Diyala Journal of Engineering Sciences

Journal homepage: https://djes.info/index.php/djes

ISSN: 1999-8716 (Print); 2616-6909 (Online)

A Data-Driven Approach for Demand Side Flexibility: Reducing Peak Demand and CO_2 Emissions in Smart Grid

Seema Pal[1,2*], Kumar Shantanu[1], Niraj Kumar Choudhary[1], Nitin Singh[1]

[1] Department of Electrical Engineering, MNNIT, Allahabad, U.P, India.
[2] Department of Electrical Engineering, JEC, Jabalpur, M.P, India.

ARTICLE INFO

Article history:
Received: 30/04/2026.
Revised: 15/08/2026.
Accepted: 17/08/2026.
Available online: 15/09/2026.

Keywords:
Demand Response
Load Shifting
Clustering Analysis,
Time-of-Use Pricing
Carbon Reduction
Smart Grid

ABSTRACT

The volatility in energy supply and the flexible demand require more efficient demand management techniques. The residential demand response has the ability to provide a promising solution. It becomes difficult to identify as households differ widely in how they utilize electricity and how flexible their usage patterns are. To address this issue, this study suggests a data-driven dual-clustering framework that utilizes both the electricity consumption behaviour and its flexibility potential for targeting residential consumers. Consumption clusters are formed using cluster validity indices that are used along with a majority voting mechanism, while the flexibility clusters are formed using a novel Demand Response Separability Score that emphasizes operational relevance over relying solely on statistical criteria. The cross-cluster analysis of these clusters reveals how consumption patterns relate to flexibility levels and facilitates the identification of most suitable households for the demand response programs. A simulation to achieve demand response is conducted by shifting loads from peak-hour to off-peak periods for selected flexible groups under a time-of-use pricing scheme. The results show that the suggested strategy achieves \$12,928.77 saving in the electricity costs and a reduction of 25,857.54 kg in CO_2 emissions over the study period. The proposed framework offers improved practical feasibility by matching DR participation with household-specific flexibility capabilities. The importance of integrating behavioral and operational characteristics in consumer segmentation provides useful information to design demand response program for the efficient, scalable, and sustainable smart grid.

1. INTRODUCTION

The stability and reliability of grids are constantly being affected by the volatility of the global energy market and the increased penetration of renewable energy sources. The recent global and geopolitical events have highlighted the fragility of energy supply chains, leading to wide variation in fuel prices and increasing concerns about energy security [1] [2]. Many countries like Japan, South Korea, Taiwan, Singapore , Italy, Spain, Belgium , India, Germany and European Union rely heavily on imported fossil fuels to produce electricity[3][4] [5],making their power systems vulnerable to disruptions of global energy markets. The effect is also extended to the cleaner options, such as LNG, with supply disruptions causing a sharp increase in the price and reduced availability. The stability and cost of making energy of the system are directly affected due to this [6].

The renewable energy, demand response, storages, flexible loads and improving energy efficiency are helping countries reduce import dependence and strengthen energy security.

Many countries are being forced to switch from gas-based to coal-based electricity to meet demand in the current scenario. However, this doesn't align with the sustainability goals as it increases greenhouse gas emissions and operational costs, but makes things more dependable in the short-term. Demand-Side Management (DSM) contributes to reduce peak electricity demand by shifting flexible electrical and thermal loads, thermal and electrical storages and coordinating energy consumption across multiple energy carriers. This reduces the reliance on coal-fired generation during gas shortages, improves the

*Corresponding author's E-mail: seema@mnnit.ac.in | ORCID: 0000-0003-3275-2832
DOI: 10.24237/djes.2026.19303

utilisation of renewable energy, lowers carbon emissions, and ensures reliable system operation. Many governments have also took actions with demand-side in response to the recent crisis, and demand-side measures issued, that governments, businesses and households can use to support energy security [7].

To reduce the reliance on fossil fuels and to enhances energy flexibility, AI based energy management system are playing crucial role [8]. Renewable energy sources are highly variable and depend on the weather, which can cause substantial variations in the supply of electricity and requires fast and selective protection [9]. This makes it harder to balance supply and demand and impacts the electricity markets [10]. In this condition, making the use of electricity more flexible and efficient has become very crucial for keeping the power system running efficiently and reliably [11] [12].

Demand response (DR) has become an important way to deal with these problems. Demand response techniques reduce the need for expensive and carbon-heavy generation sources, as well as make the system more stable by shifting the use of energy away from peak time. Demand response programs are one of the significant factors for making the grid more flexible, lowering peak demand, and adding variable renewable power [13].To ensure the reliability with integration of renewable sources, cloud based energy storages are emerging as a potential solution[14].

Similar control actions for diverse consumer populations are often applied in the classical demand response strategies. Although these methods may deliver certain profits to some of the users, they can pose substantial operational challenges for power system management. In particular, the difference between the actual and the expected reduction in the load may arise due to users with less flexible consumption patterns, as they may find it difficult to act according to demand response (DR) signals, which can lead to defeating the overall purpose of the demand response (DR) program. The variations in consumer flexibility may result in grid operational instability, as system operators depend on flexibility forecasts when scheduling generation. Forecasting flexibility is challenging because it depends on variable factors such as consumer behavior, electricity consumption patterns, and weather conditions[15]. Afterward, the anticipated demand response capacity may not actually occur during peak loads, which would negatively impact the reliability and operation of the demand-side resources and the grid's operational stability. The total effectiveness of the demand response programs is decreased by such inconsistencies [16]. Therefore, to create and categorize consumers into effective clusters based upon their consumption patterns and flexibility to enable more targeted and successful demand response strategies (DR), data-driven methods for clustering will be needed.

Aggregated consumer pools can form clusters for demand response, but there are challenges associated with those due to the mixed-use nature of residential load patterns. Since there are various types of consumption and load flexibility, accurately defining and predicting those cluster groups based on their responses to grid signals will be difficult.

A careful analysis of the individual customer's consumption behaviour, time-of-use load profiles, and level of flexibility is the key to accurately defining each of the groups. If these issues can be successfully addressed, the complete potential of cluster-based demand response will be realized in improving grid flexibility, decreasing peak load, lowering operating costs, and reducing CO_2 emissions.

Whereas past research has clustered consumers by consumption patterns alone, the dual-clustering framework combines consumption behaviors with flexibility features. With the demand response separability score (DSS) to separate demand response participants as well as identify actionable demand response participants, rather than just consumers with similar characteristics, the work connects the clustering results with the practical need to operate the electric grid.

The main contributions of this work are as follows:

- Comprehensive Feature-Based Clustering: To enable a novel classification of households based on both their load profile similarity and their flexibility potential, the work proposes a dual clustering framework, based on consumption patterns and flexibility characteristics, for the demand response mechanism.
- Advanced Cluster Validation and Selection: The classical cluster validity indices (Silhouette, Calinski-Harabasz, Davies-Bouldin) are combined with the majority-based decision to make robust consumption clusters, and a Demand Separability Score (DSS) for flexibility clusters ensures that identified clusters using K-Means are both statistically sound and operationally meaningful for DR purposes. In addition, the results are compared with other clustering algorithms.
- Cross-Analysis of Consumption and Flexibility Clusters: The proposed work identified households that contribute to peak demand and show flexibility. This helps to better target demand response efforts.
- Automated Demand Response Simulation: An approach to apply peak-to-off-peak load shifting to automatically selected flexible clusters simulates the DR program that maintains energy balance while reducing carbon emissions and cost savings.

- Quantification of Economic and Environmental Impacts: The framework calculates cost savings and CO2 emission reductions at both system and household levels, compared to the baseline, providing a holistic assessment of the benefits of residential flexibility. In addition, the results are compared with the uniform DR scenario.

The structure of the paper is as follows: Section 2 describes the literature review used in the research. Section 3 presents the methodology of the proposed work. Section 4 shows the obtained results using the methodology. Section 5 presents the discussion on the results obtained. The last section presents the conclusion and future scope.

2. LITERATURE REVIEW

The existing literature has shown the various efficient techniques of clustering algorithms like Hierarchical-based clustering algorithm, Partition-based clustering algorithm, Density-based clustering algorithm, Fuzzy theory-based clustering algorithm, Distribution-based clustering algorithm, Graph theory-based clustering algorithm, Grid-based clustering algorithm, Fractal theory-based clustering algorithm, Model-based clustering algorithm, etc [17] [18] [19] [20]. The capacity of deep learning models to identify the patterns in unlabeled data makes it more useful in consumer segmentation[21]. The metaheuristic algorithms are widely used in data mining and power system optimisation. They are used to improve clustering performance, e.g., feature selection and centroid optimisation, and also in other applications like, to find the optimal placement and sizing of distributed generators (DGs) in distribution networks[22] [23] [24] [25].

The goodness of the formed clusters is measured by the various cluster validity indices (CVIs). The main purpose of CVIs is to determine the validity of clustering algorithms. Mostly internal and external indicators are used, based on test results during the clustering procedure. Some internal indicators are the Davis-Bouldin indicator, Dunn indicator, Silhouette coefficient, Calinski–Harabasz Index, Within-Cluster Sum of Squares (WCSS), and Xie–Beni Index. The Rand indicator, the F indicator, the Jaccard indicator, and the Fowlkes-Malkows indicator are external evaluation indicators [26]. The main question arises as to how many optimal clusters should be formed to ensure the validity and correctness of the clusters [27].

As the clustering requires the fulfillment of multiple objectives, a single CVI cannot satisfy the competing criteria. Therefore, dependence on a single CVI is basically not suitable for robust cluster validation. A principled approach requires the integration of multiple CVIs and domain-specific interpretability considerations. The "majority-based decision" provides a fast solution where the final decision is made to determine the optimal clusters by the majority of the total CVIs' votes [28].

Existing demand response customer segmentation research focuses on grouping customers by consumption patterns using clustering algorithms validated through statistical quality metrics (Silhouette Index, Davies-Bouldin Index, Calinski-Harabasz Index), but doesn't provide whether customers within the same pattern cluster possess different load-shifting capabilities [29] [30] [31]. The recent methods of clustering will ensure that the final clusters formed have well-defined and compact groups that are well separated by a minimum amount of outliers but will not identify specific members of a cluster that have increased flexibility in shifting their consumption from use during peak periods to use during off-peak periods. Two customers may belong to the same "evening peak" cluster because both consume electricity primarily from 5:00 to 10:00 PM, achieving identical clustering validation scores. However, Customer A's evening consumption comes from the loads that cannot be shifted to different times, while Customer B's evening consumption comes from EV charging and water heating, etc., that can be easily shifted to overnight. Current clustering methods treat both customers identically for demand enrollment, despite Customer B having more peak reduction potential than Customer A. Literature provides no framework for distinguishing flexible from inflexible customers within validated consumption clusters. When utilities select customers for demand response programs based only on consumption patterns, they enroll many customers who cannot shift loads despite fitting the cluster perfectly. This may lead to problems for distribution system operators (DSO).

