

Detection of Hazardous Gas Composition in Sewer Pipeline and to Identifying Safe Working Condition in Sewer Pipelines

J. Arunnehru
DEPT. of CSE
SRM University, Chennai

C. Sabarinathan
DEPT. of CSE
SRM University, Chennai

R. N. Naresh
DEPT. of CSE
SRM University, Chennai

M. Shreeram
DEPT. of CSE
SRM University, Chennai

R. Vaidyanathan
DEPT. of CSE
SRM University, Chennai

ABSTRACT

In this paper, we formulated around a certifiable issue identified with sewer-pipeline gas detection using the approaches which classified the problem into two categories, also known as classification-based approach. The primary objective of this project is to recognize the dangerous gases present inside sewer-pipeline to offer protected access to sewer-pipeline with the goal that the human fatalities, which happen due to presence of toxic gases can be avoided. The dataset is created through mixing all the different gases present in sewer pipelines under different situation to make sure all the cases are taken into consideration, these datasets were sorted out to plan a predictive model that could distinguish/characterize hazardous and non-hazardous circumstance of sewer-pipeline. To design such prediction model, classification algorithms were utilized and their performance were assessed and thought about, both exactly and factually, over the collected dataset. In addition, the project also will predict the composition of gases present in the sewer pipeline and display the values for the better understanding of the condition present in sewer pipeline. The final observation of this study demonstrated that the performed of instance based-learning algorithm were superior to numerous different algorithms, for example, multi-layer perceptron, support vector machine, and so on. Also, it was observed that multi-scheme ensemble approach improved the execution of base indicators.

Keywords

classification-based approach, predictive model, multi-layer perceptron, support vector machine

1. INTRODUCTION

In this paper we talk about formulating an answer to this critical problem for humans using latest technological advancement, to avoid many human fatalities due to manual scavenging. The technology used for this should be easily usable by the workers. In this paper, we have tackled a unresolved problem related sewer-pipeline environment nature, where sewer-pipeline safe environment detection and identification (in terms of non-toxic environment) was re-

quired to allow maintenance and cleaning of the pipeline. The detection of these gases is very complex and ambitious because of many different factors such as presence of mixture of several toxic gases, and there is no single gas sensor which will offer reliable result alone, we need to make use of multiple sensors to achieve desired results. Therefore, we observe the behaviour of the complex mixture of gases. The primary goal of this work is to offer a mechanism with high efficiency and accuracy to differentiate the toxic gases in a simple and straightforward way such that the end result is as simple as "safe or not safe". We make use of classification to tackle detection problem.

Sewer-pipeline are usually not opened for a long time due to which sunlight does not to enter the pipelines. This leads to generation of toxic gases from the organic waste present inside these pipelines. Sewer-pipeline contains a composite toxic gases such as hydrogen sulphide (H₂S), ammonia (NH₃), methane (CH₄), carbon dioxide (CO₂), nitrogen oxides (NO_x), etc. These gases occurs because of decomposition of the organic matter in anaerobic decomposition inside the sewer-pipeline. These gases are fatal for humans who come in contact with these gases or get exposure to these gases. Following this, a distressingly high number of human fatalities are occurring each year. The authorities who are responsible for maintaining and clearing of the waste in sewer pipelines are provided various electronic portable gas detectors available to determine the safety of the sewer-environment before entering the sewer pipelines physically for maintaining and cleaning. However, the available electronic portable gas detectors are not providing satisfactory results. Its evident from the incidents mentioned in the Indian Express article [1] where it has mentioned about the cause of deaths of workers while cleaning drain pipes. The Supreme Court of India expressed, "entering sewer lines without security equipment ought to be made a wrongdoing even in crisis circumstances." This persuaded us to complete our exploration in this domain and to turn out with a straightforward arrangement so that without having the base learning of the details of gas structure and safety limits, a man can comprehend nature of a sewer system before entering.

