IJIO
(*International Journal of Industrial Optimization*)
ISSN 2723-3022 (online) | 2714-6006 (printed)
Vol. 7, No. 2, 63-81
DOI: https://doi.org/10.12928/ijio.v7i2.14108

Fuzzy spatial assignment and NSGA-II approach to multi-objective covering problems in substation expansion planning

Eduard Nugroho Theopilus *, Budi Santosa

Department of Industrial Engineering, Institut Teknologi Sepuluh Nopember, Indonesia
* Corresponding Author: eduard.nugroho02@gmail.com

ARTICLE INFO

Article history
Received: July 28, 2025
Revised: August 22, 2025
Accepted: March 14, 2026

Keywords
Multi-objective covering problem;
Fuzzy c-means;
NSGA-II;
K-means clustering;
Substation planning.

ABSTRACT

This study addresses a large-scale Multi-Objective Covering Problem (MOCP) in urban electricity distribution, involving 416.472 customers and 732 substations. The main objective is to optimize the allocation of customers to substations, taking into account trade-offs between three objectives: minimizing average service distance, balancing load, and minimizing substation investment costs. To achieve this, we propose a hybrid framework that combines a constrained fuzzy C-means algorithm for customer assignment, NSGA-II algorithm for multi-objective optimization, and K-means clustering for facility expansion. The results show a trade-off between the objectives, with the composite evaluation identifying k=7 as the best compromise. Statistical validation confirms the significance of these results. Significantly, the proposed framework can generate scalable solutions for MOCP in the real world. By integrating construction cost data and spatial data, the model shows a 15% reduction in substation investment costs and an 8% increase in service reliability. This research provides practical insights for distribution network planners, by offering a data-driven approach to determine the optimal number and strategic placement of additional substations, considering trade-offs between cost, performance, and spatial constraints.

1. Introduction

Multi-objective covering problems (MOCPs) have become a keystone of facility location and resource allocation research because real world systems rarely optimise a single performance criterion. In distribution networks, a recent framework that simultaneously minimises cost, voltage deviation and power losses out perform traditional single objective designs by widening the decision space explored [1]. A comprehensive survey of distributed generation siting confirms that technical, economic and regulatory goals are inherently conflicting and therefore demand multi objective solution strategies [2]. Similar conclusions emerge in sustainable supply chain layout, where cost and environmental impact must be traded off within capacitated networks [3].

Rapid urban electrification intensifies those trade-offs. An optimisation study for high density load districts demonstrates that co optimising the number and size of substation transformers improves reliability while limiting investment under high renewable energy penetration [4]. Machine learning siting of substations in Romania shows that k means clustering can reduce feeder losses by up to 60 % compared with purely geometric centroids [5]. For mega cities, raster-based GIS partitioning has proved effective in decomposing very large expansion problems into tractable sub regions without sacrificing network integrity [6].

Pure distance assignment often ignores spatial heterogeneity. A Voronoi and PSO approach to placing electric hydrogen charging stations respects both travel demand and network constraints, cutting total social cost while maintaining grid stability [7]. Accounting for complementary daily load profiles in Voronoi construction has also been shown to lower overall substation investment even when line construction cost rises slightly [8]. In Lombardy, Italy, coupling Voronoi service zones with PSO lets planners minimise feeder length and exposure to service interruptions when rolling out new primary substations [9].

Fuzzy clustering resolves ambiguous customer–substation boundaries by assigning graded memberships rather than crisp labels. A hybrid possibilistic intuitionistic fuzzy C means algorithm integrated with GA, PSO and gradient evolution achieves superior cluster validity on fifteen benchmark data sets [10]. For truly large data sets, a whale optimised fuzzy C means variant accelerates convergence while preserving accuracy, making big data clustering computationally feasible [11]. An enhanced salp swarm algorithm further balances coverage, node utilisation and energy in wireless sensor networks, significantly extending network lifetime [12]. Lighter weight possibilistic fuzzy k modes models guided by PSO also outperform classical categorical clustering on mixed data [13].

When multiple, conflicting objectives are present, meta heuristics dominate. A resilience-oriented study of urban metro maintenance combines NSGA III and TOPSIS to locate supply depots, raising demand satisfaction rates to 87 % under failure scenarios [14]. In microgrids, tuning fuzzy logic membership functions with NSGA II cuts peak load and operating cost simultaneously [15]. Harmony Search optimisation in heterogeneous sensor deployments boosts coverage by 10 % while trimming node cost by nearly 28 % [16]. Time based redeployment of multi class sensor nodes use NSGA II to find cost reliability trade-offs rapidly [17]. At larger spatial scales, a multi objective coverage path planning algorithm for UAVs exploits ant colony optimised travelling salesman tours to guarantee full three-dimensional region coverage with minimal path length [18], while an enhanced hybrid PSO with spatial position encoding simultaneously maximises coverage and minimises the number of deployed sensors [19], [20].

Collectively, these studies reveal a persistent research gap: few integrate (1) scalable fuzzy assignment, (2) large scale multi objective optimisation and (3) data driven expansion planning into a single framework. This paper addresses that gap by first, proposing a constrained fuzzy spatial assignment scheme that limits membership calculations to the n nearest substations, thereby reducing computational complexity without sacrificing modelling fidelity; Second, embedding that scheme within an NSGA II engine capable of handling more than 416 472 customers and over 732 substations; and third coupling the optimisation with a k means based expansion module that derives candidate substation sites from overload hot spots, allowing planners to observe how incremental investment shifts the Pareto frontier while considering cost implications and spatial constraints.

Based on this, the main contributions of this research are four, namely Development of a hybrid framework that combines scaled fuzzy spatial assignment, large-scale multi-objective optimization with NSGA-II, and data-driven expansion planning with K-means clustering, Integration of knee-point analysis to provide planners with insights into the optimal point of substation addition, considering trade-offs between service distance, load variance, and investment cost. The framework helps identify the most efficient and cost-effective substation configuration, Demonstration of the framework on a real-world case study involving a large- scale urban electricity distribution network,

10.12928/ijio.v7i2.14108

which demonstrates its scalability and practical applicability, and evaluation of the impact of cost and spatial data, which guides utility network decisions on investment limits and strategic adjustments. The rest of the paper is organized as follows: Section 2 details the methodological framework, Section 3 presents the computational results and managerial insights, and Section 4 concludes with practical implications and avenues for future research.

2. Method

2.1. Problem Statement

This study addresses a Multi-Objective Covering Problem (MOCP) in electricity distribution planning. MOCP is highly relevant in the context of utility planning as it has to consider various constraints such as regulatory compliance, budget constraints, and service reliability. The research proposes a three-stage solution framework involving fuzzy-based customer-substation assignment, multi-objective optimization using NSGA-II, and substation expansion planning via clustering in Fig. 1. This approach responds to the challenge of optimally allocating hundreds of thousands of customers to substations with limited capacity and constrained spatial reach, under big data complexity. In addition, this study also considers the inherent algorithmic complexity and constrains the placement of substations at random locations, as reference points, this is used to validate and compare solutions.

