CURRENT TRENDS IN INFORMATION COMMUNICATION TECHNOLOGY RESEARCH

ISSN: 2814-1709

CTICTR 2(1): 1 - 12 (June 2023)
Received:00-04-2023
Accepted:00-06-2023

Voting Ensemble Learning Model (VELM) for Harmful Gas Detection Systems: A Proposed Model

Alfred Udosen Akpan[a]*, Oluwabukola Ajayi Funmilayo[b], Opeyemi Adelowo Joshua[c]

[a]*Babcock University, Ilishan Remo, Ogun State, 121003, Nigeria.*
[b] *Babcock University, Ilishan Remo, Ogun State, 121003, Nigeria.*
[c] *Babcock University, Ilishan Remo, Ogun State, 121003, Nigeria.*

[a]*udosena@babcock.edu.ng*
[b]*ajayio@babcock.edu.ng*
[c]*adelowoo@babcock.edu.ng*

Abstract

An increase in industrialization has caused the release of harmful gases into the atmosphere and this has resulted in environmental pollution that greatly affects the health of both plants and animals alike. Over time, sensor-based technology has been used to develop gas detection systems but they were faced with some major challenges, such as sensor failure and poor performance that have greatly affected their performance. Also, the use of machine learning techniques has been developed for this purpose by leveraging the powers of sensor-based technology. However, they were faced with issues, such as poor feature selection criteria and missing data, which resulted in delays and inaccurate prediction performance of the models. This study, therefore, seeks to survey existing ensemble learning models and to design a Voting Ensemble Learning Model (VELM) for harmful gas detection systems. The methodology adopted for this study is a comparative analysis of the classification models used for the study. The study concludes that there is a need to develop advanced machine learning models to be used for detecting harmful gas and other detection systems such as earthquake, volcano, and landslide detection systems, in real-time that can inform future research. Also, a recommendation was made for the implementation of the developed model based on its low variance and low bias when dealing with widely dispersed data points in a dataset in addition to its ability to deal with local minima and overfitting that affect classic machine learning.

Keywords: Detection System, Ensemble Learning, Machine Learning, Sensor

Alfred Udosen Akpan
----
* Corresponding author.

1. Introduction

The most recent development in industrialization and technological advancements paved the path for cost-effective, faster, and easier-to-implement innovation that has influenced lives in many spheres such as business, education, and manufacturing industry. However, these increasing industrial activities have resulted in cases of gas leaks in both residential and industrial areas that have resulted in the loss of several lives over the years based on exposure to the poisonous gases in the air [1]. Currently, the rising atmospheric pollution level has affected human physical development due to the inhaling of contaminated air and this poisoning happens as a result of the connection between human health and the quality of air owing to the increasing concentration of harmful gases like sulphur dioxide (SO_2), different Nitrogen oxides (NO_x), and carbon monoxide (CO) in the air [2] that have negatively affected humanity.

Furthermore, [3] identified other gases like Methane, Liquid Petroleum Gas (LPG), and other flammable gases, and maintained that if not handled properly could result in accidents or disastrous consequences, such as death. According to [4], burning fossil fuel resulted in about half of the estimated 645,000 premature deaths in the world as a result of air pollution. Since human senses have a poor capacity of detecting these gases and their fumes due to their colourless, tasteless, and odourless nature [3], several studies, leveraging the development of the Internet of Things (IoT) using gas sensors with safety alarm systems for the sole purpose of detecting these harmful gases in real-time have been carried out. These sensor-based systems have been used for analyzing environmental data at intervals to automatically detect these gases and raise a safety alarm. Also, each sensor was manufactured for the detection of just one type of gas with only a few enabled to sense more than one gas and this was not cost-effective as there arose the need to acquire different sensors for different types of gases [5]. However, these techniques did not yield the much expected result as a result of sensor failure, high operating temperature, and poor system calibration among other issues with chemical sensing [1].

Hence, obtaining a sensing system that could accurately predict the existence of toxic gases was challenging, and this is largely due to the occasion of mixed gases or poor sensing systems amongst other issues. With several failed attempts at getting this problem solved, there is still the need to seek ways to effectively detect the presence of these gases. Several attempts have been made to proffer a solution to developing an effective sensing system and they include the use of chemical and thermal sensors and integrated sensing systems that did not achieve much results. However, the advent of big data gave room for further studies in fuzzy logic, classic machine learning, and data mining for recognizing the hidden patterns from previously collected large sensor datasets used in detecting and reporting cases of pollutant gases in real time. The use of conventional machine learning techniques, including the use of homogeneous ensemble techniques like random forest and bagging, was also exploited [6] in the interpretation of the patterns in data. Nevertheless, these methods of gas detection had some limitations and could not adaptively learn with changing environmental conditions resulting in poor performance accuracy [7]. These issues were largely due to the problems of overfitting and local minima which could be solved by ensemble classifier algorithms [8].

