© Volume 2, Issue 11, November 2026 | JATIR
e-ISSN: 2839-5786, p-ISSN: 3139-8464

Machine Learning Driven Analysis and Prediction of Hazardous Gas Concentrations: A Comprehensive Review of Sensing Data, Algorithms, and Emerging Applications

[1]Kavya Shivanand Kandgal, [2]Dr. Annamalail, [3]Dr. A. Suresh Kumar
[1,2,3]*Department of Computer Science, JAIN (Deemed to be University), Bengaluru, India*
https://doi.org/10.64643/jatir.141288

***Abstract*—Hazardous gases—carbon monoxide, hydrogen sulphide, ammonia, nitrogen oxides, methane, and volatile organic compounds—pose severe risks to health, industrial safety, and environmental quality, making continuous monitoring and early prediction a critical engineering priority. Conventional sensing systems, based on metal-oxide-semiconductor (MOS), electrochemical, and infrared transducers, suffer from cross-sensitivity, baseline drift, and non-linear response, limiting the reliability of threshold-based alarms. Over the past decade, machine learning (ML) and deep learning (DL) have emerged as powerful tools for extracting selective, drift-resilient, and temporally predictive information from sensor signals, enabling accurate gas identification and multi-step-ahead concentration forecasting. This review surveys thirty recent studies (2020–2026) coupling hazardous gas sensing hardware with ML pipelines across domains including IoT-based indoor and rural air quality monitoring, underground coal mine safety, electronic nose (eNose) gas mixture identification, power transformer dissolved gas analysis, industrial emission control, and cyber-physical anomaly detection. Methodologies include classical algorithms (SVM, Random Forests, k-NN), ensembles (XGBoost, Gradient Boosting, Stacking), neural networks, and sequence-to-sequence models (LSTM, GRU, BiLSTM, BiGRU, Transformer, Autoformer). The comparison proves that hybrid and bidirectional architectures deliver better performance in multi-step forecasting tasks compared to a single model, while tree-based ensembles continue to be relevant for low-latency edge computation tasks. The review highlights the difficulties connected with sensor drifting, dataset limitations, cross-sensitivity, and the challenge of transferring models from simulated environment to the real world, suggesting future research directions in federated learning, benchmarks, explainable AI, and edge-AI compression.**

***Index Terms*—Machine learning; Deep learning; Hazardous gas detection; Gas concentration prediction; Internet of Things; Electronic nose; Time-series forecasting; Edge computing**

I. INTRODUCTION

The presence of dangerous gaseous species in the air environment, such as CO, H_2S, NH_3, NO_x, SO_2, CH_4, and many others, is one of the most persistent sources of danger to human health, industrial safety, and environmental sustainability [1]. Acute and chronic exposure to such toxic substances as a rule causes respiratory diseases, cardiovascular diseases, asphyxia, and, in closed premises, the explosion and fire hazards [2]. Every type of dangerous gas has its regulatory threshold limit value (TLV), and violation of this threshold even at once leads to health problems, which is the reason for the necessity of monitoring system to be timely, selective, and sensitive [3].

Gas sensor devices, such as metal oxide semiconductor (MOS) Chemi resistors, electrochemical cells, non-dispersive infrared (NDIR) sensors, and cheap MQ-series sensors have made the realization of monitoring infrastructures cost-effective [4]. But still, the above-mentioned transducers have well-known drawbacks in terms of strong cross-sensitivity between similar structural gases, the tendency to baseline drift due to surface poisoning and ageing, sensitivity to humidity and temperature changes, and non-linearity of concentration-responsiveness functions [4]. As a result, simple threshold-based alarming algorithms, though being relatively easy to develop, are prone to false alarms and delayed reactions especially in dynamic multi-gas atmosphere, for example, in coal mines, chemical plants, or vapor recovery units [3].

Machine Learning (ML), and to a lesser degree – Deep Learning (DL) – has revolutionized the approach to transforming unprocessed data coming from sensors into reliable gas detection and concentration estimation. Due to the complexity of non-linear relationships learned by ML algorithms, they are capable of solving the problem of sensor drifts, decoupling overlapping signatures of gases, and predicting future concentration trajectory of several minutes before the hazard threshold[5]. It is especially useful in the safety-related field of underground coal mining, in which the process of gas outbursts is characterized by its shortness and non-linearity[6], and in asset monitoring of power systems where dissolved gas analysis (DGA) of transformer oil is used as a proxy to insulation faults prediction [7].

The advent of IoT, edge computing and TinyML architectures opened up an opportunity to use pre-trained ML models on resource-constrained microcontrollers and perform real-time inference without the need of always-on connection to cloud servers TinyML[8]. Additionally, thanks to the progress made in the field of electronic nose (eNose) technologies, which include cheap sensors and pattern recognition algorithms, there have been a variety of applications for the eNoses developed, including quality assessment of food and non-invasive diagnostic health care procedures[9]. Moreover, there are a number of important methodological papers that have shown the overestimation of accuracy in publicly available benchmark gas-sensor datasets due to improperly treated drift artifacts[10].

Despite the rapid growth of this research area, the literature remains fragmented across disparate application domains, sensor types, and modelling paradigms, making it difficult for researchers and practitioners to obtain a unified perspective on which combinations of sensing hardware and

ML architecture are best suited to a given hazardous gas monitoring scenario. Motivated by this gap, the present review systematically surveys thirty recent studies (2020-2026) that integrate hazardous gas sensing with machine learning for concentration analysis and prediction. The specific objectives of this review are to: (i) classify the sensing technologies and data acquisition strategies employed across the surveyed literature; (ii) categorize and critically compare the machine learning and deep learning methodologies applied to hazardous gas concentration analysis and forecasting; (iii) organize the reviewed studies according to their principal application domain; (iv) quantitatively compare the reported performance metrics across studies; and (v) identify the prevailing challenges and outline future research directions capable of advancing the field toward robust, field-deployable, and explainable hazardous gas intelligence systems.

II. SOURCES, HEALTH IMPACT, AND MONITORING REQUIREMENTS OF HAZARDOUS GASES

Hazardous gases relevant to the reviewed literature originate from a broad spectrum of sources: combustion and vehicular emissions (CO, NO_x, SO_2), biological decomposition and sewage (H_2S, NH_3, CH_4), industrial and petrochemical processing (VOCs, benzene, toluene), underground coal seams (CH_4, CO), and electrical insulation ageing in oil-filled power transformers (H_2, C_2H_2, C_2H_4, C_2H_6)[11] . Table I summarizes representative gases discussed across the reviewed studies together with their principal sources, typical health or safety impact, and the monitoring context in which they are most frequently studied.

TABLE I REPRESENTATIVE HAZARDOUS GASES DISCUSSED IN THE REVIEWED LITERATURE, THEIR PRINCIPAL SOURCES, HEALTH/SAFETY IMPACT, AND MONITORING CONTEXT.

Gas Species	Principal Sources	Health / Safety Impact	Representative Monitoring Context
CO / NO_x / SO_2	Combustion, vehicular exhaust, coal-fired power plants	Impaired oxygen transport, respiratory and cardiovascular stress	Industrial stack emissions [21]; IoT ambient monitoring [2,3]
H_2S / NH_3	Sewage, agriculture, chemical/refrigeration industries, brownfield sites	Chemical asphyxiation, corrosive and irritant effects	Toxic-gas field monitoring [26]; MOF/MOx sensing [12,23]
CH_4 (methane)	Underground coal seams, natural gas leakage	Explosion and asphyxiation hazard in confined spaces	Coal mine safety forecasting [9,16,20,24,25,27]

VOCs (acetone, ethanol, methanol, benzene, toluene)	Industrial solvents, essential oils, breath biomarkers, vapour recovery units	Narcotic, carcinogenic, and respiratory effects at elevated exposure	eNose classification [10,17,23]; vapour recovery control [28]
Dissolved gases (H_2, C_2H_2, C_2H_4, C_2H_6)	Thermal and electrical degradation of transformer insulating oil	Precursor to catastrophic transformer failure	Transformer dissolved gas analysis (DGA) [18,22]
Radon / CO_2	Soil radioactivity, indoor occupancy, poor ventilation	Long-term carcinogenic risk; indoor air quality degradation	Rural/indoor decision-support platforms [1,7]

Because each of these gas families exhibits distinct chemical reactivity, concentration range, and temporal dynamics, the choice of sensing hardware and the corresponding ML modelling strategy differ considerably across domains, motivating the structured review presented in the following sections.

