

Review

Applications of Clustering Methods for Different Aspects of Electric Vehicles

Masooma Nazari ¹, Akhtar Hussain ¹ and Petr Musilek ^{1,2,*} ¹ Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2G2, Canada² Department of Applied Cybernetics, University of Hradec Králové, 500 03 Hradec Králové, Czech Republic

* Correspondence: pmusilek@ualberta.ca

Abstract: The growing penetration of electric vehicles can pose several challenges for power systems, especially distribution systems, due to the introduction of significant uncertain load. Analysis of these challenges becomes computationally expensive with higher penetration of electric vehicles due to various preferences, travel behavior, and the battery size of electric vehicles. This problem can be addressed using clustering methods which have been successfully used in many other sectors. Recently, there have been several studies published on applying clustering methods for various aspects of electric vehicles. To summarize the existing efforts and provide future research directions, this contribution presents a three-step analysis. First, the existing clustering methods, including hard and soft clustering, are discussed. Then, the recent literature on the application of clustering methods for different aspects of electric vehicles is reviewed. The review concentrates on four major aspects of electric vehicles: the behavior of the user, driving cycle, used batteries, and charging stations. Then, several representative studies are selected from each category and their merits and demerits are summarized. Finally, gaps in the existing literature are identified and directions for future research are presented. They indicate the need for further research on the impact on distribution circuits, charging infrastructure during emergencies, equity and disparity in rebate allocations, and the use of big data with cluster analysis to assist transportation network management.

Keywords: clustering; charging station; electric vehicles; greenhouse gasses; transportation electrification; user behavior



Citation: Nazari, M.; Hussain, A.; Musilek, P. Applications of Clustering Methods for Different Aspects of Electric Vehicles. *Electronics* **2023**, *12*, 790. <https://doi.org/10.3390/electronics12040790>

Academic Editors: Zita Vale, João Soares and Hugo Morais

Received: 3 January 2023

Revised: 28 January 2023

Accepted: 1 February 2023

Published: 4 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the last decade, governments around the globe have implemented significant policy reforms to establish countermeasures and corrective steps to address the problem of climate change caused by humans [1]. The European Union proposed the European Green Deal in December 2019, in which the majority of member states pledge to zero net greenhouse gas emissions by 2050. The decarbonization of society is one of the foundations of the green deal [2]. There is global concern about climate change, which is typically associated with human influence on the environment caused by greenhouse gas emissions [3]. Global carbon dioxide (CO₂) from fossil fuel usage of fossil fuels surged from 6 billion tons in 1950 to 36.4 billion tons in 2021 [4,5]. As the second largest producer of CO₂ emissions, the transportation industry is responsible for 22.67% of the total emission. Figure 1 shows the distribution of CO₂ emissions in different sectors [4]. The global population is rising. It is expected that there will be 1.5 billion automobiles on the planet by 2025, and 2 billion by 2040 [6]. This would cause more carbon dioxide in the world.

To limit global warming to 1.5 °C or at least below 2 °C [7], it is crucial to stop using fossil fuels. China (31%), the United States (14%), the EU27 (7%) and India (7%) contributed the most to global fossil CO₂ emissions in absolute terms in 2020. These four areas are responsible for 59% of global CO₂ emissions. However, the rest of the world contributes 41% which also includes marine bunker fuels and international aviation [5]. The United

States, the European Union, and the United Kingdom aim to achieve net zero emissions by 2050, China and Russia by 2060, and India by 2070 [8]. In addition to climate change, energy security and the future of oil supply pose a significant threat. Car manufacturers are becoming more aware of their involvement in achieving the goals of decarbonizing the economy and reducing oil dependence [3]. Around 97% of the European Union (EU) oil consumption is met by imports, with a quarter of these supplies coming directly from Russia. The European Commission has the plan to eliminate Russian oil, gas, and coal imports by 2027. The road transportation sector uses approximately 60% of the total oil consumption in the EU [9].

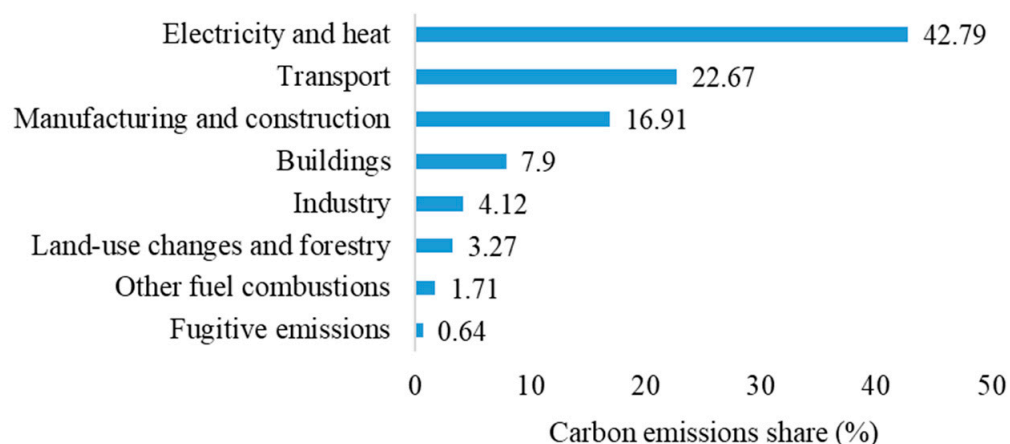


Figure 1. Distribution of CO₂ emissions across different sectors [4].

Electric vehicles (EVs) are one of the practical means of significantly and immediately decarbonizing transportation [10]. In 2021, sales of EVs hit a record high of 6.6 million, doubling from the previous year. Only 120,000 electric cars were sold worldwide in 2012. Each week in 2021, sales exceeded that amount. In 2021, around 10% of global automotive sales were electric, four times the percentage in 2019. This increased the total number of electric cars on the world's roadways to nearly 16.5 million, or three times the 2018 level. Two million electric cars were sold worldwide in the first quarter of 2022, a 75% increase over the same time in 2021. China and Europe accounted for more than 85% of the worldwide sales of EVs in 2021, followed by the United States (10%) [11].

Higher penetration of EVs can bring several benefits in terms of renewable consumption and reducing CO₂ emissions, as discussed in previous paragraphs. In addition, EVs are beneficial for distribution systems, microgrids, and nano grids in several ways. For example, the authors in [12] analyzed different architectures and concluded that AC-DC hybrid architecture is the most suitable for EV integration in micro and nano grids. Similarly, different challenges and enablers for using EVs as a service are discussed in [13]. Different aspects, such as technical, economic, behavioral, and regulatory aspects of integrating EVs with distribution systems are discussed in [14]. Finally, the useability of EVs for providing reliability as a service for different building types is analyzed in [15] and fault estimation methods in [16]. However, with the increased penetration of EVs, several challenges arise—for example, planning and management of power systems considering highly uncertain loads due to EVs [17]. In addition, the driving preferences and patterns of different users are different. This further complicates the management of power system loads and significantly increases the computational burden. One practical solution to deal with this problem is to group EVs and other related aspects of EVs by using different clustering methods. Clustering methods are widely used in different disciplines to arrange and group datasets, and then analysis of representative samples from each cluster can be carried out. These methods can be used to cluster EVs and will eliminate the need to analyze all EVs individually.

There are several studies in the literature on clustering different aspects of EVs to reduce the computational complexity of the network during analysis. Some of the main areas studied in the existing literature include modeling the behavior of the EV user [18], EV driving cycle [19], used EV batteries [15] and clustering [20], and EV charging stations [21]. However, other aspects of EVs also need to be further analyzed using clustering techniques. For example, the impact of EVs on different distribution circuits [22], charging infrastructure during emergencies [23], equity and disparities in rebate allocations [24], and the use of big data with cluster analysis to assist transportation network management [25]. Cluster analysis is a potential solution to reduce the complexity of the network under higher penetration of EVs while preserving the diversity of user behavior and EV traits.

The main objective of this study is to analyze the current literature on the application of clustering methods for various aspects of EVs. In addition, the shortcomings of the existing literature will be identified along with future research directions needed to facilitate rapid analysis of systems under higher penetration of EVs. Therefore, the analysis in this study is divided into three parts. In the first part, different clustering methods are analyzed, which include both hard and soft clustering methods. In addition, different categories of hierarchical and partitional clustering methods are also discussed. Each section is followed by the merits and demerits of different clustering methods. In the second part, the existing literature is analyzed on cluster analysis of different aspects of EVs, specifically the application of clustering methods for modeling the behavior of EV users and EV driving cycle, used EV battery clustering, and EV charging station clustering. Each section is followed by a summary of the methods used in these studies and the major consideration in each study. Finally, in the third part, the shortcomings of existing studies are summarized, and future research directions are presented. Specifically, the need for further research is discussed on the application of cluster analysis to different related fields. These fields include the EV impact on different distribution circuits, charging infrastructure during emergencies, equity and disparities in rebate allocations, and the use of big data with cluster analysis to assist in transportation network management.

2. Clustering Methods

Clustering, or cluster analysis, is an unsupervised learning technique for assigning data into separate groups based on a predetermined set of criteria. It helps users understand the grouping in a data set. Data from the same class are often similar, while data from other classes are typically dissimilar [26]. There are two major types of clustering techniques: crisp (hard) clustering and soft (flexible) clustering. In the case of hard clustering, a data point only belongs to a single cluster, while in the case of fuzzy clustering, each point may belong to two or more groups [27]. An overview of different clustering methods is presented in Figure 2.



Figure 2. An overview of clustering methods [27].

Hard clustering algorithms can be divided into hierarchical algorithms and partitional algorithms. The dataset is split into a single partition in case of partitional algorithm. Contrarily, the dataset is divided into a series of partitions (nested inside one another) in case of hierarchical algorithms [27]. A generalized dendrogram for hierarchical clustering algorithms is shown in Figure 3.

