# Evaluation Of Multiple Optimal Cluster And Gateway Placement Implementation In Iot Sensor Network Using Silhouette Method And K-Means Algorithm

[1] **Ezema D.C.**
Department of Electrical and Electronic Engineering
[3]State University of Medical and Applied Sciences(SUMAS),Igbo-Eno,Enugu State, Nigeria
donatus.ezema@sumas.edu.ng

[2]**Ifeagwu E.N.**
Department of Electrical and Electronic Engineering
Federal University Otuoke, Bayelsa State.
ORCID:0009-0005-7448-0187
ifeagwuen@fuotuoke.edu.ng

[3] **Geku Diton**
Department of Electrical and Electronic Engineering
Federal University Otuoke, Bayelsa State.
Email: gekudd@fuotuoke.edu.ng

***Abstract***— **In this paper, evaluation of multiple optimal cluster and gateway placement implementation in IoT sensor network using Silhouette method and K-means algorithm is presented. The IoT sensor network used in the simulation had 2000 sensor nodes with network coverage area of 800 m by 800 m. The Silhouette score plot showed that the optimum number of clusters is 5. As such, K value of 5 was used in the k-means algorithm for optimal gateway placement. The K-means was implemented three times on the same dataset. The results show that the mean distance for the entire network was 139 m in the first implementation, 140 m in the second implementation, and 141 m in the third implementation. The maximum sensor to gateway distance in the entire network was 269 m in the first implementation, 267 m in the second implementation, and 274 in the third implementation. Again, since the energy consumption in sensor network is proportional to the communication distance, it means that the third implementation will have the first sensor node to die due to power outage given the long distance of 274 m. On the other hand, the battery life of the critical node with the longest distance in the second implementation will stay longer than that of the third implementation since the distance is 267m. In all, it can be concluded that the K-means approach studied can be effective in optimal gateway placement. However, several implementations may be required and the implementation with the best communication distance result can be adopted.**

***Keywords— Gateway Placement, Clustering, Internet of Things (IoT), Sensor Network, Silhouette Method, K-Means Algorithm***

## 1. Introduction

Wireless sensors are very vital components of the present day smart system [1,2,3]. The sensor enable networks and systems to monitor, measure and communicate essential parameters of the environment or system and hence facilitate automated controls, remote controls and various kinds of intelligent operations via the interconnected network of thing; sensors, actuators and system being monitored [4,5]. In any case, sensors are usually resource constrained; many of the sensors are powered using battery [6,7]. In such case, effort is made to enhance the battery life span through various forms of energy efficient mechanisms [8,9].

Over the years, clustering has been identified as one of the methods to enhance battery life span be reducing the transmission path length through the grouping of the sensors together in clusters that have a base station or gateway to relay the communications from the various sensor to the ultimate destination which in some cases is the server [10,11,12]. Again, implementation of clustering can be done using various approaches but the focus of this work is to use K-means for the clustering [13,14]. The K-means algorithm is used to determine the optimal potion of the gateways in each of the clusters such the mean distance of all the sensors in each of the cluster is minimal [15,16]. This will in turn minimize the energy consumed in the process of transmitting data to the gateway.

While K-means can effectively achieve the purpose, it has been noticed that when the network is densely populated, it is possible to have different gateway location for different implementations of the K-means on the same dataset. Accordingly, this study seek to assess to what extent such variation can affect the mean distance between the gateway and the sensors and also the maximum distance of the farthest sensor in the network. Such parameters are essential in determining the network life span for battery powered wireless sensor network.

## 2. Methodology

### 2.1 The description of the IoT sensor network and the research process

The IoT sensor network considered in this study has Ns number of sensors distributed randomly over a rectangular network coverage area with dimensions given as L x W, where L is the length and W is the width. The sensor nodes are expected to be grouped into clusters and are also expected to communicate wirelessly to the internet via the gateways or base stations in the sensor's respective clusters. The first task in this work is to determine the optimum number of clusters that is suitable to the given sensor network and also use that number to determine the optimal location coordinates of the gateways within each of the clusters.

Secondly, the study seeks to assess how multiple implementations of the optimal gateway placement using the same dataset and number of clusters can affect the mean distance and the maximum distance between the sensor and the gateway in the entire network. The required number of clusters for the network is determined using the Silhouette score method while K-means algorithm is used to determine the optimal location for the gateway. The flow diagram used for the research procedure is given in Figure 1.

