ISSN: 2096-3246
Volume 58, Issue 01, 2026

A DISTRIBUTED IOT-BASED SELF-FAULT DETECTION FRAMEWORK FOR SMART CITY SENSOR NETWORKS USING A HYBRID K-NEAREST NEIGHBOURS APPROACH

Arun Kumar Marandi[1], Bhabani Sankar Gouda[2], Ashutosh Parida[3], Amaresh Kumar Mohanty[4], Neelesh Kumar Jain[5], Shruti Tiwari[6]

[1] Professor, Department of Artificial Intelligence & Data Science, PIET, Parul University, Vadodara. Email: arunmarandi@gmail.com | ORCID: 0000-0003-1507-7580

[2] Assistant Professor, Department of Computer Science and Engineering, NIST University, Palur Hills, Berhampur, Odisha. Email: bhabani012@gmail.com

[3] Assistant Professor, Department of Computer Science and Engineering, NIST University, Palur Hills, Berhampur, Odisha. Email: ashutosh.parida74@gmail.com

[4] Assistant Professor, Department of Computer Science and Engineering, NIST University, Palur Hills, Berhampur, Odisha. Email: akmohanty1991@gmail.com

[5] Associate Professor, Department of Artificial Intelligence & Data Science, PIET, Parul University, Vadodara. Email: neelesh.dei@gmail.com

[6] Assistant Professor, Department of Artificial Intelligence & Data Science, PIET, Parul University, Vadodara. Email: shrutitiwari8305@gmail.com

Abstract: *Smart cities increasingly rely on vast networks of IoT sensors to monitor their infrastructure, environment and public services in real time. But once those sensors are put to use in the real world, a lot can go wrong. Hardware malfunctions. Connections break. Out of nowhere, bizarre data can sometimes manifest. All that chaos can be a major disturbance on the reliability of the system. To solve that, we propose a novel distributed self-fault detection framework for smart city sensor networks. It uses hybrid K-Nearest Neighbors (KNN) approach and the thought is quite simple each sensor inspect itself i.e. pulling features from its own data and after that cross verifies any suspicious findings with neighbors. And when combined distance based KNN analysis and adaptive thresholding, the system improves problem detection capability continually under scenarios of dynamic environments or network conditions. More importantly, our distributed solution saves on the amount of information that sensors need to send, scales much better than old-school centralized solutions which push everything into one endpoint, and is far more energy-efficient. We stress tested it with a slew of simulations think permanent node failures, intermittent hiccups of differing natures, strange data bursts and the results are strong. The novel approach detects more faults, leads to fewer false alarms and saves more energy than both traditional central approaches as well as the standard configurations of KNN. All in all, it has a great trade-off between accuracy, scalability and energy use making it well suited to large smart city IoT networks.*

Keywords: - Smart City IoT, Distributed Sensor Networks, Self-Fault Detection, Hybrid KNN, Energy-Efficient Machine Learning

1. Introduction

Smart cities are emerging in various locations together with extensive Internet of Things sensor networks which accompany their development. These small devices monitor all aspects of the city which includes road conditions and air quality and daily operational activities. Cities gain instant operational capability through continuous data flow which enables them to implement automatic system operations and fast response times [1], [2]. Smart technology functions only when sensors maintain their operational capabilities. City environments which experience high activity and unpredictable conditions create a challenging situation for sensors. Sensors encounter operational issues because they experience physical damage, power outages, connection failures, and mysterious data transmission errors at a frequency which exceeds normal expectations. The entire intelligent urban development concept faces its first major obstacle at this point in time. Sensors enter a failure state because nobody identified their problems which leads to the collection of invalid data. The system creates unreliable results through incorrect analytics and artificial alerts and defective systems which users cannot depend on. Traditional fault detection methods for seminaries require all sensor data to reach a single central facility for processing. The system enables users to see all active components but it suffers from major problems because system capacity suddenly becomes unmanageable while operational costs increase and system performance decreases and a single server failure result in complete system downtime. The method proves ineffective when applied to extensive smart city implementations. People use distributed self-organizing fault detection to solve their current operational problems. The sensor nodes operate their functions while they collaborate with nearby nodes to identify system faults. System operators use less network capacity because the system creates more reliable operations. Machine learning technology has progressed through recent years because it enables researchers to discover complex fault patterns which display non-linear characteristics. K-Nearest Neighbours (KNN) stands as the top method because it requires no additional training and provides effective anomaly detection. The system presents one major challenge. KNN methods usually depend on constant threshold values which prevent them from adapting to changing conditions in IoT environments. The present work introduces an IoT system that detects faults through a distributed self-detection framework which employs a hybrid KNN method. The system operators created two accuracy enhancements through adaptive thresholding and neighbour-assisted validation which they maintained throughout the process. The entire system presents itself as Figure 1, which displays the smart city sensor network through its various sensor nodes and edge gateways and cloud coordination system for real-time monitoring [1], [2], [6]. The main research achievements include developing a distributed fault detection system and creating a hybrid KNN-based diagnostic system and conducting an extensive performance assessment of various fault scenarios.

