Performance Enhancement of Second-Life Lithium-ion Batteries Based on GMM Clustering and Simulation based Evaluation for Energy Storage System (ESS) Applications

Abdul Shakoor Akram and Woojin Choi *

Posted Date: 30 May 2025

doi: 10.20944/preprints202505.2428.v1

Keywords: lithium-ion battery; clustering; gaussian mixture model; repackaged module

 doi:10.20944/preprints202505.2428.v1

Article

Performance Enhancement of Second-Life Lithium-ion Batteries Based on Gaussian Mixture Model Clustering and Simulation based Evaluation for Energy Storage System Applications

Abdul Shakoor Akram and Woojin Choi *

School of Electrical Engineering, Soongsil University, Seoul, 06978, Republic of Korea

* Correspondence: cwj777@ssu.ac.kr

Abstract: Lithium-ion batteries (LIBs) are widely deployed in electric vehicles (EVs) due to their high energy density and long cycle life. Even after retirement typically at around 80% of their rated capacity LIBs can still be repurposed for second-life applications such as residential energy storage systems (ESS). However, effectively grouping these heterogeneous cells is crucial to optimize performance of the module. Retired LIBs can be effectively repurposed for numerous second-life applications such as energy storage systems (ESS), and other power backups. In this paper, we compare four clustering approaches including random grouping, equal-number Support Vector Clustering (SVC), K-means, and equal-number Gaussian Mixture Model (GMM) to organize 60 retired cells into 48 V modules consisting of 15-cell groups. We verify the performance of each method via simulations of a 15S2P configuration, focusing on standard deviation of final charge voltage, average charge throughput, delta capacity, and coulombic efficiency. Based on the evaluation metrics analyzed after regrouping the battery cells and simulating them for second-life ESS applications, the GMM based clustering method demonstrates better performance.

Keywords: lithium-ion battery; clustering; gaussian mixture model; repackaged module

1. Introduction

With the increasing global emphasis on decarbonization and sustainable energy solutions, lithium-ion batteries have emerged as the leading choice for energy storage in electric vehicles (EVs) due to their high energy density, long cycle life, and favorable performance characteristics [1]. As the EV market continues to grow, the worldwide production and deployment of Li-ion batteries are expected to surge, resulting in a parallel increase in the number of batteries reaching their end of life [2]. However, many of these batteries retain sufficient residual capacity and can be repurposed for less demanding applications rather than being prematurely discarded. This concept of second-life usage provides both economic and environmental benefits by extending battery lifespans and reducing the environmental impact associated with battery manufacturing and disposal.

Among various second-life applications, stationary Energy Storage Systems (ESS) have gained significant attention. Residential and commercial ESS solutions can buffer fluctuations in renewable energy sources and help to stabilize the power grid [3]. In particular, the demand for standardized battery module configurations is growing as it simplifies the integration process, lowers development costs, and ensures compatibility with existing battery management systems (BMS). One widely adopted standard for residential ESS is the 48V module configuration, which provides a safe and modular building block [4]. Despite its practical advantages, reassembling used cells into battery modules still faces challenges related to cell-to-cell variability, state of health (SOH) discrepancies, and capacity mismatches. These issues necessitate a robust and effective grouping or clustering strategy to ensure consistent module performance and prolonged operating life.

 doi:10.20944/preprints202505.2428.v1

In light of these challenges, there has been extensive research on methods to optimally regroup cells for second-life usage. Beyond simple heuristic or rule-based approaches, data-driven clustering and classification methods are increasingly being adopted to handle complex, multidimensional battery performance data [5]. For instance, Support Vector Clustering (SVC) leverages the kernel trick to capture non-linear relationships in the data, making it suitable for identifying tightly knit clusters even when the boundaries are not trivially separable. K-means, a classic and widely used partitioning algorithm, remains attractive due to its simplicity and computational efficiency [6]. However, K-means typically assumes spherical or near-spherical cluster shapes, which may not always align with battery performance data distributions [7]. Gaussian Mixture Models (GMM), on the other hand, offer a probabilistic framework that can model more flexible cluster geometries and allow for soft clustering [8], where data points can belong to multiple clusters with varying degrees of membership.

Despite their theoretical differences, there is a need to systematically compare these clustering methods under realistic operational scenarios, especially for second-life battery modules intended for residential ESS applications. In this study, we build a 48V module simulation platform designed to replicate real-world conditions and thoroughly evaluate SVC, K-means, and GMM in terms of their ability to form well-balanced groups of second-life Li-ion cells. By simulating the module's operational metrics such as standard deviation in final cell charging voltages, energy throughput, and delta capacity over time we demonstrate how different clustering strategies can significantly influence overall module performance and battery longevity. Our findings show that while all three methods have their merits, GMM-based clustering tends to yield more balanced cell groupings and better overall performance under a range of load scenarios.

