Science and Technology Publishing (SCI & TECH)
ISSN: 2632-1017
Vol. 8 Issue 4, April - 2024

Comparative Evaluation Of Performance Of Improved Silhouette Method For Sensor Nodes Clustering Applications

Nnamonso O. Akpabio[1]
Department of Computer Engineering Ibom Metropolitan Polytechnic,
Uyo, Akwa Ibom State Nigeria

Chikezie Samuel Aneke[2]
Department Of Computer Engineering
University Of Uyo, Akwa Ibom State

Emmanuel Nsese Udoiwod[3]
Department Of Computer Engineering
University Of Uyo, Akwa Ibom State
emmanueludoiwod@uniuyo.edu.ng

***Abstract*— In this paper, comparative evaluation of performance of improved Silhouette method for sensor nodes clustering applications is presented. Specifically, the improved Silhouette method is compared with three other methods, namely, the Elbow method, the Gap Statistics method and the classical Silhouette method. The comparison is made in terms of the optimum number of clusters obtained and the execution time of the algorithm for any given set of sensor nodes randomly distributed in a given network coverage area. Simulations were conducted for different numbers of sensor nodes in a network coverage area of 1000 m by 1000 m. Each of the four methods were employed to determine the optimum number of clusters for the 5000 sensor nodes and the Elbow method arrived at four as the optimal number whereas the other three methods gave five as the optimal number. Again the results shows that with 5000 sensors, the classical Silhouette method has the highest mean execution time of 6.68 s followed by the gap statistics method with execution time of 6.55 s. The improved Silhouette method had the lowest execution time of 6.10s. Again, execution time of the improved Silhouette is compared to the execution time of the Gap statistics method. The results showed that with 5000 nodes, the improved Silhouette method is 0.46 s faster than the Gap statistics method. With 10000 nodes, the improved Silhouette method is 1.04 s faster than the Gap statistics method and with 15000 nodes, the improved Silhouette method is 2.04 s faster than the Gap statistics method. In all, the improved Silhouette provided the fastest algorithm for computing the optimal number of clusters.**

Keywords*— *Silhouette Method, Elbow Method, Sensor Node, Gap Statistics Method, Clustering

1. INTRODUCTION

Across the globe, wireless sensors are increasingly being deployed in diverse applications [1,2,3,4]. The emerging smart systems like smart homes, smart city, smart transport, smart grid , smart farming among others are all relying on in concept of Internet of sensors for timely data collection and communication as well as control and remote management of devices and systems [5, 6,7,8]. With the growing applications of wireless sensors and the resource constraint nature of such sensors, effort is being made to provide techniques to optimize the resource utilization and maximize the battery life span of such sensors [9,10,11]. One method of achieving managing the energy efficiency of the sensors in a large sensor network is clustering [2,13,14]. In this case, the sensors are grouped into clusters and cluster head is assigned to each cluster to serve as the base station or gateway for communicating with the other remote systems connected to the network.

In a clustered network, determining the appropriate number of clusters that will be required is always a challenge [15,16]. However, this is accomplished using some methods. Some of the popular methods include the Elbow method, the Gap statistics, and the Silhouette method [17,18,19]. However, of these methods require the computing resource of the sensor network. The execution time of the methods can affect the performance of the system, especially when large number of nodes are involved. Accordingly, in this study, an improved Silhouette method is presented and the performance of the improved method is compared with the three listed approaches. The focus of the study is to demonstrate through simulations, the ability of the improved method to enhance the timely determination of the required optimal number of clusters in the face of large number of nodes to be considered.

2. METHODOLOGY

This paper focused on evaluating the performance of an improved Silhouette method of determining the optimal number of clusters required for a given distribution of sensor nodes in a given network coverage area. The improved Silhouette method is compared with three other methods, namely, the Elbow method, the Gap Statistics method and the classical Silhouette method. This is an essential step in the clustering of sensors. It helps to minimize the mean energy consumption of the sensors when they transmit to the cluster heads or gateway located within the cluster. The key concept behind each of the three methods is to determine the optimum number of clusters with the minimum total within-cluster sum of square (WSS). Increasing the number of clusters beyond the optimal value does not show any significant reduction in the value of the WSS. On the other hand, reducing the number of clusters below the optimal value will significantly increase the WSS value. Moreover, having the number of clusters above the optimum value increases the number of cluster heads or gateways required for the given network which amounts to additional cost that does not significantly improve on the network performance.

Notably, the three listed methods, Elbow method, Gap Statistics method and Silhouette method have been widely studied and are implemented in this work just as the other works have published. Therefore, in this work, the flowchart of the classical Silhouette method is presented. The improved Silhouette method is developed from the classical Silhouette method. As such, the flowchart of the improved Silhouette method is also presented as it shows the modifications the have been made and the relationship with the classical Silhouette method. Furthermore, the simulation of optimal number of cluster determination using each of the various method listed is presented.

The flowchart of the Silhouette method for determination of optimal number of clusters is presented in Figure 1. Similarly, the flowchart of the improved Silhouette method for determination of optimal number of clusters is presented in Figure 2.

The improvement is based on addressing the inherent challenge in the classical Silhouette method. Notable, it requires initialization of centroid and expected number of clusters. Given that the choice of the initial value for the centroid can significantly affect the implementation time for determination of the optimal number of clusters, the improved Silhouette method employed some smart initialization concepts among other salient features.

First, the improved Silhouette me uses smart initialization technique which employs probability distribution technique which takes points located far apart as the initial centroid. Secondly it transforms data into scale between 0 and 1 before the data is loaded into the Silhouette algorithm. Thirdly, it assigns sensor nodes to the centroid determined using the triangle inequality approach where $\|x\| + \|y\| \geq \|x + y\|$ as presented in Figure 2. This approach significantly reduces the computation time for the distance.

