Design of a hybrid artificial intelligence system for real-time quantification of impurities in gas streams: application in CO2 capture and storage.

AMINAHO, E.N., AMINAHO, N.S., HOSSAIN, M., FAISAL, N.H. and AMINAHO, K.A.

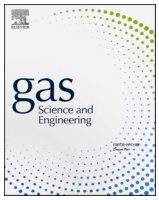
2025

134 (2025) 205546

Contents lists available at ScienceDirect

Gas Science and Engineering

journal homepage: www.journals.elsevier.com/gas-science-and-engineering



Design of a hybrid artificial intelligence system for real-time quantification of impurities in gas streams: Application in CO_2 capture and storage

Efenwengbe Nicholas Aminaho [a,b,*], Ndukaegho Sabastine Aminaho [b,c], Mamdud Hossain [a], Nadimul Haque Faisal [a], Konyengwaehie Augustus Aminaho [b,d]

[a] Robert Gordon University, School of Engineering, Aberdeen, United Kingdom
[b] A-Class Academic Consults Limited, Energy Research Group, Port Harcourt, Rivers State, Nigeria
[c] Efortlex Limited, Port Harcourt, Rivers State, Nigeria
[d] Konyens Enterprise, Port Harcourt, Rivers State, Nigeria

ARTICLE INFO

Keywords:
Hybrid
Artificial intelligence
Machine learning
Multi/binary components gas streams
Carbon dioxide
Cross-sensitivity

ABSTRACT

The concentration of gases in gas streams can be monitored using sensors. However, gas sensors can lose their response accuracy due to mechanical wear or damage, and environmental factors such as exposure to unusual temperature and pressure conditions. Therefore, it is paramount to design a hybrid artificial intelligence system to identify any sensor malfunction and the need to recalibrate or replace a sensor or a suite of sensors for real-time quantification of compositions of gas streams. Hence, this study proposes a hybrid artificial intelligence system for real-time monitoring of gas concentrations in a gas stream and recalibration of gas sensors. This system provides remote access for monitoring gas concentrations predicted by a machine learning model and sensor readings programmed in a wireless device or an application in a wireless device, enabling users to identify when certain set thresholds of gas concentrations are exceeded and to identify malfunction of sensors when predetermined deviations between the sensor readings and the machine learning predictions are exceeded, for quality control and assurance. The design also signals the need for recalibration or replacement of sensors, for more accurate readings. Therefore, this study developed a methodology for the design of the hybrid system and demonstrated the feasibility of operating the system in a nitrogen gas stream. Application of the system for carbon dioxide capture and storage was also explored. Machine learning models were developed for binary and multi-component gas mixtures using Python programming language. The findings of the study revealed that the error in quantification of gas concentrations for the binary gas mixture is less than the errors for the multi-component gas mixtures, using machine learning models. Therefore, while operating the hybrid artificial intelligence system for real-time quantification of impurities in gas streams, higher deviations in gas concentration between the sensors' readings and the machine learning model predictions should be allowable for the multi-component gas mixture compared to the binary gas mixture, as long as their set level of tolerance for the gas mixture is not exceeded.

1. Introduction

Increase in gas emissions into the atmosphere, compared to preindustrial levels, have resulted in tightened emissions standards in different parts of the world (Shindell et al., 2011). Carbon dioxide (CO_2) emission into the atmosphere is a major concern and is seen as one of the main greenhouse gases causing increased global warming and climate change (Yoro and Daramola, 2020). Therefore, meeting the world's net-zero carbon target is paramount, to achieve a carbon-neutral environment. To meet the net-zero target, it is important to develop effective carbon dioxide removal (CDR) methods to capture CO_2 from the atmosphere or industrial facilities (David and Dosunmu, 2011; Agunbiade et al., 2023; Saharudin et al., 2023; Bouaboula et al., 2024).

Carbon dioxide can be captured in a system where gas emissions from the atmosphere or industrial facilities (such as the steel or coal industry) are collected into a gas separation chamber designed to

* Corresponding author. Robert Gordon University, School of Engineering, Aberdeen, United Kingdom.
E-mail addresses: aminahonicholas@gmail.com, e.aminaho@rgu.ac.uk (E.N. Aminaho), n.s.aminaho@aclassconsults.com (N.S. Aminaho), m.hossain@rgu.ac.uk (M. Hossain), N.H.Faisal@rgu.ac.uk (N.H. Faisal), k.a.aminaho@aclassconsults.com (K.A. Aminaho).

https://doi.org/10.1016/j.jgsce.2025.205546
Received 9 June 2024; Received in revised form 9 November 2024; Accepted 9 January 2025
Available online 11 January 2025

capture CO_2 through a membrane, while other gas impurities (mixture) are ejected through another outlet in the separation chamber (Agunbiade et al., 2023). The ejected gas impurities can be re-injected into the separation chamber to capture most of the CO_2 gas left in the effluent gas mixture. This cycle can be repeated several times until nearly all the CO_2 in the mixture is captured. Furthermore, the captured CO_2 gas is transported through pipelines or ships for geological storage (Metz et al., 2005). Carbon dioxide can be stored in aquifers, depleted petroleum reservoirs, and salt caverns (Zhang et al., 2011; Aminaho et al., 2024). The storage of CO_2 in geological formations is possible with a good lateral seal (aquitard or cap rock) above the reservoir layer, to prevent the gas from migrating to the earth's surface during geo-sequestration (Aminaho et al., 2024).

Carbon dioxide may be compressed with trace amounts of other acid gases (such as H_2S, SO_2, or NO_2) and transported for underground storage in reservoirs (Zhang et al., 2011). Hence, the gases react with the formation water (or brine) and rock minerals, thereby altering the reservoir properties. Different studies have shown that CO_2 co-injection with SO_2 or NO_2 decreases the porosity and permeability of the reservoir (Bolourinejad and Herber, 2014; Aminu et al., 2018; Pearce et al., 2019). Moreover, Aminaho and Hossain (2023) have found that co-injection of CO_2 with up to 2.5 mol percent of SO_2 significantly alters the brittleness of some reservoir rocks, making the rocks more ductile, thereby limiting their potential to fracture and create more flow path for fluid flow in the reservoir. Therefore, conscious efforts must be taken to control the amount of acid gases in a CO_2 gas stream transported for storage. Also, in the separation chamber for CO_2 capture, real-time measurements of the concentration of the gas components in the mixture should be taken. Fig. 1 illustrates the process of CO_2 capture and storage. After CO_2 is captured, it can be injected into geological formations (saline aquifers, depleted oil or gas reservoirs, and coal seams), the deep ocean, or minerals, as well as terrestrial uptake by plants and soil (Nagireddi et al., 2024).

One major method of real-time monitoring of gas concentrations is using sensors. A sensor is a device that converts a physical measure into a signal that can be read by an observer or instrument (Chen et al., 2012). Sensors are devices that measure certain properties of materials by their response to those materials. Sensors can provide accurate measurements by proper calibration (Vilanova et al., 2003; Schultealbert et al., 2017; Wang et al., 2022). Several studies have found that sensors designed to measure gas concentrations in certain gas environments might not be effective in another gas environment. Walsh et al. (2011) found that electrochemical sensors predicted oxygen concentration wrongly when helium gas was present in the mixture.

Very few studies have considered the measurement of gas concentrations in a CO_2 system using sensors (Vilanova et al., 2003; Stankova et al., 2004; Shwetha et al., 2022). Most of the studies are based on a binary gas mixture of CO_2. Moreover, no study has considered recalibration of sensors using machine learning techniques. Calibration or recalibration of sensors is paramount as sensors can lose their response accuracy due to mechanical wear or damage and environmental factors such as exposure to unusual temperature and pressure conditions (Vilanova et al., 2003; Schultealbert et al., 2017; Wang et al., 2022). A hybrid artificial intelligence system will help to identify any sensor malfunction and the need to recalibrate or replace a sensor or a suite of sensors.

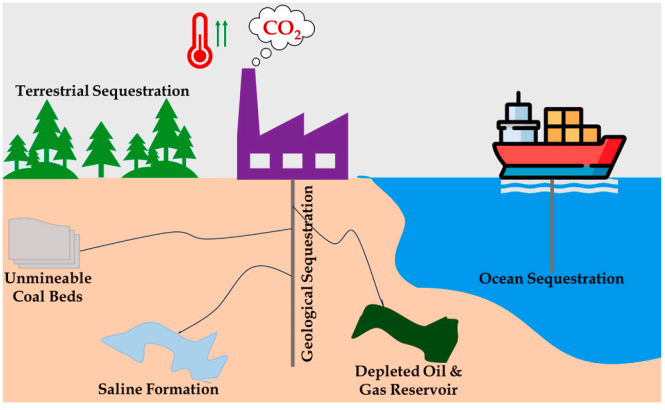


Fig. 1. Schematic representation of CO_2 capture and storage (Nagireddi et al., 2024).

Several studies have applied machine learning data generated from sensor experiments for different gas concentrations (Antanasijević et al., 2018; Casey et al., 2019; Barriault et al., 2021; Djedidi et al., 2021; Martinez et al., 2021; Javed et al., 2022; Wang et al., 2022). However, no study has been conducted on machine-learning-assisted design or recalibration of sensors. Therefore, this study proposes a hybrid artificial intelligence system and evaluates sensors' response in binary and multi-component gas mixtures, and how this hybrid technology can be applied in carbon capture and storage. This study further explores how impurities in a CO_2 gas stream can be quantified in real-time during carbon capture and storage (CCS), and the process of sensors' recalibration based on the hybrid artificial intelligence system.

2. Theoretical framework

Sensors are used in the measurement of physical quantities. A sensor responds to a material based on its change in physical quantities such as voltage, resistance, or conductance. For instance, an analog temperature sensor measures the change in the temperature of a body or material that corresponds to a change in the sensor's physical quantity such as resistance or voltage. Similarly, microheaters (as temperature sensors) respond to changes in their resistance to measure the temperature of a material (Shwetha et al., 2022) as expressed in Equation (1).

$$R_T = R_0[1 + \alpha(T - T_0)] \tag{1}$$

where R_0 represents resistance calculated at room temperature (T_0), R_T represents resistance at temperature (T), and α is the microheater temperature coefficient of resistance. Thus, the sensor is calibrated based on the above equation.

Furthermore, gas components in a gas mixture can be identified or detected using gas sensors. The common gas sensors include non-dispersive infrared (NDIR) sensors, semiconductor sensors, and electrochemical sensors (Walsh et al., 2011; McGrath and Scanaill, 2013; Schultealbert et al., 2017; Tsui et al., 2017; Bauke et al., 2018; Barriault et al., 2021). NDIR gas sensors (a type of optical sensor) operate based on the fact that gas molecules absorb infrared rays at specific wavelengths. Thus, the gas concentration is calculated based on the difference in the amounts of infrared radiation reached (Bauke et al., 2018). In semiconductor sensors (solid-state sensors), oxygen is absorbed by a metal oxide and reacts with the gas to be identified (or detected); hence, the sensor resistance value changes in response to the reaction (Schultealbert et al., 2017; Barriault et al., 2021; Shwetha et al., 2022). Electrochemical gas sensors measure gas concentration through oxidation-reduction reactions. An electrochemical sensor is made up of a working or sensing electrode, a reference electrode, and in most cases a counter electrode. The design enables these electrodes to be placed in contact with either a solid or a liquid electrolyte. Electrochemical sensors convert information received during electrochemical reactions between the electrodes and the sample (analyte) of interest into useful electrical signals or electrical parameters (Walsh et al., 2011; Tsui et al., 2017; Javed et al., 2022; Noor et al., 2024).

Electrochemical sensors are less expensive and have a longer lifespan than optical sensors. However, they are less sensitive and respond slower (Basyooni et al., 2017). Also, the responses of electrochemical sensors are not stable. Their responses change over time, thereby resulting in the recalibration of the sensors (Dhall et al., 2021). Optical

sensors have high sensitivity and selectivity and are commonly used to detect CO_2 (Singh and Marvili, 2018). Solid-state or semi-conductor sensors have high sensitivity, fast response time, low detection limit, simple operational principle, and consume less power (Zaki et al., 2019; Dhall et al., 2021). Therefore, they are more suitable for portable and wireless applications (Zaki et al., 2019). Solid-state sensors made of nanomaterials have a high surface area-to-volume ratio, which results in their increased selectivity and sensitivity towards CO_2 (Zaki et al., 2019; Basyooni et al., 2024). However, semiconductor sensors have low selectivity in the presence of various gases (Lagutin and Vasil'ev, 2022).

Gas sensors are designed for certain gases and under specified environmental (relative humidity, temperature, and pressure) conditions (Barcelo-Ordinas et al., 2019). Moreover, the gas mixtures emitted in the steel industry might be different from gas mixtures emitted in the coal or petroleum industry. Interference of other gases in an environment can limit the accuracy of gas sensors. Therefore, the same array of sensors may not be effective for use in different industries. A sensor used in a particular industry might be suitable in another industry if the sensor is recalibrated by considering the components of the gas mixture in the industry or environment where it would be mounted. Furthermore, to enhance reliability and long service life, gas sensors must be designed to withstand and resist attacks from corrosive chemicals in the environment. (TE Connectivity, 2020). Gas sensors must be properly fabricated to preserve their sensing components from corrosive materials that could limit their accuracy and service life. In such corrosive environments, sensors should be properly calibrated to avoid wrong output signals.

Sensor calibration refers to a set of adjustments performed on the sensor to increase its accuracy and functionality (Vilanova et al., 2003; Schultealbert et al., 2017; Wang et al., 2022). Calibration or recalibration of sensors can be performed by experimental and numerical simulation methods. In large-scale networks, major challenges faced with sensor calibration are the numerous number of sensors to be calibrated and the physical challenge of accessing the sensors as they may be deployed in harsh or hostile environments (Barcelo-Ordinas et al., 2019). The lack of a prompting system to signal sensors' malfunction results in organisations following routine timelines for sensors' calibration or recalibration. The sensors are checked regularly against standard instruments to maintain the quality of their responses to gas or gases in the environment, and for periodic recalibration (Barcelo-Ordinas et al., 2019). This approach to identifying malfunctioned sensors is not cost-effective. Most of the expenses on the regular checks on the numerous sensors are unnecessary, as in some cases the sensors might have been functioning properly when inspected. Therefore, this study proposes a new method of identifying sensors' malfunction and calibrating or recalibrating sensors based on hybrid artificial intelligence systems for real-time quantification of impurities in gas streams. Once sensors malfunction in any environment, they may be calibrated or recalibrated using experimental or numerical simulation methods.