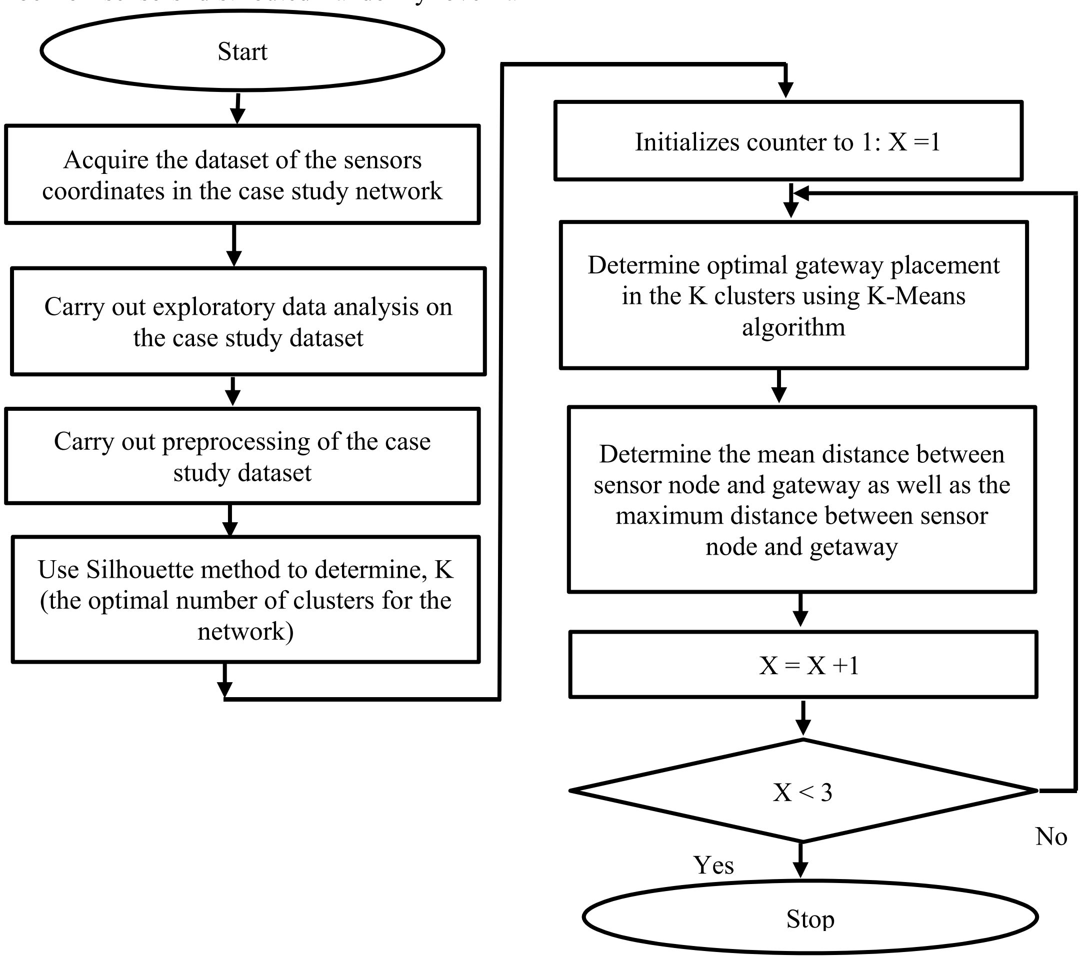


Figure 1: The flow diagram used for the research procedure

### 2.2 Data preprocessing and the determination of the optimum number of clusters using Silhouette score

The sensor nodes coordinates are the key data used in the study. Hence, the X and Y coordinates of the sensor nodes are normalized to have values between 0 and 1 using the minmax normalization approach. Also, descriptive statistical analysis is conducted to check for outliers, missing data and also determine the mean and standard deviation of the dataset.

The Silhouette method is used to determine the optimum number of clusters applicable to the case study network. The flow diagram of the Silhouette method is presented in Figure 2.

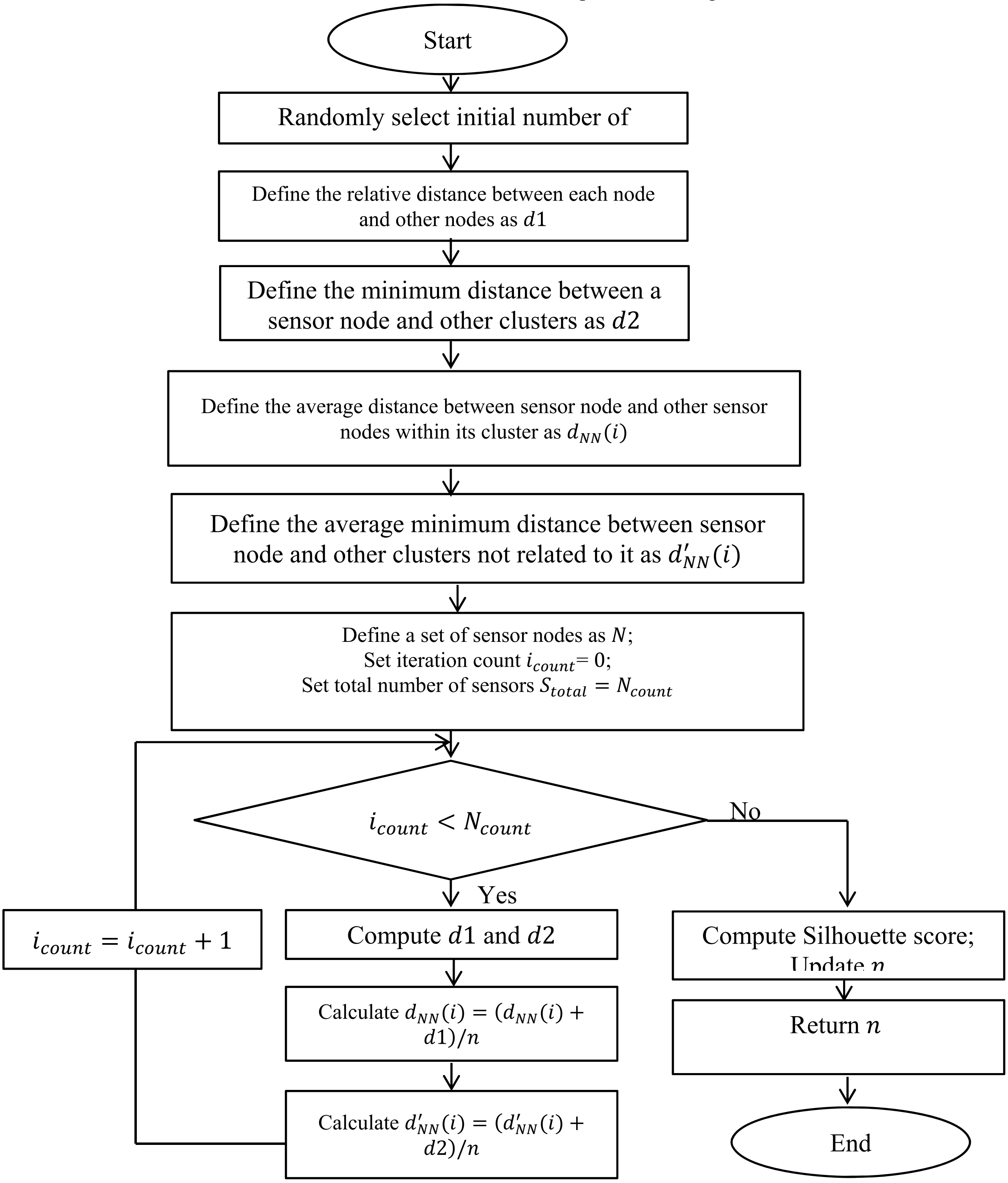


Figure 2: The flowchart for determining the number of clusters using Silhouette score.

