AUT Journal of Electrical Engineering
10.22060/EEJ.2026.24892.5788

New Approach to Assembling Used 18650 Cells by the DBSCAN Clustering Algorithm

Mohamed Redha Rezoug*, Laid Khettache, Abdeslam Benmakhlouf, Djalal Djarah

Department of Electrical Engineering, Kasdi Merbah University – Ouargla 30000, Algeria
*Corresponding author's email: mr.rezoug@univ-ouargla.dz

Abstract:

This study presents a diagnostic and regrouping approach for used 18650 lithium-ion cells using the DBSCAN clustering algorithm. A total of 154 cells were recovered from discarded laptop batteries. After electrical testing (constant-current discharge measuring capacity, voltage, and internal resistance), 120 cells (78%) were identified as healthy and suitable for reuse. The DBSCAN algorithm (eps=0.3, min_samples=10) was applied to cluster these 120 cells based on their electrical characteristics. The algorithm successfully formed three homogeneous clusters of 40 cells each. These clusters were assembled in a 3S40P configuration (three parallel packs of 40 cells connected in series) to form a second-life battery of approximately 81 Ah capacity at 10.9 V. Comparative analysis shows that DBSCAN outperforms K-means and hierarchical clustering for this application, achieving a silhouette coefficient of 0.62 versus 0.48 for K-means. The proposed method achieves a 78% cell recovery rate, demonstrating its effectiveness for battery recycling and second-life applications.

.

Keywords:

DBSCAN, Reconditioning, Recycling, lithium battery, 18650 cell, Clustering

AUT Journal of Electrical Engineering
10.22060/EEJ.2026.24892.5788

1. Introduction

Due to their widespread use in various fields, lithium batteries have become particularly noteworthy in diverse domains where the 18650 cell serves as the basic unit of construction [1]. It is essential to produce a significant quantity of 18650 cells to meet the growing demand for energy storage and portable power solutions [2, 3]. The daily production of these cells is carried out by manufacturing plants, primarily in Asia, which employs automated processes and advanced technologies to ensure optimal quality and efficiency, as mentioned in this reference [4]. Significant investments have been made in research and development to enhance the capacity, safety, and lifespan of 18650 cells, making them a crucial component of the global energy transition towards more sustainable and environmentally friendly solutions.

The expenses related to the production of these 18650 lithium cells are the result of various interdependent factors. Materials such as Lithium Cobalt Oxide (LCO) for the cathode, graphite, an allotrope of carbon for the anode, electrolytes, and microporous separators are the main components [5]. The manufacturing steps include gathering, coating, drying, and cutting of electrodes, as well as cell assembly, rolling, and electrolyte filling, as stated by the authors of this article [6]. The costs are consequently increased by the formation and aging of the cells, which require charge and discharge cycles.

On the other hand, expenses related to infrastructure, equipment, and their maintenance are considerable. It is crucial to have a skilled workforce, including training costs, as well as investments in research and development to promote innovation and compliance with international standards, as mentioned in [7]. Finally, economies of scale have a significant impact on unit costs, which allows for reducing expenses by increasing production volumes. Consequently, the costs are related to the production of 18650 lithium cells typically range between 100 and 300 USD per kilowatt-hour (kWh) [8], depending on raw material prices and technological advancements.

10.22060/EEJ.2026.24892.5788

In addition to all these expenses, this type of industry leads to significant end-of-life waste production. This waste, mainly composed of heavy metals and toxic chemical substances, poses major environmental challenges. For instance, lithium-ion batteries, commonly used today, contain lithium, cobalt, nickel, and other valuable metals that, if not properly recycled, can lead to soil and groundwater contamination [9]. Furthermore, the increasing demand for these batteries has proportionally increased the amount of waste produced, highlighting the need to develop more efficient and sustainable recycling technologies.

Research gap and novelty: While previous works [3, 10, 11] have applied clustering techniques to battery data, and DBSCAN was originally introduced in [12], the specific contribution of this paper is threefold: the specific contribution of this paper is threefold: (1) we propose a complete electrical recycling device for 18650 cell diagnostics, (2) we apply DBSCAN specifically to the problem of regrouping used cells into homogeneous packs for second-life batteries, and (3) we provide a practical validation on 154 real recovered cells with a detailed comparison of clustering algorithms. Unlike prior studies that rely on simulated data, our work uses experimentally measured capacity, voltage, and internal resistance from actual recycled cells. This constitutes the main novelty: a complete, experimentally validated methodology from cell recovery to final pack assembly.

This work was organized into three distinct stages. Firstly, a simple concept to correct is that when a lithium battery is defective, it does not mean that all the 18650 cells that compose it are also damaged. In reality, lithium batteries are often composed of several 18650 cells connected in series or parallel to ensure the desired voltage and capacity. Therefore, an electrical recycling device must be implemented to minimize costs and efforts as well as the negative impact on the environment. Provide an effective model to follow in order to reduce manufacturing costs by using damaged batteries. These are typically recycled efficiently using an electronic device that studies the actual state of each lithium cell and determines its capacity according to the reconditioning approach.

Detecting the actual state of each cell allows for sorting into two groups: the first represents completely deteriorated cells, while the second is the group of healthy cells that will be used to form new packs to meet the demand for different and efficient batteries. These healthy cells will be sorted according to a new and efficient arithmetic to benefit each from its properties. The final step is to connect and classify these cells using a simple matrix that links any cell with the rest of the other cells, either in series or in parallel according to the DBSCAN algorithm. As a researcher, the primary objective is to fully harness the maximum power of any 18650 lithium cell and extend its lifespan by identifying the best group in which it is considered the basic unit composing the overall battery. This is achieved through electronic recycling instead of using chemical methods, thereby preserving natural resources.

2. Methodical Description of the Approach

The main objective of this study is to employ a specific clustering method after recycling 18650 lithium-ion cells in a research space to optimize the management and creation of well-standardized new packs. These cells are recovered from various sources such as laptop batteries, power banks, and electric bike batteries. The general schematic diagram provided in the document illustrates and summarizes the complex process of recycling and reconditioning the cells using advanced algorithms to maximize the efficiency and lifespan of the recovered cells. Each recovered 18650 cell represents the basic unit of the process.

2.1 Experimental Setup and Measurement System

The logic behind this development is to assemble several sub-blocks that differ in their function but share a common factor, which is to observe and measure certain parameters. The most essential phase is to collect information on current, voltage, and resistance regarding the 18650 cell during the operation process. This information must be saved in real-time as a database on the SD card in an Excel file. As a researcher, the core of this achievement features an electrical block designed with a mini embedded system, the “Arduino board” which works in conjunction with a “voltage and current sensor” module. The Arduino measures the

10.22060/EEJ.2026.24892.5788

current consumed and dissipated in the form of temperature in a 10 Watt power resistor on one hand, and on the other hand, it measures the voltage across the terminals of the 18650 cell. The “MAX471” voltage and current sensor is ideal for this type of application due to its characteristics.

Experimental conditions: Discharge current = 0.5 A (constant current), cutoff voltage = 3.0 V, temperature = 25°C ± 2°C. Each cell underwent one full charge-discharge cycle. Internal resistance was measured using the voltage drop method at 1 kHz AC. Sampling frequency = 1 Hz. Measurement uncertainty: capacity ±3%, voltage ±0.5%, resistance ±5%. Each cell was tested twice, and the average values were used. Repeatability analysis on 10 randomly selected cells showed a coefficient of variation <4% for capacity measurements and <5% for resistance measurements.

The presence of the MAX471 module is essential in systems where monitoring the battery of the DC power line is considered crucial. Monitoring the high-side power line is particularly useful in battery-powered systems as it does not interfere with the ground paths of battery chargers or monitors often found in "smart" batteries. The MAX471 has an internal current sensing resistor of 35 mΩ and measures battery currents up to ±3 Amperes. For applications requiring higher current or increased flexibility in our case, and to avoid the operating limits of this module, all cells used in this application do not exceed a capacity of 2800mAh.

For measuring voltage, the test principle is simple and is based on the resistive voltage divider principle that reduces the voltage at the terminal interface input by five times. The Arduino analog input voltage level is 5 Volts, so the input voltage of the voltage detection module cannot exceed 5 Volts × 5 = 25 Volts. The MAX471 module allows measuring the current of a circuit by delivering an analog value that we can send to the Arduino board connected to the analog input A_0 and A_1 appropriately. For each ampere measured, the board will produce one Volt (ratio 1 V/A, maximum of 3 A).

2.2 Data Collection and Preprocessing

10.22060/EEJ.2026.24892.5788

In this experiment, approximately 25 used laptop batteries were collected. The number of cells recovered varies from one battery to another depending on the manufacturer. A total of 154 cells were recovered, of which nearly 120 tested positive, representing a success rate of 78%. This ratio varies from one model to another.

The nominal voltage of a Lithium-ion 18650 cell is 3.6 or 3.7 Volts, depending on the technology, as provided by the manufacturer. The charging voltage for each cell should never exceed 4.2 Volts (exceeding this voltage not only has severe consequences for this type of cell but can also be very dangerous). The minimum supported voltage is practically 3 Volts. Technically, any voltage below 2.5 Volts leads to significant degradation of the cell. Therefore, a minimum threshold of 3 Volts should always be maintained, below which the cell may experience stress situations that will quickly degrade its performance.

For the data backup block, this application involves storing collected information as a database on an external memory card to be processed later by another algorithm in Python. For this purpose, we selected an SD adapter module for the Arduino board. The micro SD module used aims to store all the measurements in an Excel file. This module is a Micro SD (TF) card reader, equipped with an SPI interface and a 5 Volts power supply compatible with Arduino, allowing the addition of storage space to the Arduino board's microcontroller. It is connected to the digital inputs 10, 11, 12, and 13 of an Arduino board.

2.3 Clustering Algorithm and Parameter Selection

Once 154 cells are sorted, the healthy cells are subjected to the application of an algorithm that performs common tasks in unsupervised machine learning [13]. The DBSCAN used is a clustering algorithm to identify the best groups or clusters of cells with similar characteristics for this case [12, 14]. This algorithm creates clusters based on the density of data points by identifying high-density areas or regions separated by low-density areas, which is particularly useful for handling noisy or non-linear data [15, 16]. By applying DBSCAN, the good cells are grouped based on their relationships with each other to ensure maximum

10.22060/EEJ.2026.24892.5788

homogeneity within each formed group. The formed clusters represent groups of cells with identical electrical characteristics such as capacity, voltage, and internal resistance.

