IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 September 2023, accepted 23 October 2023, date of publication 25 October 2023, date of current version 1 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3327640

TOPICAL REVIEW

K-Means and Alternative Clustering Methods in Modern Power Systems

SEYED MAHDI MIRAFTABZADEH, (Member, IEEE),
CRISTIAN GIOVANNI COLOMBO, (Student Member, IEEE),
MICHELA LONGO, (Member, IEEE), AND
FEDERICA FOIADELLI, (Senior Member, IEEE)
Department of Energy, Politecnico di Milano, 20156 Milan, Italy

Corresponding author: Seyed Mahdi Miraftabzadeh (seyedmahdi.miraftabzadeh@polimi.it)

This work was supported in part by the Il Centro Nazionale per Mobilitaì Sostenibile (MOST)—Sustainable Mobility Center; and in part by the European Union Next-Generation EU (Piano Nazionale di Ripresa e Resilienza (PNRR)—Missione 4 Componente 2, Investimento 1.4—D.D. 1033 17/06/2022) under Grant CN00000023.

ABSTRACT As power systems evolve by integrating renewable energy sources, distributed generation, and electric vehicles, the complexity of managing these systems increases. With the increase in data accessibility and advancements in computational capabilities, clustering algorithms, including K-means, are becoming essential tools for researchers in analyzing, optimizing, and modernizing power systems. This paper presents a comprehensive review of over 440 articles published through 2022, emphasizing the application of K-means clustering, a widely recognized and frequently used algorithm, along with its alternative clustering methods within modern power systems. The main contributions of this study include a bibliometric analysis to understand the historical development and wide-ranging applications of K-means clustering in power systems. This research also thoroughly examines K-means, its various variants, potential limitations, and advantages. Furthermore, the study explores alternative clustering algorithms that can complete or substitute K-means. Some prominent examples include K-medoids, Time-series K-means, BIRCH, Bayesian clustering, HDBSCAN, CLIQUE, SPECTRAL, SOMs, TICC, and swarm-based methods, broadening the understanding and applications of clustering methodologies in modern power systems. The paper highlights the wide-ranging applications of these techniques, from load forecasting and fault detection to power quality analysis and system security assessment. Throughout the examination, it has been observed that the number of publications employing clustering algorithms within modern power systems is following an exponential upward trend. This emphasizes the necessity for professionals to understand various clustering methods, including their benefits and potential challenges, to incorporate the most suitable ones into their studies.

INDEX TERMS Clustering algorithms, K-means clustering, power systems.

I. INTRODUCTION

In recent years, advances in Machine Learning (ML) methods have captured significant attention and interest in both academic and industry domains [1], [2]. Unsupervised learning ML models have attracted considerable interests due to their remarkable ability to extract new knowledge from datasets without relying on prior information [3], [4]. This powerful capability opens up vast opportunities for discovering hidden patterns and insights that might have otherwise remained unnoticed or unsolved. Clustering algorithms are crucial as they categorize vast datasets into understandable and meaningful groups, uncovering hidden patterns and structures [5]. Especially in modern power systems, clustering algorithms facilitate automated, data-driven decision-making by managing enormous data points. They assist in predicting load demands, optimizing grid management, anomaly detections, and formulating power management strategies, enhancing overall operational efficiency and reliability [6], [7], [8], [9], [10].

The associate editor coordinating the review of this manuscript and approving it for publication was Kashif Munir.

A power system, or electrical power system, is a complex network responsible for generating, transmitting, and distributing electrical energy to various consumers. This infrastructure, comprising power plants, substations, transmission lines, and devices, is essential for powering modern society and technological advancements, supporting diverse applications and industries.

The K-means algorithm is a popular choice for clustering analysis in power systems due to its simplicity, efficiency, scalability, and ease of implementation [11], [12]. It is an iterative algorithm that divides a dataset into k non-overlapping subgroups or clusters. It can handle large datasets effectively, making it a practical choice for numerous applications such as load profiling, fault detection and diagnosis, renewable energy forecasting and management, smart grid management, customer segmentation, optimal siting and sizing of distributed generation, Electric Vehicle (EV) charging infrastructure planning, Phasor Measurement Unit (PMU) placement, Power Quality (PQ) analysis, and energy management [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23].

The widespread utilization of k-means clustering in power systems [24], along with its diverse applications, underlines the importance of conducting a thorough bibliometric analysis and a combination of scoping and systematic literature reviews concerning the applications of the k-means algorithm in modern power systems. This clustering method has become a vital tool in the field, employed in load forecasting, condition monitoring, fault detection, and renewable energy integration, among many other applications [25], [26], [27]. These applications highlight the essential role that k-means clustering plays in improving the reliability, efficiency, and sustainability of power systems. Consequently, an exploratory examination of existing literature is not only beneficial but crucial to gain insights into the current usage trends, identifying the advantages and limitations of this clustering method in power systems, and guiding future research directions. This knowledge can also inform the development of more effective strategies and tools to enhance power systems operation and management.

On the other hand, the K-means algorithm has certain limitations that may make it less suitable for some specific applications [11]. While it is widely used and appreciated for its simplicity and efficiency, it does not always meet the subtle requirements of complex domains like power systems [28]. In the literature, numerous other clustering algorithms have been developed, used, and can be applied to modern power systems, providing various benefits that may not be achievable with K-means. These alternative methods offer more sophisticated clustering techniques and might handle particular challenges, such as noise, nonlinearity, or the presence of outliers, more robustly [29], [30]. The choice of the best clustering algorithm should be guided by the specific requirements and characteristics of the data and the goals of the analysis in power systems.

To the best knowledge of the authors, while the applications of K-means algorithm have been explored in other domains, no study has yet provided a review of this well-known algorithm within the context of power systems. Moreover, in power systems, studies have yet to thoroughly review and study the alternatives to the K-means clustering algorithm. This gap in the literature highlights a crucial area for future research, as exploring various clustering methods could uncover new insights and techniques specifically adapted to the unique challenges and demands of modern power system analysis and optimization. To fill this gap, more than 490 articles published by 2022 have been studied and analyzed in this work. The main contributions of this paper are as follows:

1) Conducting a bibliometric analysis of K-means clustering algorithms to understand their historical development, application, and future trend.
2) Providing a comprehensive analysis of K-means that includes its various variants, highlights its applications within the power systems, and examines both its advantages and disadvantages.
3) Investigating alternative clustering algorithms that can be considered in place of the K-means algorithm, such as K-medoids, Time-series K-means, BIRCH, Bayesian clustering, HDBSCAN, CLIQUE, SPECTRAL, and swarm-based clustering algorithms, thereby broadening the understanding of clustering methodologies in the context of power systems.
4) Analyzing and evaluating the performance of predominantly used clustering algorithms in modern power systems.

The rest of this paper is organized as follows: Section II presents the bibliometric analysis of the K-means algorithm. Section III discusses the K-means algorithm in detail, including different configurations for using this algorithm. Section IV explores the various applications of K-means in different aspects of modern power systems. Section V summarizes the application of K-means, discussing its advantages, limitations, and other variants. Section VI presents alternative clustering methods to K-means and their applications in modern power systems. Section VII analyzes and evaluates the performance of the most commonly used clustering algorithms in modern power systems. Section VIII delves into the findings, including additional analysis. Finally, Section IX concludes the paper.

II. BIBLOMETRIC ANALYSIS

The production of new reliable scientific data makes it possible to identify the relevance of a topic within a complex framework of bibliographic references. At the same time, identifying the progress of a subject within the general framework allows for determining new areas of research and interest, starting from the understanding of the evolution of the topic. Bibliometric analysis is among the most effective methods for analyzing these data and trends. With the development of scientific databases, this methodology has grown in popularity and fields of application [31], [32], [33], [34], [35]. Bibliometric analysis techniques vary according to the objective. A performance analysis considers the authors' contribution to topic-related research. On the other hand, science mapping is focused on correlating research topics and components. The latter technique is used in five different ways [36], [37], [38], [39]. Through the use of one of these methods or their combination, it is possible to indicate the relevance of a topic and monitor its evolution over time, until arriving at an accurate prediction of its trend for the future. Therefore, it is analyzed the K-means clustering in modern power system applications to identify the relevance of the different correlated elements. Fig. 1 reports the K-means clustering in power systems bibliometric trend using the Scopus database.

In Fig. 1 are annually collected the number of publications made involving the topic "K-means" in the power system field. The publications are differentiated as:

Conference papers;
Journal articles;
Conference reviews;
Book chapters;
Reviews.

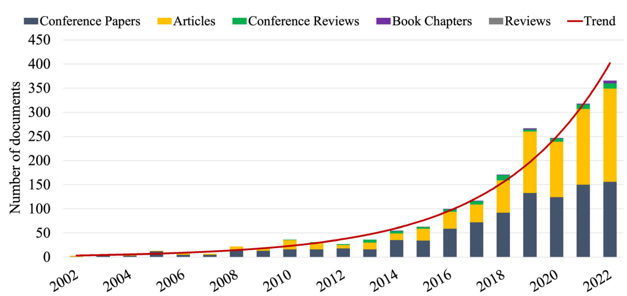


FIGURE 1. K-means clustering and power system bibliometric trend.

As presented in Fig. 1, the trend of interest in the k-means clustering topic has increased exponentially, growing the number of publications almost every year, reaching 366 in 2022. Following this trend and using the same classification, it is worth investigating the most relevant sources of publications related to the topics. The processing of the sources results in Fig. 2.

Fig. 2 highlights that conference papers are the primary source of publication related to K-means (984), and the most interesting conference is "Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics," which published 38 conference papers until 2022. Additionally, journal articles account for a significant proportion of publications related to K-means, comprising 44% of the total, which equates to 851 articles. The journals with the most publications are IEEE Access and Energies, with 44 and 43 articles, respectively. Both of these are open-access journals. On a completely different scale, conference reviews and

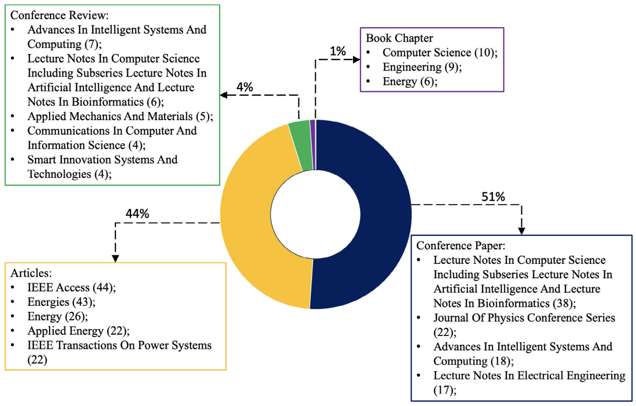


FIGURE 2. K-means clustering and power system most relevant sources.

book chapters account for 72 and 18 publications, respectively. In this context, "Advances in Intelligent Systems and Computing" is the most publishing conference of conference review, while "Computer Science" accounts for the highest number of book chapter publications. Finally, article reviews are only 2, making them almost negligible in the framework. Considering the increasing trend of the K-means clustering topic, it is predictable that this number is going to increase.

Moreover, the origins of publications are of interest in order to understand which countries are more actively engaged in working on K-means clustering in the power system field. Therefore, Fig. 3 presents the publications related to this topic by country, focusing on the ten most relevant nations in terms of publication numbers.

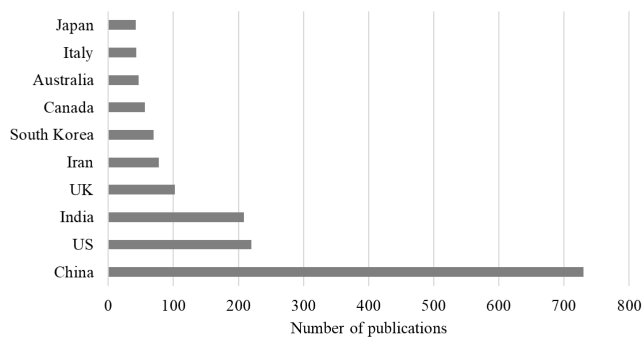


FIGURE 3. K-means clustering and power system bibliometric trend per country.

Fig. 3 distinctly illustrates that China leads in research on K-means clustering in the power system, boasting 730 indexed publications up to 2022. The United States and India are next, with 220 and 208 indexed publications, respectively. The remaining countries in the top 10, in terms of publications related to K-means clustering in the power system, include the UK, Iran, South Korea, Canada, Australia, Italy, and Japan.

Lastly, a co-word analysis is performed to explore the relationship between the main subject and the broader research landscape. This examination reveals the significance of each topic connected to K-means clustering in power systems. Fig. 4 presents a graphical representation of the co-word analysis related to K-means clustering, portrayed through a heatmap, using VOSviewer software. This analytical approach clarifies how K-means clustering integrates with other key research themes and domains.

The heatmap uses more intense colors to signify proximity to the reference topic, reflecting the high number of keyword connections, while the size of a word depends on the frequency with which the keyword is mentioned. As anticipated, 'K-means clustering' is the most significant keyword and is positioned in the midst of the most intense color, as it represents the main keyword of the research, often coupled with 'power grid' to define the general context. Several other connected topics are frequently mentioned, such as forecasting and energy efficiency. These occurrences are unsurprising since K-means clustering is often utilized to create scenarios or clusters in forecasting models, and enhancing energy efficiency is a common goal. The intensity of the color indicates that K-means clustering is a widely used technique, extensively linked to various subjects, as evidenced by connections to terms like stochastic system and forecasting.

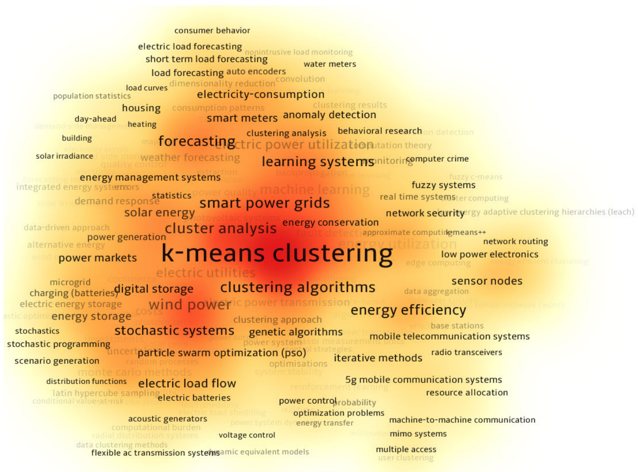


FIGURE 4. Co-word analysis for K-means clustering in power system.

Considering Fig. 4 and the domain knowledge of the authors, several key areas have been identified for further exploration in the context of applying K-means clustering to power systems. These topics represent diverse aspects of power systems, ranging from different energy sources to operational and managerial considerations as presented in Table 1. Each of these areas presents unique challenges and opportunities, and understanding how K-means clustering can be adapted to meet the specific needs of these aspects is investigated in section IV. The selection of papers published up to 2022 was based on these keywords, aiming to conduct a comprehensive review that covers a diverse range of applications of K-means in power systems.

The VOSviewer software is used to identify the connections between the selected keywords and K-means clustering, considering the publications in the literature from the Scopus database by 2022, as presented in Fig. 5.

The rules for Fig. 5 are the same as those for Fig. 4; a larger name represents a higher number of occurrences, while a more intense color indicates a greater number of connections to the keywords. In this case, K-means clustering also is the main topic, signified by the largest size and most intense color. However, other keywords are also noteworthy. Forecasting and energy efficiency remain two of the most frequent and correlated keywords, but other terms such as wind power, electric power utilization, electric utilities, and PV-related keywords also stand out. This reflects that K-means clustering is widely used in literature for forecasting renewable energy sources (RESs) and identifying common energy demand patterns. Further descriptions of K-means applications in the power system field for each selected topic are provided in Section IV.

TABLE 1. Selected keywords.

"Wind Power"	"Solar Power"
"Electric Power Plant Loads"	"Electric Lines"
"Electric Load Flow"	"Fault Detection"
"Reactive Power"	"Electricity Consumption"
"Energy Utilization"	"Smart Meters"
"Scheduling"	"Network Security"
"Outages"	"Risk Assessment"
"Power Markets"	"Load Forecasting"
"Smart Grids"	"Energy Efficiency"
"Energy Storage"	"Energy Management"
"Optimization"	"Power Quality"
"Electric Vehicles"	

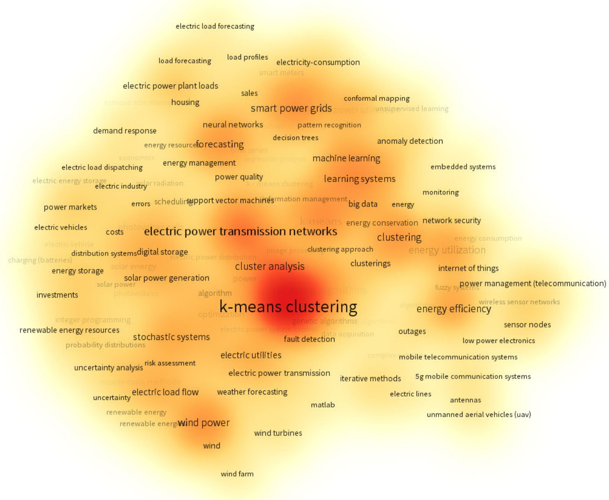


FIGURE 5. Co-word analysis for K-means clustering in power system restricted to main applications.

III. K-MEANS CLUSTERING

The K-means clustering algorithm is among the most popular unsupervised machine learning algorithms. It was first proposed by Stuart Lloyd [40] in 1957; However, James

MacQueen [41], a decade later in 1967, not only popularized this effective algorithm but also named it "K-means," a term now used worldwide [42]. The K-means algorithm has gained popularity because it is easy to understand, implement, and relatively efficient, especially when dealing with large datasets.

A. K-MEANS ALGORITHM

The primary role of the K-means algorithm is to divide a dataset into K distinct clusters, which are non-overlapping groups formed around the K centroids calculated by the algorithm [11]. The K-means algorithm aims to minimize the within-cluster variance, forcing the centroids to move to positions surrounded by samples closer to them than any other centroid. The measure of similarity or dissimilarity between data points, typically determined using the Euclidean distance (1), plays a crucial role in this process [43], although other distance measures can also be utilized.

$$\boldsymbol{d}\left(\boldsymbol{x}^{p}, \boldsymbol{c}^{q}\right)=\sqrt{\sum_{i=1}^{n}\left(x_{i}^{p}-c_{i}^{q}\right)}=\left\|\boldsymbol{x}^{p}-\boldsymbol{c}^{q}\right\|_{2} \tag{1}$$

where $\boldsymbol{x}$ is the p^{th} sample from $\boldsymbol{X}$, $\boldsymbol{c}$ is the q^{th} centroid of the qth cluster, and $\boldsymbol{x}$ and $\boldsymbol{c}$ are sets of n-dimensional vectors.

The K-means, algorithm 1, is initialized by randomly selecting k data points from the input $\boldsymbol{X}$ as initial centroids. The process then enters a loop that continues for a specified number of iterations or until the centroids cease to shift significantly, indicating convergence. During each iteration of the algorithm, every data point in $\boldsymbol{X}$ is assigned to a closet cluster with a minimum Euclidean distance to its centroid. Following this assignment, the centroids are updated based on the current composition of the clusters. The new position for each centroid is computed as the mean of all data points currently within the respective cluster [42], [44], [45]. After completing the iterative loop or convergence, the algorithm ultimately yields the final set of k centroids and the corresponding clusters.

Convergence in k-means is typically defined as the point at which the assignments of data points to clusters no longer change or the changes are below a predefined threshold. This indicates that the algorithm has found an optimal solution (local optimum) given the current centroids. The mathematical definition of convergence in K-means can be expressed by 2:

$$\sum_{i=1}^{k}\left\|\boldsymbol{c}_{t}^{i}-\boldsymbol{c}_{t-1}^{i}\right\|^{2} \leq \boldsymbol{\epsilon} \tag{2}$$

where $\boldsymbol{c}_t^i$ is the i^{th} centroid at the t^{th} iteration, and $\boldsymbol{c}_{t-1}^i$ is the same centroid at the $(t-1)^{th}$ iteration.

The K-means algorithm begins by randomly assigning data points to clusters and determining their centroids. The algorithm aims to minimize the variance or distance of points within the same cluster. This process, however, does not assure an efficient convergence or an optimal clustering outcome. Due to its random initialization, the K-means algorithm can suffer from poor convergence and suboptimal clustering dependent on initial conditions. Moreover, it is prone to settling into local minima instead of finding the global optimum [12], [46], [47].

Algorithm 1 Original K-Means Algorithm

Input:

Input samples: $\boldsymbol{X}=\left\{\boldsymbol{x}^{1}, \boldsymbol{x}^{2}, \cdots, \boldsymbol{x}^{m}\right\}, \boldsymbol{x}^{i} \in \boldsymbol{R}^{\boldsymbol{n}}$

Number of clusters: k

number of iterations: I

Output:

Cluster centroids: $\boldsymbol{C}=\left\{\boldsymbol{c}^{1}, \boldsymbol{c}^{2}, \cdots, \boldsymbol{c}^{k}\right\}$

Associated clusters: $S=\left\{S^{1}, S^{2}, \cdots, S^{k}\right\}$

Initialization: Randomly select k data points from $\boldsymbol{X}$ as the initial centroids

For $iter = 1$ to I **or** until convergence **do**

  For $i = 1$ to m **do**

    $j^{*} \leftarrow \underset{j}{\operatorname{argmin}}\left\|\boldsymbol{x}^{i}-\boldsymbol{c}^{j}\right\|_{2}^{2}$

    $S^{j^{*}} \leftarrow S^{j^{*}} \cup\left\{\boldsymbol{x}^{i}\right\}$

  For $i = 1$ to k **do**

    $\boldsymbol{c}^{i}=\frac{1}{\left|\boldsymbol{c}^{i}\right|} \sum_{\boldsymbol{x}^{j} \in \boldsymbol{c}^{i}} \boldsymbol{x}^{j}$

Return:

$\boldsymbol{C}$ and S

K-means++, along with Principal Component Analysis (PCA) initialization, serves as a robust strategy for enhancing the performance of the K-means clustering algorithm [12], [46], [48]. K-means++, by providing a more intelligent initialization method, substantially reduces the algorithm's sensitivity to the initial cluster centroids. This minimizes the convergence risk to suboptimal solutions, making the K-means algorithm more consistent and reliable across different runs. On the other hand, the PCA components are used as the initial cluster centroids for K-means, allowing K-means to start with a more informed initialization rather than random centroids [49], [50]. This approach can be useful in scenarios where PCA is applied to preprocess data, and the resulting components are assumed to provide a reasonable starting point for clustering. While several initialization techniques are available for the K-means algorithm, K-means++ stands out as one of the most widely adopted and effective methods due to its ability to consistently generate high-quality initial cluster centroids.

B. K-MEANS++ ALGORITHM

K-means++ introduced to improve the initialization process and significantly reduces the likelihood of converging to a local minimum, thus offering a more reliable path to optimal clustering [12], [51]. The K-means++ algorithm is described with the following steps:

1. Randomly select a data point from the dataset to serve as the first cluster center.

2. For each remaining data point in the dataset, calculate its distance squared to the nearest pre-existing cluster center.
3. Select the next cluster center from the remaining data points with a probability proportional to the calculated squared distance. This means that a point further away from the existing cluster centers is more likely to be selected as the next center.
4. Repeat steps 2 and 3 until a total of K cluster centers have been chosen.

After selecting the initial cluster centers during the initialization phase, the K-means algorithm resumes its usual course, proceeding with the standard iteration process for optimal cluster formation.

Repeating the standard K-means algorithm multiple times while offering opportunities to avoid poor initializations can be computationally expensive and does not guarantee improved clustering outcomes due to the randomness of initialization [51], [52]. On the other hand, K-means++ provides a substantial advantage by adopting a deterministic approach to initialization, improving the probability of finding the global optimum and reducing the likelihood of converging to local minima.

K-means++ provides computational efficiency by enhancing initialization in a single run, making it especially beneficial for large datasets or when many clusters are required [53]. It strategically selects initial centroids to increase the likelihood of achieving a global optimum, reducing the chance of convergence to local minima. Moreover, K-means++ ensures more consistent and stable results than the traditional K-means algorithm.

C. PCA INITIALIZATION

Principal Component Analysis (PCA), a widely used dimensionality reduction technique, can also be used in the initialization step of the K-means algorithm. PCA finds the dimensions displaying maximum data variance. Thus, assigning the initial cluster centers to the points that align with the first K principal components potentially provides a beneficial starting point for the clustering procedure. This approach is particularly effective when clusters are distinguishable and linearly separable along the axes of high variance.

However, PCA initialization has some limitations. Mainly, it assumes a linear structure in the data, which is not always met. If the principal components do not align well with the actual cluster structure in the data, this could lead to an ineffective initialization and potentially poor clustering results.

D. CONFIGURATIONS

The K-means clustering algorithm is implemented in three distinct configurations [54], [55], [56], as shown in Fig. 6. Since K-means employs Euclidean distance to evaluate the similarity and dissimilarity within each cluster, normalizing the data in the preprocessing phase is essential [11]. This step guarantees that each feature holds the same weight in determining the cluster centroids. The main configurations are as follows:

- Vanilla configuration: K-means can be directly applied to data following the preprocessing step, as presented in Fig. 6 a).
- Feature-extracted configuration: Another potential configuration involves extracting new features, such as statistical attributes from the dataset, to enhance the performance of the K-means algorithm [22]. These newly extracted features may necessitate normalization once again. The newly extracted features can be used independently or in combination with the original features as inputs to the model, as depicted in the Fig. 6 b).
- Dimensionality-reduction configuration: The performance of the K-means algorithm can be significantly impacted by the so-called "curse of dimensionality," a phenomenon where high-dimensional data can lead to computational or analytical difficulties for machine learning models. Dimensionality reduction techniques, such as PCA, Kernel PCA (KPCA), or neural network autoencoder can be strategically utilized to alleviate the challenges posed by high-dimensionality in data before initiating the clustering process [43], [57], [58], [59], as shown in Fig. 6 c). In recent years, due to their efficient and effective compression capabilities, autoencoders have been a state-of-the-art choice for dimensionality reduction, particularly when dealing with complex or large-scale datasets.

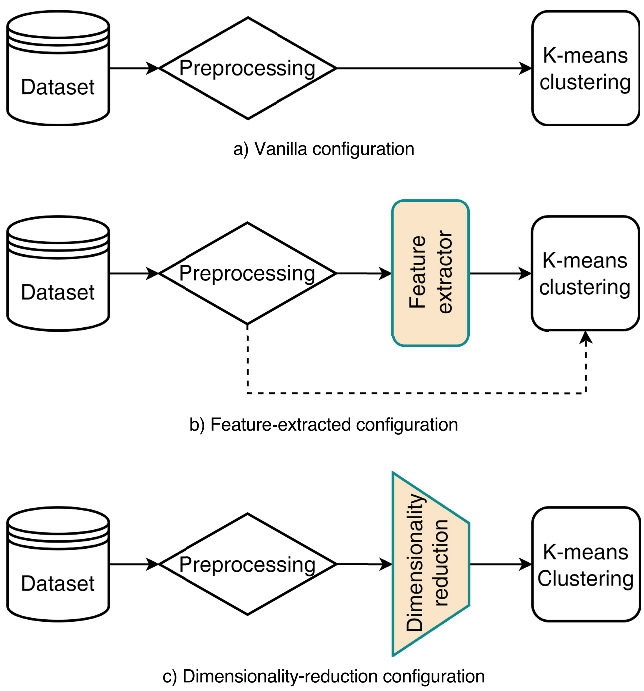


FIGURE 6. K-means clustering configurations: a) Vanilla, b) Feature-extracted, and c) Dimensionality-reduction configurations.

In the second configuration, it is crucial to emphasize the computation of distances or similarities between data points as a powerful feature extraction step before applying clustering algorithms in specific scenarios. These measures, such as Euclidean distance, Manhattan distance, Cosine similarity, and Dynamic Time Warping (DTW), quantify the relationships or dissimilarities among data points within a dataset. These computations create a distance or similarity matrix, which serves as input for clustering algorithms [60]. This approach can lead to efficient clustering, as it establishes the foundation for identifying clusters based on the inherent relationships within the data.

IV. K-MEANS APPLICATIONS IN MODERN POWER SYSTEMS

This section comprehensively reviews and presents the applications of K-means clustering across various fields within modern power systems.

A. POWER GENERATION

1) WIND POWER

The K-Means clustering method is used to cluster the days into distinct clusters based on factors such as wind power, weather conditions, and other relevant impact factors. Subsequently, a separate model is created for each cluster to predict wind power generation within that specific cluster [61], [62], [63], [64], [65].

A wind farm frequency regulation technique is developed by incorporating the k-means clustering algorithm, which considers wind uncertainties for spatially grouping generators [66]. Reference [67] proposed a method based on clustering algorithms, including K-means, to coordinate the scheduling of generators under wind power uncertainty. References [68] and [69] introduces a novel power curve modeling technique that utilizes the k-means algorithm.

In [70], the K-means clustering method is employed to identify outliers within each cluster. Subsequently, these outliers are then filtered out using an autoencoder applied to SCADA measurements. Reference [13] proposes using K-means clustering with meteorological data to strategically allocate wind power capacities in an interconnected power system, aiming to optimize the allocation of wind power resources.

2) SOLAR ENERGY

Using solar output, weather conditions and time as parameters, the K-means algorithm divides data into multiple clusters. Subsequently, a unique photovoltaic power prediction model is formed for each individual group [20], [71], [72], [73]. Also, this algorithm is utilized to cluster sky images, enhancing photovoltaic power forecasting accuracy [74], [75].

The K-means clustering method has been applied to identify high-quality and diverse reduced scenarios within a distribution network characterized by high photovoltaic penetration [76], [77], [78], [79]. Leveraging this algorithm, [80] developed a method for rating rooftop PV system capacity for buildings. Similarly, this algorithm clusters different cities based on their installed PV capacity [81], [82].

A K-means algorithm clusters PV systems with similar Effective Capacity Factor (ECF) behaviors to enhance Extreme Value Analysis (EVA) applicability in real-life scenarios of power systems with inadequate extreme PV data [83]. The K-means algorithm is used by [84] to identify isolated points within solar cells.

