Asia-Pacific Journal of Science and Technology: Volume: 28. Issue: 06. Article ID.: APST-28-06-21. Research Article

Asia-Pacific Journal of Science and Technology

https://www.tci-thaijo.org/index.php/APST/index

Published by the Research and Graduate Studies Division,
Khon Kaen University, Thailand

Clustering optimisation using fuzzy c-means clustering and artificial bee colony algorithm for wireless sensor network

Mohammed S.H. Thekiya[1,*] and Mangesh D. Nikose[1]

[1]Department of Electrical and Electronics Engineering, Sandip University, Maharashtra, India
[*]Corresponding author: shahidthekiya@gmail.com

Received 20 May 2022
Revised 24 September 2022
Accepted 25 October 2022

Abstract

Wireless sensor networks (WSNs) play an important role in numerous applications such as industrial automation, commercial, robotics, environmental monitoring, landslide detection, earthquake detection, transport and logistics, and habitat monitoring. WSN clustering provides an efficient way to improve the network lifetime, throughput, scalability, and packet delivery ratio. However, the performance of WSN is limited because of low-power battery-operated sensor nodes and improper positioning of the cluster heads (CHs) during cluster formation. This study presents a fuzzy c-means algorithm (FCM) for WSN clustering and an artificial bee colony (ABC) algorithm for optimal selection of CHs. The proposed ABC considers various factors for clustering, such as the energy Gini coefficient, CH energy balancing, inter- and intra-cluster distance, and CH load balancing factors. The proposed algorithm provides optimised cluster selection that provides better network lifetime and throughput than traditional FCM.

Keywords: Artificial bee colony algorithm, Data aggregation, Fuzzy c-mean clustering, Wireless sensor net work

1. Introduction

Wireless sensor networks (WSN) are widely used in several industrial, commercial, and social applications owing to the vast Internet of Things. WSN is a group of homogeneous or heterogeneous sensor nodes distributed over a plane. It consists of a sensor node, processing unit, memory, transmitter, receiver, and battery. The lifetime of the network is usually short because it uses a low-power battery [1-3]. WSN distributed over a larger area requires the clustering of sensor nodes in the group to provide efficient routing and data aggregation. The selection of cluster head (CH) is crucial because it accepts data from the sensor nodes and transmits it to the base station. Often, CH is selected based on its position in the cluster, and chances are often given to a centrally located node irrespective of its energy, load balancing capacity, connectivity, and distance from the base station [4-7].

Previous studies have reported various clustering techniques for WSN data aggregation and routing. Raam et al [8] presented a study on a parallel ant colony optimisation (ACO) algorithm and k-means clustering algorithm for grouping sensor nodes to detect the optimal path in WSN routing; it provided a longer lifespan of the network and an efficient routing algorithm. Liu et al. [9] proposed an unsupervised clustering problem based on the ACO algorithm, wherein they used ACO with a stochastic best solution, kept-ESacc. Gupta et al. [10] proposed an adapted version of the ACO base low-energy adaptive clustering. hierarchy (LEACH) clustering algorithm for effective CH selection. In their work, data transmission took place in three phases: from the sensor node to the CH, from the CH to the cluster leader, and from the cluster leader to the base station (BS). This resulted in an improvement in the average energy consumption. However, they did not consider data redundancy, which affects data aggregation. Aadil et al. [11] proposed a clustering-based ACO for vehicular ad hoc networks (VANETs; CACONET) for clustering; they performed extensive experimentation by varying the network size, sensor nodes, coverage area of the network, sensor node's transmission range, speed, and direction of the VANET nodes. The performance of the CACONET algorithm was compared with that of the multi-objective particle swarm algorithm and clustering algorithm-based ACO. The proposed algorithm outperformed the other algorithms. Yang et al. [12]

presented an ACO along with a dynamic clustering-based multipath routing protocol (MRP) for burst event monitoring in a reactive WSN to maximise the sensor network lifetime and minimise energy consumption. In the MRP algorithm, CHs were selected based on the remaining energy, and multiple paths between the sensor node and CH were selected using ACO. This resulted in efficient data aggregation, better load balancing, maximised sensor network lifetime, and lowered energy consumption in the WSN. The authors found the parameter setting for the algorithm to be challenging and slower speech owing to parameter tuning. Many researchers have focused on clustering routing protocols for static and dynamic WSNs using various optimisation techniques. However, the performance of the algorithm is limited owing to cluster imbalance, heavy traffic toward CH, shorter network lifetime, huge data packet losses, and poor network performance under dynamic conditions [22-25]. Hence, there is a need to provide energy-efficient clustering to achieve high network throughput and longer lifetimes.

