Journal of Theoretical and Applied Information Technology
31st May 2024. Vol.102. No. 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CLUSTERING ALGORITHM FOR ELECTRICAL LOAD PROFILING ANALYSIS: A SYSTEMATIC REVIEW OF MACHINE LEARNING APPROACHES FOR IMPROVED CLUSTERING ALGORITHMS

DINE TIARA KUSUMA[1,2] , NORASHIKIN AHMAD[1], SHARIFAH SAKINAH SYED AHMAD[1]

[1]Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka, Malaysia

[2]Faculty of Energy Telematics, Institut Teknologi PLN, Jakarta, Indonesia

E-mail: [1]dinetiarakusuma@gmail.com, [2]norashikin@utem.edu.my, [3]sakinah@utem.edu.my

ABSTRACT

The objective of this study is to examine a range of research studies conducted between 2017 until 2023 that focus on the analysis of Electrical Load Profiles (ELPs) using clustering algorithms within a machine learning framework. The methodology used in this research is Preferred Reporting for Systematic Review and Meta-analysis (PRISMA) framework. According to this study, it was discovered that the process of formation using the clustering algorithm can be categorized into two distinct approaches. The first approach involves the utilization of a specified number of clusters, while the second approach does not necessitate the explicit determination of the number of clusters. Additionally, it has been observed that the method employed to determine the number of clusters has a significant impact on the performance and quality of clustering, as it influences the features involved. This study explores various aspects related to clustering, including techniques for measuring the distance between data points, strategies for initializing cluster centers, approaches for reducing the dimensions of initial data, and methods for identifying and addressing outliers. The findings of this study offer insights into the various technological obstacles and emerging patterns in the analysis of ELPs, as well as investigate potential prospects for the future.

Keywords: *Clustering, Machine Learning, Load Profiles, Pattern Recognition, PRISMA*

1. INTRODUCTION

The energy sector has embraced the big data trend, as evidenced by the growing interest of researchers in gathering and analyzing energy data [1], [2]. This shift towards big data analytics for flexible energy sharing signifies a significant evolution in how the industry approaches data, emphasizing the importance of comprehensive data collection and advanced analytics for more effective energy management. Energy consumption data can be obtained from several sources such as Energy Meters that use recording devices on energy meters. Automatic Meter Reading (AMR) is a tool that can be used to record electrical energy consumption data in medium-voltage and high-voltage electricity groups [3] and Energy Meter using GSM in low-voltage electricity groups [4]. The results of AMR recording in the electrical system have the potential to produce Electrical Load Profiles (ELPs) records. ELPs are real-time data on electrical energy usage that can be recorded from electricity meter data installed in electricity customer buildings every 10, 15, or 30 minutes [5]. Research on ELPs is a crucial focus required by utility companies for formulating strategic steps in running various business processes, including price and tariff planning, distribution network operation planning, electricity production planning, load management, customer service, and public authorities. Additionally, it can be employed to identify energy consumption patterns, enabling the forecasting of future energy demand, designing energy efficiency programs, and planning power grid capacity [6], [7] . To achieve these goals, clustering methods can be employed as effective tools for identifying patterns and trends within large and complex ELPs data. This helps in the analysis and understanding of load profiles [8] . The analysis of ELPs also presents numerous potentially beneficial opportunities across various aspects of energy consumption systems [9], [10] including electricity usage behavior, electrical energy pricing, and forecasting energy demand management for the future [11].

In the context of machine learning, clustering is often utilized in exploratory data analysis, where the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

process involves learning structures and patterns existing in data without prior class labels. Therefore, clustering is typically categorized as an 'unsupervised learning' method in machine learning [8], [12], [13]. Applications of clustering methods in machine learning encompass various fields, including customer segmentation, document grouping, anomalous detection, gene or protein grouping, image grouping, superconducting clustering, recommendation systems, big data analysis [14] and even specifically in the analysis of energy load profiles [6], [7], [15].

Considering the theme of load profile analysis, this work focuses on observing an enhanced clustering approach. For instance, the FCM improvement performed by Mingyang (2020) modifies the objective function by incorporating cluster volumes to overcome the effects of distribution imbalances in the data [16]. In contrast, Qaiyum et al. (2019) utilized Ant Colony Optimization (ACO) and Maximum Residual Sampling (MRS) techniques to accelerate time, reduce space, and address time complexity issues in big data dimensioning problems related to clustering algorithm [17] s. The subsequent challenge was to define clustering methods for complex load profile data analysis and determine appropriate techniques to enhance clustering methods in terms of efficiency and accuracy.

A compilation of findings from pertinent previous studies can aid in identifying current research trends. This identification process can be accomplished through a Systematic Literature Review (SLR), a tool that encourages researchers to investigate their research subject using a broad search strategy, predetermined search terms, and straightforward inclusion and exclusion criteria. Theoretically, employing SLR increases the likelihood of obtaining clearer and more objective research answers [18].

The purpose of this study is to conduct a Systematic Literature Review (SLR) on the subject of Energy Load Profile research, evaluate various clustering methods concerning different problems, and identify potential gaps in the existing literature between 2017 and 2023 that focus on the analysis of Electrical Load Profiles (ELPs) using clustering algorithms within a machine learning framework. For this purpose, the research database utilized is a Scopus-indexed database with Quartile 1 to Quartile 4, which is considered an important database of papers with reviewed publications. Scopus offers a comprehensive overview of global research in diverse and impactful subject areas in the scientific journals within the academic community [19].

The need for this research is to examine and evaluate various clustering algorithms that have been applied in Load Profile analysis so that the performance of each clustering algorithm that has been used can be easily understood and identify potential gaps that can be filled in this literature such as challenges in improving the performance of clustering methods. in the analysis of ELPs. Our contribution in this research primarily addresses the latest developments in the research field, clustering techniques, and the focus of studies for each application. To achieve this contribution, we modified our review procedure to emphasize the engineering details of each selected study rather than the results of each paper. Considering the objectives and constraints of its application, this study will significantly assist researchers in identifying the most prominent domains of ELPs analysis, as well as the most popular clustering methods.

2. RESEARCH METHODOLOGY

The method employed in this study is Systematic Literature Review (SLR). SLR is a technique for managing information sources related to a predetermined topic [20]. In this study, SLR was utilized to assess various clustering methods, their associated challenges, and to identify potential gaps in Load Profiling Analysis. Relevant research was searched using the term "Load Profile," present in the title, abstract, and keywords of articles, with the analysis method being "Clustering".

The SLR method employed in this study follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. PRISMA is utilized to select reporting elements for systematic reviews and meta-analyses, providing a tool to evaluate the reliability of publications for systematic or literature reviews [20]. The following steps outline the preparation of checklist items for systematic review and meta-analysis:

1. Title: Identifies systematic reviews and meta-analyses.
2. Structured abstract, comprising several parts, namely introduction, materials and methods, results, and conclusions.
3. Introduction: Addresses the urgency of systematic review, outlines objectives, and introduces the meta-analysis of the systematic review
4. The literature search method is executed by exploring literature portal sources based on research questions in the scientific article database. This involves eliminating data

Journal of Theoretical and Applied Information Technology
31st May 2024. Vol.102. No. 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

duplication, applying inclusion and exclusion criteria, and selecting appropriate papers.
5. The results are presented with an overview diagram illustrating the article selection process.
6. The discussion section outlines the limitations and gaps in previous studies.
7. The conclusion summarizes the findings and provides recommendations for future research.