**2.3The optimal gateway placement based on K-Means Algorithm**

The K-means algorithm is used to determine the best locations for the gateways such that the mean distance between the sensor nodes and the gateway is minimized. In the K-means algorithm, the Euclidian distance is used for the computation of the dissimilarities measure and the Within-Cluster Sum of Squares (WCSS) is used to compute the total within-cluster variation.

Now, consider a sensor network with K clusters where $xC_n, yC_n$ are the X and Y coordinates of the nth clusters respectively and $x_{j,n}, y_{j,n}$ are the X and Y coordinates of the sensor j in cluster n (where n ≤ K), then, the Euclidian distance, $d_{j,n}$ of each of the sensors from their respective centroids is computed as follows;

$$d_{j,n} = \sqrt{\left(xC_n - x_{j,n}\right)^2 + \left(yC_n - y_{j,n}\right)^2} \quad for\ n = 1,2,3, \ldots, k \qquad (1)$$

The total within-cluster variation which is a measure of how compact (or how good)the clustering is can be defined in respect of the WCSS where;

$$WCSS = \sum_{n=1}^{k}\left(d_{j,n}\right) = \sum_{j=1}^{j=k}\left(\sum_{x \in n, j=1}^{j=N_n}\left(\sqrt{\left(xC_n - x_{j,n}\right)^2 + \left(yC_n - y_{j,n}\right)^2}\right)\right) \qquad (2)$$

Where $N_n$denoted the number of sensor nodes assigned to cluster $n$.

For each sensor node in each cluster n, the distance, $d_{j,n}$to the gateway or centroid is computed. Then, the means distance, $d_{mean(n)}$ for each cluster, n and the mean distance , $d_{mean}$for the entire network are computed, as well as the maximum distance, $d_{max(n)}$for each the cluster and the maximum distance, $d_{max}$for the entire network, as follows;

$$d_{mean(n)} = \left(\frac{1}{Nn}\right)\left(\sum_{j=1}^{j=Nn}\left(d_{j,n}\right)\right) \qquad (3)$$

$$d_{mean} = \left(\frac{1}{K}\right)\left(\sum_{n=1}^{n=K}\left(d_{mean(n)}\right)\right) \qquad (4)$$

$$d_{max(n)} = Maximum\left(d_{j,n}\right) for\ j = 1,2,3, \ldots, \frac{1}{Nn} \qquad (5)$$

$$d_{max} = Maximum\left(d_{max(n)}\right)\ for\ n = 1,2,3, \ldots, k \qquad (6)$$

The K-means algorithm is presented in the flow diagram of Figure 3.

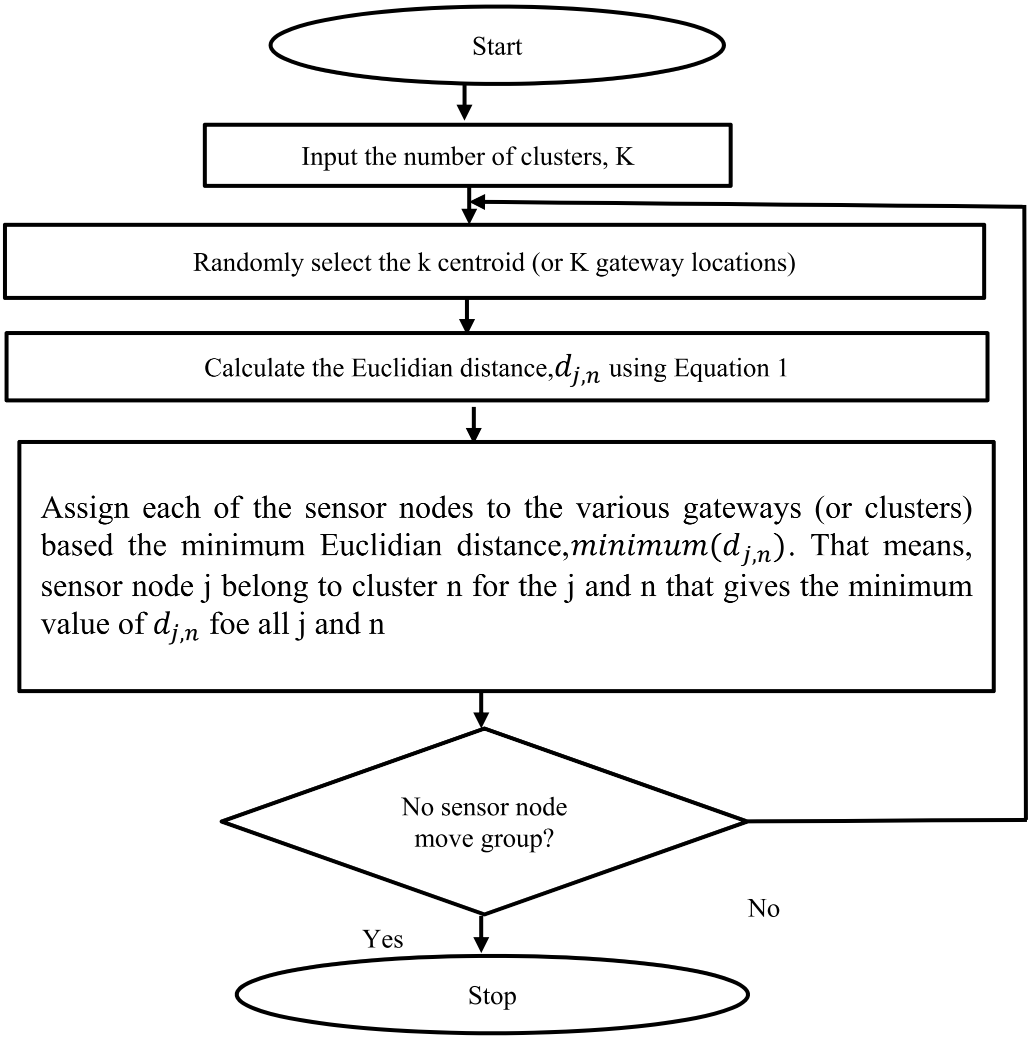


Figure 3: The flow diagram of the K-means algorithm

### 3. Results and Discussion

The IoT sensor network used in the simulation has 2000 sensor nodes with network coverage area of 800 m by 800 m. The result of the Silhouette score plot for optimum cluster number selection is presented in Figure 4. Based on Figure 4, the optimum number of clusters by the Silhouette score method is 5. As such, K value of 5 was used in the k-means algorithm for optimal gateway placement. The K-means was implemented three times on the same dataset.

