2023

Distributed Intermittent Fault Diagnosis in Wireless Sensor Network Using Likelihood Ratio Test

Bhabani Sankar Gouda
Biju Patnaik University of Technology (BPUT),Rourkela, Odisha 769015, India

Meenakshi Panda
International Institute of Information Vadodara, Gandhingar, Gujarat 382028, India

Trilochan Panigrahi
National Institute of Technology Goa, Ponda 403401, India

See next page for additional authors

Funder: European Union's Horizon 2020 Research and Enterprise Ireland under the Marie Skłodowska-Curie under Grant 847402.

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 December 2022, accepted 22 December 2022, date of publication 13 January 2023, date of current version 24 January 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3236880

RESEARCH ARTICLE

Distributed Intermittent Fault Diagnosis in Wireless Sensor Network Using Likelihood Ratio Test

BHABANI SANKAR GOUDA[1], MEENAKSHI PANDA[2], TRILOCHAN PANIGRAHI[3], (Senior Member, IEEE), SUDHAKAR DAS[4], (Member, IEEE), BHARGAV APPASANI[5], OMPRAKASH ACHARYA[5], (Member, IEEE), HOSSAM M. ZAWBAA[6,7,8], AND SALAH KAMEL[9]

[1]Department of Computer Science and Engineering, Biju Patnaik University of Technology (BPUT), Rourkela, Odisha 769015, India
[2]International Institute of Information Technology Vadodara, Gandhinagar, Gujarat 382028, India
[3]Department of Electronics and Communication Engineering, National Institute of Technology Goa, Ponda 403401, India
[4]Department of Electronics and Communication Engineering, National Institute of Science and Technology, Berhampur, Odisha 761008, India
[5]School of Electronics Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar 751024, India
[6]Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni-Suef 2722165, Egypt
[7]CeADAR Irelands Center for Applied AI, Technological University Dublin, Dublin, D07 EWV4 Ireland
[8]Applied Science Research Center, Applied Science Private University, Amman 11931, Jordan
[9]Electrical Engineering Department, Faculty of Engineering, Aswan University, Aswan 81542, Egypt

Corresponding authors: Trilochan Panigrahi (tpanigrahi@nitgoa.ac.in), Hossam M. Zawbaa (hossam.zawbaa@gmail.com), and Salah Kamel (skamel@aswu.edu.eg)

The work of Hossam M. Zawbaa was supported by the European Union's Horizon 2020 Research and Enterprise Ireland under the Marie Skłodowska-Curie under Grant 847402.

ABSTRACT In current days, sensor nodes are deployed in hostile environments for various military and commercial applications. Sensor nodes are becoming faulty and having adverse effects in the network if they are not diagnosed and inform the fault status to other nodes. Fault diagnosis is difficult when the nodes behave faulty some times and provide good data at other times. The intermittent disturbances may be random or kind of spikes either in regular or irregular intervals. In literature, the fault diagnosis algorithms are based on statistical methods using repeated testing or machine learning. To avoid more complex and time consuming repeated test processes and computationally complex machine learning methods, we proposed a one shot likelihood ratio test (LRT) here to determine the fault status of the sensor node. The proposed method measures the statistics of the received data over a certain period of time and then compares the likelihood ratio with the threshold value associated with a certain tolerance limit. The simulation results using a real time data set shows that the new method provides better detection accuracy (DA) with minimum false positive rate (FPR) and false alarm rate (FAR) over the modified three sigma test. LRT based hybrid fault diagnosis method detecting the fault status of a sensor node in wireless sensor network (WSN) for real time measured data with 100% DA, 0% FAR and 0% FPR if the probability of the data from faulty node exceeds 25%.

INDEX TERMS Wireless sensor network, intermittent fault, likelihood ratio test, fault diagnosis, distributed algorithm.

I. INTRODUCTION

Wireless sensor networks (WSNs) are comprises of tiny autonomous low cost sensor nodes with computing, processing and wireless communication capability. In the recent past the WSNs have been used in many remote sensing applications for military, civilians and industries [1], [2]. This is because of its capability to collect the data where they deployed, process data locally with limited storing and processing capability and communicate data smartly with the

The associate editor coordinating the review of this manuscript and approving it for publication was Gongbo Zhou.

central processor or known as sink node. The WSNs is an interface between the physical and digital world.

In most of the real time WSNs applications, sensor nodes work together. In general, the sensor nodes are deployed in an unattended, harsh environment such as: deep forest, underwater, volcanoes. Furthermore, the sensor is an electronic device that is vulnerable to many failures [3]. WSNs are susceptible to many failures, which can be differentiated as hardware, software and communication failures. Malfunction in hardware may happen due to a problem in sensing unity, power unity, location unity and processing unit. On the other hand software failure may happen because of problems in sensor programs [4]. Problems in the transceiver, improper estimation of channel states may cause communication failures in the sensor network. Among many faults in WSNs, intermittent faults are more frequent. It is due to low battery, improper channel state information (CSI) estimation, impulsive noise in the channel, loose contacts in hardware etc. Sometimes spikes may occur either randomly or in regular intervals. In this scenario, one should observe the outcomes over a period and find the faulty status of the node [5], [6], [7], [8]. If the faulty nodes are not detected and isolated from other nodes in the network in time, their presence leads to data unreliability, affects network bandwidth because of suspicious data communication, causes route congestion, and at the end reduces a network's lifetime [9], [10]. Thus it motivates the researchers to do the fault diagnosis in WSNs to improve the data reliability, network lifetime and efficient bandwidth utilization.

Several fault detection methods are available in literature based on cooperation strategy among sensor nodes and methodology followed to detect the faulty sensor node in WSNs [7]. Depending upon the cooperation of sensor nodes and where the decision of sensor node's fault status is made, there are three fault detection approaches known as centralized, distributed and hybrid. In a centralized approach a sink node or base station collects data from all the nodes in the network and then diagnoses to detect the faulty node. Sharing all the data to the central node through the multihop Communication paradigm is not energy efficient and has a relatively higher diagnosis latency. In fact, for finding intermittent faults we need to share observed data over a certain period. Thus a centralized approach cannot be applied in large-scale networks for finding intermittent faults. On the other hand, in a distributed approach each sensor node collects the data from the neighbors and then detects the fault [12]. Most of the literature uses a majority voting method to find the fault. Although a distributed method is energy efficient compared to that of a centralized approach, repeated testing is required for detecting intermittent faulty sensor nodes. All the nodes in WSNs may not be equipped for diagnosing the status of all the neighbours including self. Further, detection accuracy of distributed algorithms rely on the performance of neighboring nodes, but over time nodes may become faulty and then have a negative effect on performance.

In order to overcome the limitation of both centralized and distributed approaches, the hybrid approach is proposed here that combines the advantages of both centralized and distributed approaches. Each sensor node now shares the measured data to anchor nodes having better processing capability and energy efficiency. In reality, the sensor nodes are connected to anchor nodes for data transmission and localization [13]. Therefore in the proposed hybrid approach each sensor node shares the data to the connected anchor node or sink node. Then the anchor node follows the fault diagnosis algorithm to detect the faulty status of the node. The hybrid approach bridges the gap between the centralized, and distributed approaches.

In literature, statistical methods and machine learning (ML) approaches including fuzzy logic are used to identify different kinds of faults in WSNs. In statistical based approaches, the mean, variance and robust statistical parameters like mean of absolute deviation (MAD) are computed and then compared with a threshold to identify the status of a node [2]. The residue number system (RNS) is used as an error detection and correction tool in ad-hoc network [14]. In order to diagnose the intermittent faults, these testing processes need to be repeated as many times as data collected during the observation period. Whereas in the ML approach, the data over an observation period is used together to identify its status [5]. Because of the huge computational cost, the ML methods may be used at the base station. Further, we know that the centralized methods are not energy efficient for large WSNs due to more communication overhead. Therefore, a statistical method is proposed here using a hybrid approach and without repeating the test statistics.

We know that intermittent faulty nodes provide very noisy data sometimes over a period. This leads to the change in mean and variance from the statistics of the actual data [6]. In fact a good node also shares the real time data with noise depending upon the sensor used and the channel noise, but with tolerable limits. Now, if most of the data are suspicions and having more noise, then the changes in mean and variance is beyond the tolerance limit.Therefore, in this paper we tried to utilize the statistical test procedure to detect the changes in mean and variance simultaneously [15]. In the changes of statistics of the measured data, if either mean or variance is changed, then also a procedure is introduced under normality [16]. Thus we proposed a fault diagnosis method using likelihood ratio test where the method is capable of detecting the changes in mean and variance or one at a time in the measured data. A hybrid approach is incorporated here where each sensor node shares the measured data with the Anchor node. Then the anchor combined with the actual data, determined the mean and variance that are used to estimate the likelihood ratio. Then it is compared with a threshold value that depends upon the tolerance limit to estimate the faulty status of a sensor node. The highlights of the paper are:

- A hybrid method is proposed to diagnose the intermittent fault status of a sensor node. In this approach, an anchor node accumulates data from the sensor node connected to it and then follows the algorithm for fault diagnosis. The advantages of both centralized and distributed approaches are achieved in this approach.

- A novel simultaneous likelihood ratio test (LRT) statistics based fault diagnosis algorithm that senses the changes in mean and variance of the measure data is developed to detect the intermittent faults (both random and spike faults) in WSNs.
- The proposed algorithm detects the fault status in one shot. Repeated tests are not required as used in the existing fault diagnosis methods.
- The fault diagnosis algorithm is tested over the real time measured data available in literature. The simulation results show that the LRT based fault diagnosis algorithm outperforms the existing robust three sigma edit test [2] with minimum computational cost.

The rest of the paper is organized as follows. An overview of the research work carried out for intermittent fault diagnosis is presented in II. The system model along with the sensor fault model and taxonomy is provided in Section III. The problem formulation and the proposed likelihood ratio test (LRT) with proper threshold is explained in Section IV. The hybrid approach for fault diagnosis algorithm using LRT along with cost analysis is presented in Section V. The analysis of the algorithm is presented in VI. The results and discussion are made in Section VII. Finally, the work is concluded with future direction in Section VIII.

II. RELATED WORK

In this section we presented the existing literature for intermittent fault diagnosis in WSNs along with their advantages and disadvantages.

Intermittent faults of sensor nodes are proposed where the number of faults in a specified period is calculated [17], [18], [19]. The centralized naïve Bayes detector is proposed in [20] to classify sensor nodes by analyzing the end-to-end transmission time collected at the sink. In [21], the authors modeled the network as a graph using an extended algorithm known as a neighborhood hidden conditional random field (NHCRF). The NHCRF method finds a faulty sensor node in WSNs by diagnosing its received signal strength, frequency, and signal delay. These fault detection methods based on probabilistic approach cannot distinguish various faults in WSNs.

