



Data Descriptor

MultimodalGasData: Multimodal Dataset for Gas Detection and Classification

Parag Narkhede ^{1,*}, Rahee Walambe ^{1,2,*}, Pulkit Chandel ^{1,†}, Shruti Mandaokar ^{1,†} and Ketan Kotecha ^{1,2}

¹ Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune 412115, India

² Symbiosis Centre of Applied Artificial Intelligence, Symbiosis International (Deemed University), Pune 412115, India

* Correspondence: parag.narkhede@sitpune.edu.in or ns.parag@yahoo.com (P.N.);
rahee.walambe@scaa.siu.edu.in (R.W.)

† These authors contributed equally to this work.

Abstract: The detection of gas leakages is a crucial aspect to be considered in the chemical industries, coal mines, home applications, etc. Early detection and identification of the type of gas is required to avoid damage to human lives and the environment. The MultimodalGasData presented in this paper is a novel collection of simultaneous data samples taken using seven different gas-detecting sensors and a thermal imaging camera. The low-cost sensors are generally less sensitive and less reliable; hence, they are unable to detect the gases from a longer distance. A thermal camera that can sense the temperature changes is also used while collecting the present multimodal dataset to overcome the drawback of using only the sensors for detecting gases. This multimodal dataset has a total of 6400 samples, including 1600 samples per class for smoke, perfume, a mixture of smoke and perfume, and a neutral environment. The dataset is helpful for the researchers and system developers to develop and train the state-of-the-art artificial intelligence models and systems.



Citation: Narkhede, P.; Walambe, R.; Chandel, P.; Mandaokar, S.; Kotecha, K. MultimodalGasData: Multimodal Dataset for Gas Detection and Classification. *Data* **2022**, *7*, 112. <https://doi.org/10.3390/data7080112>

Academic Editor: Joaquín Torres-Sospedra

Received: 17 June 2022

Accepted: 10 August 2022

Published: 12 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Innovations in engineering designs are helping mankind solve industrial and social problems. Technology is solving many problems in chemical industries; however, industrial hazards can cause damage to the nearby environment. Gas leakage is a common problem in the chemical industries. Explosions, leaks, waste emissions, fires, etc. are some of the common reasons behind the industrial disasters. The carelessness in residential waste disposal also leads to fumes and hazardous gas leakages. The burning of woods is also one of the major causes of air pollution. The gases leaked during mining operations are also the reason behind the deaths of the workers. Even though the machines undergo gas leak tests before installments, gas leakage incidences were reported. Gas-detecting sensors are generally employed near the leakage-prone areas. However, the sensors are sensitive to particular gas and unable to identify leakage when the mixture of gases is present. The manual identification of gases using the chemical apparatus is always not the feasible solution due to the hazardous nature of the chemicals; for example, the smoke leakage blocks the visibility. Hence, the automatic identification of the gas is of extreme importance in the gas-leaking situations. It helps to save human lives as well as machines.

In the literature, many techniques have been employed to identify gas leakage detection. In recent years, the Internet of Things (IoT)-enabled systems that can detect gas leakage using various low-cost sensors have been developed [1–4]. However, these systems

have limited capabilities due to the use of low-sensitive sensors. The detection of gas in a mixed environment is a challenging task and requires the use of technological advancements. Chemical methodologies that include Colorimetric Tape and Gas Chromatography are popular methodologies for detecting particular gas concentrations in a mixed environment [5–7]. The least square-based methodology is proposed by Khalaf for gas classification and estimating the gas concentration [8]. machine learning techniques are employed for gas detection purposes in [9]. Deep Neural Network-based methodologies are also applied for accurate gas detection purposes in [10–13]. However, the methodologies proposed in all these articles are based on the input data from multiple gas-detecting sensors. The array of sensors sensitive to different gases is formed and considered in experimentation.

However, there are a few concerns with utilizing just a gas sensors-based recognition and detection methodology. Generally, the concentration of gas in air is comparatively low; in some cases, the gases are not detectable with a standard set of gas sensors. This leads to ambiguous detection and further affects the detection accuracy of the framework. In addition, cheaper sensors are normally less accurate and may not give exact estimations. When any leakage happens, the surrounding temperature increases; this temperature change can be identified using the thermal camera [14] as well. Using a thermal camera provides the advantage of determining the gas leakage from a longer distance.

