Multimodal Fusion Techniques for Gas Detection Systems

Ansah Siddiqui, Seba Mokdad, and Aleena Lifiya

***Abstract*—Multimodal gas detection has gained increasing attention as the limitations of single-sensor systems become more apparent in real-world applications. By integrating heterogeneous data sources—such as gas sensor readings, thermal imagery, and environmental parameters, deep learning-based fusion methods have demonstrated enhanced accuracy and robustness. This survey reviews recent research on multimodal fusion strategies for gas detection, with a focus on early, intermediate, late, and hybrid fusion approaches. We analyze over 30 studies, categorizing them by their fusion architecture, input modalities, and deployment contexts across domains like industrial monitoring, fire hazard detection, and smart environments. Intermediate fusion is the most widely adopted strategy, leveraging modality-specific encoders and attention mechanisms to extract and integrate complementary features. However, it often introduces high computational costs and requires full input availability. In contrast, early fusion offers simplicity but struggles with incomplete data, while late and hybrid fusion approaches—though less explored—show promise in scenarios involving uncertain or asynchronous data streams. This survey highlights each strategy's strengths and trade-offs and proposes future research directions within the context of multimodal gas detection.**

***Index Terms*—Multimodal fusion, gas detection, deep learning, sensor fusion, environmental safety.**

I. Introduction

Gas leaks remain a persistent and dangerous challenge in industrial, domestic, and mining environments, often leading to explosions, fires, and toxic emissions. Despite precautionary measures, incidents continue to occur due to equipment failure, improper waste handling, or the unpredictable nature of chemical processes. To address these risks, there is a growing need for reliable gas detection systems that can automatically identify hazardous leaks before they escalate [1]. Artificial intelligence is transforming gas sensing by enabling sensors to accurately detect and distinguish between multiple gases, even those with similar chemical signatures. By integrating AI with gas sensors, these systems can now adapt to changing environments and provide faster, more reliable gas detection for applications like environmental monitoring and industrial safety [2].

However, unimodal gas detection systems that rely only on gas sensors or thermal imaging often face issues like low sensitivity, false positives or negatives, and difficulty distinguishing between different gases, especially in mixed environments. These limitations come from environmental interference, weak signal strength, or the lack of context that a single sensor can provide. Multimodal fusion helps overcome these problems by combining data from different sources, like thermal images and multiple gas sensors. This allows AI models to learn more meaningful features, leading to better accuracy, stronger performance, and more reliable gas detection, even in real-time and safety-critical situations [3]. Furthermore, multimodal data fusion enhances gas detection accuracy by combining complementary information from thermal imaging and gas sensors, allowing one modality to compensate when another is limited. This integrated approach reduces the chances of false positives and negatives, offering a more robust and reliable detection system for real-world industrial environments [4].

Multimodal fusion in gas detection, while powerful, faces real-world challenges such as inconsistent sensor formats (e.g., scalar gas readings vs. 2D thermal images), sensor-specific limitations across environments, and external factors like pipeline bending or corrosion that affect detection accuracy [5]. Ensuring reliable performance requires handling noisy data, sensor drift, and integrating robust modeling and denoising techniques to maintain system integrity in dynamic conditions [3].

This survey aims to categorize and analyze early, intermediate, and late fusion strategies for gas detection using modalities such as gas sensors, thermal imaging, and environmental parameters. It focuses on fusion techniques, architectural innovations, and deployment challenges across applications including smart homes, industrial monitoring, and fire hazard detection.

Through this work, we aim to answer the following research questions:

- **RQ1:** Which multimodal fusion strategies are most effective in enhancing gas detection accuracy across different application domains?
- **RQ2:** What types of neural network architectures and fusion mechanisms are commonly employed for integrating heterogeneous data such as gas sensor values and thermal imagery?
- **RQ3:** What are the key limitations and deployment challenges such as sensor noise, data imbalance, or real-time constraints—associated with each fusion strategy?

Addressing these questions will help identify promising directions for future research and inform the design of robust, efficient gas detection systems across real-world environments. Consequently, this survey will provide an overview of methodologies employed in detecting gases by the use of varying modalities like sensor data and thermal images. Each section will explore the specific deep learning architectures and fusion strategies used, elucidating their strengths and challenges. Additionally, the survey will explore the impact of each fusion approach, identify avenues for future research, and highlight existing gaps in the current literature.

II. BACKGROUND

A. Sensor Modalities for Gas Detection

Gas detection goes all the way back to 1815 when the Davy lamp was invented by Sir Humphry Davy. It was used to detect methane in coal mines. Few decades later, in 1926-1927, Dr. Oliver Johnson made the first practical 'electric vapor indicator' meter [6]. Since then, gas detection has been evolving and now protect workers in industries from petrochemicals to mining by detecting flammable, poisonous, and oxygen-deficient atmospheres before they reach hazardous levels.

Gas sensors are classed as catalytic bead, metal oxide semiconductor (MOS), electrochemical, photoionization detector (PID), non-dispersive infrared (NDIR), photoacoustic spectroscopy (PAS), thermal conductivity, surface acoustic wave (SAW), and optical fiber. Although MOS and catalytic beads sensors are essential components of flammable gas detection, they are vulnerable to environmental drift and poisoning. At the expense of limited lifespans or light maintenance, electrochemical and PID sensors, respectively, reach ppb-level sensitivity for hazardous and volatile organic chemicals. Although PAS provides high sensitivity through light-induced pressure waves, NDIR sensors provide exceptional selectivity for gases such as CO_2. Simple, arrayable, or remote measurements are made possible by thermal conductivity, SAW, and optical fiber sensors, although they frequently need intricate signal processing [7].

TABLE I
COMPARISON OF GAS SENSOR TYPES

Sensor	Principle	Gases	Advanatages	Disadvantages
Catalytic Bead	Oxidation on catalyst bead	CH_4, H_2	Robust	Poisoning
MOS	Conductivity change	CO, VOCs	Low cost	Drift
Electrochemical	Redox current	NO_2, SO_2	ppb-level	Lifetime
PID	UV-ionization	VOCs	Fast	Lamp upkeep
NDIR	IR absorption	CO_2	Selective	Bulky
PAS	Photoacoustic	VOCs	Sensitive	Complex
Thermal	Heat dissipation	H_2	Simple	Low sensitivity
SAW	Acoustic-wave change	VOCs	Sensitive	DSP need
Optical-Fiber	Evanescent wave	Various	Remote	Complex

B. Gas Concentration Sensors

Gas-concentration sensors encompass the electrically-transduced classes discussed in Table 1—MOS, catalytic bead, electrochemical, PID, NDIR, PAS, thermal-conductivity, and SAW—since all directly measure a target gas. These directly transduce gas presence into an electrical signal.

1) Metal-Oxide Semiconductor (MOS) Sensors (MQ Series): Metal-oxide semiconductor (MOS) sensors, including the MQ-2 and MQ-5 series, use a heated tin-oxide (SnO_2) or similar sensing film that increases conductivity when combustible gases adsorb and react on its surface. The MQ series modules are commonly used in household and industrial leak alarms to detect CH_4, LPG, CO, H_2, and VOCs since they are inexpensive and simple to interface. Their main advantages are their affordability and wide gas coverage, but they have drawbacks such as baseline drift, cross-sensitivity (temperature/humidity), and the necessity for frequent calibration [8].

2) Electrochemical Sensors (CO,CO_2): Electrochemical sensors were utilized to determine the concentration of various analyte chemicals in experimental samples. Electrochemical sensors employ a three-electrode cell where target gases undergo redox reactions, producing a current proportional to concentration. They achieve ppb-level limits of detection and high specificity for toxic gases such as CO, NO_2, SO_2, and CO_2 [8].

C. Sensor-Based Systems Limitations

We group the limitations of sensor-based gas detection systems into three key areas: semiconductor limits, nanostructure challenges, and material stability.