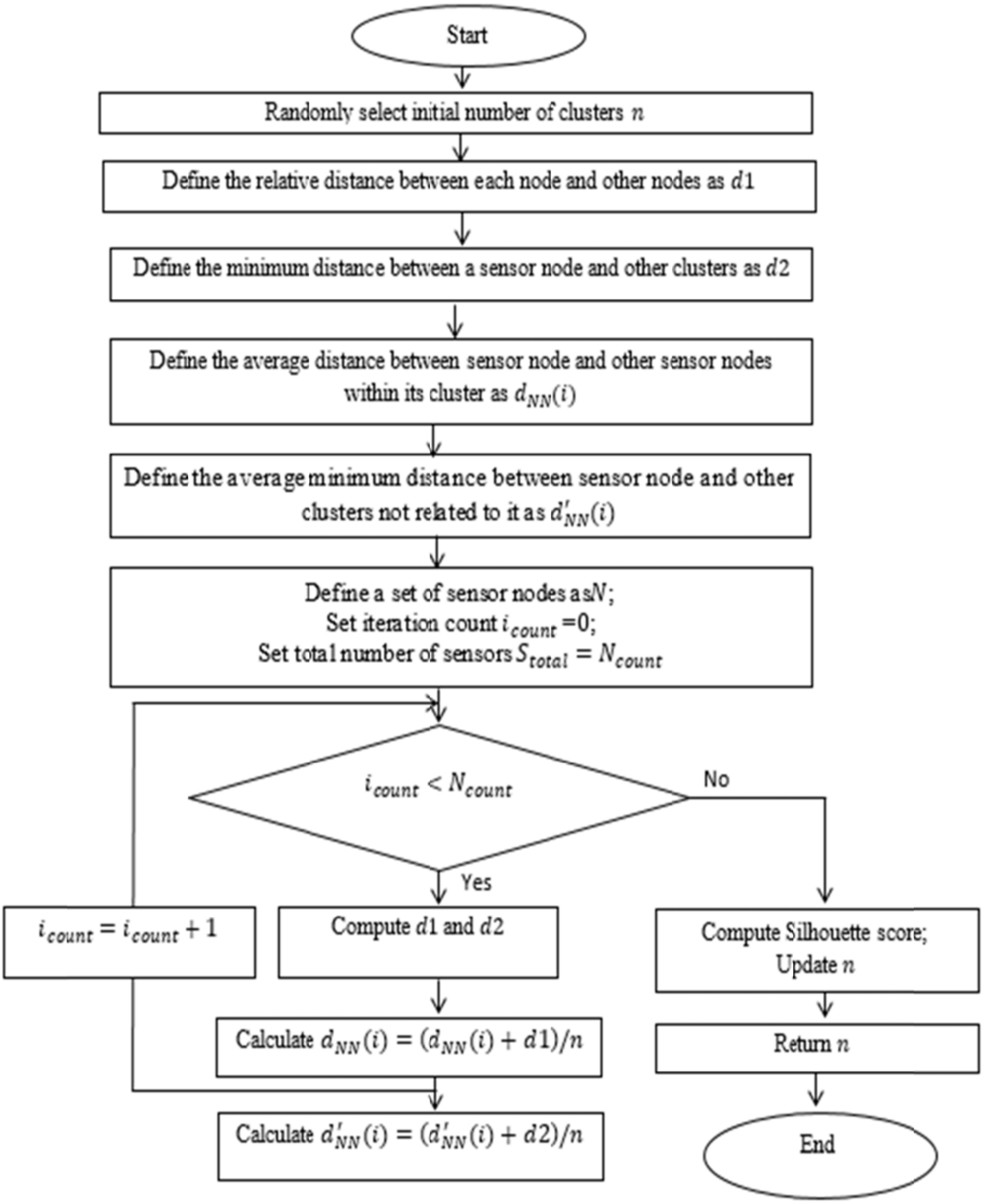


Figure 1 The flowchart of the Silhouette method for determination of optimal number of clusters