As a result of the increase in the usability of artificial intelligence and data analytic methods, the need for the availability of an accurate training dataset is increasing. The availability of the dataset will not only help to train the system but also will provide the platform and a base for the new dataset generation. The existing datasets for the detection of gases are majorly collected using the sensors array only. The various datasets available for gas detection are discussed in the next section. However, all these datasets lack multimodal information.

This paper discusses a novel multimodal dataset collected using the gas sensors along with a thermal camera. The key aspects of the present dataset are mentioned below:

- The dataset is logged using two modalities; images from the thermal camera and numerical values from gas sensors.
- The dataset is collected using two gas sources (smoke and perfume), which are used to generate data for four classes: smoke, perfume, mixture of smoke and perfume, and neutral environment.
- As per the authors' knowledge, this is the only open-source multimodal dataset for gas detection purposes.
- Low-cost sensors are generally less sensitive and may not detect the gas emissions from longer distances. Hence, the use of multimodal data (thermal images along with sensors' measurements) helps to detect the presence of gas from a long distance and even with less concentration.
- The dataset is of interest to the researchers and professionals working in the domain of gas detection and electronic nose. It is also useful for the system designers developing an e-nose for robotic and autonomous systems.
- The dataset can be used to train machine learning and deep learning models and then deploy the algorithms in real-time systems.
- The current version of the dataset also provides the basis for the further extension of the dataset where more gases and their mixtures can be taken into consideration.

Furthermore, this paper is organized as follows: Section 2 provides the information and links for the existing sensors-based gas datasets. A detailed description of the proposed MultimodalGasData and its collection procedure is presented in Section 3. The downloadable link for the dataset is also mentioned in Section 3. Section 4 concludes the paper.

2. Gas Detection Datasets

Various datasets are available for the researchers to take the advancements further. This section provides the information and links for the available datasets.

2.1. Gas Sensor Array Drift Dataset Dataset

During the years 2007 to 2011, Alexander Vergara collected the dataset from gas sensors. It consists of 13910 measurement samples. A total of 16 metal oxide gas sensors were used in the process, and data were collected for six gases at different concentration levels. This open-source dataset is available at <https://archive.ics.uci.edu/ml/datasets/gas+sensor+array+drift+dataset> (accessed on 5 June 2022) for research purposes. This dataset is primarily collected for the development of gas sensor drift compensation purposes and is discussed in [15]. The extension of this dataset with concentration information is provided at <http://archive.ics.uci.edu/ml/datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations> (accessed on 5 June 2022).

2.2. Gas Sensor Array under Dynamic Gas Mixtures

Jordi Fonollosa collected the dataset using gas sensors for the mixed environment. Various gases considered in this experimentation are Ethanol, Ethylene, Ammonia, Acetaldehyde, Acetone, and Toluene. An array of 16 gas-detecting sensors was formed, and the dataset was collected for about 12 h. Two different mixtures are considered in this experiment; one of them is Ethylene and Methane in air, and another one is Ethylene and CO in air. This dataset is open source and available at <https://www.kaggle.com/uciml/gas-sensor-array-under-dynamic-gas-mixtures/version/2> (accessed on 5 June 2022). The experimentation using the dataset is presented in [16].

2.3. Gas Sensors for Home Activity Monitoring Data Set

This dataset consists of data collected using eight different metal oxide gas sensors for home activity monitoring. It consists of 100 recordings taken from the sensor array. The complete dataset can be accessed using <http://archive.ics.uci.edu/ml/datasets/gas+sensors+for+home+activity+monitoring> (accessed on 5 June 2022). The experimentation using this dataset is presented in [17].

2.4. Gas Sensor Array Exposed to Turbulent Gas Mixtures Dataset

This dataset is collected using eight different chemo-resistive gas sensors. It provides the measurements when a mixture of Ethylene and Methane is present at different levels of concentrations. It provides the measurements taken at 180 instances (<http://archive.ics.uci.edu/ml/datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures> (accessed on 5 June 2022)). The experimental validation of the collected dataset is presented in [18].