1) Metal-Oxide Semiconductor (MOX) Constraints:

- **Cross-sensitivity & Nonlinearity:** MOX sensors often respond to multiple gases and ambient humidity/temperature, resulting in poor selectivity and nonlinear calibration curves [9].
- **Operating Temperature Dependence:** sensing mechanism in MOX sensor is highly affected by high tempratures [10].
- **Humidity Interference:** MOX sensors suffer from low selectivity in harsh and humid environments. use [11].

2) 2. Nanostructure Challenges:

- **Humidity sensitivity:** humidity severely affects performance, increasing response time and decreasing amplitude. [12].
- **Poisoning:** – Presence of CO, NOx, and sulfuric compounds can poison Pd-based sensors[12].

3) Material Stability Issues:

- **Grain-Boundary Effects:** Changes in grain-boundary conductivity affect sensitivity and long-term stability [10].
- **Environmental Degradation:** Exposure to particulates, corrosion, or UV light can damage sensing materials [13].

D. Multimodal Fusion

Multimodal fusion refers to the process of integrating information from multiple sensing modalities, such as gas sensors and visual images, to improve the reliability, precision and robustness of decision-making in complex environments. Each modality captures different aspects of the physical world, and their fusion leverages complementary strengths to mitigate individual limitations. Traditional gas detection methods based solely on single-sensor modalities—such as metal oxide semiconductors or electrochemical sensors—face well-documented limitations, including cross-sensitivity, environmental interference, and limited specificity. These difficulties increase in dynamic and high-risk contexts, such as chemical plants, where prompt and precise detection of gas leaks is significant.

In the case of gas detection, this involves integrating low-level signals from conventional gas concentration sensors with visual information like thermal imaging, which can record heat signatures linked to gas leaks or fires, and environmental characteristics like temperature and humidity. Fusion techniques can be broadly categorized into three levels. Figure 1 describes the three different fusion strategies visually [14].

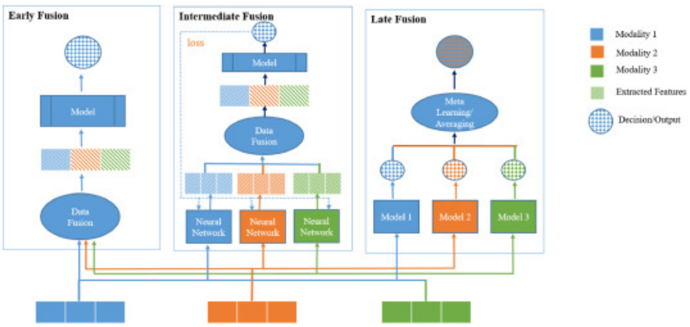

Fig. 1. An illustration of different fusion techniques. Adapted from [14]

1) Early Fusion: Also known as data-level fusion, this technique combines raw data or low-level features from multiple modalities before any learning or decision process. For instance, normalized thermal image pixel values can be concatenated with time-series sensor readings to form a joint feature vector. While data-level fusion retains the most information, it may suffer from modality imbalance and noise sensitivity.

Example: Concatenating thermal image features with real-time CO_2 concentration values as input to a deep neural network shown in Fig.2.

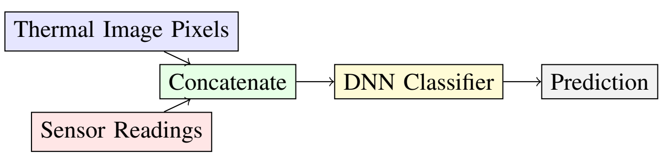

Fig. 2. Data-level fusion: raw modalities concatenated before modeling.

2) Intermediate Fusion: Also called feature-level fusion, this method extracts modality-specific features first, such as deep image embeddings from a CNN or statistical features from gas sensors, and then combines them through concatenation, attention, or dimensionality reduction methods (e.g., PCA, autoencoders).

Example: A CNN extracts thermal patterns, while an LSTM encodes temporal sensor data; the two feature sets are then fused and fed into a classifier shown in Fig.3.

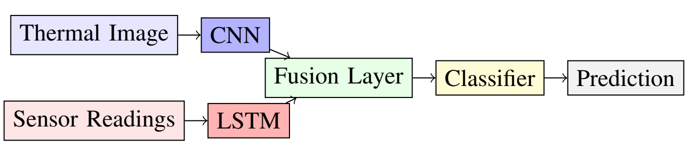

Fig. 3. Feature-level fusion: modality-specific features combined before classification.

Feature-level fusion balances the richness of representation with manageable dimensionality and is popular in deep learning-based fusion frameworks.

3) Late Fusion: Known as decision-level fusion, this approach fuses the outputs of individual classifiers or models—each trained on a separate modality—using voting, weighted averaging, or meta-learners. It is suitable when modalities are processed independently or when data synchronization is difficult.

Example: A decision tree classifies sensor readings, while a CNN processes thermal images; their outputs are fused via majority voting to trigger an alert shown in Fig.4.

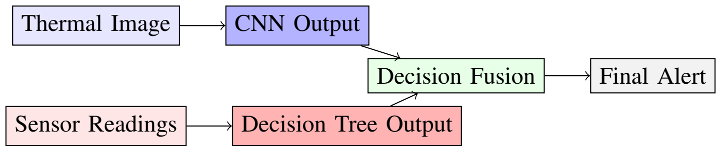

Fig. 4. Decision-level fusion: predictions combined after individual modality processing.

Each fusion level offers trade-offs between computational complexity, flexibility, and information richness. Recent advances in deep learning have enabled more sophisticated architectures, such as attention-based fusion, transformer-based multimodal encoders, and graph fusion networks, which can dynamically weigh modality importance depending on context. This and all other fusion types is summarized in Table II.

TABLE II
COMPARISON OF MULTIMODAL FUSION TECHNIQUES

Fusion Level	Integration Point	Advantages	Challenges	Typical Use Cases
Early	Raw or low-level features	Preserves full information	Sensitive to noise and modality mismatch	Sensor + raw thermal pixel vectors
Intermediate	Modality-specific feature embeddings	Balanced, flexible	Requires feature engineering or compatible embeddings	Deep learning with CNN + LSTM or attention
Late	Output probabilities or decisions	Modular and interpretable	Ignores interaction between modalities	Voting between sensor model and thermal model

III. METHODOLOGY

We used a structured process that included literature search, screening, inclusion criteria, and analysis to guarantee a thorough and scientific evaluation. The goal of this procedure was to document the most recent developments in multimodal fusion methods for gas detection and fire prediction that are based on sensors and images.

A. Paper Collection

We searched peer-reviewed articles and conference proceedings using databases such as IEEE Xplore, Elsevier ScienceDirect, and Google Scholar. The following keywords and their combinations were used:

- `("gas detection" OR "gas sensing")` AND `("multimodal fusion" OR "data fusion" OR "sensor fusion")`
- `("thermal imaging" OR "infrared image")` AND `("fire detection" OR "hazardous gas" OR "early fire detection")`

- `("deep learning" OR "machine learning")` AND `("gas sensor" OR "thermal camera")`

These keywords were used to ensure a comprehensive retrieval of studies that integrate both sensing technologies and thermal imagery. The distribution of publications over years is shown in Fig. 5

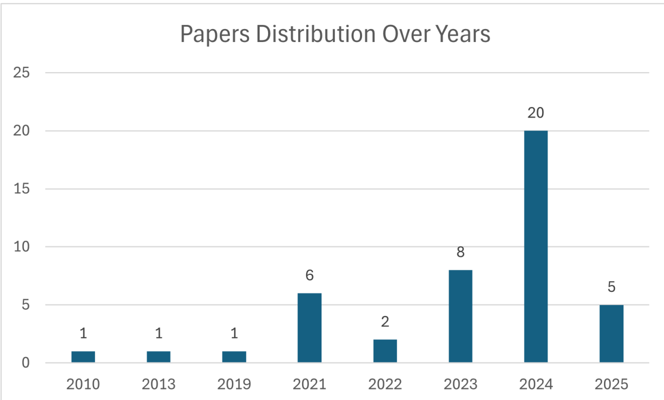

Fig. 5. Distribution of reviewed papers by year (Total N=44).

