Development Of Fuzzy Inference System (FIS) For Detection Of Outliers In Data Streams Of Wireless Sensor Networks

Felix . A. Njoku[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

Fina Otosi Faithpraise[3]
Department of Computer Engineering
University of Calabar Cross River State Nigeria

Abstract— This paper focuses on the investigation of an Artificial Intelligent model for the detection of outliers in the data streams of wireless sensor network. For this study atmospheric parameters sensor in Ikeja metropolis of Lagos state, Nigeria was studied for outlier detection with the Fuzzy Inference System (FIS) model. The data was an hourly data obtained for a duration of one week with 35 outlier data points introduced into the data stream. The centroid for each parameter was obtained with Fuzzy C-Means and the percentage tolerance from the centroid used in this study to obtain the index values was 20%. The results showed that the FIS model gave accuracies of 88.6% outliers' detection for temperature, 94.3% for pressure and 88.6% for relative humidity. The overall purpose of the paper is to detect anomalous data points in wireless sensor networks due to several causes such as erroneous records from sensors due malicious attacks and stringent resource availability to wireless sensor networks. A conventional clustering technique (k-means) clustering algorithm was also explored for the same purpose: the performance obtained by the two models are compared. The results show that fuzzy inference clustering algorithm gave better detection capability than the K-means algorithm. This indicates that the artificial intelligent clustering algorithm is superior in terms of outlier detection of atmospheric parameters than the conventional clustering algorithm.

Keywords— Outlier Detection, Wireless sensor networks, and Fuzzy Inference System, Fuzzy C-Means, clustering technique, centroid, K-means clustering.

I. Introduction

Over the years, increasing adoption and diverse applications of Wireless Sensor Network (WSN) has been posing running challenges to WSN designers and users of WSN services [1,2,3]. A typical WSN consists of a number of modules each with sensors and communications circuitry which are deployed to different locations within the desired environment and intended to monitor and transmit certain environmental or system parameters it acquired from diverse locations [4,5,6]. Some of the parameters commonly monitored using WSN include temperature, pressure, humidity, wind direction and speed, vibration intensity, sound intensity, voltage, chemical concentrations, and pollutant levels [7,8,9,10].

In the early days of WSNs, they are mainly used in military operations however recently; WSNs are used in many common applications such as structural health monitoring, environmental monitoring, agriculture and industrial application [11,12,13]. In any case, the sensors have limitations with respect to memory, communications bandwidth, battery power and computational capacity. Also, WSNs are exposed to faults and malicious attacks which do cause unreliable and inaccurate readings and poor data quality [14]. Furthermore, WSNs data streams do have outliers which are data with anomaly or deviation from the general data distribution pattern over a given period [15,16,17,18]. Due to the negative impact of outliers on the data quality, studies on the mechanisms for identification of outliers in WSN data streams has been on the increase since the advent of the WSN technology [19,20,21,22]. More so, sometimes, outliers are considered more interesting than the normal data, and this is the reason why outliers need to be identified since they may contain important information about the WSN [23,24,25]. Consequently, in this study, fuzzy inference system (FIS) [26,27] was applied to a WSN data stream to detect the outliers in the dataset. The result obtained from the FIS model was compared with that obtained using K-means clustering algorithm [28,29]. The essence of the study is to develop a more effective outlier detection mechanism for WSN used to acquire atmospheric parameter data stream.

II. METHOD

The major focus in this paper is the detection of outliers in data streams of a wirelesses sensor network using fuzzy inference model and K-means algorithm clustering algorithm. As the initial step, site survey was conducted within Ikeja metropolis of Lagos State in Nigeria to obtain the atmospheric temperature, pressure and relative humidity of the region. Then, a Sugeno rule-based Fuzzy Inference Model [30,31] was developed to detect outlier data from the atmospheric parameters datasets. Finally, outliers were detected using the Fuzzy Inference System (FIS) model and K-means algorithm, all simulated in MATLAB software. The flow diagram of the research process is shown in Figure 1. According to the methodology, there are few main major tasks performed in the research process. They are acquisition of wireless sensor data, determination of number of outliers inherent in the acquired data, detection of outliers with Fuzzy Inference System and discussion of result.