The X and Y coordinates of the centroid of clusters in each of the three implementations of the K-means are presented in Table 1 while the scatter plot of the coordinates of the centroids is presented in Figure 8. The results show that in each implementation of the K-means algorithm the selected gateway locations for the five clusters are different. Although some of the getaway locations are repeated but the

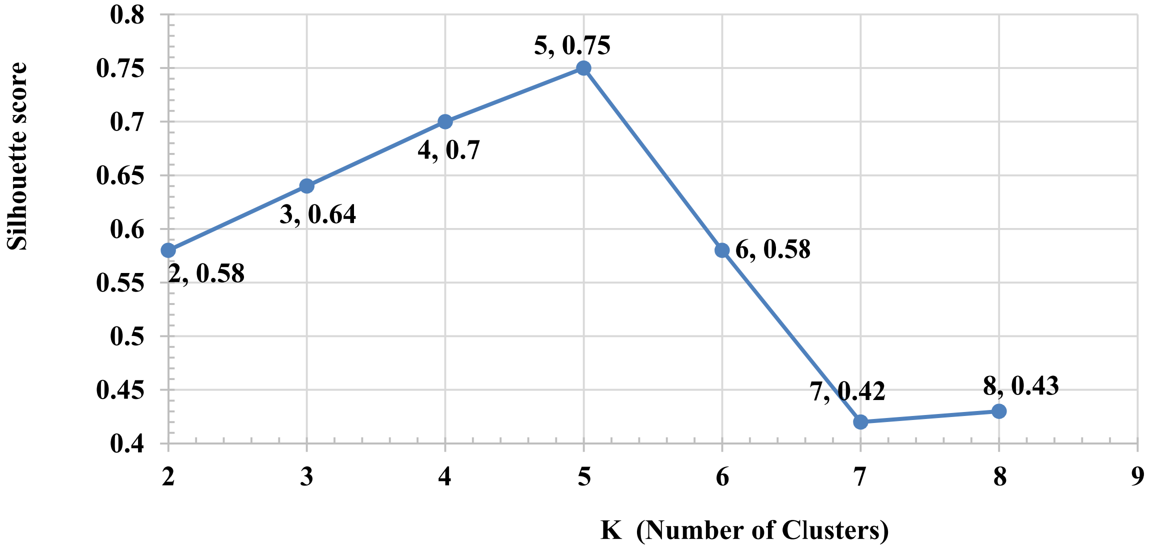


Figure 4: The result of the Silhouette score plot for optimum cluster number selection

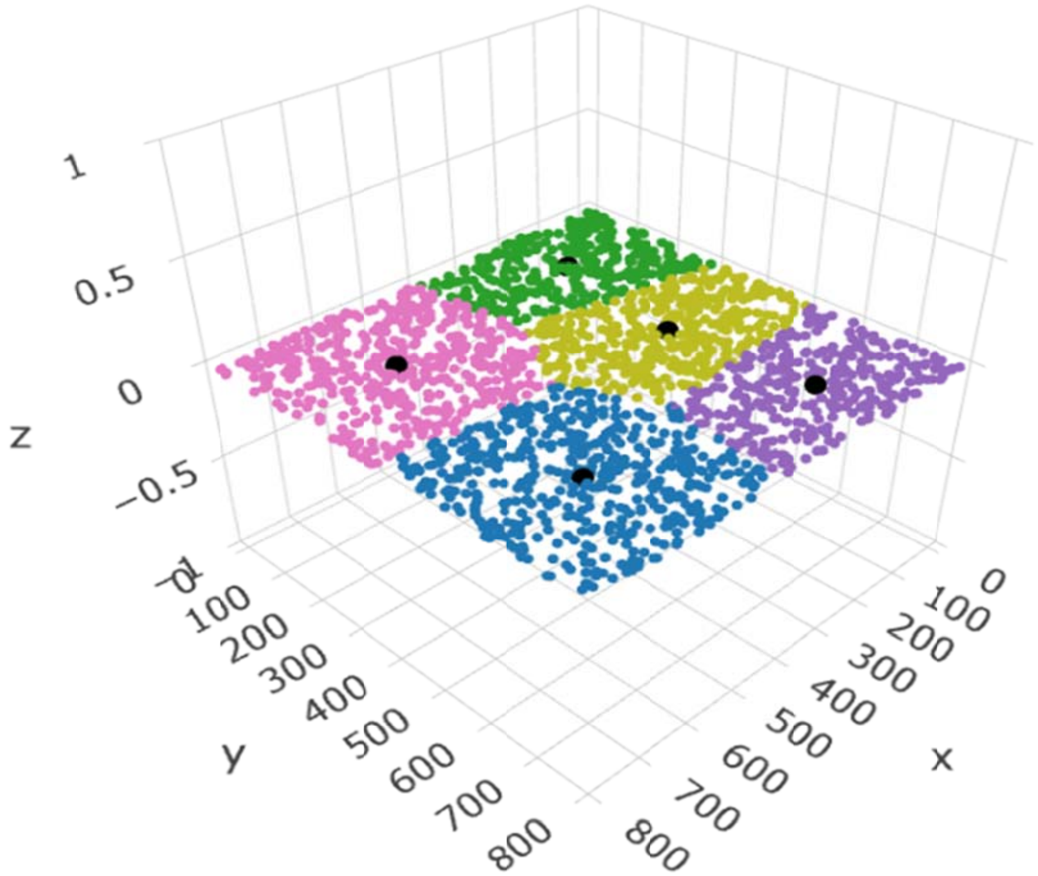


Figure 5:The scatter plot of the first implementation (implementation 1) of the K-means clustering on the dataset