The existing research doesn't provide a systematic method for dual clustering to first create consumption pattern clusters and clusters based on flexibility, using validated clustering techniques as well as with DR-specific features. After obtaining the consumption and flexibility clusters, a cross-cluster analysis can help to identify the distribution of households with similar consumption patterns across different flexibility groups. This would allow utilities to target high-flexibility customers within each consumption pattern, rather than assuming a similar demand response potential for all the customers having a similar consumption profile. This integrated approach would bridge the gap between what clustering identifies (consumption timing patterns) and what demand response requires (load shifting capability). It will enable utilities to achieve substantially better outcomes by distinguishing flexible from inflexible members within validated consumption clusters. The systematic methodology would help identify which similar consuming customers can actually perform the load shifting that modern grid operations and renewable integration fundamentally need. To deal with the variations between the load patterns and the flexibility of the consumers and to reduce these variations, a dual cluster-based segmentation method is used for the demand response.

By classifying households into quality clusters based on the consumption patterns and flexibility potential, utilities can formulate a strategy that can make smart decisions for consumer selection for participation in DR, provide cost savings, and emissions reductions while considering consumer heterogeneity. Table 1 shows the comparative analysis of different clustering algorithms used for residential demand segmentation. Among all the studied clustering techniques, K-Means has shown a superior suitability for demand response (DR) due to its balance between computational efficiency, interpretability, and stability of cluster formation. While DBSCAN identified noise points and irregular cluster structures, it produced meaningless segmentation for flexibility analysis. Agglomerative clustering yielded comparable results but with reduced cluster separation and higher computational complexity.

The centroid-based representation in K-Means enables straightforward interpretation of cluster characteristics, which is critical for demand response decision-making. Therefore, K-Means is selected as the final clustering method for both consumption and flexibility segmentation in this study.

Table 1. Comparative Analysis of Clustering Algorithms for Residential Demand Segmentation

Algorithm	Key Strengths	Limitations	Use	Scalability
K-Means	Computationally efficient, easy to implement, and produces well-separated clusters.	Requires a predefined number of clusters (k) and is sensitive to initialization	Large-scale datasets demand response segmentation with clear cluster structures.	High
DBSCAN	Identifies arbitrarily shaped clusters and is robust to noise and outliers	Sensitive to parameter selection and struggles with varying densities	Outlier detection, datasets with noise, and irregular cluster shapes	Medium
Agglomerative Hierarchical	Provides a hierarchical structure, and no need to predefine k	Computationally expensive and less scalable for large datasets	Small to medium datasets, and exploratory analysis of cluster relationships	Low

3. METHODOLOGY

The work proposed the clustering-based demand response (DR) among residential electricity consumers. Clustering is performed on residential electrical demand data to identify distinct consumer load profiles having high consumption and high flexibility, both for the targeted DR in the first phase. In the second phase, DR simulation was carried out on an automatically selected highly flexible cluster for peak reduction, cost saving, and reduction in CO_2.

Conventional demand response programs often treat all consumers alike, assuming uniform flexibility and identical responsiveness to price signals or incentive schemes. In reality, consumers differ considerably in their usage habits, peak consumption timing, and capacity to shift loads - and these differences matter.

A uniform DR strategy risks poor participation, uneven load distribution, and missed opportunities to engage consumers who could contribute meaningfully under the right conditions. Clustering analysis addresses this by revealing distinct consumer groups, each interacting with DR signals in its own way. Some clusters respond readily to time-of-use pricing, while others, constrained by fixed routines, need different approaches altogether. Clusters that are not compact and are not properly separated from other clusters can misrepresent consumer segments and ultimately misdirect DR strategies.

To overcome this, the present study constructs clusters using multiple cluster validity indices (CVIs) that are robust to range sensitivity and resolve disagreements among them through a majority voting approach. This ensures that the resulting segments are not an artifact of the chosen metric but reflect genuine behavioral groupings in the data.

Once reliable clusters are established, the cross-tabulation of flexible and consumption clusters identifies the best cluster for DR application to reduce the load at peak hours, cost savings, and CO_2 reduction. The approach moves the analysis from what the data shows to what it actually means for DR program design. In this study, hourly electricity consumption data for 200 households over a full year (8760 hours) were analyzed to explore residential demand flexibility for demand response applications. The workflow consisted of feature extraction, clustering, consumption-flexibility cross-analysis, cross-tabulation, and DR simulation. The framework for optimal cluster formation (consumption behavior clusters and flexibility related clusters) for cross-cluster analysis from initial raw data is shown in Figure 1. The flow chart for the DR simulation based on the information from the formed clusters is shown in Figure 2.

3.1 Feature Extraction

Feature extraction in clustering or in deep learning based models, refers to techniques that identify, select, or weight features to improve the quality and interpretability of cluster assignments by mitigating the impact of irrelevant or noisy variables [32] [33]. To effectively characterize household electricity usage and demand response potential, two categories of features are extracted from the load profiles, as shown in Table 2.

- Consumption behaviour features
- Flexibility behaviour features

These features transform the raw load data into a compact and informative representation of household demand characteristics. Consumption behaviour features describe the general electricity usage patterns of households. These features summarize the statistical properties and time-of-use characteristics of electricity consumption. These features determine how electricity is typically consumed throughout the day, and help identify consumers with similar demand patterns. Flexibility features show the ability of consumers to shift or adjust their electricity demand, required for participation in demand response programs. These features provide useful information on the dynamic characteristics of electricity consumption, enabling the identification of consumers with higher potential for contributing to the demand response program.

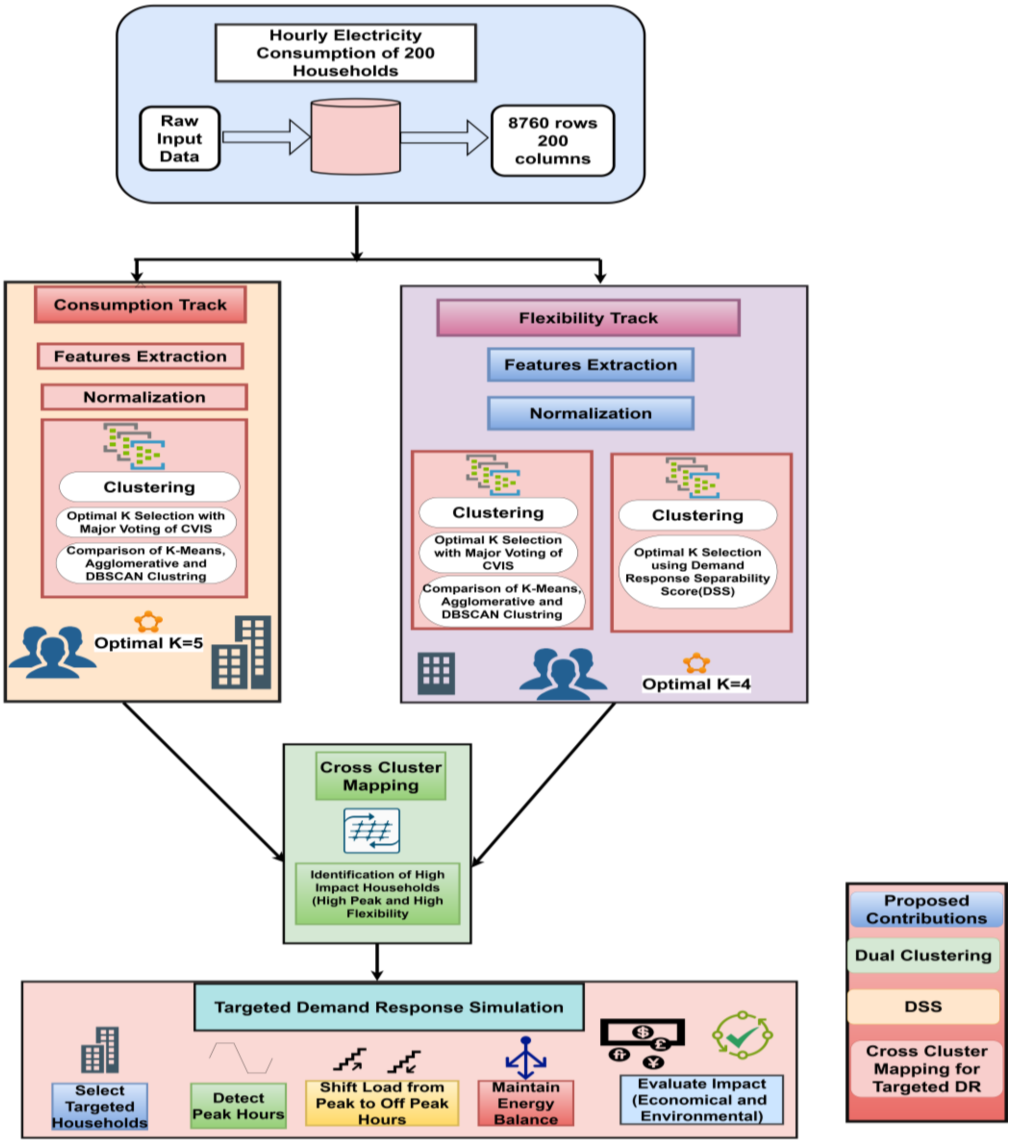

Figure 1. Proposed Dual-Clustering Framework for Identifying High-Impact Households for Targeted Demand Response

In total, 34 consumption behaviour features and 9 flexibility behaviour features are extracted from the hourly load profiles to represent household demand characteristics for clustering. Since the extracted features have different scales and units, z-score normalization is applied as shown in (1) to standardize the feature values [34].

$$x_{norm} = \frac{x-\mu}{\sigma} \quad (1)$$

where μ represents the mean and σ represents the standard deviation of the feature. Normalization ensures that all features contribute equally to the clustering process.

