# A COMPARATIVE ANALYSIS OF SELF-ORGANIZING MAP AND K-MEANS MODELS FOR SELECTION OF CLUSTER HEADS IN OUT-OF-BAND DEVICE-TO-DEVICE COMMUNICATION

**Florence Kingsley Atakpo[1]**
Department of Electrical/Electronic and Computer Engineering, University of Uyo, Akwa Ibom, Nigeria

**Ozuomba Simeon[2]**
Department of Electrical/Electronic and Computer Engineering, University of Uyo, Akwa Ibom, Nigeria
simeonoz@yahoomail.com
simeonozuomba@uniuyo.edu.ng

**Stephen Bliss Utibe-Abasi[3]**
Department of Electrical/Electronic and Computer Engineering, University of Uyo, Akwa Ibom, Nigeria

*Abstract*— Device-to-device communication is necessary to enhance signal transmission and to ensure effective communication. For this to be achieved, clustering of the cellular devices was necessary. In this paper, a comparative analysis of Self-organizing map (SOM) and K-means clustering algorithms for selection of cluster heads in out-of-band device-to-device communication is presented. Particularly, about 1000 cellular devices were considered in this study, where the devices were randomly spread over a region of 2000 $m^2$. The x-y coordinate position of the devices and the hardware capacity of the device were generated via normal distribution data generation format in MATLAB. The base station occupied a 1000 m by 0 m x-y coordinate. Out of the 1000 network devices (nodes) considered in the study area, the number of cluster heads selected with SOM cluster algorithm was 100. The remaining devices (nodes) not considered as cluster heads (cluster slaves) were clustered to the cluster heads. Similarly, K-means, being a conventional clustering algorithm was also used to cluster the system and was only able to select only two cluster heads out of the 1000 network devices (nodes) considered. Accordingly, the SOM algorithm performed much better than the K-means. SO, the SOM algorithm is recommended for selection of cluster heads in out-of-band device-to-device communication in order to enhance transmission of signal and quality of service in such communication systems.

***Keywords***— **Device-to-device communication, clustering, Self-organizing map (SOM), K-means clustering, 5G network, Cluster Head, Hardware Capacity**

## I. INTRODUCTION

The wide use of wireless applications, such as computing, surfing the Internet, and downloading and watching digital multimedia has created a large demand for high-speed and efficient wireless communication technology [1,2]. Fifth generation (5G) is the next-generation mobile communication system that is being developed for the expected demand of information and communication after 2020. It will have higher spectrum utilization and transmission rate, significantly improved transmission delay and quality of service (QoS) perception, and an increased number of access links and security [3]. The use of device to device (D2D) communication allows the increase in system effectiveness of cellular communication; moreover, D2D directly influences at system level both efficiency and energy [4,5,6]. The users are distributed on the base stations (BS) coverage area randomly. Generally, network planning takes into account distribution of nodes in the geographical area letting operator provide at least wanted coverage and required throughput and QoS.

Integration of D2Dcommunication technology became a mainstream direction for 5G communication networks. Driven by a huge increase in demand of multimedia traffic transfer, D2D communication allows saving scarce network resources by transferring data directly between devices either through in-band or out-of-band, and it also allows significant reduction in traffic, between base stations and end-user device [7,8,9].

The combination of D2D communication technology and a cognitive radio (CR) can effectively reduce interference [10, 11]. CR technology, through interactions with the external environment in terms of multidimensional spectrum detection, as well as real-time and interactive environments, is able to perceive any interference and make subsequent judgments so that cognitive users can choose the most appropriate communication frequency to avoid interference to primary users under the condition of the spectrum with the primary users sharing. How to manage the spectrum resources of the community as a whole, reasonably determine the communication power of each device, and minimize the interference between devices have become the main bottleneck for D2D communication to enter the practical stage. It is worthwhile to try to achieve power stability or learn from software reliability prediction through context sensitive rate Boolean control network [10, 11].

## II. REVIEW OF RELATED WORKS

According to the 3rd-Generation Partnership Project (3GPP), D2D is a flexible paradigm of direct communication between devices which is open for use and based on cellular communication technologies (in-band D2D communication) and also WLAN technologies which are IEEE 802.11 standardization (out-of-band D2D communication) device [7,8,9]. This significantly complicates feasibility of in-band D2D wide scale implementation at least for the time being. Out-of band D2D can be easily implemented with network assistance option; hence cellular operators are able to control out-of band sessions. For the obvious reasons, IEEE 802.11-based Wireless Fidelity (Wi-Fi) is taken as the transmission technology for implementation of out-of-band D2D functionality [7,8,9]. Operators of communication networks can encourage regular users to use D2D technology in order to improve the overall performance of the communication system in return for rewards proportional to their contributions.

Typically, geographically, D2D nodes can form a cluster (see Figure 1), where traffic circulates between cluster nodes directly, and outside-of-cluster traffic is forwarded to BS via relay node, so-called cluster head. A number of algorithms for cluster head selection are available today [7,8,9].