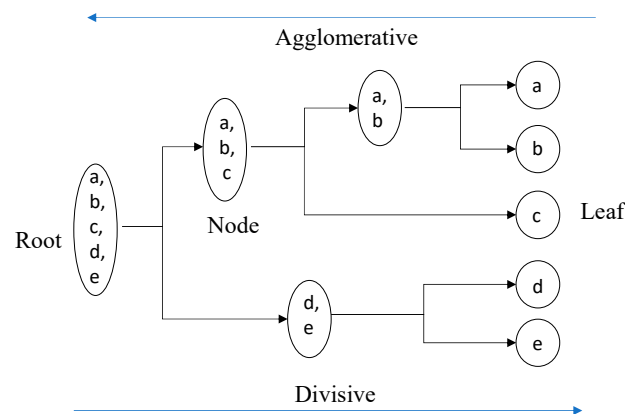


Figure 3. An example dendrogram for hierarchical clustering algorithms [27].

2.1. Hierarchical Clustering Algorithms

Hierarchical algorithms can be categorized as agglomerative and divisive algorithms. A divisive hierarchical algorithm divides data into smaller clusters, while an agglomerative algorithm merges data points into larger clusters from the bottom to the top. Contrarily, partitioning algorithms establish a one-level division of the dataset [28]. Hierarchical clustering is often shown using a dendrogram, a specific tree structure, as shown in Figure 3 [27].

2.1.1. Agglomerative Hierarchical Clustering

In the case of agglomerative hierarchical clustering, each data point starts in its own cluster. Comparable clusters are then merged to form a hierarchy [29]. Agglomerative hierarchical methods can be categorized into graph and geometric methods. Graph methods can be further divided into complete, single, average, and weighted average linkage methods. Similarly, geometric methods include Ward, median, and centroid methods [27]. An overview of hierarchical clustering methods is shown in Figure 4. There are several subcategories of the agglomerative hierarchical clustering algorithms:

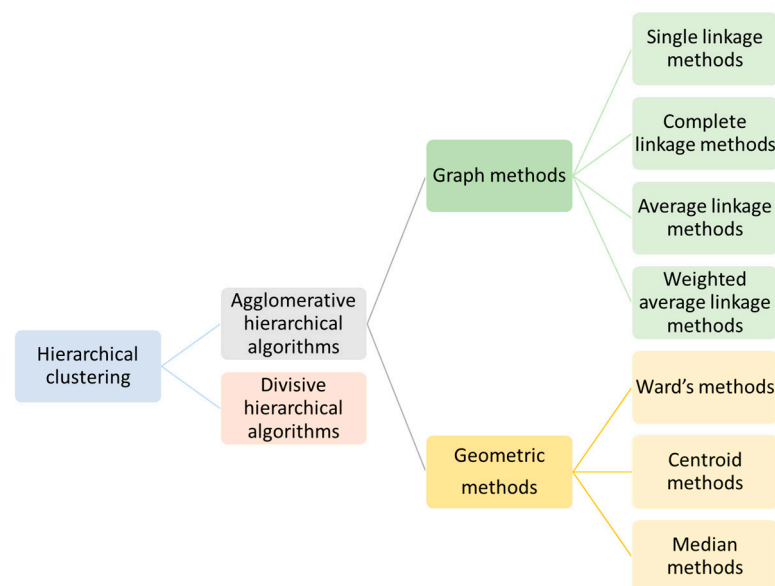


Figure 4. An overview of hierarchical clustering methods [27].

Single-Link Method: Single-link hierarchical clustering, also referred to as nearest-neighbor, is one of the most straightforward methods [27]. The single-linkage criterion is the lowest difference between two objects. The vicinity between two clusters is determined

by the minimum distance between any two objects of each cluster [29]. A single linkage may efficiently cluster non-elliptical elongated-shaped groupings of data objects. A significant disadvantage of this approach is that it is susceptible to noise and outliers in the data set [28,29].

Complete Link Method: This method is also known as the farthest neighbor method and it determines the most prominent dissimilarity between two objects. The maximum distance between any objects that belong to separate clusters defines the proximity of the two clusters [29]. This linkage method considers the structure of the cluster, exhibits non-local behavior, and typically produces clusters with compact shapes [28]. These clusters are more compact than clusters based on the single linkage method [30]. However, this linkage method is also vulnerable to outliers [29].

Group Average Method: The group average method or the unweighted pair group method uses arithmetic averages to determine the mean or median distances among all the objects between clusters [27,30]. Compared to single and complete links, an average linkage method offers the best balance between reducing the variance within the clusters and increasing the variance between clusters [29]. However, one of the main disadvantages of this method is that it is likely that elongated clusters divide and parts of neighboring elongated clusters combine as a result of average link clustering [31].

Weighted Group Average Method: This method is also known as the ‘weighted pair group method’ and it uses arithmetic average. It first constructs a dendrogram that contains information on a similarity matrix. The nearest two clusters are combined into a higher-level cluster at each step. Then, its distance to another cluster is calculated. It is the arithmetic mean of the average distances between members of clusters.

Centroid Method: The centroid method computes the distance between centroids of two clusters. Compared to previous linkage methods, it is more tolerant of outliers and performs better when dealing with clusters of various sizes [29]. Centroid linkage clustering employs only the centroid of the cluster to determine the similarity between two clusters. In contrast, the group average method considers all pairs of datasets to calculate the average pairwise similarity [28].

Median Method: This method is also known as the weighted pair group method and uses centroids or weighted centroids. It was first introduced by Gower in 1967 [28]. Although the median and centroid methods are relatively similar, there is a difference. The centroid of the new group does not depend on the size of the groups that make up that group [32]. The major drawback of this method is that it is not suitable for metrics, as it cannot be interpreted geometrically [28].

Ward’s Method: This method, also known as Ward minimum variance method, was proposed by Ward in 1963 to compute the minimum increase in the within-cluster sum of squares as a result of the merging of two clusters. The objective of the Ward technique is to combine these two clusters into a group with minimal variations [33].

The advantages and disadvantages of different agglomerative hierarchical methods are listed in Table 1.

Table 1. Advantages and disadvantages of agglomerative hierarchical clustering methods.

Method	Advantages	Disadvantages	Ref.
Single-link Method	Can differentiate between non-elliptical shapes as long as the gap between the two clusters is not small.	Susceptible to noise and outliers in the dataset.	[27]
Complete-link method	Provides well-separated clusters even if there is some noise present between clusters.	Biased towards globular clusters and tends to break large clusters.	[29]

Table 1. Cont.

Method	Advantages	Disadvantages	Ref.
Group average method	Offers the best balance of reducing within-cluster variance and increasing between-cluster variance.	Due to average-link clustering, likely for elongated clusters to get divided and for parts of neighboring elongated clusters to get combined.	[30]
Weighted group average method	Unbiasedness towards middle value and unaffected by outliers or extreme values.	Difficult to understand when the number of observations increases.	[29]
Centroid method	More tolerant to outliers and performs better when dealing with clusters of various sizes.	Updates may cause large changes throughout the cluster hierarchy.	[29]
Median method	The new group's centroid is independent of the size of the groups that make up that group.	Not suitable for metrics since it cannot be interpreted geometrically.	[32]
Ward's method	Good at recovering cluster structure and yields unique and exact hierarchy.	Sensitive to outliers and poor at recovering elongated clusters.	[33]

2.1.2. Divisive Hierarchical Clustering

Another variant of hierarchical clustering is a top-down approach known as divisive hierarchical clustering [34]. At the beginning, all items belong to the same, single cluster. The cluster is then split into sub-clusters and subdivided into still smaller sub-clusters. This procedure is repeated until the appropriate cluster structure is achieved [31]. There are two types of divisive clustering: monothetic and polythetic methods. Unlike the monothetic technique, which is focused on a single feature, polythetic approaches consider the values of all characteristics within a data set [27]. To highlight the similarities between the two instances, polythetic divisive clustering considers all elements concurrently. When many variables are present, scalability concerns may arise. The best results will be achieved with monothetic clustering when the focus is on a single character throughout the time [30].

2.2. Partitional Clustering Algorithms

Partitional clustering techniques partition the data set into a defined number of clusters without any hierarchical structure [35]. The benefits of hierarchical algorithms are the drawbacks of partitional algorithms and vice versa. Partitional clustering approaches are more prevalent in pattern recognition than hierarchical algorithms [36]. They are advantageous when constructing a dendrogram would be computationally prohibitive for an application requiring an extensive data set. Figure 5 shows the clustering pattern of the partitional clustering method for 145 data points into four clusters [30]. However, in general, selecting the number of desired output clusters is challenging using a partitional method [35]. An overview of different partitional clustering methods is shown in Figure 6. The following sections describe several partitioning approaches.

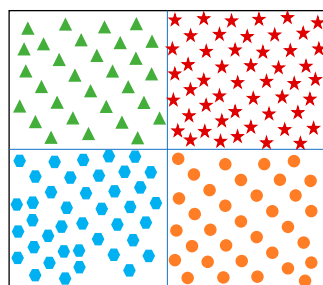


Figure 5. A partition with $n = 154$ and $k = 4$ [35].

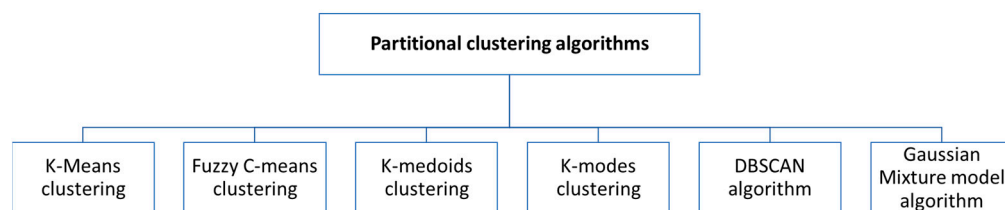


Figure 6. An overview of partitional clustering methods [30].

2.2.1. K-Means Clustering

The K-means clustering technique is the most widely used partitional clustering algorithm [32,33]. The K-means clustering technique was first proposed by Steinhaus in 1956 and has since been used in many domains, including psychology, marketing research, medicine, and biology [29]. The fundamental objective of this approach is to split an n -dimensional dataset into k clusters such that the sum of squares inside each partition is as low as possible. K-means generates a flatter grouping structure than hierarchical methods. The Euclidean distance is the most common distance metric used to determine the similarity between two objects. There must be at least one item in each k group partitioned by the partitioning algorithm [36].

Despite its popularity, there are some limitations to K-means clustering [37]. For example, there is no efficient and universal approach to determine the initial partitions and the number of clusters. In addition, the K-means algorithm is susceptible to noise and outliers. Even if an item is far from the cluster's center, it is nevertheless compelled to join the cluster, distorting its structure [38,39].