2.1. Experimental calibration method

Experiments can be performed to calibrate sensors. The response of an array of sensors to different gas mixtures can be measured in a data acquisition system during the experiment (Chen et al., 2012). For a particular amount of a base gas (for example, up to 10% of oxygen gas by volume if required, which is constant for each case in the experiment) and a balance gas (making up the remaining volume percent), a baseline measurement of the sensor is achieved by injecting the base gas and the balance gas in the gas mixing chamber (a fixed volume stainless steel manifold), mixed properly under computer control (Javed et al., 2022) and controlled to flow through the array of sensors. The gas mixture flows through each of the sensors (in a suite of sensors) and the sensors' measurement of physical, thermal, or electrical property for which it was designed is recorded in a data acquisition system such as a LABVIEW program running in a computer. This sensor measurement represents the baseline reading of the sensor.

To measure the response of the sensors to the base gas and balance gas in the presence of a gas impurity or gas impurities of known concentration(s), the gases are mixed (in a dry condition, to avoid having water vapour content in the mixtures which could result in a minor voltage shifts between dry and water-wet conditions) properly with the base gas and balance gas and allowed to flow at a specified rate through the array of sensors, while the sensors readings are recorded (Stankova et al., 2004; Javed et al., 2022). The data acquisition system records the sensors' response when the base gas (constant amount) and balance gas alone flow through the sensor and when an impurity or impurities together with the base gas and balance gas flow through the same array of sensors. In other words, the value of the physical quantity change recorded when only the base gas and balance gas flow through the sensor could be set to zero, while in the presence of another impurity or impurities, the property change recorded could be negative or positive, representing the difference in the property when a base gas and a balance gas are injected and when they are injected together with the gas impurities and allowed to flow through the sensors.

The experimental setup to record the response of sensors to different gas compositions and concentrations is shown in Fig. 2. Different gas cylinders are connected to a gas mixer (or a gas mixing chamber), where different gas compositions and concentrations are dry-mixed. For instance, the gas cylinders represent possible components from gas emissions in an industrial environment. In a carbon capture system, the gas of interest to capture is CO_2 (while a constant amount of base gas, say oxygen, is maintained in the system [if required in the carbon capture facility receiving the gas mixtures from an exhaust]) in the presence of the other gases (in this case referred to as pollutants or impurities). In this case, the balance gas is CO_2 with a constant amount of a base gas (say up to 10% oxygen, by volume, if required). So, the base gas (constant amount) and CO_2 are injected into the gas mixing chamber and allowed to flow into the chamber containing a suite of sensors at a specified flow rate. The sensors' response is recorded and set as the baseline measurement of the physical quantity (say, resistance, conductance, or voltage). Then, the CO_2 gas (with a base gas) is changed over to the base gas (say up to 10% of oxygen, by volume) with a given gas impurity (sharing the remaining volume percent with the balance gas, CO_2 injected) mixed in the mixing chamber and allowed to flow into the sensors testing chamber, while the data acquisition system records the sensors' response to the gas mixture (Stankova et al., 2004). The same procedure is followed for different gas components (multiple gases) and concentrations with the base gas, while the sensors' responses to the gas mixtures are recorded.

The gas mixture is allowed to flow through the suite of sensors at a specified rate set in a computer program or controlled from a flowmeter at the gas mixing chamber. The sensors are arranged in series such that the same gas mixture enters from one of the sensors, passes through the other, and exits from the last sensor in the array (Javed et al., 2022), enabling each sensor to be contacted by the gas mixture and its response to the gas mixture is recorded in the data acquisition system. Throughout the experiments, It is a good practice to check for any possible drift in the sensors' response by testing the sensors periodically with a particular amount or concentration of each gas impurity (Javed et al., 2022).

The sensors are carefully selected to minimize cross-sensitivity in their response to the different gas components (Martinez et al., 2021; Acharyya et al., 2022; Shwetha et al., 2022). For example, pure Tungsten (VI) oxide (WO_3) is very sensitive to H_2S in CO_2, and exhibits low sensitivity to SO_2 in CO_2; while Platinum (Pt) doped Tungsten (VI) oxide ($WO_3 + Pt$) is very sensitive to SO_2 in CO_2, and exhibits low sensitivity to H_2S in CO_2 (Stankova et al., 2004). Studies have found that sensors pairing Pt and $La_{0.8}Sr_{0.2}CrO_3$ (LSCO) have high sensitivity to hydrocarbons (such as C_3H_8 and C_3H_6) at open circuit, and nitrogen oxides (NO and NO_2) under bias; while sensors pairing Pt and gold (Au)

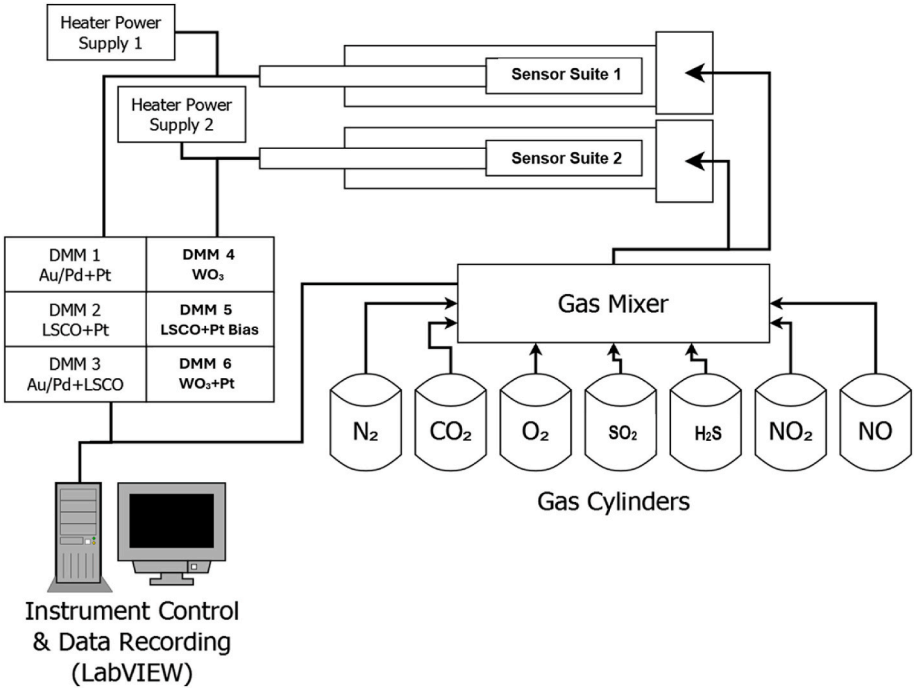


Fig. 2. Experimental setup to record the response of sensors to different gas compositions and concentrations (Modified from Tsui et al., 2017).

electrodes or Au alloys such as Au and Palladium (Pd) which can be represented as (Au/Pd + Pt) are sensitive to both carbon (II) oxide (CO) and ammonia (Tsui et al., 2017). However, cross-sensitivity and interference in sensor responses in the presence of multiple gas impurities make the selection of sensors critical in properly identifying gas concentrations in a mixture. For instance, a sensor made up of Au/Pd + Pt electrodes (with solid electrolyte, 3 mol% yttria-stabilized zirconia [YSZ]) has shown unique sensitivity towards ammonia (NH_3) with no interference from mixtures of C_3H_8, CO, NO, and NO_2. In other words, the sensor exhibits very high sensitivity to NH_3 in the presence of C_3H_8, CO, NO, and NO_2 with very low or no sensitivity to any of the other gas components in the mixture. However, when NH_3 was not in the mixture, the sensor showed some sensitivity to the other analyte gases in the mixture, which could result in false positives when using the sensor (Javed et al., 2022). Therefore, it is important to understand the gas components from a particular gas environment through gas chromatography and mass spectrometry (Vilanova et al., 2003) before selecting an array of sensors to test their responses to different gas concentrations in the environment. Based on the sensors' responses during the experiments, the existing sensors can be recalibrated to include prediction of the concentrations of gas components in a gas mixture, thereby enhancing the sensors' response functionalities.

2.2. Numerical simulation calibration method

Numerical modelling techniques can be used to simulate sensor response in a system or an engineering material. Numerical modelling is performed based on mathematical models (Thekkethil et al., 2016; Ghommem et al., 2022) developed to predict physical, thermal, hydrological, mechanical, and chemical phenomena in real-life processes and systems. For example, gas mole fractions and temperature can be measured by a direct Tunable Diode Laser Absorption Spectroscopy (TDLAS) technique which is based on the Beer-Lambert law or model, relating optical absorbance to gas parameters (Diemel et al., 2019; Zhang et al., 2023). So, by applying the Beer-Lambert mathematical model, sensors are created that can measure gas mole fraction and temperature with reference to the gas optical absorbance. Most numerical simulations are achieved by discretizing differential equations into discrete difference equations in time and space (Scherer, 2013). For example, the diffusion equation in one spatial dimension

$$\frac{\partial f(x,t)}{\partial t} = D\frac{\partial^2}{\partial x^2} f(x,t) + S(x,t) \tag{2}$$

Ecan be discretized (Fig. 3) in the forward in time, centred in space (FTCS) scheme as

$$\left(f_m^{n+1} - f_m^n\right) = D\frac{\Delta t}{\Delta x^2}\left(f_{m+1}^n + f_{m-1}^n - 2f_m^n\right) + S_m^n \Delta t \tag{3}$$

where f_m^{n+1}, f_m^n, and Δt represent the concentration term at the current time step, the concentration term at the previous time step, and the time step or interval (between the previous and the current time, respectively); Δx, f_m^n, f_{m+1}^n, f_{m-1}^n , and S_m^n represent the space interval (between two consecutive discretized points), the concentration term at the point of interest in space, the concentration term at a point that is a space interval ahead of the point of interest, the concentration term at a point that is a space interval before the point of interest, and the source or sink term at the point of interest – all the terms in the previous time step; while D represents the constant diffusion coefficient.

Numerical simulations have been performed in several studies to validate sensor measurements (Walsh et al., 2011; Bellegoni et al., 2022; Shwetha et al., 2022; Xu et al., 2023). Walsh et al. (2011) conducted a study to investigate how electrochemical sensors (with either acid or alkaline electrolytes) respond to gas mixtures made up of air (about 21% of oxygen, by volume) and enhanced levels of carbon dioxide, nitrogen, helium or argon (Fig. 4). They conducted experiments using the

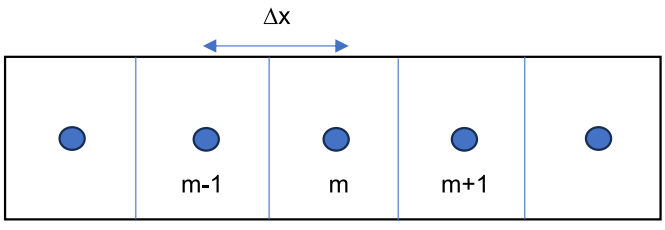


Fig. 3. Numerical solution approach.

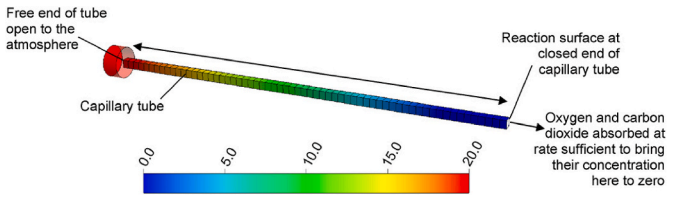


Fig. 4. Numerical simulation of sensor responses to gas mixtures (Source: Walsh et al., 2011).

electrochemical sensors and performed numerical simulations using Ansys CFX (a Computational Fluid Dynamics [CFD] simulation software). Based on the numerical simulations, they found that the electrochemical sensors over-read the oxygen concentrations when high levels of helium are in the atmosphere. They also found that the electrochemical sensors with alkaline electrolytes underestimate the severity of the oxygen in atmospheres containing high amounts of carbon dioxide. Overall, they found that the relative percentage difference between the CFD model prediction and recorded sensor responses is less than 10% for almost all the groups of gas mixtures tested. Therefore, using an appropriate numerical simulation tool and setting the numerical model parameters and conditions correctly, poor sensor responses to certain conditions (for instance, the electrochemical sensors over-reading oxygen concentrations when high levels of helium are in the atmosphere) can be detected. Thus, flaws in the sensor measurements would necessitate recalibration of the sensor to measure certain properties and conditions correctly.

3. Methodology

This study proposed a new sensor calibration methodology and the design of a hybrid artificial intelligence system for real-time quantification of impurities in gas streams. Furthermore, machine learning models were developed in this study to explore how impurities in gas streams can be quantified. In the case study for the machine learning models, a nitrogen gas (N_2) system with impurities was used for the demonstration. The study compares the performance of sensors in quantifying the concentrations of component gases in binary gas and multi-component gas mixtures. The binary gas mixture is made up of N_2 and NO_2, while the multi-component gas mixtures are mixtures of N_2 with two-gas, three-gas, or four-gas combinations of NO, NO_2, C_3H_8, and NH_3; each gas mixture has a constant 10% O_2 by volume (as oxygen reduction is a requirement in electrochemical sensors). The hybrid artificial intelligence system of quantifying gas concentrations is scalable and can be applied in complex gas mixtures found in different industries. For instance, in a CO_2 gas stream, the binary gas mixture could represent a mixture of CO_2 with NO_2; while the multi-component gas mixtures could represent CO_2 with two-gas, three-gas, or four-gas combinations of NO, NO_2, C_3H_8, and NH_3.