3) ELECTRIC POWER PLANT LOADS

Distributed energy systems represent the key elements of the new perspective of the smart grid. If, in the past, the flow of energy followed a linear and hierarchical logic, including generation, transmission, dispatching, and sale, this is no longer the case [85], [86]. The landscape has changed, and users have now become an active part of the generation phase, reflecting a more decentralized and participatory approach to energy management. This new structure has led to the creation of energy hubs, allowing consumers to function as power plants through renewable distributed generation [87], [88]. Still, they have to face several issues related to both load-side and source-side. To face this issue, several publications work to connect energy consumption to energy production [25], [89], [90]. Therefore, short-term predictions became a relevant topic of interest in the electric power plant loads field [14], [91], [92], [93], [94]. Finally, the inclusion of EVs in household loads can also be considered in the case studies performed, even though they add another degree of uncertainty to the load due to their potential integration into energy hubs [95], [96].

In this context, K-means clustering is combined with other algorithms and techniques to predict the load at both sides, load and source, to identify consumption patterns to match power generations [97], [98], [99], [100]. Currently, combinations of learning strategies are frequently employed, as they are considered more accurate and feasible than individual predictive models [101]. Moreover, with the high amount of available data, K-means clustering is also used to filter out data that are considered noise [102].

B. POWER TRANSMISSION AND DISPATCHING

1) ELECTRIC LINES

In the electric lines case study, K-means algorithms are primarily used in cascade fault prevention, allowing the decision maker to have an early warning [103]. Meteorological conditions, such as ice, can influence electric line performances, so K-means algorithms are utilized to perform an anti-ice prediction strategy [104].

Moreover, as in many other cases, the K-means algorithm is used during the PMU data processing process along the line, with the aim of solving the numerical dispersion of the line loss rate in the area covered by the transformer, also with high accuracy [105], [106], [107]. Linked to the operational monitoring of electric lines, the K-means algorithm is utilized to minimize the number of inspection teams dedicated to the maintenance of an electric line [108]. Focusing on the reliability of the electric lines, K-means algorithms are used to process data for the identification of the level of influence of lighting resistance on lightning arresters [109].

2) ELECTRIC LOAD FLOW

Electric load flow in modern applications is closely connected to the coordination of the elements within a grid, a coordination that becomes especially important in cases involving renewable energy sources. References [110], [111], and [112] provide a model for a direct probabilistic power flow, including renewables. In this context, [67], [113], [114] focus on predicting renewable production, and this estimation is achieved through the use of K-means algorithms. Moreover, the integration of a significant number of distributed energy resources into the electrical grid introduces new substantial challenges, where the K-means is employed to address them. In [115] the ultimate goal of using K-means is to identify the impact of harmonics generated by the integration of PV systems. Similarly, it is used in maximum frequency assessment, specifically to group the samples into multi-clusters by maximizing the membership degree of each training sample [116]. In [117] and [118] K-means is used again in cooperation with the distributed renewable energy sources, first for generation balancing, then for scenario selection, preserving simultaneous and chronological combinations of different loads and distributed energy resources. Reference [119] uses the K-means algorithm to cluster and provide a suitable linearization of the power-flow equations reducing the computational burden involved in determining terminal voltages for the clusters. Last, with the electrification of transport systems, the optimization of power flow became a primary interest also in this field [120].

3) FAULT DETECTION

In a hybrid anomaly detection model for electricity theft in smart meter systems, the K-means algorithm clusters customers exhibited similar behavior [121]. Similarly, this algorithm is utilized for anomaly detection in PMU streaming data [122]. Reference [123] proposed an ensemble learning model, which includes K-means, to detect attacks and anomalies in the power system. The same clustering algorithm is employed to categorize single phase-to-ground fault instances and identify the defective segments within the distributed network [124].

Damaged areas in PV systems are identified using an algorithm based on the K-means clustering technique [125]. This algorithm is employed to identify faults within the PV data using PVs and weather data [15], [126]. In [127], a novel anomaly detection technique that leverages the K-means algorithm to cluster behavior patterns of industrial components is proposed for application in hydropower plants. Reference [128] introduced a hybrid model employing the K-means clustering algorithm to identify and locate anomalous cells within lithium-ion battery packs.

Alternatively, [16], [17], [129] employed a K-means clustering approach, effectively trimming the original sample size of the dataset and thus enhancing the computational speed for fault detection in wind energy conversion and islanding approach for power systems.

4) REACTIVE POWER

With the increment of renewables in energy distribution networks, reactive power management became a relevant issue in many power systems applications [7], [130], [131]. With the aim of quantitatively identifying the influence of renewables access on voltage and reactive power operation, the K-means algorithm is used to process data on renewables energy production and load demand [132]. Thus, reactive power optimization problems with large-scale distributed generation are performed in [133], and the K-mean clustering algorithm is employed to identify the better initial clustering center. In this framework, [134] proposes an approximated AC model for simultaneous transmission expansion planning and reactive power planning depending on wind power investment, and the K-means clustering technique is used to reduce the scenario numbers.

The K-means clustering is also utilized with the final aim of compensation method for unbalanced voltage by active and reactive power control through a smart inverter [135]. Moreover, with the purpose of fluctuation reduction, K-means cluster, combined with logic operation techniques, are used to process offline data for training support vector machine [16], [136]. The K-means clustering is incorporated to process data linked to the State of Charge analysis in hierarchical planning in charging electric vehicles and the interaction between the power grid and transportation sector; this led to the minimization of reactive power losses [137].

C. POWER SYSTEM UTILITIES

1) ELECTRICITY CONSUMPTION

The K-means algorithm is utilized to determine groups of customers exhibiting similar patterns in electricity consumption, aiding in understanding the various types of regular behavior [121], [138], [139], [140]. Alternatively, [141] also incorporates temperature alongside electricity consumption data to analyze regional energy usage patterns.

K-means clustering is used to derive daily electricity consumption profiles from the electric energy data acquisition system based on the centroid centers [142], [143], [144], [145]. Similarly, household electricity consumption patterns have been identified by [18], [95], [146], [147], and [148]. By employing consumption pattern curves derived from this clustering technique, the typical use of appliances has been distinguished in [149].

2) ENERGY UTILIZATION

The variability of renewable energy sources and the electric load represents a major concern in modern power systems; providing a precise forecast of power output can help to reduce the cost of power system operation. References [62], [70], and [150] provide a model for short-term wind power forecasting, utilizing the K-means algorithm to cluster the inputs into different groups and to reduce noise in measurements. Similarly, [18], [146], [151], and [152] develop models to enhance energy efficiency and reduce electricity consumption in household applications. In these works, the K-means algorithm is employed to categorize typical electricity consumption patterns, types of users, and house characteristics.

Also, [89] proposed an optimization-designed distributed energy system based on load forecasting, and K-means clustering is used to predict the load both on the source and load sides to improve efficiency. However, these predictive solutions are not used exclusively for household applications. Reference [143] provides a methodology for discovering electricity consumption patterns in smart cities using this algorithm for energy utilization systems. The push to expand electrification in developing countries often focuses on microgrid applications. K-means clustering plays a role in this effort, being employed to identify dominant trends within the energy generation sector [153].

3) SMART METERS

The K-means algorithm has been utilized by [154] for monitoring appliance-level energy consumption in homes, deriving energy consumption profiles from smart meter data. By analyzing voltage measurements of smart meters, K-means clustering is deployed for identifying the phase (single or three phases) of the connected network [155], [156]. The clustering algorithm has been utilized to determine the optimal location for the Data Aggregation Point (DAP) to collect data from smart meters [157].

In [158] and [159], K-means is employed to identify neighbors with similar energy consumption patterns. Reference [160] proposes a method that uses the K-means algorithm and smart meter data to establish household load profiles for energy analytics. K-means is also applied to smart meter data from a university campus to establish representative monthly load profile patterns [161]. This algorithm is utilized in a study by [162], where they aggregate household energy consumption data sourced from smart meters. The study introduces a model aimed at reducing the peak demand of customers. I [19], [163] n, the centroids of each cluster serve as representatives for each group within the smart meter energy consumption data.

D. POWER SYSTEM PLANNING AND CONTROL

1) SCHEDULING

With the integration of distributed renewable energy sources and the consequent bidirectionality of the electricity grid, the planning and scheduling strategy becomes a topic of primary importance within the research field, for local distributors, and for infrastructure management [164]. In this context, several studies proposed coordinated scheduling of generators in order to relieve the grid from the randomness and volatility of renewables, such as wind and solar [67], [165], [166], [167], [168]. Moreover, the significance of energy hubs in the field is steadily increasing, as they embody a multi-carrier energy system catering to diverse types of energy demands. However, the unpredictability of renewable energy sources remains a major concern. Consequently, the scheduling process has become a key area of interest in exploring solutions to this challenge [169], [170]. Therefore, several applications, also different from renewable distributed power plants and energy hubs, such as smart buildings and EVs, play a relevant role in power flow scheduling. The K-means clustering also serves as a useful technique for data clustering in scheduling problems in these cases [101]. Additionally, K-means clustering is employed within this framework to classify typical scenarios, providing a foundation to propose a predictive model [171].

The integration of PHEVs (Plug-in Hybrid Electric Vehicles) can also present challenges for the electric network due to the unpredictability of their connections. Utilizing K-means clustering can help organize them into distinct fleets, mitigating this issue [172], [173]. The topic of scheduling is closely related to the aggregation and remuneration of services, where K-means clustering plays a key role in defining tariffs for specific periods of the week [174], [175].

2) NETWORK SECURITY

Following the trend of interest in power system applications, K-means clustering is being employed within the network security framework. In this context, the false data injection attack has emerged as a recent and relevant tendency in the state estimation field, where aspects like frequency regulation and market operations can be dramatically affected [176], [177], [178], [179]. Moreover, intrusion detection systems are of significant interest for identifying data anomalies in control signals and sensor measurements. Thus, K-means clustering algorithms are combined with power predictions to create a cluster-driven ensemble learning algorithm to address this issue [180]. The literature also presents examples of methods to identify vulnerable elements: determining the most vulnerable components is crucial for establishing a robust defense for the power grid [181]. K-means clustering is also employed in energy and trust management routing algorithms for mobile ad-hoc networks. This approach leads to identifying unstable clusters, which can be replaced by others, thereby implementing a self-configurable cluster method [182]. Lastly, using K-means clustering enhances network security and energy conservation in wireless communication systems [183].

3) OUTAGES

The reliability of the electrical network has become a crucial concern, with outages emerging as a significant issue to address in order to prevent losses of data, money, and services. In this context, prognostic models able to combine model-based and data-driven techniques are used both for anomaly detection and remaining useful life [8], [184], [185], [186]. Moreover, self-healing is becoming an essential feature for self-organizing networks to detect faults and provide a response and a recovery [187]. Also, transmission line outages caused by overtemperature are studied and determined using K-means clustering [188]. Failure predictions in power systems are therefore treated with supervised and unsupervised learnings techniques, such as K-means clustering, or using a combination of them [181], [189], [190].

In the Quality of Service (QoS) field, cross-layer resource allocation is also noteworthy; in [191], a model is proposed based on K-means that can define a statistical quality of service based on a fixed power allocation method. Moreover, in edge networks, QoS often involves bringing network resources closer to end devices. The speed, ease of deployment, and cost efficiency of Unmanned Aerial Vehicles (UAVs) make them promising technologies for the network. UAVs provide computational capability, enhance services to edge devices, and establish a better line-of-sight link with ground devices. Therefore, identifying the optimal height for UAVs to hover has become a valuable application [192].

Thanks to big data management, the ability to identify and eliminate erroneous data within the system has become more accessible. K-means clustering is employed to categorize the data gathered for various contingencies, such as blackout prevention and islanding. This approach has been proposed for use in multi-bus and multi-node systems [193], [194], [195], [196], [197], [198]. Additionally, K-means clustering algorithms are used in predicting load shedding to alleviate pressure on the supply during peak times [199]. Lastly, joint clustering, including using K-means, is involved in power allocation for managing crossroad congestion in cooperative vehicular networks [200].

4) RISK ASSESSMENT

Another challenge to face in the reliability of power system planning and control is risk assessment. Currently, in this research field, machine learning algorithms are employed for state evaluation and risk assessment for relay protections, ensuring the stable operation of the power system [201], [202], [203]. Addressing high safety risks in the new integrated energy system requires sophisticated computations for safe and smart electricity consumption. K-means clustering is utilized to create a smart and safe electricity consumption model for the integrated energy system, leveraging big data management [204], [205], [206]. Using a risk-based clustering method to identify Near Misses among safe scenarios is important since the possibility of recovering the combinations of failures in a tolerable time allows to avoid deviations to accident, reducing the downtime and its risk to the system [207]. K-means clustering methodology is also applied in the risk assessment of power transformers, which occupy a critical role in power systems, considering technical conditions and strategic importance of the units [208]. Risk assessment is a critical aspect of the electricity market, as it introduces uncertainties that cause fluctuations in electricity prices and complicate the measurement of losses and gains. Therefore, K-means clustering is employed to quantize the time series, accurately reflecting real-world conditions in the electricity market [209], [210].

E. ELECTRICITY MARKET

1) POWER MARKETS

The K-means clustering algorithm is employed to identify representative scenarios for optimizing the short-term electricity market [210]. To represent an entire year, [211], [212] utilized this algorithm and identified eight representative days, capturing solar irradiance, offshore and onshore wind speeds, as well as demand patterns. The K-means clustering algorithm is employed to categorize diverse loads, including EVs [213], [214], thereby simplifying the size of bidding optimization problems for an aggregator in the day-ahead energy market [215].

In a specific case study of Black Friday, the K-means algorithm was implemented for market segmentation, identifying potential customer zones to facilitate the formulation of effective marketing strategies [216]. Similarly, in [217], this algorithm is utilized to categorize all participating companies in the market based on their similarities, facilitating strategic involvement in competitive electricity markets. References [218], [219],and [220] employs a K-means clustering model to segment similar price zones into clusters, aiming to improve energy price prediction accuracy by creating separated models for each cluster. Likewise, In the study [221], the authors applied the K-means algorithm to group electric vehicles (EVs) based on their travel behavior patterns. This grouping allowed the creation of a predictive model for each cluster, aiming to forecast the day-ahead energy demand and manage uncertainty in the electricity market [169], [222].

2) LOAD FORECASTING

The K-means clustering algorithm is utilized to categorize users based on weather conditions or consumption levels, aiming to identify similarities among different load patterns [223], [224], [225], [226], [227]. Most studies employ this clustering algorithm to cluster users based on their load profiles. The goal is to construct separate models for each cluster, thereby enhancing the accuracy of their predictive models. In [6], [226], [228], [229], [230], and [231], the authors applied the same methodology of clustering similar energy consumption patterns to develop a prediction model for residential load consumption. Additionally, temperature data was utilized to cluster users in order to build load forecasting models [224].

Conversely, some studies utilize the K-means algorithm to extract new features based on consumption similarities, ultimately building a unified model for all clusters [223], [225], [232], [233], [234]. In this case, the new features derived from K-means serve as additional inputs to the model. This approach requires the machine learning model to interpret the relationship between each cluster and the output. The same approach was utilized in [235] and [236], where a household load consumption prediction model was developed by clustering similar load profiles.

References [102] and [237] employed the K-means algorithm to detect noise and outliers in load forecasting models. Reference [238] used this algorithm to decrease the complexity of consumption scenarios for more efficient load forecasting.

F. SMART POWER GRIDS

1) SMART GRIDS

The continuous exchange of information is a key aspect of smart grid implementation. The ability to collect and process vast amounts of data promotes the integration of renewable energy sources and enables proper distribution grid management. Additionally, with continuous population growth and constant urbanization, pressure on the national grid increases, leading to potential failures. In this context, Demand-Response (DR) programs can alleviate electrical demand during periods of stress, contributing to a more resilient and adaptive energy system. In this application, K-means clustering is applied to identify the clusters of DR power consumption [239], [240], [241]. Cluster analysis, including methods like K-means clustering, is widely used to analyze smart meter electricity demand data. By identifying patterns in electricity consumption, these techniques enhance predictions for DR programs, contributing to more efficient and responsive energy management [152], [242]. DR-based methods are also applicable for optimal scheduling applications within the smart distributed grid, particularly in cases involving the integration of EVs. Consequently, it helps coordinate charging and discharging schedules, balance the load, and maintain the grid's stability [173].

Considering the complex management requirements of smart grids, support management systems and operation planning tools, such as aggregators, become noteworthy in the research framework. K-means clustering is often utilized for scenario reduction, helping to simplify and focus the analysis on critical patterns and trends within the data [175], [243]. One of the key aspects of the smart grid is the renewables integration, and in this context, K-means clustering is widely used for predictions in renewable power production, especially in short terms cases [20], [73]. Also, in smart grids, the applications of clustering techniques, such as K-means, to synchrophasor data to determine the number of clusters formed for grid management [244]. Moreover, In residential applications, K-means clustering is used to identify user categories and minimize energy expenditure, facilitating targeted interventions for more efficient and equitable energy usage [151], [245], [246]. Finally, the energy pricing aspect is not negligible considering smart grid applications, where K-means is also used for the determination of a real-time pricing [247].

2) ENERGY EFFICIENCY

Energy efficiency improvement is a major concern in industrial, residential, and building applications, aiming to minimize costs, emissions, and losses while ensuring grid reliability. Demand response programs and storage systems are crucial instruments for achieving this goal. In [248], a model of an improved residential micro energy grid is presented, focusing on enhancing energy efficiency. The K-means algorithm is utilized for scenario reduction in this context. Similarly, identifying electricity consumption patterns can minimize wasted energy in the residential sector, with K-means being widely used for clustering typical electricity consumption patterns [146]. The necessity to improve energy efficiency is also highlighted because of renewable integration in the smart grid. Reference [249] proposes a model to increase wind farm efficiency by analyzing the parameters which influence the wake effect of wind turbines using K-means clustering. Additionally, dynamic user clustering and the utilization of UAVs for optimal power allocation are employed to enhance energy efficiency [250], [251]. There is a significant interest in smart building applications specifying innovative strategies to operate HVAC systems and reduce energy consumption, as they are among the main energy-consuming components in these applications. Thus, K-means clustering can be utilized to identify new energy-saving opportunities in high-efficiency buildings, such as offices [252]. By analyzing the data and patterns related to HVAC usage, K-means clustering can help optimize energy consumption and enhance overall energy efficiency in these buildings.

In [253], a joint resource allocation model and clustering algorithm are proposed for machine-to-machine communication systems. They convert the optimization problem into an optimal location problem and then utilize K-means clustering to obtain an effective clustering strategy. This approach aims to enhance the overall efficiency of the system by optimizing resource allocation and communication strategies.

G. ENERGY MANAGEMENT SYSTEMS

1) ENERGY STORAGE

With the increase in the number of renewables connected to the electricity grid, the reliability of the energy supply has become more volatile due to the unpredictable inhomogeneity of distributed generation [254]. To increase grid flexibility and to satisfy grid ancillary services, the addition of energy storage systems has become of primary importance [165], [255], [256], [257]. Energy storage applications are used both for household and industrial applications [215]. The appropriate and optimal sizing is a key aspect of the design of an efficient model. In this context, K-means clustering is used on customer net meter electricity data to limit the input net/gross meter energy data for the optimal sizing [159]. The management of energy hubs requires the presence of storage systems to enhance their flexibility, and K-means clustering is employed to reduce consumption scenarios [25].

This approach helps optimize energy usage and improve the overall performance of energy hubs by effectively managing and utilizing the storage systems. K-means clustering is applied to identify optimal storage device locations, reducing the number of scenarios while maintaining data correlation, leading to efficient placement for improved energy management [76], [78].

2) ENERGY MANAGEMENT

Energy management has become a crucial topic due to the growing use of renewable resources and distributed generation. The unpredictability of these sources, along with varying consumption patterns, poses challenges for the electrical grid and market. As a result, it has sparked interest among industries and researchers in effective energy management strategies to optimize and balance the system. Energy hubs must implement scheduled energy management to reduce uncertainties. Clustering methods like K-means are utilized to manage these uncertainties effectively [169], [190], [258]. Furthermore, energy hubs must coordinate the simultaneous operations of various generators and transmission infrastructures. Their goal is to optimally choose energy carriers to minimize costs while complying with environmental regulations and concerns [259], [260]. Also, the use of smart transformers integrated into the hybrid energy hub improves the quality of energy management, enhancing efficiency and control [261]. However, energy monitoring is primarily conducted through smart meters, which provide detailed energy consumption profiles when combined with data analysis algorithms [154].

Beyond buildings, energy management strategies are also utilized in the field of electric vehicles. For instance, in [262], K-means clustering is used with an adaptive energy management strategy to extend the range of electric logistics vehicles by classifying driving blocks through intelligent driving pattern recognition. In [95], a model is proposed that schedules the effects of EV movement and implements machine learning-based load forecasting to provide the electricity cost for a single household application. Additionally, K-means clustering techniques are used to identify extreme atmospheric conditions and their impact on electricity load profiles [263].

H. OPTIMIZATION

As power systems continue to advance with heightened levels of sensorization, the need for optimization becomes paramount in boosting energy efficiency. The K-means clustering algorithm has become a valuable tool within optimization, providing the capacity to identify distinct clusters, thereby facilitating predictions. This clustering method enhances optimization modeling by providing both reliability and the capacity to maximize performance. Optimization is a transversal field that includes PV and wind generation predictions, EVs and load consumption forecasting, and other different kind of stochastic models [70], [89], [158], [204], [214], [264], [265], [266]. K-means clustering, combined with other algorithms, helps to provide reliable and optimized forecasting. For instance, [72] proposed a model integrated with Differential Evolution Grey Wolf Optimizer to predict PV power generation.

Moreover, multi-objectives models are widely used combined with K-means clustering for the implementation of complex structures such as independent energy hubs: K-means clustering is used for the scenarios identification, and multi-objective model is utilized for the system optimization [258]. Similarly, [267] proposed a stochastic optimization approach for multi-energy microgrids after using K-means clustering for scenario reduction. A similar model is proposed in [259], where the micro-grid is powered by solar energy, and the optimization is conducted using the grasshopper optimization algorithm. due to the uncertainties associated with renewable energy sources, the optimization of integrated storage systems has become a significant area of interest.

In [255], a model for optimal sizing of storage systems is proposed. Following the clustering of the load profile by K-means, a bi-level optimization is performed that considers both cost minimization and power deviation. This optimization is carried out using a combination of meta-heuristic algorithms and mixed-integer programming for the intended purpose. Also, K-means clustering is used combined with Monte Carlo simulation for the optimal location of the distributed generation [268]. Additionally, the reduction of real power loss is achieved by combining K-means clustering with another optimization algorithm, specifically the enhanced brainstorm optimization algorithm, as detailed in [269].

Moreover, with the integration of electric mobility, EVs are no more negligible in power systems field [96], [270], [271]. Reference [23] proposed an optimized model for allocating EV charging stations, integrating them with renewables, where K-means clustering is used to highlight the connection between charging distance and user satisfaction degree. Similarly, [272] proposed optimized energy management strategies for fuel cell hybrid vehicles, where K-means clustering is employed to identify the optimal data set in clusters for the rules extraction of the rule learned-based energy management strategy.

Lastly, economic optimization for energy systems can also result from combining K-means clustering and optimization algorithms. K-means and discretization methods are used to represent scenarios affected by uncertainties, utilizing the modified non-dominated sorting genetic algorithm [273].

I. POWER QUALITY

The K-means method is utilized to identify the most critical scenarios for voltage dip assessment in [274]. In [275], voltage variation patterns at the sub-10-minute scale are extracted using the K-means algorithm. Similar harmonic voltage and current distortions in distribution networks are recognized and clustered using this algorithm in [9]. Reference [10] proposed a new model that uses the K-means algorithm to cluster and segment oscillographic records. K-means clustering is used to determine the optimal level of PV penetration, considering harmonic power quality constraints [115].

In [276], [277], [278] and [279], the K-means clustering analysis algorithm is applied to categorize and detect voltage sag, utilizing the historical data gathered from a large-scale grid. Likewise, [280] used this clustering algorithm to identify and locate the flicker sources in a non-radial power system. Power quality disturbances (PQDs) are determined and clustered using the k-means method [21], [281] with SVD [282].

J. ELECTRIC VEHICLES

The K-means clustering algorithm is used to categorize EVs battery cells [283], considering their charging and discharging states, to optimize their electrochemical performance. Similarly, spatiotemporal uncertainties associated with EVs are managed by K-means to maximize the expected marginal revenue of the Distribution System Operator (DSO) by using charging data in [284]. This algorithm was utilized in [285] to categorize driving styles to develop strategies to minimize the EV energy consumption. Also, K-means clustering was used to classify plug-in hybrid electric vehicles (PHEVs) based on their daily mileage and arrival and departure times [96], [172], [270], facilitating the integration of PHEVs.

In [286], the K-means clustering method is applied to group batteries based on their State of Charge (SOC) and State of Health (SOH) characteristics, facilitating the construction of an Energy Storage System (ESS) for microgrids. Reference [243] used the same clustering algorithm to reduce scenarios, taking into account factors such as renewable energy, EVs, energy demand, and energy storage. The objective was to enhance the electricity market within microgrids. Alternatively, this algorithm has been employed by [287] for generating electric vehicle charging scenarios to understand the distribution of EV charging current.

The K-means clustering was utilized by [120], [288], [289], and [290] to identify the optimal number, positioning and configurations of charging stations. The same algorithm was employed to depict the association between the distance of charging stations and the degree of user satisfaction [23]. Considering travel time, solar access availability, and energy consumption for solar-powered electric vehicles, [291] propose a unique route merging method that utilizes the K-means clustering algorithm to extract the most representative Pareto routes.

V. K-MEANS VALUES, ADVANTAGES, AND CHALLENGES IN MODERN POWER SYSTEMS

The K-means clustering is applied to various aspects of power systems, highlighting its unique value contributions, its benefits in solving complex problems, and the potential challenges or limitations that professionals should consider when implementing this technique.

A. APPLICATIONS AND VALUES

Table 2 summarizes the principal applications of K-means clustering in modern power systems based on the literature review presented in this paper. From this abstract summary, it is possible to identify the main applications of the K-means algorithm in various domains of power systems, such as energy management, forecasting, optimization, and power system planning. The main applications of K-means clustering in different power systems topics can be summarized as follows:

- Generation prediction: Clustering methods like K-means can segment data into distinct groups, facilitating the creation of specialized models for each cluster. This is particularly beneficial for predicting power generation from various energy sources, like solar or wind, where different clusters might represent different weather conditions or geographical regions.
- Outlier and anomaly detection: This algorithm can be used to detect abnormalities in data, such as faults, islanding conditions, or unexpected changes in load or generation. By identifying the typical patterns and deviations, maintenance can be more proactive, enhancing grid reliability.
- Scenario reduction and generation: Clustering helps in reducing the complexity of large datasets by categorizing them into manageable subsets. In various power system applications, it can be used to decrease sample size, improve computational speed, and facilitate more accurate simulations and optimizations. Moreover, K-means can be used to generate scenarios based on the clustering results, representing the entire dataset in a more manageable and structured form.
- Scheduling and demand-supply matching Clustering techniques assist in efficiently scheduling energy production and consumption. It helps in understanding the patterns of demand and aligning them with the generation, thereby aiding in the balance of supply and demand.
- Identifying dominant patterns: K-means can be used to identify and understand dominant trends in energy consumption, generation, and demand. By clustering similar behaviors, insights can be derived to inform energy-saving strategies and intelligent grid management.
- Optimal sizing and hosting capacity: For RESs like PV, wind, or even EV charging stations, clustering algorithms can help in determining the optimal size and location of installations. It can also assist in battery sizing, ensuring energy storage is aligned with consumption patterns and generation capabilities.
- Electricity pricing and optimization: Clustering enables a better understanding of consumption patterns, allowing more sophisticated pricing strategies. It can also be used in downtime management, minimizing losses by identifying optimal times for maintenance and upgrades.
- PQD identification and localization: Clustering methods can help in the precise detection and location of

TABLE 2. Summary of K-means applications in modern power systems.

Key area	Topic	Applications	
Power generation	Wind power	• Production prediction [61]–[65] • Frequency regulation [66] • Scheduling [67]	• Power curve modeling [68], [69] • Outlier detection [70] • Optimal allocation of onshore [13]
	Solar power	• Production prediction (k models for each clusters, sky image clustering) [22], [71]–[75] • Scenarios reduction [76]–[79]	• Hosting capacity clustering [80] • Data missing curation [83] • Isolated detection on solar cells [84]
	Electric power plant loads	• Energy hubs [25], [89], [90] • Short-term prediction [91]–[94], [14] • Electricity consumption and generation matching [95], [96]	
Power transmission and dispatching	Electric lines	• Fault prevention [103] • Line loss identification [105]–[107]	• Inspection optimization [108] • Lightning resilience [109]
	Electric load flow	• Probabilistic load flow [110]–[112] • Grid coordination [67], [113], [114] • RES integration [120]	• Harmonic mitigation [115] • EV integration [117], [118]
	Fault detection	• Smart meters and PMU [122] • Fault in distributed networks [123] • PV systems [125] • Hydropower plant [127]	• Lithium-ion battery [128] • Reducing sample size and improve computational speed in energy conversion systems [16], [17], [129]
	Reactive power	• Energy distribution network district [133] • Voltage fluctuations [134]	• Smart inverter [135] • Vehicle-Grid interaction [137]
Power system utilities	Electricity consumption	• Electricity consumption patterns [121], [138]–[140] • Typical use of appliances [18], [95], [146]–[148]	
	Energy utilization	• Power output forecast [62], [70], [150] • Energy efficiency improvement [18], [146], [151], [152] • Household applications [89]	• Smart cities applications [143] • Microgrids [153]
	Smart meters	• Optimal location for the Data [155], [156] • Aggregation Point (DAP) [157]	• Reducing the peak demand [161] • Energy consumption representative [19], [163]
Power system planning and control	Scheduling	• Distributed energy sources [67], [165]–[168] • Energy Hubs [101]	• Aggregation [169], [170] • Remuneration [174], [175]
	Network security	• False data injection attack [176]–[179] • Cluster-driven ensemble learning [180] • Identification of unstable cluster heads [182]	
	Outages	• Anomaly detection [8], [184]–[186] • Failure predictions [181], [189], [190]	• Quality of Service [191] • Islanding [193]–[198]
	Risk assessment	• Safety risk [201]–[203] • Near Misses [207] • Downtime reduction [207]	• Power transformers [208] • Electricity price [209], [210]
Electricity market	Power markets	• Short-term electricity market optimization [210] • Demand patterns clustering [211], [212]	• Energy price prediction [169], [222] • Day-ahead energy demand forecasting [215] • Strategic market involvement [216]
	Load forecasting	• Building model for each cluster [6], [226], [228]–[231] • Feature extractions [223], [225], [232]–[239]	• Data cleaning and outlier detections [102], [237] • Scenario reduction [238]
Smart power grids	Smart grids	• Demand-Response programs [239]–[241] • Electricity patterns consumptions [152], [242]	• Real-time electricity pricing [247]
	Energy efficiency	• Demand-Response programs [248] • HVAC [252] • Optimal location [253]	
Energy management systems	Energy storage	• Grid flexibility [165], [255]–[257] • Battery sizing [159] • Optimal location [76], [78]	
	Energy management	• Energy Hub [259], [260] • Electricity cost [95]	
Optimization	Optimization	• Optimal siting [265] • Optimal sizing [255] • Power loss minimization [269]	• EVs optimal integration [269] • Renewable integration [70], [89], [154], [204], [214], [264]–[266]

TABLE 2. ***(Continued.)*** **Summary of K-means applications in modern power systems.**

Power quality	Power quality	• Identifying and locating the Power quality disturbances (PQDs) such as Flicker, voltage sage, and voltage dip [21], [281] • Voltage variation patterns extraction [275] • Clustering harmonic voltage and current distortions [9]	
Electric vehicles	Electric vehicles	• Battery cells [283] • Categorize driving styles [285] • Optimal number, positioning and configurations of charging stations [120], [288], [289], [290] • Energy Storage System [286]	• Electricity market [243] • Reduce scenario [96], [172], [270] • Generate scenario [287] • Ruote representative [291]

disturbances in power quality. By understanding the typical patterns and clustering anomalies, prompt actions can be taken to rectify issues, maintaining the quality of power delivery.

