Computation Of Optimal Gateway Location Using Silhouette Method And Support Vector Machines Technique

Ahiara, W . C.[1]
Department of Computer Engineering
Michael Okpara University of Agriculture Umudike, Abia State
ahiara.wilson@mouau.edu.ng

Afolayan Jimoh Jacob [2]
Department of Electrical/Electronics Engineering,
University of Uyo, Akwa Ibom State, Nigeria
afolayan.jimoh@yahoo.com

Ofonime Dominic Okon[3]
Department of Computer Engineering
University of Uyo, Akwa Ibom State Nigeria

***Abstract*— In this paper, computation of optimal gateway location using Silhouette method and support Vector Machines (SVM) technique is presented. The Silhouette method is used to compute the optimum number of cluster for any given number of sensor nodes distributed within a given network coverage area. On the other hand, the SVM technique is used to determine the optimal location of the gateways or cluster heads within each of the clusters. Also, the SVM is used to assign the each of the sensor nodes to their respective cluster. The algorithms and analytical expressions associated with the Silhouette method and support Vector Machines (SVM) technique are presented. The models are implemented in Python 3.12 using a case study 5000 sensor nodes distributed randomly in an area of $800m \times 800m \; or \; 640km^2$ .The Silhouette method identified five clusters as the optimal value. Then, the SVM was used to determine the optimal locations of the five gateways in four different distributions of the 5000 sensor nodes. The computation time was also determined. The results show that different random distribution of the 5000 sensor nodes give rise to different location for the gateways. Also, the results show that the Silhouette method and SVM technique took an average execution time of 991.7825 seconds to compute the optimum number of clusters using the Silhouette method, determine the optimal gateway locations for each of the clusters and also cluster the 5000 sensor nodes to their respective cluster gateway.**

***Keywords*— Optimal Gateway Location, Sensor Network, Silhouette Method, Clustering Algorithm, Support Vector Machines Technique**

1. INTRODUCTION

Today, advancements in both electronic and communication technologies have resulted in sensors and sensor networks that have gained wide applications across the globe [1,2,3]. The rising smart systems and Internet of Things (IoT) and their applications depends majorly on the wireless sensor network [4,5,6]. In this wise, researcher have devoted much time to proffer solutions pertaining to wireless sensor networks.

One of the concerns of this work is to utilize two methods, the first method which is Silhouette method is used to determine the optimum number of clusters in a clustered network [7,8,9] while the second method which is Support Vector Machines (SVM) technique is used to determine the optimal placement or location of the gateways within the clusters [10,11,12]. The study also examined the effect of random distribution of the sensor nodes on the gateway placement locations. Again the execution time of the models are evaluated. Essentially, this study seeks to provide insights into the execution time for computing the optimal number of clusters and the optimal gateway placement in a clustered sensor network. The optimal gateway placement solutions will minimize cost and energy consumption in the network [13,14,15]. The energy consumption is always a big concern in sensor networks especially given that sensors are resource constrained and mainly battery powered [16,17,18]. So, this study provides requisite ideas for sensor network designers and researchers seeking to enhance network lifespan through optimized cluster and gateway placement solutions.

2. METHODOLOGY

This work seek to first determine the optimum number of clusters for a given number of sensor nodes. Afterwards, the optimum number of clusters is used to determine the optimal gateway or cluster head placement with each of the clusters. Specifically, Silhouette method is used for the optimum number of clusters while SVM technique is used for the optimal gateway placement.

2.1 Computation of the Optimum Number of Clusters based on Silhouette Method

The optimum number of clusters must be determine first and then applied to determine the optimal gateway location using support vector Machines (SVMs) method. In this work the Silhouette score technique to compute optimal number of clusters. The Silhouette method is used to compute the optimum number of clusters. The flowchart for the Silhouette method is presented in Figure 1.