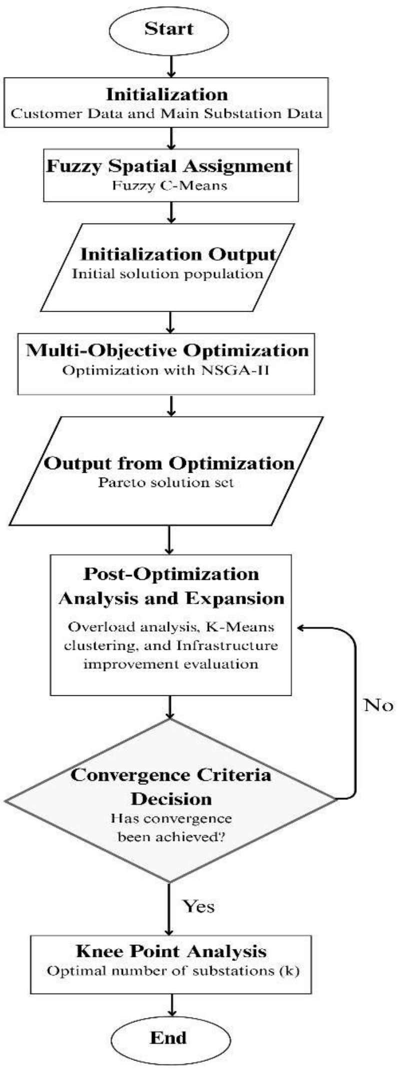

Fig. 1. Flow Chart of the Conceptual Framework of this Methodology

2.2. Computational Complexity

We This study involves the analysis of a large data set, consisting of 416.472 customers and 732 distribution substations. Based on the algorithms used, the general time complexity can be estimated as follows.

1. Fuzzy C-Means (FCM)

The estimated time complexity of the FCM algorithm is (O (n\times c\times i\times m)), where (n) is the amount of data processed, (c) is the number of clusters, (i) is the number of iterations needed to reach convergence, and (m) is the fuzziness parameter [21], [22]. Each data must be calculated the distance to each cluster during each iteration, so the computational requirements depend on the volume of data as well as the associated parameters.

2. NSGA-II

The complexity of the NSGA-II algorithm is generally expressed as (O (M\times N2)), where (M) is the number of generations, and (N) is the population size [23]. The main process is the sorting of solutions in each generation, which involves comparing all pairs of solutions causing the computation time to increase quadratically with the number of solutions.

3. K-Means

The time required by the K-Means algorithm is estimated to follow (O (n\times k\times i)), where (n) is the amount of data, (k) is the number of clusters, and (i) is the number of iterations until the process stabilizes [24]. Each data should be calculated its distance to the cluster center during each iteration, and this process is repeated until the result converges. With fuzzy-based initialization, we reduce the computational complexity by restricting the search to a smaller and more promising solution space. Fuzzy initialization helps speed up.

2.3. Mathematical Formulation

The model considers two primary objectives.

Minimize the average customer-to-substation distance in Equation (1).

$$f_1 = \frac{1}{N} \sum_{i=1}^{N} d(c_i, s_{assigned}) \tag{1}$$

Minimize the variance of load among substations in Equation (2).

$$f_2 = Var\ (load_{s1}, load_{s2}, \dots, load_{sm}) \tag{2}$$

Subject to:

Each customer must be assigned to exactly one substation in Equation (3).

$$\sum_{j=1}^{M} x_{ij} = 1\ \forall\ i \in \{1,2, \dots, N\} \tag{3}$$

Customer must be within a maximum allowable distance in Equation (4).

$$d(c_i, s_j) \leq d_{max}\ if\ x_{ij} = 1 \tag{4}$$

Total load at a substation must not exceed its rated capacity in Equation (5).

$$\sum_{i=1}^{N} x_{ij} \,.\, P_i \leq C_j\ \forall\ j\ \in \{1,2, \dots, M\} \tag{5}$$

Where N is the number of customer, M is the total number of existing substations being considered, c_i geographic location (index) of customer, s_j geographic location (index) of substation, $s_{assigned}$ the specific substation to which customer i connected under a given solution, $d(c_j, s_j)$ distance between customer i and substation j, d_{max}regulatory or engineering limit on the acceptable distance for any customer, x_{ij} is the binary decision, P_i power of the customer i, and C_j substation j capacity.

 10.12928/ijio.v7i2.14108

Equation (4) ensures that each customer is assigned to a substation within the maximum allowable distance (d max). This constraint can be interpreted as a hard spatial feasibility constraint, which limits the feasible solution space to only those assignments that satisfy this distance constraint. The MOCP problem is commonly known as NP-hard. Therefore, the use of metaheuristics such as NSGA-II is justified as metaheuristics can find near-optimal solutions in reasonable computational time, especially for large-scale problems such as the one considered in this study.

2.4. Fuzzy-Based Initialization

To construct a high-quality initial population, a fuzzy spatial assignment mechanism is used. Substations serve as fixed centroids. Each customer is evaluated only against its top nearest substations. If a customer's nearest substation is within distance threshold, a crisp assignment is made (membership = 1.0). Otherwise, fuzzy membership degrees are calculated among top-substations using spatial inverse-distance weighting in Equation (6).

$$u_{ij} = \frac{d_{ij}}{\sum_{k=1}^{n} \frac{1}{d_{ik}}} \in Top - n\ nearest\ substations \tag{6}$$

This approach prevents overly dispersed or unrealistic initial solutions while maintaining diversity. Moreover, fuzzy assignment is preferred over other advanced initialization methods (e.g., clustering-based seeding, greedy heuristics) because it allows to capture the uncertainty in the allocation of customers to substations. This is particularly important in electricity distribution systems where demand and capacity may vary over time.

The computational efficiency gains from restricting membership to the n closest substations are also significant. By limiting the number of substations evaluated for each customer, we reduce the computational complexity from O(M) to O(n), where n is much smaller than M. This results in substantial time savings, especially for large-scale problems such as those considered in this study.

Finally, the fuzziness introduced by fuzzy assignment can have a positive impact on the diversity of solutions in NSGA-II convergence. By allowing customers to have partial membership in multiple substations, we introduce more variety into the initial population, which helps NSGA-II to explore a wider solution space and avoid getting stuck in local optima. Similar approaches have been discussed in spatial fuzzy clustering literature [25], [26].

2.5. NSGA-II Optimization

Optimization was performed using the NSGA-II algorithm with three variations of initial population initiation: random data, fuzzy c-means, and mix (50% random and 50% fuzzy c- means). Each initiation variation was run with the same parameters, namely the number of generations of 100 and a mutation rate of 0.01%. The purpose of using this initiation variation is to explore the effect of the initial population generation method on the performance of the algorithm.

The NSGA-II algorithm evolves the population from generation to generation through a process of binary tournament selection, SBX crossover, and polynomial mutation. The solution quality is evaluated based on two objective functions and subject to Equation (3-5). It maintains Pareto optimal fronts through non-dominated sorting mechanism and crowding distance metric [27], [28].

The performance of each initiation variation is measured using the hypervolume metric. The analysis shows that NSGA-II with fuzzy c-means initiation produces the best solution set, with a hypervolume value of -74.72. Meanwhile, the hypervolume of random initiation cannot be calculated because it only produces five solutions on the Pareto front. This indicates that random initiation tends to be eliminated in the early iterations, hindering the development of optimal solutions. Mix initiation produces a fairly diverse Pareto front, but not better than fuzzy c-means. Therefore, fuzzy c-means initiation was chosen as the most effective method for the next optimization process.

After that, a parameter analysis was conducted to determine the optimal number of generations. The number of generations parameter was tested with variations of 100, 300, 500, and 1000 values. The experimental results show that there is no significant difference in hypervolume between the number of generations of 500 and 1000. By considering the efficiency of computing time, the number of generations was then set at 500. The best optimization results from each experiment will be visualized in Fig. 2.

