KeAi CHINESE ROOTS GLOBAL IMPACT

International Journal of Innovation Studies

AN INTELLIGENT FAULT DETECTION FOR IoT-EMBEDDED VITAL MONITORING IMPLANTS USING PRM

S. Nandini[1] & Dr. A.R. Arunachalam[2]

[1]Research Scholar, Department of Information Technology, Dr. M.G.R Educational and Research Institute, Maduravoyal, Chennai, Tamilnadu, India

[2]Professor & Dean, Department of Computer Science, Dr. M.G.R Educational and Research Institute, Maduravoyal, Chennai, Tamilnadu, India,

E-Mail: sn.nandini21@gmail.com[1] arunachalam.cse@drmgrdu.ac.in[2]

Abstract

The Internet of Things (IoT) - embedded vital monitoring revolutionizes the healthcare industry by enabling the personalized, continuous time and data driven patient care system. The incorporation of the IoT-embedded in the medical systems, promotes the early detection of disease, enhances the patient outcomes, and reduces the cost along with the enhanced level of accuracy by marking a significant shift towards the preventive and precision medicine. The major challenge in the IoT-embedded vital monitoring system is the fault occurrence, leading to critical concerns affecting the health of the patients, device functionality and the system performance. The fault in the system leads to increase in False Positive (FP) and False Negative (FN) rate leading to inaccurate data reading leading to failure in personalized treatment plans. To overcome this concern, this research work of detecting the fault in the IoT-embedded vital monitoring system using the hybrid Pearson Correlation, Random Forest (RF) classifier and the Multi linear Regression (MR) technology. The proposed work is trained using two different datasets namely the Kaggle dataset and PhysioNet dataset. Both the datasets are the open access datasets and the performance of the proposed work was compared with the performance of the state of the art methodologies.

Keywords: Internet of Things (IoT), Embedded Vital Monitoring system, Pearson Correlation, Random Forest classifier, Multi-Linear Regression (MR), Accuracy.

1. INTRODUCTION

In the present era characterized by incredible technological skill and advancements, the healthcare application is transitioning into the realm of digital connectivity through the Internet of Things (IoT) [1], which comprises smart wearable sensors and sophisticated Machine Learning (ML) [2] algorithms. The advent of the IoT enormously pioneered the advancements of healthcare applications. The Internet of Things (IoT) [3] integrated with the wearable sensors for performing the continuous health monitoring process the IoT- Embedded vital monitoring system was introduced, ease the remote monitoring of the patients along with the earlier detection of the human diseases. The IoT connects the wearable sensors and communicates the measurable data for the real time processing like treatment by the physicians. The major need for the IoT-Embedded vital monitoring system [4] arises due to the several factors reflecting the landscape of healthcare, advancements in the technology and the various desires of the patients. Most countries possess one third of aged people leading to the high prevalence chronic diseases like diabetes, cardiovascular diseases, blood pressure, and respiratory diseases. These diseases necessitate the continuous monitoring of the vital signs, for an effective management of health conditions of the patients. The number of IoT connected devices has been exponentially increasing over the past one decade as depicted in Figure 1 [5].

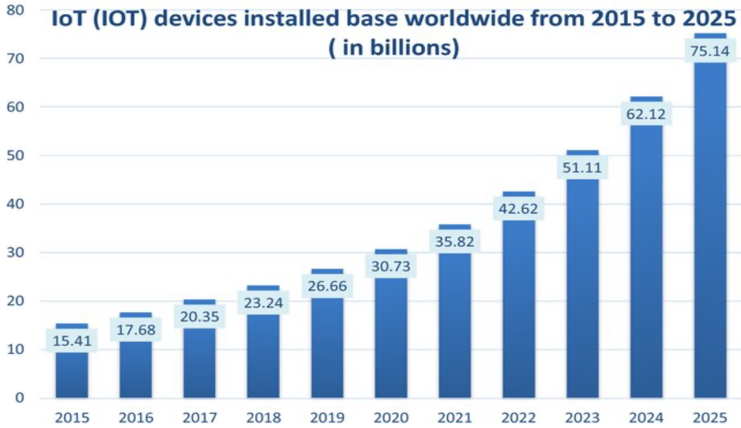

Figure 1: Statistics on Number of IoT devices connected worldwide [5]

The analysis from the Figure 1, discloses that 75.14 billions of IoT devices were connected worldwide which covers wide range of applications namely, healthcare application, monitoring purpose, smart home applications etc. One such finest application of the IoT is the IoT-Embedded vital monitoring system is a technology driven solution designed for the continuous monitoring of the patient's vital signs namely the heart rate [6], blood pressure [7], oxygen saturation [8], temperature [9], blood sugar level [10] and the respiratory rate [11]. The vital signs have been monitored using the wearable and wireless sensors, which are capable of collecting the physiological data and communicates it with the central processing unit such as cloud servers or applications for executing the real time analysis. Despite of fruitful applications of the IoT, it is essential and challenging that the detection of fault in the IoT-embedded systems, as the fault in the system or the devices, results in measurement of incorrect readings or may not measure the vitals due to sensor failure, resulting in inappropriate decision or delayed medical responses. The case of failure of systems leads to the failure in alerting the healthcare professionals during the emergency conditions, leading to the critical condition or serious risk to the patients. This serious effect results and necessitates the regular detection of fault to ensure that the data received are consistent, reliable and is accurate for performing the effective monitoring and medical decision making processes.

Fault detection process [12] ensures that the received alert signals were triggered at the real time and is created on necessary condition. This assists in maintaining the trust in the IoT-embedded vital monitoring system. Numerous methods of fault detection have been introduced with wide range of technologies namely the rule based fault detection, model based fault detection, signal processing based fault detection, Neural Network [13] based fault detection, Fuzzy logic based fault detection. However, the major challenging factors associated with these fault detection methods are the limited computational resources, environmental noise and interferences, false positive and negatives. To overcome these challenges, this proposed work employs the hybrid version of Pearson Correlation, Random Forest (RF) classifier [14] and the

Multi linear Regression (MR) technology [15] for the fault detection in IoT-embedded vital monitoring system. The contributions of this proposed work is as follows.

- Pearson Correlation (PC) can be used in the preprocessing phase to determine which health parameters (such as heart rate, oxygen levels, etc.) are most relevant for fault detection. By filtering out less significant features, the system can focus on key variables, which improves the accuracy and efficiency of fault detection models.
- Random Forest (RF) can accurately classify different types of faults or health anomalies based on historical sensor data.
- Random Forest's ensemble approach reduces the chances of overfitting and can achieve high fault detection accuracy by combining multiple decision trees. This is particularly useful in identifying faults that may be missed by simpler models.
- Multi-Linear Regression (MLR) can help quantify the extent or severity of a fault by modeling the relationship between sensor deviations and system performance. This is useful for prioritizing critical faults over minor ones.