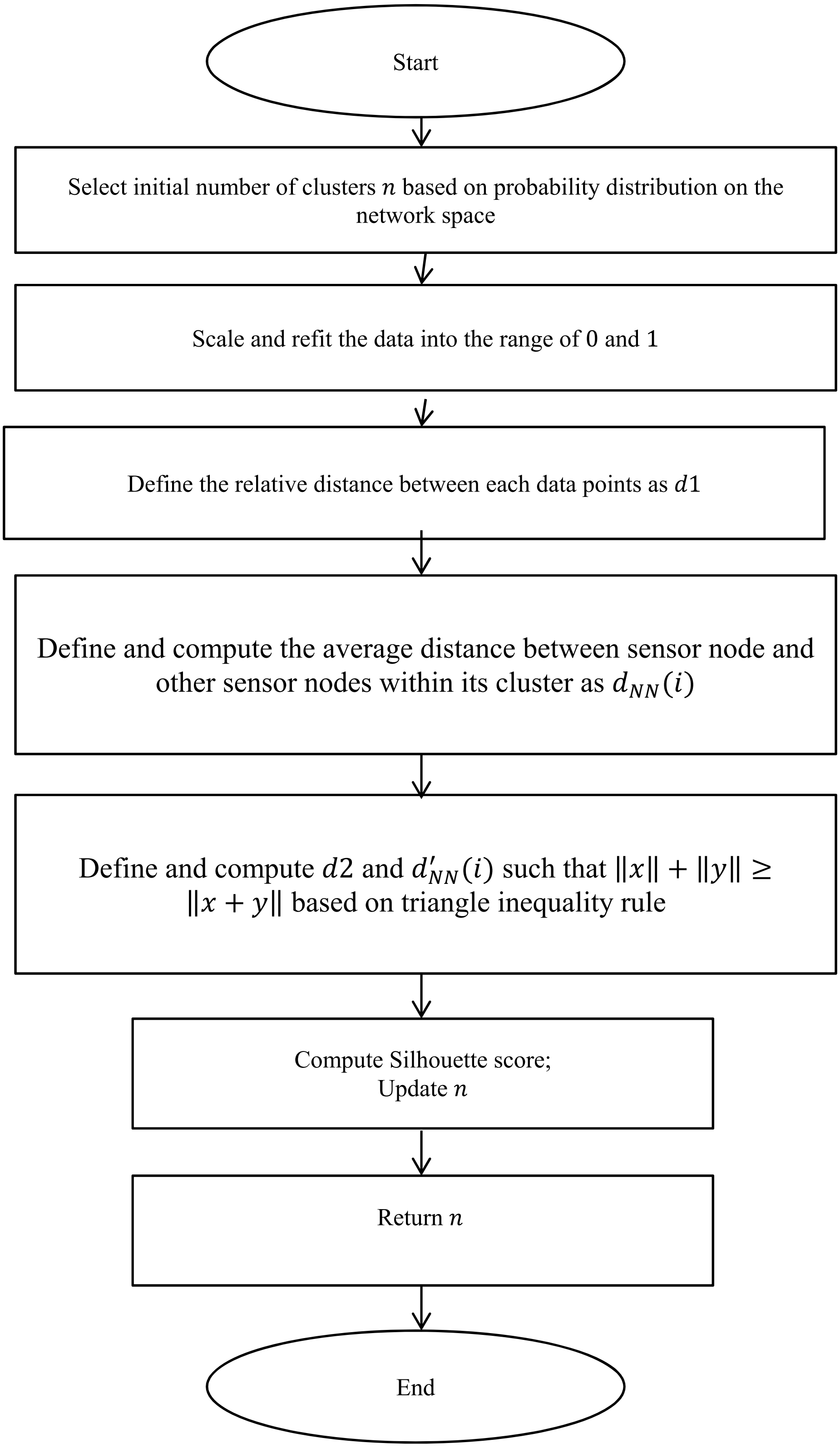


Figure 2 The flowchart of the improved Silhouette method for determination of optimal number of clusters

3. RESULTS AND DISCUSSIONS

Simulations were conducted for different numbers of sensor nodes in a network coverage area of 1000 m by 1000 m. The case of 5000 sensor nodes distributed randomly within the 1000 m by 1000 m area is shown in Figure 3. Each of the four methods were employed to determine the optimum number of clusters for the 5000 sensor nodes and the Elbow method arrived at 4 as the optimal number, as shown in Figure 4. However, the Gap statistics, the classical Silhouette and the improved Silhouette method gave 5 as the optimal number, as shown in Figure 5 for the Gap Statistic method, in Figure 6 for the classical Silhouette method and in Figure 8 for the classical Silhouette method. The bar chart that shows the summary of the optimal number of clusters for the four different methods is presented in Figure 8.