III. SENSING TECHNOLOGIES AND DATA ACQUISITION STRATEGIES

Across the reviewed studies, hazardous gas data acquisition relies predominantly on four hardware categories: low-cost MQ-series/MOS Chemi resistors [2], engineered interdigital-electrode (IDE) and nanostructured Chemi resistive/field-effect transistor (FET) sensors [9], dissolved gas or oil-chromatography analysers for power-system assets [7], and multi-sensor eNose arrays coupled with dynamic thermal or pulse-width-modulated excitation to enhance selectivity from a single sensing channel [12]. Regardless of hardware type, three data acquisition paradigms recur throughout the literature: (i) static/laboratory chamber exposure, in which known gas concentrations are injected under controlled temperature and humidity to build labelled training datasets [6]; (ii) continuous field or industrial deployment, generating long, often noisy, real-world time-series requiring cleaning and interpolation prior to model training [3]; and (iii) simulation-based data generation, used notably for electromagnetic metamaterial gas sensors where full-wave simulations substitute for costly physical prototyping[9] .

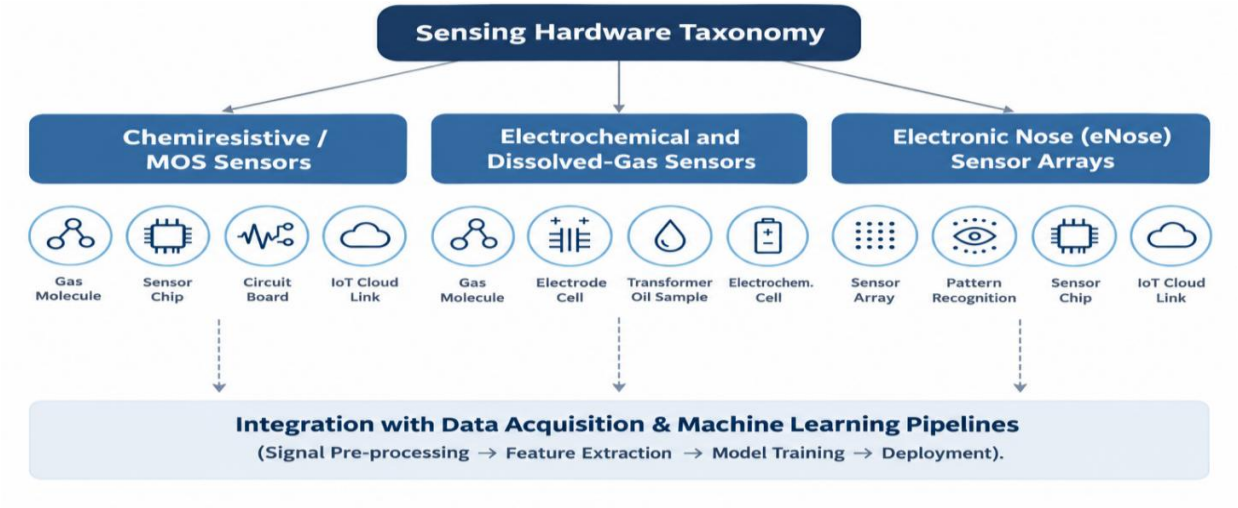


Fig. 1. Conceptual taxonomy of hazardous gas sensing technologies (MOS/Chemi resistive, electrochemical, dissolved-gas/oil-chromatography, and eNose sensor arrays) and their integration with data acquisition and machine learning pipelines, reproduced in a style consistent with the layout of Fig. 1.

A common pre-processing challenge identified across the reviewed studies is the correction of missing or corrupted samples in continuous industrial time-series, typically addressed through statistical outlier rejection (e.g., the Pauta/Laida 3-sigma criterion) followed by Lagrange interpolation [9,20], Z-score/median imputation [21], or exponential-moving-average-based imputation embedded directly within the forecasting architecture [18]. Feature engineering approaches range from simple statistical descriptors (mean, variance, slope) computed over sliding windows [10] to automated time-series feature extraction libraries such as tsfresh [13,23], Pearson correlation-based dimensionality reduction [25], and unsupervised dimensionality-reduction diagnostics such as t-distributed stochastic neighbour embedding (t-SNE) used to expose drift artefacts in benchmark datasets [11].

IV. MACHINE LEARNING METHODOLOGIES FOR HAZARDOUS GAS CONCENTRATION ANALYSIS AND PREDICTION

The thirty reviewed studies apply a broad spectrum of machine learning and deep learning methodologies to two related but distinct tasks: (a) gas type identification/classification, particularly relevant to multi-gas or mixed-gas eNose scenarios, and (b) concentration regression/forecasting, relevant to time-series prediction of a known gas species ahead of a hazard threshold. Fig. 2 summarizes the relative frequency with which each broad technique family appears across the reviewed literature.

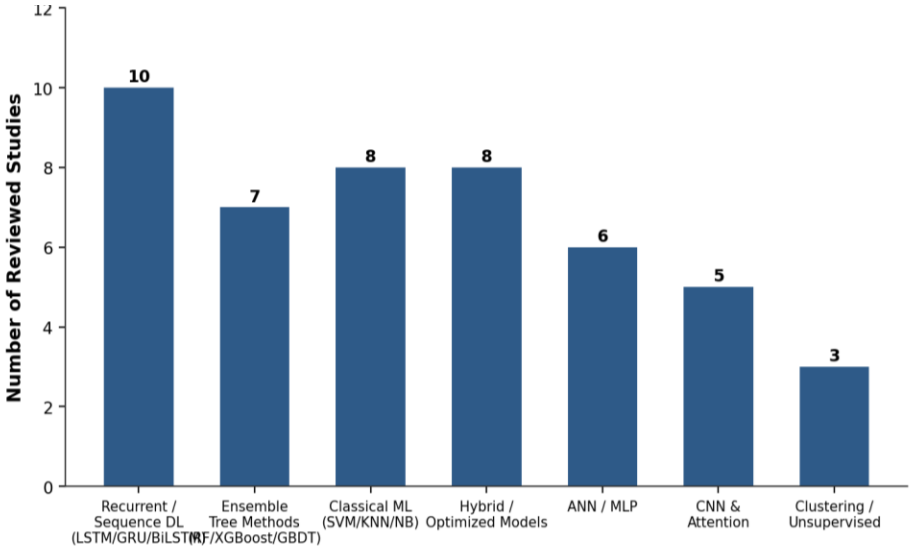


Fig. 2. Distribution of machine learning technique families across the thirty reviewed studies on hazardous gas concentration analysis and prediction.

A. Classical Machine Learning and Ensemble Tree-Based Methods

Support vector machines (SVM), k-nearest neighbours (KNN), random forests (RF), and gradient-boosted tree ensembles (XGBoost, LightGBM, GBDT) remain widely used, particularly where dataset sizes are moderate and interpretability or low inference latency is prioritized [1]. Random forest classifiers have been reported to outperform simpler regression and neural baselines for field-calibrated low-cost sensor networks, achieving 82.8% precision against 76.1% for an equivalent multilayer perceptron [9], while RF-based dual regression/classification pipelines achieved $R^2 \approx 0.90$ and classification accuracy exceeding 90% for industrial vapour-recovery outlet gas concentration prediction [7]. Stacking ensembles that combine random forest, extra-trees, and GBDT base learners through a meta-regressor, tuned via automated grid search, have been shown to substantially reduce regression error for overlapping mixed-gas concentration estimation relative to any single base learner [13]. Similarly, a selective ensemble methodology based on the Yager fuzzy triangular norm, combining autoregressive and back-propagation neural base learners, achieved lower forecasting error across five evaluation metrics than any individual base model for coal mine gas forecasting [6].