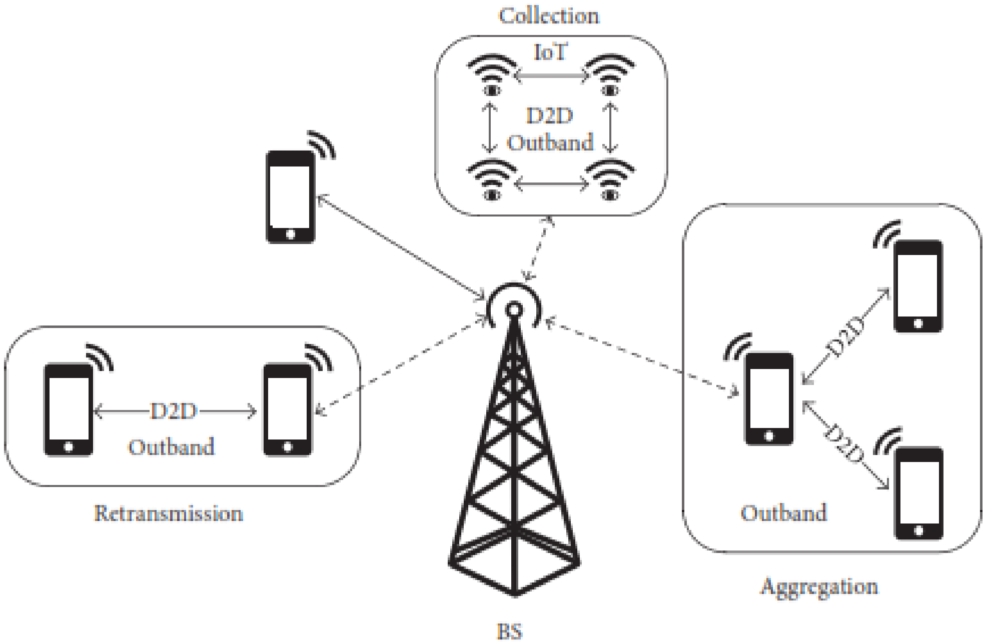


**Figure 1: Out of Band Clustering of D2D components**
**Source: Paramonov, *et al*. [9].**

The decision on selection of a particular cluster member as a cluster head affects at least network efficiency, energy expenditures, and quality of service (QoS) offered to all members within the cluster. Generally, if all data transfer where end-user device shall be equipped with appropriate members are the members of the same cluster, then the cluster can operate off-line, (that is, without connection to a BS). A larger number of cluster members are expected to lead to larger savings of the network resources. However, the maximum number of members in a cluster is restricted by coverage of selected D2D technology, channel throughput of cluster head and cluster traffic intensity, and cluster members physical location towards cluster head. Existing studies show that D2D clustering in 5G leads to reduction of signaling traffic and provides higher spectral efficiency and better energy performance than conventional cellular systems [12,13].

Thus, efficient D2D clustering in 5G networks especially with high density of devices is of a importance. Some past works have concentrated on quantitative and qualitative analysis of cluster algorithms for D2D communications. In [14] the authors provide comprehensive analysis of D2D communications. The use of out-of-band D2D communications and D2D clustering is discussed in detail given criteria of cluster head selection based on channel quality between cluster head and BS.

In [15] the authors designed clustering algorithm for in-band D2D case, which increases system-level spectral efficiency. Numerical analysis and simulation modeling have shown that this proposal gives66%gain in terms of through put compared to traditional solutions, in the case where 20% of users use D2D communication. The authors derived the probability density formula (pdf) for the optimal number of repeater units in the cluster and have come up with the cross-cluster interaction scheme.

Also, via simulation the authors show that the proposed algorithm provides gains up to 40% in terms of network efficiency of resource use. uplink/downlink functionality. In case of in-band D2D communications, the transmission power should be properly regulated so that the D2D transmitter does not interfere with different aspects of the out-of-band D2Dcommunication are presented in [16]. In [17] the authors developed analytical model of the network unloading for different D2D scenarios using stochastic geometry. The authors estimated potential opportunities of the out-of-band D2D communication, using both the system level and the mathematical analysis. They show that at 30% of clustering productivity and energy network performance increase up to four and two times, respectively.

In [18] the authors studied problems of implementation of network-assisted D2D communication while interacting in social networks. Besides, they use the existing experimental LTE test bed for implementation of D2D system and show its performance evaluation in terms of latency and users' satisfaction [19].

D2D transmission technology selection is still rather limited to Wi-Fi and Bluetooth due to wide implementation of those in consumer devices. Recent works consider D2D devices forming clusters by using WiFi Direct (see Figure 2) [8].

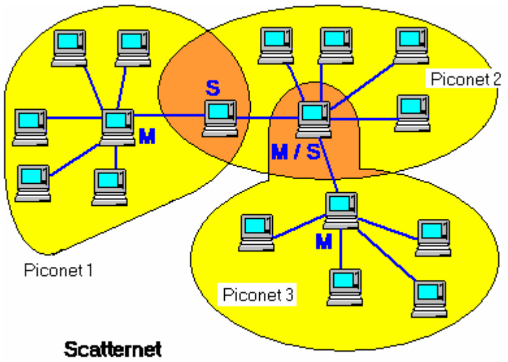


**Figure 2: Bluetooth piconet and scatternet structure**
**Source: Nitti *et al*., [8]**

Due to features of radio channel, resources of channel between cluster member and cluster head may drastically vary for different nodes within one cluster. Therefore, while forming cluster, the way cluster heads are selected, shall be based primarily on anticipated QoS parameters, not distance alone suggested in recent studies [8]. The same applies to selection of cluster members. Meanwhile, clustering algorithm can be implemented for different target parameters such as cumulative throughput of cluster as a whole, maximum number of cluster nodes, and quality of service. Faced with the great prospect of applications with wireless D2D transmission in personal, public and industrial areas, many competitive out-band D2D technologies have already been developed. In this paper, the distance and the hardware capability of the nodes are considered simultaneously in order to determine if the node device is suitable to serve as a cluster head.

## III. METHODOLOGY

The methodology involves simulation with stochastically generated data. Generally, the data needed in the research includes the location of the mobile (handset) device with respect to the base station, the device (hardware) capacity and the received signal strength. Specifically, the location of the mobile (handset) device with respect to the base station is stochastically generated using random number that is exponentially distributed. Also, the device (hardware) capacity is stochastically generated using random number that is exponentially distributed. However, the received signal strength is computed using the statistically generated device location (distance) from the base station.

In the research two different clustering algorithms were employed to select the cluster heads from among devices in a given cellular network and the clustering algorithms were also employed to assign different devices in the network to the selected cluster heads. The selection of the cluster heads is based on the distance of the nodes from the base station and the strength of the signal they received along with the hardware capacity of the phone. The flow diagram for the research process is presented in Figure 3.