All these datasets discussed above are collected using the gas sensors only. However, to the authors' knowledge, there does not exist any dataset simultaneously collected using gas sensors and thermal cameras. Using a thermal camera along with gas sensors increases the performance accuracy even in low gas concentration areas. Hence, this paper introduces the multimodal dataset collected for the accurate identification and classification of gases based on various gas sensors and thermal camera.

3. Multimodal Gas Detection Dataset

The current work focuses on the dataset collected using multiple gas detecting sensors and a thermal camera.

3.1. Setup for Dataset Collection

The system used while collecting the dataset consists of seven metal-oxide gas detecting sensors along with a thermal camera. Figure 1 shows the framework used for data collection. Various sensors used are MQ2, MQ3, MQ5, MQ6, MQ7, MQ8 and MQ135, along with a thermal camera: Seek Compact Thermal Imaging Camera (UW-AAA).

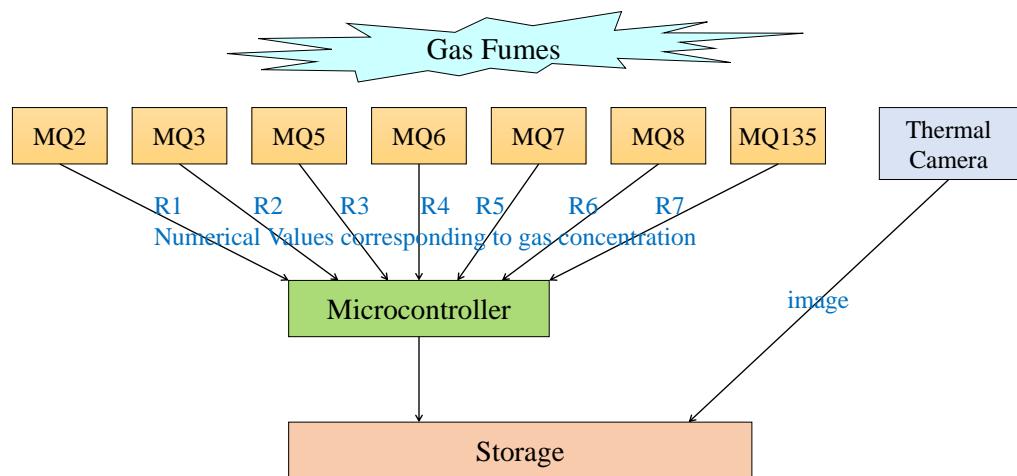


Figure 1. Block level connections for dataset collection setup (figure reproduced from [19]).

The experimentation setup completed while collecting the present dataset is shown in Figure 2. Using the two gas sources, four classes are generated.