3.1. Proposed sensor calibration methodology for the design of hybrid artificial intelligence systems for real-time quantification of impurities in a CO_2 gas stream

This study proposes a hybrid artificial intelligence system for real-time quantification of impurities in gas streams. This gas monitoring technology can be applied in different gas systems as well as in carbon capture, utilization, and storage. For example, during CO_2 capture from the atmosphere, different impurities in the mixture are passed into the separation chamber, where CO_2 is captured preferentially over other gas impurities, while a small amount of CO_2 is collected together with other gas impurities (Agunbiade et al., 2023). These impurities can be re-injected into the separation chamber, following the same process, until the CO_2 gas in the effluents (impurities) is negligible. Thus, the concentration or mole fraction of the CO_2 gas in the separation chamber during the different separation cycles is different. Similarly, in the beverage industry where CO_2 utilization is common, the procedures adopted in producing the CO_2 used in the beverage preparation make it possible to have a trace amount of impurities such as sulphur compounds and hydrocarbons. To certify the purity of the CO_2 used in the beverage industry, rigorous techniques such as gas chromatography and mass spectrometry are adopted by CO_2 suppliers. However, during transportation and storage the CO_2 gas can become contaminated (Vilanova et al., 2003). Therefore, it is paramount to develop a real-time monitoring system in carbon capture facilities, CO_2 utilization facilities such as beverage processing plants or storage units, and gas pipelines transporting CO_2 gas for underground storage in aquifers or depleted oil or gas reservoirs.

A hybrid artificial intelligence (AI) system will help quantify the concentrations or mole (or mass) fractions of the gas components in a mixture in real time, enabling users to monitor the concentrations on their cell phones or wireless devices at a low cost. The system will also promote quality control and assurance and reduce non-productive time (NPT), as early detection of malfunction of gas sensors is assured for the recalibration or replacement of the sensors. The hybrid AI system can be designed using sensors and machine learning. It is important to understand the gas components expected in the facility or medium where the suite of sensors will be installed. Also, the array of sensors in a suite [of sensors] should be carefully selected such that each sensor has preferential selectivity to a particular gas in the gas mixture. For example, a sensor that is very sensitive to SO_2 gas should have very little or no sensitivity to H_2S gas, while another sensor that is very sensitive to H_2S gas has very little or no sensitivity to SO_2 gas to minimize sensor interference or cross-sensitivity. The procedures for designing the hybrid artificial intelligence system are explained below (Fig. 5).

A gas mixture (for example, CO_2 gas with some gas impurities properly mixed) of known concentration or mole (or mass) fractions of component gases should be allowed to flow through the array of sensors (ensuring the same gas composition flows through each of the sensors). The sensors' response [of physical quantity change] to different concentrations of the gas components in a mixture is recorded through a data acquisition system (such as a LABVIEW program installed in a computer). Where possible, numerical simulations are performed to validate (or corroborate) the experimental results, using relevant software that can model the concentration or mole (or mass) fractions of the component gases in the mixture and output relevant physical, thermal, or electrical properties of the mixture. If the numerical simulation results properly fit the experimental results, then further numerical simulations can be performed to cover a larger range of data and increase the amount of data available to enhance the design of new sets of sensors.

The experimentally validated numerical simulation data (or the experimental data alone) can be used to calibrate another sensor or calibrate the existing sensor with more functionality to measure the physical, thermal, or electrical properties as well as the concentration of the gas components in the gas mixture. It is worth noting that the same dataset source (experimental data or numerically validated experimental data) should be used for the sensor calibration and the development of the machine learning model to ensure the reliability of the data.

Some of the properties of the original sensors' (before recalibration) response to gas mixtures could be resistance, conductance, capacitance, voltage, or dielectric constant. The real-time sensor response is recorded in a data acquisition system and deposited in a cloud-based system (a container or bucket in a cloud platform like Microsoft Azure, Amazon Web Services, or Google Cloud Platform) at the same time, including the concentrations of the gas components in the gas mixture.

The experimental data or experimentally validated numerical simulations data from which the sensors are calibrated to improve their functionalities are processed to develop a machine learning (ML) model that uses the sensors' property response to the gas mixture to predict the

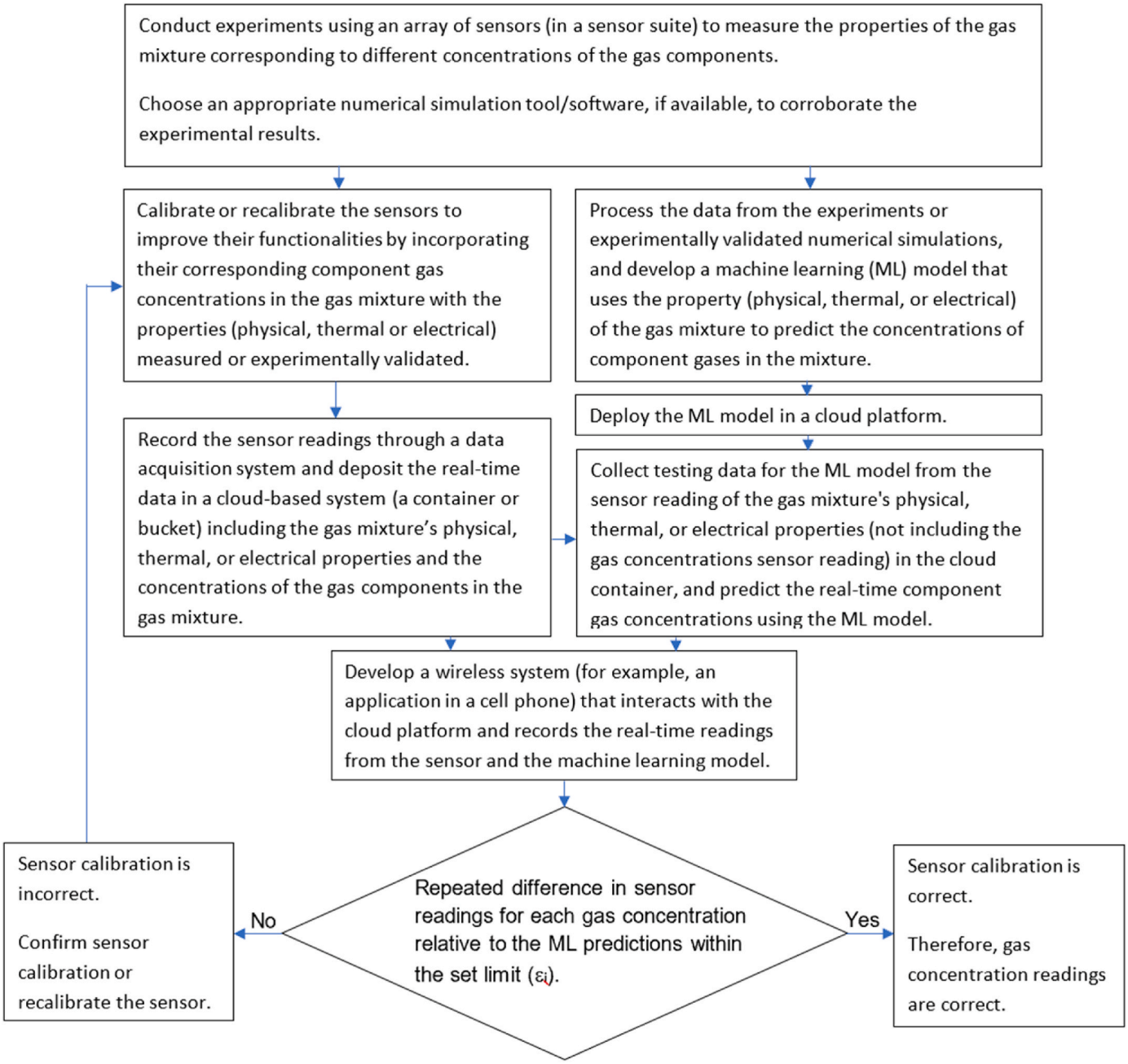


Fig. 5. Proposed methodology for the design of hybrid artificial intelligence systems for real-time quantification of impurities in a gas stream and recalibration of sensors.

concentrations of component gases in the mixture. In other words, the input features include each sensor's property response (for example, voltage) when the gas mixture flows through the array of sensors. In some cases, the input features could be the changes in the property reading (Vilanova et al., 2003). For example, in a CO_2 gas stream with different gas impurities, the voltage change (ΔV_{gas_ij}) is the difference in voltage of sensor j in the presence of CO_2 (with or without oxygen as a base gas) with impurities (V_{gas_ij}) and the voltage of sensor j (V_{0j}) in the presence of pure CO_2 (with or without oxygen as a base gas).

$$\Delta V_{gas\_ij} = V_{gas\_ij} - V_{0j} \tag{4}$$

The developed machine learning model is deployed in a cloud platform. The test data applied in the machine learning model in the cloud platform to predict the real-time component gas concentrations are collected from the sensors' property response (for example, voltage) to the gas mixture's (which can be automatically converted to the property change in the machine learning model, if the model is trained using the property change) without including the gas concentrations reading of the calibrated sensors in the cloud system (container or bucket) (Narayana et al., 2024).

To connect the sensors' reading and the machine learning predictions of gas concentrations to enable a hybrid artificial intelligence system, a wireless system (for example, an application in a cell phone or a wireless device) can be developed (Fig. 6), which can interact with the cloud platform and record the real-time readings from the array of sensors and the machine learning model. The wireless device is also able to measure the absolute error (ΔX_i) in the prediction of the concentration (or mole fraction) of each gas component in the mixture. The absolute error in the prediction of the concentration of each gas composition (ΔX_i) is the absolute difference in the gas concentration

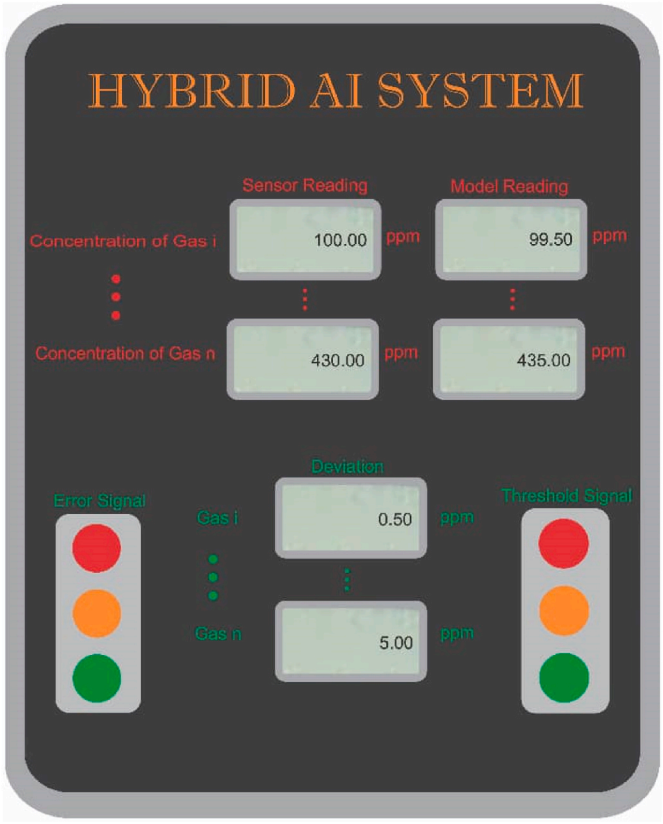


Fig. 6. Proposed design of the wireless system to quantify impurities in a gas stream.

predicted by the suite of sensors (X_{i_Sensor}) based on the sensors' response (physical, thermal, or electrical) and the gas concentrations predicted by the machine learning model (X_{i_ML}). An error limit (ε_i) can be critically set based on technical judgments (it could be based on the maximum error, mean, or median absolute error in the machine learning model predictions of each gas component concentration), to measure the performance of the hybrid artificial intelligence system. That is, to test whether the improved sensor design is properly calibrated or not. At least two suites of sensors should be in place – one of them only recording its response to the different gas concentrations in the mixture, and the other being a high functionality suite of sensors calibrated based on experiments or experimentally validated numerical simulations. These two suites of sensors help to ensure the reliability of the sensors' response (original physical quantity for which it was designed, for example, voltage) based on which the machine learning model makes its predictions of the component gas concentrations in the mixture.

If several (say, 50% of the time) differences in sensor readings over a period (say a day or a month for different physical quantity readings) for each gas concentration relative to the machine learning predictions are not within the set error limit ($\Delta X_i > \varepsilon_i$), the sensors' calibration (for the high functionality sensor designed) is incorrect. Then, the sensors' calibration should be confirmed or a responsible person (for example, an Engineer or a Technologist in charge) should proceed to recalibrate the sensors. Otherwise, the sensors' calibration is correct. Thus, users of the application or wireless device can monitor the sensor and the machine learning model predictions of the gas concentrations in real-time (and remotely).

$$\Delta X_i = \left| X_{i\_Sensor} - X_{i\_ML} \right| \tag{5}$$

For quality control and assurance purposes, the error and threshold signals can be designed in the wireless system and colour-coded, following the traffic light signs (green, amber, and red) as shown in Fig. 6. For the threshold signals, the green light activates when the concentrations of each of the gases are well below the set thresholds, amber activates when the concentrations of some of the gases are very close to or at the set thresholds, while red activates when some or all the gases in the mixture exceed the set thresholds. As soon as the gas concentration threshold is reached, the gas supply could be shut down to avoid excess concentrations of a specific gas or gases in the mixture.

Similarly, for the error signals, the green light activates when the deviations in concentration in the sensor readings and the machine learning model predictions are well below the set thresholds, amber activates when the deviations in concentration in the sensor readings and the machine learning model predictions are very close to or at the set thresholds, while red activates when deviations in concentration in the sensor readings and the machine learning model predictions for some or all the gases in the mixture exceed the set thresholds. The activation of the red sign of the error signals could suggest sensor malfunction and the need for sensor recalibration or replacement. Then the sensor should be replaced or recalibrated.

However, to maintain accurate readings, the sensors should be periodically tested, since any error in the property readings, will result in an error in the machine learning model predictions. This is because the machine learning model depends on the physical, thermal, or electrical properties of the gas mixture (sensors' response) to make predictions of the concentrations of the gases. It is also important to ensure that the operating temperature and pressure conditions of the sensors are maintained in the gas system monitored, to ensure accuracy in sensor and machine learning predictions.