2.2.2. Fuzzy C-Means Clustering

J.C. Dunn developed fuzzy c-means (FCM) clustering in 1973, and J. C. Bezdek improved it in 1981 [40,41]. FCM is an unsupervised clustering algorithm [42] in which a single data point may belong to two or more clusters [43,44].

FCM can be used to solve various feature analysis, clustering, and classifier construction issues. It has been widely used in diverse fields [42]. When compared with K-means, FCM allocates each pattern with some degree of membership to a cluster, i.e., it yields a fuzzy clustering. When there are some overlaps between clusters in the data set, it is more appropriate for real-world applications than K-means.

2.2.3. K-Medoids Clustering

K-medoid also seeks to minimize the sum of squared error (SSE) [31]. One of the cluster points is used to represent it in k-medoids approaches. The objective function is the averaged distance or another measure of dissimilarity between a point and its medoid when medoids are chosen. Clusters are subsets of points near their corresponding medoids [38].

This method is quite similar to the K-means algorithm. The K-medoids approach, like K-means, aims to discover a clustering solution that minimizes a given objective function. Like the K-means clustering technique, the K-medoids algorithm iterates until each representative data point becomes the cluster medoid [29]. Since the placement of most of the points within a cluster determines the choice of medoids, it is less vulnerable to outliers. Therefore, the K-medoids approach is more robust for noise and outliers as compared to the K-means algorithm. However, compared to the K-means approach, it is computationally more expensive [31,38].

2.2.4. K-Modes Clustering

Huang (1997) proposed the K-modes clustering algorithm for categorical data by presenting a new dissimilarity metric. The K-modes algorithm is an improved version of the K-means algorithm. Due to the improvements to the K-means method, the K-modes algorithm can cluster very large categorical data sets from real-world databases effectively [45,46]. Another benefit of the K-modes technique is that the modes provide

distinctive cluster descriptions. These descriptions are crucial to the user's ability to comprehend clustering findings. The K-modes method is faster than the K-means algorithm because it requires fewer iterations to achieve convergence [46].

The K-modes method employs the same clustering procedure as the K-means algorithm, except for the clustering cost function, which has the same limitations. The K-modes algorithm has several additional shortcomings. For example, inability to detect the number of clusters, inability to converge to the global optimum, and prone to outliers [4,47].

2.2.5. DBSCAN Algorithm

Ester proposed the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, a density-based clustering algorithm to discover arbitrarily shaped clusters in 1996 [27]. Clusters are determined using the DBSCAN technique by examining the point density. The presence of clusters is indicated by the density of points. Similarly, regions with a low density of points represent noise clusters or outlier clusters. This technique is well-suited for dealing with large datasets that include noise. In addition, it can distinguish clusters of different sizes and forms.

The essential concept of the DBSCAN algorithm is that, for each point in a cluster, the neighborhood of a specific radius must have a minimum number of points, i.e., the density in the neighborhood must surpass a set threshold [34].

2.2.6. Gaussian Mixture Model Algorithm

Gaussian Mixture Model (GMM) is a probabilistic model that indicates the existence of subclusters in every observation. Mixture models are used to identify subcluster characteristics. For developing mixed models, approaches such as unsupervised learning and clustering are used. These, however, do not apply to all feature extraction processes. Combinational models may be assumed for mixture models. In combinational models, the members of the cluster are specified arbitrarily, while the total size of the clusters in mixture models is fixed at 1 [48]. GMM is an estimation method for probability density distributions [49]. GMM may be seen as an extension of the Vector Quantization (VQ) model. The clusters in this model overlap. A feature vector is not allocated to the cluster that is closest to it. Nonetheless, the probability value determined from cluster observations is not zero [50,51].

The expectation-maximization (EM) technique for Gaussian mixtures and the K-means algorithm are comparable in many ways [52]. Instead of assigning each data point to a single cluster in a rigid way, as the K-means algorithm does, the EM method assigns data points based on posterior probabilities. The K-means method can be derived from the EM for Gaussian mixtures as a specific limit [53].

The advantages and disadvantages of different partitional clustering methods are listed in Table 2.

Table 2. Advantages and disadvantages of partitional clustering methods.

Method	Advantages	Disadvantages	Ref.
K-means clustering	Most widely used method; it generates a flatter grouping structure than hierarchical methods.	No universal approach for determining the initial partitions and the number of clusters; susceptibility to noise and outliers.	[39]
Fuzzy C-means clustering	More appropriate for datasets with some overlaps between clusters.	Poor performance for clusters with unequal sizes/densities and sensitive to noise and outliers.	[44]
K-medoids clustering	Less vulnerable to outliers; therefore, it is more resilient than the K-means algorithm in the face of noise and outliers.	Compared to the K-means approach, it is more computationally expensive.	[28]

Table 2. Cont.

Method	Advantages	Disadvantages	Ref.
K-modes clustering	Better cluster very large real-world datasets in an effective manner; it provides distinctive cluster descriptions.	It cannot detect the number of clusters, converge to global optimum (owing to the random selection of initial seed artifacts), distinguish clustering tendency, recognize outliers, etc.	[46]
DBSCAN algorithm	Well suited for dealing with large datasets that include noise; can identify clusters of various sizes and forms.	Fails to identify clusters of density if the dataset is too sparse.	[34]
Gaussian mixture model algorithm	Less sensitive to scale; provide estimates of the probability that each data point belongs to each cluster.	Difficult to incorporate categorical variables; struggle with numeric variables that are not normally distributed.	[52]

3. Application of Clustering in Electric Vehicles

Cluster analysis can be applied to different aspects of EVs. Common examples include EV user behavior, EV driving cycle, and EV battery charging. In addition, clustering can also be applied to group EV charging stations and to analyze the impacts of EVs on power distribution systems. It should be noted that the application of clustering analysis for EV battery charging and impacts on distribution systems are specific to EVs. However, user behavior and driving cycle analysis are common for both EVs and internal combustion engine vehicles (ICEVs). Although the same clustering methods can be applied for both EVs and ICEVs, the driving behavior of EV owners and ICEV owners has major differences mainly due to the difference in charging/fueling mechanisms. Similarly, the driving cycle of EVs and non-EVs is different, as discussed in [54]. Therefore, the outcome of the clustering method could be different. The following sections cover these aspects of EV clustering in detail.

3.1. EV User Behavior Clustering

To ensure the reliability of the power supply, it is necessary to anticipate the EV's behavior in advance. However, the activity of individual EVs is very unpredictable, and their daily behavior patterns can vary considerably. This makes it challenging to create a model that simultaneously predicts the actions of all EVs operating in a system or area. To solve this problem, the results of the cluster analysis can be used to model and forecast the behavior of a group of similar EVs. The collection of similar EV activities is expected to minimize unpredictability and improve the behavior prediction accuracy [55]. The following subsections summarize different studies conducted on EV user behavior clustering.

3.1.1. K-Means Algorithm

The K-means clustering algorithm has been commonly used to cluster the behavior of EV users due to its simplicity and many other advantages [56–58]. For example, Hu et al. [18] used the K-means and DBSCAN clustering algorithms to classify EV consumers. They categorize 7426 EV users into six classes, which includes lost users, possible users, new users, key users to develop, key users to sustain, and high-value users. An overview of the proposed method [18] is shown in Figure 7. The suggested technique was compared with the standard clustering algorithm and fuzzy c-means method, showing that the new method is more robust than the other approaches. Similarly, Xiong et al. [59] proposed a new method that integrates K-means clustering with multilayer perceptron. First, historical charging data are processed using K-means clustering to establish assumptions about EV user behavior for EV charging schedules. Then, a multilayer perceptron is used to analyze the EV user charging record data and generate classifications based on clustering labels from the K-means algorithm and manual labeling through data visualization. The

suggested technique automates the labeling of the data sets. In addition, it is not required to perform clustering when a new user connects to the charging network. After training, the method may be used concurrently with real-time control.

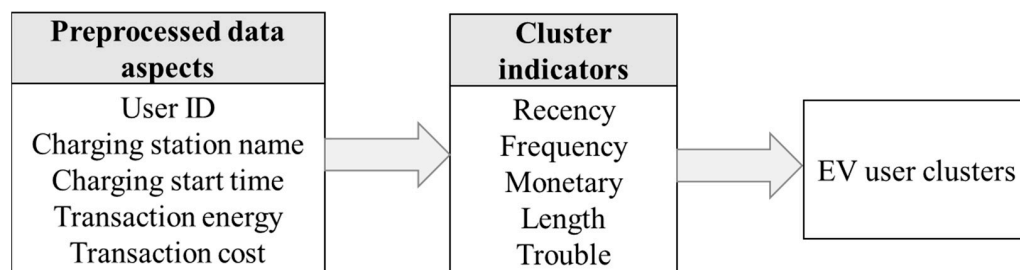


Figure 7. Overview of K-means based EV clustering method proposed in [18].

3.1.2. DBSCAN Algorithm

The DBSCAN algorithm has also been used by a number of studies to cluster EVs based on user behavior. For example, Fan et al. [60] clustered EV users by combining the coefficient matrix with density clustering. The grouping comprises clustering based on user preferences and clustering based on item similarity. User preference clustering refers to grouping based on trust between users. It first computes the correlation coefficient matrix between users and then performs density clustering based on the similarity matrix. The premise of density clustering is that samples of the same category are strongly connected; that is, samples of the same category must be close to each sample in this category. A cluster category is generated by grouping closely related samples into one category. When all samples are divided into distinct groups, the findings of all clustering groups can be collected. Item similarity clustering also refers to the setup of similarity clustering for new EVs. Based on the findings of the user clustering, the score of the new EV configurations in each user group is determined by item similarity. This way, the degree of preference of each type of user for various configurations can be easily understood. An overview of the approach used in [60] for EV clustering is shown in Figure 8.