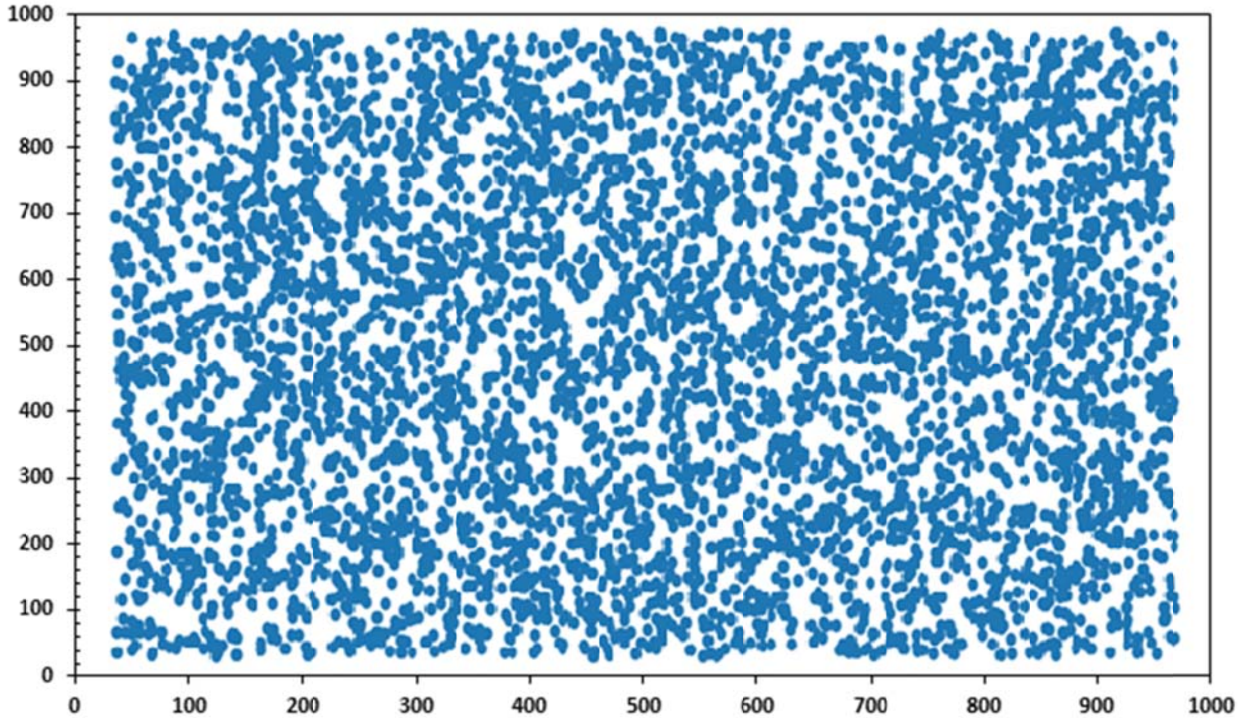


Figure 3 The 1000 m by 1000 m network area with 5000 sensor nodes

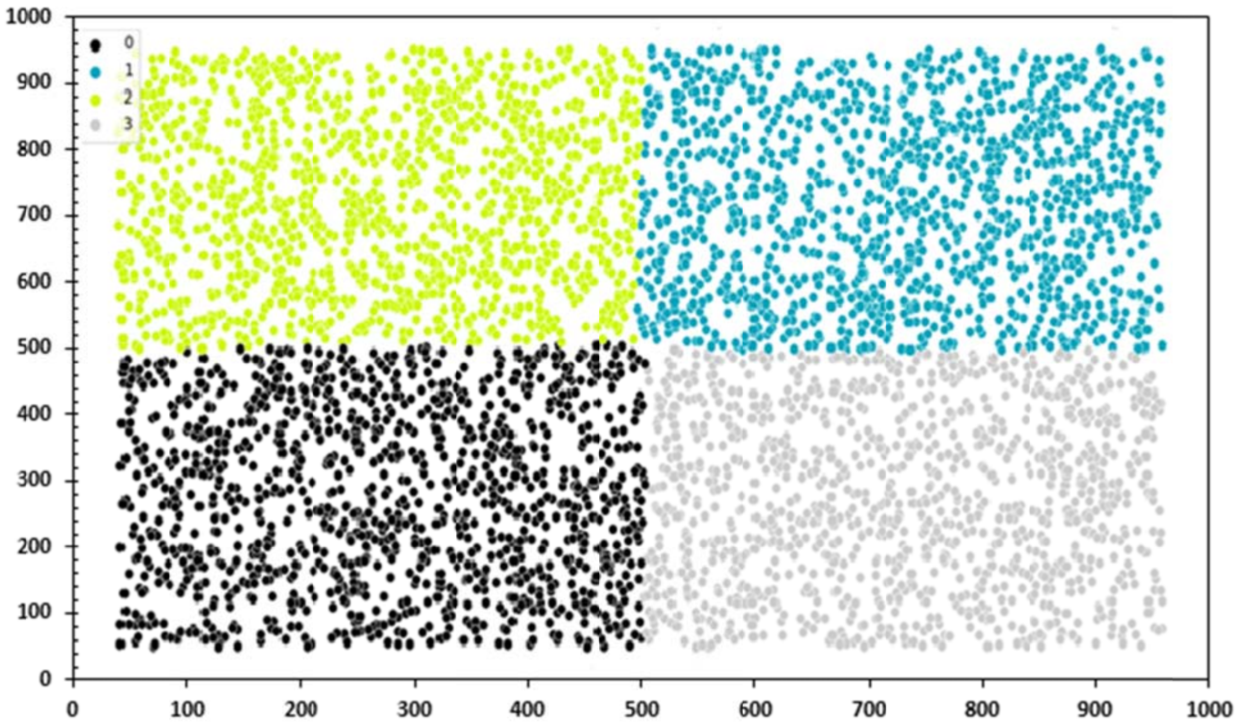


Figure 4: Results of the Elbow method for the optimal number of cluster

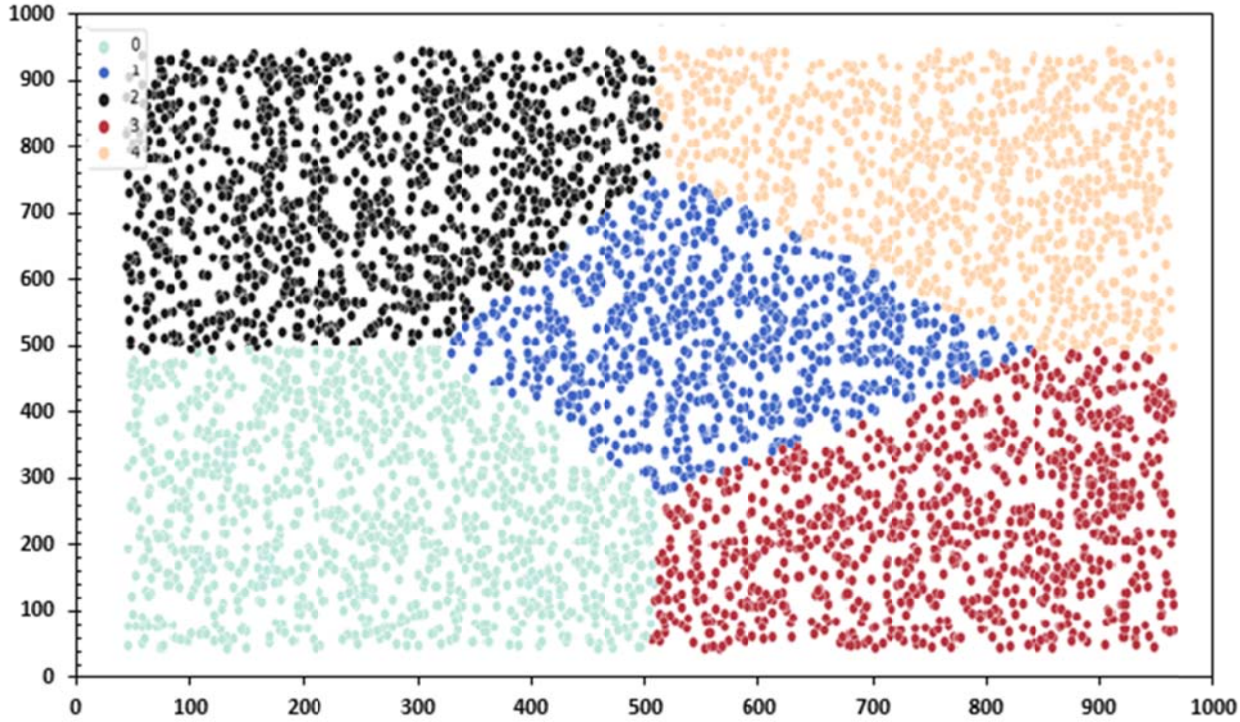


