International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Fuzzy based Reliable Data Collection and Communication in Artificial Intelligence of Things (AIoT) Networks

B. Maria Joseph*[1], Dr. K. K. Baseer[2]

Submitted: 22/08/2023 **Revised**: 07/10/2023 **Accepted**: 21/10/2023

Abstract: Enhancing the efficiency of Internet of Things (IoT) operations is the primary objective of Artificial Intelligence for Things. In harsh environments, IoT nodes are prone to failures due to hardware faults, battery depletion and external events etc. By analyzing the correctness and quality of received data, the IoT device's quality can be assessed. In this paper, Fuzzy based reliable data collection and communication (FRDCC) is proposed. For detecting the data faults, each device applies machine learning based Autoencoder technique. The fault detection module receives the device readings as input which is then used to monitor the data correctness. In order to ensure reliable data collection and transmission, a set of data collectors are determined by applying Fuzzy logic model. The variables queue size, total energy consumption and reliability index are considered as fuzzy inputs and the selection probability is returned as the Fuzzy output. From experimental results, it was shown that the proposed technique has an higher fault detection rate and data correctness with reduced packet drop and recovery delay.

***Keywords**: Artificial Intelligence, Internet of Things (IoT), Wireless Sensor Networks (WSN), Data Collection, Autoencoder, Fuzzy decision model,Fault-tolerant,Reliable*

1. Introduction

IoT gives billions of devices omnipresent wireless connectivity. This network is often made up of massive clusters of devices that are dispersed spatially over large geographic areas. It enables the management and connection of many public services and infrastructure in smart cities [1]. These IoT devices have a variety of sensors to measure, monitor, and report on some physicaloccurrences.

Devices with Artificial Intelligence (AI) capabilities are more intelligent and hence help to conserve lot of time and resources. While IoT provides a framework for device communication and data collection, AI gives the system a "brain" and improves its capacity to handle the available data. Artificial Intelligence of Things (AIoT) aims to improve data management and analytics capabilities and make IoT operations more efficient [2]. As IoT systems operate continuously and produce large amounts of multi-modal data, it is essential to guarantee the correct functioning of the IoT's devices. Nodes in IoT applications that operate in harsh environments are vulnerable to failure, due hardware issues, battery drain, and other external factors [7]. Therefore, an accurate monitoring process should be applied to confirm the behaviour and effectiveness of the IoT devices [9].

1.1 Objectives and Major Contributions

To ensure reliable data collection, it is required to check the correctness of data and the reliability of data collectors. But unfortunately, there are very limited works which can handle boththese issues.

In this paper, Fuzzy based reliable data collection and communication (FRDCC) in AIoT sensor networks is proposed.

The main objectives of this work are

i) To remove faulty or inconsistent data

ii) To ensure reliable data collection

Before training a network, prior deep learning techniques required random initialization of parameters. "As a result of the enormous amount of time needed to train prior deep neural networks (DNNs), the technology was deemed unusable for practical purposes. But in Autoencoder (AE), each layer will have its own pre-training stage. Hence the first objective is satisfied bythe AE based data fault detection.

The major reasons for the failure of data collectors are overloading of queue, high energy depletion and poor reliability. Hence the second objective is satisfied by the Fuzzy based reliable data collection method, which considers the queue size, energy consumption and reliability index as inputs and the returns the selection

[1] Research Scholar, CSE Department, Jawaharlal Nehru Technological university Anantapur, Ananthapuramu-515002, Andhra Pradesh., India
** Corresponding Author Email: josephmariab@gmail.com*
[2] Professor in Department of Information Technology, Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College (Autonomous), Affiliated to Jawaharlal Nehru Technological University Anantapur, Ananthapuramu),Tirupati-517102. Andhra Pradesh, India. Email Id: drkkbaseer@gmail.com

probability as output.

Further in Section II we discuss about the related work of data collections. In Section III proposed solution is conversed in detail. Section IV, consist of the simulation outcome and discussion on the results with PBRB, Section V gives the conclusion of proposed work in AIoT networks.

2. Related Works

An Intelligent Proficient Data Collection Approach (IPDCA) has been suggested by Walid Osamy et al. [2] to provide public data in a smart city setup. IPDCA makes use of public transportation as D-collectors that retrieve data from numerous Access Points (APs) and transmit it back to the main Base Station (BS). In addition, IPDCA uses an altered version of the Bat algorithm to solve our discrete optimization problem when determining the path of D-collectors.