In the previous studies, gases were only detected when their concentration levels exceeded the pre-defined thresholds set on the sensors with the use of IoT sensor-based technology [6]. However, this method has not been effective enough because it was faced with several challenges such as the difficulty of detecting the gases in real-time, the issues of selectivity, and performance issues due to the non-uniformity of the invented gas sensors [2]. Also, [1] argued that the other identified issues with the sensor-based gas detection systems included sensor failure and system calibration. These identified issues prevented this sensor-based system from optimally solving the problem of detecting harmful gases as intended.

Subsequently, some other researchers included the application of fuzzy logic, machine learning techniques, and data mining in recognizing the hidden patterns and irregularities in gas sensor data but there was still the need for a more proactive detection system. In order to classify lung illnesses, [9] suggested using an ensemble system that used a semi-supervised learning technique with chest X-rays. Also, [2] applied ensemble learning strategies for predicting the prevalence of harmful gases like SO_2 in the air. As a result of the wide use of ensemble learning in many areas including sensing systems, this research seeks to design the use of ensemble learning algorithms in classifying harmful gases. The aforementioned challenges call for sophisticated machine learning techniques such as a heterogeneous ensemble learning algorithm as a result of the expansion of urban areas which are more vulnerable to this problem.

This necessitated designing a Voting Ensemble Learning Model (VELM) for harmful gas detection systems that would be used for detecting the presence of harmful gas in the air. This design could help in performance improvement in harmful gas detection systems given that sensors are typically selective in part.

2. Literature Review

2.1 Gas Pollution

Although petroleum fuels are a useful source of energy for manufacturing industries, residential and commercial use, the emission of exhaust gases due to incomplete combustion of these fuels, in the absence of oxygen [10], releases trace amounts of nitrogen dioxide (NO2), carbon monoxide (CO), nitric oxide (NO), sulphur (SO_2), carbon dioxide (CO_2), and other chemicals in the atmosphere [11]. Several harmful effects and even deaths have been associated with exposure to harmful gases such as ammonia, carbon monoxide, and toluene even at low concentrations in the air [5].

These gases are harmful to the environment and vehicles alone contribute up to 70% of these air pollutants (mainly CO and SO_2) while the rest is generated from industrial emissions [12]. In a similar study, [13] outlined some of these effects including eye, skin, and respiratory tract irritation amongst others. It was shown in [10] that the addition of CO would likely develop a close bond with haemoglobin within the blood that may reduce the blood's circulation of oxygen to all parts of the body, thus may cause dizziness, nausea, headaches, central nervous system abnormalities, reduced body movements, irregular lung and heart function, unconsciousness, respiratory crisis, and even death at high concentrations [14].

2.2 Machine Learning Algorithms

Machine learning was defined in [15] as the answer to questions by the computer through reliance on pattern recognition and repetition through different types of algorithms. The processing strength of machine learning techniques allows almost any application to be adapted to the basic requirements of any dataset [16]. Machine Learning (ML) is used to forecast values more correctly without explicitly programming them [17]. Many human systems are currently gaining from the power of machine learning algorithms as they can adaptively learn and automatically remove hidden patterns from specific datasets that traditionally had to be identified by domain experts [18]. Used in prediction systems, machine learning is used in identifying patterns, and other uniformities in neighbouring data points that are connected with huge weight edges that are similar [6]. Some examples of machine learning algorithms for classification include:

1. *K-Nearest Neighbours (KNN): This classification technique is based on the similarity of features to determine the worth of any new data. It groups data into clusters by calculating the disparity between each query data and each of the dataset's data points, and then the K picks the values close to the query, choosing the label with the highest frequency (for categorization), or averaging the labels (in regression) [19]. As an adaptable classification technique, KNN is very suitable for multi-modal classes; it is simple to implement and fast in computation on a large dataset. However, classification based on unknown records could be relatively costly as it is required to calculate the distance between k-nearest neighbours. Also, as the size of the training dataset increases, its computational cost is greatly influenced and the presence of noisy or irrelevant features reduces the accuracy of the algorithm [4]. KNN, a non-parametric approach, was used for training because it is a proximity-based classifier technique that classifies the dataset based on the distance measured within those numbers in the dataset [7].*
2. *Logistic regression: As a type of linear classifier, this algorithm is used to predict the probabilistic outcome having only boolean values. The prediction could be done on both numerical and categorical data. The basis of the classification used in logistic regression is the logistic function which is used to create a probability distribution conforming to the weighted feature vectors [18]. It is worth noting that Logistic Regression is a generalized (linear or non-linear) function [20]. The logistic function according to [21] is given as:*

$$h_w(X) = \frac{1}{1 + e^{-v^T X}} \qquad (1)$$

Where x, the independent variable, is the feature vector with each component as a weighted independent characteristic of the object along with a bias. Hence, the total input into the logistic function can be written as:

$$X = V^T X \quad (2)$$

Where V is a vector of associated weights inclusive of the bias term, x is the n-dimensional feature vector, and T is a parameter of the model.

$$V = (vn, vn - 1 \ldots., v2, v1, 1) \quad (3)$$

$$X = (xn, xn - 1 \ldots., x2, x1, 1) \quad (4)$$

3. *Support Vector Machines (SVMs): They are equivalent to traditional perceptron neural networks with several layers. They are centered around a margin that lies on a hyperplane's side dividing two data classes. Exploiting the margin would create a large gap between the dividing hyperplane and the data points beside it on both sides; this would reduce the maximum limit on the anticipated generalization error [21].*
4. *Decision Trees: They classify data points according to their feature values and they are models for making predictions used to map observations about an item to its target value. During classification, the features in a dataset are represented using nodes in a decision tree, and every branch depicts a value that the node may assume. The performance evaluation of decision tree classifiers is based on post-pruning techniques by using a validation set [10]. To avoid overfitting, the trees are usually pruned after being allowed to grow larger than necessary [22].*
5. *Random Forest: Compared to single base statistics and machine learning techniques, random forest performs better as a homogenous ensemble classifier [7]. Random Forests were developed by [23] to further Breiman's bagging idea [24]. They are an ensemble of trees in which each tree is dependent on a number of random factors. They can be used to classify categorical variables or for a continuous response also known as regression. They are also good at handling categorization (with multiple classes) that is quick to learn and predict based on the use of only one or two tuning parameters; good at estimating generalization error and for processing problems with high dimensions, while allowing parallel implementation of tasks [13].*

2.3 Ensemble Learning

Ensemble learning decreases bias and discrepancy between the values in the dataset while also boosting the resilience of the newly developed technique. It is applied to provide an estimate with fewer prediction mistakes that could have affected the performance of the base classifiers when used individually [25].

2.3.1 Types of Ensemble learning

[26] asserts that there are primarily four different types of ensembles and they include:

1. Bayes Optimal Classifier: This classifier is an ensemble of all the hypotheses that are present in the hypothesis space. That is without a doubt the best ensemble strategy. It cannot be used since there are problems with implementation when data points are spaced widely apart. Moreover, the result of the method is a predictive classification rather than a set of probabilistic values.
2. Bootstrap aggregating (bagging): In this method, every model in the ensemble is given the same weight while voting. Bagging also increases the technique's variance

by training each model in the ensemble using a randomly selected portion of the training data. [27] added that there is a replacement in the sample of the dataset that is taken for the training. As an illustration, the random forest technique can be used to combine bagging and random decision trees to dramatically improve classification accuracy. [26] added that bagging was used to develop multiple (same type) models using different subsamples of the training dataset. [27] identified Bagged Decision Trees, Extra Trees, and Random Forest as the types of bagging that exist.

3. Boosting: This iterative ensemble strategy emphasizes the training cases that earlier model's prediction errors used to train every new model. Boosting has occasionally outpaced bagging. Notwithstanding, it has the problem of overfitting the training technique. Adaboost is one example, the other higher-performing methods are Gradient Boosting and XGBoost (Extreme Gradient Boosting). In addition, [86] argued that boosting was used to develop multiple (same type) techniques where each technique learns to repair the forecast errors of a previous model in the categorization of the models.
4. Stacking: Stacking, sometimes referred to as stacked generalization, is the process of developing a learning strategy that incorporates the predictions of various different base classifiers. The process begins with the use of the available dataset to train the basic classifiers, after which an integrated classifier is trained to produce a final forecast using the combination of all the forecasts from the other participants as additional contributions. Although stacking might theoretically be one of the ensemble strategies mentioned previously if a random integrator technique is employed, in practice the integrator is frequently a single-layer logistic regression model. In most cases, the outcome of the stacking is superior to that of any single classifier that was utilized. To this degree, stacking has been utilized to enhance the predictions of the supervised learning tasks for both classification and regression.