As the bar chart illustrates, there has been a significant increase in the number of reviewed papers in recent years, particularly in 2024. This trend may indicate a growing interest in the research area or an increase in the submission rate. Further analysis could explore the reasons behind this increase.

B. Inclusion/Exclusion Criteria

Studies were selected based on the following inclusion criteria:

- Focus on gas detection, hazardous gas leak monitoring, or fire prediction.
- Utilization of at least two sensing modalities, including thermal imagery.
- Employ multimodal fusion at the data, feature, or decision level.
- Published between 2019 and 2025.
- Written in English and accessible in full text.

Papers were excluded if they:

- Focused solely on unimodal sensing (e.g., only gas sensors).
- Did not apply or analyze fusion strategies.
- Were surveys or tutorials without experimental results.
- Duplicated papers.

The initial search returned 44 papers. After removing duplicates and screening titles and abstracts, 39 papers were selected for full-text review. Of these, 32 studies met all inclusion criteria and were included in this review.

IV. Fusion Approaches for Gas Detection

A. Early Fusion Approaches

In early fusion, also known as data or feature-level fusion, multiple input modalities are combined before training a single model, with raw or preprocessed features integrated at the input stage. This approach requires modality-specific feature encoding to ensure that the semantic meaning aligns across different data types for effective fusion [14].

A number of studies have explored the use of thermal imaging alongside gas sensor data for improved detection accuracy. The authors in [3] designed a system combining thermal images from a Seek Thermal Camera with data from seven MOS gas sensors (MQ2, MQ3, MQ5, MQ6, MQ7, MQ8, MQ135). LSTM was used for temporal feature extraction from gas sensors, and CNN for spatial feature extraction from thermal images. The fusion occurs at the early stage—immediately after initial feature extraction. The features were concatenated and passed through a dense layer for classification. The early fusion approach achieved a testing accuracy of 96%, outperforming individual modalities (gas sensors at 82%, thermal images at 93%). However, this system required a large training set to prevent overfitting and was limited by its sensitivity to sensor noise. The paper [15] used a similar setup but replaced the CNN-LSTM framework with VGG16-based feature extraction and employed classical classifiers like Random Forest, SVM, and KNN. Their goal was to classify four classes—Smoke, Perfume, No Gas, and Mixture—using the Multimodal Gas Data dataset, which included 6400 samples. Early fusion was performed by concatenating the extracted image features with the processed sensor data to form a unified feature vector. This fused representation was then passed into classical machine learning models. Among these, RFC achieved the highest validation accuracy of 96.41%, followed by KNN and SVM. Despite the promising results, the study acknowledged limitations such as the restricted variety of gas types in the dataset and lack of testing under different environmental conditions, which could affect generalizability and robustness. Furthermore, authors in [16] combined ConvNet features from thermal images and BiLSTM outputs from simulated gas profiles. Early fusion occurred through concatenation of these spatial and temporal features. The model achieved 100% validation accuracy under controlled conditions but performed poorly on truncated inputs and lacked generalization to real-world environments.

Some studies applied early fusion within deep learning pipelines explicitly focused on time-aware modeling. The authors in [17] developed a fire detection system using eight metal oxide gas sensors and temperature readings, which were preprocessed using moving averages, normalization, and sliding windows before being fed into an LSTM model. The system achieved 99.88% accuracy with only one false negative, using controlled data from PVC combustion at various temperatures. However, the model's performance was highly dependent on data collected in a lab environment and may not generalize well to other fire sources or real-world conditions. It also did not address long-term sensor drift, and its parameters—such as window size and labeling ratio—were manually tuned and may not be optimal across settings. Building on this, [18] proposed a BiLSTM-based system that incorporated trend features using Kendall's tau correlation in addition to raw sensor values. This approach improved robustness and accuracy, achieving 99.93% accuracy with a moving average filter and fire threshold. While the inclusion of trend data enhanced prediction, the system was evaluated over a short observation period and lacked validation on long-term datasets, leaving questions about its durability and resistance to sensor degradation.

Other studies implemented early fusion using traditional machine learning algorithms. The authors in [19] focused on ethanol, CO, and hydrogen detection, combining gas readings with environmental data like temperature and humidity. These were fused at the feature level and input into a Random Forest classifier, which achieved 100% accuracy. Despite strong performance, the model was evaluated only under clean, controlled conditions. Authors in [20] used normalized outputs from MQ-2, MQ-7, and DHT-22 sensors, fused and input into an SVM classifier. The approach achieved over 85% accuracy and outperformed K-means and Random Forest in some settings, though performance declined under variable room conditions. Lastly, [21] implemented sensor-level fusion of CO, temperature, humidity, and flame detectors in smart home environments. The fused data were used to train classifiers such as CART and ANN, achieving over 95% accuracy. However, testing was limited and the system did not account for sensor drift or missing data. A separate line of work has used early fusion within shallow neural networks or simulation-based environments. The paper [22] proposed a GA-optimized backpropagation neural network using early-fused temperature, smoke, and CO data. Fusion occurred at the input layer of a three-layer neural network trained on simulation data generated via FDS. While the model achieved perfect classification accuracy in the simulated environment, it was not tested in real-world deployments, limiting its practical relevance.

In real-time embedded and IoT-based systems, early fusion has been implemented through direct sensor integration at the hardware level. Authors in [23] integrated MiCS-5524 and MQ-series sensors for detecting CO, CH_4, SO_2, and CO_2, along with Winsen MG812 for temperature and humidity. Fusion occurred on the Arduino Uno microcontroller, and sensor values were transmitted via LoRa for remote alerts. While simple and scalable, the system lacked adaptive intelligence and relied heavily on network quality. Authors in [24] developed an IoT-based freon detection system for refrigerators where sensor signals were processed and alerts triggered via buzzer, LCD, and a mobile app. Fusion was performed at the input level, offering a low-cost and user-friendly system but with limited adaptability. The paper [21] also fits within this category, where the fused sensor inputs were used in a smart home system for fire detection using classical models. Finally, [25] proposed an IoT-based cloud monitoring framework for building fires. It integrated MQ-series gas sensors (MQ-2, MQ-7, MQ-135), temperature and humidity sensors (HTS221), and pressure/light sensors (LPS22HBTR, APDS-9960). Sensor-level fusion occurred before input into a shallow ANN classifier. The system achieved over 95% accuracy and performed well for detecting wood and gas fires. However, it was constrained by a small dataset (2,000 samples), dependence on stable connectivity, and limited generalization across diverse fire types.

Consequently, early fusion has proven to be an effective approach for improving gas detection by combining information from multiple sensors at the feature level. Studies such as [3], [15], [16] show that integrating thermal images with gas sensor data leads to higher classification accuracy than using either modality alone. This is particularly useful for detecting complex gas environments or mixtures. Other works like [17], [18] demonstrate that early fusion with deep learning models like LSTM and BiLSTM can capture temporal patterns in sensor data, resulting in better detection capabilities. Traditional machine learning approaches also benefit from early fusion, as seen in [20], [19], [21], where combining gas and environmental data improves performance with models like SVM, Random Forest, and CART. Even shallow neural networks trained on simulated data, such as in [22], show promise when fed fused sensor inputs. In embedded and IoT-based systems, early fusion at the hardware or microcontroller level enables real-time, low-cost monitoring solutions, as demonstrated by [23], [24], [25]. Overall, these studies highlight that early fusion makes the input data more informative, leads to better classification results, and works well with both advanced AI models and simpler real-time systems—all while helping to keep computational costs low.