Figure 5: Results of the Gap Statistic method for the optimal number of cluster

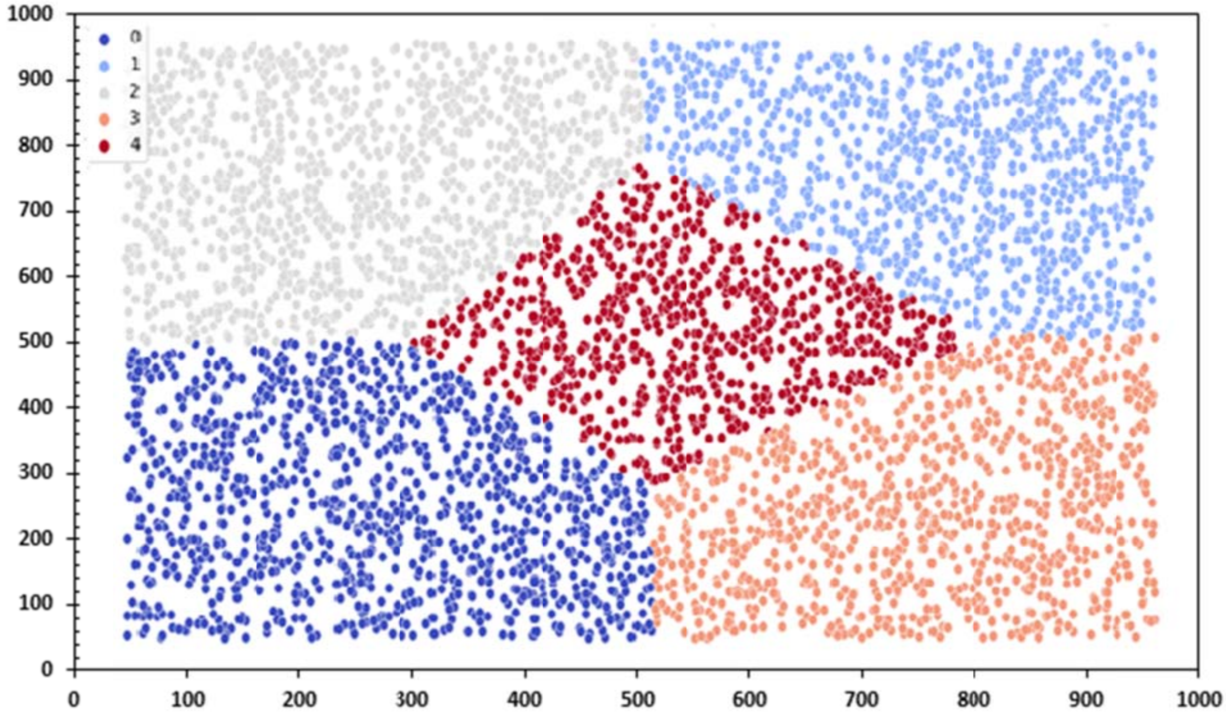


Figure 6: Results of the Silhouette method for the optimal number of cluster

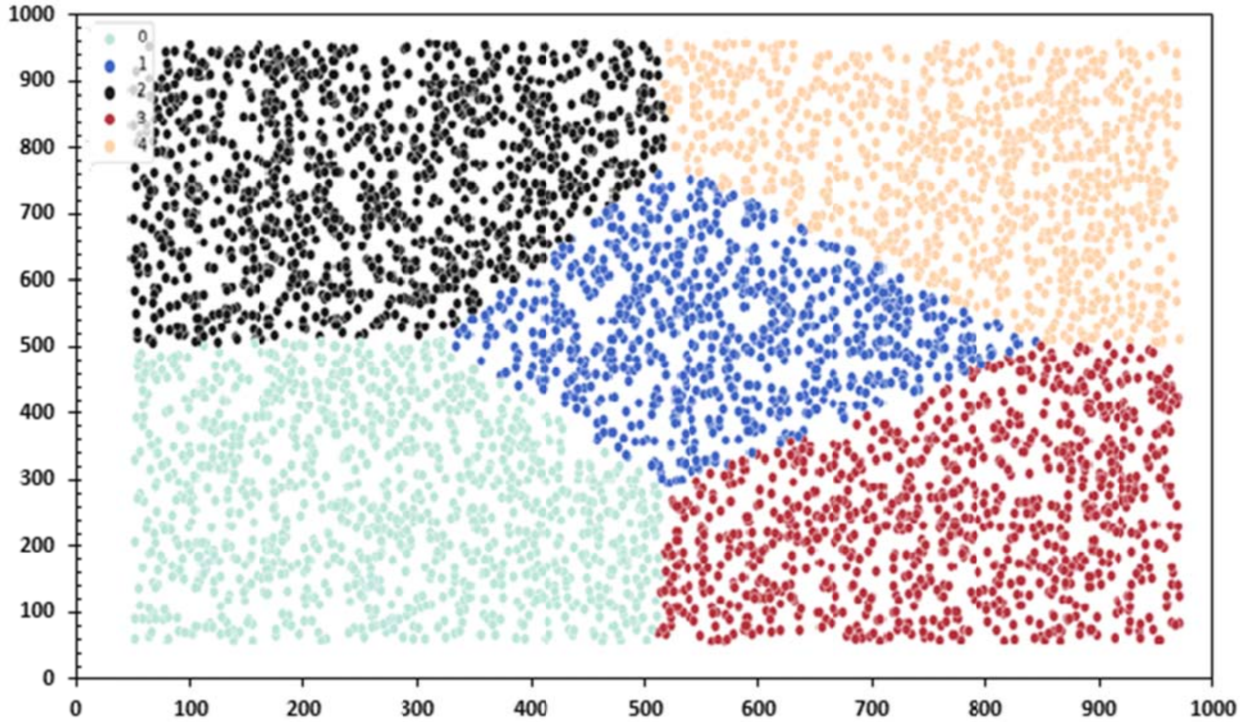


Figure 7: Results of the improved Silhouette method for the optimal number of cluster