2.3.2 Application of Ensemble Learning

Many researchers have worked on the use of ensemble learning techniques in professing solutions to diverse problems. To overcome the gap of inconsistent performance accuracy, [30] used it to develop a model for checking air quality with the use of Bagging and Boosting for pollution forecasting. Compared to SVM, the developed model had enhanced performance accuracy. However, the dataset used for training the model was not large enough to ensure high-level performance. In order to detect hypertension and type 2 diabetes early with the use of patient's risk factors information, [31] developed an ensemble model using prehypertension, hypertension, chronic kidney disease, and type 2 diabetes datasets using SMOTE Tomek link. Using, the isolation forest (iForest) algorithm [32] for detecting and removing outliers in the datasets. The results from the study showed a better performance accuracy when compared to the previously used methods with accuracies of 75.8%, 85.8%, 96.7%, and 100% when predicting prehypertension, hypertension, type 2 diabetes, and chronic kidney disease respectively.

[33] developed an ensemble learning model that efficiently predicted the presence of the COVID-19 virus using AdaBoost that performed at an accuracy of 99.3%, and [34] developed an ensemble learning model that used XGBoost algorithm for the early detection of COVID-19. The model performed with an accuracy of 99.9%. Furthermore, [35] developed an ensemble learning by combining a support vector machine, random forest, and k-nearest neighbours. The model was used to detect heart disease through the use of different

combination schemes such as weighted average voting, and average voting, and majority voting. For the feature selection, Boruta feature selection technique was used and the result showed that the weighted average voting method gave the highest accuracy of 100%.

Furthermore, for detecting credit card fraud, [36] proposed an advanced ensemble method that combined bagging and boosting techniques to produce a better result. AdaBoost technique was used for feature engineering, and extra tree classifier and random forest algorithms were used to develop the learning models. Combining AdaBoost with the extra tree classifier gave an accuracy of 99.1% while combining AdaBoost with the random forest classifier gave an accuracy of 99.2% and both outperformed the other conventional models. Also, for sentiment analysis involving the assessment of the degree of humor in text data, [37] used the ensemble method through the use of stacking technique on 6 machine learning models that included deep learning algorithms such as recurrent convolutional neural network (R-CNN), as long short-term 1200 memory (LSTM), BERT, Text-CNN, feed-forward neural network (FFNN), and BiLSTM. SemEval-2020 dataset, used to assess humor in edited news headlines, was used for training the models with the aim of achieving six separate text embeddings by stacking and feeding the representation of the text into a meta-classifier which produces the prediction outcome.

The results from the reviews showed that the use of ensemble learning-based models based on the integration of several models outperformed the use of conventional machine learning models by selecting the most important features of the dataset used. In most of the related studies reviewed, when the results from classic ensemble learning methods like gradient boosting ensemble learning methods (for example AdaBoost and Stochastic Gradient Boosting), Bootstrap Aggregation (Bagging) ensemble methods such as Random Forest (RF), and voting ensemble methods are compared, the prediction result from the voting ensemble methods had a better performance accuracy because it combined the strengths of all the algorithms used, thereby, complementing the weaknesses of each other [35]. The success recorded in the application of ensemble learning techniques in many different fields is an indication that its application in environmental science would yield a good result.

2.4 Voting Ensemble Learning Machine

The operation of the homogeneous ensemble learning algorithms is based on finding the weighted average of a group of decision trees to get a more accurate prediction accuracy in a classification problem. A voting ensemble learning machine is used to improve the learning efficiency of the integrated models. An example of a homogeneous ensemble learning algorithm is Random Forest (an integration of decision tree classifiers) [28]. For example, if f is an ensemble constructed as a collection of so-called base classifiers, h1(x), ... hN(x) that are blended to provide a collective predictor, f(x). Within classification, f(x) would be the class that is most often foretold (voting) and would be illustrated as:

$$\mathrm{f(x)} = \mathrm{argmax}\, \mathrm{y} \in \mathrm{Y} \sum_{n=1}^{N} \mathrm{I}\big(\mathrm{y} = \mathrm{hj(x)}\big) \qquad (5)$$