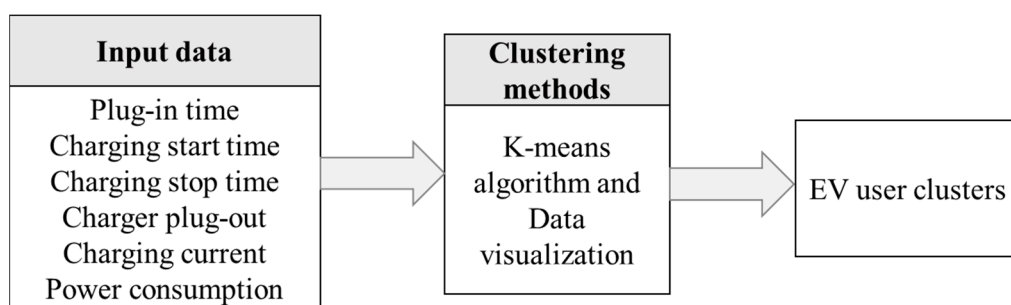


Figure 8. Overview of DBSCAN-based EV clustering method proposed in [60].

3.1.3. Hybrid Methods

Hybrid methods can provide better results due to their ability to overcome the demerits of different algorithms they combine. Therefore, several researchers have combined different methods for EV clustering. For example, Helmus et al. [61] classified the charging behavior of EV users using a two-stage clustering technique. First, a Gaussian mixture model is used to cluster charging sessions, revealing 13 unique charging session categories (including seven types of daylight charging sessions and six types of nocturnal charging sessions). The Partition Around Medoids method yields nine user classes based on their separate portfolio of charging session types. Three types of daytime charging users, three types of night-time charging types, and three types of irregular users. An overview of the hybrid method (Gaussian mixture model and Partition Around Medoids) proposed in [61] is shown in Figure 9.

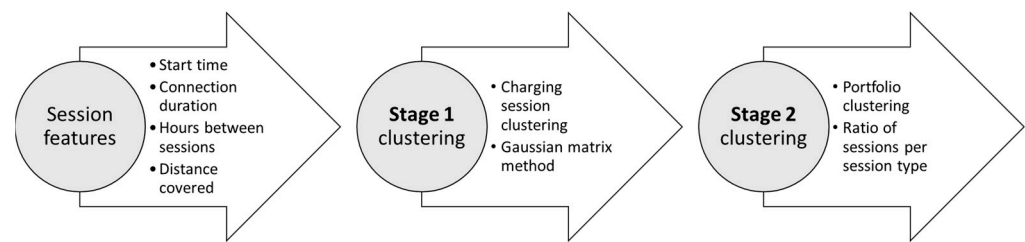


Figure 9. Overview of hybrid method for EV clustering proposed in [61].

3.1.4. Other Methods

There have been other clustering methods used to group EVs according to user behavior. For example, Powell et al. [62] used agglomerative clustering to classify drivers in ascending order. Each driver is assigned to its cluster at initialization, and the method joins two clusters at each step. The selection of clusters to merge is based on minimizing the increase in the sum of squares within the cluster. Similarly, Campbell et al. [3] used the Wards cluster analysis approach to census data (based on age, income, automobile ownership, property ownership, socioeconomic status, and education) to discover prospective drivers of alternative fuel vehicles in Birmingham, United Kingdom. The sum of squares (distance) between an item in the first cluster and an object in the second cluster is calculated using Ward's approach and then totaled over all variables. This strategy maximizes the formation of clusters of roughly equal proportions. An overview of the clustering approach used in [3] is shown in Figure 10.

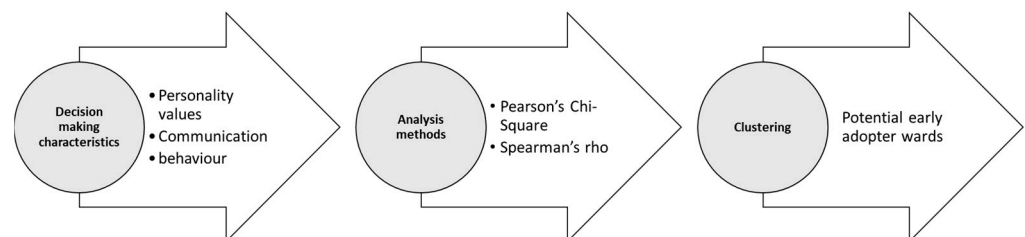


Figure 10. Overview of Ward's method for EV clustering proposed in [3].

Finally, a summary of the different clustering methods used for EV clustering based on user behavior is shown in Table 3.

Table 3. Summary of clustering methods along with their objectives for user behavior-based clustering of EVs.

Method Used	Clustering Objective	Ref.
K-means and DBSCAN	Classify EV consumers to improve profitability and user loyalty	[18]
K-means clustering with multilayer perceptron	EV user behavior for charging schedules	[59]
Coefficient matrix with density clustering	Grouping based on the trust among users	[60]
Gaussian Mixture Model	User classes based on their separate charging session portfolios	[61]
Agglomerative clustering	To estimate EV charging load for long-term planning	[62]
Wards cluster analysis	Find prospective drivers of alternative fuel vehicles	[3]

3.2. EV Driving Cycle Clustering

Driving cycles are speed-time profiles representing real-world driving conditions in a particular city or region [63]. They can be used during laboratory chassis dynamometer

simulation tests and in automotive simulation research to evaluate fuel consumption and exhaust emissions. In addition, driving cycles can be used to monitor energy consumption and estimate the driving range of EVs. Moreover, driving cycles are essential for realistic life cycle studies and evaluating the impacts of EVs on the power system [63,64]. Therefore, several researchers have clustered EV driving cycles using different methods to reduce the computational burden of the analysis. An overview of these studies is presented in the following sections.

3.2.1. Hard Clustering Approaches

In hard clustering, a data point only belongs to a single cluster. Several researchers have used hard clustering methods to analyze the EV driving cycle. For example, EV driving cycles are developed by Berzi et al. [19] where driving sequence analysis is performed to group different EV clusters. Similarly, different microsegment parameters are estimated by Brady and O'Mahony [63], and driving cycle synthesis is carried out using data segmentation and classification techniques. An overview of the clustering process proposed in [63] is shown in Figure 11. Fotouhi and Montazeri-Gh [64] used K-means clustering to group vehicles into four clusters considering two driving features, such as the average speed and idle time percentage. K-means clustering was used by Yuhui et al. [65] to design a target driving cycle using six characteristic parameters. Driving time and instantaneous velocity are used by Zhou et al. [66] and Chen and Xiong [66] to develop the driving cycle of EVs using K-means clustering. Principal component analysis and K-means clustering are used by Zhang et al. [67] for driving cycle estimation of special-purpose EVs. The K-means algorithm can also be used for cluster analysis of EV driving cycles. Zhao et al. [68,69] classified driving segments using a hybrid classification technique combining K-means and support vector machine (SVM). In [69], the SVM model training sets comprised the top 10% of optimal driving segments from the K-means clustering results.

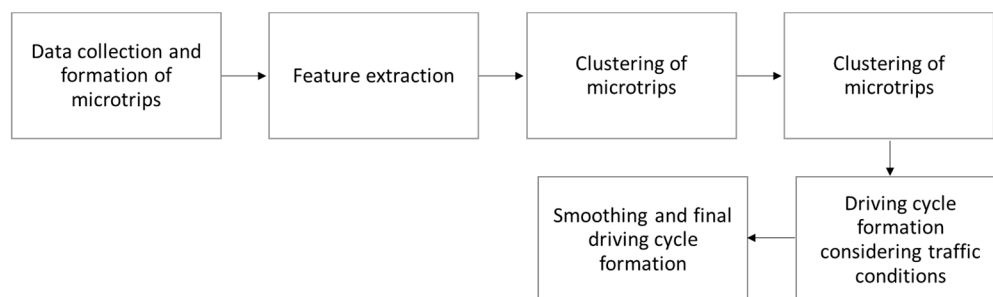


Figure 11. Overview of EV clustering based on microsegmentation proposed in [63].

The K-means clustering method employs the Euclidean distance between the driving cycles and the cluster center as a classification metric and offers the benefits of efficiency and simplicity. However, it is a hard clustering approach; when driving segment clustering includes numerous classes or the distance between cluster centers is short, the clustering impact is weak and it is possible to slip into a local optimum that cannot be clustered incrementally. Therefore, other researchers have used soft clustering methods to conduct studies summarized in the following subsection.

3.2.2. Soft Clustering Approaches

In soft clustering, data points can belong to more than one cluster. In contrast to the K-means method, the fuzzy C-means clustering algorithm calculates the degree to which each sample point is similar to each class; this value is referred to as the class membership degree. Zhao et al. [70] used this method to cluster the driving segments. In this scheme, a membership degree matrix reflects the likelihood that samples correspond to specific classes. Therefore, the hard clustering of the K-means algorithm is changed into fuzzy clustering based on soft membership, which has a greater chance of achieving

global optimality. The fundamental concept of the FCM method is to iteratively search for the membership degree matrix and clustering centers to achieve the minimum value of the objective function. An overview of the representative cycle selection based on soft clustering is presented in Figure 12.

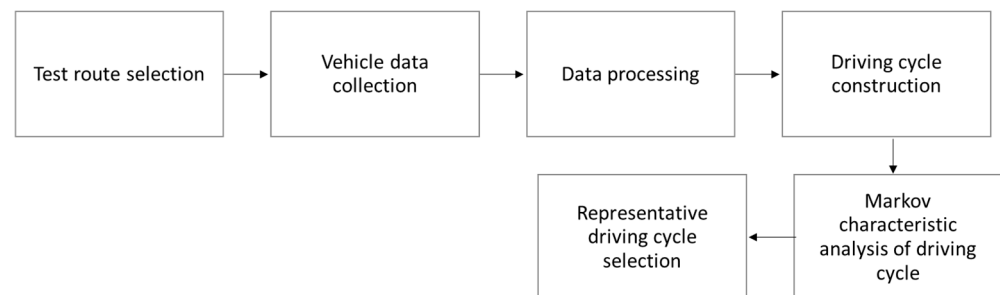


Figure 12. Overview of EV clustering based on soft clustering proposed in [70].

3.2.3. Other Approaches

Apart from hard and soft clustering approaches, other methods have also been used for estimating the driving cycles of EVs. For example, Chen et al. [71] classified various driving cycles into six distinct types of driving cycle using the K-Shape clustering technique, a new algorithm that maintains the forms of the driving cycle data. This algorithm is suitable for clustering time series; therefore, it is not essential to extract features describing driving cycle characteristics. The authors compared this new approach to the common K-means algorithm for grouping driving cycles and concluded that the K-Shape approach performs better.