Merim Dzaferagic et al. [6] have suggested a generative model for fault detection and classification purposes for Intelligent IoT (IIoT). In order to ensure that the performance of the monitoring system is unaffected by missing data, the missing sensor measurements are computed and replaced. They used Generative Adversarial Networks (GANs) in particular to produce missing sensor readings and they proposed fine-tuning the GAN's training based on how the generated data affected the modules for fault detection and classification.

A belief rule base with power set (PBRB) fault detection technique has been put forth by Guo-Wen Sun et al. [7]. In this approach, the reasoning process is evidential reasoning (ER), the parameter optimization algorithm is projection covariance matrix adaptive evolution strategy (P-CMA-ES), and the fuzzy information is represented using the power set identification framework.

A fault detection and error correction method based on redundant residue arithmetic has been proposed by Chinmaya Mahapatra et al. [8]. They demonstrated the benefits of the proposed solution in raising the quality of data transmission by contrasting it with the existing approach in terms of perceived packet loss rate and anticipated delivery latency.

A framework for online sensor fault detection has been put forth by Yu Liu et al [9]. They found inspiration for their method in the issue of data value mismatches and event detection. They contained the most recent sensor data using Statistics Sliding Windows (SSW) and regressed each window using a Gaussian distribution. Its outcome can be used to find the problem with the data value. Production devices might have varying workloads, and their associated sensors might be in varying states of operation.

We separate the sensors into a number of status groups in accordance with the various production flow chats. This links a sensor's status to those of the other sensors in its group. They generated a group trend vector by fixing the values in the Status Transform Window (STW) to determine the slope.

3. Proposed Solution

3.1 Overview

In this paper, FRDCC in AIoT sensor networks is proposed. For data fault detection, each device applies the Autoencoder (AE) technique. The fault detection module of AE receives sensor readings as input which is used to monitor the data correctness. In order to ensure, reliable data collection and transmission, a set of data collectors are estimated by applying Fuzzy Logic Model (FLM). The variables queue size, total energy consumption and reliability index are treated as input for the FLM and the selection probability (Sprob) is returned as the output.

3.2 Data Fault Detection

The method for extracting the data features has been developed. In WSN, there are correlations in both time and space between the data gathered by various sensors. The time and space correlation components will be altered by WSN node fault. As a result, it is necessary to analyse the sensor-collected raw data and extract these components that can be used as model input attributes. The following is a description of the extraction process:

$$\chi = \frac{\sum_{k=0}^{T}\left[x_i(t-k) - \overline{x_i}(t)\right]\left[x_j(t-k) - \overline{x_j}(t)\right]}{\sqrt{\sum_{k=0}^{T}\left[x_i(t-k) - \overline{x_i}(t)\right]^2 \sum_{k=0}^{T}\left[x_j(t-k) - \overline{x_j}(t)\right]^2}}$$

The data gathered by the sensor i at time [t-T,t] is indicated by $x_i(t-k)$ where K is equal to 0 to T. Its average value from (t –T) to t is indicated by $\overline{x_i(t)}$. The Autoencoder (AE) utilized in this paper helps in data fault detection.

3.2.1 Autoencoders (AE)

An AE is a feature learning method that employs backpropagation to set output values equal to input values. This particular type of neural network (NN) has a symmetric structure that can be split into 2 replicated sub-structures. In other words, the layers of AE can be divided into symmetrical numbers of encoder and decoder layers, and sometimes parameters.

It was developed to aid in the discovery of data structure without the need for labels, hence facilitating unsupervised learning and ensuring that output matched input.

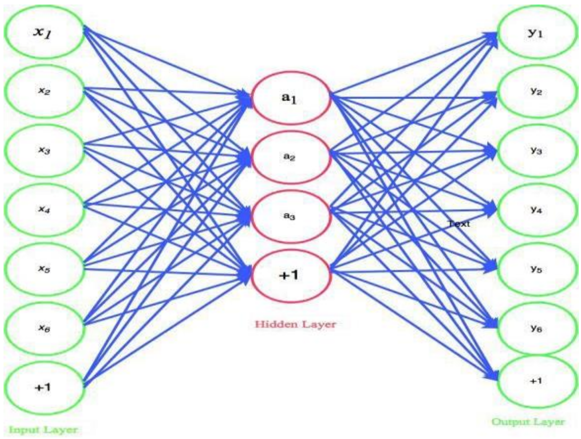

Fig. 1 Autoencoders

$$Encoding: a^{l+1} = f\left(\sum_{j=1}^{n} W_{i,j}^{(l)} a_j^{l} + b_i^{(l)}\right) \quad (1)$$

$$Decoding: a^{n+l+1} = f\left(\sum_{j=1}^{n} W_{i,j}^{(n-l)} a_j^{n+l} + b_i^{(n-l)}\right) \quad (2)$$

$$f(x) = \frac{1}{1+e^{x}} \quad (3)$$

The AE is shown in Figure 2. It performs compression for dimension reduction.