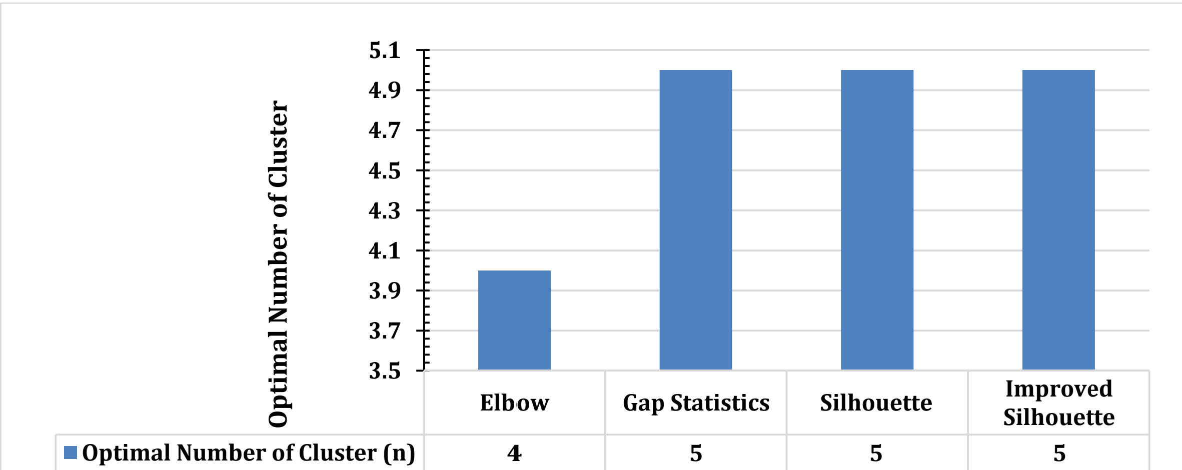


Figure 8 The bar chart that shows the summary of the optimal number of clusters for the four different methods

Basically, the essence of the modification of the classical Silhouette method is to improve on the speed of implementation of the algorithm. As such, the execution time of each of the methods captured and compare. Specifically, the simulation was conducted for 5000 sensors using a random number generator to determine the location of each of the 5000 nodes within the 1000 m by 1000 m area. The random distribution of the sensor nodes was conducted four times and for each random distribution the optimal number of cluster is determined using each of the methods and the execution time is recorded.

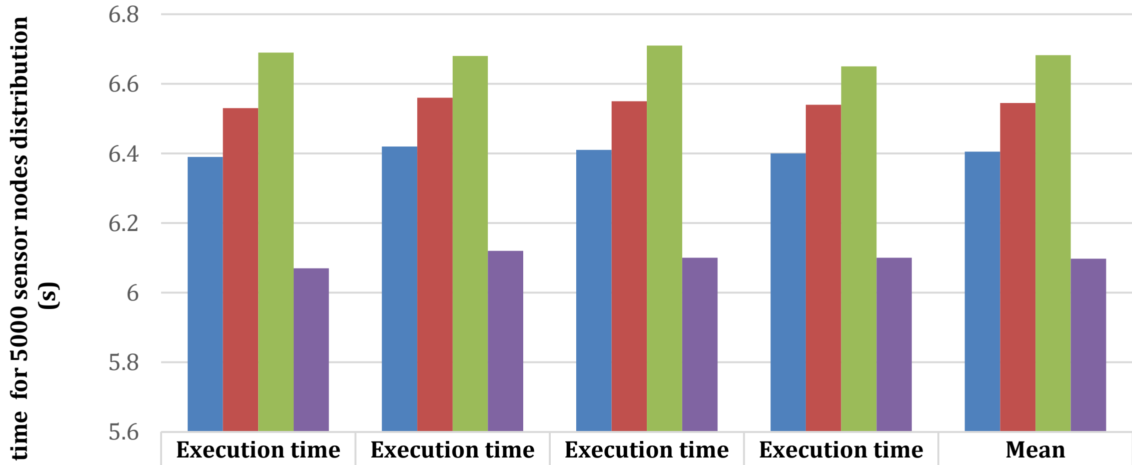
The results obtained for the execution time of the four methods in the four different random distributions of the sensor node is captured in Figure 9. The results shows that the classical Silhouette method has the highest mean execution time of 6.68 s followed by the gap statistics method with execution time of 6.55 s. The improved Silhouette method had the lowest execution time of 6.10s.

The bar chart that shows the summary of the execution times of the four different methods for each of the four different random distributions of the 10000 sensors is captured in Figure 10. The results shows that the classical Silhouette method has the highest mean execution time of 9.23 s followed by the gap statistics method with execution time of 8.89 s. The improved Silhouette method had the lowest execution time of 7.86 s.

The bar chart that shows the comparison of the execution times of the four different methods one of the four different random distributions of the 5000, 10000 and 15000 sensors is captured in Figure 12. The results shows that the differences in execution times of the four methods increase with increase in the number of sensor nodes considered. With 5000 nodes, the improved Silhouette method is0.32 s faster than the Elbow method. With 10000 nodes, the improved Silhouette method is 0.46 s faster than the Elbow method and with 15000 nodes, the improved Silhouette method is 1.34 s faster than the Elbow method.

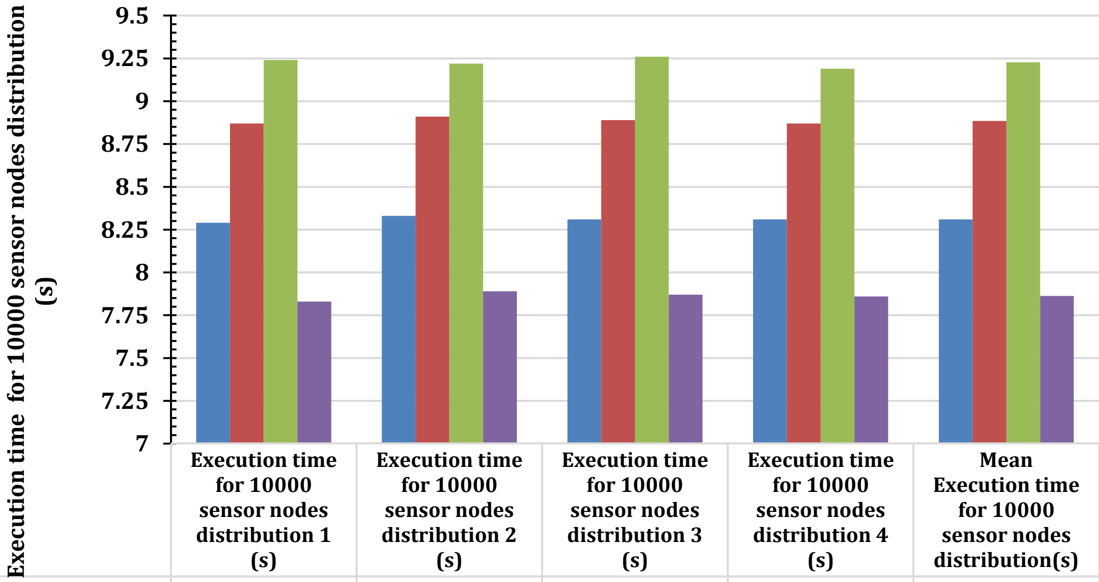
Again, execution time of the improved Silhouette is compared to the execution time of the Gap statistics method. The results showed that with 5000 nodes, the improved Silhouette method is 0.46 s faster than the Gap statistics method. With 10000 nodes, the improved Silhouette method is 1.04 s faster than the Gap statistics method and with 15000 nodes, the improved Silhouette method is 2.04 s faster than the Gap statistics method.