In literature RNS is proposed as an error detection and correction tool and has been used in ad-hoc network [14]. The authors used a new system by mixing redundant RNS, multi level RNS and multiple valued logic RNS which supports the system for secure data communication and has the ability of error detection and correction. A centralized transient fault detection algorithm for WSNs based on comparisons between neighboring nodes and their own central node is proposed in [2] and [19]. The RNS system is used to compute the residues and compare with the predefined threshold to detect the fault [14], [19]. The advantage of the RNS based fault detection algorithm is that it is faster but uses a threshold and the accuracy depends on it. Further to detect intermittent fault, repeated testing is needed. To avoid this issue, a one time statistical method is proposed in this paper.

In [22], the authors proposed a fault detection algorithm based on the Debraj De algorithm in long-thin WSNs. A correlation parameter of two nodes is used to detect nodes with faulty readings. This method reduces computational complexity of the correlation algorithm, but we need to repeat the test for intermittent fault nodes in WSNs.

A centralized method based on fuzzy logic and majority voting technique to detect faulty nodes in WSNs is proposed in [9] and [10]. The authors used fuzzy logic to identify the ratio of the faulty nodes in the network and in each sub-network. Detection of faulty nodes is done by using the calculated ratio and majority voting. In [11], Fuzzy rule-based faulty node classification and management scheme to improve the quality of service for large scale WSNs is proposed. The fuzzy based algorithm increases the percentage of detecting faulty nodes with minimal computational complexity leading to minimizing end to end delay and energy consumption in WSN. The problems in fuzzy classification is it requires a better understanding of the network and aprior information about fault type.

These methods are centralized where the fusion center is deciding the fault status of each sensor node in WSNs. The centralized method provides good accuracy as data from all the sensor nodes are available while making decisions, but involves more communication overhead for sharing the data to the fusion center. To minimize communication overhead in the network, a distributed approach is reported in [2] and [22] where each sensor node is capable of detecting their fault status by cooperating with the neighbours. In this way, it reduces traffic overhead, improves the robustness, and minimizes the end-to-end delay over the network [23]. Although a distributed method is energy efficient, this method has a limitation of not having all the network data while making decisions.

Because of energy inefficiency and higher diagnosis latency a centralized approach cannot be applied in large-scale WSNs. Statistical methods are used to detect the intermittent data faults in the sensor network [7], [24]. Sensor nodes are diagnosed for a specific period. Once the data is collected from its neighbours, the mean, variance and higher order statistics are computed and then a z-test or modified three sigma edit test is followed for self diagnosis. Each data is then decided either suspicious or not. Over a period if more data (usually more than 50%) of a sensor node is suspicious, the node is considered as faulty [25]. The authors in [26] used a hypothesis testing method to find the fault status of a node. In these self diagnosis methods, each sensor node is required to compare its observed data with the neighbors. Fault status of the sensor node is determined by the neighbors and final decision is taken based on majority voting. These algorithms are good in the early stage of deployment because most of the sensor nodes are fault free. The detection performance will degrade with time progress. It is because a self-diagnosing approach depends on the data from neighboring nodes. Further, all these methods are tested by using synthesized data. When real time data sets are used, there may be a wrong decision due to masking effect and that has been shown in our simulation results.

A fault diagnosis protocol using majority neighbors coordination based approach for WSNs is proposed in [27]. To detect intermittent fault the mean difference and standard error comparison with predefined threshold are used. A statistical method to classify the faults in power transmission lines [28]. The authors used a decision tree method and the measured parameters are the root mean square current duration, voltage dip, and discrete wavelet transform. A novel distributed fault detection algorithm for WSNs in smart grid based on credibility and cooperation is presented in [12]. Based on temporal data correlation, a credibility model of a sensor is established first. In this approach all the sensor nodes are not going for fault diagnosis. The suspicious sensor is then chosen as per the credibility model and then requested for fault diagnosis. This approach improves the efficiency of fault detection and uses neighbor cooperation. The problem in majority voting is when the majority of nodes in the neighbourhood are faulty (sometimes due to intentional attack), all may be detected as fault free and will have an impact on the performance of the algorithm. Further this testing to be repeated for finding intermittent faulty sensor nodes in WSNs.

To overcome this disadvantage, neural networks and machine learning algorithms are used for fault diagnosis. Composite types of fault are diagnosed in WSN using neural networks [29], [30]. In recent years fault diagnosis based on deep learning has been proposed because of its strong feature representation capability. Machine learning is used for detecting drift fault of sensors in cyber-physical systems which may not be suitable for WSN scenario [31]. A Naive Bayes classifier and convolution neural network (CNN) are used to classify the faults in distributed WSN. These deep learning methods are used to improve the convergence performance over other neural networks [32]. Support vector machine (SVM) has been used to classify the faulty sensor node in WSN [5]. Fault diagnosis based on extremely randomized trees in wireless sensor networks is proposed in [33] and [34]. The authors have compared the results with other grid search approaches such as multilayer perceptron, SVM, decision tree and random forest. Although the accuracy of detection is improved, the computational complexity is also increased. We are looking for a method that improves the detection accuracy when less data is corrupted with less computational complexity. The method should not depend on more data that is more diagnostic period.

Sensor nodes in WSNs often operate under adverse conditions and the probability of different kinds of faults is also different. This leads to the extreme class imbalance and long-tailed distribution between different faults [35]. In literature, imbalance learning is used for fault [36] diagnosis. Numerous researches have been carried out and it is observed that the imbalanced classification is a challenging issue in the fault diagnosis [37]. A deep convolutional Generative adversarial networks (DCGAN) model is presented in [37] for simulating the original distribution from minority classes and generating new data to solve the imbalance problem. Recently, the extreme learning machine (ELM) has attracted much attention because of faster speed and better generalization performance [38], [39]. Whereas, ELM rarely involves strategies for imbalanced data distributions in a sensor network with few nodes being faulty. A class-specific cost regulation extreme learning machine (CCR-ELM), together with its kernel based extension, for binary and multiclass classification problems with imbalanced data distributions is proposed in [39]. In order to alleviate the learning bias towards the majority class and to have better classification performance, a weighted extreme learning machine (W-ELM) is used. The difficulty in W-ELM is that we may not have the optimal weights for the samples from different classes. To overcome this issue, multi-objective optimization-based adaptive class-specific cost extreme learning machine (MOAC-ELM) is the best alternative to W-ELM [38].

An effective deep reinforcement learning for faulty node detection and recovery schemes in large WSNs is given in [4]. The method is integrated with the hosted cuckoo-based optimal routing scheme and enhances the network life with least energy utilization. The experimental results are compared with recent optimization methods used in this purpose such as harmony search algorithm, optimal emperor penguin optimization and particle swarm optimization algorithm.

A transmission rate control method based on its traffic loading information and Multiclassification using an effective data science technique, namely support vector machine (SVM) is discussed in [40]. Swarm intelligence algorithms like differential evolution (DE), genetic algorithm (GA) and grey wolf optimization (GWO) are used to tune the SVM in order to get less misclassification error. Enhanced random forest based approach has detected the faults with better accuracy and outperformed over the other classifiers in fault detection. These methods are centralized based as it requires an efficient processor.

In literature, hybrid approaches originated due to major disadvantages in centralized and distributed approaches. Extra equipment, such as an anchor node, needs to be added in the hybrid approach to achieve diagnosis reliability, robustness, energy efficiency, and minimization of traffic overhead. A mobile sink-based distributed fault detection scheme, which identifies the health status of each software and hardware component separately is proposed in [41]. In fact the sensor nodes near the base station die earlier than those far away. To deal with this a mobile sink-based adaptive immune energy-efficient clustering protocol by using a controlled mobile sink is proposed in [42]. The main problem in these methods lies in the path planning of mobile sink nodes that is related to the algorithm's performance.

It is observed that in literature the statistical methods have been used where mean, variance, z-score, robust median and MAD are computed and compared with a threshold to find the faulty node in WSNs. Majority voting and repeated testing are major criteria for intermittent fault diagnosis. Centralized ML methods are introduced to classify the faulty nodes in WSNs which need more computational complexity. Therefore in this paper we proposed a simple simultaneous likelihood ratio statistics (that can detect change in mean and variance or one of them when fault occurs) method to identify the intermittent fault in one shot with minimum computational complexity. The proposed method follows a hybrid approach which has the advantage of both centralized and distributed algorithms.

III. SYSTEM MODEL

Here the model for the sensor network and the topology of the network along with communication methods between the sensor nodes is explained. The fault model where the behaviour of the various kinds of faulty sensor nodes is described.

A. ASSUMPTIONS

In order to develop sensor network, without losing a generality, the following assumption are taken [25]

1) Sensor nodes are of homogeneous nature and have equal energy.
2) In the sensor network, the sensor nodes are capable of communicating with themselves.
3) If any node fails to communicate refereed as cut or hard fault.
4) If any error occurs in the communication link, then the MAC layer will take care of that.
5) The UDP/IP communication protocol is used by the sensor nodes for communication with others.

B. SENSOR NETWORK MODEL

Let us consider N number of sensor nodes deployed randomly in a square terrain of area RXR. In a network, every sensor nodes s_i, $i = 1, 2, 3 \ldots, N$ have known unique identifier. The position of s_i is $P_i(xc_i, yc_i)$, where $0 \leq xc_i \leq R$, $0 \leq yc_i \leq R$. In the sensor network, sensor node s_i interacts with the immediate neighbours in their basic transmission mode. Let us define T_r be the transmission range of every sensor node. Uniform energy and transmission range are the common assumption in homogeneous sensor networks. Any sensor node lies within the transmission range T_r of s_i is assumed to be connected. All the connected sensor nodes are called the neighbours of it. In general a sensor network may be considered as a graph $G(S, C)$ where S is the set of sensors and C represent the set of communication links among the sensors.

In wireless sensor networks, each sensor node is connected through a wireless link. Every sensor node sends as well as receives messages from the immediate neighbouring nodes within a bounded time period in synchronous WSNs. The IEEE 802.15.4 MAC layer protocol used here for data communication of sensor nodes among themselves using one-to-one primitives.

