity was only 0.7536 in our case, whereas in Ref.[42] ρ reached 0.9235, indicating that the two criteria chosen did not correlate as well with capacity as they did on their data so leading to less favorable static performance, i.e. capacity consistency.

Our proposed method yields superb results because it captures a critical feature of the module stemming from battery pack degradation, namely the SOC interval, which is both a consequence and a cause of degradation. It reflects the differences among the modules to some extent, which, after all, have almost identical values of the SOC interval initially. During the pack-level testing, the accumulated ampere-hours are the same for all series-connected modules. If the module has a larger capacity, the changes of SOC (ΔSOC) will be smaller. Conversely, if the capacity of the module is smaller, the ΔSOC is larger. Therefore, if some criteria representing the ΔSOC can be found, these criteria can be used to classify the modules into different capacity ranges. Our proposed criteria not only reflect the changes in SOC but also contain information about the starting SOC, which further describes the operating SOC interval resulting from the degradation. Due to exposure to an inhomogeneous external environment, the SOC intervals of the modules in the pack gradually drift away from their initials. Batteries aged in different SOC intervals can trigger different aging modes [48], reflected in dynamic performance, such as terminal voltage during operation. Batteries aged in similar SOC intervals tend to have similar aging modes, meaning similar dynamic performance consistency. Except for the absolute value of the terminal voltage, a vital feature to characterize the drifting SOC interval of the module is the electrochemical reactions corresponding to the IC curve. Via dynamic characteristic-based criteria that represent the "consequence" of the degradation: the phase transformation reactions number and voltage proportion during the pack-level testing, the modules are classified into groups with similar SOC intervals and potentially experience similar aging modes. Thus, by classifying the modules using our proposed dynamic characteristic-based criteria, good consistency in static and dynamic performance can be achieved.

4.3.2. Adaptability of the proposed approach

The proposed approach performed well in fast screening with pack-level testing data under 0.1C, prompting us to further increase the C-rate to 0.3C to verify its adaptability, as this C-rate is closer to practical applications of EVs and reduces the screening testing time. The module quantities in each class are the same for both 0.1C and 0.3C, respectively (Fig.8) and the modules in each class are also the same. The unchanged results suggest that the proposed approach is relatively robust. The quantified results for static (cap) and dynamic (OCV-SOC) performance are shown in Supplementary Table S1 and Fig.S4.

4.3.3. Applications and outlook

The proposed fast screening approach is promising for large-scale applications of retired LIBs. Compared with the approaches that are based on module-level testing, it is cost-, effort- and time-saving as there is no need to disassemble the pack into modules to test them separately to obtain corresponding screening criteria. A battery pack testing equipment containing auxiliary voltage measurements or the battery management system is enough to conduct the screening in this study, while it may take much longer to measure the screening criteria for approaches based on criteria that require module-level testing. Not to mention the labor and the cost. Furthermore, the adAP algorithm employed is unsupervised and does not need to train the model beforehand, further saving time from an algorithmic perspective. These characteristics of the proposed framework facilitate fast, accurate, stable screening and accelerate the progress of second use, contributing to the sustainability of renewable energy and the reliability of battery storage systems.

Although the proposed approach has been verified on one practical dataset, there are still several issues that could not be further investigated in this paper due to resource and time limitations. First, the adaptability of the proposed approach on battery packs with varying degrees of degradation and inconsistency. Second, research extending to other battery chemistries, such as nickel-manganese-cobalt and lithium-titanate-oxide, is needed in the future. In addition, the effect of measurement errors of different testing equipment in large-scale screening should be studied. Last but not least, future research is expected on how to regroup the retired LIBs to suit different application scenarios.

5. Conclusions

Aiming at accelerating the progress of retired lithium-ion batteries for the second use, a fast and

accurate screening approach based on pack-level testing is proposed for evaluating and classifying module-level aging. The main conclusions are summarized:

1) Static characteristic-based criterion alone does not necessarily guarantee good dynamic performance consistency of retired batteries in second use, and therefore it is essential to incorporate dynamic characteristic-based criteria during screening.
2) Our proposed dynamic characteristic-based criteria capture the comprehensive performance of retired modules, making the approach applicable for battery packs with up to 30% module-to-module SOC inconsistencies. Its adaptability to large C-rate shows promises for large-scale applications in second-use screening.
3) The efficiency of pack-level testing in screening criteria measurements is increased by at least 50% compared with approaches based on criteria that require module-level testing. Adaptive affinity propagation clustering, a pre-training-free algorithm, further accelerates the screening progress.
4) Our proposed approach can effectively divide the 95 modules into four classes with separated static and dynamic performances, achieving reasonable static performance consistency and better dynamic performance consistency as well as higher clustering stability, with average overall performance improvements of 18.94%, 4.83% and 34.41% compared with the other three benchmarks. Yet, there is still a large gap between research and industrial implementation. Hence our future work is planned to explore the regrouping strategies for retired batteries from different manufacturers to suit various application scenarios.