- Integrating RESs: As renewable energy becomes a significant part of the energy mix, clustering helps in forecasting generation, matching supply with demand, and integrating these sources into the grid in a stable manner.
- Energy efficiency enhancement: By using clustering to understand consumption and generation patterns, systems can be optimized to minimize waste and improve energy efficiency. DR programs, storage system integration, and smart building applications.

The application of K-means or other clustering algorithms in power systems leads to enhancements in both modernization and optimization, supplying essential insights for creating a more resilient and intelligent grid. As technological innovation advances, these techniques are poised to become increasingly integral in shaping smart, sustainable, and efficient energy systems.

B. ADVANTAGES AND LIMITATIONS

K-means clustering has been applied successfully in numerous power system applications. Table 3 summarizes the advantages and disadvantages or challenges associated with this algorithm [11], [292]. The K-means algorithm is simple to understand and easy to implement, which also is ranked among the most rapid clustering methodologies. Moreover, in the context of big data, it presents a favorable time complexity of big $O(n)$ [11], [58], [293]. This feature indicates the capability of K-means to scale efficiently with larger datasets.

The K-means algorithm— capable of handling various types of features, including numeric and binary— yields clusters that are straightforward to interpret. Moreover, the algorithm exhibits an inherent ability to parallelize, enabling it to leverage multi-core processors or distributed computing systems. This attribute is particularly advantageous in big data applications.

The primary challenge with the K-means algorithm is the requirement to pre-determine the correct number of clusters (k), which is not automatically determined [11], [294]. This is particularly problematic when dealing with complex multi-dimensional power system data, where the ideal number of clusters is not always obvious. Several methodologies are available to estimate the optimal number of clusters, including the "elbow" method, a commonly utilized approach that considers measures like inertia, or silhouette score [44], [295]. However, in specific applications, the suitable number of clusters is frequently determined following a post-processing stage or via domain-specific knowledge. This introduces a certain level of subjectivity to the process. It highlights the importance of integrating reliable statistical techniques with a nuanced understanding of the context to achieve the most effective results.

The k-means algorithm is sensitive to the initial selection of centroids and can get stuck in local optima [11], [52]. This could lead to inconsistent results in power system applications, where system states can significantly vary based on time and environmental conditions. To mitigate these issues, the K-means++ algorithm was proposed, which employs a more ingenious initialization technique to choose the initial centroids, thereby enhancing the quality and stability of the final clustering solution. Furthermore, given that this algorithm relies on Euclidean distance, it exhibits sensitivity to outliers. A few extreme data points can skew the centroids, resulting in potentially incorrect cluster formations.

K-means assumes clusters are spherical (or globular), which may not hold true for many power system scenarios [11], [296]. For example, load profiles or power generation patterns could form clusters of various shapes, not necessarily spherical. Also, it assumes linear boundaries between

TABLE 3. **Advantages and disadvantages of K-means clustering algorithm.**

Advantages	Disadvantages
Simplicity	Choice of number of cluster (k)
Ease of implementation	Initialization sensitivity
Speed	Sensitivity to outliers
Efficiency	Spherical clusters
Scalability	Linear boundaries
Versatility	Equal variance assumption
Parallelizable	Local optima
Interpretability	High-dimensional data challenges

clusters, which may not accurately capture the complex relationships and transitions between different power system states or operating conditions. Moreover, this algorithm operates under the assumption that each cluster exhibits approximately equal variance. However, in power system applications, this assumption may not always accurately represent the data, potentially affecting the performance and interpretability of K-means clustering results. For example, consumer behavior can introduce significant variability within clusters. Usage patterns can vary greatly depending on the type of consumer (residential, commercial, or industrial), time of day, day of the week, or seasonal factors.

While K-means is generally regarded as scalable, it can face performance challenges when dealing with exceedingly high-dimensional data. Such scenarios are not uncommon in power systems, particularly with the rise of smart grids and the explosion of big data. Therefore, several solutions have been proposed in the literature to handle high-dimensional data in power systems before clustering, as it shown in Fig. 6 c). PCA, Kernel PCA, and autoencoders help reduce the dimensionality of the dataset, preserving as much relevant information as possible. Methods such as Singular Value Decomposition (SVD) [58], factor analysis, and feature selection techniques can simplify the data by extracting critical features, thereby improving the K-means algorithm's efficiency and performance.

C. K-MEANS VARIANTS

There are several variants of the K-means algorithm, as shown in Fig. 7, each with its own enhancements or adaptations to address some of the limitations of the original algorithm. For instance, Mini-Batch K-means efficiently handles larger datasets by leveraging a subset of data, also known as a "mini-batch," in each iteration, which accelerates computation significantly [297].

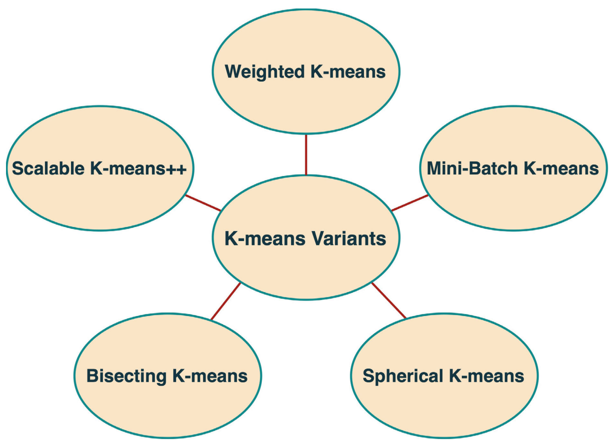


FIGURE 7. K-means variants.

The Spherical K-means variation is particularly effective with text data or any dataset transformable into a "tf-idf" representation, adjusting the standard Euclidean distance to cosine similarity [298]. Bisecting K-means starts with a single all-encompassing cluster and iteratively divides it into two, determining the split cluster based on the potential to reduce the sum of squared errors [299]. the Weighted K-means modification considers the varying importance of data points by assigning weights to each point, contributing to the calculation of cluster centroids [300]. It is noteworthy that K-means++ or PCA can be utilized as an initialization technique across all variants of the K-means algorithm. Finally, scalable K-means++, also known as parallel K-means, provides a scalable and parallelizable initialization algorithm that efficiently handles many clusters and dataset sizes [53].

VI. ALTERNATIVE CLUSTERING METHODS

Although the k-means algorithm is widely used in the literature, other clustering methods can surpass its performance in specific tasks in modern power systems. The principal categories of clustering algorithms encompass Centroid-Based Clustering, Hierarchical Clustering, Distribution-Based Clustering, Density-Based Clustering, Grid-Based Clustering, Graph-Based Clustering, and others, as shown in 7. Each category provides distinctive approaches and benefits, and their effectiveness can vary depending on the specific requirements of the data analysis task.

A. CENTROID-BASED CLUSTERING

Similar to K-Means, these types of algorithms assign data objects to the closest cluster based on the distance to the cluster's centroid. Some of the most commonly used algorithms are:

1) K-MEDOIDS

K-medoids, or PAM (Partitioning Around Medoids), is a variant of K-Means, offers greater resilience to noise and outlier data points. Unlike K-Means, which uses the mean value as the cluster's center, PAM designates the most centrally positioned data point within each cluster, known as the medoid, as the cluster representative. This method significantly mitigates the impact of outliers on the clustering process. For instance, in power systems, PAM could be utilized to categorize consumers based on their electricity consumption patterns [29], [301], [302]. As power usage data often contain noise and outliers due to varying factors such as weather conditions, time of day, or unexpected appliance use, PAM's robustness to such irregularities makes it an ideal choice for this application [303], [304].

2) CLARA

CLARA (CLustering for Large Applications) is an extension of the PAM method and is particularly suited for large datasets. Instead of finding medoids over the entire dataset, CLARA draws a small sample of the dataset, applies PAM on the sample, and defines the best medoids over the samples. After several iterations, the best clustering is chosen over all samples is reported as the final result. For example, in power systems with vast smart meter data [305], CLARA could

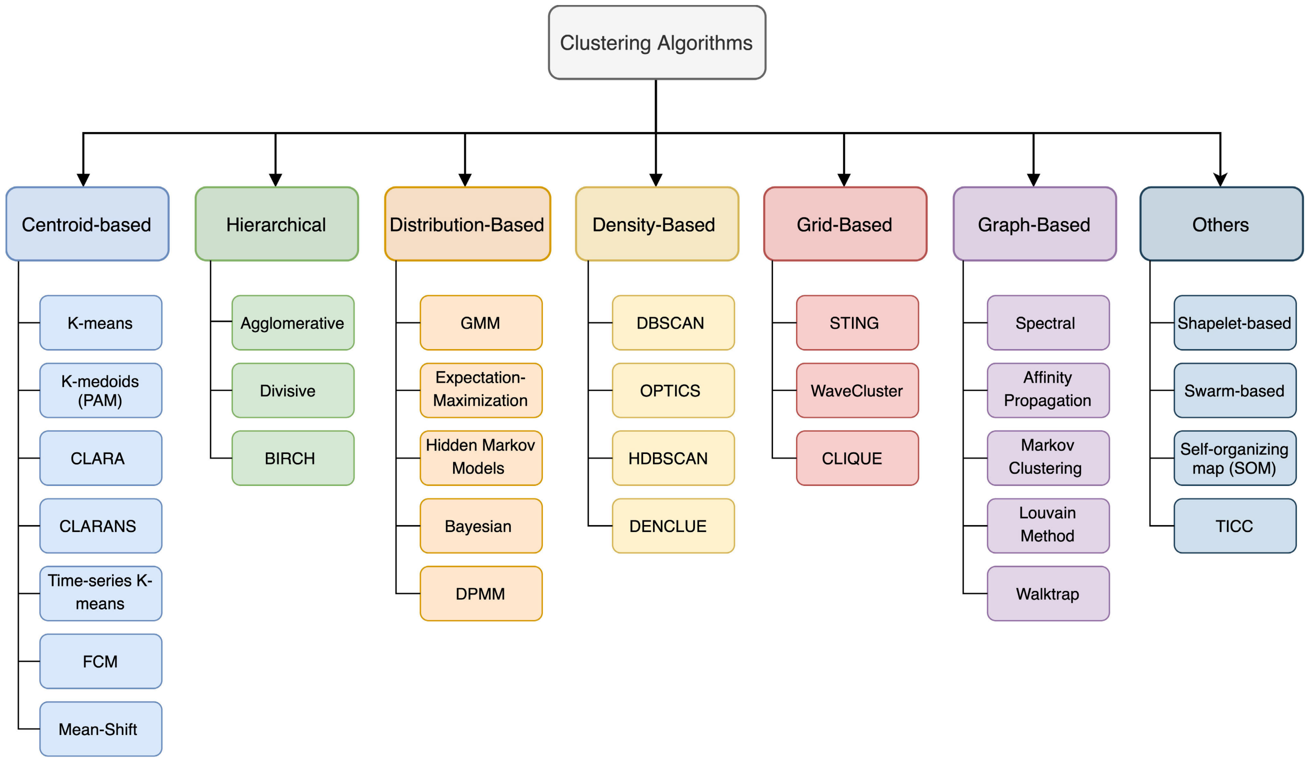


FIGURE 8. Clustering algorithms category in modern power systems: Centroid-Based Clustering, Hierarchical Clustering, Distribution-Based Clustering, Density-Based Clustering, Grid-Based Clustering, Graph-Based Clustering, and others.

efficiently identify unique consumption patterns by sampling and processing smaller subsets of the large dataset [306], [307], [308].

3) CLARANS

CLARANS (CLustering Large Applications based upon RANdomized Search) is an improvement over CLARA which considers a more randomized search instead of a sample-based search. It offers the flexibility of choosing the number of neighbors in each search step, leading to the possibility of faster convergence. In a power system context, CLARANS could simplify the identification of customer load profiles [309] or fault zones [310], [311] by quickly converging on distinctive patterns within the large, multidimensional datasets common in smart grid applications [28].

4) TIME-SERIES K-MEANS

This is another variant of the K-means algorithm explicitly designed to handle time-series data. It uses Dynamic Time Warping or other appropriate distance measures to compare time-series data which can vary in length and may exhibit shifts in time. In power systems, Time-series K-means can be applied to analyze consumption data or load profiles over time, aiding in demand forecasting [312], [313] and peak load management by identifying patterns and trends that might be obscured in a traditional K-means analysis [126], [314].

5) FUZZY C-MEANS (OR FCM)

Fuzzy C-Means (FCM), K-means variation, permits data points to have varying degrees of membership across multiple clusters, which proves beneficial when cluster boundaries aren't well-defined. In power systems, FCM could be useful in load profiling [315], [316] where consumption patterns might not distinctly fall into specific categories, thereby allowing consumers to be classified under multiple profiles based on their electricity usage habits [317], [318], [319], [320].

6) MEAN-SHIFT CLUSTERING

This algorithm updates candidates for centroids to be the mean of the points within a given region. These candidates are then shifted towards regions of the highest density, identified using a kernel density estimate. In power system applications, Mean Shift could be beneficial for detecting areas of high energy consumption or demand hotspots [321], [322], [323], providing valuable insights for power distribution and load management strategies [324], [325], [326].

B. HIERARCHICAL CLUSTERING

Hierarchical clustering algorithms progressively generate nested clusters, either by merging or splitting them. The primary types of hierarchical clustering are as follows.

1) AGGLOMERATIVE HIERARCHICAL CLUSTERING

Agglomerative Hierarchical Clustering employs a "bottom-up" methodology, initially treating each data point as an individual cluster and successively merging pairs as we ascend the hierarchy. In power system applications, such as smart grid analytics, this approach can facilitate grouping power consumers with similar energy usage patterns [327], [328], [329], thereby enabling better demand forecasting and load balancing strategies [26].

2) DIVISIVE HIERARCHICAL CLUSTERING

This method implements a "top-down" strategy. Initially, all data points are classified under a single cluster, and recursive divisions occur as one moves down the hierarchy. In power systems, this method could be valuable for breaking down large-scale power grid data into manageable subsets, such as regional [330] or local clusters, aiding in tasks like fault detection [331], load management [332], [333], and infrastructure planning [328].

3) BIRCH

BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) algorithm is designed to perform hierarchical clustering over large data sets. A standout feature of BIRCH is its use of a tree structure known as a Clustering Feature (CF) tree, which serves as an in-memory summary of the data distribution. This enables BIRCH to handle much larger data than the available memory. BIRCH works by scanning the dataset and updating the CF tree in one pass, with clustering decisions made without needing to revisit the actual data, which makes BIRCH a significantly efficient algorithm. BIRCH's efficiency in handling large datasets would be beneficial in a modern power system, where smart grid technologies generate vast amounts of data. For instance, BIRCH could be applied to group together customers based on their electricity usage patterns [334], [335]. By aggregating customers into clusters, utility companies could gain a more comprehensive understanding of demand patterns [335], [336], [337], which facilitates more efficient planning and operation of the power grid.

Linkage methods [338] in hierarchical clustering are strategies employed to calculate distances between clusters, influencing their arrangement in the hierarchy and determining which clusters to merge (in agglomerative hierarchical clustering) or divide (in divisive hierarchical clustering). The common linkage methods include [56], [338], [339], [340] Single Linkage, defining cluster distance as the shortest distance between points in each cluster; Complete Linkage, using the longest distance between points; Average Linkage, using the average distance between each point in one cluster to every point in the other cluster; Centroid Linkage, utilizing the distance between centroids of clusters; and Ward's Method [339], calculating distance as the increase in the summed square error upon merging two clusters. Each of these methods shapes the clusters distinctly, and their application relies on the unique characteristics of the dataset and the objectives of the clustering process.

C. DISTRIBUTION-BASED CLUSTERING

Distribution-based clustering algorithms model the data as if it was generated from a mixture of probability distributions. This type of clustering model assumes that the dataset is an outcome of mixed probability distributions [341], and the goal of the algorithm is to identify these distributions and their parameters. In this model, each cluster is represented by the particular distribution that generated it. A common examples of a distribution-based clustering algorithm are as follows.

1) GAUSSIAN MIXTURE MODELS

Gaussian Mixture Models (GMMs) operate on the probabilistic assumption that each data point arises from a mix of a finite number of Gaussian distributions, each with unknown parameters. In the context of power systems, GMM can be employed to model various scenarios like electricity consumption patterns [342], [343] or system reliability analyses [344], [345], [346], [347], given that these patterns often follow Gaussian or near-Gaussian distributions.

2) EXPECTATION-MAXIMIZATION

Expectation-Maximization (EM) algorithm is used with GMMs to estimate the parameters of the Gaussian distributions and to assign cluster memberships. In power systems, EM clustering can prove advantageous for tasks like system state estimation or load forecasting [27], [246], [348] where underlying data might align well with Gaussian distributions, enabling more precise identification of distinctive states or patterns.

3) HIDDEN MARKOV MODEL

Hidden Markov Model (HMM) is a statistical model, that assumes the system being modeled to operate as a Markov process with unseen or "hidden" states. They are often used in time series data to find underlying states or clusters. HMMs can be valuable in power systems for tasks like outage prediction [349], anomaly detection [350], [351], or identifying state transitions in load profiles over time [352], [353], [354], [355], as these tasks often involve underlying temporal patterns or states that may not be immediately observable.

4) BAYESIAN CLUSTERING

This algorithm leverages a probabilistic model to characterize data, using Bayesian statistical principles to estimate the model parameters [356], [357]. The Bayesian approach allows uncertainty within the modeling and can integrate prior knowledge. In power systems, Bayesian Clustering could be applied in scenarios where prior knowledge about consumer behavior, equipment performance, or energy generation patterns is available, helping to improve the accuracy of tasks such as demand forecasting [358], [359], failure prediction, or the detection of energy consumption patterns [360], [361].

5) DPMM (DIRICHLET PROCESS MIXTURE MODELS)

Dirichlet Process Mixture Models (DPMMs) are a type of infinite mixture model where the number of clusters is not predetermined and can grow with the data. They are often used when the number of clusters is unknown in advance. DPMM could be valuable in scenarios like identifying consumer groups or grid segments in power systems, where the precise number of clusters may be unknown or expected to change with new data [362], [363]. By employing DPMM, utilities could dynamically adjust their models to account for evolving patterns in energy usage or grid performance [364], [365].

D. DENSITY-BASED CLUSTERING

Density-based clustering algorithms group data points into clusters based on regions of high data density, distinguishing them from sparse regions. These methods, effective with spatial data or irregularly shaped clusters, identify clusters as areas where data point concentration surpasses a predetermined threshold. Their strength lies in their ability to discover clusters of any shape and effectively handle noise and outliers. Common density-based clustering algorithms include:

1) DBSCAN

DBSCAN (Density-Based Spatial Clustering of applications with Noise) algorithm defines clusters as continuous regions of high density. As one of the most widely-used density-based clustering algorithms, DBSCAN stands out for its ability to discover clusters of varied shapes and sizes, making it particularly qualified for handling complex data structures. For instance, in the power systems domain, DBSCAN can be used in identifying clusters of power consumption patterns [366], regions on the power grid that display similar characteristics [146], [367], [368], or anomaly detection [369] in grid operations, as it effectively discerns dense regions indicative of similar behavior or performance, in the presence of noise [370], [371].

2) OPTICS

OPTICS (Ordering Points To Identify the Clustering Structure) is a more advanced version of DBSCAN that can find clusters of varying densities. It orders data points in a way that spatially closest points become neighbors in the ordering. Unlike DBSCAN, it doesn't require a single density threshold, instead generating an augmented ordering of the dataset representing its density-based clustering structure. This characteristic allows it to adapt to the varying density of clusters, providing superior performance in scenarios with complex spatial distributions. OPTICS could be applied in power systems to detect clusters of consumers with varying energy usage densities. It could potentially identify high-consumption neighborhoods or commercial districts versus low-consumption areas, aiding in demand forecasting and grid management [372], [373], [374]. It can also be helpful in identifying clusters of power grid failures or anomalies [375] that might indicate issues requiring maintenance or infrastructure upgrades.

3) HDBSCAN (HIERARCHICAL DBSCAN)

HDBSCAN (Hierarchical DBSCAN) is an extension of DBSCAN that converts it into a hierarchical clustering algorithm, enabling it to find clusters of varying densities. Unlike traditional DBSCAN, which requires a predefined density threshold, HDBSCAN adjusts dynamically, leading to improved cluster identification, especially in datasets with diverse density distributions. In power systems, HDBSCAN could be advantageous in analyzing varying power consumption densities across different regions or times [376], [377]. For example, it could be used to discover clusters of households with similar usage patterns throughout the year [377], despite significant variability in consumption between summer and winter months. Moreover, it can assist in identifying irregularities or anomalies in power usage or system performance data, providing valuable insights for power grid maintenance and optimization [378].

4) DENCLUE

DENCLUE (DENsity CLUstEring) method is based on the concept of the density distribution function and can identify clusters of varied shapes and sizes [379]. This capability gives it an edge in handling complex datasets where clusters are not necessarily spherical or of similar size [380]. In power systems, DENCLUE could be particularly useful in tackling intricate datasets, such as those from smart grids or renewable energy sources. For example, it might be employed to cluster wind or solar power generation data based on varying output patterns, which could be irregular and of diverse scale due to fluctuating environmental conditions [381]. Additionally, it can assist in creating distinct user profiles for energy consumption, even when these patterns are inconsistent or irregular, thus aiding in efficient energy management and planning.

E. GRID-BASED CLUSTERING

Grid-based clustering is a technique that segments the data space into a finite number of cells (or set of cells), forming a grid-like structure [382]. Instead of performing operations directly on the data objects, the technique operates on the grid cells, significantly expediting processing time. As the time complexity is dependent solely on the number of cells in the quantized space rather than the number of data objects, this approach is especially effective for handling large spatial databases. The most common grid-based clustering algorithms include:

1) STATISTICAL INFORMATION GRID

Statistical Information Grid (STING) uses a hierarchical grid structure where each cell at a given level of the grid is partitioned into several smaller cells at the next level [383]. Statistical information is stored at each cell in the grid, allowing for efficient processing of queries. STING can be valuable in power systems for detecting patterns or anomalies in large spatial data sets, such as the geographical distribution of power line faults or the spread of energy consumption across a large region [382], [384]. By breaking down the data into a grid structure, this algorithm can quickly process and identify high-fault frequency or heavy power usage clusters, thus aiding in system maintenance and planning decisions.

2) WaveCluster

This grid-based algorithm employs wavelet transformation to identify dense regions in the feature space for clustering [385]. The transformation allows for multi-resolution clustering and noise removal, enhancing the effectiveness and robustness of the algorithm. WaveCluster can be leveraged in power systems for multi-resolution analysis of power consumption data. For instance, it can identify clusters of consumers exhibiting similar consumption patterns at different temporal resolutions (hourly, daily, monthly) [386], assisting demand forecasting. Moreover, its noise removal capability can help filter out transient spikes in the data, providing a more accurate representation of typical consumption patterns [387].

3) CLIQUE (CLUSTERING IN QUEst)

CLIQUE (CLustering In QUEst) identifies dense cells in a grid structure and then finds adjacent dense cells to form clusters. It is unique in its ability to find subspaces of the highest dimensionality such that high-density clusters exist in these subspaces. In power systems, CLIQUE can be used in identifying high-density areas of power usage, essentially identifying clusters of consumers with similar power usage behaviors. Thus, it can help in effective load forecasting, optimal power distribution, and identifying potential areas for the expansion of the grid infrastructure. By finding high-dimensionality subspaces, it can also aid in understanding multi-dimensional relationships between different factors affecting power consumption, such as time, weather conditions, and type of usage [388], [389], [390].

F. GRAPHED-BASED CLUSTERING

Graph-based clustering, or network clustering, is a method where data is represented as a graph or a network. Each data point is considered a node, and the relationships or similarities between these nodes are represented as edges. The aim is to find clusters or communities within this network structure. These clusters often represent groups of nodes more highly connected to each other than the rest of the network. Here are some common graph-based clustering algorithms:

1) SPECTRAL CLUSTERING

This method uses the spectrum (eigenvalues) of the similarity matrix of the data to perform dimensionality reduction before clustering in fewer dimensions. The similarity matrix is provided as an input and consists of measures of the similarity of each pair of points in the dataset. In power systems, Spectral Clustering can be highly useful. Assessing the similarity in power usage [391], [392] or power generation patterns across different nodes in the power grid, it can facilitate the identification of groups of nodes with analogous behaviors [393], [394], [395]. These insights can assist operators in enhancing grid efficiency, planning maintenance, or predicting and mitigating potential disruptions.

2) AFFINITY PROPAGATION

This algorithm uses a graph-based approach where each data point sends messages to other points, communicating its 'affinity' for being its exemplar (representative). This results in a collection of exemplars that represent the centers of each cluster and the respective assignments of data points. The process concludes with a set of identified exemplars that signify the center of each cluster and the corresponding association of individual data points to these clusters. In power systems, Affinity Propagation can be used in identifying patterns [396], [397] or groupings within smart grid data, such as usage patterns across various customer segments or the identification of similar performance metrics across different regions of the grid, thereby aiding in more efficient system management and resource allocation [398], [399], [400].

3) MARKOV CLUSTERING

Markov Clustering (MCL) algorithm simulates random walks within a graph through an alternating process involving two operations: expansion and inflation. It has been widely used in bioinformatics but can be applied to any graph [401]. In power systems, MCL could be used for network analysis, identifying closely connected nodes or components in the system based on usage or functionality. Therefore, it can aid in system optimization, fault detection, and network upgrade or expansion planning [402], [403], [404].

4) LOUVAIN METHOD

This method is specifically used for community detection in large networks. It optimizes the modularity of the network iteratively to identify communities. In power systems, this method can be used to determine groups of interconnected assets or subsystems based on their mutual interactions, dependencies, or commonalities [405], [406]. It can facilitate an understanding of network dynamics, simplify fault localization, and help develop efficient energy distribution or maintenance scheduling strategies [407], [408].

5) WALKTRAP

This algorithm finds communities by performing random walks. The basic idea is that if you perform random walks on the graph, then short walks are likely to stay in the same community because only a few edges lead outside a given community [409], [410]. In power systems, Walktrap can help identify tightly interconnected groups of assets or nodes. Such information can be useful in various ways, such as optimizing load distribution, enhancing network resilience, or understanding localized effects of faults or disturbances. It can also aid in the planning and management of renewable energy sources integration within traditional power grid infrastructure.

G. OTHERS

Aside from the previously discussed categories of clustering algorithms, there are other commonly used algorithms, such as shapelet-based, swarm-based, and Self-Organizing Maps (SOM) clustering, that do not neatly fall into the predefined groups.

1) SHAPELET-BASED CLUSTERING

This algorithm represents a distinct approach predominantly used for time series data [411], [412]. This technique identifies shapelets, which are representative subsequences within a time series, facilitating the discovery of inherent patterns in the data [413]. Such methodology can offer insightful applications in various fields, including power systems, where it could be deployed for identifying patterns in power consumption or predicting anomalies, thereby enhancing overall system efficiency and reliability [414], [415], [416].

2) SWARM-BASED CLUSTERING

Swarm clustering, also known as swarm intelligence-based clustering [417], [418], [419], applies the concepts of swarm intelligence—inspired by decentralized, self-organized systems like ant colonies or flocks of birds—to cluster data. It employs swarm algorithms, wherein numerous simplistic agents or "particles" maneuver through the data space, adjusting their positions based on a basic rule until a satisfactory clustering solution is achieved. Popular swarm optimization techniques such as Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) can be employed to enhance the results of clustering operations [420], [421]. Swarm clustering can be used in power systems to improve load forecasting by segmenting users or regions based on electricity usage, simplifying state estimation by grouping similar electrical buses, optimizing power flow by clustering power generation and load points, and assisting in fault detection by analyzing historical fault data clusters [422], [423], [424], [425].

3) SELF-ORGANIZING MAP CLUSTERING

Self-Organizing Map Clustering (SOM) is an unsupervised neural network used for clustering and visualization. SOMs employ artificial neural networks to map high-dimensional data into a low-dimensional (usually 2D or 3D) space. This technique arranges similar data points close to each other in this reduced space, thus forming clusters of similar data points. In power systems, it enhances load forecasting by clustering customers based on their energy usage and electricity consumption patterns [426], [427], [428]. Additionally, it can be used in fault detection systems, maintain system security, and improve power quality [429], [430], [431], [432].

4) TOEPLITZ INVERSE COVARIANCE-BASED CLUSTERING

Toeplitz Inverse Covariance-based Clustering (TICC), a technique proposed in 2017, offers a unique approach to handling time series data. It is designed to discover communities or clusters in multivariate time series by leveraging the inverse covariance structure of the data. An essential feature of this method is that it assumes the time series data to be stationary and the covariance structure to be Toeplitz [433], [434]. In power systems, TICC can be employed to tackle various challenges such as load forecasting and profiling [435], anomaly detection, and integration of renewable energy [436].