Equations (1), (2), and (3) depict the AE encoding and decoding process.

In AE, each layer will have its own pre-training stage. The primary goal in pre-training is to keep the training costs as low as possible.

One way we accomplish this is by setting the weight and bias parameters (W, b) as small values based on their respective values of W and b.

The AE is made up of three layers: a hidden HA layer with ReLu activation functions, an input layer with N inputs, and an output layer with N outputs. The first two hidden layers and the input layer are found in the encoder, and the final two hidden layers and the output layer are found in the decoder. The resulting AE has 12 hidden layers with the following sizes:

[52, 52, 48, 47, 46, 45, 46, 47, 48, 52, and 52].

3.2.2 Fault detection using AE

The data inconsistencies will be detected by AE technique and by eliminating those faulty data. N sensor readings are input to the fault detection module of the AE, which is used to check the accuracy of the data. The AE reconstructs the outputs from the input. It is important to remember that measurements without errors represent the system's typical operation and make up the majority of the data gathered.

In order to learn a notation for the input data and filter the noise, the model minimises the Root Mean Square Error (RMSE) of the recovered values.

The AE learns the patterns of typical working patterns by training the model only on data that is free from errors. In this manner, the corresponding reconstruction error for defective data will be higher than the error for faulty-free data.

Therefore, once sufficient measurements have been gathered, the data fault detection can be quickly performed. After the training, a threshold value is selected to minimize the RMSE because it will alert us to data errors, such as missing sensor measurements or irrelevant measurements.

3.3 Fuzzy based Reliable Data collection

3.3.1 Estimation of Parameters

The total energy consumption (TEj) of a node Nj can be derived as the sum of energy consumption during sensing (Ese), processing (Epr) , packet transmission from to BS (Eptx) and packet reception from sensor nodes (Erx) , which is represented as

$$\mathbf{TE_j = E_{se} + E_{pr} + E_{ptx} + E_{rx}} \quad \mathbf{(4)}$$

It is assumed that the packets are sent in bursts of n-bit data between the sender node and the receiver node.

The receiver's packet loss rate is given by [6].

$$P_e^L = 1-\left(1-\sum_{i=t+1}^{n}\binom{n}{i} p_b^i (1-p_b)^{n-i}\right)^{\left\lceil \frac{Lp}{k} \right\rceil} \quad (5)$$

Where pb is the bit error rate, Lp is the packet size of the payload in a single transmission

The SINR is estimated from the RSSI using the following equation

$$\text{SINR} = \frac{RSSI}{Bn} \quad (6)$$

where Bn is the background noise.

Then the transmission reliability of data transmission among the nodes Ni and Nj is given by

TR(Ni,,Nj) = c1. SINR + c2. (1-PLe) (7)

Where c1 and c2 are weighting constants in the range of (0,1).

The Reliability Index (RIj) of a node Nj is given by the product of transmission reliability (TR) and hardware reliability (HR) as

$$RI_j = \sum_{j=1}^{N} HR(N_j)TR(N_i, N_j) \quad (8)$$

The Queue size of the node Nj at time period Tn is given by

$$QS_j = SD_j(T_n) + \sum_{i \in Ne_i} AD_i(T_n) \quad (9)$$

Where SDj denotes the sensed data of Nj and ADi denotes

the size of aggregated data from each Neighbor Nei of Nj.

3.3.2 Fuzzy Logic Model (FLM)

In order to ensure, reliable data collection and transmission, a set of data collectors are selected using FLM. The variables QS ,TE and RI (estimated in the previous sub-section) are treated as input for the FLM and the selection probability (Sprob) is returned as the output.

