

Fuzzy-Based Flat Anomaly Diagnosis and Relief Measures in Distributed Wireless Sensor Network

Usman Barakkath Nisha¹ · Natarajan Uma Maheswari¹ · Ramalingam Venkatesh² · Rabi Yasir Abdullah³

Received: 13 January 2016/Revised: 7 September 2016/Accepted: 12 September 2016/Published online: 22 September 2016
© Taiwan Fuzzy Systems Association and Springer-Verlag Berlin Heidelberg 2016

Abstract This paper bestows a distributed adaptive scheme for diagnosing inaccurate data (anomaly) in wireless sensor networks. Faults occurring in sensor nodes are routine owing to the sensor device itself and the harsh environment in which the sensor nodes are deployed. It is mandatory for the WSNs to discover the anomaly and take actions to avoid further seediness of the network lifetime for confirming data accuracy. In this standpoint, we propose two perspectives for diagnosing and alleviating anomalies. The first view depicts input space partitioning by subtractive clustering method with robust density measure. Later, Takagi–Sugeno fuzzy inference model is applied for selection of several parameters and its membership functions, and rule-based construction is practiced to spot anomalies in distributed clustering wireless sensor network. By exploring combined correlation analysis with second perspective, the eliminated anomalous data are replaced by imputed data. Experimental results infer

accuracy and reliability with a reduced amount of energy consumption than the state-of-the-art techniques.

Keywords Accuracy · Anomaly detection · Subtractive clustering · Fuzzy · Sensor networks · Takagi–Sugeno model

1 Introduction

Wireless sensor networks (WSNs) applications surge very fast over the past by collecting information and monitoring the solutions. They are enumerated as group of interlinked sensors of various forms around the specified area [1, 2]. Battery power is an important limitation learned by various designers in WSNs since it hammers the overall performance of the whole system and be handled with care. The data deviating from its peers are coined by the term “Outliers” or “Anomalies” which originates from the field of statistics [3]. All sensors may fail in unreliable and harsh environments. Faulty sensors result in noisy and random readings due to interferences. The main reasons for anomalies are

1. Failure of sensor nodes/node anomaly
 - a. Battery power drains to 0 %.
 - b. Bugs may short-circuit sensor nodes.
 - c. Environmental changes.
2. Miscomputation process/data aggregation anomaly
 1. Inner cluster
 2. Outer cluster
- b. Inability to analyze huge data

✉ Natarajan Uma Maheswari
numamahi@gmail.com
Usman Barakkath Nisha
ubnisha@gmail.com
Ramalingam Venkatesh
rlvenkatesh@gmail.com
Rabi Yasir Abdullah
yasirsince1984@gmail.com

¹ Department of Computer Science and Engineering, PSNA College of Engineering and Technology, Dindigul, Tamilnadu, India

² Department of Information Technology, PSNA College of Engineering and Technology, Dindigul, Tamilnadu, India

³ Department of Computer Science and Engineering, SSCET, Palani, Tamilnadu, India

3. Miscommunication process/network anomaly

- a. Original values of sensor nodes may be routed in wrong directions due to loss of connectivity

Large numbers of highly correlated data are demanded by redundant data, and energy is drained in huge amount which in turn will be processed and collected by the base station. Network lifetime is boosted by providing fused information and purging redundant transmission through data aggregation [4]. If anomalies are not spotted during the data aggregation process, the data inaccuracy ensues [6]. Data aggregation follows cluster structure, where cluster head performs data aggregation process. The data collected by cluster head from one or more sensing nodes are smeared to aggregation functions. Sink often receives the aggregated value [5] [7]. In this context, some faulty nodes may be present, producing incorrect readings that deviate from the exact output. Cluster head (aggregator) may perform inaccurate calculation, resulting in anomaly.

The existence of any anomaly will end in inaccurate query results and thwart the system's efficiency. Thus, it is acute in identifying and replacing incorrect information so as to improve the query accuracy and reliability. In this paper, we focus at the snag of determining inaccurate readings with high reliability in sensor networks. Obviously, an ingenious approach to this problem is to accumulate all readings to a sink, where fuzzy-based analysis is performed to finalize anomalous data. In this proposed work, we spotlight the anomaly diagnosis process in cluster head in view of the fact that data accuracy may be increased gradually by concentrating on every individual sensor's reading. The anomaly disclosure of each sensor must be broadcasted to a cluster center for shaping a core decision. The cluster center performs data aggregation beyond considering the readings from anomalous sensor nodes [8]. In this lap, base station receives accurate data, though the reliability is depraved. The cluster center forecasts reliability of each received detection result throughout the making of exclusive final decision. When the final decision is not significant, the cluster center will mandate the sensor which has sent the acknowledged result with the lowermost reliability to retransmit. To rise above this issue, the imputed data are predicted based on correlation between sensor's readings. This prospective fuzzy-based model will scrutinize the anomalies originated by sensor nodes, and the readings are mitigated by the clean imputed value. Accordingly, the global accuracy and reliability of the distributed sensor network are elevated.

Pertaining to the above observation in the proposed system, two strategies in finding anomaly diagnosis and relief measures are dealt. The first standing concentrates on forming optimal clusters with subtractive clustering method proceeding from robust density measures that focus

on creating input space for the fuzzy inference model. The second phase puts across the fuzzy logic technique with Takagi–Sugeno (TS) fuzzy inference model [9]. It receives inputs from first juncture for identifying parameters and its relationships. Anomaly will be uncovered by crossing second juncture with increase in accuracy and short in curing anomalous data. Anomalous data are removed in aggregation process which worsens the performance of entire network. Missing of data in data stream lowers reliability while receiving fused data at base station. To overcome this issue, fuzzy rules are applied to replace anomalous data by estimating imputed data exploiting correlation technique.

The leftovers of the paper are organized as follows: Section 2 is granted with some significant related literature review for our proposed algorithm. Section 3 clarifies the network model and problem statement. Section 4 mounts the proposed approach and methodology of the three junctures. Section 5 concludes the experimental evaluation based on both unreal and real datasets with comparison results. Finally, Sect. 6 concludes this paper.

2 Related Work

Wireless sensor networks face challenge as long as base station requires accurate data. Anomalies are cultured by cumulating data accuracy in WSNs. The norms for anomaly detection are to establish a normal profile of the monitored object, and anomaly is the significant deviation from this normal profile [3]. Anomaly detection techniques are generally ordered into: statistical detection methods, data-mining-based methods, and rule-based (fuzzy logic) methods [10]. Statistical anomaly detection methods build both normal profile during a training phase and current profile during the detection phase. Data-mining-based methods can systematize the process of finding meaningful activities and fascinating features like classification-based intrusion detection. Mostly, they are computational intensive and yield very high false alarm rates. In particular, statistical-based techniques provide good detection rate with less false alarm rate by involving a mathematical model that requires more computational time. In WSNs, all sensor nodes monitor the surroundings, store data in the main memory, and interconnect the data between neighbor nodes by evaluating the computation process where energy is required for all the operations mentioned above. Meanwhile, capacities of the sensor nodes restrict its use for essential threads. In [11], Pottie postulates that the energy anticipated for communication is more than the energy required for computation. Data aggregation consumes cluster structure in hierarchical fashion by transporting data from source to sink in saving computational memory

though summarizing data from sensor nodes. Throughout this process, annoying anomalous data are detected in making meaningful data aggregation process.

Chitradevi et al. [12] projected anomaly detection centered on distributed agglomerative clustering method by removing anomalies at both local and global levels. Using cluster distance and density measures, optimal clusters are grouped keeping anomalies lower with cheap computational and communication complexity. Zhang et al. [13] have anticipated an ellipsoidal-based support vector machine in classifying sensor node data as anomaly using ellipsoidal SVM-based online anomaly detection and adaptive anomaly detection for multivariate data. They used the time window concept for classifying deviations from normal behavior of the system. This system hurts computational complexity by periodically updating a normal profile. Zhang et al. [14] recommended statistical-based outlier detection based on time series and geostatic analysis using spatial and temporal correlation theories. Variogram model develops a way of exhibiting temporal correlation by decent auto-regressive moving average (ARMA) and spatial correlation model.

Kapitanova's et al. [15] offered overall fuzzy logic system by means of spatial and temporal semantics for event detection by decreasing the number of rules in merging simple rules and pruning undesirable rules in the rule-based system. They utilized fuzzy logic as an alternative in taking fixed thresholds and crisp values, where the accuracy of fire event detection is increased. Liang et al. [16] suggested dual sliding window detection for increasing the revealing rate of event discovery. Conversely, they elaborate the outcome of fuzzy logic and the power of spatial and temporal possessions of the data in classifying detection rate.

In [17], the authors deliberated inputs similar to packet delivery ratio, energy, distance, packet loss and receiving signal strength to determine the jamming attacks in WSN using fuzzy-based optimization techniques. Finding anomalies after these attacks are difficult. In [18], imperialist clustering algorithm (ICA) is altered using fuzzy logic controller, which is implemented to vary the incorporation operator in the competition phases of ICA achieving less detection rate on an average of 87 %. Kumaragea et al. [19] recommended a fuzzy data modeling for distributed anomaly detection in dissimilar real datasets. Scalability and sensitivity are less while employing large number of nodes. In our previous work [20], we used relative correlation clustering technique for distinguishing anomaly. This model employs clustering and re-clustering process with relative combinational correlation technique [20].

A fuzzy prototype suitable for estimating schemes and functions is the Takagi and Sugeno's fuzzy model which is connected with fuzzy rules and possesses a job special

format with a functional-type consequent. This model concentrates on nonlinear dynamic systems with best proven results. In [21], TS fuzzy modeling algorithm using input–output data online in the presence of noise, extended Kalman filter (EKF) is used which deals with the minimum-variance state estimator in case of linear dynamic systems having white noise with zero-mean value. Likewise, switched and hybrid nonlinear systems use observer-based fault-tolerant control and robust fault-tolerant tracking control, respectively [22] [23]. For a real linear drive system in [24], Kalman filter also called as the residual generator is effective in fault diagnosis and fault-tolerant problems. In [25], authors present a well-organized hybrid fuzzy clustering style for Takagi–Sugeno (TS) fuzzy modeling based on a hybrid fuzzy clustering scheme. The method travels through various phases for constructing an optimal Takagi–Sugeno fuzzy model from sample data. Gustafson–Kessel clustering algorithm is developed, resulting in an optimal input–output space fuzzy partition matrix.

In [26], MIMO TS model is used to select suitable cluster. MIMO TS models are constructed from numerical data by using fuzzy cluster algorithm based on weighted fuzzy expected value (WFEV). Results obtained from this model achieve high performance against its peers if TS model is applied for adaptive fuzzy model sensor node (ATSFMSN) routing protocol. This model focuses only on fuzzy clustering algorithm based on weighting factor and fails in identifying inconsistent data. In [27], subtractive clustering-based clustering routing algorithm for wireless sensor networks is used to choose the cluster heads and generate the cluster heads where the node density is high. Their results provide the reasonable arrangement of cluster head, longer lifetime of the first node, and even longer lifetime of the network to balance the energy consumption of all the nodes in the network. Here data density measure is calculated based on Euclidean distance which will not yield optimal cluster head. In our proposed system, we use the Mahalanobis distance for shaping the cluster radius in factor space. Mahalanobis distance generates hyper ellipsoid clusters with similar volume. In [28], the authors focus on the routing algorithm based on subtractive clustering so as to reduce communication of WSNs and terminated flooding. Cluster head selection adopts subtractive clustering to produce cluster node with new cost function. Though it provides effective survival time and saves network energy, it fails in providing data accuracy whenever anomaly occurs. In our proposed system, input space partitioning is performed by using subtractive clustering with Mahalanobis distance by dealing with spatial relationships among the various sensor nodes. Optimal clusters are formed using this spatial correlation analysis with mean and standard deviation.

