

Research Article

IoT-Based Harmful Toxic Gases Monitoring and Fault Detection on the Sensor Dataset Using Deep Learning Techniques

J. Praveenchandar,¹ D. Vetrithangam ,² S. Kaliappan,³ M. Karthick,⁴ Naresh Kumar Pegada,⁵ Pravin P. Patil,⁶ S. Govinda Rao,⁷ and Syed Umar ⁸

¹Department of Computer Science and Engineering, Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences, Chennai 600027, Tamil Nadu, India

²Department of Computer Science & Engineering, Chandigarh University, Mohali 140413, India

³Department of Mechanical Engineering, Velammal Institute of Technology, Chennai 601204, Tamil Nadu, India

⁴Department of Mechanical Engineering, Velammal Engineering College, Velammal New-Gen Park, Ambattur-Redhills Road, Chennai 600066, Tamil Nadu, India

⁵Department of Computer Science & Engineering, KG Reddy College of Engineering and Technology, Telangana, Hyderabad, India

⁶Department of Mechanical Engineering, Graphic Era Deemed to Be University, Bell Road, Clement Town, Dehradun 248002, Uttarakhand, India

⁷Department of Computer Science Engineering, Gokaraju Rangaraju Institute of Engineering and Technology (GRIET), Bachupally 500090, Hyderabad, India

⁸Department of Computer Science, College of Engineering and Technology, Oromiya, Wollega University, Nekemte, Ethiopia

Correspondence should be addressed to D. Vetrithangam; vetrigold@gmail.com and Syed Umar; syedumar@wollegauniversity.edu.et

Received 7 June 2022; Revised 12 July 2022; Accepted 18 July 2022; Published 8 August 2022

Academic Editor: Punit Gupta

Copyright © 2022 J. Praveenchandar et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

One of the main reasons for accidents among workers is harmful gas leakage. Many people die in chemical industries and their surrounding areas. The present invention is responsible for monitoring and controlling hazardous toxic gases like nitrogen dioxide (NO₂), carbon monoxide, ozone (O₃), sulfur dioxide (SO₂), LPG, hydrocarbon gases, silicones, hydrocarbons, alcohol, CH₄, hexane, benzene, as well as environmental conditions, such as temperature and relative humidity to prevent industrial accidents. The Arduino UNO R3 board is used as the central microcontroller. It is connected to the Cloud via AQ3 sensor, Minipid 2 HS PID sensor, IR5500 open path infrared gas detector, DHT11 Temperature and Humidity Sensor, MQ3 sensor, and ESP8266 and WIFI Module, which can store real-time sensor data and send alert messages to the industry's safety control board. Machine learning and artificial intelligence will be used to make an intelligent prediction (AI). The information gathered will be examined in real-time. The real-time data provided through the sensor can be accessed worldwide. Sensor data quality is critical in the Internet of Things (IoT) applications because poor data quality renders them useless. Error detection in sensor data improves the IoT-based toxic gas monitoring, controlling, and prediction system. Live data from sensors or datasets should be analyzed properly using appropriate techniques. Hence, hybrid hidden Markov and artificial intelligence models are applied as an error detection technique in the sensor dataset. This technique outperformed the dataset gas sensor array under dynamic gas mixtures and lived data. Our method outperformed harmful gas monitoring and error detection in sensor datasets compared to other existing technologies. The hybrid HMM and ANN fault detection methods performed well on the datasets and produced 0.01% false positive rate.

1. Introduction

The expansion of the chemical industry in India and around the world has increased the possibility of events involving hazardous chemicals. The chemical industry in India and other parts of the world produces many products. Chemical mishaps can have fatal consequences for both humans and the environment. When a hazardous substance is discharged unrestrained, it threatens public health and the environment. In conjunction with unintentional or deliberate liveliness, natural circumstances can proceed in chemical mishaps. Globally, air quality is deteriorating, and indoor air pollution poses a significant hazard to human health. The pollutants in indoor air are often made up of several volatile organic compounds (VOCs) that are detrimental to the human body, particularly those with low molecular weights. Furthermore, in some cases, more than one type of VOC is present. Therefore, a device that can detect one or more VOCs simultaneously would be ideal [1]. It is difficult to monitor harmful gases in the ambient air, however, the industry needs to maintain a safe atmosphere. Too much oxygen or a hazardous gas, such as sulfide, hydrogen, chlorine, or ammonia, could endanger persons and equipment nearby. It is critical to have a reliable, accurate, automated, and continuous toxic gas monitoring, controlling, and prediction system.

On the other hand, finding the proper sensor and monitoring system for industrial area and the environment can help automate the gas detection process [2]. The Bhopal gas tragedy happened in the year of 1984, and the Visakhapatnam (Vizag) gas leak, as well as a number of other industrial accidents, occurred at night or when the plants were about to restart after a period of inactivity [3]. Sensor fault detection methods were developed to detect the faults in the sensor dataset [4]. It is vital to utilize as much data as available while evaluating their influence by measuring their information quality to execute the utmost accurate detection of crucial developing defects. Time series analysis and forecasting approaches are utilized to evaluate the course of critical performing defects and estimate future values to discover them as early as possible [5]. A remote monitoring terminal detects the site's environment and gas concentrations. Remote sense terminals are used to detect scene circumstances and gas concentration states. The control station manages the main station's connection to the network of remote detection terminal information and sends timely alarm data to your phone via the GSM module tracking terminal [6]. To identify the presence of numerous different types of gases in a sample or in the environment, an IoT-based dangerous gas monitoring system is designed that will include several sensors and components, and the fault data detection method is applied to the sensor dataset to detect the faults that exist in the real-time sensor dataset.

This research paper typically contains an introduction, a literature review, related terminologies, a proposed system, and results and discussion.

2. Literature Review

Shi et al. [7] developed Gaseous cataluminescence (CTL) sensor equipment that uses aluminum/iron oxide composites to detect and measure harmful gases, such as ethyl ether, acetone, hexane, and chloroform. The output presented that the CTL sensor equipment could respond to ethyl ether at temperatures as low as 180°C, significantly less than most detailed CTL reaction temperatures. The decision also revealed that the linear detection extent was extensive, spanning from 10 ppm to 58 ppm ($R = 0.998$, $n = 7$), and that the detection boundary was low (4.30 ppm). Furthermore, the CTL sensor had a reaction time of 4 s and a resumption time of 8 s for ethyl ether detection, which was fairly quick. The word error refers to soft faults discovered in sensor information, typically identified in the framework, just as missing values, outliers, bias, drifts, and uncertainty, which should be identified, measured, and eliminated or rectified to enhance the value of the sensor data. To explore the various types of sensor data errors that contribute to sensor information contamination, as well as existing solutions for detecting and correcting those mistakes that can be implemented at the layer of the IOT architecture. This research work considered a limited number of gases. Sun et al. [8] proposed a system to monitor the harmful gas for the tunnel. The dispersion of NO_x and CO in the tunnel over time and space was studied. NO_x absorptions were lower than 2 mg/m³ and CO absorptions were lower than 3 mg/m³ about 4 minutes after the trains passed the monitoring stations. The combinations of CO and NO_x in the yangbajing tunnel, which is 3000 meters above sea level, diesel-hauled, and has a length of more than 3 kilometers, can meet the specifications in the code for design on managing the ventilation of railway tunnels (tb10068 - 2010) during natural lighting circumstances, as per the outcomes of on-site selections of hazardous gases. Only CO and NO_x were evaluated in this study.