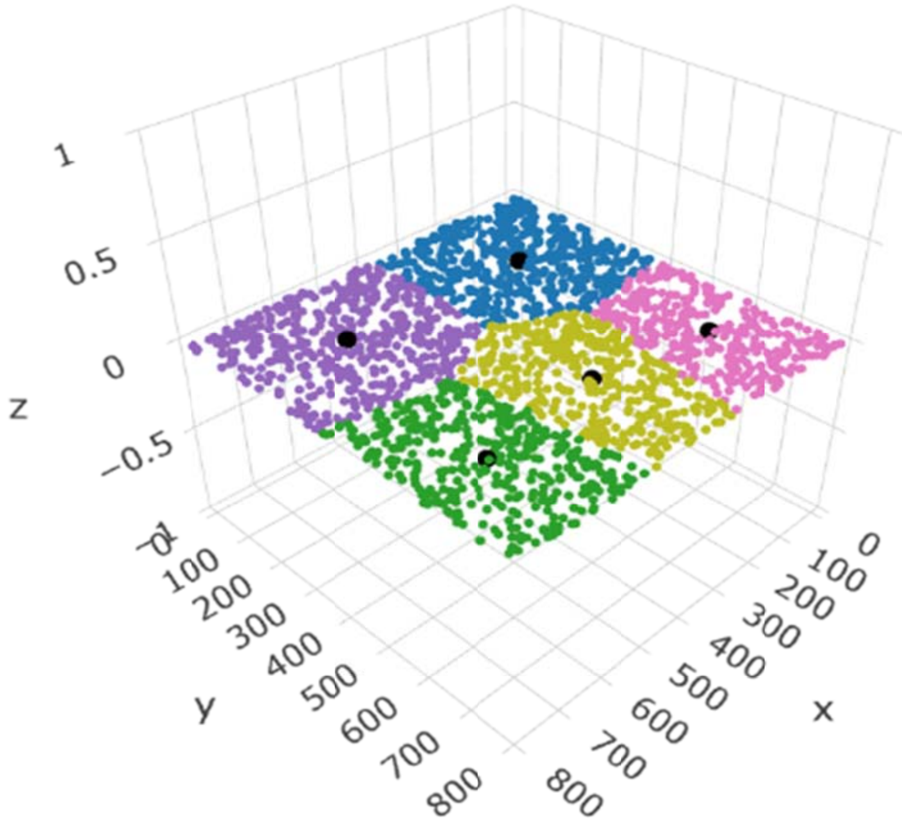


Figure 6: The scatter plot of the second implementation (implementation 2) of the K-means clustering on the dataset

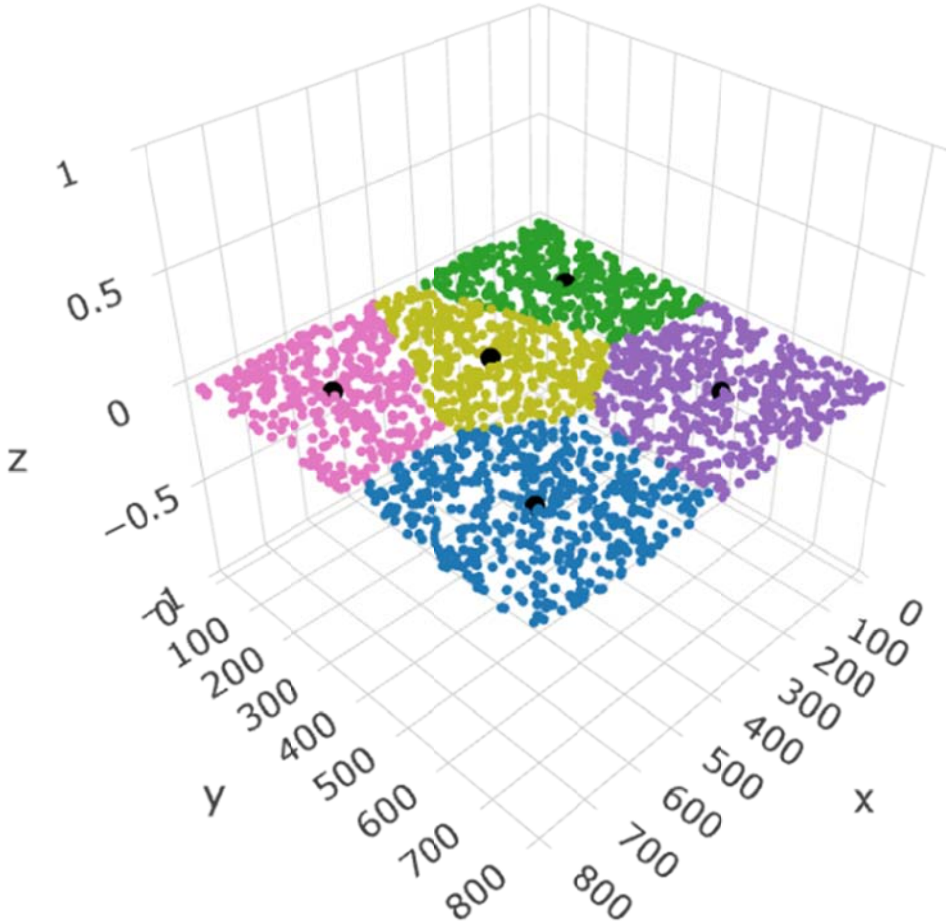


Figure 7: The scatter plot of the third implementation (implementation 3) of the K-means clustering on the dataset

**Table 1 The X and Y coordinates of the centroid of clusters in each of the three implementations of the K-means**

| | X coordinate of the centroid of clusters in implementation 1 (m) | Y coordinate of the centroid of clusters in implementation 1 (m) | X coordinate of the centroid of clusters in implementation 2 (m) | Y coordinate of the centroid of clusters in implementation 2 (m) | X coordinate of the centroid of clusters in implementation 3 (m) | Y coordinate of the centroid of clusters in implementation 3 (m) |
|---|---|---|---|---|---|---|
| Cluster 0 | 185.334347 | 136.610942 | 129.931677 | 611.431677 | 133.781081 | 200.597297 |
| Cluster 1 | 199.535168 | 676.58104 | 185.233405 | 198.768737 | 201.115217 | 621.373913 |
| Cluster 2 | 230.639474 | 407.742105 | 386.612536 | 567.059829 | 399.002907 | 283.43314 |
| Cluster 3 | 601.852878 | 192.132196 | 595.115139 | 186.452026 | 619.578838 | 610.726141 |
| Cluster 4 | 616.060606 | 605.90101 | 659.675192 | 600.946292 | 648.776163 | 172.197674 |