Proactive correlated fuzzy system with spatial, temporal, and attribute act is applied with fuzzy C-means clustering for refining data accuracy [29]. Similarly, the proposed system spots anomaly and determines imputed data in WSNs. It is evident from the literature survey that an ideal anomaly detection system should harvest high accuracy with negligible energy consumption. The proposed system integrates fuzzy logic techniques like subtractive clustering and TS fuzzy inference model for the improvement of reliability and accuracy in WSNs. The main contributions of the proposed work are summarized:

1. Input space portioning is performed for creating optimal clusters using well-conditioned subtractive clustering.
2. Diagnosing anomalies with the help of robust TS fuzzy model increase data accuracy.
3. Replacing anomaly with imputed data increases data integrity.

3 Network Model and Problem Statement

We put on the wireless sensor network as a large connected network with totally n sensor nodes denoted by $1, 2, 3 \dots n$. The nodes are dispersed randomly in some physical domain and are stationary after deployment. The transmission range for each node is static, and link between the nodes is bidirectional. The system can be modeled as communication graph $G = \{S, E\}$, where $S = \{1, 2 \dots n\}$, and $E = \{(S_x, S_y): S_x \text{ and } S_y \text{ are any two nodes in same cluster}\}$. A cluster is a unit disk having radius equivalent to center node's transmission range. A suitable clustering protocol for implementing distributed clustered wireless sensor network is anticipated by grouping several clusters into deployment area. The center node is the aggregator though the node that is a one hop neighbor of the aggregators of two different clusters is the outshine node. After the self-ruling cluster formation, only aggregator and outshine nodes got elected in a fully distributed fashion which participates in the inter-cluster communication despite the fact that sensing nodes in each cluster communicate with their aggregator node (and to other nodes when required). Figure 1 portrays the distributed sensor clustered architecture.

As discussed earlier, WSNs consist of many types of anomaly (like node or link failure) and avoids erroneous calculation in aggregation. Earlier, we worked on finding the animalized or normal node in the sensor network. In this critique, the total number of nodes was alienated into a number of clusters. Each cluster has an aggregator, and some acts as outshine nodes for dispatching the message from cluster to base station. The anomaly is discovered by

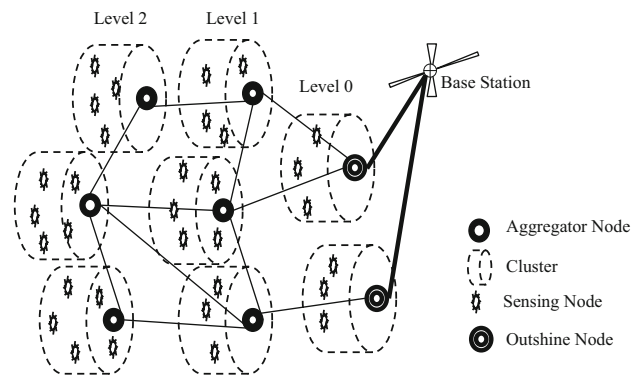


Fig. 1 Clustered network framework

the aggregator node in respective clusters, and the message is being forwarded to all nodes of the clusters and other aggregators. Entire clusters will be operating simultaneously. Each aggregator accomplishes data aggregation process, and the fused information is spawned and advanced to the base station. In this recommended model, optimized clusters are formed initially followed by replacing anomalous data with imputed data to be done by the aggregator in each cluster.

4 Proposed Methodology

The fuzzy-based anomaly detection and alleviation system starts by input space partitioning, where subtractive clustering is employed and accompanied by detecting erroneous data perfectly and replacing completely using TS fuzzy modeling. The proposed model is embodied in Fig. 2. A fuzzy model structure can be epitomized by a set of fuzzy IF-THEN rules. A rule-based fuzzy model obliges rule antecedent, rule consequent, and assembly of membership functions. The stages in the system comprise.

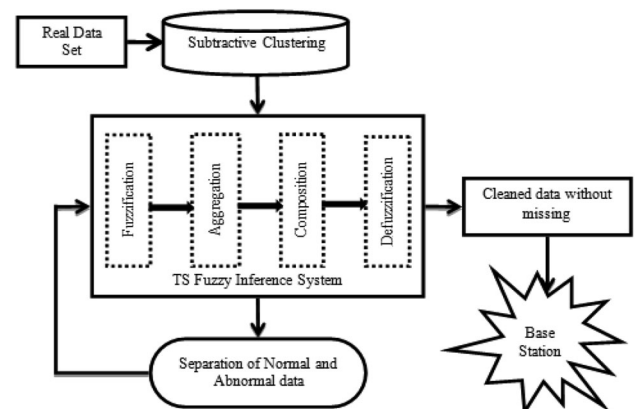


Fig. 2 Schematic representation of anomaly diagnosis and relief measure model

1. Fuzzification: Input and output variables are defined by mapping the crisp input into linguistic values.
2. Inference: Fuzzy inference holds number of fuzzy IF–THEN rules. Rule-based database delineates the membership functions of the fuzzy sets used in the fuzzy rules. Main processes executed by inference engine are:
 - a. Aggregation: Compute the IF part (antecedent) of the rules. The antecedent variables replicate information about the process operating environments.
 - b. Composition: Compute the THEN part (consequence) of the rules. The consequent variables are a linear regression model nearby the given functioning condition.
3. Defuzzification: The output variable calculated in the composition juncture is transformed to real output stage.

In analyzing the conventional techniques, it is inferred whether the data are incorrect or not and lack in replacing the incorrect data. To solve this concern, we use a number of correlation analysis for identifying imputed data and are substituted at the place of spotted data as anomaly.

Data accuracy will be upgraded by recognizing the anomaly with the support of aforesaid techniques, thereby dropping the computational power in sensors which reduces the sensor's battery power utilization rate by increasing the lifetime of all sensors. At the end of TS anomaly diagnosis model, base station reaches maximum accuracy and reliability is increased by using additional rules for calculating accurate imputed data.

4.1 Subtractive Clustering Method

Clustering slices a dataset into few groups in such a way that the similarities inside the groups are bigger among its peers. Clustering is to separate the data space into a few groups with kindred data, and grouping strategies are utilized broadly to sort out and classify data, as well as valuable for data compression and model development. Moreover, the greater part of the information congregated in numerous issues appear to have some immutable properties that lend themselves to standard groupings [30]. Discovering these groupings or attempting to arrange the information is not a straightforward job for people unless the information is of low dimension. Generally, clustering process is well arranged into off-line and online process. Online grouping is a technique in which every data vector is employed to overhaul the cluster as indicated by its vector position. In off-line mode, the structure is given a preparing dataset, which is utilized to locate the bunch by

examining all the data vectors in the preparing set. When the cluster focuses are found settled, they are utilized later to order new input vectors.

In this proposed work, input space model is erected with subtractive clustering methods which are chiefly utilized as a part of off-line clustering procedure. Subtractive clustering method (SCM) is similar to mountain clustering [31, 32], except in determining the data density measure at every reasonable position in the factor space. It employs the positions of the data focusing to figure the density measure, lessening the quantity of counts essentially. SCM comprehends computational intricacy utilizing cluster information rather than lattice problem as focused in mountain clustering [33, 34].

The computation is currently relative to the issue size of the input which is the sensed values from the sensors. On the other hand, the real cluster centers are not always situated at one of the data points. SCM decreases computational complexity and gives a suitable conveyance among cluster centers.

Consider assortment of data collected from sensor nodes $\{X_1, X_2, \dots, X_n\}$ which are the vectors in the dimensional space. Without forfeiture of sweeping statement, we expect that the element space is standardized so that all the data are confined by a unit of hyper solid shape. We assume every data point as a possibility for cluster centers which describes data density measure (*DDM*) of the point to serve as a cluster center. A data density measure for cluster center is computed by Eq. (1). The possibility of cluster center X_i is designated as DDM_i . After the density measure of each data point, it will have a high DDM_i value that is selected as the dominant cluster center. Let X_{cc_1} be the possibility point of cluster center and DDM_{c_1} be its density measure. DDM_i for every data point is revised by Eq. (2).

$$DDM_i = \sum_{i=1}^n \left\{ e^{-\left(\frac{4\sqrt{md}}{r_a^2}\right)} \right\} \quad (1)$$

$$DDM_i = DDM_i - DDM_{c_1} \left\{ e^{-\left(\frac{4\sqrt{md}}{r_b^2}\right)} \right\} \quad (2)$$

where, Mahalanobis distance is denoted as *md* and $r_a r_b$ are positive constants. r_a is outlining hyper solid shape cluster radius in factor space [35]. Usually $r_b = 1.5r_a$ where r_b is used to isolate from maliciously spaced cluster center. *md* is designed by the given following Eq. (3), where μ_m is the mean and σ_m is the standard deviation of X_i or X_{cc_1}

$$md = (X_i - \mu_m)' \sigma_m^{-1} (X_i - \mu_m) \quad (3)$$

where,

$$\mu_m = \frac{1}{n} \sum_{i=1}^n X_i \quad \sigma_m = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \mu_m)^2}$$

Succeeding the DDM, every data point is analyzed and the highest possibility point is elected as a next cluster center. The process produces sufficient numbers of cluster centers by performing n number of iterations. At last, the cluster center is fixed and the data density measure relates to the remaining updated data points. Halting condition for the likelihood of cluster center calculation can be established in [31].

Input space structuring is prepared by subtractive clustering with cluster centers. These cluster centers would be sensibly employed as the centers for the fuzzy reasoning in a Sugeno fuzzy model. We expect the center for the i th cluster is C_i in factor space. C_i can be divided into segment vectors IP_i and OP_i where IP_i is the input part holding the input elements of C_i ; OP_i is the output part containing yield elements of C_i . At this point, given a data vector X , the extent to which fuzzy rule i is satisfied by the membership function ϑ_i Where $\| \dots \|$ denotes the Euclidean distance.

$$\vartheta_i = e^{-\left(\frac{4x-IP_i^2}{r_a^2}\right)} \tag{4}$$

4.2 Takagi–Sugeno Fuzzy Modeling

Fuzzy-based modeling chiefly based on Takagi–Sugeno (TS) fuzzy systems permit in gaining highly accurate models with trivial number of rules [9]. TS models are universal approximators and achieve high accuracy with a small number of rules [36, 38]. Whenever number of rules in TS model is increased, lower approximation error rises and the quality of the modeling algorithm can be critical. It is relatively easy to alter them into nonlinear state model by supporting formal analysis to be used in control engineering. A fuzzy model construction can be characterized by a set of fuzzy IF–THEN rules. A rule-based fuzzy model involves rule antecedent, rule consequent, and structure of membership functions.