The flow of PRISMA implemented in this study can be observed in Figure 1:

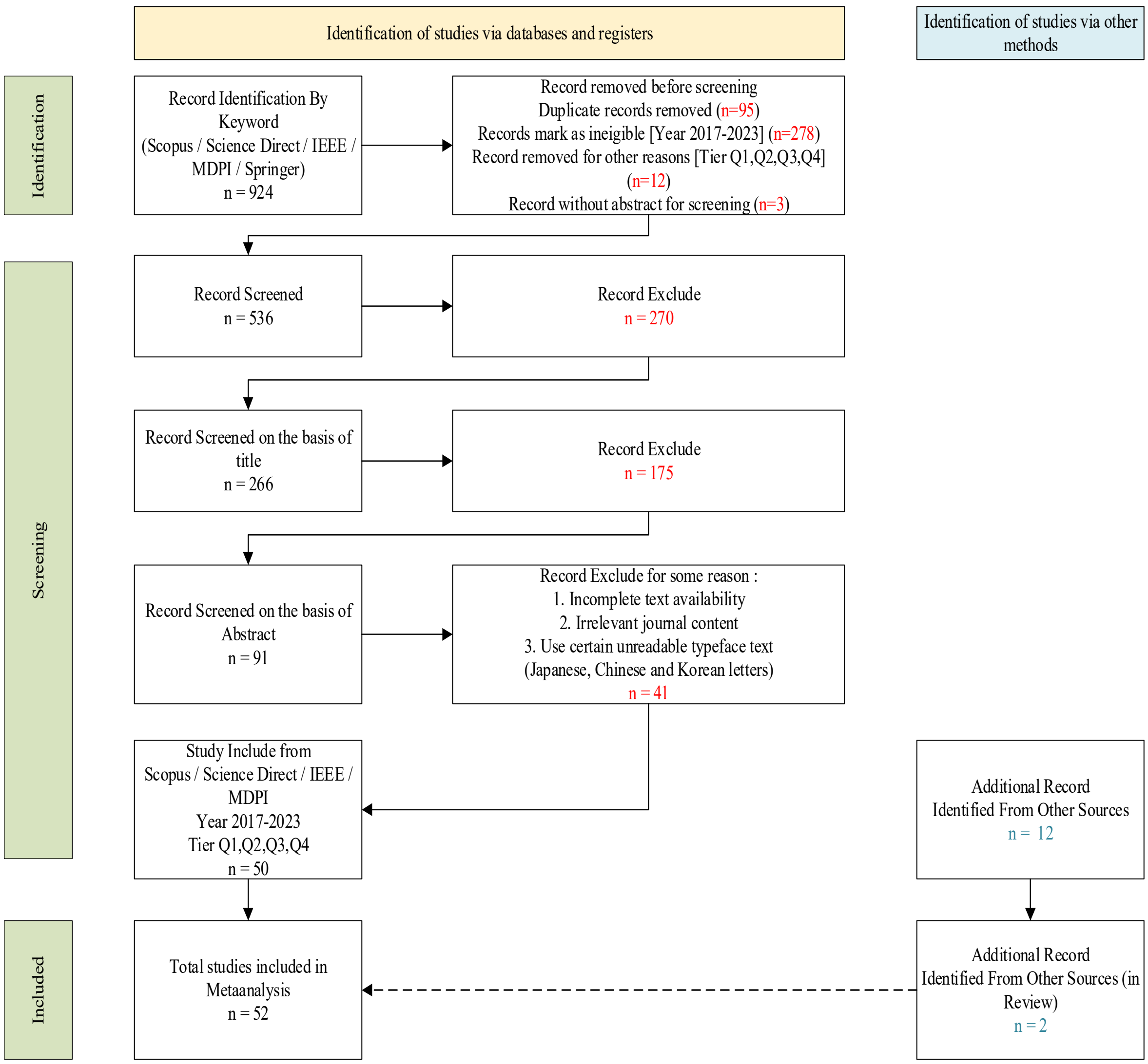

Figure 1 Process of Literature Review with PRISMA

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.1 Literature Review

The parameters used in paper selection are explained based on the exclusion criteria employed for document screening. The objectives related to the analytical components of this Systematic Literature Review (SLR) are carried out through the stages of identification, screening (excluded), and inclusion (included). SLR is conducted by systematically discovering, critically analyzing, and broadly interpreting relevant and applicable research findings, as shown in Table 1 and Fig. 1 illustrates the application of the first filter in the paper selection process on a chosen database, resulting in a total of 52 papers

Table 1 Observation of Literature Review

Observation	Literature Review
Research Question (RQ)	1. What clustering methods are often used in the analysis of ELPs? 2. What features are needed to improve the performance of clustering methods? 3. What evaluation methods are often used to assess the performance of clustering techniques?
Literature Selection	1. Journal Publication, Review Paper, Original Papers 2. The Publication Period 2017 – 2023 3. Potentially answer research questions 4. The Scopus database contains indexed publications, including information such as title, affiliation, year published, source, abstract, and quartile rank (Q1-Q4 and Non-Q). 5. The focus of the literature is on clustering methods for load profiling analysis based on machine learning trends in the field of clustering. 6. Publications are written in alphabet and in English
Literature Source	Scopus, Science Direct, IEEE Xplore, Multidisciplinary Digital Publishing Institute (MDPI), Springer
Keywords	Load profile AND (Clustering analysis OR Clustering Method OR Data clustering OR Electricity load OR Pattern recognition OR Unsupervised learning OR Supervised learning OR Data mining OR Machine learning OR Feature extraction OR Data preprocessing OR Validation Clustering OR Optimal cluster number OR Time series analysis OR K-Means OR Fuzzy C-Mean OR Fuzzy Subtractive Clustering OR Self Organizing Map OR Hierarchical Clustering OR eXplainable Artificial Intelligence OR Artificial Intelligence)

Table I above explains the research questions, literature selection, literature sources, and keywords used, encompassing the phases involved in conducting the Systematic Literature Review (SLR) as outlined in the PRISMA stages in Figure 1

2.1.1 Identification

Based on the search keywords used in the literature sources employed for this SLR (Scopus, Science Direct, IEEE Xplore, Multidisciplinary Digital Publishing Institute - MDPI, and Springer), a total of 924 papers were initially identified. Before the screening process, 95 duplicate papers were removed. Additionally, 278 papers were deemed ineligible as they fell outside the timeframe of 2017-2023, 12 papers lacked a Quartile rating of 1 to Q4, and 3 papers without abstracts were excluded prior to the screening process.

2.1.2 Screening

Following the identification and examination process, a total of 536 papers were acquired. Subsequently, the screening of these papers involved a selection process based on titles related to improving clustering methods and load profile analysis, resulting in a total of 266 papers. Further screening was conducted by selecting papers based on abstracts, yielding 91 papers that met the inclusion criteria. A more thorough screening process involved careful reading of the full text of these 91 papers. Out of these, 41 papers were included in the exclusion criteria due to incomplete text, irrelevant journal content, or content in Japanese, Chinese, and Korean letters that could not be read. Consequently, 50 papers were obtained that aligned with the research objectives. However, 12 papers lacked a Quartile rating. They were carefully reviewed to extract relevant information, leading to the identification of 2 non-Q journals that could be utilized for this study.

2.1.3 Included

Following the screening process, a total of 52 articles were identified for further analysis through meta-analysis. The objective is to examine data patterns and research trends.