This study presents WSN clustering for the clustering of sensor nodes to improve the network performance and lifetime. The major contributions of this study are summarised as follows:

- Implementation of optimal CH selection using fuzzy C-means (FCM) and biologically inspired ABC algorithm based on several aspects such as energy Gini coefficient, CH energy balancing, inter- and intra-cluster distance, and CH load balancing factors.
- The performance of the proposed algorithm was estimated based on the network throughput, lifetime, and packe delivery ratio.
- The remainder of this study is organised as follows: Section 2 provides detailed information about the proposed clustering and CH selection strategy. Section 3 focuses on simulation results and a discussion of the results. Section 4 provides a concise conclusion and future scope.

2. Materials and methods

The anticipated methodology comprises four main phases: initialisation, clustering, cluster selection, and data aggregation (Figure 1). The initialisation phase consists of network parameter initialisation, such as simulation area, several sensor nodes, base station position, sensor node energy, and locations, as well as initialisation of the radio model parameters such as the number of bits in the transmission frame, energy required for transmission and reception of a single bit, amplification factor, traffic pattern, and media access control protocol.

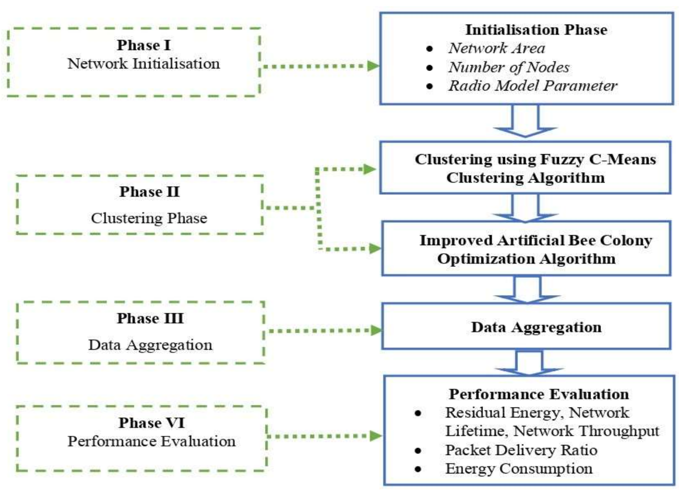

Figure 1 Flow diagram of the proposed system.

In the clustering phase, an FCM clustering algorithm is used to form a cluster of sensor nodes having the same initial energy based on the position of the sensor node in the simulation area. FCM is a data clustering method in which raw data are categorised into clusters that provide the degree to every member node based on its distance from the centre of the cluster. The nodes closer to the centre of the cluster have a high degree, whereas those far away from the centre of the cluster have a lower degree. The degree of sensor nodes is used to compute the membership of sensor nodes to their respective clusters [13-14].

Initialise the sensor node set as X = {x1, x2, x3..., xn} and the set of centres of clusters as V = {v1, v2, v3..., vc}.

Step 1: Arbitrarily choose ‘c’ cluster centres.

Step 2: Compute the fuzzy membership uij using Equation 1.

$$\mu_{ij} = 1 \Big/ \sum_{k=1}^{c} (d_{ij}/d_{ik})^{(2/m-1)} \tag{1}$$

Step 3: The fuzzy clusters 'vj' are computed usingμ_{ij} as given in Equation 2.

$$V_j = \left(\sum_{i=1}^{n} (\mu_{ij})^m x_i\right) \Big/ \left(\sum_{i=1}^{n} (\mu_{ij})^m\right), \forall j = 1,2, \ldots.. c \tag{2}$$

Step 4: Repeat Steps 2 and 3 until the minimum 'J' value is achieved, or $||U(k+1) - U(k)|| < \beta$.

where k stands for the iteration step, U = (μij)n × c represents the fuzzy membership matrix, β signifies the termination condition between [0,1], and j defines the objective function.