2. Data Description and Preprocessing

The experimental data used in this work is derived from the MIT-Stanford Toyota Research Center battery dataset, originally measured by Severson et al. [9]. The dataset focuses on the APR18650M1A battery cells, manufactured by A123 Systems. Table I provides an overview of the key parameters and usage recommendations for the APR18650M1A battery. Notably, the testing was conducted at a constant temperature of 30°C, employing a one-step or two-step fast charging strategy. The charging protocol follows a constant-current–constant-voltage (CCCV) approach with an upper cutoff voltage of 3.6 V and a lower cutoff voltage of 2.0 V, as recommended by the manufacturer.

Table 1. This is a table. Tables should be placed in the main text near to the first time they are cited.

Parameter	Value
Nominal Capacity	1.1Ah
Nominal Voltage	3.3V
Recommended standard charge method	1.5 A to 3.6 V (CCCV)
Recommended charge and cut-off voltage at 25°	3.6 to 2.0V
Operating temperature range	-30 ~ 60 °C

For our study, we select 60 battery samples from the dataset and use the data of their first 400 charge–discharge cycles to validate the proposed clustering approach. All samples undergo 4C-discharge rates to ensure a uniform discharge process. The fundamental goal is to repurpose these retired or partially used cells for second-life applications, making it critical to quantify their current condition accurately and group them accordingly. From the available measurements, we focus on three key parameters [10]. Capacity, internal resistance (IR), and remaining useful life (RUL) as these factors play a dominant role in determining battery health, performance, and viability for second-life usage.

Battery capacity is a primary indicator of the SOH for lithium-ion batteries. The discharge capacity C_n of the nth cycle can be expressed by:

$$C_n = C_n^H - C_n^L \quad (1)$$

 doi:10.20944/preprints202505.2428.v1

where C_n^H denotes the total capacity after the battery is fully discharged, and C_n^L is the capacity at the start of discharge for cycle n. As the number of cycles increases, C_n typically declines, adversely affecting the battery's usable lifetime. Figure 1 illustrates how the discharge capacity decreases over the cycles.

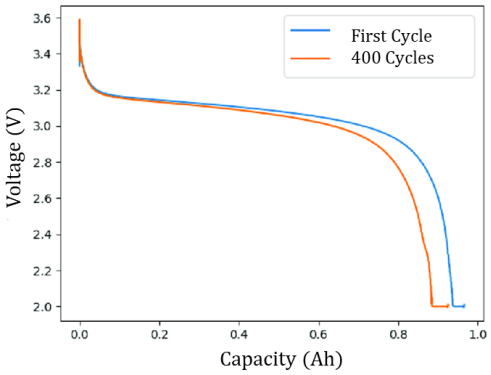


Figure 1. Relationship between capacity and voltage.

Internal resistance (IR) is another critical parameter that impacts the performance of battery and thermal characteristics. A higher IR leads to increased energy losses and faster degradation. When the battery starts discharging, there is an instantaneous voltage drop [11]. Let t be the moment the battery begins discharging, and U be the interpolated voltage at that instant. If t_1 and t_2 are two closely spaced sampling points with voltages U_1 and U_2, the ohmic internal resistance R_n during the nth discharge cycle can be approximated by:

$$R_n = \frac{U_n}{I_n} \quad where\ U_n = \left|\frac{U_2 - U_1}{t_2 - t_1}\right|_t \tag{2}$$

here, U_n is the effective voltage change caused by internal resistance, and I_n is the discharge current of the nth cycle. Figure 2 illustrates the rapid voltage-current fluctuations occurring during instantaneous discharge of the power lithium-ion batteries due to internal resistance.

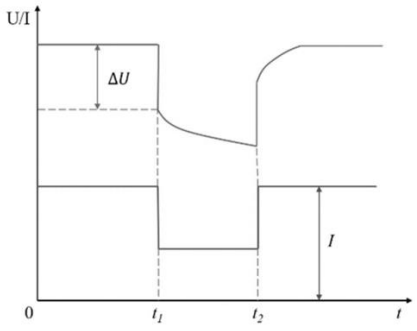


Figure 2. Concept of internal resistance measurement by discharge pulse.

RUL represents the projected service life or the number of additional cycles a battery can deliver before reaching a specified end-of-life threshold. In this dataset, we can calculate RUL directly from the measured capacity. When a battery's capacity decays to 80% of its rated capacity C_R, it is considered retired [12]. That is, the condition for retirement is:

$$C \leq 0.8C_R \tag{3}$$

where C is the actual capacity of the battery. Figure 5 shows the variation trend of battery capacity with the number of cycles.

 doi:10.20944/preprints202505.2428.v1

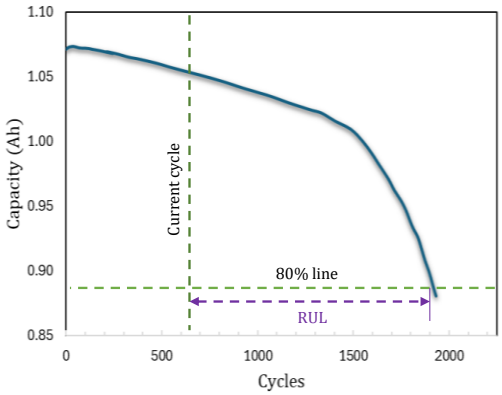


Figure 3. Battery capacity degradation and RUL approximation.