As noted, the results show that the Elbow is not accurate in determining the optimal number of clusters when compared with the other three methods. As such, though the Elbow is slightly slower than the improved Silhouette method, it made not be recommended since the computation accuracy is poor. In all, the improved Silhouette provided the fastest algorithm for computing the optimal number of clusters.



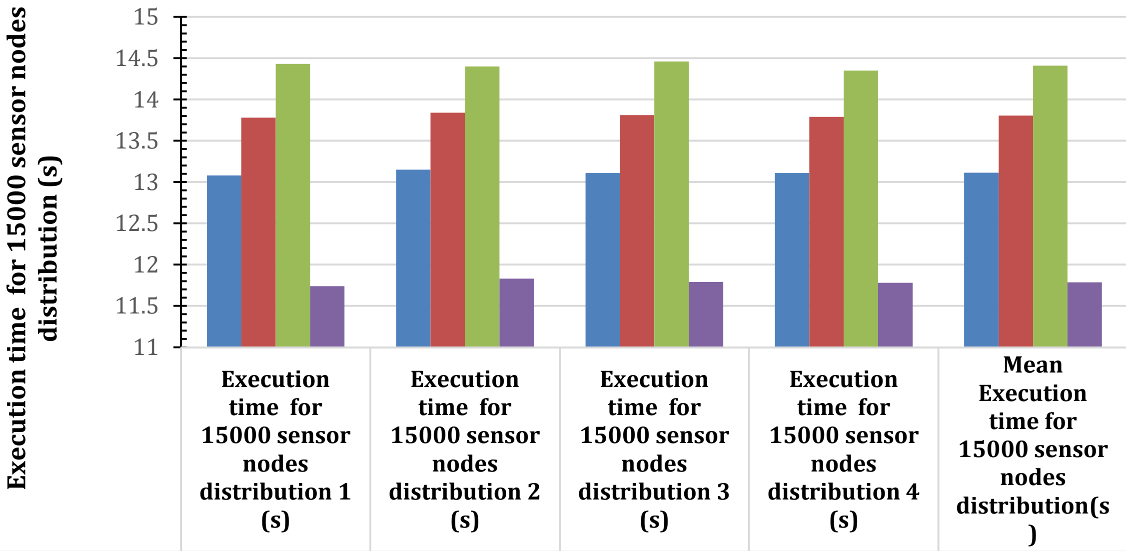
	Execution time for 5000 sensor nodes distribution 1 (s)	Execution time for 5000 sensor nodes distribution 2 (s)	Execution time for 5000 sensor nodes distribution 3 (s)	Execution time for 5000 sensor nodes distribution 4 (s)	Mean Execution time for 5000 sensor nodes distribution(s)
Elbow	6.39	6.42	6.41	6.4	6.41
Gap Statistics	6.53	6.56	6.55	6.54	6.55
Silhouette	6.69	6.68	6.71	6.65	6.68
Improved Silhouette	6.07	6.12	6.1	6.1	6.10

Figure 9 The bar chart that shows the summary of the execution times of the four different methods for each of the four different random distributions of the 5000 sensors



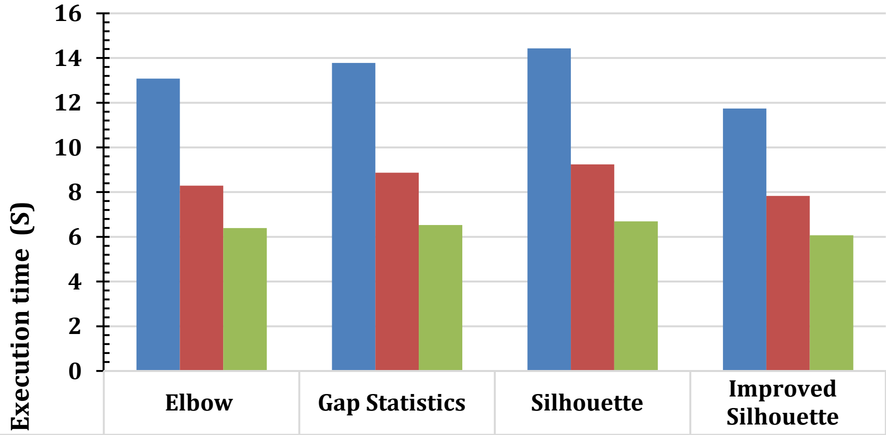
	Execution time for 10000 sensor nodes distribution 1 (s)	Execution time for 10000 sensor nodes distribution 2 (s)	Execution time for 10000 sensor nodes distribution 3 (s)	Execution time for 10000 sensor nodes distribution 4 (s)	Mean Execution time for 10000 sensor nodes distribution(s)
Elbow	8.29	8.33	8.31	8.31	8.31
Gap Statistics	8.87	8.91	8.89	8.87	8.89
Silhouette	9.24	9.22	9.26	9.19	9.23
Improved Silhouette	7.83	7.89	7.87	7.86	7.86