Takami et al. [9] suggested sensor data analysis with machine learning. Machine learning can enable advanced maintenance, such as degradation diagnostics and predictive maintenance when used for sensor data analysis in industrial automation. Even though factors that cause mistakes and influence the sensor data quality are well-understood, it has easy solutions to data quality issues, including using industry-level sensors that are more precise, sustainable, and robust. These are not attainable for applications that demand the formation of huge and dense-based sensor networks, and so are several IoT applications. The concept of analyzing the data of pH sensors using machine learning was considered, and fault data detection is not included in this paper [9]. Singhania et al. [10] suggested dense-based sensor networks. In horticulture, for instance, sensors must be positioned to provide heavy coverage and precision via huge and dense-based sensor networks. Utilizing many eminently precise but valuable sensors will surpass the deployment rate. As a result, uttermost IoT appliance employs low-price sensors, albeit with the consumption of data quality. Using either corporation degree or low-price sensors leads to tremendous time and repair activities because a skilled person must fade

out into the area to assess and renovate the whole network of the sensors to support the data. Aside from that, retransmitting data when correct data errors (lacking information) occur does not perform well in an IoT-based application. It is because the devices in the network are mechanized by restricted battery and recollection. It is costly to undermine the capability and computation power to send back the missing data over the network, particularly if there is a large pack of data to retransmit. The process of retransmission furthermore causes slow determination, which can point to defective results. The presentation of wireless gas observed on an internet server will allow users to closely monitor the value of the construction or commercial air environment from a distance.

Namuduri et al. [11] suggested the deep learning techniques that can be used on the data collected by these sensors. It is likely that information from electrochemical and solid-state sensors will be used more frequently for predictive maintenance. A potential agricultural sensor for resolving the quality of water is one example. In a typical predictive maintenance scenario, the inceptive data produced from the destination sensor would be combined with information from extra environmental sensors that could influence the target sensor's degradation. In the analysis, only a few variables, such as moisture, temperature, pH, pressure, and other metals and ions are considered. Zhang et al. [12] proposed a system to identify potentially harmful gas combinations in the smart kitchen using a wireless transmission network software layout. Six commercial MOS sensors were used in the sensor array, each of which was cross-sensitive to three different types of dangerous gases. To apprehend the awareness stages of three dangerous gases, SVM models were trained using features acquired using two methods. A five-fold cross and validation were employed for all target gases to assess and examine the precision of various methods. Compared to the traditional feature extraction method, the results showed that wavelet time dispersal can extract features more successfully. The CO, CH₄, and CH₂O models based on wavelet time dispersing characteristics had a correctness of 98.73 percent, 100 percent, and 97.46 percent, respectively. It gives a realistic popularity technique and detection stage for multigas sensing applications. This paper considered only three numbers of gases. Makiko Kawada et al. [13] developed a new method to enhance the correctness of the leakage price computation. The gas strain sensor has sufficient features, a determination of 20 Pa, and a stableness of 0.004%, consistent with the year. The new set of rules of the leakage price computation has eliminated intervention because of solar radiation and climate. This algorithm helped the users find the gas leakage only from gas-insulated switchgear (GIS).

Chen [14] proposed a fault detection technique simulation entrenched ordinary statistics evaluation in wireless sensor networks. It has a spatial clustering optimization method and generally groups to find the facts of float within the wireless community time window over the clustering method. It has determined the clustering information to discern the findings of ordinary community occasions. It is able to keep the traits of events and facilitate categorize the

odd information occasions. This paper will justify the reliability and predominance of the advanced optimization set of rules over the simulation era. Experiments demonstrate that the fault detection fee primarily depends on peculiar information analysis as excessive as ninety-seven percent, which is five per cent greater than the usual fault detection charge. Simultaneously, the comparative fault false detection value is low and restrained below one percent. The capability of this set of rules is around ten percentage greater than that of the conventional set of rules. The comparable data include humidity, temperature, voltage, and light and current values, and this method was not applied to the dataset, which deals with different harmful gases.

Tsai et al. [15] proposed a detection system to avoid irregularity amidst the sensors depending on the machine getting to know. The resumption precision relies on the Bayesian method precision because the recovery method is standardized with the Bayesian model prediction and sensor assessment. The precision of resumption is 95.5% for the dataset network aquatic microbial observing system. The author considered only a few parameters, such as humidity, temperature, and light. Xu et al. [16] proposed a gas concentration prediction set of rules based on a new version. The random regression tree and gradient boosting selection tree regression technique are decided on because of the fact the base learning devices help the new set of rules to utilize the output of every base getting to know the tool as an input to educate a new model to supply a final output via the new version. The grid search set of rules is examined to routinely advance the parameters so that the overall accomplishment of the complete gadget can extend the ideal values. The new version exhibited the first-class prediction impact for ethylene and carbon monoxide with the decency of in-shape data of up to 0.99 and 0.991. This model is the appropriate handiest for predicting carbon monoxide and ethylene.

3. Related Terminologies

An artificial neural network is a basis loosely formed after a brain's artificial neural systems. Its major purpose is to help in modelling techniques from difficult processes, including pattern realization. ANNs are made of a deep convolutional network of neurons, termed as a perceptron, in which every unit gets many original-valued inputs (associated utilizing a guidance function) and runs them by way of its activation value (e.g., sigmoid, linear, and reorganized linear unit), and they provoke an original-valued output. Each particular input is associated with a weight, which decides the input data's participation in the outcome. The basis for training an ANN to gain knowledge is to understand the density of the input rates so that it generates the appropriate output rate. There are numerous methods for achieving this, such as the perceptron principle for binary classification gradient descent datasets, nonlinear datasets, and back propagation [11].