After removing noise points (cells identified as outliers by DBSCAN), the healthy cells are grouped into three homogeneous clusters (Clusters 1, 2, and 3), each containing 40 cells. These clusters correspond to the three parallel packs (40P) that will be assembled in series (3S).

Each cluster is optimized for uniform performance, which is crucial for the stability and efficiency of the resulting battery. Each cell belonging to one of these homogeneous packs is then integrated into a switch matrix where it can be dynamically controlled and connected in series or parallel with others by activating or deactivating a transistor [17]. This matrix allows for maximum flexibility by enabling the connection of any lithium cell with the others, thereby optimizing the overall configuration of the battery. The switch matrix plays a key role in dynamic reconfiguration, ensuring that the cells always operate under optimal conditions [18]. The operating cycle of the optimized battery thus formed is continuously monitored by a switch control system.

Fig. 1, provides an overview of the recycling and grouping process of 18650 lithium-ion cells. Fig. 2 presents the overall hardware architecture.

10.22060/EEJ.2026.24892.5788

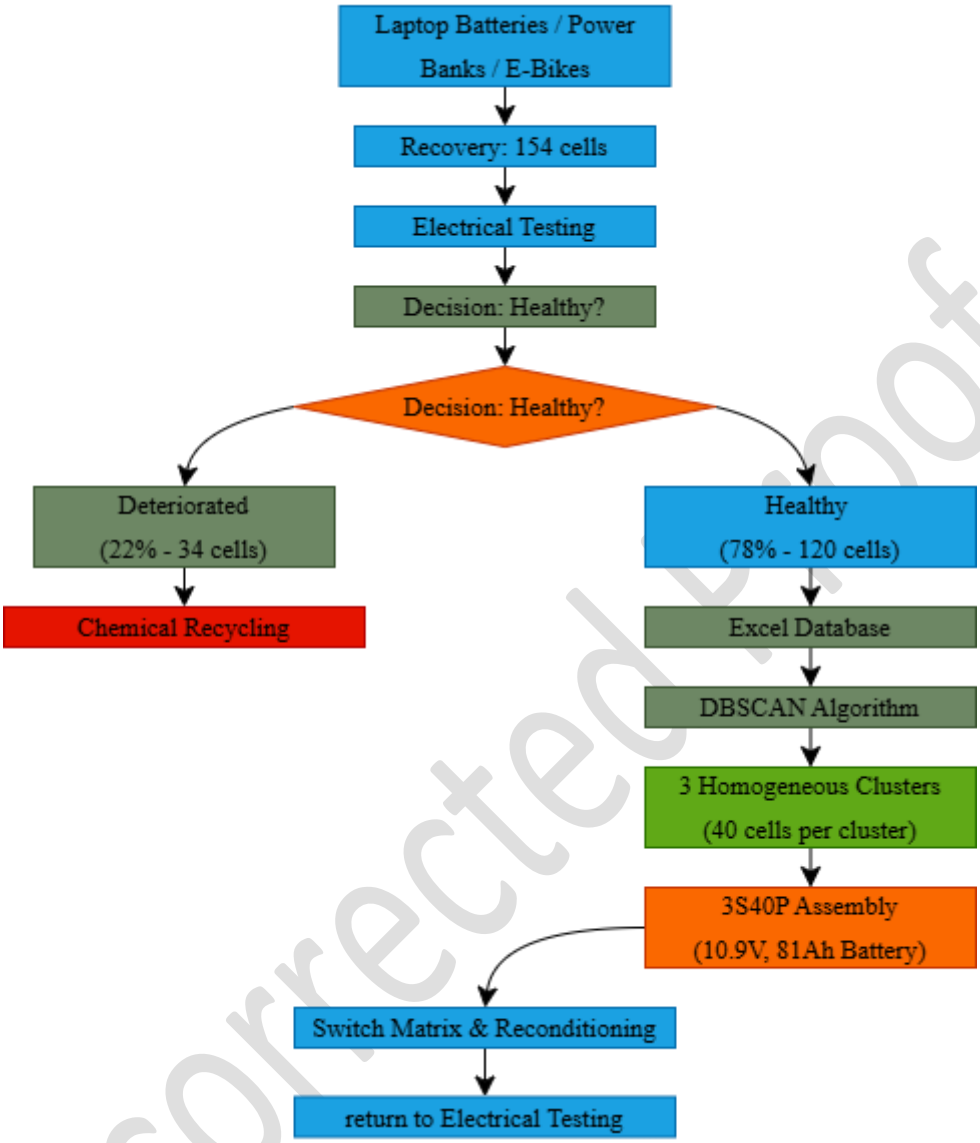


Fig. 1. General block diagram

10.22060/EEJ.2026.24892.5788

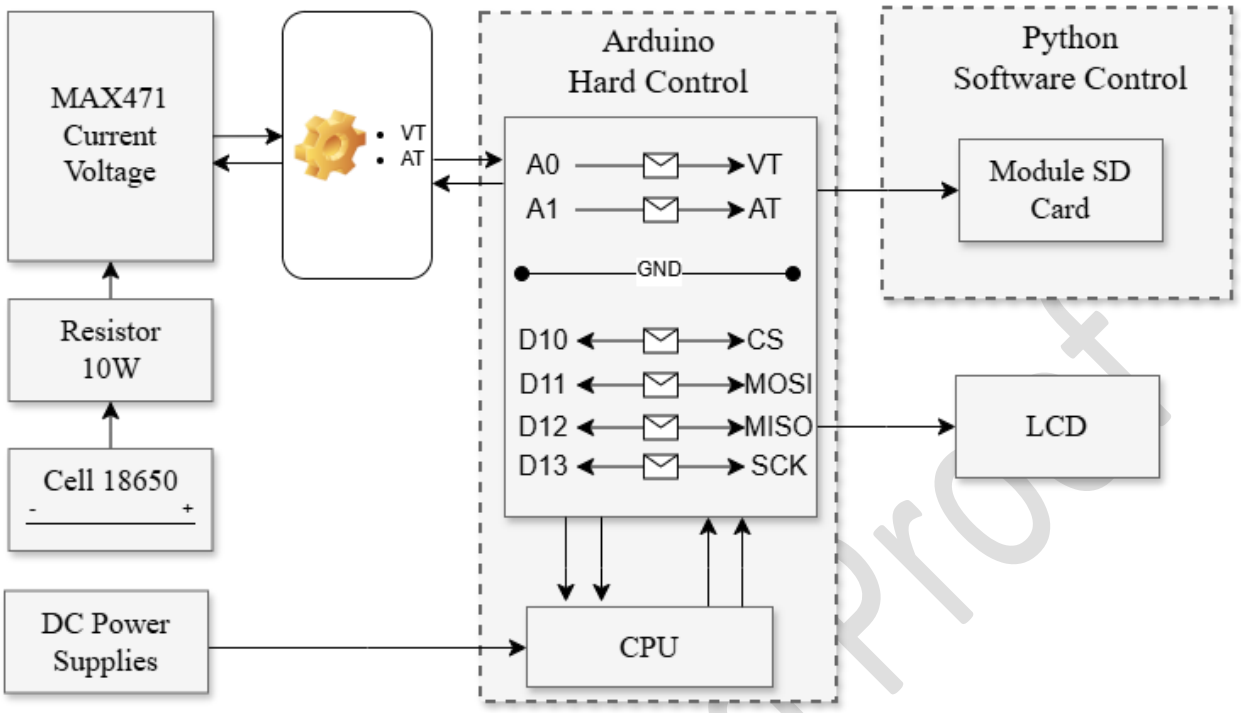


Fig. 2. Overall Presentation of the Developed Device

3. Unit Diagnostic

Knowing the capacity of a battery is essential in many situations. Meanwhile, 18650 cells operate according to variable charge and discharge cycles that are not stable over time, making it necessary to measure the actual capacity experimentally rather than relying on nominal values.

The application of the reorganization approach to the diagnosis of 18650 cells aims primarily to improve and enhance diagnostic accuracy by using advanced technologies to measure critical parameters such as voltage, current, and internal resistance [10]. Secondly, it seeks to reduce maintenance costs by implementing advanced detection systems for potential failures. Finally, the approach aims to extend the lifespan of these cells by optimizing the charge and discharge cycles and monitoring the state of the cells in real-time, which will be achieved with the DBSCAN algorithm [19].

10.22060/EEJ.2026.24892.5788

3.1 Capacity Measurement Method

The capacity of a cell is determined by integrating the discharge current over time until the cutoff voltage is reached:

$$C = \int I(t)dt \quad from \quad t = 0 \quad to \quad t\_end \tag{1}$$

where:

C = measured capacity (mAh or Ah), I(t) = discharge current at time t (A) and t_end = time when voltage reaches 3.0 V cutoff (hours).

For constant-current discharge, this simplifies to:

$$C = I\_discharge \times t\_discharge \tag{2}$$

3.2 Internal Resistance Measurement

The internal resistance R_int is calculated using the voltage drop method:

$$R\_\text{int} = \Delta V / I\_ac \tag{3}$$

where:

ΔV = voltage drop measured across the cell terminals (V) and I_ac = alternating current at 1 kHz (A).

All measurements were performed at 50% state of charge and 25°C ± 2°C to ensure consistency.

3.3 Experimental Procedure

The diagnostic device consists of several sub-blocks that differ in their function but share a common goal: observing and measuring cell parameters. The core of this achievement features an electrical block designed with a mini embedded system, the “Arduino board”, which works in conjunction with a voltage and current

10.22060/EEJ.2026.24892.5788

sensor module. The Arduino measures the current consumed and dissipated as heat in a 10 Watt power resistor, and simultaneously measures the voltage across the terminals of the 18650 cell. The MAX471 voltage and current sensor is ideal for this type of application.

To make the module work, the VIN pin is connected to the positive power line (the positive terminal of the 18650 cell) and our VOUT is connected to the load (the 10 Watts power resistor). The other end of the MAX471 contains four pins (two GND and two pins VT and AT) where the (VT) where:

- The VT pin carries the voltage measurement information linked to the Arduino board's analog pin A_0.
- The AT pin carries the discharge current measurement information, connected to the Arduino via pin A_1.
- One of the two GND pins must be connected to the Arduino board's GND.