2. LITERATURE REVIEW

The increase in the IoT connections and the necessity for the fault detection in the IoT-monitoring system motivates various researchers to actively involve in designing solutions for the fault detection. The fault detection in IoT-embedded vital monitoring system is a novel method introduced in this proposed work and some of the related research results were presented here for the definition of the objectives of the proposed work.

C.Hou et al. (2024) had designed a fault detection system in the sensors by formulating the sensors as the Single Input Single Output (SISO) model [16] by sensing the error as the input and the output. The model derives the relation between the energy supply and the occurrence of the fault in the smart sensors. The algorithm employed the sound theoretical foundations based on the optimal energy supply. In addition, the proposed model employed the cyber physical system to sense the fault in the sensors. G. Stamatakis et al. (2024) had performed the fault detection of the connected IoT devices by Partially Observable Markov Decision Process [17]. This process computes the efficiency of the sensor by probing the energy of the system and the communication models. In addition, the process adopted the aging information of the sensors and the communication models to determine the fault in IoT models.

M.Ragnoli et al. (2024) had employed the Artificial Intelligence (AI) [18] for the detection of fault in the Computer Numerical Control (CNC) machines. The author implemented the fault detection model in detecting the fault in the machineries of the industry. The developed model aimed in realizing the multi purpose machine for detection of fault in advance. A.Prakash Rawal et al. (2024) had proposed a novel Mobile sink based fault detection model using Q-learning method to enhance the Network lifetime of the sensors in the WSN. The model employed the Genetic algorithm [19] for performing the fault detection process and had enhanced the lifetime of the network by performing the prior prediction of the faults in the sensor.

B.Zhou et al. (2024) had proposed a Resilient Sensor Data Dissemination (RSDD) model [20] for disseminating the sensor data and to detect the fault in the sensors deployed in random manner. The accuracy of the fault prediction is enhanced by minimizing the end to end failure rate in the packet delivery rate. This model employed the hop to hop retransmission process for the efficient detection of the fault in the IoT sensors. A.Sinha et al. (2023) had proposed XAI-LCS model [21] for the early detection of the faults in the sensor to provide an uninterrupted monitoring and to supply the vital data. This model is composed of Artificial Intelligence (AI) technique to mitigate the high computational burden in the fault detection process. This model identified four different types of faults in the sensor and the accuracy is measured to 99.8%.

A.Sinha et al. (2023) had proposed a Deep Reinforcement Learning (DRL) model [22] for the detection of the various types of the faults occurring in the sensors. This model classified the faults in the sensor as bias, drift, complete failure and the precision degradation. The DRL model diagnosed and classified the failure under these four categories and had provided an accuracy of 96.17% with a minimal computational time. This model had proved its efficiency in terms of accuracy, despite of noisy environments. H.Darvishi et al. (2023) had addressed the sensor fault detection (SFD) issue by designing a novel framework composed of Deep Recurrent Graph Convolutional Network (DRGC) [23] which is the integrated version of graph neural network along with convolutional neural networks to detect the faults in the sensors at the earlier stage. The framework was trained using the publicly available datasets to exhibit an accuracy of 94.13%.

X.Yang et al. (2023) had introduced a Solar Insecticidal Lamp Internet of Things (SIL-IoT) model and had implemented fault self detection scheme [24] for computing the fault at the earlier stages. The fault detection scheme is composed of binary based sliding window mechanism, which is proved to be possessing minimal false prediction rate with a minimal computation time. The accuracy of this model is reported to 99.14% but the model is capable of detecting the fault rather than classifying it. G.Kaur et al. (2023) had introduced the Artificial Intelligence (AI) model in the earlier detection of the fault in the IoT sensors. The model is composed of AI based hyper-parameter tuned least square support vector machine for the diagnosis of the fault with an high level of accuracy. The AI model implemented the Reinforcement Learning (RL) [25] method for the earlier prediction of the fault. This model exhibited its supremacy with minimal false alarm rate and with better F1 score and minimal energy consumption.

Despite of various novel methodologies in detecting the fault in the IoT sensors, the drawback of sensor/ node failures persist exhibiting the following challenges in the existing state of the art methodologies.

- IoT devices may capture noisy data due to various factors, such as sensor malfunction, poor connectivity, environmental conditions, or device wear and tear.
- Traditional fault detection mechanisms often struggle with high rates of false positives or false negatives, leading to either unnecessary alerts or failure to detect actual faults.
- Monitoring systems often involve large and complex data streams, making fault detection harder to analyze accurately.
- IoT devices often have limited computational and battery power, which makes it difficult to implement complex fault detection algorithms without draining resources.
- Traditional systems may not adapt well to dynamic changes in the environment, user behavior, or device performance, leading to inaccurate fault detection.

To overcome these challenges, this research work was proposed with the following objectives.

- To Identify and filter out noisy or irrelevant data by focusing on highly correlated variables using Pearson Correlation (PC).
- To employ Pearson Correlation method for detecting abnormal correlations between multiple sensor readings, which might indicate faults or malfunctioning sensors
- To employ Random Forest (RF) classifier method for distinguishing between faulty and non-faulty conditions by learning from labeled datasets.
- To predict potential faults based on the relationships between multiple sensor variables and the likelihood of fault occurrence using Multi-Linear Regression (MLR)

3. PROPOSED WORK

The proposed work of detecting the fault in the IoT-embedded vital monitoring system is of three folded method composed of Pearson Correlation for noise filtering process, Random Forest (RF) classifier for performing the classification process among the faulty and non-faulty nodes and Multi Linear Regression (MLR) for finest classification among the critical and normal faults. The proposed model for the fault detection in the IoT-Embedded vital monitoring system is initially composed of 'n' number of sensor nodes, which is capable of measuring the vital signs of the human body. The data captured by the sensor nodes were communicated to the cloud dataset through the IoT gateway. The cloud server acts as a central repository, from which the fault diagnosis and the corresponding follow up actions by the physicians were performed. The architecture of the proposed fault detection model is depicted in Figure 2.