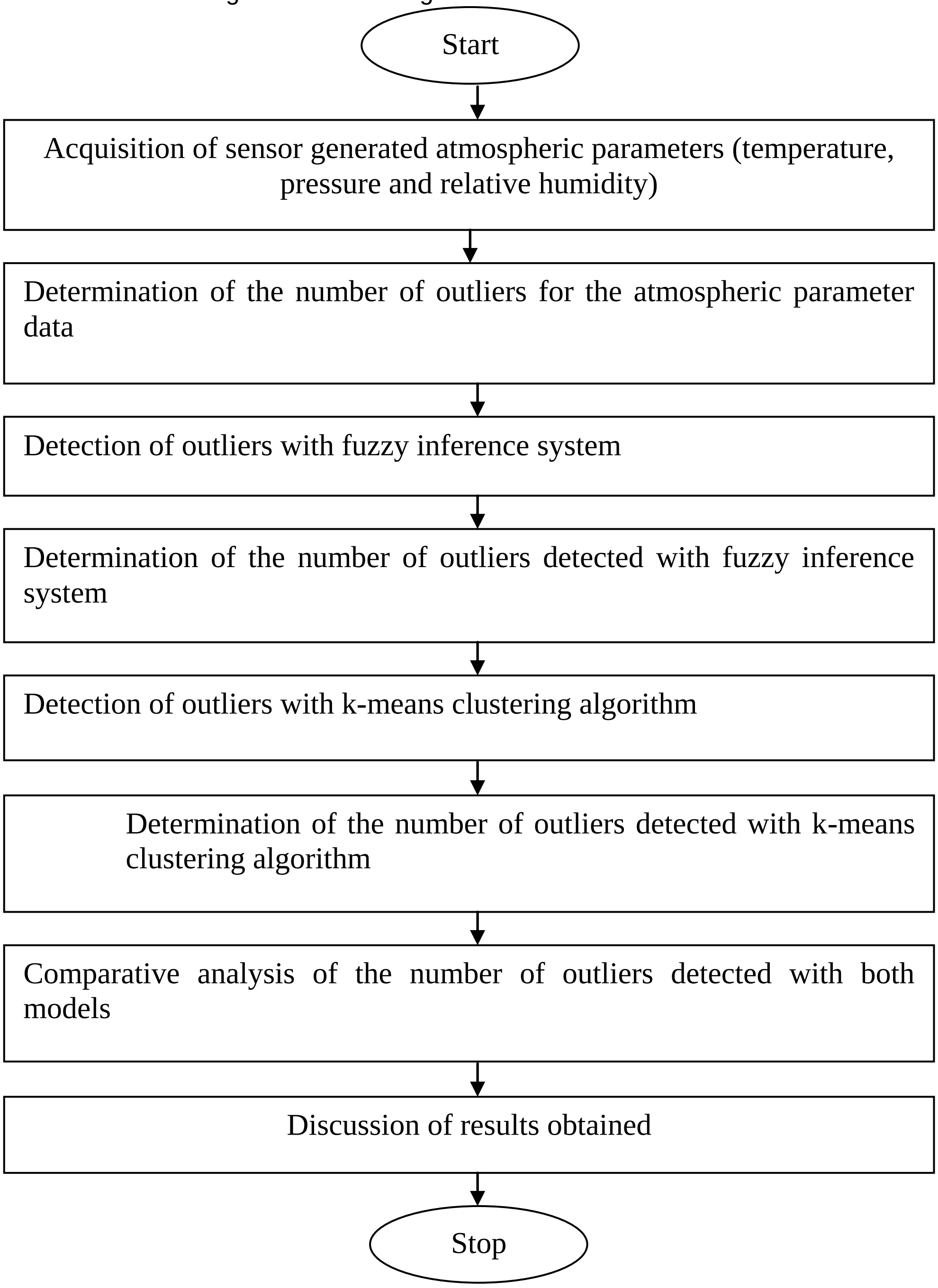

Figure 1: Flow diagram of the research process.

A. Acquisition of the wireless sensor network data

The wireless sensor network data considered for this paper was obtained from the Nigeria meteorological department located in Ikeja, Lagos state, Nigeria. Particularly, hourly atmospheric data was obtained for a period of one week. The atmospheric parameters considered were temperature, pressure and relative humidity. The sample of data obtained is displayed in Table 1

Table 1: The Atmospheric Parameter Data

Time	Temperature (^oC)	Pressure (mb)	Relative humidity (%)
00:00	25	1011	84
01:00	26	1011	85

02:00	26	1011	86
03:00	26	1010	86
04:00	25	1010	86
05:00	25	1010	87
06:00	25	1011	87
07:00	25	1011	87
08:00	26	1012	81
09:00	28	1012	74
10:00	29	1012	67
11:00	29	1011	67
12:00	29	1010	67
13:00	29	1010	67
14:00	29	1009	68
15:00	29	1008	68
16:00	28	1008	69
17:00	28	1008	72
18:00	27	1009	75
19:00	27	1009	78
20:00	27	1010	79
21:00	27	1010	80
22:00	26	1011	81
23:00	26	1011	82

B. Procedure for Detection of Outliers with Fuzzy Inference System (FIS) model and K-means algorithm

The FIS model is based on some notable features of outliers in any given dataset, namely;

i. If a dataset has a centroid, m the outlier value is usually far from the centroid of the dataset than the other data items in the set. Essentially, the outlier value is either too far above or too far below the centroid. The extent to which the difference between the data point value is from the centroid can be expressed numerically and then a threshold is set at which point the data point is considered an outlier.
ii. For any given dataset, the outlier usually has few points within the neighbourhood.
iii. For any given dataset, the outlier usually has low degree of membership to the cluster to which it could be assigned.

Therefore, the FIS model developed in this work exploits both the consideration of neighbouring points and the capability of clustering algorithms for pointing out outliers. In particular, two preliminary operations were performed: the calculation of the centroid that is simply the mean vector of the data distribution, and the clustering operation, that is performed by means of the FIS model.

Fuzzy logic model was used to detect the outliers in the data of Table 1. The flow diagram for the application of Fuzzy logic toolbox in MATLAB to detect the outlier in the given data stream is outlined in Figure 2. The model used in the Fuzzy logic rules is the Sugeno model. Also, the outliers are detected using K-means algorithm. The performance of the two methods; Fuzzy logic and K-means algorithm in detecting the outliers is compared in terms of the number of outliers detected for a given scenario.

The data acquired from the sensors monitoring the atmospheric parameters (temperature, pressure, and relative humidity) is keyed into the system and saved as an m-file MATLAB file. In this paper, the atmospheric parameters namely; temperature, pressure and relative humidity were considered with respect to time in four hours time interval. Specifically, the data now in the MATLAB software is grouped into six (6) sets of 4 hours interval groups to account for diurnal and nocturnal variation in temperature. This will help the model to account for outliers' detection within the daytime and the night time, as temperature within these periods vary differently.

For the grouped data, the size of each of the dataset has to be determined so as to know the exact number of data points that appear in each of the atmospheric parameter values. In this case, 28 data points were generated for temperature, pressure and relative humidity. The variable x yields a matrix of 1x28 dataset.

For this paper, the membership function used is the triangular membership function. The triangle is inter-twined to provide the least, central and the upper range. With the specified values of input data on the x-axis, the three triangular membership functions equalized to fit on the x-axis. With the membership functions, the fuzzy logic automatically sets the least, central, and upper values for each triangle to map an output variable for it. This process is governed by the fuzzy inference rule.