Although there is great potential for relatively new clustering algorithms to be applied in modern power systems, they are not currently utilized. Reference [437] presents an innovative density-based clustering algorithm that classifies data elements based on similarity. Inspired by the K-medoids method, this approach focuses solely on the distances between data points. Like DBSCAN and the mean-shift method, it can identify non-spherical clusters and automatically determine the appropriate number of clusters. The algorithm defines cluster centers, similar to the mean-shift method, as local maxima in the density distribution of data. Also, it automatically detects and excludes outliers from the analysis. This method offers a versatile and effective approach to cluster analysis with applications spanning various domains, including astronomy, bioinformatics, bibliometrics, and pattern recognition [11], [438].

This paper [439] introduces a novel non-parametric clustering method based on the concept that each latent cluster consists of layers surrounding its core, with the outer layers acting as border points that effectively delineate the clusters. Unlike previous approaches like DBSCAN, which directly define cluster cores by density, this method uncovers latent cores by gradually peeling border points. By examining the density of local neighborhoods, it identifies border points and links them to inner-layer points. This peeling mechanism adapts to the local density and inherent characteristics, effectively distinguishing neighboring clusters, even when they exhibit varying densities [440].

Reference [441] introduces the Multistep Three-Way Clustering (M3W) algorithm, which addresses limitations in existing three-way clustering approaches. M3W employs a progressive erosion strategy to build a multilevel data structure, enabling lower levels to gather more information from higher levels. Additionally, it incorporates a multistep three-way allocation strategy that considers neighborhood information for eroded instances. These combined techniques enhance the algorithm's ability to adaptively capture a dataset's inherent clustering structure by gradually acquiring more knowledge, thereby increasing the likelihood of accurate assignments.

VII. ANALYSIS OF CLUSTERING ALGORITHMS PERFORMANCE

This section evaluates the performance of clustering algorithms other than K-means. The selected clustering algorithms from each category include agglomerative, Bayesian, DBSCAN, CLIQUE, Spectral, SOM, and time-series K-means, which are among the most commonly used.

Seven distinct time-series waveform patterns commonly observed in power systems were synthesized to create the input dataset, with each pattern consisting of sub-10-minute intervals and one data point recorded per second. The primary patterns consist of the sinusoidal, sawtooth, square, linear ramp down, exponential decay wave, and constant value. From each pattern, a random selection of 500-800 samples is generated by altering both the phase (along the x-axis) and the amplitude (along the y-axis) while adding Gaussian noise to them. These approaches preserve the temporal relationships and patterns in the data, which is crucial for time-series analysis. Fig. 9 demonstrates the random samples derived from the sinusoidal pattern.

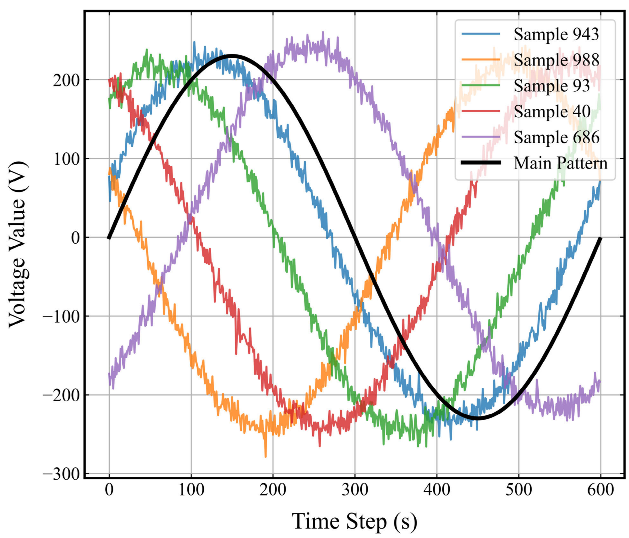


FIGURE 9. Five random samples derived from the sinusoidal pattern.

The final dataset, comprising 5037 samples with seven different classes, is employed to evaluate the performance of the selected clustering algorithms. The dataset is normalized using the standard z-score method before modeling. Fig. 10 illustrates the dataset in three dimensions using the t-SNE visualization technique, which is utilized to project the high-dimensional input data (with 600 dimensions) onto a 3-dimensional space, resulting in a more compact representation. As seen in this figure, each cluster exhibits a distinctive

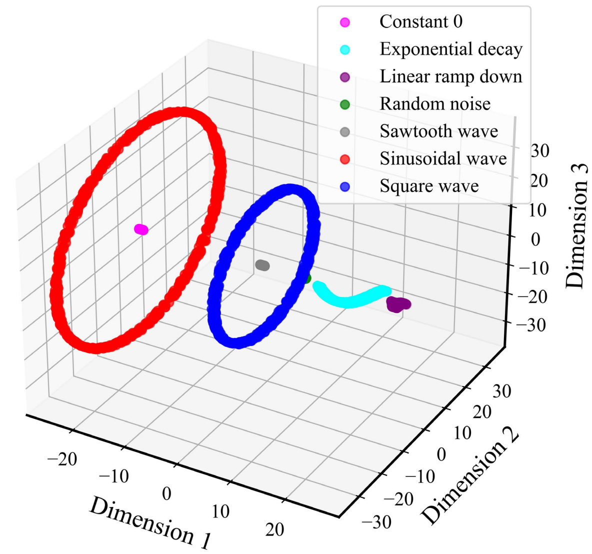


FIGURE 10. The input dataset in three dimensions. Each class is depicted using a unique color.

shape, which can pose a challenge for clustering methods to identify them accurately.

The number of clusters is pre-determined and remains consistent at seven for all the clustering models; this is the only information provided to the models. The hyperparameters of all models are optimized using the Bayesian optimization method. For algorithms such as DBSCAN, which automatically determine the number of clusters, their hyperparameters are tuned and optimized to identify precisely seven clusters within the dataset.

The accuracy of models is assessed using the Normalized Mutual Information (NMI) metric since the true class labels for the clusters are available. A normalized score ranging from 0 to 1 is provided by NMI, with higher values indicating a stronger alignment between the clustering results and the ground truth. Importantly, it should be noted that the NMI score remains unchanged when the class or cluster labels are shuffled or permuted in any way.

The performance of the chosen clustering algorithms is evaluated with and without KPCA using an RBF kernel for dimensionality reduction, as presented in Table 4. Both accuracy and elapsed time are considered as evaluation metrics.

Fig. 11 showcases the clustering outcomes achieved by eight selected clustering algorithms after applying KPCA, utilizing the initial 50 principal components. In this paper, the CLIQUE algorithm is specifically designed for processing two-dimensional data. This CLIQUE implementation segments the 2D space into a grid of cells or partitions. In contrast, when the CLIQUE deals with higher dimensions, the grid comprises hyperrectangular cells, each being defined by a set of coordinates in n-dimensional space.

As presented in 4, the introduction of KPCA has shown noticeable enhancements in both computation time and accuracy across the models. Notably, Bayesian Gaussian Mixture

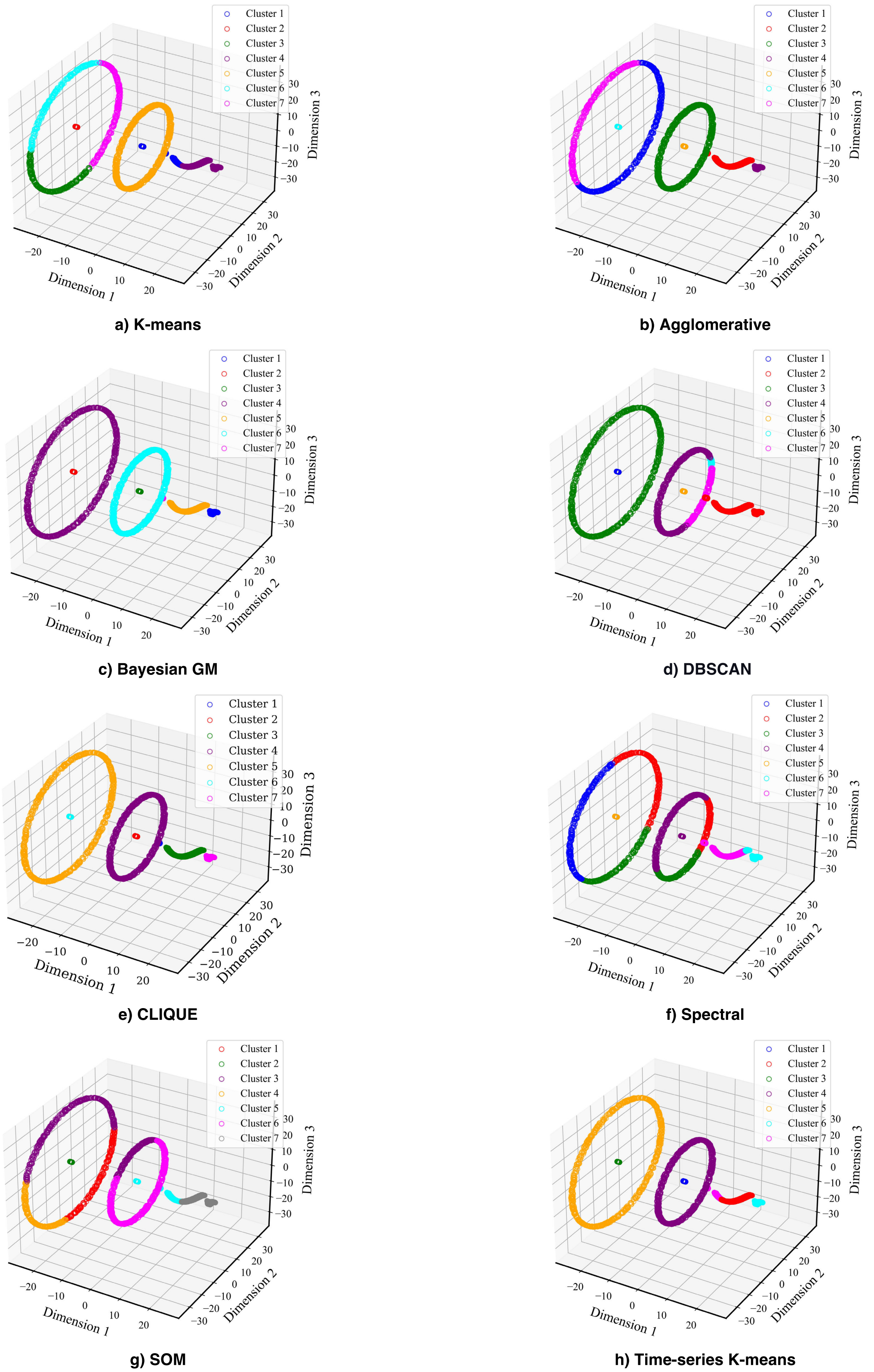


FIGURE 11. The clustering results using KPCA and selected algorithms: a) K-means, b) Agglomerative, c) Bayesian GM, d) DBSCAN, e) CLIQUE, f) Spectral, g) SOM, and h) Time-series K-means.

TABLE 4. Performance evaluation of clustering methods.

Algorithm	Without KPCA		With KPCA	
	Accuracy	Time (s)	Accuracy	Time (s)
K-means	0.6769	2.33	0.7962	1.45
Agglomerative	0.7168	3.50	0.9212	0.54
Bayesian GM	0.7297	8.11	1	2.53
DBSCAN	0.6632	1.39	0.8300	0.24
CLIQUE *	-		1	0.12
Spectral	0.4702	15.29	0.7728	10.97
SOM	0.4696	0.52	0.7506	0.08
Time-series K-means	0.9245	13410	0.9561	292

* The CLIQUE used in this paper works only with 2-dimensional data.

(Bayesian GM) achieved its highest accuracy post-KPCA integration. In terms of computational efficiency, the Time-series K-means model experienced the most significant benefit from input dimensionality reduction, with a remarkable reduction from nearly 4 hours to just 5 minutes—representing over a 97% reduction in processing time. However, its accuracy exhibited a more modest improvement, advancing from 0.9245 to 0.9561.

As shown in Fig. 11, it becomes evident that numerous algorithms encountered challenges when it came to accurately distinguish the random noise cluster, which is situated in close proximity to both the exponential decay and sawtooth clusters.

Both the Bayesian GM and CLIQUE clustering algorithms demonstrated precision in correctly clustering all data points, whereas Time-series K-means exhibited strong performance both with and without KPCA. The table emphasizes that several algorithms surpass the performance of the traditional K-means algorithm. This suggests that considering alternative clustering algorithms to K-means is essential when tackling clustering tasks.

In this study, Python served as the primary programming language and a range of libraries and packages were utilized. These included NumPy and Pandas for data pre-processing, Scikit-learn for modeling and optimization, and Matplotlib and Seaborn for crafting visualizations. For this research, the computational resources employed consisted of an i7-8700K CPU with 16 GB of RAM and an Nvidia RTX GeForce 2080 GPU with 16 GB of RAM.

VIII. DISCUSSIONS AND FUTRE TRENDS

Clustering algorithms, specifically K-means, hold a crucial role in modern power systems. Their wide-ranging applications highlight their importance, including load forecasting, fault detection, power quality analysis, and system security

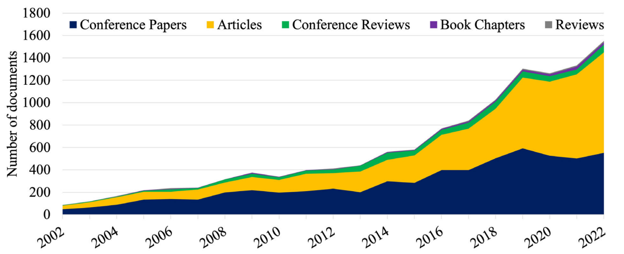


FIGURE 12. Publications in power systems using clustering algorithms.

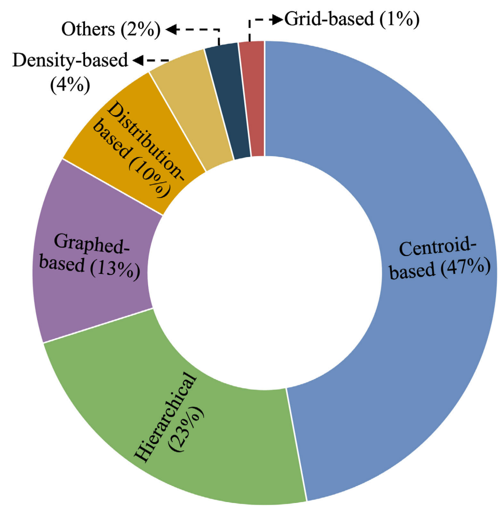


FIGURE 13. Contributions of the various clustering categories to the power systems literature up until 2022.

assessment. Furthermore, as power systems become more complex with the integration of RESs, EVs, and DRs, the significance of clustering algorithms in identifying patterns, segmenting the system, and optimizing operations is only set to grow. Moreover, since more data has become available and data-driven methodologies have demonstrated promising solutions in recent years, there has been a corresponding growth in the use of these methods within the power system domain. Therefore, as shown in Fig. 12, the number of publications using clustering algorithms in power systems is growing as well. This growth has surged by 4.5 times between 2012 and 2022. The results presented in Fig. 12 are based on the keywords "clustering" and "power systems," sourced from the Scopus library. This figure also contains all clustering algorithms, including customized ones or those specifically proposed for particular problems in power systems.

Fig. 13 illustrates the distribution of different categories of clustering algorithms used in the power system literature. As observed in this figure, centroid-based algorithms are prominent, contributing to 47% of publications, primarily thanks to the K-means algorithm, which is the most frequently used. Following that, hierarchical clustering algorithms (with 23%) have been employed extensively, succeeded by graph-based algorithms. The other categories have

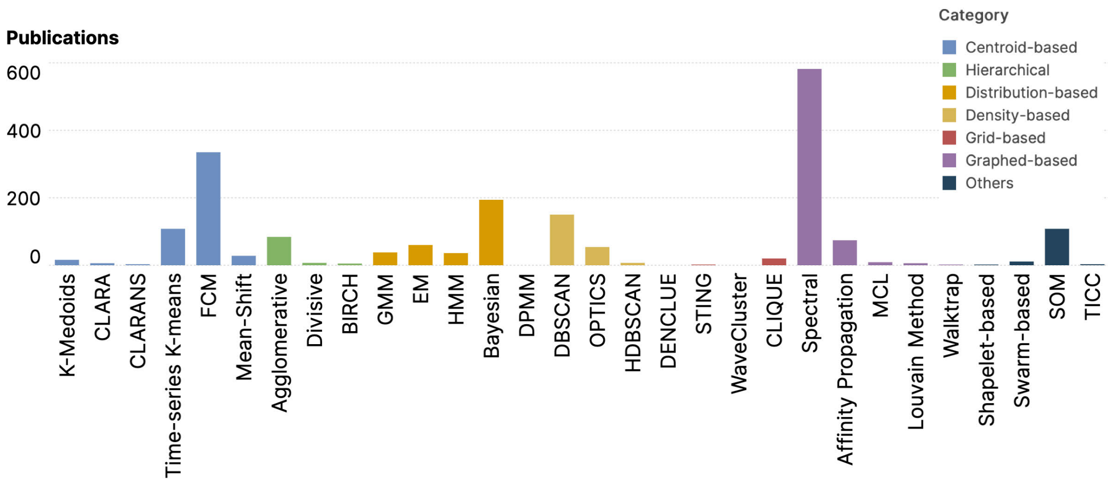


FIGURE 14. The contributions of various clustering algorithms to the power systems literature until 2022 based on the Scopus database. In this figure, the K-means algorithm has been excluded as it skewed the scale of the figures with a substantial number of 1,929 publications.

fewer contributions, indicating that they are not as extensively explored or utilized in the literature. For instance, grid-based algorithms make up only 1% of contributions to publications in the field of power systems, emphasizing their relatively limited adoption in this domain despite their potential applicability. This distribution not only highlights the popularity of certain methods but also suggests potential areas for future research and development in underexplored clustering techniques within power systems.

Fig. 14 illustrates the contributions of various clustering algorithms to the power systems literature up until 2022, as derived from the Scopus database and discussed in this article. In this figure, the K-means algorithm has been intentionally excluded, as its substantial count of 1,929 publications skewed the scale of the figures, thus potentially affecting the comparative analysis. After K-means, the graph-based spectral clustering algorithm has been employed most frequently, with almost 600 publications, and it is closely followed by Fuzzy C-means, which has been featured in more than 300 publications. As shown in this figure, other clustering algorithms are less widely used, indicating a prevailing tendency toward utilizing K-means over alternative methods. This observation highlights the dominant position of K-means within the clustering methodologies applied to modern power systems.

However, while K-means proves to be powerful, it has inherent limitations, such as sensitivity to initial starting conditions and the requirement of pre-determining the number of clusters, often suggesting the exploration and utilization of alternative methods for the highest clustering solution. In the case study presented in Section VII, the CLIQUE algorithm demonstrated superior performance over K-means, achieving 100% accuracy in identifying all clusters and samples. This performance highlights the importance of delving into less typical algorithms within the power system domain, for example, grid-based algorithms, which constitute only 1% of contributions in this research area.

Therefore, one of the emerging trends in the application of clustering algorithms in power systems could be the exploration and implementation of less common or newer clustering techniques. By tackling conventional methods and applying these innovative algorithms to power system problems, there may be an opportunity to enhance the overall performance of the systems. This approach not only fosters innovation but also encourages researchers and engineers to challenge existing paradigms, potentially leading to breakthroughs in efficiency, reliability, and adaptability within the power systems domain.

Another possible trend may be the application of clustering algorithms specifically designed to work with time-series datasets. The time-series nature of many electrical variables in power systems, such as voltage and current, emphasizes the importance of utilizing specific clustering algorithms designed to handle time-series data. These specialized algorithms can capture the temporal dependencies and patterns inherent in power system data, which standard clustering methods may overlook. This capacity for handling time-series data expands their utility in various power system applications, from load forecasting and anomaly detection to integrating and managing renewable energy sources.

Several clustering algorithms are recognized in the literature for effectively handling time-series datasets. These include K-means, PAM, Time-series K-means, Fuzzy C-means, DBSCAN,Hidden Markov Models, Spectral Clustering, and Shapelet-Based Clustering. Additionally, Self-Organizing Maps and the innovative Toeplitz Inverse Covariance-based Clustering have been recognized for their applicability in time-series analysis.

Dynamic time warping is critical in time-series analysis and clustering because it can align temporally shifted or distorted sequences. It provides flexibility in comparing different lengths and warped sequences and can handle non-linear alignments. Its robustness to noise and ability to integrate with other methods make it applicable across various fields like speech recognition, finance, and healthcare. By considering the entire sequence and its dynamics, DTW enhances clustering and enables more accurate and insightful data groupings.

Time-series K-means and Shapelet-Based Clustering utilize dynamic time warping as a distance metric among these algorithms. While other methods like PAM, DBSCAN, Hidden Markov Models, TICC, SOM, and Spectral Clustering do not typically employ DTW in their original versions; however, DTW could potentially be integrated as a distance measure in these algorithms, thereby adapting these algorithms for time-series data. As demonstrated in Section VII, the Time-series K-means algorithm outperformed the original K-means clustering algorithm, achieving an NMI accuracy of 0.9245 compared to the NMI accuracy of K-means with 0.6769 when applied to the original dataset. This improvement can be attributed to the utilization of the DTW similarity measurement.

K-means also should be used as the benchmark to compare the performance of other clustering algorithms. Its simplicity and efficiency make it a good starting point for any clustering task, especially with large datasets. As one of the most widely used clustering algorithms, K-means is often chosen for comparison to measure the relative performance of newer or less-known algorithms. Its results are easy to interpret, contributing to its ubiquity and utility in explaining clustering tasks' outcomes. It provides robust baseline performance, setting a high standard that any advanced algorithm must exceed to demonstrate its effectiveness. Furthermore, its adaptability to various data types and applications in diverse domains indicates its role as a fundamental reference point in clustering analysis.

IX. CONCLUSION

The application of K-means in modern power systems provides essential technical insights into load forecasting, fault detection, power quality analysis, system security assessment, and other applications. Its simplicity, efficiency, and adaptability make it a benchmark in clustering tasks. However, while K-means offers several advantages, its limitations and challenges must be addressed. It can be sensitive to initialization and may converge to local optima, which presents concerns in complex and large-scale systems. These inherent challenges drive the investigation and implementation of alternative clustering methods, such as K-medoids, Time-series K-means, BIRCH, Bayesian clustering, HDBSCAN, CLIQUE, SPECTRAL, TICC, and SOM, among others. These alternatives provide a broader perspective and flexibility in addressing the unique demands of modern power systems.

The exploration of time-series clustering in modern power systems emphasizes the importance of techniques such as dynamic time warping, allowing non-linear alignment of temporally distorted sequences. It provides a significant advantage in handling complex time-series data, which is becoming increasingly common in modern energy systems. Moreover, continual data availability and computational power growth indicate a promising future for clustering algorithms in power systems. The trends suggest a growing tendency towards exploring newer clustering techniques and an enhanced focus on time-series data analysis. The future of K-means and other clustering algorithms in modern power systems is poised to play a crucial role in shaping sustainable, efficient, and intelligent energy systems, providing a platform for continuous innovation and optimization.

ACKNOWLEDGMENT

This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

REFERENCES

[1] M. S. Ibrahim, W. Dong, and Q. Yang, "Machine learning driven smart electric power systems: Current trends and new perspectives," *Appl. Energy*, vol. 272, Aug. 2020, Art. no. 115237.

[2] S. M. Miraftabzadeh and M. Longo, "High-resolution PV power prediction model based on the deep learning and attention mechanism," *Sustain. Energy, Grids Netw.*, vol. 34, Jun. 2023, Art. no. 101025, doi: 10.1016/j.segan.2023.101025.

[3] S. M. Miraftabzadeh, F. Foiadelli, M. Longo, and M. Pasetti, "A survey of machine learning applications for power system analytics," in *Proc. IEEE Int. Conf. Environ. Electr. Eng. IEEE Ind. Commercial Power Syst. Eur. (EEEIC/ICPS Europe)*, Jun. 2019, pp. 1–5, doi: 10.1109/EEEIC.2019.8783340.

[4] L. Cheng and T. Yu, "A new generation of AI: A review and perspective on machine learning technologies applied to smart energy and electric power systems," *Int. J. Energy Res.*, vol. 43, no. 6, pp. 1928–1973, May 2019.

[5] S. M. Miraftabzadeh, M. Longo, F. Foiadelli, M. Pasetti, and R. Igual, "Advances in the application of machine learning techniques for power system analytics: A survey," *Energies*, vol. 14, no. 16, p. 4776, Aug. 2021, doi: 10.3390/en14164776.

[6] C. Zhang, Z. Chen, and J. Zhou, "Research on short-term load forecasting using K-means clustering and CatBoost integrating time series features," in *Proc. 39th Chin. Control Conf. (CCC)*, Shenyang, China, 2020, pp. 6099–6104, doi: 10.23919/CCC50068.2020.9188856.

[7] R. Li, W. Wang, and M. Xia, "Cooperative planning of active distribution system with renewable energy sources and energy storage systems," *IEEE Access*, vol. 6, pp. 5916–5926, 2018, doi: 10.1109/ACCESS.2017.2785263.

[8] S. Paul, M. R. Haq, A. Das, and Z. Ni, "A comparative study of smart grid security based on unsupervised learning and load ranking," in *Proc. IEEE Int. Conf. Electro Inf. Technol. (EIT)*, May 2019, pp. 310–315, doi: 10.1109/EIT.2019.8834059.

[9] L. Marrero, L. García-Santander, L. Hernandez-Callejo, P. Bañuelos-Sánchez, and V. J. González, "Harmonic distortion characterization in groups of distribution networks applying the IEEE standard 519–2014," *IEEE Latin Amer. Trans.*, vol. 19, no. 4, pp. 526–533, Apr. 2021.

[10] V. H. Ferreira, A. D. C. Pinho, D. S. de Souza, and B. S. Rodrigues, "A new clustering approach for automatic oscillographic records segmentation," *Energies*, vol. 14, no. 20, p. 6778, Oct. 2021, doi: 10.3390/en14206778.

[11] K. P. Sinaga and M.-S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[12] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, p. 1295, Aug. 2020.

[13] C. Zhang, X. Lu, G. Ren, S. Chen, C. Hu, Z. Kong, N. Zhang, and A. M. Foley, "Optimal allocation of onshore wind power in China based on cluster analysis," *Appl. Energy*, vol. 285, Mar. 2021, Art. no. 116482, doi: 10.1016/j.apenergy.2021.116482.

[14] Y. Chen and Z. Chen, "Short-term load forecasting for multiple buildings: A length sensitivity-based approach," *Energy Rep.*, vol. 8, pp. 14274–14288, Nov. 2022, doi: 10.1016/j.egyr.2022.10.425.

[15] S. Park, S. Park, M. Kim, and E. Hwang, "Clustering-based self-imputation of unlabeled fault data in a fleet of photovoltaic generation systems," *Energies*, vol. 13, no. 3, p. 737, Feb. 2020, doi: 10.3390/en13030737.

[16] Y. Li, N. Lu, X. Wang, and B. Jiang, "Islanding fault detection based on data-driven approach with active developed reactive power variation," *Neurocomputing*, vol. 337, pp. 97–109, Apr. 2019, doi: 10.1016/j.neucom.2019.01.054.

[17] K. Dhibi, M. Mansouri, K. Bouzrara, H. Nounou, and M. Nounou, "Reduced neural network based ensemble approach for fault detection and diagnosis of wind energy converter systems," *Renew. Energy*, vol. 194, pp. 778–787, Jul. 2022, doi: 10.1016/j.renene.2022.05.082.

[18] G. Trotta, "An empirical analysis of domestic electricity load profiles: Who consumes how much and when?" *Appl. Energy*, vol. 275, Oct. 2020, Art. no. 115399, doi: 10.1016/j.apenergy.2020.115399.

[19] Z. Wang and T. Hong, "Generating realistic building electrical load profiles through the generative adversarial network (GAN)," *Energy Buildings*, vol. 224, Oct. 2020, Art. no. 110299, doi: 10.1016/j.enbuild.2020.110299.

[20] Y. Wang, Y. Shen, S. Mao, X. Chen, and H. Zou, "LASSO and LSTM integrated temporal model for short-term solar intensity forecasting," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2933–2944, Apr. 2019, doi: 10.1109/JIOT.2018.2877510.

[21] H. Gharavi and B. Hu, "Space-time approach for disturbance detection and classification," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5132–5140, Sep. 2018, doi: 10.1109/TSG.2017.2680742.

[22] Y. Mohammadi, S. M. Miraftabzadeh, M. H. J. Bollen, and M. Longo, "Voltage-sag source detection: Developing supervised methods and proposing a new unsupervised learning," *Sustain. Energy, Grids Netw.*, vol. 32, Dec. 2022, Art. no. 100855, doi: 10.1016/j.segan.2022.100855.

[23] J.-P. Liu, T.-X. Zhang, J. Zhu, and T.-N. Ma, "Allocation optimization of electric vehicle charging station (EVCS) considering with charging satisfaction and distributed renewables integration," *Energy*, vol. 164, pp. 560–574, Dec. 2018, doi: 10.1016/j.energy.2018.09.028.

[24] A. K. Gangwar, O. P. Mahela, B. Rathore, B. Khan, H. H. Alhelou, and P. Siano, "A novel k-means clustering and weighted k-NN-regression-based fast transmission line protection," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6034–6043, Sep. 2021.

[25] A. Naghikhani and S. M. H. Hosseini, "Optimal thermal and power planning considering economic and environmental issues in peak load management," *Energy*, vol. 239, Jan. 2022, Art. no. 121955, doi: 10.1016/j.energy.2021.121955.

[26] W. Sheng, K.-Y. Liu, Y. Liu, X. Meng, and Y. Li, "Optimal placement and sizing of distributed generation via an improved nondominated sorting genetic algorithm II," *IEEE Trans. Power Del.*, vol. 30, no. 2, pp. 569–578, Apr. 2015, doi: 10.1109/TPWRD.2014.2325938.

[27] P. Ferraro, E. Crisostomi, M. Tucci, and M. Raugi, "Comparison and clustering analysis of the daily electrical load in eight European countries," *Electr. Power Syst. Res.*, vol. 141, pp. 114–123, Dec. 2016.