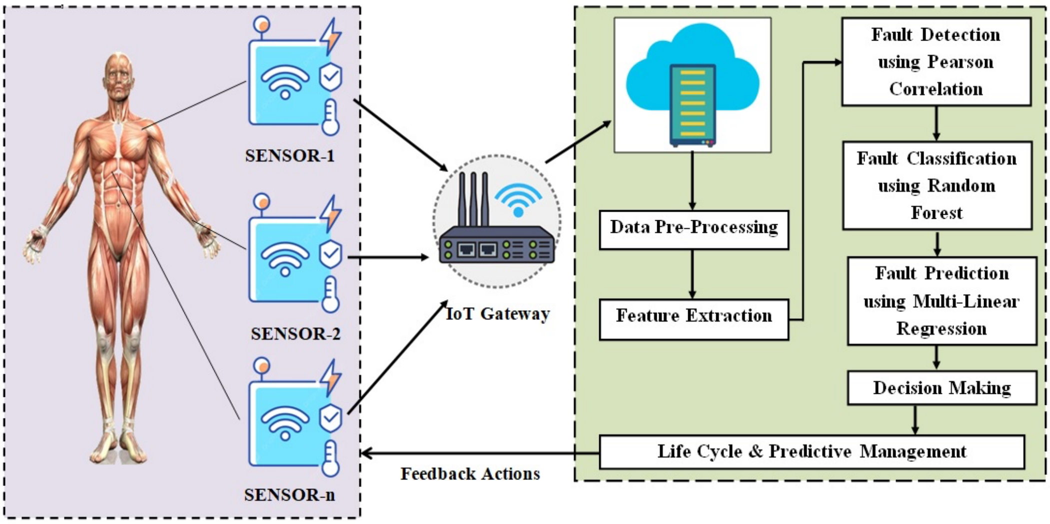

Figure 2: Architecture of the Proposed Fault Detection Model in IoT-Embedded Vital Monitoring System

The proposed fault detection model is composed of seven consecutive processes namely the data pre-processing, feature extraction, fault detection using Pearson Correlation (PC), fault classification using Random Forest (RF) classifier, fault prediction using Multi-Linear Regression (MLR) method followed by the decision making process. The final phase of the proposed model is the life cycle and predictive management process, which is the feedback action of the proposed work to replace or maintain the embedded nodes. The sensor nodes for the vital monitoring systems were placed on the human body which was communicating to the cloud server using LoRaWAN protocol, which is capable of communicating for long range. On an average, 'n' number of different types of sensors have been placed in contact with the human body to measure the common vitals namely, human blood pressure, heart rate, body temperature, respiratory rate, along with the critical vitals blood oxygen saturation, electrocardiogram, blood glucose level, sleep pattern and accelerometer. The different types of sensors were physically connected to the human body and are made to communicate to the cloud server. The cloud server provides access to the physicians, guardians for proceeding with the further actions. In addition to these follow up actions, it is essential to identify the lifetime of the sensor nodes along with the working condition namely the fault detection in the vital measuring sensor nodes.

3.1 Data Pre-processing

The initial process of the fault detection is the pre-processing stage, concentrating on the data cleaning, data normalization, time synchronization and finally ends up with the feature extraction process. The significance of performing the data pre-processing is the cleaning of data which has been recorded in the real time environments. The sensors are capable of performing continuous time monitoring of various physiological parameters namely the heart rate, blood pressure and blood oxygen levels. The data captured through these sensor nodes are highly prone to the noise and interference concerns due to the existing environmental factors. The initial process is the data cleaning, involves mathematical technique for handling the missing values, smoothening the noise and removal of outliers. The most common approach of handling the missing values from the sensor nodes is the mean imputation process, in which the missing values are determined and placed using the mathematical linear interpolation of observed values as defined in equation 1. Let the data communicated by the various nodes to the cloud server through the IoT gateway is the D_N, where D_N= { D_{N1}, D_{N2}, D_{N3},… D_{Nn}}. The D_N is the cumulative data stored in the cloud server, is a mixture of multiple types of data observed and communicated by various nodes N1, N2.. Nn (n number of nodes).

$$D_i = \frac{1}{\acute{\eta}}\sum_{j=1}^{n} D_{Nj} \tag{1}$$

Where, D_i is the incomplete of missing values and D_{Nj} is the available observed values. The term $\acute{\eta}$ is the total number of non-missing observations. The values were observed at various time intervals and were linked using the linear interpolation process as defined in equation 2.

$$D_{Nj} = D_{Nj}(t) = D_{Nj}(t_1) + \frac{\{D_{Nj}(t_2) - D_{Nj}(t_1)\}}{(t_2 - t_1)} \times (t - t_1) \tag{2}$$

Where, $D_{Nj}(t)$ is the observed values over a defined period and is considered as the interpolation of values between two different time intervals. The data handling process is followed by the noise reduction process using the Exponentially Moving Average (EMA) method. The EMA is a technical analysis method of data present in the cloud server. This method assigns weightage to the sensor data, to make it more responsive for the variations in the underlying data. The EMA of the data is measured at various time intervals and it is the function of the previous EMA function as defined in equation 3.

$$EMA_D(t) = \alpha D_{Nj} + (1-\alpha)EMA_D(t-1) \tag{3}$$

Where, $EMA_D(t)$ is the exponential moving average value of the data measured at time ‘t’, while $EMA_D(t-1)$ is the exponential moving average value of the data measured at time ‘t-1’. The term ‘α’ is the smoothing factor ranges from (0,1) and is determined through

$\alpha = 2/(W+1)$, in which the 'W' is the chosen window size. The data cleaning process is followed by the outlier removal process using the Z score. The Z score method employs the threshold value in the data D_N and the data which are deviating from the threshold value were removed as defined in equation 4.

$$Z_D = \frac{D_{Nj} - \mu}{\sigma} \quad (4)$$

Where, D_{Nj} is the data of the sensor node with μ as the mean and σ is the standard deviation. The data cleaning process is preceded by the data normalization process to ensure that all the data from the sensor are on a similar scale. This normalization process assists in enhancing the performance of the proposed PC, RF and MLR algorithms. The normalization process is performed using the Min-Max normalization process as defined in equation 5.

$$D_{norm} = \frac{D_N - \min(D_N)}{\max(D_N) - \min(D_N)} \quad (5)$$

Where, D_{norm} is the normalized data, while $\max(D_N)$ & $\min(D_N)$ are the maximum and minimum values of the observed data. The data normalization process is followed by the feature extraction process, performed on the basis of time and frequency domain analysis to extract the feature on the basis of the time domain. The reason for choosing time and frequency domain basis over the statistical basis, is the behavior and the life time of the sensor nodes varies from time to time as the statistical analysis is not suitable for the fault detection process. The features of the sensor data were measured on the basis of Root Mean Square (RMS), Peak to Peak amplitude, and zero crossing rate. The fourier transformation and wavelet transformation techniques were employed for the analysis in terms of frequency domain. The mathematical representation of the time domain analysis is defined in equations 6.

$$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^{n} D_{norm}(i)} \quad (6)$$

The peal to peak amplitude is defined using the maximum and the minimum value of the normalized data as defined in equation 7.

$$A_{P-P} = \max(D_{Norm}) - \min(D_{Norm}) \quad (7)$$

The zero crossing rate of the time series data is measured based on the consecutive points of related data and is defined in equation 8.

$$R_{ZC} = \frac{1}{(n-1)} \sum_{i=1}^{n} 1(D_{norm}(i)) \times (D_{norm}(i+1)) \quad (8)$$