For data backup, this application involves storing collected information as a database on an external memory card to be processed later by Python algorithms. A micro SD module is used, connected to the digital inputs 10, 11, 12, and 13 of the Arduino board.

3.4 Results of Cell Diagnosis

A total of 25 used laptop batteries were collected for this study. The number of cells recovered varies from one battery to another depending on the manufacturer. Table 1 summarizes the recovery results:

Table 1: Cell recovery and diagnostic results

Parameter	Value
Total batteries collected	25
Total cells recovered	154
Healthy cells identified	120
Deteriorated cells	34
Success rate (healthy cells)	78%

For the 120 healthy cells, the average characteristics were:

- Mean capacity: 2015 mAh (range: 1760 mAh to 2621 mAh)
- Mean nominal voltage: 3.65 V (range: 3.6 V to 3.7 V)
- Mean internal resistance: 35.2 mΩ (range: 20 mΩ to 57 mΩ)

It is essential that the packs made are as balanced as possible in terms of capacity. Therefore, the most effective method to assemble a battery of three packs containing 40 cells in a 3S40P configuration is presented in the following section.

4. Application of DBSCAN and Interpretation

After collecting 120 lithium-ion 18650 cells and diagnosing them with the recycling device, the final step consists of assembling three homogeneous groups of cells. Each group represents a pack of 40 cells connected in parallel. When these three packs are assembled in series, an 11.1 V battery is obtained with a capacity of about 80 Ah, according to the values previously measured and distributed using the DBSCAN algorithm.

Capacity calculation: Average cell capacity from experimental data = 2015 mAh. For 40 cells in parallel: 2015 mAh × 40 = 80,600 mAh ≈ 80 Ah. The series configuration (3S) multiplies voltage while maintaining capacity: 3 × 3.65 V = 10.95 V ≈ 11.1 V.

4.1 DBSCAN Algorithm Overview

Clustering is a branch of machine learning that aims to separate collected data into homogeneous groups. The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm is based on the concept of point density and generally uses the Euclidean distance

Given two data points x_i and x_j in an n-dimensional space, the distance is defined as :

10.22060/EEJ.2026.24892.5788

$$d(x_i, x_j) = \sqrt{\sum (x_{ik} - x_{jk})^2} \tag{4}$$

Each observation is characterized by several measured attributes (capacity, voltage, internal resistance). To compare these attributes on an equal basis, normalization is applied:

$$z = \frac{x - \mu}{\sigma} \tag{5}$$

The DBSCAN algorithm depends on two key parameters: the neighborhood radius ε and the minimum number of neighbors *MinPts*. The neighborhood of a point x_i is defined as:

$$N_\varepsilon(x_i) = \{x_j \in D \quad | \quad d(x_i, x_j) \leq \varepsilon\} \tag{6}$$

A point is a core point if $N_\varepsilon(x_i) \geq MinPts$. A border point is not a core point but belongs to the neighborhood of a core point. A point is considered noise if it is neither a core point nor a border point.

Thus, DBSCAN clusters correspond to density-connected sets of points, i.e., regions where the local density exceeds athreshold. The volume of a d-dimensional ball of radius ε is :

$$V_d(\varepsilon) = \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)} \varepsilon^d \tag{7}$$

Which allows associating DBSCAN with an implicit minimum density:

$$\rho \min = \frac{MinPts}{V_d(\varepsilon)} \tag{8}$$

4.2 Synthetic Data Analysis: Impact of Cluster Standard Deviation

10.22060/EEJ.2026.24892.5788

This subsection presents results on synthetic data generated using the *make_blobs* function. These illustrations serve to demonstrate the behavior of DBSCAN and the influence of its parameters on clustering quality. The results on real battery data are presented in Sections 4.5 and 4.6.

Fig. 3 illustrates the result of applying the DBSCAN algorithm implemented in Python on synthetic data characterizing lithium-ion 18650 cells [11, 20]. The real data are first imported from an Excel file, then transformed using the "make_blobs" function to generate synthetic clusters with three predefined centers: [1, 1], [-1, -1], and [1, -1]. Afterwards, the "StandardScaler" function is applied to normalize the data, ensuring equal contribution of each dimension during clustering [21].

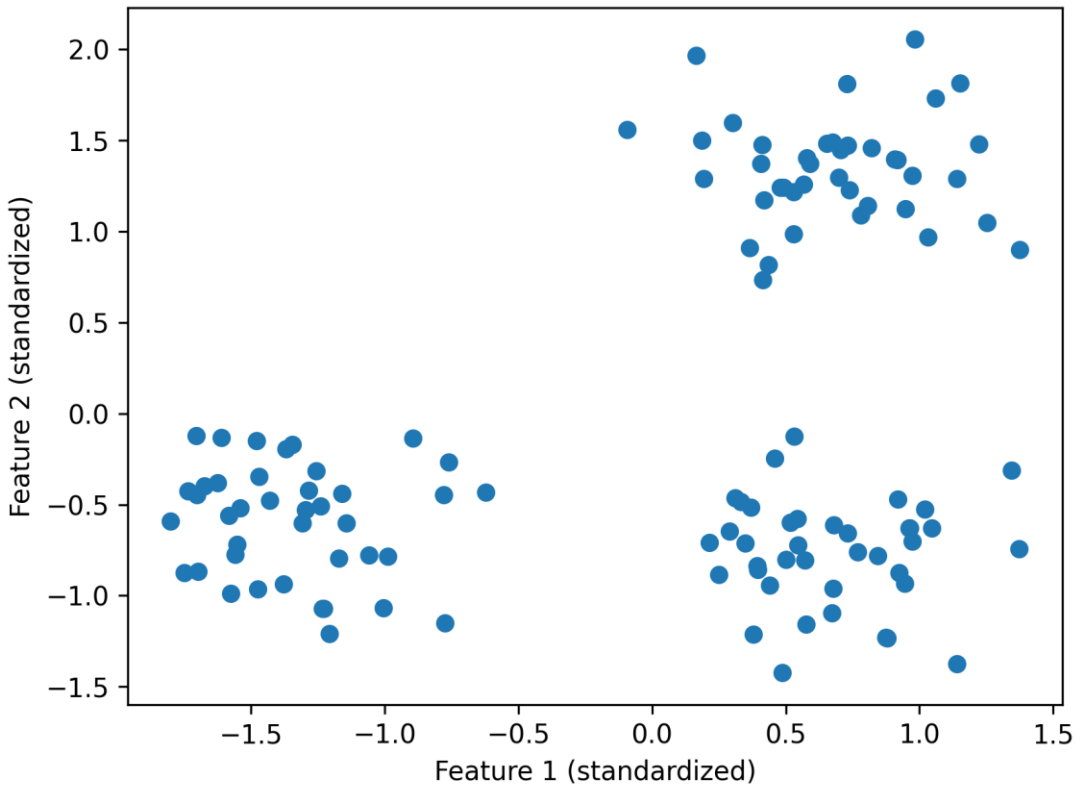


Fig. 3. Creation of Synthetic Clusters (synthetic data for methodological illustration)

10.22060/EEJ.2026.24892.5788

This subsection (Sections 4.2 to 4.7) presents results on synthetic data generated using the make_blobs function for methodological illustration. Results on real battery data are presented in Section 4.8.

In order to analyze the characteristics of the 120 lithium 18650 cells, the application of the DBSCAN algorithm generates synthetic clusters created by the ‘make_blobs’ function. This algorithm identifies core samples in high-density areas and extends clusters suitable for our data with similar density clusters. The labels assigned by the DBSCAN algorithm are accessible using the ‘labels_’ attribute, where noise points, which are shown in orange in Fig. 4, receive the label (-1).

The analysis and evolution show that, for standard deviations ‘cluster_std’ [22] of 0.10 and 0.20, DBSCAN identifies three clusters with a noise point value equal to (0 and 2 points respectively), indicating well-defined and dense clusters. At the ‘cluster_std’ value of 0.30, DBSCAN identifies three clusters with a significant increase in noise points up to 23, indicating increasing dispersion of data points. For the ‘cluster_std’ value of 0.40, the number of detected clusters and noise points changes significantly, rising to four clusters and 60 noise points; however, for the value 0.50, there is a single cluster and 106 noise points. This demonstrates that DBSCAN is sensitive to variations in cluster density, and its ability to identify precise clusters diminishes with an increase in standard deviation, leading to a notable increase in noise points.

10.22060/EEJ.2026.24892.5788

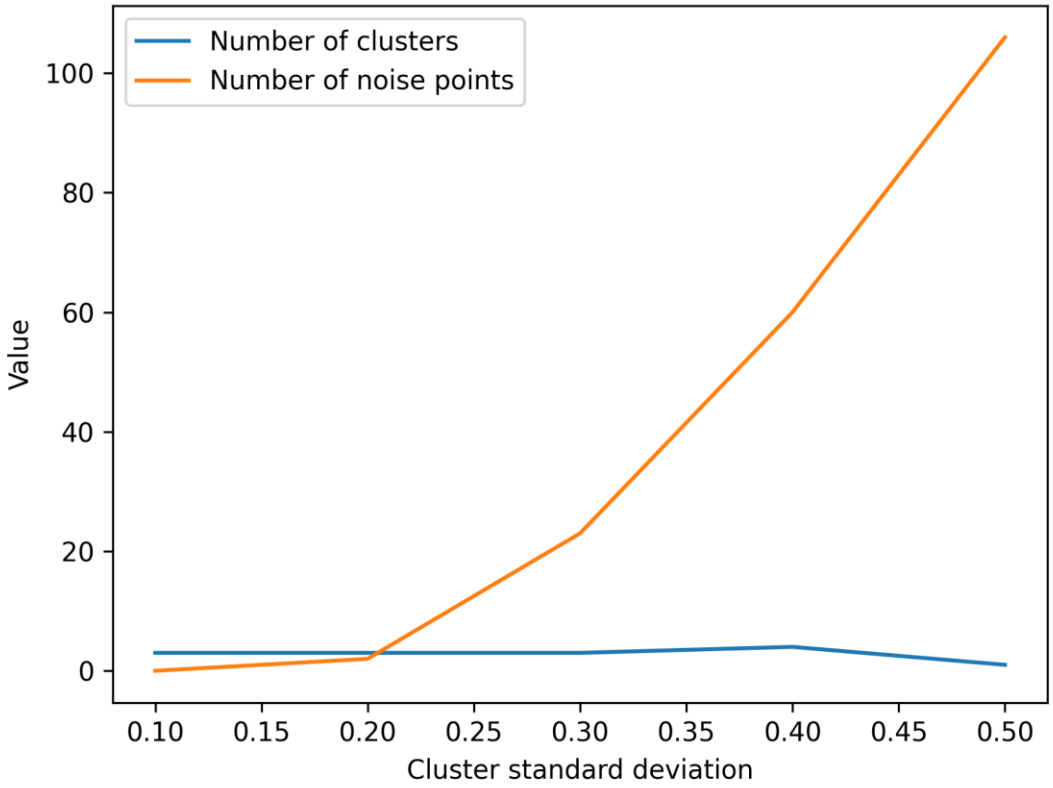


Fig. 4. Impact of Cluster Standard Deviation on Number of Clusters and Noise Points (synthetic data)