A support vector machine is a machine learning-based technique that seeks a hyperplane in an F-dimensional space to distinguish and categorize data points. The characteristics are derived precisely as parables or through a process known

as remodeled engineering, which generates new features, depending on the evidence and its port of independent values. The hyperplane is a determination dividing line, in which information points on one part are about a particular class and information points on the other side correspond to the next. The goal is to identify a hyperplane with the greatest boundary that is the greatest distance of 2 information points from various groups. Support vectors are information scores nearer to the hyperplane and are used to determine the hyperplane by maximizing the classifier's margin. The determination boundary of the general data is identified for the intrusion detection system, containing most of the data in the peculiarity space. Then, new data deviating from the border are labeled exceptions [12]. As the number of internet users grows and new technologies arise, such as the Internet of Things (IoT concept), new and ongoing methods to pervade computer systems and networks emerge. Some organizations are investing more in the investigation to identify these attacks. By correlating the best value of correctness, institutions are using clever approaches to test and verify [17]. Support vector machines are a group of supervised learning techniques for classifying and predicting multidimensional datasets. This class of classifiers can minimize the empirical classification error and maximize the geometric margin. A maximum margin classifier is another name for an SVM [18].

Clustering is an unsupervised fault detection method that does not necessitate precedent information about the systematic technique or underlying information distribution. It comprises several processes, beginning with preprocessing, in which the fixed-width clustering method segregates and categorizes the data in the dataset. Each corresponds to a cluster in the corrected clustering process, and the information score is into the interior part before the distance from the cluster's center. A new one is formed if no such group exists, with that piece of data as its center. Following that, the anomaly identification process classifies each group as "normal" or "outlier." The user must, therefore, decide the ideal number of clusters or cluster width. It is accomplished by computing the Euclidean distance amidst two clusters. If a cluster's typical intercluster distance is more than one standard error out of the way from the average cross distance, it is considered an outlier. The variables in the outlier groups are then examined carefully with the help of nearest neighbors and timestamps to establish whether they are events or true abnormalities. Harmful gases in the industrial area are detected and shown on display. The appropriate decisions have been taken to reduce pollution in the commercial plant and ensure a healthy job atmosphere for the employees [19].

A Bayesian Network, often called a belief network, is a problematic graphics method based on Bayesian inference. It employs a counselled acyclic graph to represent several elements and their provisional relations. It can be utilized to estimate the probability values of an unknown point utilizing data from those other parameters. As per the chain rule of probability, the joint possibility distribution of the elements a, b, c, and d is expressed [20]. It also fulfils the local Markov substance, which expresses that provided its parental values, each parameter is unrelated to its nondescendants, reducing

the chain rule. This architecture uses the sensor node idea for the gas parameter layer. The attributes and parameters for the hazardous gases that the gadget assesses are all the same in principle and customized for each gas (CO and NO₂). A "gas sensor" and a "gas sensor driver" will be included in each sensor node. The driver board comprises multiple components that enhance and transform the sensor's output current. A connector board, external resistor, and low pass filter have been used to regulate the voltage [21].

Anyhow, the statistical testing of a slop distribution is essential for the reliable detection of gas leaks [13]. In the gas leakage detection system [13], based on the filtered data of diminished gas leak rate, the gas pressure was determined using the regression analysis. This method is generally used to determine the rate of gas leakage. The GBDT regression method employs the boosting model, which consists of two main parts: the core values of the GBDT regression technique and the decision tree values. Among the frame values of GBDT, the sampling portion "subsample" and the learning data "learning rate" are the most important. The GBDT regression algorithm takes the biological sample, and when the sampling proportion equals 1, it uses the learning rate. In other words, the samples are all sampled, identical to the nonsampling process. In regression analysis, one or more variables are modeled and analyzed to determine the relationship between the dependent and independent variables. The objective of regression analysis is to determine how the characteristic value of the established variable changes whilst any impartial variable is varied, at the same time as the other impartial variables stay regular [22].

Fault data in sensor measurements to infuse NOISE faults: we take a series of models and inject an arbitrary value sapped from general distribution to every sample in the severity of NOISE faults (Table 1). For median and high-level intensity faults, as well as problems with a reasonably long duration generally, the 1-step ahead forecasting situated approach is beneficial for identifying concise to intermediate duration faults. Still, the 1-step ahead anticipating positioned technique is essential [23]. When faults last for an extended period [4], as expected, methodologies can diminish the number of false positives and negatives. The positives are not true. Detecting errors with two (or more) procedures in succession can result in more false positives. Outliers and events are two types of outliers. In natural tracking, sensor knowledge can vary from predicted rates because of an unforeseen event or external part of familiar explanations (outliers). If the faulty specimens do not meet some established fault techniques, it is adorable that the result has attended different or are merely extreme values. Contingent information around the sensors and the phenomenon being observed, if these data result from an exploring alternative in this setting.

4. Proposed System

In multilayer artificial neural network, Figure 1 depicts the system architecture of an automated IOT-based smart harmful toxic gases and environment monitoring system that measures the concentration of a certain gas in the air to

TABLE 1: Comparison of algorithms with used dataset.

Author	Algorithm/method	Dataset	Metrics	Purpose
Sharma et al. [4]	Rule-based (heuristic) methods, histogram-based method, and hidden Markov model (HMM)	Berkeley dataset, Great Duck Island (GDI) dataset, NAMOS dataset	Accuracy, false positive, false negative	For diagnosing faults, heuristic rules are developed to detect abnormal sensor behavior based on sensor correlations. Anomaly sensor measurements are flagged as faults by leveraging this information
Chen [19]	Centralized anomaly data detection	Real time dataset	Fault detection rate	To enhance the fault detection technique of correlated irregular information analysis and abnormal event detection
Tuan Anh Nguyen [23]	Fault-injection algorithms	Sensor Scope dataset	Fault detection rate	To detect and classify faults in sensor data
Mahmudul Hasan	Artificial neural network Support vector machine Decision tree	DS2OS	Accuracy, precision, recall	Identify the attack and detection anomalies in the internet of objects (IoT) infrastructure
Tsai et al. [15]	Automatic threshold selection, impermanent faults, fault recovery	Airbox open data, Intel lab data	Accuracy, fault detection rate	Design the architecture of detection system and prevent anomaly amongst sensors based on machine learning