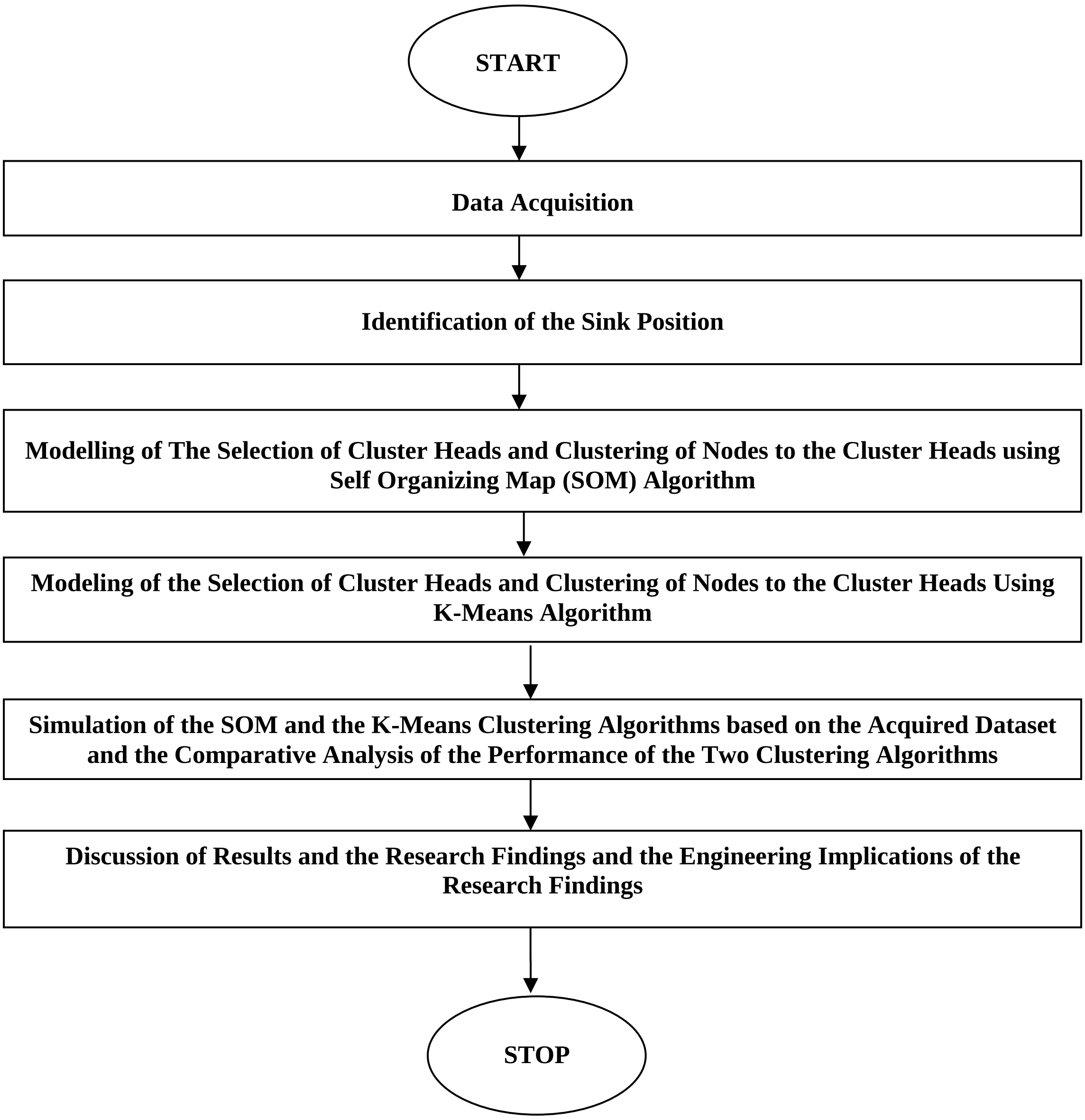


**Figure 3: Flow diagram for the Research Process.**
**Source: Formulated by the researcher (2019).**

### *A. The Case Study Data*

In this research, 1000 cellular device users are considered and they are located within a square distance of 2000 metres around the base station. The position of the base station was placed at the origin of the y-axis and the center of the x-axis (0 by 1000 metres) with the different cellular users scattered around it (in order words the x-y position of the base station is 1000 m by 0 m (1000,0). The location of each of the 1000 cellular device users from the base station was generated in MatLab using the random number generator with normal distribution. The expression for the random number generator with normal distribution is given as follows:

$$randnd = lb + (ub - lb) \times rand(1000,1) \quad (1)$$

where $randnd$ is the randomly generated data for normal distribution. Lb is the lower boundary which indicates the minimum point of the location of the devices from the base station (0m) and ub is the upper indicating the maximum distance of the device from the base station (2000m). The graphical representation of the generated x-y coordinate position of each of the cellular user device is shown in Figure 4. Also, the MatLab codes for generating and plotting the device location data in the command window of MATLAB 2015a is as shown in Figure 5.