2. RELATED WORKS

Literature survey was conducted in the perspective of Electronic nose and gas-detection system available to cover broad area of research in the field, using intelligent computing techniques. Although there were not much research about sewage gas detection some few important contributions were found. VK Ojha et al.[5] reported a noticeable work on identification of hazardous gas in sewage pipeline where discussed about performance of various classification algorithms that has been evaluated.

VK Ojha et al.[6] have worked on Convergence analysis of back-propagation algorithm for designing an intelligent System for sensing manhole gases used back-propagation method to train neural network in order to identify concentration of gases using sensor with cross sensitivity. VK Ojha et al.[7] worked on performance of neuro swarm optimization algorithm to measure on detection of gases with cross sensitivity. The neural network is trained in supervised mode by particle swarm algorithm. VK Ojha et al.[8] reported work on portable gas sensor array to classify different gas inside sewage pipeline using pattern recognition technique. VK Ojha et al.[9] guides about use of sensor array instead of single sensor to effectively detect gas and also gives information on concentration of each gas present in the mixture. VK Ojha et al.[10] reported a work on comparing performance of neurogenic (neural network + genetic algorithm) with previous works results to detect various gas in sewage pipeline. Kelly J. Donham et al [11] reported an important work on exposure limit for various gases and aftermath after inhalation of various gas at certain concentration is discussed. Srivatsava et al [12] discussed about design issue of Electronic nose and also discussed about various machine learning algorithm that are used for detection of gases using sensors. Won So et al [13] discussed about the rate at various gas is released to the surrounding or system affects the sensitivity of sensor and detection of E-Nose. Eduard Llobet et al [14] discussed about single tin oxide sensor for checking air quality of a surrounding or system. They also empathize on using DWT instead of traditional fourier transform.

Lindel k weaver [15] has given a valuable data on Carbon monoxide poisoning and aftermath of CO inhalation at various concentration. All the above works talks about the use of machine learning techniques to detect gas in the environment and identify is it safe to enter the environment.

Limitations of the proposals

- (1) The above proposals does not talk about the oxygen concentration in the environment which is essential for human life.
- (2) The above proposals does not talk about the pressure inside the sewer system that can also can be cause of concern.
- (3) The above proposals does not concern with the external factors which can combustion of the gases present inside the sewer pipelines.
- (4) The above proposals does not deal with real life implementation or large scale implementation.
- (5) The above proposal does not have a network related element or networking concept which will help with interconnection between the GSA placed at different openings of same sewer pipeline.

In this project we look at the sewer pipeline condition as safe to work in and not safe to work in. Hence the classification method is a better approach than other approaches available. Since we are dealing with multiple gases at the same time, we need to take cross-sensitivity of each sensors towards multiple gases into account to avoid any miscalculations. In our cases we mainly deal with 4 natu-

ral bio toxic gases, they are Methane (CH₄), Ammonia (NH₃), Carbon Monoxide (CO) and Hydrogen Sulphide (H₂S). To tackle the cross-sensitivity problem we handle each gas with its own separate sensor. The GSA (Gas Sensor Array) will be attached to a computationally intelligent model learned to tackle the cross-sensitivity problem. This GSA will make the process to detection and calculating the gasses presence much easier and simpler.

3. METHODOLOGY

In the following section, we discuss in detail about background study conducted, data collection methods and mechanisms used to attain the data set used to train the prediction model and also discuss about classification methods and their definitions. The background study helps us to understand the need for the proposed system and will help us understand the different side effects that different gases have on our bodies when we are exposed to them. The data collection methods help us understand the various sources that were referred to get valid data which helps with building the data set.

3.1 Data Collection and pre-processing

Before we dive into the data collection methods and pre-processing methods, we have to grasp the basic design and purpose of our work. The main purpose of our project and work is to develop a portable device that can be used by the workers to ensure their safety before entering the pipelines. The system consist of an input unit i.e. the sensors, predicting model and output unit i.e. alarm light.