The following list contains the phases that make up theFDM:

- **Fuzzification**: This entails getting determining how well the hard inputs fit into each relevant fuzzy set using the hard inputs from the chosen input variables.
- **Rule evaluation**: The enter variables are utilised by the fuzzy rules' forerunners. The resulting membership function is then used.
- **Collection of the rule outputs**: This includes aggregation of the result of the complete set of rules.
- **Defuzzification:** In this phase, the aggregated output set is taken in
- to consideration and a specific crisp value is acquired as the result.

The FLM is shown in Figure 2.

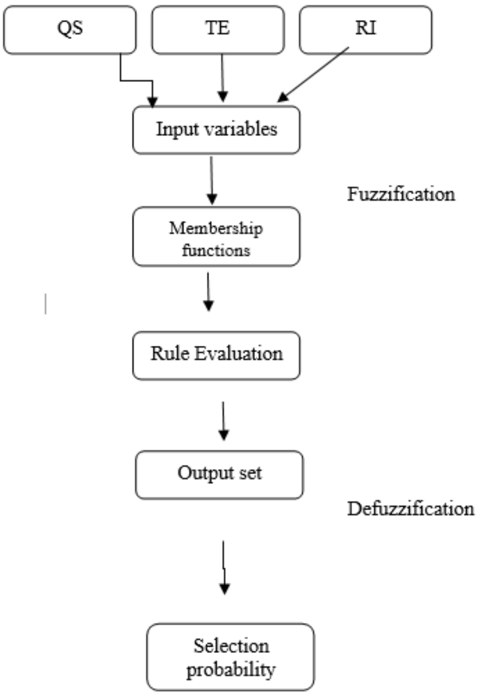

Fig 2 Architecture of proposed FDM

Fuzzification: The input variables QS ,TE and RI are provided a degree to suitable fuzzy sets. We take the possibilities of High, Medium and Low for the input variables. The output variable Sprob assumes Low, medium and High values. Figure 3 to 6 display the membership functions for the input and output variables.