C. SENSOR FAULT MODEL AND TAXONOMY

In this paper, we are diagnosing different types of intermittent faults such as spike, or random intermittent faults. These faults are either due to the noise while collecting data or due to the malfunction of the sensor system. These faults in WSN are classified into two main categories such as time-span and location based faults [6]. In the timespan-based faults, there are transient and persistent type faults based on the period of the fault. At first, transient faults are temporary and occur for a short period due to weather conditions, network congestion, etc. On the other hand, persistent faults are permanent and exist until recovery is carried out. The entire WSN is not normally defective.

For a fault free node, the measurement of a nth sensor at ith time is $x_n(i)$ modeled as

$$x_n(i) = d_n(i) + v_n(i), \quad n = 1, 2, 3, \ldots, N \tag{1}$$

where $d_n(i)$ is actual value of the nth sensor at ith instant of time and $v_n(i)$ is the error or called the noise due to sensor calibration, noise added during transmission or because of other malfunction in the node and network. But it is assumed that the error is within the tolerance limit, then only we can say that the measurement is fault free. $v_n(i)$ is normally distributed zero mean random and variance σ^2. Now, when a node becomes faulty, the sensor output is given as [5], [6]

$$x_n^{fault}(i) = b + gd_n(i) + v_n(i) \tag{2}$$

where b is the additive bias (which is a constant) and called offset, g is the multiplicative constant called gain. Two different kinds of intermittent faults such as random or spike are defined here based on different values of b, g and the noise variance. We define the data for a faulty or fault free node over a period of T and measure in the interval of δT. Thus the number of discrete measurements is $K = \lceil \frac{T}{\delta T} \rceil$. Now the different types of intermittent fault are defined as follows.

1) RANDOM/INTERMITTENT FAULT

The intermittent faulty sensor nodes provide arbitrary data for some time duration and behave as a good node for the remaining time. In order to model the arbitrary behavior of the intermittently faulty sensor nodes, the Bernoulli distribution function is used. It is assumed that a sensor node gives faulty reading at time i with probability α and fault free reading with probability $1 - \alpha$. To formalize the idea of approximate normality, one can imagine that a proportion $1 - \alpha$ of the observations is generated by the normal model, while a proportion α is generated by an unknown mechanism [25].

This may be described by supposing

$$F = (1 - \alpha)G + \alpha H \tag{3}$$

where, $G = \mathcal{N}(A, \sigma^2)$ and H may be any distribution; for instance, another normal with a larger variance [30]. In general, F is called a mixture of G and H, and is called a normal mixture when both G and H are normal. If G and H have probability density functions g and h, respectively, then F has probability density function defined as

$$f = (1 - \alpha)g + \alpha h \tag{4}$$

Therefore, the data model of an intermittent faulty sensor node is given as

$$x_n^{Intermittent}(i) = [d_n(i) + v_n(i)] + b_i v_n^{imp}(i) \tag{5}$$

where, $v_n(i)$ and $v_n^{imp}(i)$ are independent zero mean Gaussian random variable with variances σ^2 and σ^2_{imp}, respectively; b_i is a switch sequence of ones and zeros and is modeled as an independent and identically distributed Bernoulli random process with occurrence probability $P_r(b_i = 1) = \alpha$ and $P_r(b_i = 0) = 1 - \alpha$ [43]. The value of α is in between 0 to 1. The variance of $v_n^{imp}(i)$ is chosen to be very large than that of $v_n(i)$ so that when $b_i = 1$, a large error is experienced in $x_i(n)$. The b_i is given in (6).

$$b_i = \begin{cases} 1, & r \geq \alpha \\ 0, & r < \alpha \end{cases} \tag{6}$$

where, r is a random variable in between 0 and 1 having a uniform probability density function (pdf).

2) SPIKE FAULT

In sensor node, this kind of fault is observed intermittently in the form of high-amplitude spikes. The probability of the occurrence of the spike fault is p_s. Randomly with the defined probability, a constant b_s is added to the ith element of the non-faulty signal to obtain spike fault. It is modeled as [5]:

$$x_n^{Spike}(i) = d_n(i) + b_i b_s + v_n(i) \tag{7}$$

where $b_i = 1$ with probability p_s and 0 with probability $1 - p_s$. Therefore, b_i is binary random variable with Bernoulli distribution defined before.

IV. FAULT DETECTION USING LIKELIHOOD TEST

Let us consider a sensor network with N sensor nodes distributed in an unattended geographical area for specific application. The sensor nodes are having one or more sensors to measure the physical environment such as humidity and temperature. The sensor nodes are collecting data in the regular interval of time ΔT for time duration T known as diagnosis period. The random erroneous data are temporally and spatially independent and have the same distribution function at each node. It follows that the observation $x_n(1), x_n(2), \ldots, x_n(K)$ are independent with common distribution function and can say that the $x_n(i)$'s are *i.i.d.* i.e *independent and identically distributed*. A conventional way to represent "well-behaved" data, i.e. data from a fault free sensor node, is assumed to be normal distribution with mean A and variance σ^2 which implies $\mathcal{N}(A, \sigma^2)$ [25].

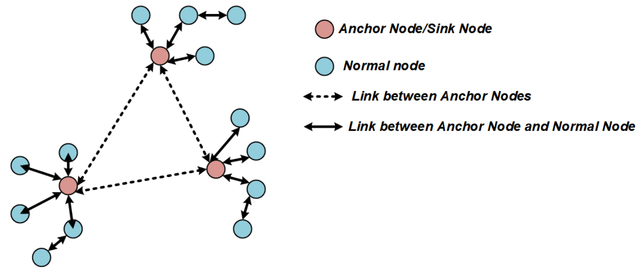


FIGURE 1. Network model where nodes are connected with anchor node.

Now each sensor node is to be diagnosed whether it is faulty or not. The node has measured data over a certain period. Let the measured data by nth node is represented as $\mathbf{Y}_n$. The diagnosis can be done at each node if nodes are well equipped with sufficient memory and a good processing unit. As the technology is advanced and also sensor technology, nodes are capable of that, but price may be high. Otherwise, the best way to diagnose the sensor network is at the sink node. Anchor node is the intermediate node to send the information to the fusion center in a network [45]. Since the anchor/sink node has higher energy, memory and processing capability than the normal node and is also connected to neighbors, it is the better option for diagnosing the network locally. The network model is given in Fig. 1.

Once the anchor node has the measured data from the connected neighbor node, it tests the faulty status by following the Likelihood Ratio Test. The details of LRT are given in the following section.

A. LIKELIHOOD RATIO (LR) STATISTICS FOR THE SIMULTANEOUS TEST

Let's we have K independent measurements of a nth sensor represent as

$$\mathbf{X}_n = \{x_n(1), x_n(2), \ldots, x_n(K)\}$$

. Now fault is introduced as per the model defined above and can have the same K number of output data. The faulty output of the nth sensor is

$$\mathbf{Y}_n = \{y_n(1), y_n(2), \ldots, y_n(K)\}$$

Now, let us assume that the mean and variance of the above population by considering the distribution is Gaussian as $\mathcal{N}(\mu, \sigma^2)$ and $\mathcal{N}(\mu_1, \sigma_1^2)$ for fault free and faulty data respectively.

Now we are going to apply the likelihood ratio (LR) principle for testing the hypothesis H_0 and H_1. For the we need to compute the following statistics as

$$\bar{X}_n = \frac{1}{K} \sum_{k=1}^{K} x_n(k) \tag{8}$$

$$\bar{Y}_n = \frac{1}{K} \sum_{k=1}^{K} y_n(k) \tag{9}$$

$$\sigma^2_{x_n} = \frac{1}{K} \sum_{k=1}^{K} (x_n(k) - \bar{X}_n)^2 \tag{10}$$

$$\sigma_{y_n}^2 = \frac{1}{K}\sum_{k=1}^{K}(y_n(k) - \bar{X}_n)^2 \tag{11}$$

Now combine the data $\mathbf{X}_n$ and $\mathbf{Y}_n$ as $\mathbf{Z}_n = [\mathbf{X}_n\mathbf{Y}_n]$, and the compute the mean and variance as

$$\bar{Z}_n = \frac{\bar{X}_n + \bar{Y}_n}{2} \tag{12}$$

$$\sigma_{z_n}^2 = \frac{1}{2K}\sum_{k=1}^{2K}(z_n(k) - \bar{Z}_n)^2 \tag{13}$$

Now $\bar{X}_n$, $\bar{Y}_n$, $\sigma_{x_n}^2$ and $\sigma_{y_n}^2$ are LR statistics under $H_0 \cup H_1$. Further $\bar{Z}_n$ and $\sigma_{z_n}^2$ are also LR under H_0. Now the LR statistics $LR(\mathbf{X}_n, \mathbf{Y}_n; \mu, \mu_1, \sigma^2, \sigma_1^2)$ to test

$$H_0 : (\bar{X}_n = \bar{Y}_n) \cap (\sigma_{x_n}^2 = \sigma_{y_n}^2) \tag{14}$$

$$H_1 : (\bar{X}_n \neq \bar{Y}_n) \cup (\sigma_{x_n}^2 \neq \sigma_{y_n}^2) \tag{15}$$

in the following way [15], [46]

$$LR_n(\mathbf{X}_n, \mathbf{Y}_n; \mu, \mu_1; \sigma^2, \sigma_1^2) = \frac{\sup\left\{f\left(\mathbf{X}_n, \mathbf{Y}_n; \mu, \mu_1; \sigma^2, \sigma_1^2\right) : H_0 \cup H_1\right\}}{\sup\left\{f\left(\mathbf{X}_n, \mathbf{Y}_n; \mu, \mu_1; \sigma^2, \sigma_1^2\right) : H_0\right\}} \tag{16}$$

The $LR_n(\mathbf{X}_n, \mathbf{Y}_n; \mu, \mu_1; \sigma^2, \sigma_1^2)$ expressed in therms of the statistics of the actual data and the combined data with the measured as

$$LR_n = \frac{\left(2\pi\sigma_{x_n}^2\right)^{-K/2}\left(2\pi\sigma_{y_n}^2\right)^{-K/2}\exp(K)}{\left(2\pi\sigma_{z_n}^2\right)^{-K}\exp(K)} = \frac{\left(\sigma_{z_n}^2\right)^{K}}{\left(\sigma_{x_n}^2\right)^{K/2}\left(\sigma_{y_n}^2\right)^{K/2}} \tag{17}$$

As per the fact, one may reject H_0 in favor of H_1 for some large value of LR in (17) in the view of the likelihood principle. Taking logarithm on both side of the above equation, we get

$$\log(LR_n) = \frac{K}{2}\left[\log\left(\frac{\sigma_{z_n}^2}{\sigma_{x_n}^2}\right) + \log\left(\frac{\sigma_{z_n}^2}{\sigma_{y_n}^2}\right)\right] \tag{18}$$

Now, the logarithm of the LR_n is to be compared with the threshold value in order to decide either the node n is faulty or faulty free.