3. Methodology

This study focuses on the survey of the currently existing machine learning models for gas detection systems, and the development of a Voting Ensemble Learning Model (VELM) to be used in harmful gas detection systems. The flowchart and conceptual model for the VELM for harmful gas detection systems adopts five machine learning classification algorithms, namely K-Nearest Neighbours (KNN), Decision Trees (DT), Logistic Regression (LR), Support Vector Machines (SVM), and Random Forest (RF), a homogeneous ensemble learning algorithm. Amongst others, these five classification algorithms were considered for this study based on their ability to process widely dispersed data points in a dataset, and their accuracy and timeliness in processing large datasets. The developed model includes three stages: the preparation stage, the training stage, and the validation and testing stage. Figure 3.1 shows the logic flow of the model from the input stage through the processing stage to the output stage.

3.1 Design of the Voting Ensemble Learning Model (VELM)

This section discusses the outlined steps on how the proposed model would be designed. This would involve the data preparation stage, the training stage, and the test and validation stage. Figure 3.1 presents the conceptual design of the proposed Voting Ensemble Learning Model (VELM). The system, after it had learned the hidden patterns in the dataset, it would then detect the presence of pollutant gases.

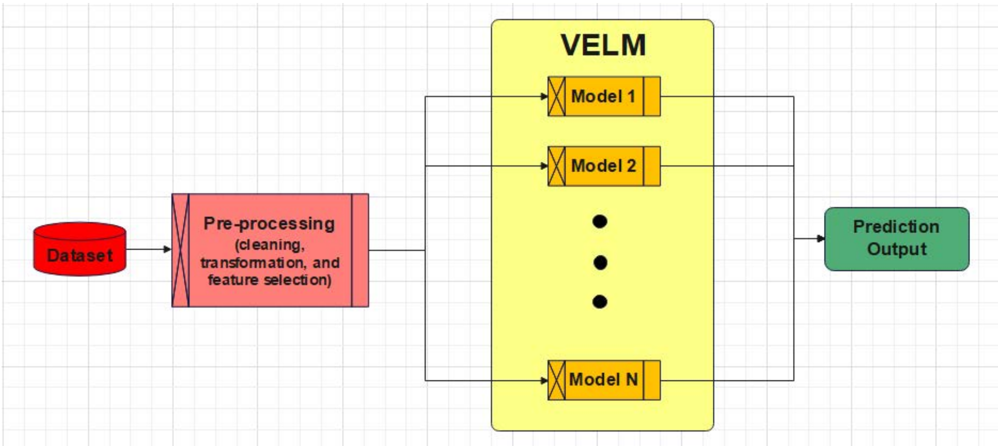

Figure 3.1: Voting Ensemble Learning Model (VELM)

3.1.1 Voting Ensemble Learning Model (VELM) for Harmful Gas Detection Systems

This section of the study shows the design of the ensemble learning model for harmful gas detection systems. The proposed model would learn the patterns from the real-time sensor datasets by passing the dataset through the different classification algorithms used and finding the weighted average of algorithms to ensure better classification accuracy. Thereafter, the prediction value would be used by gas detection systems to detect the presence of harmful gases. After a comprehensive study of the existing models for gas detection with the consideration of their weaknesses in the reviewed literature, such as sensor failure, high operating temperature, poor system calibration, non-uniformity of the sensors, local minima, and data overfitting, the ensemble learning model in this study was then designed to improve upon the existing systems with better performance accuracy. Figure 3.2 that shows the logic flow through the different stages of the proposed model.