To ensure fair and consistent comparison among the three parameters (capacity, internal resistance, and RUL), we apply a min–max normalization scheme. This transformation maps each parameter to a uniform range of 0 to 1, thereby preventing any single parameter with larger numeric values from dominating the clustering algorithms. The min–max normalization for a given parameter x can be expressed as follows [13]:

$$x_{norm} = \frac{x - x_{min}}{x_{\max} - x_{min}} \tag{4}$$

where x_{min} and x_{max} are the minimum and maximum observed values of x in the dataset, respectively. By performing this step, the capacity, internal resistance, and RUL metrics all lie on a comparable scale, which is beneficial for methods such as SVC, K-means, and GMM.

3. Clustering Methods

After extracting and normalizing the key battery parameters like capacity, IR and RUL, we employ three different clustering algorithms to regroup second-life batteries: SVC, K-means, and GMM. This section offers an overview of each approach and their application to our standardized dataset.

3.1. Support Vector Clustering (SVC)

Support Vector Clustering (SVC) is a kernel-based approach that adapts the principles of Support Vector Machines (SVM) to unsupervised learning. In SVC, the normalized data points are mapped to a high-dimensional feature space using a kernel function, typically the radial basis function (RBF) [10], which is defined as

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2} \tag{5}$$

where γ is a hyperparameter controlling the width of kernel. In this feature space, SVC determines the smallest hypersphere that encloses the majority of the data points.

 doi:10.20944/preprints202505.2428.v1

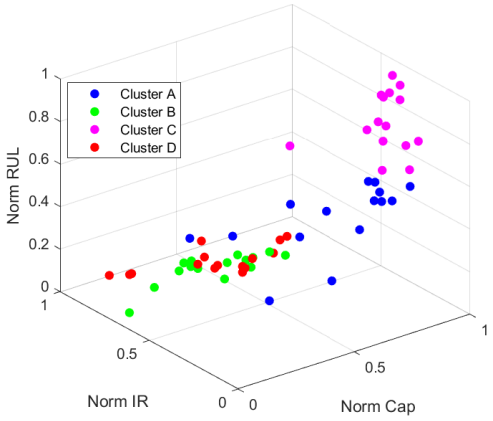


Figure 4. Support Vector Clustering Results.

Upon convergence, the projection of this hypersphere back into the original input space produces a decision boundary. A data point x can be tested by evaluating

$$f(x) = \|\Phi(x) - a\|^2 - R^2 \tag{6}$$

so that $f(x) < 0$ indicates x is inside the sphere, and $f(x) > 0$ signals it is an outlier or lies in another region. When different enclosed regions emerge in the input space, multiple clusters can be formed. In the context of second-life battery regrouping, equal number SVC applies an additional post-processing step to ensure each cluster contains a fixed number of cells. Although SVC excels at discovering arbitrarily shaped clusters, it demands careful parameter tuning.

3.2. K-Means

In contrast, the K-means algorithm partitions the dataset into a predetermined number of clusters k, by minimizing the within-cluster sum-of-squares. Which can be represented by

$$J(c,u) = \sum_{i=1}^{M} \|x - u_i\|^2 \tag{7}$$

where $\|x_i - u_i\|^2$ is the squared Euclidean distance between a data point x and the centroid u_i [14]. To improve the stability and convergence speed of the algorithm, we employ K-means++ initialization, which selects initial centroids in a way that maximizes their separation. This approach helps the clusters converge to a more optimal configuration than random initialization would typically allow.

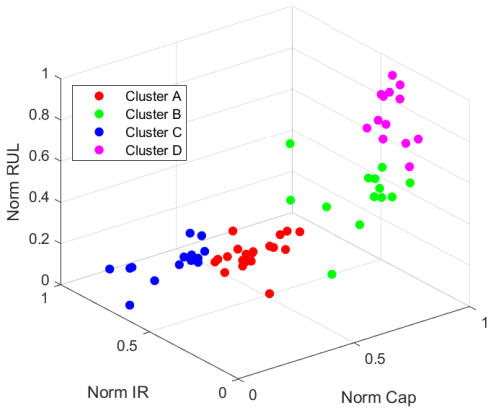


Figure 5. K-Means Clustering Results.

 doi:10.20944/preprints202505.2428.v1

After initialization, each data point is iteratively assigned to the nearest centroid [15]. The centroids are then recalculated as the mean of the assigned points, and this process is repeated until the assignments stabilize or a maximum iteration limit is reached. While K-means++ reduces the risk of poor initial cluster placements, the algorithm itself still assumes roughly spherical clusters and relies on Euclidean distance for defining cluster boundaries. Consequently, K-means struggles to handle overlapping or non-spherical distributions, as is often the case in second-life battery data, where cell parameters may not neatly conform to uniform cluster shapes.