2.2 Research trends in Electricity Load Profile

The examination of the dataset, which encompasses 52 scientific articles published between 2017 and 2023, reveals the distribution of papers over the research timeline, as visually depicted in Figure 2:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

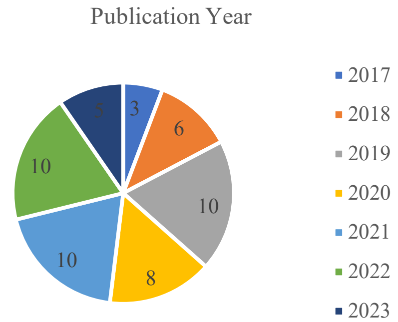

Figure 2 Literature Distribution from 2017 until 2023

Figure 3 and Figure 4 illustrate the distribution of scientific papers categorized by their reputation at the quartile level. The data shown shows that 63% of the papers included in the analysis fall into the Q1 category. In the context of quartile ranking, Q1 often includes high-quality papers that have made significant contributions to the scientific literature or have relevance in a particular study domain [21]. To clarify, it can be said that more than 50% of the articles examined are classified in the Q1 group.

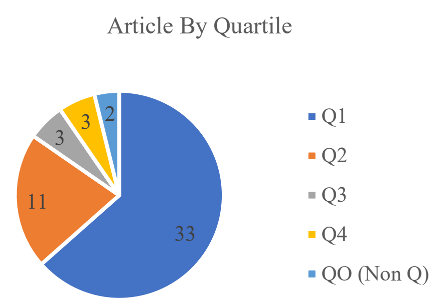

Figure 3 The Distribution of Literature Based on Quartile

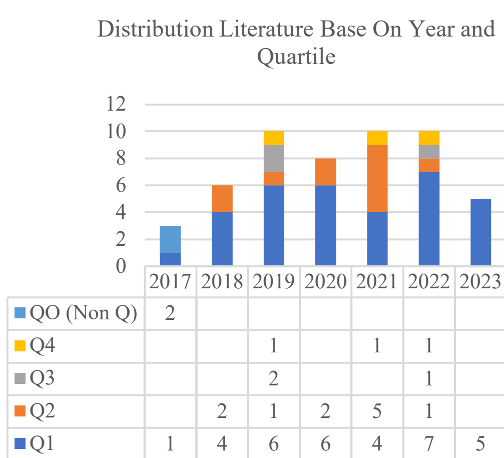

Figure 4 Distribution Literature Based On Year and Quartile

Clustering is an approach used in unsupervised machine learning models [14], [22], [23]. A literature review focusing on ELPs analysis using machine learning and artificial intelligent approaches has been conducted based on several existing papers. Table II presents a comparative analysis of surveys similar to those conducted in this study, aiming to assess the uniqueness and originality of the current research.

Table 2 Comparison With Comparable Surveys In The Existing Literature

Ref.	Research Material	The Present Study
Cembranel, 2019 [13]	This paper examines various data mining techniques used in clustering electricity customers. The main emphasis lies on the process of knowledge discovery in databases (KDD), which includes various stages including data selection, pre-processing, data mining, evaluation, and application of knowledge.	This study analyzes clustering techniques commonly used in ELPs research and looks for opportunities to improve the effectiveness of certain clustering methods in load profile analysis.
Kewo, 2023 [24]	This paper reviews residential electrical load profiles, identifying current methods for modeling such profiles. The study also addresses the advantages and disadvantages of various approaches, focusing on data characteristics, validation, and quality scores. Most of the research focuses on load profile development and load disaggregation.	This study analyzes what features are needed to improve the performance of clustering methods in ELPs analysis.
Ramoke, 2021 [25]	The paper discusses the advantages and disadvantages of AI-based models and compares them with conventional non-AI-based models to determine energy consumption patterns by time series data analysis.	This study analyzes the methods that can be used to evaluate clustering methods

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Several studies have explored the application of clustering methods with a machine learning approach on Energy Load Profiles (ELPs). However, the aforementioned study does not specifically investigate the use of ELPs data in various tariff categories, including Household, Business, Industrial, Social, and Public. Conducting research studies on ELPs using clustering with a machine learning approach within each tariff group—Household, Business, Industrial, Social, and Public—offers new opportunities for ELPs research analysis.

Identified potential research gaps in employing a machine learning approach to clustering methods for ELPs analysis include:

1. Challenges in reducing dimensions in ELPs data.
2. Challenges in determining which clustering methods can be improved for ELPs analysis.
3. Challenges in improving the performance of clustering methods in ELPs analysis.
4. Challenges in determining the optimal number of clusters in clustering analysis.
5. Challenge of interpreting improved clustering performance

The novelty of this research lies in addressing new opportunities for the advancement of clustering methods in Energy Load Profiles (ELPs) analysis, aiming for improved optimality, efficiency, and accuracy. Trends in load profiling analysis continue to evolve, driven by technological advancements, increasingly complex energy demands [24], and a growing emphasis on energy efficiency and renewable energy [26]. These aspects offer significant benefits across various contexts, particularly within the energy industry and energy resource management. Therefore, Fig. 4 provides a detailed overview of publications mentioning terms related to data analysis in the energy domain.

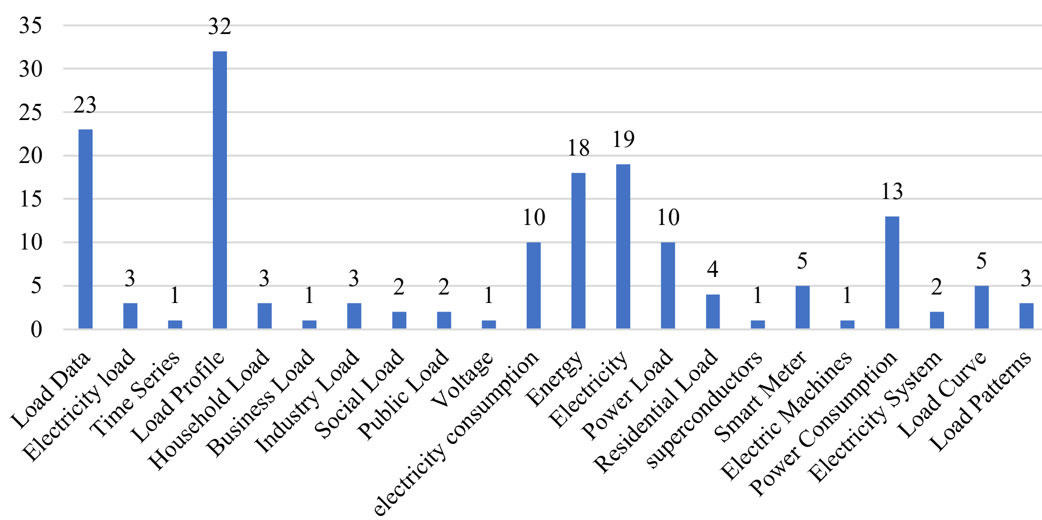

Figure 5 Number Of Articles In The Sample That Mention A Particular Research Term

Figure 5 displays the number of entities in the energy industry, as confirmed through scientific research that has been carried. The of discussion in the selected article is the load profile, which is indicated by the frequent occurrence of the term

2.3 Clustering Research In Load Profile

In this section, we provide a comprehensive review of selected articles, evaluating them based on their year of publication, keyword appearance, and relevance of reference journals in answering research questions.In the context of clustering procedures, data is arranged or partitioned based on similarities and differences. These techniques are sometimes referred to as unsupervised learning approaches [13]. Unsupervised learning techniques rely solely on information attached to data to group data sets. Table 3 below describes the clustering methods used for ELPs analysis.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3 Studies Of The Use Of Machine Learning And AI To Enhance Clustering Method In Load Profiling Analysis