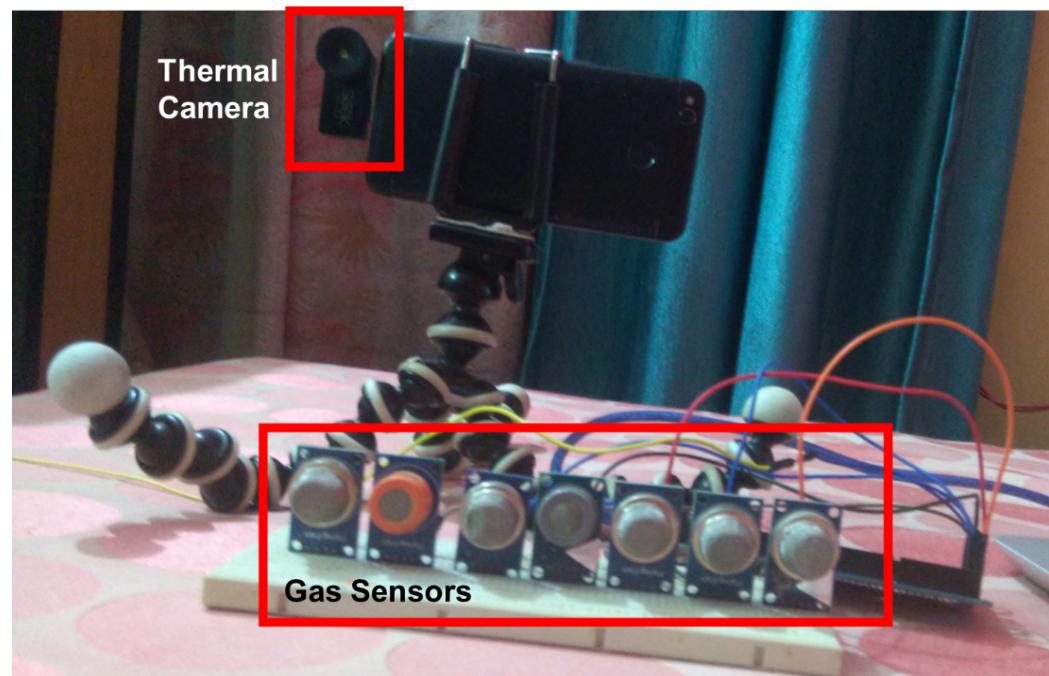


Figure 2. Experimental setup for data collection.

The considered seven gas-detecting sensors are reactive to different and multiple gases.

3.2. Gas Sensors

Gas sensors distinguish the existence of a particular gas to generate electrical signals. Gas sensors based on Metal Oxide Semiconductor (MQ) technology are the suitable choice, as these sensors are compact in size, have a quick response time, and also have a longer life [20,21]. Every sensor comprises a warming component that delivers the simple result voltage corresponding to the gas concentration. Table 1 provides the information about the different sensors and their respective sensitive gases.

Table 1. Gas sensors and sensitive gases.

Sensor	Sensitive Gas
MQ2	LPG, Butane, Methane, Smoke
MQ3	Smoke, Ethanol, Alcohol
MQ5	LPG, Natural Gas
MQ6	LPG, Butane
MQ7	Carbon Monoxide
MQ8	Hydrogen
MQ135	Air Quality (Smoke, Benzene)

3.3. Thermal Camera

A thermal imaging camera is a gadget that recognizes variations in temperature based on infrared light. Each pixel of a camera is basically an infrared temperature sensor; all such pixel sensors measure the temperature of all points simultaneously. The RGB images are then produced by temperature configuration. Dissimilar to typical imaging cameras, the performance of a thermal camera is not affected by dim environmental factors and can work with any climate [22]. The considered Seek Compact Thermal Imaging Camera is compact in size. It has a resolution of 206×156 . It can measure temperature in the range of -40 to 330 $^{\circ}\text{C}$, and it has a frame rate of <9 Hz frequency.

In this work, thermal cameras and gas sensors are used simultaneously to collect the dataset. For a better understanding of the actual sensors used in this work, the pictures of the thermal camera and sensors are shown in Figure 3.

**Figure 3.** Actual pictures of sensors and camera.

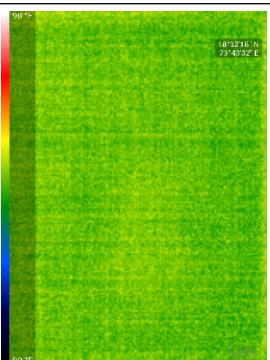
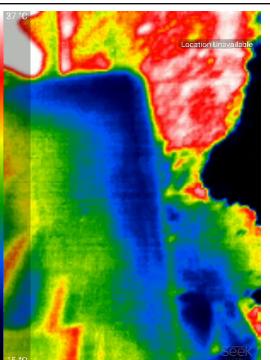
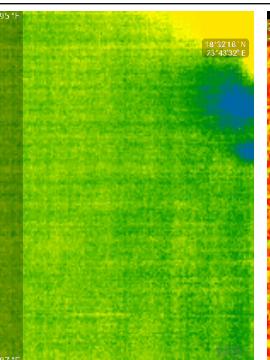
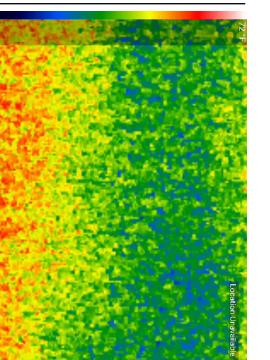
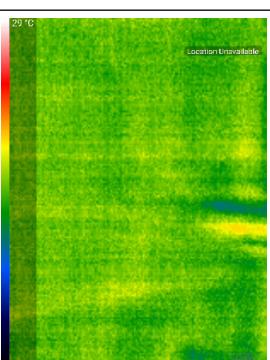
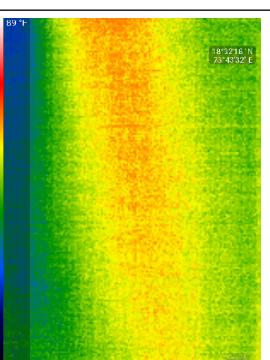
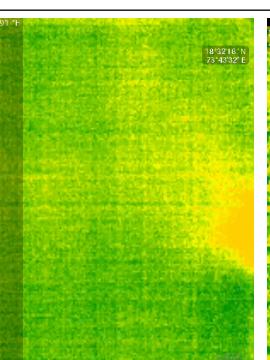
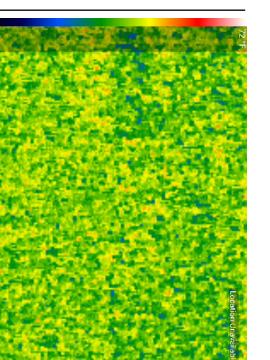
3.4. Process of Dataset Collection