The modified inference approach is a universal approximation of any smooth nonlinear system which was proposed by Takagi and Sugeno [9, 37]. TS fuzzy model is embodied by a small set of fuzzy IF–THEN rules that refer local input–output functions of nonlinear systems. Rule of continuous TS fuzzy model is of the following form:

If $x_1(t)$ is M_{k1} and $x_2(t)$ is M_{k2} ... and $x_n(t)$ is M_{kn}

$$\text{Then } \begin{cases} \dot{x}(t) = A_i x(t) + B_i p(t) \\ y(t) = Cx(t) \end{cases}, i = 1, 2, \dots, r;$$

where x_1, x_2, x_n are input variables and $y(t)$ is the output variable, $M_{k1}, M_{k2}, \dots, M_{ij}$ are fuzzy sets, $A, B,$ and C are matrices of proper dimensions, and r is the number of fuzzy IF–THEN rules. Suppose W_i is a firing strength of the logical expression x_1 is M_{k1} and x_2 is M_{k2}, \dots and x_n is M_{kn} is Rule i , then the overall output is obtained via weighted mean value, by escaping the time-consuming process of

defuzzification which is essential in Mamdani model [39, 40]. In this model, defuzzification is performed by various methods, such as centroid of area, bisector of area, mean of maximum, smallest of maximum, and largest of maximum. These operations are consuming more time for its calculation. To overcome this issue, we use TS model which never requires defuzzification.

$$y = \frac{\sum_{i=1}^n W_i f_i(x_1, x_2, \dots, x_n)}{\sum_{i=1}^n W_i} \tag{5}$$

Consider the nonlinear system below:

$$\begin{pmatrix} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_1^2 + x_2^2 + y \end{pmatrix}$$

here x_1 and x_2 are premises variables and input variable limit are denoted as $x_1 \in [a, m_1, b]$ and $x_2 \in [c, m_2, d]$, where $a, b, c, d, m_1,$ and m_2 are literal constants. Common steps for two inputs fuzzy inference system are described as follows:

- Step 1 To determine the fuzzy variables and membership functions or fuzzy sets
 - Step 1.1 For simplicity, assume two fuzzy variables x_1 and x_2
 - Step 1.2 To acquire membership functions, we must calculate the minimum and maximum values of $x_1(t)$ and $x_2(t)$ where $x_1 \in [a, m_1, b]$ and $x_2 \in [c, m_2, d]$, and are clearly obtained as follows:

$$\begin{aligned} \text{lower } x_1(t) &= a, & \text{Middle } x_1(t) &= m_1 & \text{higher } x_1(t) &= b \\ \text{lower } x_2(t) &= c, & \text{Middle } x_2(t) &= m_2 & \text{higher } x_2(t) &= d \end{aligned}$$

Therefore, x_1 and x_2 can be represented by membership functions T_1, T_2, T_3 and H_1, H_2, H_3 , respectively, and are denoted by:

$$\begin{aligned} x_1(t) &= T_1(x_1(t))a + T_2(x_1(t))b + T_3(x_1(t))m_1 \\ x_2(t) &= H_1(x_2(t))c + H_2(x_2(t))d + H_3(x_2(t))m_2. \end{aligned}$$

- Step 2 Model rule evaluation
 - Rule 1 If $x_1(t)$ is high and $x_2(t)$ is high, then $\dot{x}(t) = A_1 x(t)$
 - Rule 2 If $x_1(t)$ is high and $x_2(t)$ is low, then $\dot{x}(t) = A_2 x(t)$
 - Rule 3 If $x_1(t)$ is high and $x_2(t)$ is middle, then $\dot{x}(t) = A_3 x(t)$
 - Rule 4 If $x_1(t)$ is low and $x_2(t)$ is high, then $\dot{x}(t) = A_4 x(t)$
 - Rule 5 If $x_1(t)$ is low and $x_2(t)$ is low, then $\dot{x}(t) = A_5 x(t)$
 - Rule 6 If $x_1(t)$ is low and $x_2(t)$ is middle, then $\dot{x}(t) = A_6 x(t)$
 - Rule 7 If $x_1(t)$ is middle and $x_2(t)$ is high, then $\dot{x}(t) = A_7 x(t)$
 - Rule 8 If $x_1(t)$ is middle and $x_2(t)$ is low, then $\dot{x}(t) = A_8 x(t)$
 - Rule 9 If $x_1(t)$ is middle and $x_2(t)$ is middle, then $\dot{x}(t) = A_9 x(t)$

Step 3 Defuzzification process of the system incorporates with $A_i x(t)$

$$A_i x(t) = \begin{bmatrix} 0 & m_1 \text{ or } m_2 & 1 \\ \max(\text{high}) & \text{mid}(\text{middle}) & \min(\text{low}) \end{bmatrix} \quad (6)$$

$$\dot{x}(t) = \sum_{i=1}^r u_i(x(t)) A_i x(t) \quad (7)$$

Step 4 Final outputs of fuzzy model evaluation

$$\dot{x}(t) = \sum_{i=1}^r u_i(x(t)) \{A_i x(t) + B_i p(t)\} \quad (8)$$

$$y(t) = \sum_{i=1}^r u_i(x(t)) C_i x(t) \quad (9)$$

where weighting function w_i is standardized as

$$u_i(x(t)) = \frac{w_i x(t)}{\sum_{i=1}^r w_i x(t)} \quad (10)$$

We think through some input variables from real dataset for evaluating the overall performance of our proposed methodology. Three linguistic variables are declared as low (L), middle (M), and high (H). Input variable membership functions are created by using above-said three linguistic variables and are used to analyze the sensing range of sensor's attributes. Figure 3 displays the temperature membership functions. Figure 4 depicts the humidity membership functions. Similarly, we can consider any input variable from the real datasets and its membership function is designed based on fuzzy limits. Figures 5, 6, 7, and 8 show voltage, ambient temperature, surface temperature, and relative humidity, respectively. Fuzzy rules are generated based on fuzzy limits that are specified in the membership functions.

Input values for our fuzzy inference system are the node distance from cluster head, time difference of each sensor readings, and input variable correlation. The fuzzy input

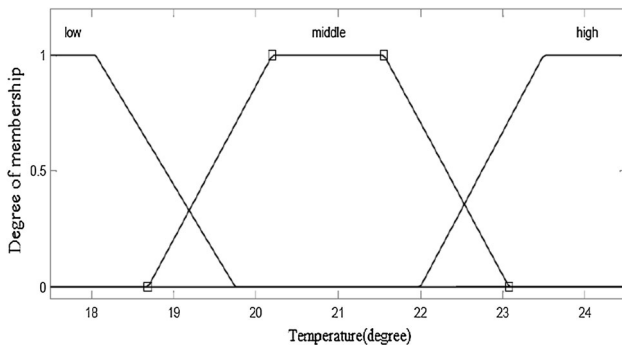


Fig. 3 Temperature membership function

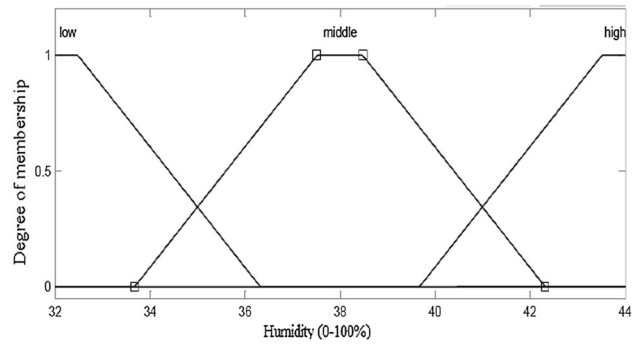


Fig. 4 Humidity membership function

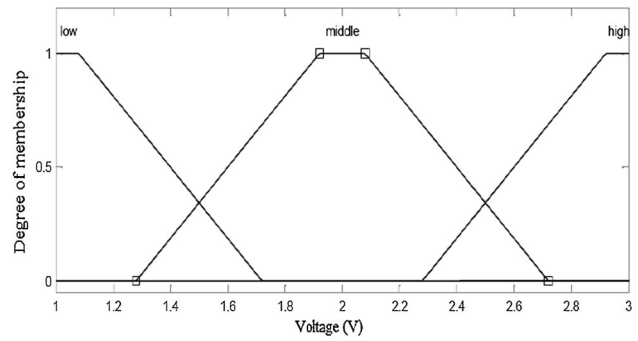


Fig. 5 Voltage membership function

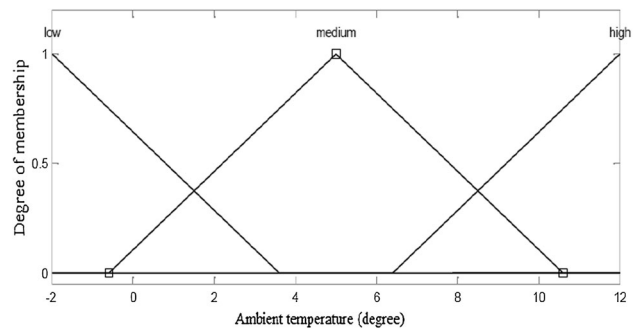


Fig. 6 Ambient temperature membership function

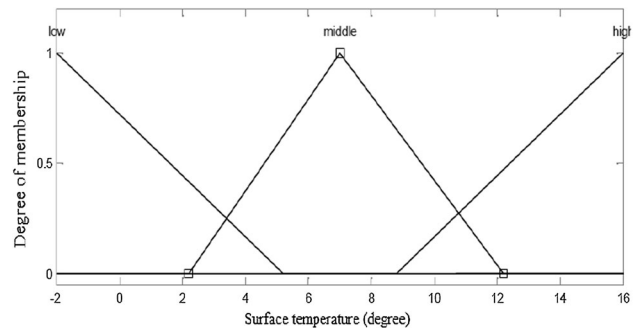


Fig. 7 Surface temperature membership function

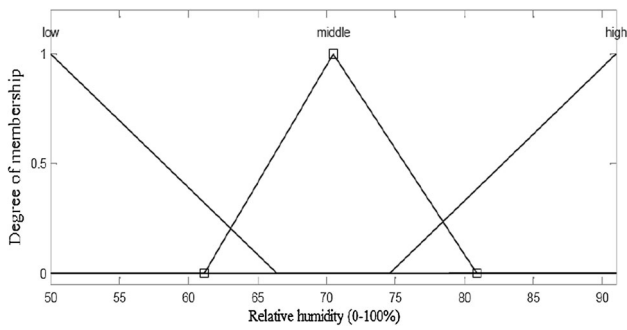


Fig. 8 Relative humidity membership function

distance is represented using triangular functions as shown in Fig. 9, where the linguistic terms *close*, *medium*, and *far* represent the range of contribution for that input. The fuzzy input time is denoted using triangular functions as shown in Fig. 10, where the terms *small*, *average*, and *long* represent the level of involvement for that input. The fuzzy input *variable correlation* is fuzzified using triangular functions as shown in Figs. 3, 4, 5, 6, 7, and 8, where the gears *low*, *middle*, and *high* represent the magnitude of participation for that input.

Distance, time, and variable correlation are affirmed as input variables for declaring result in anomaly detection system in Table 1. The anomaly detection confidence levels of the proposed system are classified as normal and anomaly. Normal data are directed to the cluster head or base station. Anomalous data are detached, and its corresponding imputed data are replaced by combined correlation analysis of particular sensor nodes [40, 41]. To finish, accurate imputed data are inserted and anomaly free dataset is analyzed by the cluster head for aggregation process.