3.2. Case study

Experimental data used in this study were from the work of Javed et al. (2022). They measured the responses of mixed-potential electrochemical sensors (MPES) to gas mixtures. The patented Los Alamos National Laboratory (LANL) MPES devices are made up of a solid electrolyte (3 mol percent yttria-stabilized zirconia [YSZ]), Pt pseudo-reference electrode that has higher reaction rates for oxygen reduction, and Au/Pd (also represented as AuPd) or $La_{0.8}Sr_{0.2}CrO_3$ (LSCrO) working electrode which has significantly different redox reaction rates from those at the Pt electrode. The packaged electrochemical sensors were mounted in a steel compartment with a gas inlet and outlet. The different electrochemical sensors are labelled by their compositions and operating temperatures. LSCrO|YSZ|Pt which operated at 450 °C, 470 °C, and 545 °C are labelled as Cr450, Cr470, and H545 (developed mainly for hydrogen sensing), respectively; while AuPd|YSZ|Pt which operated at 475 °C is labelled as Au475. The four sensors are arranged in a series connection, allowing gas mixtures to enter the array of sensors through Cr450, before passing through Au475 and Cr470, and exiting the array after contacting the H545 sensor. Before the gas mixture was introduced to the suite of sensors, the component gases were mixed in a mixing chamber (stainless steel manifold) of constant volume. Each gas mixture formulated was maintained at a constant volume percent of oxygen (10% O_2) as the base gas while varying the composition of the balance gas (N_2) and the analyte gases (single gas, two-gas, three-gas, or four-gas combinations of NO, NO_2, C_3H_8, and NH_3). The balance gas was used to maintain a constant flow rate of 1000 cm^3/min. Throughout the experiment, no significant drift in the sensor response was observed, as the sensors were tested routinely with fixed concentrations of each analyte gas (Javed et al., 2022).

In the present study, a regression model based on Artificial Neural Network (ANN) algorithm was developed, using Python programming language, to quantify the concentrations of gas components using data from the work of Javed et al. (2022) and to demonstrate the design of a hybrid artificial intelligence system for real-time quantification of impurities in gas streams. The algorithm involves multiple input-multiple output (MIMO) regression neural networks as shown in Fig. 7

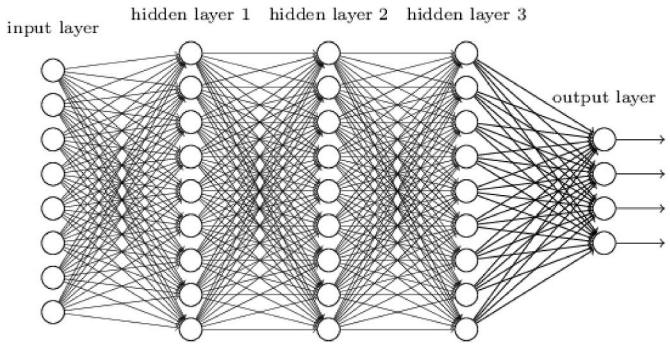


Fig. 7. Fully connected MIMO neural network (adapted from McNaughton, 2019).

(Antanasijević et al., 2018; McNaughton, 2019).

ANN model is based on the architecture of the human brain, made up of neurons, which receive and transmit (after processing) information to all adjacent neurons (He et al., 2022). The connections between these neurons are defined by weights. ANN model is structured in layers (input, hidden, and output layers) with nodes in one layer connected to nodes in the following layer. The nodes (or neurons) adopt the weighted connection to learn the dataset and utilize an activation function to transmit their signal to the output layer (Kannaiah and Maurya, 2023). In other words, the learning process of the ANN model is in its adjustment of weights (iteratively) between neurons and the bias of each neuron in the way of repeated input and output (a process referred to as model training); thus, making it possess excellent non-linear fitting abilities (He et al., 2015, 2022; Yao et al., 2023). The weights are updated as follows:

$$w_i = w_{i-1} - \alpha\left(\frac{dLoss}{dw_{i-1}}\right) \tag{6}$$

where w_{i-1} and w_i represent the old weight and updated weights, respectively; α is the learning rate, while $dLoss/dw_{i-1}$ represents the derivative of error (or loss function) with respect to weight.

Some parameters that control the performance of the neural network are optimizers, batch size, and epochs. Optimizers are algorithms used to minimize loss function or error during model training. This is achieved by adjusting the weights and learning rate during training. Some common optimizers are root mean square propagation (RMSprop), stochastic gradient descent (SGD), and adaptive moment estimation (Adam). Furthermore, to improve the fitness of the model, optimum batch size and epochs are chosen. Batch size (controls how many observations in the training data pass through the algorithm at a time, until the entire training data pass through the algorithm in an epoch. Epochs control the number of times the entire training data goes (or passes) through the algorithm during the training. The parameters of the neural network model are updated with each epoch (Kannaiah and Maurya, 2023).

3.2.1. Data preparation

Data from experiments conducted by Javed et al. (2022) are used in developing the machine learning model. The original data separated into training and testing data were concatenated into a single dataset with 107,103 data points. The sensor responses (in voltage) are the input features (Cr450 [V], Au475 [V], Cr470 [V], and H545 [V]); while the gas concentrations (NO [ppm], C_3H_8 [ppm], NO_2 [ppm], and NH_3 [ppm]) are the output features in this case of multiple-input and multiple-output deep learning neural network. Outliers were detected in the input features and removed from the dataset using the Inter Quartile Range (IQR) method (Dash et al., 2023). IQR is defined as the 75th percentile minus the 25th percentile of the input feature. Based on the IQR method, any value that lies outside the range of (25th percentile – 1.5(IQR)) to (75th percentile + 1.5(IQR)) is referred to as an outlier (Dash et al., 2023). After cleaning the data for the multicomponent gas mixture, the number of data points (rows) was reduced to 37967. The distribution of the features is presented in Table 1.

The correlation coefficients between each feature and another were determined using Pearson correlation coefficient (r), expressed as

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{\left[n\sum x^2 - (\sum x)^2\right]\left[n\sum y^2 - (\sum y)^2\right]}} \tag{7}$$

where, n is the number of observations, and x and y represent the features correlated.

The Pearson correlation coefficient is more suitable for measuring the degree of linear correlation between two variables (Profillidis and Botzoris, 2024). In this study, the dependent and independent variables correlated linearly. Therefore the Pearson correlation coefficient was suitable, based on the variables considered in this study. A strong correlation between dependent and independent variables is preferred and acceptable, while a strong correlation between two independent variables is undesirable. When two independent variables strongly correlate with each other, it suggests that excluding one of them from the training data would be beneficial to prevent redundancy and improve model performance (Kannaiah and Maurya, 2023). Hence, one of the features is removed for two independent variables in the dataset perfectly correlating (r = 1.0) with each other. The resulting correlation coefficient matrix is shown in Fig. 8.

The voltage readings [V] of sensors Cr450 and Cr470 positively correlate with the concentrations [ppm] of NO, C_3H_8, NO_2, and NH_3; the voltage reading [V] of sensor H545 positively correlates with the concentration [ppm] of C_3H_8 and NH_3 but did not correlate significantly with the concentration [ppm] of NO and NO_2; while no significant correlation was observed between the voltage reading [V] of sensor Au475 and any of the gas components, except NH_3. The correlations suggest that the higher the voltage reading from the Cr450 and Cr470 sensors, the higher the chance of predicting more concentrations [ppm] of each of the gases; while the higher the voltage reading of sensor H545, the higher the chance of predicting a higher concentration [ppm] of C_3H_8 and NH_3 in the mixture. The higher the voltage reading of the Au475 sensor, the higher the chance of predicting more concentration of NH_3 compared to other gases. Hence, the correlation suggests that there is interference or cross-sensitivity in all the sensors (Cr450, Cr470, H545, and Au475) in predicting NH_3 concentration in the gas mixture.

To create a dataset of the binary gas mixture (where N_2 is the balance gas and NO_2 is the analyte gas or impurity), a Python code was written to select only data points (or rows) where the concentrations of other analyte gases are simultaneously zero. Hence, the number of observations becomes 192. The distribution of the features is shown in Table 2. For the binary gas mixture, H545 [V], Cr450 [V], and Cr470 [V] correlate with each other with r = 0.99–1.0, therefore, H545 [V] and Cr450 [V] were removed. The resulting correlation coefficient matrix is shown in Fig. 9.

The voltage reading [V] of sensor Cr470 negatively correlates with the concentration [ppm] of NO_2 significantly, while the voltage reading [V] of sensor Au475 positively correlates with the concentration [ppm] of NO_2 in the binary gas mixture. The correlations suggest that the higher the voltage reading from the Cr470 sensor, the lower the concentration [ppm] of NO_2; while the higher the voltage reading of sensor Au475, the higher the concentration [ppm] of NO_2 in the mixture. Thus, it appears that there might be very little or no interference or cross-sensitivity in sensors Cr470 and Au475 in predicting NO_2 concentration in the binary gas mixture, due to the opposite correlation of gas concentrations [ppm] with the sensors' voltage readings [V].

The statistical distribution of the datasets shows that the mean and median (50th percentile) are different (skewed distribution) for most of the features. Therefore, the input and output features (variables) were

Table 1
Statistical parameter for the multicomponent mixture dataset cleaned.

	Count	Mean	Std	Min	25%	50%	75%	Max
Cr450 (V)	37967.0	0.066500	0.018592	0.012272	0.054670	0.069713	0.081083	0.099234
Au475 (V)	37967.0	0.139850	0.026158	0.037733	0.123306	0.144076	0.160185	0.186824
Cr470 (V)	37967.0	0.065718	0.018187	0.013952	0.053246	0.068132	0.079926	0.101352
H545 (V)	37967.0	0.068193	0.012848	0.015095	0.060823	0.071684	0.078021	0.088066
NO (ppm)	37967.0	115.181303	60.582273	19.831350	60.120000	120.020314	159.824010	223.205700
C_3H_8 (ppm)	37967.0	342.617031	240.894968	20.088368	149.637001	279.966044	482.358971	1107.686963
NO_2 (ppm)	37967.0	35.587152	22.828174	4.048490	16.050036	31.949990	48.000048	88.655815
NH_3 (ppm)	37967.0	162.795553	114.516446	4.947795	69.930045	134.877409	239.732935	498.438661

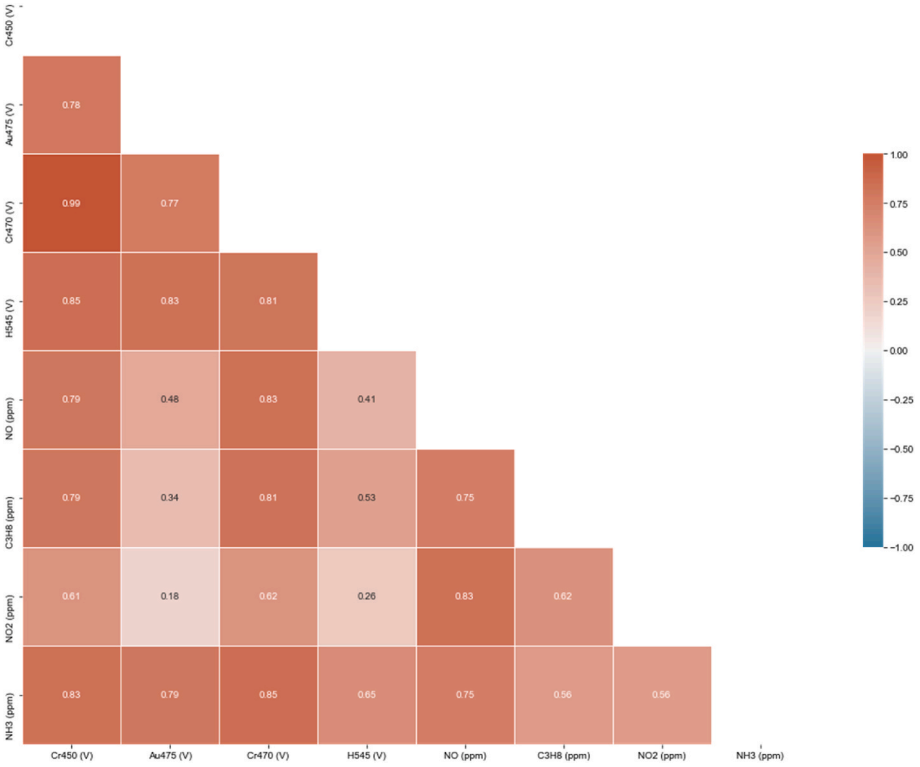


Fig. 8. Correlation coefficient matrix of features for the multicomponent mixture dataset.

Table 2
Statistical parameter for the binary gas mixture dataset.

	Count	Mean	Std	Min	25%	50%	75%	Max
Au475 (V)	192.0	0.002157	0.005052	−0.005641	−0.001889	0.001990	0.007349	0.00903
Cr470 (V)	192.0	−0.019486	0.008669	−0.034985	−0.026628	−0.016244	−0.012481	−0.01030
NO_2 (ppm)	192.0	24.752169	21.390178	4.019612	7.914937	15.925900	39.955507	65.11670

scaled. The features were scaled using a standard scaler, to prevent any feature with high values from overfitting the model as other equally important features might be neglected if not scaled. Thus, scaling the features improves the training accuracy (Kannaiah and Maurya, 2023). Standard scaler standardizes input features as follows:

$$z = \frac{x - \mu}{\sigma} \tag{8}$$

where, x represents the input variable, while μ and σ are the mean and standard deviation of the variable dataset, respectively.