B. Artificial Neural Networks and Multilayer Perceptron's

Feed-forward artificial neural networks (ANN) and multilayer perceptrons (MLP) continue to serve as effective baseline regressors for both concentration prediction and sensor calibration, especially where the underlying relationship between raw sensor voltage and true concentration is strongly non-linear[14] . A multi-layer feed-forward neural network trained on chamber-collected fecal odour gas mixtures achieved high correlation ($R^2 > 0.90$) for quantifying trace sulphur-compound and hydrogen concentrations in a non-invasive smart-toilet health monitoring context [15], illustrating the versatility of ANN-based regression beyond conventional industrial settings.

C. Recurrent and Sequence Deep Learning Models

Given the inherently sequential and non-stationary nature of gas concentration time-series, recurrent architectures, particularly long short-term memory (LSTM), gated recurrent unit (GRU), and their bidirectional variants (BiLSTM, BiGRU), dominate the forecasting-oriented subset of the reviewed literature [7,9,18,20,21,22,25,26,27]. A GRU-based model for underground coal mine methane forecasting outperformed LSTM, RNN, back-propagation neural networks, and support vector regression, reducing prediction error by 7.9% relative to LSTM while improving computation speed by over 13%[16] . Building on this, an Adamax-optimized bidirectional GRU (Adamax-BiGRU) architecture further reduced test error by 25.6% relative to a standard RNN, 12.5% relative to LSTM, and 3.0% relative to unidirectional GRU, by exploiting both forward and backward temporal context together with a numerically stable optimizer [17]. A stacked bidirectional-unidirectional LSTM model deployed within a rural IoT decision-support platform achieved R^2 above 0.94 for indoor CO_2 and radon forecasting, outperforming GRU and Transformer baselines on a comparatively small (17,236-sample) dataset [11]. LSTM models have likewise proven effective for industrial stack-gas emission forecasting from continuous emission monitoring system (CEMS) data, achieving R^2 of 0.87 for NO_x and 0.85 for SO_2, although volatile CO emissions remained comparatively harder to predict ($R^2 = 0.67$) [18].

D. Hybrid, Attention-Augmented, and Optimization-Enhanced Architectures

A significant and rapidly growing share of the reviewed literature (eight of thirty studies) proposes hybrid architectures that combine signal decomposition, attention mechanisms, and metaheuristic hyperparameter optimization to further improve forecasting accuracy. The NRBO-VMD-AM-BiLSTM framework decomposes non-stationary dissolved-gas signals into stationary intrinsic mode functions using Newton-Raphson-based-optimized variational mode decomposition, before an attention-augmented BiLSTM predicts each component; this framework achieved a hydrogen prediction RMSE of only 0.51 μL/L (MAPE = 1.27%) for transformer oil dissolved-gas analysis[7] . A related EMA-Autoformer architecture combines exponential-moving-average missing-value imputation with an Auto-Correlation-based Autoformer, achieving over 15% error reduction relative to a standard Transformer for the same application domain [19]. In coal mine gas forecasting, a Sparrow-Search-Algorithm-optimized CNN-BiLSTM-Attention (SCBA) model achieved the slowest error-growth rate across twenty-step-ahead multi-step forecasting, improving RMSE by 23.3% relative to an attention-LSTM baseline and 30.2% relative to a Transformer-LSTM baseline [20]. A hybrid Fuzzy-XGBoost model, designed to jointly exploit fuzzy logic's tolerance of environmental (temperature/humidity) uncertainty and XGBoost's gradient-boosting power, achieved R^2 in the range 0.97-0.99 for acetone, ethanol, and methanol concentration prediction using custom nano-hematite molecularly imprinted sensors, outperforming either standalone model [5]. A big-data streaming framework (SPARS), combining ARIMA for linear temporal trend extraction with SVM for non-linear residual regression on an Apache Spark Streaming backend, achieved sub-second forecasting throughput for continuous coal mine gas monitoring [6].

E. Edge, TinyML, and Anomaly-Detection Approaches

A distinct and increasingly important subset of the literature targets deployment of trained models directly on resource-constrained microcontrollers and edge nodes. A TinyML pipeline combining polynomial regression and a lightweight MLP predicted indoor CO_2 one, fifteen, and thirty minutes ahead directly on an ESP32 microcontroller with minimal computational overhead (RMSE 0.5-7 ppm, R^2 0.80-0.99)[8] . An edge-assisted gas-sensing network (Ea-GNet), combining tsfresh feature extraction with random forest classification and support vector regression at the network edge, reduced end-to-end latency from 18 seconds to 52 milliseconds while achieving 98% gas identification accuracy and an R^2 of 0.84 for concentration regression, representing a throughput gain of over 600% relative to cloud-only inference [13]. From a cybersecurity perspective, a two-tier architecture combining a deterministic rule engine with an unsupervised Isolation Forest anomaly detector achieved approximately 98.5% detection accuracy for both sudden threshold breaches and subtle false-data-injection cyber-attacks on ESP32-based smart-building gas monitoring prototypes, with negligible computational overhead [5].

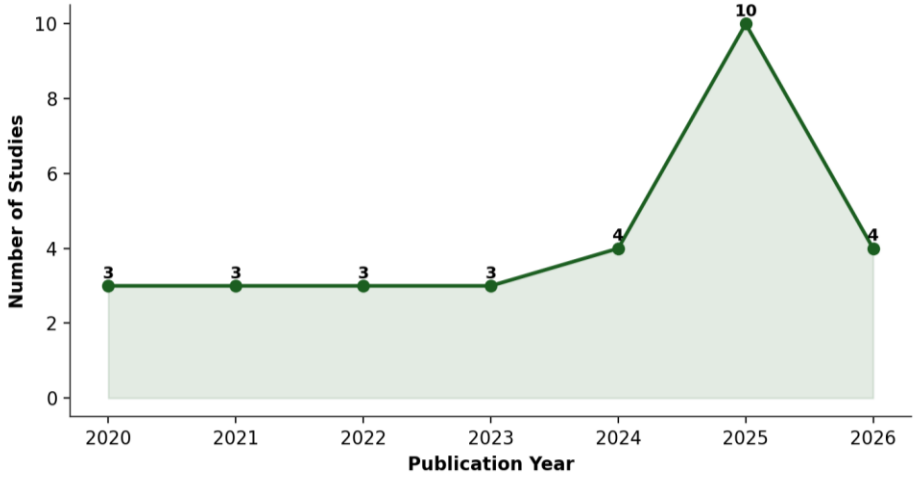


Fig. 3. Publication year trend of the thirty reviewed studies, illustrating the accelerating research interest in machine-learning-based hazardous gas concentration analysis and prediction between 2020 and 2026.

V. APPLICATION DOMAINS OF MACHINE-LEARNING-BASED HAZARDOUS GAS MONITORING

The reviewed studies naturally cluster into six overlapping application domains, summarized quantitatively in Fig. 4 and discussed in detail in the following subsections: (i) IoT- and edge-based indoor/ambient air quality monitoring; (ii) underground coal mine and industrial safety forecasting; (iii) electronic nose (eNose) based gas mixture identification and classification; (iv) power system and transformer dissolved gas analysis; (v) industrial process and emission control; and (vi) cyber-physical security and anomaly detection in gas monitoring networks.