Acknowledgements

This work is supported by the National Natural Science Foundation of China [Grant No.51977007, Grant No.52007006], the Natural Science Foundation of Beijing under Grant No.3212033 and the research project "Model2Life" (03XP0334) funded by the German Federal Ministry of Education and Research (BMBF).

References

[1] Edge, J.S., O'Kane, S., Prosser, R., Kirkaldy, N.D., Patel, A.N., Hales, A., et al., Lithium ion

battery degradation: what you need to know. Phys Chem Chem Phys 2021; 23: 8200–8221. https://doi.org/10.1039/d1cp00359c.

[2] Duong, T.Q., USABC and PNGV test procedures. J Power Sources 2000; 244–248.

[3] Käbitz, S., Bernhard, J., Ecker, M., Yurdagel, Y., Emmermacher, B., André, D., et al., Cycle and calendar life study of a graphite|LiNi1/3Mn1/3Co1/3O2 Li-ion high energy system. Part A : Full cell characterization. J Power Sources 2013; 239: 572–583. https://doi.org/10.1016/j.jpowsour.2013.03.045.

[4] statista, Global second life battery capacity 2023-2030. 2021; https://www.statista.com/statistics/876624/global-second-life-battery-capacity/.

[5] International Energy Agency (IEA), Global EV Outlook 2022:securing supplies for an electric future. 2022; https://www.iea.org/reports/global-ev-outlook-2022.

[6] Zhu, J., Mathews, I., Ren, D., Li, W., Cogswell, D., Xing, B., et al., End-of-life or second-life options for retired electric vehicle batteries. Cell Reports Phys Sci 2021; 2: 100537. https://doi.org/10.1016/j.xcrp.2021.100537.

[7] Electric car sales this year resist Covid-19's blow to global car market. IeaOrg 2020; https://www.iea.org/news/electric-car-sales-this-year-resist-covid-19-s-blow-to-global-car-market (accessed April 20, 2022).

[8] Börner, M.F., Frieges, M.H., Späth, B., Spütz, K., Heimes, H.H., Sauer, D.U., et al., Challenges of second-life concepts for retired electric vehicle batteries. Cell Reports Phys Sci 2022; 101095. https://doi.org/10.1016/j.xcrp.2022.101095.

[9] Yu, W., Guo, Y., Shang, Z., Zhang, Y., Xu, S., A review on comprehensive recycling of spent power lithium-ion battery in China. ETransportation 2022; 11: 100155. https://doi.org/10.1016/j.etran.2022.100155.

[10] Shahjalal, M., Roy, P.K., Shams, T., Fly, A., Chowdhury, J.I., Ahmed, M.R., et al., A review on second-life of Li-ion batteries: prospects, challenges, and issues. Energy 2022; 241: 122881. https://doi.org/10.1016/j.energy.2021.122881.

[11] Muhammad, M., Ahmeid, M., Attidekou, P.S., Milojevic, Z., Lambert, S., Das, P., Assessment of spent EV batteries for second-life application. 2019 IEEE 4th Int Futur Energy Electron Conf IFEEC 2019 2019; https://doi.org/10.1109/IFEEC47410.2019.9015015.

[12] Hossain, E., Murtaugh, D., Mody, J., Faruque, H.M.R., Sunny, M.S.H., Mohammad, N., A

Comprehensive Review on Second-Life Batteries: Current State, Manufacturing Considerations, Applications, Impacts, Barriers Potential Solutions, Business Strategies, and Policies. IEEE Access 2019; 7: 73215–73252. https://doi.org/10.1109/ACCESS.2019.2917859.

[13] Zhang, Y., Li, Y., Tao, Y., Ye, J., Pan, A., Li, X., et al., Performance assessment of retired EV battery modules for echelon use. Energy 2020; 193: 116555. https://doi.org/10.1016/j.energy.2019.116555.

[14] Liao, Q., Mu, M., Zhao, S., Zhang, L., Jiang, T., Ye, J., et al., Performance assessment and classification of retired lithium ion battery from electric vehicles for energy storage. Int J Hydrogen Energy 2017; 42: 18817–18823. https://doi.org/10.1016/j.ijhydene.2017.06.043.

[15] Dubarry, M., Pastor-Fernández, C., Baure, G., Yu, T.F., Widanage, W.D., Marco, J., Battery energy storage system modeling: Investigation of intrinsic cell-to-cell variations. J Energy Storage 2019; 23: 19–28. https://doi.org/10.1016/j.est.2019.02.016.