Incomplete data are common in WSNs and may rise due to hardware glitches, packet collisions, signal strength diminishing, and environmental nosiness. These incomplete data are created manually by removing anomalous data. The removal of anomalous data at cluster head leads to missing data [42]. To solve this, fuzzy inference systems discover imputed data by analyzing spatial, temporal, and

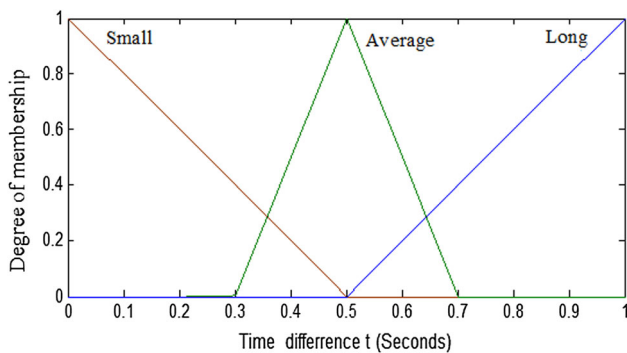


Fig. 9 Time membership function

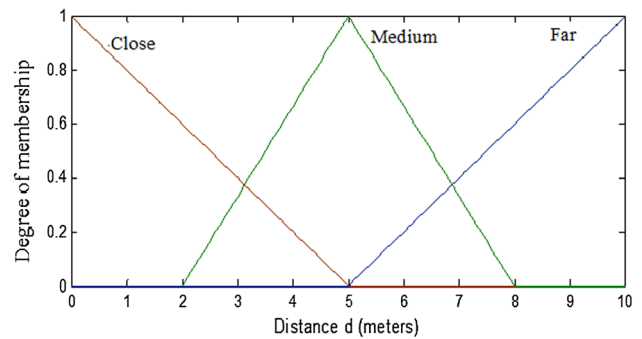


Fig. 10 Distance membership function

attribute correlation among the sensor nodes. Suppose sensor node X generates anomalous data and sensor node Y generates normal data and if the correlation between sensor nodes X and Y is high, then Y can impute data from X and vice versa. Apart considering from spatial distance between sensor nodes, time difference and variable correlation of two sensor nodes and its neighbor nodes are considered too.

5 Experimental Results

The proposed anomaly diagnosis system is tested by using both unreal and real datasets, and the results are obtained by incessant experiments. This is performed in terms of anomaly detection rate, specificity and false alarm rate for both unpolluted and polluted datasets. The proposed methodology is experimented with the help of datasets acquired from Intel Berkeley Research Lab (IBRL) [43] which uses 54 Mica2Dot sensors with 4 attributes (temperature, humidity, voltage, and light). Fifty-four sensor nodes were deployed in the laboratory between February 28, 2004, and April 5, 2004. The nodes collected approximately 2.3 million readings. The data in the dataset are selected at the time of exportation, namely March 2004 (30-day period) during the time interval 00:00 am to 03:59 am, and the dataset acquired from SensorScope project, located at the Grand-St-Bernard (GSB) pass at 2400 m between Switzerland and Italy [44] during the period of 2 months between September 2007 and October 2007. The 23 sensor nodes sense the environment with several attributes such as ambient temperature, humidity, soil moisture, wind direction, and wind speed accordingly.

5.1 Performance Evaluation

The performance of proposed methodology is usually calculated by a confusion matrix. An illustration of the confusion matrix is designed in Table 2 which depicts columns (predictable class) and rows (actual class). In the table, true

Table 1 Anomaly detection rule structure

| Rule [®] | Distance | Time | Variable correlation | Confidence level |
|-------------------|----------|---------|----------------------|------------------|
| 1 | CLOSE | SMALL | LOW | ANOMALY |
| 2 | CLOSE | SMALL | MIDDLE | ANOMALY |
| 3 | CLOSE | SMALL | HIGH | ANOMALY |
| 4 | CLOSE | AVERAGE | LOW | ANOMALY |
| 5 | CLOSE | LONG | LOW | ANOMALY |
| 6 | CLOSE | AVERAGE | MIDDLE | NORMAL |
| 7 | CLOSE | AVERAGE | HIGH | ANOMALY |
| 8 | CLOSE | LONG | MIDDLE | NORMAL |
| 9 | CLOSE | LONG | HIGH | ANOMALY |
| 10 | MEDIUM | SMALL | LOW | ANOMALY |
| 11 | MEDIUM | SMALL | MIDDLE | NORMAL |
| 12 | MEDIUM | SMALL | HIGH | ANOMALY |
| 13 | MEDIUM | AVERAGE | LOW | NORMAL |
| 14 | MEDIUM | LONG | LOW | ANOMALY |
| 15 | MEDIUM | AVERAGE | MIDDLE | NORMAL |
| 16 | MEDIUM | AVERAGE | HIGH | NORMAL |
| 17 | MEDIUM | LONG | MIDDLE | ANOMALY |
| 18 | MEDIUM | LONG | HIGH | NORMAL |
| 19 | FAR | SMALL | LOW | ANOMALY |
| 20 | FAR | SMALL | MIDDLE | ANOMALY |
| 21 | FAR | SMALL | HIGH | NORMAL |
| 22 | FAR | AVERAGE | LOW | ANOMALY |
| 23 | FAR | LONG | LOW | ANOMALY |
| 24 | FAR | AVERAGE | MIDDLE | NORMAL |
| 25 | FAR | AVERAGE | HIGH | NORMAL |
| 26 | FAR | LONG | MIDDLE | NORMAL |
| 27 | FAR | LONG | HIGH | NORMAL |

Table 2 Confusion matrix

| | Predictable negative | Predictable positive |
|-----------------|----------------------|----------------------|
| Actual negative | TN | FP |
| Actual positive | FN | TP |

negative (TN) signifies the number of normal data correctly classified, and true positive (TP) is the number of abnormal data properly classified. False positive (FP) is the number of normal data classified as abnormal, and false negative (FN) is the number of abnormal data classified as normal. Other common evaluation metrics are accuracy which is a measure of the proposed system to predict the anomaly correctly. Sensitivity or recall (detection rate) is a degree of a system to detect positive abnormal cases and specificity is the ability of the system to spot negative normal cases. False alarm rate (FAR) is the capability of the system to detect positive normal cases. The evaluation metrics are described with the help of the confusion matrix. Positive projection rate (PPR) or precision is defined as the

proportion of positive test results that are true positives. The measures are described below:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} & \text{Sensitivity} &= \frac{TP}{TP + FN} \\
 \text{Specificity} &= \frac{TN}{TN + FP} & \text{False Alarm Rate} &= \frac{FP}{FP + TN} \\
 \text{Positive Projection Rate} &= \frac{TP}{TP + FP}
 \end{aligned}$$

5.2 Estimation on Datasets

For assessing the proposed method, the datasets are normalized by identifying and removing extreme values initially. Scatter plot and Chi-square tests are used to prompt cleaned data as customary data. Initially, data are cleaned manually by identifying extreme values and missing values. Most of the values are either missing or damaged. The rest of the data is labeled as normal for evaluations. Specifically, in IBRL dataset, 30-day period between March 1, 2004, and March 30, 2004, is considered as normal data. On completion, two sets of data are considered for evaluation such as normal data without corruption

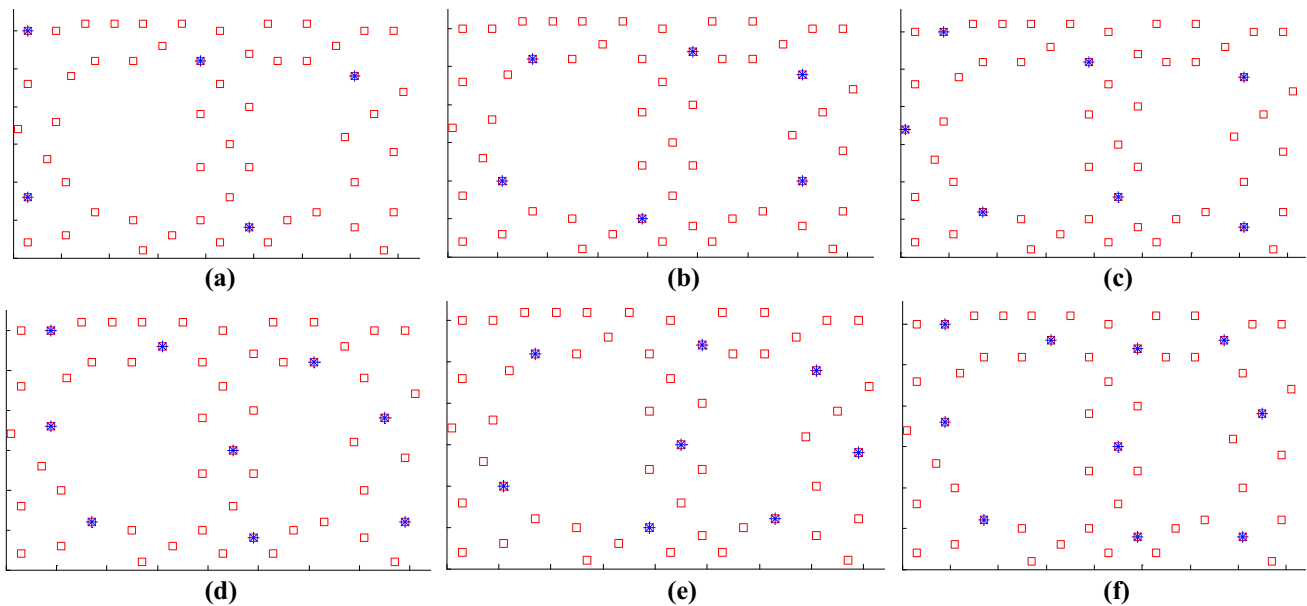


Fig. 11 Subtractive clustering in IBRL **a** 5 clusters, **b** 6 clusters, **c** 7 clusters, **d** 8 clusters, **e** 9 clusters, **f** 10 clusters

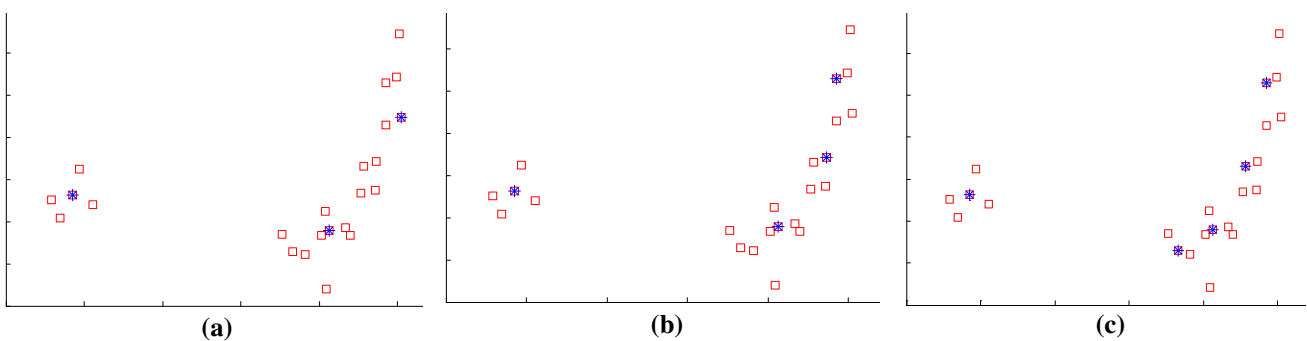
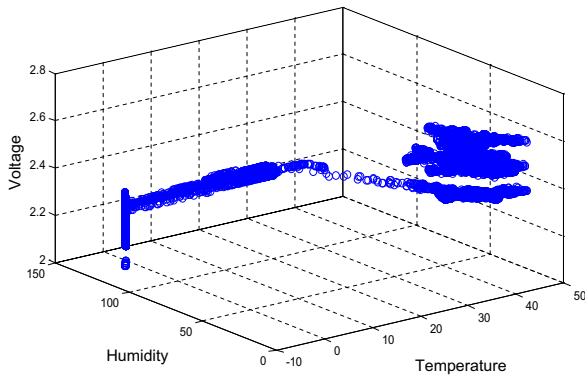