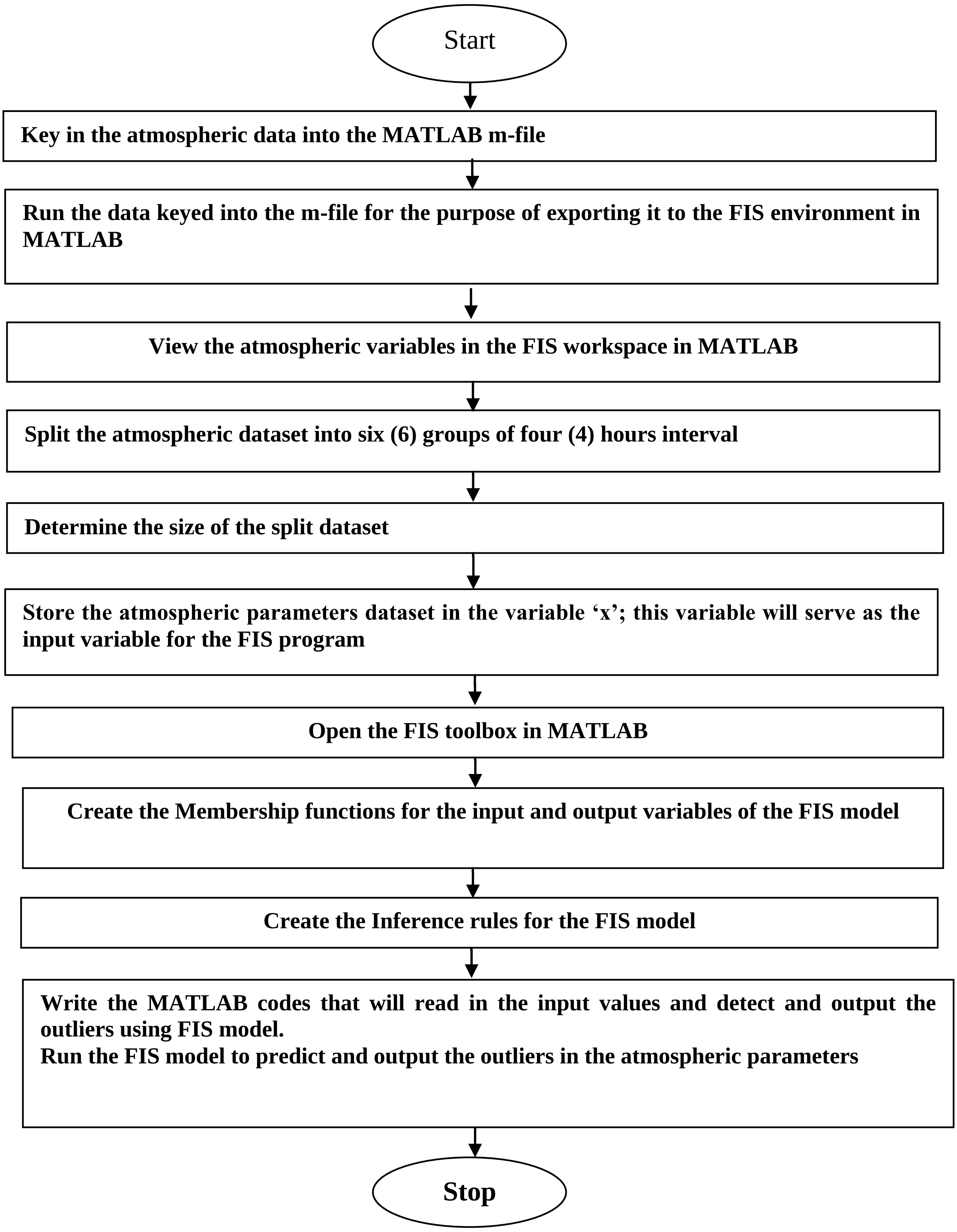

Figure 2: The flow diagram for the detection of outliers with fuzzy inference system (FIS).

C. The Inference rules for the FIS model

The fuzzy inference system includes 3 rules and uses the Sugeno inference model. The antecedent of the all rules has three sections; these include low (L), moderate (M) and high (H). In this case, in1mf1, in2mf2, and in3mf3 are for the input and out1mf1, out2mf2 and out3mf3 are for the output. The AND operator is used to combine the antecedent parts of the rules. The degree of truth (R) of the rule is determined by evaluating the non-zero minimum value using the AND operator.

The detailed rule base comprising of 3 rules for each atmospheric parameter is obtained from the sensors of Ikeja metropolis for temperature, pressure and relative humidity and shown in Table 2. The inference uses the maximum operator acting on the rule base. The rule base condition for the temperature in Table 2 states that if the first membership function (with 'L' marking) is low, then the output is the first membership function of the output variable (out1mf1 of the temperature). If the membership function of the input is medium (M), the marked output was out1mf2 and if the membership function of the input is High (H), the membership function for the output mapped out was out1mf3.

Table 2 The Rule base for inference on temperature, pressure and Relative Humidity

S/N	Time	Temperature	Pressure	Relative Humidity
1	L	Out1mf1	Out1mf1	Out1mf1
2	M	Out1mf2	Out1mf2	Out1mf2
3	H	Out1mf3	Out1mf3	Out1mf3

In Table 2, the rule base condition for pressure states that if the first membership function is low, then the output is the first membership function of the output variable (out1mf1). If the membership function of the input is medium (M), the marked output was out1mf2 and if the membership function of the input is High (H), the membership function for the output mapped out was out1mf3.

Also, in Table 2, the rule base condition for relative humidity states that if the first membership function (with ‘L’ marking) is low, then the output is the first membership function of the output variable (out1mf1 of the temperature). If the membership function of the input is medium (M), the marked output was out1mf2 and if the membership function of the input is High (H), the membership function for the output mapped out was out1mf3.

D. Determination of the center index using Fuzzy C-means

Fuzzy C-means is used to simulate the system in an iterative process to determine the centre index point as its objective function. Table 3 shows the centroid index values of the temperature measured with respect to time. The tolerable distance values from the index are 20% of each determined index. The values above the tolerable index range were considered as outliers.

Table 3 : Index point of the atmospheric data.

Time	Index values for temperature	Index values for pressure	Index values for relative humidity
00:00-04:00	26	1010	83
05:00-08:00	25	1010	86
09:00-12:00	27	1012	78
13:00-16:00	28	1009	71
17:00-20:00	26	1008	77
21:00-23:00	25	1010	82

III. RESULTS

A. Outlier Detection using Fuzzy Inference System

The simulation of the rule base inference of Fuzzy Logic was done using MATLAB 2015a software which yielded graphical results illustrating the number of datasets recorded and outliers within it. The FIS temperature outliers detected with FIS for 7 days between 00:00 to 04:00 for each day are shown in Figure 3.

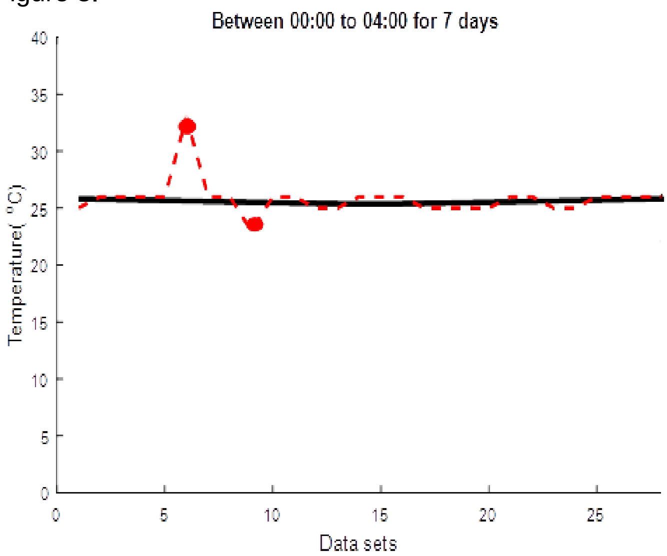

Figure 3: Temperature outliers for the time interval 00:00 to 04:00.