[28] E. Schubert and P. J. Rousseeuw, "Fast and eager k-medoids clustering: $O(k)$ runtime improvement of the PAM, CLARA, and CLARANS algorithms," *Inf. Syst.*, vol. 101, Nov. 2021, Art. no. 101804.

[29] F. Luo, G. Ranzi, W. Kong, Z. Y. Dong, and F. Wang, "Coordinated residential energy resource scheduling with vehicle-to-home and high photovoltaic penetrations," *IET Renew. Power Gener.*, vol. 12, no. 6, pp. 625–632, Apr. 2018.

[30] S. M. Miraftabzadeh, M. Longo, and M. Brenna, "Knowledge extraction from PV power generation with deep learning autoencoder and clustering-based algorithms," *IEEE Access*, vol. 11, pp. 69227–69240, 2023, doi: 10.1109/ACCESS.2023.3292516.

[31] J. Luo, Y. Hu, and Y. Bai, "Bibliometric analysis of the blockchain scientific evolution: 2014–2020," *IEEE Access*, vol. 9, pp. 120227–120246, 2021, doi: 10.1109/ACCESS.2021.3092192.

[32] W. Zuo, B. Mu, H. Fang, and Y. Wan, "User experience: A bibliometric review of the literature," *IEEE Access*, vol. 11, pp. 12663–12676, 2023, doi: 10.1109/ACCESS.2023.3241968.

[33] H. Liao, M. Tang, L. Luo, C. Li, F. Chiclana, and X.-J. Zeng, "A bibliometric analysis and visualization of medical big data research," *Sustainability*, vol. 10, no. 2, p. 166, Jan. 2018, doi: 10.3390/su10010166.

[34] C. Newman, D. Edwards, I. Martek, J. Lai, W. D. Thwala, and I. Rillie, "Industry 4.0 deployment in the construction industry: A bibliometric literature review and U.K.-based case study," *Smart Sustain. Built Environ.*, vol. 10, no. 4, pp. 557–580, Nov. 2021, doi: 10.1108/SASBE-02-2020-0016.

[35] S. M. Mirmazloumi, A. Moghimi, B. Ranjgar, F. Mohseni, A. Ghorbanian, S. A. Ahmadi, M. Amani, and B. Brisco, "Status and trends of wetland studies in Canada using remote sensing technology with a focus on wetland classification: A bibliographic analysis," *Remote Sens.*, vol. 13, no. 20, p. 4025, Oct. 2021.

[36] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *J. Bus. Res.*, vol. 133, pp. 285–296, Sep. 2021, doi: 10.1016/j.jbusres.2021.04.070.

[37] E. Ç. Akay, N. T. Y. Soydan, and B. K. Gacar, "Bibliometric analysis of the published literature on machine learning in economics and econometrics," *Social Netw. Anal. Mining*, vol. 12, no. 1, p. 109, Dec. 2022, doi: 10.1007/s13278-022-00916-6.

[38] C. Chen and M. Song, "Visualizing a field of research: A methodology of systematic scientometric reviews," *PLoS ONE*, vol. 14, no. 10, Oct. 2019, Art. no. e0223994, doi: 10.1371/journal.pone.0223994.

[39] G. Zurita, C. Mulet-Forteza, J. M. Merigó, V. Lobos-Ossandón, and H. Ogata, "A bibliometric overview of the IEEE transactions on learning technologies," *IEEE Trans. Learn. Technol.*, vol. 15, no. 6, pp. 656–672, Dec. 2022, doi: 10.1109/TLT.2022.3204457.

[40] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982, doi: 10.1109/TIT.1982.1056489.

[41] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, Oakland, CA, USA, 1967, pp. 281–297.

[42] H. H. Bock, "Clustering methods: A history of k-means algorithms," in *Selected Contributions in Data Analysis and Classification* (Studies in Classification, Data Analysis, and Knowledge Organization), P. Brito, G. Cucumel, P. Bertrand, and F. de Carvalho, Eds. Berlin, Germany: Springer, 2007, pp. 161–172, doi: 10.1007/978-3-540-73560-1_15.

[43] S. M. Miraftabzadeh, M. Longo, M. Brenna, and M. Pasetti, "Data-driven model for PV power generation patterns extraction via unsupervised machine learning methods," in *Proc. North Amer. Power Symp. (NAPS)*, Oct. 2022, pp. 1–5.

[44] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-means clustering method and elbow method for identification of the best customer profile cluster," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 336, Apr. 2018, Art. no. 012017.

[45] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image segmentation using K-means clustering algorithm and subtractive clustering algorithm," *Proc. Comput. Sci.*, vol. 54, pp. 764–771, Jan. 2015.

[46] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci.*, vol. 622, pp. 178–210, Apr. 2023.

[47] P. Fränti and S. Sieranoja, "How much can k-means be improved by using better initialization and repeats?" *Pattern Recognit.*, vol. 93, pp. 95–112, Sep. 2019.

[48] M. E. Celebi, H. A. Kingravi, and P. A. Vela, "A comparative study of efficient initialization methods for the k-means clustering algorithm," *Expert Syst. Appl.*, vol. 40, no. 1, pp. 200–210, Jan. 2013.

[49] C. Ding and X. He, "K-means clustering via principal component analysis," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, p. 29.

[50] Y. Liang, M.-F. Balcan, and V. Kanchanapally, "Distributed PCA and k-means clustering," in *Proc. Big Learn. Workshop NIPS*. Princeton, NJ, USA: Citeseer, 2013, pp. 1–8.

[51] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern Recognit.*, vol. 36, no. 2, pp. 451–461, Feb. 2003.

[52] D. Arthur and S. Vassilvitskii, "k-means++ the advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.

[53] B. Bahmani, B. Moseley, A. Vattani, R. Kumar, and S. Vassilvitskii, "Scalable k-means++," 2012, *arXiv:1203.6402*.

[54] M.-S. Yang and K. P. Sinaga, "A feature-reduction multi-view k-means clustering algorithm," *IEEE Access*, vol. 7, pp. 114472–114486, 2019, doi: 10.1109/ACCESS.2019.2934179.

[55] D. Feldman, M. Schmidt, and C. Sohler, "Turning big data into tiny data: Constant-size coresets for k-means, PCA, and projective clustering," *SIAM J. Comput.*, vol. 49, no. 3, pp. 601–657, 2020.

[56] P. Govender and V. Sivakumar, "Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)," *Atmos. Pollut. Res.*, vol. 11, no. 1, pp. 40–56, Jan. 2020.

[57] Y. Mohammadi, S. M. Miraftabzadeh, M. H. J. Bollen, and M. Longo, "An unsupervised learning schema for seeking patterns in rms voltage variations at the sub-10-minute time scale," *Sustain. Energy, Grids Netw.*, vol. 31, Sep. 2022, Art. no. 100773, doi: 10.1016/j.segan.2022.100773.

[58] M. M. Fard, T. Thonet, and E. Gaussier, "Deep k-means: Jointly clustering with k-means and learning representations," *Pattern Recognit. Lett.*, vol. 138, pp. 185–192, Oct. 2020.

[59] J. Chen, X. Qi, L. Chen, F. Chen, and G. Cheng, "Quantum-inspired ant lion optimized hybrid k-means for cluster analysis and intrusion detection," *Knowl.-Based Syst.*, vol. 203, Sep. 2020, Art. no. 106167.

[60] A. S. Shirkhorshidi, S. Aghabozorgi, and T. Y. Wah, "A comparison study on similarity and dissimilarity measures in clustering continuous data," *PLoS ONE*, vol. 10, no. 12, Dec. 2015, Art. no. e0144059.

[61] B. Zhou, X. Ma, Y. Luo, and D. Yang, "Wind power prediction based on LSTM networks and nonparametric kernel density estimation," *IEEE Access*, vol. 7, pp. 165279–165292, 2019, doi: 10.1109/ACCESS.2019.2952555.

[62] Y. Zhang, Y. Li, and G. Zhang, "Short-term wind power forecasting approach based on Seq2Seq model using NWP data," *Energy*, vol. 213, Dec. 2020, Art. no. 118371, doi: 10.1016/j.energy.2020.118371.

[63] W. Dong, Q. Yang, and X. Fang, "Multi-step ahead wind power generation prediction based on hybrid machine learning techniques," *Energies*, vol. 11, no. 8, p. 1975, Jul. 2018, doi: 10.3390/en11081975.

[64] M. B. Ozkan and P. Karagoz, "A novel wind power forecast model: Statistical hybrid wind power forecast technique (SHWIP)," *IEEE Trans. Ind. Informat.*, vol. 11, no. 2, pp. 375–387, Apr. 2015, doi: 10.1109/TII.2015.2396011.

[65] R. Praveena and K. Dhanalakshmi, "Wind power forecasting in short-term using fuzzy K-means clustering and neural network," in *Proc. Int. Conf. Intell. Comput. Commun. Smart World (I2C2SW)*, Erode, India, 2018, pp. 336–339, doi: 10.1109/I2C2SW45816.2018.8997350.

[66] C. Changqing, L. Xinran, and L. Hua, "Frequency regulation control strategy of wind farms based on temporal and spatial uncertainty," *Sustain. Energy Technol. Assessments*, vol. 53, Oct. 2022, Art. no. 102679, doi: 10.1016/j.seta.2022.102679.

[67] H. Zhang, X. Hu, H. Cheng, S. Zhang, S. Hong, and Q. Gu, "Coordinated scheduling of generators and tie lines in multi-area power systems under wind energy uncertainty," *Energy*, vol. 222, May 2021, Art. no. 119929, doi: 10.1016/j.energy.2021.119929.

[68] M. Yesilbudak, "Implementation of novel hybrid approaches for power curve modeling of wind turbines," *Energy Convers. Manage.*, vol. 171, pp. 156–169, Sep. 2018, doi: 10.1016/j.enconman.2018.05.092.

[69] Y. Dai, Z. Ren, K. Wang, W. Li, Z. Li, and W. Yan, "Optimal sizing and arrangement of tidal current farm," *IEEE Trans. Sustain. Energy*, vol. 9, no. 1, pp. 168–177, Jan. 2018, doi: 10.1109/TSTE.2017.2719042.

[70] M. Neshat, M. M. Nezhad, E. Abbasnejad, S. Mirjalili, D. Groppi, A. Heydari, L. B. Tjernberg, D. Astiaso Garcia, B. Alexander, Q. Shi, and M. Wagner, "Wind turbine power output prediction using a new hybrid neuro-evolutionary method," *Energy*, vol. 229, Aug. 2021, Art. no. 120617, doi: 10.1016/j.energy.2021.120617.

[71] F. Najibi, D. Apostolopoulou, and E. Alonso, "Enhanced performance Gaussian process regression for probabilistic short-term solar output forecast," *Int. J. Electr. Power Energy Syst.*, vol. 130, Sep. 2021, Art. no. 106916, doi: 10.1016/j.ijepes.2021.106916.

[72] D. Liu and K. Sun, "Random forest solar power forecast based on classification optimization," *Energy*, vol. 187, Nov. 2019, Art. no. 115940, doi: 10.1016/j.energy.2019.115940.

[73] Y. Liu, "Short-term prediction method of solar photovoltaic power generation based on machine learning in smart grid," *Math. Problems Eng.*, vol. 2022, pp. 1–10, Sep. 2022, doi: 10.1155/2022/8478790.

[74] Z. Zhen, J. Liu, Z. Zhang, F. Wang, H. Chai, Y. Yu, X. Lu, T. Wang, and Y. Lin, "Deep learning based surface irradiance mapping model for solar PV power forecasting using sky image," *IEEE Trans. Ind. Appl.*, vol. 56, no. 4, pp. 3385–3396, Jul. 2020, doi: 10.1109/TIA.2020.2984617.

[75] X. Yang, M. Xu, S. Xu, and X. Han, "Day-ahead forecasting of photovoltaic output power with similar cloud space fusion based on incomplete historical data mining," *Appl. Energy*, vol. 206, pp. 683–696, Nov. 2017, doi: 10.1016/j.apenergy.2017.08.222.

[76] V. B. Pamshetti and S. P. Singh, "Coordinated allocation of BESS and SOP in high PV penetrated distribution network incorporating DR and CVR schemes," *IEEE Syst. J.*, vol. 16, no. 1, pp. 420–430, Mar. 2022, doi: 10.1109/JSYST.2020.3041013.

[77] S. R. Ghatak, S. Sannigrahi, and P. Acharjee, "Multiobjective framework for optimal integration of solar energy source in three-phase unbalanced distribution network," *IEEE Trans. Ind. Appl.*, vol. 56, no. 3, pp. 3068–3078, May 2020, doi: 10.1109/TIA.2020.2968046.

[78] J. M. Home-Ortiz, M. Pourakbari-Kasmaei, M. Lehtonen, and J. R. S. Mantovani, "Optimal location-allocation of storage devices and renewable-based DG in distribution systems," *Electr. Power Syst. Res.*, vol. 172, pp. 11–21, Jul. 2019, doi: 10.1016/j.epsr.2019.02.013.

[79] S. Lin, C. Liu, Y. Shen, F. Li, D. Li, and Y. Fu, "Stochastic planning of integrated energy system via Frank–Copula function and scenario reduction," *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 202–212, Jan. 2022, doi: 10.1109/TSG.2021.3119939.

[80] M. Lee, T. Hong, J. Jeong, and K. Jeong, "Development of a rooftop solar photovoltaic rating system considering the technical and economic suitability criteria at the building level," *Energy*, vol. 160, pp. 213–224, Oct. 2018, doi: 10.1016/j.energy.2018.07.020.

[81] V. Shahbazbegian, S.-M. Hosseini-Motlagh, and A. Haeri, "Integrated forward/reverse logistics thin-film photovoltaic power plant supply chain network design with uncertain data," *Appl. Energy*, vol. 277, Nov. 2020, Art. no. 115538, doi: 10.1016/j.apenergy.2020.115538.

[82] H. Zhang, Y. Cao, Y. Zhang, and V. Terzija, "Quantitative synergy assessment of regional wind-solar energy resources based on MERRA reanalysis data," *Appl. Energy*, vol. 216, pp. 172–182, Apr. 2018, doi: 10.1016/j.apenergy.2018.02.094.

[83] Z. Tang, P. Zhang, K. Muto, M. Sawasawa, M. Simonelli, C. Gutierrez, J. Yang, M. Astitha, D. A. Ferrante, J. N. Debs, R. Manning, and J. Mader, "Extreme photovoltaic power analytics for electric utilities," *IEEE Trans. Sustain. Energy*, vol. 11, no. 1, pp. 93–106, Jan. 2020, doi: 10.1109/TSTE.2018.2884500.

[84] H. Chen, Y. Cui, S. Li, J. Liu, and K. Liu, "An improved GMM-based algorithm with optimal multi-color subspaces for color difference classification of solar cells," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 4, pp. 503–513, Nov. 2018, doi: 10.1109/TSM.2018.2867254.

[85] F. Borghetti, M. Longo, S. M. Miraftabzadeh, G. Mizzoni, and G. Giudici, "A quantitative method to assess the vehicle-to-grid feasibility of a local public transport company," *IEEE Access*, vol. 11, pp. 55644–55656, 2023, doi: 10.1109/ACCESS.2023.3279713.

[86] C. G. Colombo, S. M. Miraftabzadeh, A. Saldarini, M. Longo, M. Brenna, and W. Yaici, "Literature review on wireless charging technologies: Future trend for electric vehicle?" in *Proc. 2nd Int. Conf. Sustain. Mobility Appl., Renewables Technol. (SMART)*, Nov. 2022, pp. 1–5.

[87] S. A. Jees and V. Gomathi, "Load forecasting for smart grid using non-linear model in Hadoop distributed file system," *Cluster Comput.*, vol. 22, no. S6, pp. 13533–13545, Nov. 2019, doi: 10.1007/s10586-018-1997-2.

[88] S. M. Miraftabzadeh, C. G. Colombo, M. Longo, and F. Foiadelli, "A day-ahead photovoltaic power prediction via transfer learning and deep neural networks," *Forecasting*, vol. 5, no. 1, pp. 213–228, Feb. 2023, doi: 10.3390/forecast5010012.

[89] J. Guo, P. Zhang, D. Wu, Z. Liu, X. Liu, S. Zhang, X. Yang, and H. Ge, "Multi-objective optimization design and multi-attribute decision-making method of a distributed energy system based on nearly zero-energy community load forecasting," *Energy*, vol. 239, Jan. 2022, Art. no. 122124, doi: 10.1016/j.energy.2021.122124.

[90] A. Kell, A. S. McGough, and M. Forshaw, "Segmenting residential smart meter data for short-term load forecasting," in *Proc. 9th Int. Conf. Future Energy Syst.*, New York, NY, USA, Jun. 2018, pp. 91–96, doi: 10.1145/3208903.3208923.

[91] Z. Zhao, J. Tang, J. Liu, G. Ge, B. Xiong, and Y. Li, "Short-term microgrid load probability density forecasting method based on k-means-deep learning quantile regression," *Energy Rep.*, vol. 8, pp. 1386–1397, Aug. 2022, doi: 10.1016/j.egyr.2022.03.117.

[92] A. I. Arvanitidis, D. Bargiotas, A. Daskalopulu, D. Kontogiannis, I. P. Panapakidis, and L. H. Tsoukalas, "Clustering informed MLP models for fast and accurate short-term load forecasting," *Energies*, vol. 15, no. 4, p. 1295, Feb. 2022, doi: 10.3390/en15041295.

[93] X. Pang, W. Sun, H. Li, Y. Wang, and C. Luan, "Short-term power load forecasting based on gray relational analysis and support vector machine optimized by artificial bee colony algorithm," *PeerJ Comput. Sci.*, vol. 8, p. e1108, Sep. 2022, doi: 10.7717/peerj-cs.1108.

[94] P. Borthakur and B. Goswami, "Short term load forecasting: A hybrid approach using data mining methods," in *Proc. Int. Conf. Emerg. Frontiers Electr. Electron. Technol. (ICEFEET)*, Jul. 2020, pp. 1–6, doi: 10.1109/ICEFEET49149.2020.9187009.

[95] S. Arens, K. Derendorf, F. Schuldt, K. Maydell, and C. Agert, "Effect of EV movement schedule and machine learning-based load forecasting on electricity cost of a single household," *Energies*, vol. 11, no. 11, p. 2913, Oct. 2018, doi: 10.3390/en11112913.

[96] M. A. Acquah and S. Han, "Online building load management control with plugged-in electric vehicles considering uncertainties," *Energies*, vol. 12, no. 8, p. 1436, Apr. 2019, doi: 10.3390/en12081436.

[97] C. Ye, Y. Ding, P. Wang, and Z. Lin, "A data-driven bottom-up approach for spatial and temporal electric load forecasting," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 1966–1979, May 2019, doi: 10.1109/TPWRS.2018.2889995.

[98] A. Almalaq and G. Edwards, "A review of deep learning methods applied on load forecasting," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 511–516, doi: 10.1109/ICMLA.2017.0-110.

[99] C. Zhang, Z. Chen, and J. Zhou, "Research on short-term load forecasting using K-means clustering and CatBoost integrating time series features," in *Proc. 39th Chin. Control Conf. (CCC)*, Jul. 2020, pp. 6099–6104, doi: 10.23919/CCC50068.2020.9188856.

[100] Y. He, F. Luo, G. Ranzi, and W. Kong, "Short-term residential load forecasting based on federated learning and load clustering," in *Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm)*, Oct. 2021, pp. 77–82, doi: 10.1109/SmartGridComm51999.2021.9632314.

[101] K. Li, J. Zhang, X. Chen, and W. Xue, "Building's hourly electrical load prediction based on data clustering and ensemble learning strategy," *Energy Buildings*, vol. 261, Apr. 2022, Art. no. 111943, doi: 10.1016/j.enbuild.2022.111943.

[102] W. Lizhen, Z. Yifan, W. Gang, and H. Xiaohong, "A novel short-term load forecasting method based on mini-batch stochastic gradient descent regression model," *Electr. Power Syst. Res.*, vol. 211, Oct. 2022, Art. no. 108226, doi: 10.1016/j.epsr.2022.108226.

[103] J. Geng, X. Sun, F. Li, and X. Wu, "Prediction method of important nodes and transmission lines in power system transactive management," *Electr. Power Syst. Res.*, vol. 208, Jul. 2022, Art. no. 107898, doi: 10.1016/j.epsr.2022.107898.

[104] Y. Hao, Z. Yao, J. Wang, H. Li, R. Li, L. Yang, and W. Liang, "A classification method for transmission line Icing process curve based on hierarchical K-means clustering," *Energies*, vol. 12, no. 24, p. 4786, Dec. 2019, doi: 10.3390/en12244786.

[105] L. Liu, B. Li, Y. Li, J. Yi, J. Zhang, and Q. Wang, "Calculation of line loss in low-voltage transformer district based on BP network model optimized by LM algorithm," in *Proc. IEEE Elect. Power Energy Conf. (EPEC)*, Oct. 2016, pp. 1–6, doi: 10.1109/EPEC.2016.7771740.

[106] J. Wu, S. Shao, D. Jiang, H. Yao, C. Ma, and K. Xu, "Research on adaptive power transmission line fault inspection," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 787, no. 1, Jun. 2021, Art. no. 012050, doi: 10.1088/1755-1315/787/1/012050.

[107] J. Wang, T. Zhang, X. Xue, and L. Chen, "Real-time recognition of transmission line insulators under complex backgrounds: A YOLOV5s approach," in *Proc. 4th Int. Conf. Power Energy Technol. (ICPET)*, Jul. 2022, pp. 77–83, doi: 10.1109/ICPET55165.2022.9918361.

[108] Z.-L. Hu, J.-H. Li, A. Chen, F. Xu, R. Jia, F.-L. Lin, and C.-B. Tang, "Optimize grouping and path of pylon inspection in power system," *IEEE Access*, vol. 8, pp. 108885–108895, 2020, doi: 10.1109/ACCESS.2020.3001435.

[109] D. Yang, T. He, M. Chen, Z. Zhang, and L. Shen, "Analysis of lightning arrester operating current based on multidimensional neural network for transmission lines," *Evol. Intell.*, vol. 16, no. 5, pp. 1581–1588, Nov. 2022, doi: 10.1007/s12065-022-00791-2.

[110] X. Deng, P. Zhang, K. Jin, J. He, X. Wang, and Y. Wang, "Probabilistic load flow method considering large-scale wind power integration," *J. Mod. Power Syst. Clean Energy*, vol. 7, no. 4, pp. 813–825, Jul. 2019, doi: 10.1007/s40565-019-0502-0.

[111] A. Oshnoei, R. Khezri, M. T. Hagh, K. Techato, S. Muyeen, and O. Sadeghian, "Direct probabilistic load flow in radial distribution systems including wind farms: An approach based on data clustering," *Energies*, vol. 11, no. 2, p. 310, Feb. 2018, doi: 10.3390/en11020310.

[112] S. Song, C. Han, S. Jung, M. Yoon, and G. Jang, "Probabilistic power flow analysis of bulk power system for practical grid planning application," *IEEE Access*, vol. 7, pp. 45494–45503, 2019, doi: 10.1109/ACCESS.2019.2909537.

[113] P. Mirhoseini and N. Ghaffarzadeh, "Economic battery sizing and power dispatch in a grid-connected charging station using convex method," *J. Energy Storage*, vol. 31, Oct. 2020, Art. no. 101651, doi: 10.1016/j.est.2020.101651.

[114] S. Madadi, B. Mohammadi-Ivatloo, and S. Tohidi, "A data clustering based probabilistic power flow method for AC/VSC-MTDC," *IEEE Syst. J.*, vol. 13, no. 4, pp. 4324–4334, Dec. 2019, doi: 10.1109/JSYST.2019.2918234.

[115] I. C. Barutcu, E. Karatepe, and M. Boztepe, "Impact of harmonic limits on PV penetration levels in unbalanced distribution networks considering load and irradiance uncertainty," *Int. J. Electr. Power Energy Syst.*, vol. 118, Jun. 2020, Art. no. 105780, doi: 10.1016/j.ijepes.2019.105780.

[116] H. Li, C. Li, and Y. Liu, "Maximum frequency deviation assessment with clustering based on metric learning," *Int. J. Electr. Power Energy Syst.*, vol. 120, Sep. 2020, Art. no. 105980, doi: 10.1016/j.ijepes.2020.105980.

[117] N. Y. Yürüşen, B. Uzunoğlu, A. P. Talayero, and A. L. Estopiñán, "Apriori and K-means algorithms of machine learning for spatio-temporal solar generation balancing," *Renew. Energy*, vol. 175, pp. 702–717, Sep. 2021, doi: 10.1016/j.renene.2021.04.098.

[118] B. P. Ferraz, M. Resener, L. A. Pereira, F. A. B. Lemos, and S. Haffner, "MILP model for volt-var optimization considering chronological operation of distribution systems containing DERs," *Int. J. Electr. Power Energy Syst.*, vol. 129, Jul. 2021, Art. no. 106761, doi: 10.1016/j.ijepes.2021.106761.

[119] V. Purba, B. B. Johnson, S. Jafarpour, F. Bullo, and S. V. Dhople, "Dynamic aggregation of grid-tied three-phase inverters," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1520–1530, Mar. 2020, doi: 10.1109/TPWRS.2019.2942292.

[120] C. F. Heuberger, P. K. Bains, and N. M. Dowell, "The evolution of the power system: A spatio-temporal optimisation model to investigate the impact of electric vehicle deployment," *Appl. Energy*, vol. 257, Jan. 2020, Art. no. 113715, doi: 10.1016/j.apenergy.2019.113715.

[121] A. Maamar and K. Benahmed, "A hybrid model for anomalies detection in AMI system combining K-means clustering and deep neural network," *Comput., Mater. Continua*, vol. 60, no. 1, pp. 15–39, 2019, doi: 10.32604/cmc.2019.06497.

[122] A. L. Amutha, R. A. Uthra, J. P. Roselyn, and R. G. Brunet, "Anomaly detection in multivariate streaming PMU data using density estimation technique in wide area monitoring system," *Expert Syst. Appl.*, vol. 175, Aug. 2021, Art. no. 114865, doi: 10.1016/j.eswa.2021.114865.

[123] S. D. Roy and S. Debbarma, "A novel OC-SVM based ensemble learning framework for attack detection in AGC loop of power systems," *Electr. Power Syst. Res.*, vol. 202, Jan. 2022, Art. no. 107625, doi: 10.1016/j.epsr.2021.107625.

[124] F. Shi, L. Zhang, H. Zhang, K. Xu, and T. Vladimir, "Diagnosis of the single phase-to-ground fault in distribution network based on feature extraction and transformation from the waveforms," *IET Gener., Transmiss. Distrib.*, vol. 14, no. 25, pp. 6079–6086, Dec. 2020, doi: 10.1049/iet-gtd.2020.0877.

[125] A. Et-Taleby, Y. Chaibi, M. Boussetta, A. Allouhi, and M. Benslimane, "A novel fault detection technique for PV systems based on the K-means algorithm, coded wireless orthogonal frequency division multiplexing and thermal image processing techniques," *Sol. Energy*, vol. 237, pp. 365–376, May 2022, doi: 10.1016/j.solener.2022.04.017.

[126] G. Liu, L. Zhu, X. Wu, and J. Wang, "Time series clustering and physical implication for photovoltaic array systems with unknown working conditions," *Sol. Energy*, vol. 180, pp. 401–411, Mar. 2019, doi: 10.1016/j.solener.2019.01.041.

[127] P. Calvo-Bascones, M. A. Sanz-Bobi, and T. M. Welte, "Anomaly detection method based on the deep knowledge behind behavior patterns in industrial components. Application to a hydropower plant," *Comput. Ind.*, vol. 125, Feb. 2021, Art. no. 103376, doi: 10.1016/j.compind.2020.103376.

[128] Q. Xue, G. Li, Y. Zhang, S. Shen, Z. Chen, and Y. Liu, "Fault diagnosis and abnormality detection of lithium-ion battery packs based on statistical distribution," *J. Power Sources*, vol. 482, Jan. 2021, Art. no. 228964, doi: 10.1016/j.jpowsour.2020.228964.

[129] M. Mansouri, K. Dhibi, M. Hajji, K. Bouzara, H. Nounou, and M. Nounou, "Interval-valued reduced RNN for fault detection and diagnosis for wind energy conversion systems," *IEEE Sensors J.*, vol. 22, no. 13, pp. 13581–13588, Jul. 2022, doi: 10.1109/JSEN.2022.3175866.

[130] A. Gantayet and D. K. Dheer, "Multiobjective planning strategy for a distribution network integrated with wind power system considering solid state transformer," in *Proc. Int. Conf. Comput. Perform. Eval. (ComPE)*, Dec. 2021, pp. 692–697, doi: 10.1109/ComPE53109.2021.9752251.

[131] W. Jiang and M. Zhong, "Dynamic equivalent modeling of wind farm based on dominant variable hierarchical clustering algorithm," *Math. Problems Eng.*, vol. 2021, pp. 1–10, Nov. 2021, doi: 10.1155/2021/7629414.

[132] Y. Ji, H. Xiao, S. Shi, J. Kang, J. Wang, and S. Zhang, "A novel combinational evaluation method of voltage and reactive power in regional power grid containing renewable energy," *Arch. Elect. Eng.*, vol. 70, no. 4, pp. 925–942, 2021, doi: 10.24425/aee.2021.138270.

[133] Y. Liu, L. Li, Y. Li, M. Li, M. Huang, H. Yu, and W. Liu, "Distributed photovoltaic cluster partition and reactive power optimization strategy based on BAS-IGA algorithm," in *Proc. 7th Asia Conf. Power Electr. Eng. (ACPEE)*, Apr. 2022, pp. 2198–2203, doi: 10.1109/ACPEE53904.2022.9784084.

[134] A. Arabpour, M. R. Besmi, and P. Maghouli, "Transmission expansion and reactive power planning considering wind energy investment using a linearized AC model," *J. Electr. Eng. Technol.*, vol. 14, no. 3, pp. 1035–1043, May 2019, doi: 10.1007/s42835-019-00085-1.

[135] R. Shigenobu, A. Nakadomari, Y.-Y. Hong, P. Mandal, H. Takahashi, and T. Senjyu, "Optimization of voltage unbalance compensation by smart inverter," *Energies*, vol. 13, no. 18, p. 4623, Sep. 2020, doi: 10.3390/en13184623.