3.2 Clustering and Cluster Validation

Clustering groups the data based on their characteristics. In the data-rich environment of the smart grid era, data is gaining greater importance due to the availability of vast amounts from smart meters. Three clustering methods (K-Means, DBSCAN, and Agglomerative Hierarchical Clustering) are used in the proposed work for comparison of clustering performance across different clustering paradigms. K-Means clustering is a partition-based method that partitions households into *k* clusters by minimizing the within-cluster variance.

Table 2. Consumption and Flexibility Features Used for Residential Demand Characterization and Clustering Analysis

Consumption Behaviour Features	Flexibility Features
Mean Load	Ramp Rate
Standard Deviation	Load Factor
Minimum Load	Peak Ratio
Maximum Load	Curtailment Headroom
Median Load	Predictability (Lag-24)
Peak-to-Average Ratio	Mean Flexible Load
Hourly Avg. Consumption (24 features)	Peak Load
Time-of-Use Ratios (4 Features)	Off-Peak Load
	Load Variability

The DBSCAN method is a density-based clustering method that is capable of identifying clusters of arbitrary shapes and detecting noise points. Agglomerative Hierarchical Clustering is a bottom-up clustering approach that iteratively merges similar clusters [35] [36] [37]. To determine the optimal number of clusters, multiple CVIs are used, including the Silhouette score, Calinski–Harabasz index, and Davies–Bouldin index [38].

The use of domain-informed feature engineering enhances the interpretability and improve the effectiveness of the model [39]. The framework in the proposed work, combines a domain-specific separability metric with statistical clustering validation to make sure that both the mathematical robustness and the operational relevance are in good order. In the proposed work, clustering based on consumption behaviour uses elbow-based majority voting on traditional CVIs, while clustering based on flexibility uses the DR-Separability Score (DSS) metric, which clearly shows how separable demand response characteristics are.

This two-stage approach ensures that the formed clusters are both structurally sound and useful for managing grid dispatch and flexibility. To determine the optimal number of clusters (k) and mitigate monotonic drift toward the search range ceiling, an elbow-based consensus of the CVIs framework using discrete second-order differences is implemented. This method provides the point of maximum curvature across various validation indices. A Savitzky-Golay filter is used to smooth out the curves so they are less affected by noise in the time series.

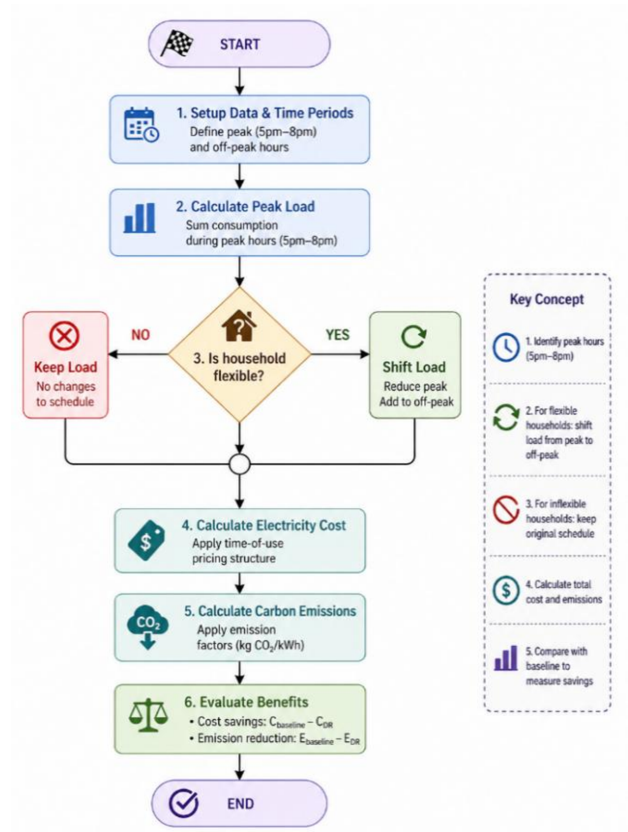

Figure 2. Decision Flow of Flexibility-Driven Demand Response for Peak Load Reduction and Cost–Emission Reduction

Traditional metrics for validating clustering mainly look at how close and far apart the clusters are. But these metrics might not show how useful clusters are for demand response applications. So, the Demand Response Separability Score (DSS) is used to assess how well clusters are separated based on their flexibility in demand response operations [40]. The DSS metric shows how different levels of demand response flexibility are represented by clusters. The four significant demand response features: peak ratio, curtailment headroom, predictability, and ramp rate are used in the proposed work.

Each feature is normalized by an approximate full-scale range to make sure they are all comparable. To find the optimal number of clusters, the DSS is assessed with increasing values of k, and its marginal gains are analyzed. First-order differences show small improvements in separability, while second-order differences show changes in these improvements. The optimal k is chosen at the point where the second-order difference is most negative, which is when DSS gains start to level off. This is when adding more clusters doesn't make a meaningful difference in separation.

Although traditional clustering validation indices like Silhouette, Calinski–Harabasz, and Davies–Bouldin are calculated, they are used as references, and they do not affect the final selection in this work. This prioritization shows a planned move away from purely statistical criteria and toward a domain-centric view. This makes sure that the resulting clusters show different and useful flexibility profiles for running the distribution system. Adding DSS to the process of evaluating clusters makes the clustering results more useful for a power system operator, making sure that the identified clusters help with good demand response management.

3.3 Consumption Flexibility Cross Analysis

Each household gets a label for a consumption cluster and a flexibility cluster. A cross-tabulation of consumption and flexibility clusters helped find the households that were most likely to cause peak demand and would be good candidates for DR interventions. After the consumption and flexibility clusters are made, a cross-cluster analysis is done to see how these two clustering results are related. This analysis helps to understand how households with similar consumption patterns are spread out among different groups of flexibility.

This analysis helps find groups of households with certain demand patterns, like those with high consumption but low flexibility or those with moderate consumption and high flexibility. This kind of information is very important for making demand response strategies that work.

Let C_i denote the i^{th} consumption cluster, and F_j denote the j^{th} flexibility cluster. The number of households belonging to each cluster combination is computed as per (2).

$$N_{i,j} = \sum_{h=1}^{H} I(C_h = i, F_h = j) \qquad (2)$$

where:

- H shows the total number of households,
- C_h is the label for the consumption cluster of household h,
- F_h is the label for the flexibility cluster of household h,
- $I(\cdot)$ is an indicator function that equals when the condition is true and 0 otherwise
- The resulting matrix $N_{i,j}$ forms a table that shows the distribution of households across consumption and flexibility clusters. The household cluster mapping step organizes the clustering results by assigning each household to its corresponding cluster group. This process enables a structured representation of the clustering output, making it easier to analyze and interpret the characteristics of each consumer segment. Cross-tabulation helps to identify which consumption groups are more flexible. Furthermore, the mapping enables utilities to target specific consumer segments when designing demand response strategies, thereby improving the practical applicability of the clustering framework. The resulting clusters are analyzed to understand the consumption patterns and flexibility characteristics of different household groups.

3.4 DR Simulation on Selected Flexible Clusters

The two flexibility clusters with the highest peak-to-total consumption ratio were automatically assigned for DR. During peak hours (17:00 to 20:00), a 20% load shift was performed, and the load that was cut was spread out evenly across off-peak hours. This method simulates a residential DR program that uses cluster-level flexibility while maintaining the energy balance. Post-DR electricity costs and carbon emissions with a clustered approach were calculated and compared with baseline and uniform DR scenarios.

4. RESULTS

4.1 Data Description

The dataset consists of residential electricity demand profiles for 200 households from 01-01-2010 , 00:00:00 to 31-12-2010, 23:50:00 at 10 min intervals. The dataframe consists of 52562 rows and 200 columns. The profiles shows total electricity use, in watts [41]. The proposed dual clustering based demand response algorithm was implemented in Python Version : 3.12.13 and executed on Google Colab using a cloud-based virtual machine equipped with an Intel(R) Xeon(R) CPU @ 2.20GHz, 2 vCPUs, and 12.67 GB RAM.The data is checked for null values and text in the columns which shows the cleaned dataset. After the hourly aggregation and unit conversion, it resulted in 8760 rows × 200 columns. After that features were included and normalized. The raw load data is organized in a matrix as per (3).

$$L = \{l_{t,i}\} \qquad (3)$$

where, $l_{t,i}$ represents the electricity consumption of household i at hour t.

4.2 Dual Cluster Formation

The proposed framework adopts a dual clustering strategy to capture both consumption behavior and demand response (DR) flexibility characteristics of residential demand of 200 houses Consumption and flexibility features were extracted from the household smart meter dataset, resulting in feature matrices of size 200×34 for consumption characteristics and 200×9 for flexibility indicators. However, because many CVI metrics tend to change monotonically as the number of clusters increases, selecting the optimal k using simple maximum or minimum values can lead to biased results, typically favouring the upper limit of the search range.

To address this, the elbow approach was used as the determination method for cluster count. The Savitzky-Golay filter is applied to smooth out the CVI score curves in order to minimize noise. Following this, the first and second differences were calculated on the scores, and the elbow point is defined as the location on this score curve at which there is the greatest rate of decrease in the amount of additional improvement in the metric, or the point along the curve with the greatest amount of curvature. Any additional clusters will provide less benefit. The elbow point is a good estimate for the maximum number of clusters.

Here, the Silhouette score, the Calinski–Harabasz (CH) index, and the Davies–Bouldin (DB) index are used to find the optimal number of clusters for consumption profiles. The final best number of clusters was found using a majority voting method, where the cluster number that got the most agreement among the three indices was chosen as the final answer. This method of voting with multiple metrics gives a better idea of the right cluster structure than just using one validation index.

Each CVI contributes one vote for its selected k. The cluster number with the highest vote count is selected. If multiple values tie, the value with the highest Silhouette score among the tied candidates is chosen. Here, k=5 is taken as the optimal number of clusters, as shown in Figure 3. This approach provides robustness by combining multiple cluster validation perspectives. As per the figure in the consumption clustering plots, the Silhouette Score attains its highest value around k=2, suggesting strong separation when the dataset is partitioned into two clusters. The Calinski–Harabasz Index reaches its peak near k=4, indicating that the ratio of between-cluster variance to within-cluster variance is most favorable at this cluster number.

For the Davies–Bouldin Index, the lowest values appear around k=2 and k=7, reflecting improved cluster compactness at those points according to its conventional interpretation. However, instead of selecting cluster numbers directly based on these extrema, an elbow-based approach derived from the elbow Method is applied to identify points where further increases in k provide diminishing improvement in clustering quality.