Accepted:00-06-2023

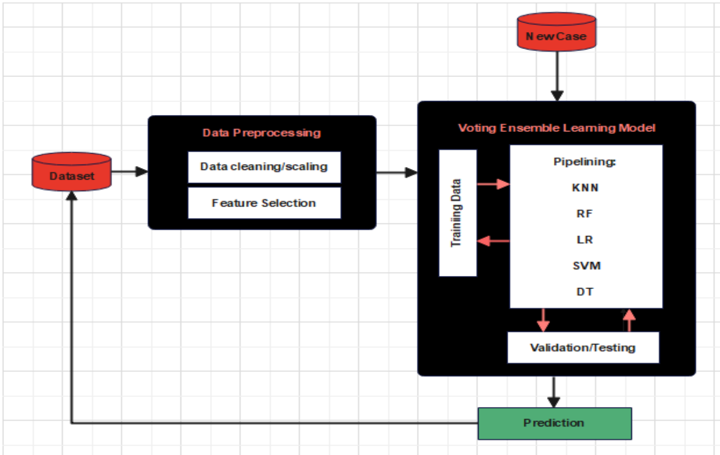

Figure 3.2: Voting Ensemble Learning Model for harmful gas detection systems

Figure 3.2 shows the proposed model and how the dataset would flow from the preparation stage to the output. Before the dataset would be passed into the model, it would be cleaned ad transformed for uniformity after which the important features of the dataset that would influence the performance of the model would be selected for training and testing the model. After the pre-processing is done, the dataset would be made to pass through all of the five models pipelined to create the heterogenous ensemble learning model and the weighted average of the different techniques would be used to detect the presence of harmful gases in the air in real-time. Figure 3.3 shows the flowchart of the model's operations. Here, we have an indefinite loop that exists as long as the model needs to work. If the aggregated classification result is 0 then no harmful gas is detected and class is set to 0. However, if the result is greater than or equal to 1 then a harmful gas is detected and class is set to 1 after which the statement "harmful gas is detected" is printed on the display screen.

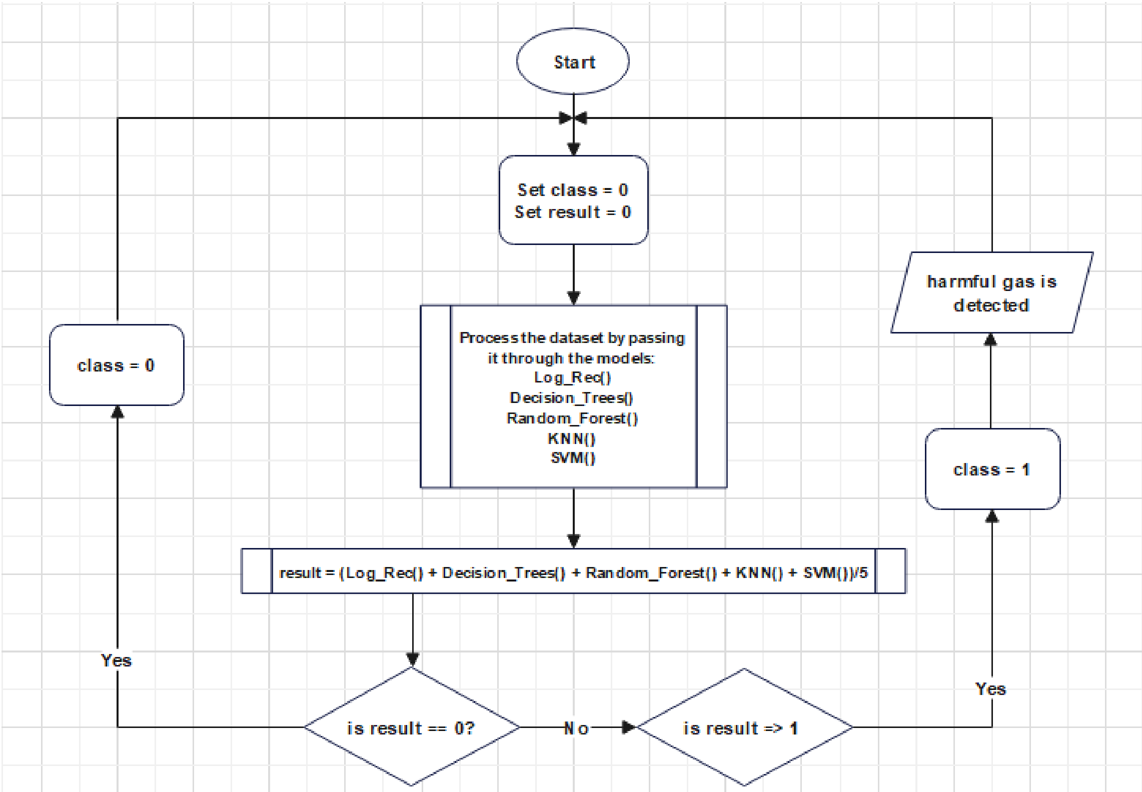

Figure 3.3: Flowchart of the Proposed Technique

When implemented with the aim of detecting the presence of harmful gases in the air per time, the loop runs indefinitely for as long as the system remains turned on.