3.2.2. Model architecture

This study applied the pre-processed datasets for the model development. For the multi-component gas mixtures, the input features are Cr450 [V], Cr470 [V], Au475 [V], and H545 [V]; while the output

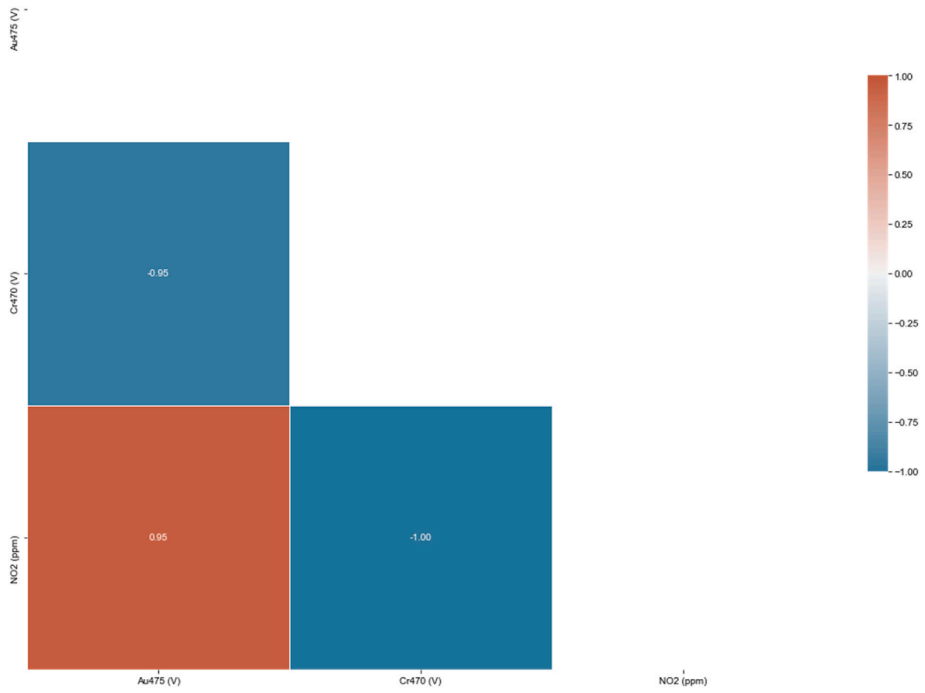


Fig. 9. Correlation coefficient matrix of features for the binary gas mixture.

features are the concentrations of the individual gases in the mixtures. Similarly, for the binary gas mixture, Cr470 [V] and Au475 [V] are the input features; while the output feature is the concentration of NO_2 [ppm]. The structure of the ANN model is made up of 2–4 nodes in the input layer (representing the input features), 3 hidden layers (64 neurons each), and 1 or 4 nodes in the output layer (representing the output features – gas concentrations). The neural network is fully connected. In each layer (except the output layer), a rectified linear unit (ReLU) was used as an activation function, while He normal variance scaling initializer was used to generate the weights (He et al., 2022). No transformation is needed at the output layer as numerical values (gas concentrations) are expected.

3.2.3. Model evaluation

To gauge the accuracy of the machine learning models, some evaluation criteria (Chicco et al., 2021) including mean absolute error (MAE), median absolute error (MdAE), maximum error (ME), mean absolute percentage error (MAPE), mean square error (MSE), root mean square error (RMSE), and coefficient of determination (R^2 or R-squared score). In the model evaluation, N is the total number of observations (or the total number of data values), $\hat{y}_i$ is the predicted ith value, y_i is the actual ith value, and $\bar{y}$ is the mean of actual value (Equations (9)–(15)).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{9}$$

$$MdAE = Median\ (|y_i - \hat{y}_i|) \tag{10}$$

$$ME = Maximum\ (|y_i - \hat{y}_i|) \tag{11}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i} * 100 \tag{12}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{13}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{14}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y_i})^2} \tag{15}$$

3.2.4. Model development

To build the multiple-input and multiple-output neural network models for the multi-component gas mixture and a general artificial neural network (ANN) model for the binary gas mixture, the pre-processed datasets were used. The optimal machine learning model was achieved by dividing the data into three sets: 70% of the dataset was selected as the training set, 12% as the validation set, and 18% as the testing set. The training set was used to build the model, while the validation set was employed to ensure the model was not overfitting or underfitting. The testing set was reserved to test the developed and validated model, to ensure there was no form of data leakage during the model development stage, and to reveal the accuracy of the model in estimating gas concentrations. Satisfactory tuning of optimizer hyperparameters (the batch size and epochs) of training was performed, using the grid search approach, to obtain better predictions of gas concentrations. The hyperparameter tuning process was performed using different hidden layers (single hidden layer, two hidden layers, and three hidden layers, in different cases), as well as batch sizes and epochs. The batch sizes considered are 16, 32, 64, and 128; while the number of epochs considered are 10, 20, 50, and 100. The loss function and optimizer employed in this study are mean square error and Adam, respectively. During the hyperparameter tuning, the model was cross-validated. For the multicomponent gas mixture, a combination of three hidden layers, a batch size of 16, and 50 epochs, gave the minimum mean absolute percentage error (2.89% for the validation dataset, and 2.80% for the training dataset) with a very small deviation between the validation and training MAPE. Similarly, for the binary gas mixture a combination of three hidden layers, a batch size of 16, and 10 epochs,

gave the minimum mean absolute percentage error (0.89% for the validation dataset, and 0.80% for the training dataset). Therefore, the model is generalizable, and neither overfitting nor underfitting (Figs. 10 and 11). The model parameters are shown in Table 3.

3.2.5. Extreme gradient boosting

Extreme Gradient Boosting (XGBoost) was used in this study to compare the performance of the deep learning models for the multi-component and binary gas mixtures. In the development of machine learning models using a limited dataset, XGBoost is less prone to overfitting and has lower processing requirements compared to deep learning (Jeong et al., 2021; Karbassiyazdi et al., 2022). XGBoost transforms many weak learners (for example, decision trees) into strong learners through its enhanced gradient-boosting algorithm (Karbassiyazdi et al., 2022).

Similar to the deep learning models developed in this study, the multiple-input and multiple-output algorithm of the XGBoost was adopted. The same datasets used for the deep learning models were used for the XGBoost models. The input features are Cr450 [V], Cr470 [V], Au475 [V], and H545 [V]; while the output features are the concentrations of the individual gases in the mixtures. Also, for the binary gas mixture, Cr470 [V] and Au475 [V] are the input features; while the output feature is the concentration of NO_2 [ppm]. The datasets were split (70% of the dataset was selected as the training set, 12% as the validation set, and 18% as the testing set). Hyperparameters (gamma, maximum depth of tree, and learning rate) were selected using a grid search approach on a 10-fold cross-validation of the XGBoost model using the training dataset. The input and output features were not scaled to achieve good model performance, as all outliers were removed from the datasets. Moreover, XGBoost can handle complex datasets (including datasets with missing information) without the need for imputation preprocessing (Rusdah and Murfi, 2020; Karbassiyazdi et al., 2022).

The optimum hyperparameters (Table 4) were selected from the grid search for the multicomponent gas mixture (gamma = 0.0, learning rate = 0.15, maximum depth = 10, and number of estimators = 150) and the binary gas mixture (gamma = 0.1, learning rate = 0.15, maximum depth = 5, and number of estimators = 100).

4. Results

4.1. Prediction of gas concentrations in multi-component gas mixtures

4.1.1. Deep learning model prediction of the multicomponent gas mixture

The ANN model developed for the multicomponent gas mixture predicted the component gas concentrations with mean absolute error, median absolute error, and root mean square error of 3.35 [ppm], 2.40 [ppm], and 6.18 [ppm], respectively; while the R^2 value and the mean absolute percentage error are 99.79% and 3.13%, respectively. Therefore, 99.79% of the variance in the gas concentrations is explained by the model (Table 5) and the model predicts the gas concentrations with good precision. The median absolute error of 2.40 [ppm] is less than the mean absolute error and indicates that at least 50% of the model's predictions of gas concentrations had an overall absolute error of less than 2.40 [ppm]. Also, the negligible difference in the performance of the model based on training dataset and testing dataset, confirms that the model is neither overfitting nor underfitting.

Furthermore, the model's predictions of concentrations of each gas component were evaluated and shown in Table 6. The model predicted gas concentrations with the least error in NO_2 concentration (MAE = 0.78 [ppm], MdAE = 0.61 [ppm], and RMSE = 1.24 [ppm]) and the highest error in C_3H_8 concentration (MAE = 7.72 [ppm], MdAE = 5.44 [ppm], and RMSE = 11.00 [ppm]). The R^2 value for all the gas concentrations predicted is greater than 99%. Therefore, over 99% of the variance in concentrations for each of the gases is explained by the model. The maximum error in the predictions of concentrations is in tens of ppm, compared to the median absolute errors (each less than 6 [ppm]). This suggests that the chance of predicting gas concentrations with errors up to tens of ppm is low.

The input features' importance in predicting the gas concentrations was evaluated based on the absolute sum of weights between the neurons in the first hidden layer of the fully connected neural network. The features with higher weights might have contributed more to the final predictions of the model. The order of importance of the input features in predicting gas concentration using the machine learning model is shown in Fig. 12.

The most important feature in the prediction of the gas concentrations is Cr450 [V], followed by Cr470 [V], while the least important feature in the prediction of the gas concentrations is Au475 [V]. The relatively low importance of Au475 [V] and H545 [V] might be attributed to the Cr450 [V] and Cr470 [V] sensors' interference or cross-sensitivity in the prediction of gas concentrations. Therefore, the voltage readings [V] of the Cr450 and Cr470 sensors which appeared to have exhibited significant interference or cross-sensitivity with any of the sensors were more important in predicting reliable gas concentrations of the multi-component gas mixtures. There is very little or no interference (or cross-sensitivity) between the Au475 and H545 sensors while taking measurements, evidenced by a very small difference in their absolute sum of weights in predicting gas concentrations (close level of feature importance).

4.1.2. XGBoost model prediction of multicomponent gas mixtures

The XGBoost model developed for the multi-component gas mixture predicted the component gas concentrations with mean absolute error, median absolute error, and root mean square error of 0.53 [ppm], 0.13 [ppm], and 2.04 [ppm], respectively; while the R^2 value and the mean absolute percentage error are 99.96% and 0.33%, respectively. Therefore, 99.96% of the variance in the gas concentrations is explained by the model (Table 7). The median absolute error of 0.13 [ppm] is less than the mean absolute error and indicates that at least 50% of the model's predictions of gas concentrations had an overall absolute error of less than 0.13 [ppm].

Furthermore, the model's predictions of concentrations of each gas

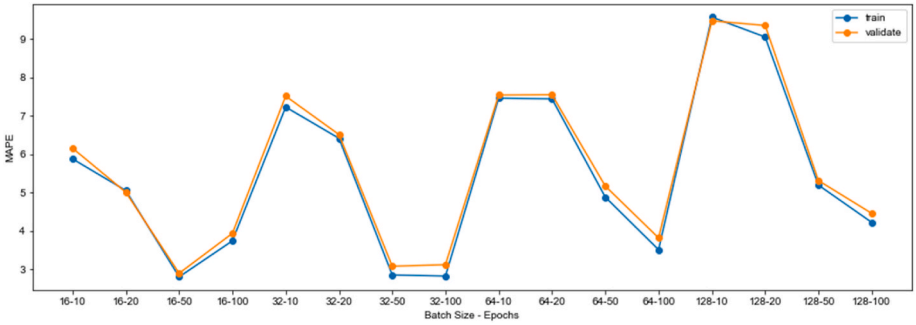


Fig. 10. Hyperparameter tuning for the multi-component gas mixture.

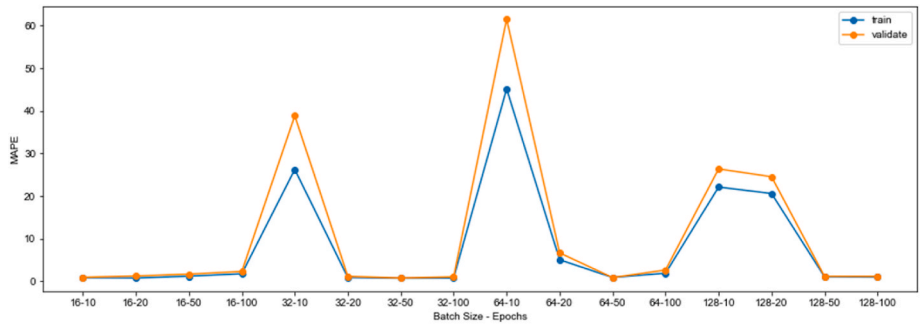


Fig. 11. Hyperparameter tuning for the binary gas mixture.

Table 3
Deep learning model structure and parameters.

Parameters	Binary Mixture	Multicomponent Mixture
Number of hidden layers	3	3
Number of neurons in the hidden layers	64	64
Number of neurons in the output layer	1	4
Number of output feature(s)	1	4
Number of input features	2	4
Activation function in the hidden layers	ReLU	ReLU
Kernel initializer	He normal	He normal
Seed value	42	42
Loss function	Mean square error	Mean square error
Optimizer	Adam	Adam
Batch size	16	16
Number of epochs	10	50

Table 4
XGBoost model hyperparameters.

XGBoost hyperparameters	Binary Mixture	Multicomponent Mixture
Gamma (0.0, 0.1, 0.2, 0.3)	0.1	0.0
Learning rate (0.01, 0.05, 0.10, 0.15)	0.15	0.15
Maximum depth (5, 6, 8, 10)	5	10
Number of estimators (20, 50, 100, 150)	100	150

Table 5
Performance measurements of ANN model for the multi-component gas mixtures.

	Performance measures				
	Gas concentration				
	R^2 [%]	MAPE [%]	MAE [ppm]	MdAE [ppm]	RMSE [ppm]
Overall training set	99.83	3.21	2.93	2.17	5.07
Overall testing set	99.79	3.13	3.35	2.40	6.18

component were evaluated and shown in Table 8. The model predicted component gas concentrations with the least error in NO_2 concentration (MAE = 0.14 [ppm], MdAE = 0.03 [ppm], and RMSE = 0.65 [ppm]) and the highest error in C_3H_8 concentration (MAE = 1.14 [ppm], MdAE = 0.28 [ppm], and RMSE = 3.36 [ppm]). The R^2 value for the prediction of each of the gases is greater than 99%. Therefore, over 99% of the

Table 6
Test performance measurements of ANN model for the multi-component gas mixtures.

