Optimal Gateway Placement In Clustered Sensor Network Using Fuzzy C Means Algorithm

Chibuzor Henry Amadi[1]
Department of Electrical and Electronic Engineering
Imo State University Owerri, Imo State

Akaniyene Benard Obot[2]
Department of Electrical /Electronic Engineering,
University of Uyo, Akwa Ibom State, Nigeria
akaniyeneobot@uniuyo.edu.ng

Kufre Monday Udofia[3]
Department of Electrical /Electronic Engineering,
University of Uyo, Akwa Ibom State, Nigeria
kmudofia@uniuyo.edu.ng

Kingsley M. Udofia[4]
Department of Electrical /Electronic Engineering,
University of Uyo, Akwa Ibom State, Nigeria

***Abstract*— In this paper, optimal gateway placement in clustered sensor network using Fuzzy C Means algorithm is presented. The Fuzzy C Means (FCM) method is particularly useful when overlapping occurs among the clusters in sensor network. Also, before optimal gateway placement can be done, the optimal number of clusters required is determined based on the spatial distribution of the sensor nodes. As such, in this study Elbow method and the Silhouette method are employed to determine the optimal number of clusters in the network. The Fuzzy C Means algorithm is then applied in each case to determine the optimal placement of the gateway. Some simulations suing Python program were conducted in a sensor network with 5000 nodes that are randomly distributed in a spatial coverage area with dimension of $800m \times 800m$. The results show that the Silhouette method yielded five clusters and the application of the Fuzzy C Means on the five clusters gave the optimal gateway location coordinates as (410.69, 430.49), (160.32, 180.88), (160.28, 600.64), (650.19, 610.92) and (620.77, 200.39). On the other hand, The results show that the Elbow method yielded four clusters and the application of the Fuzzy C Means on the four clusters gave the optimal gateway location coordinates as (215.09,205.19), (215.30,605.48), (600.22,608.14) and (600.10,200.22). In all, results of the Fuzzy C Means is greatly affected by the choice of the number of clusters which is in turn dependent on the method used to determine the optimal number of clusters and the spatial distribution of the sensor nodes in the network coverage area considered in the study.**

Keywords*— *Optimal Gateway Placement, Clustered Sensor Network, Fuzzy C Means Algorithm, Elbow Method, Silhouette Method, Sensor Node

1.0 Introduction

Wireless sensor network (WSN) has become the main backbone of today's Internet of Things (IoT) and smart systems applications [1,2]. It is also generally believed that the applications of wireless sensors will continue to grow as new areas of applications will continue to emerge and as technologies continue to advance [3]. As regards communication in wireless sensor network, the energy consumption is affected by the communication distance [4,5]. Also, clustering has been employed over the years to address the issue of reordering the overall communication distance within clustered sensor nodes so as to reduce the mean energy consumption in the network [6,7,8,9]. In doing so, the optimal number of clusters must be determined based on the spatial distribution of the sensor nodes within the network coverage area [8,10]. Next, the optimal placement of the gateways with each of the identified clusters must be done [11,12]. The determination of optimal number of clusters can be done using gap statistics method, Elbow method or the silhouette method [13,14,15]. On the other hand, K-means or Fuzzy C means among other methods can be used to determine the optimal gateway placement [16,17,18].

Generally, crisp clustering method is useful when sensor nodes are distributed on different clusters without overlapping. However, if overlapping occurs among the clusters, Fuzzy C means is more useful [19,20]. In this case, each sensor node is assigned to clusters, and each cluster must know the degree to which the assigned nodes belong to it. Accordingly, in this paper, the details of the Fuzzy C mean approach is presented. In the simulation, the Elbow method and the silhouette methods are used for the determination of the optimal number of clusters in the network.

2. Methodology

In this section the details of optimal gateway location based on Fuzzy C Means method are presented.