Further, the CHs are selected using an improved artificial bee colony (ABC) algorithm. The improved ABC considers the Gini coefficient, CH energy balancing, inter- and intra-cluster distance, and CH load balancing factors to improve the clustering and CH position. ABC is inspired by the biological phenomenon and was invented by Dervis Karaboga in 2005. It involves three sets of bees: employed, onlooker, and scout bees [15-16].

Employed bees: These bees search for food sources.

Onlooker bees: These reside in the nest and make food source selection decisions using the waggle dance.

Scout bees: These are reserved bees that search for food in a random direction when employed bees are abandoned.

The number of employed bees was set to be the same as the number of food sources around the hive. Employed bees whose food source is finished become scout bees. The ABC algorithm is illustrated in Figure 2.

ABC initially creates an arbitrarily dispersed population with possible SN solutions that are analogous to food sources.

Let x be the initial population, where i = 1, 2, 3…, SN

where

X_i: i^{th} food source in the population

SN signifies the swarm size (food sources)

Onlooker bees search the food based on the probability function given in Equation 3.

$$\rho_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \tag{3}$$

Here, fit_i is the fitness of i^{th} the solution, which is proportional to the nectar quantity of the food source at position i.

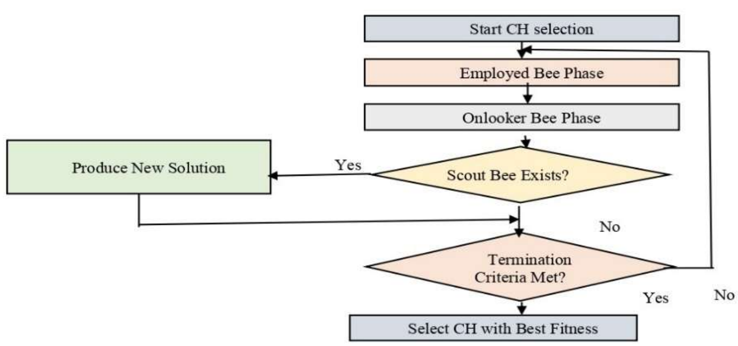

Figure 2 Flochart of the proposed ABC-based CH selection algorithm.

The fitness function for CH selection is designed as follows, based on energy Gini coefficient, CH energy balancing, inter- and intra-cluster distance, and CH load balancing factors, as given in Equation 4. Here, d_{ik} is the distance between CH and BS, C_i is the connectivity of a node (number of neighbouring nodes connected to a particular node), and RE represents residual energy. The weight factors $w_1, w_2, w_3,$ and w_4 are selected as 0.2, 0.3, 0.3, and 0.2, respectively, such that $w_1 + w_2 + w_3 + w_4 = 1$.

$$\mathrm{fit_i} = w_1 * f_1 + w_2 * f_2 + w_3 * f_3 + w_4 * f_4 \quad (4)$$

The fitness function f_1 is related to the Gini coefficient, which indicates the energy distribution equilibrium in the cluster. A generalised Gini coefficient provides the degree of uneven income for the population, as represented by equation 5. This equation is adopted to characterise cluster balancing, as given in Equation 6.

$$G = \frac{1}{2n^2\mu}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i - x_j\right| \quad (5)$$

where n is the total population in the region, x_i is the earnings of the i^{th} person, x_j represents the earnings of the j^{th} person, and μ is the average earnings of the population in the region.

$$E_{s(G)} = \frac{1}{2num^2(s)E_{ave}(s)}\sum_{i=1}^{num(s)}\sum_{j=1}^{num(s)}|E(i) - E(j)| \quad (6)$$

where $E_{s(G)}$ is the energy Gini coefficient of the s^{th} cluster, num(s) is the number of sensors in the s^{th} cluster, $E_{ave}(s)$ is the average remaining energy of the s^{th} cluster, and E(i) signifies the residual energy of node i.