3.3. Gaussian Mixture Model (GMM)

Gaussian Mixture Models (GMM) offer a probabilistic framework that models the data as a mixture of *K* Gaussian distributions. Each component in the mixture is characterized by a mean vector μ_k and a covariance matrix Σ_k, and a mixing coefficient π_k and the overall probability density function is given by [16]

$$p(x \mid \Theta) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x; \mu_k, \Sigma_k) \tag{8}$$

- $\pi_k \geq 0 \quad and \quad \sum_{k=1}^{k} \pi k = 1$
- $\mathcal{N}(x; \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} e^{exp\left(-\frac{1}{2}(x-U_1)^T \Sigma_k (x-\mu_k)\right)}$
- $\Theta = \{\pi_k, u_k, \Sigma_k\}_{k=1}^{K}$

where π_k represents the mixing coefficient for the *kth* component and Θ denotes the set of all model parameters.

A crucial step in fitting a GMM is choosing initial parameter values, since poor initialization can lead to suboptimal clustering or slow convergence in the EM algorithm. In our case, we initialize the GMM means u_k using K-means++ which was proposed by S. Vassilvitskii and D. Arthur in 2007 [17]. Specifically, we first use K-means++ to select *K* well-spaced centroids. These centroids then serve as the initial means u_k. The covariance matrices Σ_k can be initialized (for instance) as the identity matrix or estimated from the local neighborhoods of these initial means, while the mixing coefficients π_k are set to $\frac{1}{K}$ by default. This initialization strategy helps avoid local maxima often encountered by random initialization.

Once we have initial parameter estimates, the GMM parameters are iteratively refined using the Expectation-Maximization (EM) algorithm [18]. Each iteration of EM comprises two steps:

3.3.1. Expectation Step

Compute the posterior probability γ_{ik} that data point x_i belongs to cluster k:

$$\gamma_{ik} = p(z_k \mid x_i) = \frac{\pi_k N(x_i; \mu_k, \Sigma_k)}{\sum_{i=1}^{N} \pi_j N(x_i; \mu_k, \Sigma_k)} \tag{9}$$

where z_k denotes the event that x_i comes from the kth Gaussian component. The responsibilities γ_{ik} quantify soft cluster membership, allowing a single data point to have fractional membership in different clusters.

3.3.2. Maximization Step

Update the model parameters $\{\pi_k, u_k, \Sigma_k\}$ to maximize the expected complete-data log-likelihood. The standard formulas for the updates are:

 doi:10.20944/preprints202505.2428.v1

$$\left\{ \begin{array}{c} \pi_k^{new} = \frac{1}{N}\sum_{i=1}^{N}\gamma_{ik}, \quad \mu_k^{new} = \frac{\sum_{i=1}^{N}\gamma_{ik}x_i}{\sum_{i=1}^{N}\gamma_{ik}}, \\ \sum_k^{new} = \frac{\sum_{i=1}^{N}\gamma_{ik}(x_i-\mu_k^{new})(x_i-\mu_k^{new})^{\mathrm{T}}}{\sum_{i=1}^{N}\gamma_{ik}} \end{array} \right\} \tag{10}$$

where N is the total number of data points. Each update step re-estimates the mixing coefficients, means, and covariances based on the current responsibilities.

These two steps (E and M) alternate until convergence, typically when changes in the log-likelihood or parameter values fall below a chosen threshold.

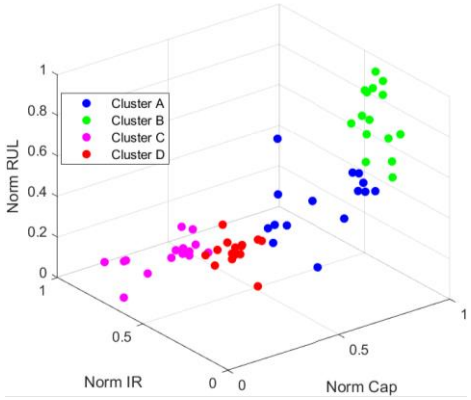


Figure 6. GMM Clustering Results.

Soft clustering is a defining characteristic of Gaussian Mixture Models (GMM) that proves especially valuable in second-life battery applications, where cells often exhibit overlapping performance characteristics. Unlike hard clustering methods, which assign each cell to exactly one cluster, a GMM outputs membership probabilities that indicate how strongly each cell belongs to each cluster. In our case, we leverage these soft memberships to form equal-sized modules, by sorting the cells of each cluster in descending order of membership probability and assigning them until the desired module size is reached. If any clusters become over- or underfilled, cells near the boundary of cluster membership are reassigned to maintain the overall equal distribution. This probabilistic approach ensures that each 15-cell module remains compositionally uniform while still respecting the practical requirement of equal-sized modules.