Despite its effectiveness, frameworks utilizing early fusion presents certain limitations across different implementations. Many studies, including [3], [15], [16], [17], [18], rely on controlled datasets collected in lab environments, which may not generalize well to real-world scenarios involving diverse fire sources, environmental conditions, or sensor drift. Small dataset sizes, as noted in [25], increase the risk of overfitting and limit robustness. Moreover, early fusion can struggle with missing or incomplete data, since most architectures depend on all modalities being present at the point of fusion. This is particularly problematic in IoT-based systems [23], [24], where data transmission issues or sensor failure are common. Imbalanced modalities, where some sensors provide more informative signals than others, can cause bias or assign weights in an unfair manner during fusion. Several neural network-based systems [16], [22], [17] also face challenges with interpretability, making them less suitable for safety-critical applications. Additionally, many early fusion methods involve manual tuning of parameters like window size or thresholds [17], [18], which may not be optimal across settings. These factors highlight that while early fusion offers strong performance in ideal conditions, its real-world reliability, adaptability to missing data, and interpretability remain ongoing challenges. The next section would explore intermediate fusion and may address some of the challenges associated with early fusion.

B. Intermediate Fusion Approaches

Intermediate fusion, also known as feature-level fusion, involves processing each modality independently to extract meaningful features before combining them for joint learning. This method allows the model to leverage modality-specific encoders tailored to each data type, improving representation learning before integration. It is particularly useful when fusing temporally and spatially diverse modalities such as

TABLE III
SUMMARY OF EARLY FUSION IN MULTIMODAL GAS DETECTION

Study	Input Modalities	Fusion Operation	Model	Task/Output	Key Results & Limitations
Narkhede et al. (2021) [3]	Thermal images, MQ2–MQ135	Feature concatenation (CNN + LSTM outputs)	CNN + LSTM	Gas classification (4 types)	96% accuracy; large data requirement; sensitive to sensor noise
Arya et al. (2024) [15]	Thermal images, MQ sensors	Feature concatenation (CNN + sensor vectors)	VGG16 + RFC/SVM/KNN	Gas classification (Perfume, Smoke, Mixture, No Gas)	96.41% (RFC); limited gas types; not tested in variable environments
Kopbayev et al. (2022) [16]	Simulated thermal images, gas profiles	Feature concatenation (ConvNet + BiLSTM)	ConvNet + BiLSTM	Gas leak detection	100% accuracy on simulation; poor generalization to truncated input
Ro & Lee (2024) [17]	MOS gas sensors, temperature	Raw input fusion at sequence level	LSTM	Early fire detection	99.88% accuracy; tested on PVC gas only; parameter sensitivity
Kim et al. (2024) [18]	Gas sensors + trend features	Raw + statistical fusion at input	BiLSTM	Fire prediction	99.93% accuracy; robust to short-term variation; lacks long-term validation
Tejaswi et al. (2024) [19]	CO, H_2, Ethanol, temp, humidity	Sensor feature concatenation	Random Forest	Hazard classification	100% accuracy; trained in controlled settings only
Chen et al. (2022) [20]	MQ2, MQ7, DHT-22	Sensor-level feature fusion	SVM	Fire/gas hazard classification	85%+ accuracy; performance drops in variable conditions
Salhi et al. (2019) [21]	CO, temperature, humidity, flame	Sensor-level fusion	CART + ANN	Smart home fire detection	95% accuracy; lacks handling of sensor drift
Wen et al. (2021) [22]	Simulated CO, smoke, temperature	Raw input fusion for neural network	GA-optimized NN	Fire detection (simulation)	Perfect simulated accuracy; lacks real-world testing
Sakthimohan et al. (2024) [23]	MQ-series, MiCS-5524, temp/humidity	Input-level hardware fusion (Arduino)	Rule-based (LoRa)	Real-time gas/fire alerts	Scalable, low-cost; lacks adaptive learning or intelligence
Divya et al. (2023) [24]	Freon sensor, microcontroller signals	Sensor fusion with rule-based logic	Arduino threshold-based	Refrigerator leak detection	Cost-effective; lacks contextual learning or automation
Pan et al. (2025) [25]	MQ2, MQ7, MQ135, HTS221, LPS22HBTR, APDS-9960	Input-level sensor fusion into ANN	Shallow ANN	Building fire classification	95% accuracy; only 2000 samples; limited generalization

gas sensor time-series data and thermal images. This section outlines and compares a selection of works that adopt intermediate fusion for gas detection and fire hazard prediction.

Several studies have adopted parallel CNN and RNN feature extraction paths for thermal imagery and gas sensor signals. For instance, the work by [26] employs ShuffleNetV2 for spatial feature extraction from thermal images and a combination of 1D-CNN and BiGRU for capturing temporal dynamics from seven gas sensors. These features are fused using a cross-attention mechanism and passed to a classifier. This intermediate fusion approach (SCGA) achieved 99.22% accuracy, outperforming both single-modality and simple concatenation-based fusion baselines. However, the study did not assess robustness to environmental noise or mixed gas concentrations. Similarly, [27] explored DWT- and DCT-based fusion schemes using CNNs (e.g., ResNet-50, MobileNet) and BiLSTM for independent modality encoding. Their intermediate fusion achieved 98.47% accuracy on a four-class gas dataset, demonstrating strong detection performance, though real-time responsiveness and robustness to dynamic environments were noted as areas for improvement. Another notable work is by [28], who explored multimodal gas detection using thermal imaging and metal-oxide sensor data, processed via deep learning and federated learning frameworks. Thermal images were encoded using CNN variants like VGG16 and DenseNet, while gas sensor values were fed to LSTM/BiLSTM networks. Intermediate fusion was performed by merging features from both modalities before final classification. The study used a novel dataset of 6400 samples across four gas classes and reported up to 99.7% validation accuracy using federated learning, outperforming centralized and unimodal models. While the paper demonstrated strong performance, it did not explicitly evaluate robustness to environmental variability or sensor drift, and focused primarily on accuracy under clean test conditions.

Building on this multimodal direction, authors in [29] developed the SPIRo system, which combines features extracted from thermal images via CNNs (AlexNet, ResNet-50, MobileNet) with normalized gas sensor data using a Deep Forest Classifier. This intermediate fusion model achieved 98.5% accuracy on the MultiModalGasData dataset and 88.1% on SPIRo's custom field-collected dataset, demonstrating robustness to controlled and semi-realistic conditions. However, performance dropped in low-gas-concentration scenarios due to limited signal strength. Another recent model, MFIA-Net by [30], fused visible and IR imaging data for VOC detection using a dual-stream YOLOv5 architecture enhanced by attention modules. Despite achieving strong visual classification performance, the model faced challenges in computational efficiency and deployment on constrained platforms due to its high parameter count and FLOPs.