4. Conclusion

The model was designed by leveraging the combined powers of the integrated classifiers, namely Logistic Regression, Support Vector Machine, K-nearest Neighbours, Decision Trees, and Random Forest. The model was designed to overcome the issues of data overfitting and local minimal which are associated with classic machine learning algorithms. Also, the successes recorded with the use of the voting ensemble learning technique in solving classification problems informed why it was used in designing this model to be used in harmful gas detection systems. By complementing the detection capacity of gas sensors, this designed model, when used in gas detection systems, would help to accurately detect the presence of harmful gases in the air in real-time. After implementation, model would enable gas detection systems to detect the presence of harmful gases in the air in real time. Therefore, it is recommended that this model be adopted and implemented in gas and other detection systems as well in order to ensure environmental safety.

References

[1] J. Fonollosa, I. Rodríguez-Luján, R. Huerta, "Chemical gas sensor array dataset," Data Article, 2015, pp. 85-89.

[2] A. Masih, "Application of ensemble learning techniques to model the atmospheric concentration of SO_2," Global Journal of Environmental Science and Management, vol. 5, issue. 3, 2019, pp. 309-318.

[3] J. Oh, H. Hwang, Y. Nam, M. I. Lee, M. J. Lee, W. Ku, et al., "Machine Learning-Assisted Gas Specific Fingerprint Detection/Classification Strategy Based on Mutually Interactive Features of Semiconductor Gas Sensor Arrays," Electronics, 2022, pp. 3884. Available https://doi.org/10.3390/electronics11233884

[4] V. Shetha, U. Tripathi, A. Sharma, 2 A Comparative Analysis of Machine Learning Algorithms for Classification Purpose," 4th International Conference on Innovative Data Communication Technology and Application, 2022, pp. 422–431.

[5] Z. Wu, X. Chen, S. Zhu, Z. Zhou, Y. Yao, W. Quan, B. Liu, "Sensors and Actuators," B Chem, 2013, pp. 178, 485.

[6] J. R. Kumar, R. K. Pandey, B. K. Sarkar, "Pollutant Gases Detection using the Machine learning on Benchmark Research Datasets," International Conference on Pervasive Computing Advances and Applications, Procedia Computer Science, issue. 152, pp. 360–366, 2019.

[7] H. Yuan, C. Xiao, Y. Wang, X. Peng, Y. Wen, Q. Li, "Maritime Vessel Emission Monitoring by an UAV Gas Sensor System," Ocean Eng. pp. 218, 2020.

[8] U. Brunelli, V. Piazza, L. Pignato, F. Sorbello, S. Vitabile, "Two days ahead prediction of daily maximum concentrations of SO2, O3, PM10, NO2, CO in the urban area of Palermo, Italy," Atmos. Environ. Vol 41, pp. 2967-2995.

[9] P. Narkhede, R. Walambe, S. Mandaokar, P. Chandel, K. Kotecha, G. Ghinea, "Gas Detection and Identification Using Multimodal Artificial Intelligence Based Sensor Fusion," Applied System Innovation, 2021, vol. 4, issue. 3. Available: https://doi.org/10.3390/asi4010003

[10] S. Marathe, "Leveraging Drone Based Imaging Technology for Pipeline and RoU Monitoring Survey," Society of Petroleum Engineers, 2019.

[11] A. Jadon, M. Omama, A. Varshney, M. S. Ansari, R. Sharma, "FireNet: A Specialized Lightweight Fire & Smoke," 2019.

[12] Rajith, S. Soki, and M. Hiroshi, “Real-time Optimized HVAC Control System on top of an IoT Framework,” Third International conference on Fog and Mobile Edge Computing, 2018, pp. 181-186.

[13] S. Huang, A. Croy, L. A. Panes-Ruiz, V. Khavrus, V. Bezugly, B. Ibarlucea, G. Cuniberti, “Machine Learning-Enabled Smart Gas Sensing Platform for Identification of Industrial Gases,” Advanced Intelligent Systems, vol. 4, 2022, pp. 1-11.

[14] S. M. Irid, M. Hadjila, and H. E. Adardour, “Design and Implementation of an IoT Prototype for the Detection of Carbon,” 6th International Conference on Image and Signal Processing and their Applications (ISPA), Mostaganem, Algeria, 2019, pp. 1-5, doi:10.1109/ISPA48434.2019.8966882.

[15] D. Stenvatten, “A Comparative Study for Classification Algorithms on Imbalanced Datasets. An Investigation into the Performance Of RF, GBDT And MLP,” Examensarbete inom huvudområdet, 2020.