The standard deviation of the Gini coefficient was computed using Equation 7. A lower value of E_s indicates that the energy balance is similar in the cluster and can be used to form an effective cluster.

$$E_\sigma = \sqrt{\frac{\sum_{s=1}^{k}(E(s) - E_{ave}0)^2}{k}} \quad (7)$$

where K represents the total CHs of the present network, E(s) is the residual energy of the s^{th} cluster, E_{ave} is the average residual energy of each cluster, and E_σ represents a measure of the degree of distribution of $E_{s(G)}$ of k clusters.

The fitness function corresponding to the Gini coefficient is given in Equation 8.

$$f_1 = \frac{e^k}{K}.E_\sigma \quad (8)$$

The fitness function f_2 corresponds to CH energy balance (f_{21}) and CH energy proportion (f_{22}), as given in Equations 9 and 10. The overall fitness function represents the energy balance degree and energy ratio of CHs, as given in Equation 11.

$$f_{21} = 1 - [\frac{1}{k}\sum_{s=1}^{k}(\frac{E_{CH}(s)}{E_{ave}(CH)})^{1-\varepsilon}]^{1/1-\varepsilon} \quad (9)$$

where k is the total CHs in the network, $E_{CH}(s)$ is the residual energy of the s-th CH, $E_{ave}(CH)$ denotes the average residual energy of the CH, and ε characterises the disparity aversion parameter, which is 0.5. A lesser f_{21} depicts better energy balance between the CHs.

$$f_{22} = \frac{\sum_{i=1}^{n}E(i)}{\sum_{s=1}^{k}E_{CH}(s)} \quad (10)$$

n is the number of surviving nodes in the existing network, E(i) is the total energy of the current network, and ECH(s) is the total residual energy of the CHs.

$$f_2 = \omega_1 f_{21} + \omega_2 f_{22} \quad (11)$$

The fitness function f_3 is subjected to the sum of the distance between all CHs (f_{31}) and the total distance between the node and its corresponding CH (f_{32}), as given in Equations 12, 13, and 14. A larger (f_{31}) denotes that the clusters are evenly distributed. A smaller value of (f_{32}) indicates that the distance between nodes and CH is lesser than is used to indicate the compactness of the cluster.

$$f_3 = \frac{f_{32}}{f_{31}} \quad (12)$$

$$f_{31} = \sum_{s=1}^{k-1}\sum_{m=s+1}^{k}d_{CH}(s,m) \quad (13)$$

$$f_{32} = \sum_{i=1}^{k} \sum_{j=1}^{num(i)} d_{CN}(i,j) \tag{14}$$

Fitness function f_4 is based on load balancing in CH, which is computed using Equations 15-18. For WSNs, balancing the load of the CHs is critical. CHs must undertake additional operations and spend more energy than their members. As a result, lowering the load on the CH can significantly increase the network speed. The load on the CH corresponds to the number of members in the cluster.

$$num_{ave} = \frac{n-k}{k} \tag{15}$$

$$Th_{max} = num_{ave} + \frac{num_{max} - num_{min}}{k} \tag{16}$$

$$Th_{max} = num_{ave} + \frac{num_{max} - num_{min}}{k} \tag{17}$$

$$f_4 = \left(\frac{num_{max} - num_{ave}}{num_{max}}\right) \frac{num_h - num_u}{k} \tag{18}$$

Here, num_{ave} represents the average count of sensor nodes in each cluster, num_{max} and num_{min} depict the count of sensor nodes in the largest and smallest clusters, num_h characterises several clusters with a greater number of sensor nodes than Th_{max}, and num_u is the number of clusters with fewer sensor nodes than Th_{min}.

It produces a new solution (position) from the old one in memory (position update equation for jth direction of the ith candidate) using Equation 19.

$$v_{ij} = \chi_{ij} + \emptyset_{ij}(\chi_{ij} - \chi_{kj}) \tag{19}$$

where

- $\emptyset_{ij}$ $(\chi_{ij} - \chi_{kj})$ is called step size and ($i \neq k$)
- k€ (1,2,......SN)
- j€ (1,2,......D)
- $\emptyset_{ij}$: random number within [-1, 1]

If the position bees remain un-updated for a predefined duration (limit), then the food source χ_i is abandoned.

3. Results and discussion

The proposed system is simulated using MATLAB R2018b on a Windows environment using a personal coputer with 8 GB RAM and a Core i5 processor with 2.64 GHz speed. The network parameters and specifications are listed in Table 1. Results of FCM and FCM-ABC for WSN clustering and CH selection are presented in Figure 3.