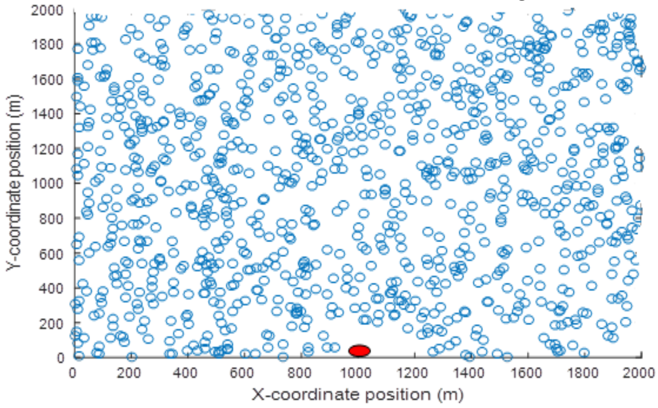


**Figure 4: Device to Device network**

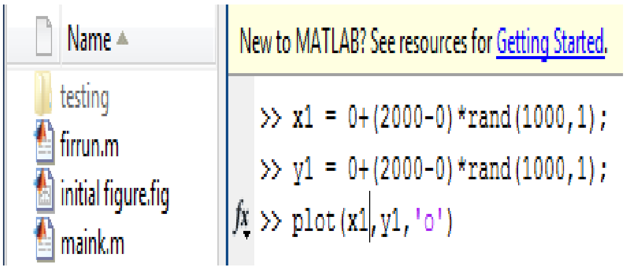


**Figure 5: The MATLAB codes for generating the device location random number generator with normal distribution.**

The distance of each of the devices from the base station was computed using Pythagoras formula , given as;

$$d = \sqrt[2]{x^2 + y^2} \quad (2)$$

where x any are the coordinates of the location of the device with respect to the base station, (x and y are distance in metres) while d is the straight-line distance from the device to the base station, again, d is in meter. Also, the received signal strength of each of the devices was computed using link budget equation. Specifically, the Received Signal Strength (RSS) value at each of the distances from base station is denoted as ($PL_{m(dB)}$) and is computed as follows [20];

$$P_r(\text{dBm}) = EIRP_t(\text{dBm}) - PL_{m(dB)} \quad (3)$$

where: $PL_{m(dB)}$ is the path loss foreach location at a distance *d* ( km).

In this research, the path loss is the free-space path loss ($PL_{FSPL(dB)}$), $EIRP_t$ is the Effective Isotropic Radiated Power in dBm and $P_r$ is the mean Received Signal Strength (RSS) in dBm. The free-space path loss ($PL_{FSPL(dB)}$) is given as[20];

$$PL_{FSPL(dB)} = 32.5 + 20 \log f + 20 \log d \quad (4)$$

where: f is the frequency in MHz, and d is the link distance in Km. The effective isotropic radiated power $EIRP_t$ (dBm) is given as [20];

$$EIRP_t = P_{BTS} + G_{BTS} + G_{MS} - L_{FC} - L_{AB} - L_{CF} \quad (5)$$

where: $P_{BTS}$ = Transmitter Power (dBm), $G_{BTS}$ = Transmitter Antenna Gain (dBi), $G_{MS}$ = receiver antenna gain (dBi), $L_{FC}$ = feeder cable and connector loss (dB), $L_{AB}$ = Antenna Body Loss (dB) and $L_{CF}$ = Combiner and Filter Loss (dB). The values of these parametres are given as [20];

$P_{BTS}$ = 40 W = [30 + 10Log$_{10}$ 40] = 46 dBm, $G_{BTS}$ = 16 dBd = [16 + 2.15] = 18.15 dBi, $G_{MS}$ = 0 dBi, $L_{FC}$ = 3 dB, $L_{AB}$ = 3 dB, $L_{CF}$ = 4.7 dB. Hence, the $EIRP_t$ becomes; $EIRP_t$ = 46 + 18.15 – 3 – 3 – 4.7 = 53.5 dBm. Therefore,

$$P_r(\text{dBm}) = 53.5 - PL_{m(dB)} \quad (6)$$

Furthermore, the hardware capacity of each of the 1000 cellular device users was generated in MatLab using random number generator with normal distribution. In practice, the hardware capacity of mobile devices ranges from 1 to 5 as given by the following composite relations;

$$\left.\begin{array}{l} 4 < a \leq 5;\ \text{means a very strong capacity.} \\ 3 < a \leq 4;\ \text{means strong capacity} \\ 2 < a \leq 3;\ \text{means fairly strong hardware capability} \\ 1 < a \leq 2;\quad \text{means poor capability} \\ a \leq 1;\quad \text{very poor capability} \end{array}\right\} \quad (7)$$

where a is the coded value of the hardware capability of the cellular devices. In respect of the range of values for the hardware capacity of the mobile devices, the expression used for the random number generator with normal distribution is given as follows:

$$randnd = 1 + (5 - 1) \times rand(1000{,}1) \quad (8)$$

***B. Clustering With Self-Organizing Map and K-Means Algorithms***

The steps used in this paper to implement the SOM and K-Means algorithms are shown in Figure 6.