Due to the absence of detailed methodology for driving cycle analysis in [71], Wang et al. [72] proposed a “dimension reduction, clustering”-based driving cycle construction method that uses an advanced machine learning method for the offline solution of the SDP problem. It is based on the identification of driving conditions and its objective is to minimize the number of driving cycle characteristics while preserving the travel information included in the driving cycle data. An overview of the offline training and online testing method is shown in Figure 13.



Figure 13. Overview of offline and online method for EV clustering proposed in [72].

Finally, a summary of the different clustering methods used for the clustering of EVs for driving cycle analysis is shown in Table 4.

Table 4. Summary of clustering methods along with their objectives for driving cycle analysis of EVs.

Method Used	Clustering Objective	Ref.
K-means	Group different EV clusters for driving sequence analysis	[19]
	To devise a target driving cycle	[65]
	To develop the driving cycle of EVs	[66]
K-means and support vector machine	To perform clustering even with numerous classes or with a short distance between cluster centers	[69]

Table 4. Cont.

Method Used	Clustering Objective	Ref.
Fuzzy C-means clustering	To cluster the driving segments	[70]
K-Shape clustering technique	To choose driving cycle characteristics	[71]
Dimension reduction clustering	To reduce the driving cycle features required for clustering while preserving the travel information	[72]

3.3. EV Battery Clustering

Lithium ion batteries (LIBs) are commonly used in EVs because of their benefits, such as extended service life, high safety, and substantial specific energy. However, with usage, their capacity will degrade. As soon as the capacity drops to 80% of its original rated capacity, LIBs will meet the criteria for being retired from EVs. There will be an immediate need to address the optimal usage of decommissioned power LIBs due to the fast growth of the number of EVs. It is expected that more than 12 million tons of LIBs will be retired by 2030. There has been a surge of attention to the retirement of power LIBs in electric cars to reduce resource waste and pollution. Clustering and regrouping large-scale decommissioned LIBs are now the most important ways to achieve optimal use of the echelon usage [73].

3.3.1. Fuzzy Clustering Methods

Several studies [15,44,68,69] used fuzzy clustering methods to classify LIBs. Hu and Sun [20] proposed a new model to evaluate the state of charge (SOC) of lithium ion batteries used in EVs. The fuzzy c-means and subtractive clustering combination approach are used to perform fuzzy partitioning of data vectors, including the temperature, load voltage, and current of the lithium-ion battery pack under the urban dynamometer driving schedule. Then, the multi-model support vector regression (SVR) approach was used to estimate the SOC of a lithium-ion battery pack. The synthesized model was evaluated using 2000 training data and 3500 validation data. Simulation results indicate that the mean validation error of the fuzzy clustering-based multimode SVR technique is less than that of the conventional SVR model.

In addition, using machine learning, Hu et al. [74] created a state-of-charge indicator for LIB modules used in EVs. To identify the model's topology and antecedent parameters, a novel fuzzy C-means, they use a clustering technique based on a genetic algorithm. This reduced the risk of being trapped in local minima, and the number of fuzzy clusters was then estimated using a fast one-pass algorithm called the subtractive clustering algorithm. The second stage uses the backpropagation learning technique to improve the model's antecedent and consequent parameters.

Similarly, Tian et al. [49] clustered batteries using an enhanced fuzzy clustering approach based on a genetic algorithm. In addition, they used the Kernel Function (KF) to optimize the clustering center. The KF turned the samples of the original space into the feature space. Samples in the feature space were separated to obtain the best partition of the original space, enhancing the efficiency of clustering. Finally, nine months of data from EVs was compiled to verify the suggested algorithms. The simulation results demonstrated that the proposed technique clusters batteries more effectively. An overview of the battery clustering method proposed in [49] is shown in Figure 14. The fuzzy c-means clustering algorithm was also used by Wang et al. [75] to estimate the state of function (SOF) of the power LIBs.

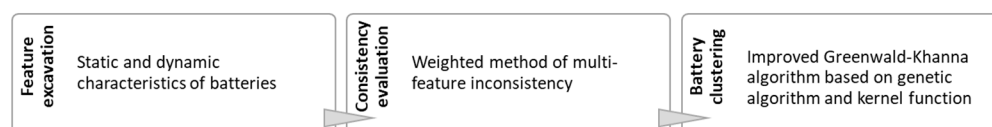


Figure 14. Overview of used battery clustering method proposed in [49].

3.3.2. Support Vector Machine-Based Methods

SVM-based methods have also been used for grouping of used EV batteries. For example, Li et al. [73] developed an SVN-based approach for clustering and regrouping retired LIBs. Preliminary screening (based on battery capacity, internal resistance, and remaining useful life) was used to eliminate batteries with no echelon usage value. On the basis of the SVC, they developed an equal-number clustering technique. Using a publicly available validation data set, the proposed method correctly split 60 batteries into four even clusters. The proposed algorithm was compared with K-means and Gaussian mixture models clustering methods, and the results indicated that the equal-number SVC technique is very promising. An overview of the EV battery clustering method proposed in [67] is shown in Figure 15.

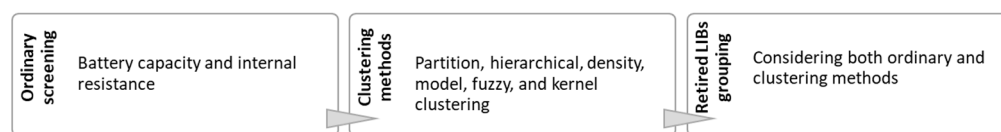


Figure 15. Overview of EV battery clustering based on capacity and internal resistance proposed in [67].

Li et al. [26] used K-means and SVC methods to group battery cells with similar performance to construct battery modules with improved electrochemical performance. The results of the cluster analysis were experimentally validated by monitoring the cell temperature increase during a specified period in an air-conditioned environment.

3.3.3. Other Methods

Apart from fuzzy clustering and SVM-based methods, other clustering methods have also been used for grouping used EV batteries. For example, Liu [76] used a modified K-means clustering algorithm to classify EVs with different battery states of charge and different average daily vehicle travel (AVDT). The principle of this battery clustering method is shown in Figure 16.

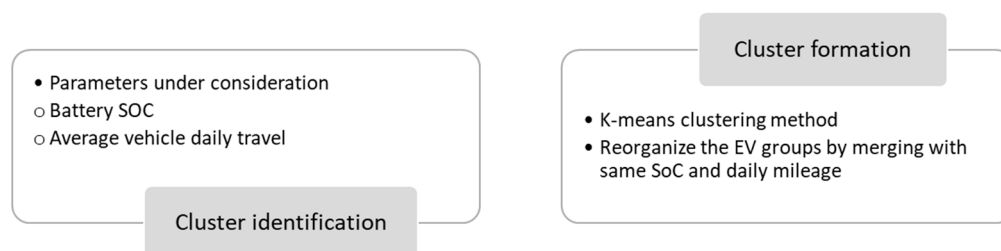


Figure 16. Overview of EV battery clustering based on SOC and daily mileage [76].

Similarly, Xu et al. [77] introduced a new clustering approach for retired batteries based on traversal optimization to reduce computation time and increase clustering accuracy. This approach does not need predefined cluster numbers and centers, and the clustering outcome is independent of outliers. In addition to avoiding repeated computation, this approach completes clustering by visiting all target locations. This way, the optimization process is not iterative and scales well to large sample sets. Compared to existing clustering methods, the new algorithm generates partitions with high disparity between clusters and the lowest differences between points within clusters.

A summary of different clustering methods used for EV battery clustering is shown in Table 5.

Table 5. Summary of clustering methods along with their objectives for EV battery grouping.

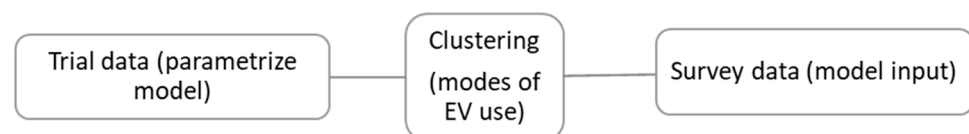
Method Used	Clustering Objective	Ref.
Fuzzy c-means and subtractive clustering	SOC estimation of different EV clusters	[20]
Fuzzy C-means based on a genetic algorithm	SOC indicator for lithium-ion battery modules used in EVs	[74]
Innovative equal-number support vector clustering	Clustering and regrouping retired LIBs	[73]
K-means and support vector clustering	Cluster battery cells with similar performance to construct battery modules with improved performance	[26]
Modified K-means clustering	Classify EVs based on battery SOC and different average vehicle daily travel	[76]
A novel clustering approach	Grouping retired batteries based on traversal optimization	[77]

3.4. EV Charging Station Clustering

As EV ownership expands, the number of charging stations also increases. The construction of a charging station requires a significant investment. Only with an optimum placement can charging stations save a substantial amount of money, offer users convenience, and increase their operational efficiency. Therefore, it is crucial to also include relevant studies on this topic [44]. Clustering can eliminate the need for analysis of individual charging stations by grouping stations with similar profiles together. Several studies on clustering EV charging stations are analyzed in the following sections.

3.4.1. K-Means Clustering

K-means is the most widely used clustering technique in general and has also been used for charging station clustering. For example, Sánchez et al. [21] developed a clustering technique based on the K-means algorithm to partition consumers into small zones and identify potential locations of EV charging station. Hence, each centroid of the partition indicates a possible location for a charging station, while each cluster represents a customer's region. An overview of the clustering method used in [21] for charging station clustering is shown in Figure 17. Similarly, Chen et al. [78] employed the K-means clustering technique to compute the number of charging stations for EVs and their locations.

**Figure 17.** Overview of charging station clustering method proposed in [21].

3.4.2. Hierarchical Clustering

Hierarchical clustering has also been used for grouping EV charging stations. For example, Zhang et al. [48] used a hierarchical clustering method and a quadratic division based on K-means to group the charging demand location for EVs. Similarly, Catalbas et al. [50] estimated the optimal charging station locations of EVs for Ankara using various clustering approaches such as spectral clustering and the Gaussian Mixture Model. Ip et al. [79] implemented a two-step framework. First, road traffic data, such as traffic flows, were converted into data points. Then, an agglomerative hierarchical approach was used for the data points to produce different levels of clusters. The stations were then assigned to demand clusters using linear programming for optimization purposes. An overview of the proposed method for demand estimation based on charging station clustering is shown in Figure 18.