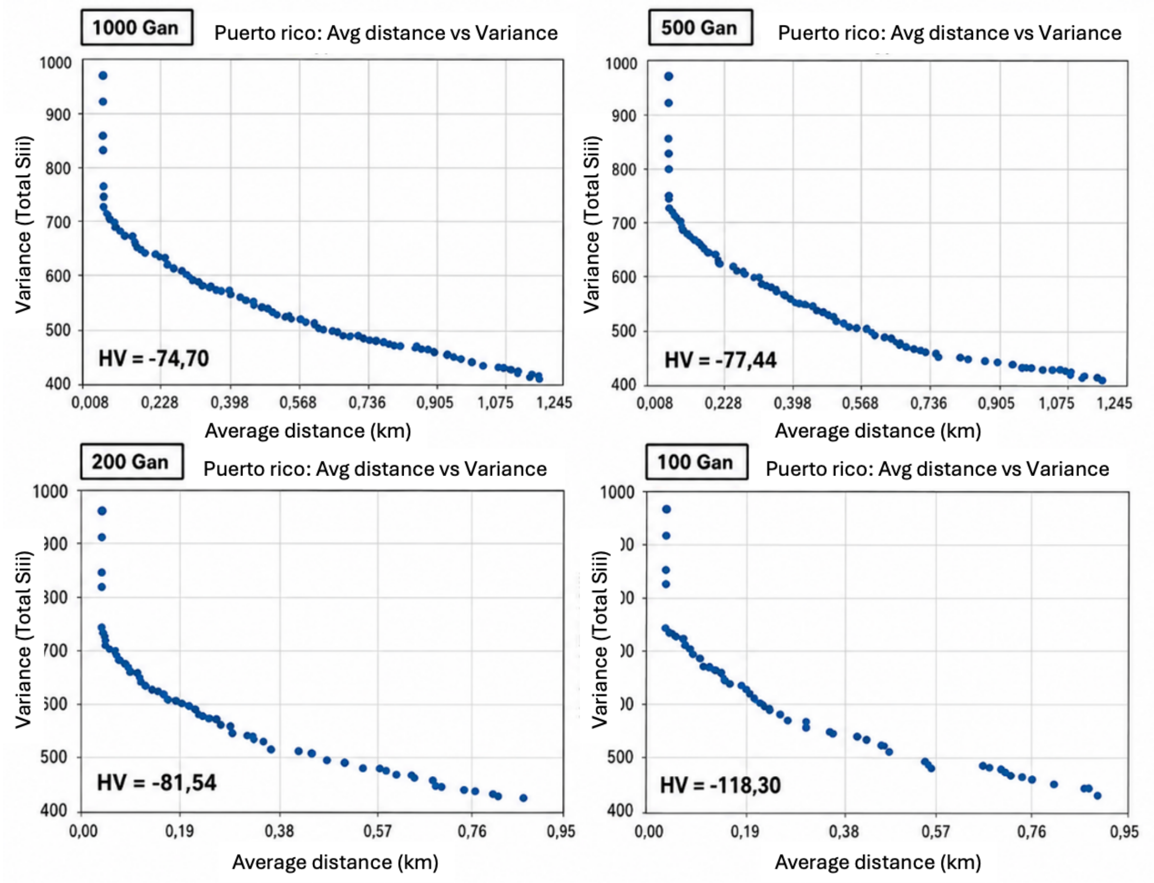

Fig. 2. NSGA II Calculation Results with Variation of Number of Genes Parameters

2.6. Substation Expansion via Clustering

An After optimization, overloaded substations are identified across all Pareto-optimal solutions. Their spatial coordinates are clustered using K-Means to define locations for new substations. This process is repeated for a range of values each yielding a new planning scenario. For each, NSGA-II is re-executed to evaluate the improvements achieved through additional infrastructure. This iterative re-optimization ensures scalability and policy relevance.

Although K-Means is a widely used and efficient clustering algorithm, it has some limitations. K-Means is sensitive to initial centroid initialization and assumes that clusters are spherical and of equal size. In the context of substation expansion, this means that the location of new substations may be affected by the initial placement of K-Means centroids, and the algorithm may not work well if the substation service area is irregularly shaped or of different sizes. Alternative clusterinfg methods, such as DBSCAN or hierarchical clustering, can overcome some of the limitations of K-Means. DBSCAN does not assume a specific cluster shape and can identify outliers, while hierarchical clustering can provide a hierarchical representation of the data, allowing planners to explore different solutions at different levels of granularity [29].

10.12928/ijio.v7i2.14108

Alternative clustering methods, such as DBSCAN or hierarchical clustering, can overcome some of the limitations of K-Means. DBSCAN does not assume a specific cluster shape and can identify outliers, while hierarchical clustering can provide a hierarchical representation of the data, allowing planners to explore different solutions at different levels of granularity. However, it is also more computationally intensive than K-Means, especially for large-scale datasets.

K-Means was chosen for this study due to its simplicity, computational efficiency, and scalability. Given the large dataset size and the need to evaluate many expansion scenarios, K- Means provides a good trade-off between solution quality and computation time. Moreover, since the main objective of the clustering is to identify reasonable candidate locations for new substations, and not to determine exact service boundaries, the limitations of K-Means are considered acceptable.

2.7. Substation Expansion via Clustering

To validate the robustness and effectiveness of the proposed optimization framework, several statistical and performance evaluation techniques were employed.

2.7.1. Hypervolume Metric for Pareto Front Evaluation

The quality of the obtained Pareto fronts across different parameter settings and initialization methods was evaluated using the hypervolume (HV) indicator. The hypervolume measures the volume in the objective space covered by the Pareto front relative to a defined reference point.

Let $P = \{(f_1^i, f_2^i)\}_{i=1}^{n}$ be a Pareto front obtained from the optimization algorithm, where f_1 and f_2 represent the objective values for each solution. Given a reference point $r = r_1 + r_2$, the hypervolume is computed as in Equation (7).

$$HV\ (P) = Volume\ \left(\cup_{i=1}^{n}[f_1^i, r_1]\ [f_2^i, r_2]\right) \tag{7}$$

This value reflects the convergence rate and diversity of solutions in the Pareto front.

Table 1. shows the hypervolume values obtained for the variation of the number of generations parameter. Based on these results, NSGA-II with fuzzy c-means initiation gives higher hypervolume than other initiations. For example, the hypervolume for the fuzzy c-means initiation is -74.72, while the random initiation cannot be calculated because it produces a very limited solution on the Pareto front.

Table 1. Hypervolume Calculation with Variance Parameter Number of Genes

Population	Gene	Trial To-		
		1	2	3
100	1000	74.70154	79.15128	80.4762
	500	77.4429	81.42.185	79.17449
	300	81.54913	82.19113	125.6473
	100	118.4976	119.0104	131.4511

2.7.2. Statistical Comparison of Initialization Methods

Three initialization strategies were compared: random, fuzzy-based, and mixed. The performance of these strategies was evaluated using non-parametric tests:

1. Friedman Test assesses overall differences across methods.
2. Wilcoxon Signed-Rank Test evaluates pairwise significance between methods.

The Friedman test statistic is Equation (8).

$$\chi_f^2 = \frac{12N}{k(k+1)}\left(\sum_{j=1}^{k} R_j^2\right) - 3N(k+1) \tag{8}$$

Where N is the number of experiments, k is the number of methods compared, and R_j is the average of method j.? If the Friedman test indicates significant, the Wilcoxon test compares paired differences d_i between two methods across n observations:

The Wilcoxon test evaluates ranked differences Equation (9).

$$T = \min(R^+, R^-) \tag{9}$$

Where $R^+$and $R^-$are the ranks of positive and negative differences, respectively [12],[25]

The Friedman test results show that there is a statistically significant difference between the initialization methods ($\chi^2 = 10.7935$, p-value = 0.004). Wilcoxon analysis further revealed that fuzzy initialization was significantly better than random initialization (p = 0.028, effect size = 0.56) and mixed initialization (p = 0.041, effect size = 0.48).

The implication of these results for planners is that the use of fuzzy initialization may result in better solutions than using other initialization methods. Although there is no significant difference between fuzzy and mixed initialization, fuzzy initialization is preferred due to its simplicity and computational efficiency.

Moreover, the lack of significant differences in average distance (p=0.5467) implies that planners can prioritize other factors in their decision-making alongside distance, such as cost or reliability considerations, without substantially compromising the spatial performance of the network.