To obtain stable curvature estimates, the metric curves were first smoothed using the Savitzky–Golay Filter, after which first and second differences were computed to detect the elbow point for each metric [42]. For computational consistency, the Davies–Bouldin values were temporarily sign-inverted during this step so that all metrics followed a common increasing-trend interpretation.

The statistical CVIs showed that there were more clusters for flexibility clustering ($k_{Sil} = 5$, $k_{CH} = 7$, and $k_{DB} = 8$), as shown in Figure 4. However, these metrics showed a consistent trend as k increased, which can cause too much segmentation and clusters that aren't very useful for operations. To get around this problem, a Demand Response Separability Score (DSS) that takes into account the domain is used to rate how well clusters can be separated based on important demand response features like peak ratio, curtailment headroom, predictability, and ramp rate.

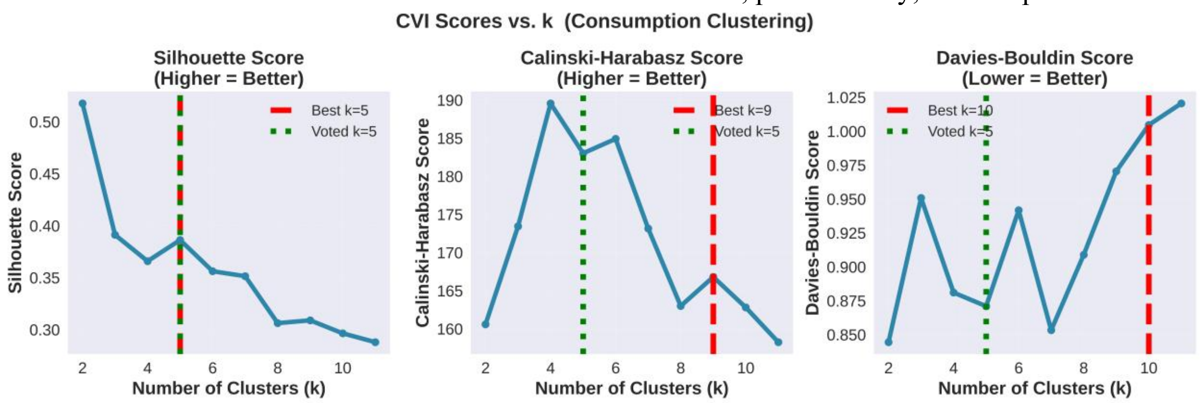

Figure 3. Consensus-Based Selection of Optimal Cluster Number (k = 5) for Consumption Clustering Using Multiple Cluster Validity Indices

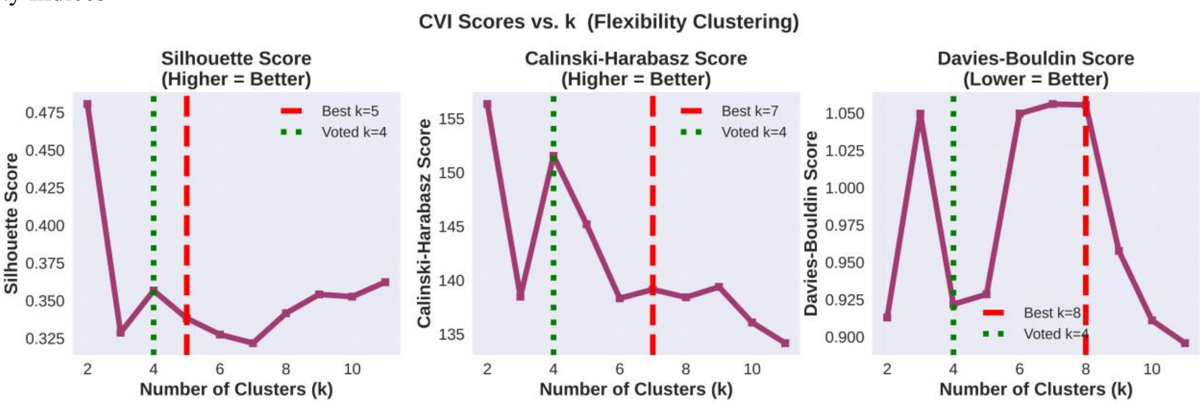

Figure 4. Consensus-Based Selection of Optimal Cluster Number for Flexibility Clustering Using Multiple Cluster Validity Indices

The main advantage of DSS is that it transforms cluster selection from a purely statistical optimisation problem into an operational decision-making framework. Instead of maximising the quality of geometric clusters, it finds the minimum number of clusters that captures all practically relevant flexibility behaviours for DSO dispatch. For comparison, traditional CVIs are still reported, but they are used as supportive statistical indicators rather than the main criterion for selecting clusters.

For each candidate number of clusters k, the Demand Response (DR) features' mean values were calculated for each cluster.

These features capture operational characteristics relevant for demand response dispatch decisions. The demand response separability score (DSS) quantifies the degree to which clusters differ along DR-relevant dimensions, as shown in (4). It is defined as:

$$DSS\ (k) = \sum_{f=1}^{4} \frac{\max(\mu_f) - \min(\mu_f)}{Range_f} \quad (4)$$

where:

- μ_f represents the mean value of feature f within each cluster,
- $\max(\mu_f)$ and $\min(\mu_f)$ denote the maximum and minimum cluster means for feature f across all clusters,
- $Range_f$ is the approximate full-scale range of feature f

The normalization by $Range_f$ ensures that each feature contributes comparably to the overall score regardless of its numerical scale.

A higher DSS value indicates stronger separation between clusters in terms of operational flexibility characteristics. The optimal number of flexibility clusters is determined by identifying the point at which improvements in DSS begin to plateau. The marginal gain in separability when increasing the number of clusters from k to $k + 1$ is shown in (5).

$$\Delta DSS_k = DSS_{k+1} - DSS_k \quad (5)$$

To detect the plateau point, the second-order difference of DSS gains is computed as per (6).

$$\Delta^2 DSS_k = \Delta DSS_{k+1} - \Delta DSS_k \quad (6)$$

The elbow of the DSS values corresponds to the largest negative value of $\Delta^2 DSS_k$, which indicates the strongest reduction in marginal improvement. This point marks the onset of diminishing returns in cluster separability. Consequently, this value of k is selected as the optimal number of clusters for flexibility analysis, as additional clusters do not produce new operationally meaningful demand response groups.

The Table 3 shows the gains started to slow down after k = 4, which meant that a separability plateau was starting to form. So, k = 4 was chosen as the best number of flexibility clusters, each of which represents a different level of demand response capability.

The practical significance of the Demand Response Separability Score (DSS) is in identifying flexibility clusters that are directly applicable to real-world demand response (DR) applications, rather than merely providing good statistical separation. DSS maximises the differences in key DR characteristics (peak ratio, curtailment headroom, predictability, and ramp rate) to ensure that each cluster represents a unique level of flexibility and response capability. This allows utilities and aggregators to accurately identify customers with high, medium or low DR potential, enabling targeted incentive schemes, optimised load control strategies and more reliable flexibility aggregation. Moreover, DSS assists the power system operators to estimate the available flexible capacity for the peak demand reduction, renewable energy integration, and ancillary services. Choosing the right number of clusters at the DSS elbow also avoids over-segmentation resulting in clusters that are interpretable and operationally relevant. Therefore, DSS provides a more reliable foundation for decision making in demand response planning and implementation than only using traditional statistical clustering validity indices. A combined visualization for optimal clusters is shown in Figure 5.

Table 3. Optimal cluster selection with DSS

Number of Clusters ((k))	DSS	Marginal Gain
2	1.968	–
3	2.848	0.879
4	4.653	1.805
5	5.332	0.679
6	5.493	0.161
7	5.843	0.350
8	5.814	−0.029
9	5.952	0.138
10	6.446	0.494
11	6.769	0.323

Three clustering algorithms were subsequently applied to both datasets: K-Means, DB-SCAN, and Agglomerative hierarchical clustering. For consumption clustering, K-Means produced five clusters with a Silhouette score of 0.386, a CH score of 183.02, and a DB score of 0.871. Agglomerative clustering provided results with lower scores. DBSCAN provided two dense clusters with five noise points, which shows a higher silhouette score, but a lower CH score. For flexibility-related clustering, both K-Means and Agglomerative clustering made four clusters, which is in line with the DSS-based selection. The Silhouette score for K-Means was 0.357, the CH score was 151.53, and the DB score was 0.922. Agglomerative clustering got results that were similar, but the groups were a little less separate. DBSCAN found two clusters and nine noise points, which means that the structure of the flexibility features is less dense. K-Means gave the most stable clustering results for both consumption and flexibility analysis, so it was chosen for the final household segmentation.

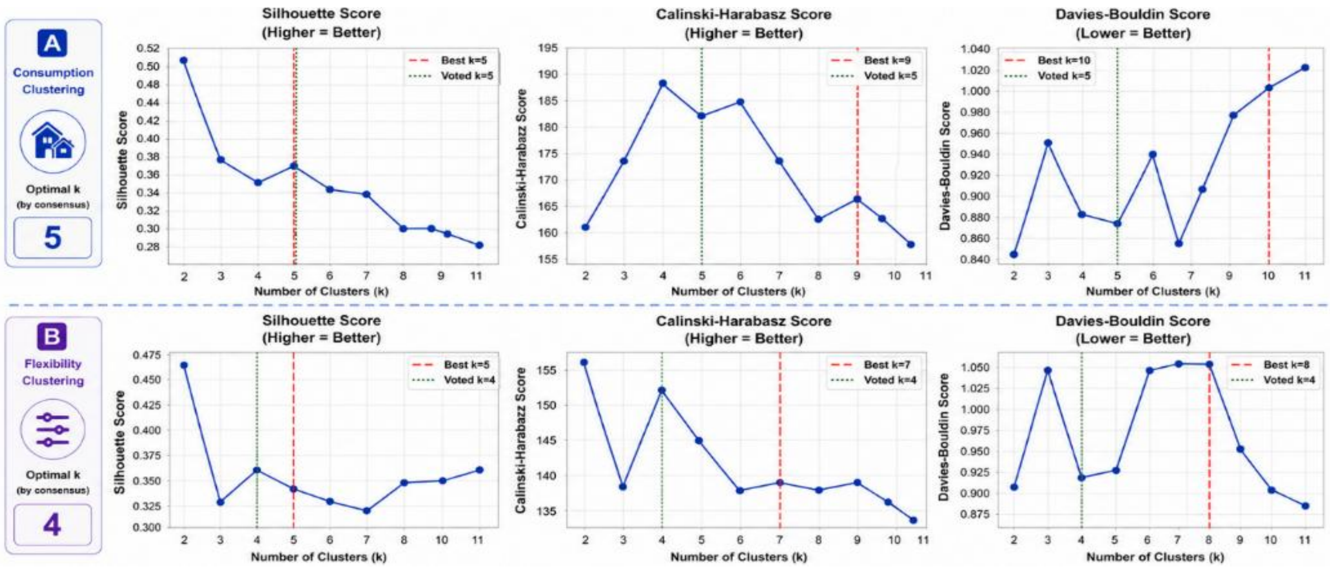

Figure 5. Elbow-based consensus selection of optimal cluster number using multiple cluster validation indices for Combined Analysis

4.3 Selection of Final Clustering Configuration

The K-Means clustering is selected for household segmentation. This was based on a comparison of clustering algorithms and validation metrics as given in section 4.2 .