In the case show in Figure 3, only two temperature outliers were detected. This was so because the temperature centre index point during this period occurred at 26^{o}C. The temperature outlier detection between 05:00 to 08:00 is shown in Figure 4. Seven outliers were detected as shown in Figure 4. The temperature centre index point during this period was 25^{0}C.

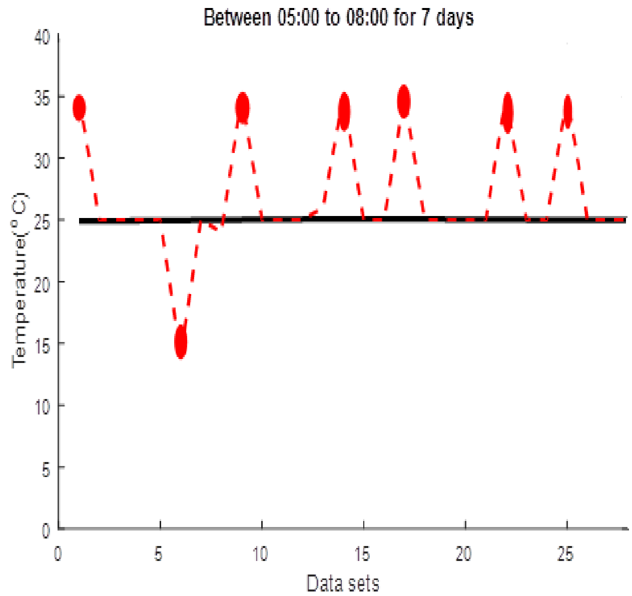

Figure 4: Temperature outlier for the time interval 05:00 to 08:00.

The temperature outlier detection between 09:00 to 12:00 for 7days is shown in Figure 5. The number temperature outliers detected at this time was seven. The temperature centre index point during this period occurred at 26^{0}C.

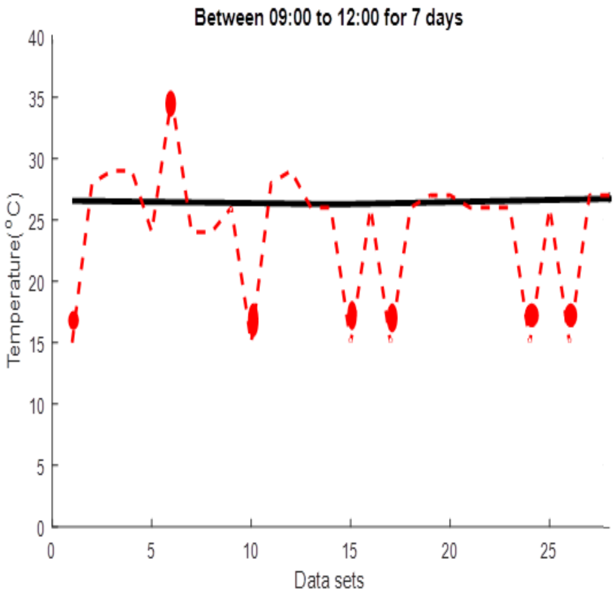

Figure 5: Temperature outlier detection between 09:00 to 12:00 for 7 days.

The temperature outlier detection between 13:00 to 16:00 for 7days is shown in Figure 6. From Figure 6, the number of outliers detected was six. As such the sensor at this point is very critical because the temperature centre index point during this period occurred at 28^{o}C.

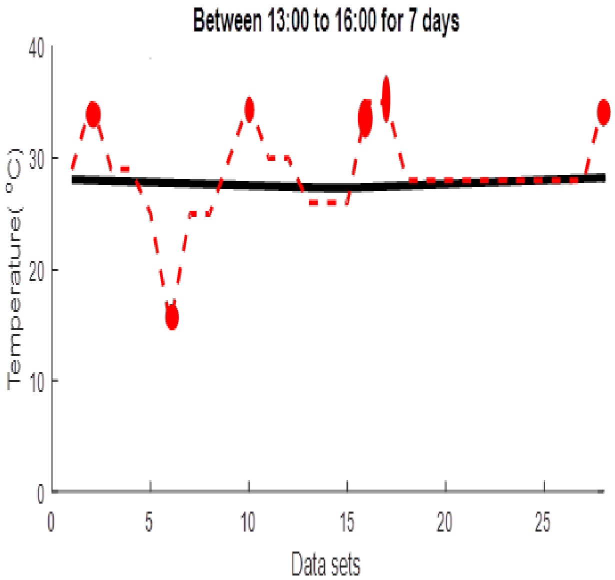

Figure 6: Temperature outlier detection between 13:00 to 16:00 for 7 days.

The temperature outlier detection between 17:00 to 20:00 for 7 days is shown in Figure 7. The number of outliers detected in Figure 7 was five. The temperature centre index point during this period occurred at 26^{0}C.

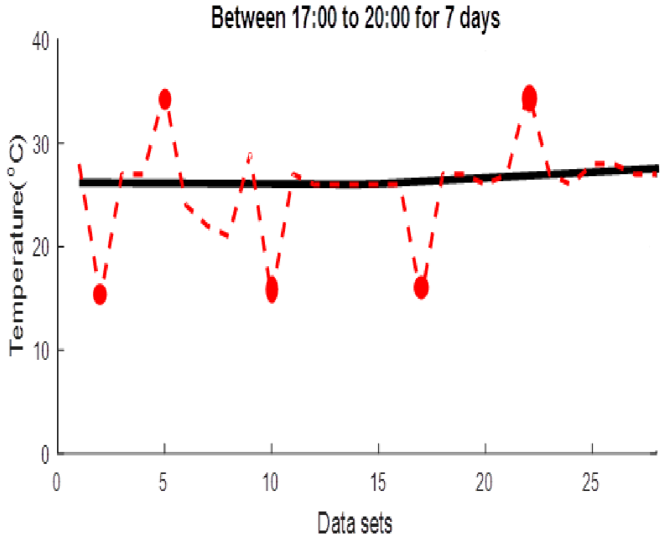

Figure 7: Temperature outlier detection between 17:00 to 20:00 for 7 days.

The temperature outlier detection between 21:00 to 24:00 for 7days is shown in Figure 8. In Figure 8, only four outliers were detected. The temperature centre index point during this period occurred at 26^{0}C. In all, for the six time intervals, a total of thirty one (31) outliers were detected for temperature.