3. Results and Discussion

3.1. Overview of Distribution Network Conditions

This study focuses on an electricity distribution system that currently serves approximately 416.472 customers via 732 distribution substations. Given budgetary and operational constraints, the utility prioritizes load redistribution before opting for infrastructure expansion. However, when redistribution proves insufficient to alleviate overload, the construction of new substations becomes a necessary intervention.

3.2. Statistical Comparison of Optimization Outcomes

To evaluate the effectiveness of substation expansion, a statistical analysis was performed across planning scenarios with additional substations, denoted by different values of. The two primary objectives analyzed are the average customer-to-substation distance and the variance in substation loads.

3.2.1. Friedman Test on Average Distance

The results of the Friedman test on the average distance of the customer resulted in a value $\chi^2 = 10.7935$, p-value = 0.5467. A p-value greater than 0.05 indicates that there is no statistically significant difference between group k in the context of the average distance of customers to substations. In other words, the increase in the number of substations does not provide a significant change in the distance of customer service when viewed from the median optimization results. Recent studies confirm that the Friedman test is a non-parametric statistical method that is effective in handling paired and non-normally distributed data, making it suitable for use in the comparative evaluation of multiple conditions [30].

 10.12928/ijio.v7i2.14108

3.2.2. Friedman Test on Load Variance

The Friedman test was also performed on substation load variance to determine whether the number of additional substations affected the equal distribution of load between substations. The test results showed $\chi^2 = 85.1279$, p-value = 4.3×10^{-13}. A p-value much smaller than 0.05 indicates that there is a statistically significant difference between groups in the number of additional substations in the context of load variance. Therefore, it is necessary to carry out a follow-up analysis with the signed-rank Wilcoxon test to find out which pairs have significant differences. Other studies have also shown that the Friedman test is effective in detecting significant differences in various conditions, especially when the data does not meet parametric assumptions [31].

3.2.3. Wilcoxon Signed-Rank Test for Pairwise Comparison

To Wilcoxon signed-rank test was performed for each pair in a row (k vs k+1) to find out in more detail the significant differences between variations in the number of substations in Table 2.

Table 2. Wilcoxon Signed Rank Test Score Calculation Results (K4-K16)

Comparison	Statistic (stat)	p-value
k=4 vs k=5	461.0	0.281
k=5 vs k=6	295.0	0.004
k=6 vs k=7	496.0	0.479
k=7 vs k=8	430.0	0.159
k=8 vs k=9	448.0	0.224
k=9 vs k=10	468.0	0.315
k=10 vs k=11	539.0	0.797
k=11 vs k=12	471.0	0.331
k=12 vs k=13	463.0	0.290
k=13 vs k=14	514.0	0.604
k=14 vs k=15	473.0	0.341
k=15 vs k=16	421.0	0.132

From these results, the comparison between k = 5 and k = 6 showed a p value below 0.05, namely p = 0.004. This indicates that the addition of substations from 5 to 6 has a significant impact in reducing load variance between substations. This can be caused by changes in substation configuration, which at k=6 value is able to be more optimal in covering customers with high loads that were previously unevenly distributed at k=5 in Table 2. Other comparisons show no significant differences, although some of them are close to the threshold.

3.3. Visualization-Based Interpretation

Visual analysis was performed to provide a further understanding of the optimization results obtained from NSGA-II with variations in the number of additional substations. Visualizations in the form of box plots are used to show the value distribution characteristics of two main objective functions the average distance of the customer to the substation and the load variance between substations for each k-value from 4 to 16. This visual interpretation complements the results of statistical analysis in Fig. 1, and provides insight into trends, variability, and the possibility of diminishing returns.

In addition to the box plot and knee point analysis, Fig. 3 shows Pareto fronts for the k = 4, k = 7, and k = 11 scenarios. These Pareto front plots illustrate the trade-off between average customer distance and load variance. As can be seen, the solution with the lowest average distance tends to have a higher load variance, and vice versa. Recent studies have shown that visualizations such as

Pareto plots and box plots effectively assist in understanding complex trade-offs in multi-objective optimization problems, as well as facilitate better decision-making [32]. Moreover, another study confirmed that visual interpretation can improve the general understanding of the distribution of results and support the identification of optimal solutions in a practical manner [33]–[35].

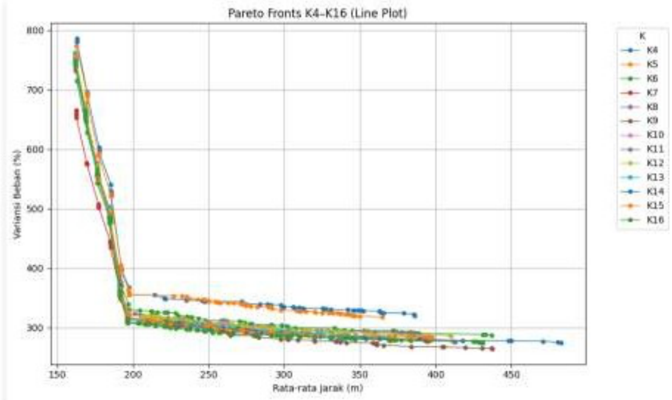

Fig. 3. Pareto Front at K = 4 - 16 Values

3.3.1. Performance Comparison of NSGA-II Fuzzy C-Means and Nearest Neighbor

At this stage, validation is carried out to measure the effectiveness of the NSGA-II optimization model with fuzzy c-means initiation compared to the commonly used approach, namely the Nearest Neighbor (KNN) algorithm with maximum substation load limitations. In the NSGA-II model with fuzzy c-means, initiation is done using fuzzy c-means with membership thresholding, followed by optimization using the NSGA-II algorithm with predefined parameters (number of generations = 500, mutation rate = 1%). While in the Nearest Neighbor (KNN) approach with Maximum Substation Load Limit, each customer is paired with the nearest distribution substation in a greedy manner. If the nearest substation is overloaded, the customer is transferred to the next closest substation until finding a feasible substation [36], [37].

3.3.2. Comparison of NSGA II and KNN

Table This section will explain the results of the comparison of the objective function in NSGA II with fuzzy c-means and Nearest Neighbor (KNN) with maximum substation load restrictions. Comparison results are shown in Fig. 4 and Table 3. Based on Fig. 4, the KNN algorithm produces the lowest distance of 149.43 meters and the highest of 164.91 meters, while NSGA II with fuzzy c-means produces the lowest result of 164.78 meters and the highest result of 265.55 meters. This happens because the KNN approach prioritizes the absolute closest distance without considering the overall load impact. While the NSGA II substation load variance produces more stable results with a value range of 339.97 to 575.76, while the KNN calculation produces a wider range with values of 773.13 to 1432.03.

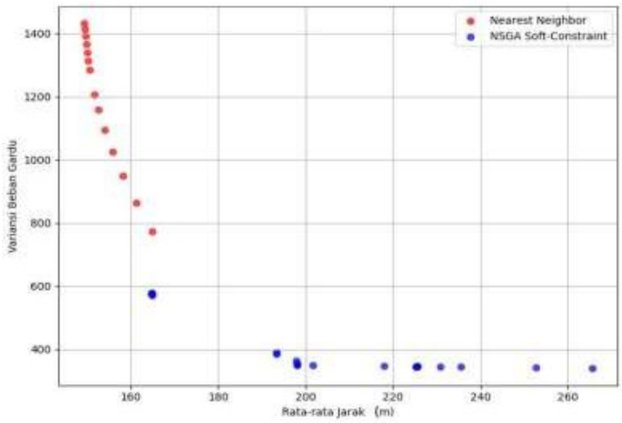

Fig. 4. Comparison of NSGA II and KNN Results

10.12928/ijio.v7i2.14108

Table 3. Comparison of NSGA II and KNN Results

Description	Distance			Variance		
	Min	Max	Stdv	Min	Max	Stdv
NSGA -II	164,79	265,55	29,21	339,97	578,75	92,62
KNN	149,44	164,91	4,91	773,13	1432,04	215,1207

From the above results, it can be concluded that although KNN is better in the aspect of local distance minimization, the solution provided by NSGA II with fuzzy c-means initiation is superior in the aspect of more even load distribution between substations, and produces a broader solution. This is in line with the initial objective of the research, which is to create a solution that is both efficient and balanced in terms of load [38], [39].