B. DEFINING THRESHOLD

Let us define the variables S_x and S_y as

$$S_x = \frac{\sigma_{z_n}^2}{\sigma_{x_n}^2}$$

and

$$S_y = \frac{\sigma_{z_n}^2}{\sigma_{y_n}^2}$$

Now the Eq. (18) is defined as

$$\log(LR_n) = \frac{K}{2}\left[\log(S_x) + \log(S_y)\right] = \frac{K}{2}\left[\log(S_xS_y)\right] \tag{19}$$

Now let us assume the ratios defined are nearly equal, that is $S_x \approx S_y$ when the measured data is very close to the actual data when the sensor node is fault free. Then the above equation (19) is written as

$$\log(LR_n) \approx K\left[\log(S_x)\right] \tag{20}$$

Now, let the threshold λ_{th} is used to compare the LR_n and the following hypothesis are defined

$$H_0 : K\left[\log(S_x)\right] < \lambda_{th}$$
$$H_1 : K\left[\log(S_x)\right] \geq \lambda_{th} \tag{21}$$

In fact, when the measured data set is exactly the same as that of the actual, the ratio S is one and the threshold will be $\lambda_{th} = 0$. But the measurement is always noisy with acceptable levels. It depends upon the environment, the sensor is calibrated (usually more costly) or uncalibrated, the channel estimation algorithm, accordingly the tolerance limit is defined. Let's define the tolerance in terms of change in the variance ratio $S_x = 1 \pm p$, where p is the percentage of increase or decrease in the S_x and $p < 1$.

Therefore, the λ_{th} is defined as

$$\lambda_{th} = \frac{K}{2}[\log(1+p) + |\log(1-p)|] \tag{22}$$

where $|\bullet|$ denotes the absolute value. The threshold depends upon two parameters known as the number of data used for diagnosis K and the tolerance limit in percentage p. The impact of these parameters in the performance will be analyzed in the result section.

V. HYBRID FAULT DIAGNOSIS ALGORITHM USING LR TEST

In this section, the distributed fault diagnosis method based on the statistical likelihood ratio test as discussed in the above section will be explained. Each sensor node shares the measured data over a certain duration T with a time interval of ΔT to the cluster head. In this way the number of data to be shared by each sensor node to the cluster center is $K = \frac{T}{\Delta T}$. Let the data is denoted as $\mathbf{Y}_n$. The anchor computes mean and variance and they are denoted as $\bar{\mathbf{Y}}_n$ and $\sigma_{y_n}^2$ respectively as per the formula defined in (11).

Let us assume that the anchor node has the actual measured data without fault, that is $\mathbf{X}_n$ from the nth node and the statistics of that data too. Let the mean and variance are $\bar{\mathbf{X}}_n$ and $\sigma_{x_n}^2$ respectively. Now the anchor node compares the mean and variance of the actual data with the mean and variance of the measured data in the diagnosis period by following the likelihood ratio test as discussed in Section IV-A. Before going to test the status of the node whether it is faulty or not, first fix the tolerance limit of the deviation of measured data from the actual data without fault. That probability of tolerance P and the number of data used to diagnose the node are used to compute the threshold λ_{th} as defined in (22).

Now the anchor node computes the likelihood ratio as given in (18). If the absolute likelihood ratio is less than the predefined threshold, then the fault status of the node is considered as fault free and we make $FS_j(n) = 0$, where FS_j contains the fault status of the jth anchor node. On the other hand, if the likelihood ratio is more than the λ_{th}, the node is determined as faulty and $FS_j(n) = 1$.

In this way a sink/anchor node diagnoses the faulty status of each node connected to it. Similarly, all the anchor nodes now can find the faulty status of all the nodes present in a sensor network that is provided in the set FS that is the union of all the fault status determined and stored in FS_j. The detailed algorithm is given in Algorithm 1

Algorithm 1 Distributed Fault Diagnosis Algorithm Using LR Test

Data: Observed time period T, sensed data $\mathbf{Y}_n = \{y_n(i)\}_{i=1}^K$, actual data $\mathbf{X}_n = \{x_n(i)\}_{i=1}^K$ at each node, allowed tolerance limit p
Result: Calculate threshold λ_{th}, Fault status of sensor nodes FS

```
Determine threshold
λ_th = (K/2)[log(1 + p) + |log(1 − p)|]
for j = 1 ··· M do
    for l = 1 to|N_j| do
        A_j collects the measured data Y_n from each of
          the node n ∈ N_j.
        Compute mean Ȳ_n and σ²_{y_n}
        if Ȳ_n == X̄_n and σ²_{y_n} == σ²_{x_n} then
            FS_j(n) = 0
        else
            Combine actual and measure data as
              Z_n = {X_n Y_n}
            Measure the mean Z̄_n as in (12)
            Measure the variance σ²_{z_n} as in (13)
            Measure the LR(n) as defined in (18)
            Then the fault status are defined as per
              Hypothesis defined in (21)
        end
        if |LR(n)| ≥ λ_th then
            FS_j(n) = 1
        else
            FS_j(n) = 0
        end
    end
    end
    Each anchor node provides the fault status to its
      connected node.
    Combine all the fault status of the nodes of all
      anchor node neighbors
    FS = ⋃_{j=1}^{M} FS_j
end
```

A. COST ANALYSIS

In this subsection, the computational complexity of the proposed algorithm is discussed and then compared with existing modified three sigma edit test (Mod3Sigma) [25]. The proposed likelihood ratio statistics for the simultaneous test to detect the fault status needs to estimate the mean and variance of the individual data X_n and Y_n first. The mean and sample variance statistics of the concatenated data Z_n is to be determined. The required computational complexity for each node to calculate mean and variance is $\mathcal{O}(2K)$. On the other hand the Mod3Sigma algorithm needs to compute the median of data accumulated from the neighbouring nodes twice for testing. The best computational complexity for computing the median is $\mathcal{O}(d \log d)$, where d is the average degree of the nodes in a sensor network. Since the robust threshold value needs to be computed for each observation, the total complexity will be $\mathcal{O}(Kd \log d)$. Therefore, the proposed algorithm has less computational cost compared to that of existing methods. On the other hand, the machine learning algorithms required much more computational complexity for their learning.

VI. MATHEMATICAL ANALYSIS OF THE ALGORITHM

In this section a mathematical analysis is presented to justify why likelihood ratio statistics for the simultaneous test is providing a good detection performance. The null-hypothesis given in (14) has used union-intersection procedure. In order to investigate the LR statistic given in (17) is to be decomposed to two components. The LR_n is rewritten as

$$LR_n = \left(\frac{\sigma_{z_n}^2}{\sigma_{x_n}^2}\right)^{K/2} \left(\frac{\sigma_{z_n}^2}{\sigma_{y_n}^2}\right)^{K/2} \tag{23}$$

Now, let us define the term used to find variance of the data $x_n(k)$ with respect to the new mean $\bar{Z}$ as

$$\sum_{k=1}^{K}(x_n(k) - \bar{Z}_n)^2 = \sum_{k=1}^{K}(x_n(k) - \bar{X})^2 + K(\bar{X}_n - \bar{Z}_n)^2 \tag{24}$$

After using (12), above equation is rewritten as

$$\sum_{k=1}^{K}(x_n(k) - \bar{Z}_n)^2 = \sum_{k=1}^{K}(x_n(k) - \bar{X}_n)^2 + \frac{K}{4}(\bar{X} - \bar{Y}_n)^2 \tag{25}$$

In similar way, the term used to find variance of the data $y_n(k)$ with respect to $\bar{Z}$ is given as

$$\sum_{k=1}^{K}(y_n(k) - \bar{Z}_n)^2 = \sum_{k=1}^{K}(y_n(k) - \bar{X}_n)^2 + \frac{K}{4}(\bar{X}_n - \bar{Y}_n)^2 \tag{26}$$

Now, the variance of $z_n(k)$ can be expressed by using equations (25) and (26) in (13) as

$$\sigma_{z_n}^2 = \frac{1}{2K}\sum_{k=1}^{K}(x_n(k)-\bar{X}_n)^2 + \frac{1}{2K}\sum_{k=1}^{K}(y_n(k)-\bar{X}_n)^2 + \frac{1}{4}(\bar{X}_n-\bar{Y}_n)^2 \tag{27}$$

Let us use the definition of $\sigma_{x_n}^2$ and $\sigma_{y_n}^2$ in (27), we get

$$\sigma_{z_n}^2 = \frac{1}{2}\left\{\sigma_{x_n}^2+\sigma_{y_n}^2+\frac{1}{2}(\bar{X}_n-\bar{Y}_n)^2\right\} \tag{28}$$

The first product term in the (23) can be computed as

$$\left(\frac{\sigma_{z_n}^2}{\sigma_{x_n}^2}\right)^{\frac{K}{2}} = \left(\frac{1}{2}\right)^{\frac{K}{2}}\left\{1+\frac{\sigma_{y_n}^2}{\sigma_{x_n}^2}\frac{1}{2}\frac{(\bar{X}_n-\bar{Y}_n)^2}{\sigma_{x_n}^2}\right\}^{\frac{K}{2}} \propto \left(1+\frac{\sigma_{y_n}^2}{\sigma_{x_n}^2}\right)^{\frac{K}{2}}\left\{1+\frac{1}{2}\frac{(\bar{X}_n-\bar{Y}_n)^2}{\sigma_{x_n}^2+\sigma_{x_n}^2}\right\}^{\frac{K}{2}} \tag{29}$$

Similarly, the second product term in the (23) is written as

$$\left(\frac{\sigma_{z_n}^2}{\sigma_{y_n}^2}\right)^{\frac{K}{2}} \propto \left(1+\frac{\sigma_{x_n}^2}{\sigma_{y_n}^2}\right)^{\frac{K}{2}}\left\{1+\frac{1}{2}\frac{(\bar{X}_n-\bar{Y}_n)^2}{\sigma_{x_n}^2+\sigma_{x_n}^2}\right\}^{\frac{K}{2}} \tag{30}$$

Now, the equation (23) is

$$LR_n \propto \left(1+\frac{\sigma_{y_n}^2}{\sigma_{x_n}^2}\right)^{\frac{K}{2}}\left(1+\frac{\sigma_{x_n}^2}{\sigma_{y_n}^2}\right)^{\frac{K}{2}}\left\{1+\frac{1}{2}\frac{(\bar{X}_n-\bar{Y}_n)^2}{\sigma_{x_n}^2+\sigma_{y_n}^2}\right\}^{K} \tag{31}$$