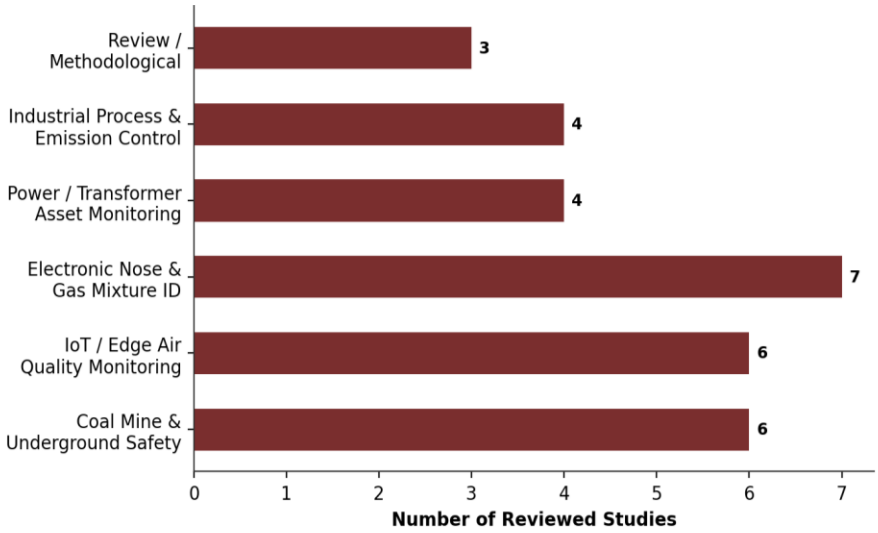


Fig. 4. Application domain distribution of the thirty reviewed machine-learning-based hazardous gas concentration studies.

A. IoT- and Edge-Based Indoor and Ambient Air Quality Monitoring

A recurring theme across several studies is the design of low-cost IoT sensor nodes, calibrated through machine learning against reference instruments, to democratize access to air quality monitoring beyond sparse and expensive regulatory-grade stations [8]. Bindhu et al. demonstrated that a random-forest-calibrated MQ-sensor IoT node achieved higher precision than an equivalent ANN-MLP model when predicting atmospheric CO, NO_2, and particulate trends against a reference CO monitor [1]. Complementary work has coupled such low-cost hardware with cloud SCADA/HMI dashboards (Thinger.io, Blynk) for real-time visualization and automated actuation, for example triggering ventilation fans within seconds of a threshold breach [1]. At the rural and remote end of this spectrum, a decision-support IoT platform (RuraLTHINGS) incorporating a stacked BiLSTM model has been proposed specifically to address the health risks of unmonitored indoor radon and CO_2 accumulation in isolated dwellings lacking conventional infrastructure [7].

B. Underground Coal Mine and Industrial Safety Forecasting

Underground coal mine methane forecasting constitutes the single largest and most methodologically mature cluster within the reviewed literature, comprising six studies that collectively trace a clear evolution from simple recurrent networks toward increasingly sophisticated hybrid and ensemble architectures [16]. Early work established GRU as a computationally efficient alternative to LSTM for this task [9], while subsequent studies progressively incorporated bidirectionality and adaptive optimization (Adamax-BiGRU) [17], selective fuzzy ensembling [5], Pearson-correlation-guided feature selection enabling fifteen-minute advance forecasting [21], distributed Spark-Streaming real-time processing for high-throughput sensor networks [22] , and, most recently, Sparrow-Search-Algorithm-optimized CNN-BiLSTM-Attention hybrids explicitly designed for robust multi-step-ahead forecasting

under multivariate spatio-temporal conditions (gas, temperature, wind speed, rock pressure) [20]. Complementing this cluster, an IoT-GSM collaborative warning system integrating regression-based prediction with cellular SMS alerting has been proposed for general industrial gas leak scenarios independent of Wi-Fi infrastructure, emphasizing rapid, network-independent emergency response [23].

C. Electronic Nose Based Gas Mixture Identification and Classification

A substantial share of the reviewed literature addresses the classification of overlapping or structurally similar gases using arrays of partially selective, low-cost sensors combined with pattern-recognition algorithms, an approach broadly termed the electronic nose (eNose) [24]. Notable results include a KNN-based eNose achieving flawless 100% multiclass accuracy for essential-oil authentication and adulteration detection using an array of seven MQ-series sensors[9] , and a dynamic pulse-width-modulated thermal-signature approach that achieved 98.3% classification accuracy for binary CO-NO_2 mixtures using only a single MOS sensor channel, substantially reducing hardware cost relative to conventional multi-sensor arrays [8]. Bae et al. provided an important theoretical contribution by quantitatively demonstrating a proportional relationship between an individual sensor's selectivity factor S(X) and the resulting machine learning classification accuracy, showing that pairing a low-selectivity sensor with a high-selectivity sensor can substantially boost combined-array accuracy above 95-98%, offering practical design guidelines for minimizing redundant sensors in future eNose arrays [4]. Two comprehensive review articles within this cluster further synthesize the broader Chemi resistive/FET and interdigital-electrode sensor literature, concluding that dynamic transient feature extraction (rather than static steady-state readings) is essential to achieving true selectivity and long-term drift compensation, with ML-assisted sensors routinely reported to achieve 95-100% accuracy across the surveyed body of work [15]. A methodologically important cautionary study demonstrated, however, that a widely used public benchmark dataset suffers from severe temporal/spatial drift artefacts that allow models to classify gases with spuriously high accuracy even before gas injection occurs, highlighting the urgent need for standardized, drift-corrected benchmark protocols across the eNose research community [10].

D. Power System and Transformer Dissolved Gas Analysis

Two studies specifically address dissolved gas analysis (DGA) in oil-immersed power transformers, a well-established early-warning indicator of incipient electrical or thermal faults [22]. Both studies confront the twin challenges of missing measurement data and long-range non-linear temporal dependencies inherent to transformer oil chromatography records, addressing them respectively through an EMA-Autoformer architecture [18] and a three-stage NRBO-VMD-AM-BiLSTM soft-sensor framework that achieved state-of-the-art hydrogen prediction accuracy [22], underscoring the value of signal decomposition and attention mechanisms for this high-stakes asset-monitoring application.

E. Industrial Process and Emission Control

Machine learning has not been limited to coal mines and transformers but has been successfully used in the industrial emission control scenario and process control domain such as stack gas emission prediction in coal-fired power plants based on continuous emission monitoring systems (CEMS) data [21] and vapour concentration prediction in petrochemical depots using vapour recovery units [25], both being directly related to environmental monitoring. Another case study of terahertz metamaterial absorber design via simulation also highlights the recent methodological development of using machine learning regressors as computationally cheaper proxies for full wave electromagnetic simulation in order to reduce the design time of hazardous gas optical sensors by 85% while delivering absorptivity of 99.96% with 905 GHz/RIU sensitivity [19].

F. Cyber-Physical Security and Health-Oriented Applications

Finally, two studies extend hazardous gas machine learning into adjacent but important niches: a hybrid rule-based and Isolation-Forest anomaly detection framework safeguarding IoT gas monitoring networks against both physical sensor faults and cyber-attacks such as false data injection [5], and a neural-network-based smart-toilet gas sensing module capable of non-invasively quantifying trace sulphur and hydrogen concentrations in fecal odour for continuous, wearable-free health telemetry [14]. Together, these studies illustrate the expanding reach of hazardous-gas machine learning beyond conventional industrial safety into cybersecurity and personalized healthcare diagnostics.

TABLE II COMPREHENSIVE SUMMARY OF THE THIRTY REVIEWED STUDIES ON MACHINE-LEARNING-BASED HAZARDOUS GAS CONCENTRATION ANALYSIS AND PREDICTION, INCLUDING THE PRINCIPAL ML/DL MODEL(S), BEST REPORTED PERFORMANCE, AND KEY CONTRIBUTION OF EACH STUDY. REFERENCE NUMBERS CORRESPOND TO THE NUMBERED REFERENCE LIST.

Ref.	Research Focus / Application	ML / DL Model(s) Used	Best Reported Performance	Key Contribution
1	To explore the integration of low-cost SCADA/HMI with TinyML for real-time monitoring and short-term prediction of indoor CO_2…	TinyML, Neural Networks (MLP), Polynomial Regression	RMSE: 0.5 - 7 ppm; R^2 score: 0.80 - 0.99	Demonstrated a complete low-cost SCADA/HMI and TinyML pipeline for real-time indoor CO_2 forecasting.
2	To develop an IoT-driven, low-cost sensor node system calibrated via machine learning to accurately monitor and predict…	Random Forest (RF), Artificial Neural Network (ANN / MLP), Linear Regression, Decision Trees	RF Precision: 82.84%; ANN-MLP Precision: 76.05%	Created an integrated IoT-ML atmospheric monitoring node with field-validated low-cost sensor calibration.