3.3.3. Distribution of Average Customer Distance to Substation

The box plot for the median value is shown in Fig. 5. This visualization provides information about the tendency of the middle value, the distribution of data (interquartile), and the outliers that appear in each variation in the number of additional substations (k).

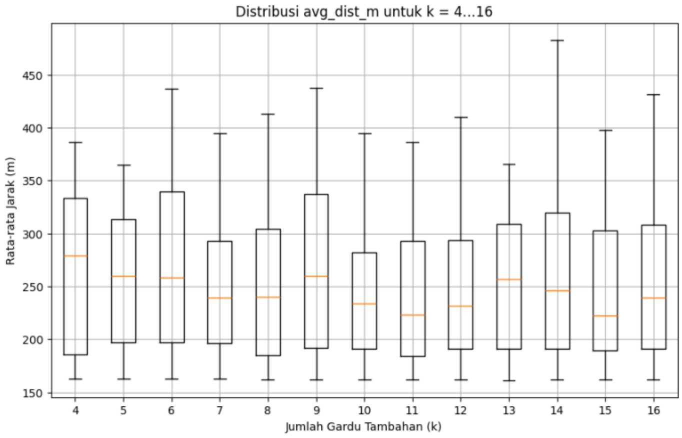

Fig 5. Box Plot Average Customer Distance to Substation for k = 4 to 16

The resulting box plot in Fig. 5 shows some important characteristics of the average distribution of customer distance to substations with the following results.

The median average distance tends to be stable in the range of 220–280 meters for the number of additional substations ranging from k = 4 to k = 16. This indicates the absence of a significant downward trend in average customer service distances as the number of additional substations increases. Recent research has shown that visual distribution analysis such as box plots are very useful in identifying trends and variability in data that are not apparent from descriptive statistics alone, as well as assisting in data-driven decision making [40].

The median decrease in the average distance of customers to substations is most noticeable when the number of additional substations increases from k = 4 to k = 7. After reaching k = 7, the increase in the number of additional substations provides only a minimal improvement in the median distance of the average, reflecting the phenomenon of diminishing returns. Another study confirmed that distribution visualization such as box plots can help identify optimal points and the phenomenon of diminishing returns in the network planning process [41], [42].

The top-down quartiles and whiskers in each group k show a significantly overlapping range. This finding is in line with the results of the Friedman test which did not show a significant median difference between the number of additional substations.

3.3.4. Load Variance Distribution Between Substations

The second box plot illustrates the distribution of load variance between substations for each value k. This visualization is presented in Fig. 6.

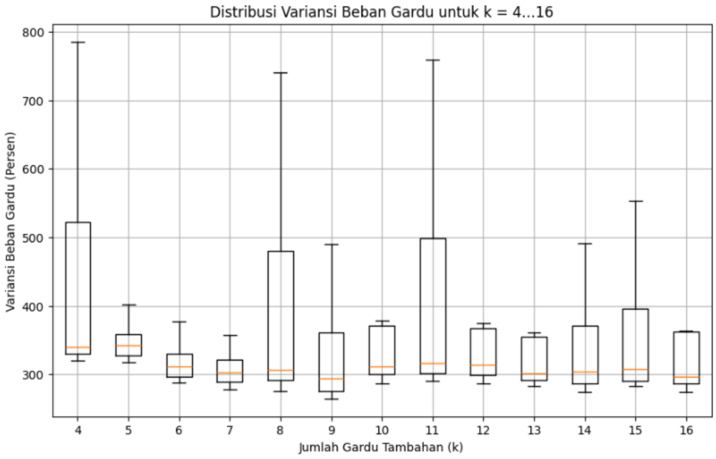

Fig. 6. Substation Load Variance Plot Boxes for k = 4 to 16

From the box plot "Substation Load Variance Distribution for k=4 to 16" several interpretations are obtained. High Variance Rate at k=4, the median value of the variance of substation charging at k=4 shows the highest number, which is around 340–350%. In addition, the wide IQR indicates that the solutions generated in this scenario are highly variable and unstable. This indicates that the addition of substations is still very much needed at this point.

Significant Decrease in k=5 to k=7, there was a sharp decrease in the median value of substation load variance between k=5 to k=7. This shows that the number of substations in that range is more effective in distribute the load evenly throughout the substation. Bump at k=8 and k=11, It was found that the median value and IQR increased again at k=8 and k=11. This phenomenon may be caused by suboptimal centroid configurations or overlapping service areas between substations which actually magnify the load imbalance. Plateau after k=7, Starting from k=7 to k=16, the median load variance tends to be stable in the range of 290–320%. The decrease in variance became smaller and smaller even though the number of substations continued to grow, suggesting that the additional benefits of new substations were saturated.

3.4. Knee Point Analysis

Knee point analysis was carried out to identify the optimal point for the number of additional substations (k) that provide maximum benefits to both objective functions, namely minimizing the average distance of customers to substations and minimizing load variance between substations. The concept of knee point refers to the point at which the addition of resources (in this case the number of substations) no longer provides a significant increase in performance, making it considered a cost- and outcome-efficient solution. Recent research has shown that visual and numerical methods for knee point detection are highly effective in identifying optimal boundaries in multi-objective performance curves, which supports more efficient decision-making [43].

3.4.1. Knee Point Detection Methodology

The method used in this analysis is a visual and numerical approach to the performance curve of each objective function to the value of k. Each point on the graph represents the median value of the objective function for each additional number of substations (k = 4 to 16). Other studies emphasize the importance of a combination of visual analysis and numerical methods to identify appropriate knee points, particularly in the context of infrastructure planning and resource optimization [44].

Knee point is defined as the point at which the rate of change of the objective function is significantly decelerated, characterized by a marked inflection or gradient change. In this case, two main curves are analyzed, namely the median curve of the average distance of customers to substations vs the addition of substations (k) and the minimization of load variance between substations vs the addition of substations (k). The analysis was also strengthened by the elbow/knee point visualization approach and the normalization of the two objective functions for the combined analysis. Research supports the use of a combination of visual and analytical approaches to detect this optimal point, which is important in the development of data-driven decision-making models [45].

3.4.2. Knee Point on the Average Customer Distance

The results of the elbow method analysis on the objective function of the average distance of the customer to the substation are shown in Fig. 7. shows that the knee point is located at a certain number of substations, which represents the optimal point in terms of maximum efficiency and benefits. The elbow method has been visually and numerically proven to be an effective approach in determining the optimal point in a multi-objective performance curve, especially in the context of infrastructure planning where the number of substations needs to be selected efficiently [46]. Recent research also emphasizes that the use of geometric approaches such as maximum distance and curve graph analysis can improve the accuracy of elbow point detection, as well as assist in strategic decision making [47].