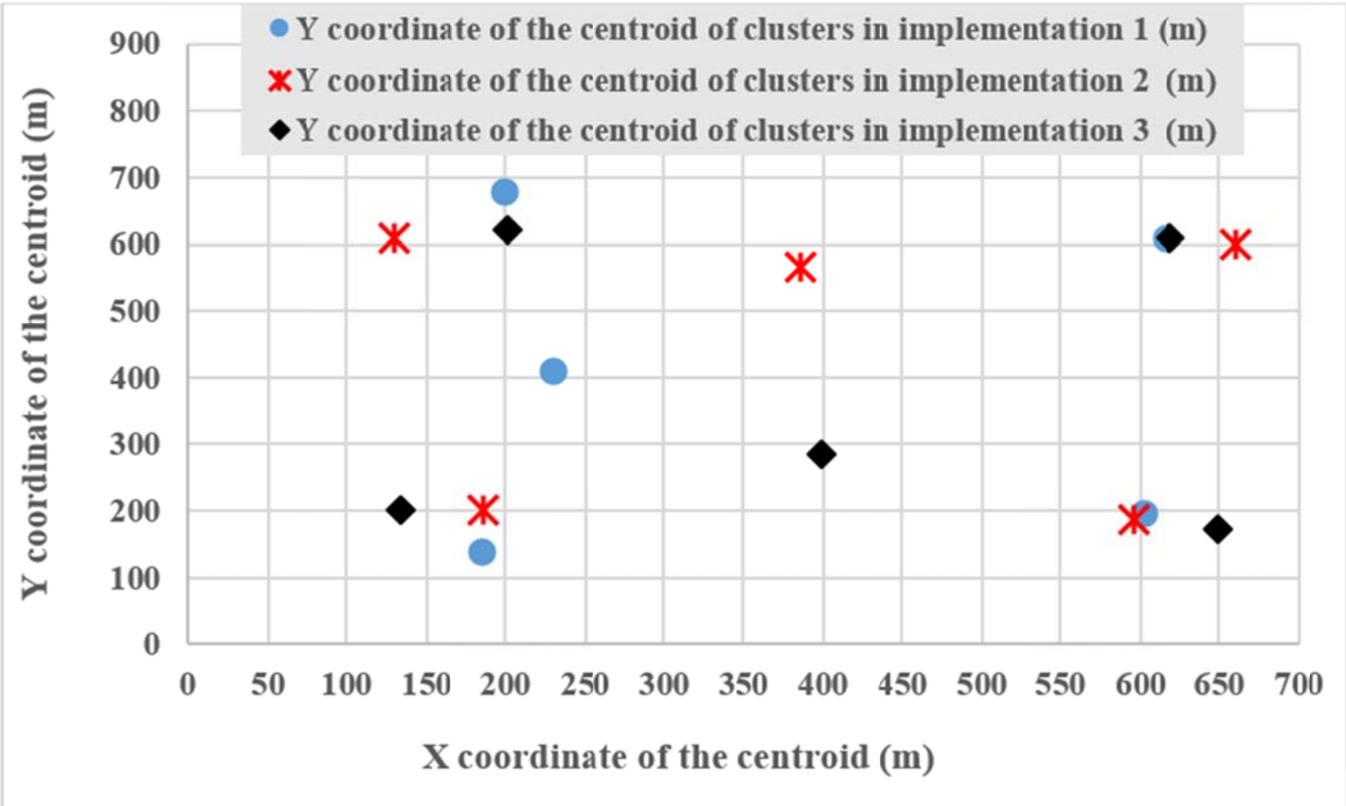


Figure 8: The scatter plot of the X and Y coordinates of the centroid of clusters in each of the three implementations of the K-means

The summary of the distribution of nodes per cluster and the distance between gateway and sensor nodes for k-mean implementation 1 is presented in Table 2, for k-mean implementation 2 is presented in Table 3, and for k-mean implementation 3 is presented in Table 4. Also, the bar chart comparing the mean distance between gateway and sensor node and the maximum distance from gateway for the three implementations of the K-means algorithm are presented in Figure 9. The results show that the mean distance for the entire network was 139 m in the first implementation, 140m in the second implementation, and 141 m in the third implementation. Since the energy consumption in sensor network is proportional to the communication distance, it means that the first implementation will give the lowest energy demand and hence longer battery life span for the network while the third implementation will have the highest and hence worst energy consumption and shortest battery life span.

Again, the bar chart comparing the maximum distance between gateway and sensor node for the three implementations of the K-means algorithm is presented in Figure 10.The results show that the maximum sensor to gateway distance in the entire network was 269 m in the first implementation, 267 m in the second implementation, and 274 in the third implementation. Again, since the energy consumption in sensor network is proportional to the communication distance, it means that the third implementation will have the first sensor node to die due to power outage given the long distance of 274 m. On the other hand, the battery life of the critical node with the longest distance in the second implementation will stay longer than that of the third implementation since the distance is 267m. In all, it can be concluded that the K-means approach studied can be effective in optimal gateway placement. However, several implementations may be required and the implementation with the best communication distance result can be adopted.