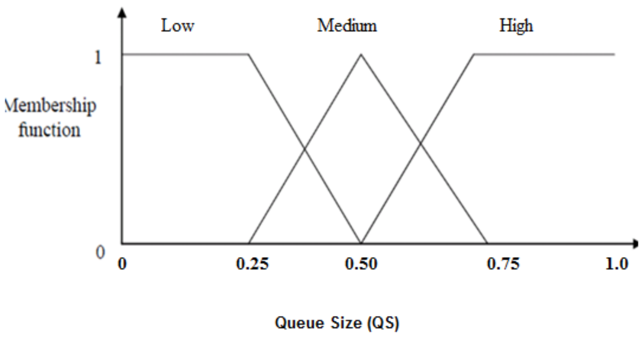

Fig 3 Membership function of QS

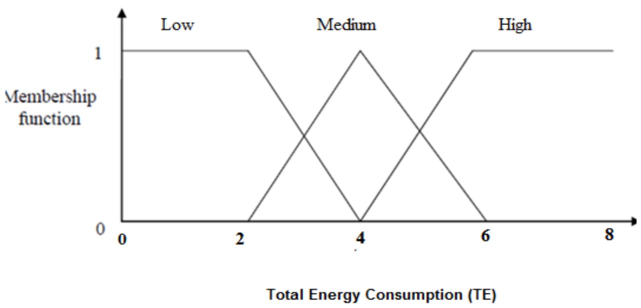

Fig 4 Membership Function of TE

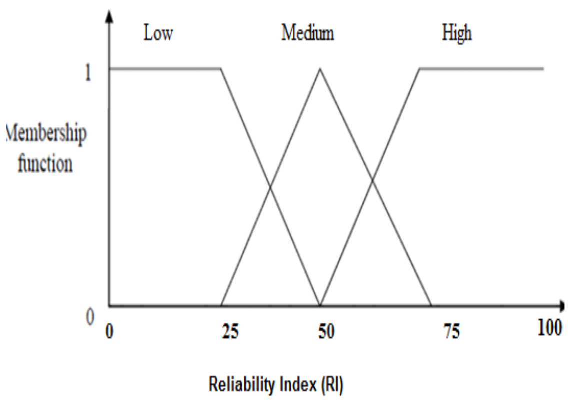

Fig 5 Membership Function of RI

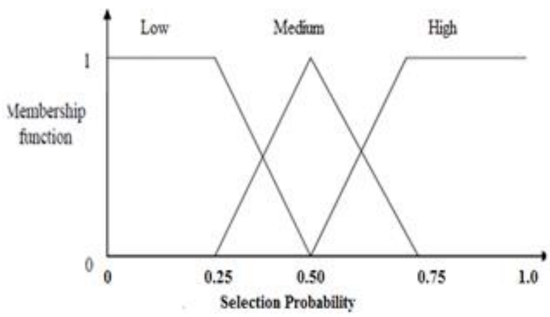

Fig 6 Membership Function of S_{prob}

Table 1 shows the defined fuzzy rules

S.No	QS	TE	RI	S_{prob}
1	High	High	Low	Low
2	High	High	Medium	Low
3	High	High	High	Medium
4	High	Low	Low	Low
5	High	Low	Medium	Medium
6	High	Low	High	Medium
7	High	Medium	Low	Low
8	High	Medium	Medium	Medium
9	High	Medium	High	Medium
10	Low	High	Low	Low
11	Low	High	Medium	Low
12	Low	High	High	Medium
13	Low	Medium	Low	Low
14	Low	Medium	Medium	Medium
15	Low	Medium	High	High
16	Low	Low	Low	Low
17	Low	Low	Medium	High
18	Low	Low	High	High
19	Medium	High	Low	Low
20	Medium	High	Medium	Low
21	Medium	High	High	Medium
22	Medium	Low	Low	Low
23	Medium	Low	Medium	Medium
24	Medium	Low	High	High
25	Medium	Medium	Low	Low

S.No	QS	TE	RI	S_{prob}
26	Medium	Medium	Medium	Medium
27	Medium	Medium	High	Medium

Defuzzification

It is the process of removing a stale value that serves as a symbol value from a fuzzy group. For defuzzification, the centroid of area approach is considered.

$$F = [\sum_{allrules} z_i * \lambda(z_i)] / [\sum_{allrules} \lambda(z_i)] \qquad (10)$$

Where F is utilised to identify the degree of choice making, zi denotes all fuzzy rules and variable λ(zi) is its association function. The output of F is modified to a crisp value using this technique.

4. Simulation Results

4.1 Simulation Settings

The simulation of FRDCC technique is performed in NS2.

The simulation settings are shown in Table 2.

Nodes	100
Topology size	50m X 50m
MAC protocol	IEEE 802.15.4
Traffic model	Constant Bit Rate
Number of data flows	10
Data traffic rate	50Kb
Fault occurrence probability	0.2 to 1.0
Initial Energy	12 Joules
Transmit power	0.5 watts
Receive power	0.3 watts

Table 2 Simulation settings

4.2 Results & Discussion

4.2.1 Comparison of Training Accuracy

The Adam optimiser was used to train the neural network with a batch size of 103 samples over 100 epochs at a learning rate of 0.001.. The training dataset consists of 300 samples of the total 500 samples for the fault free scenario. From the remaining 200 samples, 100 is used as validation.

In this section, the training accuracy of the proposed Autoencoder(AE) of FRDCC is compared with ANN, Random Forest (RF) and Support Vector Machine (SVM) techniques for the training percentage of 35 to 75.

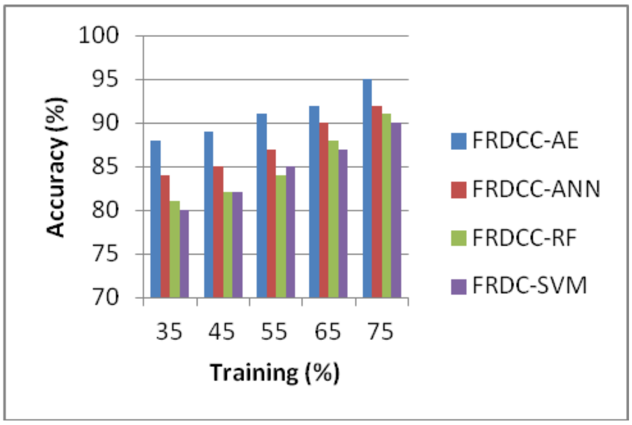

Fig 7 Comparison of Training Accuracy

From Figure 7, we can infer that the proposed FRDCC-AE achieves the highest accuracy (around 95%), followed by FRDCC-ANN (around 92%), FRDCC-RD (around 91%) and FRDCC-SVM (around 90%).

4.2.2 Performance Comparison with Existing Technique

In this section the performance of FRDCC has been compared with Belief Rule Base with powerset (PBRB) protocol [7]. The performance is evaluated with respect to fault detection rate, average packet drop, data correctness, residual energy and recovery delay, by varying the fault occurrence probability from 0.2 to 1.0.