Other studies focused on enhancing robustness in adverse conditions. The co-learning framework proposed by [31] used CNNs for thermal features and DNNs for gas sensor readings, fusing them at the feature level for classification. This setup was evaluated under missing modality and noise scenarios, showing that intermediate fusion outperformed single-sensor models, although multitask fusion performed even better. The system's resilience to noise and dropout (up to 90% for thermal, 20% for gas) makes it promising, though it assumes the availability of at least partial input from each modality. Likewise, the SPIRo system by [32] fused CNN-extracted features from thermal images with normalized gas sensor data before passing them into a Deep Forest Classifier. Tested on both clean and noisy datasets, the model achieved 98.5% accuracy in controlled settings and 88.1% on more realistic data, though accuracy declined under minimal gas presence and poor sensor calibration. Intermediate fusion has also been used in specialized detection tasks, where combining different types of sensor features helps improve decision-making for more complex scenarios than basic gas classification.The authors in [33] proposed the rTPNN model for fire detection using an intermediate fusion strategy. Time-series data from five different sensors (CO, CO_2, O_2, temperature, and smoke) were first individually processed to extract three features per sensor: trend, level, and raw value. These extracted features were then concatenated to form a fused input vector, which was fed into a fully connected neural network for final classification. This intermediate-level fusion allowed the model to incorporate both dynamic and static characteristics of each sensor stream. The rTPNN achieved 96% accuracy, triggered alarms within 11 seconds, and outperformed several baseline models in terms of F1 score, AUC, and alarm delay. However, the model was evaluated on a single, imbalanced dataset that favored non-fire cases, raising questions about its generalizability across more diverse fire scenarios. Similarly, a co-attention based framework for natural gas micro-leakage detection by [34] fused hyperspectral and thermal features from vegetation, achieving a 94.6% accuracy and 0.93 kappa score. The feature fusion via multi-attentive networks offered good interpretability, though constrained by limited spatial coverage and reliance on manually collected field data. Several systems have applied fusion within industrial safety contexts. A mine hazard monitoring system proposed by [35] fused gas sensor thresholds with visual structural analysis via MobileNet CNNs. The model, evaluated on 1,000 annotated images, achieved 86.81% accuracy and real-time dashboard alerts, although its applicability was limited by hardware constraints (e.g., battery life) and dataset size. Similarly, [36] developed a multi-index fusion method using AHP and fuzzy logic to weight and combine features from CO, SO_2, radon, temperature, and air pressure sensors. This expert-rule-based intermediate fusion generated a fire risk index (IFZD), validated through field deployments in coal mines. While interpretable and effective in expert-assessed scenarios, its reliance on manually assigned weights and lack of dynamic learning presented clear limitations.

Some advanced models have been developed for complex sensing tasks that involve multiple data types. The MAM-Net model proposed by [37] was designed for fault detection using both spatial features from ResNet with CBAM and temporal gas sensor data from a BiLSTM. These features were fused in a multi-task learning setup and the model achieved F1 scores above 99% even on imbalanced datasets, showing the strength of intermediate fusion. Another system,by [38] , used over 50 chemical and environmental sensors to detect narcotics and contaminants. Although the fusion method was not clearly described, the use of many sensor types suggests intermediate-level integration. However, this system is not directly focused on gas or fire detection and faced challenges with data scarcity and high input dimensionality.

Other examples include papers like [39], where the authors proposed the FFBGD framework for LNG detection using LWIR and MWIR optical gas imagery. Foreground and original thermal image features were independently extracted and fused using deformable CNNs and a Foreground Fusion Network. The model demonstrated high AP scores under both scene-dependent and independent settings, although computational load and domain generalization remain areas for improvement.

Overall, intermediate fusion has shown strong potential in gas detection and safety-related tasks by allowing each modality—such as thermal images and gas sensor data—to be processed separately before combining their meaningful features. Across the reviewed studies [27], [28], [39], [29], [30], [26], [37], [35], [33], [36], [34], common techniques include the use of CNNs for spatial feature extraction from thermal or visual data, and RNNs like BiLSTM or GRU for capturing temporal patterns in gas sensor streams. These features are then fused using mechanisms such as concatenation, attention layers, or task-specific networks. This modular structure allows the models to better handle complex scenarios, including noisy environments, missing data, and imbalanced datasets. Attention-based models like SCGA [26] and co-attention frameworks [34] have further enhanced the fusion process by assigning weight to more informative features. Compared to early fusion, which combines raw data directly and struggles with modality imbalance, intermediate fusion provides better interpretability, flexibility, and robustness by preserving modality-specific structure before integration.

Despite its advantages, intermediate fusion also presents several challenges. Many models rely on deep architectures with multiple processing branches (e.g., CNNs for thermal data and RNNs for gas sensors), which significantly increase computational requirements and may limit deployment on resource-constrained platforms [30], [37]. Models that use attention mechanisms or large feature vectors—can suffer from high dimensionality, leading to slower inference and memory overhead [28], [34]. Some studies reported performance drops in real-world conditions, particularly when sensor calibration was suboptimal or gas concentrations were low [32], [29]. Moreover, intermediate fusion frameworks often assume that all modalities are present and of sufficient quality. While a few models address missing data [39], most require complete inputs and may not perform well when sensor streams are partially degraded or lost. Lastly, interpretability can still be limited in highly complex fusion networks, especially those that use deep feature transformations without explicit reasoning modules. Despite these challenges, intermediate fusion still offers an improvement over early and late fusion by providing a more flexible and context-aware approach to multimodal integration. It still demands careful handling of feature interaction and computational resources. Ultimately, the decision to utilize intermediate fusion should consider the specific data types involved, the required depth of interaction, and the interpretability of the resulting model.

C. Late and Hybrid Fusion Approaches

While early and intermediate fusion strategies have seen widespread application in multimodal gas detection, comparatively fewer studies have explored late fusion, where the outputs of separate models are integrated at the decision level. Hybrid fusion approaches—which combine both intermediate (feature-level) and late (decision-level) integration—have also emerged to combine the strength of different fusion strategies. This section discusses both categories together, highlighting their frameworks, results, and associated limitations.

Late fusion has been explicitly implemented in only a few studies. For instance, [40] presented a weighted ensemble model that combined the outputs of a ResNet-50 CNN (processing thermal images) and a k-Nearest Neighbors (kNN) model (processing gas sensor data). The final decision was made via weighted averaging of individual classifier outputs. This late fusion approach achieved an accuracy of 99.32%, significantly outperforming the individual models (CNN: 97.28%, kNN: 97.56%). Despite its effectiveness, the approach relied on offline processing and lacked robustness against environmental fluctuations and blended gas concentrations. Another study by [41], proposed a hybrid ensemble method for fire detection in smart buildings using MQ sensors, temperature, and humidity data. Four classifiers—Logistic Regression, SVM, Decision Tree, and Naive Bayes—were trained independently and their predictions were fused using average voting. This late fusion model achieved a precision of 0.9752 and an AUC of 0.996 on fire datasets from NIST. While accuracy was high, the model exhibited vulnerability to class imbalance and potential overfitting or underfitting depending on the feature representation.

Several studies opted for hybrid approaches, which made use of different fusion mechanisms fusion mechanisms. The authors in [42] introduced a system that used fuzzy logic for fusing electric wire current and temperature features (intermediate fusion) to detect potential electrical fires, followed by late fusion of threshold-based outputs from gas and smoke sensors with the fuzzy model's output. Visual fire detection was handled using a MobileNet CNN trained on real and synthetic fire imagery. This multi-modal pipeline achieved over 98% accuracy, but required further refinement in data diversity and integration of non-visual sources. In another hybrid example, [43] proposed a smart industrial safety system that fused electrochemical, catalytic bead, and NDIR sensor outputs using machine learning techniques. Flame signatures captured via UV, IR, and visible imaging were processed independently using CNNs, with alerts triggered through a final late fusion step. Though the system reached 95% detection accuracy with only 5% false alarms, its performance was affected by sensor degradation and ambient interferences in real-world deployments. Additional hybrid strategies include those by [44], who fused smoke, gas, IR, and acoustic data at both feature and decision levels. Their system outperformed traditional single-sensor approaches and demonstrated robustness in detecting fires during early stages, including smoldering and concealed fires. Another study by [45] proposed a multi-sensor hybrid fusion framework for indoor and outdoor fire detection. The system integrated gas, smoke, flame, temperature, humidity, light, and GPS data with real-time image analysis using a Raspberry Pi camera. It employed a hybrid architecture where anomaly detection was first performed using Z-score analysis on sensor data, followed by early fusion through a Fuzzy Logic system and Random Forest Classifier (RFC). This multi-stage design—combining early decision-level fusion and late-stage visual confirmation—allowed the system to incorporate both environmental sensing and visual verification. The sensor-based models achieved approximately 99% accuracy, while the CNN component reached around 90%. However, the system relied on artificial, simulated sensor data modeled on historical records due to lab constraints, as well as stable internet connectivity and accurate sensor calibration, which may limit its robustness in real-world, diverse fire conditions.