**Table 2 The summary of the distribution of nodes per cluster and the distance between gateway and sensor nodes for k-mean implementation 1**

| | Cluster 0 | Cluster 1 | Cluster 2 | Cluste r 3 | Cluster 4 | Min | Max | Average |
|---|---|---|---|---|---|---|---|---|
| No. of nodes | 495 | 329 | 327 | 469 | 380 | 327 | 495 | 400 |
| Percentage of total nodes | 24.8 | 16.5 | 16.4 | 23.5 | 19.0 | 16.4 | 24.8 | 20.0 |
| Minimum Distance from the Gateway (m) | 1 | 2 | 24 | 19 | 6 | 1 | 24 | 10 |
| Maximum Distance from the Gateway (m) | 254 | 252 | 244 | 269 | 242 | 242 | 269 | 252 |
| Average Distance from the Gateway (m) | 146 | 131 | 131 | 150 | 135 | 131 | 150 | 139 |

**Table 3 The summary of the distribution of nodes per cluster and the distance between gateway and sensor nodes for k-mean implementation 2**

| | Cluster 0 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Min | Max | Average |
|---|---|---|---|---|---|---|---|---|
| No. of nodes | 467 | 391 | 469 | 322 | 351 | 322 | 469 | 400 |
| Percentage of total nodes | 23.4 | 19.6 | 23.5 | 16.1 | 17.6 | 16.1 | 23.5 | 20.0 |
| Minimum Distance from the Gateway (m) | 15 | 5 | 10 | 6 | 7 | 5 | 15 | 9 |
| Maximum Distance from the Gateway (m) | 264 | 267 | 267 | 242 | 253 | 242 | 267 | 259 |
| Average Distance from the Gateway (m) | 150 | 135 | 150 | 132 | 138 | 132 | 150 | 141 |

**Table 4 The summary of the distribution of nodes per cluster and the distance between gateway and sensor nodes for k-mean implementation 3**

| | Cluster 0 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Min | Max | Average |
|---|---|---|---|---|---|---|---|---|
| No. of nodes | 482 | 370 | 460 | 344 | 344 | 344 | 482 | 400 |
| Percentage of total nodes | 24.1 | 18.5 | 23.0 | 17.2 | 17.2 | 17.2 | 24.1 | 20.0 |
| Minimum Distance from the Gateway (m) | 7 | 15 | 4 | 15 | 21 | 4 | 21 | 12 |
| Maximum Distance from the Gateway (m) | 258 | 262 | 270 | 274 | 269 | 258 | 274 | 267 |
| Average Distance from the Gateway (m) | 144 | 141 | 147 | 135 | 133 | 133 | 147 | 140 |

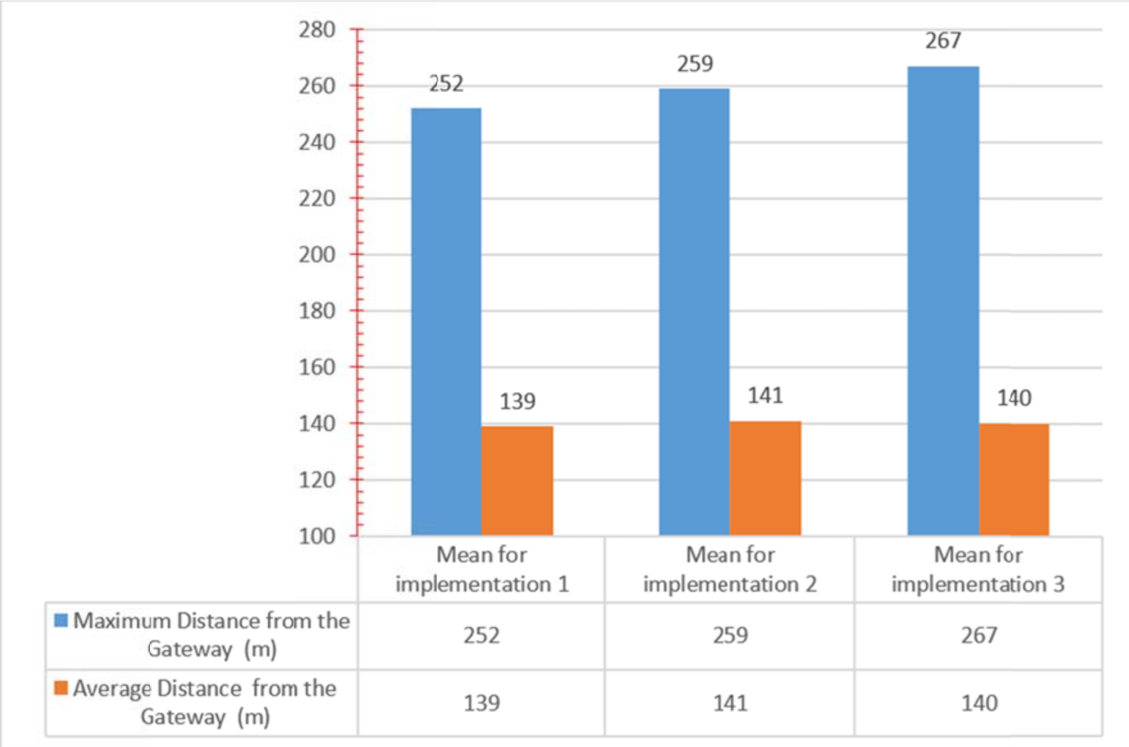


Figure 9: The bar chart comparing the mean distance between gateway and sensor node and the maximum distance from gateway for the three implementations of the K-means algorithm