[16] Li, S., Zhang, C., Du, J., Cong, X., Zhang, L., Jiang, Y., et al., Fault diagnosis for lithium-ion batteries in electric vehicles based on signal decomposition and two-dimensional feature clustering. Green Energy Intell Transp 2022; 1: 100009. https://doi.org/10.1016/j.geits.2022.100009.

[17] Reinhardt, R., Christodoulou, I., Gassó-Domingo, S., Amante García, B., Towards sustainable business models for electric vehicle battery second use: A critical review. J Environ Manage 2019; 245: 432–446. https://doi.org/10.1016/j.jenvman.2019.05.095.

[18] Lai, X., Huang, Y., Deng, C., Gu, H., Han, X., Zheng, Y., et al., Sorting, regrouping, and echelon utilization of the large-scale retired lithium batteries: A critical review. Renew Sustain Energy Rev 2021; 146: 111162. https://doi.org/10.1016/j.rser.2021.111162.

[19] Chen, H., Shen, J., A degradation-based sorting method for lithium-ion battery reuse. PLoS One 2017; 12: 1–15. https://doi.org/10.1371/journal.pone.0185922.

[20] Lyu, C., Song, Y., Wang, L., Li, J., Zhang, B., Liu, E., A new method for lithium-ion battery uniformity sorting based on internal criteria. J Energy Storage 2019; 25: 100885. https://doi.org/10.1016/j.est.2019.100885.

[21] Xu, Z., Wang, J., Lund, P.D., Fan, Q., Dong, T., Liang, Y., et al., A novel clustering algorithm for grouping and cascade utilization of retired Li-ion batteries. J Energy Storage 2020; 29:

101303. https://doi.org/10.1016/j.est.2020.101303.

[22] Schneider, E.L., Oliveira, C.T., Brito, R.M., Malfatti, C.F., Classification of discarded NiMH and Li-Ion batteries and reuse of the cells still in operational conditions in prototypes. J Power Sources 2014; 262: 1–9. https://doi.org/10.1016/j.jpowsour.2014.03.095.

[23] Li, R., Zhang, H., Li, W., Zhao, X., Zhou, Y., Toward group applications: A critical review of the classification strategies of lithium-ion batteries. World Electr Veh J 2020; 11: 1–23. https://doi.org/10.3390/wevj11030058.

[24] Salinas, F., Kowal, J., Classifying aged Li-ion Cells from notebook batteries. Sustain 2020; 12:. https://doi.org/10.3390/su12093620.

[25] Ananda, S., Lakshminarasamma, N., Radhakrishna, V., Pramod, M., Srinivasan, M.S., Sankaran, M., A cell sorting Algorithm for series-parallel configured Lithium ion battery packs in spacecrafts. 9th IEEE Int Conf Power Electron Drives Energy Syst PEDES 2020 2020; 10–14. https://doi.org/10.1109/PEDES49360.2020.9379500.

[26] Salinas, F., Krüger, L., Neupert, S., Kowal, J., A second life for li-ion cells rescued from notebook batteries. J Energy Storage 2019; 24: 100747. https://doi.org/10.1016/j.est.2019.04.021.

[27] Garg, A., Yun, L., Gao, L., Putungan, D.B., Development of recycling strategy for large stacked systems: Experimental and machine learning approach to form reuse battery packs for secondary applications. J Clean Prod 2020; 275: 124152. https://doi.org/10.1016/j.jclepro.2020.124152.

[28] Jiang, T., Sun, J., Wang, T., Tang, Y., Chen, S., Qiu, S., et al., Sorting and grouping optimization method for second-use batteries considering aging mechanism. J Energy Storage 2021; 44: 103264. https://doi.org/10.1016/j.est.2021.103264.

[29] Lai, X., Qiao, D., Zheng, Y., Ouyang, M., Han, X., Zhou, L., A rapid screening and regrouping approach based on neural networks for large-scale retired lithium-ion cells in second-use applications. J Clean Prod 2019; 213: 776–791. https://doi.org/10.1016/j.jclepro.2018.12.210.

[30] Enache, B.-A., Seritan, G.-C., Grigorescu, S.-D., Cepisca, C., Adochiei, F.-C., Argatu, V.-V., et al., A Battery Screening System for Second Life $LiFePO_4$ Batteries, in: 2020 Int. Conf. Expo. Electr. Power Eng., IEEE, 2020: pp. 298–301.

https://doi.org/10.1109/EPE50722.2020.9305538.

[31] Yan, N., Li, X., Zhao, H., Zhong, Y., Ma, S., Stratified Sorting Method of Battery Module Considering SOH in Echelon Utilization. IEEE Trans Appl Supercond 2021; 31: 1–4. https://doi.org/10.1109/TASC.2021.3107803.

[32] Yin, H., Li, Y., Kang, Y., Zhang, C., A two-stage sorting method combining static and dynamic characteristics for retired lithium-ion battery echelon utilization. J Energy Storage 2023; 64: 107178. https://doi.org/10.1016/j.est.2023.107178.