Author (Citation)	Clustering Method	Application	Feature Extraction	Evaluation Method
Shi, 2020 [15]	K-Means	Load Profile	PCA	SIL, CHI, DBI
Yilmaz, 2019 [27]	K-Means	Load Profile	-	SIL
Duarte, 2022 [9]	K-Means	Load Profile	PCA	-
Choksi, 2020 [6]	K-Means, SOM	Load Profile	-	-
Zang, 2023 [28]	FCM	Consumer Power Load Data	-	SIL, DBI, CHI, IN
Bian, 2023 [29]	FCM, Spectral Clust	Power Load	PCA	-
Zhao, 2023 [30]	Spectral Clust	Power Load	PAA	-
Cen, 2022 [31]	K-Means, Hierarchical Clust, FCM	Load Profile	PCA	SIL, SSE
Jessen, 2022 [32]	K-Means	Load Profile	-	-
Kim, 2022 [11]	K-Menas	Electrical Load	MultiTag2Vec	-
Liu, 2021 [16]	FCM	Load Profile	-	-
Unal, 2021 [33]	DBSCAN	Load Profile	MDS	RMSSTD Idx
Jeong, 2021 [34]	K-Means	Load Profile	PCA	-
Valdes, 2021 [35]	Time Series Clustering	Electrical Load	-	-
Jain, 2021 [36]	K-Means, FCM	Electrical Load	PCA, FA	SIL, CHI, DI, DBI, XB
Wang, 2022 [37]	GMM	Load Profile	PCA	-
Ruhang, 2020 [38]	Agglomerative Clustering	Electrical Load	Overlapping Sliding Window	-
Zang, 2020 [39]	K-Means	Load Profile	SVD	-
Lin, 2019 [7]	Spectral Clust	Load Profile	PPA	-
Lin, 2019 [10]	Evolutionary Clust	Load Profile	-	SQ, HQ
Vahedi, 2023 [40]	Auto Clustering	Load Profile	TUBERCULOSIS	-
Binh, 2018 [41]	K-Means, SC, SOM	Electricity Consumption	-	-
Ray, 2019 [2]	Adaptive Clust	Load Profile	-	NMI
Damayanti, 2017 [5]	K-Means, FCM, KHM	Load Profile	-	DBI
Yang, 2022 [42]	K-Means	Electrical Load	LSTM-AE	SIL
Eiraudo, 2023 [43]	K-Means,	Load Profiles	-	WCSS, SIL, DBI, CHI
Senen, 2023 [44]	FCM	Load Profile	-	SIL
Jain, 2022 [45]	K-Means, FCM, Agglomerative Clustering	Load Profile	t-SNE	CVIs, CH
Flor, 2021 [46]	K-Means	Load Profile	-	SIL
Mares, 2020 [47]	Hierarchical Clust	Load Profile	-	-
Kim, 2018 [48]	MSC	Load Profile	-	ISD
Llanos, 2017 [49]	SOM	Load Profile	-	DBI

Acronyms (Alphabetical) : CHI (CalinskiN Harabasz Index); CNN (Convolutional Neural Network); CVIs (Cluster Validation Indices); DBI (Davies Bouldin Index); DBSCAN (Density Based Spatial Clustering of Applications with Noise); DI (Dunn Index); FA (Factor Analysis); FCM (Fuzzy C-Means); GMM (Gaussian Mixture Model); History Quality (HQ); Scattering Density Index (ISD) KHM (K-Harmonic Means); LSTM-AE (Long Short Term Memory AutoEncoder); MDS (Multidimensional Scaling); MultiTag2Vec (Multidimensional Tag to Vector); MSC (Mean Shift Clustering); NMI (Normalized Mutual Information), PCA (Principal Component Analysis); PAA (Piecewise Aggregate Approximation); RMSSTD Idx (Root Mean Square Standard Deviation Index); SC (Subtractive Clustering), SIL (Silhouette Index); SOM (Self-Organizing Maping); SSE (Sum of Squared Errors); SVD (Singular Value Decomposition); SQ (Snapshot Quality), TB (Temperature Based Clustering); t-SNE (t-ditribution Stochastic Neighborhood Embedding; VMD (Variational Mode Decomposition); WCSS (Within Cluster Sum of Square); XB (Xie and Beni Index).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.4 Features To Improve The Performance Of Clustering Methods

The determination of the number of clusters, data normalization, selection of distance functions between data points, initialization of cluster centers, dimension reduction, and handling of outliers are highly influential in enhancing clustering performance. Table IV below outlines the clustering methods detailed in Table III and the features utilized to improve clustering performance [12], [14], [50].

Table 4. Features That Influence The Formation And Improvement Of A Cluster

Clustering Method	Cluster Number Method	Method Distance Function	Cluster Center Initiation	Outlier detection and Handling Methods	References
K-Means	Elbow	Euclidean	Random (K-Means++)	Deletion outliers that surpass the variables' standard deviation.	Shi, 2020 [15]
	Random Based on Indicator	-	-	-	Duarte, 2022 [9]
	Gap Evaluation	Euclidean distance, Manhattan distance, and Cosine similarity	Cluster Point Average	-	Choksi K, 2020 [6]
	DBSCAN	Euclidean, Cityblock, Cosine, Chebychev,	-	DBSCAN	Roter, 2022 [14]
	Elbow & Silhouette	Euclidean	Random (K-Means++)	Deletion, Imputation, Transformation	Shi, 2021 [23]
	Random Based on Entropy	Euclidean	Cluster Point Average	-	Chowdhury, 2021 [51]
	Elbow and Silhouette	Euclidean	Cluster Point Average	DBSCAN	Nguyen, 2020 [52]
	Elbow	Euclidean	Cluster Point Average	Deletion outliers	Liang, 2020 [53]
	-	Euclidean	Cluster Point Average	Deletion outliers	Jain, 2021 [36]
	Elbow and Silhouette	Euclidean	Cluster Point Average	-	Pooya, 2021 [54]
	Elbow	Euclidean	Cluster Point Average	-	Jeong, 2021 [34]
	Elbow and Silhouette	Euclidean	Cluster Point Average	-	Kim, 2022 [11]
	Gap Evaluation	Euclidean	-	-	Xie, 2022 [22]
	Index Validity	Inter Class distance	KICIC	-	Zang, 2020 [39]
	-	Euclidean	SC	-	Binh, 2018 [41]
	Elbow and Calinski-Harabasz scores	Euclidean	Cluster Point Average	-	Cen, 2022 [31]
	Elbow	Euclidean	-	-	Jessen, 2022 [32]
	-	Euclidean	-	Deletion outliers	Yang, 2022 [42]
	Elbow and Gap Statistic	Euclidean	Cluster Point Average	-	Jain, 2022 [45]
	-	DTW	-	-	Flor, 2021 [46]