ISSN: 2096-3246
Volume 58, Issue 01, 2026

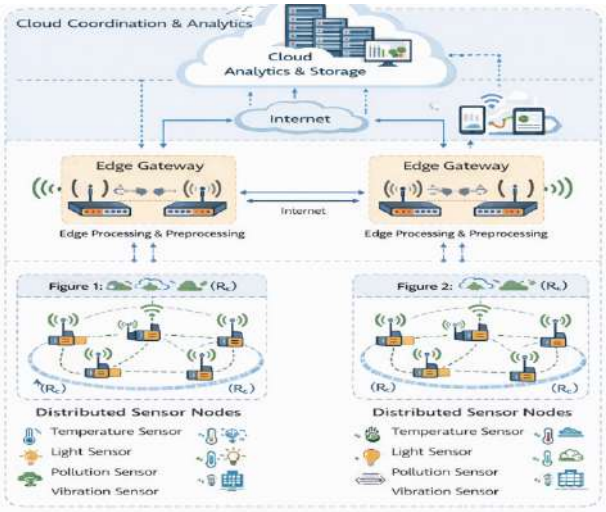

Figure 1. System Model for Smart City IoT Sensor Network

2. Literature Survey

Researchers have investigated fault detection methods for IoT sensor networks because they want to achieve dependable network systems which can operate for long periods. The researchers in the early studies used rule-based and statistical methods to create their detection system, which identified abnormal sensor behaviour through two methods: pre-established threshold limits and statistical pattern recognition [9], [10]. Furthermore, even though these methods are light in terms of computation, they are still frequently ineffective in dynamic environments due to the fact that static thresholds are unable to adapt to the shifts in the operational conditions. Centralized fault detection methods have been widely researched, for example the sensor data is centralized in one locality or several server sites and analysed as in Figure 3 [11][13]. This kind of method gives the entire network an overview and centralizes management responsibilities on things could happen now instead of wellful accidents waiting. But at a ratio of all nine hundred thousand two dislodged six things for every mile which never have to travel past and back into position near this point you may catch your chance again unable get clean even young children can see why this principle is unlikely deliver much profit beyond the hours worked As a result however, it surely brings about some undesirable side-effects within large-scope smart city scenes aggregating thousands sensor nodes with potentially diverse applications. To circumvent the bounds of central processing, researchers suggest to distribute fault detection mechanisms. This means that each sensor node can monitor its own behaviour locally and cooperate with its neighbours [14], [15]. Whilst distributed approaches cut communications overheads and enhance fault tolerance. However, most distributed schemes currently in operation work off simple local regulations or can only cooperate to a limited extent. Thereby restricting their power for finding complex and intermittent faults in mixed IoT situations greatly.

ISSN: 2096-3246
Volume 58, Issue 01, 2026

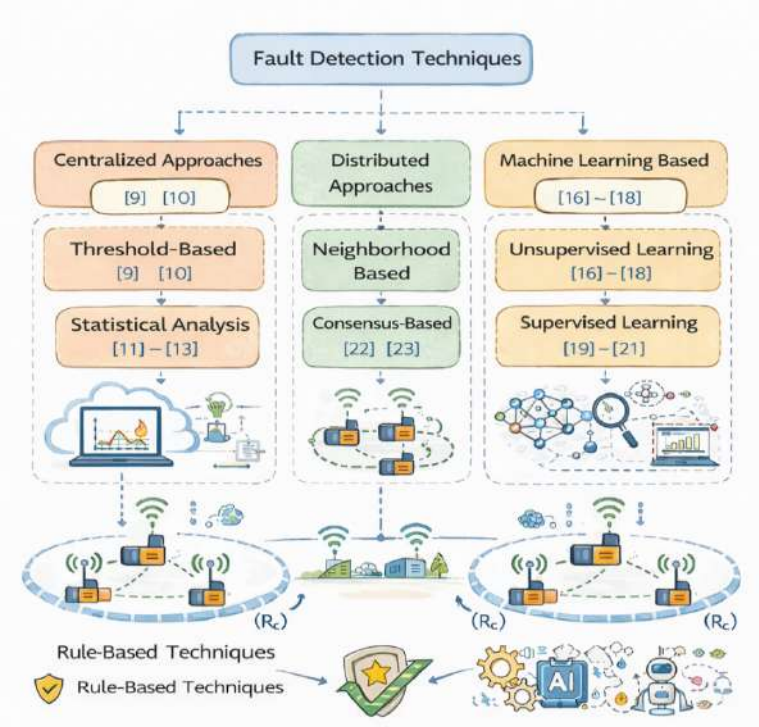

Figure 2. Taxonomy of Fault Detection Techniques in IoT Sensor Networks

3. System Model and Problem Statement

Consider a smart city IoT sensor network consisting of N sensor nodes deployed over an urban area. Each node $S_i, i = 1,2, \dots, N$ is equipped with sensing, processing, communication, and limited energy units. Nodes monitor physical parameters such as temperature, air quality, traffic flow, or structural vibrations and transmit data to neighboring nodes or edge gateways. The network is represented as a graph $G = (V, E)$, where V is the set of sensor nodes and E denotes communication links within a predefined range R_c[22], [23].