K-Means consistently generated stable cluster structures for both consumption and flexibility datasets, yielding competitive Silhouette, Calinski–Harabasz, and Davies–Bouldin scores. DBSCAN, on the other hand, only found a few dense clusters and labeled some households as noise. This could make it harder to understand in large-scale demand response applications.Agglomerative clustering gave results that were similar to K-Means, but with less scores. The K-Means clustering created five consumption clusters and four flexibility clusters, as shown in Figure 6 and Figure 7.

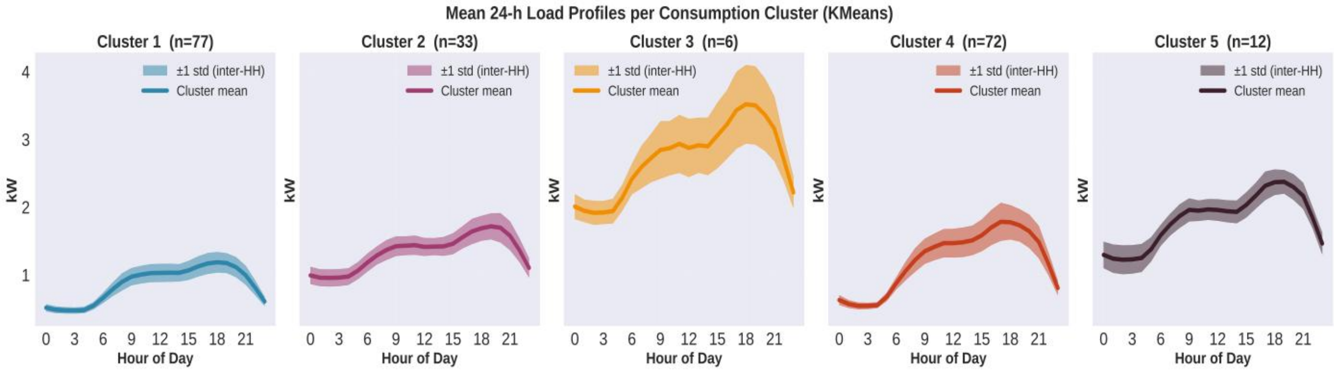

Figure 6: Mean 24-hour electricity load profiles for consumption-based clusters derived using K-Means: (a) low and stable demand, (b) moderate increasing usage, (c) high consumption with pronounced peaks, (d) evening-dominant peak behavior, and (e) sustained high baseline consumption

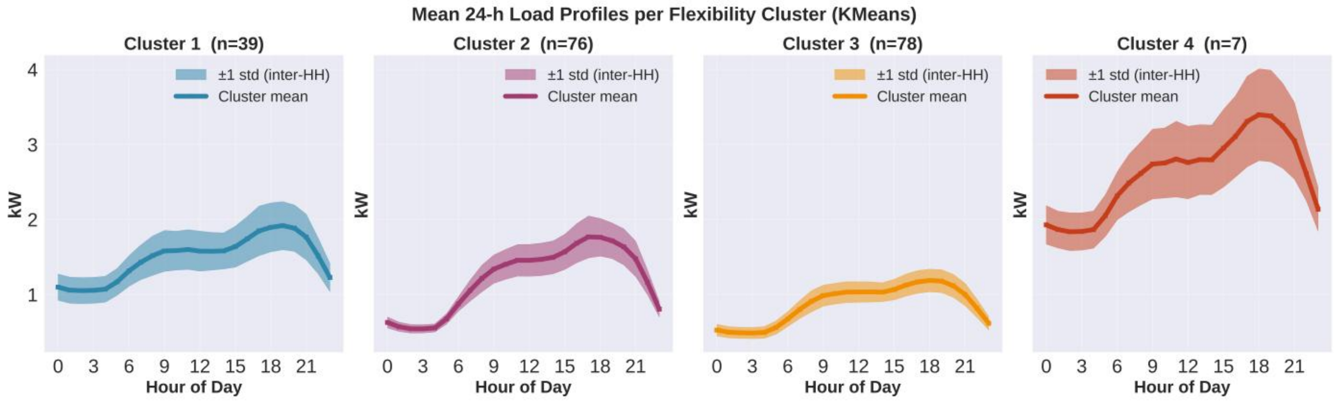

Figure 7: Mean 24-hour load profiles for flexibility-based clusters derived using K-Means: (a) moderate and stable demand, (b) increasing load with peak sensitivity, (c) low and minimally variable demand, and (d) high demand with strong peak flexibility

The clustering results reveal clear and practically meaningful heterogeneity in residential electricity use, as evidenced by the distinct 24-hour load profiles across the five K-means clusters. Cluster 1 represents low-demand households with flat, stable profiles, indicating limited peak contribution and correspondingly low demand response (DR) potential. Cluster 2 shows moderate consumption with a gradual daytime increase, reflecting predictable behavior and moderate flexibility. In contrast, Clusters 3 and 4 exhibit pronounced peak-hour demand, with Cluster 3 characterized by overall high consumption and Cluster 4 by sharp evening-dominant peaks; both groups contribute disproportionately to system peak load and therefore offer the highest value for targeted DR interventions. Cluster 5 maintains a high baseline load throughout the day with moderate peaks, suggesting consistent usage with potential for partial load shifting rather than aggressive peak reduction. Overall, these patterns demonstrate that peak demand is not uniformly distributed across households but concentrated within specific segments, validating the need for segmentation-driven DR strategies. Targeting high-impact clusters—particularly those with elevated and time-concentrated peaks—can achieve substantial peak reduction and cost–emission benefits while avoiding unnecessary intervention in low-impact groups. This reinforces the effectiveness of the proposed clustering-based framework in enabling precise and operationally relevant demand response planning.

The flexibility-based clustering reveals distinct patterns in households' ability to adjust and redistribute electricity demand over the day. Cluster 1 exhibits relatively stable consumption with moderate variation, indicating limited but consistent flexibility potential. Cluster 2 demonstrates a gradual increase in load with noticeable peak-hour concentration, suggesting moderate flexibility with opportunities for controlled load shifting. Cluster 3, characterized by lower overall consumption and smoother profiles, reflects households with minimal variability and consequently lower demand response (DR) potential due to limited shiftable load. In contrast, Cluster 4 displays the highest consumption levels along with pronounced peak-hour demand, indicating substantial flexibility capacity and strong suitability for DR interventions.

These results highlight that flexibility is not solely dependent on total consumption but is strongly influenced by temporal load variability and peak concentration. Households in high-flexibility clusters - particularly those with elevated and time-sensitive peaks - offer the greatest potential for effective load shifting and peak reduction. Conversely, low-variability clusters contribute less to peak stress and are less responsive to DR strategies. This segmentation underscores the importance of distinguishing between consumption magnitude and operational flexibility, reinforcing the need for targeted, cluster-driven demand response approaches to maximize system-level efficiency and impact.

The consumption clusters show how different homes use electricity, while the flexibility clusters show how well different homes can adapt to changes in demand. This dual clustering framework gives a complete picture of households by looking at both how they use energy and how easy it is for them to change or cut back on their load. This kind of segmentation is very helpful for energy aggregators and distribution system operators because it helps them plan for and respond to changes in demand more quickly.

4.4 Joint Distribution of Consumption and Flexibility Clusters

To understand the relationship between electricity consumption patterns and demand response flexibility, a cross-tabulation analysis was performed between the consumption clusters (five groups) and flexibility clusters (four groups). Table 4 summarizes the number of households belonging to each cluster combination.

Table 4. Cross-Cluster Distribution of Households Across Consumption and Flexibility Segments, Highlighting Dominant Behavioral Patterns and Demand Response Potential

Consumption Cluster	**Flexibility Clusters**				**Total**	**Dominant Flexibility Cluster**
	0	**1**	**2**	**3**		
0	0	3	**74**	0	77	Flex 2
1	**28**	1	4	0	33	Flex 0
2	0	0	0	**6**	6	Flex 3
3	0	**72**	0	0	72	Flex 1
4	**11**	0	0	1	12	Flex 0
Total	39	76	78	7	200	—

The findings indicate distinct correlations between consumption and flexibility attributes. Most of the homes in consumption cluster 0 are also in flexibility cluster 2 (74 homes). This means that a lot of homes that don't consume energy much or at all have the potential to be flexible. Consumption cluster 3 is also closely related to flexibility cluster 1, which has 72 households. This means that households that have similar consumption also have similar flexibility traits. The findings indicate distinct correlations between consumption and flexibility attributes. Most of the homes in consumption cluster 0 are also in flexibility cluster 2 (74 homes). This means that a lot of homes that don't consume energy much or at all have the potential to be flexible. Consumption cluster 3 is also closely related to flexibility cluster 1, which has 72 households. This means that households that have similar consumption also have similar flexibility traits.On the other hand, consumption cluster 2 is only connected to flexibility cluster 3 (6 households). This is a small but distinct group of households that act differently when it comes to flexibility. The distribution shows that not all consumption groups have the same amount of flexibility traits. This shows how important it is to think about both how people consume energy and how flexible they can be when making plans for demand response. To determine which households are most suitable for demand response interventions, the peak electricity usage ratio was calculated for each flexibility cluster. Clusters with a higher share of electricity consumption during peak hours were selected as candidates for load shifting. With this analysis, flexibility clusters 1 and 2 were automatically identified as the most suitable for demand response participation. These clusters show households with higher peak consumption proportions, indicating greater potential for peak load reduction through load shifting strategies.