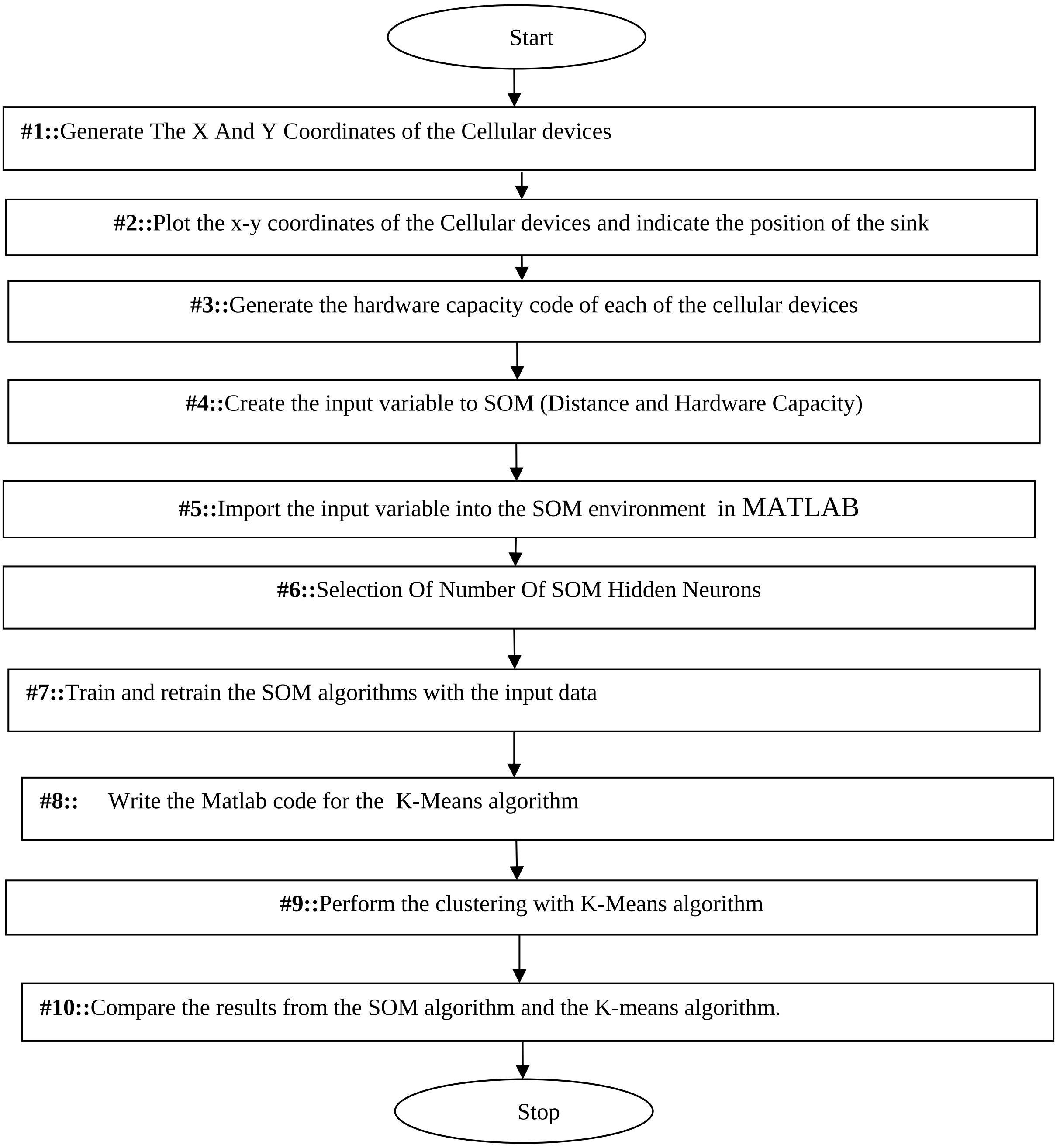


**Figure 6: The steps used in this paper to implement the SOM algorithm.**

**Step One: Generate the X and Y Coordinates of the Cellular devices**

In step one, the location of each of the 1000 cellular devices is generated using the random number expression given in Equation 1 where the value of ub = 1000 and lb = 0. The implementation in MATLAB generated an array of 1000 rows by one column data of the cellular devices location for the x and y coordinates which are then plotted in step two of the flowchart using the plot(x,y) commanc in MATLAB.

**Step Three: Generate the hardware capacity code of each of the cellular devices**

In step three, the hardware capacity code (a) is determined based on Equation 8. The value of ranges from 1 to 5, as shown in Equations 7 and 8 which are implemented in MATLAB as shown in Figure 7.

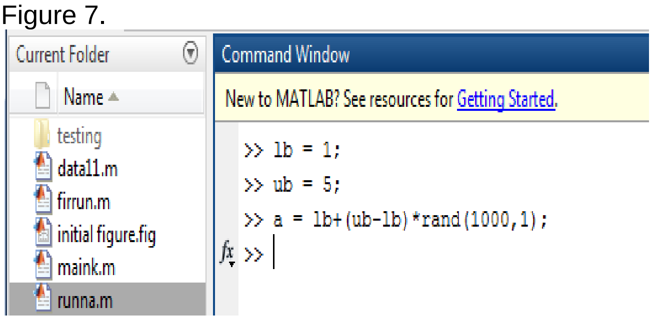


**Figure 7 : The code in MATLAB for generating the values of hardware capacity code (a) of each of the cellular devices.**

In step four the input variable to SOM (distance and hardware capacity) is created. Specifically, the x and y coordinates of the location of the cellular devices relative to the origin is used to determine the distance d, based on Equation 2. At the same, the distance, d is used to compute the received signal strength, Pr(dBm). Then, the variable for the input dataset to be used for the SOM algorithm (in this case, r as shown in Figure 8) is created. The variable r will contain the 1000 two column matrix with one column containing the generated value of d and the other column containing the generated value of hardware capacity code (a), as shown in Figure 8, where d is the distance and a is the hardware capacity.

The MATLAB SOM toolbox GUI is called up using nctool command, as shown in Figure 8. The architectural network used for the SOM has the input layer, the hidden SOM neuron layer and the output layer. In step five the input variable is imported into the SOM environment in MATLAB.

The variable r that was created in Figure 8 contains the input 1000 x 2 matrix dataset for the SOM training. In this section, the dataset is loaded for the training of the cluster algorithms. In step six, number of SOM hidden neurons is selected. In this study, 10 hidden neurons is selected which gave 100 output neurons for the SOM algorithm.

Train and retrain of the SOM algorithms with the input data is performed in step seven. After the data is loaded and the neurons are set, the SOM algorithms is trained and retrained until acceptable minimal error value is obtained. After the training and retraining were done, the cluster slaves are automatically attached to the cluster heads by the SOM algorithm.

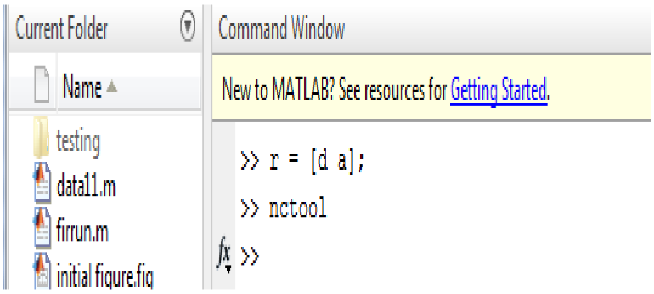


**Figure 8: The Input variable to SOM for the selection of cluster heads from the cellulardevices.**

In step eight, the clustering with K-Means algorithm is conducted. Since the data is already load in the MATLAB, the relevant MATLAB toolbox and code for the K-means clustering algorithm are used to implement the selection of the cluster heads and at the same time clustering of the slave devices to the cluster heads. The screenshot for the MATLAB code used to implement the K-means clustering algorithm is shown in Figure 9.