2.1 Determination of the Optimal Gateway Location Using Fuzzy C Means Method

Let a cluster be defined as C_j in which a sensor node S_i belong to, that is: $C_j \in S_i$. Let w_{ij} denote the weight in which S_i belong to C_j. Note that S_i can also belong to another cluster say C_i. The application of weight of belonging is possible since Fuzzy logic can handle ambiguity or numbers in-between 0 and 1. Let a set of sensor node be defined as $S = \{S_1, S_2, S_3, \dots, S_N\}$, where, N denotes the number of sensor nodes; and a set of clusters is defined as $\mathbb{C} = C_1, C_2, C_3, \dots, C_k$, where k, denotes the number of clusters. Let d be the dimension of the data, then $S_i = \{S_{i1}, S_{i2}, S_{i3}, \dots, S_{id}\}$. The membership weight w_{ij} must be assigned to each clusters and the values varies between 0 and 1.

Supposed four clusters exist as C_1, C_2, C_3, and C_4; and three sensor nodes are distributed within these clusters with various weights of belonging as shown in Table 1, then, w_{ij}, can be computed for the distribution, where i denotes the sensor node, and j denotes the cluster. Thus, $w_{11} = 0.02$, $w_{12} = 0.94$, $w_{13} = 0.03$, and $w_{14} = 0.01$.

Table 1: Sensor nodes distribution with weights within clusters

	C_1	C_2	C_3	C_4
S_1	0.02	0.94	0.03	0.01
S_2	0.01	0.10	0.08	0.09
S_3	0.97	0.01	0.01	0.01

Note that the distribution of the membership weight is valid for the following conditions:

1. The sum of all the weights in any given row r_i must be equal to 1. Mathematically, $\sum_{j=1}^{k} w_{ij} = 1$
2. Each cluster C_j must contain a weight between 0 and 1. Mathematically, $0 < \sum_{i=1}^{m} w_{ij}$

FCM clustering algorithm is presented in Algorithm 1

Algorithm 1: Gateway optimal location using FCM

1: **Begin**
2: Define the sensor dataset, S
3: Define number of clusters, k
4: Define membership weights w_{ij}, $1 \le i \le m, 1 \le j \le k$
5: Define the clusters centers, C_j;
6: Assign random values to all membership weights w_{ij}, $1 \le i \le m, 1 \le j \le k$
7: Use the Fuzzy pseudo-partition to compute the cluster center
8: Assign w_{ij} using Fuzzy pseudo-partition
9: **if** the cluster centroid changes **then**
10: **goto** 7
11: else
12 **End**

The sum squared error (SSE) can be computed based on Equation 1

$$SSE = \sum_{j=1}^{k} \sum_{i=1}^{m} w_{ij}^{P} \cdot d(x_i, C_j); \qquad 1 < P < \infty \tag{1}$$

Where, P is the weight inverse, d is the distance between x_i and C_j

In Algorithm 1, all definition and initialization are performed in Step 1 to Step 6. Step 7 computes the centroid based on Equation (2);

$$C_j = \frac{\sum_{1}^{m} w_{ij}^{P} \cdot x_i}{\sum_{1}^{m} w_{ij}^{P}} \tag{2}$$

Note that if $P = 0$, then Equation (2) is reduced to K-Means algorithm. In Step 8 of Algorithm 1, w_{ij} is updated using Equation (3);

$$w_{ij} = \frac{\left({}^{1}/_{d(x_i, C_j)} \right)^{\frac{1}{P-1}}}{\sum_{q=1}^{k} \left({}^{1}/_{d(x_i, C_q)} \right)^{\frac{1}{P-1}}} \tag{3}$$

If $P = 2$, then the simplified version of Equation (4) is:

$$w_{ij} = \frac{\left({}^{1}/_{d(x_i, C_j)} \right)}{\sum_{q=1}^{k} \left({}^{1}/_{d(x_i, C_j)} \right)} \tag{4}$$

From Equation 4, it is seen that the value of the membership weight w_{ij} depends on the distance between x_i and C_j. To demonstrate the application of this method, let $P = 2$, the sensor dataset $S = \{(1,2), (2,3), (9,4), (10,1)\}$, $k = 2$. The dataset S can be distributed as shown in Table 2.