4.5 Demand Response Impact on System Cost and Emissions

A demand response strategy was implemented by shifting 20% of peak-hour electricity consumption (17:00–20:00) to off-peak periods for households belonging to the automatically selected flexibility clusters. To evaluate the effectiveness of demand response (DR) strategies, several operational parameters are defined. The peak demand period is assumed to occur between 17:00 and 20:00 hours, which typically corresponds to periods of high electricity demand, increased electricity prices, and higher carbon intensity in power generation.

A time-of-use tariff structure is considered in the analysis [43]. The electricity price during peak hours is assumed to be 0.30 USD/kWh, while the off-peak electricity price is 0.10 USD/kWh and it provides the price ratio of 3:1. This ratio comes within the range commonly used for residential TOU tariffs, where peak rates have been shown two to three times higher than off-peak rates[44] [45].In addition, carbon emission factors are incorporated to estimate environmental impacts with same step pattern for simplicity. The carbon intensity during peak periods is assumed to be 0.7 $kgCO_2$/kWh, whereas the off-peak carbon intensity is 0.3 $kgCO_2$/kWh. These values are representative of high- and low-carbon operating conditions for power systems. Peak demand is generally met through dispatching fossil fuel based generation, while off-peak demand is more likely to be supplied from lower carbon generation sources.

Actual grid carbon intensity is more variable, but this simplification allows a direct comparison of cost and carbon savings using identical time bounds. These assumptions allow the proposed model to be assessed simultaneously for cost savings and carbon savings using a consistent pricing and emissions framework.

A load shifting strategy is implemented in which a fraction of the electricity consumption during peak hours is shifted to off-peak hours [46] [47]. In this study, it is assumed that 20% of the peak household load can be shifted through demand response (DR) strategies. However, residential demand flexibility is inherently heterogeneous, depending on occupants' daily routines, appliance usage, willingness to participate, and the availability of smart energy management systems. This value is adopted as a representative planning assumption due to the flexible loads[48], [49].

Let s represent the shift percentage , defined in (7).

$$s = 0.20 \tag{7}$$

This implies that 20% of peak electricity usage is shifted to off-peak periods. However, only households belonging to highly flexible clusters participate in the demand response program in targeted DR Strategy

4.5.1 Cluster-Based Demand Response Simulation

A cluster-based demand response simulation is conducted to evaluate the impact of targeted load shifting. A new structured dataset is created from formed clusters by mapping each household to its corresponding consumption and flexibility cluster labels and organizing the results into a tabular format.

To evaluate the impact of demand response strategies, a mathematical formulation is developed to model the load shifting process, electricity cost, and carbon emissions.

Let H denote the set of households and T denote the set of hourly time periods in a year. The electricity consumption of household h at hour t is represented as $L_{h,t}$. To determine peak and off-peak periods, the time horizon T is divided into two subsets as given in (8).

$$T = T_{peak} \cup T_{off} \tag{8}$$

where T_{peak} represents peak hours, and T_{off} represents off-peak hours. In this study, the peak period is defined as per (9).

$$T_{peak} = \{17, 18, 19, 20\} \tag{9}$$

which corresponds to the time interval from 5 PM to 8 PM. For each household h and day d, the total electricity consumption during peak hours is calculated according to (10)

$$E_{peak}^{(h,d)} = \sum_{(t \in T_{peak})} L_{h,t} \tag{10}$$

where $L_{h,t}$ represents the hourly electricity consumption. A predefined fraction of peak load is shifted to off-peak hours. Let's s denote the load shifting percentage as shown in (11).

$$E_{shift}^{(h,d)} = s \times E_{peak}^{(h,d)} \tag{11}$$

The adjusted peak-hour load after demand response is calculated as per (12).

$$L_{h,t}^{DR} = (1 - s)L_{h,t}, \quad \forall t \in T_{peak} \tag{12}$$

The shifted energy is redistributed uniformly across off-peak hours according to (13).

$$L_{h,t}^{DR} = L_{h,t} + \frac{E_{shift}^{(h,d)}}{|T_{off}|}, \quad \forall t \in T_{off} \tag{13}$$

where $|T_{off}|$ represents the number of off-peak hours.

4.5.2 Cluster-Based Participation

Let F_h denote the flexibility cluster assigned to household h. Demand response is applied only if the household belongs to a predefined set of flexible clusters F^*, as shown in (14).

$$h \in F^* \tag{14}$$

If $h \notin F^*$ the household load remains unchanged and can be shown as (15).

$$L_{DR}^{\{h,t\}} = L_{h,t} \tag{15}$$

4.5.3 Electricity Cost Calculation and Carbon Emission Calculations

The total electricity cost is calculated as per (16) based on time-of-use pricing.

$$C = \sum_{(t \in T_{peak})} P_{peak} L_t + \sum_{(t \in T_{off})} P_{off} L_t \tag{16}$$

where:

- P_{peak} = electricity price during peak hours
- P_{off} = electricity price during off-peak hours
- L_t = aggregated load at hour t

Total carbon emissions, as shown in (17), are estimated using time-dependent emission factors.

$$E_{CO_2} = \sum_{(t \in T_{peak})} \gamma_{peak} L_t + \sum_{(t \in T_{off})} \gamma_{off} L_t \tag{17}$$

where,

- γ_{peak} = emission factor during peak hours ($kgCO_2$ /kWh)
- γ_{off} = emission factor during off-peak hours

4.5.4 Performance Metrics

To evaluate the effectiveness of the demand response strategy, the following metrics are computed. The Electricity cost savings are calculated as per (18).

$$C_{save} = C_{baseline} - C_{DR} \tag{18}$$

Carbon emission reduction is calculated as per (19).

$$E_{save} = E_{baseline} - E_{DR} \tag{19}$$

These metrics quantify the economic and environmental benefits achieved through cluster-based demand response.

The targeted DR strategy resulted in a cost reduction of \$12,928.77 and a CO_2 emission reduction of 25,857.54 kg over the study period. The cluster-wise reduction of peak load as per the proposed approach is shown in Figure 8.

These results demonstrate that selectively applying demand response to households with higher flexibility potential can effectively reduce both operational costs and environmental impacts.

The impact of demand response was also evaluated at the household level by comparing baseline and post-DR energy costs and emissions. The results indicate that households participating in the demand response program achieved measurable cost and emission savings.

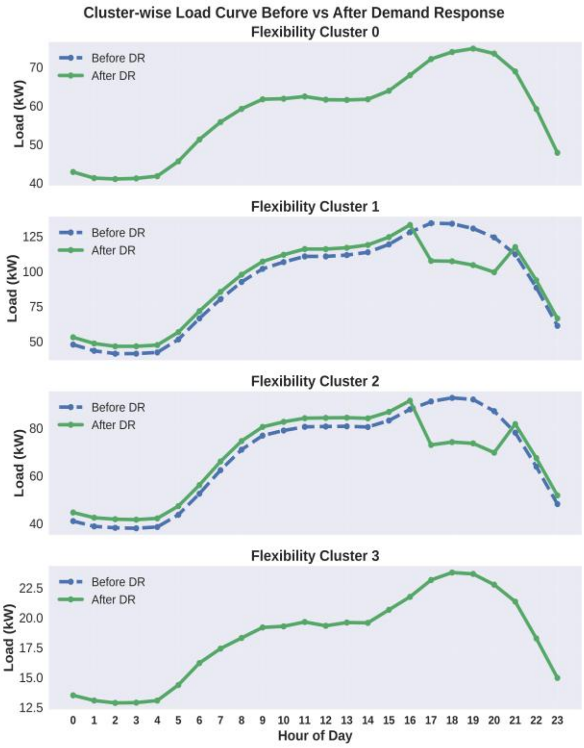

Figure 8: Comparison of load profile for 24-hour load before and after the DR program across different Flexibility clusters.

4.5.5 Uniform Demand Response

To evaluate the effectiveness of the cluster-based demand response strategy, a comparison scenario called uniform demand response is implemented. In this scenario, the same load shifting strategy is applied to all households regardless of their flexibility characteristics.

The same peak-hour definition and shift percentage are used, but participation is not restricted to specific clusters. This scenario represents a non-targeted demand response strategy in which all consumers contribute equally to peak load reduction.

To assess the effectiveness of targeted demand response, a uniform DR scenario was also simulated in which the same 20% peak load shifting strategy was applied to all households regardless of their flexibility characteristics. The results show the uniform DR total Cost and carbon emission are \$280,433.87 and 766,954.28 kg, respectively.

4.5.6 Comparative Evaluation of Demand Response Strategies

Finally, three scenarios are compared to assess the performance of the proposed framework.

- Baseline Scenario: No demand response is applied.
- Cluster-Based Demand Response: Only households belonging to flexible clusters participate in load shifting.
- Uniform Demand Response: All households participate in load shifting.

Through a comparison analysis of each scenario, the key performance indicators are total electricity cost,

cost savings, and emission reduction. Performance results using baseline and DR scenarios are presented in Table 5. The results demonstrate that there are substantial system-wide benefits from targeted DR actions in households with significant flexibility, even if not all households participate in the action at the same time.
The proposed cluster-based demand response strategy achieves significant annual reductions of 12,928.77 USD in electricity cost and 25,857.54 kg of CO_2 emissions, demonstrating its effectiveness in improving both economic and environmental performance.
The cost reduction reflects both decreased peak consumption and the beneficial shift to lower-priced off-peak periods.

Table 5. Aggregate Annual Performance: Baseline vs. Demand Response

Metric	Baseline	With DR	Absolute Reduction	Relative Reduction (%)
Total Cost (USD)	299,020.70	286,091.93	12,928.77	4.32%
Total CO_2 Emissions (kg)	804,127.94	778,270.40	25,857.54	3.21%

The uniform DR scenario resulted in greater overall system-level savings compared to the targeted DR strategy. However, targeted demand response provides several practical advantages, including improved feasibility, reduced participant burden, and better alignment with household flexibility capabilities.
By focusing on households with higher peak consumption and high flexibility shares, the targeted strategy enables more realistic and implementable demand response programs while still achieving high cost savings and emission reductions.
A sensitivity analysis was conducted to investigate the effect of different load shifting percentages on electricity cost, and CO_2 emissions.
The sensitivity analysis with varying load shift percentages is shown in Table 6 which demonstrates that increasing the load shifting percentage enhances electricity cost savings and CO_2 emission reductions.