The data collection for the training data was done by the following steps. First, side effects and danger limits of gases of sewer pipeline gas mixture were collected and analysed. Secondly, three different levels

- (1) The high concentration limit level
- (2) At above safe concentration level and
- (3) Below safety limit for each sewer pipeline gas were recognized separately

Thirdly samples from sewer pipeline gases are collected and analysed to get their ppm levels and these values are also recorded. Our collected dataset contains seven input features which will provide all data in CSV (comma separated value) and an output class which will be in either a flag format or binary format. Each sample are labelled as safe or unsafe according to the limits which will help with training the prediction model more efficiently. Table 1 (Dataset) consist of dataset and table 2 (Description of data) consist of description of data.

3.2 Classifiers

We used four different kinds of classifiers from which one is selected for the prediction model that has to be deployed. We make the selection based on the performance of the classifier. The performance is measured in terms of their accuracy and efficiency which are the Output of the prediction models.

3.2.1 Naive Bayes. Naive Bayes is approach which is a gathering of basic and simple probabilistic classifiers which applies Bayes hypothesis with solid (Naive) freedom suppositions between highlights. They are very versatile, requiring various parameters straight in the quantity of factors highlights/indicators in a learning issue. Most extreme probability preparing should be possible by assessing a shut shape articulation, which takes direct time, instead of

by costly iterative estimate as utilized for some different kinds of classifiers.

3.2.2 Support Vector Machines. Support vector machines (SVMs, likewise support vector networks) are regulated learning models with related learning calculations that break down information utilized for arrangement and relapse examination. Given an arrangement of preparing cases, each set apart as having a place with either of two classes, a SVM preparing calculation fabricates a model that allocates new cases to one class or the other, making it a non-probabilistic double straight classifier (despite the fact that strategies, for example, Platt scaling exist to utilize SVM in a probabilistic order setting). A SVM show is a portrayal of the cases as focuses in space, mapped with the goal that the cases of the different classes are partitioned by a reasonable hole that is as wide as could be expected under the circumstances. New illustrations are then mapped into that same space and anticipated to have a place with a class in light of which side of the hole they fall.

Table 1.

Sno	Temp	Humidity	NO	CO	H2S	NH3	CH	Class
1	20	65	0	25	10	25	1000	0
2	20	65	0	25	10	25	2000	0
3	20	65	0	25	10	25	5000	1
4	20	65	0	25	10	25	10000	1
5	20	65	0	25	10	35	1000	0
6	20	65	0	25	10	35	2000	0
.								
24069	50	75	10	25	10	35	1000	1
24070	50	75	10	25	10	35	2000	1
24071	50	75	10	25	10	35	5000	1
24072	50	75	10	25	10	35	10000	1
24073	50	75	10	25	10	40	1000	1
24074	50	75	10	25	10	40	2000	1
24075	50	75	10	25	10	40	5000	1
.								
46361	20	73	20	110	75	50	600	1
46362	20	73	20	110	75	50	800	1
46363	20	73	20	110	75	50	7500	1
46364	20	73	20	110	75	50	8000	1
46365	20	73	20	110	75	75	600	1
46366	20	73	20	110	75	75	800	1
46367	20	73	20	110	75	75	7500	1
46368	20	73	20	110	75	75	8000	1

Table 2.

Rubric	Description
Temperature	Temperature of the environment in celcius
Humidity	Relative Humidity of environment
NO	Concentration of nitrogen oxides in ppm
CO	Concentration of carbon monoxide in ppm
H2S	Concentration of Hydrogen sulphide in ppm
NH3	Concentration of Ammonia in ppm
CH3	Concentration of methane in ppm

3.2.3 Artificial Neural Network. Artificial neural systems (ANNs) or connectionist systems are processing systems dubiously inspired by the natural biological networks that constitute wild animals brains. Such systems "learn" (i.e. logically enhance execution on tasks by thinking about illustrations and image recognition) without undertaking particular programming. For instance, in image recognition, they may figure out how to distinguish pictures that contain cats by analysing sample pictures that have been manually named as "feline" or "not feline" and utilizing the outcomes to recognize felines in different images. They do this with no a prior knowledge about felines, e.g., that they have hide, tails, whiskers and feline like features. Rather, they develop their own set of relevant characteristics from the learning material that they use.