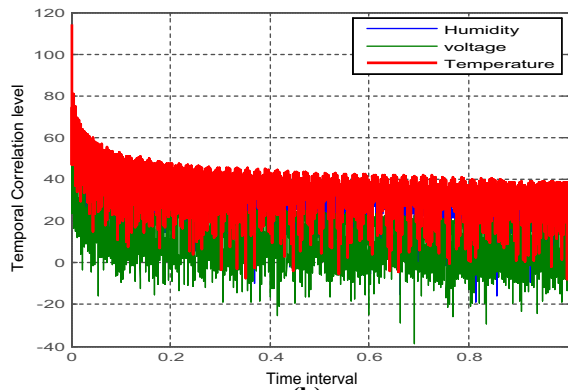
Fig. 12 Subtractive clustering in GSB **a** 3 clusters, **b** 4 clusters, **c** 5 clusters

ratio and normal data with corruption ratio. For this purpose, random sets of corrupted data are injected at various clusters. Anomalies were interleaved often in one or many nodes in each cluster at the frequency of data corruption between 5 and 70 %. We implement the proposed algorithm using the MATLAB package for a simulated wireless sensor node on the two datasets generated already. On visualizing the IBRL dataset, the proposed system is smeared with several numbers of clusters ranging from 5 to 10 from subtractive clustering algorithm for finding optimal number of clusters in the dataset. Figure 11 portrays the number of clusters used for evaluating anomaly detection technique from IBRL. By eagle-eyeing the GSB dataset, the subtractive clustering is applied for gathering optimal number of clusters oscillating from 3 to 5. Figure 12 displays the finest cluster used for assessing anomaly detection technique in GSB.

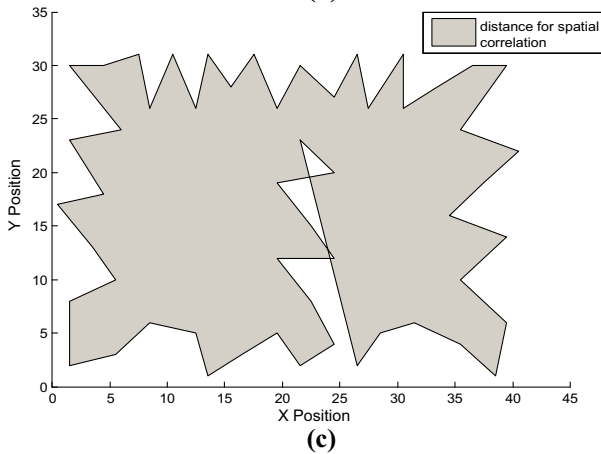
Each cluster reading is experimented by using TS fuzzy inference system with fuzzy rules. The fuzzy results are sectored into normal and abnormal cases. Abnormal data are removed, and its corresponding imputed data are predicted based on fuzzy rules with correlation level of sensor nodes. By applying subtractive clustering, we obtain several numbers of optimal clusters ranging from low to high in the IBRL dataset. At least 19 and to a maximum of 75 iterations are performed with optimal accept and reject ratio yielding suitable cluster centers. Gradual increase in number of clusters increases computational complexity and communicational complexity. Inadequate number of clusters could not analyze anomalous data optimally. By considering both datasets, optimal clusters range is selected based on objective function of subtractive clustering. The computational complexity of subtractive clustering is varied based on number of clusters. Thus, computational



(a)



(b)

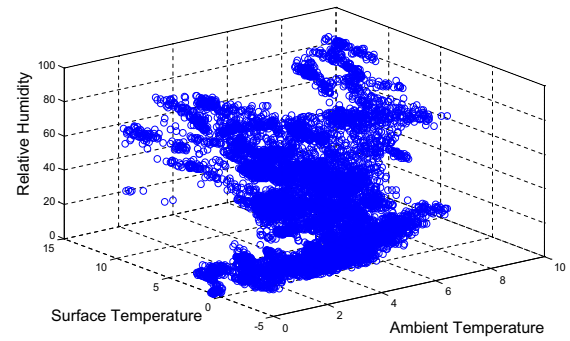


(c)

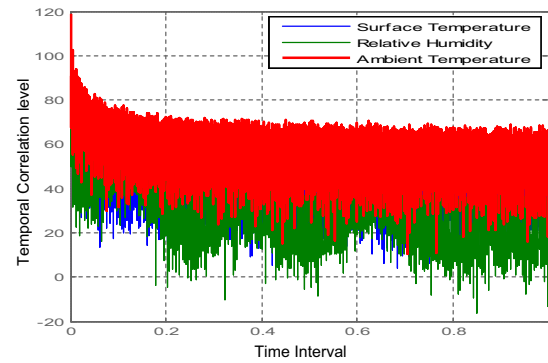
Fig. 13 Correlation analysis of IBRL dataset without corruption **a** attribute correlation, **b** temporal correlation, **c** spatial correlation

complexity is denoted by $O(ns)$ where n is the number of iterations and s is the number of nodes. The memory space requirement for the proposed approach is $O(I)$ for storing the value of cluster center.

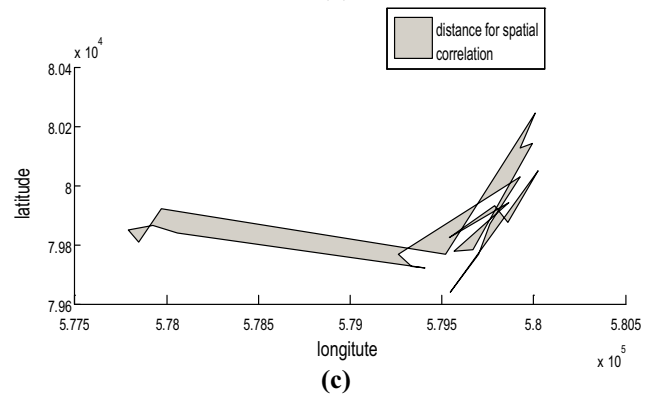
Generally, multivariate analysis technique is used to detect outliers based on correlation, i.e., by identifying relationship among the variables that are participated in the outlier detection process. Multivariate outlier deviates from the usual correlation structure in multi-dimensional space defined by the variables. Figures 13 and 14 gives a brief



(a)



(b)



(c)

Fig. 14 Correlation analysis of GSB dataset without corruption **a** attribute correlation, **b** temporal correlation, **c** spatial correlation

depiction of attribute correlation, temporal and spatial correlation involved in the evaluation using the real dataset without any corruption ratio. Three attributes (temperature, humidity, and voltage) and other attributes (ambient temperature, surface temperature, and relative humidity ranging from 0 to 100 %) are multivariate and are normalized from IBRL and GSB datasets, respectively. To show spatial correlation, we have calculated the correlation coefficients between the two sensors at each observation based on distance measure, where Mahalanobis distance [27] is applied for estimating nearest less spacious sensor nodes. Temporal correlation is illustrated with different time intervals at time $t - 1$ and t of a particular sensor node.

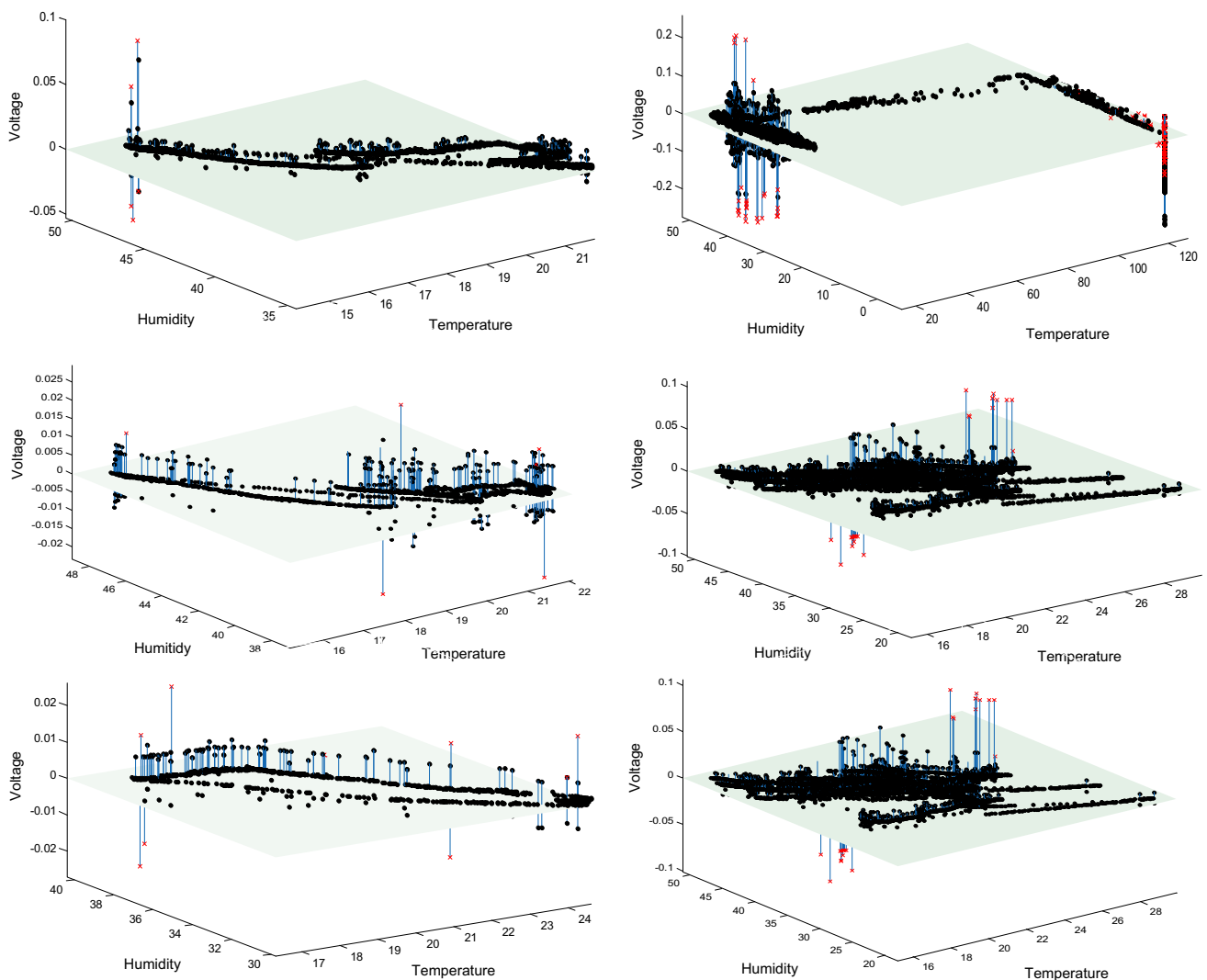


Fig. 15 Anomaly detection variation of IBRL dataset with corruption level. Six cases are considered and evaluated by fuzzy inference engine. Lines with times symbol indicates anomaly data, lines with

bullets indicates normal data. Five clusters in 1st left row, 6 clusters in 2nd left row, 7 clusters in 3rd left row, 8 clusters in 1st right row, 9 clusters in 2nd right row, 10 clusters in 3rd right row

Additionally, attribute correlation illustrates internal relationship among the physical phenomena attributes.