Fault Occurrence Probability	**FRDCC (%)**	**PBRB (%)**
0.2	95.72	92.13
0.4	94.05	90.37
0.6	92.35	87.89
0.8	91.01	86.45
1.0	90.47	85.27

Table 3 Results of Fault Detection Rate

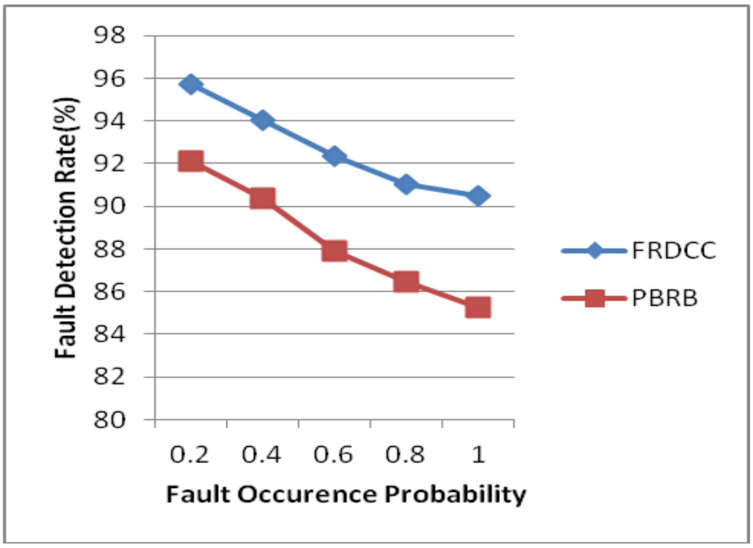

Fig 8 Fault occurrence probability Vs Fault Detection Rate

From figure 8, it is observed that the fault detection rate of FRDCC is 5% higher than PBRB.

Fault Occurrence Probability	FRDCC	PBRB
0.2	2719	3603
0.4	4096	5263
0.6	5706	6341
0.8	5840	7412
1.0	6340	9584

Table 4 Results of Packet Dropped

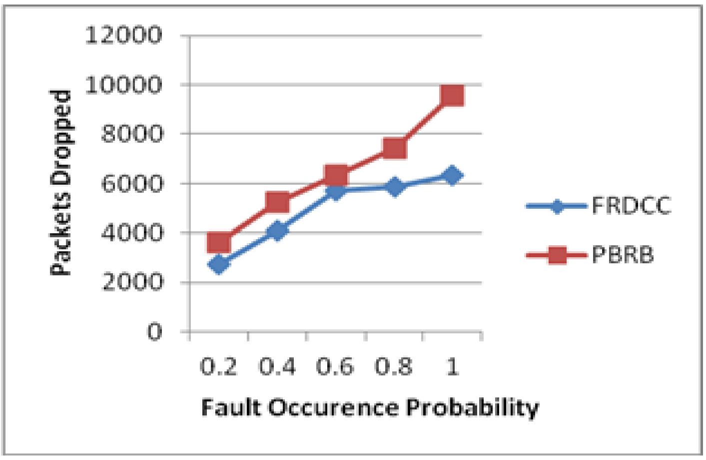

Fig 9 Fault occurrence probability Vs Packets Dropped

From figure 9, it is observed that the packets dropped of FRDCC is 22% lesser than PBRB.

Fault Occurrence Probability	FRDCC (%)	PBRB (%)
0.2	96.0	94.8
0.4	94.7	92.7
0.6	94.2	90.9
0.8	93.6	90.5
1.0	93.3	88.5

Table 5 Results of Correctness of data

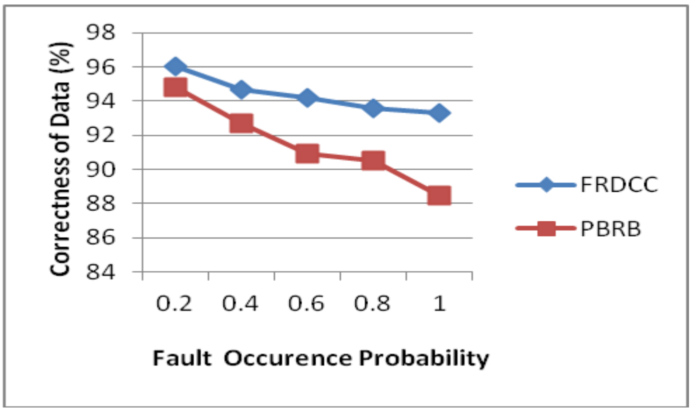

Fig 10 Fault occurrence probability Vs Correctness of Data

From figure 10, it is observed that the correctness of data of FRDCC is 3% higher than PBRB.