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Clustering Method	Cluster Number Method	Method Distance Function	Cluster Center Initiation	Outlier detection and Handling Methods	References
SOM	Gap Evaluation	Euclidean, Manhattan	Cluster Point Average	Median Value, Statistic Method and Deletion	Choksi, 2020 [6]
	-	Euclidean	Cluster Point Average	-	Llanos, 2017 [49]
FCM	-	Non-Euclidean distance Metrics	Cluster Point Average	Non-Euclidean distance Metrics	Hashemzadeh, 2019 [2]
	Index Validity	Euclidean	Find Primary Value	-	Sing, 2019 [55]
	-	-	Cluster Point Average	-	Jain, 2021 [36]
	Index Validity	DTW	Find Primary Value	-	Liu, 2021[16]
	Index Validity	Hyperbolic Correlation based Distance	Cluster Point Average	-	PeerJ, 2018 [56]
	Elbow &; Calinski Harabasz Score	Euclidean	Cluster Point Average	-	Cen, 2022 [31]
	Index Validity	-	-	-	Nguyen, 2022 [57]
	Elbow, Silhouette, Gap Evaluation	Euclidean, Manhattan, Cosine Similarity	Cluster Point Average	Median Value &; Deletion	Bian, 2023 [29]
	-	Euclidean	-	-	Amane, 2023 [58]
	Automatically based on data density	-	-	-	Mola, 2021 [59]
	Silhouette	Euclidean	Find Primary Value	-	Senen, 2023 [44]
	FPC	Euclidean	Cluster Point Average	-	Jain, 2022 [45]
Spectral Clust	Matrix Perturbation Theory	A multi-scale similarity metric consisting of Euclidean distance, shape fluctuation, and shape trend	Eigenvectors	-	Lin, 2019 [7]
Hierarchical Clust	Cut Debdogram	Single Linkage	There is no concept of "cluster center"	-	Roter, 2022 [14]
	The data is clustered and the two most similar clusters are combined using distance	Single linkage, complete linkage, average linkage, and Ward's method	There is no concept of "cluster center"	-	Li, 2020 [53]
	Elbow &; Calinski Harabasz	Ward Linkage	There is no concept of "cluster center"	-	Cen, 2022 [31]
DBSCAN	Does not Require a Predetermined	Euclidean	Density Based	Data points outside the main cluster & Deletion	Unal, 2021 [33]

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Clustering Method	Cluster Number Method	Method Distance Function	Cluster Center Initiation	Outlier detection and Handling Methods	References
	Number of Clusters				
	Does not Require a Predetermined Number of Clusters	Data point density	Density Based	Data points outside the main cluster & Deletion	Chen, 2021 [60]
	Does not Require a Predetermined Number of Clusters	Data point density	Has MinPts within maximum epsilon distance	-	Guan, 2019 [61]
	Does not Require a Predetermined Number of Clusters	Euclidean, Cityblock, Cosine, Chebychev	Average or Median of cluster points	Distant points from the cluster, separated, and Deletion	Roter, 2022 [14]
Time Series Clustering	-	-	-	-	Valdes, 2021 [35]
GMM	BIC	-	Data Points are Assigned to the cluster with the Highest Probability		Wang, 2022 [37]
Agglomerative Clustering	Try a certain cluster count	Cosine Similarity	Embedding Vector Average	-	Ruhang, 2020 [38]
	Dendrogram and cut it off at a certain point	Single Linkage, Complete Linkage, and Average Linkage	There is no concept of "cluster center"		Jain, 2022 [45]
Evolutionary Clust	-	-	One data point represents each potential group	-	Lin, 2019 [10]
SC	Does not Require a Predetermined Number of Clusters	-		-	Binh, 2018 [41]
	Does not Require a Predetermined Number of Clusters	Radius Value	Data point with highest density	-	Dhanalakshmi, 2019 [62]
	Does not Require a Predetermined Number of Clusters	PSO	Data point with highest density	-	False, 2019 [63]
	Does not Require a Predetermined Number of Clusters	Radius Value	Data point with highest density	-	Zeng, 2019 [64]
	-	-	Data point with highest density	-	Abdolkarimi, 2020 [65]
	Does not Require a	Radius Value	Data point with highest density	-	Mola, 2021 [59]

Journal of Theoretical and Applied Information Technology
31st May 2024. Vol.102. No. 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Clustering Method	Cluster Number Method	Method Distance Function	Cluster Center Initiation	Outlier detection and Handling Methods	References
	Predetermined Number of Clusters				
Adaptive Clust	Probabilistic	DTW	Average Load in each cluster	-	Le, 2019 [2]
KHM	Index Validity	Euclidean	Cluster Point Average	-	Damayanti, 2017 [5]
MSC	Does not Require a Predetermined Number of Clusters	SPPC	Average Load in each cluster	-	Kim, 2018 [48]

3. RESULT AND DISCUSSION

Based on a comprehensive analysis conducted on a total of 52 major studies published between 2017 and 2023, it is revealed that 50 papers are classified into Quartile 1 to 4, while the remaining 2 papers are categorized as non-Q. This review has yielded several important findings. Table III offers a comprehensive overview of clustering methods used in the analysis of Energy Load Profiles (ELPs), including evaluation methods utilized in the last six years. Table IV provides a summary of the clustering methods used, highlighting influential features that contributed to the development of these methods. Furthermore, data related to each research question is compiled and presented in Figures 5 and 6.

3.1 RQ 1 : Clustering methods for the analysis of ELPs

Based on the information presented in Table IV, it was found that the formation of the number of clusters can be categorized into two different approaches: an approach that involves a predetermined number of clusters and another approach that does not require an explicit determination of the number of clusters. Its distribution is shown in Fig. 6-7 below

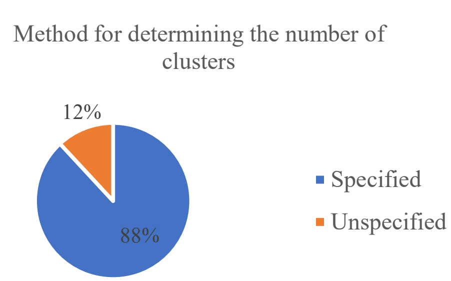

Figure 6 Method for determining the number of clusters

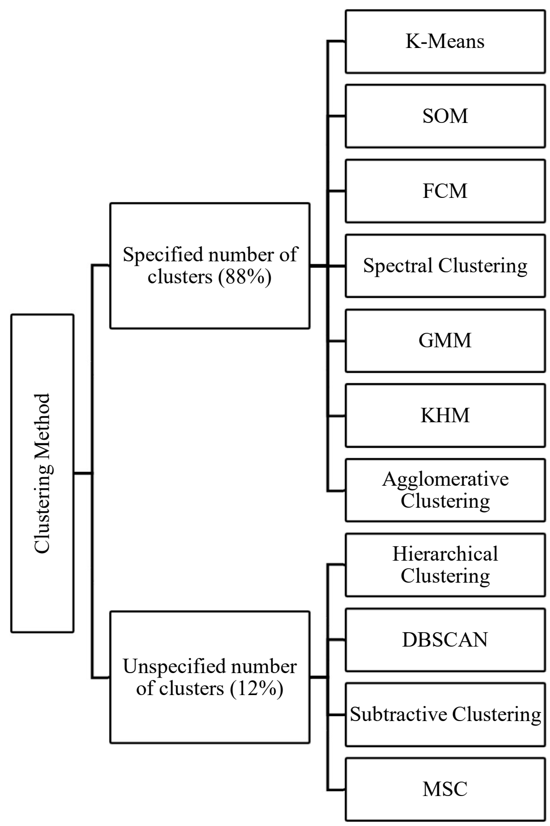

Figure 7 Distribution of clustering methods based on the method of forming the number of clusters

The primary approach (88%) to analyzing Energy Load Profiles (ELPs) involves utilizing a specified number of clusters. This approach primarily includes the application of the K-Means method (38%), followed by the FCM clustering method (19%). Spectral Clustering and SOM each account for a proportion of 7%, while Agglomerative Clustering comprises 5%. Adaptive Clustering, Evolutionary Clustering, GMM, KHM,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and Time Series Clustering each contribute with a proportion of 2%. The smaller proportion of approaches (12%) does not require the explicit determination of cluster numbers and consists of Subtractive Clustering, DBSCAN, Hierarchical Clustering, and MSC methods.