[136] R. Karandeh, M. Davoudi, and V. Cecchi, "Distributed control of behind-the-meter energy resources for multiple services," *IEEE Trans. Power Del.*, vol. 37, no. 3, pp. 2050–2059, Jun. 2022, doi: 10.1109/TPWRD.2021.3103661.

[137] S. Sun, Q. Yang, J. Ma, A. J. Ferré, and W. Yan, "Hierarchical planning of PEV charging facilities and DGs under transportation-power network couplings," *Renew. Energy*, vol. 150, pp. 356–369, May 2020, doi: 10.1016/j.renene.2019.12.097.

[138] N. J. Williams, P. Jaramillo, K. Campbell, B. Musanga, and I. Lyons-Galante, "Electricity consumption and load profile segmentation analysis for rural micro grid customers in Tanzania," in *Proc. IEEE PES/IAS PowerAfrica*, Jun. 2018, pp. 360–365.

[139] P. Liu et al., "Hybrid features based K-means clustering algorithm for use in electricity customer load pattern analysis," in *Proc. 37th Chin. Control Conf. (CCC)*, Wuhan, China, 2018, pp. 8851–8857, doi: 10.23919/ChiCC.2018.8483451.

[140] M. Emmanuel and J. Giraldez, "Net electricity clustering at different temporal resolutions using a SAX-based method for integrated distribution system planning," *IEEE Access*, vol. 7, pp. 123689–123697, 2019, doi: 10.1109/ACCESS.2019.2938212.

[141] R. R. Rathod and R. D. Garg, "Regional electricity consumption analysis for consumers using data mining techniques and consumer meter reading data," *Int. J. Electr. Power Energy Syst.*, vol. 78, pp. 368–374, Jun. 2016, doi: 10.1016/j.ijepes.2015.11.110.

[142] J. Song, Z. Mao, X. Li, T. Liu, W. Zhong, and Y. Cui, "A classification and synthesis method for load characteristics of typical industry based on daily electricity consumption curves," in *Proc. 5th Int. Conf. Power Renew. Energy (ICPRE)*. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, Sep. 2020, pp. 78–84, doi: 10.1109/ICPRE51194.2020.9233159.

[143] R. Pérez-Chacón, J. Luna-Romera, A. Troncoso, F. Martínez-Álvarez, and J. Riquelme, "Big data analytics for discovering electricity consumption patterns in smart cities," *Energies*, vol. 11, no. 3, p. 683, Mar. 2018, doi: 10.3390/en11030683.

[144] M. Mora-Alvarez, P. Contreras-Ortiz, X. Serrano-Guerrero, and G. Escrivá-Escriva, "Characterization and classification of daily electricity consumption profiles: Shape factors and k-means clustering technique," in *Proc. E3S Web Conf.*, 2018, p. 6, doi: 10.1051/e3sconf/20186408004.

[145] I. P. Panapakidis and G. C. Christoforidis, "Implementation of modified versions of the K-means algorithm in power load curves profiling," *Sustain. Cities Soc.*, vol. 35, pp. 83–93, Nov. 2017, doi: 10.1016/j.scs.2017.08.002.

[146] F. Wang, K. Li, N. Duić, Z. Mi, B.-M. Hodge, M. Shafie-Khah, and J. P. S. Catalão, "Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns," *Energy Convers. Manage.*, vol. 171, pp. 839–854, Sep. 2018, doi: 10.1016/j.enconman.2018.06.017.

[147] H. W. Choi, N. M. F. Qureshi, and D. R. Shin, "Comparative analysis of electricity consumption at home through a silhouette-score prospective," in *Proc. 21st Int. Conf. Adv. Commun. Technol. (ICACT)*. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, Feb. 2019, pp. 589–591, doi: 10.23919/ICACT.2019.8701923.

[148] Y. He, Z. Jiao, F. Chen, F. Guang, P. You, and Q. He, "Residential power user segmentation based on k-means clustering method in the context of big data," in *Proc. E3S Web Conf.* Les Ulis, France: EDP Sciences, Sep. 2018, p. 4, doi: 10.1051/e3sconf/20185302006.

[149] M. Amayri, C. S. Silva, H. Pombeiro, and S. Ploix, "Flexibility characterization of residential electricity consumption: A machine learning approach," *Sustain. Energy, Grids Netw.*, vol. 32, Dec. 2022, Art. no. 100801, doi: 10.1016/j.segan.2022.100801.

[150] Y. Wang, Y. Liu, L. Li, D. Infield, and S. Han, "Short-term wind power forecasting based on clustering pre-calculated CFD method," *Energies*, vol. 11, no. 4, p. 854, Apr. 2018, doi: 10.3390/en11040854.

[151] N. Iliopoulos, M. Esteban, and S. Kudo, "Assessing the willingness of residential electricity consumers to adopt demand side management and distributed energy resources: A case study on the Japanese market," *Energy Policy*, vol. 137, Feb. 2020, Art. no. 111169, doi: 10.1016/j.enpol.2019.111169.

[152] S. Yilmaz, J. Chambers, and M. K. Patel, "Comparison of clustering approaches for domestic electricity load profile characterisation—Implications for demand side management," *Energy*, vol. 180, pp. 665–677, Aug. 2019, doi: 10.1016/j.energy.2019.05.124.

[153] M. Numata, M. Sugiyama, and G. Mogi, "Barrier analysis for the deployment of renewable-based mini-grids in Myanmar using the analytic hierarchy process (AHP)," *Energies*, vol. 13, no. 6, p. 1400, Mar. 2020, doi: 10.3390/en13061400.

[154] B. Buddhahai, W. Wongseree, and P. Rakkwamsuk, "An energy prediction approach for a nonintrusive load monitoring in home appliances," *IEEE Trans. Consum. Electron.*, vol. 66, no. 1, pp. 96–105, Feb. 2020, doi: 10.1109/TCE.2019.2956638.

[155] F. Olivier, A. Sutera, P. Geurts, R. Fonteneau, and D. Ernst, "Phase identification of smart meters by clustering voltage measurements," in *Proc. Power Syst. Comput. Conf. (PSCC)*. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, Jun. 2018, pp. 1–8, doi: 10.23919/PSCC.2018.8442853.

[156] Y. Ma, X. Fan, R. Tang, P. Duan, Y. Sun, J. Du, and Q. Duan, "Phase identification of smart meters by spectral clustering," in *Proc. 2nd IEEE Conf. Energy Internet Energy Syst. Integr. (EI2)*, Oct. 2018, pp. 1–5.

[157] A. Hassan, Y. Zhao, L. Pu, G. Wang, H. Sun, and R. M. Winter, "Evaluation of clustering algorithms for DAP placement in wireless smart meter network," in *Proc. 9th Int. Conf. Modeling, Identificat. Control (ICMIC)*, Jul. 2017, pp. 1085–1090.

[158] S. Walker, V. Bergkamp, D. Yang, T. A. J. van Goch, K. Katic, and W. Zeiler, "Improving energy self-sufficiency of a renovated residential neighborhood with heat pumps by analyzing smart meter data," *Energy*, vol. 229, Aug. 2021, Art. no. 120711, doi: 10.1016/j.energy.2021.120711.

[159] R. Tang, B. Yildiz, P. H. W. Leong, A. Vassallo, and J. Dore, "Residential battery sizing model using net meter energy data clustering," *Appl. Energy*, vol. 251, Oct. 2019, Art. no. 113324, doi: 10.1016/j.apenergy.2019.113324.

[160] A. Marlen, A. Maxim, I. A. Ukaegbu, and H. S. V. S. K. Nunna, "Application of big data in smart grids: Energy analytics," in *Proc. 21st Int. Conf. Adv. Commun. Technol. (ICACT)*. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, Feb. 2019, pp. 402–407, doi: 10.23919/ICACT.2019.8701973.

[161] A. G. Kazaki and T. A. Papadopoulos, "Cluster analysis of university campus smart meter data," in *Proc. 53rd Int. Univ. Power Eng. Conf. (UPEC)*. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, Sep. 2018, pp. 1–6, doi: 10.1109/UPEC.2018.8541941.

[162] G. Swathi, P. S. Varma, S. R. Donepudi, and K. R. Kumar, "Consumer-centric rate design for peak time energy demand coincidence reduction at domestic sector level—A smart energy service for residential demand response," *J. Electr. Comput. Eng.*, vol. 2022, pp. 1–16, Oct. 2022, doi: 10.1155/2022/8294369.

[163] P. Laurinec, M. Lóderer, M. Lucká, and V. Rozinajová, ''Density-based unsupervised ensemble learning methods for time series forecasting of aggregated or clustered electricity consumption,'' *J. Intell. Inf. Syst.*, vol. 53, no. 2, pp. 219–239, Oct. 2019, doi: 10.1007/s10844-019-00550-3.
[164] S. M. Miraftabzadeh, M. Longo, and F. Foiadelli, ''A-day-ahead photovoltaic power prediction based on long short term memory algorithm,'' in *Proc. Int. Conf. Smart Energy Syst. Technol. (SEST)*. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, Sep. 2020, pp. 1–6, doi: 10.1109/SEST48500.2020.9203481.
[165] C. Silva, P. Faria, A. Fernandes, and Z. Vale, ''Clustering distributed energy storage units for the aggregation of optimized local solar energy,'' *Energy Rep.*, vol. 8, pp. 405–410, Jun. 2022, doi: 10.1016/j.egyr.2022.01.043.
[166] N. Zhu, D. Jiang, P. Hu, and Y. Yang, ''Honeycomb active distribution network: A novel structure of distribution network and its stochastic optimization,'' in *Proc. 15th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, Nov. 2020, pp. 455–462, doi: 10.1109/ICIEA48937.2020.9248161.
[167] N. N. Hong and Y. Nakanishi, ''Optimal scheduling of an isolated wind-diesel-energy storage system considering fast frequency response and forecast error,'' *Energies*, vol. 12, no. 5, p. 843, Mar. 2019, doi: 10.3390/en12050843.
[168] Y. Yoo, G. Jang, J. Lee, and S. Jung, ''Stochastic forecast error estimation of high PV penetration system considering net-load/PV decoupling for microgrid operation,'' *IET Renew. Power Gener.*, vol. 16, no. 10, pp. 2152–2162, Jul. 2022, doi: 10.1049/rpg2.12506.
[169] Y. Allahvirdizadeh, S. Galvani, and H. Shayanfar, ''Data clustering based probabilistic optimal scheduling of an energy hub considering risk-averse,'' *Int. J. Electr. Power Energy Syst.*, vol. 128, Jun. 2021, Art. no. 106774, doi: 10.1016/j.ijepes.2021.106774.
[170] Y. Allahvirdizadeh, S. Galvani, H. Shayanfar, and M. P. Moghaddam, ''Risk-averse scheduling of an energy hub in the presence of correlated uncertain variables considering time of use and real-time pricing-based demand response programs,'' *Energy Sci. Eng.*, vol. 10, no. 4, pp. 1343–1372, Apr. 2022, doi: 10.1002/ese3.1104.
[171] J. Yang, H. Gao, S. Ye, L. Lv, Y. Liu, and J. Liu, ''Applying multiple types of demand response to optimal day-ahead stochastic scheduling in the distribution network,'' *IET Gener., Transmiss. Distrib.*, vol. 14, no. 20, pp. 4509–4519, Oct. 2020, doi: 10.1049/iet-gtd.2020.0158.
[172] S. Zeynali, N. Nasiri, M. Marzband, and S. N. Ravadanegh, ''A hybrid robust-stochastic framework for strategic scheduling of integrated wind farm and plug-in hybrid electric vehicle fleets,'' *Appl. Energy*, vol. 300, Oct. 2021, Art. no. 117432, doi: 10.1016/j.apenergy.2021.117432.
[173] F. K. Abo-Elyousr, A. M. Sharaf, M. M. F. Darwish, M. Lehtonen, and K. Mahmoud, ''Optimal scheduling of DG and EV parking lots simultaneously with demand response based on self-adjusted PSO and K-means clustering,'' *Energy Sci. Eng.*, vol. 10, no. 10, pp. 4025–4043, Oct. 2022, doi: 10.1002/ese3.1264.
[174] C. Silva, P. Faria, and Z. Vale, ''Multi-period observation clustering for tariff definition in a weekly basis remuneration of demand response,'' *Energies*, vol. 12, no. 7, p. 1248, Apr. 2019, doi: 10.3390/en12071248.
[175] P. Faria, J. Spínola, and Z. Vale, ''Distributed energy resources scheduling and aggregation in the context of demand response programs,'' *Energies*, vol. 11, no. 8, p. 1987, Jul. 2018, doi: 10.3390/en11081987.
[176] Y. Ding, X. Wang, D. Zhang, X. Wang, L. Yang, and T. Pu, ''Research on key node identification scheme for power system considering malicious data attacks,'' *Energy Rep.*, vol. 7, pp. 1289–1296, Nov. 2021, doi: 10.1016/j.egyr.2021.09.135.
[177] S. D. Roy, S. Debbarma, and J. M. Guerrero, ''A data-driven algorithm to detect false data injections targeting both frequency regulation and market operation in power systems,'' *Int. J. Electr. Power Energy Syst.*, vol. 143, Dec. 2022, Art. no. 108409, doi: 10.1016/j.ijepes.2022.108409.
[178] P. Wang, M. Govindarasu, A. Ashok, S. Sridhar, and D. McKinnon, ''Data-driven anomaly detection for power system generation control,'' in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 1082–1089, doi: 10.1109/ICDMW.2017.152.
[179] P. Shukla, ''ML-IDS: A machine learning approach to detect wormhole attacks in Internet of Things,'' in *Proc. Intell. Syst. Conf. (IntelliSys)*, Sep. 2017, pp. 234–240, doi: 10.1109/IntelliSys.2017.8324298.
[180] S. D. Roy, S. Debbarma, and A. Iqbal, ''A decentralized intrusion detection system for security of generation control,'' *IEEE Internet Things J.*, vol. 9, no. 19, pp. 18924–18933, Oct. 2022, doi: 10.1109/JIOT.2022.3163502.
[181] S. Poudel, Z. Ni, X. Zhong, and H. He, ''Comparative studies of power grid security with network connectivity and power flow information using unsupervised learning,'' in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 2730–2737, doi: 10.1109/IJCNN.2016.7727542.
[182] C. G. Krishnan, A. H. Nishan, and P. Theerthagiri, ''k-means clustering based energy and trust management routing algorithm for mobile ad-hoc networks,'' *Int. J. Commun. Syst.*, vol. 35, no. 9, Jun. 2022, doi: 10.1002/dac.5138.
[183] G. Dhand and S. S. Tyagi, ''SMEER: Secure multi-tier energy efficient routing protocol for hierarchical wireless sensor networks,'' *Wireless Pers. Commun.*, vol. 105, no. 1, pp. 17–35, Mar. 2019, doi: 10.1007/s11277-018-6101-y.
[184] Y. Lu and A. Christou, ''Prognostics of IGBT modules based on the approach of particle filtering,'' *Microelectron. Rel.*, vol. 92, pp. 96–105, Jan. 2019, doi: 10.1016/j.microrel.2018.11.012.
[185] P. Yu, X. Yang, F. Zhou, H. Li, L. Feng, W. Li, and X. Qiu, ''Deep reinforcement learning aided cell outage compensation framework in 5G cloud radio access networks,'' *Mobile Netw. Appl.*, vol. 25, no. 5, pp. 1644–1654, Oct. 2020, doi: 10.1007/s11036-020-01574-8.
[186] G. C. Krishnan, S. GomathiGomathi, and A. M. A. Bamini, ''High energy efficient lifetime management system and trust management framework for manet using self-configurable cluster mechanism,'' *Peer-to-Peer Netw. Appl.*, vol. 14, no. 3, pp. 1229–1241, May 2021, doi: 10.1007/s12083-021-01094-6.
[187] Y. Jie, A. Alsharoa, A. E. Kamal, and M. Alnuem, ''A CoMP-based outage compensation solution for heterogeneous femtocell networks,'' *Comput. Netw.*, vol. 168, Feb. 2020, Art. no. 107061, doi: 10.1016/j.comnet.2019.107061.
[188] R. Yao, K. Sun, F. Liu, and S. Mei, ''Efficient simulation of temperature evolution of overhead transmission lines based on analytical solution and NWP,'' *IEEE Trans. Power Del.*, vol. 33, no. 4, pp. 1576–1588, Aug. 2018, doi: 10.1109/TPWRD.2017.2751563.
[189] M. Dong, ''Combining unsupervised and supervised learning for asset class failure prediction in power systems,'' *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 5033–5043, Nov. 2019, doi: 10.1109/TPWRS.2019.2920915.
[190] N. Salehi, H. Martínez-García, and G. Velasco-Quesada, ''Networked microgrid energy management based on supervised and unsupervised learning clustering,'' *Energies*, vol. 15, no. 13, p. 4915, Jul. 2022, doi: 10.3390/en15134915.
[191] Y. Yin, M. Liu, G. Gui, H. Gacanin, H. Sari, and F. Adachi, ''Cross-layer resource allocation for UAV-assisted wireless caching networks with NOMA,'' *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3428–3438, Apr. 2021, doi: 10.1109/TVT.2021.3064032.
[192] T. Bose, A. Suresh, O. J. Pandey, L. R. Cenkeramaddi, and R. M. Hegde, ''Improving quality-of-service in cluster-based UAV-assisted edge networks,'' *IEEE Trans. Netw. Service Manage.*, vol. 19, no. 2, pp. 1903–1919, Jun. 2022, doi: 10.1109/TNSM.2021.3139892.
[193] S. A. Shafaat, P. V. Teja, P. S. K. Reddy, R. Manoj, S. B. Daram, and O. Obulesu, ''Application of big data analytics in power system under single transmission line outage condition,'' in *Proc. 3rd Int. Conf. Trends Electron. Informat. (ICOEI)*, Apr. 2019, pp. 377–382, doi: 10.1109/ICOEI.2019.8862685.
[194] A. Esmaeilian and M. Kezunovic, ''Controlled islanding to prevent cascade outages using constrained spectral k-embedded clustering,'' in *Proc. Power Syst. Comput. Conf. (PSCC)*, Jun. 2016, pp. 1–6, doi: 10.1109/PSCC.2016.7541017.
[195] B. A. Hemad, N. M. A. Ibrahim, S. A. Fayad, and H. E. A. Talaat, ''Hierarchical clustering-based framework for interconnected power system contingency analysis,'' *Energies*, vol. 15, no. 15, p. 5631, Aug. 2022, doi: 10.3390/en15155631.
[196] A. Abbasi, H. K. Karegar, and T. S. Aghdam, ''Optimal adaptive protection using setting groups allocation based on impedance matrix,'' in *Proc. 14th Int. Conf. Protection Autom. Power Syst. (IPAPS)*, Dec. 2019, pp. 77–82, doi: 10.1109/IPAPS49326.2019.9069377.
[197] A. Esmaeilian and M. Kezunovic, ''Prevention of power grid blackouts using intentional islanding scheme,'' *IEEE Trans. Ind. Appl.*, vol. 53, no. 1, pp. 622–629, Jan. 2017, doi: 10.1109/TIA.2016.2614772.
[198] Y. Wang and Y. Zhao, ''Research on classification method of voltage sag based on scenario analysis,'' in *Proc. IEEE 3rd Int. Conf. Electron. Technol. (ICET)*, May 2020, pp. 518–522, doi: 10.1109/ICET49382.2020.9119706.

[199] A. Alkilany, A. Ahmed, H. Said, and A. A. Bakar, ''Application of the k-means clustering algorithm to predict load shedding of the southern electrical grid of Libya,'' in *Proc. 4th, Ed., Int. Conf. Innov. Comput. Technol. (INTECH)*, Aug. 2014, pp. 244–247, doi: 10.1109/INTECH.2014.6927763.

[200] H. Xiao, Y. Chen, Q. Zhang, A. T. Chronopoulos, Z. Zhang, and S. Ouyang, ''Joint clustering and power allocation for the cross roads congestion scenarios in cooperative vehicular networks,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2267–2277, Jun. 2019, doi: 10.1109/TITS.2018.2866236.

[201] L. Ying, Y. Jia, and W. Li, ''Research on state evaluation and risk assessment for relay protection system based on machine learning algorithm,'' *IET Gener., Transmiss. Distrib.*, vol. 14, no. 18, pp. 3619–3629, Sep. 2020, doi: 10.1049/iet-gtd.2018.6552.

[202] S. Rezaeian-Marjani, H. Ebrahimi, and B. Tousi, ''Probabilistic multi-objective optimization method for interline power flow controller (IPFC) allocation in power systems,'' *IET Gener., Transmiss. Distrib.*, vol. 16, no. 24, pp. 4951–4962, Dec. 2022, doi: 10.1049/gtd2.12645.

[203] F. A. Assis, A. J. C. Coelho, L. D. Rezende, A. M. L. D. Silva, and L. C. Resende, ''Unsupervised machine learning techniques applied to composite reliability assessment of power systems,'' *Int. Trans. Electr. Energy Syst.*, vol. 31, no. 11, Nov. 2021, Art. no. e13109, doi: 10.1002/2050-7038.13109.

[204] Q. Liu, B. Tong, D. Li, Y. Lu, Y. Fu, L. Chen, and K. Zhao, ''A smart and safe electricity consumption model for integrated energy system based on electric big data,'' *Int. J. Saf. Secur. Eng.*, vol. 10, no. 4, pp. 529–534, Aug. 2020, doi: 10.18280/ijsse.100412.

[205] O. A. Ansari, Y. Gong, W. Liu, and C. Y. Chung, ''Data-driven operation risk assessment of wind-integrated power systems via mixture models and importance sampling,'' *J. Mod. Power Syst. Clean Energy*, vol. 8, no. 3, pp. 437–445, May 2020, doi: 10.35833/MPCE.2019.000163.

[206] C.-M. Huang, Y.-C. Huang, K.-Y. Huang, S.-J. Chen, and S.-P. Yang, ''Deterministic and probabilistic wind power forecasting using a hybrid method,'' in *Proc. IEEE Int. Conf. Ind. Technol. (ICIT)*, Mar. 2017, pp. 400–405, doi: 10.1109/ICIT.2017.7913264.

[207] F. D. Maio, M. Vagnoli, and E. Zio, ''Risk-based clustering for near misses identification in integrated deterministic and probabilistic safety analysis,'' *Sci. Technol. Nucl. Installations*, vol. 2015, pp. 1–29, Aug. 2015, doi: 10.1155/2015/693891.

[208] D. A. Zaldivar, A. A. Romero, and S. R. Rivera, ''Risk assessment algorithm for power transformer fleets based on condition and strategic importance,'' *Algorithms*, vol. 14, no. 11, p. 319, Oct. 2021, doi: 10.3390/a14110319.

[209] A. Doostmohammadi, N. Amjady, and H. Zareipour, ''Day-ahead financial loss/gain modeling and prediction for a generation company,'' *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3360–3372, Sep. 2017, doi: 10.1109/TPWRS.2017.2654363.

[210] A. Papavasiliou, ''An overview of probabilistic dimensioning of frequency restoration reserves with a focus on the Greek electricity market,'' *Energies*, vol. 14, no. 18, p. 5719, Sep. 2021, doi: 10.3390/en14185719.

[211] A. J. M. Kell, M. Forshaw, and A. S. McGough, ''Long-term electricity market agent based model validation using genetic algorithm based optimization,'' in *Proc. 11th ACM Int. Conf. Future Energy Syst.* New York, NY, USA: Association for Computing Machinery, Jun. 2020, pp. 1–13, doi: 10.1145/3396851.3397682.

[212] O. Motamedisedeh, B. Ostadi, F. Zagia, and A. H. Kashan, ''A novel optimization model for biding in the deregulated power market with pay as a bid settlement mechanism, based on the stochastic market clearing price,'' *Electr. Power Syst. Res.*, vol. 210, Sep. 2022, Art. no. 108122, doi: 10.1016/j.epsr.2022.108122.

[213] J. Iria and F. Soares, ''A cluster-based optimization approach to support the participation of an aggregator of a larger number of prosumers in the day-ahead energy market,'' *Electr. Power Syst. Res.*, vol. 168, pp. 324–335, Mar. 2019, doi: 10.1016/j.epsr.2018.11.022.

[214] J. L. Crespo-Vazquez, C. Carrillo, E. Diaz-Dorado, J. A. Martinez-Lorenzo, and M. Noor-E-Alam, ''Evaluation of a data driven stochastic approach to optimize the participation of a wind and storage power plant in day-ahead and reserve markets,'' *Energy*, vol. 156, pp. 278–291, Aug. 2018, doi: 10.1016/j.energy.2018.04.185.

[215] H. C. Jeong, J. Jung, and B. O. Kang, ''Development of operational strategies of energy storage system using classification of customer load profiles under time-of-use tariffs in South Korea,'' *Energies*, vol. 13, no. 7, p. 1723, Apr. 2020, doi: 10.3390/en13071723.

[216] P. D. Hung, N. D. Ngoc, and T. D. Hanh, ''K-means clustering using R A case study of market segmentation,'' in *Proc. 5th Int. Conf. E-Business Appl.*, Feb. 2019, pp. 100–104, doi: 10.1145/3317614.3317626.

[217] T. Pinto and F. Falcão-Reis, ''Strategic participation in competitive electricity markets: Internal versus sectorial data analysis,'' *Int. J. Electr. Power Energy Syst.*, vol. 108, pp. 432–444, Jun. 2019, doi: 10.1016/j.ijepes.2019.01.011.

[218] F. Feijoo, W. Silva, and T. K. Das, ''A computationally efficient electricity price forecasting model for real time energy markets,'' *Energy Convers. Manage.*, vol. 113, pp. 27–35, Apr. 2016, doi: 10.1016/j.enconman.2016.01.043.

[219] G. Poyrazoglu, ''Determination of price zones during transition from uniform to zonal electricity market: A case study for Turkey,'' *Energies*, vol. 14, no. 4, p. 1014, Feb. 2021, doi: 10.3390/en14041014.

[220] K. Zheng, H. Chen, Y. Wang, and Q. Chen, ''Data-driven financial transmission right scenario generation and speculation,'' *Energy*, vol. 238, Jan. 2022, Art. no. 122056, doi: 10.1016/j.energy.2021.122056.

[221] H. Jahangir, S. S. Gougheri, B. Vatandoust, M. A. Golkar, A. Ahmadian, and A. Hajizadeh, ''Plug-in electric vehicle behavior modeling in energy market: A novel deep learning-based approach with clustering technique,'' *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 4738–4748, Nov. 2020, doi: 10.1109/TSG.2020.2998072.

[222] F. Wang, K. Li, L. Zhou, H. Ren, J. Contreras, M. Shafie-Khah, and J. P. S. Catalão, ''Daily pattern prediction based classification modeling approach for day-ahead electricity price forecasting,'' *Int. J. Electr. Power Energy Syst.*, vol. 105, pp. 529–540, Feb. 2019, doi: 10.1016/j.ijepes.2018.08.039.

[223] M. Savi and F. Olivadese, ''Short-term energy consumption forecasting at the edge: A federated learning approach,'' *IEEE Access*, vol. 9, pp. 95949–95969, 2021, doi: 10.1109/ACCESS.2021.3094089.

[224] K. E. Farfar and M. T. Khadir, ''A two-stage short-term load forecasting approach using temperature daily profiles estimation,'' *Neural Comput. Appl.*, vol. 31, no. 8, pp. 3909–3919, Aug. 2019, doi: 10.1007/s00521-017-3324-x.

[225] X. Tang, Y. Dai, Q. Liu, X. Dang, and J. Xu, ''Application of bidirectional recurrent neural network combined with deep belief network in short-term load forecasting,'' *IEEE Access*, vol. 7, pp. 160660–160670, 2019, doi: 10.1109/ACCESS.2019.2950957.

[226] M. A. Acquah, Y. Jin, B.-C. Oh, Y.-G. Son, and S.-Y. Kim, ''Spatiotemporal sequence-to-sequence clustering for electric load forecasting,'' *IEEE Access*, vol. 11, pp. 5850–5863, 2023, doi: 10.1109/ACCESS.2023.3235724.

[227] T. Malatesta and J. K. Breadsell, ''Identifying home system of practices for energy use with k-means clustering techniques,'' *Sustainability*, vol. 14, no. 15, p. 9017, Jul. 2022, doi: 10.3390/su14159017.

[228] A.-N. Khan, N. Iqbal, A. Rizwan, R. Ahmad, and D.-H. Kim, ''An ensemble energy consumption forecasting model based on spatial–temporal clustering analysis in residential buildings,'' *Energies*, vol. 14, no. 11, p. 3020, May 2021, doi: 10.3390/en14113020.

[229] J. Chen, D. Zhang, and Y. Nanehkaran, ''Research of power load prediction based on boost clustering,'' *Soft Comput.*, vol. 25, no. 8, pp. 6401–6413, Apr. 2021, doi: 10.1007/s00500-021-05632-5.

[230] X. Cheng, T. Wang, Y. Li, J. Pi, M. Li, and B. Liu, ''Reactive power load forecasting based on K-means clustering and random forest algorithm,'' in *Proc. Int. Conf. Power Syst. Technol. (POWERCON)*. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, Dec. 2021, pp. 925–928, doi: 10.1109/POWERCON53785.2021.9697840.

[231] X. Gong, J. L. Cardenas-Barrera, E. Castillo-Guerra, B. Cao, S. A. Saleh, and L. Chang, ''Bottom-up load forecasting with Markov-based error reduction method for aggregated domestic electric water heaters,'' *IEEE Trans. Ind. Appl.*, vol. 55, no. 6, pp. 6401–6413, Nov. 2019, doi: 10.1109/TIA.2019.2936330.

[232] S. Huang, N. A. M. Ali, N. Shaari, and M. S. M. Noor, ''Multi-scene design analysis of integrated energy system based on feature extraction algorithm,'' *Energy Rep.*, vol. 8, pp. 466–476, Sep. 2022, doi: 10.1016/j.egyr.2022.03.161.

[233] R. Jiao, T. Zhang, Y. Jiang, and H. He, ''Short-term non-residential load forecasting based on multiple sequences LSTM recurrent neural network,'' *IEEE Access*, vol. 6, pp. 59438–59448, 2018, doi: 10.1109/ACCESS.2018.2873712.

[234] I. S. Jahan, M. Prilepok, S. Misak, and V. Snasel, ''Intelligent system for power load forecasting in off-grid platform,'' in *Proc. 19th Int. Sci. Conf. Electr. Power Eng. (EPE)*, May 2018, pp. 1–5.