In turn, to analyze the available sensor data in the cloud server, the fourier transformation method is employed as represented in equation 9.

$$X(D_{norm}(f)) = \int_1^n D_{norm}(t) e^{-j2\pi f t} dt \quad (9)$$

Where, $D_{norm}(t)$ is the normalized data measured in time domain and $D_{norm}(f)$ is the converted normalized data in frequency domain. It is essential to determine the behavior of the sensor in the discrete domain and hence the Discrete Wavelet Transformation (DWT) is employed to define the normalized data in the wavelet basis function as represented in equation 10.

$$D_{norm}(t) = \sum_{i,j}^{m,n} c_{i,j}(D_{norm}(t)) \varphi_{i,j}(t) \quad (10)$$

$$A\left(D_{norm}(t)\right) = \frac{1}{n-\beta}\sum_{i=1}^{n}\left(D_{norm}(i)\right) \times \left(D_{norm}(i+\beta)\right)$$

(11)

Where, $c_{i,j}\left(D_{norm}(t)\right)$ is the coefficient of the normalized data, $\varphi_{i,j}(t)$ is the wavelet basis function, $A\left(D_{norm}(t)\right)$ is the autocorrelation among the normalized data and β is the lag factor. The preprocessed data is fed to the feature selection using the Pearson Correlation (PC) method.

3.2 Fault Detection using Pearson Correlation (PC) method

Pearson Correlation (PC) is the statistical measure quantifying the linear relationship among the two variables ranging from (-1, +1). The value in the range '-1' defines the negative correlation, while '0' represents the no correlation process and '+1' defines the positive correlation as depicted in Figure 3.

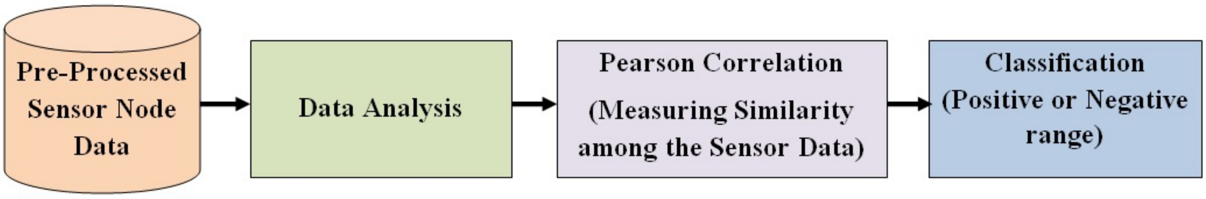

Figure 3: Pearson Correlation for fault detection using data similarity identification

The term correlation in this PC defines that one variable is increasing while the other decreases in the negative correlation, while the positive correlation mentions if both the variables are increasing in nature. The PC method detects the faults in the IoT embedded vital monitoring system by determining the correlation among the readings observed by the sensors at various time intervals. The correlation analysis among the sensor reading determined by the Pearson Correlation is performed by determining the coefficients between the pair of the sensor readings. The Pearson correlation coefficient between two sensor data streams D_i and D_j are defined in equation 12.

$$r_{ij} = \frac{\sum_{i=1}^{m}(D_i - \overline{D_i})\sum_{j=1}^{n}(D_j - \overline{D_j})}{\sqrt{\sum_{i=1}^{m}(D_i - \overline{D_i})^2 \sum_{j=1}^{n}(D_j - \overline{D_j})^2}}$$

(12)

Where, D_i and D_j are the individual data points from the same sensor at two different instant, $\overline{D_i}$ is the mean of D_i value and $\overline{D_j}$ is the mean of D_j value.

Similarly, $\overline{D_i} = \frac{1}{m}\sum_{i=1}^{m} D_i$

(13)

And $\overline{D_j} = \frac{1}{n}\sum_{j=1}^{n} D_j$

(14)

The range of correlation coefficient is classified in the range [-1,+1] defining,

- $r_{ij} = -1$; Perfect Negative linear relationship
- $r_{ij} = 0$; No linear relationship
- $r_{ij} = 1$; Perfect Positive linear relationship.

The variance of the two data points of the similar sensor is determined using the covariance function as defined in equation 15.

$$Cov(D_i, D_j) = \frac{1}{m \times n} \sum_{i=1}^{m} (D_i - \bar{D}_i) \sum_{j=1}^{n} (D_j - \bar{D}_j)$$

(15)

The variance of the D_i and D_j are determined using the mathematical relationship defined in equation 16 and 17.

$$Var(D_i) = \frac{1}{m} \sum_{i=1}^{m} (D_i - \bar{D}_i)^2$$

(16)

$$Var(D_j) = \frac{1}{n} \sum_{j=1}^{n} (D_j - \bar{D}_j)^2$$

(17)

The Pearson Correlation Coefficient in terms of covariance and variance among the two data points is defined in equation 18.

$$r_{ij} = \frac{Cov(D_i, D_j)}{\sqrt{Var\,(D_i) \times Var(D_j)}}$$

(18)

The fault of the IoT-embedded vital monitoring system is detected using the variance exhibited by the sensors. The classification of the fault is performed using the Random Forest (RF) classifier.

3.3 Fault Classification using Random Forest Classifier

The Random Forest (RF) classifier is used to classify the fault identified by the Pearson Correlation method. The fault is classified into critical fault and normal/minor fault in the IoT-embedded vital monitoring system. The Random Forest is the widely employed Machine Learning (ML) algorithm employed for the classification application and is well known for its ability to handle the complex datasets with noisy and missing values. The RF algorithm functions on the basis of decision tree process as depicted in Figure 4.