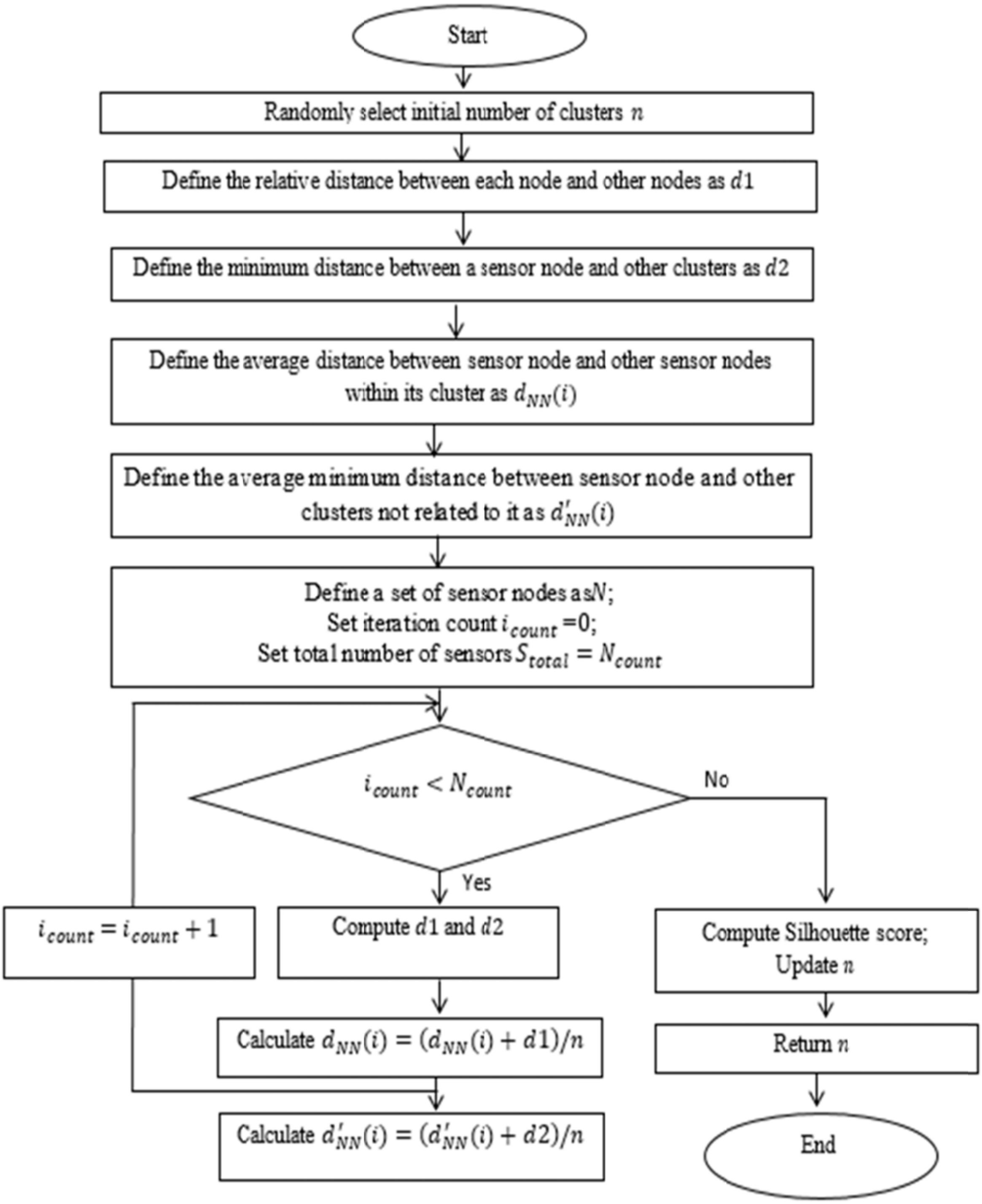

Figure 1 The flowchart for the Silhouette method

2.2 Determination of Optimal Gateway Location using Support Vector Machines (SVMs)

Supposed a dataset contains n number of points and can be represented as: $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, any specific feature y_i, in which a p dimensional vector x_i belongs can take the value of 1 or -1. For each data point's segment, $s_i \in S$, the goal is to locate a margin that differentiate all x_i corresponding to $y_1 = 1$ from all x_i corresponding to $y_1 = -1$. The difference is defined to maximize the Euclidean distance between x_i and the margin, from either side of the plane. Within the network space, the thick margin can be written in terms of x such that Equation 1 is satisfied

$$\eta^T x_i - b = 0 \quad (1)$$

where, η represents the normalization vector and b is the intercept vector.

The deviation of the thick margin from the center, with respect to η can be defined as $\frac{b}{\|\eta\|}$. Considering the fact that there could be spontaneous data fall off towards the thick margin, Equation 1 is constrained by the expression in Equation 2:

$$\begin{cases} \eta^T x_i - b \leq -1; & if\ y_i = -1 \\ \eta^T x_i - b \geq 1; & if\ y_i = 1 \end{cases} \quad (2)$$

This ensures that all the sensor nodes locations are maintained around the thick margin. The authentic expression of Equation 2 can be written as:

$$y_i(\eta^T x_i - b) \geq 1; \quad \forall\, 1 \leq i \leq n \quad (3)$$

Hence, the optimization problem can be defined as:

$$\underbrace{minimize}_{\eta, b} \|\eta\|_2^2 \quad (4)$$

The Equation 4 is subject to the constraint in Equation 3. The value of η and b for which the solution to Equation 5 is obtained determines the central and optimal location where the gateway must lie.

$$x \rightarrow \sigma(\eta^T x - b) \quad (5)$$

Where $\sigma(\cdot)$ denotes the sigmoid function. Two support vectors are extracted, each from both the positive and the negative sides of the boundary. The positional deviation vector between these two support vectors v_1 and v_2 can be computed as $\Delta P = v_2 - v_1$.