The LR value in simultaneous testing in above equation (VI) contains two terms such as the ratio between the variances $\frac{\sigma_{x_n}^2}{\sigma_{y_n}^2}$ (and its reciprocal) and the differences between their means $(\bar{X}_n-\bar{Y}_n)^2$. Let us denote these variables as

$$S = \frac{\sigma_{y_n}^2}{\sigma_{x_n}^2} \tag{32}$$

and

$$M = \frac{(\bar{X}_n-\bar{Y}_n)^2}{\sigma_{x_n}^2+\sigma_{y_n}^2} \tag{33}$$

Both the random variables are $S > 0$ and $M > 0$ and considered as joint likelihood ratio statistics. Now, using these statistically dependent variables S, and M defined in above equations (32) and (33), equation can be rewritten as

$$LR_n \propto \frac{(1+S)^K}{S^{\frac{K}{2}}}\left(1+\frac{M}{2}\right)^K = f(S)g(M) \tag{34}$$

where,

$$f(S) = \frac{(1+S)^K}{S^{\frac{K}{2}}} \tag{35}$$

and

$$g(M) = \left(1+\frac{M}{2}\right)^K \tag{36}$$

Thus, the LR_n is the product of two functions, $f(S)$ and $g(M)$ and each of which function as a random variable S and M respectively. The maximization of the simultaneous LR_n or its equivalent in (34) can be obtained by maximizing $f(S)$ and $g(M)$ independently. Now consider the Roy's union-intersection (UI) test [44] based on $f(S)$ and $g(M)$ or equivalently S and M.

Let us use the variable M for the likelihood ratio test for testing the sub-null hypothesis $H_{01} : \bar{X}_n = \bar{Y}_n$ against $H_{11} : \bar{X}_n \neq \bar{Y}_n$. The LR testing rule may reject $H_{01} : \bar{X}_n = \bar{Y}_n$ for some large value of M. It is because the function $g(M)$ defined in (36) is a strictly increasing function in M. As we know that M follows F-distribution with 1 and $2(K-1)$ d.f. and usually considered K is large. Therefore, the probability of getting a large value is very less. It means the probability of rejecting H_{01} when the mean of the observed data is nearly the same as that of actual measurement.

Similarly, we can use the variance ratio S for testing the sub-null hypothesis $H_{02} : \sigma_{x_n}^2 = \sigma_{y_n}^2$ against $H_{12} : \sigma_{x_n}^2 \neq \sigma_{y_n}^2$. It can be shown that the function $f(S)$ defined in (35) is achieving its minimum at $S = 1$. This can be shown by differentiating the function $f(S)$ with respect to S. Therefore, it is observed that the function $f(S)$ attains its maximum value when S is closer to zero or larger than 1. The LR testing rule may reject $H_{02} : \sigma_{x_n}^2 = \sigma_{y_n}^2$ for the value either in between 0 to 1 or greater than 1 [44]. When the variance of the observed data X_n is nearly same to that of actual measurement Y_N, then the variance ratio is nearly. This leads to the point that the probability of rejecting H_{01} against H_{02} is less as the function is having minima at 1. In fact when intermittent fault occurs the mean or variance or both of the measured data deviate from the actual values. Then at least one of the functions $f(S)$ or $g(M)$ will be more dependent upon which statistics deviate more. Therefore Roy's union-intersection test ensures that the detecting H_1 against H_0 is more probable.

VII. RESULTS AND DISCUSSIONS

In this section the evaluation of our proposed fault detection technique by using LR test is described. Description of real time data used is given first [47]. Next, the simulation results of the new method are presented, analyzed and compared with other techniques in the literature like three sigma edit test.

Three performance parameters are used to analyze the fault diagnosis algorithms. These are [25].

- Detection accuracy (DA): The DA is defined as the ratio between the number of faulty sensor nodes detected as faulty to the total number of faulty nodes in the network.
- False alarm rate (FAR): The FAR is defined as the ratio between the number of fault free sensors detected as faulty and the total number of fault free sensors in the network.

- False positive rate (FPR): The FPR is defined as the ratio of number of faulty sensors detected as faulty free to total number of fault free sensors in the network.

A. DATA SET

The labeled data set used for fault diagnosis is based on an existing dataset published in [47]. The researchers collected the measured humidity and temperature data from a simple single-hop and a multi-hop WSNs by using TelosB motes. They measure data in every five seconds during six hours by introducing steam of hot water to increase the humidity and the temperature. The labeled WSNs data set that we have prepared consists of a set of sensor measurements where we have injected different types of faults. We have used both single hop and multihop measured data. Different faults are introduced in the normal measurements. The impact of the faults is introduced by changing the fault probability, additive noise with varying noise variance, additive bias and gain parameters. We have generated data for 512 sensor nodes and 40% sensor nodes are introduced as intermittent faulty nodes.

B. DETECTION OF INTERMITTENT/SPIKE FAULTY NODE

It is one of the difficult tasks in diagnosing the intermittent fault of the sensor node. It is because the nodes behave well sometimes and have suspicious behavior at other times. Sometimes the suspicious data is completely random and different from the normal distribution. Other kinds of intermittent behavior like a spike either occur in regular intervals or randomly occur with the same constant value. In our simulation study, we consider the random location of suspicious data in both scenarios.

Let the probability of occurrence of suspicious data is p_r. In fact, if the p_r is high, then it will be easy to detect the faulty status of the node using any conventional techniques. It is challenging when it is very less. In literature, most of the authors have diagnosed the intermittent behaviors when the p_r is more than 0.5. But in this paper we analyzed the performance of the proposed fault diagnosis algorithm for p_r is from 0.05 to 0.50. The DA and FAR performance is provided in Fig. 2 when the tolerance of the change in data from the original is ± 0.05. It is observed that the new methodology detects the fault status of the faulty node correctly, that is DA = 1 when the p_r is greater than 0.25. This is for both the scenarios, either the suspicious data is random or a fixed bias known as spike. In literature, as per our knowledge, no fault diagnosis algorithm is providing this detection accuracy when p_r is 0.25. The better detection accuracy is When fault occurs the position and spread of the measured data varies. The simultaneous mean and variance comparison between measured and the actual data provides better accuracy.

In literature, the performances are measured by varying the percentage of faulty sensor nodes in a network. It is because, in a distributed approach a sensor node diagnoses its fault status by accumulating data from the neighbours. When more neihbours are faulty, then the ability of a node to identify its own status is reduced and that leads to degradation in the detection performance. Whereas in the proposed work, a completely hybrid approach is adapted. Here the anchor node is diagnosing the sensor node independently without any other node data. Therefore, in this paper the performances are measured by varying the percentage of faulty data in when a sensor node underlying intermittent fault. When more data are faulty in a faulty sensor node, the deviation in mean and variance will be more. That helps the LRT based algorithm to classify the faulty node with more noisy data as faulty sensor node easily. Therefore, the DA, FAR and FPR performance improves with increase in probability of faulty data in a faulty sensor node. Please note this is different from the percentage of faulty nodes in WSNs.

The DA, FAR and FPR are determined using the proposed LRT method for $L = 30$ and plotted in Fig. 3. It is observed that the performance is improved here over the scenario when $L = 20$ data are used for the diagnosis. When more data that has a longer diagnosis period is used for diagnosis, the probability of having faulty data points of intermittent faulty sensor nodes increases. Then the LRT algorithm can detect the faulty status easily with high accuracy. But, the computational cost along with required memory size is increased when more data is used for diagnosis.

After having the DA, FAR and FPR performance analysis, then the other performance parameters like positive likelihood ratio (PLR) and negative likelihood ratio (NLR) are determined and compared. These parameters are defined as

$$PLR = \frac{TPR}{FAR} = \frac{DA}{FAR}, \tag{37}$$

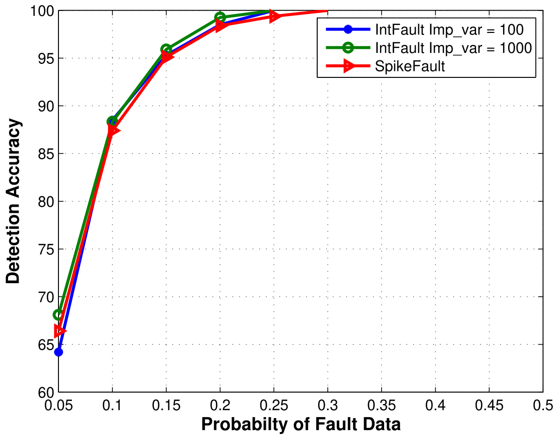
and

$$NLR = \frac{FAR}{TNR} = \frac{1 - DA}{1 - FAR} \tag{38}$$

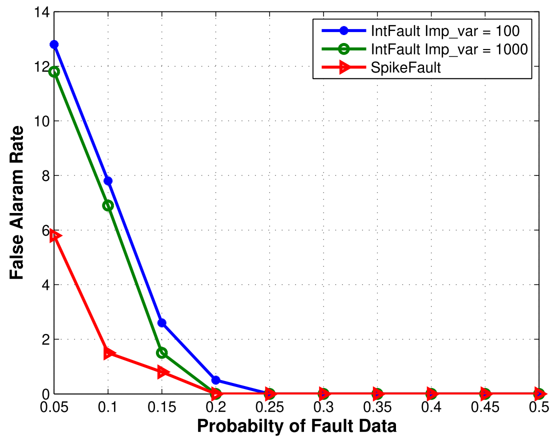
where TPR is the true positive rate and TNR is the true negative rate. Likelihood ratios are used to summarise diagnostic accuracy. Each LRT result has its own likelihood ratio. It summarizes how many times more (or less) likely the nodes are faulty to have that particular result for a fault free node. The PLR and NLR are plotted in Figs. 4 and 5 for 20 and 30 numbers of samples used for diagnosis respectively. In general, when the $PLR > 1$ indicates that the test LRT result is associated with the presence of faulty sensor nodes, whereas the $NLR < 1$ indicates that the test result is associated with the fault free sensor node. PLR above 10 and NLR below 0.1 are considered as strong evidence of the test rule in most circumstances.