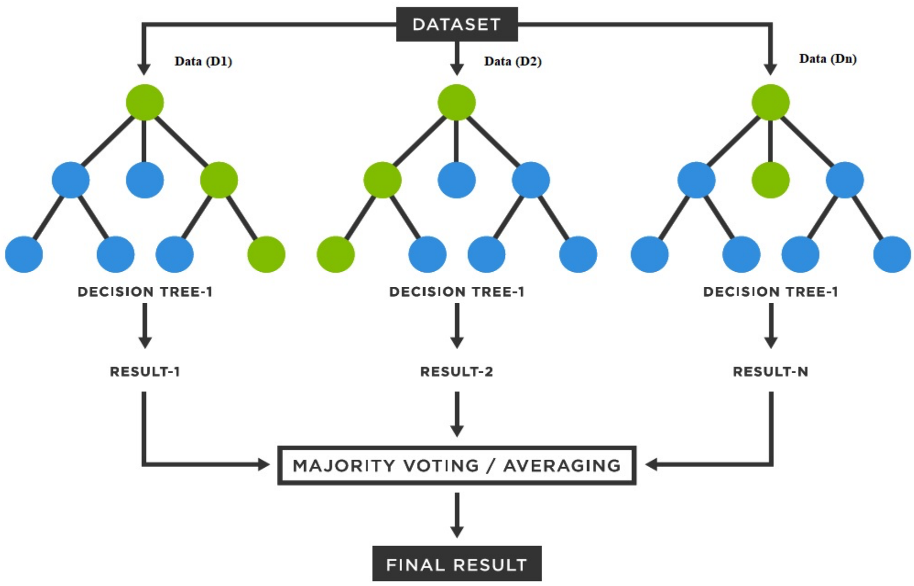

Figure 4: Random Forest in classifying the Fault from the dataset

Random Forest is considered as the suitable classifier for this method of fault detection, as the cloud dataset is composed of complex types of data received by diversified sensor types and the data may be with noise and consists of incomplete or missing values. The RF classifier operates to perform decision making process by constructing the multiple decision trees and the mean prediction from the decision trees. In the RF method of fault classification process, the bootstrapping method is the initial process, in which each decision tree T_i for the bootstrap sample D_i from the normalized original dataset $D_{norm.}$ The decision tree T_i is created by performing the sampling process of the data with replacement mechanism. The tree construction process is performed by training the subset of features in the cloud data. The major criteria employed for tree construction process is through the computation of the gini impurity or entropy as defined in equation 19.

$$Gini\left(r_{ij}\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij} * r_{(i-1)(j-1)}$$

(19)

Where, $r_{ij} * r_{(i-1)(j-1)}$ is the proportion of the samples corresponding to the samples 'i' and 'j'. The entropy of the data from any node is determined as in equation 20.

$$H\left(r_{ij}\right) = -\sum_{i=1}^{m}\sum_{j=1}^{n}\left(r_{ij}\right) log_2\left(r_{ij}\right)$$

(20)

The primary objective for dividing the data at each sensor node is to minimize the gini impurity or the entropy. Based on the gini index value, the fault has been classified as represented in equation 21.

$$F_c = \begin{cases} Gini\left(r_{ij}\right) > Gini_T; Critical\ Fault \\ Gini\left(r_{ij}\right) \leq Gini_T; Normal\ Fault \end{cases}$$

(21)

The fault classification is performed on the basis of the gini index, and is the gini index is greater than the threshold gini, then the fault is classified as the critical fault, else it is classified as the normal fault.

3.4 Fault prediction using Multi-Linear Regression

The final phase of the proposed model is the fault prediction of the Multi-Linear Regression (MLR) method, determines the linear or non linear relationship between the multiple variables of the sensor observed values, which were independent in nature with the dependent variable of fault probability. The MLR method assumes the linear relationship among the input variables $\{D_1, D_2, \ldots D_n\}$ which were recorded by the various types of sensors and the output $P\{r_{ij}\}$, which the probability of occurrence of the fault. The linear relationship for the occurrence of the fault is defined in equation 22.

$$P\{r_{ij}\} = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \cdots . \alpha_n D_n + \tau$$

(22)

Where, $P\{r_{ij}\}$ is the probability of the occurrence of the fault in the IoT embedded vital monitoring systems, D_1, .. D_n are the independent data variables recorded by various types of sensors, α_0 is the intercept while $\alpha_1, \alpha_2 \ldots \alpha_n$ are the multi-linear regression coefficients representing the contribution of the each embedded node towards the output. The term τ is the error term, representing the residuals. The coefficients of the MLR are estimated using the least square estimation method minimizing the sum of squared residuals as defined in equation 23.

$$\min_{\alpha_1,\alpha_2\dots\alpha_n} \alpha = \sum_{i=1}^{n}\{f_a - (\alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \cdots . \alpha_n D_n)^2\}$$

(23)

Where, the f_a is the actual fault, $\alpha_0, \alpha_1 D_1, \alpha_2 D_2, \dots . \alpha_n D_n$ are the MLR coefficients of the independent data for 'n' number of samples. The algorithm for detecting the fault in the IoT-embedded vital monitoring system is presented in Table 1.

Table 1: Algorithm for fault detection and classification using PRM method

Algorithm: Fault Detection and Classification using Pearson Correlation, Random Forest and Multi Linear Regression (PRM) method

Input: Embedded Sensor data D_N= { D_{N1}, D_{N2}, D_{N3},… D_{Nn}}

Output: Fault Classification; Probability of fault occurrence P{r_{ij}}

Processes:

1: Initialize the number of sensor nodes 'n'

2: Initialize the time interval for data measurement 't'

3: Predict missing values using $D_{Nj} = D_{Nj}(t) = D_{Nj}(t_1) + \frac{\{D_{Nj}(t_2) - D_{Nj}(t_1)\}}{(t_2 - t_1)} \times (t - t_1)$

4: Determine Exponental Moving Average function: $EMA_D(t) = \alpha D_{Nj} + (1-\alpha) EMA_D(t-1)$

5: Normalize the input data: $D_{norm} = \frac{D_N - \min(D_N)}{\max(D_N) - \min(D_N)}$

6: Perform Outlier detection

7: Determine RMS of normalized data: $RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} D_{norm}(i)}$

8: Determine peak to peak amplitude: $A_{P-P} = \max(D_{Norm}) - \min(D_{Norm})$

9: Determine Zero crossing rate: $R_{ZC} = \frac{1}{(n-1)}\sum_{i=1}^{n} 1\left(D_{norm}(i)\right) \times (D_{norm}(i+1))$

10: For node (i=1, i++, i<n)

11: Determine Pearson Correlation Coefficient: $r_{ij} = \frac{\sum_{i=1}^{m}(D_i - \overline{D_i})\sum_{j=1}^{n}(D_j - \overline{D_j})}{\sqrt{\sum_{i=1}^{m}(D_i - \overline{D_i})^2 \sum_{j=1}^{n}(D_j - \overline{D_j})^2}}$

12: If ($r_{ij} = -1$), then Return, "Negative linear relationship"

13: Else if ($r_{ij} = 0$), then Return, "No Linear Relationship"

14: Else if ($r_{ij} = 1$); then Return, "Positive Linear Relationship"

15: Determine Covariance among data: $Cov(D_i, D_j) = \frac{1}{m \times n}\sum_{i=1}^{m}(D_i - \overline{D_i})\sum_{j=1}^{n}(D_j - \overline{D_j})$

16: Determine Gini index: $Gini\left(r_{ij}\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij} * r_{(i-1)(j-1)}$

17: Entropy of gini index: $H(r_{ij}) = -\sum_{i=1}^{m}\sum_{j=1}^{n}(r_{ij})\, log_2(r_{ij})$

18: If ($Gini\left(r_{ij}\right) > Gini_T$); Return, "Critical Fault"

19: Else if ($Gini\left(r_{ij}\right) \leq Gini_T$), Return, "Normal Fault"

20: Determine Linear Relationship: $P\{r_{ij}\} = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \cdots . \alpha_n D_n + \tau$

21: If ($P\{r_{ij}\} > r_T$), Probability >0.5; Fault occurrence

22: Else if ($P\{r_{ij}\}$<r_T), Probability <0.5; Least Fault occurrence
23: End if
24: End if
25: End for
26: End processes

The fault occurrence in detected followed by the decision making process, which involves two processes namely the life cycle management and predictive management process which involves the replacement of the dead nodes with the new ones or to perform the repairing mechanisms in the embedded monitoring nodes.