An ANN depends on an accumulation of associated units or nodes called artificial neurons (a simplified form of organic neurons in a creature brain). Every association (a simplified version of a neurotransmitter) between artificial neurons can transmit a flag or signals starting with one then onto the next. The artificial neuron that gets or receives the flag or signal can process it and afterward signal artificial neurons associated with it. In general, ANN implementations, the signal at a connection between artificial

neurons is a genuine real number, and the yield of each artificial neuron is calculated by a non-linear function of the of its data sources. Artificial neurons and connections typically have a weight that changes as learning continues. The weight increases or reduces the quality of the signals at a connection. Artificial neurons may have a maximum limit (threshold limit) such that only if the cluster of signal crosses that threshold is the output signal sent. Regularly, artificial neurons are composed in multiple layers. Distinctive layers may perform various types of transforming processes on their sources of input. Signals go from the primary input, to the last (yield) layer, conceivably after going through the multiple layers under different circumstances.

3.2.4 *Random Forest.* Random forests or random decision forest is a gathering learning technique for arrangement and classification , regression and diverse such assignments and process, that work by building up an expansive number of decision trees at the time of training (training time) and yielding the class that is the most occurring type of the classes (classification) or regression of the particular decision tree. Random Decision forest is used to modify the decision tree tendency to overfit the training set data.

4. EXPERIMENT AND RESULTS

As mentioned in the paper the main interpretation of our work is to obtain highly accurate result to ensure the safety for the workers in sewer pipeline. The collected dataset represents sewer environment. The results of classifiers are compared to deploy that model to identify whether it's safe or unsafe to enter the pipeline. The parameter setting of each classifiers are given in table 3 (classifiers and Their parameters)where the dataset is divided into two part training set and test set where 85% of dataset is used for training and remaining 15% is used as test data. Three methodologies were used to select a consistent classifier, the accuracy of the classifiers, Cohen's kappa and ROC curve are used to rank the classifiers. The Cohen's kappa measures inter-annotator agreement.

Table 3.

Classifiers	Parameters
Naive bayes	<i>priors=None</i>
ANN	<i>hidden_layer_sizes=(2,2)</i>
Random forest	<i>random_state=2</i>
SVM	<i>kernel='linear'</i>
SVM	<i>kernel='polynomial'</i>

The formula for Cohen's kappa as follows

$$\kappa = (p_0 - p_e) / (1 - p_e)$$

Where,

- (1) p_0 is the probability of agreement on label
- (2) p_e is the expected agreement when assigned randomly
- (3) The κ value ranges from -1 to 1 where 1 means two annotators agree with each other and 0 or less means chances are less.

Lets discuss about the experimental setup used in the paper. The experimental setup consist of a GSA with 4 gas sensors which are specifically sensitive to a particular gas and will not interfere with the working of other sensors within the GSA and this prevents the cross-sensitivity problem which was a big problem. The GSA is connected with the Arduino Mega 2560 board using a breadboard and the power supply is given to breadboard from the Arduino Board. The LED lights are connected to the Breadboard and the power supply is given from the similar power supply of the GSA. The GSA detects the gases and sends the data to microcontroller and the Ethernet or Bluetooth is used to send the data to a server where prediction model is running. The result from the prediction model results are sent back to microcontroller and are displayed through the blinking or glowing of LED light assigned to each type of the output. Lets get back to what method we used to select the prediction model , The Relative operation characteristic curve is the other name of ROC curve because the curve is plotted by comparing two operating characteristics of TPR (true positive rate) and FPT (false positive rate) as they keep changing over time. A classification model is a mapping of results between different classes and groups. The classifier result is usually a real value (nonstop output),