In fact, the PLR gives the change in the odds of having a diagnosis in sensor nodes with a positive test and the change is usually greater than 1. The larger the PLR value, the more informative the test result. From the Figs. 4(a) and 5(a) shows that the PLR is more than 10 when $p_r \geq 0.1$ and give very high after the $p_r \geq 0.25$ for $L = 20$ and 30 respectively. This is due to the fact that the FAR decreases with increase in p_r and nearly zero after $p_r = 0.25$.

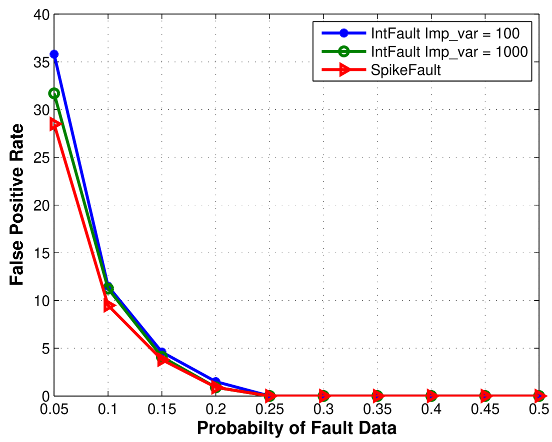
Similarly, the NLR gives the change in the odds of having a diagnosis in sensor nodes with a negative test. The change



(a) Detection Accuracy

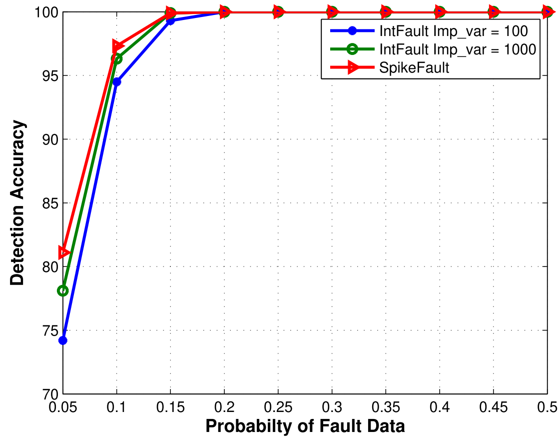


(b) False Alarm Rate

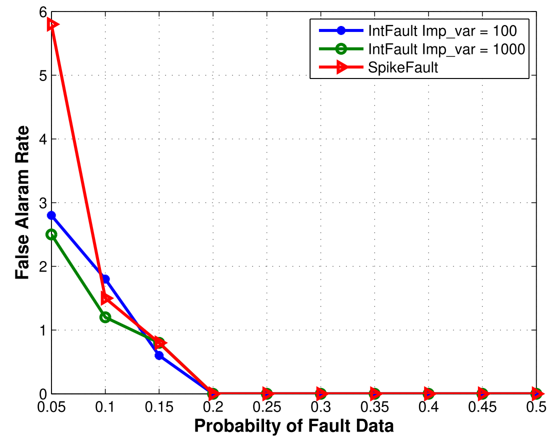


(c) False Positive Rate

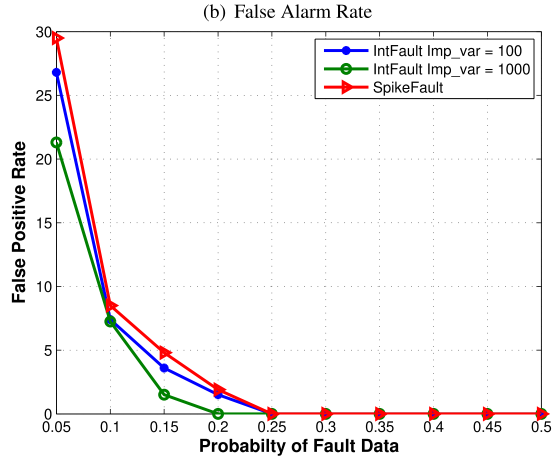
FIGURE 2. DA, FAR and FPR versus fault probabilities plots of intermittent/spike fault when the tolerance probability is ±0.05 and by suing the LRT criterion. The number of data used for diagnosis $L = 20$.



(a) Detection Accuracy



(b) False Alarm Rate

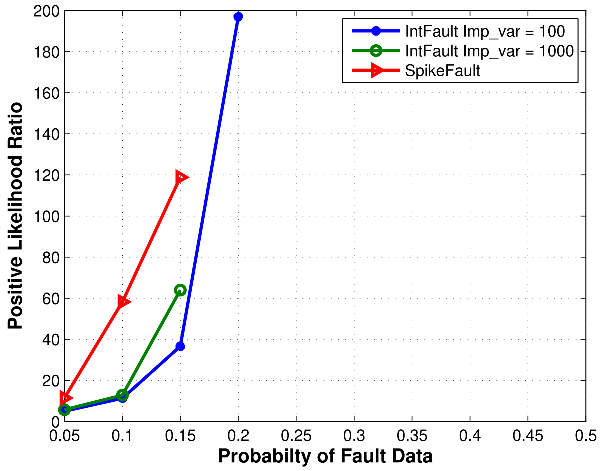


(c) False Positive Rate

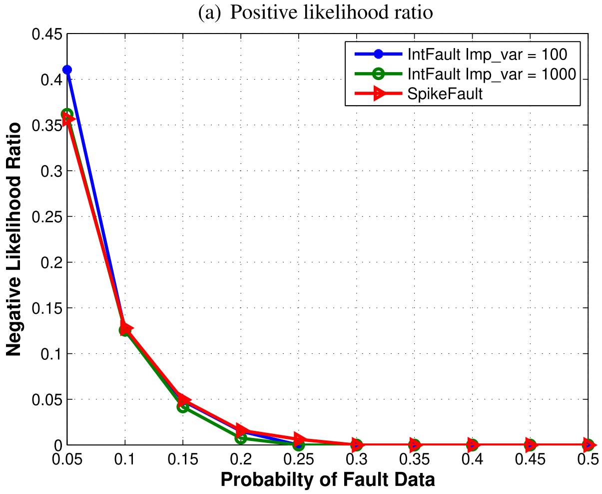
FIGURE 3. DA, FAR and FPR versus fault probabilities plots of intermittent/spike fault when the tolerance probability is ±0.05 and by suing the LRT criterion. The number of data used for diagnosis $L = 30$.

is usually less than 1. The test is more informative for the smaller value of NLR. It is observed from the Figs. 5(b) and Figs. 4(b) that the NLR value is asymptotically approaching zero when the $p_r = 0.25$ for both random and spike intermittent faults. The NLR value decreases when the diagnosis is carried out by taking 30 samples rather than 20, which indicates improvement of detection accuracy with an increase in the number of samples.

The likelihood ratio performance parameters shows that the proposed simultaneous LR statistics is providing better detection accuracy. The probability of detection accuracy increases with increase in the fault probability of the observed

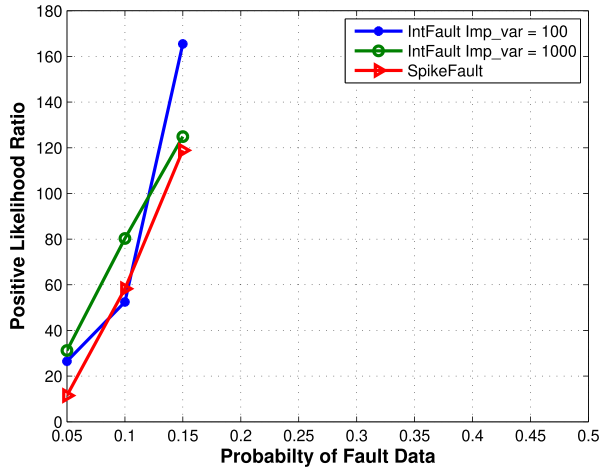


(a) Positive likelihood ratio

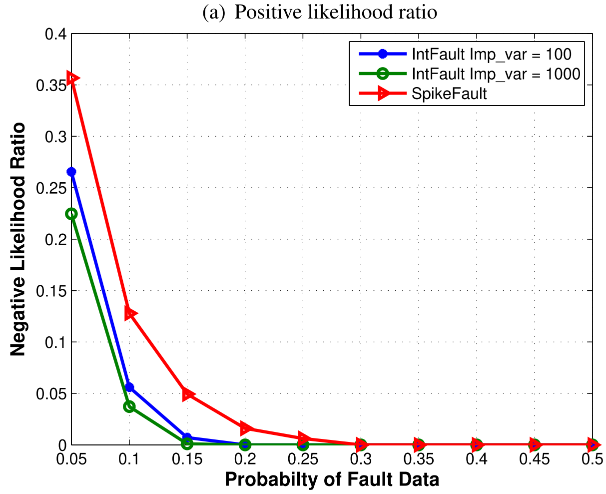


(b) Negative likelihood ratio

FIGURE 4. PLR and NLR versus fault probabilities plots of intermittent/spike fault by using the LRT criterion. The tolerance probability is ±0.05 and the number of data used for diagnosis $L = 20$.



(a) Positive likelihood ratio



(b) Negative likelihood ratio

FIGURE 5. PLR and NLR versus fault probabilities plots of intermittent/spike fault by using the LRT criterion. The tolerance probability is ±0.05 and the number of data used for diagnosis $L = 30$.

data in a sensor node. It is also evident from the results that the test rule achieves 100% accuracy when the $p_r \geq 0.25$.

C. COMPARISON WITH MODIFIED THREE SIGMA EDIT TEST

In this subsection, the performance of the LRT method is compared with a robust statistical method known as *modified three sigma edit test* (Mod3Sigma) [25]. In Mod3Sigma method, the robust statistics parameters such as median, median of absolute difference between data and median and robust standard deviation are computed first from the measured data set. Then each individual data is tested using the Modified Three Sigma Edit Test algorithm. It is a repeated test and the number of repetition is the same as the number of data used for diagnosis. Let us assume that if 20% of data are found suspicious, then the node is considered as faulty, otherwise it is fault free as against the 50% in [25]. in. Thus we have added fault the measure data with $p_r = 0.3$ to 0.6 with step size of 0.1. Both the fault detection algorithms are used to diagnose the faulty status. It is already observed that when the pr is 0.25 or more the LRT algorithm attains DA =1, FAR = 0 and FPR is also 0. Now, the same performance for Mod3Sigma is provided in Tables 1 and 2.

TABLE 1. Performance comparison between LRT and Mod3Sigma algorithms when $L = 20$.