In this work, the dataset is generated using two gas sources: namely, perfume and smoke. Park Avenue Deodorants contains 95% alcohol and is used as a source of perfume. Incense sticks are used to generate smoke. Smoke is majorly a combination of Carbon Monoxide, Carbon Dioxide, Nitrogen Dioxide and Sulfur Dioxide as well as other gases in small quantity [23].

The MultimodalGasData dataset presented in this work is gathered through a variety of seven gas sensors along with utilizing the thermal camera. During experimentation, the gas sensors were kept at a distance of 1 mm from each other. In the dataset collection process, two explicit gas sources were identified and considered. One of the sources was the gases generated after spraying a perfume spray, and another one was the gases generated after lighting the incense sticks.

Thermal images and readings of the gas sensors were logged for each of the gas sources as an individual class. Another class considered was the mixture of gases generated after generating gases from both sources together. The data are logged at a frequency of 2 s continuously for the duration of 90 min. During the process, the gas was spread in the environment at an interval of 15 s for the first 30 min; then, a 30 s interval is followed for the next 30 min, and a 45 s interval is considered for the last 30 min. For better understanding and representation purposes, a few data samples are shown in Table 2. The table provides two samples of thermal images and the corresponding gas sensors' measurements for each considered class. The images shown are the data samples collected using a thermal camera, whereas the numerical values represent the readings collected using considered gas sensors. The numerical values are the digital equivalents of the analog signals from gas sensors obtained after using the 10-bit analog to digital converter.

Table 2. Data samples for thermal images and their corresponding gas sensors measurements. Gas sensor measurements are in the sequence of MQ2, MQ3, MQ5, MQ6, MQ7, MQ8, and MQ135.

	No Gas	Perfume	Smoke	Mixture
Thermal Image				
Gas Sensor Measurements	725, 756, 529, 420, 428, 585, 641	739, 526, 488, 459, 662, 716, 502	523, 345, 375, 396, 574, 563, 292	505, 383, 356, 326, 396, 223, 336
Thermal Image				
Gas Sensor Measurements	771, 533, 423, 419, 550, 624, 487	809, 527, 533, 510, 696, 785, 583	539, 351, 331, 370, 572, 548, 278	545, 426, 382, 359, 487, 348, 418

The way the gases are discharged while collecting the dataset could affect their thermal signatures. Hence, to avoid conflicts and maintain uniformity, the gases were spread all the time similarly. In addition, appropriate care has been taken to spread a constant amount of gas every time.

3.5. Dataset Description

A MultimodalGasData dataset has 6400 samples in total. These 6400 samples are divided equally in four classes. The dataset contains 1600 samples of perfume class, 1600 samples belonging to the smoke class, and 1600 samples for a mixture of perfume and smoke class. The remaining 1600 samples were collected for the neutral environment (No gas) class. To showcase the variation in the developed dataset, the statistical analysis is also carried out, and the obtained statistical properties for the collected data from gas sensors are presented in Table 3.

Table 3. Statistical Properties of the Measurements from Gas Sensors.