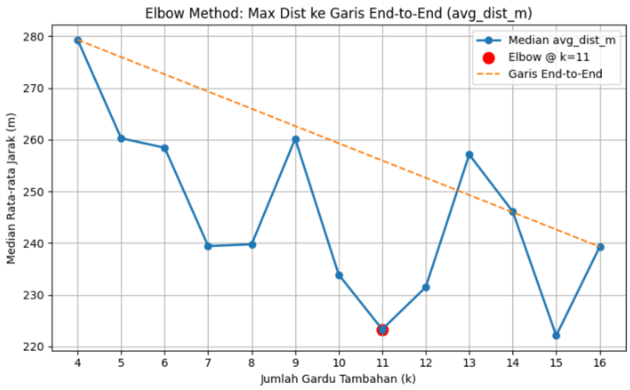

Fig. 7. Median Curve of Average Customer Distance to Substation to Number of Additional Substations (k = 4 to 16)

The blue curve depicts the change in the median value of the average distance of the customer to the substation as the number of additional substations increases from k = 4 to k = 16. On top of the curve, the geometric elbow method is applied, which is to measure the maximum distance between points on the curve and a straight line connecting the starting and ending points (k = 4 and k = 16). The point that has the maximum distance from the line is marked as the elbow point. From these results, it was obtained that the elbow point was located at k = 11, which means that the addition of substations to this point provided the most significant improvement to the average decrease in the distance of the customer to the substation. After k = 11, the performance improvement begins to slow down and shows a tendency to stabilize (diminishing returns). This makes k = 11 a structural optimum point in the context of customer service distance efficiency. Nonetheless, the average distance of customers to substations does not show a completely linear downward trend. There are several fluctuations, for example at k = 9 and k = 13, which show an increase in the average distance compared to the previous point. This indicates that not all substation addition configurations result in consistent performance improvements. The causative factor can come from the position of the centroid resulting from clustering that is less strategic or the uneven distribution of customers.

3.4.3. Knee Point at Substation Load Variance

Furthermore, an analysis was carried out on the median curve of load variance between substations. A visualization of the curve is shown in Fig. 8 next.

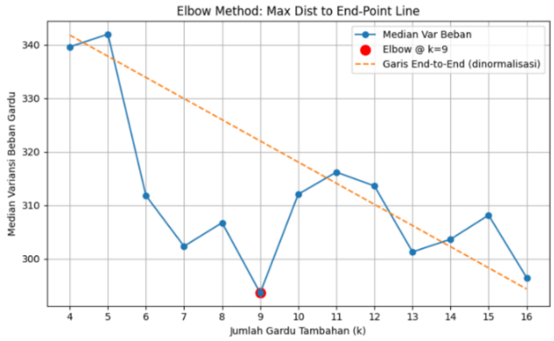

Fig. 8. Median Curve of Substation Load Variance to Number of Additional Substations (k = 4 to 16)

The blue curve depicts the change in the median value of substation load variance for each additional number of substations (k = 4 to 16). The dotted orange line represents a straight line connecting the start and end points of the curve (k = 4 to k = 16), used as a reference to detect elbow points geometrically. Based on the analysis, the elbow point was identified at k = 9, marked by a red circle on the graph.

This elbow point indicates the position where the decrease in load variance begins to slow down significantly. In other words, the addition of substations from k = 4 to k = 9 results in a noticeable improvement in the equal distribution of the load between substations, but after k = 9, the addition of substations tends to result in smaller or less stable improvements. This can be seen from the curve trend that begins to fluctuate and is inconsistent after k = 9, even at some points (for example, k = 10 to 13), there is a return to the median value of substation load variance.

3.4.4. Combined Knee Point (Normalization of Two Objective Functions)

To get a more comprehensive perspective, an analysis was carried out on the combined normalization value of the average distance of customers to substations and the minimization of load variance between substations. Normalization is carried out so that both objective functions are on an equal scale, so that the knee point can be evaluated simultaneously. Based on Fig. 9, it can be seen that from k = 4 to k = 7 there was a sharp decline in both objective functions. This shows that the addition of substations in the early stages provides great benefits, both in reducing the service distance and in evening the load between substations. After k = 7, both curves show non-uniform fluctuations, the average value of the customer's distance to the substation increases at some point (e.g., k = 9 and 13), and the value of load variance also shows instability (e.g., k = 10 to 12). This shows an indication of a trade-off between the two objective functions. In addition, fluctuations in value indicate diminishing returns and even potential inefficiencies due to excessive or unstrategic addition of substations.

10.12928/ijio.v7i2.14108

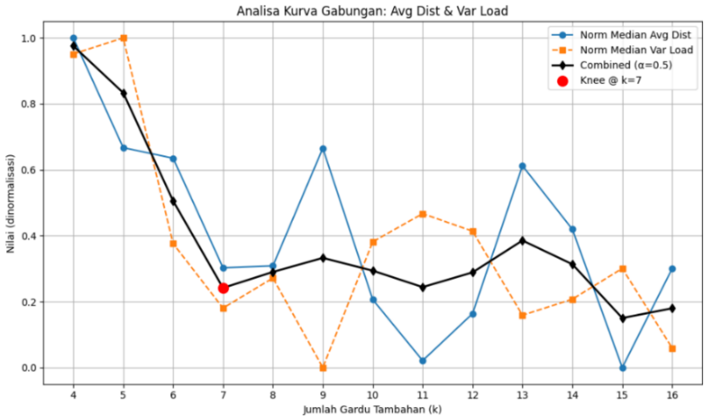

Fig. 9. Normalization Curve of Combined Two Objective Functions Against Additional Substations (k = 4 to 16)

Meanwhile, the combined curve, calculated as the average of the two objective functions, shows a minimum value at k = 7, which is also marked as the combined knee point. The addition of substations with k = 7 is the most balanced number of additional substations in optimizing both objective functions simultaneously.

3.5. Heatmap Analysis Visualization of Relationships Between Objective Functions

To deepen our understanding of the relationship between the number of additional substations (k) and the two main objective functions, the average distance of the customer to the substation and the variance of substation loads, visualizations were carried out using heatmaps and 3D scatter plots.

3.5.1. Visualization of Objective Function Median Heatmap

The heatmap visualization in Fig. 10 illustrates the median values of both objective functions for each of the k values (4 to 16). The horizontal axis represents the value of k, while the vertical axis represents the metric being analyzed. Darker colors indicate better performance.

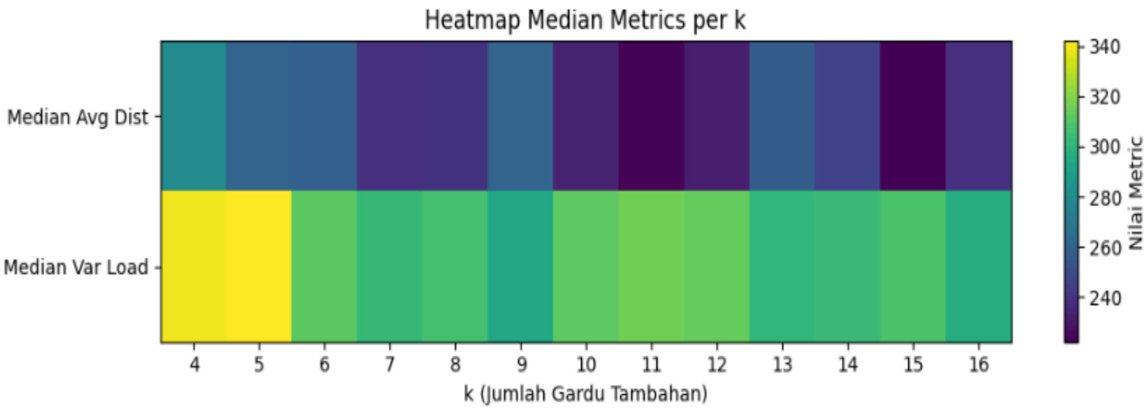

Fig. 10. *Heatmap* Median Average Customer Distance to Substation and Load Variance to Number of Additional Substations (k = 4–16)

For the average distance of customers to substations, there was a darker color gradient from k=4 to k=7, indicating a significant decrease in the average distance of customers. Meanwhile, for the variance of substation loads, a decrease also occurred from k=4 to k=7, but after that there were slight fluctuations. The color does not become much lighter, indicating that the addition of substations after k=7 provides a relatively constant improvement. Both lines of heatmaps show a synchronous trend to k=7, then stabilization or even diminishing returns afterwards. This reinforces the previous results that k = 7 be the critical point for both metrics.