in which case boundary between classes is fixed using a threshold value and this is called classifier boundary. A discrete class label can also be used to indicate each and every class. The ROC Curve is plotted by the true positive rate (TPR) and false positive rate (FPR) at different limit and threshold values on different axis. In machine learning, TPR (true positive rate) is also known as sensitivity, recall or probability of detection. The FPR (false-positive rate) is also called as fall-out or probability of false alarm. The FPR is calculated as (1 - specificity). It is similar to the plot of the Power as a function of the Type I Error of the decision rule (when the execution is computed from only an example of the populace , it can be considered of as estimators of these quantities).When the ROC curve is along these lines the sensitivity as a function of fall out. As a rule, if detection and false alarm are known, the ROC curve can be produced by plotting the y-axis with cumulative distribution function of detection probability and x-axis with cumulative distribution function of false-alarm probability. AUROC (Area Under Receiver Operating Characteristic Curve) is metric used to determine the best classifier for that problem.

The Table 4 (Accuracy Table) contains algorithm ranked based on accuracy and table 5 (Rate of efficiency in Cohen's Kappa) contains algorithm ranked based on Cohens kappa. In table 6 the confusion matrix is calculated for test data and using which the ROC curve is drawn. The algorithm with more than 90% accuracy is selected, due to the dynamic nature of the pipeline where 100% accuracy is not feasible.Now From Figure 6 we can understand that Naive Bayes has good performance based on the AUROC score.

Table 4.

Algorithm	Train data accuracy	Test data accuracy
Svmpoly	0.9487	0.95241
ANN	0.9487	0.9524
SVMlinear	0.9465	0.9463
Nave Bayes	0.9371	0.9406
random forest	0.9544	0.9243

Table 5.

Algorithm	Train data kappa score	Test data kappa score
Naive Bayes	0.726	0.743
Svmpoly	0.716	0.739
SVMlinear	0.722	0.726
Random forest	0.757	0.6
ANN	0.457	0.52

The Figure 1 shows the workflow of the entire process. After selection model the prototype of device is made using sensors and Arduino board as micro-controller. The Arduino board is used to collect data from sensor and transmit to the local server where the prediction model is used predict whether the condition is safe or unsafe and return a flag, which is used by Arduino to provide alert based on the flag. The following Figures shows the sensors and board used.

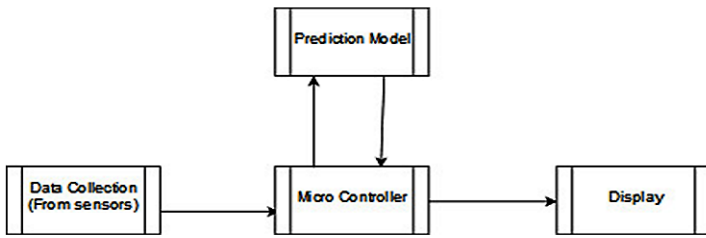


Fig. 1. Workflow Diagram



Fig. 2. Arduino

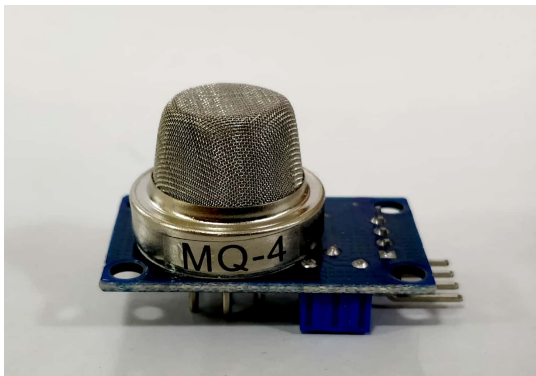


Fig. 3. MQ4

5. CONCLUSION

In this project, we researched a certifiable issue concerning classification, where we simplified the approach by making the choice binary decision named as safe and unsafe condition with respect to the issue. We looked into and explored the issue dealing with presence and identification of these dangerous bio toxic gases inside a sewer pipeline environment. This is greatly critical and imperative