Table 1 System and network parameters and specifications.

System Parameter	Specifications
Number of Node	50,100,200,300,…500
Node Position	Fixed and Mobile
Simulation Area	100 m × 1000 m
Base Station Position	Fixed and Mobile
Initial Energy (Eo)	2 J
MAC Protocol	802.11
Traffic Patterns	Constant Bit Rate (CBR)
Threshold Distance (do)	$\sqrt{E_{fs} / E_{mps}}$
Energy Dissipated Per Bit (Eelec)	50 nJ/bit
Receiver Power Dissipation (ERX)	50 nJ/bit
Transmission Power Dissipation (ETX)	50 nJ/bit
Amplification Factor for Moltipath (Emp)	0.0013 pJ/bit/m4
Amplification Factor for Free Space (Efs)	10 pJ/bits/m2
Message Bit (K)	2000 bits

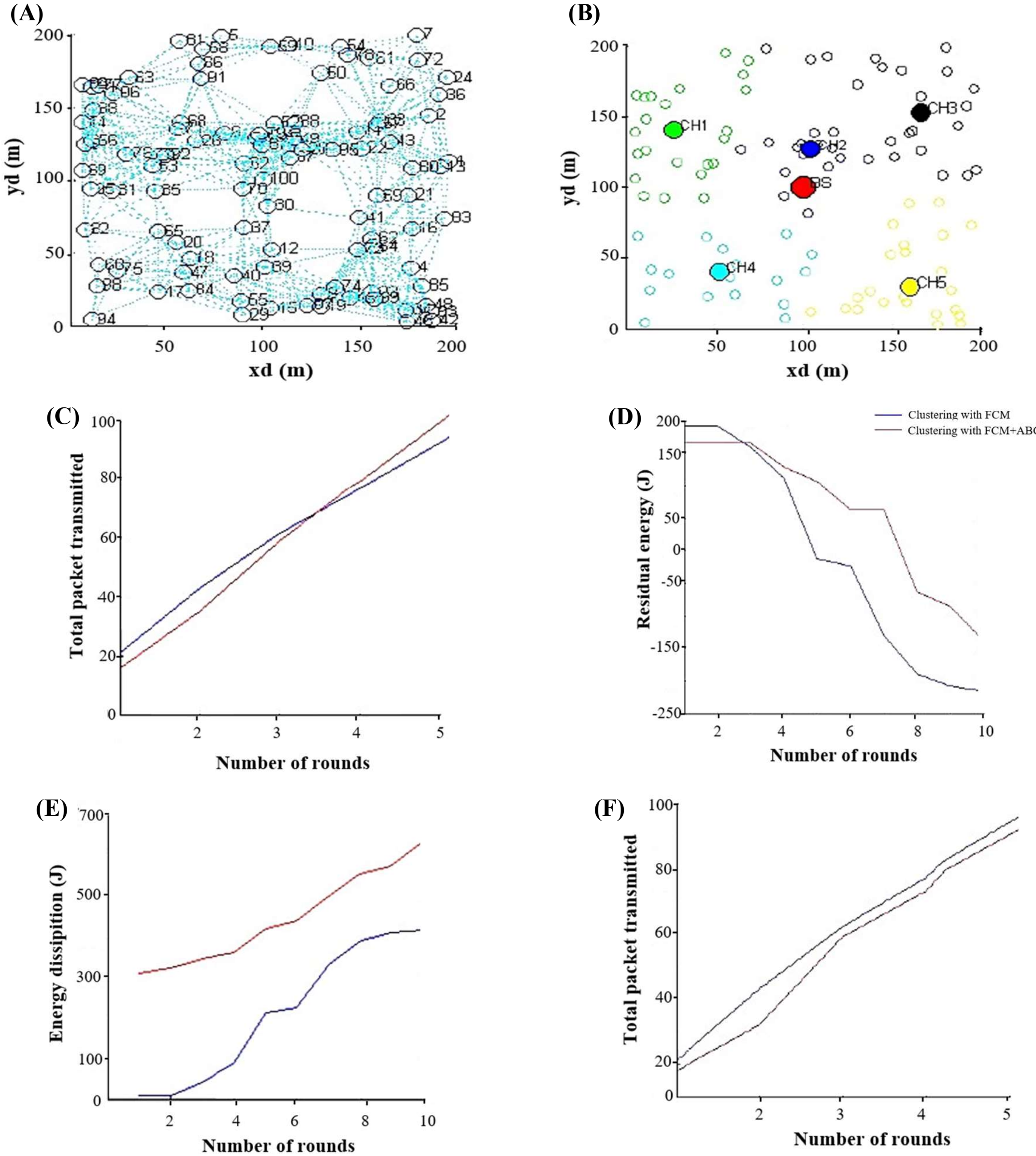

Figure 3 Simulation performance of proposed algorithm: (A) Network scenario for N=100, and (B) Clustering and CH optimization using FCM-ABC, (C) packet transmitted to CH, (D) Residual energy, (E) Energy dissipation, and (F) Packet transmittion to BS.

Table 2 provides information regarding network packet delivery for FCM and FCM-ABC for different nodes over a simulation area of 1000 m × 1000 m. An FCM-ABC shows significant improvement over the FCM for data transmission toward the base station.

The experimental results show that optimal selection of CH improves the network lifetime, resulting in higher packet throughput for CH and BS and higher residual energy. The variability in the number of nodes indicates that the proposed algorithm can perform efficiently in dense networks.

In recent years, machine and deep learning have attracted more attention for various signal processing and data analytics applications such as face recognition, speech recognition, and object detection. Deep learning algorithms have shown significant improvement over the traditional benchmark methods in terms of their performance. However, deep learning has rarely been addressed in WSN applications for routing, clustering, and data aggregation [17-21]. In the future, deep learning schemes can be presented for WSN clustering, routing, and data aggregation to improve the performance.

Table 2 Network Packet delivery performance.

Algorithm	Number of node	Total packet transmitted to base station	The total inrail energy of the network (J)