4.3 Synthetic Data Analysis: Impact of Epsilon (eps)

The determination of the optimal value of ‘eps’ primarily depends on the data distribution on one hand and on what is considered a (good) clustering result on the other hand. Generally, statistical methods such as cross-validation can be employed to evaluate the performance and generalization of a machine learning model and to search for the optimal ‘eps’ value that maximizes a specific clustering quality measure. Fig. 5 presents the variations of several clustering performance metrics as a function of the ‘eps’ value, a crucial parameter of the DBSCAN algorithm, applied to a dataset of 120 electrically tested lithium 18650 cells. The evaluated metrics include the number of clusters, the number of noise points, homogeneity, completeness, V-measure, adjusted Rand index, adjusted mutual information (AMI), and silhouette

10.22060/EEJ.2026.24892.5788

coefficient. Analyzing these metrics provides better insight into how ‘eps’ influences the quality of clusters formed by DBSCAN in the specific context of lithium cell data.

According to Fig. 5, the values of the adjusted Rand index and adjusted mutual information, stable around 0.7, show a good agreement between the predicted clusters and the true cluster labels for the 18650 lithium cells. The silhouette coefficient, around 0.5, indicates well-separated clusters. The ‘eps’ values around 0.3 maximize these performance metrics, offering a good compromise between a low number of noise points and high, stable clustering quality. Increasing ‘eps’ from 0.25 to 0.50, the number of clusters remains stable around three, while the noise points decrease from 40 to 0, suggesting that higher ‘eps’ values incorporate more points into the existing clusters. The curves of homogeneity (0.8), completeness (0.65), and V-measure (0.7) indicate consistent and well-formed clusters. Thus, DBSCAN with optimal parameters reliably identifies significant groups of cells with similar properties, crucial for the analysis of 18650 lithium cells.

10.22060/EEJ.2026.24892.5788

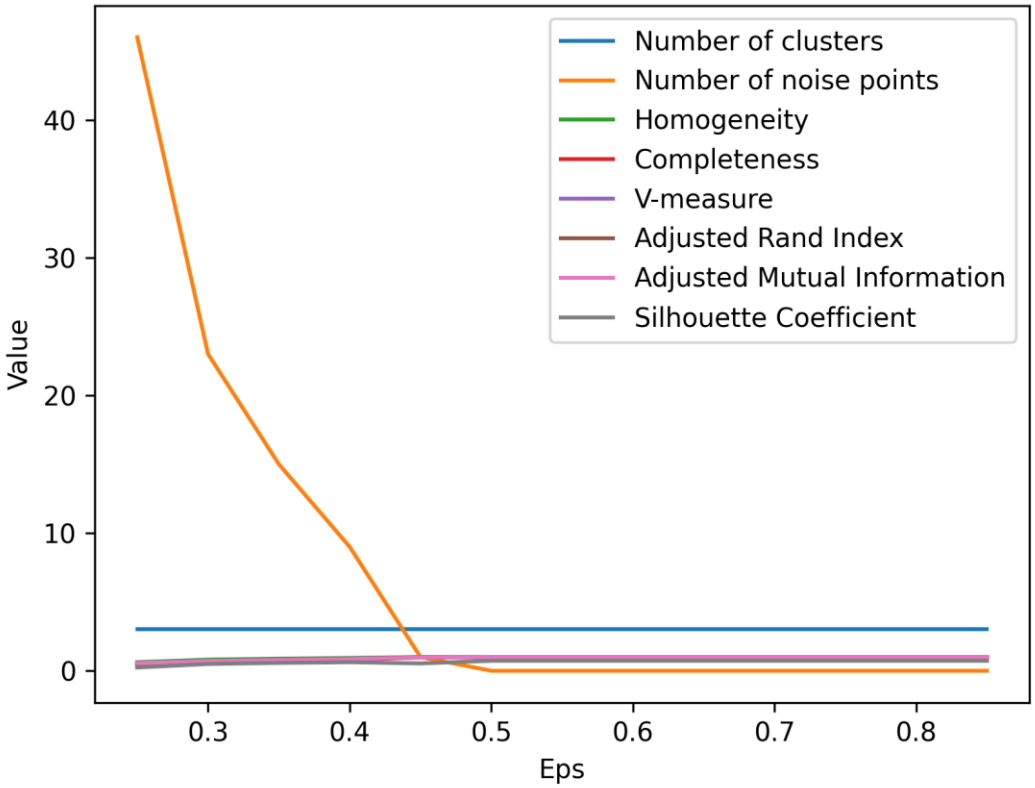


Fig. 5. Impact of Epsilon Variation on Clustering Metrics for Lithium 18650 Cells using DBSCAN (synthetic data)

4.4 Synthetic Data Analysis: Impact of Min Samples

In the context of data clustering, it is crucial to consider the distribution and allocation of data as well as the definition of a good clustering result to determine the optimal value of a critical parameter such as 'min_samples,' which represents the minimum number of points required for an area to be considered a dense cluster. Similar to the 'eps' parameter, using the same technique (cross-validation) is preferable to determine the optimal value of 'min_samples' that maximizes a specific quality measure of clustering. Additionally, a visual analysis of the results for various 'min_samples' values can help choose the value that generates the most homogeneous and meaningful clusters. By using a database consisting of lithium 18650 cells, it was observed that performance indicators such as the number of clusters, the number of noise

10.22060/EEJ.2026.24892.5788

points, homogeneity, completeness, the V-measure, the adjusted Rand index, the adjusted mutual information, and the silhouette coefficient vary depending on the value of ‘min_samples’, highlighting the importance of this parameter in clustering optimization.

According to Fig. 6, it is observed that increasing the value of ‘min_samples’ leads to a significant increase in the number of noise points, while the number of clusters remains relatively stable. This indicates that more data points are classified as noise with higher values of ‘min_samples’. The metrics of homogeneity, completeness, and V-measure remain stable, indicating internally consistent and well-formed clusters. Similarly, the adjusted Rand index and adjusted mutual information maintain values around 0.7, leading to a good match between the predicted clusters and the true labels. The silhouette coefficient, varying slightly around 0.5, demonstrates that the clusters are well-separated. These results show that despite the variations in ‘min_samples’, DBSCAN successfully forms distinct and coherent clusters, facilitating a reliable analysis of the properties of lithium cells.

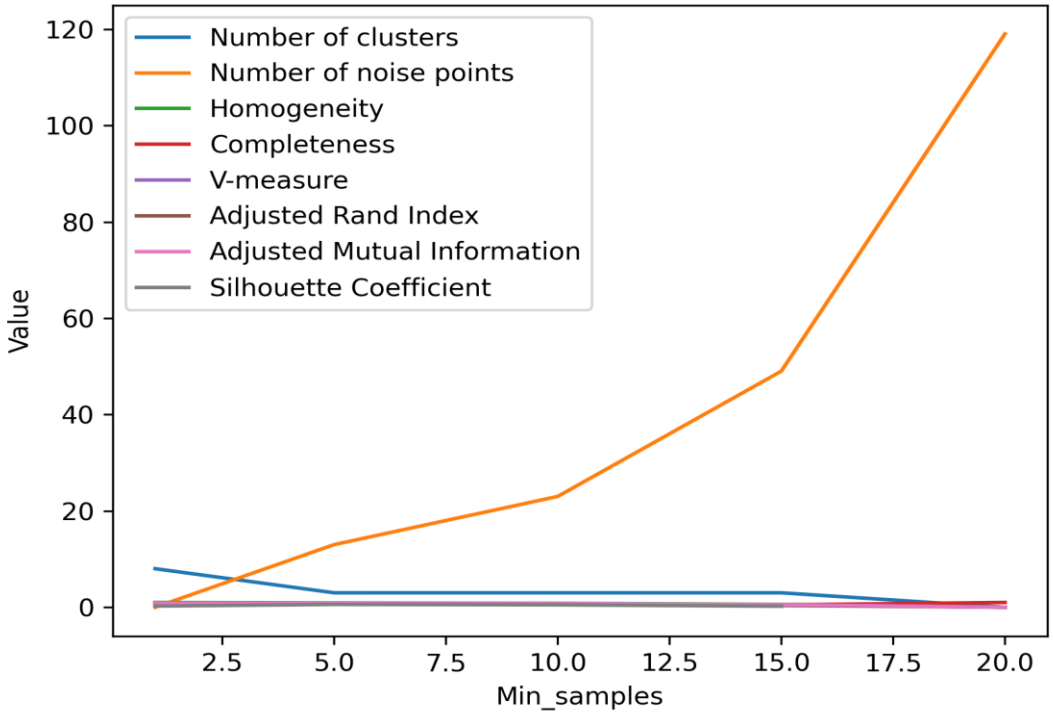


Fig. 6. Impact of ‘min_samples’ Variation on Clustering Metrics for Lithium 18650 Cells using DBSCAN (synthetic data)

4.5 Synthetic Data Analysis: Combined Impact of eps and min_samples

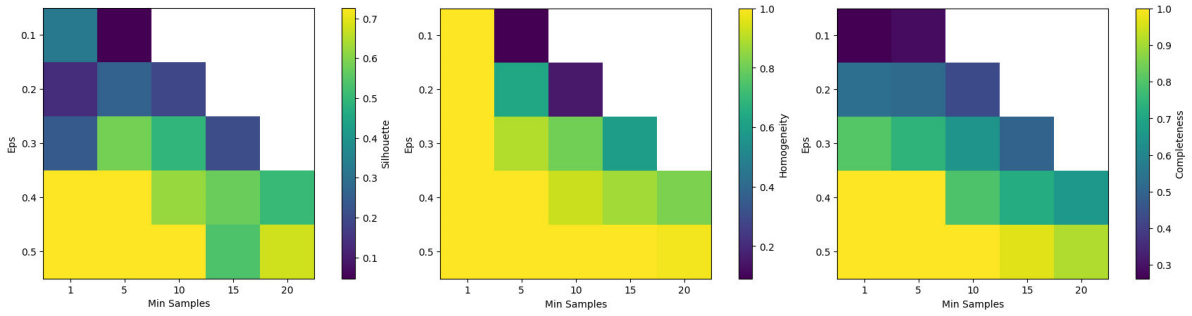
Fig. 7-a illustrates the influence of the ‘eps’ and ‘min_samples’ parameters of the DBSCAN algorithm on the silhouette coefficient, which measures the quality of the clusters formed for 18650 lithium cells. The silhouette coefficient, indicated by a color scale, ranges from 0.1 to 0.7, where higher values suggest more coherent and well-separated clusters. According to Fig. 7-a, lower ‘eps’ values (0.1 to 0.3) associated with intermediate ‘min_samples’ values (5 to 15) tend to produce higher silhouette coefficients, indicating better-defined clusters.