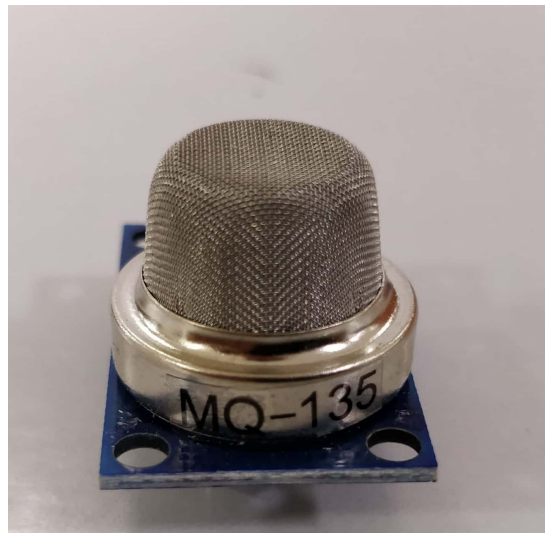


Fig. 4. MQ135

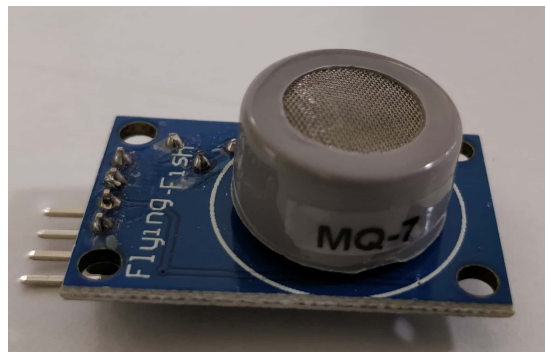


Fig. 5. MQ7

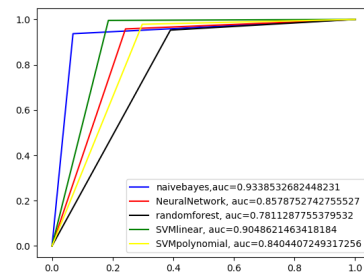


Fig. 6. ROC Curve

issue since recognition of gases is connected with the security and wellbeing of the individual who need to work under the deadly a lethal state of the sewer-pipeline. Generally, a sewer-pipeline condition contains mixture of certain deadly bio-toxic gases. In this way, we accumulated sample informational data set and related information from a few diverse sewer pipelines in various zones. By then we analysed those cases and samples to recognise different and diversified data set for our project. We made an extensive dataset by gathering gas sensor results and outputs from lab tests, literature survey and scaled the accumulated gas sensor results to

Table 6.

n=6956	TN=672	FP=50
NaiveBayes	FN=393	TP=5841
n=6956	TN=547	FP=175
NeuralNetwork	FN=261	TP=5940
n=6956	TN=440	FP=282
Randomforest	FN=294	TP=5940
n=6956	TN=555	FP=167
SVMLinear	FN=198	TP=6036
n=6956	TN=507	FP=215
SVMPolynomial	FN=133	TP=6101

a dataset where safe samples were stamped 0 and unsafe samples were checked 1. Finally, we made use of 4 distinct classifiers over the recognized dataset and their observational and factual performance were compared and analysed. Based on the analysed results we chose the algorithm and deployed it as a local server which collects the information and sends the result, which is displayed in Arduino using led lights. Two led light colour are used one red for unsafe condition and another green for safe condition which are made to blink based on the result of the classifiers.