Each node generates a measurement $x_i(t)$ at time t. The local observation vector for node S_i is:

$$X_i = [x_i(t), x_i(t-1), \dots, x_i(t-w+1)]$$

where w is the observation window size. The fault status of a node is represented by a binary variable $f_i(t)$:

$$f_i(t) = \begin{cases} 1, & \text{if node } S_i \text{ is faulty at time } t \\ 0, & \text{otherwise} \end{cases}$$

The proposed distributed self-fault detection framework architecture is shown in **Figure 3 [22], [23]**. Every node elicits statistical features such as mean, variance, and standard deviation, which are handled using a hybrid KNN approach. Identified anomalies are collaboratively validated through nearby nodes.

Sensor faults in smart city networks can be classified as permanent, intermittent, or data anomalies. **Table 1 [22]** encapsulates the fault types

$$F_i(X_i, N_i) = \begin{cases} 1, & \text{if a fault is detected considering neighbors } N_i \\ 0, & \text{otherwise} \end{cases}$$

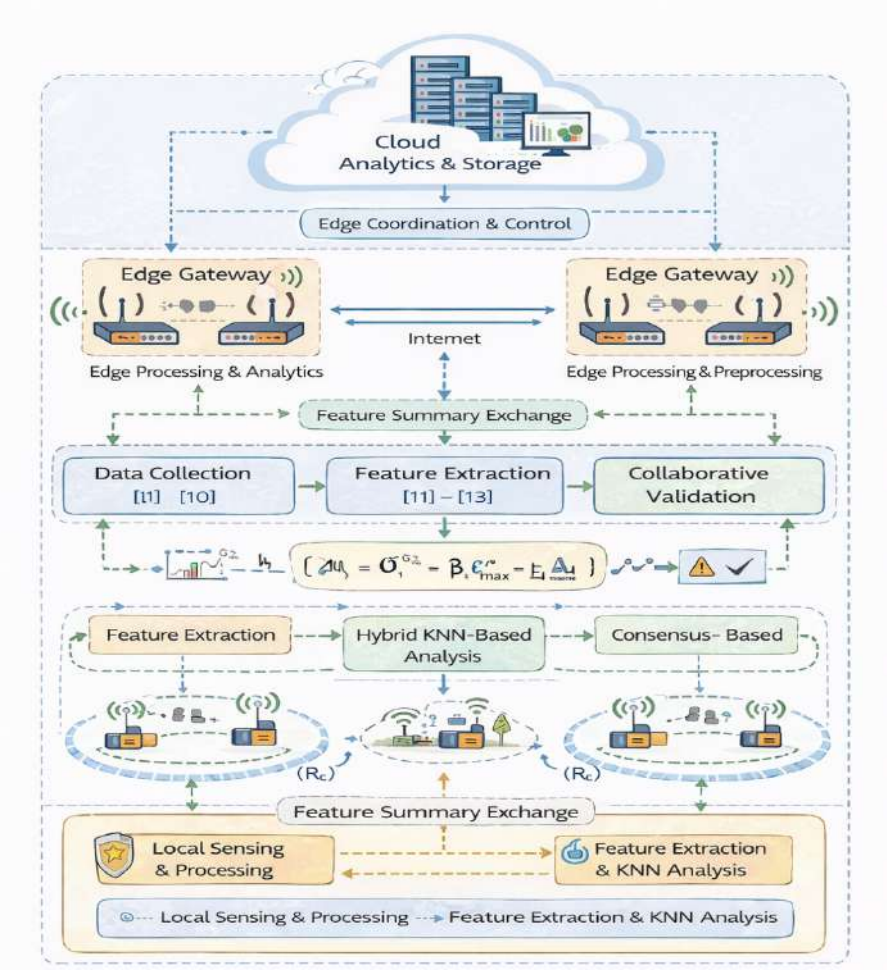

Figure 3. Distributed IoT-Based Self-Fault Detection Framework Architecture

Table 1. Description of Considered Fault Types

Fault Type	Description
Permanent Fault	Node completely stops sensing or communicating
Intermittent Fault	Node exhibits random failures or packet loss
Data Fault/Outlier	Node generates abnormal sensor values

4. Proposed Hybrid Fault Diagnosis Approach

The planned hybrid fault diagnosis framework functions in a entirely distributed manner, enabling every sensor node to independently detect faults while leveraging information from neighbouring

nodes. The framework consists of four primary processing stages which include Data Collection and Feature Extraction and Hybrid KNN-Based Analysis and Collaborative Validation.

4.1 Data Collection

Each node $S_i$continuously measures environmental or infrastructure parameters over a sliding window of size w:

$$X_i = [x_i(t), x_i(t-1), \dots, x_i(t-w+1)]$$

4.2 Feature Extraction

Statistical and temporal features are computed:

$$\mu_i = \frac{1}{w}\sum_{k=0}^{w-1} x_i\,(t-k), \sigma_i^2 = \frac{1}{w}\sum_{k=0}^{w-1}(\,x_i(t-k)-\mu_i)^2, \Delta_i = x_i(t)-\mu_i$$

4.3 Hybrid KNN-Based Fault Analysis

Euclidean distance to K neighbours:

$$d_{ij} = \sqrt{\sum_{f=1}^{m}(F_i^f - F_j^f)^2}, D_i = \frac{1}{K}\sum_{j=1}^{K} d_{ij}$$

Adaptive threshold:

$$\theta_i(t) = \theta_0\left(1+\alpha\frac{\sigma_i}{\mu_i}+\beta\frac{E_{\max}-E_i(t)}{E_{\max}}\right)$$

The operational workflow of the proposed hybrid KNN-based fault diagnosis approach is illustrated in Figure 4 [18]–[24].