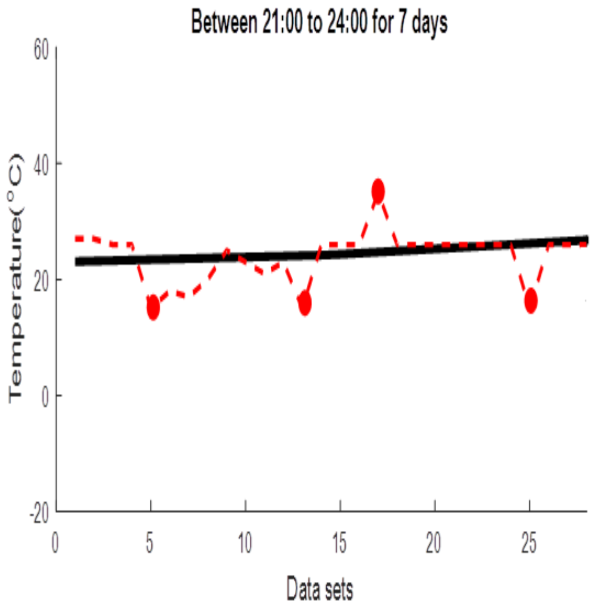

Figure 8: Temperature outlier detection between 21:00 to 24:00 for 7 days.

B. Outlier detection for Pressure with FIS

Fuzzy inference system was employed in detection of outliers in the atmospheric pressure data stream. Figure 9 shows the pressure outlier detected for the data stream in the time interval between 00:00 to 04:00 for 7 days. From Figure 9, the number of outliers detected was six. The pressure centre index point during this period was 1011mb.

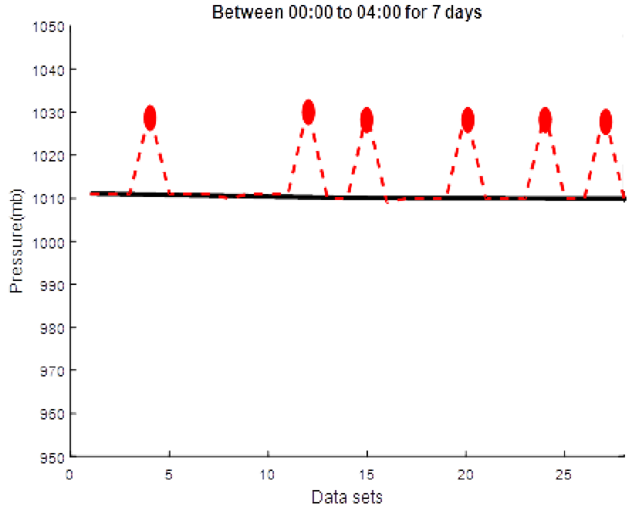

Figure 9: Pressure outlier detection between 00:00 to 04:00 for 7 days.

The pressure outlier detected between 05:00 to 08:00 for 7 days are as shown in Figure 10 (six outliers were detected and the pressure centre index point during this period was 1011mb).

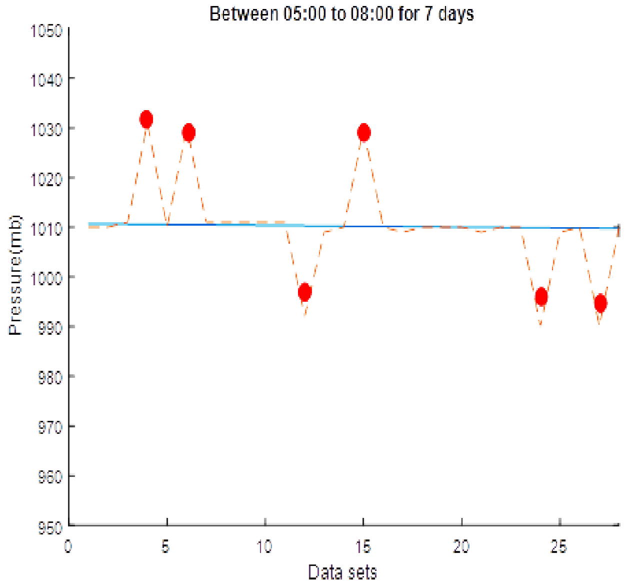

Figure 10: Pressure outlier detection between 05:00 to 08:00 for 7 days.

The pressure outlier detected between **09:00 to 12:00** for 7 days are as shown in Figure 11 (six outliers were detected and the pressure centre index point during this period was 1012mb).

The pressure outlier detected between 13:00 to 16:00 for 7 days are as shown in Figure 12 (six outliers were detected and the pressure centre index point during this period was 1009mb).

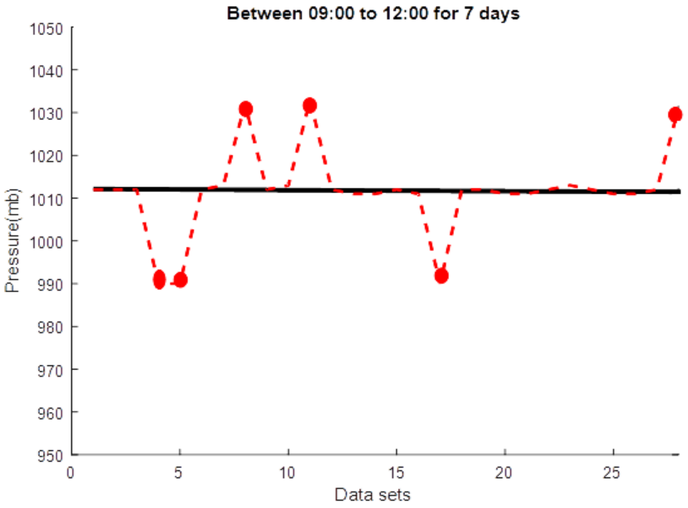

Figure 11: Pressure outlier detection between 09:00 to 12:00 for 7 days.

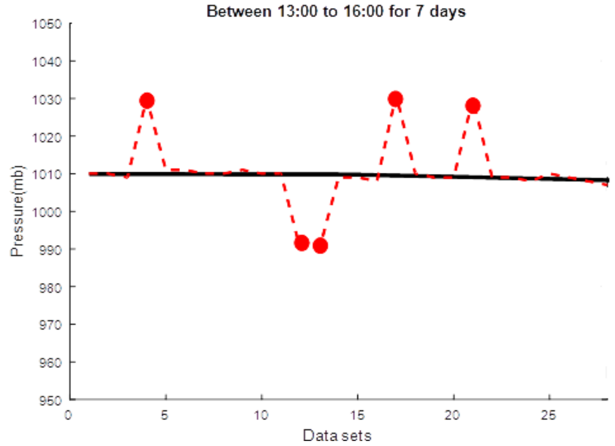

Figure 12: Pressure outlier detection between 13:00 to 16:00 for 7 days.