4. Conclusion

This study addresses the Multi-Objective Covering Problem (MOCP) in substation expansion planning. By integrating spatial fuzzy assignment, NSGA-II, and K-means clustering, the proposed approach minimizes the service distance and balances the load effectively. The evaluation shows that the addition of new substations up to k = 7 results in optimal network performance. Significantly, this framework offers a scalable solution to real-world problems while considering the trade-off between cost, performance, and operational constraints. The results of this study provide practical guidance for distribution network planners in making informed decisions on substation placement and resource allocation.

For future research, (1) integration of a comprehensive cost model, (2) simulation-based scenario analysis to address demand uncertainty, and (3) exploration of multi-criteria decision-making methods to select optimal solutions are recommended. In addition, it is also necessary to conduct a comparative study of the proposed framework with state-of-the-art algorithms in the literature and consider spatial constraints in system planning.

Supplementary Materials: It is not applicable.

Author Contribution: Each author contributed equally to this paper. The final paper is read and approved by all authors.

Funding: It is not applicable.

Acknowledgment: It is not applicable.

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

References

[1] R. S. F. Ferraz, and A. C. R. Medina, "A Comprehensive Multi-objective Optimization Framework for Balanced Distribution System Planning," *J. Control. Autom. Electr. Syst.*, vol. 36, no. 1, pp. 117–133, Feb. 2025, doi: 10.1007/s40313-024-01134-5.

[2] M. Kumar, A. Soomro, W. Uddin, and L. Kumar, "Optimal Multi-Objective Placement and Sizing of Distributed Generation in Distribution System: A Comprehensive Review," *Energies*, vol. 15, no. 21, p. 7850, Oct. 2022, doi: 10.3390/en15217850.

[3] M. A. Brahami, M. Dahane, M. Souier, and M. Sahnoun, "Sustainable capacitated facility location/network design problem: a Non-dominated Sorting Genetic Algorithm based multiobjective approach," *Ann. Oper. Res.*, vol. 311, no. 2, pp. 821–852, Apr. 2022, doi: 10.1007/s10479-020-03659-9.

[4] J. Huang, "Optimal substation capacity planning method in high-density load areas considering renewable energy," *Energy Reports*, vol. 9, pp. 1520–1530, Sep. 2023, doi: 10.1016/j.egyr.2023.04.104.

[5] M. E. Ţiboacă-Ciupăgeanu and D. A. Ţiboacă-Ciupăgeanu, "Optimal Substation Placement: A Paradigm for Advancing Electrical Grid Sustainability," *Sustainability*, vol. 16, no. 10, p. 4221, May 2024, doi: 10.3390/su16104221.

[6] S. Vahedi, M. Banejad, and M. Assili, "GIS-Based Substation Expansion Planning," *IEEE Syst. J.*, vol. 15, no. 1, pp. 959–970, Mar. 2021, doi: 10.1109/JSYST.2020.3010565.

[7] X. Tian, H. Yang, Y. Ge, and T. Yuan, "Site Selection and Capacity Determination of Electric Hydrogen Charging Integrated Station Based on Voronoi Diagram and Particle Swarm Algorithm," *Energies*, vol. 17, no. 2, p. 418, Jan. 2024, doi: 10.3390/en17020418.

 10.12928/ijio.v7i2.14108

[8] N. Zheng, K. Chen, Q. Cai, Y. Li, P. Shi, and K. Xiang, “Active Distribution Network Substation Planning Method Considering Complementation of Load Characteristic,” in *2021 6th Asia Conference on Power and Electrical Engineering (ACPEE)*, Apr. 2021, pp. 1730–1734, doi: 10.1109/ACPEE51499.2021.9437113.

[9] A. Bosisio, A. Berizzi, M. Merlo, A. Morotti, and G. Iannarelli, “A GIS-Based Approach for Primary Substations Siting and Timing Based on Voronoi Diagram and Particle Swarm Optimization Method,” *Appl. Sci.*, vol. 12, no. 12, p. 6008, Jun. 2022, doi: 10.3390/app12126008.

[10] R. J. Kuo, C. C. Hsu, T. P. Q. Nguyen, and C. Y. Tsai, “Hybrid multi-objective metaheuristic and possibilistic intuitionistic fuzzy c-means algorithms for cluster analysis,” *Soft Comput.*, vol. 28, no. 2, pp. 991–1008, Jan. 2024, doi: 10.1007/s00500-023-09367-3.

[11] S. E. Hashemi, F. Gholian-Jouybari, and M. Hajiaghaei-Keshteli, “A fuzzy C-means algorithm for optimizing data clustering,” *Expert Syst. Appl.*, vol. 227, p. 120377, Oct. 2023, doi: 10.1016/j.eswa.2023.120377.

[12] D.-D. Yang, M. Mei, Y.-J. Zhu, X. He, Y. Xu, and W. Wu, “Coverage Optimization of WSNs Based on Enhanced Multi-Objective Salp Swarm Algorithm,” *Appl. Sci.*, vol. 13, no. 20, p. 11252, Oct. 2023, doi: 10.3390/app132011252.

[13] R. J. Kuo, Y. R. Zheng, and T. P. Q. Nguyen, “Metaheuristic-based possibilistic fuzzy k-modes algorithms for categorical data clustering,” *Inf. Sci. (Ny).*, vol. 557, pp. 1–15, May 2021, doi: 10.1016/j.ins.2020.12.051.

[14] L. Tang, S. Chen, and Q. Li, “Optimizing Maintenance Resource Scheduling and Site Selection for Urban Metro Systems: A Multi-Objective Approach to Enhance System Resilience,” *Systems*, vol. 12, no. 7, p. 262, Jul. 2024, doi: 10.3390/systems12070262.

[15] T. T. Teo, “Optimization of Fuzzy Energy-Management System for Grid-Connected Microgrid Using NSGA-II,” *IEEE Trans. Cybern.*, vol. 51, no. 11, pp. 5375–5386, Nov. 2021, doi: 10.1109/TCYB.2020.3031109.

[16] Q. T. Vu, P. T. Nguyen, T. H. Nguyen, T. T. B. Huynh, V. C. Trinh, and M. Gidlund, “Striking the perfect balance: Multi-objective optimization for minimizing deployment cost and maximizing coverage with Harmony Search,” *J. Netw. Comput. Appl.*, vol. 232, p. 104006, Dec. 2024, doi: 10.1016/j.jnca.2024.104006.

[17] N. Chakrabarty, K. M. Sullivan, and D. B. Lopes da Silva, “Time-based redeployment of multi-class nodes for reliable wireless sensor network coverage,” *Comput. Ind. Eng.*, vol. 197, p. 110549, Nov. 2024, doi: 10.1016/j.cie.2024.110549.

[18] A. Majeed and S. O. Hwang, “A Multi-Objective Coverage Path Planning Algorithm for UAVs to Cover Spatially Distributed Regions in Urban Environments,” *Aerospace*, vol. 8, no. 11, p. 343, Nov. 2021, doi: 10.3390/aerospace8110343.

[19] Y. Tong, L. Lin, L. Tian, Z. Wang, W. Wu, and J. Wu, “Coverage optimization and node minimization in WSNs: an enhanced hybrid PSO approach with spatial position encoding,” *Sci. Rep.*, vol. 15, no. 1, p. 25332, Jul. 2025, doi: 10.1038/s41598-025-09849-4.

[20] M. Wasid and R. Ali, “A frequency count approach to multi-criteria recommender system based on criteria weighting using particle swarm optimization,” *Appl. Soft Comput.*, vol. 112, p. 107782, Nov. 2021, doi: 10.1016/j.asoc.2021.107782.