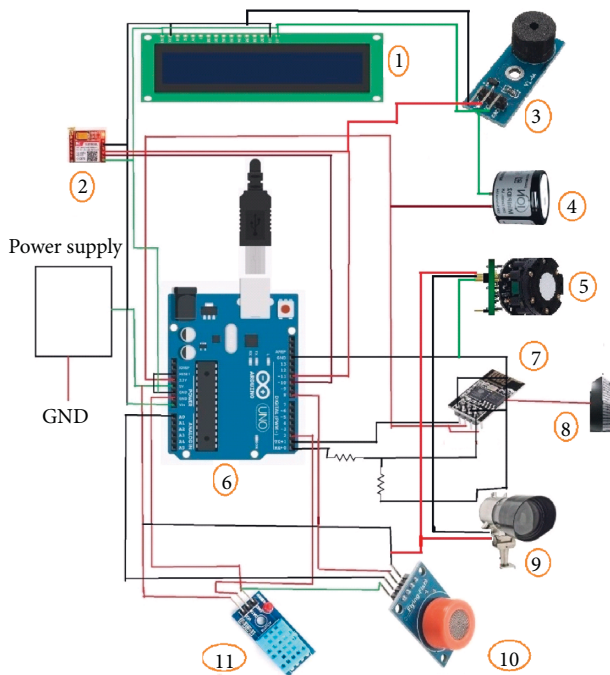


FIGURE 1: Hardware components of automated IOT-based smart harmful toxic gases and environment monitoring system. 1. LCD display, 2. SIM, 3. DC active buzzer, 4. MiniPID 2 HS PID sensor, 5. AQ3 sensor, 6. Arduino board, 7. ESP8266 WIFI module, 8. Cloud server, 9. IR5500 open path infrared gas detector, 10. MQ3 sensor, and 11. Temperature and humidity sensor.

monitor toxic gases and environmental conditions [24]. This system is designed for stationary gas detection using remote or local sensors. Depending on the different gases involved and the size of the work area, one or more types of gas detectors may be utilized in the same area. Globally, air quality is deteriorating, and indoor air pollution poses a severe hazard to the health of humans. Pollutants in indoor air are often made of a variety of volatile organic compounds (VOCs) that are generally toxic to humans, particularly those with low molecular weights [25].

Furthermore, multiple VOCs may be present in some conditions. Hence, a device that can detect one or more VOCs simultaneously would be the most advantageous. This system combines various types of sensors and technology, such as the AQ3 gas sensor, Grove gas sensor (MQ3), DHT11 sensor, MiniPID 2 HS PID sensor, open path infrared (OPIR) industrial gas sensors, ESP8266 WIFI module, active buzzer, and so on. Because it has been designed to operate in a wide range of environments with extreme conditions, it is highly specific and intended to measure the gas [26]. The AQ3 gas sensor functions in various situations and challenging circumstances. AQ3 gas sensors detect gases, such as nitrogen dioxide, ozone, carbon monoxide, and sulfur dioxide (SO₂). They will, nevertheless, respond to various gases to some extent, achieving individual temperature and cross-sensitivity adjustment. The irregularity detection manner may also easily assess or contain a greater complex analysis. Therefore, the power signature seize is coordinated with the chosen IoT operation using the microcontroller program. The abnormality detection manner knows which operation is related to the power signature [27]. Therefore, it no longer wants to perceive the IoT operation from the obtained signature. It simply assesses if the received power signature is in the ordinary power intake variety of the comparable IoT operation [28].

The central detection of anomalies relies mainly on the spatiotemporal association standard of anticipated data. The hypothesis of the anomalous data detection technique is the feasibility of information failure of all nodes within wireless sensor networks. The respective single node within the wireless sensor network is served by data v_i [29]. The concerned a_i denotes the sensing data node like the respective single node v_i . The respective parameters of the respective detection accuracy rate and the respective warning rate are C and so on [19]. They will respond to various gases, achieving individual temperature and cross-sensitivity adjustment. The Grove gas sensor module can be utilized to find gas leaks. It can find benzene, CH₄, hexane, alcohol, LPG, and carbon monoxide. Measurements can be directed very soon to their high-level sensitivity and quick

reaction time. The potentiometer can be utilized to revise the sensitivity value of the sensor [30].

DHT11 is a low-cost digital temperature and humidity sensor. The connection to this sensor is made simply to any microcontroller, such as an Arduino, to compute the value of humidity and temperature of the environment in real-time [31]. It can detect humidity levels ranging from 20% to 90%. It is one of the sensors that belong to the DHT series. Its purpose is to read the sensor's humidity and temperature. The digital number from which the output of the DHT sensor must be read is called PIN. Industrial gas sensors with open path infrared (OPIR) detect hydrocarbon gases, silicones, hydride gases, and halogenated hydrocarbons in vast, open regions. Because open path infrared gas detectors work similarly to point infrared gas detectors, with the exception that their detection path can be extended to more than 100 meters, they can identify big and small gas leaks by monitoring LIL-m and ppm-m ranges [32]. Because the MiniPID 2 HS PID sensor is more sensitive and capable of detecting trace quantities of VOCs, such as carbon, excluding carbon dioxide, carbon monoxide, carbonates or metallic carbides, carbonic acid, and ammonium carbonate, it is employed in this invention. Because of its straightforward connectivity, ESP8266 is utilized as an access point (AP Mode) to provide wireless internet access to any microcontroller-based design [33].

An active buzzer alarm module for Arduino is an audio signaling device that is more expensive than a passive buzzer but easier to manage. Buzzers are frequently utilized for timers, alarms, and the affirmation of human input, such as a keystroke or mouse click. ThingSpeak is a cloud-oriented IoT analytics platform that enables users to gather, view, and analyze real-time data streams [34]. Data may be supplied to ThingSpeak from our devices, and web services can be used to build rapid representations of live data and deliver warnings/alerts. To set up ThingSpeak, a user account and a channel would be created. Install Thing Speak Communication Library for Arduino and configure with my Channel Number and my Write API Key variables. Every 20 seconds, it takes an analogue voltage from pin 0 and publishes it to a ThingSpeak channel. We utilized ThingSpeak to communicate several values to ThingSpeak from Arduino. For each value to send, we used set field (#, value) and Thing Speak. With the help of ThingSpeak, data can be sent from devices/sensors to the cloud and stored in either a private or public channel.

Figure 2 depicts the various types of sensors that continuously monitor the gases in the environment and transfer data to a server for storage and future use. While continuously monitoring, an alert will be raised if any gas levels in the air exceed the normal range. An alert notification will be sent to the organization's safety control board, workers' mobile stations, and the nearest police station alone. If necessary, it sends the data to the cloud and then plots the sensor data in the graphical form [35].