[33] Dubarry, M., Baure, G., Anseán, D., Perspective on State-of-Health Determination in Lithium-Ion Batteries. J Electrochem Energy Convers Storage 2020; 17: 1–8. https://doi.org/10.1115/1.4045008.

[34] Fan, L., Haihong, H., Haixin, W., A Fast Screening and Recombinant Method Based on Short-Time Pulse Discharge and Electrochemical Impedance Spectroscopy for Decommissioned Power Batteries. J Electrochem Energy Convers Storage 2022; 19:. https://doi.org/10.1115/1.4053866.

[35] Dubarry, M., Berecibar, M., Devie, A., Anseán, D., Omar, N., Villarreal, I., State of health battery estimator enabling degradation diagnosis: Model and algorithm description. J Power Sources 2017; 360: 59–69. https://doi.org/10.1016/j.jpowsour.2017.05.121.

[36] Yang, S., Zhang, C., Jiang, J., Zhang, W., Gao, Y., Zhang, L., A voltage reconstruction model based on partial charging curve for state-of-health estimation of lithium-ion batteries. J Energy Storage 2021; 35: 102271. https://doi.org/10.1016/j.est.2021.102271.

[37] Jiang, Y., Jiang, J., Zhang, C., Zhang, W., Gao, Y., Guo, Q., Recognition of battery aging variations for LiFePO4 batteries in 2nd use applications combining incremental capacity analysis and statistical approaches. J Power Sources 2017; 360: 180–188. https://doi.org/10.1016/j.jpowsour.2017.06.007.

[38] Zhou, Z., Duan, B., Kang, Y., Shang, Y., Cui, N., Chang, L., et al., An efficient screening method for retired lithium-ion batteries based on support vector machine. J Clean Prod 2020; 267: 121882. https://doi.org/10.1016/j.jclepro.2020.121882.

[39] Jiang, B., Dai, H., Wei, X., Incremental capacity analysis based adaptive capacity estimation for lithium-ion battery considering charging condition. Appl Energy 2020; 269: 115074. https://doi.org/10.1016/j.apenergy.2020.115074.

[40] Lu, X., Tarascon, J.-M., Huang, J., Perspective on commercializing smart sensing for batteries. ETransportation 2022; 14: 100207. https://doi.org/10.1016/j.etran.2022.100207.

[41] Zhang, Y., Zhou, Z., Yang, X., Gu, P., Zhang, C., Duan, B., A Novel Screening Approach Based on Neural Network for the Second Usage of Retired Lithiumion Batteries, in: 2020 Chinese Autom. Congr., IEEE, 2020: pp. 1193–1197. https://doi.org/10.1109/CAC51589.2020.9327171.

[42] Zhang, Y., Zhou, Z., Kang, Y., Zhang, C., Duan, B., A Quick Screening Approach Based on Fuzzy C-Means Algorithm for the Second Usage of Retired Lithium-Ion Batteries. IEEE Trans Transp Electrif 2021; 7: 474–484. https://doi.org/10.1109/TTE.2020.3032289.

[43] Lai, X., Deng, C., Li, J., Zhu, Z., Han, X., Zheng, Y., Rapid Sorting and Regrouping of Retired Lithium-Ion Battery Modules for Echelon Utilization Based on Partial Charging Curves. IEEE Trans Veh Technol 2021; 70: 1246–1254. https://doi.org/10.1109/TVT.2021.3055068.

[44] Yang, S., Zhang, C., Jiang, J., Zhang, W., Zhang, L., Wang, Y., Review on state-of-health of lithium-ion batteries: Characterizations, estimations and applications. J Clean Prod 2021; 314: 128015. https://doi.org/10.1016/j.jclepro.2021.128015.

[45] Frey, B.J., Dueck, D., Clustering by passing messages between data points. Science (80-) 2007; 315: 972–976. https://doi.org/10.1126/science.1136800.

[46] Wang, K.J., Zhang, J.Y., Li, D., Zhang, X.N., Guo, T., Adaptive affinity propagation clustering. Zidonghua Xuebao/Acta Autom Sin 2007; 33: 1242–1246. https://doi.org/10.1360/aas-007-1242.

[47] Kaijun Wang, Adaptive Affinity Propagation clustering. MATLAB Cent File Exch 2009; https://www.mathworks.com/matlabcentral/fileexchange/18244-adaptive-affinity-propagation-clustering (accessed April 22, 2022).

[48] Collath, N., Tepe, B., Englberger, S., Jossen, A., Hesse, H., Aging aware operation of lithium-ion battery energy storage systems: A review. J Energy Storage 2022; 55: 105634. https://doi.org/10.1016/j.est.2022.105634.