Algorithm	Pr	Random fault			Spike fault		
		DA	FAR	FPR	DA	FAR	FPR
LRT	0.3	1	0	0	1	0	0
Mod3Sigma		68.5	5.6	37.4	77.4	1.9	29.7
LRT	0.4	1	0	0	1	0	0
Mod3Sigma		88.9	4.3	11.2	80	4.8	17.4
LRT	0.5	1	0	0	1	0	0
Mod3Sigma		82.1	5.4	19.9	80.5	4.5	15.3
LRT	0.6	1	0	0	1	0	0
Mod3Sigma		42.4	7.4	60.2	86.8	5.7	16.2

It is found in the table that the detection performance of the Mod3Sigma algorithm is improving first with p_r increasing from 0.3 to 0.4, then decreasing. In particular the FPR is increasing when more data points are faulty. It may be due to the masking effect. When more data points are becoming faulty, then the estimated robust standard deviation is very high. Any suspicious data near the median is considered as good data. Whereas, the LST is detecting the faulty node with

TABLE 2. Performance comparison between LRT and Mod3Sigma algorithms when $L = 30$.

Algorithm	p_r	Random fault			Spike fault		
		DA	FAR	FPR	DA	FAR	FPR
LRT	0.3	1	0	0	1	0	0
Mod3Sigma		79.4	2.4	18.5	88	4.9	11.2
LRT	0.4	1	0	0	1	0	0
Mod3Sigma		98.2	3.5	1.2	95.2	2.8	5.6
LRT	0.5	1	0	0	1	0	0
Mod3Sigma		87.2	2.8	15.5	80.9	5.5	22.2
LRT	0.6	1	0	0	1	0	0
Mod3Sigma		43.4	4.8	49.4	85.4	1.2	23.4

TABLE 3. Performance comparison between LRT and machine learning algorithms [33].

Algorithms	Pr	Precission	
		Random Intremitent faults	Spike faults
LRT	0.2	0.98	0.99
ET		1	0.88
DT		0.65	1
RF		0.98	0.92
SVM		0.98	0.99
MLP		0.96	0.26
LRT	0.4	1	1
ET		1	1
DT		0.97	0.91
RF		1	0.98
SVM		1	1
MLP		0.92	0.08
LRT	0.6	1	1
ET		1	1
DT		1	0.91
RF		1	1
SVM		1	1
MLP		0.98	0.28
LRT	0.8	1	1
ET		1	1
DT		1	0.94
RF		1	1
SVM		1	1
MLP		1	0.92

100% accuracy when the pr is more than 0.3%. In general the performance improves when the number of data used for diagnosis L is increased as expected.

D. COMPARISON WITH MACHINE LEARNING ALGORITHMS

In the recent past machine learning (ML) algorithms are used to classify different kinds of faults that occur in WSNs [5], [33]. The algorithms like multi-layer perception (MLP), random forest (RF), decision tree (DT), support vector machine (SVM) and extreme tree (ET). It has been reported that among all the machine learning algorithms, ET and RF perform better than the SVM, MLP and DT. We have defined the precision of the fault detection as

$$Precision = \frac{DA}{DA + FPR} \tag{39}$$

The precision of the proposed algorithm is compared with all the above said machine learning algorithms such as MLP, SVM, ET, DT and RF. The comparison result is provided in Table 3. It has been observed from the table that the performance improves with the increase in p_r value. With the same real time testing data, all the methods perform well with high percentage faulty data. Overall, the performance of SVM, ET and RF is at par with the proposed LRT test and outperformed over other classifiers. Another observation is that detecting accuracy of the random intermittent faults using ML methods is better than the spike intermittent fault.

Although the ML methods like ET are providing at par detection accuracy with the proposed algorithm, the computational complexity and processing time is very less for the LRT test based fault detection algorithm over the ML approach. Because the LRT test is a one shot calculation of likelihood ratio. Thus, the proposed method is one of the best candidates to diagnose intermittent faulty status of the sensor node in WSN.

VIII. CONCLUSION

In this paper we proposed a one shot likelihood ratio test to find the intermittent fault status of a sensor node in a wireless sensor network. The proposed method measures the mean and variance of the received data combined with the stored data set and then compares the likelihood ratio with the threshold value with certain tolerance. When the nodes are underlying intermittent fault, the statistics of the data is changing and hence the new LRT based fault diagnosis algorithm detecting the faulty node with 100% detection accuracy, 0% false alarm rate and 0% false positive rate when the probability of the faulty data is more than 25%. Whereas the robust modified repeated three sigma edit test fails to perform when the probability of faulty data is less than 30%. Further, the precision of the LRT algorithm is at par with machine learning approaches with very less computational complexity. In the future, we can extend the analysis to other kinds of faults in wireless sensor networks and improve the false positive rate performance.

REFERENCES

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Comput. Netw.*, vol. 38, no. 4, pp. 393–442, Mar. 2002.

[2] M. Panda, B. Gouda, and T. Panigrahi, "Distributed online fault diagnosis in wireless sensor networks," in *Design Frameworks for Wireless Networks* (Lecture Notes in Networks and Systems Series), S. Das, S. Samanta, N. Dey, and R. Kumar, Eds. Singapore: Springer, 2020, pp. 197–221.

[3] J. Xiao, Y. Zhu, Y. Zhong, and Z. Lin, "A review on fault diagnosis in wireless sensor networks," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 428, no. 1, 2020, Art. no. 012070.

[4] R. Palanikumar, K. Ramasamy, and P. S. Ragavan, "Faulty node detection and recovery scheme for large-scale wireless sensor network using hosted cuckoo optimization algorithm," *Int. J. Commun. Syst.*, vol. 35, no. 9, Jun. 2022, Art. no. e5143.

[5] S. Zidi, T. Moulahi, and B. Alaya, "Fault detection in wireless sensor networks through SVM classifier," *IEEE Sensors J.*, vol. 18, no. 1, pp. 340–347, Nov. 2018.

[6] T. Muhammed and R. A. Shaikh, "An analysis of fault detection strategies in wireless sensor networks," *J. Netw. Comput. Appl.*, vol. 78, pp. 267–287, Jan. 2017.

[7] Z. Zhang, A. Mehmood, L. Shu, Z. Huo, Y. Zhang, and M. Mukherjee, "A survey on fault diagnosis in wireless sensor networks," *IEEE Access*, vol. 6, pp. 11349–11364, 2018.

[8] T.-B. Dang, D.-T. Le, T.-D. Nguyen, M. Kim, and H. Choo, "Monotone split and conquer for anomaly detection in IoT sensory data," *IEEE Internet Things J.*, vol. 8, no. 20, pp. 15468–15485, Oct. 2021.

[9] A. Barati, S. J. Dastgheib, A. Movaghar, and I. Attarzadeh, "An effective fuzzy based algorithm to detect faulty readings in long thin wireless sensor networks," *Int. J. Tech. Phys. Problems Eng.*, vol. 3, no. 1, pp. 2077–3528, 2012.

[10] S. Javanmardi, "A novel approach for faulty node detection with the aid of fuzzy theory and majority voting in wireless sensor networks," *Int. J. Adv. Smart Sensor Netw. Syst.*, vol. 2, no. 4, pp. 1–10, Oct. 2012.

[11] P. Chanak and I. Banerjee, "Fuzzy rule-based faulty node classification and management scheme for large scale wireless sensor networks," *Expert Syst. Appl.*, vol. 45, pp. 307–321, Mar. 2016.

[12] S. Shao, S. Guo, and X. Qiu, "Distributed fault detection based on credibility and cooperation for WSNs in smart grids," *Sensors*, vol. 17, no. 5, p. 983, Apr. 2017.

[13] W. Liu, X. Luo, G. Wei, and H. Liu, "Node localization algorithm for wireless sensor networks based on static anchor node location selection strategy," *Comput. Commun.*, vol. 192, pp. 289–298, Aug. 2022.

[14] A. Barati, M. Dehghan, A. Movaghar, and H. Barati, "Improving fault tolerance in ad-hoc networks by using residue number system," *J. Appl. Sci.*, vol. 8, no. 18, pp. 3273–3278, Sep. 2008.

[15] P. Choudhari, D. Kundu, and N. Misra, "Likelihood ratio test for simultaneous testing of the mean and the variance of a normal distribution," *J. Stat. Comput. Simul.*, vol. 71, no. 4, pp. 313–333, Dec. 2001.

[16] H.-I. Park, "On tests detecting difference in means and variances simultaneously under normality," *Commun. Statist. Theory Methods*, vol. 46, pp. 10025–10035, Oct. 2017.

[17] M.-H. Lee and Y.-H. Choi, "Fault detection of wireless sensor networks," *Comput. Commun.*, vol. 31, no. 14, pp. 3469–3475, Sep. 2008.

[18] A. Mahapatro and P. M. Khilar, "Detection and diagnosis of node failure in wireless sensor networks: A multiobjective optimization approach," *Swarm Evol. Comput.*, vol. 13, pp. 74–84, Dec. 2013.

[19] E. Khazaei, A. Barati, and A. Movaghar, "Improvement of fault detection in wireless sensor networks," in *Proc. ISECS Int. Colloq. Comput., Commun., Control, Manage.*, Aug. 2009, pp. 644–646.

[20] B. C. P. Lau, E. W. M. Ma, and T. W. S. Chow, "Probabilistic fault detector for wireless sensor network," *Expert Syst. Appl.*, vol. 41, no. 8, pp. 3703–3711, Jun. 2014.

[21] P. Tang and T. W. S. Chow, "Wireless sensor-networks conditions monitoring and fault diagnosis using neighborhood hidden conditional random field," *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 933–940, Jun. 2016.

[22] A. Barati, S. J. Dastgheib, A. Movaghar, and I. Attarzadeh, "An optimised algorithm to detect faulty readings along the substrate access wireless long-thin sensor networks," in *Proc. UKSim 5th Eur. Symp. Comput. Modeling Simulation*, Nov. 2011, pp. 372–377.

[23] Z. Zhang, L. Shu, A. Mehmood, L. Yan, and Y. Zhang, "A short survey on fault diagnosis in wireless sensor networks," in *Wireless Internet*, M. Huang, Y. Zhang, W. Jing, and A. Mehmood, Eds. Cham, Switzerland: Springer, Dec. 2018, pp. 21–26.

[24] N. Wang, J. Wang, and X. Chen, "A trust-based formal model for fault detection in wireless sensor networks," *J. Sensors*, vol. 19, no. 8, p. 1916, Apr. 2019.

[25] M. Panda and P. M. Khilar, "Distributed self fault diagnosis algorithm for large scale wireless sensor networks using modified three sigma edit test," *Ad Hoc Netw.*, vol. 25, pp. 170–184, Feb. 2015.