[235] K. Liang, F. Liu, and Y. Zhang, "Household power consumption prediction method based on selective ensemble learning," *IEEE Access*, vol. 8, pp. 95657–95666, 2020, doi: 10.1109/ACCESS.2020.2996260.
[236] F. Han, T. Pu, M. Li, and G. Taylor, "Short-term forecasting of individual residential load based on deep learning and K-means clustering," *CSEE J. Power Energy Syst.*, vol. 7, no. 2, pp. 261–269, Mar. 2021, doi: 10.17775/CSEEJPES.2020.04060.
[237] M. Cui, J. Wang, and M. Yue, "Machine learning-based anomaly detection for load forecasting under cyberattacks," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5724–5734, Sep. 2019, doi: 10.1109/TSG.2018.2890809.
[238] S. Yan and M. Hu, "A multi-stage planning method for distribution networks based on ARIMA with error gradient sampling for source–load prediction," *Sensors*, vol. 22, no. 21, p. 8403, Nov. 2022, doi: 10.3390/s22218403.
[239] X. Yu and S. Ergan, "Estimating power demand shaving capacity of buildings on an urban scale using extracted demand response profiles through machine learning models," *Appl. Energy*, vol. 310, Mar. 2022, Art. no. 118579, doi: 10.1016/j.apenergy.2022.118579.
[240] L. Hernández, C. Baladrón, J. Aguiar, B. Carro, and A. Sánchez-Esguevillas, "Classification and clustering of electricity demand patterns in industrial parks," *Energies*, vol. 5, no. 12, pp. 5215–5228, Dec. 2012, doi: 10.3390/en5125215.
[241] N. Schöne, K. Greilmeier, and B. Heinz, "Survey-based assessment of the preferences in residential demand response on the island of Mayotte," *Energies*, vol. 15, no. 4, p. 1338, Feb. 2022, doi: 10.3390/en15041338.
[242] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 911–918, Mar. 2015, doi: 10.1109/TSG.2014.2364233.
[243] I. L. R. Gomes, R. Melicio, and V. M. F. Mendes, "A novel microgrid support management system based on stochastic mixed-integer linear programming," *Energy*, vol. 223, May 2021, Art. no. 120030, doi: 10.1016/j.energy.2021.120030.
[244] A. R. Kulkarni and M. S. Ballal, "Synergism of synchrophasor measurements and data analytics for enhancing situational awareness of power grid," *Comput. Electr. Eng.*, vol. 93, Jul. 2021, Art. no. 107231, doi: 10.1016/j.compeleceng.2021.107231.
[245] G. Mohi-Ud-Din, A. K. Marnerides, Q. Shi, C. Dobbins, and A. MacDermott, "Deep COLA: A deep competitive learning algorithm for future home energy management systems," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 5, no. 6, pp. 860–870, Dec. 2021, doi: 10.1109/TETCI.2020.3027300.
[246] A. C. Varghese, V. Padmini, G. Kumar, and S. A. Khaparde, "Smart grid consumer behavioral model using machine learning," in *Proc. IEEE Innov. Smart Grid Technol. Asia (ISGT Asia)*, May 2018, pp. 734–739, doi: 10.1109/ISGT-Asia.2018.8467824.
[247] H. T. Javed, M. O. Beg, H. Mujtaba, H. Majeed, and M. Asim, "Fairness in real-time energy pricing for smart grid using unsupervised learning," *Comput. J.*, vol. 62, no. 3, pp. 414–429, Mar. 2019, doi: 10.1093/comjnl/bxy071.
[248] Y. Ma, C. Li, J. Zhou, and Y. Zhang, "Comprehensive stochastic optimal scheduling in residential micro energy grid considering pumped-storage unit and demand response," *J. Energy Storage*, vol. 32, Dec. 2020, Art. no. 101968, doi: 10.1016/j.est.2020.101968.
[249] E. T. Al-Shammari, S. Shamshirband, D. Petković, E. Zalnezhad, P. L. Yee, R. S. Taher, and Ž. Ćojbašić, "Comparative study of clustering methods for wake effect analysis in wind farm," *Energy*, vol. 95, pp. 573–579, Jan. 2016, doi: 10.1016/j.energy.2015.11.064.
[250] M. Katwe, K. Singh, P. K. Sharma, C.-P. Li, and Z. Ding, "Dynamic user clustering and optimal power allocation in UAV-assisted full-duplex hybrid NOMA system," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2573–2590, Apr. 2022, doi: 10.1109/TWC.2021.3113640.
[251] J. Yu, R. Zhang, Y. Gao, and L.-L. Yang, "Modularity-based dynamic clustering for energy efficient UAVs-aided communications," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 728–731, Oct. 2018, doi: 10.1109/LWC.2018.2816649.
[252] H. Talei, D. Benhaddou, C. Gamarra, H. Benbrahim, and M. Essaaidi, "Smart building energy inefficiencies detection through time series analysis and unsupervised machine learning," *Energies*, vol. 14, no. 19, p. 6042, Sep. 2021, doi: 10.3390/en14196042.
[253] R. Chai, C. Liu, and Q. Chen, "Energy efficiency optimization-based joint resource allocation and clustering algorithm for M2M communication systems," *IEEE Access*, vol. 7, pp. 168507–168519, 2019, doi: 10.1109/ACCESS.2019.2954713.
[254] S. M. Miraftabzadeh, M. Longo, and F. Foiadelli, "Estimation model of total energy consumptions of electrical vehicles under different driving conditions," *Energies*, vol. 14, no. 4, p. 854, Feb. 2021, doi: 10.3390/en14040854.
[255] Y. He, S. Guo, J. Zhou, G. Song, A. Kurban, and H. Wang, "The multi-stage framework for optimal sizing and operation of hybrid electrical-thermal energy storage system," *Energy*, vol. 245, Apr. 2022, Art. no. 123248, doi: 10.1016/j.energy.2022.123248.
[256] H. Zhu, R. Hou, T. Jiang, and Q. Lv, "Research on energy storage capacity configuration for PV power plants using uncertainty analysis and its applications," *Global Energy Interconnection*, vol. 4, no. 6, pp. 608–618, Dec. 2021, doi: 10.1016/j.gloei.2022.01.004.
[257] Y. Zeng, C. Li, and H. Wang, "Scenario-set-based economic dispatch of power system with wind power and energy storage system," *IEEE Access*, vol. 8, pp. 109105–109119, 2020, doi: 10.1109/ACCESS.2020.3001678.
[258] L. Li, J. Wang, X. Zhong, J. Lin, N. Wu, Z. Zhang, C. Meng, X. Wang, N. Shah, N. Brandon, S. Xie, and Y. Zhao, "Combined multi-objective optimization and agent-based modeling for a 100% renewable island energy system considering power-to-gas technology and extreme weather conditions," *Appl. Energy*, vol. 308, Feb. 2022, Art. no. 118376, doi: 10.1016/j.apenergy.2021.118376.
[259] P. Wen, Y. Xie, L. Huo, and A. Tohidi, "Optimal and stochastic performance of an energy hub-based microgrid consisting of a solar-powered compressed-air energy storage system and cooling storage system by modified grasshopper optimization algorithm," *Int. J. Hydrogen Energy*, vol. 47, no. 27, pp. 13351–13370, Mar. 2022, doi: 10.1016/j.ijhydene.2022.02.081.
[260] P. Ifaei, A. Farid, and C. Yoo, "An optimal renewable energy management strategy with and without hydropower using a factor weighted multi-criteria decision making analysis and nation-wide big data—Case study in Iran," *Energy*, vol. 158, pp. 357–372, Sep. 2018, doi: 10.1016/j.energy.2018.06.043.
[261] A. Gantayet and D. K. Dheer, "A multi-objective optimisation strategy exploring the energy routing capability of a smart transformer while integrating hybrid energy hub into a distribution network," *Sustain. Energy, Grids Netw.*, vol. 32, Dec. 2022, Art. no. 100956, doi: 10.1016/j.segan.2022.100956.
[262] C. Wei, Y. Chen, X. Li, and X. Lin, "Integrating intelligent driving pattern recognition with adaptive energy management strategy for extender range electric logistics vehicle," *Energy*, vol. 247, May 2022, Art. no. 123478, doi: 10.1016/j.energy.2022.123478.
[263] S. H. Jessen, Z. G. Ma, F. D. Wijaya, J. C. Vasquez, J. Guerrero, and B. N. Jørgensen, "Identification of natural disaster impacted electricity load profiles with k means clustering algorithm," *Energy Informat.*, vol. 5, no. S4, p. 59, Dec. 2022, doi: 10.1186/s42162-022-00237-0.
[264] A. Khatibi, R. Ghasempour, R. Shirmohammadi, A. Farahmand, H. Alizadeh, and E. Bellos, "Optimization of multi-layer absorbing systems in solar flat-plate collectors using cluster analysis," *Sustain. Energy Technol. Assessments*, vol. 36, Dec. 2019, Art. no. 100538, doi: 10.1016/j.seta.2019.100538.
[265] X. Wang, H. Yu, Y. Lin, Z. Zhang, and X. Gong, "Dynamic equivalent modeling for wind farms with DFIGs using the artificial bee colony with K-means algorithm," *IEEE Access*, vol. 8, pp. 173723–173731, 2020, doi: 10.1109/ACCESS.2020.3024212.
[266] A. D. Martino, S. M. Miraftabzadeh, and M. Longo, "Strategies for the modelisation of electric vehicle energy consumption: A review," *Energies*, vol. 15, no. 21, p. 8115, Oct. 2022.
[267] K. Li, F. Yang, L. Wang, Y. Yan, H. Wang, and C. Zhang, "A scenario-based two-stage stochastic optimization approach for multi-energy microgrids," *Appl. Energy*, vol. 322, Sep. 2022, Art. no. 119388, doi: 10.1016/j.apenergy.2022.119388.
[268] M.-R. Yaghoubi-Nia, H. Hashemi-Dezaki, and A. H. Niasar, "Optimal stochastic scenario-based allocation of smart grids' renewable and non-renewable distributed generation units and protective devices," *Sustain. Energy Technol. Assessments*, vol. 44, Apr. 2021, Art. no. 101033, doi: 10.1016/j.seta.2021.101033.
[269] L. Kanagasabai, "Shrinkage of real power loss by enriched brain storm optimization algorithm," *IAES Int. J. Artif. Intell.*, vol. 8, no. 1, p. 1, Mar. 2019, doi: 10.11591/ijai.v8.i1.pp1-6.

[270] N. Nasiri, S. Zeynali, and S. N. Ravadanegh, "A tactical transactive energy scheduling for the electric vehicle-integrated networked microgrids," *Sustain. Cities Soc.*, vol. 83, Aug. 2022, Art. no. 103943, doi: 10.1016/j.scs.2022.103943.

[271] F. Foiadelli, M. Longo, and S. Miraftabzadeh, "Energy consumption prediction of electric vehicles based on big data approach," in *Proc. IEEE Int. Conf. Environ. Electr. Eng. IEEE Ind. Commercial Power Syst. Eur. (EEEIC/ICPS Europe)*, Jun. 2018, pp. 1–6, doi: 10.1109/EEEIC.2018.8494573.

[272] Y. Liu, J. Liu, Y. Zhang, Y. Wu, Z. Chen, and M. Ye, "Rule learning based energy management strategy of fuel cell hybrid vehicles considering multi-objective optimization," *Energy*, vol. 207, Sep. 2020, Art. no. 118212, doi: 10.1016/j.energy.2020.118212.

[273] H. Yu, C. Zhang, Z. Deng, H. Bian, C. Sun, and C. Jia, "Economic optimization for configuration and sizing of micro integrated energy systems," *J. Modern Power Syst. Clean Energy*, vol. 6, no. 2, pp. 330–341, Mar. 2018, doi: 10.1007/s40565-017-0291-2.

[274] R. Torkzadeh, J. B. M. van Waes, V. Cuk, and J. F. G. Cobben, "Clustering technique for scenario reduction in post-energy transition voltage dips assessment," *Renew. Energy Power Qual. J.*, vol. 19, pp. 241–245, Sep. 2021, doi: 10.24084/repqj19.267.

[275] Y. Mohammadi, S. M. Miraftabzadeh, M. H. J. Bollen, and M. Longo, "Seeking patterns in rms voltage variations at the sub-10-minute scale from multiple locations via unsupervised learning and patterns' post-processing," *Int. J. Electr. Power Energy Syst.*, vol. 143, Dec. 2022, Art. no. 108516, doi: 10.1016/j.ijepes.2022.108516.

[276] R. C. Duan, F. H. Wang, J. Zhang, R. H. Huang, and X. Zhang, "Data mining & pattern recognition of voltage sag based on K-means clustering algorithm," in *Proc. IEEE Power Energy Soc. Gen. Meeting*. Washington, DC, USA: IEEE Computer Society, Jul. 2015, pp. 1–5, doi: 10.1109/PESGM.2015.7286079.

[277] W. A. Febrianto, I. Gunartono, O. Penangsang, and R. S. Wibowo, "Fault location and voltage sag analysis in electric distribution network," in *Proc. Int. Seminar Intell. Technol. Appl. (ISITIA)*, Aug. 2018, pp. 71–75.

[278] J. Mora-Flórez, J. Cormane-Angarita, and G. Ordóñez-Plata, "k-means algorithm and mixture distributions for locating faults in power systems," *Electric Power Syst. Res.*, vol. 79, no. 5, pp. 714–721, May 2009, doi: 10.1016/j.epsr.2008.10.011.

[279] Y. Mohammadi, A. Salarpour, and R. C. Leborgne, "Comprehensive strategy for classification of voltage sags source location using optimal feature selection applied to support vector machine and ensemble techniques," *Int. J. Electr. Power Energy Syst.*, vol. 124, Jan. 2021, Art. no. 106363.

[280] A. Dejamkhooy, A. Dastfan, and A. Ahmadifard, "K-means clustering and correlation coefficient based methods for detection of flicker sources in non-radial power system," *Russian Electr. Eng.*, vol. 85, no. 4, pp. 251–259, Apr. 2014, doi: 10.3103/S1068371214040142.

[281] P. Shinde, P. Sonwane, and R. Munje, "Machine learning based clustering for identifying power quality events," in *Proc. IEEE 5th Int. Conf. Converg. Technol. (I2CT)*, Mar. 2019, pp. 1–6.

[282] L. Priyadarshini, E. N. V. D. V. Prasad, and P. K. Dash, "Diagnosis of PQ disturbances using local mean decomposition based SVD entropy and modified K-means clustering," in *Proc. Int. Conf. Adv. Power, Signal, Inf. Technol. (APSIT)*. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, Oct. 2021, pp. 1–6, doi: 10.1109/APSIT52773.2021.9641298.

[283] W. Li, S. Chen, X. Peng, M. Xiao, L. Gao, A. Garg, and N. Bao, "A comprehensive approach for the clustering of similar-performance cells for the design of a lithium-ion battery module for electric vehicles," *Engineering*, vol. 5, no. 4, pp. 795–802, Aug. 2019, doi: 10.1016/j.eng.2019.07.005.

[284] X. Lei, Y. Shang, Z. Shao, Y. Jia, and L. Jian, "Grid integration of electric vehicles for optimal marginal revenue of distribution system operator in spot market," *Energy Rep.*, vol. 8, pp. 1061–1068, Nov. 2022, doi: 10.1016/j.egyr.2022.08.107.

[285] X. Lin, Y. Li, and B. Xia, "An online driver behavior adaptive shift strategy for two-speed AMT electric vehicle based on dynamic corrected factor," *Sustain. Energy Technol. Assessments*, vol. 48, Dec. 2021, Art. no. 101598, doi: 10.1016/j.seta.2021.101598.

[286] N. Yan, H. Zhao, S. Ma, and T. Yan, "Research on energy management and control method of microgrid considering health status of batteries in echelon utilization," *Energy Rep.*, vol. 7, pp. 389–395, Apr. 2021, doi: 10.1016/j.egyr.2021.01.055.

[287] F. Yang, S. Yin, S. Zhou, D. Li, C. Fang, and S. Lin, "Electric vehicle charging current scenario generation based on generative adversarial network combined with clustering algorithm," *Int. Trans. Electr. Energy Syst.*, vol. 31, no. 8, Aug. 2021, Art. no. e12971, doi: 10.1002/2050-7038.12971.

[288] Z. Bendiabdellah, S. M. Senouci, and M. Feham, "A hybrid algorithm for planning public charging stations," in *Proc. Global Inf. Infrastructure Netw. Symp. (GIIS)*. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, Sep. 2014, pp. 1–3, doi: 10.1109/GIIS.2014.6934262.

[289] X. Jiang, L. Zhao, Y. Cheng, S. Wei, and Y. Jin, "Optimal configuration of electric vehicles for charging stations under the fast power supplement mode," *J. Energy Storage*, vol. 45, Jan. 2022, Art. no. 103677, doi: 10.1016/j.est.2021.103677.

[290] D. S. Kumar, H. B. A. Jefri, A. Sharma, and W. L. Woo, "Planning of distributed energy resources in Singapore," in *Proc. IEEE Innov. Smart Grid Technol. Asia (ISGT Asia)*, May 2019, pp. 3953–3958.

[291] L. Jiang, Y. Hua, C. Ma, and X. Liu, "SunChase: Energy-efficient route planning for solar-powered EVs," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, Jun. 2017, pp. 383–393, doi: 10.1109/ICDCS.2017.214.

[292] A. Kapoor and A. Singhal, "A comparative study of K-means, K-means++ and fuzzy C-means clustering algorithms," in *Proc. 3rd Int. Conf. Comput. Intell. Commun. Technol. (CICT)*, Feb. 2017, pp. 1–6.

[293] E. Schubert and P. J. Rousseeuw, "Faster k-medoids clustering: Improving the PAM, CLARA, and CLARANS algorithms," in *12th Int. Conf. Similarity Search Appl. (SISAP)*, Newark, NJ, USA, Oct. 2019, pp. 171–187.

[294] D. T. Pham, S. S. Dimov, and C. D. Nguyen, "Selection of K in K-means clustering," *Proc. Inst. Mech. Eng. C, J. Mech. Eng. Sci.*, vol. 219, no. 1, pp. 103–119, 2005.

[295] K. R. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," in *Proc. IEEE 7th Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2020, pp. 747–748.

[296] S. Ahmadian, A. Norouzi-Fard, O. Svensson, and J. Ward, "Better guarantees for k-means and Euclidean k-median by primal-dual algorithms," *SIAM J. Comput.*, vol. 49, no. 4, p. 61, 2019.

[297] A. Feizollah, N. B. Anuar, R. Salleh, and F. Amalina, "Comparative study of k-means and mini batch k-means clustering algorithms in Android malware detection using network traffic analysis," in *Proc. Int. Symp. Biometrics Secur. Technol. (ISBAST)*, Aug. 2014, pp. 193–197.

[298] Y. Endo and S. Miyamoto, "Spherical k-means++ clustering," in *Modeling Decisions for Artificial Intelligence*. Skövde, Sweden: Springer, Sep. 2015, pp. 103–114.

[299] S. M. Savaresi and D. L. Boley, "A comparative analysis on the bisecting K-means and the PDDP clustering algorithms," *Intell. Data Anal.*, vol. 8, no. 4, pp. 345–362, Oct. 2004.

[300] H. Liu, J. Wu, T. Liu, D. Tao, and Y. Fu, "Spectral ensemble clustering via weighted K-means: Theoretical and practical evidence," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 5, pp. 1129–1143, May 2017.

[301] A. Andrews and R. K. Jain, "Beyond energy efficiency: A clustering approach to embed demand flexibility into building energy benchmarking," *Appl. Energy*, vol. 327, Dec. 2022, Art. no. 119989, doi: 10.1016/j.apenergy.2022.119989.

[302] M. S. Yildiz, K. Dogansahin, and B. Kekezoglu, "Cascaded clustering analysis of electricity load profile based on smart metering data," in *Proc. 13th Int. Conf. Electr. Electron. Eng. (ELECO)*, Nov. 2021, pp. 62–66, doi: 10.23919/ELECO54474.2021.9677826.

[303] P. K. Adom and W. Bekoe, "Conditional dynamic forecast of electrical energy consumption requirements in Ghana by 2020: A comparison of ARDL and PAM," *Energy*, vol. 44, no. 1, pp. 367–380, Aug. 2012, doi: 10.1016/j.energy.2012.06.020.

[304] M. Babaei, A. Abazari, M. M. Soleymani, M. Ghafouri, S. M. Muyeen, and M. T. H. Beheshti, "A data-mining based optimal demand response program for smart home with energy storages and electric vehicles," *J. Energy Storage*, vol. 36, Apr. 2021, Art. no. 102407, doi: 10.1016/j.est.2021.102407.

[305] F. Z. Abera and V. Khedkar, "Machine learning approach electric appliance consumption and peak demand forecasting of residential customers using smart meter data," *Wireless Pers. Commun.*, vol. 111, no. 1, pp. 65–82, Mar. 2020.

[306] N. Mararakanye, A. Dalton, and B. Bekker, "Incorporating spatial and temporal correlations to improve aggregation of decentralized day-ahead wind power forecasts," *IEEE Access*, vol. 10, pp. 116182–116195, 2022, doi: 10.1109/ACCESS.2022.3219602.

[307] O. Perpiñán, M. A. Sánchez-Urán, F. Álvarez, J. Ortego, and F. Garnacho, "Signal analysis and feature generation for pattern identification of partial discharges in high-voltage equipment," *Electr. Power Syst. Res.*, vol. 95, pp. 56–65, Feb. 2013, doi: 10.1016/j.epsr.2012.08.016.

[308] W. Cao, S. Wang, and M. Xu, "Optimal scheduling of virtual power plant based on Latin hypercube sampling and improved Clara clustering algorithm," *Processes*, vol. 10, no. 11, p. 2414, Nov. 2022.

[309] Z. Li, F. Yu, and Q. Guan, "BiLSTM network-based approach for electric load forecasting in energy cell-tissue systems," in *Proc. 1st Int. Conf. Cyber-Energy Syst. Intell. Energy (ICCSIE)*, Jan. 2023, pp. 1–5, doi: 10.1109/ICCSIE55183.2023.10175285.

[310] Y. Ou, B. Deng, X. Liu, and K. Zhou, "Local outlier factor based false data detection in power systems," in *Proc. IEEE Sustain. Power Energy Conf. (iSPEC)*, Nov. 2019, pp. 2003–2007.

[311] M. Dey, S. P. Rana, C. V. Simmons, and S. Dudley, "Solar farm voltage anomaly detection using high-resolution μPMU data-driven unsupervised machine learning," *Appl. Energy*, vol. 303, Dec. 2021, Art. no. 117656.

[312] D. Han, H. Bai, Y. Wang, F. Bu, and J. Zhang, "Day-ahead aggregated load forecasting based on household smart meter data," *Energy Rep.*, vol. 9, pp. 149–158, Sep. 2023, doi: 10.1016/j.egyr.2023.04.317.

[313] K. G. Ranjan, B. R. Prusty, and D. Jena, "Review of preprocessing methods for univariate volatile time-series in power system applications," *Electr. Power Syst. Res.*, vol. 191, Feb. 2021, Art. no. 106885.

[314] G. A. Susto, A. Cenedese, and M. Terzi, "Chapter 9—Time-series classification methods: Review and applications to power systems data," in *Big Data Application in Power Systems*, R. Arghandeh and Y. Zhou, Eds. Amsterdam, The Netherlands: Elsevier, 2018, pp. 179–220, doi: 10.1016/B978-0-12-811968-6.00009-7.

[315] Y. Qi, B. Luo, X. Wang, and L. Wu, "Load pattern recognition method based on fuzzy clustering and decision tree," in *Proc. IEEE Conf. Energy Internet Energy Syst. Integr. (EI2)*, Nov. 2017, pp. 1–5.

[316] P. A. Adedeji, S. Akinlabi, N. Madushele, and O. O. Olatunji, "Wind turbine power output very short-term forecast: A comparative study of data clustering techniques in a PSO-ANFIS model," *J. Cleaner Prod.*, vol. 254, May 2020, Art. no. 120135, doi: 10.1016/j.jclepro.2020.120135.

[317] F. Liu, T. Dong, T. Hou, and Y. Liu, "A hybrid short-term load forecasting model based on improved fuzzy C-means clustering, random forest and deep neural networks," *IEEE Access*, vol. 9, pp. 59754–59765, 2021.

[318] D. Wang, F. Yang, L. Gan, and Y. Li, "Fuzzy prediction of power lithium ion battery state of function based on the fuzzy c-means clustering algorithm," *World Electr. Vehicle J.*, vol. 10, no. 1, p. 1, Jan. 2019.

[319] Z. Wu, X. Dong, Z. Liu, X. Kong, Y. Zeng, Q. Hu, and Y. Chen, "Power system bad load data detection based on an improved fuzzy C-means clustering algorithm," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2017, pp. 1–5.

[320] K. R. Sudha, Y. B. Raju, and A. C. Sekhar, "Fuzzy C-means clustering for robust decentralized load frequency control of interconnected power system with generation rate constraint," *Int. J. Electr. Power Energy Syst.*, vol. 37, no. 1, pp. 58–66, May 2012.

[321] H. Tan, Y. Li, L. Yu, J. Liu, L. Ni, and X. Diao, "Estimation method of line loss rate in low voltage area based on mean shift clustering and BP neural network," *J. Phys., Conf. Ser.*, vol. 1754, no. 1, Feb. 2021, Art. no. 012225.

[322] K. Blazakis and G. Stavrakakis, "Efficient power theft detection for residential consumers using mean shift data mining knowledge discovery process," 2019, *arXiv:1902.03296*.

[323] Z. Wang and G. Zheng, "Residential appliances identification and monitoring by a nonintrusive method," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 80–92, Mar. 2012, doi: 10.1109/TSG.2011.2163950.

[324] M. Netzer, J. Michelberger, and J. Fleischer, "Intelligent anomaly detection of machine tools based on mean shift clustering," *Proc. CIRP*, vol. 93, pp. 1448–1453, Jan. 2020.

[325] L. Rego, J. Sumaili, V. Miranda, C. Francês, M. Silva, and Á. Santana, "Mean shift densification of scarce data sets in short-term electric power load forecasting for special days," *Electr. Eng.*, vol. 99, no. 3, pp. 881–898, Sep. 2017.

[326] N. Kim, S. Park, J. Lee, and J. Choi, "Load profile extraction by mean-shift clustering with sample Pearson correlation coefficient distance," *Energies*, vol. 11, no. 9, p. 2397, Sep. 2018.

[327] F. AlMahamid and K. Grolinger, "Agglomerative hierarchical clustering with dynamic time warping for household load curve clustering," in *Proc. IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, Sep. 2022, pp. 241–247, doi: 10.1109/CCECE49351.2022.9918481.

[328] Y. Liu, R. Sioshansi, and A. J. Conejo, "Hierarchical clustering to find representative operating periods for capacity-expansion modeling," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3029–3039, May 2018.

[329] K. Li, Z. Ma, D. Robinson, and J. Ma, "Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering," *Appl. Energy*, vol. 231, pp. 331–342, Dec. 2018, doi: 10.1016/j.apenergy.2018.09.050.

[330] D. Kiran, A. R. Abhyankar, and B. K. Panigrahi, "Hierarchical clustering based zone formation in power networks," in *Proc. Nat. Power Syst. Conf. (NPSC)*, Dec. 2016, pp. 1–6.

[331] K. Manivinnan, C. L. Benner, B. D. Russell, and J. A. Wischkaemper, "Automatic identification, clustering and reporting of recurrent faults in electric distribution feeders," in *Proc. 19th Int. Conf. Intell. Syst. Appl. Power Syst. (ISAP)*, Sep. 2017, pp. 1–6.

[332] P. Barmpas et al., "A divisive hierarchical clustering methodology for enhancing the ensemble prediction power in large scale population studies: The ATHLOS project," *Health Inf. Sci. Syst.*, vol. 10, no. 1, p. 6, Apr. 2022.

[333] Z. A. Khan, D. Jayaweera, and M. S. Alvarez-Alvarado, "A novel approach for load profiling in smart power grids using smart meter data," *Electr. Power Syst. Res.*, vol. 165, pp. 191–198, Dec. 2018.

[334] A. D. Fontanini and J. Abreu, "A data-driven BIRCH clustering method for extracting typical load profiles for big data," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2018, pp. 1–5.

[335] Y.-L. He, H.-L. Qin, J. Z. Huang, and Y. Jin, "Novel electricity pattern identification system based on improved I-nice algorithm," *Comput. Ind. Eng.*, vol. 150, Dec. 2020, Art. no. 106908, doi: 10.1016/j.cie.2020.106908.

[336] L. Wang, S. Mao, and B. Wilamowski, "Short-term load forecasting with LSTM based ensemble learning," in *Proc. Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Jul. 2019, pp. 793–800.

[337] R. Yu, J. Gao, M. Yu, W. Lu, T. Xu, M. Zhao, J. Zhang, R. Zhang, and Z. Zhang, "LSTM-EFG for wind power forecasting based on sequential correlation features," *Future Gener. Comput. Syst.*, vol. 93, pp. 33–42, Apr. 2019.

[338] Z. Šulc and H. Řezanková, "Comparison of similarity measures for categorical data in hierarchical clustering," *J. Classification*, vol. 36, no. 1, pp. 58–72, Apr. 2019.

[339] Vijaya, S. Sharma, and N. Batra, "Comparative study of single linkage, complete linkage, and ward method of agglomerative clustering," in *Proc. Int. Conf. Mach. Learn., Big Data, Cloud Parallel Comput. (COMITCon)*, Feb. 2019, pp. 568–573.

[340] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: An overview. II," *WIREs Data Mining Knowl. Discovery*, vol. 7, no. 6, Nov. 2017, Art. no. e1s219.

[341] J. Zhang, J. Yan, D. Infield, Y. Liu, and F.-S. Lien, "Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model," *Appl. Energy*, vol. 241, pp. 229–244, May 2019.

[342] C. S. Lai, Y. Jia, M. D. McCulloch, and Z. Xu, "Daily clearness index profiles cluster analysis for photovoltaic system," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2322–2332, Oct. 2017, doi: 10.1109/TII.2017.2683519.

[343] M. Sun, I. Konstantelos, and G. Strbac, "C-vine copula mixture model for clustering of residential electrical load pattern data," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2382–2393, May 2017.