Compared to early and intermediate fusion, relatively fewer studies have explored late fusion in gas and fire detection systems. Late fusion, which combines the outputs of separate models at the decision level, has been shown to improve overall accuracy by using the strengths of individual classifiers [40], [41]. However, most of these systems remain limited to offline analysis and are often sensitive to class imbalance or feature noise. In contrast, hybrid fusion strategies—used more frequently—integrate different fusion techniques to enhance model robustness and adaptability [42], [43], [44], [45]. These multi-stage frameworks allow systems to process sensor data first and use that to inform or trigger secondary decisions such as image-based fire detection. Hybrid models have demonstrated high accuracy and improved early-stage detection of complex fire scenarios. However, challenges persist, including increased computational complexity and longer response times. Late or hybrid fusion is especially useful for

TABLE IV
SUMMARY OF INTERMEDIATE FUSION IN MULTIMODAL GAS DETECTION

Study	Input Modalities	Fusion Operation	Model	Task/Output	Key Results & Limitations
Wang et al. (2024) [26]	Thermal (ShuffleNetV2), Gas (7 MOX sensors)	Cross-attention intermediate fusion (SCGA)	1D-CNN + Bi-GRU	4-class gas leak classification	99.22% acc.; lacks robustness testing under noise or mixed gases
Attallah (2023) [27]	Thermal + MOX gas sensors	DWT/DCT fusion after CNN encoding	ResNet, MobileNet + BiLSTM	Gas classification (perfume, smoke, mixture, neutral)	98.47% acc.; strong in controlled settings; lacks real-time validation
Sharma et al. (2024) [28]	Thermal (VGG, DenseNet), Gas (LSTM/BiLSTM)	Feature fusion before final classifier	CNN + BiLSTM (federated)	Multiclass gas detection	99.7% acc.; not tested under noise, drift or edge deployment
Zhang and Zhang (2024) [29]	Thermal (CNN) + Gas sensors	Feature concatenation + ensemble learning	Deep Forest Classifier	Real-time leak detection	98.5% (clean), 88.1% (field); poor performance in low-concentration cases
Kang et al. (2024) [30]	Visible + IR imaging	Dual-stream attention-based fusion	YOLOv5 + 2D Transformer	VOC visual detection	F1 = 0.601; accurate but computationally intensive
Yan et al. (2024) [32]	CNN (thermal) + DNN (gas)	Feature-level co-learning	Parallel CNN + DNN	Fire/gas detection under partial input loss	Resilient to 90% thermal data dropout
Nakip et al. (2021) [33]	Temp, Smoke, CO, CO_2, O_2	Trend, level, and raw feature concatenation	Recurrent Trend Predictive NN	Binary fire classification	96% acc.; tested only on imbalanced public dataset
Li et al. (2025) [34]	Hyperspectral + Thermal vegetation imagery	Multi-attentive co-fusion	CBMAFNet + 3D CNN	Micro-leakage detection	94.6% acc.; good interpretability; field scope limited
Alam et al. (2024) [35]	Gas thresholds + Roof imagery	Rule + CNN feature fusion	MobileNet + decision logic	Mine hazard alerting	86.81% acc.; real-time alerts; hardware/battery limited
Wang et al. (2021) [36]	CO, SO_2, Radon, Temp, Pressure	AHP + fuzzy logic index (IFZD)	Expert-rule-based model	Fire zone classification	Field-validated; static weights; lacks adaptability
Huang et al. (2023) [37]	14 MOX gas sensors	2D spatial + 1D temporal fusion	MAM-Net (CBAM + BiLSTM)	Fault detection/classification	F1 99%; generalizes well; high dimensionality
Jang et al. (2023) [38]	56 sensors + Temp/Hum	Implicit feature fusion	MCGCN	Narcotics detection	90–98% acc.; high input complexity; not fire/gas specific
Bin et al. (2023) [39]	IR thermal + foreground maps	Foreground fusion (FFBGD)	DCN + FFN + ROI head	LNG leak detection via video	Strong AP; lacks domain generalization; high computation
Rahate et al. (2023) [31]	CNN (thermal) + DNN (gas)	Modality-co-learning (feature fusion)	CNN + DNN	Robust gas/fire classification	Resilient fusion; assumes at least partial modality input

fire detection, where different signals like gas levels, temperature changes, and camera images need to be combined or checked in sequence to make accurate decisions. The relatively sparse literature in this area also suggests a need for further benchmarking and real-time validation of such architectures across varied environmental conditions and use cases. Overall, while hybrid fusion is gaining popularity, late fusion remains underexplored and presents opportunities for more resilient system designs.

V. DISCUSSION

The reviewed studies show that multimodal models generally outperform unimodal ones in gas and fire detection tasks. However, no single fusion strategy is universally superior—each is suited to specific use cases [46].

Early fusion is effective when all modalities are reliably available and provide complementary information. It is commonly used in lightweight systems, such as embedded or IoT devices, where sensor data like gas, temperature, and humidity are combined at the input level for fast decision-making [23], [24]. However, it struggles when data is missing or noisy, and often lacks flexibility in handling heterogeneous data types.

Intermediate fusion is more adaptable and powerful for scenarios involving different types of data, such as thermal images and time-series gas sensor readings. It allows each modality to be processed separately using specialized encoders (e.g., CNNs for images, RNNs for sequences), and then merges

TABLE V
LATE AND HYBRID FUSION IN MULTIMODAL GAS DETECTION

Study (Year)	Input Modalities	Fusion Operation	Model(s)	Task/Output	Key Results & Limitations
Azizian et al. (2024) [40]	Thermal images, MQ gas sensors	Late fusion: Weighted averaging of ResNet-50 and kNN classifier outputs	ResNet-50 + kNN	4-class gas detection	99.32% accuracy; better than unimodal; limited to offline use; lacks robustness to blended gases
Jana and Shome (2023) [41]	MQ sensors, temperature, humidity	Late fusion: Average voting of Logistic Regression, SVM, Decision Tree, and Naive Bayes classifiers	LR, SVM, DT, NB	Fire classification	Precision 0.9752, AUC 0.996; vulnerable to feature imbalance and overfitting
Gaur et al. (2023) [42]	Wire current, temperature, gas, smoke, thermal image	Hybrid: Intermediate fusion of current/temp using fuzzy logic; late fusion of sensor thresholds with fuzzy output; CNN for visual input	Fuzzy Logic + CNN (MobileNet)	Electrical and general fire detection	98% accuracy; lacks diverse visual data; multi-stage complexity
Samson (2025) [43]	Electrochemical, catalytic bead, NDIR sensors; IR/UV/Vis flame imaging	Hybrid: Feature fusion for gas via ML; independent CNNs for flame analysis; final decision via late-stage integration	CNNs + ML classifiers	Industrial fire/leak detection	95% accuracy; 5% false alarms; affected by real-world sensor degradation
Yusuff et al. (2021) [44]	Gas, smoke, IR, acoustic sensors	Hybrid: Feature-level fusion of sensor streams followed by late fusion across classifier outputs	Ensemble logic	Early-stage fire detection	Effective in smoldering fires; outperforms unimodal systems; needs deeper fusion models
Burnias (2024) [45]	Gas, temp, humidity, flame, light, GPS, image	Hybrid: Early fusion via Z-score anomaly detection, fuzzy logic, and RFC; followed by late visual confirmation using ResNet50 CNN	Fuzzy Logic + RFC + CNN (ResNet50)	Fire detection (indoor/outdoor)	99% accuracy on sensors; 90% on images; based on simulated data; requires stable connectivity

the extracted features for classification [28], [30], [33]. This makes it ideal for capturing complex spatial-temporal patterns. Still, intermediate fusion can be computationally intensive, and performance may drop if inputs are incomplete or poorly calibrated [39], [29].