On the other hand, higher ‘eps’ values (0.4 to 0.5) show a significant decrease in the silhouette coefficient, especially marked for lower ‘min_samples’ values. This observation suggests that the DBSCAN algorithm generates higher quality clusters in terms of separation and internal cohesion when ‘eps’ parameters are kept at moderate levels and ‘min_samples’ at intermediate levels. In conclusion, to optimize the quality of clusters in this analysis, it is recommended to choose ‘eps’ values around 0.2 to 0.3 and ‘min_samples’ values around 10.

Fig. 7-b shows the effect of the ‘eps’ and ‘min_samples’ parameters of the DBSCAN algorithm on the homogeneity of the clusters formed for 18650 lithium cells. Homogeneity, indicated by a color scale, ranges from 0 to 1, where higher values represent clusters with points that are more similar to each other. It is observed that lower ‘eps’ values (0.1 to 0.3) combined with lower ‘min_samples’ values (1 to 10) produce higher homogeneity, reaching almost 1, suggesting that the clusters formed are very homogeneous. As ‘eps’ increases beyond 0.3 and ‘min_samples’ increases beyond 10, homogeneity gradually decreases, indicating a reduction in the similarity of points within clusters. These observations suggest that to maximize the homogeneity of clusters, it is optimal to choose low ‘eps’ values (around 0.1 to 0.2) and low to intermediate ‘min_samples’ values (between 1 and 10). This allows the formation of clusters where the points are very similar, which is crucial for analyses where the internal consistency of clusters is important.

10.22060/EEJ.2026.24892.5788

Fig. 7-c shows the effect of the ‘eps’ and ‘min_samples’ parameters of the DBSCAN algorithm on the completeness of the clusters formed for 18650 lithium cells. Completeness, measured on a scale from 0 to 1, indicates the extent to which all points that should be grouped into a cluster are actually included. Higher completeness means that the clusters effectively capture all relevant points. In this figure, we observe that lower ‘eps’ values (0.1 to 0.2) combined with low to intermediate ‘min_samples’ values (1 to 10) tend to produce higher completeness, indicating that these parameter combinations effectively capture relevant points in the clusters. In contrast, as ‘eps’ increases beyond 0.3, completeness starts to decrease, particularly for higher ‘min_samples’ values. This suggests that higher values of ‘eps’ and ‘min_samples’ reduce the algorithm’s ability to capture all relevant points in the clusters, resulting in less complete clusters.



(a) (b) (c)

Fig. 7. Impact of ‘eps’ and ‘min_samples’ Parameters; (a) on Silhouette Coefficient for DBSCAN Clustering, (b) on Homogeneity for DBSCAN Clustering, (c) on Completeness for DBSCAN Clustering (synthetic data)

To optimize the completeness of the clusters, it is recommended to choose lower ‘eps’ values (around 0.1 to 0.2) and low to intermediate ‘min_samples’ values (between 1 and 10). This configuration allows the formation of clusters that include most of the relevant points, thereby maximizing completeness and ensuring that the clusters accurately reflect the data structure of the 18650 lithium cells. The three figures have been presented as heatmaps to facilitate the visualization of variations and trends, demonstrating the relationship between two continuous variables (‘min_samples’ and ‘eps’) and illustrating the performance of the models based on different parameters. According to the ‘make_blobs’ function, which provides a

basic function for generating synthetic data in the field of machine learning, its purpose is to create datasets with clusters and simulate three distinct clusters. This function is part of the datasets module of the (Scikit-learn library in Python). According to Fig. 7: the best value of the ‘eps’ parameter was selected as 0.3, with ‘min_samples’ set to 10, where the DBSCAN algorithm produced the following results as shown in Table 2.

Table 2. Context and Parameters Algorithm used: DBSCAN with ‘eps’=0.35 and ‘min_samples’=10 (synthetic data)

cluster_std	0.20	0.22	0.25	0.28	0.30
Estimated number of clusters	3	3	3	3	3
Estimated number of noise points	0	1	7	13	15
Homogeneity	1.000	1.000	0.946	0.893	0.875
Completeness	1.000	0.965	0.827	0.741	0.718
V-measure	1.000	0.982	0.882	0.810	0.789
Adjusted Rand Index	1.000	0.987	0.909	0.826	0.798
Adjusted Mutual Information	1.000	0.982	0.880	0.806	0.785
Silhouette Coefficient	0.819	0.609	0.683	0.604	0.574

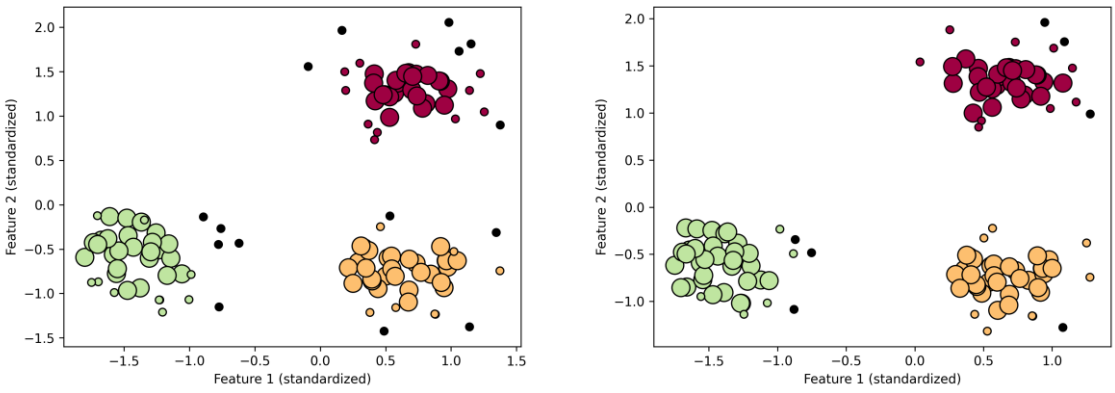
4.6 Synthetic Data Analysis: Optimal Clustering Results

By adjusting the value of the cluster standard deviation ‘cluster_std’ as an operational scenario, we can observe that for a ‘cluster_std’ value of 0.20, the clusters are perfectly identified without any noise points, with all performance metrics at their maximum, showing perfect alignment between the predicted clusters and the true cluster labels. However, as the ‘cluster_std’ increases, there is a progressive introduction of noise points and a decrease in performance metrics. At a ‘cluster_std’ of 0.22, only one noise point appears, where homogeneity remains perfect, but completeness and other metrics slightly decrease. When the ‘cluster_std’ reaches 0.25, the number of noise points increases to 7, with less defined and more dispersed clusters. This trend continues with a ‘cluster_std’ of 0.28, where the number of noise points rises to 13, and performance metrics continue to decrease, indicating even more diffused clusters. At a ‘cluster_std’ value of 0.30, the number of noise points reaches 15, with reduced homogeneity and completeness values, indicating significantly less coherent and more dispersed clusters. Two graphical presentations illustrate

10.22060/EEJ.2026.24892.5788

the negative impact of a poor choice of ‘cluster_std’, as shown in Figures 8 (a and b), where noise points appear.

Fig. 9 shows the result of applying the DBSCAN algorithm on synthetic data representing 18650 lithium cells. The graph and numerical results clearly illustrate three distinct, well-separated, and cohesive clusters, confirming the (Estimated number of clusters: 3). The points are colored according to the clusters they belong to, with no noise points present, as indicated by (Estimated number of noise points: 0). The three distinct clusters demonstrate that DBSCAN successfully identified well-separated groups of points. Each cluster is visually compact, indicating good separation and internal cohesion.