Ref.	Research Focus / Application	ML / DL Model(s) Used	Best Reported Performance	Key Contribution
3	To build a collaborative IoT-based industrial gas leak detection and warning framework with predictive machine learning…	Linear Regression, Random Forest, Decision Tree	High trend prediction accuracy (exact numeric metric not specified)	Designed a reliable IoT-GSM collaborative warning system integrated with predictive analytics for industrial gas safety.
4	To adapt the Recency, Frequency, Monetary (RFM) framework alongside clustering techniques to categorize and evaluate urban air…	RFM Modeling, K-Means Clustering, Fuzzy C-Means, RFM Analysis	Optimal cluster separation achieved with K-Means (k=3 to 5)	Pioneered the adaptation of RFM analytics for urban air pollution spatial comparison and cluster segmentation.
5	To propose a dual-layer cybersecurity and fault detection framework combining rule engines and edge unsupervised AI for smart…	Isolation Forest (iForest), Rule-Based Engine	AI Anomaly Detection Accuracy: ~98.5% (outperforming static rules)	Established a cyber-physically resilient IoT gas monitoring architecture combining rule engines and edge Isolation Forest.
6	To construct an integrated real-time IoT solution for gas leak detection, air quality monitoring, and automated safety actuation.	Rule-Based Threshold Control Logic, Descriptive Analytics	100% actuation reliability under laboratory test conditions	Implemented an active closed-loop IoT safety monitoring and automatic ventilation control system.
7	To build a digital decision support platform (RuraLTHINGS) incorporating deep learning to monitor and predict indoor hazardous…	Stacked Bi-LSTM (Bidirectional + Unidirectional LSTM), GRU, Transformer, ARIMA	Stacked Bi-LSTM RMSE: 0.018, R^2 score: >0.94	Developed an end-to-end decision support platform combining rural IoT infrastructure with stacked BiLSTM deep forecasting.
8	To achieve highly accurate selective classification of CO and NO_2 gas mixtures using dynamic thermal modulation on a single MOS…	Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), Gradient Boosting	Classification Accuracy: 98.3% (with SVM)	Proved that dynamic PWM thermal fingerprints enable high-accuracy gas mixture classification on a single MOS sensor.

Ref.	Research Focus / Application	ML / DL Model(s) Used	Best Reported Performance	Key Contribution
9	To establish a highly accurate and computationally efficient time-series prediction model for underground coal mine gas…	Gated Recurrent Unit (GRU), LSTM, Recurrent Neural Network (RNN), BPNN, SVR	Prediction Error Reduction vs LSTM: 7.9%; Execution Speed Improvement: +13.05%	Validated GRU as a superior, low-latency deep learning alternative to standard LSTM for coal mine gas prediction.
10	To design an eNose system based on an array of low-cost MQ gas sensors combined with ML to classify pure and adulterated…	K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Artificial Neural Network (ANN),…	Classification Accuracy: 100% (KNN with 2 averaged features or 15 raw features)	Demonstrated a low-cost 7-MQ sensor eNose capable of 100% accurate essential oil classification and adulteration identification.
11	To systematically investigate how sensor drift and flawed experimental data collection order in public MOx datasets cause…	Deep Neural Networks (DNN), CNN, t-SNE, Zero-Offset Subtraction, OOD Classifier	Uncorrected Accuracy: ~98-99% (shown to be artifically inflated by drift artifacts)	Exposed fundamental temporal-spatial drift bias in benchmark eNose datasets and established baseline correction guidelines.
12	To review and synthesize advancements in Interdigital Electrode (IDE) smart gas sensors, microfabrication techniques, and…	PCA, KNN, SVM, ANN, Random Forest, Naïve Bayes, Deep Learning (CNN/LSTM)	ML-assisted IDE sensors routinely achieve >95-99% gas identification accuracy across…	Delivered an authoritative review linking IDE sensor hardware configurations directly with tailored ML pattern recognition…
13	To design an edge-assisted gas-sensing network (Ea-GNet) using low-cost MOS sensors and ML models deployed at edge nodes for…	Random Forest, Support Vector Regression (SVR), Edge ML, tsfresh Feature Extractor	Gas Classification Accuracy: 98%; Concentration Regression R^2 Score: 0.84; Throughput…	Pioneered edge-intelligence architecture (Ea-GNet) achieving 98% gas identification and sub-100ms network latency.

Ref.	Research Focus / Application	ML / DL Model(s) Used	Best Reported Performance	Key Contribution
14	To implement a non-invasive smart toilet healthcare monitoring system that identifies fecal odor gas components and predicts…	Multi-Layer Perceptron Neural Networks (MLP-NN), Deep Neural Networks	Gas Concentration Estimation Accuracy: High correlation (R^2 > 0.90, low MAE)	Developed a novel smart toilet gas-sensing module powered by neural networks for non-invasive health diagnostics.
15	To evaluate 14 deep learning neural network architectures for long-term hourly campus electricity demand forecasting using…	CNN-BiLSTM, CNN-BiGRU, FNN, RNN, LSTM, GRU, CNN, Hybrid CNN-RNN models	CNN-BiLSTM Test Accuracy: 0.92 (Train: 0.98); CNN-BiGRU Test Accuracy: 0.93 (Val Loss:	Conducted a benchmark study proving hybrid CNN-BiGRU/BiLSTM superiority for long-term institutional energy demand forecasting.
16	To develop a selective ensemble learning model based on the Yager fuzzy norm to improve coal mine gas concentration forecasting…	Selective Ensemble Learning, Yager Norm, Forward Sequential Selection, Autoregressive (AR),	Lowest overall prediction error (MSE/MAPE significantly lower than single BPNN or SVR)	Formulated a Yager norm-based selective ensemble methodology for robust coal mine gas forecasting.
17	To fabricate nano-hematite molecularly imprinted gas sensors and evaluate XGBoost, Fuzzy Logic, and Hybrid models for predicting…	XGBoost, Fuzzy Logic, Hybrid Fuzzy-XGBoost Model	R^2 Score: ~0.97 - 0.99 (Hybrid Model outperforming standalone XGBoost and Fuzzy Logic)	Integrated custom molecularly imprinted nano-hematite sensors with a novel hybrid Fuzzy-XGBoost ML model for precise VOC…
18	To develop an EMA-Autoformer deep learning algorithm for predicting dissolved oil chromatography gas concentrations in electric…	EMA-Autoformer, Autoformer, Transformer, LSTM, Informer	Prediction Accuracy Improvement: >15% error reduction over standard Transformer	Introduced the EMA-Autoformer model to solve missing data and long-term trend forecasting in transformer dissolved gas analysis.

Ref.	Research Focus / Application	ML / DL Model(s) Used	Best Reported Performance	Key Contribution
19	To design a compact THz metamaterial absorber (MMA) and use ML regressors to bypass computationally expensive full-wave…	K-Nearest Neighbors (KNN) Regressor, Extra Trees Regressor (ERT), Random Forest Regressor	Absorptivity: 99.96%; Sensitivity: 905 GHz/RIU; KNN Adjusted R^2: 0.9993; RMSE: 0.4576;…	Created an ultra-sensitive THz gas sensor design framework accelerated 8.5x by KNN surrogate machine learning.
20	To construct an Adamax-optimized Bidirectional Gated Recurrent Unit (Adamax-BiGRU) model for high-precision mine gas…	Adamax-BiGRU, BiGRU, GRU, LSTM, RNN, Adam, SGD	Test Error Reduction: 25.58% vs RNN, 12.53% vs LSTM, 3.01% vs GRU; Adamax achieved…	Established Adamax-BiGRU as a benchmark deep learning model for coal mine gas concentration forecasting.
21	To evaluate machine learning and deep learning models for predicting stack gas emissions (CO, SO_2, NO_x, O_2, dust) from a…	Long Short-Term Memory (LSTM), Light Gradient Boosted Machine (LightGBM), Multi-Layer…	LSTM R^2: 0.87 (NO_x), 0.85 (SO_2), 0.67 (CO), 0.78 (Dust); LightGBM R^2: 0.85 (O_2)	Provided a comprehensive benchmark of CEMS stack gas emission forecasting using real-world coal power plant data.
22	To develop an NRBO-VMD-AM-BiLSTM hybrid soft sensor framework for highly accurate dissolved gas concentration prediction in…	NRBO-VMD-AM-BiLSTM, Variational Mode Decomposition (VMD), Newton-Raphson-Based Optimization…	Hydrogen (H_2) Prediction RMSE: 0.51 μL/L; MAPE: 1.27%	Pioneered an NRBO-optimized VMD-AM-BiLSTM hybrid soft sensor architecture setting state-of-the-art accuracy in transformer DGA…
23	To review recent material developments in chemiresistive/FET gas sensors and analyze the transformative role of ML in solving…	PCA, LDA, KNN, SVM, Decision Trees (CART), Naïve Bayes, Random Forest, MLP/ANN, Deep Learning	ML classifiers routinely achieve 95% - 100% accuracy across evaluated eNose and FET…	Synthesized the state-of-the-art in chemical gas sensing hardware and established a structured framework for ML-driven gas…