As illustrated in Figure 3, it is important to ensure that the sensor data is in the correct format or structure before applying an ML algorithm [36]. Preprocessing is the term for this. Denoising and dimensionality reduction are the

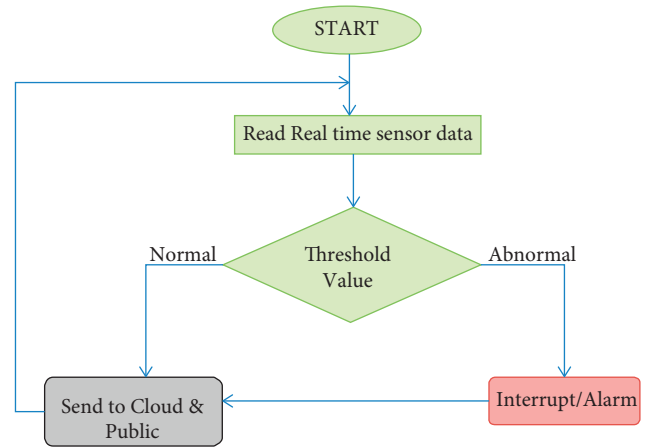


FIGURE 2: Process flow diagram of a Gas monitoring and controlling system.

examples of preprocessing procedures. Eliminating noise from a signal is known as denoising to increase the signal ratio to noise. By obtaining a set of principal variables, dimensionality reduction reduces the number of random variables considered. It can be split into two categories: feature extraction and feature selection. Dimensionality reduction is lowering the number of independent variables in a dataset while preserving as much information as possible. The data collected/gathered will be analyzed continuously.

4.1. Error Detection. Multilayer artificial neural networks (MLANNs) and the hidden Markov model are utilized to detect mistakes in the sensor dataset. Hidden layers are found in multilayer artificial neural networks. These will, of course, be utilized for more complex tasks than the perceptron. ANNs with one hidden layer, and it simplifies with the feed forward back propagation algorithm. Feed-forward networks are those one can input on the left and reproduce in advance to have an output. An ANN with 2 sigmoid parts in the hidden layer is shown below. All of the units' weights have been set arbitrarily. It is a multilayer feed-forward ANN that analyzes a signal's sliding windows to permit translating between the past and current values. The examples are based on the same real-time-oriented datasets we are using to assess the occurrence of sensor defects. More information on these datasets is provided posterior in the substance. In real datasets, some sensor reading errors were discovered. These examples provide readers with a visual understanding of the types of defects that can occur in practice and motivation for the fault model.

The real-time data that must be included in the sensor can be obtained from the internet from all over the world, and the time series sensor data will be included in forecasting future beliefs and detecting abnormalities, with the variance being found to be comparable to a specified limit and being unconnected to the fundamental physical phenomena. There are three types of faults: the first type is a constant fault, which defines the high number of consecutive samples, and the sensor reports a consistent value. Compared to "typical"

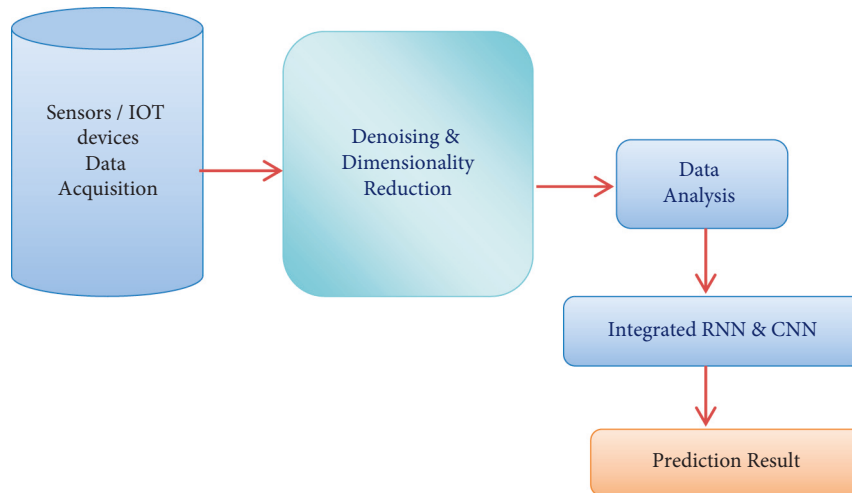


FIGURE 3: Stages of data preprocessing, analysis, and prediction.

sensor readings, the described constant rate is very low or very high. The second type is a short fault, which is the difference between two consecutive data points, with a significant variation in the actual values. The third type is NOISE fault. The measured value becomes more variable. NOISE faults, except short faults, influence a single sample simultaneously with the impact of a series of samples. SHORT and NOISE faults were first distinguished and personalized but only for a single dataset to categorize the three categories previously defined as expressing the “data-centric” perception of fault classification, i.e., the fault kinds are expressed in terms of incorrect data characteristics.

Convolutional neural networks (CNNs) can be used to discover anomalies in this system, which involves using time series sensor data to predict future values and comparing the deviation to a predetermined threshold. While it is unimaginable to determine the main driver for sensor blames all the time, a few framework (equipment and programming) flaws have been known to bring about sensor issues. The run-of-the-mill equipment has been seen to motivate sensor issues to incorporate harmed sensors, short-circuited associations, low battery, and adjustment mistakes. A learning-based strategy may be more suitable for phenomena that are not spatially. As an illustration, if the sequence of “normal” sensed data and the impact of the sensor’s faults on the described information for a sensor is well-comprehended, we can develop a framework for the dimensions observed by that sensor utilizing learning-based techniques, such as integrated hidden Markov models and artificial neural networks. Fault detection plays an important role in identifying permanent (very long-lasting) missing defects (lack of sensor readings). Long term irregular detection components save the data from real-time irregular detection components, while the real-time irregular detection components continue to determine whether each sensor is malfunctioning. As many parameters must be trained on a regular basis and because these operations must operate with enormous amounts of data, they must be batch processing jobs.

4.2. The Types of Faults and Threshold Selection. The threshold is performed by comparing the real measurement value to the predicted value to determine if the data is erroneous. The threshold can easily influence the accuracy. False negatives rise if the threshold is set too low, and the value of a false positive is enhanced if the threshold is set too high. As a result, the threshold could be carefully chosen to balance both false negatives and positives. Because the dataset contains incorrect data and the estimated error may be disparate during the validation phase, we must first discard the highest variability estimation error and choose other maximal estimation errors as the threshold.

- (1) Sorting estimation errors in ascending order
- (2) Sorting the estimation error sequence into categories.
- (3) Determine the slope of each group.
- (4) Determine the difference in slope between the two groups. The impermanent fault sensor showing is periodically good, and sometimes, it is terrible. This problem shows up as very high or very low information. The data from the sensor is combined with noise information. Hence, a noise fault is recognized if there are no long continuous faults in the sensor’s log.

Successive error detection is typical in persistent faults. Thus, when a successive error is identified, all we have to do is identify it as (by observing standard deviation ST) the following:

- (i) Absolute (continuous) faults: compute the standard deviation of sequential sensor readings, where ST is zero.
- (ii) Bias drift faults: while the ST should be modest, the bias was introduced as an offset in the output.
- (iii) Degradation faults: a defect that constantly changes and is mostly driven by drift velocity. Analyze the differences between the measurement and the

predicted value to see if standard deviation has to be raised.

The verification of event anomaly detection optimization techniques falls under the second category. The offset reading of fault types precalibration algorithm has been verified, and the random reading fault types algorithm has detected. The experiment mainly validates the algorithm's detection accuracy and the common error detection value for the performance test of the algorithm before calibration.