	100	80	200
	200	180	400
FCM	300	270	600
	400	380	800
	500	513	1000
	100	95	200
	200	192	400
FCM-ABC	300	284	600
	400	421	800
	500	586	1000

4. Conclusion

This study presents optimal clustering and CH selection using FCM and an improved ABC optimisation algorithm to tackle the problems of low energy efficiency and poor network lifetime of WSN. Cluster selection was based on connectivity, distance from the base station, residual energy, and position in the cluster. It has shown significant improvement over centralised CH selection using the FCM algorithm for various network density and scalability conditions. The proposed algorithm can provide better performance in real-time scenarios because of its adaptability to environmental changes. In the future, the performance of the proposed scheme will be validated using real-time mobility models of WSNs.

5. Acknowledgments

I would like to express our sincere thanks to Marathwada Shikshan Parasarak Mandal's Shri Shivaji Polytechnic Institute, Parbhani for the continuous support of my research work.

6. References

[1] Kirubasri G. A contemporary survey on clustering techniques for wireless sensor networks. Turk J Comput Math Educ. 2021;12(11):5917-5927.

[2] Kaur A, Gupta P, Garg R. Soft computing techniques for clustering in WSN. IOP Conf Ser: Mater Sci Eng. 2021;1022:012041.

[3] Daanoune I, Abdennaceur B, Ballouk A. A comprehensive survey on LEACH-based clustering routing protocols in Wireless Sensor Networks. Adv Hoc Netw. 2021;114:102409.

[4] Rawat P, Chauhan S. Clustering protocols in wireless sensor network: A survey, classification, issues, and future directions. Comput Sci Rev. 2021;40:100396.

[5] Ramani KPL, Badholia A. Cluster based routing protocols In WSN Sensor. Information. Technol Interface Industry. 2021;9(1):198-206.

[6] Rathore PS, Chatterjee JM, Kumar A, Sujatha R. Energy-efficient cluster head selection through relay approach for WSN. J Supercomput. 2021;77(7):7649-7675.

[7] Mishra PK, Verma SK editors. A survey on clustering in wireless sensor network. 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT); 2020 July 1-3; Kharagpur, India. Kharagpur: IEEE Publications; 2020. p. 1-5.