2.3 Parameters Selection for the Kernel: The kernel function used for the cluster head selection problem is the sigmoid function which is expressed as [19,20,21]:

$$\gamma(x_i, y_i) = \tanh(\gamma x_i \cdot y_i + c) \quad (6)$$

Where, γ denotes the training influencer, and c denotes the penalty factor. A higher c aids a more correct optimal point for the cluster head. Hence, γ must be properly selected to achieve the desired goal by imposing higher c. Considering an imperfect separable problem which is applicable to the real cases, certain data points are allowed to stay at some distance d_i from the thick margin, then Equation 4 can be written in expanded form as:

$$\underbrace{minimize}_{\eta, b, d} \frac{1}{2}\eta^T \eta + C \sum_{i=1}^{n} d_i \quad (7)$$

Subject to:

$$y_i(\eta^T x_i - b) \geq 1 - d_i; \quad \forall\, 1 \leq i \leq n; \quad d_i \geq 0\ \forall 1 \leq i \leq n \quad (8)$$

Equation 8 aims at reducing the thick margin which corresponds to $\eta^T \eta$. In this process, penalty is measured for misclassification. The weight of the penalty is controlled by the parameter C in Equation 7. Considering the two sides of the thick margin, the optimization problem can be defined as:

$$\underbrace{minimize}_{\alpha} \frac{1}{2}\alpha^T \alpha Z - \alpha e^T \quad (9)$$

Subject to;

$$y^T \alpha = 0; \quad 0 \leq \alpha_i \leq C,\ i = 1, \dots, n \quad (10)$$

Where, Z denotes a finite matrix having $n \times n$ dimension, e denotes vector in which all values are 1, α denotes the two-side coefficient. The finite matrix Z can be defined in terms of the kernel as:

$$Q_{ij} = x_i y_i \gamma(x_i, y_i) \quad (11)$$

The output of the optimization function yields the decision function for a given set of sample. The decision function is defined as:

$$f_d = \sum y_i \alpha_i \gamma(x_i, x) + b \quad (12)$$

In the implementation of the SVM model, the system parameters are exposed by the SVM object properties which are: $dual\_coeff$, $support\_vector$, and $intercept$. The complete algorithm is presented in Algorithm 1.

Algorithm 1: Optimal location of gateway using SVM

1: Begin
2: Define input dataset $\mathcal{X}$
3: Define η, b
4: Represent $\mathcal{X}$ in vector form and select x_i vector which satisfies Equation 1.
5: define the margin m
6: for each data-point in the network space
7: compute the deviation of each data point from the defined margin
8: if the deviation is greater the a threshold ε
9: Compel the data-point towards the margin using Equation 3 and the constraint in Equation 2
10. end if
11: end for
12: Compute $\gamma(x_i, y_i)$ based on Equation 6
13: end

3. RESULTS AND DISCUSSION

In order to evaluate the SVM technique, the design model is implemented in Python 3.12. The core libraries used for the implementation include: numpy, pandas, SVC, sklearn and Matplotlib. The experiment is performed for 5000 sensor nodes in an area of $A = 800m \times 800m = 640km^2$, as presented in Figure 2 for the case of 5000 sensor nodes.

The scatter plot of the results of the computation of the optimum number of clusters based on Silhouette method is shown in Figure 3. According to the result in Figure 3, there are five clusters identified by the Silhouette method. Hence, the SVM is used to determine the optimal location for the placement of the gateway ineach of the five clusters. The SVM is also used to assign each of the 5000 sensor nodes to their respective cluster gateway. The results for execution time for the optimal gateway location based on Silhouette method and SVM technique for 5000 sensor nodes are shown in Figure 4. In this case distribution means the random distribution of the sensor node. Specifically, the specific x and y coordinates of each of the 5000 sensor nodes are randomly generated a and the Silhouette method and SVM technique were applied to determine the optimum number of clusters and the optimal gateway location for each of the clusters and then the clustering of each of the sensor node to their respective cluster gateway (or cluster head). The random distribution of the 5000 sensor nodes is implemented again and the Silhouette method and SVM technique processes are repeated and this was done four time, hence the results for the four distribution shown in Figure 4. The results in Figure 4 show that the Silhouette method and SVM technique took an average execution time of 991.7825 seconds per distribution. In essence, it took approximately 991.7825 seconds to compute the optimum number of clusters using the Silhouette method, determine the optimal gateway locations for each of the clusters and also cluster the 5000 sensor nodes to their respective cluster gateway.