Table 2: The initial distribution of the sample dataset, S

	X	Y
S_1	1	2
S_2	2	3
S_3	9	4
S_4	10	1

Step 1: Begin // Begin the process of applying or implementing the Fuzzy C Means algorithm

Step 2: Let the membership weights w_{ij} be randomly initialized as shown in Table 3.

Table 3: The randomly initialized membership weights w_{ij} of the sample dataset, S

	C_1	C_2
S_1	0.4	0.6
S_2	0.88	0.12
S_3	0.41	0.59
S_4	0.27	0.73

Step 3: Apply Equation (2) to obtain C_j. First, consider the denominator for the first cluster:

$$\sum_{1}^{4} w_{i1}^2 = (0.4)^2 + (0.88)^2 + (0.41)^2 + (0.27)^2 = 1.18$$

The denominator for the second cluster is computed as:

$$\sum_{1}^{4} w_{i2}^2 = (0.6)^2 + (0.12)^2 + (0.59)^2 + (0.73)^2 = 1.25$$

$$C_{11} = \frac{(0.4)^2 \times 1 + (0.88)^2 \times 2 + (0.41)^2 \times 9 + (0.27)^2 \times 10}{1.18} = \frac{3.97}{1.18} = 3.38$$

$$C_{12} = \frac{(0.4)^2 \times 2 + (0.88)^2 \times 3 + (0.41)^2 \times 4 + (0.27)^2 \times 1}{1.18} = \frac{3.39}{1.18} = 2.88$$

$$C_{21} = \frac{(0.6)^2 \times 1 + (0.12)^2 \times 2 + (0.59)^2 \times 9 + (0.73)^2 \times 10}{1.25} = \frac{8.85}{1.25} = 7.08$$

$$C_{22} = \frac{(0.6)^2 \times 2 + (0.12)^2 \times 3 + (0.59)^2 \times 4 + (0.73)^2 \times 1}{1.25} = \frac{2.69}{1.25} = 2.15$$

The cluster centers can now be expressed as:

$$C_1 = [C_{11} \quad C_{12}] = [3.38 \quad 2.88]$$
$$C_2 = [C_{21} \quad C_{22}] = [7.08 \quad 2.15]$$

Step 4: Apply Equation (3). to compute w_{ij}. First, the Euclidean distances between the sensor node location dataset and the cluster centers are computed using $d = \sqrt{(C_{j1} - S_{i1})^2 + (C_{j2} - S_{i2})^2}$, and this results in the matrix given in Table 4.

Table 4: The Euclidean distances between the sensor node location dataset and the cluster centers

	C_1	C_2
S_1	2.54	6.03
S_2	1.38	5.10
S_3	5.73	2.71
S_4	6.88	3.19

Then, the weights can be computed as follows;

$$w_{11} = \frac{1/2.54}{1/2.54 + 1/6.03} = \frac{0.39}{0.56} = 0.7$$
$$w_{12} = \frac{1/6.03}{1/2.54 + 1/6.03} = \frac{0.17}{0.56} = 0.3$$
$$w_{21} = \frac{1/1.38}{1/5.10 + 1/1.38} = \frac{0.72}{0.91} = 0.79$$
$$w_{22} = \frac{1/5.10}{1/5.10 + 1/1.38} = \frac{0.20}{0.91} = 0.21$$
$$w_{31} = \frac{1/5.73}{1/2.71 + 1/5.73} = \frac{0.17}{0.54} = 0.32$$
$$w_{32} = \frac{1/2.71}{1/2.71 + 1/5.73} = \frac{0.37}{0.54} = 0.68$$
$$w_{41} = \frac{1/6.88}{1/3.19 + 1/6.88} = \frac{0.15}{0.46} = 0.32$$
$$w_{42} = \frac{1/3.19}{1/3.19 + 1/6.88} = \frac{0.31}{0.45} = 0.68$$

After seven iterations, the final membership weights metric is given in Table 5. The final weights shows that S_1 and S_2 belong to C_1, while S_3 and S_4 belong to C_2.