[344] L. Zhang, L. Wan, Y. Xiao, S. Li, and C. Zhu, "Anomaly detection method of smart meters data based on GMM-LDA clustering feature learning and PSO support vector machine," in *Proc. IEEE Sustain. Power Energy Conf. (iSPEC)*, Nov. 2019, pp. 2407–2412, doi: 10.1109/iSPEC48194.2019.8974989.

[345] H. Zhu, H. Dai, S. Yang, Y. Yan, and B. Lin, "Estimating power consumption of servers using Gaussian mixture model," in *Proc. 5th Int. Symp. Comput. Netw. (CANDAR)*, Nov. 2017, pp. 427–433.

[346] R. Li, Z. Wang, C. Gu, F. Li, and H. Wu, "A novel time-of-use tariff design based on Gaussian mixture model," *Appl. Energy*, vol. 162, pp. 1530–1536, Jan. 2016.

[347] T. Ahmad and N. Senroy, "Statistical characterization of PMU error for robust WAMS based analytics," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 920–928, Mar. 2020, doi: 10.1109/TPWRS.2019.2939098.

[348] A. V. Pankratov, N. L. Batseva, E. S. Polyakova, A. S. Tavlintsev, I. L. Lapatin, and I. Y. Lipnitskiy, "Application of expectation maximization algorithm for measurement-based power system load modeling," in *Proc. Int. Siberian Conf. Control Commun. (SIBCON)*, Apr. 2019, pp. 1–5.

[349] R. Bonfigli, E. Principi, M. Fagiani, M. Severini, S. Squartini, and F. Piazza, "Non-intrusive load monitoring by using active and reactive power in additive factorial hidden Markov models," *Appl. Energy*, vol. 208, pp. 1590–1607, Dec. 2017, doi: 10.1016/j.apenergy.2017.08.203.

[350] J. Jiang, R. Chen, M. Chen, W. Wang, and C. Zhang, "Dynamic fault prediction of power transformers based on hidden Markov model of dissolved gases analysis," *IEEE Trans. Power Del.*, vol. 34, no. 4, pp. 1393–1400, Aug. 2019, doi: 10.1109/TPWRD.2019.2900543.

[351] F. Xiao and Q. Ai, "Data-driven multi-hidden Markov model-based power quality disturbance prediction that incorporates weather conditions," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 402–412, Jan. 2019, doi: 10.1109/TPWRS.2018.2856743.

[352] S. Lu, G. Lin, H. Liu, C. Ye, H. Que, and Y. Ding, "A weekly load data mining approach based on hidden Markov model," *IEEE Access*, vol. 7, pp. 34609–34619, 2019, doi: 10.1109/ACCESS.2019.2901197.

[353] J. P. Hermias, K. Teknomo, and J. C. N. Monje, "Short-term stochastic load forecasting using autoregressive integrated moving average models and hidden Markov model," in *Proc. Int. Conf. Inf. Commun. Technol. (ICICT)*, Dec. 2017, pp. 131–137, doi: 10.1109/ICICT.2017.8320177.

[354] Y.-Y. Hong and M.-Y. Wu, "Markov model-based energy storage system planning in power systems," *IEEE Syst. J.*, vol. 13, no. 4, pp. 4313–4323, Dec. 2019, doi: 10.1109/JSYST.2019.2900081.

[355] B. Stephen, X. Tang, P. R. Harvey, S. Galloway, and K. I. Jennett, "Incorporating practice theory in sub-profile models for short term aggregated residential load forecasting," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1591–1598, Jul. 2017, doi: 10.1109/TSG.2015.2493205.

[356] F. Landini, J. Profant, M. Diez, and L. Burget, "Bayesian HMM clustering of x-vector sequences (VBx) in speaker diarization: Theory, implementation and analysis on standard tasks," *Comput. Speech Lang.*, vol. 71, Jan. 2022, Art. no. 101254, doi: 10.1016/j.csl.2021.101254.

[357] S. Wade and Z. Ghahramani, "Bayesian cluster analysis: Point estimation and credible balls (with discussion)," *Bayesian Anal.*, vol. 13, no. 2, pp. 559–626, Jun. 2018, doi: 10.1214/17-BA1073.

[358] E. Ni and P. B. Luh, "Forecasting power market clearing price and its discrete PDF using a Bayesian-based classification method," in *Proc. IEEE Power Eng. Soc. Winter Meeting*, Feb. 2001, pp. 1518–1523, doi: 10.1109/PESW.2001.917338.

[359] M. Sun, T. Zhang, Y. Wang, G. Strbac, and C. Kang, "Using Bayesian deep learning to capture uncertainty for residential net load forecasting," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 188–201, Jan. 2020, doi: 10.1109/TPWRS.2019.2924294.

[360] M. Gilanifar, H. Wang, L. M. K. Sriram, E. E. Ozguven, and R. Arghandeh, "Multitask Bayesian spatiotemporal Gaussian processes for short-term load forecasting," *IEEE Trans. Ind. Electron.*, vol. 67, no. 6, pp. 5132–5143, Jun. 2020, doi: 10.1109/TIE.2019.2928275.

[361] S. Wang, X. Sun, and U. Lall, "A hierarchical Bayesian regression model for predicting summer residential electricity demand across the U.S.A.," *Energy*, vol. 140, pp. 601–611, Dec. 2017, doi: 10.1016/j.energy.2017.08.076.

[362] Y. Jinsong, L. Shuang, T. Diyin, and L. Hao, "Remaining discharge time prognostics of lithium-ion batteries using Dirichlet process mixture model and particle filtering method," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 9, pp. 2317–2328, Sep. 2017, doi: 10.1109/TIM.2017.2708204.

[363] R. Granell, C. J. Axon, and D. C. H. Wallom, "Clustering disaggregated load profiles using a Dirichlet process mixture model," *Energy Convers. Manage.*, vol. 92, pp. 507–516, Mar. 2015, doi: 10.1016/j.enconman.2014.12.080.

[364] Y. Gan, S. Ye, and P. Guo, "Abnormal wind turbine data identification using a Dirichlet process Gaussian mixture model," in *Proc. 34th Chin. Control Decis. Conf. (CCDC)*, Aug. 2022, pp. 529–534, doi: 10.1109/CCDC55256.2022.10034077.

[365] H. Liu, Z. Duan, C. Chen, and H. Wu, "A novel two-stage deep learning wind speed forecasting method with adaptive multiple error corrections and bivariate Dirichlet process mixture model," *Energy Convers. Manage.*, vol. 199, Nov. 2019, Art. no. 111975, doi: 10.1016/j.enconman.2019.111975.

[366] L. Wang, S. Mao, B. M. Wilamowski, and R. M. Nelms, "Ensemble learning for load forecasting," *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 2, pp. 616–628, Jun. 2020, doi: 10.1109/TGCN.2020.2987304.

[367] K. Aurangzeb, M. Alhussein, K. Javaid, and S. I. Haider, "A pyramid-CNN based deep learning model for power load forecasting of similar-profile energy customers based on clustering," *IEEE Access*, vol. 9, pp. 14992–15003, 2021, doi: 10.1109/ACCESS.2021.3053069.

[368] X. Wang, C. Zhou, Y. Yang, Y. Yang, T. Ji, J. Wang, J. Chen, and Y. Zheng, "Electricity market customer segmentation based on DBSCAN and k-means: A case on Yunnan electricity market," in *Proc. Asia Energy Electr. Eng. Symp. (AEEES)*, May 2020, pp. 869–874, doi: 10.1109/AEEES48850.2020.9121413.

[369] P. Wang and M. Govindarasu, "Anomaly detection for power system generation control based on hierarchical DBSCAN," in *Proc. North Amer. Power Symp. (NAPS)*, Sep. 2018, pp. 1–5, doi: 10.1109/NAPS.2018.8600616.

[370] L. Zhang, S. Deng, and S. Li, "Analysis of power consumer behavior based on the complementation of K-means and DBSCAN," in *Proc. IEEE Conf. Energy Internet Energy Syst. Integr. (EI2)*, Nov. 2017, pp. 1–5, doi: 10.1109/EI2.2017.8245490.

[371] K. Li, B. Wang, Z. Wang, F. Wang, Z. Mi, and Z. Zhen, "A baseline load estimation approach for residential customer based on load pattern clustering," *Energy Proc.*, vol. 142, pp. 2042–2049, Dec. 2017, doi: 10.1016/j.egypro.2017.12.408.

[372] Y. Qu, J. Xu, Y. Sun, and D. Liu, "A temporal distributed hybrid deep learning model for day-ahead distributed PV power forecasting," *Appl. Energy*, vol. 304, Dec. 2021, Art. no. 117704, doi: 10.1016/j.apenergy.2021.117704.

[373] M. Seera, C. P. Lim, C. K. Loo, and H. Singh, "A modified fuzzy min–max neural network for data clustering and its application to power quality monitoring," *Appl. Soft Comput.*, vol. 28, pp. 19–29, Mar. 2015, doi: 10.1016/j.asoc.2014.09.050.

[374] Z. Luo, S. Hong, and Y. Ding, "A data mining-driven incentive-based demand response scheme for a virtual power plant," *Appl. Energy*, vol. 239, pp. 549–559, Apr. 2019, doi: 10.1016/j.apenergy.2019.01.142.

[375] F. Fathnia and M. H. J. D. Bayaz, "Anomaly detection in smart grid with help of an improved OPTICS using coefficient of variation," in *Proc. Iranian Conf. Electr. Eng. (ICEE)*, May 2018, pp. 1044–1050, doi: 10.1109/ICEE.2018.8472534.

[376] F. Znidi, H. Davarikia, M. Arani, and M. Barati, "Coherency detection and network partitioning based on hierarchical DBSCAN," in *Proc. IEEE Texas Power Energy Conf. (TPEC)*, Feb. 2020, pp. 1–5, doi: 10.1109/TPEC48276.2020.9042511.

[377] A. Rajabi, M. Eskandari, M. J. Ghadi, L. Li, J. Zhang, and P. Siano, "A comparative study of clustering techniques for electrical load pattern segmentation," *Renew. Sustain. Energy Rev.*, vol. 120, Mar. 2020, Art. no. 109628, doi: 10.1016/j.rser.2019.109628.

[378] W. J. Farmer and A. J. Rix, "Evaluating power system network inertia using spectral clustering to define local area stability," *Int. J. Electr. Power Energy Syst.*, vol. 134, Jan. 2022, Art. no. 107404, doi: 10.1016/j.ijepes.2021.107404.

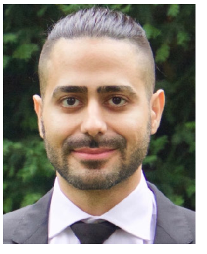
[379] H. Rehioui, A. Idrissi, M. Abourezq, and F. Zegrari, "DENCLUE-IM: A new approach for big data clustering," *Proc. Comput. Sci.*, vol. 83, pp. 560–567, Jan. 2016, doi: 10.1016/j.procs.2016.04.265.

[380] A. Idrissi, H. Rehioui, A. Laghrissi, and S. Retal, "An improvement of DENCLUE algorithm for the data clustering," in *Proc. 5th Int. Conf. Inf. Commun. Technol. Accessibility (ICTA)*, Dec. 2015, pp. 1–6, doi: 10.1109/ICTA.2015.7426936.

[381] H.-Z. Xia, L.-M. Chen, F. Qi, X.-D. Mao, L.-Q. Sun, and F.-Y. Xue, "DP-denclue: An outlier detection algorithm with differential privacy preservation," in *Proc. IEEE 24th Int. Conf. High Perform. Comput. Commun., 8th Int. Conf. Data Sci. Syst., 20th Int. Conf. Smart City, 8th Int. Conf. Dependability Sensor, Cloud Big Data Syst. Appl. (HPCC/DSS/SmartCity/DependSys)*, Dec. 2022, pp. 2264–2269, doi: 10.1109/HPCC-DSS-SmartCity-DependSys57074.2022.00334.

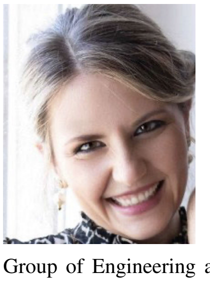
[382] I. A. Venkatkumar and S. J. K. Shardaben, "Comparative study of data mining clustering algorithms," in *Proc. Int. Conf. Data Sci. Eng. (ICDSE)*, Aug. 2016, pp. 1–7, doi: 10.1109/ICDSE.2016.7823946.
[383] V. K. Gonugunta, T. Sakai, V. Pokatayev, K. Yang, J. Wu, N. Dobbs, and N. Yan, "Trafficking-mediated STING degradation requires sorting to acidified endolysosomes and can be targeted to enhance anti-tumor response," *Cell Rep.*, vol. 21, no. 11, pp. 3234–3242, Dec. 2017, doi: 10.1016/j.celrep.2017.11.061.
[384] T. Qian, S. Sun, X. Shan, X. Wei, C. Tai, and C. Liu, "Distributed-swarm: A real-time pattern detection model based on density clustering," *IEEE Access*, vol. 10, pp. 59832–59842, 2022, doi: 10.1109/ACCESS.2022.3179367.
[385] G. Sheikholeslami, S. Chatterjee, and A. Zhang, "WaveCluster: A wavelet-based clustering approach for spatial data in very large databases," *VLDB J. Int. J. Very Large Data Bases*, vol. 8, nos. 3–4, pp. 289–304, Feb. 2000, doi: 10.1007/s007780050009.
[386] Z. Gui, D. Peng, H. Wu, and X. Long, "MSGC: Multi-scale grid clustering by fusing analytical granularity and visual cognition for detecting hierarchical spatial patterns," *Future Gener. Comput. Syst.*, vol. 112, pp. 1038–1056, Nov. 2020, doi: 10.1016/j.future.2020.06.053.
[387] L. Ye, J. Zhao, P. Lu, M. Pei, M. Chen, and B. Wang, "Combined approach for short-term wind power forecasting considering meteorological fluctuation and feature extraction," in *Proc. IEEE/IAS Ind. Commercial Power Syst. Asia (ICPS Asia)*, Jul. 2020, pp. 1334–1343, doi: 10.1109/ICPSAsia48933.2020.9208615.
[388] H. Wenrui, L. Wei, and Z. Hongli, "Extraction method of renewable energy output interval based on CLIQUE clustering," in *Proc. Power Syst. Green Energy Conf. (PSGEC)*, Aug. 2022, pp. 159–163, doi: 10.1109/PSGEC54663.2022.9881185.
[389] O. P. Dahal, S. M. Brahma, and H. Cao, "Comprehensive clustering of disturbance events recorded by phasor measurement units," *IEEE Trans. Power Del.*, vol. 29, no. 3, pp. 1390–1397, Jun. 2014, doi: 10.1109/TPWRD.2013.2285097.
[390] N. Nastaiinullah, K. Adiwijaya, and A. P. Kurniati, "Anomaly detection on intrusion detection system using CLIQUE partitioning," in *Proc. 2nd Int. Conf. Inf. Commun. Technol. (ICoICT)*, May 2014, pp. 7–12, doi: 10.1109/ICoICT.2014.6914031.
[391] S. Gao, B. Hu, K. Xie, T. Niu, C. Li, and J. Yan, "Spectral clustering based demand-oriented representative days selection method for power system expansion planning," *Int. J. Electr. Power Energy Syst.*, vol. 125, Feb. 2021, Art. no. 106560, doi: 10.1016/j.ijepes.2020.106560.
[392] C. Dinesh, S. Makonin, and I. V. Bajić, "Residential power forecasting using load identification and graph spectral clustering," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 11, pp. 1900–1904, Nov. 2019, doi: 10.1109/TCSII.2019.2891704.
[393] L. Ding, F. M. Gonzalez-Longatt, P. Wall, and V. Terzija, "Two-step spectral clustering controlled islanding algorithm," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 75–84, Feb. 2013, doi: 10.1109/TPWRS.2012.2197640.
[394] M. Dabbaghjamanesh, B. Wang, A. Kavousi-Fard, S. Mehraeen, N. D. Hatziargyriou, D. N. Trakas, and F. Ferdowsi, "A novel two-stage multi-layer constrained spectral clustering strategy for intentional islanding of power grids," *IEEE Trans. Power Del.*, vol. 35, no. 2, pp. 560–570, Apr. 2020, doi: 10.1109/TPWRD.2019.2915342.
[395] Y. Li and S. Wu, "Controlled islanding for a hybrid AC/DC grid with VSC-HVDC using semi-supervised spectral clustering," *IEEE Access*, vol. 7, pp. 10478–10490, 2019, doi: 10.1109/ACCESS.2018.2886533.
[396] X. Zhang, J. L. Ramírez-Mendiola, M. Li, and L. Guo, "Electricity consumption pattern analysis beyond traditional clustering methods: A novel self-adapting semi-supervised clustering method and application case study," *Appl. Energy*, vol. 308, Feb. 2022, Art. no. 118335, doi: 10.1016/j.apenergy.2021.118335.
[397] Q. Gao, Y. Wang, X. Cheng, J. Yu, X. Chen, and T. Jing, "Identification of vulnerable lines in smart grid systems based on affinity propagation clustering," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5163–5171, Jun. 2019, doi: 10.1109/JIOT.2019.2897434.
[398] J. Ortiz-Bejar, M. R. A. Paternina, A. Zamora-Mendez, L. Lugnani, and E. Tellez, "Power system coherency assessment by the affinity propagation algorithm and distance correlation," *Sustain. Energy, Grids Netw.*, vol. 30, Jun. 2022, Art. no. 100658, doi: 10.1016/j.segan.2022.100658.
[399] H. Hassani, R. Razavi-Far, and M. Saif, "Fault location in smart grids through multicriteria analysis of group decision support systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 12, pp. 7318–7327, Dec. 2020, doi: 10.1109/TII.2020.2977980.
[400] P. Li, W. Gu, L. Wang, B. Xu, M. Wu, and W. Shen, "Dynamic equivalent modeling of two-staged photovoltaic power station clusters based on dynamic affinity propagation clustering algorithm," *Int. J. Electr. Power Energy Syst.*, vol. 95, pp. 463–475, Feb. 2018, doi: 10.1016/j.ijepes.2017.08.038.
[401] E. Ferrario, N. Pedroni, and E. Zio, "Evaluation of the robustness of critical infrastructures by hierarchical graph representation, clustering and Monte Carlo simulation," *Rel. Eng. Syst. Saf.*, vol. 155, pp. 78–96, Nov. 2016, doi: 10.1016/j.ress.2016.06.007.
[402] M. Zeng, Y. Pan, D. Zhang, Z. Lu, and Y. Li, "Data-driven location selection for battery swapping stations," *IEEE Access*, vol. 7, pp. 133760–133771, 2019, doi: 10.1109/ACCESS.2019.2941901.
[403] S. Wei, H. Lei, and X. Jian, "Fault detection identification and location based on measurement data from finite synchronized phasor measurement," *Power Syst. Technol.*, vol. 39, no. 4, pp. 1067–1074, 2015.
[404] X. Fan, M. Gong, and Y. Wu, "Markov clustering regularized multi-hop graph neural network," *Pattern Recognit.*, vol. 139, Jul. 2023, Art. no. 109518, doi: 10.1016/j.patcog.2023.109518.
[405] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, "Generalized Louvain method for community detection in large networks," in *Proc. 11th Int. Conf. Intell. Syst. Design Appl.*, Nov. 2011, pp. 88–93, doi: 10.1109/ISDA.2011.6121636.
[406] M. E. Eddin, M. Massaoudi, H. Abu-Rub, M. Shadmand, and M. Abdallah, "Novel functional community detection in networked smart grid systems-based improved Louvain algorithm," in *Proc. IEEE Texas Power Energy Conf. (TPEC)*, Feb. 2023, pp. 1–6, doi: 10.1109/TPEC56611.2023.10078573.
[407] G. Lin, S. Liu, A. Zhou, J. Dai, B. Chai, B. Zhang, H. Qiu, K. Gao, Y. Song, and R. Chen, "Community detection in power grids based on Louvain heuristic algorithm," in *Proc. IEEE Conf. Energy Internet Energy Syst. Integr. (EI2)*, Nov. 2017, pp. 1–4, doi: 10.1109/EI2.2017.8245596.
[408] L. Wang, A. Dubey, A. H. Gebremedhin, A. K. Srivastava, and N. Schulz, "MPC-based decentralized voltage control in power distribution systems with EV and PV coordination," *IEEE Trans. Smart Grid*, vol. 13, no. 4, pp. 2908–2919, Jul. 2022, doi: 10.1109/TSG.2022.3156115.
[409] D. D. Hieu and P. T. H. Duong, "Detecting communities in large networks using the extended walktrap algorithm," in *Proc. RIVF Int. Conf. Comput. Commun. Technol. (RIVF)*, Dec. 2022, pp. 100–105, doi: 10.1109/RIVF55975.2022.10013880.
[410] Z. Halim, H. M. Sargana, Aadam, Uzma, and M. Waqas, "Clustering of graphs using pseudo-guided random walk," *J. Comput. Sci.*, vol. 51, Apr. 2021, Art. no. 101281, doi: 10.1016/j.jocs.2020.101281.
[411] O. Bahri, P. Li, S. F. Boubrahimi, and S. M. Hamdi, "Shapelet-based temporal association rule mining for multivariate time series classification," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2022, pp. 242–251, doi: 10.1109/BigData55660.2022.10020478.
[412] J. Hills, J. Lines, E. Baranauskas, J. Mapp, and A. Bagnall, "Classification of time series by shapelet transformation," *Data Mining Knowl. Discovery*, vol. 28, no. 4, pp. 851–881, Jul. 2014, doi: 10.1007/s10618-013-0322-1.
[413] L. Zhang, H. Su, E. Zio, L. Jiang, L. Fan, and J. Zhang, "A graph structure feature-based framework for the pattern recognition of the operational states of integrated energy systems," *Expert Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 119039, doi: 10.1016/j.eswa.2022.119039.
[414] H. Yu, C. Xu, G. Geng, and Q. Jiang, "Multi-time-scale shapelet based feature extraction for non-intrusive load monitoring," *IEEE Trans. Smart Grid*, early access, Jun. 12, 2023, doi: 10.1109/TSG.2023.3285117.
[415] L. Zhu and D. J. Hill, "Networked time series shapelet learning for power system transient stability assessment," *IEEE Trans. Power Syst.*, vol. 37, no. 1, pp. 416–428, Jan. 2022, doi: 10.1109/TPWRS.2021.3093423.
[416] L. Zhu, C. Lu, and Y. Luo, "Time series data-driven batch assessment of power system short-term voltage security," *IEEE Trans. Ind. Informat.*, vol. 16, no. 12, pp. 7306–7317, Dec. 2020, doi: 10.1109/TII.2020.2977456.
[417] D. W. van der Merwe and A. P. Engelbrecht, "Data clustering using particle swarm optimization," in *Proc. Congr. Evol. Comput. (CEC)*, vol. 1, 2003, pp. 215–220, doi: 10.1109/CEC.2003.1299577.

[418] A. Abraham, S. Das, and S. Roy, "Swarm intelligence algorithms for data clustering," in *Soft Computing for Knowledge Discovery and Data Mining*, O. Maimon and L. Rokach, Eds. Boston, MA, USA: Springer, 2008, pp. 279–313, doi: 10.1007/978-0-387-69935-6_12.
[419] M. Dash, N. D. Londhe, S. Ghosh, V. K. Shrivastava, and R. S. Sonawane, "Swarm intelligence based clustering technique for automated lesion detection and diagnosis of psoriasis," *Comput. Biol. Chem.*, vol. 86, Jun. 2020, Art. no. 107247, doi: 10.1016/j.compbiolchem.2020.107247.
[420] S. Kalyani and K. S. Swarup, "Particle swarm optimization based K-means clustering approach for security assessment in power systems," *Expert Syst. Appl.*, vol. 38, no. 9, pp. 10839–10846, Sep. 2011, doi: 10.1016/j.eswa.2011.02.086.
[421] M. Y. Arafat and S. Moh, "Localization and clustering based on swarm intelligence in UAV networks for emergency communications," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8958–8976, Oct. 2019, doi: 10.1109/JIOT.2019.2925567.
[422] M. T. Hagh, P. Amiyan, S. Galvani, and N. Valizadeh, "Probabilistic load flow using the particle swarm optimisation clustering method," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 3, pp. 780–789, Feb. 2018.
[423] A. A. Munshi and Y. A.-R.-I. Mohamed, "Photovoltaic power pattern clustering based on conventional and swarm clustering methods," *Sol. Energy*, vol. 124, pp. 39–56, Feb. 2016, doi: 10.1016/j.solener.2015.11.010.
[424] D. Gupta, A. Khanna, L. Sk, K. Shankar, V. Furtado, and J. J. P. C. Rodrigues, "Efficient artificial fish swarm based clustering approach on mobility aware energy-efficient for MANET," *Trans. Emerg. Telecommun. Technol.*, vol. 30, no. 9, Sep. 2019, Art. no. e3524.
[425] Y. Wang, D. Wang, and Y. Tang, "Clustered hybrid wind power prediction model based on ARMA, PSO-SVM, and clustering methods," *IEEE Access*, vol. 8, pp. 17071–17079, 2020, doi: 10.1109/ACCESS.2020.2968390.
[426] A. Ullah, K. Haydarov, I. U. Haq, K. Muhammad, S. Rho, M. Lee, and S. W. Baik, "Deep learning assisted buildings energy consumption profiling using smart meter data," *Sensors*, vol. 20, no. 3, p. 873, Feb. 2020.
[427] J. Lee, J. Kim, and W. Ko, "Day-ahead electric load forecasting for the residential building with a small-size dataset based on a self-organizing map and a stacking ensemble learning method," *Appl. Sci.*, vol. 9, no. 6, p. 1231, Mar. 2019.
[428] B. Zheng, P. He, L. Zhao, and H. Li, "A hybrid machine learning model for range estimation of electric vehicles," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6, doi: 10.1109/GLOCOM.2016.7841506.
[429] S. V. Verdu, M. O. Garcia, C. Senabre, A. G. Marin, and F. J. G. Franco, "Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1672–1682, Nov. 2006, doi: 10.1109/TPWRS.2006.881133.
[430] I. Jahan, F. Mohamed, V. Blazek, L. Prokop, S. Misak, and V. Snasel, "Power quality parameters forecasting based on SOM maps with KNN algorithm and decision tree," in *Proc. 23rd Int. Sci. Conf. Electr. Power Eng. (EPE)*, May 2023, pp. 1–6, doi: 10.1109/EPE58302.2023.10149269.
[431] J.-L. Wu, P.-C. Chang, C.-C. Tsao, and C.-Y. Fan, "A patent quality analysis and classification system using self-organizing maps with support vector machine," *Appl. Soft Comput.*, vol. 41, pp. 305–316, Apr. 2016, doi: 10.1016/j.asoc.2016.01.020.
[432] G. Wang, R. Jia, J. Liu, and H. Zhang, "A hybrid wind power forecasting approach based on Bayesian model averaging and ensemble learning," *Renew. Energy*, vol. 145, pp. 2426–2434, Jan. 2020, doi: 10.1016/j.renene.2019.07.166.
[433] D. Hallac, S. Vare, S. Boyd, and J. Leskovec, "Toeplitz inverse covariance-based clustering of multivariate time series data," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2017, pp. 215–223.
[434] V. Kapp, M. C. May, G. Lanza, and T. Wuest, "Pattern recognition in multivariate time series: Towards an automated event detection method for smart manufacturing systems," *J. Manuf. Mater. Process.*, vol. 4, no. 3, p. 88, Sep. 2020.
[435] Z. Wei, X. Shen, G. Qiu, Y. Liu, and J. Liu, "An integrated transfer learning method for power generation prediction of run-off small hydropower in data-scarce areas," *IEEE Trans. Smart Grid*, early access, May 18, 2023, doi: 10.1109/TSG.2023.3276390.
[436] H. Ouyang, X. Wei, and Q. Wu, "Discovery and prediction of stock index pattern via three-stage architecture of TICC, TPA-LSTM and multivariate LSTM-FCNs," *IEEE Access*, vol. 8, pp. 123683–123700, 2020.
[437] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.
[438] C. G. K. Ziegler et al., "SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues," *Cell*, vol. 181, no. 5, pp. 1016–1035, 2020.
[439] H. Averbuch-Elor, N. Bar, and D. Cohen-Or, "Border-peeling clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 7, pp. 1791–1797, Jul. 2020.
[440] X. Ran, Y. Xi, Y. Lu, X. Wang, and Z. Lu, "Comprehensive survey on hierarchical clustering algorithms and the recent developments," *Artif. Intell. Rev.*, vol. 56, no. 8, pp. 8219–8264, Aug. 2023.
[441] M. Du, J. Zhao, J. Sun, and Y. Dong, "M3W: Multistep three-way clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 29, 2022, doi: 10.1109/TNNLS.2022.3208418.



SEYED MAHDI MIRAFTABZADEH (Member, IEEE) received the M.Sc. degree in electrical engineering from Politecnico di Milano, Italy, in 2017, and the Ph.D. degree, in 2021. He is currently an Assistant Professor with the Department of Energy, Politecnico di Milano. His current research interests include the application of artificial intelligence and deep learning in electrical systems and smart cities to improve power grid stability and promote e-mobilities.

CRISTIAN GIOVANNI COLOMBO (Student Member, IEEE) received the Master of Science degree in electrical engineering from Politecnico di Milano, where he is currently pursuing the Ph.D. degree with the Department of Energy. He is a member of the Transportation Research Group, Department of Energy. His research interests include the e-mobility and how these interfaces with the electricity grid, optimization, optimal management, and siting of charging stations for electric vehicles.



MICHELA LONGO (Member, IEEE) received the M.Sc. degree in information engineering and the Ph.D. degree in mechatronics, information, innovative technologies and mathematical methods from the University of Bergamo, Bergamo, Italy, in 2009 and 2013, respectively. She is currently an Associate Professor with the Department of Energy, Politecnico di Milano. Her research interests include electric power systems and electric traction. She is also a member of the Italian Group of Engineering about Railways (CIFI) and the Italian Association of Electrical, Electronics, Automation, Information and Communication Technology (AEIT).

FEDERICA FOIADELLI (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from Politecnico di Milano, Milan, Italy, in 2003 and 2008, respectively. She is currently an Associate Professor with the Department of Energy, Politecnico di Milano. Her research interests include electric power systems and electric traction. She is a member of the Italian Group of Engineering about Railways (CIFI) and the Italian Electric Association (AEIT).

Open Access funding provided by 'Politecnico di Milano' within the CRUI CARE Agreement