(a) (b)

Fig. 8. Optimal Clustering of Lithium 18650 Cells with DBSCAN; (a) (‘cluster_std’=0.3), (b) (‘cluster_std’=0.25) (synthetic data)

10.22060/EEJ.2026.24892.5788

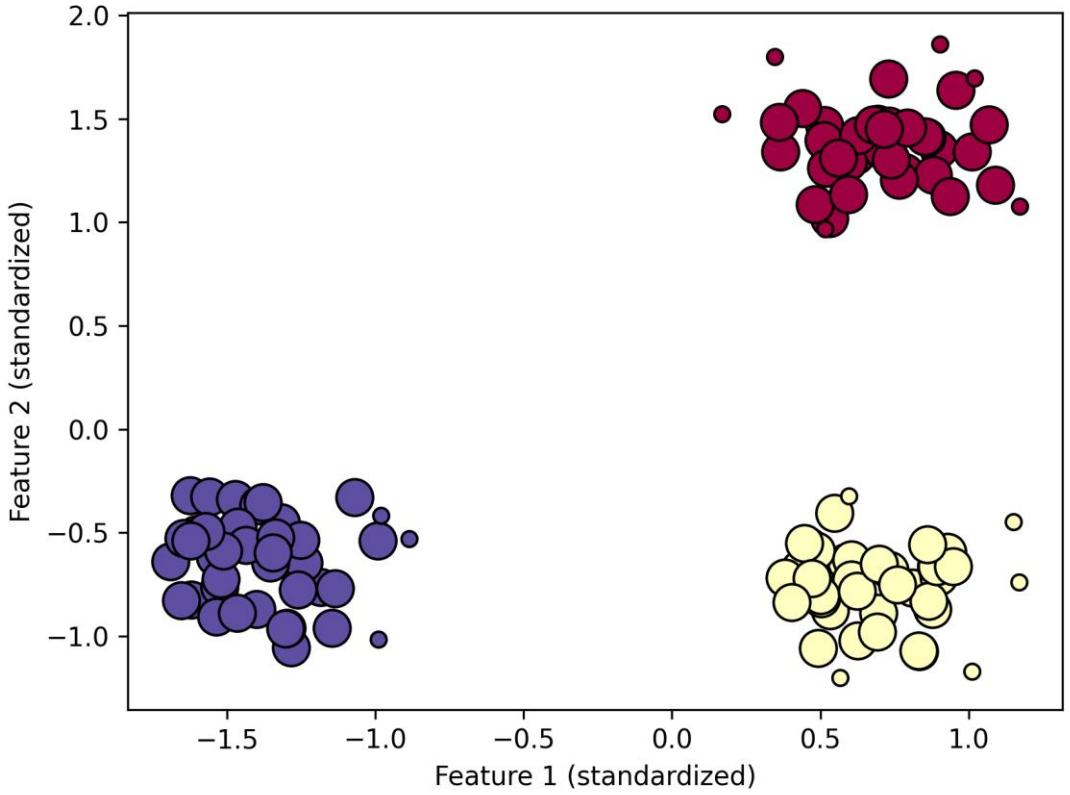


Fig. 9. Optimal Clustering of Lithium 18650 Cells with DBSCAN (cluster_std=0.2) (synthetic data)

4.7 Synthetic Data Analysis: Random Labeling Performance

Fig. 10 illustrates the clustering performance metrics of the 120 18650 cells as a function of the number of clusters, comparing a uniform random labeling to a reference assigned with 10 classes, where each cell is characterized by its nominal voltage, capacity, and internal resistance. The two plots mentioned in blue and orange show that these two evaluation metrics (the adjusted Rand index and the adjusted mutual information) remain close to zero regardless of the number of clusters, indicating a poor correspondence between the random labeling and the true reference classes. Conversely, the three metrics (homogeneity, completeness, and V-measure) significantly increase when the number of clusters rises from 0 to 20 before stabilizing. We observe a homogeneity of approximately 0.8, suggesting that the points within the clusters

10.22060/EEJ.2026.24892.5788

exhibit high similarity concerning nominal voltage, capacity, and internal resistance. Achieving a completeness of about 0.5 means that the clusters encompass a large portion of points from the same class, but not entirely. The V-measure, which combines homogeneity and completeness, shows continuous improvement up to about 0.6, reflecting better clustering quality when the number of clusters more closely matches the true classes of the cells. Fig. 10 highlights the importance of correctly adjusting the number of clusters to enhance clustering quality and accurately reflect the underlying structure of the 18650 lithium cell data.

Fig. 11 presents the clustering performance measures of 18650 lithium cells as a function of the number of clusters, using two uniform random labelings with an equal number of clusters. The cell characteristics include nominal voltage, capacity, and internal resistance. The curves show that the first two static measures (adjusted Rand index and adjusted mutual information) remain close to zero regardless of the number of clusters, indicating minimal correspondence between the random labelings and the true reference classes. On the other hand, the three metrics (homogeneity, completeness, and V-measure) significantly increase when the number of clusters rises from 0 to 40, then stabilize around 0.8 to 0.9.

According to homogeneity, the points within each cluster exhibit very similar characteristics, while completeness indicates that the clusters encompass a large portion of the points from the same class. The V-measure, which combines homogeneity and completeness, also reaches high values, reflecting good clustering quality. These results suggest that, despite random labeling, an adequate number of clusters allows capturing the underlying structure of the 18650 lithium cell data, thereby maximizing the coherence and representativeness of the clusters based on their essential characteristics.

10.22060/EEJ.2026.24892.5788

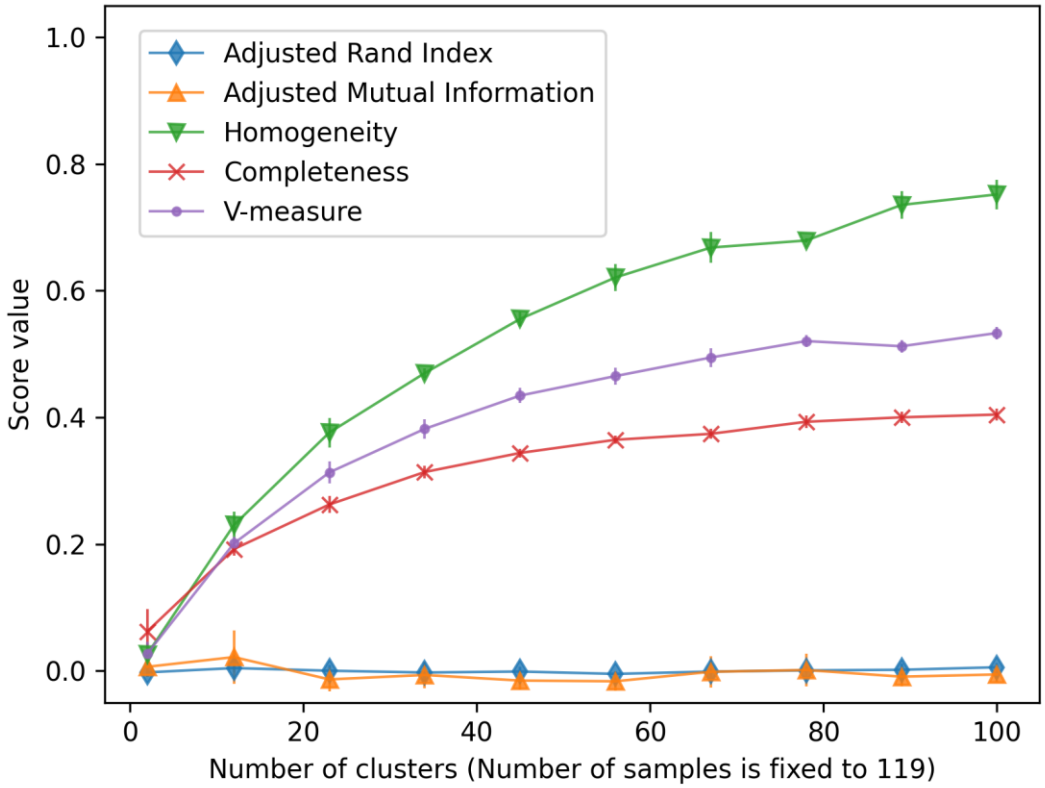


Fig. 10. Clustering measures for random uniform labeling against reference assignment with 10 classes (synthetic data)

10.22060/EEJ.2026.24892.5788

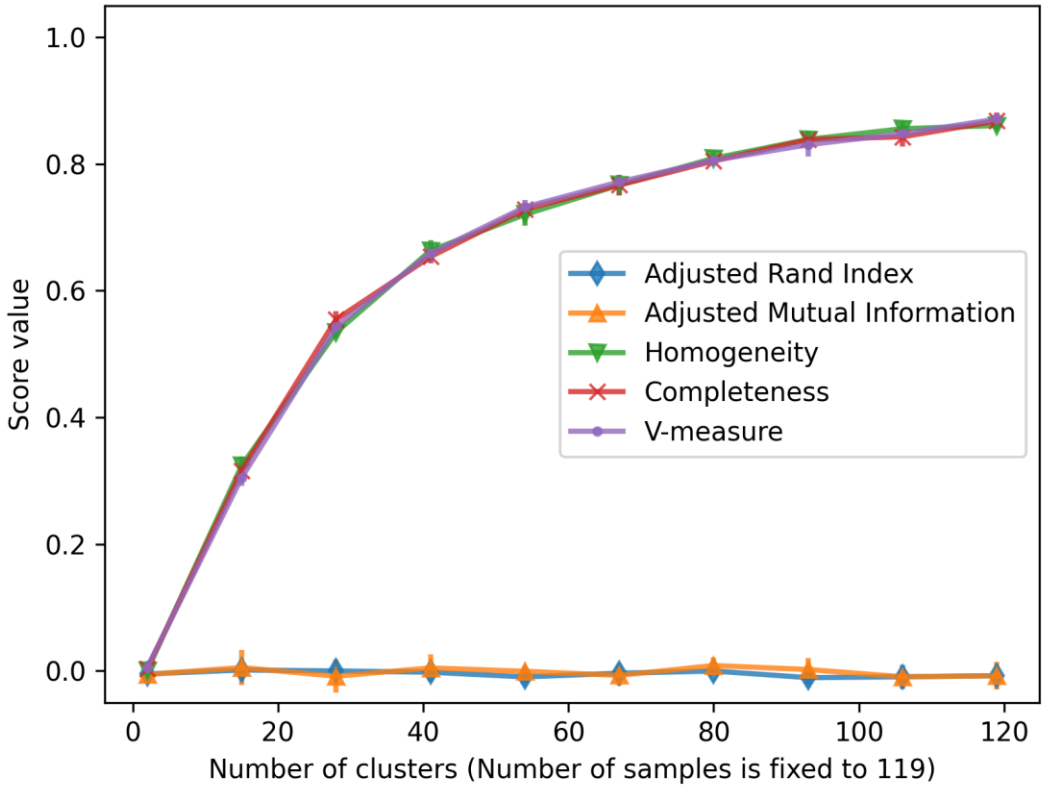


Fig. 11. Clustering measures for 2 random uniform labelings with equal number of clusters (synthetic data)

4.8 Results on Real Battery Data

This subsection presents results on real battery data from 120 experimentally measured 18650 cells. Fig. 12 shows the distribution of the 120 healthy cells in terms of capacity and internal resistance, with color indicating the nominal voltage (3.6 V or 3.7 V). The figure reveals a clear separation between cells with different voltage ratings.