4. PERFORMANCE ANALYSIS AND DISCUSSION

The proposed work was trained using two different datasets namely the Kaggle dataset and the PhysioNet dataset for the training process. The dataset is composed of various data from different types of sensors like heartbeat sensor, temperature sensor, blood pressure sensor etc. Among the entire data, 80% of data have been used for training process while 20% data have been used for testing process. The performance of the proposed work is measured in terms of accuracy, precision, recall and F score. In addition to these vital parameters, the proposed work is analyzed in terms of Root Mean Square Error and is compared with the existing state of art methods like, Multi-Scale Fusion Neural Network (D.K. Reddy Basani, 2024), Convolutional Neural Networks (B. Aljafari, 2024), Recurring Neural Networks (I.S. Ramírez, 2024), Support Vector Machine (L.Hou, 2024), and CNN+BiLSTM (R. Laythkhaleel, 2024).

The accuracy is the vital parameter defining how perfectly the proposed model predicts the fault. In the realm of this fault detection framework designed for an IoT-integrated vital monitoring system, accuracy signifies the fraction of overall forecasts generated by the model that hit the mark. This vital performance indicator measures the model's prowess in accurately discerning both typical and defective conditions derived from the sensor data gathered from the monitoring apparatus. The mathematical expression for determining the accuracy is defined in equation 24.

$$Accu_y = \frac{TP + TN}{TP + TN + FP + FN} \tag{24}$$

Where, TP is the True Positive, TN is the True Negative, FP is the False Positive, FN is the False Negative. The parameter next to the accuracy is the precision, defines the measure of rate of correct prediction of faulty instances out of all occurring instances predicted as the faulty. The mathematical expression of the precision is defined in equation 25.

$$Prec_n = \frac{TP}{TP + FP} \tag{25}$$

Recall assesses the fraction of genuine faulty occurrences that the model successfully recognizes. It emphasizes the model's prowess in spotting faults when they arise and aids in reducing false negatives (overlooked faults). A high recall signifies that the model excels in fault detection and diminishes the tally of undetected faults (overlooked faults or false negatives). Conversely, low recall suggests that the model is overlooking numerous faulty instances, which could be critical in essential monitoring systems where failure to detect a fault could lead to dire outcomes. The F1 Score represents the harmonic average of precision and recall. It delivers a unified performance indicator that harmonizes both recall and precision, particularly when there is a lopsided distribution of faulty and normal instances. A lofty F1 score indicates that the model boasts both high precision (minimal false alarms) and high recall (few missed faults). This becomes crucial in skewed datasets (where normal instances significantly outnumber faulty ones), as it guarantees the model performs admirably in both fault detection and avoiding excessive false alarms. A low F1 score denotes that the model is either neglecting numerous faults (low recall) or generating an abundance of false alarms (low precision). The mathematical expression for the recall and F1 score are defined in equation 26 and 27.

$$Reca_l = \frac{TP}{TP+FN} \tag{26}$$

$$F1\,Score = \frac{2 \times Prec_n \times Reca_l}{Prec_n + Reca_l} \tag{27}$$

In addition to these parameters, the Root Mean Square (RMS) error is measured to determine the average of the squared difference among the actual and the predicted values as presented in equation 28.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(D_{ai} - D_{pi}\right)^2} \tag{28}$$

The proposed model is analyzed for the two different types of datasets and the observed values are listed in Table 2.

Table 2: Performance analysis- Two different datasets

Parameters	Kaggle	PhysioNet
$Accu_y$	96.21%	95.14%
$Prec_n$	95.59%	94.26%
$Reca_l$	95.31%	94.06%
F1 Score	94.16%	93.39%
RMSE	0.0821	0.0974

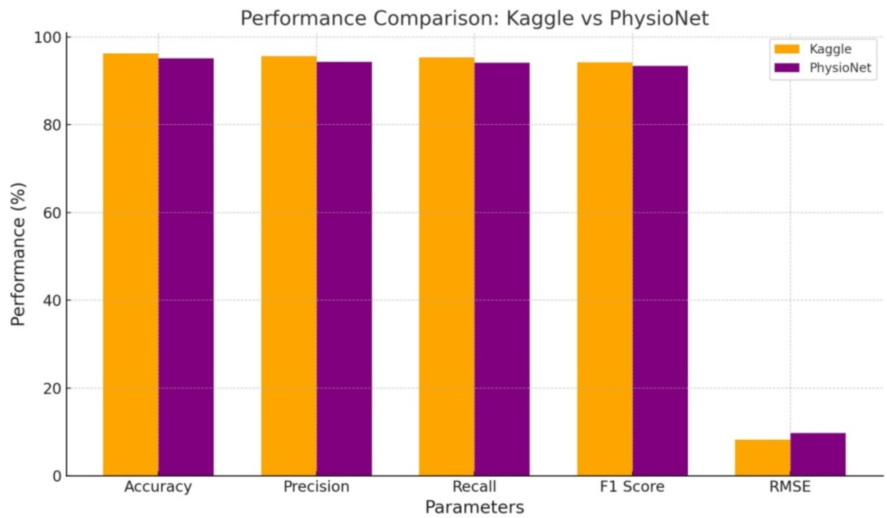

Figure 5: Performance comparison- Two different datasets

The Kaggle dataset seems to surpass the PhysioNet dataset in identifying faults within an IoT-integrated vital monitoring system, as reflected by its superior scores in accuracy, precision, recall, and F1 measure. Datasets from Kaggle are frequently crafted with great care, showcasing meticulously organized data and accurate labeling. This meticulousness aids in the development of more precise models, as there are fewer erroneous or inaccurately labeled data points, resulting in enhanced fault detection capabilities. While PhysioNet datasets are esteemed for their medical insights, they may carry more noise or discrepancies stemming from the processes of real-world data collection, potentially impacting model efficacy. The performance of the proposed work is compared with the existing state of the art methodologies to prove the supremacy of the proposed work. The observed values were listed in Table 3.