4. Comparative Analysis

4.1. Clustering Methods Evaluation

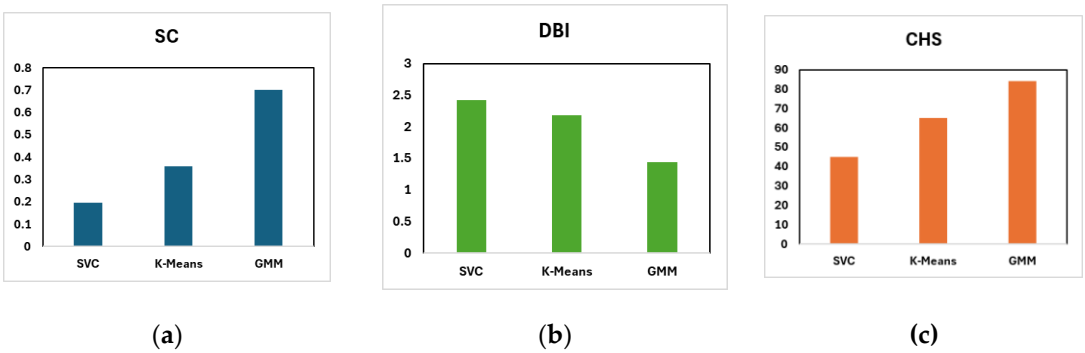
In our evaluation, we quantitatively compared the clustering performance of SVC, K-means, and GMM using three widely recognized indices: the Silhouette Coefficient (SC), Calinski-Harabasz Score (CHS), and Davies–Bouldin Score (DBI). The Silhouette Coefficient measures how similar an object is to its own cluster compared to other clusters, defined for each sample as

$$SC = \frac{b-a}{\max(a,b)} \tag{11}$$

where a is the average distance between a sample and all other points in the same cluster, and b is the average distance between a sample and the points in the nearest cluster. Values closer to 1 indicate well-separated clusters. The Calinski-Harabasz Score, calculated as the ratio of the between-cluster dispersion to the within-cluster dispersion, further validates the distinctness of clusters, the higher the CHS, the more distinct the clusters are [19]. In contrast, the Davies–Bouldin Score, which evaluates the average similarity between each cluster and its most similar one, rewards lower values that imply better clustering performance [20].

 doi:10.20944/preprints202505.2428.v1

Figure 7 shows a bar chart comparing SC, CHS, and DBI for SVC, K-means, and GMM. GMM consistently outperforms the other two methods across all three metrics. The ability of GMM to provide soft assignments helps it adapt to inherent variations in the battery data, allowing each cell to be assigned in a way that better balances the 15-cells-per-module requirement while accommodating natural differences in capacity, resistance, and RUL.



(**a**) (**b**) (**c**)

Figure 7. Comparison of evaluation indices for the clustering methods: (**a**) Silhouette Coefficient; (**b**) Davies-Bouldin Score; (**c**) Calinski-Harabasz Score.

In contrast, K-means, which strongly depends on the Euclidean distance to cluster centroids, can struggle with clusters that have non-spherical shapes or uneven data densities. SVC, particularly in its equal-number variant, does allow for more flexible boundaries compared to K-means, but the forced equal-cluster-size constraint means that if the natural distribution of battery parameters is strongly imbalanced, the clustering may not fully capture all nuances of the data. As a result, both SVC and K-means can exhibit slightly lower SC, lower CHS, and higher DBI scores than GMM.

4.2. Clustering Performance Evaluation

After determining appropriate cell clusters using SVC, K-means, or GMM, it is essential to verify that these groupings also perform well under practical operating conditions. A purely data-driven clustering approach can overlook system-level constraints such as voltage, current, and thermal behavior during charge–discharge cycles. Therefore, a MATLAB/Simulink-based simulation block diagram shown in Figure 8 is a valuable bridge between theoretical clustering outcomes and actual battery pack performance. By modeling the battery modules in software, we can accurately estimate how each cluster will behave as part of a 48V battery pack.

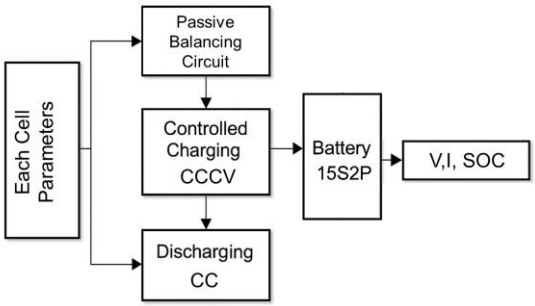


Figure 8. Simulation Block Diagram.