Property	MQ2	MQ3	MQ5	MQ6	MQ7	MQ8	MQ135
Count	6400	6400	6400	6400	6400	6400	6400
Mean	677.59	462.02	404.58	399.76	565.95	542.47	416.73
Standard Deviation	92.91	70.28	55.67	45.09	83.13	151.02	76.68
Minimum	502	337	291	311	361	220	275
25th Percentile	591	405	366	366	524	447	354
50th Percentile	701	486	400	393	576	576	437
75th Percentile	756	529	443	426	629	642	473
Maximum	824	543	596	524	796	794	589

The dataset is published online at the Mendeley Data repository. The direct link for the dataset is <https://data.mendeley.com/datasets/zkwgkjkjn9/2> (accessed on 10 June 2022). This repository contains two folders namely

1. Gas Sensors Measurements
2. Thermal Camera Images

The *Gas Sensors Measurements* folder contains a CSV file having measurements taken using seven gas sensors. In the file, the first column represents the serial number; the following seven columns consist of measurements from gas sensors in the sequence mentioned in the tables. The ninth column denotes the class to which the measurements belong (no gas, perfume, smoke, mixture). The tenth column represents the name of the corresponding thermal image. The *Thermal Camera Images* folder consists of four different zip files belonging to four classes (no gas, perfume, smoke, mixture). Each zip file consists of images belonging to a particular class. Thermal images are named following the nomenclature ‘serialNumber_Class’. The same value of the serial number from the CSV file and thermal images folders signifies that the measurements are taken simultaneously.

4. Conclusions

The presented MultimodalGasData is the dataset collected for the researchers and system designers to work on the gas detection and classification application. It is a novel multimodal dataset consisting of data samples collected using multiple gas detecting sensors and a thermal imaging camera. It provides the advantage of detecting the presence of certain gas using two modalities: images and numerical values. The dataset can train machine learning and deep learning models and deploy the algorithms in real-time systems. The current version of the dataset also provides a basis for the researchers to enhance and increase the sample size by considering more gases and their combinations.