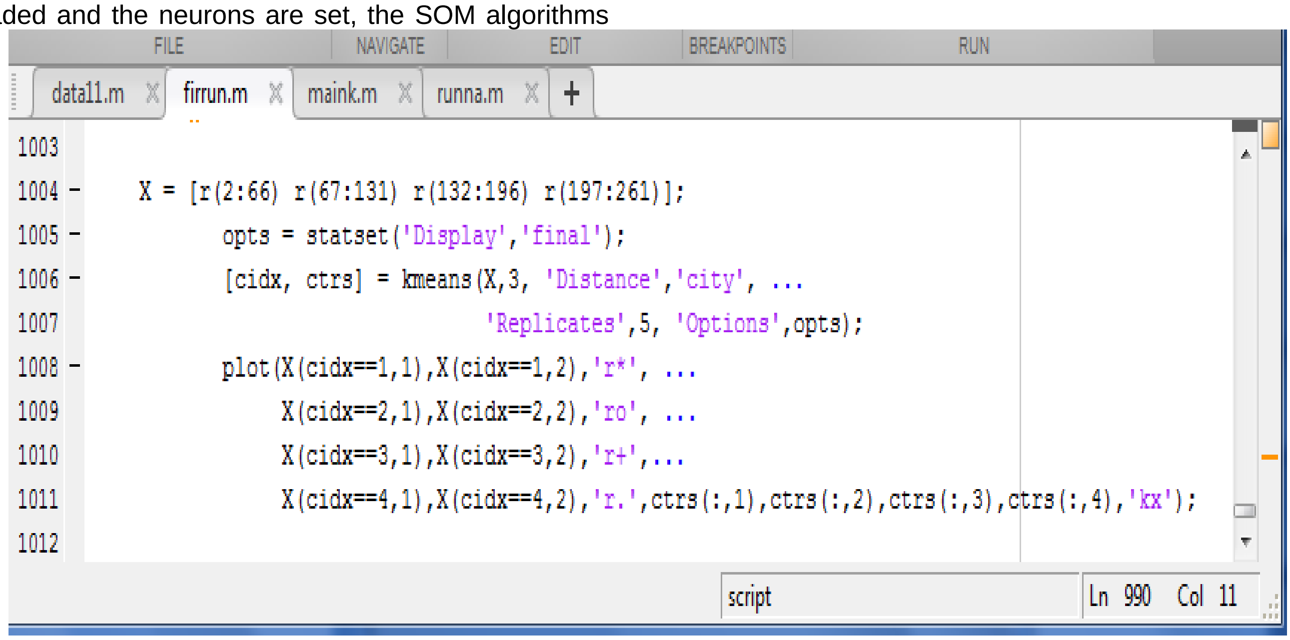


**Figure 9: The screenshot for the MATLAB code used to implement the K-means clustering algorithm**

## IV. RESULTS AND DISCUSSION

The graph in Figure10 shows the randomly generated clustered data points plotted with the number of cellular device users against the distance from the sink (base station).

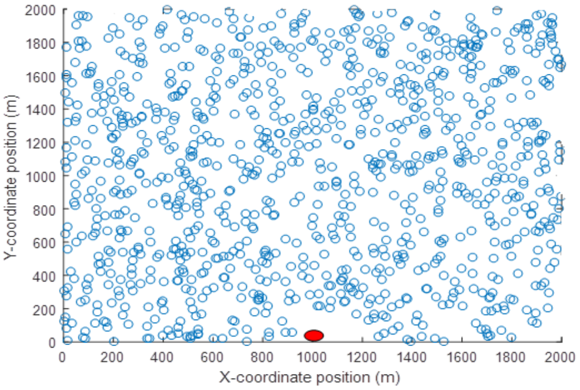


**Figure 10: The obtained data of cellular device users with their distance from the sink.**

A total of 1000 cellular device users were randomly clustered within a distance of 0 to 2000 meters away from the base stations (for y-coordinates) and 0 to 1000 meters from the base station for x-coordinates in a 2000 x 2000 $m^2$ region. The base station (sink) is situated at the base of the region at 0 m of y-coordinates and 1000m (center) of x-coordinates. With the SOM algorithm a total of 100 cluster heads were generated and the remaining cluster slaves were assigned to each cluster head. The generated cluster heads topology is shown in Figure 11.

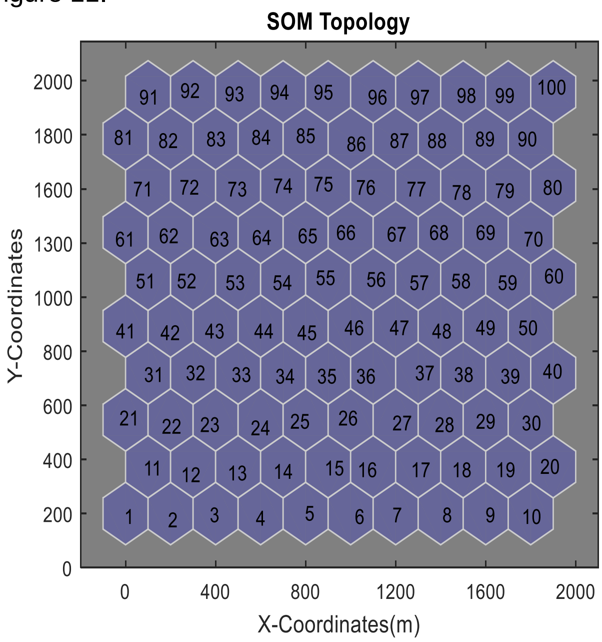


**Figure 11: SOM topology of the cluster heads**

The diagram shown in Figure 11 is the topology of the 100 cluster heads after the input data had being trained and retrained through the SOM layer. The other devices not considered as cluster heads (cluster slaves) were assigned to the cluster heads with SOM algorithm. The neighbor connectivity of the cluster heads is displayed in Figure 10.