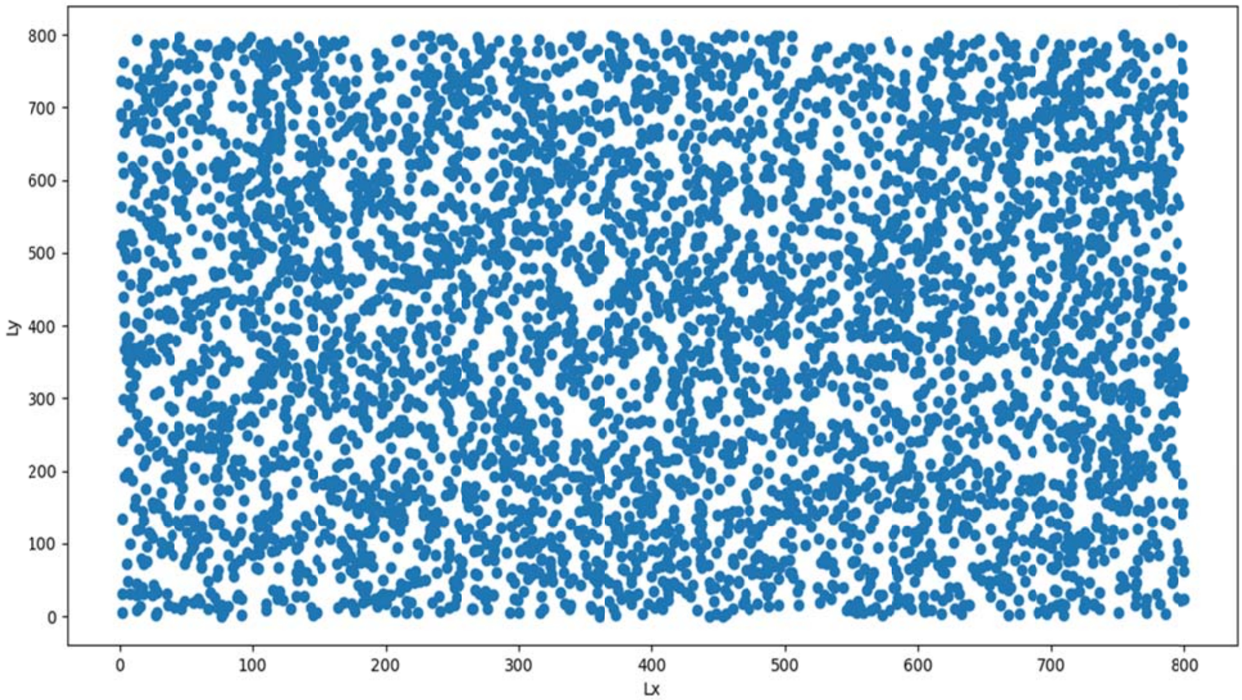

Figure 2: The distribution of 5000 sensor nodes on the case study area of $800m \times 800m = 640km^2$