Figure 15 displays data distribution based on IBRL dataset with three physical phenomena attributes. Anomalous data are randomly generated at various clusters. To provide baseline for our results, we perform fuzzy-value experiments with TS fuzzy model. Initially, data corruption level is increased from 6 to 85 %. Figure 15 shows the result of predicted anomaly data at 15 % corruption level. Anomalous data are detected and isolated from clusters ranging from 5 to 10. Detection accuracy is not degraded while increasing number of clusters. Henceforth, our

proposed fuzzy inference system detects anomaly with high accuracy based on fuzzy rules.

Figure 16 shows data dissemination based on GSB dataset. The three physical phenomena attributes disseminations with anomalous data are randomly generated at various clusters. This illustration shows the result of predicted anomaly data at 19 % corruption level. Anomalous data are detected and separated for various numbers of cluster ranging from 3 to 5. Figures 15 and 16 conclude most of the deviation in data is accurately identified by the correlation analysis among the attributes; fuzzy inference rules are framed along with this correlation exposing high

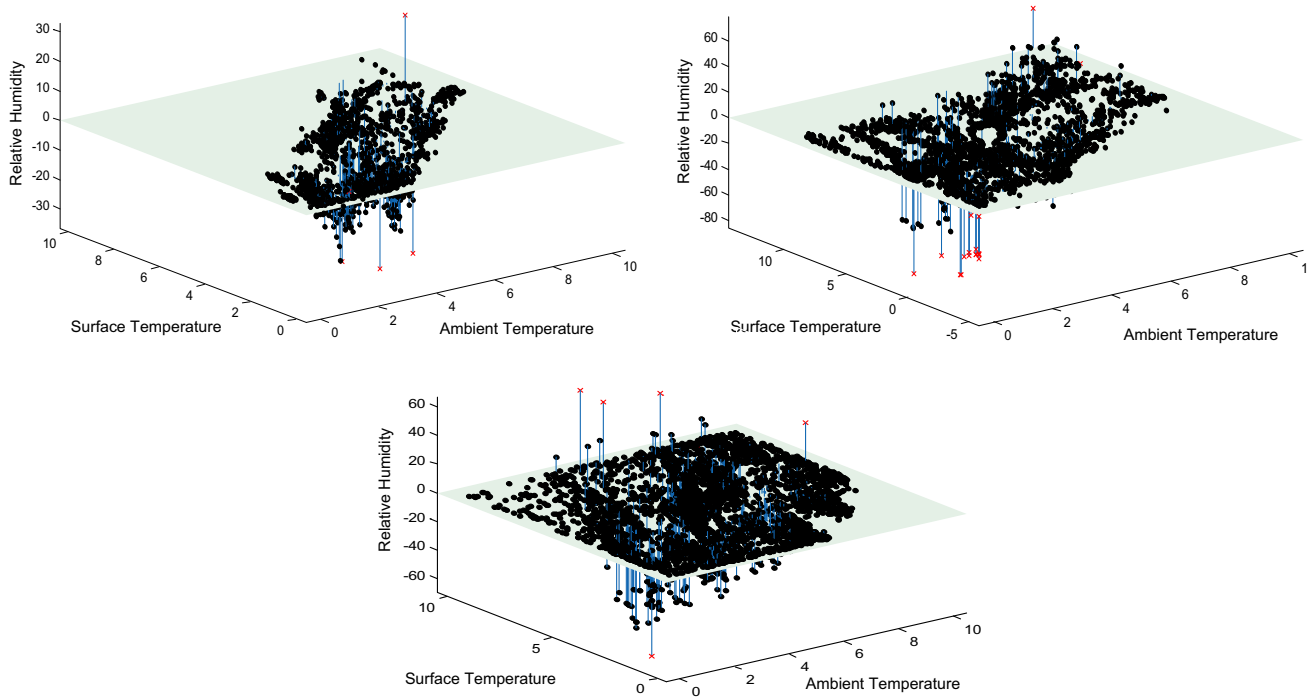


Fig. 16 Anomaly detection variation of GSB dataset with corruption level. Three cases are considered and evaluated by fuzzy inference engine. Lines with times symbols indicates anomaly data, lines with bullets indicates normal data. Three clusters in top left corner, 4 clusters in top right, 5 clusters in 2nd row

Table 3 Performances of proposed system with <6 % corruption level

| | IBRL dataset | | | | | | GSB dataset | | |
|-------------|--------------|------------|------------|------------|------------|-------------|-------------|------------|------------|
| | 5 clusters | 6 clusters | 7 clusters | 8 clusters | 9 clusters | 10 clusters | 3 clusters | 4 clusters | 5 clusters |
| Accuracy | 99.75 | 99.55 | 99.51 | 99.84 | 99.96 | 99.11 | 99.08 | 99.59 | 99.5 |
| Sensitivity | 98.14 | 98.78 | 98.85 | 98.35 | 98.00 | 98.29 | 98.58 | 98.11 | 98.02 |
| Specificity | 99 | 98.25 | 99.08 | 97.51 | 97.92 | 98.77 | 98.12 | 98.55 | 97.93 |
| PPR | 97.14 | 99.19 | 99.05 | 99.6 | 99.54 | 99.33 | 99.00 | 99.57 | 99.28 |
| FAR | 1.00 | 2.65 | 1.02 | 2.51 | 2.08 | 1.33 | 2.98 | 2.45 | 2.07 |

Table 4 Performances of proposed system with 6–25 % corruption level

| | IBRL dataset | | | | | | GSB dataset | | |
|-------------|--------------|------------|------------|------------|------------|-------------|-------------|------------|------------|
| | 5 clusters | 6 clusters | 7 clusters | 8 clusters | 9 clusters | 10 clusters | 3 clusters | 4 clusters | 5 clusters |
| Accuracy | 100 | 100 | 100 | 100 | 99.89 | 100 | 100 | 100 | 100 |
| Sensitivity | 98.14 | 98.78 | 98.85 | 98.35 | 98.00 | 98.29 | 98.58 | 98.11 | 98.02 |
| Specificity | 99.99 | 99.88 | 99.75 | 99.79 | 99.49 | 99.89 | 99.85 | 99.77 | 99.45 |
| PPR | 99.74 | 99.19 | 99.05 | 99.6 | 99.54 | 99.33 | 99.00 | 99.57 | 99.28 |
| FAR | 0.1 | 0.12 | 0.25 | 0.21 | 0.51 | 0.11 | 0.15 | 0.23 | 0.55 |

detection rate. The fuzzy inference engine plays an important role in the reduction of FAR by increasing the detection rate. Fuzzy inference engine that incorporates

with sensors location and its correlation information generated by the correlation analysis accurately classifies the anomalous and normal data and has high sensitivity with

Table 5 Performances of proposed system with 26–45 % corruption level

| | IBRL dataset | | | | | | GSB dataset | | |
|-------------|--------------|------------|------------|------------|------------|-------------|-------------|------------|------------|
| | 5 clusters | 6 clusters | 7 clusters | 8 clusters | 9 clusters | 10 clusters | 3 clusters | 4 clusters | 5 clusters |
| Accuracy | 100 | 100 | 99.9 | 99.9 | 100 | 100 | 100 | 100 | 100 |
| Sensitivity | 98.98 | 98.58 | 98.22 | 98.02 | 98.05 | 98.55 | 98.9 | 98.23 | 98.05 |
| Specificity | 99.5 | 99.2 | 99.2 | 99.6 | 99.1 | 99.6 | 99.55 | 99.32 | 99.26 |
| PPR | 99.89 | 99.70 | 99.51 | 99.87 | 99.32 | 99.31 | 99.55 | 99.54 | 99.92 |
| FAR | 0.5 | 0.8 | 0.2 | 0.4 | 0.9 | 0.4 | 0.45 | 0.68 | 0.74 |

Table 6 Performances of proposed system with 46–65 % corruption level

| | IBRL dataset | | | | | | GSB dataset | | |
|-------------|--------------|------------|------------|------------|------------|-------------|-------------|------------|------------|
| | 5 clusters | 6 clusters | 7 clusters | 8 clusters | 9 clusters | 10 clusters | 3 clusters | 4 clusters | 5 clusters |
| Accuracy | 99.93 | 99.91 | 99.94 | 99.9 | 99.84 | 99.95 | 99.91 | 99.75 | 99.69 |
| Sensitivity | 98.56 | 98.64 | 98.65 | 98.66 | 98.9 | 99 | 99.5 | 99.12 | 98.65 |
| Specificity | 99.65 | 99.05 | 99.95 | 99.68 | 99.1 | 98.04 | 98.35 | 98.91 | 98.55 |
| PPR | 99.95 | 99.65 | 99.48 | 99.65 | 99.23 | 99.02 | 99.19 | 99.65 | 99.87 |
| FAR | 1.35 | 1.95 | 1.05 | 1.32 | 1.9 | 2.96 | 2.65 | 2.09 | 2.65 |

Table 7 Performances of proposed system with 66–85 % corruption level

| | IBRL dataset | | | | | | GSB dataset | | |
|-------------|--------------|------------|------------|------------|------------|-------------|-------------|------------|------------|
| | 5 clusters | 6 clusters | 7 clusters | 8 clusters | 9 clusters | 10 clusters | 3 clusters | 4 clusters | 5 clusters |
| Accuracy | 99.65 | 99.25 | 99.41 | 99.70 | 99.06 | 99.05 | 99.15 | 99.07 | 99.23 |
| Sensitivity | 98.09 | 98.54 | 98.73 | 98.39 | 98.12 | 98.15 | 98.45 | 98.04 | 98.01 |
| Specificity | 99.94 | 99.95 | 99.75 | 99.11 | 99.92 | 99.46 | 99.16 | 98.99 | 99.94 |
| PPR | 99.45 | 99.45 | 99.12 | 99.11 | 99.21 | 99.13 | 98.89 | 99.47 | 99.65 |
| FAR | 1.06 | 1.05 | 1.25 | 1.89 | 1.08 | 1.54 | 1.84 | 2.01 | 1.06 |

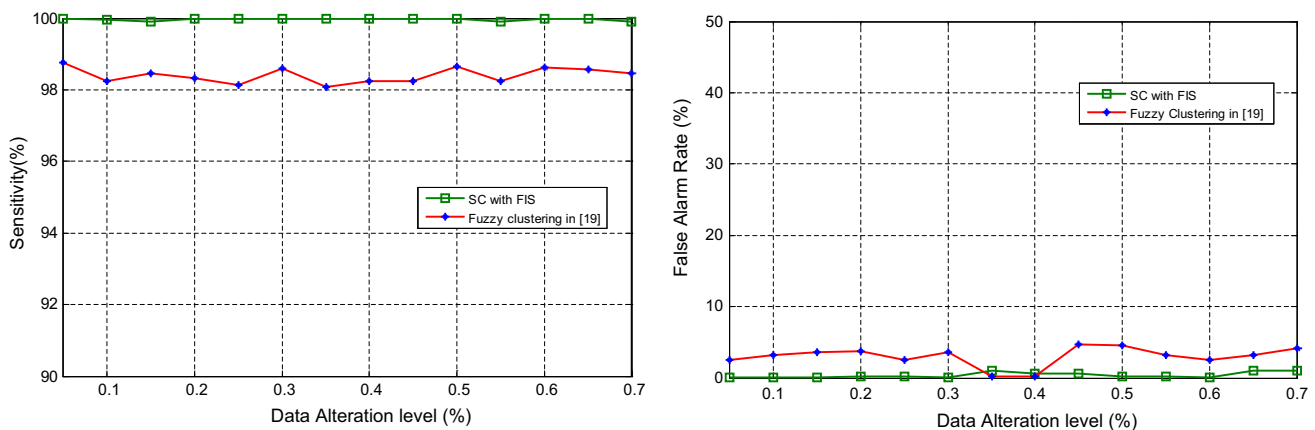


Fig. 17 Sensitivity and false alarm rate for altering corruption level in IBRL dataset

less false alarm rate. The results for the IBRL and GSB dataset with different range of corruption levels are given in Tables 3, 4, 5, 6, and 7.