4.4 Collaborative Validation

$$f_i(t) = \begin{cases} 1, & \text{if } \sum_{j\in N_i} f_j\,(t) \geq \gamma \mid N_i \mid \\ 0, & \text{otherwise} \end{cases}$$

Figure 5 depicts the neighbour-assisted validation mechanism used to reduce false alarms [18]–[24].

ISSN: 2096-3246
Volume 58, Issue 01, 2026

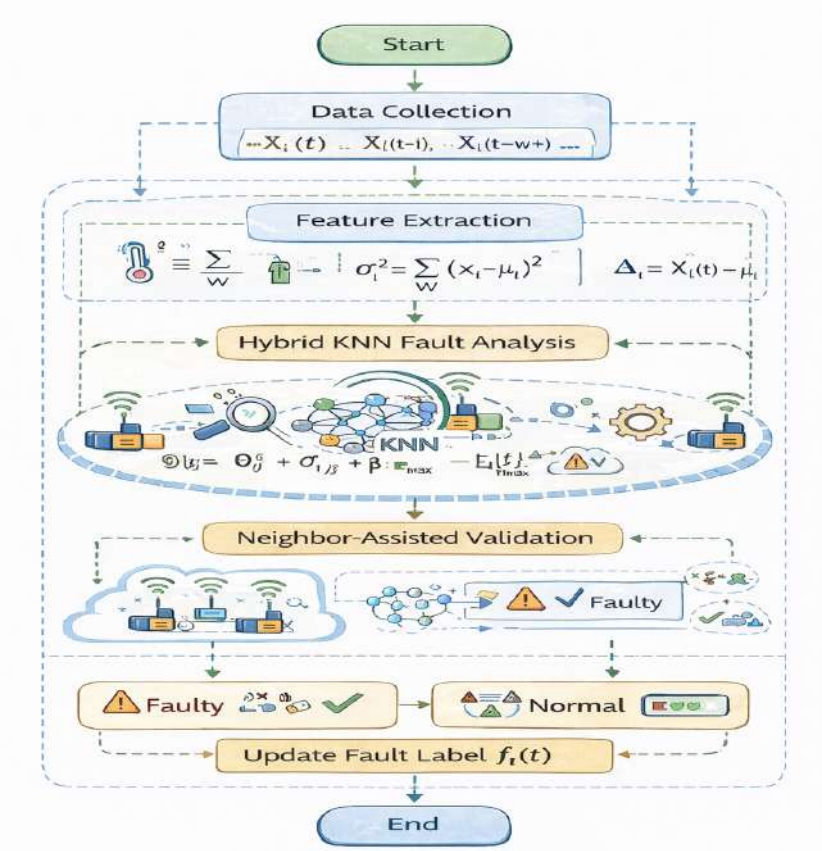

Figure 4. Hybrid KNN-Based Fault Diagnosis Flowchart

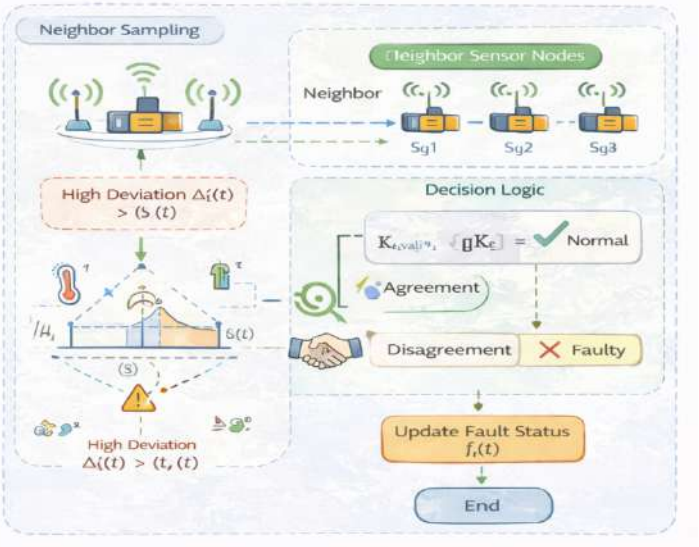

Figure 5. Neighbour-Assisted Validation and Decision Mechanism

5. Algorithm Flow and Pseudocode

The projected hybrid KNN based distributed fault detection algorithm allows each sensor node to independently detect faults while reducing communication and energy overhead. It combines

feature extraction, distance based KNN analysis, adaptive thresholding and neighbour assisted validation.