Late fusion is useful when modalities are asynchronous or have variable reliability. Since each data stream is processed independently and fused at the decision stage, it offers greater robustness to missing or degraded sensor input [40], [41]. This approach works well in modular or offline systems, but may miss deeper cross-modal feature relationships.

Hybrid fusion combines both intermediate and late fusion techniques and is especially valuable in fire detection systems, where sensor data may first trigger an alert and image-based confirmation follows [42], [43], [45]. Hybrid models offer high accuracy and flexibility but often involve multi-stage architectures, making them more complex and resource-intensive.

Ultimately, the best fusion method depends on the task requirements, data quality, and system constraints. Tasks requiring low-latency and consistent input may benefit from early fusion, while tasks involving diverse data formats or needing richer context often require intermediate or hybrid strategies. Late fusion remains less explored but holds promise for resilient and modular applications.

VI. FUTURE DIRECTIONS

Future research should focus on addressing the limitations of current fusion techniques through robust feature alignment methods, noise-aware architectures, and evaluation on diverse, real-world datasets. Additionally, further benchmarking and real-time validation of late and hybrid architectures are needed. Exploring the role of interpretability and explainable AI (XAI) in multimodal fusion models is also essential—particularly for safety-critical applications, where understanding model decisions can improve trust and deployment. Finally, more experimental work is needed to explore hybrid fusion strategies, including novel combinations of early, intermediate, and late fusion pipelines that can potentially enhance efficiency.

VII. CONCLUSION

In this paper we explored the application of machine learning and deep learning techniques in multimodal gas detection across different dimensions based on the stage of fusion – early , intermediate , late and hybrid. Analysing these approaches we have reached the following conclusions :

1) Multimodal models generally outperform unimodal models in the context of gas detection. This suggests that integrating data from multiple sensor types or modalities leads to improved detection capabilities compared to relying on a single source of information.

2) The choice of fusion type (early, intermediate, or late/hybrid) is highly task-specific, and there is no conclusive evidence that one type is universally superior. The most effective approach depends on factors such as the nature of the data, the specific detection task, and the available computational resources.
3) A significant limitation across many studies, particularly those employing early and intermediate fusion, is the reliance on synthetic or simulation-based datasets, which can limit their real-world applicability. There is a need for more evaluation on diverse, real-world datasets and under noisy, long-term operational conditions.

In conclusion, while multimodal fusion shows great promise for enhancing gas detection systems, the optimal fusion strategy is context-dependent, and further research is needed to address current limitations and improve real-world applicability.

References

[1] A. Sharma, V. Khullar, I. Kansal, G. Chhabra, P. Arora, R. Popli, and R. Kumar, "Gas detection and classification using multimodal data based on federated learning," *Sensors*, vol. 24, no. 18, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/18/5904

[2] M. A. Z. Chowdhury and M. A. Oehlschlaeger, "Artificial intelligence in gas sensing: A review," *ACS Sensors*, mar 2025. [Online]. Available: https://pubs.acs.org/doi/10.1021/acssensors.4c02272

[3] P. Narkhede, R. Walambe, S. Mandaokar, P. Chandel, K. Kotecha, and G. Ghinea, "Gas detection and identification using multimodal artificial intelligence based sensor fusion," *Applied System Innovation*, vol. 4, no. 1, p. 3, January 2021. [Online]. Available: https://www.mdpi.com/2571-5577/4/1/3

[4] P. Jadhav, V. A. Sairam, N. Bhojane, A. Singh, S. Gite, B. Pradhan, M. Bachute, and A. Alamri, "Multimodal gas detection using e-nose and thermal images: An approach utilizing srgan and sparse autoencoder," *Computers, Materials & Continua*, vol. 83, no. 2, pp. 3493–3517, 2025, open Access under CC BY 4.0. [Online]. Available: https://www.techscience.com/cmc/v83n2/60537

[5] W. Yan, W. Liu, Q. Zhang, H. Bi, C. Jiang, H. Liu, T. Wang, T. Dong, and X. Ye, "Multisource multimodal feature fusion for small leak detection in gas pipelines," *IEEE Sensors Journal*, vol. 24, no. 2, pp. 1857–1865, 2024.

[6] "The evolution of gas detection - interscan — fixed portable industrial gas detectors." [Online]. Available: https://gasdetection.com/articles/the-evolution-of-gas-detection/?utm_source=chatgpt.com

[7] S. Panda, S. Mehlawat, N. Dhariwal, A. Kumar, and A. Sanger, "Comprehensive review on gas sensors: Unveiling recent developments and addressing challenges," *Materials Science and Engineering: B*, vol. 308, p. 117616, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921510724004458

[8] S. Dhall, B. Mehta, A. Tyagi, and K. Sood, "A review on environmental gas sensors: Materials and technologies," *Sensors International*, vol. 2, p. 100116, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666351121000371

[9] M. F. R. Al-Okby, S. Neubert, T. Roddelkopf, and K. Thurow, "Mobile detection and alarming systems for hazardous gases and volatile chemicals in laboratories and industrial locations," *Sensors*, vol. 21, no. 23, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/23/8128

[10] M. Fleischer and H. Meixner, "Thin-film gas sensors based on high-temperature-operated metal oxides," *Journal of Vacuum Science Technology A*, vol. 17, pp. 1866–1872, 7 1999. [Online]. Available: /avs/jva/article/17/4/1866/980636/Thin-film-gas-sensors-based-on-high-temperature

[11] P. Barik and M. Pradhan, "Selectivity in trace gas sensing: recent developments, challenges, and future perspectives," *Analyst*, vol. 147, pp. 1024–1054, 2022. [Online]. Available: http://dx.doi.org/10.1039/D1AN02070F

[12] I. Darmadi, F. A. A. Nugroho, and C. Langhammer, "High-performance nanostructured palladium-based hydrogen sensors - current limitations and strategies for their mitigation," *ACS Sensors*, vol. 5, pp. 3306–3327, 11 2020. [Online]. Available: https://pubs.acs.org/doi/full/10.1021/acssensors.0c02019

[13] F. I. M. Ali, F. Awwad, Y. E. Greish, and S. T. Mahmoud, "Hydrogen sulfide (h2s) gas sensor: A review," *IEEE Sensors Journal*, vol. 19, no. 7, pp. 2394–2407, 2019.

[14] S. Kumar, O. Ivanova, A. Melyokhin, and P. Tiwari, "Deep-learning-enabled multimodal data fusion for lung disease classification," *Informatics in Medicine Unlocked*, vol. 42, p. 101367, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352914823002137

[15] G. Arya, A. Bagwari, S. Agarwal, J. K. Aneja, and C. Rodriguez, "A comparative analysis of early fusion architectures for multimodal gas detection using machine learning models," *Instrumentation, Mesures, Métrologies*, vol. 23, no. 4, pp. 297–306, 2024, open Access under CC BY 4.0. [Online]. Available: https://www.iieta.org/journals/i2m/paper/10.18280/i2m.230405

[16] A. Kopbayev, F. Khan, M. Yang, and S. Z. Halim, "Gas leakage detection using spatial and temporal neural network model," *Process Safety and Environmental Protection*, vol. 160, pp. 968–975, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957582022001999

[17] S.-J. Ro and K. Lee, "Early fire detection system for embedded platforms: Deep learning approach to minimize false alarms," *Journal of Sensor Science and Technology*, vol. 33, no. 5, pp. 298–304, September 2024, in Korean: : .

[18] G.-L. Kim, S.-J. Ro, and K. Lee, "A multi-sensor fire detection method based on trend predictive bilstm networks," *Journal of Sensor Science and Technology*, vol. 33, no. 5, pp. 248–254, September 2024, in Korean: : BiLSTM .