Finally form our available dataset and experimentation results, we conclude that Naive Bayes algorithm gave the best accuracy in prediction and the most accurate results to say if the condition are safe or unsafe. The Kappa score of Naive Bayes is 0.726 for Training dataset and 0.743 for Testing dataset which is the best and most consistent of all the other algorithms we used in this paper. The higher the kappa score the more reliable the result of the algorithm is and in our case the Naive Bayes gives the best score and which means that the Naive Bayes results are more consistent and reliable then other algorithm results and also the accuracy of the Naive Bayes is 0.9371 which is really very impressive and which proves the consistency of Naive Bayes algorithm. This is the reason we choose the Naive Bayes as the algorithm to build the run the prediction model.

6. FUTURE WORK

The future developments in the project can be multiple GSA circuits placed at regular intervals at certain sewer pipelines to continuously monitor the levels of these bio toxic gases and this network of sensors will be connected to a server where all the data is collected and analysed and this data can be used by the person entering the sewer pipelines to get to know the condition inside without risking his/her life.

Acknowledgment

We students would like to express our deepest gratitude to our guides, Arunnehr Sir and C.Sabarinathan Sir for their valuable guidance, consistent encouragement, personal caring, timely help and providing us with an excellent atmosphere for doing research.

All through the work, in spite of their busy schedule, they have extended cheerful and cordial support to us for completing this research work.

7. REFERENCES

- [1] <http://indianexpress.com/article/delhi/lessons-unlearnt-3-dead-1-critical-while-cleaning-sewer-in-gurugrams-khanda-4868270/>
- [2] <http://safaikarmachariandolan.org/Judgement/SafaiKaramchariAndolanvsUOI2732014.pdf>
- [3] <http://www.thehindu.com/news/national/tamil-nadu/manual-scavenger-dies-due-to-asphyxiation-another-critical/article18231747.ece>
- [4] <https://scroll.in/latest/832368/tamil-nadu-three-manual-scavengers-die-while-cleaning-underground-sewage-tank-in-cuddalore>
- [5] Identifying Hazardousness of Sewer-Pipeline Gas-Mixture using Classification Methods -Varun Kumar Ojha Parmartha Dutta Atal Chaudhuri
- [6] Convergence Analysis of Backpropagation Algorithm for Designing an Intelligent System for Sensing Manhole Gases - Varun Kumar Ojha and Paramartha Dutta and Atal Chaudhuri and Hiranmay Saha
- [7] Performance analysis of neuro swarm optimization algorithm applied on detecting proportion of components in manhole gas mixture - Varun Kumar Ojha, Paramartha Dutta
- [8] Portable Sensor Array System for Intelligent Recognizer of Manhole Gas -Sugato Ghosh / Hiranmay Saha ,Chirasree RoyChaudhuri ,Varun Kr Ojha , Paramartha Dutta
- [9] Sensor Array for Manhole Gas Analysis Sugato Ghosh/Animesh Roy/Sarat Singh/ Hiranmay Saha Varun Kr Ojha/ Paramartha Dutta
- [10] PERFORMANCE ANALYSIS OF NEURO GENETIC ALGORITHM APPLIED ON DETECTING PROPORTION OF COMPONENTS IN MANHOLE GAS MIXTURE Varun Kumar Ojha1, Paramartha Dutta1, and Hiranmay Saha
- [11] Exposure Limits Related to Air Quality and Risk Assessment Kelley J. Donham, Peter S. Thorne, George M. Breuer, Wendy Powers, Shannon Marquez, Stephen J. Reynolds
- [12] On The Design Issue of Intelligent Electronic nic Nose System A. K. Srivastava, S. K. Srivastava, K. K. Shukla
- [13] The Estimation of Hazardous Gas Release Rate Using Optical Sensor And Network, Won So , Jamin Koo , Dongil Shin , En Sup Yoon.
- [14] Carbon Monoxide Poisoning ,Lindell K.Weaver,M.D.