Ref.	Research Focus / Application	ML / DL Model(s) Used	Best Reported Performance	Key Contribution
24	To build a big-data real-time streaming prediction framework (SPARS) combining ARIMA and SVM on Apache Spark to forecast coal…	SPARS Model (Spark + ARIMA + SVM), ARIMA, Support Vector Machine (SVM)	Lowest RMSE and MAPE among evaluated models (exact numeric error minimized under optimal…	Formulated the SPARS big-data streaming framework for continuous real-time coal mine gas forecasting.
25	To develop a Pearson-LSTM feature selection and time-series prediction framework to forecast gas concentrations in coal mine…	Pearson-LSTM, Long Short-Term Memory (LSTM), Bi-RNN, Gated Recurrent Unit (GRU)	2407 Working Face MSE: 0.015 (error range 0.005-0.04); 2409 Working Face (15-min advance…	Validated a Pearson correlation-guided LSTM model achieving 15-minute advance forecasting for coal mine safety.
26	To build a multi-gas monitoring hardware device and a Feature-Aware LSTM (FA-LSTM) model for predicting multi-component toxic…	Feature-Aware LSTM (FA-LSTM), LSTM, ARIMA, ETS (Exponential Smoothing), BP Neural Network	FA-LSTM achieved lowest RMSE/MAE and fastest training convergence across all 6 toxic…	Developed custom multi-gas sensing hardware and pioneered the Feature-Aware LSTM model for complex toxic gas prediction.
27	To develop a hybrid SCBA model (SSA-optimized CNN-BiLSTM-Attention) for multi-step spatio-temporal coal mine gas concentration…	SCBA Model (SSA + CNN + BiLSTM + Attention), Attention-LSTM, SSA-LSTM-Attention,	RMSE: 0.0171; MAPE: 0.084 (RMSE improved by 23.3% vs Attention-LSTM and 30.2% vs…	Established state-of-the-art accuracy in multi-step coal mine gas forecasting using the SSA-optimized CNN-BiLSTM-Attention…
28	To predict outlet oil and gas concentrations from vapor recovery units in oil depots using multi-parameter process inputs and…	Random Forest (RF), Support Vector Machine (SVM), XGBoost, Decision Trees, Multi-Layer…	Regression R^2: ~0.90 (Accuracy Error ~30%); Classification Accuracy: >90% (High…	Demonstrated the practical utility of Random Forest models for monitoring and controlling industrial oil and gas vapor recovery…

Ref.	Research Focus / Application	ML / DL Model(s) Used	Best Reported Performance	Key Contribution
29	To design an automated parameter-tuned Stacking ensemble regression algorithm for precise concentration prediction of mixed gases.	Stacking Ensemble, Random Forest (RF), Extreme Random Forest (Extra Trees), Gradient Boosting…	Stacking Model achieved lowest RMSE/MAE and highest R^2 among all evaluated ensemble…	Proved that multi-tree Stacking ensemble learning with Grid Search optimization provides superior accuracy for mixed gas…
30	To systematically investigate how sensor selectivity factors S(X) and diverse MOx sensor array combinations impact ML gas…	Support Vector Machine (SVM), Selectivity Factor S(X) Formulation	Gas Recognition / Prediction Accuracy: >95-98% when complementary high-selectivity…	Discovered the mathematical proportional link between sensor selectivity S(X) and ML accuracy, optimizing sensor array…

VI. COMPARATIVE PERFORMANCE ANALYSIS

Direct numerical comparison across the reviewed studies is inherently constrained by the heterogeneity of datasets, gas species, sensor hardware, and reported metrics; nonetheless, several consistent trends emerge. First, for multi-step temporal forecasting tasks, bidirectional and attention-augmented recurrent architectures consistently outperform their unidirectional or non-augmented counterparts, with reported error reductions ranging from 3% to over 30% depending on the baseline used for comparison [9,20,22,27]. Second, for gas type classification and mixture identification tasks, both classical algorithms (SVM, KNN) and tree-based ensembles are capable of achieving accuracies in the 95-100% range under controlled laboratory conditions, indicating that model architecture is often less critical than feature extraction quality and sensor selectivity for this task category [8,10,30]. Third, edge-optimized architectures, while occasionally trading a small amount of peak accuracy for latency, achieve dramatic throughput improvements (of the order of 600% or more) that are essential for real-time safety-critical deployment [1,13]. Fig. 5 presents a representative, illustrative comparison of the best-reported accuracy or R^2 values for a selection of high-performing models spanning classification and edge-deployment tasks across the reviewed literature.

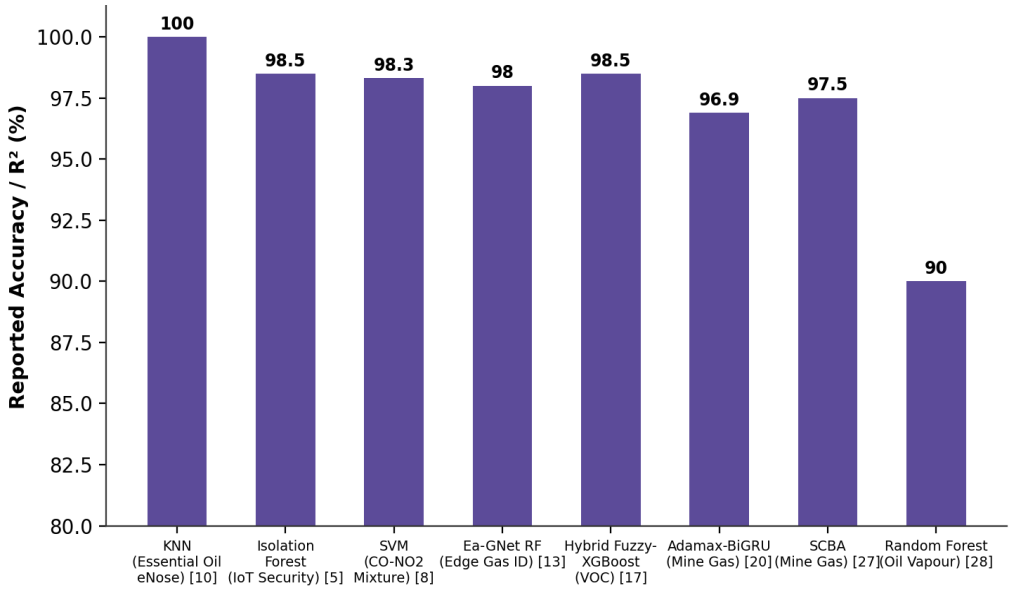


Fig. 5. Representative best-reported accuracy or R^2 performance of selected high-performing machine learning models across the reviewed studies. Values are drawn directly from the original publications and are not directly comparable across differing datasets, gas species, and experimental conditions; they are presented to illustrate the general performance range achievable with current methodologies.