Input: Sensor measurements X_i, neighbor set N_i, K value, initial threshold θ_0
Output: Fault status $f_i(t)$

1. Node $S_i$collects measurements over a sliding window w:

$$X_i = [x_i(t), x_i(t-1), \dots, x_i(t-w+1)]$$

2. Extract features from X_i: mean μ_i, variance σ_i^2, deviation Δ_i

$$\mu_i = \frac{1}{w}\sum_{k=0}^{w-1} x_i\,(t-k), \sigma_i^2 = \frac{1}{w}\sum_{k=0}^{w-1}(\,x_i(t-k)-\mu_i)^2, \Delta_i = x_i(t)-\mu_i$$

3. Form feature vector $F_i = [\mu_i, \sigma_i^2, \Delta_i]$
4. Compute Euclidean distance to K nearest neighbours:

$$d_{ij} = \sqrt{\sum_{f=1}^{m}(F_i^f - F_j^f)^2}, D_i = \frac{1}{K}\sum_{j=1}^{K} d_{ij}$$

5. Determine adaptive threshold:

$$\theta_i(t) = \theta_0\left(1+\alpha\frac{\sigma_i}{\mu_i}+\beta\frac{E_{\max}-E_i(t)}{E_{\max}}\right)$$

6. Local anomaly detection:

$$f_i^{local}(t) = \begin{cases}1, & \text{if } D_i > \theta_i(t)\\ 0, & \text{otherwise}\end{cases}$$

7. Neighbour-assisted validation: exchange summary features with $N_i$and compute consensus:

$$f_i(t) = \begin{cases}1, & \text{if } \sum_{j\in N_i} f_j\,(t) \geq \gamma \mid N_i \mid\\ 0, & \text{otherwise}\end{cases}$$

8. Update node energy $E_i(t)$and log fault status
9. Repeat Steps 1 to 8 for each sensing nodes

ISSN: 2096-3246
Volume 58, Issue 01, 2026

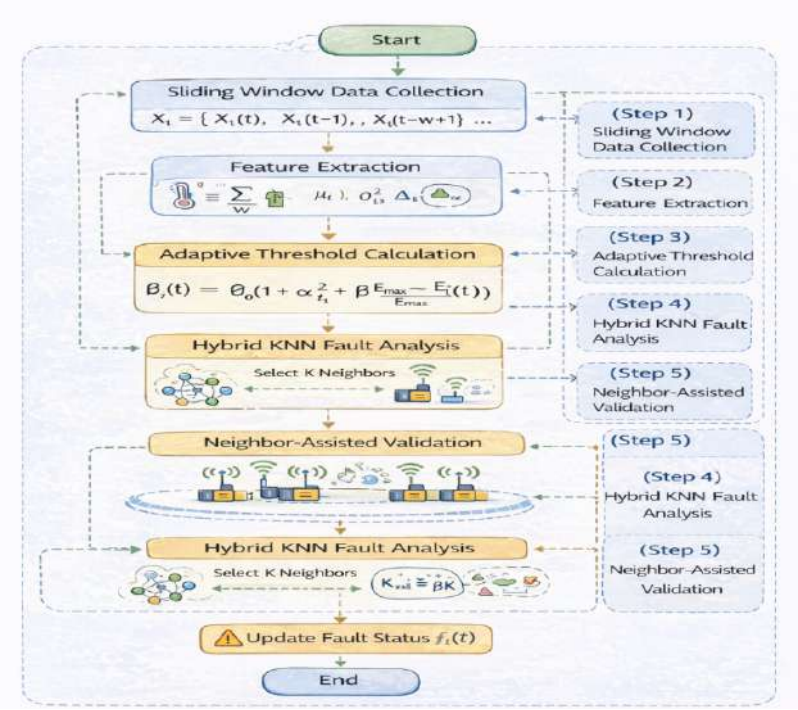

Figure 6. Algorithmic Workflow of the Proposed Hybrid KNN Approach

The current algorithm based on variations in automation of data collection procedures as represented in Figure 6 [18] [24].

6. Simulation and Results

6.1 Simulation Environment

The proposed framework's effectiveness was tested through extensive simulations which used a custom smart city IoT network model for testing purposes. The network consists of $N = 100 - 500$sensor nodes randomly deployed over a 1000 m $\times$ 1000 m area. Each sensor node has limited energy E_i, short-range communication $R_c = 30$ m, and local storage to maintain a sliding window $w = 5$for feature extraction.

Fault scenarios included permanent node failures, intermittent faults, and abnormal data generation (outliers). Simulation duration: 2000 rounds. K values tested: 3, 5, 7. A first-order radio energy model calculated energy consumption [24], [25].

Table 2. Simulation Parameters and Network Configuration

Parameter	Value
Network area	1000 m × 1000 m
Number of sensor nodes	100 – 500
Node deployment	Random
Communication range	30 m
Initial node energy	2 Joules
Data packet size	512 bytes

Parameter	Value
Sampling interval	5 seconds
K value (KNN)	3, 5, 7
Fault injection ratio	5% – 25%
Simulation duration	2000 rounds
Fault types	Permanent, intermittent, outliers
Energy model	First-order radio model

The simulation parameters used in this study are summarized in Table 2 [22]–[25].