10.22060/EEJ.2026.24892.5788

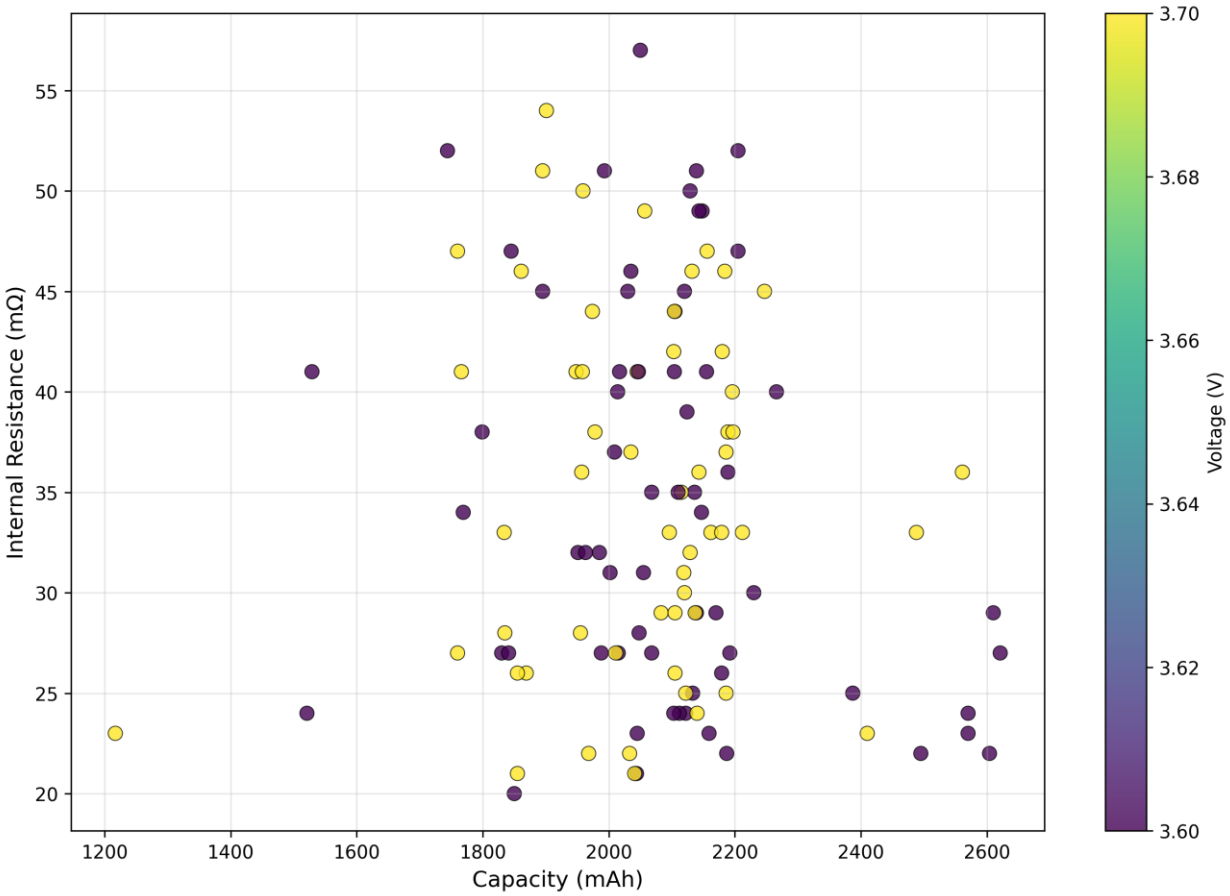


Fig. 12. Distribution of 120 real 18650 cells: capacity vs internal resistance (color indicates nominal voltage) (real battery data)

A systematic parameter sweep was performed on the real battery dataset to determine the optimal DBSCAN parameters. The search grid covered eps values from 0.3 to 1.0 and min_samples values from 3 to 15. Table 3 summarizes the best parameter combinations.

Table 3: DBSCAN parameter optimization results on real battery data

eps	Min_samples	Number of clusters	Noise points	silhouette
0.3	3	12	0	0.016
0.4	3	9	0	0.162
0.5	3	4	0	0.215
0.5	5	3	0	0.285
0.6	3	3	0	0.361
0.7	3	3	0	0.378
0.8	3	3	0	0.385
0.9	3	3	6	0.392

10.22060/EEJ.2026.24892.5788

The optimal parameters were found to be eps = 0.9 and min_samples = 3, yielding a silhouette coefficient of 0.392 with 3 clusters and 6 noise points.

Fig. 13 presents the DBSCAN clustering result on the real battery data using the optimal parameters (eps=0.9, min_samples=3). Three distinct clusters are identified, along with 6 noise points. The silhouette coefficient of 0.392 indicates reasonably well-separated clusters.

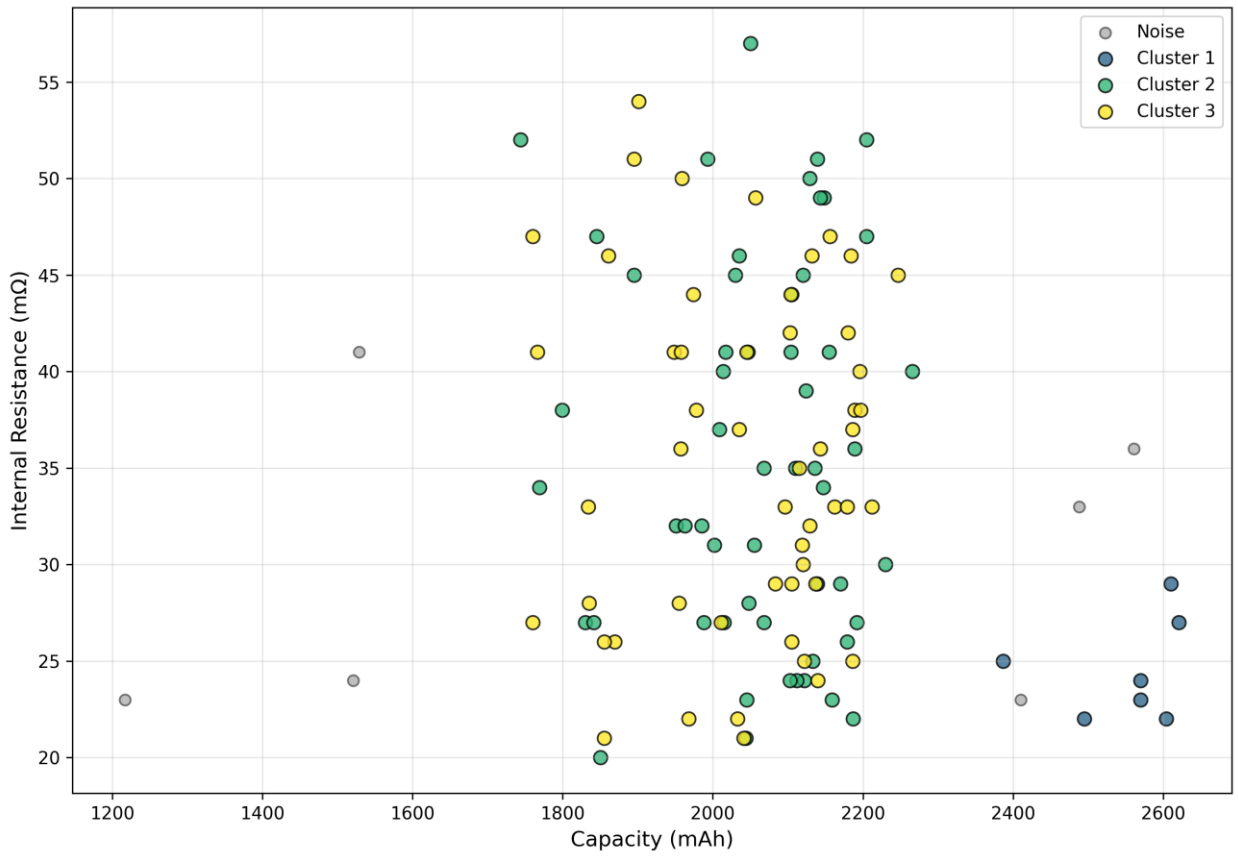


Fig. 13. DBSCAN clustering on 120 real cells (eps=0.9, min_samples=3, 3 clusters, 6 noise points, silhouette=0.39) (real battery data).

Table 4 summarizes the characteristics of each cluster identified on real data.

Table 4: Characteristics of the three clusters identified by DBSCAN on real battery data

Cluster	Number of cells Mean	capacity (mAh)	Capacity std dev (mAh)	Mean voltage (V)	Mean resistance (mΩ)
1	7	2551	84	3.60	24.6
2	55	2057	122	3.60	35.7
3	52	2041	133	3.70	35.5
Noise	3	1954	-	-	30.0

Table 5 presents a comparison of DBSCAN with K-means and hierarchical clustering on the real battery dataset.

Table 5: Performance comparison of clustering algorithms on real 18650 cell data

Algorithm	Number of clusters	Silhouette	Noise points
DBSCAN (eps = 0.9, min_samples = 3)	3	0.39	6
K-means	3	0.37	-
Hierarchical	3	0.36	-

DBSCAN achieves the highest silhouette coefficient (0.39) among the three algorithms tested. Moreover, DBSCAN uniquely identifies 6 cells as noise, which would be forced into clusters by K-means or hierarchical clustering, potentially compromising pack homogeneity. This advantage makes DBSCAN particularly suitable for battery cell regrouping applications.

4.9 Interpretation for Cell Assembly

Through DBSCAN, the 120 cells are distributed into three homogeneous clusters C_1, C_2, C_3, For the final 3S battery assembly, we consider the three identified clusters.

Capacity of a parallel pack:

$$C_{pack} = \sum_{i=1}^{n} C_i \tag{9}$$

Equivalent internal resistance (parallel):

10.22060/EEJ.2026.24892.5788

$$\frac{1}{R_{pack}} = \sum_{i=1}^{n} \frac{1}{R_i} \tag{10}$$

Nominal voltage: equal to the nominal voltage of a single cell, since in parallel the voltages are equalized. For the final 3S battery:

$$V_{total}^{(nom)} = V_{pack1} + V_{pack2} + V_{pack3} \tag{11}$$

Total capacity (in series, limited by the weakest pack):

$$C_{total} = \min\left\{C_{pack1}, C_{pack2}, C_{pack3}\right\} \tag{12}$$

Equivalent internal resistance:

$$R_{total} = R_{pack1} + R_{pack2} + R_{pack3} \tag{13}$$

Table 6 presents the calculated parameters for the 3S40P battery based on the real clusters. For packs with more than 40 cells (Clusters 2 and 3), the 40 cells with the highest capacities are selected. For Pack 3, the 7 cells from Cluster 1 (2551 mAh each) provide 17.9 Ah; the remaining 33 cells must be sourced from elsewhere to reach 40 cells. The resistance calculation assumes an average of 35 mΩ for the supplemented cells.