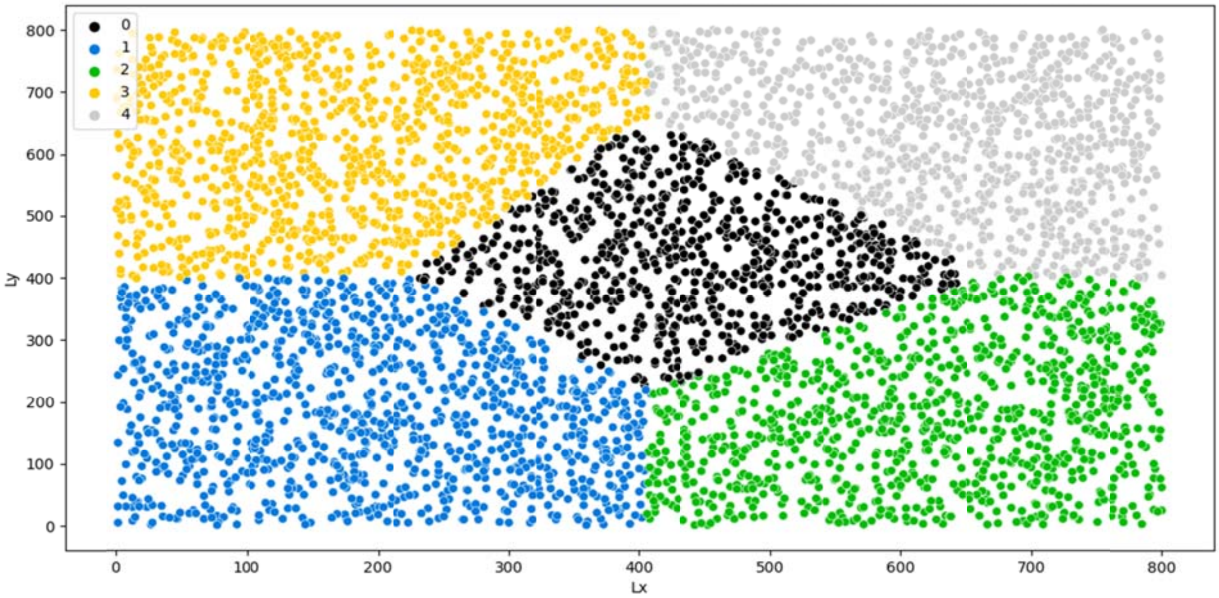

Figure 3: The scatter plot of the results of the computation of the optimum number of clusters based on Silhouette method

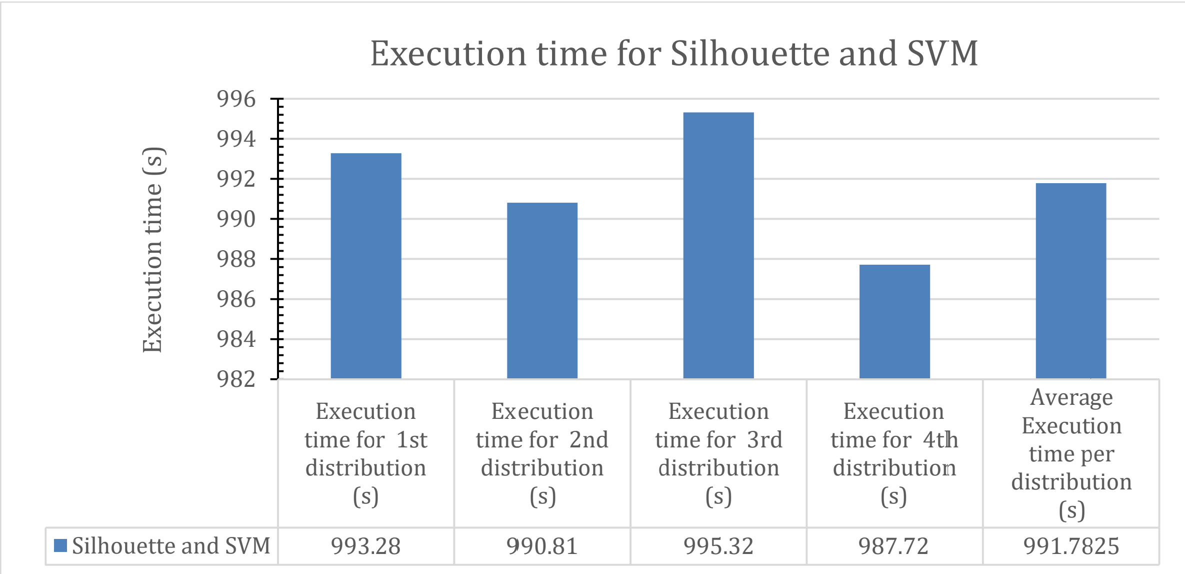

Figure 4: The bar chat for execution time optimal gateway location based on Silhouette method and SVM technique for 5000 sensor nodes

The gateway locations for sensor nodes distribution 1 are shown in Figure 5, for sensor nodes distribution 2 are shown in Figure 6, for sensor nodes distribution 3 are shown in Figure 7 and for sensor nodes distribution 4 are shown in Figure 8. Also, the locations of Gateway 1 for sensor nodes in all the 4 distributions are shown in Figure 9. In all it is evident that different random distribution of the sensor nodes give rise to different location for the gateways. However, in all the four case, the five clusters identified by the Silhouette method is used for the gateway placement implemented using the SVM technique.