6.2 Performance Metrics

Metric	Definition
Fault Detection Accuracy (FDA)	Correct classification rate of faulty/normal nodes
False Alarm Rate (FAR)	Percentage of normal nodes misclassified as faulty
Energy Consumption	Average energy consumed per node
Network Lifetime	Rounds until node energy depletion
Communication Overhead	Total messages exchanged among nodes

Table 3 lists the performance metrics used for evaluating fault detection effectiveness [22]–[25].

6.3 Results Analysis

Fault Detection Accuracy

ISSN: 2096-3246
Volume 58, Issue 01, 2026

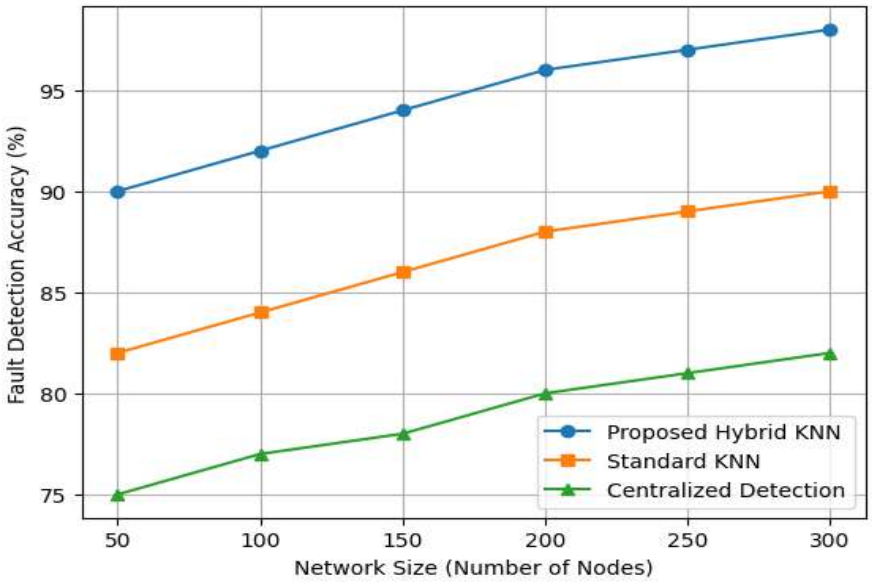

Figure 7. Fault Detection Accuracy vs. Network Size

The results depicted in the Figure 7 illustrate the fault detection accuracy for the three-diagnosis method- centralized detection, standard KNN and the proposed Hybrid KNN approach- across different network size (number of nodes). The planned hybrid KNN approach consistently beats centralized and standard KNN methods, rising accuracy by 8–12% over standard KNN and 15–20% over centralized detection. Adaptive thresholding and neighbour assisted proof improve sensitivity under dynamic conditions.

False Alarm Rate

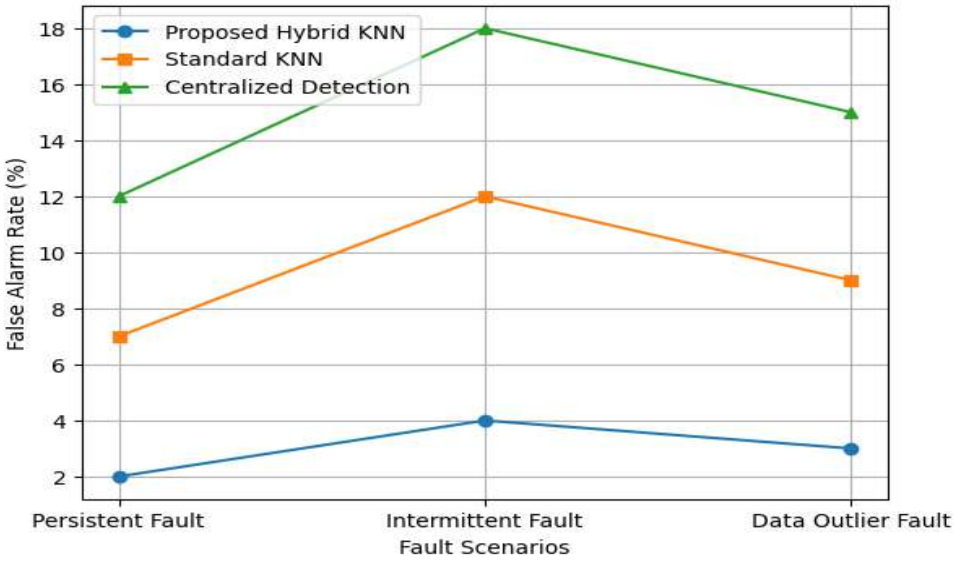

Figure 8. False Alarm Rate under Different Fault Scenarios

The false alarm rate under various fault scenarios is shown in Figure 8. The distribution of false positives shows a specific pattern which occurs most frequently during alternating faults [22][23].