[16] L. Huang, G. Liu, Y. Wang, H. Yuan, and T. Chen, “Fire Detection in Video Surveillances using Convolutional Neural Networks and Wavelet Transform,” Engineering Applications of Artificial Intelligence, pp. 110, 2022.

[17] Z, Mohammed, “Artificial Intelligence Definition, Ethics and Standards,” Electronics and Communications: Law, Standards and Practice, British University in Egypt

[18] H. C. Lin, S. Y. Chiang, K. Lee, Y. C. Kan, Sensors, vol. 15, 2015.

[19] A. Tank, “This is the year of the machine learning revolution,” Entrepreneur, 2017, Available: https://www.entrepreneur.com/article/287324.

[20] M. Setia, “Binary Cross Entropy aka Log Loss-The cost function used in Logistic Regression,” Analytics Vidhya, 2022. Available:https://www.analyticsvidhya.com/blog/2020/11/binary-cross-entropy-aka-log-loss-the-cost-function-used-in-logistic-regression/

[21] V. Sharma, “Survey of Classification Algorithms and Various Model Selection Methods,” Journal of Machine Learning Research, vol. 1, 2000, pp. 1-48

[22] L. Breiman, J. Friedman, R. Olshen, C. Stone, “Classification and Regression Trees,” Wadsworth, New York, 1984, vol. 5.

[23] L. Breiman, “Random Forests. Machine Learning,” vol. 45, issue. 1, 2001, pp. 5–32.

[24] L. Breiman, “Bagging Predictors. Machine Learning,” vol. 24, issue. 2, 2001, pp. 123–140.

[25] O. Steinki, Z. Mohammad, “Introduction to Ensemble Learning,” Educational Series, vol. 1, series. 1, 2015.

[26] X. Ying, “Ensemble Learning,” Research Gate, 2014. Available: https://www.researchgate.net/publication/262369664_Ensemble_Learning

[27] J. Brownlee, “Machine Learning Mastery with Python: Understand Your Data, Create Accurate Models and Work Projects End-To-End” Melbourne, Australia, 2016.

[28] A. Cutler, D. R. Cutler, J. R. Stevens, “Random Forests,” DBLP, 2014.

[29] W. Elmenreich, “A review on system architectures for sensor fusion applications. In IFIP International Workshop on Software Technolgies for Embedded and Ubiquitous Systems,” Springer: Berlin/Heidelberg, Germany, 2007; pp. 547–559.

[30] A. Biswal, J. Subhashini, and A. K. Pasayat, “Air Quality Monitoring System for Indoor Environments using IoT,” AIP Conference Proceedings, 2019.

[31] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, ‘‘Development of disease prediction model based on ensemble learning approach for diabetes and hypertension,’’ IEEE Access, vol. 7, pp. 144777–144789, 2019, doi: 10.1109/ACCESS.2019.2945129.

[32] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation forest,” in Proc. 8th IEEE Int. Conf. Data Mining, Dec. 2008, pp. 413–422, doi:10.1109/ICDM.2008.17.

[33] O. O. Abayomi-Alli, R. Damaševičius, R. Maskeliunas, and S. Misra, “An ensemble learning model for COVID-19 detection from blood test samples,” Sensors, vol. 22, no. 6, p. 2224, Mar. 2022, doi:10.3390/s22062224.

[34] M. AlJame, I. Ahmad, A. Imtiaz, and A. Mohammed, “Ensemble learning model for diagnosing COVID-19 from routine blood tests,” Informat. Med. Unlocked, vol. 21, 2020, Art. no. 100449, doi:10.1016/j.imu.2020.100449.

[35] D. Velusamy and K. Ramasamy, “Ensemble of heterogeneous classifiers for diagnosis and prediction of coronary artery disease with reduced feature subset,” Comput. Methods Programs Biomed., vol. 198, Jan. 2021, Art. no. 105770, doi:10.1016/j.cmpb.2020.105770.

[36] V. S. S. Karthik, A. Mishra, and U. S. Reddy, “Credit card fraud detection by modelling behaviour pattern using hybrid ensemble model,” Arabian J. Sci. Eng., vol. 47, no. 2, pp. 1987–1997, Feb. 2022, doi:10.1007/s13369-021-06147-9.

[37] J. Chen, Z. Li, and S. Qin, “Ensemble learning for assessing degree of humor,” in Proc. Int. Conf. Big Data, Inf. Comput. Netw. (BDICN), Jan. 2022, pp. 492–498, doi:10.1109/BDICN55575.2022.00096.