Residential ESS solutions commonly use a nominal 48 V battery pack design for reasons of safety, modularity, and compatibility. To achieve this nominal voltage using Li-ion cells (with each cell having a nominal voltage of 3.3 V), the typical solution is to connect around 15 cells in a series. In our case, a 15S chain ensures the battery pack can reach the required voltage range under normal conditions. While 15 cells in series fix the pack voltage, adding parallel branches increases the total

 doi:10.20944/preprints202505.2428.v1

capacity and extends the usable energy for the module. Depending on the energy needs, the number of parallel branches can be scaled up, allowing for a flexible approach to capacity sizing without changing the nominal voltage. Figure 9 shows the voltage, current and state of the charge waveforms of the 15S2P (30 cells) simulation for one of the clustering groups.

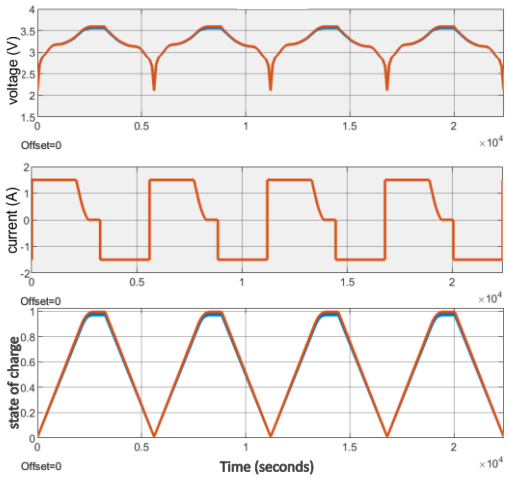


Figure 9. Voltage, current and SOC waveforms of 15s2p simulation for CCCV of regrouped cells.

The effectiveness of each clustering method was evaluated by simulating battery modules using the actual parameters of the clustered cells from the dataset, including internal resistance, actual capacity, SOC-OCV values, etc. Each module was constructed using 15 representative cells per cluster and configured as a 15S2P system. The modules were subjected to constant current (CC) charging until the voltage limit was reached, followed by constant voltage (CV) charging, and then constant current discharging.

The key performance indicators used for comparison were:

- The standard deviation of the final charging voltages measured across all 15 cells within the battery module, which indicates the degree of voltage imbalance among the cells at the end of the charging process.
- The average charge throughput of the 15 cells during the charging process, representing the total amount of charge each cell accepted on average, used as a measure of the overall charging performance and uniformity within the module.
- The difference between the maximum and minimum cell capacities within each cluster reflects the internal consistency of the clustering method in grouping cells with similar energy storage capabilities.
- The Coulombic efficiency of the battery module, calculated as the ratio of discharge capacity to charge capacity, used to evaluate how effectively the input electrical energy is converted into usable output energy.

These indicators compare the performance across modules formed via SVC, K-means, and GMM-based clustering. Each parameter was extracted from the simulation models to evaluate that the inconsistency of the LIBs group is reduced, and the performance of the groups has been improved.

In this simulation, a 15S2P battery pack underwent a 1.5A constant-current, constant-voltage (CCCV) charge. The final cell voltages within a group were then monitored, and the standard deviation of these final charge voltages was calculated to gauge the voltage spread across the cells in one group. As shown in Figure 10, comparing the random grouping, SVC, K-Means, and GMM methods, it is clear that the GMM-based module exhibits the lowest voltage spread. Specifically, methods B and D achieve standard deviations of 0.01320 and 0.01133, respectively.

 doi:10.20944/preprints202505.2428.v1

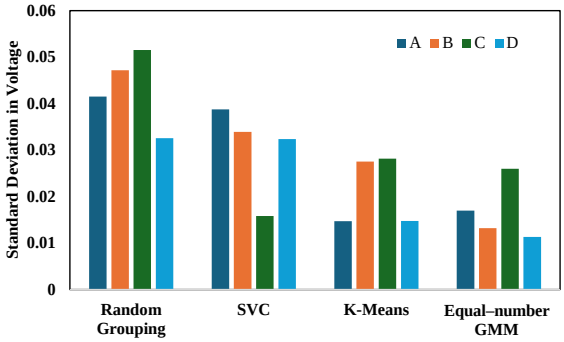


Figure 10. Standard deviation of final cell voltages of each module across different clustering methods.

In addition to analyzing the final voltage dispersion, we also examined the average charge throughput can be seen in the bar graph on Figure 11. As all cells in series string share the same current, the throughput primarily reflects how uniformly the cells accept charge over the duration of the charging period. In our simulation, we observed that modules clustered via GMM exhibit more consistent charge throughput, indicating fewer outlier cells that could reduce the overall capacity utilization as group A and B have an average charge throughput higher than 1 Ah. Conversely, random grouping or other clustering approaches tend to include cells with disproportionately lower or higher acceptance rates, which ultimately lowers the group's average throughput. By aligning cells of similar condition, GMM ensures identical charging behavior, thereby optimizing the total charge delivered to the battery pack and enhancing the reliability of second-life energy storage modules.