(D.K. Reddy Basani, 2024), (B. Aljafari, 2024), (I.S. Ramírez, 2024), (L.Hou, 2024), and (R. Laythkhaleel, 2024)

Table 3: Performance analysis- Different Methodologies

Parameters	Multi-Scale Fusion Neural Network [26]	Convolutional Neural Networks [27]	Recurring Neural Networks [28]	Support Vector Machine [29]	CNN+BiLSTM [30]	Proposed Work
$Accu_y$	80.14%	81.21%	82.06%	79.21%	83.16%	96.21%
$Prec_n$	79.54%	80.48%	81.62%	79.01%	82.30%	95.59%
$Reca_l$	78.64%	79.03%	80.14%	78.74%	81.62%	95.31%
F1 Score	79.08%	79.74%	80.87%	78.87%	81.95%	94.16%
RMSE	0.0921	0.0922	0.0984	0.0992	0.0994	0.0821

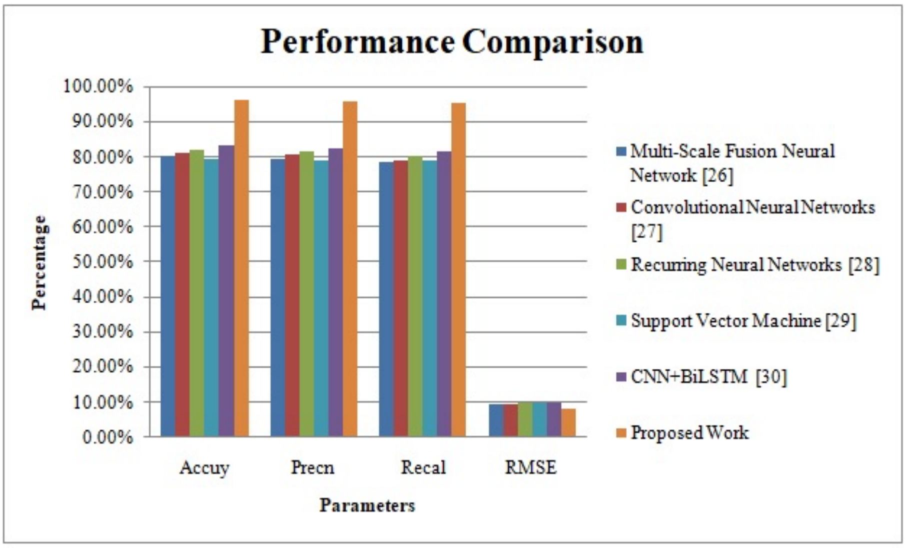

Figure 6: Performance comparison of proposed work with existing works

The Pearson correlation unveils the intricate linear connections between the input features (such as sensor outputs) and the target variable (fault or no fault). By assessing the strength of these correlations, this technique aids in pinpointing the most significant features that exhibit a strong relationship with fault occurrences, thereby filtering out the clamor of irrelevant or weakly linked attributes. By handpicking essential features based on their correlations, Pearson correlation mitigates overfitting, bolsters model interpretability, and amplifies the efficiency of Random Forest and Multilinear Regression models by engaging with more insightful data. Random Forest achieves remarkable accuracy by curtailing overfitting, a frequent pitfall in standalone decision trees. Its strategies of bootstrapping and feature randomness guarantee that the model adapts splendidly to unseen data, rendering it robust for fault identification. It adeptly manages high-dimensional data, a hallmark of IoT systems with their plethora of sensors generating continuous streams of information. Random Forest enhances accuracy, precision, and recall by skillfully navigating the complexities of high-dimensional IoT data. It shows resilience against noise and irrelevant features, particularly after Pearson correlation has sifted out the weaker elements. The model's linear characteristic ensures it operates with computational efficiency and clarity, making it straightforward to discern how sensor data contributes to fault detection. The straightforwardness and interpretability of Multilinear Regression establish a sturdy foundation for grasping the linear relationships between sensors and faults. Additionally, it complements Random Forest by illuminating linear dependencies.

5. CONCLUSION AND FUTURE WORK

In the digital era, the smart monitoring of the human health is essential, thus the IoT-embedded vital monitoring system plays a major role in performing the continuous monitoring on the vital parameters of the human body. It is essential to measure the fault and continuous management of the embedded resources to perform tireless monitoring of the human body. This proposed research work incorporates the hybrid Pearson Correlation, Random Forest and Multi-Linear Regression method of detect the fault and classify the fault more effectively than the existing methodologies. The performance analysis of the proposed work exhibits 96.21% of accuracy, 95.59% of Precision, 95.31% of recall, 94.16% of F1 score and minimum of 0.0821 Root Mean Square Error. The synergy of Pearson correlation, Random Forest, and

Multilinear Regression within fault detection frameworks fosters resilient feature selection, formidable non-linear modeling, and comprehensible linear regression. Collectively, these methodologies elevate fault detection capabilities by enhancing accuracy, minimizing overfitting, adeptly managing intricate sensor data, and delivering understandable results. This synergy renders it perfectly suited for real-time IoT-embedded systems, where both computational agility and reliability are paramount. The performance of the proposed work shall be enhanced by incorporating non-linear transformation techniques, and to incorporate boosting algorithms for a better level of accuracy in detecting the faults.

REFERENCES

[1] C. -W. Park, V. Palakonda, S. Yun, I. -M. Kim and J. -M. Kang, "OCR-Diff: A Two-Stage Deep Learning Framework for Optical Character Recognition Using Diffusion Model in Industrial Internet of Things," in *IEEE Internet of Things Journal*, vol. 11, no. 15, pp. 25997-26000, 1 Aug.1, 2024

[2] H. Yildirim, "Novel Statistical Regularized Extreme Learning Algorithm to Address the Multicollinearity in Machine Learning," in *IEEE Access*, vol. 12, pp. 102355-102367, 2024

[3] W. A. N. A. Al-Nbhany, A. T. Zahary and A. A. Al-Shargabi, "Blockchain-IoT Healthcare Applications and Trends: A Review," in *IEEE Access*, vol. 12, pp. 4178-4212, 2024.

[4] Z. Zhang, R. Zhang, C. -W. Chang, Y. Guo, Y. -W. Chi and T. Pan, "iWRAP: A Theranostic Wearable Device With Real-Time Vital Monitoring and Auto-Adjustable Compression Level for Venous Thromboembolism," in *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 9, pp. 2776-2786, Sept. 2021

[5] M. Khasawneh, A. Azab, S. Alrabaee, H. Sakkal and H. H. Bakhit, "Convergence of IoT and Cognitive Radio Networks: A Survey of Applications, Techniques, and Challenges," in *IEEE Access*, vol. 11, pp. 71097-71112, 2023.

[6] Y. Akamatsu, T. Umematsu and H. Imaoka, "CalibrationPhys: Self-Supervised Video-Based Heart and Respiratory Rate Measurements by Calibrating Between Multiple Cameras," in *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 3, pp. 1460-1471, March 2024.