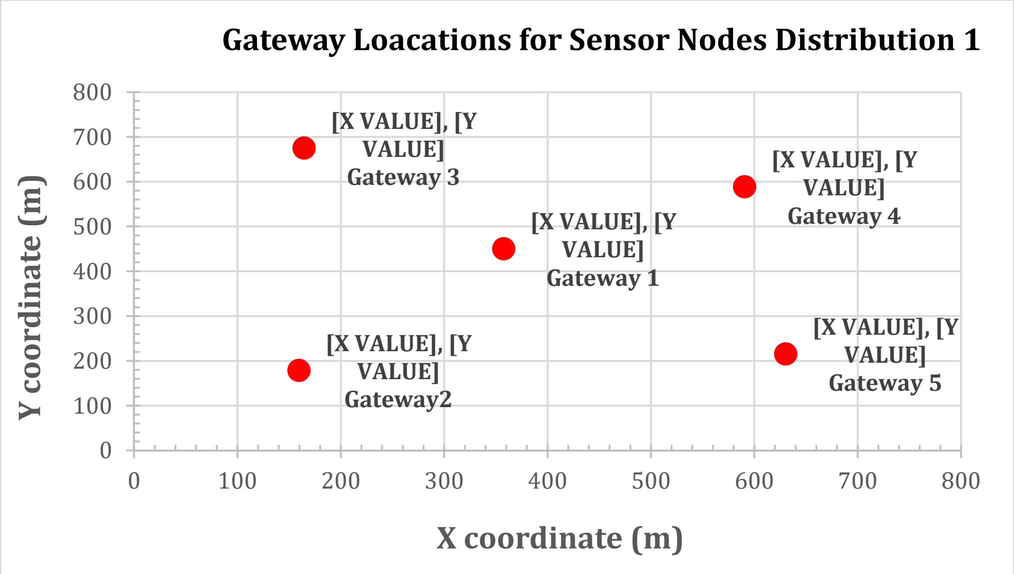

Figure 5 The gateway locations for sensor nodes distribution 1

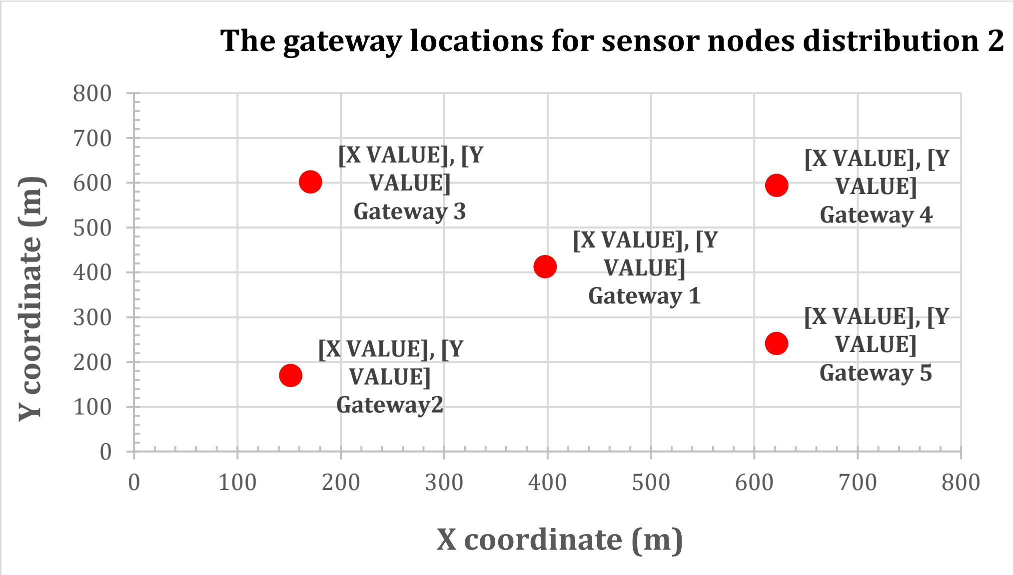

Figure 6 The gateway locations for sensor nodes distribution 2

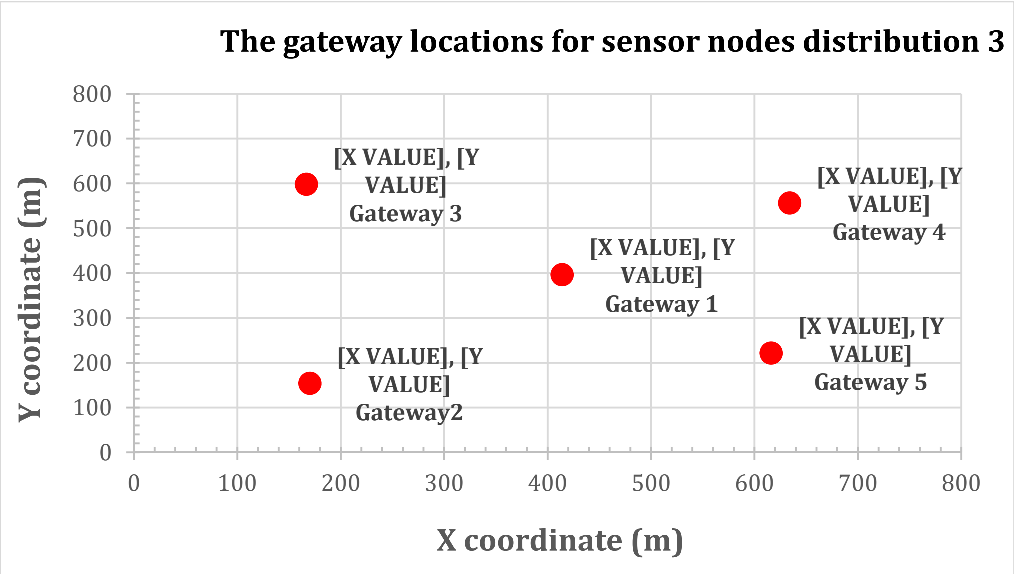

Figure 7 The gateway locations for sensor nodes distribution 3

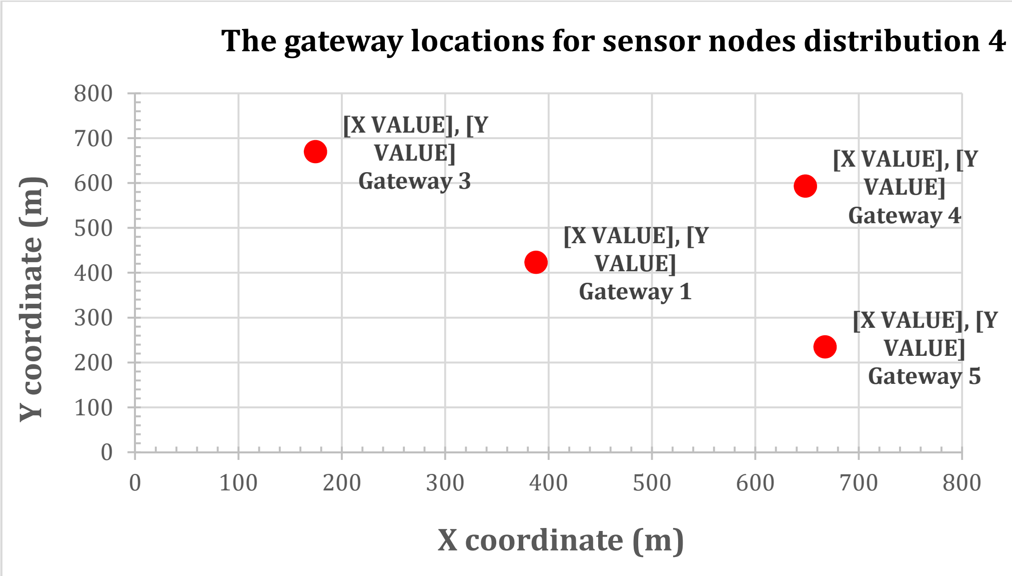

Figure 8 The gateway locations for sensor nodes distribution 4

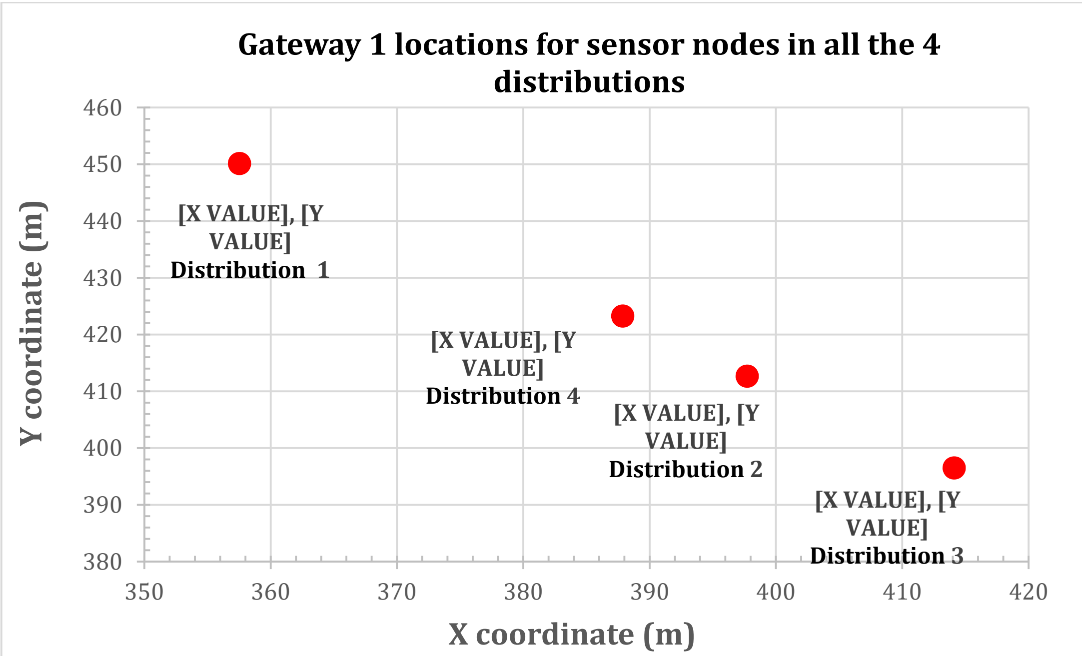

Figure 9 The locations of Gateway 1 for sensor nodes in all the 4 distributions

4. CONCLUSION

Evaluation of Silhouette method and SVM technique for application in clustering sensor nodes is presented. The study employed the Silhouette method for computing the optimum number of clusters required for set of sensor nodes distributed randomly in a given rectangular area. The Support Vector Machines is used to compute the optimal location for the gateway and also to cluster each of the sensor nodes to their respective cluster heads. The simulation was conducted for four different distribution of the sensor nodes within the same area and the execution time was determined for each of the four distribution. In all, the results obtained from the study show that the Silhouette method and SVM technique can be used for such clustered sensor nodes which is common in smart systems applications.

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. 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.

4. Malhotra, P., Singh, Y., Anand, P., Bangotra, D. K., Singh, P. K., & Hong, W. C. (2021). Internet of things: Evolution, concerns and security challenges. *Sensors*, *21*(5), 1809.