It is important to emphasize the methodological caution raised by Chang et al. [10], namely that reported classification accuracies on public benchmark datasets can be substantially inflated by uncorrected sensor drift and non-randomized data-collection order. This finding underscores that headline accuracy figures reported across the reviewed literature, while broadly indicative of relative model capability, should be interpreted with appropriate caution, and that future comparative benchmarking should adopt standardized, drift-corrected, and randomized-exposure evaluation protocols.

VII. CHALLENGES AND LIMITATIONS

Although significant progress has been made in this field, there appears to be an unvarying list of challenges which constrain the practicality and generalizability of machine-learning-enabled hazardous gas detection systems. These have been presented in Table III.

TABLE III PRINCIPAL CHALLENGES AND LIMITATIONS IDENTIFIED ACROSS THE REVIEWED LITERATURE ON MACHINE-LEARNING-BASED HAZARDOUS GAS CONCENTRATION ANALYSIS AND PREDICTION.

Challenge Category	Description	Representative Evidence
Small dataset size and heterogeneity	Most research work uses datasets that are of small duration (ranging from weeks to months) or controlled environment datasets.	[11,23]

Cross-sensitivity in mixed-gas environments	Overlapping gas signatures from structurally similar species remain difficult to fully decouple, especially with single-channel sensors under humidity variation.	[8,12,17,30]
Limited dataset size and diversity	Many studies rely on short-duration (weeks to a few months) or laboratory-controlled datasets, limiting generalization to long-term, multi-seasonal, real-world conditions.	[7,17,19]
Computational complexity versus edge deployment	Hybrid decomposition/attention/metaheuristic-optimized architectures achieve state-of-the-art accuracy but incur high computational and memory overhead unsuitable for ultra-low-power edge nodes.	[18,22,27]
Lack of adaptive online learning	Most deployed models are trained offline and lack mechanisms for continuous drift adaptation or federated retraining across distributed sensor networks.	[5,13,29]
Simulation-to-reality gap	Simulation-based sensor design and validation (e.g., electromagnetic metamaterial sensors) require experimental confirmation under real gas-flow and thermal conditions.	[19]

In addition to the challenges summarized in Table III, several studies note domain-specific limitations: coal mine forecasting models are typically validated on a single mine site and do not yet incorporate spatial gas-flow topology across multiple tunnel faces [9,20,25]; industrial CEMS-based models exhibit reduced accuracy for highly volatile species such as CO relative to more stable species such as NO_x and SO_2 [21]; and eNose systems generally require periodic recalibration and manual feature engineering, limiting fully autonomous long-term operation [10,14].

VIII. FUTURE RESEARCH DIRECTIONS

Building on the identified challenges, several promising research directions emerge from the reviewed literature. To start with, the creation of standard datasets that have been drift-corrupted and exposed in a randomized manner, for a variety of gas species, sensors, and environments, will definitely increase the validity and comparison of future research on ML-based gas sensing [11,12,23]. Next, federated and adaptive learning approaches, in which edge nodes collaborate to update model parameters in order to counteract drift without transferring raw data to a central server, are a logical follow-up to current edge-AI and security-aware solutions [5,13]. Third, hybrid architectures that combine signal decomposition, attention mechanisms, and metaheuristic optimization have demonstrated clear accuracy gains and should be further explored in conjunction with model compression techniques (pruning, quantization, knowledge distillation) to reconcile state-of-the-art accuracy with the strict memory and power constraints of field-deployable hardware [18,22,27]. Fourth, extending single-gas or binary-mixture models toward true multi-

component, spatio-temporal sensor-network architectures, potentially leveraging graph neural networks to capture spatial gas-flow dependencies across distributed sensor nodes, would address a recurring limitation noted across the coal mine and industrial monitoring clusters [9,20,27]. Finally, the integration of explainable artificial intelligence (XAI) techniques would enhance operator trust and regulatory acceptance of ML-based hazard alerts, particularly in safety-critical domains where opaque model decisions currently limit adoption relative to simpler, more interpretable threshold-based systems[4].

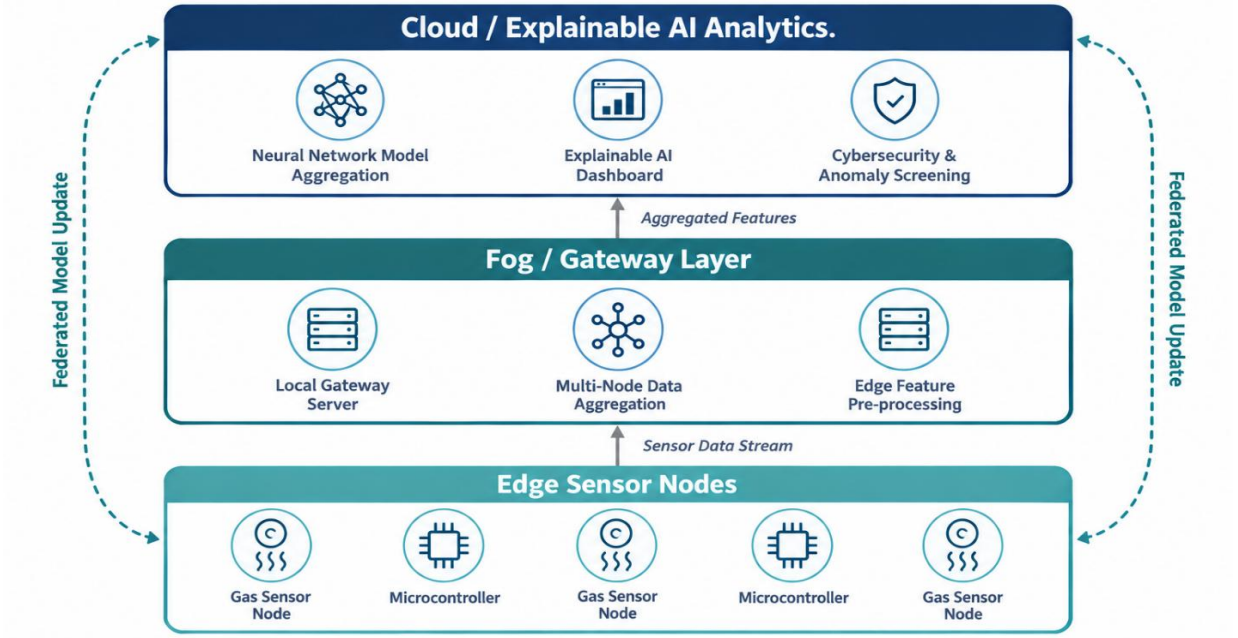


Fig. 6. Conceptual schematic of a future federated edge-fog-cloud architecture for adaptive, drift-resilient hazardous gas monitoring, integrating on-device TinyML inference, periodic federated model updates, and cloud-based explainable analytics.

IX. CONCLUSION

This review has systematically surveyed thirty recent studies (2020-2026) at the intersection of hazardous gas sensing and machine learning, spanning IoT-based ambient monitoring, underground coal mine safety, electronic nose gas mixture identification, power transformer dissolved gas analysis, industrial emission control, and cyber-physical anomaly detection. The evidence collectively demonstrates that machine learning, and particularly bidirectional and attention-augmented deep sequence models, can substantially improve both the selectivity and the temporal foresight of hazardous gas monitoring systems relative to conventional threshold-based approaches, while ensemble tree-based and lightweight neural models remain highly competitive for low-latency, resource-constrained edge deployment. At the same time, the reviewed literature exposes persistent challenges related to sensor drift, dataset scarcity, cross-sensitivity in mixed-gas environments, and the gap between algorithmic sophistication and field-deployable

computational efficiency. Addressing these challenges through standardized benchmarking, federated and online-adaptive learning, hybrid decomposition-attention architectures, spatio-temporal sensor-network modelling, and explainable artificial intelligence represents a coherent and achievable research agenda capable of advancing hazardous gas monitoring from reactive, threshold-based alerting toward proactive, intelligent, and trustworthy safety infrastructure.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this review.