Efficient output is ensued from 6 clusters of the IBRL dataset and 3 clusters of the GSB dataset having very high accuracy with less false alarm rate over the range of

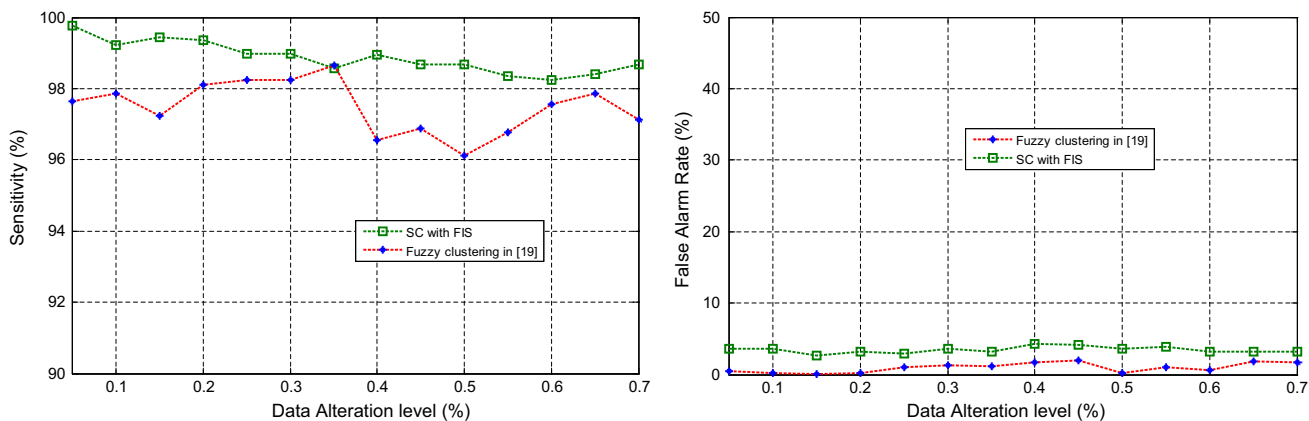


Fig. 18 Sensitivity and false alarm rate for altering corruption level in GSB dataset

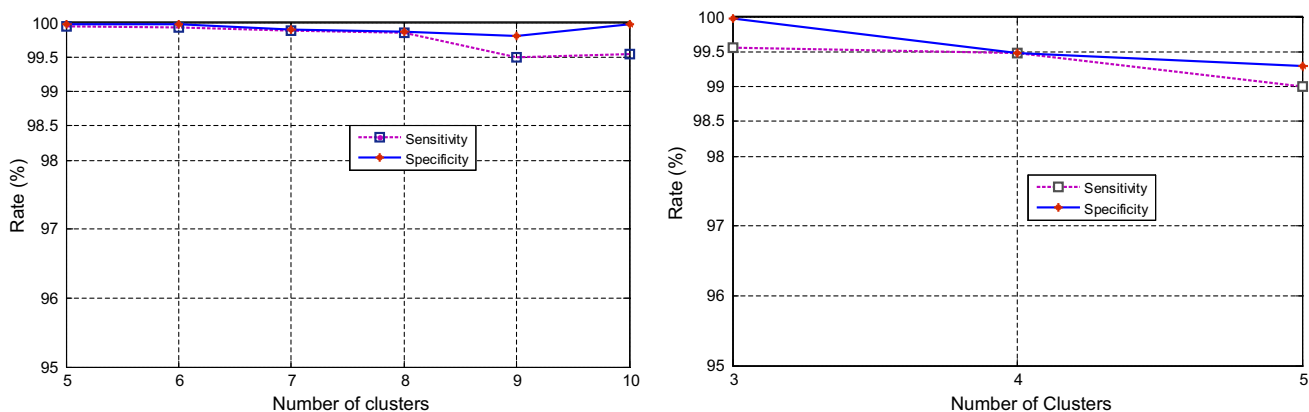


Fig. 19 Comparison of sensitivity and specificity variation by varying cluster size for dataset based on IBRL(left) and GSB(right)

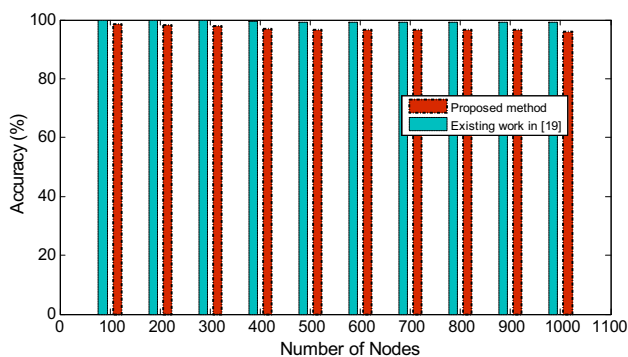


Fig. 20 Scalability comparison

clusters from 5 to 10 and 3 to 5. Corruption level range is formed into four groups, 6–25 %, 26–45 %, 46–65 %, and 66–85 %. Table 3 shows the performance of proposed methodology with less than 6 % corruption level. It is fascinated tough to predict anomaly below 6 % since the deviation from original and anomalous data is less. Percentage of prediction is increased at the various ranges of corruption levels above 6 % as shown in Tables 4, 5, 6 7.

Specifically, the proposed method considers fuzzy logic offering of 1 % false alarm and 99.87 % detection rate till 65 % of the nodes in the network. Even for corruption level between 66 and 85 %, the average false alarm created is simply 1.52 % and 1.84 in IBRL and GSB dataset, respectively.

Figures 17 and 18 illustrate the performance of the proposed system for evaluating IBRL and GSB dataset. It is observed that the approach in [19] has less sensitivity and high false alarm rate compared to subtractive clustering (SC) with fuzzy inference system (FIS). The proposed method offers 0 % false alarm rate and 100 % detection rate till 45 % of the nodes in the network are found to be anomalous. Even for corruption level between 46 and 85 %, the average false alarm created is 0.956 % in IBRL and 2.56 % in GSB dataset, respectively. Therefore, it is clear that the proposed system achieves significant improvement in detecting accuracy compared to the conventional fuzzy technique in [19].

The results are represented in Fig. 19 for comparison with the variation in sensitivity and specificity obtained for

Table 8 Complexity analysis of anomaly detection

| Techniques | Computational complexity | Communication complexity | Memory complexity |
|--------------------------------|--------------------------|--------------------------|----------------------|
| SC with FIS | $O(ndc)$ | $O(nd)$ | $O(ndr)$ |
| Fuzzy clustering in [19] | $O(nd^2)$ | $O(nvd)$ | $O(nvd^2)$ |
| Ellipsoidal SVM based | $O(nd^2)$ | $O(nvd^2)$ | $O(nvd^4)$ |
| Statistical based | $O(n(n-1)d + pq)$ | $O(nd^2)$ | $O(nd^2 + (n-1)d^2)$ |
| Agglomerative clustering based | $O(nd^6c)$ | $O(nd^3)$ | $O(nd^4)$ |

our proposed method on the same datasets. The graphs illustrate the proposed method preserves a very high degree of accuracy in defining normal data with the false alarm rate being between 0 % and 0.5 % in all occurrences. However, while the sensitivity values for the proposed scheme are more or less retained at an acceptable range between 99 % and 99.5 % for GSB data and a high 98–98.5 % for IBRL, the results for varying number of clusters gradually reach sustainable sensitivity rate. Therefore, it is evident that the proposed method achieves significant gains in detection accuracy.

Figure 20 depicts the final performance of our proposed methodology with high accuracy, while increasing number of nodes (clusters) by using unreal dataset, which is simulated for analyzing the scalability of the proposed methodology. As shown in Fig. 20, anomaly detection and misdetection fractions are attractively stable while increasing numbers of nodes from 100 to 1000. At the same time, large numbers of nodes yield moderate computational and communication complexity with the help of subtractive clustering with slight increase in space complexity. This result implies that our subtractive clustering with fuzzy logic-based anomaly detection has very fastidious scalability as it works well under different network sizes.

Table 8 explains the complexity of different state-of-the-art anomaly detection approaches. To understand the performance of the proposed approach, it is essential to relate it with state-of-the-art anomaly detection techniques. To evaluate the efficiency of the SC with FIS model, the computational complexity, communication overhead and memory complexity are considered. The computational complexity incurred by our model is $O(ndc)$ related to the calculation of n number of new of records at time, d -dimension of the observations, and c number of clusters. The communication overhead is $O(nd)$, and memory complexity is $O(ndr)$, where r represents number of rules. Table 8 shows v - intermediate values, p -spatial correlation, and q - temporal correlation in the remaining techniques. Less number of rules saves more memory space. Our method proves high accuracy compared to other available peer methods. Computational complexity, communication complexity, and memory complexity are slightly reduced when compared to other techniques.

6 Conclusion

In this paper, a system which employs fuzzy-based anomaly detection is industrialized using fuzzy logic to classify anomalous and normal data. Two real datasets are evaluated with various numbers of clusters generated by subtractive clustering. After cataloging of data, normal data are labeled as customary and incorrect readings as anomaly. The overall accuracy and reliability are increased by imputing data. The experimental result proves that the proposed anomaly diagnosis and relief measures model outperforms existing work in various aspects like anomaly detection accuracy, false alarm, sensitivity, and specificity in decision-making support.