5. Lakhwani, K., Gianey, H. K., Wireko, J. K., & Hiran, K. K. (2020). *Internet of Things (IoT): Principles, paradigms and applications of IoT*. Bpb Publications.
6. Goel, S. S., Goel, A., Kumar, M., & Moltó, G. (2021). A review of Internet of Things: qualifying technologies and boundless horizon. *Journal of Reliable Intelligent Environments*, *7*(1), 23-33.
7. Benmahdi, M. B., & Lehsaini, M. (2020). Performance evaluation of main approaches for determining optimal number of clusters in wireless sensor networks. *International Journal of Ad Hoc and Ubiquitous Computing*, *33*(3), 184-195.
8. 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.
9. Punhani, A., Faujdar, N., Mishra, K. K., & Subramanian, M. (2022). Binning-based silhouette approach to find the optimal cluster using K-means. *IEEE Access*, *10*, 115025-115032.
10. Ioannou, C., & Vassiliou, V. (2021). Network attack classification in IoT using support vector machines. *Journal of sensor and actuator networks*, *10*(3), 58.
11. Mardenov, Y., Adamova, A., Zhukabayeva, T., & Othman, M. (2023). Enhancing Fault Detection in Wireless Sensor Networks Through Support Vector Machines: A Comprehensive Study. *Journal of Robotics and Control (JRC)*, *4*(6), 868-877.
12. Bansal, M., Goyal, A., & Choudhary, A. (2022). A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. *Decision Analytics Journal*, *3*, 100071.
13. Raghavendra, Y. M., & Mahadevaswamy, U. B. (2021). Energy efficient intra cluster gateway optimal placement in wireless sensor network. *Wireless Personal Communications*, *119*(2), 1009-1028.
14. Chen, Q., Yang, L., Guo, J., Liu, X., & Chen, X. (2022). Optimal gateway placement for minimizing intersatellite link usage in LEO megaconstellation networks. *IEEE Internet of Things Journal*, *9*(22), 22682-22694.