5. Results and Discussion

If at least one of the methods flags a sample as bad, this method flags it as problematic. The goal of this strategy is to reduce false negatives. It is, however, susceptible to false positives. The accuracy and robustness of these procedures must be assessed. We should ensure no incorrect samples in the dataset before adding flaws. To achieve this, we used a real-world dataset, including observations, to introduce the defects of the sorts into sensor measurements.

Furthermore, because we did not have any field data on the defects for this dataset, we had to depend on various visual inspections and feedback that the dataset was fault-free. There are two benefits to using the same strategy. For starters, inserting flaws into a dataset provides a precise "ground truth" that can be used to comprehend how well an identification system performs. Furthermore, we may control the severity of a fault, allowing us to investigate the limits of every recognition method's efficiency and analyze various methods at low-level fault intensities. Even though several of the faults we've seen in current real-world datasets are of quite a high-intensity rate, we regard it critical to know how mistake detection techniques behave over a range of fault intensities. We also incorporate defects in the training data to build the hybrid HMM and ANN model. For each fault, we confer the detection achievements of several approaches. The number of errors discovered, false positives, and negatives are three indicators we use to assess the effectiveness of various approaches.

The AQ3 carbon monoxide gas sensor uses 4 electrode electrochemical technologies. It has a measurement range of 0 ppm CO to 10 ppm CO. Its maximal overload is 50 ppm CO without a board and 30 ppm CO with the board. Constant fault occurs when a sensor produces a constant rate for several consecutive samples. The time-series sensor data will be used to anticipate future values/data and detect irregularities by comparing deviations to a preset threshold unrelated to the underlying physical events. As illustrated in Figure 4, the highlighted data in red color are the injected data, which is valuable in the training dataset. Our paper focused on two phases: phase 1 is to detect the toxic gases from the environment using different types of effective sensors. The live data sensor reading is sent to the cloud through the internet. Phase 2 focuses on monitoring, early detection, and prediction of toxic gases in the environment. To improve the early detection and prediction of toxic

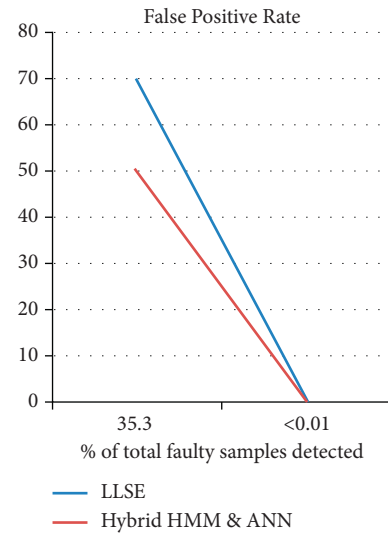


FIGURE 4: Comparison of false positive rate of the hybrid HMM and CNN with the existing method (LLSE).

gases, it is important to improve the data analysis on the given dataset. We are using a hybrid model of ANN &HMM to detect the fault data in sensor reading. We tested our hybrid model in different ways. The first way is that we took 200 sensor reading samples from AQ3 sensor and injected 100 fault data into the hybrid model of ANN &HMM. Our model performed well and produced a good result in detecting the fault data, as shown in Figure 5 and it was compared with the linear least squares estimation method and hidden Markov model. We considered the dataset gas sensor array under dynamic gas mixtures from which we took sensor readings of 50 carbon monoxide samples. We injected 20 fault data, and our model produced good results. The analyzed data will help improve the prediction accuracy.

The defective samples are detected using the ARIMA one-step and L-step approaches. The ARIMA (one-step) method is better suited for identifying SHORT defects than the ARIMA (L-step) method. The Hybrid model of ANN &HMM is compared with the ARIMA one-step model, which produced a detected percentage of total faulty samples of 96.7. Assume that the L value is 120. Its false positive rate is 0.02, which is greater than the false-positive rate of the proposed method, as shown in Figure 5, and the proposed method produced 81.2 percentage of faulty samples, which is less than the ARIMA one-step model. Hence, our model outperformed well with a good false-positive rate.

The performance of the hybrid model of ANN &HMM is compared with that of the ARIMA one-step model. The detected percentage of total faulty samples is 76.7, with a false positive rate of 3. It is greater than the false positive rate of the proposed method and the ARIMA one-step. The proposed method produced 81.2 per cent of the faulty samples, which is greater than the ARIMA L-step model. Hence, our model outperformed well, with a good false-

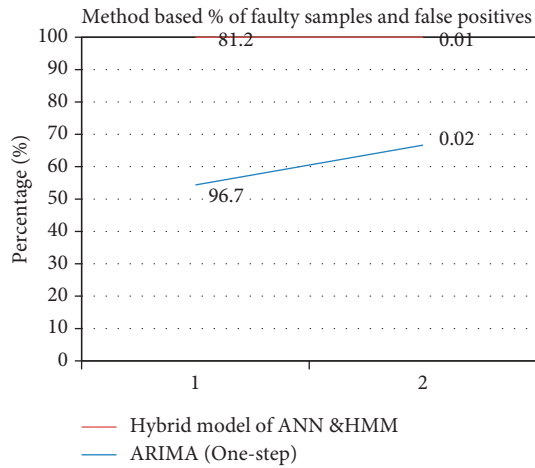


FIGURE 5: Comparison of % of faulty samples and false positive rate of the hybrid HMM and CNN with the ARIMA (one-step) model.

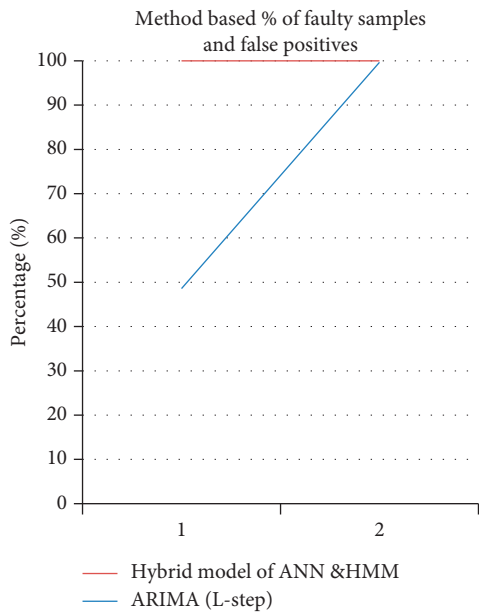


FIGURE 6: Comparison of percentage of faulty samples and false positive rate of the hybrid HMM and CNN with the ARIMA (L- step) model.

positive rate and percentage of total faulty samples detected, as shown in Figure 6. Our proposed model has increased the 0.10 percentage accuracy compared with the precalibration algorithm.