Table 5 : The membership weights metric

	C_1	C_2
S_1	0.88	0.12
S_2	0.94	0.06
S_3	0.17	0.83
S_4	0.18	0.82

3 Result and Discussion

3.1 Experimental Setup

The simulations for the model were conducted for a sensor network that has 5000 sensor nodes distributed randomly within an area of $= L(x, y) = 800m \times 800m = 640,000m^2$, as shown in Figure 4.1.

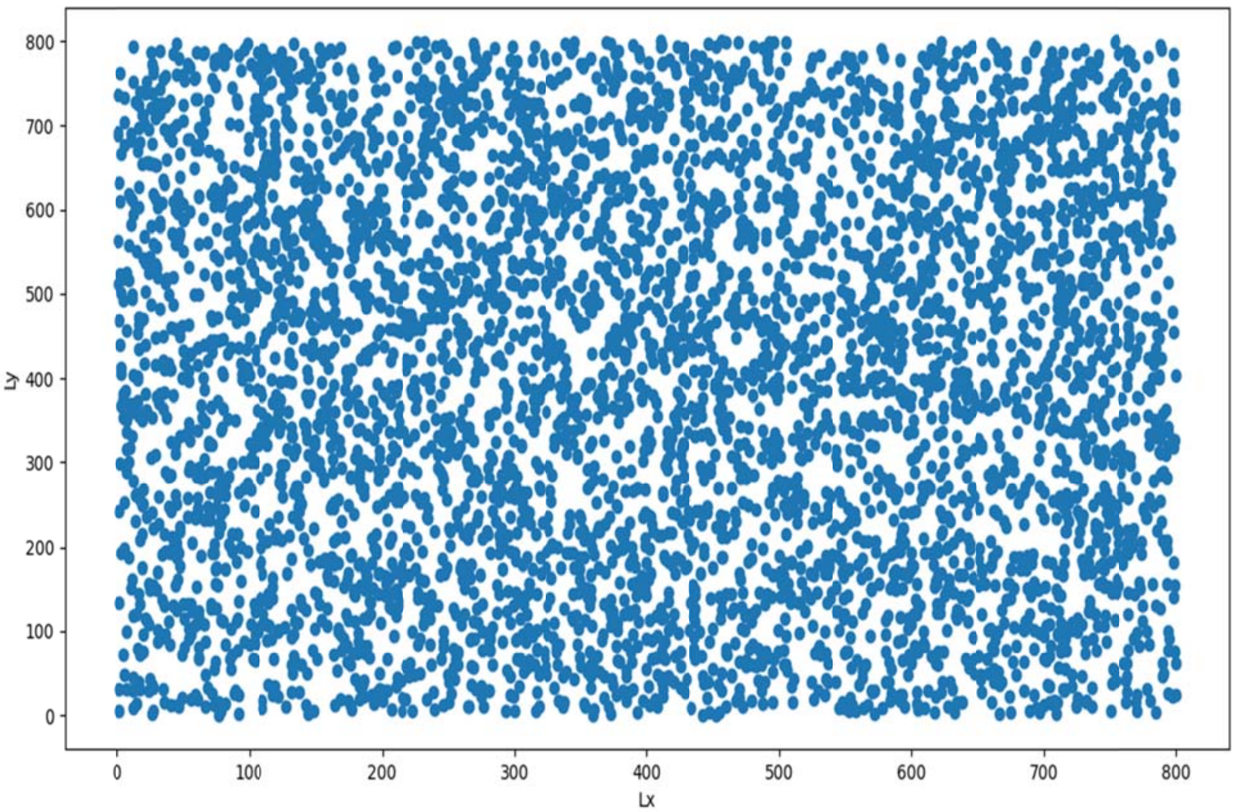

Figure .1: Sensor nodes distribution on the simulation area

3.2 Evaluation of Gateway Location using the Silhouette Method and Fuzzy C Means Method

In the first simulation, the Silhouette method is used to determine the optimal number of clusters while Fuzzy C Means is used to determine the optimal cluster head locations. The result of this combination is presented in Figure 2. From the results presented, the coordinates of the five cluster heads as determined by the Fuzzy C Means are (410.69, 430.49), (160.32, 180.88), (160.28, 600.64), (650.19, 610.92) and (620.77, 200.39).