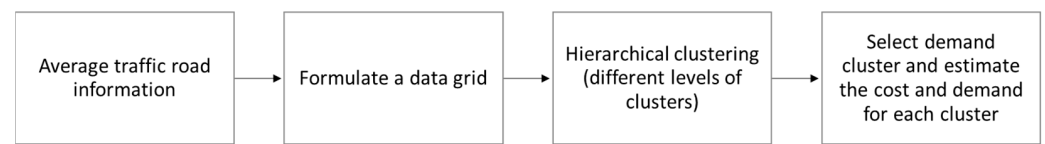


Figure 18. Overview of demand estimation based on charging station clustering [79].

3.4.3. Other Clustering Methods

Apart from K-means and hierarchical algorithms, several other methods have also been used for charging station clustering. For example, Momtazpour et al. [80] used coordinated clustering algorithms to find a collection of places that are optimal candidates for charging stations. Shi and Zheng [44] used the fuzzy C-means clustering method to investigate the optimal location of charging stations. The first stage is to collect charging information from an urban region and then measure the charging requirement areas into separate data points over a control grid. Finally, the fuzzy C-means clustering algorithm is used to group spatial data points into clusters in which the data points are similar. An overview of the coordinated charging scheme proposed in [44] based on charging station clustering is shown in Figure 19. Finally, a summary of different clustering methods used for EV charging station clustering is shown in Table 6.

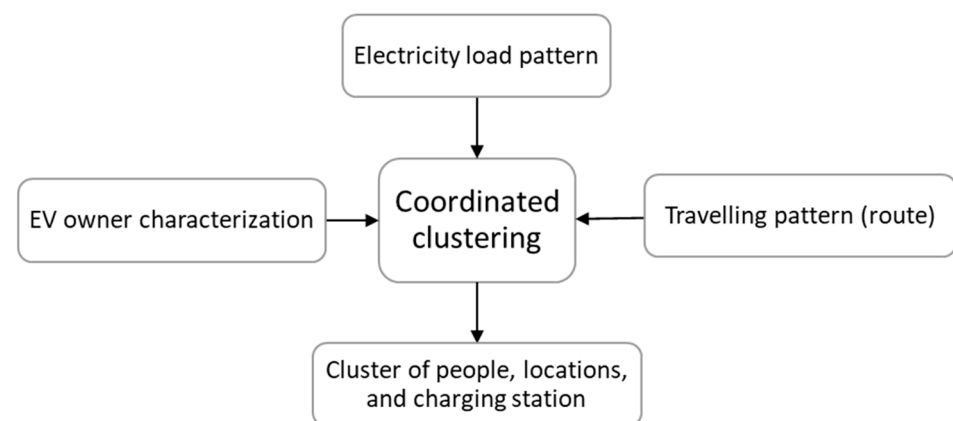


Figure 19. EV coordinated charging scheme based on charging station clustering [44].

Table 6. Summary of clustering methods along with their objectives for EV battery grouping.

Method Used	Clustering Objective	Ref.
K-means algorithm	Finding prospective recharging station locations Number of charging stations for EVs and the location of charging stations	[21] [78]
Spectral clustering and the Gaussian Mixture Model	Optimal charging station locations for EVs	[50]
Agglomerative hierarchical approach	Different levels of clusters for charging stations	[79]
Fuzzy C-means clustering method	Optimal location of charging stations	[44]
Coordinated clustering algorithms	Optimal candidates for charging stations	[80]

3.5. Summary of Selected Studies in Each Category

Four major aspects of EV clustering are discussed in the previous sections, which includes EV user behavior, driving cycle, EV battery, and charging stations. A few representative papers are selected from each category, and their advantages and drawbacks are summarized in Table 7. The drawbacks in each study are open research questions in each category and are also future research directions for researchers in these areas.

Table 7. Summary of representative papers from each EV clustering category.

EV Clustering Aspect	Advantage/Major Consideration	Drawback/Room for Further Research	Ref.
User behavior	Ability to provide user value assessment and validated based on big data	Further research is required for determining optimal weights for weighted sum method	[18]
	Granulated data are used, and users are categorized into several groups	Data of only public charging stations are used	[61]
Driving cycle	Consideration of uncertainty and scalability of the proposed model	Inclusion of holidays and irregular behavior datasets is required	[62]
	Building of synthetic driving cycles using available data	Comparison with ICEVs is required for better understanding	[19]
	Both on-board measurement and chase car methods are used	EV driving data of a specific small region are used	[70]
EV battery	Consideration of multiple clustering parameters	Economic analysis for large-scale battery packs is required	[73]
	Consideration of EV battery size and daily mileage for clustering	Further design consideration for EVs in the same class is required	[76]
	Combination of numerical and experimental methods	Large number of steps (lengthy process) before designing	[26]
Charging station	Scalable method for including future growth scenarios of EVs	Feeder and generation side limitations are not considered	[80]
	Consideration of siting of charging stations	Deterministic information of EVs is considered	[21]

4. Shortcomings in Existing Studies and Future Research Directions

As described in the previous sections, there have been a number of studies conducted on clustering EVs considering their different aspects. However, more research is required to reduce the computational burden for detailed analysis with a higher penetration of EVs. The following are some of the areas that deserve further attention, as there is limited or no research available in the current literature.

4.1. Impact of EVs on the Distribution System

With a higher percentage of EVs, the load profiles of the distribution circuits are expected to change significantly. In addition, the loading profiles of different circuits also change due to the different penetration levels of EVs in different localities. Therefore, the grouping of distribution circuits is required to reduce the computational burden during analysis. Distribution circuits can be grouped into several clusters depending on their characteristics; circuits within the same subset will have similar load profiles, while circuits from separate subsets will have different profiles. This approach decreases the variety of circuit attributes in each subset and provides a more accurate description of the features using a typical single circuit. Thus, distinct circuit groups define different system features, and a typical circuit depicts each group of related circuits.

Only a few studies have been conducted on this topic. For example, Xu et al. [52] proposed a plug-in EV impact assessment framework that uses a K-medoid clustering algorithm to select a small number of representative circuits from thousands of distribution circuits and conducts the impact study using Monte Carlo simulation in the representative circuits. The impact at the feeder level is then extrapolated to the system level. An overview of the proposed method is shown in Figure 20. Similarly, to assess the impact of electric cars on the electric power distribution system, Dow et al. [81] clustered the entire set of utility feeders using the K-medoids technique. With K clusters, each feeder in the data set is grouped into one of the clusters. However, more research is required in this area to facilitate the analysis of the distribution circuits with higher penetration of EVs.

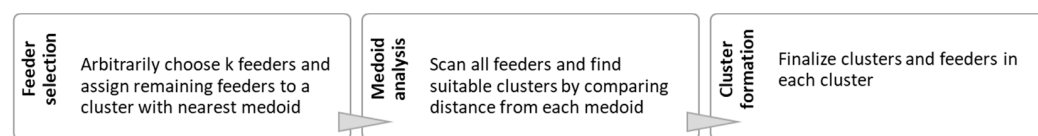


Figure 20. Overview of distribution circuit clustering method with EVs [52].

4.2. Charging Stations for Emergency Response

The intensity and frequency of natural disasters and man-made events are increasing due to climate change and increased penetration of information and the communication technology (ICT) in the power sector [23]. In the electrified transportation era, EVs will be used for emergency response and evacuation as well. Therefore, appropriate charging infrastructure is required in different locations to cope with emergencies. Cluster analysis can potentially be applied in this area as well. Specifically, grouping and clustering of different localities are required considering their ability to respond to a large-scale outage. This will help policymakers identify areas with inadequate charging infrastructure so that they can be prioritized for future development. More research is required in this area, especially considering the application of different cluster analysis techniques discussed in this paper.

4.3. Disparities and Equity in Rebate Allocations

To increase the adoption of EVs, governments around the world have introduced different rebates and incentive programs. Proper allocation of rebate programs is required to maximize their benefit, and to ensure equity and reduce disparity in different localities. For example, a study conducted in California [24] has revealed that the initial EV rebate programs were more focused on high-income groups. This study also noted that the share of rebate programs for low-income/disadvantaged communities increased later when an income cap policy was put into effect. Cluster analysis can also be applied to identify different groups and target them to make EVs affordable to all, especially to low-income groups and disadvantaged communities. A very limited number of studies have been conducted on this topic, especially with the consideration of cluster analysis.

4.4. Model-Free Analysis

The increased penetration of EVs has necessitated detailed analysis of power systems, especially distribution networks at different levels. However, modeling power systems in detail is a time-consuming task and the analysis of each region is difficult. Cluster analysis can be combined with neural networks to generate synthetic data for different regions and train the model using historical and synthetic data. Such an approach is proposed in [82] by dividing the distribution circuits into four categories (clusters). The authors note that the proposed method can produce accurate time series scenarios, under different EV penetration levels, to ensure stable power system operation. More research is required in this area to facilitate the analysis of power systems considering different levels of EV penetrations in different regions.

4.5. Clustering with Big Data for EVs

With the rapid development in electronics and ICT, all vehicles, and especially EVs, are equipped with more and more sensors and intelligence. This generates more data which can be used to manage different aspects of transportation in the electrified transportation era of the near future. For example, mitigation of transportation network congestion is proposed in [25] using big data and cluster analysis techniques to group/cluster different localities based on the traffic flow. Then, rerouting of EVs is considered to facilitate a smooth flow of traffic under different network congestion levels. More research is required on this topic to facilitate the increased penetration of EVs and to mitigate the existing congestion problems in the transportation sector.

5. Conclusions

This article presents a three-step analysis to review the application of clustering methods for different problems related to EVs. First, an overview of different existing clustering methods is provided. Then, the application of different clustering methods for diverse areas in EVs is reviewed. Finally, the research gaps in the existing literature are identified and future research directions are outlined.

The analysis has shown that the application of cluster analysis has gained popularity in the area of electromobility, and a number of studies have used clustering methods to address different related problems. The most widely applied areas identified in this study are the application of clustering methods to model EV user behavior, the EV driving cycle, the classification of used EV batteries, and clustering of EV charging stations. In addition, several potential areas have been identified in which the application of cluster analysis can bring new benefits. The prospective areas identified in this study are mitigation of the EV impact on distribution systems, development and coordination of charging infrastructure during emergencies, issues of equity and disparities in rebate allocations, and the use of big data with cluster analysis to assist transportation network management.