The fault data detection accuracy of the hybrid model of ANN &HMM proposed in this paper is significantly higher than that of the precalibration algorithm but lower than the accuracy of the random reading algorithm. It can be seen from Figure 7 that the hybrid model of ANN &HMM is compared with the error rate of existing algorithms. The algorithm proposed in this paper has low error rate than both algorithms, such as the precalibration algorithm and random reading accuracy and a lower error detection rate when dealing with ANN &HMM, as shown in Figure 8.

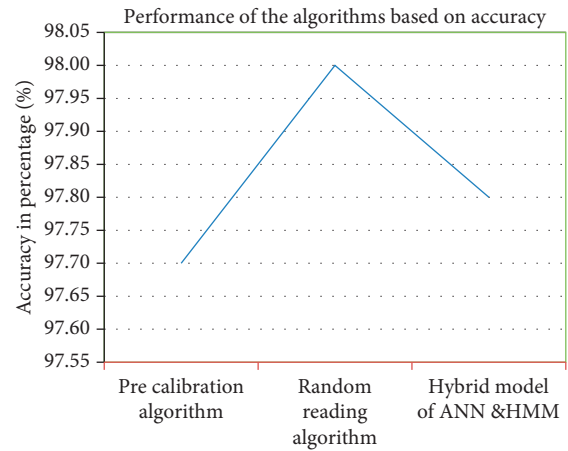


FIGURE 7: Comparison of the accuracy of the hybrid HMM and CNN with the ARIMA (one-step) model.

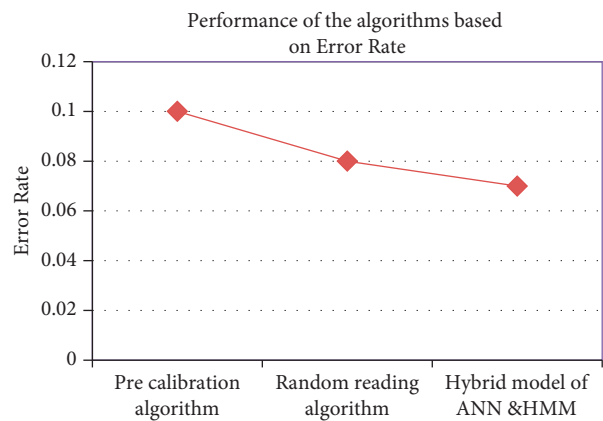


FIGURE 8: Comparison of the accuracy of the hybrid HMM and CNN with the ARIMA (L- step) model.

6. Conclusion and Future Scope

As a result of growing industrialization and concerns about the global impact of air pollution, there is a growing demand for automated gas monitoring and control systems. There are several flaws in traditional methods for evaluating and identifying the presence of numerous different types of gases in a sample. To address this issue, we designed an IoT-based dangerous gas monitoring system that will employ several sensors and components, including AQ3 gas sensors, that can sense a variety of gases: ozone (O₃), carbon monoxide (CO), nitrogen dioxide (NO₂), and sulfur dioxide (SO₂) are mostly used for environmental monitoring in the ambient environment. Our newly proposed IOT-based toxic gas monitoring system performed well. To improve the accuracy of the early prediction of toxic gas existing in the environment and to reduce the false positive rate of the model, effective data analysis is mandatory. Hence, we have applied the hybrid HMM and the ANN method to detect the fault data, which exist in the sensor dataset. The hybrid HMM and ANN fault detection methods performed well on the datasets and produced a 0.01% false positive rate. The fault data

detection accuracy of the hybrid model of ANN and HMM proposed in this paper is significantly higher than that of the precalibration algorithm. Still, it is lower than the accuracy of the random reading algorithm. The proposed system will be tested on different datasets with different parameters or features in future work.

Data Availability

The data used to support the findings of this study are included within the article. Should further data or information be required, they are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors thank Chandigarh University, Punjab, for providing characterization support to complete this research work. Authors declare that no funding was received for this research and publication.