REFERENCES

[1] I. N. K. Wardana, S. A. Fahmy, and J. W. Gardner, "Low-cost SCADA/HMI with tiny machine learning for monitoring indoor CO_2 concentration," in Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC), 2024, doi: 10.1109/I2MTC60896.2024.10561089.

[2] M. Bindhu, S. Parasuraman, S. Yogeeswaran, and S. N. Devi, "Harnessing machine learning for IoT-driven atmospheric parameter monitoring and predictive analytics," in Proc. 9th Int. Conf. Smart Struct. Syst. (ICSSS), 2023, doi: 10.1109/ICSSS58085.2023.10407051.

[3] B. R. Sathishkumar, S. Dhviya M, and S. Allirani, "IoT based collaborative industrial gas prediction and warning using machine learning," in Proc. Int. Conf. Electron. Syst. Intell. Comput. (ICESIC), 2022, doi: 10.1109/ICESIC53714.2022.9783588.

[4] J. Mohammad and M. A. Kashem, "Air pollution comparison RFM model using machine learning approach," in Proc. IEEE 7th Int. Conf. Converg. Technol. (I2CT), 2022, doi: 10.1109/I2CT54291.2022.9824248.

[5] O. Gaddour, S. B. Moallem, and M. Baklouti, "Hybrid rule-based and AI-driven anomaly detection for secure IoT gas monitoring in smart buildings," in Proc. IEEE Int. Conf. Complex Syst. Intell. Comput. (ICCSIC), 2026, doi: 10.1109/ICCSIC69301.2026.11521690.

[6] A. Ali, S. A. A. Kazmi, and M. Z. Khan, "Real-time gas and air quality monitoring and controlling by IoT solution," in Proc. 19th Int. Conf. Emerg. Technol. (ICET), 2024.

[7] E. F. Ladeira and B. M. C. Silva, "A machine learning-based platform for monitoring and prediction of hazardous gases in rural and remote areas," IEEE Access, 2025, doi: 10.1109/ACCESS.2025.3535158.

[8] D. Van Dung, N. T. N. Hoa, N. N. Viet, T. D. Giang, N. V. Chien, and N. Van Hieu, "Machine-learning-based classification of CO-NO_2 gas mixtures using dynamic thermal signatures from a single MOS sensor channel," in Proc. IEEE 18th Int. Conf. Comput. Intell. Commun. Netw. (CICN), 2026, doi: 10.1109/CICN70047.2026.11594346.

[9] P. Jia, H. Liu, S. Wang, and P. Wang, "Research on a mine gas concentration forecasting model based on a GRU network," IEEE Access, 2020, doi: 10.1109/ACCESS.2020.2975257.

[10] S. Viciano-Tudela, L. Parra, P. Navarro-Garcia, S. Sendra, and J. Lloret, "Proposal of a new system for essential oil classification based on low-cost gas sensor and machine learning techniques," Sensors, 2023, doi: 10.3390/s23135812.

[11] I.-S. Chang, S.-W. Byun, T.-B. Lim, and G.-M. Park, "A study of drift effect in a popular metal oxide sensor and gas recognition using public gas datasets," IEEE Access, 2023, doi: 10.1109/ACCESS.2023.3257414.

[12] A. Nasri, A. Boujnah, A. Boubaker, and A. Kalboussi, "A smart gas sensor using machine learning algorithms: Sensor types based on IED configurations, fabrication techniques, algorithmic approaches, challenges, progress, and limitations: A review," IEEE Sensors J., 2023, doi: 10.1109/JSEN.2023.3268342.

[13] S. Nayak, S. Misra, and S. B. Majumder, "Machine learning-based edge intelligence for reliable gas monitoring networks," IEEE Sensors J., 2025, doi: 10.1109/JSEN.2025.3595435.

[14] S. Teranishi, H. Koike, D. Ueyama, and S. Abe, "High-accuracy fecal odor gas detection using a gas sensor device and machine learning," in Proc. IEEE Int. Conf. E-health Netw., Appl. Serv. (HealthCom), 2024, doi: 10.1109/HEALTHCOM60970.2024.10880690.

[15] S. A. Alsamraee and S. Khanna, "Long-term electricity demand forecasting of a university campus based on advanced deep learning artificial neural network algorithms," Results Eng., 2025, doi: 10.1016/j.rineng.2025.106683.

[16] R. Liang and P. Jia, "Coal mine gas concentration prediction model based on selective ensemble learning," J. Xi'an Univ. Sci. Technol. / Conf. Proc., 2020.

[17] R. M. Abdelghani, W. El-Shafie, I. Morsi, T. El-Sayed, and A.-E. B. Kashyout, "Comparative analysis of XGBoost, fuzzy logic, and hybrid models for acetone, ethanol, and methanol gas concentration prediction," in Proc. Int. Telecommun. Conf. (ITC-Egypt), 2025, doi: 10.1109/ITC-EGYPT66095.2025.11186668.

[18] X. Wang, F. Niel, X. Zhang, W. Wang, and H. Lei, "Gas concentration prediction of power system based on EMA-Autoformer," in Proc. 3rd Int. Conf. Energy Power Eng., Control Eng. (EPECE), 2024, doi: 10.1109/EPECE63428.2024.00033.

[19] S. Bhuiya and S. I. Ankur, "Machine learning-enhanced polarization insensitive metamaterial absorber for hazardous gas detection," in Proc. IEEE Int. Conf. Telecommun. Photon. (ICTP), 2025, doi: 10.1109/ICTP68765.2025.11415096.

[20] R. Liang, X. Chang, P. Jia, and C. Xu, "Mine gas concentration forecasting model based on an optimized BiGRU network," ACS Omega, 2020, doi: 10.1021/acsomega.0c03417.

[21] M. Mohammadi, D. Saloglu, H. Dertli, M. Mohammadi, and M. Ghaffari-Moghaddam, "Data-driven modeling for the prediction of stack gas concentration in a coal-fired power plant in Türkiye," Water, Air, Soil Pollut., 2024, doi: 10.1007/s11270-024-07107-3.

[22] N. Wang, W. Li, and X. Li, "Optimized NRBO-VMD-AM-BiLSTM hybrid architecture for enhanced dissolved gas concentration prediction in transformer oil soft sensors," Sensors, 2025, doi: 10.3390/s25165182.

[23] U. Yaqoob and M. I. Younis, "Chemical gas sensors: Recent developments, challenges, and the potential of machine learning—A review," Sensors, 2021, doi: 10.3390/s21082877.

[24] Y. Huang, J. Fan, Z. Yan, S. Li, and Y. Wang, "A gas concentration prediction method driven by a Spark Streaming framework," Energies, 2022, doi: 10.3390/en15155335.

[25] C. Liu, A. Zhang, J. Xue, C. Lei, and X. Zeng, "LSTM-Pearson gas concentration prediction model feature selection and its applications," Energies, 2023, doi: 10.3390/en16052318.

[26] Y. Cong, X. Zhao, K. Tang, G. Wang, Y. Hu, and Y. Jiao, "FA-LSTM: A novel toxic gas concentration prediction model in pollutant environment," IEEE Access, 2022, doi: 10.1109/ACCESS.2021.3133497.

[27] W. Xing, Y. Yang, Y. Zhang, and Y. Yang, "Gas concentration prediction based on SSA algorithm with CNN-BiLSTM-attention," Sci. Rep., 2025, doi: 10.1038/s41598-025-15838-4.

[28] X. Wang, B. Zhu, H. Zheng, J. Wang, Z. Chen, and B. Hong, "Research on export oil and gas concentration prediction based on machine learning methods," Energies, 2025, doi: 10.3390/en18051078.

[29] Y. Xu, R. Meng, and X. Zhao, "Research on a gas concentration prediction algorithm based on stacking," Sensors, 2021, doi: 10.3390/s21051597.

[30] G. Bae, M. Kim, W. Song, S. Myung, S. S. Lee, and K.-S. An, "Impact of a diverse combination of metal oxide gas sensors on machine learning-based gas recognition in mixed gases," ACS Omega.