Author Contributions: Conceptualization, P.N. and R.W.; methodology, P.N.; software, P.C. and S.M.; validation, P.N., P.C., S.M., R.W. and K.K.; formal analysis, P.N.; resources, P.N. and R.W.; data curation, P.N., P.C. and S.M.; writing—original draft preparation, P.N. and R.W.; writing—review and editing, P.N. and R.W.; supervision, R.W. and K.K.; funding acquisition, P.N. and R.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Symbiosis International (Deemed University) under the grant Minor Research Project Grant 2017-18 void letter number SIU/SCRI/MRPAPPROVAL//2018/1769.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset described in this paper is available at <https://data.mendeley.com/datasets/zkwgkjkjn9/2> (accessed on 10 June 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Adekitan, A.I.; Matthews, V.O.; Olasunkanmi, O. A microcontroller based gas leakage detection and evacuation system. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *413*, 012008. [[CrossRef](#)]
2. Kodali, R.K.; Greeshma, R.; Nimmanapalli, K.P.; Borra, Y.K.Y. IOT based industrial plant safety gas leakage detection system. In Proceedings of the 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 14–15 December 2018; pp. 1–5.
3. Suma, V.; Shekar, R.R.; Akshay, K.A. Gas leakage detection based on IOT. In Proceedings of the 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 12–14 June 2019; pp. 1312–1315.
4. Evalina, N.; Azis, H. Implementation and design gas leakage detection system using ATMega8 microcontroller. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *821*, 012049. [[CrossRef](#)]
5. Fox, A.; Kozar, M.; Steinberg, P. CARBOHYDRATES | Gas Chromatography and Gas Chromatography—Mass Spectrometry. In *Encyclopedia of Separation Science*; Wilson, I.D., Ed.; Academic Press: Oxford, UK, 2000; pp. 2211–2223. [[CrossRef](#)]
6. Stauffer, E.; Dolan, J.A.; Newman, R. Gas chromatography and gas chromatography—Mass spectrometry. In *Fire Debris Analysis*; Academic Press: Oxford, UK, 2007; pp. 235–293.
7. Wang, T.; Wang, X.; Hong, M. Gas leak location detection based on data fusion with time difference of arrival and energy decay using an ultrasonic sensor array. *Sensors* **2018**, *18*, 2985. [[CrossRef](#)] [[PubMed](#)]
8. Khalaf, W.M.H. Electronic Nose System for Safety Monitoring at Refineries. *J. Eng. Sustain. Dev.* **2012**, *16*, 220–228.
9. Yin, X.; Zhang, L.; Tian, F.; Zhang, D. Temperature modulated gas sensing E-nose system for low-cost and fast detection. *IEEE Sensors J.* **2015**, *16*, 464–474. [[CrossRef](#)]
10. Liu, Q.; Hu, X.; Ye, M.; Cheng, X.; Li, F. Gas recognition under sensor drift by using deep learning. *Int. J. Intell. Syst.* **2015**, *30*, 907–922. [[CrossRef](#)]
11. Peng, P.; Zhao, X.; Pan, X.; Ye, W. Gas classification using deep convolutional neural networks. *Sensors* **2018**, *18*, 157. [[CrossRef](#)] [[PubMed](#)]
12. Pan, X.; Zhang, H.; Ye, W.; Bermak, A.; Zhao, X. A fast and robust gas recognition algorithm based on hybrid convolutional and recurrent neural network. *IEEE Access* **2019**, *7*, 100954–100963. [[CrossRef](#)]
13. Bilgera, C.; Yamamoto, A.; Sawano, M.; Matsukura, H.; Ishida, H. Application of convolutional long short-term memory neural networks to signals collected from a sensor network for autonomous gas source localization in outdoor environments. *Sensors* **2018**, *18*, 4484. [[CrossRef](#)] [[PubMed](#)]
14. Hamilton, S.; Charalambous, B. *Leak Detection: Technology and Implementation*; IWA Publishing: London, UK, 2013.
15. Vergara, A.; Vembu, S.; Ayhan, T.; Ryan, M.A.; Homer, M.L.; Huerta, R. Chemical gas sensor drift compensation using classifier ensembles. *Sens. Actuators Chem.* **2012**, *166*, 320–329. [[CrossRef](#)]
16. Fonollosa, J.; Sheik, S.; Huerta, R.; Marco, S. Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring. *Sens. Actuators Chem.* **2015**, *215*, 618–629. [[CrossRef](#)]
17. Huerta, R.; Mosqueiro, T.; Fonollosa, J.; Rulkov, N.F.; Rodriguez-Lujan, I. Online decorrelation of humidity and temperature in chemical sensors for continuous monitoring. *Chemom. Intell. Lab. Syst.* **2016**, *157*, 169–176. [[CrossRef](#)]
18. Fonollosa, J.; Rodríguez-Luján, I.; Trincavelli, M.; Vergara, A.; Huerta, R. Chemical Discrimination in Turbulent Gas Mixtures with MOX Sensors Validated by Gas Chromatography-Mass Spectrometry. *Sensors* **2014**, *14*, 19336–19353. [[CrossRef](#)] [[PubMed](#)]
19. Narkhede, P.; Walambe, R.; Mandaokar, S.; Chandel, P.; Kotecha, K.; Ghinea, G. Gas detection and identification using multimodal artificial intelligence based sensor fusion. *Appl. Syst. Innov.* **2021**, *4*, 3. [[CrossRef](#)]
20. Han, L.; Yu, C.; Xiao, K.; Zhao, X. A new method of mixed gas identification based on a convolutional neural network for time series classification. *Sensors* **2019**, *19*, 1960. [[CrossRef](#)] [[PubMed](#)]
21. Pashami, S.; Lilenthal, A.J.; Trincavelli, M. Detecting changes of a distant gas source with an array of MOX gas sensors. *Sensors* **2012**, *12*, 16404–16419. [[CrossRef](#)] [[PubMed](#)]
22. Havens, K.J.; Sharp, E.J. *Thermal Imaging Techniques to Survey and Monitor Animals in The Wild: A Methodology*; Academic Press: Cambridge, MA, USA, 2015.
23. Lin, T.C.; Krishnaswamy, G.; Chi, D.S. Incense smoke: Clinical, structural and molecular effects on airway disease. *Clin. Mol. Allergy* **2008**, *6*, 3. [[CrossRef](#)]