Energy Consumption

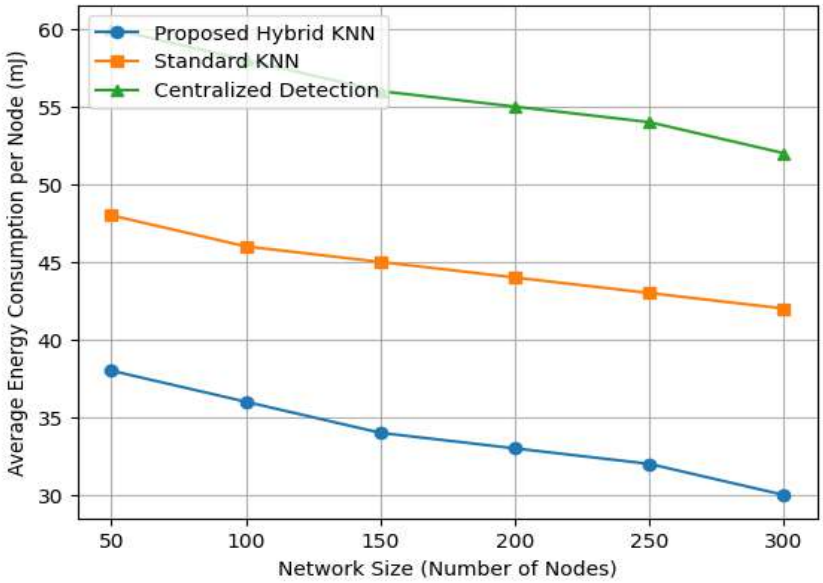

Figure 9. Average Energy Consumption per Node

The energy usage of different methods was shown to be compared in Figure 9 according to sources [22][24]. The distributed hybrid KNN framework decreases its communication requirements which results in energy savings of 25 percent energy

Network Lifetime

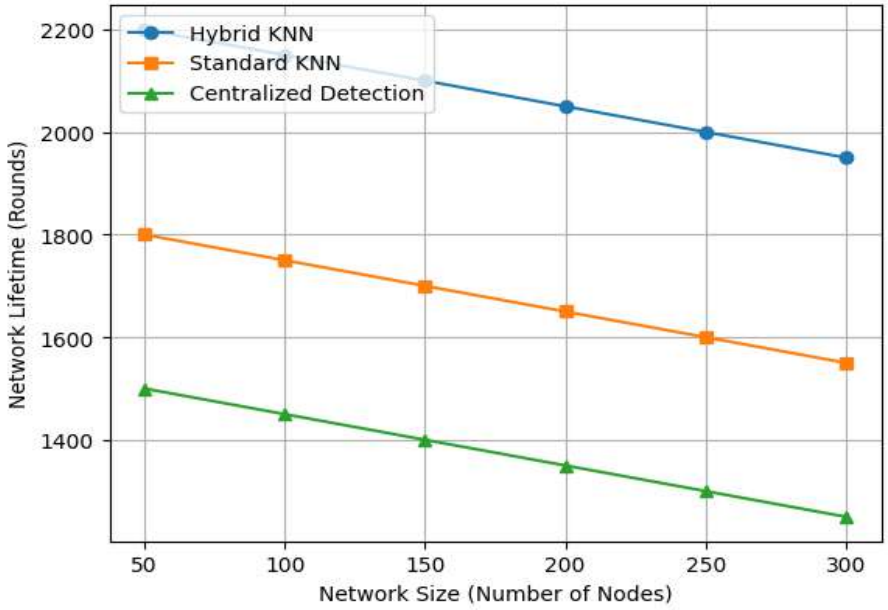

Figure 10. Network Lifetime Comparison

The network lifetime data shows that fault detection systems have their most significant impact according to the results presented in Figure 10 [22], [24]. The framework extends first node and 50% node death metrics compared to centralized methods.

Comparative Analysis

Table 4. Comparative Performance Analysis

Metric	Centralized	Standard KNN	Proposed Hybrid KNN
Detection Accuracy (%)	82.4	88.6	94.2
False Alarm Rate (%)	11.8	7.2	3.9
Avg. Energy Consumption (J)	1.48	1.26	0.94
Network Lifetime (rounds)	1240	1460	1785
Communication Overhead	High	Medium	Low

Impact of K Value

Table 5. Impact of K Value on Detection Accuracy

K	Detection Accuracy (%)	False Alarm Rate (%)
3	92.1	4.5
5	94.2	3.9
7	93.8	4.1

The hybrid architecture improves resource usage which results in better network lifetime performance. The proposed method shows an average energy consumption of 0.94 "J" which represents a 0.32 "J" decrease from the standard KNN value of 1.26 "J" and a 0.54 "J" decrease from the centralized method value of 1.48 "J". The network lifetime extends to 1785 "rounds" as a result