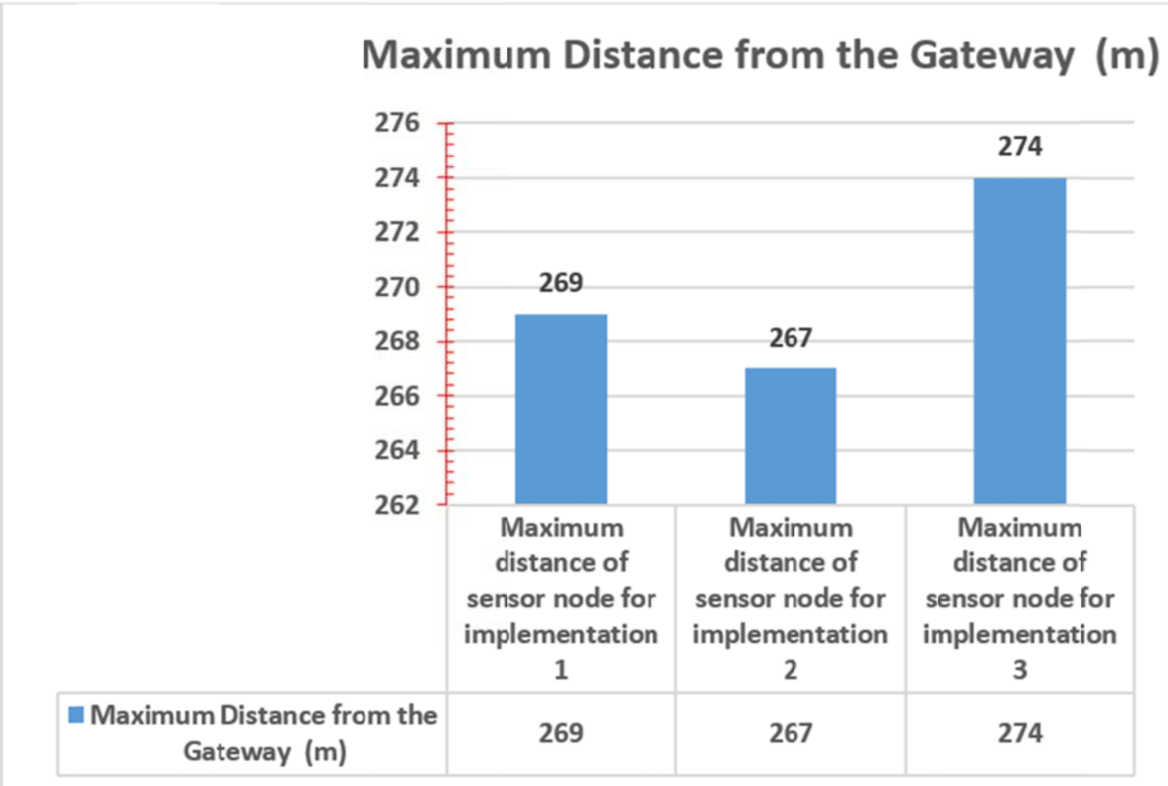


Figure 10: The bar chart comparing the maximum distance between gateway and sensor node for the three implementations of the K-means algorithm

## 4. Conclusion

The sensor node used in clustered IoT network is studied. The study focus on assessing the effect of multiple implementation of the optimal gateway placement using the K-means algorithm. The essence of the study is to see if different instances of implementation of the K-means on the same dataset will give the same centroid coordinates for the gateways and hence the same communication distance for the various sensor nodes in the network in each implementation. However, the results in this study showed that the gateway placement different implementation of the K-means on the same dataset gave rise to different gateway placement in the network. This resulted in different values for the mean distance between the gateways and the sensor node and the maximum distance for the farthest sensor node in the network. The implication of the results is that in order to get the optimal gateway placement in the clustered sensor node, several instances of the K-means implementation is required on the same dataset and the one that gives the best results in terms of mean communication distance and distance of the farthest sensor node can be selected as the optimum solution.

**References**

1. Sofi, A., Regita, J. J., Rane, B., & Lau, H. H. (2022). Structural health monitoring using wireless smart sensor network–An overview. *Mechanical Systems and Signal Processing*, *163*, 108113.
2. Jamshed, M. A., Ali, K., Abbasi, Q. H., Imran, M. A., & Ur-Rehman, M. (2022). Challenges, applications, and future of wireless sensors in Internet of Things: A review. *IEEE Sensors Journal*, *22*(6), 5482-5494.
3. Channi, H. K., & Kumar, R. (2021). The role of smart sensors in smart city. In *Smart Sensor Networks: Analytics, Sharing and Control* (pp. 27-48). Cham: Springer International Publishing.
4. Chatzigiannakis, I., Mylonas, G., &Nikoletseas, S. (2009). The design of an environment for monitoring and controlling remote sensor networks. *International Journal of Distributed Sensor Networks*, *5*(3), 262-282.
5. Harris, T. J., Seppala, C. T., &Desborough, L. D. (1999). A review of performance monitoring and assessment techniques for univariate and multivariate control systems. *Journal of process control*, *9*(1), 1-17.
6. Cai, Z., Chen, Q., Shi, T., Zhu, T., Chen, K., & Li, Y. (2022). Battery-free wireless sensor networks: A comprehensive survey. *IEEE Internet of Things Journal*, *10*(6), 5543-5570.
7. Luo, Y., Abidian, M. R., Ahn, J. H., Akinwande, D., Andrews, A. M., Antonietti, M., ...& Chen, X. (2023). Technology roadmap for flexible sensors. *ACS nano*, *17*(6), 5211-5295.
8. Chavan, S., Venkateswarlu, B., Prabakaran, R., Salman, M., Joo, S. W., Choi, G. S., & Kim, S. C. (2023). Thermal runaway and mitigation strategies for electric vehicle lithium-ion batteries using battery cooling approach: A review of the current status and challenges. *Journal of Energy Storage*, *72*, 108569.
9. Li, R., Li, W., Singh, A., Ren, D., Hou, Z., & Ouyang, M. (2022). Effect of external pressure and internal stress on battery performance and lifespan. *Energy Storage Materials*, *52*, 395-429.
10. Borkar, G. M., Patil, L. H., Dalgade, D., &Hutke, A. (2019). A novel clustering approach and adaptive SVM classifier for intrusion detection in WSN: A data mining concept. *Sustainable Computing: Informatics and Systems*, *23*, 120-135.
11. Srinivasan, D., Kiran, A., Parameswari, S., &Vellaichamy, J. (2025). Energy efficient hierarchical clustering based dynamic data fusion algorithm for wireless sensor networks in smart agriculture. *Scientific Reports*, *15*(1), 7207.
12. Bhasker, B., &Murali, S. (2024). An Energy-Efficient Cluster-based data aggregation for agriculture irrigation management system using wireless sensor networks. *Sustainable Energy Technologies and Assessments*, *65*, 103771.
13. Kodinariya, T. M., &Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, *1*(6), 90-95.
14. Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, *9*(8), 1295.
15. Gupta, M. K., & Chandra, P. (2022). Effects of similarity/distance metrics on k-means algorithm with respect to its applications in IoT and multimedia: a review. *Multimedia Tools and Applications*, *81*(26), 37007-37032.
16. Hosseinzadeh, M., Hemmati, A., &Rahmani, A. M. (2022). Clustering for smart cities in the internet of things: a review. *Cluster Computing*, *25*(6), 4097-4127