References

- [1] A. Mahapatra, S. Parida, S. S. Surana et al., "Energy auditing for efficient planning and implementation in commercial and residential buildings," *Advances in Civil Engineering*, vol. 2021, 10 pages, 2021.
- [2] V. Balaji, S. Kaliappan, D. S. Ezhumalai, S. Boopathi, P. Pravin, and S. Mani, "Combustion analysis of biodiesel-powered propeller engine for least environmental concerns in aviation industry," *Aircraft Engineering & Aerospace Technology*, vol. 94, no. 5, pp. 760–769, 2022.
- [3] K. R. Vaishali, S. R. Rammohan, L. Natrayan, D. Usha, and V. R. Niveditha, "Guided container selection for data streaming through neural learning in cloud," *International Journal of System Assurance Engineering and Management*, vol. 16, pp. 1–7, 2021.
- [4] P. Asha, L. Natrayan, B. T. Geetha et al., "IoT enabled environmental toxicology for air pollution monitoring using AI techniques," *Environmental research*, vol. 205, Article ID 112574, 2022.
- [5] C. S. S. Anupama, L. Natrayan, and L. Laxmi, "Deep learning with backtracking search optimization-based skin lesion diagnosis model," *Computers, Materials & Continua*, vol. 70, no. 1, pp. 1297–1313, 2021.
- [6] K. Seeniappan, B. Venkatesan, N. N. Kandhasamy, S. Arunachalam, R. K. Seeta, and M. V. Depoures, "A comparative assessment of performance and emission characteristics of a DI diesel engine fuelled with ternary blends of two higher alcohols with lemongrass oil biodiesel and diesel fuel," *Energy & Environment*, vol. 13, Article ID 0958305X2110513, 2021.
- [7] G. Shi, Y. He, Y. Zhang, B. Ali, and F. Ali, "Detection and determination of harmful gases in confined spaces for the Internet of Things based on cataluminescence sensor," *Sensors and Actuators B: Chemical*, vol. 296, Article ID 126686, 2019.
- [8] S. Sun, Z. Yunxia, and Y. Zhang, *Monitoring of harmful gas for Yangbajing 1# tunnel in high altitude plateau*, in *Proceedings of the 2014 Sixth International Conference on Measuring Technology and Mechatronics Automation*, pp. 410–412, Zhangjiajie, China, January 2014.
- [9] G. Takami, M. Tokuoka, H. Goto, and Y. Nozaka, "Machine learning applied to sensor data analysis," *Yokogawa Tech. Rep. English Ed*, vol. 59, no. 1, pp. 27–30, 2016.
- [10] M. E. Porter and C. v d. Linde, "Toward a new conception of the environment-competitiveness relationship," *The Journal of Economic Perspectives*, vol. 9, no. 4, pp. 97–118, 1995.
- [11] S. Namuduri, B. N. Narayanan, V. S. P. Davuluru, L. Bhansali, and S. Bhansali, "Review-deep learning methods for sensor based predictive maintenance and future perspectives for electrochemical sensors," *Journal of the Electrochemical Society*, vol. 167, no. 3, Article ID 037552, 2020.
- [12] J. Zhang, Y. Xue, T. Zhang et al., "Detection of hazardous gas mixtures in the smart kitchen using an electronic nose with support vector machine," *Journal of the Electrochemical Society*, vol. 167, no. 14, Article ID 147519, 2020.
- [13] M. Makiko Kawada, T. Tadao Minagawa, E. Eiichi Nagao, M. Mitsuhito Kamei, C. Chieko Nishida, and Koji Ueda, "Advanced monitoring system for gas density of GIS," in *Proceedings of the 2008 International Conference on Condition Monitoring and Diagnosis*, pp. 363–368, Beijing, China, April 2008.
- [14] X. Chen, "Fault detection method and simulation based on abnormal data analysis in wireless sensor networks," *Journal of Sensors*, vol. 2021, Article ID 6155630, 11 pages, 2021.
- [15] F. K. Tsai, C. C. Chen, T. F. Lin, and T. J. Lin, "Sensor abnormal detection and recovery using machine learning for IoT sensing systems," in *Proceedings of the 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA)*, pp. 501–505, Tokyo, Japan, April 2019.
- [16] Y. Xu, R. Zhao, and X. Zhao, "Research on a gas concentration prediction algorithm based on stacking," *Sensors*, vol. 21, no. 5, pp. 1597–1617, 2021.
- [17] K. A. P. da Costa, J. P. Papa, C. O. Lisboa, R. de Albuquerque, and V. H. C. de Albuquerque, "Internet of Things: a survey on machine learning-based intrusion detection approaches," *Computer Networks*, vol. 151, pp. 147–157, 2019.
- [18] W. Khalaf, C. Gaudio, and M. Gaudio, "Least square regression method for estimating gas concentration in an electronic nose system," *Sensors*, vol. 9, no. 3, pp. 1678–1691, 2009.
- [19] E. N. Sathishkumar, K. Thangavel, D. Arul, and P. Daniel, "Effective clustering algorithm for gas sensor array drift dataset," *Int. J. Comput. Intell. Informatics*, vol. 3, no. 3, 2013.
- [20] M. P. Moraru, M. Pesko, C. Fortuna, and D. Mladenic, "Using machine learning on sensor data," *Journal of Computing and Information Technology*, vol. 18, no. 4, pp. 341–347, 2010.
- [21] V. K. Ojha, P. Dutta, H. Ghosh, and S. Ghosh, "Linear regression based statistical approach for detecting proportion of component gases in manhole gas mixture," in *Proceedings of the 2012 1st International Symposium on Physics and Technology of Sensors (ISPTS-1)*, pp. 17–20, Pune, India, March 2012.
- [22] B. de Bruijn, T. A. Nguyen, D. Tei, and K. Tei, "Benchmark datasets for fault detection and classification in sensor data," in *Proceedings of the 5th International Conference on Sensor Networks*, pp. 185–195, Coventry, Great Britain, January 2016.
- [23] D. Thompson and H. Wang, "Integrated power signature generation circuit for IoT abnormality detection," *ACM Journal on Emerging Technologies in Computing Systems*, vol. 18, no. 1, pp. 1–13, 2022.

- [24] S. S. Sundaram, N. Hari Baskar, and L. Natrayan, "Smart clothes with bio-sensors for ECG monitoring," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 4, pp. 298–301, 2019.
- [25] S. Raja and A. J. Rajan, "A decision-making model for selection of the suitable FDM machine using fuzzy TOPSIS," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–15, Article ID 7653292, 2022.
- [26] S. Magesh, V. R. Niveditha, P. S. Rajkumar, and L. Natrayan, "Pervasive computing in the context of COVID-19 prediction with AI-based algorithms," *International Journal of Pervasive Computing and Communications*, vol. 16, no. 5, pp. 477–487, 2020.
- [27] S. Dhall, B. R. Mehta, A. K. Sood, and K. Sood, "A review on environmental gas sensors: materials and technologies," *Sensors International*, vol. 2, Article ID 100116, 2021.
- [28] S. K. A. I. Khan and S. K. Das, "Ensuring energy efficient coverage for participatory sensing in urban streets," in *Proceedings of International Conference on Smart Computing (SMARTCOMP)*, Hong Kong, November 2014.
- [29] P. Trivedi, D. Purohit, A. Soju, and R. R. Tiwari, *Major industrial disasters in India An ofcial newsletter of ENVIS-NIOH*, vol. 9, p. 7, 2014. Available online:http://niohenvnis.nic.in/newsletters/vol9_no4_Indian%20Industrial%20Disasters.pdf.
- [30] B. Sharma, L. Govindan, and R. Govindan, "Sensor faults," *ACM Transactions on Sensor Networks*, vol. 6, no. 3, pp. 1–39, 2010.
- [31] T. Leitner, C. Feilmayr, and W. Wöß, "Early Detection of Critical Faults Using Time-Series Analysis on Heterogeneous Information Systems in the Automotive Industry," in *Proceedings of the The Third International Conference on Data Analytics*, pp. 70–75, Rome, Italy, August 2014.
- [32] F. Wei and L. Wei, "Harmful gases wireless network monitoring system design," in *Proceedings of the 2014 International Symposium on Computer, Consumer and Control, Consumer and Control*, pp. 551–553, Taichung, Taiwan, June 2014.
- [33] G. Kanimozhi, L. Natrayan, S. Angalaeswari, and P. Paramasivam, "An effective charger for plug-in hybrid electric vehicles (PHEV) with an enhanced PFC rectifier and ZVS-ZCS DC/DC high-frequency converter," *Journal of Advanced Transportation*, vol. 2022, Article ID 7840102, 14 pages, 2022.
- [34] K. Nagarajan, A. Rajagopalan, S. Angalaeswari, and S. Natrayan, "Combined economic emission dispatch of microgrid with the incorporation of renewable energy sources using improved mayfly optimization algorithm," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 6461690, 22 pages, 2022.
- [35] L. P. Natrayan, S. S. Sakthi, and J. Elumalai, "Analyzing the Uterine physiological with MMG Signals using SVM," *International journal of pharmaceutical research*, vol. 11, no. 2, pp. 165–170, 2019.
- [36] D. K. Jain, S. K. S. Tyagi, S. Neelakandan, M. Prakash, and L. Natrayan, "Metaheuristic optimization-based resource allocation technique for cybertwin-driven 6G on IoE environment," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 7, pp. 4884–4892, 2022.