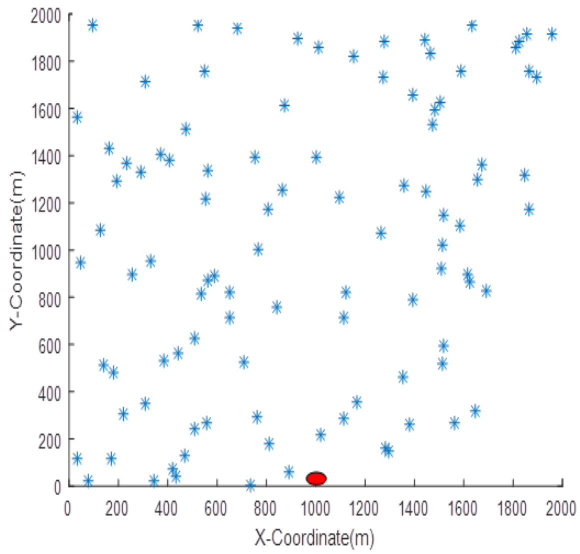


**Figure 10: X-Y coordinate position of the SOM selected cluster heads.**

The advantage of the SOM intelligent clustering algorithm is not just its efficiency is selecting higher cluster heads but its ability to identify the cluster heads with a high mobile device hardware capacity spread across the 2000 square metre of the region. The hardware capacity of each mobile device is shown in Figure 11.

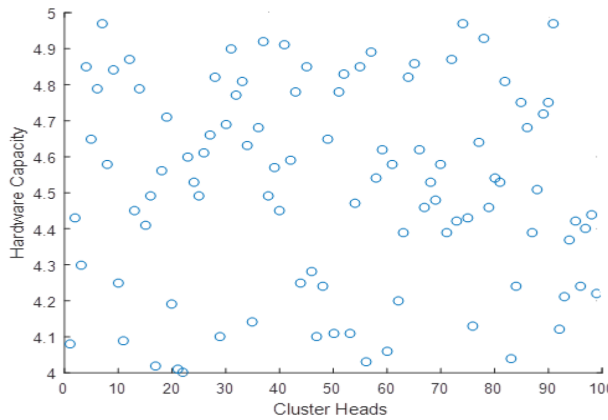


**Figure 11: Hardware capacity of the cluster heads.**

The number of cluster slaves per cluster head is shown in Figure 12 and the data is plotted in Figure 13. According to Figure 12 and Figure 13, the cluster head with highest number of cluster slaves was the twelfth cluster head whose x and y coordinate values are 78.562 m and 24.455 m respectively. The cluster head has a hardware capacity of 4.09.

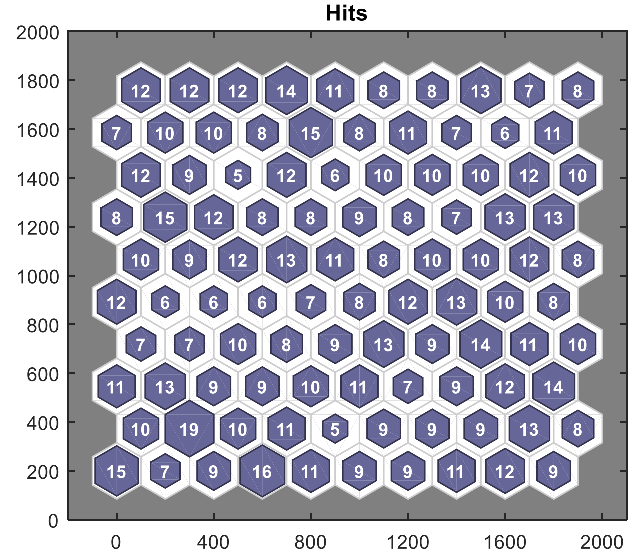

Figure 12: Number of cluster slaves to cluster heads selected by SOM

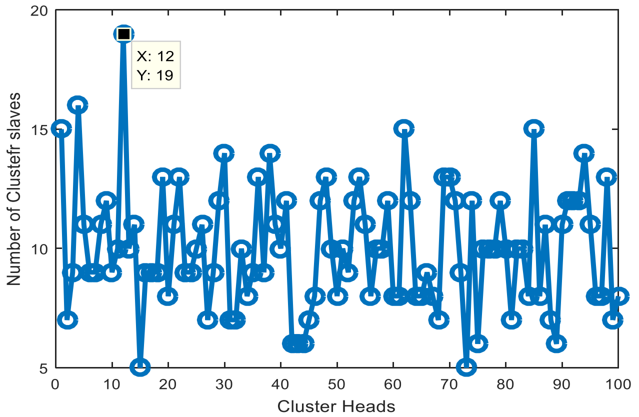

Figure 13: Plot showing the number of cluster slaves in a cluster head.

In addition to the Self-organizing map (SOM), K-means was also used to cluster the cellular devices to their cluster heads and the result is shown in Figure 14. Unlike the SOM, only two cluster heads was selected by the K-means. Also, despite having the least number of cluster heads, K-means was not able to properly and accurately cluster the cluster slaves to the cluster heads when compared with the SOM.