Author Contributions: Conceptualization, M.N., A.H. and P.M.; investigation, M.N.; resources, P.M.; writing—original draft preparation, M.N.; writing—review and editing, A.H. and P.M.; funding acquisition, P.M., supervision, P.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Future Energy Systems project under the Canada First Research Excellence Fund (CFREF) program at the University of Alberta, Canada.

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

References

1. Shek, C.L.; Manoharan, A.K.; Gampa, S.; Chandrappa, T.; Aravinthan, V. A Diversity-Based Clustering Technique for Implementing Decentralized Node Level Charge Scheduling of Electric Vehicles. In Proceedings of the 2019 North American Power Symposium (NAPS), Wichita, KS, USA, 13–15 October 2019. [CrossRef]
2. Pallonetto, F.; Galvani, M.; Torti, A.; Vantini, S. A Framework for Analysis and Expansion of Public Charging Infrastructure under Fast Penetration of Electric Vehicles. *World Electr. Veh. J.* **2020**, *11*, 18. [CrossRef]
3. Campbell, A.R.; Ryley, T.; Thring, R. Identifying the early adopters of alternative fuel vehicles: A case study of Birmingham, United Kingdom. *Transp. Res. Part A Policy Pract.* **2012**, *46*, 1318–1327. [CrossRef]
4. CO₂ and Greenhouse Gas Emissions—Our World in Data. Available online: <https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions> (accessed on 14 December 2022).
5. Friedlingstein, P.; Jones, M.W.; O’Sullivan, M.; Andrew, R.M.; Bakker, D.C.E.; Hauck, J.; Le Quééré, C.; Peters, G.P.; Peters, W.; Pongratz, J.; et al. Global Carbon Budget 2021. *Earth Syst. Sci. Data* **2022**, *14*, 1917–2005. [CrossRef]
6. The World Economic Forum. Available online: <https://www.weforum.org/> (accessed on 14 December 2022).
7. Annual report 2021 | UNFCCC. Available online: <https://unfccc.int/annualreport> (accessed on 14 December 2022).
8. Liu, Z.; Deng, Z.; Davis, S.J.; Giron, C.; Ciais, P. Monitoring global carbon emissions in 2021. *Nat. Rev. Earth Environ.* **2022**, *3*, 217–219. [CrossRef]
9. Reducing oil dependence in the EU through applied measures for trucks and buses—International Council on Clean Transportation. Available online: <https://theicct.org/publication/fs-eu-hdv-oil-imports-may22/> (accessed on 14 December 2022).
10. Why are electric vehicles the only way to quickly and substantially decarbonize transport?—International Council on Clean Transportation. Available online: <https://theicct.org/why-are-electric-vehicles-the-only-way-to-quickly-and-substantially-decarbonize-transport/> (accessed on 14 December 2022).
11. Global EV Outlook 2022—Analysis—IEA. Available online: <https://www.iea.org/reports/global-ev-outlook-2022> (accessed on 14 December 2022).
12. Yu, H.; Niu, S.; Shang, Y.; Shao, Z.; Jia, Y.; Jian, L. Electric vehicles integration and vehicle-to-grid operation in active distribution grids: A comprehensive review on power architectures, grid connection standards and typical applications. *Renew. Sustain. Energy Rev.* **2022**, *168*, 112812. [CrossRef]
13. Umoren, I.A.; Shakir, M.Z. Electric Vehicle as a Service (EVaaS): Applications, Challenges and Enablers. *Energies* **2022**, *15*, 7207. [CrossRef]
14. Gonzalez Venegas, F.; Petit, M.; Perez, Y. Active integration of electric vehicles into distribution grids: Barriers and frameworks for flexibility services. *Renew. Sustain. Energy Rev.* **2021**, *145*, 111060. [CrossRef]

15. Hussain, A.; Musilek, P. Reliability-as-a-Service Usage of Electric Vehicles: Suitability Analysis for Different Types of Buildings. *Energies* **2022**, *15*, 665. [CrossRef]
16. Sakthivel, R.; Kavikumar, R.; Mohammadzadeh, A.; Kwon, O.M.; Kaviarasan, B. Fault Estimation for Mode-Dependent IT2 Fuzzy Systems with Quantized Output Signals. *IEEE Trans. Fuzzy Syst.* **2021**, *29*, 298–309. [CrossRef]
17. Hussain, A.; Musilek, P. Resilience Enhancement Strategies For and Through Electric Vehicles. *Sustain. Cities Soc.* **2022**, *80*, 103788. [CrossRef]
18. Hu, D.; Zhou, K.; Li, F.; Ma, D. Electric vehicle user classification and value discovery based on charging big data. *Energy* **2022**, *249*, 123698. [CrossRef]
19. Berzi, L.; Delogu, M.; Pierini, M. Development of driving cycles for electric vehicles in the context of the city of Florence. *Transp. Res. Part D Transp. Environ.* **2016**, *47*, 299–322. [CrossRef]
20. Hu, X.; Sun, F. Fuzzy clustering based multi-model support vector regression state of charge estimator for lithium-ion battery of electric vehicle. In Proceedings of the 2009 International Conference on Intelligent Human-Machine Systems and Cybernetics, Hangzhou, China, 26–27 August 2009; pp. 392–396. [CrossRef]
21. Sánchez, D.G.; Tabares, A.; Faria, L.T.; Rivera, J.C.; Franco, J.F. A Clustering Approach for the Optimal Siting of Recharging Stations in the Electric Vehicle Routing Problem with Time Windows. *Energies* **2022**, *15*, 2372. [CrossRef]
22. Hussain, A.; Musilek, P. Utility-scale energy storage system for load management under high penetration of electric vehicles: A marginal capacity value-based sizing approach. *J. Energy Storage* **2022**, *56*, 105922. [CrossRef]
23. Hussain, A.; Musilek, P. Fairness and Utilitarianism in Allocating Energy to EVs During Power Contingencies Using Modified Division Rules. *IEEE Trans. Sustain. Energy* **2022**, *13*, 1444–1456. [CrossRef]
24. Guo, S.; Kontou, E. Disparities and equity issues in electric vehicles rebate allocation. *Energy Policy* **2021**, *154*, 112291. [CrossRef]
25. Lv, Z.; Qiao, L.; Cai, K.; Wang, Q. Big Data Analysis Technology for Electric Vehicle Networks in Smart Cities. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 1807–1816. [CrossRef]
26. Li, W.; Chen, S.; Peng, X.; Xiao, M.; Gao, L.; Garg, A.; Bao, N. A Comprehensive Approach for the Clustering of Similar-Performance Cells for the Design of a Lithium-Ion Battery Module for Electric Vehicles. *Engineering* **2019**, *5*, 795–802. [CrossRef]
27. Gan, G.; Ma, C.; Wu, J. 1. *Data Clustering: Theory, Algorithms, and Applications*; SAIM Publishers: Philadelphia, PA, USA, 2007; pp. 109–159. [CrossRef]
28. Reddy, A. *Data Clustering Algorithms and Applications*; Chapman & Hall: London, UK, 2014; pp. 1–150.
29. Govender, P.; Sivakumar, V. Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmos. Pollut. Res.* **2020**, *11*, 40–56. [CrossRef]
30. Ezugwu, A.E.; Ikotun, A.M.; Oyelade, O.O.; Abualigah, L.; Agushaka, J.O.; Eke, C.I.; Akinyelu, A.A. A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Eng. Appl. Artif. Intell.* **2022**, *110*, 104743. [CrossRef]
31. Maimon, O.; Rokach, L. *Data Mining and Knowledge Discovery Handbook*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2010.
32. Erman, N.; Korosec, A.; Suklan, J. Performance of Selected Agglomerative Hierarchical Clustering Methods. *Innov. Issues Approaches Soc. Sci.* **2015**, *8*, 180–204. [CrossRef]
33. Vijaya, V.; Sharma, S.; Batra, N. Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering. In Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 14–16 February 2019; pp. 568–573. [CrossRef]
34. Madhulatha, T.S. An Overview on Clustering Methods. *IOSR J. Eng.* **2012**, *2*, 719–725. [CrossRef]
35. Jain, A.K.; Murty, M.N.; Flynn, P.J. Data Clustering: A Review. *ACM Comput. Surv.* **2000**, *31*, 264–323. [CrossRef]
36. Omran, M.G.; Engelbrecht, A.P.; Salman, A. An overview of clustering methods. *Intell. Data Anal.* **2007**, *6*, 583–605. [CrossRef]
37. Saxena, A.; Prasad, M.; Gupta, A.; Bharill, N.; Patel, O.P.; Tiwari, A.; Er, M.J.; Ding, W.; Lin, C.T. A review of clustering techniques and developments. *Neurocomputing* **2017**, *267*, 664–681. [CrossRef]
38. Berkhin, P. A survey of clustering data mining techniques. In *Grouping Multidimensional Data*; Springer: Berlin, Germany, 2006; pp. 25–71. [CrossRef]
39. Steinbach, M.; Karypis, G.; Kumar, V. A Comparison of Document Clustering Techniques, Technical Report; 00-034, University of Minnesota Digital Conservancy, 2000, 1–22. Available online: <https://hdl.handle.net/11299/215421> (accessed on 14 December 2022).
40. Dunn, J.C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *J. Cybern.* **2008**, *3*, 32–57. [CrossRef]
41. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Springer: New York, NY, USA, 1981; pp. 1–272. [CrossRef]
42. Ghosh, S.; Dubey, S.K. Comparative Analysis of K-Means and Fuzzy C-Means Algorithms. *IJACSA* **2013**, *4*, 1–5. [CrossRef]
43. Wu, J. *Advances in K-means Clustering: A data Mining Thinking*; Springer: New York, NY, USA, 2012; pp. 152–235. [CrossRef]
44. Shi, Q.S.; Zheng, X.Z. Electric Vehicle Charging Stations Optimal Location Based on Fuzzy C-Means Clustering. *Appl. Mech. Mater.* **2014**, *556–562*, 3972–3975. [CrossRef]
45. Huang, Z.; Ng, M.K. A Note on K-modes Clustering. *J. Classif.* **2003**, *20*, 257–261. [CrossRef]
46. Huang, J. A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining. *Dmkd* **1997**, *3*, 34–39.
47. Sajidha, S.A.; Chodnekar, S.P.; Desikan, K. Initial seed selection for K-modes clustering—A distance and density based approach. *J. King Saud Univ. Comput. Inf. Sci.* **2021**, *33*, 693–701. [CrossRef]