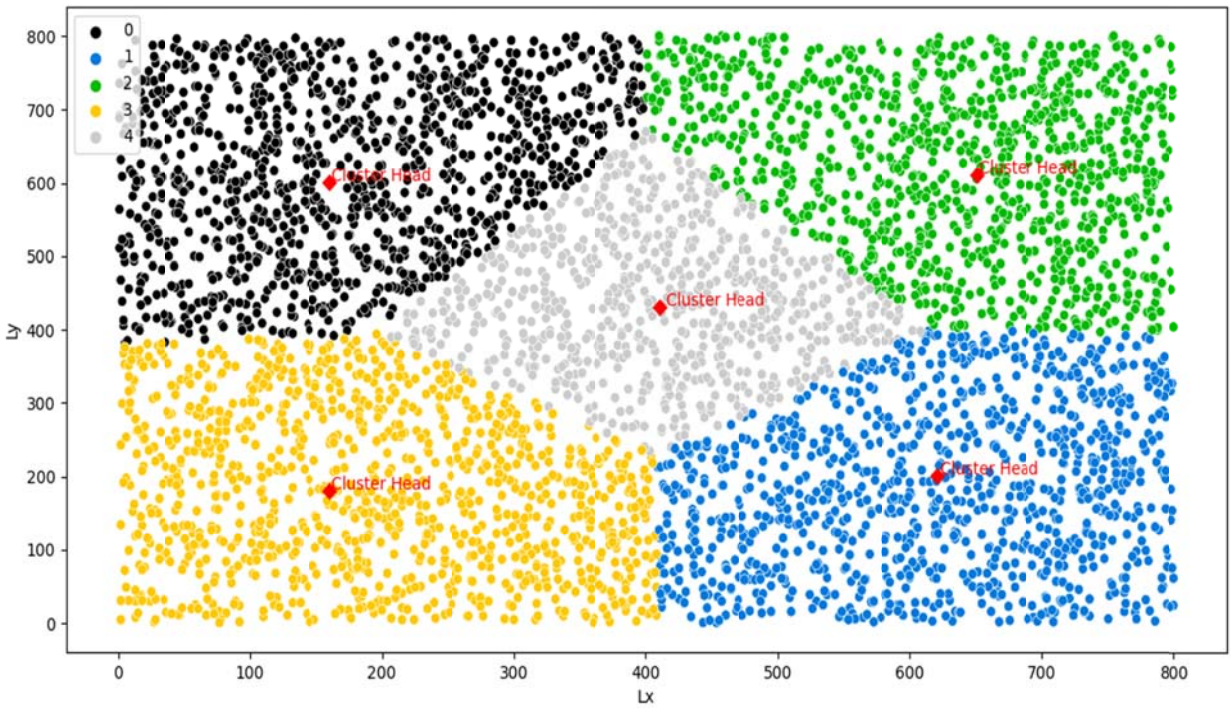

Figure 2: Optimal Gateway Location Using Silhouette Method and Fuzzy C Means

3.3 Evaluation of Gateway Location using Elbow Method and Fuzzy C Means Method

In the second simulation, the Elbow method is used to determine the optimal number of clusters while Fuzzy C Means is used to determine the cluster head location. Recall that Elbow method suggested that the optimal number of cluster is four. The result of this combination is presented in Figure 3. From the results presented, the coordinates of the four cluster heads as determined by Fuzzy C Means are (215.09, 205.19), (215.30, 605.48), (600.22, 608.14) and (600.10, 200.22).