[19] G. Tejaswi, R. Bhavani, S. Srihitha, S. Arshiha, and R. V. S. Sarayu, "Predicting fire alarms using multi sensor data: A binary classification approach," *Turkish Journal of Computer and Mathematics Education*, vol. 15, no. 1, pp. 242–255, March 2024, open Access under CC BY 4.0. [Online]. Available: https://turcomat.org/index.php/turkbilmat/article/view/14617

[20] S. Chen, J. Ren, Y. Yan, M. Sun, F. Hu, and H. Zhao, "Multi-sourced sensing and support vector machine classification for effective detection of fire hazard in early stage," *Computers and Electrical Engineering*, vol. 101, p. 108046, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S004579062200307X

[21] L. Salhi, T. Silverston, T. Yamazaki, and T. Miyoshi, "Early detection system for gas leakage and fire in smart home using machine learning," in *2019 IEEE International Conference on Consumer Electronics (ICCE)*, 2019, pp. 1–6.

[22] C. Wen, K. Li, Y. Liao, and Z. Xiao, "Design of an intelligent alarm system based on multi-sensor data fusion," in *Journal of Physics: Conference Series*, vol. 1961, no. 1. IOP Publishing, 2021, p. 012025, presented at the 2021 International Conference on Computer Engineering and Innovative Application of VR (ICCEIA VR 2021), Guangzhou, China.

[23] S. M, D. S, L. V, A. S, and E. R. G, "Pioneering solutions for advanced gas leak detection in hazardous industrial and mine environments," in *2024 9th International Conference on Communication and Electronics Systems (ICCES)*, 2024, pp. 345–350.

[24] N. Divya, K. Aravindhan, S. Gowsi, and C. B. S. Raj, "Gas leakage detection in refrigerator system using iot," *The Indian Journal of Technical Education*, vol. 46, no. Special Issue, pp. 74–76, June 2023, iSSN: 0971-3034.

[25] G. Pan, Y. Xie, and Q. Yang, "Iot-based cloud monitoring system for building fires," *International Journal of Metrology and Quality Engineering*, vol. 16, no. 1, pp. 1–10, 2025, open Access under CC BY 4.0. [Online]. Available: https://doi.org/10.1051/ijmqe/2024020

[26] X. Wang, Y. Yang, H. Tian, Y. Chen, and M. Zhang, "Gas leak detection based on cross-attention multi-source data fusion," *China Safety Science Journal*, vol. 34, no. 7, pp. 91–97, 2024. [Online]. Available: https://www.cssjj.com.cn/EN/10.16265/j.cnki.issn1003-3033.2024.07.0135

[27] O. Attallah, "Multitask deep learning-based pipeline for gas leakage detection via e-nose and thermal imaging multimodal fusion," *Chemosensors*, vol. 11, no. 7, 2023. [Online]. Available: https://www.mdpi.com/2227-9040/11/7/364

[28] A. Sharma, V. Khullar, I. Kansal, G. Chhabra, P. Arora, R. Popli, and R. Kumar, "Gas detection and classification using multimodal data based on federated learning," *Sensors*, vol. 24, no. 18, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/18/5904

[29] E. Zhang and E. Zhang, "Development of a multimodal deep feature fusion with ensemble learning architecture for real-time gas leak detection," in *2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI)*, 2024, pp. 1–6.

[30] Y. Kang, K. Shi, J. Tan, Y. Cao, L. Zhao, and Z. Xu, "Multimodal fusion induced attention network for industrial vocs detection," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 12, pp. 6385–6398, 2024.

[31] A. Rahate, S. Mandaokar, P. Chandel, and B. P. K. Reddy, "Employing multimodal co-learning to evaluate the robustness of sensor fusion for industry 5.0 tasks," *Soft Computing*, vol. 27, pp. 4139–4155, 2023, published: 07 March 2022, Issue Date: April 2023. [Online]. Available: https://doi.org/10.1007/s00500-022-06802-9

[32] W. Yan, W. Liu, Q. Zhang, H. Bi, C. Jiang, H. Liu, T. Wang, T. Dong, and X. Ye, "Multisource multimodal feature fusion for small leak detection in gas pipelines," *IEEE Sensors Journal*, vol. 24, no. 2, pp. 1857–1865, 2024.

[33] M. Nakip, C. Güzelíş, and O. Yildiz, "Recurrent trend predictive neural network for multi-sensor fire detection," *IEEE Access*, vol. 9, pp. 84 204–84 216, 2021.

[34] K. Li, K. Xiong, J. Jiang, and X. Wang, "A convolutional block multi-attentive fusion network for underground natural gas micro-leakage detection of hyperspectral and thermal data," *Energy*, vol. 319, p. 134870, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544225005122

[35] A. F. Alam, M. Tabassum, M. Shabbir, M. A. Raza, and F. Khan, "A monitoring and warning system for hazards in coal mines using cnn and sensor fusion," in *2024 5th International Conference on Advancements in Computational Sciences (ICACS)*, 2024, pp. 1–9.

[36] H. Wang, X. Fang, Y. Li, Z. Zheng, and J. Shen, "Research and application of the underground fire detection technology based on multi-dimensional data fusion," *Tunnelling and Underground Space Technology*, vol. 109, p. 103753, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0886779820307070

[37] P. Huang, Q. Wang, H. Chen, and G. Lu, "Gas sensor array fault diagnosis based on multi-dimensional fusion, an attention mechanism, and multi-task learning," *Sensors*, vol. 23, no. 18, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/18/7836

[38] Y. Jang, H. W. Noh, H. D. Park, K. H. Chung, and C.-G. Ahn, "Deep nose project: A study of multidimensional multimodal olfactory intelligence system for ultra-trace gas component detection," in *2023 IEEE SENSORS*, 2023, pp. 1–4.

[39] J. Bin, Z. Bahrami, C. A. Rahman, S. Du, S. Rogers, and Z. Liu, "Foreground fusion-based liquefied natural gas leak detection framework from surveillance thermal imaging," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 4, pp. 1151–1162, 2023.

[40] A. Azizian, B. Yousefimehr, and M. Ghatee, "Enhanced multi-modal gas leakage detection with nsmote: A novel over-sampling approach," in *2024 8th International Conference on Smart Cities, Internet of Things and Applications (SCIoT)*, 2024, pp. 94–99.

[41] S. Jana and S. K. Shome, "Hybrid ensemble based machine learning for smart building fire detection using multi modal sensor data," *Fire Technology*, vol. 59, pp. 473–496, 2023. [Online]. Available: https://doi.org/10.1007/s10694-022-01347-7

[42] A. Gaur, A. Singh, A. Verma, and A. K. and, "Artificial intelligence and multi-sensor fusion based universal fire detection system for smart buildings using iot techniques," *IETE Journal of Research*, vol. 69, no. 12, pp. 9204–9216, 2023. [Online]. Available: https://doi.org/10.1080/03772063.2022.2088626

[43] O. Samson, "Smart optimization of flame and gas detectors for enhanced industrial safety," *Preprint on ResearchGate*, February 2025. [Online]. Available: https://www.researchgate.net/publication/389356978

[44] M. Yusuff and S. Author], "Multi-sensor fusion for early fire detection," *Preprint on ResearchGate*, December 2021. [Online]. Available: https://www.researchgate.net/publication/388198459

[45] J. A. Burnias, "A multi sensor indoor/outdoor real time iot hybrid model approach for fire detection," Ph.D. Dissertation, The University of Texas at San Antonio, 2024.

[46] J. Lipkova, R. Chen, B. Chen, M. Lu, M. Barbieri, D. Shao, A. Vaidya, C. Chen, L. Zhuang, D. Williamson, M. Shaban, T. Chen, and F. Mahmood, "Artificial intelligence for multimodal data integration in oncology," *Cancer Cell*, vol. 40, pp. 1095–1110, Oct. 2022.