The pressure outlier detected between 17:00 to 20:00 for 7 days are as shown in Figure 13 (six outliers were detected and the pressure centre index point during this period was 1008mb). The pressure outlier detected between 21:00 to 24:00 for 7 days are as shown in Figure 14 (four outliers were detected and the pressure centre index point during this period was 1010 mb).

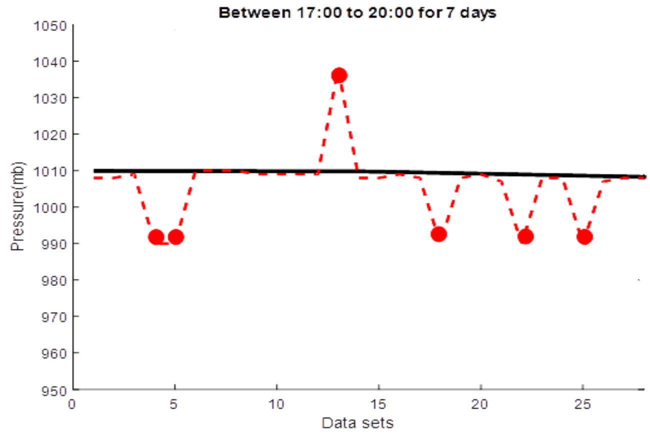

Figure 13: Pressure outlier detection between 21:00 to 24:00 for 7 days.

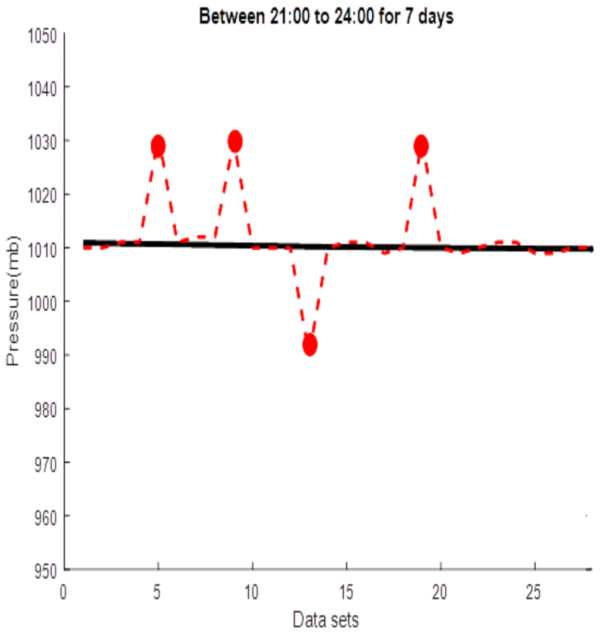

Figure 14: Pressure outlier detection between 21:00 to 24:00 for 7 days.

Source: Formulated by the paper (2019).

C. Outlier detection of Relative Humidity with FIS

Fuzzy logic model was used to detect outliers from the WSN that generated the data stream for relative humidity for 7 days. The relative humidity outlier detected between 00:00 to 04:00 for 7 days are as shown in Figure 15 (six outliers were detected and the relative humidity centre index point during this period was 83 %).

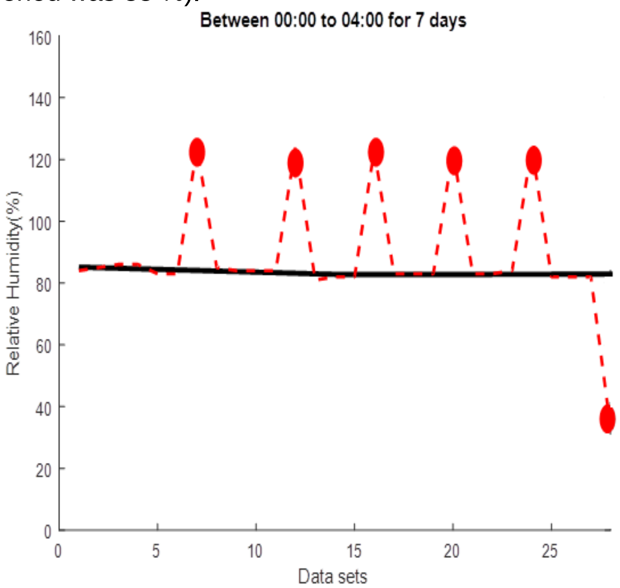

Figure 15: Relative humidity outlier detection between 00:00 to 04:00 for 7 days.

The relative humidity outlier detected between 05:00 to 08:00 for 7 days are as shown in Figure 16 (four outliers were detected and the relative humidity centre index point during this period was 86 %).

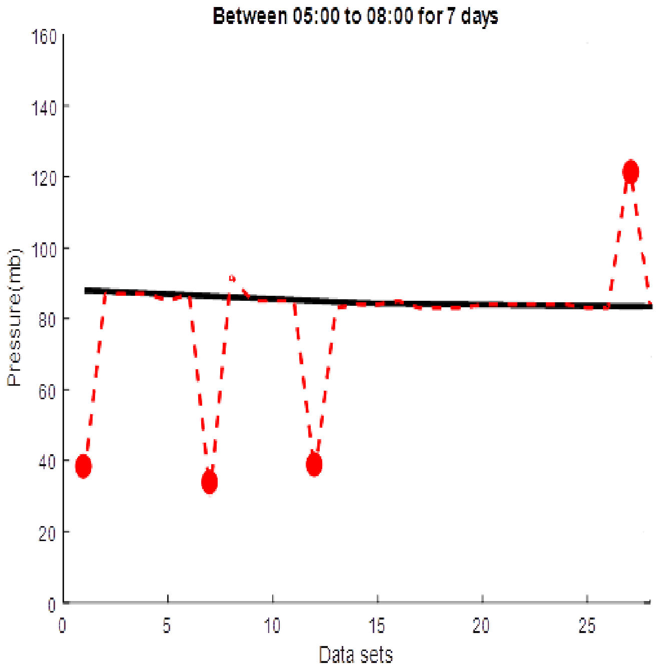

Figure 17: Pressure outlier detection between 05:00 to 08:00 for 7 days.

Source: Formulated by the paper (2019).

The relative humidity outlier detected between 09:00 to 14:00 for 7 days are as shown in Figure 17 (six outliers were detected and the relative humidity centre index point during this period was 78 %).