3.2 RQ2 : Features required to improve clustering method performance

As detailed in Table IV, the most influential features affecting clustering performance and quality are determined by the methods used for the following: determining the number of clusters, measuring distances between data points, initializing cluster centers, reducing dimensionality in initial data, and detecting and handling outliers [12], [14]. This review establishes that the Elbow Method is the most widely adopted approach (81%) in cases involving a predefined number of clusters. For approaches not requiring explicit determination of the number of clusters, results heavily depend on the method of determining distances between data points. The methods used include Radius Value (25%), Euclidean Distance (17%), Data point density (17%), and contributions of 8% each from Cosine, Cityblock, Chebychev, PSO

To attain optimal clustering results, a preprocessing procedure is necessary before data analysis. Data preprocessing incorporates various techniques such as feature extraction, outlier handling, dealing with missing values, and data normalization [9], [40]. The PCA method stands out as the most widely utilized technique (41%) for feature extraction in data analysis by clustering. Following this, the PAA method is utilized at 12%, while MultiTag2Vec, MDS, Overlapping Sliding Window, FA, SVD, TBC, LSTM-AE, and t-SNE each contribute 6%. Although the topic of outlier detection is not extensively discussed in most papers, 25% of the reviewed papers providing information about outliers suggest identifying outliers by considering data points outside the main cluster and subsequently removing them [14], [33], [60].

3.3 RQ 3: Evaluation methods used to assess clustering performance

The use of evaluation methods is crucial for assessing clustering performance, providing an objective means to determine the quality of the obtained cluster results [66], [67], [68]. Based on the conducted review, the most frequently employed cluster evaluation method for load profiling analysis is the Silhouette Index (29%), followed by the Davies Bouldin Index method (19%), Calinski Harabasz Index (16%), Dunn Index (6%), and other indices such as Cluster Validation Indices, History Quality, Scattering Density Index, Normalized Mutual Information, Root Mean Square Standard Deviation Index, Snapshot Quality, Sum of Squared Errors, Within Cluster Sum of Square, and Xie and Beni Index, each contributing 3%, as illustrated in Figure 8.

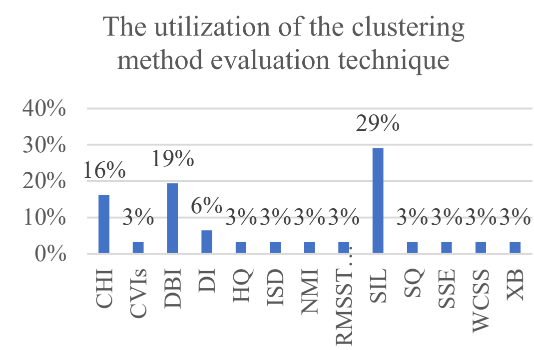

Figure 8 Distribution Clustering Evaluation Technique

4. CONCLUSION

The findings of this study contribute to the ongoing research on electrical load profiling analysis with clustering methods, building upon prior studies. A significant limitation of this review is the reliance on online repositories prioritized for literature search (Scopus, Science Direct, IEEE Xplore, Multidisciplinary Digital Publishing Institute - MDPI, Springer), with the use of additional keywords and synonyms potentially leading to further research. The implementation of Systematic Literature Review (SLR) with the PRISMA method yielded 52 articles from 2017 to 2023, discussing issues and techniques used to enhance clustering performance.

The results, as depicted in Figure 3, reveal that the most involved and influential journals in this study are those with the highest influence, primarily Q1 at 63%, with the highest distribution in 2022, constituting 21% (7 of the 33 Q1 journals were published in in 2022), as illustrated in Fig. 4. The distribution of clustering methods, based on the method of forming the largest number of clusters in Energy Load Profiles (ELPs) objects, falls into the "Specified number of clusters" group, with the K-Means method being the most frequently used, evenly distributed in the "Unspecified number of clusters" group.

To enhance the performance of a clustering method, specific techniques are required to produce optimal results. The most widely used method for the "Specified number of clusters" group is the Elbow method, accounting for 81%, while in the "Unspecified number of clusters" group, the most frequently used method involves using radius values between data points. The Silhouette Index is a commonly utilized method for evaluating clustering performance, observed in 29% of the journals analyzed.

ACKNOWLEDGMENT

The authors would like to express a sincere gratitude for the support received during this research, made possible by the Institut Teknologi PLN Ph.D. Scholarship (Agreement No. 0046.PJK/3/A0/2020) and Universiti Teknikal Malaysia Melaka (UTeM), Malaysia.

REFERENCES:

[1] M. Henderson, "Power and Energy," *IEEE Power and Energy Magazine*, vol. 16, no. 3, pp. 1–1, May 2018. doi: 10.1109/MPE.2017.2788977.
[2] G. Le Ray and P. Pinson, "Online adaptive clustering algorithm for load profiling," *Sustainable Energy, Grids and Networks*, vol. 17, Mar. 2019, doi: 10.1016/j.segan.2018.100181.
[3] A. M. Sunaya, I. N. K. Wardana, and I. N. Sukarma, *A Web-based Automatic Meter Reading for Electric Power Monitoring*. 2017.
[4] G. Peter and S. Bin Iderus, "Design of enhanced energy meter using GSM prepaid system and protective relays," in *Materials Today: Proceedings*, Elsevier Ltd, 2020, pp. 582–589. doi: 10.1016/j.matpr.2020.08.471.
[5] R. Damayanti, A. G. Abdullah, W. Purnama, and A. B. D. Nandiyanto, "Electrical load profile analysis using clustering techniques," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Mar. 2017. doi: 10.1088/1757-899X/180/1/012081.
[6] K. A. Choksi, S. Jain, and N. M. Pindoriya, "Feature based clustering technique for investigation of domestic load profiles and probabilistic variation assessment: Smart meter dataset," *Sustainable Energy, Grids and Networks*, vol. 22, Jun. 2020, doi: 10.1016/j.segan.2020.100346.
[7] S. Lin, F. Li, E. Tian, Y. Fu, and D. Li, "Clustering load profiles for demand response applications," *IEEE Trans Smart Grid*, vol. 10, no. 2, pp. 1599–1607, Mar. 2019, doi: 10.1109/TSG.2017.2773573.
[8] A. I. Károly, R. Fullér, and P. Galambos, "Unsupervised Clustering for Deep Learning: A tutorial survey," 2018.
[9] O. G. Duarte, J. A. Rosero, and M. del C. Pegalajar, "Data Preparation and Visualization of Electricity Consumption for Load Profiling," *Energies (Basel)*, vol. 15, no. 20, Oct. 2022, doi: 10.3390/en15207557.
[10] R. Lin, Z. Ye, and Y. Zhao, "OPEC: Daily load data analysis based on optimized evolutionary clustering," *Energies (Basel)*, vol. 12, no. 14, 2019, doi: 10.3390/en12142668.
[11] N. Kim, H. Park, J. Lee, and J. K. Choi, "Short-Term Electrical Load Forecasting with Multidimensional Feature Extraction," *IEEE Trans Smart Grid*, vol. 13, no. 4, pp. 2999–3013, Jul. 2022, doi: 10.1109/TSG.2022.3158387.
[12] M. Hashemzadeh, A. Golzari Oskouei, and N. Farajzadeh, "New fuzzy C-means clustering method based on feature-weight and cluster-weight learning," *Applied Soft Computing Journal*, vol. 78, pp. 324–345, May 2019, doi: 10.1016/j.asoc.2019.02.038.
[13] S. S. Cembranel, "A Short Review on Data Mining Techniques for Electricity Customers Characterization," pp. 194–199, 2019.
[14] B. Roter, N. Ninkovic, and S. V. Dordevic, "Clustering superconductors using unsupervised machine learning," *Physica C: Superconductivity and its Applications*, vol. 598, Jul. 2022, doi: 10.1016/j.physc.2022.1354078.
[15] Y. Shi, T. Yu, Q. Liu, H. Zhu, F. Li, and Y. Wu, "An Approach of Electrical Load Profile Analysis Based on Time Series Data Mining," *IEEE Access*, vol. 8, pp. 209915–209925, 2020, doi: 10.1109/ACCESS.2020.3019698.
[16] 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, doi: 10.1109/ACCESS.2021.3063123.
[17] S. Qaiyum, I. Aziz, J. Jaafar, A. Kai, and L. Wong, "Ant Colony Optimization of Interval Type-2 Fuzzy C-Means with Subtractive Clustering and Multi-Round