[21] R. F. Liji, P. Sreejaya, M. Sasikumar, and K. J. Seelan, “An Improved Image Inpainting Technique Using Fuzzy Hard C-Means Algorithm,” *Advances in Commun. Syst. and Networks*, 2020, doi: 10.1007/978-981-15-3992-3_21.

[22] F. Wang, Y. Geng, and H. Zhang, “An improved fuzzy C-means clustering algorithm based on intuitionistic fuzzy sets,” *Proceedings of the 9th Int. Conf. on Computer Engineering and Networks*, 2020, doi: 10.1007/978-981-15-3753-0_32.

[23] W. Zhao, X. Bian, and X. Mei, “An Adaptive Multi-Objective Genetic Algorithm for Solving Heterogeneous Green City Vehicle Routing Problem,” *Appl. Sci.*, vol. 14, no. 15, p. 6594, Jul. 2024, doi: 10.3390/app14156594.

[24] R. Xu and D. WunschII, “Survey of Clustering Algorithms,” *IEEE Trans. Neural Networks*, vol. 16, no. 3, pp. 645–678, May 2005, doi: 10.1109/TNN.2005.845141.

[25] M. Shalaby, A. Mohammed, and S. Kassem, “Supervised Fuzzy C-Means Techniques to Solve the Capacitated Vehicle Routing Problem,” *Int. Arab J. Inf. Technol.*, 2021, doi: 10.34028/iajit/18/3A/9.

[26] F. Fanian and M. Kuchaki Rafsanjani, “CFMCRS: Calibration fuzzy- metaheuristic clustering routing scheme simultaneous in on-demand WRSNs for sustainable smart city,” *Expert Syst. Appl.*, vol. 211, p. 118619, Jan. 2023, doi: 10.1016/j.eswa.2022.118619.

[27] C. Gao, M. Quan, C. Yang, P. Gao, and J. Zhang, “Multi-Objective Resource Optimization Allocation of Road-Domain Integrated Energy System Considering Hydrogen Energy Storage,” in *2024 International Conference on HVDC (HVDC)*, Aug. 2024, pp. 263–270, doi: 10.1109/HVDC62448.2024.10723022.

[28] Z.-A. Sergio, P. B. Tatiana, B. D. Stalin, L. G. Edwin, and F. John Fredy, “Optimal subtransmission switching using a reliability simulation-based multi-objective optimization model,” *Electr. Power Syst. Res.*, vol. 210, p. 108068, Sep. 2022, doi: 10.1016/j.epsr.2022.108068.

[29] K. Djouzi and K. Beghdad-Bey, “A Review of Clustering Algorithms for Big Data,” in *2019 International Conference on Networking and Advanced Systems (ICNAS)*, Jun. 2019, pp. 1–6, doi: 10.1109/ICNAS.2019.8807822.

[30] H. Erick and M. Dirik, “Data Analysis for Smart Grid and Communication Technologies,” *Int. Conf. Sci. Innov. Stud.*, vol. 1, no. 1, pp. 331–342, Apr. 2023, doi: 10.59287/icsis.621.

[31] V. Konstantinos, Y. Karras, J. Kohlhammer, M. Steiger, D. Tzovaras, and E. Gounopoulos, “Enhanced Visual Analytics Services for the Optimal Planning of Renewable Energy Resources Installations,” 2014, pp. 330–339. doi: 10.1007/978-3-662-44722-2_35.

[32] D. V. Nga, O. H. See, D. N. Quang, C. Y. Xuen, and L. L. Chee, “Visualization Techniques in Smart Grid,” *Smart Grid Renew. Energy*, vol. 03, no. 03, pp. 175–185, 2012, doi: 10.4236/sgre.2012.33025.

[33] H. Elba, H. Hegazy, J. Zhang, I. M. Mahdi, and H. M. Hassan, “Optimizing steel structures for solar panels: integrating artificial intelligence and web-based Decision support systems for enhanced efficiency and sustainability,” *Innov. Infrastruct. Solutions*, 2025, doi: 10.1007/s41062-024-01831-9.

[34] T. Nguyen, Y. Ahn, and B. Kim, “Integrated digital-twin-based decision support system for relocatable module allocation plan: Case study of relocatable modular school system,” *Applied Sciences*. 2025, doi: 10.3390/app15042211.

[35] D. Petraki, I. Gazoulis, M. Kokkini, M. Danaskos, and I. Travlos, “Digital tools and decision support systems in agroecology: Benefits, challenges, and practical implementations,” *Agronomy*. 2025, doi: 10.3390/agronomy15010236.

[36] R. P. Kumar and G. Karthikeyan, “A multi-objective optimization solution for distributed generation energy management in microgrids with hybrid energy sources and battery storage system,” *J. Energy Storage*, 2024, doi: 10.1016/j.est.2023.109702.

[37] A. A. K. Al-Sahlawi, S. M. Ayob, C. W. Tan, H. M. Ridha, and D. M. Hachim, “Optimal design of grid-connected hybrid renewable energy system considering electric vehicle station using improved multi-objective optimization: Techno-Economic Perspectives,” *Sustainability*. doi: 10.3390/su16062491.

[38] R. Nuvvula, E. Devaraj, and K. T. Srinivasa, “A comprehensive assessment of large-scale battery integrated hybrid renewable energy system to improve sustainability of a smart city,” *Energy sources, part a Smart City*, 2025, doi: 10.1080/15567036.2021.1905109.

[39] K. Shafiei, A. Seifi, and M. T. Hagh, “A novel multi-objective optimization approach for resilience enhancement considering integrated energy systems with renewable energy, energy storage, energy sharing, and demand-side management,” *J. Energy Storage*, 2025, doi: 10.1016/j.est.2025.115966.

[40] Q. Wang, A. Tuohy, M. Ortega-Vazquez, M. Bello, E. Ela, and R. Philbrick, “Quantifying the value of probabilistic forecasting for power system operation planning,” *Applied Energy*. 2023, doi: 10.1016/j.apenergy.2023.121254.

[41] M. Garau and B. N. Torsæter, “A methodology for optimal placement of energy hubs with electric vehicle charging stations and renewable generation,” *Energy*. 2024, doi: 10.1016/j.energy.2024.132068.

[42] M. C. Muñoz, M. A. Peñalba, and A. E. S. González, “Analysis of aggregated load consumption forecasting in short, medium and long term horizons using dynamic mode decomposition,” *Energy Reports*. 2024, doi: 10.1016/j.egyr.2024.06.040.

[43] H. Ma, Y. Zhang, S. Sun, T. Liu, and Y. Shan, “A comprehensive survey on NSGA-II for multi-objective optimization and applications,” *Artif. Intell. Rev.*, 2023, doi: 10.1007/s10462-023-10526-z.

[44] X. Wang, H. Wang, B. Bhandari, and L. Cheng, “AI-empowered methods for smart energy consumption: A review of load forecasting, anomaly detection and demand response,” *International Journal of Precision Engineering and Manufacturing-Green Technology*. 2024, doi: 10.1007/s40684-023-00537-0.

[45] S. C. Burmeister, D. Guericke, and G. Schryen, “A memetic NSGA-II for the multi-objective flexible job shop scheduling problem with real-time energy tariffs” *Flexible Services and Manufacturing Journal*. 2024, doi: 10.1007/s10696-023-09517-7.

[46] B. Cao, Y. Zhang, J. Zhao, X. Liu, and L. Skonieczny, “Recommendation based on large-scale many-objective optimization for the intelligent internet of things system,” *IEEE Internet of Things Journal*, 2021, doi: 10.1109/JIOT.2021.3104661.

[47] Z. Yang, S. Bi, and Y. J. A. Zhang, “Online trajectory and resource optimization for stochastic UAV-enabled MEC systems,” *IEEE Trans. Wirel.*, 2022, doi: 10.1109/TWC.2022.3142365.