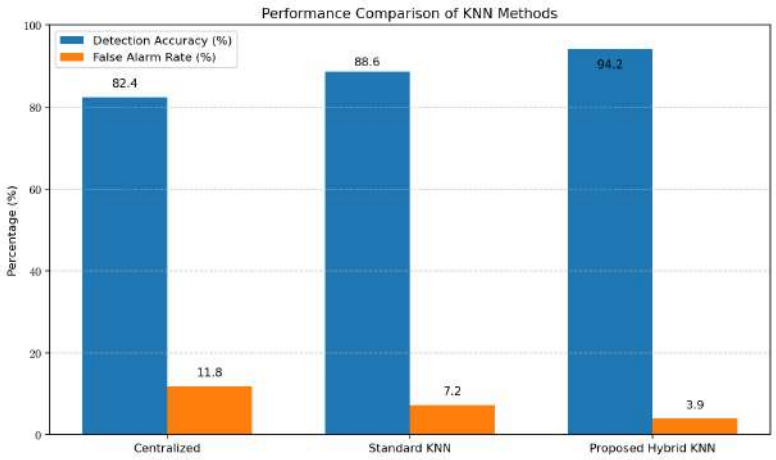

Figure 11: - Performance Comparison of KNN Model

This extends the operational period by 325 rounds over the Standard KNN 1460 rounds and 545 rounds over the Centralized approach 1240 rounds. The Proposed Hybrid KNN achieves a "Low" Communication Overhead, whereas the Standard KNN and Centralized methods result in "Medium" and "High" overheads, respectively. This suggests that the hybrid approach effectively minimizes data transmission requirements, which is consistent with the observed reduction in

energy consumption

The Proposed Hybrid KNN outperforms the other methods by achieving the highest Detection Accuracy 94.2%, the lowest False Alarm Rate 3.9%, and the longest Network Lifetime 1785 "rounds" with minimal Communication Overhead.

7. Conclusion

In this paper, the proposed mythology for a distributed IoT-enabled self-fault detection technique for smart city sensor networks, by using hybrid K-Nearest Neighbours (KNN) algorithm. The IoT-enabled data-driven architecture allows individual sensor nodes to independently analyse local feature for anomaly detection, and collaboratively verify decision through their neighbours. Combining adaptive threshold and neighbour-assisted validation, the hybrid KNN method can achieve better fault detection performance with less false alarms than centralized as well as typical KNN-based methods. Based on simulation results, it can be seen that the network framework which was illustrated to the left gets rid of communication overhead and consumes less power, so it prolongs battery life of networks. Therefore, it is very suitable for large-scale smart city deployment. Future plans include extending the framework to multiple heterogeneous sensor types. In addition, it will enable dynamic real-time adaptive learning for environments that may be inconstant over time.

8. References

1. L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.
2. K. Ashton, "That 'Internet of Things' Thing," *RFID Journal*, 2009.
3. D. Miorandi et al., "Internet of things: Vision, applications and research challenges," *Ad Hoc Networks*, vol. 10, no. 7, pp. 1497–1516, 2012.
4. F. Bonomi et al., "Fog computing and its role in the Internet of Things," in *Proc. MCC Workshop on Mobile Cloud Computing*, 2012.
5. J. Gubbi et al., "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, pp. 1645–1660, 2013.
6. H. Chen et al., "Data-driven fault detection in IoT sensor networks," *IEEE Access*, vol. 7, pp. 45700–45714, 2019.
7. Y. Xu et al., "Distributed anomaly detection for smart cities IoT," *IEEE Sensors Journal*, vol. 19, no. 12, pp. 4725–4735, 2019.
8. T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Info. Theory*, vol. 13, no. 1, pp. 21–27, 1967.
9. S. Yin et al., "Rule-based fault detection for IoT sensor networks," *Sensors*, 2018.
10. M. A. Hossain et al., "Statistical methods for fault detection in wireless sensor networks," *IEEE Commun. Surveys*, 2018.
11. W. Heinzelman et al., "Energy-efficient communication protocols for wireless microsensor networks," *Proc. Hawaii Int. Conf. Syst. Sci.*, 2000.
12. K. Akkaya and M. Younis, "A survey on routing protocols for wireless sensor networks," *Ad Hoc Networks*, 2005.

13. R. Roman et al., *Internet of Things: Architecture and Security*, Springer, 2018.
14. J. Zhang et al., "Distributed fault detection for wireless sensor networks," *IEEE Trans. Automation Science and Engineering*, 2016.
15. L. Li et al., "Collaborative fault detection in IoT sensor networks," *IEEE IoT Journal*, 2017.
16. H. Zhang et al., "Machine learning techniques for IoT fault detection," *IEEE Access*, 2019.
17. S. S. Das et al., "Support vector machine based fault detection in sensor networks," *Sensors*, 2018.
18. B. Han et al., "Hybrid KNN approach for IoT anomaly detection," *IEEE Access*, 2020.
19. R. Liu et al., "Energy-aware distributed anomaly detection in IoT," *IEEE IoT Journal*, 2020.
20. C. Wang et al., "Adaptive fault detection for wireless sensor networks," *IEEE Trans. Industrial Informatics*, 2021.
21. M. A. Hossain et al., "Collaborative KNN for distributed fault detection," *IEEE Sensors Journal*, 2020.
22. B. Gouda et al., "Distributed IoT-based self-fault detection using hybrid KNN," *IEEE Access*, 2023.
23. A. Kumar et al., "Energy-efficient fault diagnosis in smart city IoT networks," *IEEE IoT Journal*, 2022.
24. R. Singh et al., "Simulation-based evaluation of distributed fault detection," *IEEE Sensors Journal*, 2021.
25. L. Chen et al., "Performance metrics for IoT fault detection," *IEEE Trans. Network and Service Management*, 2022.