15. Zhou, D., Sheng, M., Wu, J., Li, J., & Han, Z. (2021). Gateway placement in integrated satellite–terrestrial networks: Supporting communications and Internet of Remote Things. *IEEE Internet of Things Journal*, *9*(6), 4421-4434.
16. Alaerjan, A. (2023). Towards sustainable distributed sensor networks: An approach for addressing power limitation issues in WSNs. *Sensors*, *23*(2), 975.
17. Ijemaru, G. K., Ang, K. L. M., & Seng, J. K. (2022). Wireless power transfer and energy harvesting in distributed sensor networks: Survey, opportunities, and challenges. *International journal of distributed sensor networks*, *18*(3), 15501477211067740.
18. Zhao, T., Wang, L., Chin, K. W., & Yang, C. (2020). Routing in energy harvesting wireless sensor networks with dual alternative batteries. *IEEE Systems Journal*, *15*(3), 3970-3979.
19. Perumal, T. S. R., Muthumanikandan, V., & Mohanalakshmi, S. (2021). Energy efficiency optimization in clustered wireless sensor networks via machine learning algorithms. In *Machine Learning and Deep Learning Techniques in Wireless and Mobile Networking Systems* (pp. 59-77). CRC Press.
20. Achirul Nanda, M., Boro Seminar, K., Nandika, D., & Maddu, A. (2018). A comparison study of kernel functions in the support vector machine and its application for termite detection. *Information*, *9*(1), 5.
21. Panja, S., Chatterjee, A., & Yasmin, G. (2019). Kernel functions of svm: A comparison and optimal solution. In *Advanced Informatics for Computing Research: Second International Conference, ICAICR 2018, Shimla, India, July 14–15, 2018, Revised Selected Papers, Part I 2* (pp. 88-97). Springer Singapore.