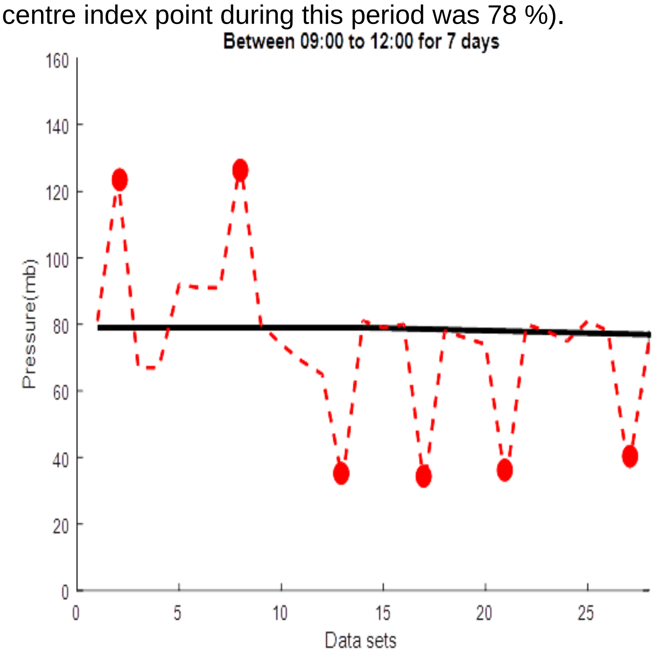

Figure 17: Pressure outlier detection between 09:00 to 12:00 for 7 days.

The relative humidity outlier detected between 13:00 to 16:00 for 7 days are as shown in Figure 18 (seven outliers were detected and the relative humidity centre index point during this period was 71 %).

The relative humidity outlier detected between 17:00 to 20:00 for 7 days are as shown in Figure 19 (six outliers were detected and the relative humidity centre index point during this period was 77 %).

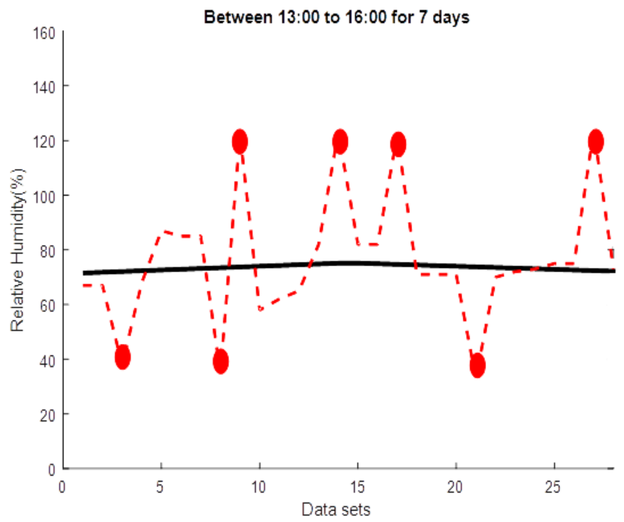

Figure 18: Pressure outlier detection between 13:00 to 16:00 for 7 days.

The relative humidity outlier detected between 21:00 to 24:00 for 7 days are as shown in Figure 20 (two outliers were detected and the relative humidity centre index point during this period was 82 %).

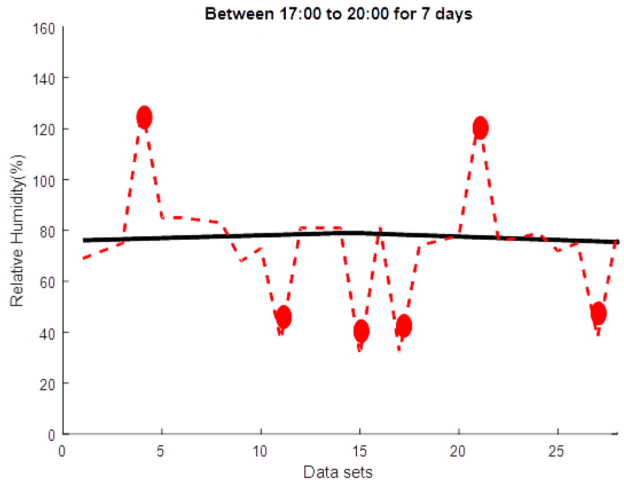

Figure 19 : Pressure outlier detection between 17:00 to 20:00 for 7 days.

Source: Formulated by the paper (2019).

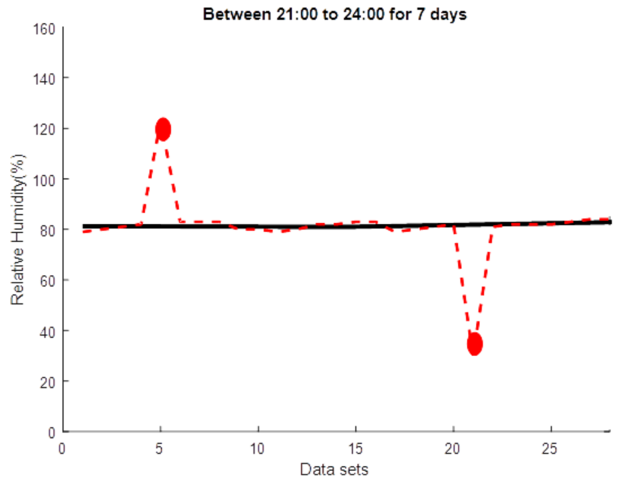

Figure 20: Pressure outlier detection between 21:00 to 24:00 for 7 days.

Similar study was conducted using K-means clustering algorithm. In all, the total number of outliers detected for the each of the three atmospheric parameters is shown in Table 4. The summary of the result as presented in Table 4 shows that Fuzzy Inference System (FIS) performs better the k-means algorithm. Specifically, FIS detected higher number of outliers than the k-means clustering. This indicates that the artificial intelligent clustering algorithm (which in this study is the FIS model) is superior in terms of outlier detection of atmospheric parameters than the conventional clustering algorithm, (which in this study is the K-means algorithm).

Table 4 : Summary of the total outliers detected for the three atmospheric parameters using the Fuzzy Inference System and the K-means

Atmospheric Parameters	K-means	Fuzzy Inference System
Temperature	10	31
Pressure	9	33
Relative Humidity	10	31

IV. CONCLUSION

In this study, atmospheric parameters data stream (temperature, pressure and relative humidity) were obtained from wireless sensors. The data was obtained hourly for a period of seven days. A fuzzy inference clustering algorithm for detecting outliers in the data stream was developed and simulated using MATLAB. The fuzzy inference system (FIS) and K-means clustering algorithm were applied to the sensor generated data to obtain outliers. A comparative analysis of the result obtained from the two algorithms was carried out to measure the effectiveness of the models in detecting outliers. The atmospheric parameters which form the basis of the sensor monitored parameters were use in the training of the clustering algorithms. The results show that fuzzy inference clustering algorithm gave better detection capability than the K-means algorithm. This indicates that the artificial intelligent clustering algorithm is superior in terms of outlier detection of atmospheric parameters than the conventional clustering algorithm.