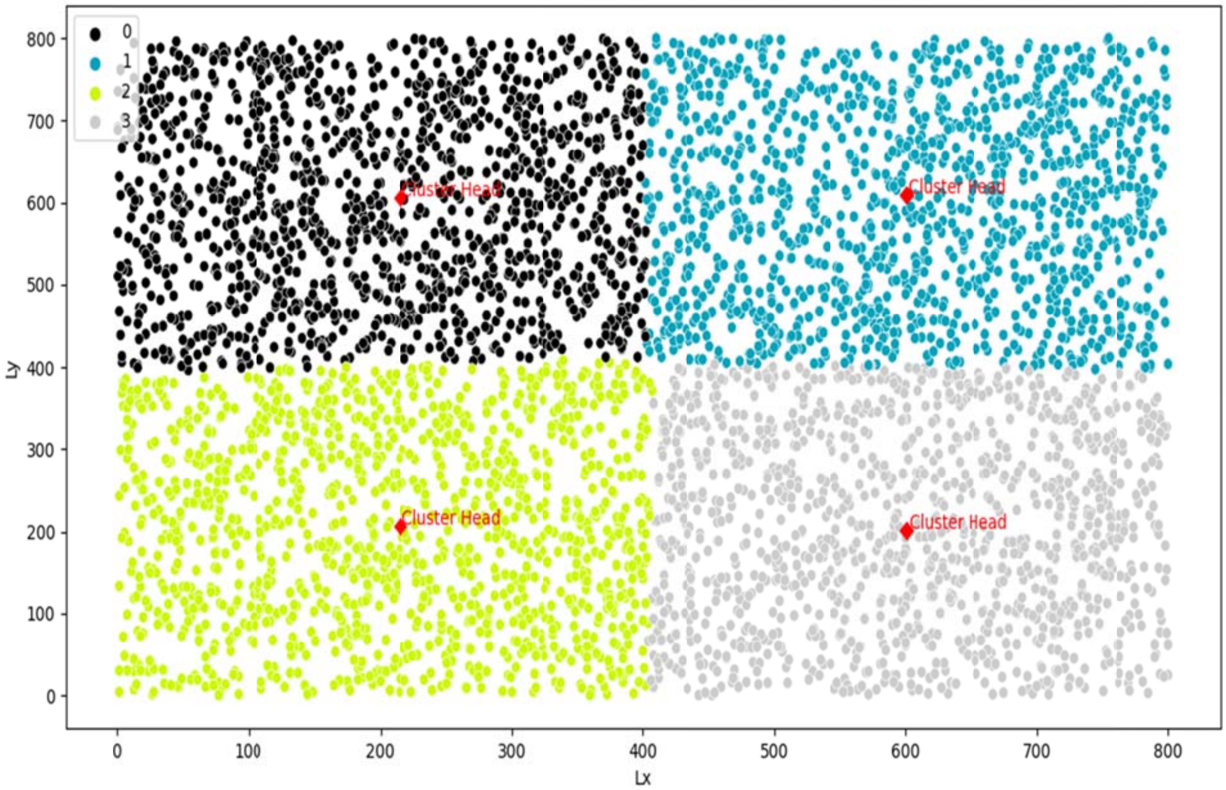

Figure 3: Optimal gateway location using Elbow method and Fuzzy C Means

4 Conclusion

This work presents a study where Fuzzy C Means approach is used to determine the optimal cluster heads locations in a sensor network. In the study, Elbow method and Silhouette method are used to determine the optimal number of clusters in the network based on the random distribution of the nodes. Next, the Fuzzy C Means method is used to determine the location of the cluster heads based on the number of clusters determined by the Elbow method and the Silhouette method. In all, the results showed that the Elbow method and Silhouette method gave different number of clusters. However, the optimal location of the Fuzzy C Means is dependent on the number of clusters returned by the Elbow method and also the number returned by the Silhouette method.

References

1. Nourildean, S. W., Hassib, M. D., & Mohammed, Y. A. (2022). Internet of things based wireless sensor network: a review. *Indones. J. Electr. Eng. Comput. Sci*, *27*(1), 246-261.
2. Majid, M., Habib, S., Javed, A. R., Rizwan, M., Srivastava, G., Gadekallu, T. R., & Lin, J. C. W. (2022). Applications of wireless sensor networks and internet of things frameworks in the industry revolution 4.0: A systematic literature review. *Sensors*, *22*(6), 2087.
3. Pahlavan, K., & Krishnamurthy, P. (2021). Evolution and impact of Wi-Fi technology and applications: A historical perspective. *International Journal of Wireless Information Networks*, *28*, 3-19.
4. Chowdhury, S. M., & Hossain, A. (2020). Different energy saving schemes in wireless sensor networks: A survey. *Wireless Personal Communications*, *114*(3), 2043-2062.
5. Tang, L., Lu, Z., & Fan, B. (2020). Energy efficient and reliable routing algorithm for wireless sensors networks. *Applied Sciences*, *10*(5), 1885.
6. Shahraki, A., Taherkordi, A., Haugen, Ø., & Eliassen, F. (2020). A survey and future