Gas	Performance measures					
	R^2[%]	MAPE [%]	MAE [ppm]	MdAE [ppm]	Max Error [ppm]	RMSE [ppm]
NO	99.84	2.24	1.71	1.37	23.27	2.40
C_3H_8	99.79	3.47	7.72	5.44	63.91	11.00
NO_2	99.71	3.74	0.78	0.61	19.92	1.24
NH_3	99.82	3.08	3.19	2.18	54.47	4.95

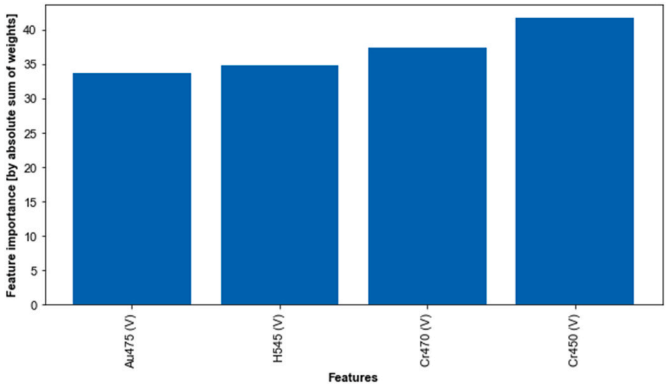


Fig. 12. Feature importance using the multiple-input and multiple-output ANN model.

Table 7
Performance measurements of XGBoost model for the multi-component gas mixtures.

	Performance measures				
	R^2 [%]	MAPE [%]	MAE [ppm]	MdAE [ppm]	RMSE [ppm]
Overall training set	100.00	0.14	0.19	0.09	0.43
Overall testing set	99.96	0.33	0.53	0.13	2.04

variance in concentrations for each of the gases is explained by the XGBoost model. The maximum error in the predictions of concentrations is in tens of ppm, compared to the median absolute errors (each less than 0.3 [ppm]). This suggests that the chance of predicting gas concentrations with errors up to tens of ppm is low.

The input features' importance in predicting each of the gas concentrations was evaluated using XGBoost algorithm. The order of importance of the input features in predicting gas concentration using

Table 8
Test performance measurements of XGBoost model for the multi-component gas mixtures.

Gas	Performance measures					
	R^2[%]	MAPE [%]	MAE [ppm]	MdAE [ppm]	Max Error [ppm]	RMSE [ppm]
NO	99.97	0.23	0.28	0.06	24.08	0.99
C_3H_8	99.98	0.36	1.14	0.28	78.56	3.36
NO_2	99.92	0.42	0.14	0.03	24.80	0.65
NH_3	99.97	0.32	0.55	0.13	41.82	1.98

the XGBoost model is shown in Fig. 13.

The most important feature in the prediction of the NO and NO_2 concentrations is Cr470 [V], followed by H545 [V], while the less important features in the prediction of the gas concentrations are Cr450 [V] and Au475 [V]. The relatively low importance of Cr450 [V], Au475 [V], and H545 [V] suggests that Cr470 [V] is more sensitive to NO and NO_2 concentrations, compared to the other sensors. The most important feature in the prediction of the C_3H_8 concentrations is Cr470 [V], followed by Cr450 [V] and Au474 [V], while the least important feature is H545 [V]. Similarly, the less important features in the prediction of NH_3 concentrations are H545 [V] and Au475 [V]; however, the most important feature is Cr450 [V], followed by Cr470 [V]. Overall, the more important features in the prediction of C_3H_8 and NH_3 concentrations are Cr450 [V] and Cr470 [V]; the more important feature in the prediction of most of the gas concentrations is Cr470 [V]; while the less important features in the prediction of most of the gas concentrations are Au475 [V] and H545 [V].

4.1.3. Summary of the multicomponent gas mixture predictions

The ANN and XGBoost models predicted the gas concentrations with minimal error. For both models, the error is lowest (MAE, MdAE, maximum error, and RMSE) in the prediction of NO_2 concentration, followed by NO concentrations, but highest in the prediction of C_3H_8 concentration. The R^2 value for both ANN and XGBoost models is greater than 99%, and tshe mean absolute percentage error for the XGBoost model is less than that of the ANN model. Also, the R^2 value of the XGBoost model is higher.

Although both models predict the gas concentration with minimal error, the results suggest that the XGBoost model performed slightly better than the ANN model in the prediction of the gas concentrations. Furthermore, for each of the models, the more important features in the prediction of the gas concentrations is Cr450 [V] and Cr470 [V], while the less important features are Au475 [V] and H545 [V]. The results suggest that Cr450 and Cr470 sensors exhibit more interference (cross-sensitivity) with most of the gas concentrations, while Au475 and H545 appear to have very little or low interference with most of the gas concentrations (see Fig. 14).

4.2. Prediction of gas concentration in a binary gas mixture

4.2.1. Deep learning model prediction of the binary gas mixture

ANN model with a single output (gas concentration of NO_2 [ppm]) was developed for the binary gas mixture. The model predicted NO_2 concentration with mean absolute error, median absolute error, maximum error, and root mean square error of about 0.22 [ppm], 0.17 [ppm], 0.75 [ppm], and 0.28 [ppm], respectively; while the mean absolute percentage error and R^2 value are 2.18% and 99.98%, indicating that the model predicts gas concentration with good precision and over 99% of the variance in the NO_2 gas concentrations is explained by the model (Table 9). The median absolute error of 0.17 [ppm] is less than the mean absolute error and indicates that at least 50% of the model's predictions of gas concentrations have an absolute error of less than 0.17 [ppm]. Also, the negligible difference in the performance of the model based on the training dataset and testing dataset, confirms that the model is neither overfitting nor underfitting.

The maximum error in the prediction of NO_2 concentration is about

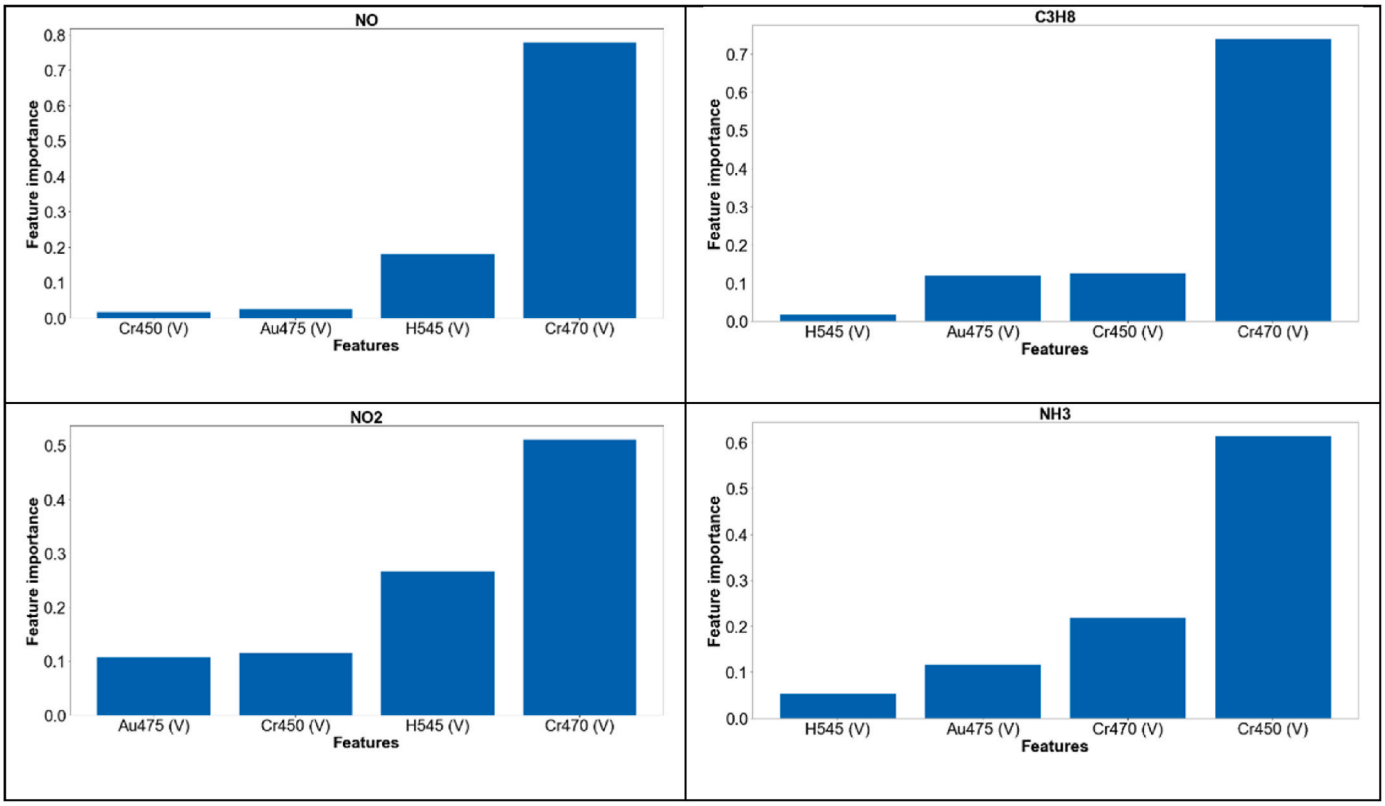


Fig. 13. XGBoost feature importance for each gas in the multicomponent mixture.

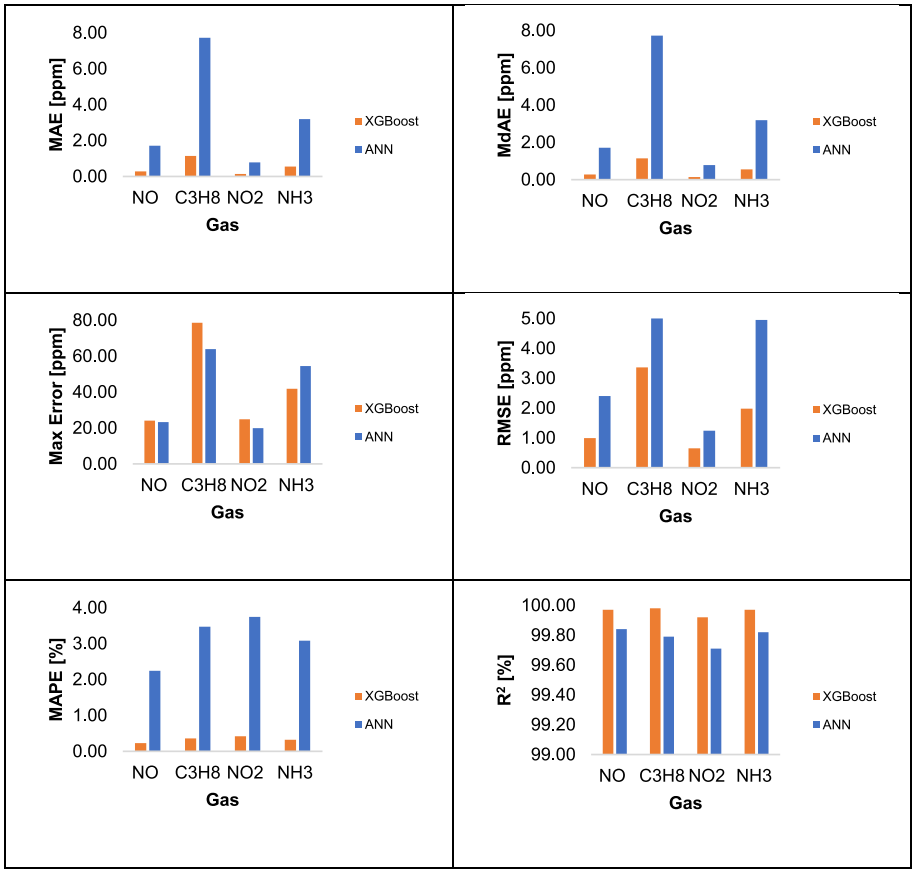


Fig. 14. Summary of the performance of XGBoost and ANN models for the multicomponent gas mixture.

Table 9
Performance measurements of ANN model for the binary gas mixture.

	Performance measures					
	R^2	MAPE [%]	MAE [ppm]	MdAE [ppm]	Max Error [ppm]	RMSE [ppm]
Overall training set	99.99	2.03	0.20	0.17	0.66	0.25
Overall testing set	99.98	2.18	0.22	0.17	0.75	0.28

0.75 [ppm], compared to the median absolute error of about 0.17. This suggests that the chance of predicting gas concentrations with an error up to 0.75 [ppm] is low. Therefore, the model predicts NO_2 concentration in a binary gas mixture with high precision.

The most important feature in the prediction of NO_2 concentration is Au475 [V], while the least important feature is Cr470 [V] as shown in Fig. 15. The feature importance of Cr470 [V] and Au475 [V] are very close (by absolute sum of weights), which implies that both sensors are sensitive to NO_2 concentrations (or cross-sensitivity).

4.2.2. *XGBoost model prediction of binary gas mixtures*

XGBoost model with a single output (gas concentration of NO_2 [ppm]) was developed for the binary gas mixture. The model predicted NO_2 concentration with mean absolute error, median absolute error, maximum error, and root mean square error of about 0.09 [ppm], 0.05 [ppm], 0.46 [ppm], and 0.14 [ppm], respectively; while the mean absolute percentage error and R^2 value are 0.52% and 100.00%, indicating that the model predicts gas concentration with minimal error and 100% of the variance in the NO_2 gas concentrations is explained by the model

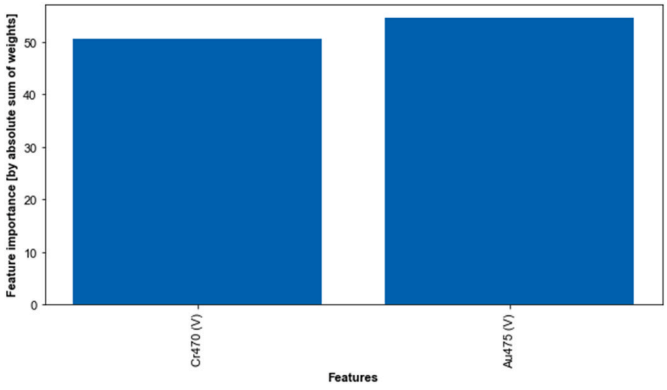


Fig. 15. Feature importance using ANN model for the binary gas mixture.