[7] M. Deutges and H. Redtel, "Autonomous Calibration of Blood Pressure Dependent Data Using Second-Order Blood Pressure Variation for a Future Mobile Diagnostic: Requirements for a Calibration," in *IEEE Access*, vol. 12, pp. 97269-97279, 2024.

[8] C. Zhang, H. Yan, Q. Liu, K. Yang, F. Liu and L. Lin, "Design and Analysis of a Through-Body Signal Transmission System Based on Human Oxygen Saturation Detection," in *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, vol. 10, no. 1, pp. 122-131, March 2024.

[9] H. Liu, N. Peng, W. Meng, Z. Zhou, Q. Liu and S. Q. Xie, "Optical Fiber Temperature Sensing Probe Based on F–P Cavity for Human Body Temperature Monitoring," in *IEEE Sensors Journal*, vol. 24, no. 10, pp. 16188-16195, 15 May15, 2024.

[10] S. Liu, R. Song and X. Lu, "Research on Blood Glucose Nonlinear Controller Based on Backstepping Adaptive Control Algorithm," *2024 IEEE 13th Data Driven Control and Learning Systems Conference (DDCLS)*, Kaifeng, China, 2024, pp. 472-477.

[11] Y. Akamatsu, T. Umematsu and H. Imaoka, "CalibrationPhys: Self-Supervised Video-Based Heart and Respiratory Rate Measurements by Calibrating Between Multiple Cameras," in *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 3, pp. 1460-1471, March 2024,.

[12] M. Guerfel and H. Messaoud, "On the use of DPCA in process fault detection and identification," *2024 International Conference on Control, Automation and Diagnosis (ICCAD)*, Paris, France, 2024, pp. 1-6.

[13] S. Kannan, Prabakaran D, Dhinehs Kumar S and Shivaraj S, "A Deep Learning-Based Convolution Neural Networks to Forecast Wind Energy," *2023 International Conference on Recent Trends in Electronics and Communication (ICRTEC)*, Mysore, India, 2023, pp. 1-6.

[14] T. Shen, C. S. Mishra, J. Sampson, M. T. Kandemir and V. Narayanan, "An Efficient Edge–Cloud Partitioning of Random Forests for Distributed Sensor Networks," in *IEEE Embedded Systems Letters*, vol. 16, no. 1, pp. 21-24, March 2024.

[15] Z. Xiao, J. Sun, Z. Zhang and Y. Liu, "Nonparametric Regression for MU-MIMO Channel Prediction: From KNN to Local Linear Regression," in *IEEE Transactions on Wireless Communications*, vol. 23, no. 4, pp. 2784-2795, April 2024.

[16] C. Hou, R. Shi, Q. Huang and Y. Wang, "Energy Harvest of Multiple Smart Sensors With Real-Time Fault-Detection," in *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 2, pp. 1592-1606, April 2024.

[17] G. Stamatakis, N. Pappas, A. Fragkiadakis, N. Petroulakis and A. Traganitis, "Semantics-Aware Active Fault Detection in Status Updating Systems," in *IEEE Open Journal of the Communications Society*, vol. 5, pp. 1182-1196, 2024.

[18] M. Ragnoli *et al*., "A Condition and Fault Prevention Monitoring System for Industrial Computer Numerical Control Machinery," in *IEEE Access*, vol. 12, pp. 20919-20930, 2024.

[19] A. Prakash Rawal and P. Chanak, "A Q-Learning-Based Fault-Tolerance Data Routing Scheme for IoT-Enabled WSNs," in *IEEE Internet of Things Journal*, vol. 11, no. 14, pp. 25283-25293, 15 July15, 2024.

[20] B. Zhou, C. Wu, Q. Yang, Y. Qian and Y. Nie, "Resilient Sensor Data Dissemination to Mitigate Link Faults in IoT Networks With Long-Haul Optical Wires for Power Transmission Grids," in *IEEE Internet of Things Journal*, vol. 11, no. 9, pp. 15919-15939, 1 May1, 2024.

[21] A. Sinha and D. Das, "XAI-LCS: Explainable AI-Based Fault Diagnosis of Low-Cost Sensors," in *IEEE Sensors Letters*, vol. 7, no. 12, pp. 1-4, Dec. 2023.

[22] A. Sinha and D. Das, "SNRepair: Systematically Addressing Sensor Faults and Self-Calibration in IoT Networks," in *IEEE Sensors Journal*, vol. 23, no. 13, pp. 14915-14922, 1 July1, 2023.

[23] H. Darvishi, D. Ciuonzo and P. Salvo Rossi, "Deep Recurrent Graph Convolutional Architecture for Sensor Fault Detection, Isolation, and Accommodation in Digital Twins," in *IEEE Sensors Journal*, vol. 23, no. 23, pp. 29877-29891, 1 Dec.1, 2023.

[24] X. Yang, L. Shu, K. Li, Z. Huo, S. Shu and E. Nurellari, "SILOS: An Intelligent Fault Detection Scheme for Solar Insecticidal Lamp IoT With Improved Energy Efficiency," in *IEEE Internet of Things Journal*, vol. 10, no. 1, pp. 920-939, 1 Jan.1, 2023.

[25] G. Kaur and M. Bhattacharya, "Intelligent Fault Diagnosis for AIT-Based Smart Farming Applications," in *IEEE Sensors Journal*, vol. 23, no. 22, pp. 28261-28269, 15 Nov.15, 2023.

[26] Dinesh Kumar Reddy Basani, Basava Ramanjaneyulu Gudivaka, Rajya Lakshmi Gudivaka, Raj Kumar Gudivaka, Enhanced Fault Diagnosis in IoT: Uniting Data Fusion with Deep Multi-Scale Fusion Neural Network, Internet of Things, 2024, 101361.

[27] Belqasem Aljafari, Priya Ranjan Satpathy, Sudhakar Babu Thanikanti, Nnamdi Nwulu, Supervised classification and fault detection in grid-connected PV systems using 1D-CNN: Simulation and real-time validation, Energy Reports, Volume 12, 2024, Pages 2156-2178.

[28] Isaac Segovia Ramírez, Fausto Pedro García Márquez, Jesús Parra Chaparro, Convolutional neural networks and Internet of Things for fault detection by aerial monitoring of photovoltaic solar plants, Measurement, Volume 234, 2024.

[29] Liqun Hou, Qianchi Huang, A smart WSNs node with sensor computing and unsupervised One-Class SVM classifier for Machine fault detection, Measurement, 2024, 115843.

[30] Riyam Laythkhaleel, Abdullahi Abdu Ibrahim, Raghda Awad Shaban Naseri, Hameed Mutlag Farhan, An efficient faults and attacks categorization model in IoT-based cyber physical systems using Dilated CNN and BiLSTM with multi-scale dense Attention module, Biomedical Signal Processing and Control, Volume 96, Part B, 2024.