Sampling for Large Data," 2019. [Online]. Available: www.ijacsa.thesai.org

[18] F. M. Assef, M. T. A. Steiner, and E. P. de Lima, "A review of clustering techniques for waste management," *Heliyon*, vol. 8, no. 1. Elsevier Ltd, Jan. 01, 2022. doi: 10.1016/j.heliyon.2022.e08784.

[19] "Content Coverage Guide," 2017.

[20] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *The BMJ*, vol. 372. BMJ Publishing Group, Mar. 29, 2021. doi: 10.1136/bmj.n71.

[21] H. I. Okagbue, S. A. Bishop, P. I. Adamu, A. A. Opanuga, and E. C. M. Obasi, "Analysis of percentiles of computer science, theory and methods journals: Citescore versus impact factor," *DESIDOC Journal of Library and Information Technology*, vol. 40, no. 1, pp. 359–365, 2020, doi: 10.14429/djlit.40.1.14866.

[22] S. Xie, A. T. Lawniczak, and C. Gan, "Optimal number of clusters in explainable data analysis of agent-based simulation experiments," *J Comput Sci*, vol. 62, Jul. 2022, doi: 10.1016/j.jocs.2022.101685.

[23] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, and J. Liu, "A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm," *EURASIP J Wirel Commun Netw*, vol. 2021, no. 1, Dec. 2021, doi: 10.1186/s13638-021-01910-w.

[24] A. Kewo, P. D. K. Manembu, and P. S. Nielsen, "A Rigorous Standalone Literature Review of Residential Electricity Load Profiles," *Energies*, vol. 16, no. 10. MDPI, May 01, 2023. doi: 10.3390/en16104072.

[25] A. Ramokone, O. Popoola, A. Awelewa, and A. Temitope, "A review on behavioural propensity for building load and energy profile development – Model inadequacy and improved approach," *Sustainable Energy Technologies and Assessments*, vol. 45, Jun. 2021, doi: 10.1016/j.seta.2021.101235.

[26] E. Proedrou, "A Comprehensive Review of Residential Electricity Load Profile Models," *IEEE Access*, vol. 9. Institute of Electrical and Electronics Engineers Inc., pp. 12114–12133, 2021. doi: 10.1109/ACCESS.2021.3050074.

[27] 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.

[28] Y. Zhang, X. Li, S. Jiang, M. L. Tseng, L. Wang, and S. Fan, "Dynamic conditional score model-based weighted incremental fuzzy clustering of consumer power load data," *Appl Soft Comput*, vol. 143, Aug. 2023, doi: 10.1016/j.asoc.2023.110395.

[29] S. Bian, Z. Wang, W. Song, and X. Zhou, "Feature extraction and classification of time-varying power load characteristics based on PCANet and CNN+Bi-LSTM algorithms," *Electric Power Systems Research*, vol. 217, Apr. 2023, doi: 10.1016/j.epsr.2023.109149.

[30] J. Zhao, B. Xie, D. Wang, L. Yan, and W. Liu, "Feature extraction and source–load collaborative analysis method for distribution network," *Energy Reports*, vol. 9, pp. 547–555, Sep. 2023, doi: 10.1016/j.egyr.2023.04.329.

[31] S. Cen, J. H. Yoo, and C. G. Lim, "Electricity Pattern Analysis by Clustering Domestic Load Profiles Using Discrete Wavelet Transform," *Energies (Basel)*, vol. 15, no. 4, Feb. 2022, doi: 10.3390/en15041350.

[32] 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 Informatics*, vol. 5, Dec. 2022, doi: 10.1186/s42162-022-00237-0.

[33] F. Ünal, A. Almalaq, and S. Ekici, "A novel load forecasting approach based on smart meter data using advance preprocessing and hybrid deep learning," *Applied Sciences (Switzerland)*, vol. 11, no. 6, Mar. 2021, doi: 10.3390/app11062742.

[34] H. C. Jeong, M. Jang, T. Kim, and S. K. Joo, "Clustering of load profiles of residential customers using extreme points and demographic characteristics," *Electronics (Switzerland)*, vol. 10, no. 3, pp. 1–10, Feb. 2021, doi: 10.3390/electronics10030290.

[35] J. Valdes and L. R. Camargo, "Synthetic hourly electricity load data for the paper and food industries," 2021, doi: 10.17632/ttx9chkdcg.1.

[36] M. Jain, T. Alskaif, and S. Dev, "Validating Clustering Frameworks for Electric Load Demand Profiles," *IEEE Trans Industr*

Inform, 2021, doi: 10.1109/TII.2021.3061470.

[37] H. Wang, Y. Tian, A. Li, J. Wu, and G. Sun, "Resident user load classification method based on improved Gaussian mixture model clustering," *MATEC Web of Conferences*, vol. 355, p. 02024, 2022, doi: 10.1051/matecconf/202235502024.

[38] X. Ruhang, "Efficient clustering for aggregate loads: An unsupervised pretraining based method," *Energy*, vol. 210, Nov. 2020, doi: 10.1016/j.energy.2020.118617.

[39] Y. Zhang, J. Zhang, G. Yao, X. Xu, and K. Wei, "Method for clustering daily load curve based on SVD-KICIC," *Energies (Basel)*, vol. 13, no. 17, Sep. 2020, doi: 10.3390/en13174476.

[40] S. Vahedi and L. Zhao, "Distributed Auto-Clustering for Residential Load Profiling Using AMI Data from The U.S. High Plains," *IEEE Trans Smart Grid*, 2023, doi: 10.1109/TSG.2023.3253824.

[41] P. T. T. Binh, T. N. Le, and N. P. Xuan, "Advanced SOM & K Mean Method for Load Curve Clustering," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 6, p. 4829, Dec. 2018, doi: 10.11591/ijece.v8i6.pp4829-4835.

[42] W. Yang, X. Li, C. Chen, and J. Hong, "Characterizing residential load patterns on multi-time scales utilizing LSTM autoencoder and electricity consumption data," *Sustain Cities Soc*, vol. 84, Sep. 2022, doi: 10.1016/j.scs.2022.104007.