REFERENCES

1. Azeem, Nahla S. Abdel, et al. "Shared Sensor Networks Fundamentals, Challenges, Opportunities, Virtualization Techniques, Comparative Analysis, Novel Architecture and Taxonomy." *Journal of Sensor and Actuator Networks* 8.2 (2019): 29.
2. Engmann, Felicia, et al. "Prolonging the lifetime of wireless sensor networks: a review of current techniques." *Wireless Communications and Mobile Computing* 2018 (2018).

3. Khalid, Zubair, Norsheila Fisal, and Mohd Rozaini. "A survey of middleware for sensor and network virtualization." *Sensors* 14.12 (2014): 24046-24097.
4. Ketshabetswe, Lucia Keleadile, et al. "Communication protocols for wireless sensor networks: A survey and comparison." *Heliyon* 5.5 (2019): e01591.
5. Yick, Jennifer, Biswanath Mukherjee, and Dipak Ghosal. "Wireless sensor network survey." *Computer networks* 52.12 (2008): 2292-2330.
6. Huang, Xiaoci, et al. "A wireless sensor network-based approach with decision support for monitoring lake water quality." *Sensors* 15.11 (2015): 29273-29296.
7. Jawad, Haider Mahmood, et al. "Energy-efficient wireless sensor networks for precision agriculture: A review." *Sensors* 17.8 (2017): 1781.
8. Zhu, Yuenong, and Kun Hua. "The development of mobile wireless sensor networks: a survey." *Technological Breakthroughs in Modern Wireless Sensor Applications*. IGI Global, 2015. 257-286.
9. Singh, Rupinder, Jatinder Singh, and Ravinder Singh. "A Novel Sybil Attack Detection Technique for Wireless Sensor Networks." *Advances in Computational Sciences and Technology* 10.2 (2017): 185-202.
10. Varma, V., Shinde, K., Shinare, R., Varpe, S., & Ingle, M. D. (2016). A Survey Paper on Reduction in Loss of Signals in Wireless Sensor Network Using Data Aggregation Techniques. *Science and Technology*, *2*(12), 1-6.
11. Shi, Xiaojie, et al. "State-of-the-art internet of things in protected agriculture." *Sensors* 19.8 (2019): 1833.
12. Erdelj, Milan, Nathalie Mitton, and Enrico Natalizio. "Applications of industrial wireless sensor networks." *Industrial wireless sensor networks: applications, protocols, and standards* (2013): 1-22.
13. Ali, Ahmad, et al. "A comprehensive survey on real-time applications of WSN." *Future internet* 9.4 (2017): 77.
14. Wang, J., Ren, X., Shen, Y. and Liu, S. (2010). A remote wireless sensor network for water quality monitoring.*IEEE Transactions on Knowledge and Data Engineering,* 11(16): 7-12.
15. Reunanen, Niko, et al. "Unsupervised online detection and prediction of outliers in streams of sensor data." *International Journal of Data Science and Analytics* (2019): 1-30.
16. Mirzaie, Mostafa, Behshid Behkamal, and Samad Paydar. "Big Data Quality: A systematic literature review and future research directions." *arXiv preprint arXiv:1904.05353* (2019).
17. Giannoni, Federico, Marco Mancini, and Federico Marinelli. "Anomaly Detection Models for IoT Time Series Data." *arXiv preprint arXiv:1812.00890* (2018).
18. Zhang, Yang, Nirvana Meratnia, and Paul JM Havinga. "Outlier detection techniques for wireless sensor networks: A survey." *IEEE Communications Surveys and Tutorials* 12.2 (2010): 159-170.
19. Djedouboum, Asside Christian, et al. "Big data collection in large-scale wireless sensor networks." *Sensors* 18.12 (2018): 4474.
20. Baniata, Mohammad, et al. "Energy-efficient unequal chain length clustering for WSN." *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. ACM, 2018.
21. Bosman, Hedde HWJ, et al. "Spatial anomaly detection in sensor networks using neighborhood information." *Information Fusion* 33 (2017): 41-56.
22. Garcia-Font, Victor, Carles Garrigues, and Helena Rifà-Pous. "A comparative study of anomaly detection techniques for smart city wireless sensor networks." *Sensors* 16.6 (2016): 868.
23. Yu, Yufeng, et al. "Time series outlier detection based on sliding window prediction." *Mathematical problems in Engineering* 2014 (2014).
24. Oluwasanya, Pelumi. "Anomaly Detection in Wireless Sensor Networks." *arXiv preprint arXiv:1708.08053* (2017).
25. Fawzy, Asmaa, Hoda MO Mokhtar, and Osman Hegazy. "Outliers detection and classification in wireless sensor networks." *Egyptian Informatics Journal* 14.2 (2013): 157-164.
26. Upadhyay, Ashish, et al. "Analysis and Comparison of Different Fuzzy Inference Systems used in Decision Making for Secondary Users in Cognitive Radio Network." *arXiv preprint arXiv:1904.11688* (2019).
27. Nugroho, Fresy, and Muhammad Faisal. "STUDI KOMPARASI FUNGSI KEANGGOTAAN BERBEDA PADA FUZZY MAMDANI UNTUK KASUS PENYIRAMAN AIR OTOMATIS." *Conference on Innovation and Application of Science and Technology (CIASTECH)*. Vol. 2. No. 1. 2019.
28. Hrosik, Romana CAPOR, et al. "Brain image segmentation based on firefly algorithm combined with k-means clustering." *Stud. Inform. Control* 28 (2019): 167-176.
29. Jothi, R., Sraban Kumar Mohanty, and Aparajita Ojha. "DK-means: a deterministic k-means clustering algorithm for gene expression analysis." *Pattern Analysis and Applications* 22.2 (2019): 649-667.
30. Santoso, Fendy, Matthew A. Garratt, and Sreenatha G. Anavatti. "T2-ETS-IE: Type-2 Evolutionary Takagi-Sugeno Fuzzy Inference Systems with Information Entropy-Based Pruning Technique." *IEEE Transactions on Fuzzy Systems* (2019).
31. Chaudhary, Alka. "Mamdani and Sugeno Fuzzy Inference Systems' Comparison for Detection of Packet Dropping Attack in Mobile Ad Hoc Networks." *Emerging Technologies in Data Mining and Information Security*. Springer, Singapore, 2019. 805-811.