48. Zhang, J.; Yang, C.; Ju, F. Optimization of ordered charging strategy for large scale electric vehicles based on quadratic clustering. In Proceedings of the 2017 4th International Conference on Information Science and Control Engineering (ICISCE), Changsha, China, 21–23 July 2017; pp. 1080–1084. [\[CrossRef\]](#)
49. Tian, J.; Wang, Y.; Liu, C.; Chen, Z. Consistency evaluation and cluster analysis for lithium-ion battery pack in electric vehicles. *Energy* **2020**, *194*, 116944. [\[CrossRef\]](#)
50. Catalbas, M.C.; Yildirim, M.; Gulten, A.; Kurum, H. Estimation of optimal locations for electric vehicle charging stations. In Proceedings of the 2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), Milan, Italy, 6–9 June 2017. [\[CrossRef\]](#)
51. Reynolds, D. Gaussian Mixture Models. In *Encyclopedia of Biometrics*; Springer: Boston, MA, USA, 2015; pp. 103–255, 827–832. [\[CrossRef\]](#)
52. Xu, L.; Marshall, M.; Dow, L. A framework for assessing the impact of plug-in electric vehicle to distribution systems. In Proceedings of the 2011 IEEE/PES Power Systems Conference and Exposition, Phoenix, AZ, USA, 20–23 March 2011. [\[CrossRef\]](#)
53. He, X.; Cai, D.; Shao, Y.; Bao, H.; Han, J. Laplacian regularized Gaussian mixture model for data clustering. *IEEE Trans. Knowl. Data Eng.* **2011**, *23*, 1406–1418. [\[CrossRef\]](#)
54. Helmbrecht, M.; Olaverri-Monreal, C.; Bengler, K.; Vilimek, R.; Keinath, A. How electric vehicles affect driving behavioral patterns. *IEEE Intell. Transp. Syst. Mag.* **2014**, *6*, 22–32. [\[CrossRef\]](#)
55. Miyazaki, K.; Uchiba, T.; Tanaka, K. Clustering to Predict Electric Vehicle Behaviors using State of Charge data. In Proceedings of the 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), Madrid, Spain, 9–12 June 2020. [\[CrossRef\]](#)
56. Bozorgi, A.M.; Farasat, M.; Mahmoud, A. A Time and Energy Efficient Routing Algorithm for Electric Vehicles Based on Historical Driving Data. *IEEE Trans. Intell. Veh.* **2017**, *2*, 308–320. [\[CrossRef\]](#)
57. Wang, H.; Wang, B.; Fang, C.; Liu, W.; Huang, H. Bidding Strategy Research for Aggregator of Electric Vehicles Based on Clustering Characteristics. In Proceedings of the 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 3–5 June 2019; pp. 6150–6156. [\[CrossRef\]](#)
58. Crozier, C.; Apostolopoulou, D.; McCulloch, M. Clustering of Usage Profiles for Electric Vehicle Behaviour Analysis. In Proceedings of the 2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), Sarajevo, Bosnia and Herzegovina, 21–25 October 2018. [\[CrossRef\]](#)
59. Xiong, Y.; Wang, B.; Chu, C.C.; Gadh, R. Electric Vehicle Driver Clustering using Statistical Model and Machine Learning. In Proceedings of the 2018 IEEE Power & Energy Society General Meeting (PESGM), Portland, OR, USA, 5–10 August 2018; pp. 1–5. [\[CrossRef\]](#)
60. Fan, Z.; Ziyi, C.; Jing, Y.; Mingxia, W. Electric Vehicle Configuration Recommendation Algorithm Using Clustering Fusion Matrix Decomposition and Pearson Calculation. In Proceedings of the 2021 IEEE International Conference on Emergency Science and Information Technology (ICESIT), Chongqing, China, 22–24 November 2021; pp. 683–693. [\[CrossRef\]](#)
61. Helmus, J.R.; Lees, M.H.; van den Hoed, R. A data driven typology of electric vehicle user types and charging sessions. *Transp. Res. Part C Emerg. Technol.* **2020**, *115*, 102637. [\[CrossRef\]](#)
62. Powell, S.; Cezar, G.V.; Rajagopal, R. Scalable probabilistic estimates of electric vehicle charging given observed driver behavior. *Appl. Energy* **2022**, *309*, 118382. [\[CrossRef\]](#)
63. Brady, J.; O'Mahony, M. Development of a driving cycle to evaluate the energy economy of electric vehicles in urban areas. *Appl. Energy* **2016**, *177*, 165–178. [\[CrossRef\]](#)
64. Fotouhi, A.; Montazeri-Gh, M. Tehran driving cycle development using the k-means clustering method. *Sci. Iran.* **2013**, *20*, 286–293. [\[CrossRef\]](#)
65. Yuhui, P.; Yuan, Z.; Huibao, Y. Development of a representative driving cycle for urban buses based on the K-means cluster method. *Clust. Comput.* **2018**, *22*, 6871–6880. [\[CrossRef\]](#)
66. Zhou, W.; Xu, K.; Yang, Y.; Lu, J. Driving Cycle Development for Electric Vehicle Application using Principal Component Analysis and K-means Cluster: With the Case of Shenyang, China. *Energy Procedia* **2017**, *105*, 2831–2836. [\[CrossRef\]](#)
67. Zhang, F.; Guo, F.; Huang, H. A Study of Driving Cycle for Electric Special-purpose Vehicle in Beijing. *Energy Procedia* **2017**, *105*, 4884–4889. [\[CrossRef\]](#)
68. Zhao, X.; Yu, Q.; Ma, J.; Wu, Y.; Yu, M.; Ye, Y. Development of a representative EV urban driving cycle based on a k-Means and SVM hybrid clustering algorithm. *J. Adv. Transp.* **2018**, 1–18. [\[CrossRef\]](#)
69. Zhao, X.; Zhao, X.; Yu, Q.; Ye, Y.; Yu, M. Development of a representative urban driving cycle construction methodology for electric vehicles: A case study in Xi'an. *Transp. Res. Part D Transp. Environ.* **2020**, *81*, 102279. [\[CrossRef\]](#)
70. Zhao, X.; Ma, J.; Wang, S.; Ye, Y.; Wu, Y.; Yu, M. Developing an electric vehicle urban driving cycle to study differences in energy consumption. *Environ. Sci. Pollut. Res.* **2018**, *26*, 13839–13853. [\[CrossRef\]](#)
71. Chen, Z.; Yang, C.; Fang, S. A Convolutional Neural Network-Based Driving Cycle Prediction Method for Plug-in Hybrid Electric Vehicles with Bus Route. *IEEE Access* **2020**, *8*, 3255–3264. [\[CrossRef\]](#)
72. Wang, P.; Pan, C.; Sun, T. Control strategy optimization of plug-in hybrid electric vehicle based on driving data mining. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2022**. [\[CrossRef\]](#)
73. Li, C.; Wang, N.; Li, W.; Li, Y.; Zhang, J. Regrouping and Echelon Utilization of Retired Lithium-Ion Batteries Based on a Novel Support Vector Clustering Approach. *IEEE Trans. Transp. Electrification* **2022**, *8*, 3648–3658. [\[CrossRef\]](#)

74. Hu, X.; Li, S.E.; Yang, Y. Advanced Machine Learning Approach for Lithium-Ion Battery State Estimation in Electric Vehicles. *IEEE Trans. Transp. Electr.* **2016**, *2*, 140–149. [[CrossRef](#)]
75. Wang, D.; Yang, F.; Gan, L.; Li, Y. Fuzzy Prediction of Power Lithium Ion Battery State of Function Based on the Fuzzy c-Means Clustering Algorithm. *World Electr. Veh. J.* **2019**, *10*, 1. [[CrossRef](#)]
76. Liu, D. Cluster Control for EVs Participating in Grid Frequency Regulation by Using Virtual Synchronous Machine with Optimized Parameters. *Appl. Sci.* **2019**, *9*, 1924. [[CrossRef](#)]
77. Xu, Z.; Wang, J.; Lund, P.D.; Fan, Q.; Dong, T.; Liang, Y.; Hong, J. A novel clustering algorithm for grouping and cascade utilization of retired Li-ion batteries. *J. Energy Storage* **2020**, *29*, 101303. [[CrossRef](#)]
78. Chen, C.; Li, T.; Wang, S.; Hua, Z.; Kang, Z.; Li, D.; Guo, W. Location Analysis of Urban Electric Vehicle Charging Metro-Stations Based on Clustering and Queuing Theory Model. *Commun. Comput. Inf. Sci.* **2022**, *1566 CCIS*, 282–292. [[CrossRef](#)]
79. Optimization for allocating BEV recharging stations in urban areas by using hierarchical clustering | IEEE Conference Publication | IEEE Xplore. Available online: <https://ieeexplore.ieee.org/document/5713494> (accessed on 14 December 2022).
80. Momtazpour, M.; Butler, P.; Hossain, M.S.; Bozchalui, M.C.; Ramakrishnan, N.; Sharma, R. Coordinated clustering algorithms to support charging infrastructure design for electric vehicles. In Proceedings of the ACM SIGKDD International Workshop on Urban Computing, Beijing, China, 12 August 2012; pp. 126–133. [[CrossRef](#)]
81. Dow, L.; Marshall, M.; Xu, L.; Agüero, J.R.; Willis, H.L. A novel approach for evaluating the impact of electric vehicles on the power distribution system. In Proceedings of the IEEE PES General Meeting, Minneapolis, MN, USA, 25–29 July 2010. [[CrossRef](#)]
82. Yang, F.; Yin, S.; Zhou, S.; Li, D.; Fang, C.; Lin, S. Electric vehicle charging current scenario generation based on generative adversarial network combined with clustering algorithm. *Int. Trans. Electr. Energy Syst.* **2021**, *31*, e12971. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.