Fault Occurrence Probability	**FRDCC (Joules)**	**PBRB (Joules)**
0.2	15.09	12.84
0.4	14.83	12.62
0.6	14.44	12.38
0.8	13.75	11.83
1.0	13.04	11.21

Table 6 Results of Residual Energy

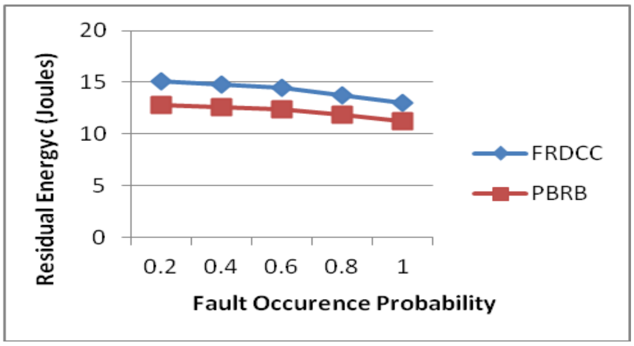

Figure 11 Fault occurrence probability Vs Residual Energy

From figure 11, it is observed that the residual energy of FRDCC is 14% higher than PBRB.

Fault Occurrence Probability	**FRDCC (ms)**	**PBRB (ms)**
0.2	19.73	24.28
0.4	21.77	27.54
0.6	23.05	29.86
0.8	25.95	33.39
1.0	27.72	35.46

Table 7 Results of Recovery Delay

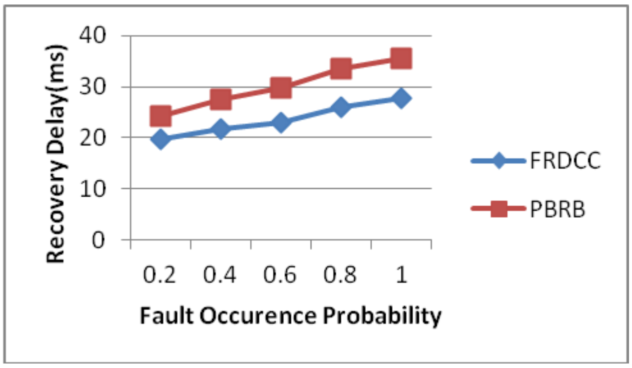

Fig 12 Fault occurrence probability Vs Recovery Delay

From figure 12, it is observed that the Recovery delay of FRDCC is 21% lesser than PBRB.

5. Conclusion

In this paper, FRDCC in AIoT sensor networks is proposed. For detecting the data faults, each device applies the Autoencoder (AE) technique. In order to ensure, reliable data collection and transmission, a set of data collectors are determined by applying FLM. The variables queue size, total energy consumption and reliability index are fed into the FLM and the selection probability (Sprob) is returned as the output. From experimental results, it was shown that the proposed FRDCC-AE achieves the highest accuracy around 95%. The simulation of FRDCC technique is performed in NS2 and its performance has been compared with PBRB protocol. Simulation results have shown that FRDCC-AE has higher fault detection rate and data correctness with reduced packet drop and recovery delay.

References

[1] Yasser Nabil, Hesham ElSawy, Suhail Al-Dharrab, Hassan Mostafa, and Hussein Attia, "Data Aggregation in Synchronous Large-scale IoT Networks: Granularity, Reliability, and Delay Tradeoffs", arXiv:2109.03563v1 [cs.IT], 2021.

[2] Walid Osamy, Ahmed M. Khedr, Ahmed A. El-Sawy, Ahmed Salim and Dilna Vijayan, "IPDCA: Intelligent Proficient Data Collection Approach for IoT-Enabled Wireless Sensor Networks in Smart Environments", Electronics, Vol-10,2021.

[3] Perigisetty Vedavalli, Deepak Ch, "Data Recovery Approach for Fault-Tolerant IoT Node", (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 13, No. 1, 2022.