(Table 10). The median absolute error of 0.05 [ppm] is less than the mean absolute error and indicates that at least 50% of the model's predictions of NO_2 concentrations had an absolute error of less than 0.05 [ppm]. Also, the negligible difference in the performance of the model based on the training dataset and testing dataset, confirms that the model is neither overfitting nor underfitting.

The maximum error in the prediction of NO_2 concentration is about 0.46 [ppm], compared to the median absolute error of about 0.05. This suggests that the chance of predicting gas concentrations with an error up to 0.46 [ppm] is low. Therefore, the model predicts NO_2 concentration in a binary gas mixture with high precision.

The most important feature in the prediction of NO_2 concentration is Au475 [V], while the least important feature is Cr470 [V] as shown in Fig. 16. The feature importance of Au475 [V] is very high, while the

Table 10
Performance measurements of XGBoost model for the binary gas mixture.

	Performance measures					
	R^2	MAPE [%]	MAE [ppm]	MdAE [ppm]	Max Error [ppm]	RMSE [ppm]
Overall training set	100.00	0.41	0.06	0.05	0.40	0.09
Overall testing set	100.00	0.52	0.09	0.05	0.46	0.14

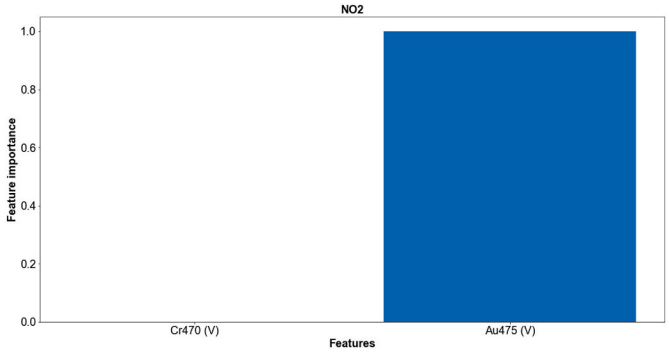


Fig. 16. Feature importance using XGBoost model for the binary gas mixture.

feature importance of Cr470 [V] is negligible. The result suggests that both sensors have very little or no interference (or cross-sensitivity) while taking measurements of NO_2 concentrations. The concentrations of NO_2 were mainly predicted by Au475 [V].

4.2.3. Summary of the binary gas mixture predictions

The ANN and XGBoost models predicted NO_2 concentrations, in the binary gas mixture, with minimal error. Also, for both models, the deviation between the metrics of performance evaluation for the training and testing datasets is negligible. However, the XGBoost model performed better (lower error values and higher R_2 value) than the ANN model in the prediction of the NO_2 concentrations (Fig. 17).

Furthermore, for each of the models, the most important feature in the prediction of the NO_2 concentrations is Au475 [V], while Cr470 [V] has very little impact on the prediction of the gas concentrations. The result suggests that both sensors have very little or no interference (or cross-sensitivity) while taking measurements of NO_2 concentrations. The concentrations of NO_2 were mainly predicted by Au475 [V].

5. Discussions

5.1. Gas concentration and feature importance for the multicomponent gas mixtures

To design a hybrid artificial intelligence system for real-time monitoring of gas concentrations, error limits must be set depending on the gas components. The XGBoost model performed slightly better than the ANN model in the prediction of each of the gas concentrations. Therefore, the performance of the XGBoost model would determine the gas concentration limits to set for the multicomponent gas mixtures.

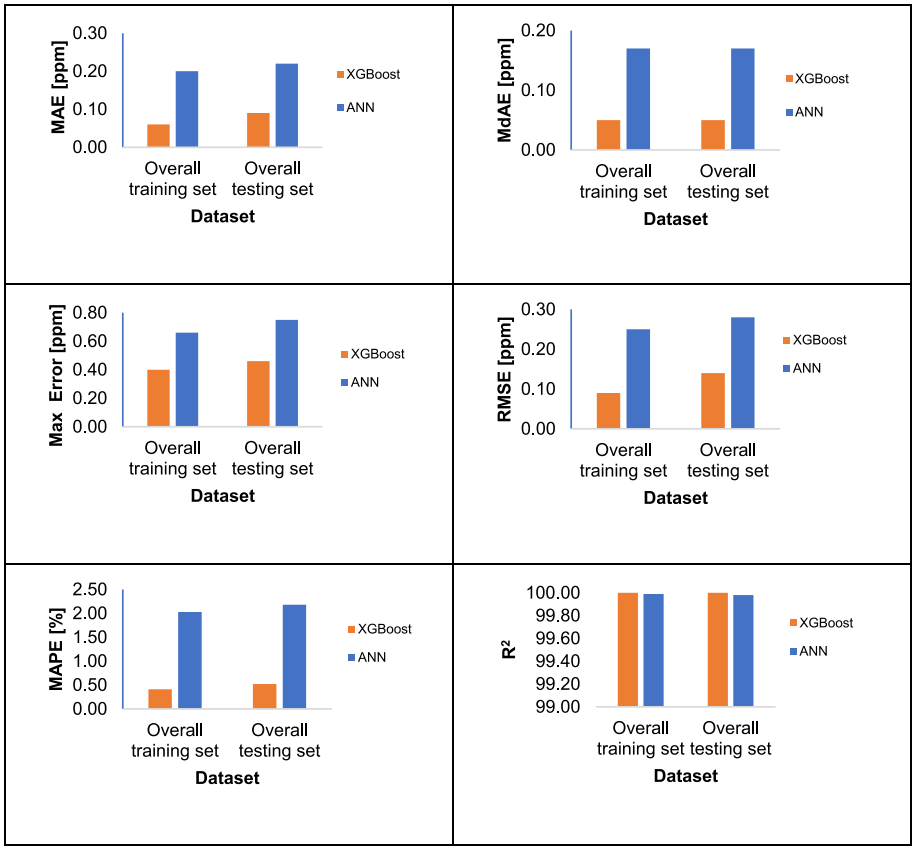


Fig. 17. Summary of the performance of XGBoost and ANN models for the binary gas mixture.

For the multicomponent gas mixture, the median absolute errors (MdAE) in the prediction of gas concentrations are less than the mean absolute errors (MAE). Therefore, setting the MAE of each gas component, as the error limit is reasonable. Over 50% of the model prediction of the component gas concentration might be less than the MAE. As already established, the maximum error cannot be used as the error limits for the multi-component gas mixture, as the chance of having up to tens of ppm (in gas concentrations) error in measurement is very low when compared with the median absolute error. Therefore, the error limits for NO, C_3H_8, NO_2, and NH_3 concentrations can be set as 0.28 [ppm], 1.14 [ppm], 0.14 [ppm], and 0.55 [ppm], respectively. In other words, $\varepsilon_{NO} = 0.28$ [ppm], $\varepsilon_{C3H8} = 1.14$ [ppm], $\varepsilon_{NO2} = 0.14$ [ppm], and $\varepsilon_{NH3} = 0.55$ [ppm].

Furthermore, the most important feature in the prediction of the NO and NO_2 concentrations is Cr470 [V], and the most important features in the prediction of C_3H_8 concentrations are Cr470 [V] and Cr450 [V]. These findings align with previous studies that confirmed that the pairing Pt and $La_{0.8}Sr_{0.2}CrO_3$ (LSCO) have high sensitivity to hydrocarbons (such as C_3H_8 and C_3H_6) at open circuit, and nitrogen oxides (NO and NO_2) under bias (Tsui et al., 2017; Javed et al., 2022). However, the less important features in the prediction of NH_3 concentrations are H545 [V] and Au475 [V]; while, the most important feature is Cr450 [V], followed by Cr470 [V]. This finding did not conform completely to the submission of Tsui et al. (2017) that sensors pairing Pt and gold (Au) electrodes or Au alloys such as Au and Palladium (Pd) which can be represented as (Au/Pd + Pt) are sensitive to both carbon (II) oxide (CO) and ammonia. Javed et al. (2022) also stated that Au/Pd + Pt sensors are highly sensitive to NH_3 in the presence of C_3H_8, CO, NO, and NO_2. The difference in the results might be attributed to the gas environment, of which in this study the gas mixtures are NO, C_3H_8, NO_2, and NH_3.

5.2. Gas concentration and feature importance for the binary gas mixture

The XGBoost model performed slightly better than the ANN model in the prediction of each of the NO_2 concentrations in the binary gas mixture. Therefore, the performance of the XGBoost model would determine the NO_2 concentration limits to set for the binary gas mixture. The MdAE is less than the MAE, and the maximum error is relatively low compared to those of the multi-component gas mixtures. Therefore, one can decide to set the error limit in the gas concentration of NO_2 based on the maximum error. Thus, the error limit (worst case) in the measurement of NO_2 concentration is 0.46 [ppm] ($\varepsilon_{NO2} = 0.46$ [ppm] for the binary gas mixture).

Furthermore, for the binary gas mixture, the most important feature in the prediction of the NO_2 concentrations is Au475 [V], while Cr470 [V] has very little impact on the prediction. The result suggests that both sensors have very little or no interference (or cross-sensitivity) while taking measurements of NO_2 concentrations. The Au475 sensor is more sensitive to NO_2 concentrations in the binary gas mixture compared to the Cr470 sensor. This result is different from the prediction of NO_2 concentrations in the multicomponent gas mixture, where the most sensitive sensor to NO_2 in the presence of the other gas impurities is Cr470. Therefore, the sensitivity of gas sensors to gas concentrations highly depends on the gas components in the environment where the sensor is deployed.

5.3. Design of the hybrid artificial intelligence system for the calibration of sensors

In the design of a hybrid artificial intelligence system, a new array of sensors can be calibrated to improve its functionality of a physical quantity response (say, voltage or resistance change) to measure the corresponding concentrations of the gas components in the mixture. The set error limits (ε_i), based on the machine learning model, would be programmed in the design of the wireless device (for example, a cell phone) or mobile application in a wireless device, recording the sensor (s) measurements and machine learning model predictions from a cloud platform.

When the errors (or deviations) in the individual component gas concentration measurements (or predictions) are more than the set limits for almost 50 out of 100 (50%, it could be higher or lower depending on the performance of the machine learning model developed) measurements or predictions over a certain period (say a day or a week), for different values of the physical quantities (say, voltage or resistance change readings) measured by the sensor or array of sensors, then the sensor(s) should be recalibrated or replaced with properly calibrated sensors (in the case of mechanical wear or damage).

For a more reliable sensor performance, the higher functionality array of sensors designed based on the experiments and/or numerical simulations can make up a suite of sensors, while another suite of sensors will maintain their original function of recording their response to a physical quantity (for instance voltage or resistance change). This way, one can be sure that the physical quantity response of the sensors from which the machine learning model is predicting the gas concentrations is accurate.

The approach of designing a hybrid artificial intelligence system in a nitrogen gas (N_2) facility in the present study can be extended to CO_2 gas facilities used for CO_2 capture and storage. During the sensor design experiments and/or numerical simulations, CO_2 becomes the balance gas, while the gas impurities represent the analyte gases such as NO_2, H_2S, or SO_2, depending on the gas environment. A hybrid artificial intelligence system would enable the concentrations of the gas impurities in the CO_2 stream to be monitored in real-time (and remotely) from a wireless device such as a cell phone or an application in a wireless device. Based on the error limits set in the measurement of gas concentrations in the multicomponent gas mixture from the sensor and machine learning model (both monitored on the wireless device), one can tell when the sensors need to be recalibrated or replaced.

In addition, when the sensor and machine learning model measurements are accurate, one can determine when certain allowable gas concentration thresholds for some gas components have been exceeded. For instance, in a CO_2 transportation facility where SO_2 concentration up to 2.5 mol% (or equivalent concentration in ppm) is dangerous for CO_2 storage in some reservoirs (Aminaho et al., 2024), the hybrid artificial intelligence system monitored remotely can help responsible users advise stoppage (or shutdown) of the CO_2 gas (mixture) transportation facility for storage. Similarly, in a separation chamber, the hybrid system would be able to monitor the gas concentrations in the chamber at any time and help to decide to stop the cycle of CO_2 capture as a reasonable amount of CO_2 has been captured.

The detection of deviations in the sensor readings and machine learning model prediction of gas concentration is much easier for a binary gas mixture (considering the relatively low maximum error in a binary gas mixture compared to the multi-component gas mixtures). Therefore, it would be easier to tell when sensors need to be recalibrated from the hybrid artificial intelligence system. Nonetheless, the hybrid artificial intelligence systems would work well for a multi-component gas mixture, as long as the machine learning model is properly trained (high-performance machine learning model), as in the present study.