Table 6 sensitivity analysis with varying load shift percentages

Load-Shifting Percentages	Cost Savings	CO_2 Reduction
10	$6,464.39	12,928.77 kg
15	$9,696.58	19,393.16 kg
20	$12,928.77	25,857.54 kg
30	19,393.16	38,786.31 kg

However, instead of economical and environmental benefits, excessive load shifting may increase rebound effects and and can affect post-demand response load recovery and overall grid performance.
The simulation is performed 5 times and the section wise execution time is shown in Table 7. By taking average of run times the final execution time is 3 minutes 30 seconds.

Table 7 Section Wise Execution Time of Simulation

Sections	1st Run	2nd Run	3rd Run	4th Run	5th Run
Data loading & cleaning	7.86 s	4.78 s	2.52 s	2.53 s	8.51 s
Hourly aggregation	169 ms	276 ms	225 ms	189 ms	377 ms
Clustering and plotting	23.9 s	21.6 s	21.6 s	21.2 s	31.9 s
Cross tabulation	10 µs+ 196 ms	12.9 µs+6 7.6 ms	18.6 µs+4 2.8 ms	16.7 µs, 49.4 ms	14.5 µs+7 4.3 ms
Targeted DR	1min 25 s	1min 15 s	1min 7 s	1min 7 s	1min 6 s
Uniform DR	1min 42 s	1min 42 s	1min 27 s	1min 29 s	1min 42 s
Average run time: 3 minutes 30 seconds					

5. DISCUSSION

The findings indicate that integrating the consumption patterns of electricity and flexibility characteristics in demand response can aid in planning. A clustering method is being used to help identify appropriate consumers for load shifting for demand response based on their electricity consumption history and the level of flexibility they have available.
The analysis of how demand response impacts on residential consumption has confirmed that the these stratagies successfully decreased electricity use during peak time, so there was less overall cost of electricity and less carbon emissions. Therefore, to effectively implement demand response programs, it is essential to cluster households based on their data-driven characteristics. In addition to providing utility and energy aggregators a practical method to identify the best households to participate in effective demand response programs, the approach will also aid them in meeting their overarching system-wide energy usage and sustainability targets. When used together, the consumption pattern clustering method and the flexibility characterization of households can allow for a more strategic and rational choice of which households to use for load-shifting activities.
The electric power consumption pattern and the potential for flexibility are not evenly distributed among the different households, as analyzed by the cross-cluster method. There is a close relation exhibited between certain consumption clusters and some specific flexibility group, which shows that the same

demand response capability is shown by similar consumption patterns. This concludes that for developing an optimal demand response strategy, both flexible features and consumption patterns are needed.
The demand response simulation results reveal further benefits to this focused approach. This technique has produced results with both economic and environmental benefits through a \$12,928.77 reduction in electricity costs and 25,857.54 kg of CO2 emissions associated with the analyzed period. A uniform demand response approach where all households took part in an identical peak load reduction program was also assessed. Although this scenario resulted in the highest overall savings (\$18,586.83 in cost reduction and 37,173.67 kg in CO_2 reduction), such an assumption is unlikely to be feasible in real-world settings due to variations in household lifestyles, appliance ownership, and operational constraints.
The overall findings reveal why using data-based segmentation techniques in household classifications is critical for today's electricity supply & distribution network technologies (i.e., smart grids). The proposed method provides examples of how it can be used effectively, which is valuable for finding suitable households willing to participate in demand side management programs, as well as assisting utilities to build less expensive & more sustainable strategies to manage demand from residential customers.
The data mining techniques with operationally relevant features in demand response in the study shows a shift from traditional DR to an intelligent, adaptive and data driven DR.
This analysis models idealized load-shifting behavior and excludes factors such as automation failures and consumer override preferences. In practice, participation rates are often lower than modeled potentials due to opt-out decisions, technical issues, and comfort preferences. While this study applies both targeted and uniform demand response strategies, future work could extend the model to include a multi-level demand response framework.

6. LIMITATIONS AND FUTURE WORK

While the proposed dual clustering-based DR framework demonstrates promising results, several limitations remain. Since the dataset used in the work is limited, an additional more diverse dataset with larger number of households will be used in futute to demonstrate scalability and generalization.
The real-world factors such as household income, occupancy patterns, geographic location, and appliance ownership can significantly influence residential load profiles. Although these factors can influence residential load profiles, they were unavailable in the adopted dataset. Future work will integrate such contextual information to investigate its impact on clustering performance, and demand response applications.
The model treats DR capacity as fixed, whereas actual customer response varies across events, Future work will consider heterogeneous flexibility. In the proposed work, the shifted load is uniformly distributed across off-peak hours, which helps reduce the potential for rebound peaks by avoiding the concentration into a single time period. However, the challenges associated with rebound effects are the main concern for the DR, the priority and flexibility based strategies will be used in future work.
The results are using fixed price and carbon intensity values (0.30/0.10 \$/kWh and 0.7/0.3 kgCO2/kWh). These results may vary by using different tariff structure and emission factors. For example, a larger difference between peak and off-peak prices would lead to larger cost savings, and a smaller difference would lead to smaller savings. Similarly, the carbon savings are dependent on the generation mix of the regions. The grids with greater renewables may have a smaller spread between peak and off-peak emissions. The future work will incorporate generation mix and different tariff structures.
However, sensitivity analysis with varying load percentages and execution time is reported in the study, statistical analysis, such as repeated simulations and confidence interval estimation, would provide additional insight into the robustness of the obtained results.
The model assumes full customer participation (in the selected flexible clusters), whereas real-world acceptance is influenced by comfort, privacy, and trust concerns, which may reduce effective participation. The framework assumes instantaneous communication between the operator and end-use devices, whereas real systems experience latency that can affect timely dispatch. Future work will address these limitations along several directions with more performance indicators like peak demand reduction percentage, load factor improvement, and peak-to-average ratio would provide a more comprehensive evaluation. Together, these extensions aim to bridge the gap between theoretical dual clustering based DR models and their practical deployment in real-world grid environments.

7. CONCLUSIONS

The proposed work clearly depicts that cluster-based targeted demand response programs targeting high-consumption and high-flexibility household groups provide substantial economic and environmental benefits in the smart grid context. Dual-clustering frameworks with DR-specific features yield high-quality clusters, thereby improving DR program targeting and effectiveness by segmenting consumers based on consumption patterns, flexibility characteristics and their cross cluster analysis . The majority voting of CVIs and Demand Separability Score (DSS) makes operational meaningful clusters and helps to better target demand response efforts. The strategy of the proposed work achieves $12,928.77 saving in the electricity costs and a reduction of 25,857.54 kg in CO_2 emissions over the study period while maintaining energy balance. Future research will incorporate integration with emerging distributed energy resources. In smart grids, the cluster-based DR frameworks will become increasingly valuable for ensuring grid flexibility for highly penetrated renewable energy systems.

REFERENCES

[1] N. P. Koirala, L. Nyiwul, Z. Hu, R. Al-Hmoud, and D. P. Koirala, "Geopolitical risks and energy market dynamics," Energy Econ., vol. 150, p. 108814, Oct. 2025, doi: 10.1016/j.eneco.2025.108814.

[2] M. Cavcic, "Middle East conflict: Energy security risks and price shocks as market volatility hits supply chains," Offshore Energy. Accessed: Mar. 26, 2026. [Online]. Available: https://www.offshore-energy.biz/middle-east-conflict-energy-security-risks-and-price-shocks-as-market-volatility-hits-supply-chains/

[3] "South Korea's renewable energy pivot can mitigate fossil fuel dependency risks | IEEFA." Accessed: Jul. 29, 2026. [Online]. Available: https://ieefa.org/resources/south-koreas-renewable-energy-pivot-can-mitigate-fossil-fuel-dependency-risks

[4] "2025_outline.pdf.pdf." Accessed: Jul. 29, 2026. [Online]. Available: https://www.enecho.meti.go.jp/en/category/whitepaper/pdf/2025_outline.pdf.pdf

[5] "Global Energy Review 2026 – Analysis," IEA. Accessed: Jul. 29, 2026. [Online]. Available: https://www.iea.org/reports/global-energy-review-2026

[6] M. Wirayuda, I. Baihaqi, and M. Sugihartanto, "Literature Review on LNG Supply Chain Risk Mapping," Int. J. Bus. Manag. Technol. Soc., vol. 1, pp. 110–117, Jan. 2026, doi: 10.12962/j30254256.v1i2.7928.

[7] IEA, "2026 Energy Crisis Policy Response Tracker – Data Tools," IEA. Accessed: Jul. 29, 2026. [Online]. Available: https://www.iea.org/data-and-statistics/data-tools/2026-energy-crisis-policy-response-tracker

[8] O. M. Neda, J. Adabi, H. Gholinezhadomran, and M. Marzband, "Machine learning-enhanced coordination of home-microgrids for resource compensation in large-scale energy systems," Sustain. Cities Soc., vol. 134, p. 106744, Nov. 2025, doi: 10.1016/j.scs.2025.106744.

[9] O. M. Neda, "Adaptive Protection Scheme Using Optimal Coordination of Directional Overcurrent Relays for Active Distribution Networks," Bul. Ilm. Sarj. Tek. Elektro, vol. 8, no. 1, pp. 165–191, Nov. 2025, doi: 10.12928/biste.v8i1.14954.

[10] M. K. Ekoue, M. Woerman, and C. Clastres, "Intermittency and uncertainty in wind and solar energy: Impacts on the French electricity market," Energy Econ., vol. 142, p. 108176, Feb. 2025, doi: 10.1016/j.eneco.2024.108176.

[11] H. Sæle, B. H. Bakken, and K. Dalen, "Flexibility as an Enabler for Value Creation: A Norwegian Perspective," IEEE Power Energy Mag., vol. 24, no. 2, pp. 43–53, Mar. 2026, doi: 10.1109/MPE.2025.3596505.

[12] "Flexibility for a secure and affordable power sector transformation." Accessed: Mar. 22, 2026. [Online]. Available: https://www.irena.org/Publications/2026/Jan/Flexibility-for-a-secure-and-affordable-power-sector-transformation

[13] M. E. Honarmand, V. Hosseinnezhad, B. Hayes, M. Shafie-Khah, and P. Siano, "An Overview of Demand Response: From its Origins to the Smart Energy Community," IEEE Access, vol. 9, pp. 96851–96876, 2021, doi: 10.1109/ACCESS.2021.3094090.