[43] S. Eiraudo *et al.*, "A Machine Learning Based Methodology for Load Profiles Clustering and Non-Residential Buildings Benchmarking," in *IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 2963–2973. doi: 10.1109/EEEIC/ICPSEurope51590.2021.9584633.

[44] A. Senen, T. W. O. Putri, J. J. Jamian, E. Supriyanto, and D. Anggaini, "Fuzzy C-means clustering based on micro-spatial analysis for electricity load profile characterization," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 1, pp. 33–45, Apr. 2023, doi: 10.11591/ijeecs.v30.i1.pp33-45.

[45] M. Jain, M. Jain, T. AlSkaif, and S. Dev, "Which internal validation indices to use while clustering electric load demand profiles?," *Sustainable Energy, Grids and Networks*, vol. 32, Dec. 2022, doi: 10.1016/j.segan.2022.100849.

[46] M. Flor, S. Herraiz, and I. Contreras, "Definition of residential power load profiles clusters using machine learning and spatial analysis," *Energies (Basel)*, vol. 14, no. 20, Oct. 2021, doi: 10.3390/en14206565.

[47] J. J. Mares, L. Navarro, M. C. G. Quintero, and M. Pardo, "A methodology for energy load profile forecasting based on intelligent clustering and smoothing techniques," *Energies (Basel)*, vol. 13, no. 15, Aug. 2020, doi: 10.3390/en13164040.

[48] N. Kim, S. Park, J. Lee, and J. K. Choi, "Load profile extraction by mean-shift clustering with sample pearson correlation coefficient distance," *Energies (Basel)*, vol. 11, no. 9, Sep. 2018, doi: 10.3390/en11092397.

[49] J. Llanos *et al.*, "Load estimation for microgrid planning based on a self-organizing map methodology," *Appl Soft Comput*, vol. 53, pp. 323–335, Apr. 2017, doi: 10.1016/j.asoc.2016.12.054.

[50] A. Mohammed Jabbar, K. Ruhana Ku-Mahamud, and R. Sagban, "Improved Self-Adaptive ACS Algorithm to Determine the Optimal Number of Clusters," vol. 11, no. 3, 2021.

[51] K. Chowdhury, D. Chaudhuri, and A. K. Pal, "An entropy-based initialization method of K-means clustering on the optimal number of clusters," *Neural Comput Appl*, vol. 33, no. 12, pp. 6965–6982, Jun. 2021, doi: 10.1007/s00521-020-05471-9.

[52] H. T. Nguyen, E. H. Lee, C. H. Bae, and S. Lee, "Multiple object detection based on clustering and deep learning methods," *Sensors (Switzerland)*, vol. 20, no. 16, pp. 1–14, Aug. 2020, doi: 10.3390/s20164424.

[53] X. Li, W. Liang, X. Zhang, S. Qing, and P. C. Chang, "A cluster validity evaluation method for dynamically determining the near-optimal number of clusters," *Soft comput*, vol. 24, no. 12, pp. 9227–9241, Jun. 2020, doi: 10.1007/s00500-019-04449-7.

[54] P. Tavallali, P. Tavallali, and M. Singhal, "K-means tree: an optimal clustering tree for unsupervised learning," *Journal of Supercomputing*, vol. 77, no. 5, pp. 5239–5266, May 2021, doi: 10.1007/s11227-020-03436-2.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[55] R. Xing and C. Li, "Fuzzy c-means algorithm automatically determining optimal number of clusters," *Computers, Materials and Continua*, vol. 60, no. 2, pp. 767–780, 2019, doi: 10.32604/cmc.2019.04500.

[56] M. L. Seghier, "Clustering of fMRI data: The elusive optimal number of clusters," *PeerJ*, vol. 2018, no. 10, 2018, doi: 10.7717/peerj.5416.

[57] S. D. Nguyen, V. S. T. Nguyen, and N. T. Pham, "Determination of the Optimal Number of Clusters: A Fuzzy-Set Based Method," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 9, pp. 3514–3526, Sep. 2022, doi: 10.1109/TFUZZ.2021.3118113.

[58] M. Amane, K. Aissaoui, and M. Berrada, "Enhancing Learning Object Analysis through Fuzzy C-Means Clustering and Web Mining Methods," *Emerging Science Journal*, vol. 7, no. 3, pp. 799–807, Jun. 2023, doi: 10.28991/ESJ-2023-07-03-010.

[59] M. Mola and R. Amiri-Ahouee, "ANFIS model based on fuzzy C-mean, grid partitioning and subtractive clustering to detection of stator winding inter-turn fault for PM synchronous motor," *International Transactions on Electrical Energy Systems*, vol. 31, no. 3, Mar. 2021, doi: 10.1002/2050-7038.12770.

[60] F. Chen, T. Zhang, and R. Liu, "An Active Learning Method Based on Variational Autoencoder and DBSCAN Clustering," *Comput Intell Neurosci*, vol. 2021, 2021, doi: 10.1155/2021/9952596.

[61] C. Guan, K. K. F. Yuen, and F. Coenen, "Particle swarm Optimized Density-based Clustering and Classification: Supervised and unsupervised learning approaches," *Swarm Evol Comput*, vol. 44, pp. 876–896, Feb. 2019, doi: 10.1016/j.swevo.2018.09.008.

[62] R. Dhanalakshmi and R. Vinodha, "Performance evaluation of fuzzy subtractive clustering scheduler in real-time spherical tank process," *J Comput Theor Nanosci*, vol. 16, no. 4, pp. 1640–1645, Apr. 2019, doi: 10.1166/jctn.2019.8090.

[63] H. Salah, M. Nemissi, H. Seridi, and H. Akdag, "Subtractive clustering and particle swarm optimization based fuzzy classifier," *International Journal of Fuzzy System Applications*, vol. 8, no. 3, pp. 108–122, Jul. 2019, doi: 10.4018/IJFSA.2019070105.

[64] S. Zeng, S. M. Chen, and M. O. Teng, "Fuzzy forecasting based on linear combinations of independent variables, subtractive clustering algorithm and artificial bee colony algorithm," *Inf Sci (N Y)*, vol. 484, pp. 350–366, May 2019, doi: 10.1016/j.ins.2019.01.071.

[65] E. S. Abdolkarimi and M. R. Mosavi, "Wavelet-adaptive neural subtractive clustering fuzzy inference system to enhance low-cost and high-speed INS/GPS navigation system," *GPS Solutions*, vol. 24, no. 2, Apr. 2020, doi: 10.1007/s10291-020-0951-y.

[66] A. Agardi and L. Kovacs, "Clustering algorithms with prediction of the optimal number of clusters," *Journal of Applied Research and Technology*, vol. 2, no. 6, pp. 638–651, Aug. 2022, doi: 10.1016/j.jart.2017.02.005.

[67] J. Rossbroich, J. Durieux, and T. F. Wilderjans, "Model Selection Strategies for Determining the Optimal Number of Overlapping Clusters in Additive Overlapping Partitional Clustering," *J Classif*, vol. 39, no. 2, pp. 264–301, Jul. 2022, doi: 10.1007/s00357-021-09409-1.

[68] A. Taufik, S. S. S. Ahmad, and N. F. E. Khairuddin, "Classification of Landsat 8 Satellite Data using Fuzzy c-means," in *Proceedings of the 2017 International Conference on Machine Learning and Soft Computing*, New York, NY, USA: ACM, Jan. 2017, pp. 58–62. doi: 10.1145/3036290.3036330.