[8] Raam KVJ, Rajkumar K. A Novel Approach using Parallel Ant Colony Optimization Algorithm for Detecting Routing Path based on Cluster Head in Wireless Sensor Network. Indian J Sci Technol. 2015;8(16):1-7.

[9] Liu X, Fu H. An effective clustering algorithm with ant Colony. J Comput. 2010;5(4):598-605.

[10] Gupta V, Sharma SK. Cluster head selection using modified ACO. Proceedings of the fourth international conference on soft computing for problem solving. Adv Intell Syst Comput. 2015;335:11-20.

[11] Aadil F, Bajwa KB, Khan S, Chaudary NM, Akram A. CACONET: Ant Colony Optimization (ACO) Based Clustering Algorithm for VANET. PLoS One. 2016;11(5):e0154080.

[12] Yang J, Xu M, Zhao W, Xu B. A multipath routing protocol based on clustering and ant colony optimization for wireless sensor networks. Sensors (Basel). 2010;10(5):4521-4540.

[13] Kalaimani D, Zah Z, Vashist S. Energy-efficient density-based Fuzzy C-means clustering in WSN for smart grids. Aust J Multidiscip Eng. 2021;17(1):23-38.

[14] Dagur A, Malik N, Tyagi P, Verma R, Sharma R, Chaturvedi R. Energy enhancement of WSN using fuzzy C-means clustering algorithm. In: Jacob IJ, Shanmugam SK, Piramuthu S, Gilski PF, editors. Data Intelligence and Cognitive Informatics. Singapore: Springer. 2021. p. 315-323.

[15] Karim SM, Ozturk C, Mahmood MK. ABC-based optimization of cluster head selection in wireless sensor networks. Int J Electr Eng Inform. 2021;13(2):287-296.
[16] Almajidi AM, Pawar VP, Alammari A, Ali NS. ABC-based algorithm for clustering and validating WSNs. In: Gunjan VK, Suganthan PN, Haase J, Kumae A, Raman B, editers. Cybernetics, Cognition and Machine Learning Applications. Singapore: Springer; 2020. p. 117-125.
[17] Bhangale K, Mohanaprasad K. Speech emotion recognition using mel frequency log spectrogram and deep convolutional neural network. In: Sivasubramanian A, Shastry PN, Hong PC, editors. Futuristic Communication and Network Technologies. Singapore: Springer; 2022. p.241-250.
[18] Bhangale KB, Mohanaprasad K. A review on speech processing using machine learning paradigm. Int J Speech Technol. 2021;24(2):367-388.
[19] Bhangale KB, Kothandaraman M. Survey of deep learning paradigms for speech processing. Wireless Pers Commun. 2022;125(2):1913-1949.
[20] Bhangale K, Ingle P, Kanase R, Desale D. Multi-view multi-pose robust face recognition based on VGGNet. In: Chen JLZ, Tavares R.S, Lin Du K, IIiyasu AM, editors. Second International Conference on Image Processing and Capsule Networks. Cham: Springer; 2022. p.414-421.
[21] Bhangale KB, Desai P, Banne S, Rajput U, editors. Neural Style Transfer: reliving art through Artificial Intelligence. 3rd International Conference for Emerging Technology (INCET). 2022 May 27-29; Belgaum, India. Belgaum: IEEE Publications; 2022.
[22] Vijayalakshmi K, Anandan P. Global levy flight of cuckoo search with particle swarm optimization for effective cluster head selection in wireless sensor network. Intell Autom Soft Comput. 2020;26(2):303-311.
[23] ShayestehTabatabaei S. A novel fault tolerance energy-aware clustering method via social spider optimization (SSO) and fuzzy logic and mobile sink in wireless sensor networks (WSNs). Comput Syst Sci Eng. 2020;35(6):477-494.
[24] Wang J, Han H, Li H, He S, Sharma PK, Chen L. Multiple strategies differential privacy on sparse tensor factorization for network traffic analysis in 5G. IEEE Trans Ind Inform. 2021;18(3):1939-1948.
[25] Wang J, Gao Y, Zhou C, Simon Sherratt R, Wang L. Optimal coverage multi-path scheduling scheme with multiple mobile sinks for WSNs. Comput Mater Continua. 2020;62(2):695-711.