Table 6: 3S40P battery calculation based on real clusters

Parameter	Pack 1 (from Cluster2)	Pack 1 (from Cluster3)	Pack 1 (from Cluster1) + supplement	Total (3S)
Capacity (Ah)	82.3	81.6	81.6	81.6
Voltage (V)	3.6	0.37	3.60	10.9
Resistance (mΩ)	0.89	0.92	1.42	3.23

Note: Pack 3 requires supplementing the 7 cells from Cluster 1 (2551 mAh each, total 17.9 Ah) with 33 additional cells from other sources to reach 40 cells. The total capacity of Pack 3 is therefore limited by the supplemented cells (estimated at 81.6 Ah).

The final 3S40P battery has a nominal voltage of 10.9 V, a total capacity of approximately 82 Ah (81.6 Ah), and an internal resistance of 3.23 mΩ. This approach provides a methodology for regrouping used 18650 cells using DBSCAN, demonstrating its practical feasibility.

5. Conclusion

In the literature concerning batteries, various research topics are addressed due to the diverse issues in this sector. This work presents an assembly of several fundamental and principal techniques, including a new recycling topology through electrical recovery based on the application of a reworking approach to diagnostics of 18650 cells. This approach aims to achieve several critical objectives such as improving diagnostic accuracy, reducing maintenance costs by implementing detection systems with a better selection criterion for chosen elements, extending cell lifespan by optimizing charge and discharge cycles, and monitoring cell state in real-time through the collaboration of an advanced algorithm such as DBSCAN.

From 154 recovered 18650 cells, electrical testing identified 120 healthy cells, representing a success rate of 78%. The average characteristics of the healthy cells were: capacity = 2015 mAh (range: 1217-2621 mAh), nominal voltage = 3.65 V (range: 3.6-3.7 V), and internal resistance = 35.2 mΩ (range: 20-57 mΩ). Parameter sensitivity analysis on synthetic data showed that eps=0.3 and min_samples=10 provide optimal clustering quality, achieving a silhouette coefficient of 0.819 with three clusters and zero noise points at cluster_std=0.20. On real battery data, the optimal parameters were found to be eps=0.9 and min_samples=3, producing three clusters and six noise points with a silhouette coefficient of 0.39, compared to 0.37 for K-means and 0.36 for hierarchical clustering. The three clusters identified on real data have the following characteristics: Cluster 1 (7 cells, 2551±84 mAh, 3.60 V, 24.6 mΩ), Cluster 2 (55 cells, 2057±122 mAh, 3.60 V, 35.7 mΩ), and Cluster 3 (52 cells, 2041±133 mAh, 3.70 V, 35.5 mΩ).

The use of the DBSCAN algorithm in this application demonstrates its robustness and power in cluster creation, as shown by the results on both synthetic and real data. The 3S40P battery assembled from these

clusters has a nominal voltage of 10.9 V, a total capacity of approximately 82 Ah, and an internal resistance of 3.23 mΩ. It is crucial to design homogeneous packs of 18650 cells to ensure the performance, safety, and longevity of the assembled battery. This approach can be applied to various devices such as electric vehicles, portable electronic equipment, and energy storage systems. The essential role of DBSCAN is to identify and group similar cells based on their performance characteristics. With this method, it is ensured that cells within the same pack exhibit similar behaviors, reducing risks of imbalance and failure. These advanced techniques help reduce the ecological footprint of 18650 cells while benefiting from their performance and reliability advantages. Future work will focus on extended cycle life testing (500+ cycles) of the assembled battery and comparison with other density-based clustering algorithms such as HDBSCAN and OPTICS. The complete dataset and Python implementation are available from the corresponding author upon reasonable request.

6. Reference

[1] N. NASAJPOUR-ESFAHANI, H. GARMESTANI, M. BAGHERITABAR, D. J. JASIM, D. TOGHRAIE, S. DADKHAH and H. FIROOZEH, Comprehensive review of lithium-ion battery materials and development challenges, Renewable and Sustainable Energy Reviews, vol. 203, 2024, p. 114783. doi: 10.1016/j.rser.2024.114783.

[2] G. CRABTREE, E. KÓCS and L. TRAHEY, The energy-storage frontier: Lithium-ion batteries and beyond, MRS Bulletin, vol. 40, no. 12, 2015, pp. 1067–1078. doi: 10.1557/mrs.2015.259.

[3] E. HOSSAIN, D. MURTAUGH, J. MODY, H. M. R. FARUQUE, M. S. H. SUNNY and N. MOHAMMAD, A comprehensive review on second-life batteries: Current state, manufacturing considerations, applications, impacts, barriers & potential solutions, business strategies, and policies, IEEE Access, vol. 7, 2019, pp. 73215–73252. doi: 10.1109/ACCESS.2019.2917859.

[4] G. ZUBI, R. S. ADHIKARI, N. E. SÁNCHEZ and W. ACUÑA-BRAVO, Lithium-ion battery-packs for solar home systems: Layout, cost and implementation perspectives, Journal of Energy Storage, vol. 32, 2020, p. 101985. doi: 10.1016/j.est.2020.101985.

[5] R. E. CIEZ and J. F. WHITACRE, Comparison between cylindrical and prismatic lithium-ion cell costs using a process based cost model, Journal of Power Sources, vol. 340, 2017, pp. 273–281. doi: 10.1016/j.jpowsour.2016.11.054.

[6] N. SHARMILI, R. NAGI and P. WANG, A review of research in the Li-ion battery production and reverse supply chains, Journal of Energy Storage, vol. 68, 2023, p. 107622. doi: 10.1016/j.est.2023.107622.

[7] OECD, Promoting green and digital innovation: The role of upskilling and reskilling in higher education, OECD Education Policy Perspectives, No. 103, OECD Publishing, Paris, 8 July 2024.

[8] U. Saleem, B. Joshi and S. Bandyopadhyay, "Hydrometallurgical Routes to Close the Loop of Electric Vehicle (EV) Lithium-Ion Batteries (LIBs) Value Chain: A Review," Journal of Sustainable Metallurgy, vol. 9, no. 3, pp. 950–971, Sept. 2023, doi: 10.1007/s40831-023-00718-w.

[9] A. L. SRIVASTAV, MARKANDEYA, N. PATEL, M. PANDEY, A. K. PANDEY, A. K. DUBEY, A. KUMAR, A. K. BHARDWAJ and V. K. CHAUDHARY, Concepts of circular economy for sustainable management of electronic wastes: Challenges and management options, Environmental Science and Pollution Research, vol. 30, no. 17, 2023, pp. 48654–48675. doi: 10.1007/s11356-023-26052-y.

[10] X. LIU, G. CHANG, J. TIAN, S. LI and C. CHEN, "Regrouping strategy of retired batteries considering SOC consistency", Energy Reports, vol. 8, 2022, pp. 218–228.

[11] D. LI, Z. ZHANG, P. LIU and Z. WANG, DBSCAN-based thermal runaway diagnosis of battery systems for electric vehicles, Energies, vol. 12, no. 15, 2019, p. 2977.

[12] M. ESTER, H. P. KRIEGEL, J. SANDER and X. XU, “A density-based algorithm for discovering clusters in large spatial databases with noise”, in Proc. KDD, vol. 96, no. 34, 1996, pp. 226–231.

[13] C. LOPEZ, S. TUCKER, T. SALAMEH and C. TUCKER, An unsupervised machine learning method for discovering patient clusters based on genetic signatures, Journal of Biomedical Informatics, vol. 85, 2018, pp. 30–39. doi: 10.1016/j.jbi.2018.07.004.

[14] A. SHARMA, R. K. GUPTA and A. TIWARI, Improved Density Based Spatial Clustering of Applications of Noise Clustering Algorithm for Knowledge Discovery in Spatial Data, Mathematical Problems in Engineering, vol. 2016, 2016, p. 1564516 (9 pages). doi: 10.1155/2016/1564516.

[15] F. ROS, S. GUILLAUME, R. RIAD and M. EL HAJJI, Detection of natural clusters via S-DBSCAN: A Self-tuning version of DBSCAN, Knowledge-Based Systems, vol. 241, 2022, p. 108288. doi: 10.1016/j.knosys.2022.108288.

[16] Y. LIN, E. GIACOUMIDIS, S. O’DUILL and L. P. BARRY, DBSCAN-Based Clustering for Nonlinearity Induced Penalty Reduction in Wavelength Conversion Systems, IEEE Photonics Technology Letters, vol. 31, no. 21, 2019, pp. 1709–1712. doi: 10.1109/LPT.2019.2942961.

[17] Y. CUI, Y. CHEN, C. WANG, C. GU, M. O’NEILL and W. LIU, Programmable ring oscillator PUF based on switch matrix, in Proc. 2020 IEEE Int. Symp. Circuits and Systems (ISCAS), 2020, pp. 1–4. doi: 10.1109/ISCAS45731.2020.9180905.

[18] S. K. SHAKTHIVEL, P. W. DAVID, S. PERIYA BACKIYAM and M. S. MURUGAN, Square dynamic reconfiguration for the partial shaded photovoltaic system–simulation and experimental analysis, Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, vol. 44, no. 3, 2022, pp. 6868–6885.

[19] W. FANG, H. CHEN and F. ZHOU, Fault diagnosis for cell voltage inconsistency of a battery pack in electric vehicles based on real-world driving data, Computers and Electrical Engineering, vol. 102, 2022, p. 108095.

[20] P. S. MOHAPATRA, Artificial Intelligence and Machine Learning for Test Engineers: Concepts in Software Quality Assurance, in Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle, vol. 4, 2025, p. 17.

[21] P. KATHIROLI and S. KANMANI, Data aggregation by enhanced squirrel search optimization algorithm for in wireless sensor networks, Wireless Networks, vol. 31, no. 3, 2025, pp. 2181–2201.

[22] A. PIVATO, F. GIROTTO, L. MEGIDO and R. RAGA, Estimation of global warming emissions in waste incineration and landfilling: An environmental forensic case study, Environmental Forensics, vol. 19, no. 4, 2018, pp. 253–264. doi: 10.1080/15275922.2018.1519741.