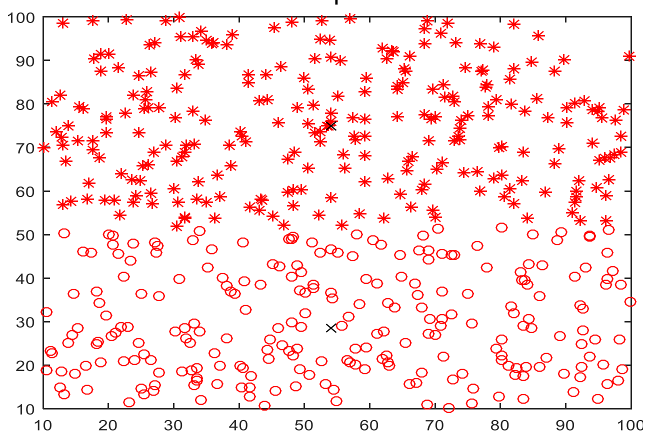

Figure 14: Clustering with k-means.

## V. CONCLUSION

The paper presented two different clustering algorithms, namely, Self-organizing map (SOM) and K-means. The two algorithms were used in selecting the cluster head for a device –to-device communication. Notably, the superiority of artificial intelligent clustering algorithm (SOM), over conventional cluster algorithm was proven. In the study, the results show that the Self-organizing map (SOM) selected higher number of cluster heads to ensure effective device to device communication to curb the inefficient direct transmission from the base station to the cellular users. On the other hand, the number of cluster heads obtained from the k-means was not sufficient in ensuring adequate signal transmission among the cellular devices. Also, SOM selected devices via the hardware capacity irrespective of the position of the device from the base station to ensure that the signal transmitted reaches the cellular devices that are far away from the base station.

## REFERENCES

1. De Rango, Floriano, Mauro Tropea, and Peppino Fazio. "Multimedia Traffic over Wireless and Satellite Networks." *Digital Video*. IntechOpen, 2010.
2. Yadav, Swati, and Sugandha Singh. "Review Paper on Development of Mobile Wireless Technologies (1G to 5G)." *Int. J. Comput. Sci. Mob. Comput* 7 (2018): 94-100.
3. Anjum, I., Razzaque, M., Hassan, M., Alelaiwi, A. and Rahman, S. (2016). Quality-of-service- aware weighted-fair medium access control protocol for coexisting cognitive radio networks. *EURASIP Journal on Wireless Communications and Networking*, 2016(1): 77.
4. Della Penda, Demia. *Device-to-Device Communication in Future Cellular Networks: Resource allocation and mode selection*. Diss. Kungliga Tekniska högskolan, 2018.
5. Chung, Yao-Liang. "Energy-efficient use of licensed and unlicensed bands in D2D-assisted cellular network systems." *Energies* 10.11 (2017): 1893.
6. Gui, Jinsong, and Kai Zhou. "Cellular throughput optimization by game-based power adjustment and outband D2D communication." *EURASIP Journal on Wireless Communications and Networking* 2018.1 (2018): 254.
7. Kar, U. and Sanyal, D. (2017). An overview of device-to-device communication in cellular networks. *ICT express,* 10(6): 20-23.
8. Nitti, M., Stelea, G., Popescu, V. and Fadda, M. (2019). When social networks meet D2D communications: A survey. *Sensors*, *19*(2): 396.
9. Paramonov, A., Hussain, O., Samouylov, K., Koucheryavy, A., Kirichek, R. and Koucheryavy, Y. (2017). Clustering optimization for out-of-band d2d

communications. *Journal of Wireless Communications and Mobile Computing*, 32(11): 54-58.

10. Amjad, M., Akhtar, F., Rehmani, M., Reisslein, M. and Umer, T. (2017). Full-duplex communication in cognitive radio networks: A survey. *IEEE Communications Surveys & Tutorials*, *19*(4): 2158-2191.
11. Jang, S. and Yoo, S. (2018). Q-learning-based dynamic joint control of interference and transmission opportunities for cognitive radio. *EURASIP Journal on Wireless Communications and Networking*, 11(8): 160.
12. Kim, S. (2019). A New Multicasting Device-to-Device Communication Control Scheme for Virtualized Cellular Networks. *Journal of Wireless Communications and Mobile Computing*, 11(15): 54-57.
13. Gandotra, P. and Jha, R. (2016). Device-to-device communication in cellular networks: A survey. *Journal of Network and Computer Applications*, *7*(14): 99-117.
14. Do, D., Van-Nguyen, M., Hoang, T. and Lee, B. (2019). Exploiting Joint Base Station Equipped Multiple Antenna and Full-Duplex D2D Users in Power Domain Division Based Multiple Access Networks. *Sensors*, 19(11): 2475.
15. Chevillon, R., Andrieux, G. and Diouris, J. (2018). *Overlay Inband D2D-e Network using Fuzzy C-Means clustering for disaster situations.*European Conference on Networks and Communications (EuCNC) held at London, United Kingdom from 7th-9th 2018, 292p.
16. Nugraha, T., Pamungkas, M. and Chamim, A. (2018). Interference Management using Power Control for Device-to-Device Communication in Future Cellular Network. *Journal of Telecommunications and Information Technology*, 12(22): 65-68.
17. Poulakis, M., Gotsis, A. and Alexiou, A. (2018). Multicell device-to-device communication: A spectrum-sharing and densification study. *IEEE Vehicular Technology Magazine*, 13(1): 85-96.
18. Safdar, G., Ur-Rehman, M., Muhammad, M., Imran, M. and Tafazolli, R. (2016). Interference mitigation in D2D communication underlaying LTE-A network. *IEEE Access*, *4*(8): 7967-7987.
19. Eckermann, F., Freudenthal, J.and Wietfeld, C. (2018). *Performance Analysis of Unsupervised LTE Device-to-Device (D2D) Communication.*88th IEEE Vehicular Technology Conference (VTC-Fall) held at Brussels, Belgium, form9th-11th July 2018, 5p.
20. Worgu, Steve, Samuel Godwin Ajumo, and Njumoke N. Odu. "Comparative Evaluation of the Pathloss Prediction Performance Hata-Okumura Pathloss Model for Urban, Suburban and Rural Areas." *International Journal of Systems Science and Applied Mathematics* 2.1 (2017): 42.