Figure 10 The bar chart that shows the summary of the execution times of the four different methods for each of the four different random distributions of the 10000 sensors



	Execution time for 15000 sensor nodes distribution 1 (s)	Execution time for 15000 sensor nodes distribution 2 (s)	Execution time for 15000 sensor nodes distribution 3 (s)	Execution time for 15000 sensor nodes distribution 4 (s)	Mean Execution time for 15000 sensor nodes distribution(s)
Elbow	13.08	13.15	13.11	13.11	13.11
Gap Statistics	13.78	13.84	13.81	13.79	13.81
Silhouette	14.43	14.4	14.46	14.35	14.41
Improved Silhouette	11.74	11.83	11.79	11.78	11.79

Figure 11The bar chart that shows the summary of the execution times of the four different methods for each of the four different random distributions of the 15000 sensors



	Elbow	Gap Statistics	Silhouette	Improved Silhouette
Execution time for 15000 sensor nodes distribution 1 (s)	13.08	13.78	14.43	11.74
Execution time for 10000 sensor nodes distribution 1 (s)	8.29	8.87	9.24	7.83
Execution time for 5000 sensor nodes distribution 1 (s)	6.39	6.53	6.69	6.07

Figure 12 The bar chart that shows the comparison of the execution times of the four different methods one of the four different random distributions of the 5000, 10000 and 15000 sensors

4. CONCLUSION

Four methods of computing the optimal number of clusters for a given random distribution of sensor nodes is presented. The methods are the Elbow method, the Gap Statistics method, the classical Silhouette method and the improved Silhouette method. The focus is to evaluate the performance of the improved Silhouette method and compare it with that of the other three existing methods. Specifically, the simulation results show that the execution time of the improved Silhouette method is the least among the four methods and the improved Silhouette method has accurate determination of the optimal number of cluster.
On the other hand, the Elbow method has small execution time which is higher than that of the improved Silhouette method. In any case, the Elbow method is not as accurate as the improved Silhouette method in determining the optimum number of clusters.

REFERENCES

1. 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.
2. Landaluce, H., Arjona, L., Perallos, A., Falcone, F., Angulo, I., & Muralter, F. (2020). A review of IoT sensing applications and challenges using RFID and wireless sensor networks. *Sensors*, *20*(9), 2495.
3. Landaluce, H., Arjona, L., Perallos, A., Falcone, F., Angulo, I., & Muralter, F. (2020). A review of IoT sensing applications and challenges using RFID and wireless sensor networks. *Sensors*, *20*(9), 2495.
4. U. Ukommi, K. Kodikara Arachchi, Safak Dogan and A.M. Kondoz, "Content-Aware Bitrates Adaptation for Robust Mobile Video Services", IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, United Kingdom, June 2013. **DOI:** 10.1109/BMSB.2013.6621696
5. Ahmad, T., & Zhang, D. (2021). Using the internet of things in smart energy systems and networks. *Sustainable Cities and Society*, *68*, 102783.
6. Essien, A., Ukommi, U., & Ubom, E. (2024). Downlink Power Budget and Bit Error Analysis for LoRa-Based Sensor Node-to-Satellite Link in the Industrial, Scientific and Medical Frequency Bands. Signals and Communication Technology. Springer Nature, Switzerland. 143-152. https://link.springer.com/chapter/10.1007/978-3-031-53935-0_14
7. Olufemi, O.I., Ukommi, U. (2024). Evaluation of Energy Consumption and Battery Life Span for LoRa IoT Multisensor Node for Precision Farming Application. Signals and Communication Technology. Springer Nature, Switzerland. 153-162. https://link.springer.com/chapter/10.1007/978-3-031-53935-0_15
8. Oduoye, O, Ukommi, U & Ubom, E (2023). Comparative Analysis of Transceiver Payload Size Impact on The Performance of LoRaBased Sensor

Node. Science and Technology Publishing (SCI & TECH), 7(8), 1559-1563.

9. Ghosal, A., Halder, S., & Das, S. K. (2020). Distributed on-demand clustering algorithm for lifetime optimization in wireless sensor networks. *Journal of Parallel and Distributed Computing*, *141*, 129-142.
10. Riaz, A., Sarker, M. R., Saad, M. H. M., & Mohamed, R. (2021). Review on comparison of different energy storage technologies used in micro-energy harvesting, WSNs, low-cost microelectronic devices: challenges and recommendations. *Sensors*, *21*(15), 5041.
11. O. Onwunah, R. Udoh and **U. Ukommi**, "Modelling of the Energy Level and Outage Analysis for Battery-powered IoT Sensor Node with Soar Energy Harvester", International Multilingual Journal of Science and Technology (IMJST), Volume 7 Issue 10, October, 2022
12. Shahraki, A., Taherkordi, A., Haugen, Ø., & Eliassen, F. (2020). Clustering objectives in wireless sensor networks: A survey and research direction analysis. *Computer Networks*, *180*, 107376.
13. Nakas, C., Kandris, D., & Visvardis, G. (2020). Energy efficient routing in wireless sensor networks: A comprehensive survey. *Algorithms*, *13*(3), 72.
14. Sinde, R., Begum, F., Njau, K., & Kaijage, S. (2020). Refining network lifetime of wireless sensor network using energy-efficient clustering and DRL-based sleep scheduling. *Sensors*, *20*(5), 1540.
15. 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*, 1-16.
16. Shahraki, A., Taherkordi, A., Haugen, Ø., & Eliassen, F. (2020). Clustering objectives in wireless sensor networks: A survey and research direction analysis. *Computer Networks*, *180*, 107376.
17. Sagala, N. T., & Gunawan, A. A. S. (2022). Discovering the optimal number of crime cluster using elbow, silhouette, gap statistics, and nbclust methods. *ComTech: Computer, Mathematics and Engineering Applications*, *13*(1), 1-10.
18. 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.
19. Wulandari, S. (2020). Analyze k-value selected method of k-means clustering algorithm to clustering province based on disease case. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *9*(3), 121-124.