directions on clustering: From WSNs to IoT and modern networking paradigms. *IEEE Transactions on Network and Service Management*, *18*(2), 2242-2274.

7. Moussa, N., Hamidi-Alaoui, Z., & El Belrhiti El Alaoui, A. (2020). ECRP: an energy-aware cluster-based routing protocol for wireless sensor networks. *Wireless Networks*, *26*, 2915-2928.
8. Umbreen, S., Shehzad, D., Shafi, N., Khan, B., & Habib, U. (2020). An energy-efficient mobility-based cluster head selection for lifetime enhancement of wireless sensor networks. *Ieee Access*, *8*, 207779-207793.
9. Idrees, A. K., Alhussaini, R., & Salman, M. A. (2020). Energy-efficient two-layer data transmission reduction protocol in periodic sensor networks of IoTs. *Personal and Ubiquitous Computing*, 1-20.
10. Sahoo, B. M., Amgoth, T., & Pandey, H. M. (2020). Particle swarm optimization based energy efficient clustering and sink mobility in heterogeneous wireless sensor network. *Ad Hoc Networks*, *106*, 102237.
11. Matni, N., Moraes, J., Oliveira, H., Rosário, D., & Cerqueira, E. (2020). LoRaWAN gateway placement model for dynamic Internet of Things scenarios. *Sensors*, *20*(15), 4336.
12. Piltyay, S., Bulashenko, A., & Demchenko, I. (2020, October). Wireless sensor network connectivity in heterogeneous 5G mobile systems. In *2020 IEEE international conference on problems of infocommunications. Science and Technology (PIC S&T)* (pp. 625-630). IEEE.
13. Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2021). A quantitative discriminant method of Elbow point for the optimal number of clusters in clustering algorithm. *Eurasip Journal on Wireless Communications and Networking*, *2021*(1), 1-16.
14. Saputra, D. M., Saputra, D., & Oswari, L. D. (2020, May). Effect of distance metrics in determining k-value in k-means clustering using Elbow and silhouette method. In *Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)* (pp. 341-346). Atlantis Press.
15. Abdullah, D., Susilo, S., Ahmar, A. S., Rusli, R., & Hidayat, R. (2022). The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data. *Quality & Quantity*, *56*(3), 1283-1291.
16. Hassan, A. A. H., Shah, W. M., Othman, M. F. I., & Hassan, H. A. H. (2020). Evaluate the performance of K-Means and the Fuzzy C Means algorithms to formation balanced clusters in wireless sensor networks. *Int. J. Electr. Comput. Eng*, *10*(2), 1515-1523.
17. Molokomme, D. N., Chabalala, C. S., & Bokoro, P. N. (2021). Enhancement of advanced metering infrastructure performance using unsupervised K-means clustering algorithm. *Energies*, *14*(9), 2732.
18. Zhu, B., Bedeer, E., Nguyen, H. H., Barton, R., & Henry, J. (2020). Improved soft-k-means clustering algorithm for balancing energy consumption in wireless sensor networks. *IEEE Internet of Things Journal*, *8*(6), 4868-4881.
19. Askari, S. (2021). Fuzzy C Means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: Review and development. *Expert Systems with Applications*, *165*, 113856.
20. Mohammadrezapour, O., Kisi, O., & Pourahmad, F. (2020). Fuzzy C Means and K-means clustering with genetic algorithm for identification of homogeneous regions of groundwater quality. *Neural Computing and Applications*, *32*, 3763-3775.