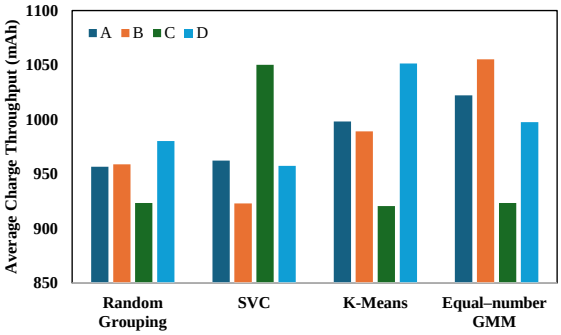


Figure 11. Average charge throughput of each module across different clustering methods.

Another key metric in analysis is the delta capacity in Figure 12, which is the difference between the highest-capacity cell and the lowest-capacity cell within each group. A smaller delta capacity indicates that the cells are more closely matched, thereby minimizing the risk of overcharging or over-discharging the weaker cells during operation. The GMM-based clusters showed notably lower delta capacity values compared to both random groupings and two other clustering methods.

 doi:10.20944/preprints202505.2428.v1

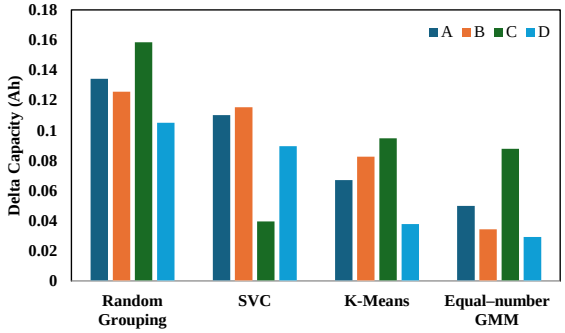


Figure 12. Delta capacity of each module across different clustering methods.

Finally, we evaluated coulombic efficiency, which measures how effectively each module converts the charged energy into usable discharge energy. The 15S2P configurations derived from GMM-based clustering demonstrated consistently higher coulombic efficiency of 98.18% relative to other methods as shown in Figure 13, underscoring more uniform energy utilization and reduced internal losses. Cells with similar capacity and internal resistance avoid mismatch driven inefficiencies that often occur in packs assembled through less adaptive clustering approaches.

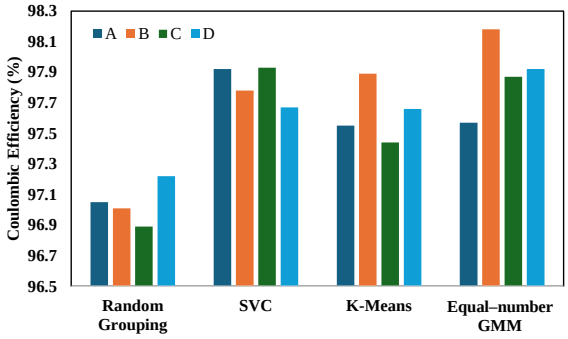


Figure 13. Coulombic efficiency of each module across different clustering methods.

Overall, these findings demonstrate that the GMM-based clustering approach provides more balanced and efficient battery modules than random grouping, SVC, or K-means. The minimized voltage spread indicates tighter cell matching, while the higher average charge-throughput reflects a more uniform charge acceptance across the entire module. In tandem, the reduced delta capacity confirms that GMM consistently groups cells of similar health, mitigating the risk of over-stressing weaker cells. Finally, the notable coulombic efficiency gain underscores the method's ability to limit internal losses and capitalize on the available capacity. Taken together, these improvements validate GMM as a robust clustering strategy for assembling second-life lithium-ion cells into reliable and long-lasting battery packs.

5. Conclusion

In conclusion, this work demonstrates that effectively grouping second-life lithium-ion cells into standardized 48 V modules can significantly enhance their performance in residential ESS applications. By comparing random grouping, equal-number SVC, K-means, and a K-means++-initialized equal-number GMM, we reveal that the soft clustering paradigm of GMM provides the most balanced clusters. Simulation results from a 15S2P configuration confirm reduced voltage deviations, higher average charge throughput, lower delta capacity, and improved coulombic efficiency for GMM-based modules. The K-means++ initialization assists in mitigating local optima, while the post-processing step ensures each group contains an equal number of cells. These findings