[14] O. M. Neda, J. Adabi, M. Marzband, and H. Gholinezhadomran, "Hierarchical energy management system for coordinated operation of multiple grid-tied home microgrids," Renew. Energy Focus, vol. 56, p. 100766, Mar. 2026, doi: 10.1016/j.ref.2025.100766.

[15] F. Plaum, R. Ahmadiahangar, A. Rosin, and J. Kilter, "Aggregated demand-side energy flexibility: A comprehensive review on characterization, forecasting and market prospects," Energy Rep., vol. 8, pp. 9344–9362, Nov. 2022, doi: 10.1016/j.egyr.2022.07.038.

[16] V. Cortez, R. Rabelo, A. Carvalho, and V. Pilloni, "Towards Consumer-Oriented Demand Response Systems," in 2022 7th International Conference on Smart and Sustainable Technologies (SpliTech), Jul. 2022, pp. 1–6. doi: 10.23919/SpliTech55088.2022.9854320.

[17] Q. Huang, R. Gao, and H. Akhavan, "An ensemble hierarchical clustering algorithm based on merits at cluster and partition levels," Pattern Recognit., vol. 136, p. 109255, Apr. 2023, doi: 10.1016/j.patcog.2022.109255.

[18] 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, p. 109628, Mar. 2020, doi: 10.1016/j.rser.2019.109628.

[19] S. Biswas and S. A. Abraham, "Identification of Suitable Consumer Groups for Participation in Demand Response Programs," in 2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Feb. 2019, pp. 1–5. doi: 10.1109/ISGT.2019.8791640.

[20] G. Chicco, R. Napoli, and F. Piglione, "Comparisons among clustering techniques for electricity customer classification," IEEE Trans. Power Syst., vol. 21, no. 2, pp. 933–940, May 2006, doi: 10.1109/TPWRS.2006.873122.

[21] O. M. Neda, "A Review on Smart Distribution Systems and the Role of Deep Learning-based Automation in Enhancing Grid Reliability and Efficiency," Bul. Ilm. Sarj. Tek. Elektro, vol. 7, no. 2, pp. 195–205, Apr. 2025, doi: 10.12928/biste.v7i2.13152.

[22] O. M. Neda, "Optimal amalgamation of DG units in radial distribution system for techno-economic study by improved SSA: Practical case study," Electr. Power Syst. Res., vol. 241, p. 111365, Apr. 2025, doi: 10.1016/j.epsr.2024.111365.

[23] A. Kaur, Y. Kumar, and J. Sidhu, "Exploring meta-heuristics for partitional clustering: methods, metrics, datasets, and challenges," Artif. Intell. Rev., vol. 57, no. 10, p. 287, Sep. 2024, doi: 10.1007/s10462-024-10920-1.

[24] O. M. Neda, "A dual placement strategy for DG and D-STATCOM units using multi-objective optimization in an

Iraqi radial distribution system," Expert Syst. Appl., vol. 299, p. 130230, Mar. 2026, doi: 10.1016/j.eswa.2025.130230.
[25] O. M. Neda, "A novel WSSA technique for multi-objective optimal capacitors placement and rating in radial distribution networks," Int. J. Appl. Power Eng. IJAPE, vol. 14, no. 4, pp. 934–950, Dec. 2025, doi: 10.11591/ijape.v14.i4.pp934-950.
[26] C. Simpson, R. J. G. B. Campello, and E. Stojanovski, "Benchmarking of Clustering Validity Measures Revisited," Stat. Anal. Data Min. ASA Data Sci. J., vol. 19, no. 1, p. e70061, 2026, doi: 10.1002/sam.70061.
[27] M. Jain, T. AlSkaif, and S. Dev, "A Clustering Framework for Residential Electric Demand Profiles," in 2020 International Conference on Smart Energy Systems and Technologies (SEST), Sep. 2020, pp. 1–6. doi: 10.1109/SEST48500.2020.9203534.
[28] H. Mahmood, T. Mehmood, and L. A. Al-Essa, "Optimizing Clustering Algorithms for Anti-Microbial Evaluation Data: A Majority Score-Based Evaluation of K-Means, Gaussian Mixture Model, and Multivariate T-Distribution Mixtures," IEEE Access, vol. 11, pp. 79793–79800, 2023, doi: 10.1109/ACCESS.2023.3288344.
[29] J.-W. Xiao, Y. Xie, H. Fang, and Y.-W. Wang, "A new deep clustering method with application to customer selection for demand response program," Int. J. Electr. Power Energy Syst., vol. 150, p. 109072, Aug. 2023, doi: 10.1016/j.ijepes.2023.109072.
[30] V. Michalakopoulos, E. Sarmas, I. Papias, P. Skaloumpakas, V. Marinakis, and H. Doukas, "A machine learning-based framework for clustering residential electricity load profiles to enhance demand response programs," Appl. Energy, vol. 361, p. 122943, May 2024, doi: 10.1016/j.apenergy.2024.122943.
[31] J. Enriquez-Loja, B. Castillo-Pérez, X. Serrano-Guerrero, and A. Barragán-Escandón, "Performance evaluation method for different clustering techniques," Comput. Electr. Eng., vol. 123, p. 110132, Apr. 2025, doi: 10.1016/j.compeleceng.2025.110132.
[32] O. M. Neda, "A hybrid LSTM-XGBoost model for multi-horizon short-term load forecasting in smart power distribution systems," E-Prime – Nexus Electr. Electron. Intell. Eng., vol. 17, p. 201211, Sep. 2026, doi: 10.1016/j.eprime.2026.201211.
[33] E. Barbierato and A. Gatti, "Decoding Urban Intelligence: Clustering and Feature Importance in Smart Cities," Future Internet, vol. 16, no. 10, p. 362, Oct. 2024, doi: 10.3390/fi16100362.
[34] Y.-S. Kim, M. K. Kim, N. Fu, J. Liu, J. Wang, and J. Srebric, "Investigating the impact of data normalization methods on predicting electricity consumption in a building using different artificial neural network models," Sustain. Cities Soc., vol. 118, p. 105570, Jan. 2025, doi: 10.1016/j.scs.2024.105570.
[35] X. Cui, M. Lee, M. N. Uddin, X. Zhang, and V. G. Zakka, "Analyzing different household energy use patterns using clustering and machine learning," Renew. Sustain. Energy Rev., vol. 212, p. 115335, Apr. 2025, doi: 10.1016/j.rser.2025.115335.
[36] D. T. Kusuma, N. Ahmad, S. S. S. Ahmad, I. B. Sangadji, and Y. Arvio, "An efficient clustering approach in electrical energy consumption patterns," Bull. Electr. Eng. Inform., vol. 14, no. 2, pp. 1168–1177, Apr. 2025, doi: 10.11591/eei.v14i2.8666.
[37] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," Energy, vol. 42, no. 1, pp. 68–80, Jun. 2012, doi: 10.1016/j.energy.2011.12.031.
[38] K. Golalipour, E. Akbari, S. S. Hamidi, M. Lee, and R. Enayatifar, "From clustering to clustering ensemble selection: A review," Eng. Appl. Artif. Intell., vol. 104, p. 104388, Sep. 2021, doi: 10.1016/j.engappai.2021.104388.
[39] O. M. Neda, "A multi-source data-driven framework for intelligent diagnostics and prognostics of power transformers," Measurement, vol. 286, p. 122354, Sep. 2026, doi: 10.1016/j.measurement.2026.122354.
[40] B. Peng, B. Cui, X. Ma, and C. Liu, "Data-Driven assessment of demand response potential of residential loads based on LSTM-WGCNA and Customer directrix load," Energy Build., vol. 356, p. 117101, Apr. 2026, doi: 10.1016/j.enbuild.2026.117101.
[41] M. Muratori, "Impact of uncoordinated plug-in electric vehicle charging on residential power demand - supplementary data." National Renewable Energy Laboratory - Data (NREL-DATA), Golden, CO (United States); National Renewable Energy Laboratory, p. 3 files, 2017. doi: 10.7799/1363870.
[42] S. Liu, T. Xu, X. Du, Y. Zhang, and J. Wu, "A hybrid deep learning model based on parallel architecture TCN-LSTM with Savitzky-Golay filter for wind power prediction," Energy Convers. Manag., vol. 302, p. 118122, Feb. 2024, doi: 10.1016/j.enconman.2024.118122.
[43] Z. Fu, K. Novan, and A. Smith, "Do time-of-use prices deliver energy savings at the right time?," J. Environ. Econ. Manag., vol. 128, p. 103054, Nov. 2024, doi: 10.1016/j.jeem.2024.103054.
[44] "Peak and Off-Peak Electricity: Cheapest Times to Use Energy and Lower Your Bills - A1 SolarStore Magazine." Accessed: Jul. 30, 2026. [Online]. Available: https://a1solarstore.com/blog/peak-and-off-peak-electricity-cheapest-time-to-use-energy.html
[45] M. Shen and J. Chen, "Optimization of peak-valley pricing policy based on a residential electricity demand model," J. Clean. Prod., vol. 380, p. 134761, Dec. 2022, doi: 10.1016/j.jclepro.2022.134761.
[46] R. Brännlund and M. Vesterberg, "Peak and off-peak demand for electricity: Is there a potential for load shifting?," Energy Econ., vol. 102, p. 105466, Oct. 2021, doi: 10.1016/j.eneco.2021.105466.
[47] S. Dahlke and M. Prorok, "Consumer Savings, Price, and Emissions Impacts of Increasing Demand Response in the Midcontinent Electricity Market," Energy J., vol. 40, no. 3, pp. 243–262, May 2019, doi: 10.5547/01956574.40.3.sdah.
[48] M. Afzalan and F. Jazizadeh, "Residential loads flexibility potential for demand response using energy consumption patterns and user segments," Appl. Energy, vol. 254, p. 113693, Nov. 2019, doi: 10.1016/j.apenergy.2019.113693.
[49] B. Dey, S. Dutta, S. Saikia, and R. S. Kumar, "An innovative hybrid load shifting and curtailing technique for operating a plug-in hybrid electric vehicle integrated microgrid system in a clean and cost-effective manner," IET Gener. Transm. Distrib., vol. 19, no. 1, p. e70020, 2025, doi: 10.1049/gtd2.70020.