[4] K. K. Baseer, B. Jaya Naga Varma, B. Harish, E. Sravani, K. Y. Kumar and K. Varshitha, "Design and Implementation of Electronic Health Records using Ethereum Blockchain," 2023 Second International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2023, pp. 784-791, doi: 10.1109/ICEARS56392.2023.10085012.

[5] Samad NAJJAR-GHABEL, Shamim YOUSEFI and Leili FARZINVASH, "Reliable data gathering in the Internet of Things using artificial bee colony", Turkish Journal of Electrical Engineering & ComputerSciences, Vol-26, pp:1710-1723,2018.

[6] Qiannan Wang and Haibing Mu, "Privacy-Preserving and Lightweight Selective Aggregation with Fault- Tolerance for Edge Computing-Enhanced IoT", Sensors, Vol-21, 2021.

[7] Joseph, B.M., Baseer, K.K. (2023). IoT-Sensed Data for Data Integration Using Intelligent Decision-Making Algorithm Through Fog Computing. In: Sharma, H., Shrivastava, V., Bharti, K.K., Wang, L. (eds) Communication and Intelligent Systems. ICCIS 2022. Lecture Notes in Networks and Systems, vol 689. Springer, Singapore. https://doi.org/10.1007/978-981-99-2322-9_34

[8] Merim Dzaferagic, Nicola Marchetti, Irene Macaluso, "Fault detection and classification in Industrial IoT in case of missing sensor data",TechRxiv.Preprint. https://doi.org/10.36227/techrxiv.14540061.v1, 2021.

[9] K. K. Baseer, S. B. A. Nas, S. Dharani, S. Sravani, P. Yashwanth and P. Jyothirmai, "Medical Diagnosis ofHuman Heart Diseases with and without Hyperparameter tuning through Machine Learning," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India,2023,pp.1-8,doi: 10.1109/ICCMC56507.2023.10084156.

[10] Guo-Wen Sun, Wei He, Hai-Long Zhu, Zi-Jiang Yang, Quan-Qi Mua and Yu-He Wang, "A wireless sensor network node fault diagnosis model based on belief rule base with power set", Heliyon 8 (2022) e10879, 2022.

[11] Chinmaya Mahapatra, Zhengguo Sheng, Victor C.M. Leung, and Thanos Stouraitis, "A Reliable and Energy Efficient IoT Data Transmission Scheme for Smart Cities based on Redundant Residue based Error Correction Coding", IEEE, 2015.

[12] Yu Liu,Yang Yang, Xiaopeng Lv, and LifengWang, "A Self-Learning Sensor Fault Detection Framework for Industry Monitoring IoT", Hindawi Publishing Corporation Mathematical Problems in Engineering, Volume 2013, Article ID 712028, 8 pages, 2013.

[13] Silpa C, Dr. S Srinivasa Chakravarthi, Jagadeesh kumar G, Dr. K.K. Baseer, E. Sandhya, “Health Monitoring System Using IoT Sensors”, Journal of Algebraic Statistics, Volume 13, No. 3, June, 2022, p. 3051-3056, ISSN: 1309-3452.

[14] Anuroop Gaddam, Tim Wilkin , Maia Angelova and Jyotheesh Gaddam, "Detecting Sensor Faults, Anomalies and Outliers in the Internet of Things: A Survey on the Challenges and Solutions", Electronics, Vol-9, No-11, 2021.

[15] K. Swetha, C. I. Shareef, G. Sreenivasulu, K. K. Baseer and M. J. Pasha, "Study on Implementation of Electronic Health Records using Blockchain Technology," 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2023, pp. 607-611, doi: 10.1109/ICESC57686.2023.10192992.

[16] Zainib Noshad, Nadeem Javaid, Tanzila Saba, Zahid Wadud,Muhammad Qaiser Saleem , Mohammad Eid Alzahrani and Osama E. Sheta, "Fault Detection in Wireless Sensor Networks through the Random Forest Classifier", Sensors, Vol-19, 219.

[17] Chaudhary, D. S. ., & Sivakumar, D. S. A. . (2022). Detection Of Postpartum Hemorrhaged Using Fuzzy Deep Learning Architecture . Research Journal of Computer Systems and Engineering, 3(1), 29–34. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/38

[18] Singh, J. ., Mani, A. ., Singh, H. ., & Rana, D. S. . (2023). Quantum Inspired Evolutionary Algorithm with a Novel Elitist Local Search Method for Scheduling of Thermal Units. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3s), 144–158. https://doi.org/10.17762/ijritcc.v11i3s.6175