 doi:10.20944/preprints202505.2428.v1

underscore the potential of GMM to accommodate subtle variations in retired EV batteries, extending their useful life and stability in second-life applications. Future work could involve expanding the proposed approach to larger-scale battery packs and validating the technique in field demonstrations to further establish its robustness and versatility.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Xu, J., Cai, X., Cai, S., Shao, Y., Hu, C., Lu, S., & Ding, S. (2023). High-energy lithium-ion batteries: recent progress and a promising future in applications. *Energy & Environmental Materials*, *6*(5), e12450, https://doi.org/10.1002/eem2.12450.
2. Jena, A., Borra, V. L., Saida, S., Venkatesan, P., Önal, M. A. R., & Borra, C. R. (2025). Synergistic hydrometallurgical recycling of Li-ion battery cathode active material, anode copper and waste SmCo magnets. *Sustainable Materials and Technologies*, e01284, doi: https://doi.org/10.1016/j.susmat.2025.e01284.
3. Sarker, M. T., Haram, M. H. S. M., Shern, S. J., Ramasamy, G., & Al Farid, F. (2024). Second-life electric vehicle batteries for home photovoltaic systems: transforming energy storage and sustainability. *Energies*, *17*(10), 2345, https://doi.org/10.3390/en17102345
4. D. Santos et al., "ESS Design and Management considering Solar PV to fed off-grid EV Charger," *2024 12th International Conference on Smart Grid (icSmartGrid)*, Setubal, Portugal, 2024, pp. 636-641, https://doi.org/10.1109/icSmartGrid61824.2024.10578284.
5. Shahjalal, M., Roy, P. K., Shams, T., Fly, A., Chowdhury, J. I., Ahmed, M. R., & Liu, K. (2022). A review on second life of Li-ion batteries: Prospects, challenges, and issues. *Energy*, *241*, 122881. https://doi.org/10.1016/j.energy.2021.122881
6. Li, W., Chen, S., Peng, X., Xiao, M., Gao, L., Garg, A., & Bao, N. (2019). A comprehensive approach for the clustering of similar-performance cells for the design of a lithium-ion battery module for electric vehicles. *Engineering*, *5*(4), 795-802, https://doi.org/10.1016/j.eng.2019.07.005.
7. Wu, S., & Chow, T. W. (2004). Clustering of the self-organizing map using a clustering validity index based on inter-cluster and intra-cluster density. *Pattern Recognition*, *37*(2), 175-188. https://doi.org/10.1016/S0031-3203(03)00237-1
8. X. He, D. Cai, Y. Shao, H. Bao and J. Han, "Laplacian Regularized Gaussian Mixture Model for Data Clustering," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 9, pp. 1406-1418, Sept. 2011, https://doi.org/10.1109/TKDE.2010.259.
9. Severson, K. A., Attia, P. M., Jin, N., Perkins, N., Jiang, B., Yang, Z., ... & Braatz, R. D. (2019). Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, *4*(5), 383-391, https://doi.org/10.1038/s41560-019-0356-8
10. C. Li, N. Wang, W. Li, Y. Li and J. Zhang, "Regrouping and Echelon Utilization of Retired Lithium-Ion Batteries Based on a Novel Support Vector Clustering Approach," in *IEEE Transactions on Transportation Electrification*, vol. 8, no. 3, pp. 3648-3658, Sept. 2022, https://doi.org/10.1109/TTE.2022.3169208
11. Zhao, S., Wu, F., Yang, L., Gao, L., & Burke, A. F. (2010). A measurement method for determination of dc internal resistance of batteries and supercapacitors. *Electrochemistry Communications*, *12*(2), 242-245, https://doi.org/10.1016/j.elecom.2009.12.004
12. "How to Measure the Remaining Useful Life of a Battery," Battery University. [Online]. Available: https://batteryuniversity.com/article/bu-901b-how-to-measure-the-remaining-useful-life-of-a-battery. [Accessed: Mar. 17, 2025].
13. Dalatu, P. I., & Midi, H. (2020). New approaches to normalization techniques to enhance K-means clustering algorithm. *Malaysian Journal of Mathematical Sciences*, *14*(1), 41-62.
14. Yuan, C., & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. *J*, *2*(2), 226-235. https://doi.org/10.3390/j2020016
15. Song, R., Pang, F., Jiang, H., & Zhu, H. (2024). A machine learning based method for constructing group profiles of university students. *Heliyon*, *10*(7). https://doi.org/10.1016/j.heliyon.2024.e29181

 doi:10.20944/preprints202505.2428.v1

16. "Gaussian Mixture Model – Understanding and Implementation," Data Flair, [Online]. Available: https://data-flair.training/blogs/gaussian-mixture-model/. [Accessed: Mar. 17, 2025].
17. Vassilvitskii, S., & Arthur, D. (2006, January). k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (pp. 1027-1035).
18. Zhang, T., Kuo, CC.J. (2001). Sound Effects Classification and Retrieval. In: Content-Based Audio Classification and Retrieval for Audiovisual Data Parsing. The Springer International Series in Engineering and Computer Science, vol 606. Springer, Boston, MA. https://doi.org/10.1007/978-1-4757-3339-6_5
19. Falah, N., Falah, N., Solis-Guzman, J., & Marrero, M. (2025). An Indicator-Based Framework of Circular Cities Focused on Sustainability Dimensions and Sustainable Development Goal 11 Obtained Using Machine Learning and Text Analytics. *Sustainable Cities and Society*, 106219, https://doi.org/10.1016/j.scs.2025.106219\
20. Sasithradevi, A., Perumal, D. A., & Persiya, J. (2024). Infrared Perspectives: Computing laptop energy dissipation via thermal imaging and the Stefan-Boltzmann equation. *Thermal Science and Engineering Progress*, *53*, 102742, https://doi.org/10.1016/j.tsep.2024.102742