References

- Akyildiz, I.F., Su, W., Sankara subramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. *Comput. Netw.* **38**, 393–422 (2002)
- Dargie W., Poellabauer C. In: Shen X., Pen D.Y. (eds.), *Fundamentals of Wireless Sensor Networks*. 3rd edn, Wiley (2010)
- Xie, Miao, Han, Song, Tian, Biming, Parvin, Sazia: Anomaly detection in wireless sensor networks: a survey. *J. Netw. Comput. Appl.* **34**, 1302–1325 (2011)
- Sun, Bo, Shan, Xuemei, Kui, Wu, Xiao, Yang: Anomaly detection based secure in-network aggregation for wireless sensor networks. *IEEE Syst. J.* **7**(1), 13–25 (2013)
- Roy, S., Conti, M., Setia, S., Jajodia, S.: Secure data aggregation in wireless sensor networks. *IEEE Inf. Forens. Secur.* **7**(3), 1040–1052 (2012)
- Mitchell, Robert, Chen, Ing-Ray: A survey of intrusion detection in wireless network applications. *Comput. Commun.* **42**, 1–23 (2014)
- Xu, H., Huang, L., Zhang, Y., Huang, H., Jiang, S., Liu, G.: Energy efficient cooperative data aggregation for wireless sensor networks. *J. Parallel Distrib. Comput.* **70**(9), 953–961 (2010)
- Forero, P., Cano, A., Giannakis, G.: Distributed clustering using wireless sensor networks. *IEEE J. Sel. Top. Signal Process.* **5**(4), 702–724 (2011)
- Takagi, T., Sugeno, M.: Fuzzy Identification of systems and its applications to modeling and control. *IEEE Trans. Syst., Man, Cybern.* **15**(1), 116–132 (1985)
- O'Reilly, C., Gluhak, A., Imran, M.A., Rajasegarar, S.: Anomaly detection in wireless sensor networks in a non-stationary environment. *IEEE Commun. Surv. Tutor.* **16**(3), 1–20 (2013)
- Pottie, G.J., Kaiser, W.J.: Wireless integrated network sensors. *ACM Commun.* **43**(5), 51–58 (2000)

12. Chitra Devi, N., Palanisamy, V., Baskaran, K., Prabeela, S.: Efficient distributed clustering based anomaly detection algorithm for sensor stream in clustered wireless sensor network. *Eur. J. Sci. Res.* **54**(4), 484–498 (2011)
13. Zhang, Yang, Meratnia, Nirvana, Havinga, P.J.M.: Distributed online outlier detection in wireless sensor networks using ellipsoidal support vector machine. *J. Ad Hoc Netw.* **11**, 1062–1074 (2012)
14. Zhang, Y., Hamm, N.A.S., Meratnia, N., Stein, A., Van de Voort, M., Havinga, P.J.M.: Statistics based outlier detection for wireless sensor networks. *Int. J. Geogr. Inf. Sci.* 1–20 (2011)
15. Kapitanova, K., Son, S.H., Kang, K.-D.: Using fuzzy logic for robust event detection in wireless sensor networks. *J. Ad Hoc Netw.* **10**, 709–722 (2011)
16. Liang Q, Wang L.: Event detection in wireless sensor networks using fuzzy logic system. In: *International Conference on Computational Intelligence for Homeland Security and Personal Safety*, IEEE, pp. 52–55 (2005)
17. Sasikala, E., Rengarajan, N.: An intelligent technique to detect jamming attack in wireless sensor networks (WSNs). *Int. J. Fuzzy Syst.* **7**(1), 76–83 (2015)
18. Shamshirband, S., Amini, A., Anur, N., Kiah, M., Teh, Y., Furnell, S.: D-FICCA: a density based fuzzy imperialist competitive clustering algorithm for intrusion detection in wireless sensor networks. *J. Meas. Elsevier* **55**, 212–226 (2014)
19. Kumaragea, Heshan, Khalil, Ibrahim, Tari, Zahir, Zomaya, Albert: Distributed anomaly detection for industrial wireless sensor networks based on fuzzy data modeling. *J. Parallel Distrib. Comput.* **73**, 790–806 (2013)
20. Barakkath Nisha, U., Maheswari, N.U., Venkatesh, R., Yasir Abdullah, R.: Robust estimation of incorrect data using relative correlation clustering technique in wireless sensor networks. In: *IEEE International Conference on Communication and Network Technologies*, Issue 1, pp. 314–318 (2014)
21. Kalman, R.E.: A new approach to linear filtering and prediction problems. *Trans. ASME-J. Basic Eng.* **82**, 35–45 (1960)
22. Yang, H., Jiang, B., Staroswiecki, M.: Observer-based fault-tolerant control for a class of switched nonlinear systems. *IET Control Theory Appl.* **5**, 1523–1532 (2007)
23. Yang, H., Cocquemot, V., Jiang, B.: Robust fault tolerant tracking control with application to hybrid nonlinear systems. *IET Control Theory Appl.* **3**(2), 211–224 (2009)
24. Huang, S., Tan, K.K., Lee, T.H.: Fault diagnosis and fault-tolerant control in linear drives using the Kalman filter. *IEEE Trans. Ind. Electron.* **59**(11), 4285–4292 (2012)
25. Chen, Shui-Li, Fang, Yuan, Yun-Dong, Wu: A new hybrid fuzzy clustering approach to Takagi-Sugeno fuzzy modeling. *Int. J. Digital Content Technol. Appl.* **6**(18), 341–350 (2012)
26. Afifi, W.A., Hefny, H.A.: Adaptive TAKAGI-SUGENO fuzzy model using weighted fuzzy expected value in wireless sensor network. In: *International Conference on Hybrid Intelligent Systems (HIS)*, IEEE, pp. 221–231 (2014)
27. Chen, J.-J., FAN, X.-P., QU, Z.-H., YANG, X., LIU, S.-Q.: Subtractive clustering based clustering routing algorithm for wireless sensor networks. *Inf. Control* **7**, 201–219 (2008)
28. Lizhe, Yu., Tiaojuan, Ren, Zhangquan, Wang, Banteng, Liu: Research on vehicle networking clustering routing algorithm based on subtractive clustering. *Appl. Mech. Mater.* **644–650**, 2366–2369 (2014)
29. Barakkath Nisha, U., Uma Maheswari, N., Venkatesh, R., Yasir Abdullah, R.: Improving data accuracy using proactive correlated fuzzy system in wireless sensor networks. *KSII Trans. Internet Inf. Syst.* **9**(9), 3515–3537 (2015)
30. Neamatollahi, P., Mashhad I., Taheri H., Naghibzadeh M., Yaghmaee M.: A hybrid clustering approach for prolonging lifetime in wireless sensor networks. *IEEE International Symposium on Computer Networks and Distributed Systems*, pp. 170–174 (2011)
31. Chiu, S.: Fuzzy model identification based on cluster estimation. *J. Intell. Fuzzy Syst.* **2**, 267–278 (1994)
32. Kim, D.W., Lee, K.Y., Lee, D., Lee, K.H.: A Kernel-based subtractive clustering method. *Pattern Recognit. Lett.* **26**, 879–891 (2005)
33. Nikhil, R.P., Chakraborty, D.: Mountain and subtractive clustering method: improvements and generalizations. *Int. J. Intell. Syst.* **15**, 329–341 (2000)
34. Yager, R.R., Filev, D.P.: Approximate clustering via the mountain method. *IEEE Trans. Syst., Man Cybern.* **24**(8), 1279–1284 (1994)
35. De Maesschalck, R., Jouan-Rimbaud, D., Massart, D.L.: The Mahalanobis distance. *J. Chemo Metrics Intell. Lab. Syst. Elsevier* **50**(1), 1–18 (2000)
36. Jang, J.-S.R., Sun, C.-T., Mizutani, E.: *Neuro-Fuzzy and Soft Computing*, 3rd edn. Prentice hall, Upper Saddle River (1997)
37. Sugeno, M., Kang, G.: Structure identification of fuzzy model. *Fuzzy Sets Syst.* **28**(1), 15–33 (1988)
38. Zadeh, L.A.: *Soft computing and fuzzy logic*. *ACM J. Softw.* **11**(6), 48–56 (1994)
39. Zimmermann, H.J.: *Fuzzy Set Theory and Its Applications*, 3rd edn. Publisher kluwer Academic Publishers Norwell, Norwell (1996)
40. Vuran, M.C., Akan, B., Akyildiz, I.F.: Spatio-temporal correlation: theory and applications for wireless sensor networks. *Comput. Netw. Int. J. Comput. Telecommun. Netw.* **45**(3), 245–259 (2004)
41. Liu, Z., Xing, W., Zeng, B., Wang, Y., Lu, D.: Distributed spatial correlation-based clustering for approximate data collection in WSNs. In: *IEEE International Conference on Advanced Information Networking and Applications*, pp. 56–63 (2013)
42. Ishibuchi, H., Nakashima, T., Kuroda, T.: A hybrid fuzzy GBML algorithm for designing compact fuzzy rule-based classification systems. In: *IEEE International Conference on Fuzzy Systems*, pp. 248–252 (1999)
43. <http://db.csail.mit.edu/labdata/labdata.html>
44. <http://lcav.epfl.ch/page-86035-en.html>



Usman Barakkath Nisha had received her Bachelor of Engineering in Computer Science and Engineering in 2008 with distinction from Anna University, Chennai. She got her Master of Engineering in Computer Science and Engineering from Anna University, Coimbatore, in 2010 as full-time student with distinction. Currently, she is working as assistant professor in Department of Computer Science and Engineering in PSNA College of Engineering of Technology,

Dindigul. Her research interests are ad hoc networks, wireless networks, information security, computer networks, sensor networks, etc. She has published 7 papers in international journals and presented 15 papers in various national, international conferences related to wireless sensor networks for the past 6 years



Natarajan Uma Maheswari received her M.E. in Computer Science and Engineering from the Madras University, Chennai, India, in 2002 and Ph.D. in Information and Communication Engineering in 2011 from Anna University, Chennai. Currently, she is working as a Professor in the Department of Computer Science and Engineering at the PSNA College of Engineering and Technology, Dindigul, India. She has totally 15 years of teaching experience

which includes 11 years of research experience. Her research interests include biometrics, image processing, compiler design, artificial intelligence, speech processing, and wireless sensor networks. She has published 20 papers in international journals, 2 papers in national journals, and presented 22 papers at international conferences, and 10 papers at national conferences. She has co-authored a book entitled “Compiler Design” published by Yes Dee Publishing. She is a recognized Ph.D. supervisor in Anna University in the area of image processing, cloud computing, network security, and networks



Ramalingam Venkatesh received his M.E. in Computer Science and Engineering from Anna University Chennai, India, in 2007 and Ph.D. in Computer Science and Engineering in 2010 at Alagappa University, Karaikudi. Currently, he is working as a Professor in the Department of Information Technology in PSNA College of Engineering and Technology, Dindigul, India. He has totally twenty years of teaching experience which includes 11 years

of research experience. He has published 25 papers in international

journals, 2 papers in national journals and presented 22 papers at international conferences and 10 papers at national conferences. His research interests include biometrics, artificial intelligence, compiler design, neural networks, soft computing, network security, and networks. He has authored a book entitled “Compiler Design” published by Yes Dee Publishing



Rabi Yasir Abdullah received his Bachelor of Engineering in Electronics and Communication Engineering in 2005 from Anna University, Chennai. He got his Master of Engineering in Computer Science as full-time student with distinction by Anna University, Coimbatore, Tamil Nadu, India, in 2009. Currently, he is working as assistant professor in Computer Science Department, SSCET, Palani, India. His research interests lie in wireless networks, information security, computer networks, sensor networks, etc. He has published various international journals and attended national, international conferences related to wireless networking for the past 6 years

6 years