[26] M. Panda and P. M. Khilar, "Distributed byzantine fault detection technique in wireless sensor networks based on hypothesis testing," *Comput. Electr. Eng.*, vol. 48, pp. 270–285, Nov. 2015.

[27] R. R. Swain, P. M. Khilar, and S. K. Bhoi, "Underlying and persistence fault diagnosis in wireless sensor networks using majority neighbors co-ordination approach," *Wireless Pers. Commun.*, vol. 111, no. 2, pp. 763–798, Mar. 2020.

[28] S. H. Asman, N. F. A. Aziz, U. A. U. Amirulddin, and M. Z. A. A. Kadir, "Decision tree method for fault causes classification based on RMS-DWT analysis in 275 kV transmission lines network," *Appl. Sci.*, vol. 11, no. 9, p. 4031, Apr. 2021.

[29] R. R. Swain and P. M. Khilar, "Composite fault diagnosis in wireless sensor networks using neural networks," *Wireless Pers. Commun.*, vol. 95, no. 3, pp. 2507–2548, Aug. 2017.

[30] M. Panda, B. S. Gouda, and T. Panigrahi, *Fault Diagnosis in Wireless Sensor Networks Using a Neural Network Constructed by Deep Learning Technique*. Singapore: Springe, Feb. 2020, pp. 77–101.

[31] S. U. Jan, U. Saeed, and I. Koo, "Machine learning for detecting drift fault of sensors in cyber-physical systems," in *Proc. 17th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2020, pp. 389–394.

[32] R. Regin, S. Rajest, and B. Singh, "Fault detection in wireless sensor network based on deep learning algorithms," *ICST Trans. Scalable Inf. Syst.*, vol. 21, Jul. 2018, Art. no. 169578.

[33] U. Saeed, S. U. Jan, Y.-D. Lee, and I. Koo, "Fault diagnosis based on extremely randomized trees in wireless sensor networks," *Rel. Eng. Syst. Saf.*, vol. 205, Jan. 2021, Art. no. 107284.

[34] M. R. C. Acosta, S. Ahmed, C. E. Garcia, and I. Koo, "Extremely randomized trees-based scheme for stealthy cyber-attack detection in smart grid networks," *IEEE Access*, vol. 8, pp. 19921–19933, 2020.

[35] Z. Chen, J. Chen, Y. Feng, S. Liu, T. Zhang, K. Zhang, and W. Xiao, "Imbalance fault diagnosis under long-tailed distribution: Challenges, solutions and prospects," *Knowl.-Based Syst.*, vol. 258, Dec. 2022, Art. no. 110008.

[36] S. Wang, L. Minku, and X. Yin, "Online class Imbalance learning and its application in fault detection," *Int. J. Comput. Intell. Appl.*, vol. 12, no. 4, 2013, Art. no. 1340001.

[37] Y. Xie and T. Zhang, "Imbalanced learning for fault diagnosis problem of rotating machinery based on generative adversarial networks," in *Proc. 37th Chin. Control Conf. (CCC)*, Jul. 2018, pp. 6017–6022.

[38] Y. Li, J. Zhang, S. Zhang, W. Xiao, and Z. Zhang, "Multi-objective optimization-based adaptive class-specific cost extreme learning machine for imbalanced classification," *Neurocomputing*, vol. 496, pp. 107–120, Jul. 2022.

[39] W. Xiao, J. Zhang, Y. Li, S. Zhang, and W. Yang, "Class-specific cost regulation extreme learning machine for imbalanced classification," *Neurocomputing*, vol. 261, pp. 70–82, Oct. 2017.

[40] H. S. Z. Kazmi, N. Javaid, M. Awais, M. Tahir, S. Shim, and Y. B. Zikria, "Congestion avoidance and fault detection in WSNs using data science techniques," *Trans. Emerg. Telecommun. Technol.*, vol. 33, no. 3, Mar. 2022, Art. no. e3756.

[41] P. Chanak, I. Banerjee, and R. S. Sherratt, "Mobile sink based fault diagnosis scheme for wireless sensor networks," *J. Syst. Softwares*, vol. 119, pp. 45–57, Sep. 2016.

[42] M. Abo-zahhad, S. M. Ahmed, N. Sabor, and S. Sasaki, "Mobile sink-based adaptive immune energy-efficient clustering protocol for improving the lifetime and stability period of wireless sensor networks," *IEEE Sensors J.*, vol. 15, no. 8, pp. 4576–4586, Aug. 2015.

[43] S. M. Ross, *Introduction to Probability Models*, 10th ed. New York, NY, USA: Academic, 2010.

[44] A. C. Rencher, *Methods of Multivariate Analysis*, 2nd ed. Hoboken, NJ, USA: Wiley, 2002.

[45] A. Kumar, D. Manjunath, and J. Kuri, "Ad hoc wireless sensor networks (WSNs)," in *Wireless Networking* (Series in Networking), A. Kumar, D. Manjunath, and J. Kuri, Eds. San Mateo, CA, USA: Morgan Kaufmann, 2008, pp. 337–374.

[46] H.-I. Park, "Test procedures for the mean and variance simultaneously under normality," *Commun. Stat. Appl. Methods*, vol. 23, no. 6, pp. 563–574, Nov. 2016.

[47] S. Suthaharan, M. Alzahrani, S. Rajasegarar, C. Leckie, and M. Palaniswami, "Labelled data collection for anomaly detection in wireless sensor networks," in *Proc. 6th Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process.*, Dec. 2010, pp. 269–274.

BHABANI SANKAR GOUDA received the M.Tech. degree in computer science and engineering (CSE) from the Biju Patnaik University of Technology (BPUT), Rourkela, Odisha, India, in 2010, where he is currently pursuing the Ph.D. degree with the Department of CSE. He worked as an Assistant Professor at the National Institute of Science and Technology (NIST), Berhampur, Odisha. He published more than 15 research articles in reputed journals, book chapters, and conference proceedings. His research interests include fault diagnosis in wireless sensor networks and robust estimation in impulsive noise environment.

MEENAKSHI PANDA received the M.Tech. degree in computer science and engineering (CSE) from Utkal University, Odisha, India, in 2006. She worked as an Assistant Professor at the Silicon Institute of Technology (SIT), Bhubaneswar, Orissa, for four year before joining Ph.D. She is currently working as an Adhoc Faculty Member with the Department of CSE, Indian Institute of Information Technology Vadodara, India. She published more than ten research articles in reputed journals, book chapters, and conference proceedings. Her research interests include fault diagnosis in distributed networks and the IoT.

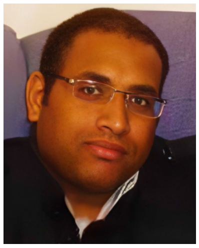
TRILOCHAN PANIGRAHI (Senior Member, IEEE) received the M.Tech. degree in electronics and communication engineering (ECE) from the National Institute of Science and Technology, Berhampur, under the Biju Patnaik University of Technology, Rourkela, India, in 2005, and the Ph.D. degree in ECE from the National Institute of Technology Rourkela, in 2012. He visited The University of Edinburgh, as a Visiting Research Scholar, during his Ph.D. time. He is currently working as an Associate Professor with the Department of ECE, National Institute of Technology Goa, India. He has published more than 90 research papers, including five book chapters in international journals and conferences of high repute. His research interests include signal processing for wireless communication, the IoT, application of evolutionary algorithms in signal processing, and nano WSN. He is a Life Member of ISTE and a member of the Institute of Engineering, India (IEI).

SUDHAKAR DAS (Member, IEEE) received the Ph.D. degree in electronics and communication engineering from the Biju Patnaik University of Technology (BPUT), India, in 2017. He is currently working as a Professor with the National Institute of Science and Technology (NIST), India. He has published more than 13 papers in the journals of repute and has presented his work in more than 15 national/international conferences. His research interests include modeling of homo/hetero-structure semiconductor devices to operate at high frequency for use in high-speed VLSI circuits and communication engineering applications and fault diagnosis in wireless sensor networks and robust distributed estimation in faulty WSN. He is a Life Member of IETE.

BHARGAV APPASANI received the Ph.D. degree from the Birla Institute of Technology, in 2018. He is currently working as an Assistant Professor with the School of Electronics Engineering, Kalinga Institute of Industrial Technology, India. He has published over 85 articles in reputed international journals and conferences. He has three patents filed to his credit and has published a book with Springer. He is the Academic Editor of the Journal of Electrical and Computer Engineering (Hindawi). His research interests include terahertz sensing, smart grid, 5G, and wireless sensor networks.

OMPRAKASH ACHARYA (Member, IEEE) received the Ph.D. degree in electronics and communication engineering from the Indian Institute of Technology, Roorkee, India, in 2016. Since 2016, he has been associated with the School of Electronics, Kalinga Institute of Industrial Technology (KIIT) Deemed to be University, Bhubaneswar, India, as an Assistant Professor. Prior to this, he worked as an Assistant Professor at the National Institute of Technology, Puducherry, India. He received the Senior Research Fellowship, from 2009 to 2012, and MHRD Fellowship, from 2012 to 2014, to carry out his research work at IIT Roorkee. His research interests include antenna arrays, array signal processing, MIMO, techniques for performance enhancement of planar antennas, electromagnetics absorber, and application of soft computing techniques in electromagnetics. On these topics, he has published more than 30 technical papers in international journals and conferences He has coauthored one book chapter on "Particle Swarm Optimization for Antenna Array Synthesis, Diagnosis and Healing."

HOSSAM M. ZAWBAA received the B.Sc. and M.Sc. degrees from the Faculty of Computers and Artificial Intelligence, Cairo University, Egypt, in 2008 and 2012, respectively, and the Ph.D. degree from Babeş-Bolyai University, Romania, in 2018. He is currently an Assistant Professor with the Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni Suef, Egypt. He has authored or coauthored over 90 research publications in peer-reviewed reputed journals and international conference proceedings. His research interests include computational intelligence, machine learning, computer vision, and natural language understanding. He has been rated as one of the top 2% scientists worldwide by Stanford in the field of AI, in 2020, 2021, and 2022. He was awarded the State Encouragement Award in the field of engineering sciences from the Academy of Scientific Research and Technology, Egypt, in 2020.



SALAH KAMEL received the international Ph.D. degree from Jaen University, Spain (Main), and Aalborg University, Denmark (Host), in January 2014. He is currently an Associate Professor with the Department of Electrical Engineering, Aswan University. He is also the Leader with the Advanced Power Systems Research Laboratory (APSR Lab), Power Systems Research Group, Aswan, Egypt. His research interests include power system analysis and optimization, smart grid, and renewable energy systems.