6. Conclusions

This study proposes a hybrid artificial intelligence system for real-time monitoring of gas concentrations in a gas stream and to identify when sensors need to be recalibrated. This system employs a machine learning-assisted method of calibrating or recalibrating sensors and the selection of arrays of sensors to minimize the challenge of sensors' interference or cross-sensitivity in measuring gas concentrations. This system also provides remote access to monitoring gas concentrations predicted by a machine learning model and sensor readings programmed in a wireless device or an application in a wireless device, enabling users to identify when certain set thresholds of gas

concentrations are exceeded for quality control and assurance. This study demonstrated the feasibility of operating the hybrid artificial intelligence system in a nitrogen gas stream and suggested how the system can be replicated for carbon dioxide capture and storage. Based on the findings of this study, the following conclusions are made:

1. The machine learning models predicted gas concentrations with minimal error. However, the XGBoost model performed slightly better than the ANN model in predicting the gas concentrations.
2. The machine learning models predicted gas concentrations with the least error in NO_2 concentration and the highest error in C_3H_8 concentration, for the multi-component gas mixtures.
3. The error in the machine learning models for the binary gas mixture is significantly less than the errors for the multi-component gas mixtures, as there is no challenge of cross-sensitivity between Cr470 and Au475 sensors.
4. Maximum error in the prediction of each gas concentration in a multi-component mixture might not be suitable criteria to set error limit for sensor recalibration, as the chance of having deviation in measurements up to the maximum error is very low. Instead, the mean or median absolute error (whichever is higher) might be a more suitable error limit for each gas concentration. However, in a binary gas mixture, the maximum error (or deviation) in the model's prediction of gas concentration might be a suitable error limit for the recalibration of sensors.
5. The machine learning-assisted design of sensors helps in determining which sensors are not required in an array of sensors, by exploring their behaviour in a correlation matrix. A perfect correlation between the responses of two different sensors indicates that one of the sensors should be dropped to minimize the challenge of sensors' interference or cross-sensitivity in the machine learning model.
6. For the multi-component gas mixtures, the more important features in the prediction of gas concentrations are Cr470 [V] and Cr450 [V], while the less important features are Au475 [V] and H545 [V]. The relatively high importance of Cr470 [V] and Cr450 [V] might be attributed to LSCO sensors having high sensitivity to NO, NO_2, and C_3H_8 gas concentrations under certain conditions.
7. In the binary gas mixture, unlike in the multicomponent gas mixture, the most important feature in the prediction of NO_2 concentrations is Au475 [V], while Cr470 [V] have very low impact in the prediction. Therefore, the gas components in the environment where sensors are deployed highly impact their sensitivity, due to gas interference.
8. The hybrid artificial intelligence system proposed in this study would perform better with a well-trained machine learning model and two suites of sensors (a suite containing an array of sensors with full functionality including measuring the normal sensor response [such as voltage change] and the gas concentrations, and another suite with an array of sensors to measure only the sensor response [say, voltage change] to a physical quantity only). This is because the machine learning models depend on the normal sensor response to a physical quantity to predict the gas concentrations. Having two suites of sensors will ensure the sensors' responses are reliable and that the machine learning model is receiving the right sensor responses for the prediction of gas concentrations.

CRediT authorship contribution statement

Efenwengbe Nicholas Aminaho: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Ndukaegho Sabastine Aminaho:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Mamdud Hossain:** Writing – review & editing, Supervision. **Nadimul Haque Faisal:** Writing – review & editing, Supervision. **Konyengwaehie Augustus Aminaho:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization.

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 paper.

Acknowledgements

The authors would like to acknowledge the funding support by Robert Gordon University, United Kingdom to conduct this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jgsce.2025.205546.

Data availability

Data will be made available on request.

References

Acharyya, S., Nag, S., Guha, P.K., 2022. Ultra-selective tin oxide-based chemiresistive gas sensor employing signal transform and machine learning techniques. Anal. Chim. Acta 1217, 339996. https://doi.org/10.1016/j.aca.2022.339996.

Agunbiade, T.O., Oluwadare, O.A., Amusan, R.O., 2023. Reservoir characterization for CO_2 sequestration and CO_2-enhanced oil recovery techniques in maje field, offshore Niger-delta, Nigeria. In: Paper SPE-217191-MS Presented at the SPE Nigeria Annual International Conference and Exhibition Held in Lagos, Nigeria 31 July - 2 August 2023. https://doi.org/10.2118/217191-MS.

Aminaho, E.N., Hossain, M., Faisal, N.H., Sanaee, R., 2024. Formation integrity evaluation for geosequestration of CO_2 in depleted petroleum reservoirs under cyclic stress conditions. Geoenergy Science and Engineering 238, 212892. https://doi.org/10.1016/j.geoen.2024.212892.

Aminaho, E.N., Hossain, M., 2023. Caprock Integrity Evaluation for Geosequestration of CO_2 in Low-Temperature Reservoirs. Robert Gordon University, Aberdeen. https://rgu-repository.worktribe.com/output/2072081.

Aminu, M.D., Nabavi, S.A., Manovic, V., 2018. CO_2-brine-rock interactions: the effect of impurities on grain size distribution and reservoir permeability. Int. J. Greenh. Gas Control 78, 168–176.

Antanasijević, D., Pocajt, V., Perić-Grujić, A., Ristić, M., 2018. Multiple-input–multiple-output general regression neural networks model for the simultaneous estimation of traffic-related air pollutant emissions. Atmos. Pollut. Res. 9, 388–397.

Barcelo-Ordinas, J.M., Doudou, M., Garcia-Vidal, J., Badache, N., 2019. Self-calibration methods for uncontrolled environments in sensor networks. A reference survey. Ad Hoc Netw. 88, 142–159. https://doi.org/10.48550/arXiv.1905.11060.

Barriault, M., Alexander, I., Tasnim, N., O'Brien, A., Najjaran, H., Hoorfar, M., 2021. Classification and regression of binary hydrocarbon mixtures using single metal oxide semiconductor sensor with application to natural gas detection. Sensors and Actuators: B. Chemical 326, 129012. https://doi.org/10.1016/j.snb.2020.129012.

Basyooni, M.A., Zaki, S.E., Tihtih, M., Boukhoubza, I., En-nadir, R., Attia, G.F., 2024. Fundamentals and classifications of CO_2 sensors. In: Ali, G.A.M., Chong, K.F., Makhlouf, A.S.H. (Eds.), Handbook of Nanosensors. Springer, Cham. https://doi.org/10.1007/978-3-031-16338-8_22-1.

Basyooni, M.A., Shaban, M., El Sayed, A.M., 2017. Enhanced gas sensing properties of spin-coated Na-doped ZnO nanostructured films. Sci. Rep. 7 (1), 1–12. https://doi.org/10.1038/srep41716.

Bauke, S., Golibrzuch, K., Wackerbarth, H., Fendt, P., Zigan, L., Seefeldt, S., Thiele, O., Berg, T., 2018. Optical sensor system for time-resolved quantification of methane concentrations: validation measurements in a rapid compression machine. J. Quant. Spectrosc. Radiat. Transfer 210, 101–110.

Bellegoni, M., Ovidi, F., Tempesti, L., Mariotti, A., Tognotti, L., Landucci, G., Galletti, C., 2022. Optimization of gas detectors placement in complex industrial layouts based on CFD simulations. J. Loss Prev. Process. Ind. 80, 104859. https://doi.org/10.1016/j.jlp.2022.104859.

Bolourinejad, P., Herber, R., 2014. Experimental and modelling study of storage of CO_2 and impurities in a depleted gas field in northeast Netherlands. Energy Proc. 63, 2811–2820.

Bouaboula, H., Chaouki, J., Belmabkhout, Y., Zaabout, A., 2024. Comparative review of Direct air capture technologies: from technical, commercial, economic, and environmental aspects. Chem. Eng. J. 484, 149411. https://doi.org/10.1016/j.cej.2024.149411.

Casey, J.G., Collier-Oxandale, A., Hannigan, M., 2019. Performance of artificial neural networks and linear models to quantify 4 trace gas species in an oil and gas production region with low-cost sensors. Sensor. Actuator. B 283, 504–514.

Chen, K.Y., Janz, K.F., Zhu, W., Brychta, R.J., 2012. Redefining the roles of sensors in objective physical activity monitoring. Med. Sci. Sports Exerc. 44 (1), 13–23. https://doi.org/10.1249/MSS.0b013e3182399bc8.

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science 7, 1–24. https://doi.org/10.7717/peerj-cs.623.

David, O., Dosunmu, A., 2011. Carbon capture for enhanced oil recovery in Niger delta: a case study of okota/okpoputa field in offshore Niger delta. In: Paper SPE 145049 Presented at the SPE Enhanced Oil Recovery Conference Held in Kuala Lumpur, Malaysia, 19–21 July 2011.

Dhall, S., Mehta, B.R., Tyagi, A.K., Sood, K., 2021. A review on environmental gas sensors: materials and technologies. Sensors International 2, 100116. https://doi.org/10.1016/j.sintl.2021.100116.

Diemel, O., Honza, R., Ding, C.-P., Böhm, B., Wagner, S., 2019. In situ sensor for cycle-resolved measurement of temperature and mole fractions in IC engine exhaust gases. Proc. Combust. Inst. 37, 1453–1460.

Djedidi, O., Djeziri, M.A., Morati, N., Seguin, J.-L., Bendahan, M., Contaret, T., 2021. Accurate detection and discrimination of pollutant gases using a temperature modulated MOX sensor combined with feature extraction and support vector classification. Sensors and Actuators: B. Chemical 339, 129817. https://doi.org/10.1016/j.snb.2021.129817.

Ghommem, M., Puzyrev, V., Sabouni, R., Najar, F., 2022. Deep learning for gas sensing using MOFs coated weakly-coupled microbeams. Appl. Math. Model. 105, 711–728.

He, Y., Zhou, Y., Wen, T., Zhang, S., Huang, F., Zou, X., Ma, X., Zhu, Y.Q., 2022. A review of machine learning in geochemistry and cosmochemistry: method improvements and applications. Appl. Geochem. 140, 1–13. https://doi.org/10.1016/j.apgeochem.2022.105273.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Paper Presented at IEEE International Conference on Computer Vision Held in Santiago, Chile, 7-13 December 2015, pp. 1026–1034. https://doi.org/10.1109/ICCV.2015.123.

Javed, U., Ramaiyan, K.P., Kreller, C.R., Brosha, E.L., Mukundan, R., Sengupta, A.M., Morozov, A.V., 2022. Quantification of gas concentrations in $NO/NO_2/C_3H_8/NH_3$ mixtures using machine learning. Sensors and Actuators: B. Chemical 359, 131589. https://doi.org/10.1016/j.snb.2022.131589.

Jeong, N., Chung, T.-H., Tong, T., 2021. Predicting micropollutant removal by reverse osmosis and nanofiltration membranes: is machine learning viable? Environ. Sci. Technol. 55 (16), 11348–11359. https://doi.org/10.1021/acs.est.1c04041.

Kannaiah, P.V.D., Maurya, N.K., 2023. Machine learning approaches for formation matrix volume prediction from well logs: insights and lessons learned. Geoenergy Science and Engineering 229, 1–20. https://doi.org/10.1016/j.geoen.2023.212086.

Karbassiyazdi, E., Fattahi, F., Yousefi, N., Tahmassebi, A., Taromi, A.A., Manzari, J.Z., Gandomi, A.H., Altaee, A., Razmjou, A., 2022. XGBoost model as an efficient machine learning approach for PFAS removal: effects of material characteristics and operation conditions. Environ. Res. 215, 114286. https://doi.org/10.1016/j.envres.2022.114286.

Lagutin, A.S., Vasil'ev, A.A., 2022. Solid-state gas sensors. J. Anal. Chem. 77 (2), 131–144.

Martinez, R.R., Santaren, D., Laurent, O., Cropley, F., Mallet, C., Ramonet, M., Caldow, C., Rivier, L., Broquet, G., Bouchet, C., Juery, C., Ciais, P., 2021. The potential of low-cost tin-oxide sensors combined with machine learning for estimating atmospheric CH_4 variations around background concentration. Atmosphere 12 (1), 107. https://doi.org/10.3390/atmos12010107.

McGrath, M.J., Scanaill, C.N., 2013. Sensing and sensor fundamentals. In: Sensor Technologies. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4302-6014-1_2.

McNaughton, B.A.P., 2019. Two-dimensional ising spin model and machine. Learning. https://doi.org/10.13140/RG.2.2.33154.43201.

Metz, B., Davidson, O., de Coninck, H., Loos, M., Meyer, L., 2005. IPCC Special Report on Carbon Dioxide Capture and Storage. Cambridge University Press, UK.

Nagireddi, S., Agarwal, J.R., Vedapuri, D., 2024. Carbon dioxide capture, utilization, and sequestration: current status, challenges, and future prospects for global decarbonization. ACS Eng. Au 4, 22–48. https://doi.org/10.1021/acsengineeringau.3c00049.

Narayana, T.L., Venkatesh, C., Kiran, A., Babu, C., Kumar, A., Khan, S.B., Almusharraf, A., Quasim, M.T., 2024. Advances in real time smart monitoring of environmental parameters using IoT and sensors. Heliyon 10, e28195. https://doi.org/10.1016/j.heliyon.2024.e28195.

Noor, A.E., Ahmad, I., Rasheed, A., Ramzan, S., Rahim, A., Khan, A., Tahir, M.Y., 2024. Electrochemical sensors and their types. In: Ahmad, A., Verpoort, F., Khan, A., Ali, S. (Eds.), Micro and Nano Technologies, Nanomaterials-Based Electrochemical Sensors: Properties, Applications and Recent Advances. Elsevier, pp. 77–87. https://doi.org/10.1016/B978-0-12-822512-7.00012-0.

Pearce, J.K., Kirste, D.M., Dawson, G.K.W., Rudolph, V., Golding, S.D., 2019. Geochemical modelling of experimental O_2–SO_2–CO_2 reactions of reservoir, cap-rock, and overlying cores. Appl. Geochem. 109, 1–19.

Profillidis, V.A., Botzoris, G.N., 2024. Statistical methods for transport demand modeling. In: Modeling of Transport Demand. Elsevier, pp. 163–224. https://doi.org/10.1016/B978-0-12-811513-8.00005-4.

Rusdah, D.A., Murfi, H., 2020. XGBoost in handling missing values for life insurance risk prediction. SN Appl. Sci. 2, 1–10. https://doi.org/10.1007/s42452-020-3128-y.

Saharudin, D.M., Jeswani, H.K., Azapagic, A., 2023. Bioenergy with carbon capture and storage (BECCS): life cycle environmental and economic assessment of electricity generated from palm oil wastes. Appl. Energy 349, 121506. https://doi.org/10.1016/j.apenergy.2023.121506.

Scherer, P.O.J., 2013. Computational Physics: Simulation of Classical and Quantum Systems, second ed. Springer International Publishing Switzerland. https://doi.org/10.1007/978-3-319-00401-3.

Schultealbert, C., Baur, T., Schütze, A., Böttcher, S., Sauerwald, T., 2017. A novel approach towards calibrated measurement of trace gases using metal oxide semiconductor sensors. Sensor. Actuator. B 239, 390–396.

Shindell, D., Faluvegi, G., Walsh, M., Anenberg, S.C., Dingenen, R.V., Muller, N.Z., Austin, J., Koch, D., Milly, G., 2011. Climate, health, agricultural and economic impacts of tighter vehicle-emission standards. Nature Nature Climate Change 1, 59–66. https://doi.org/10.1038/nclimate1066.

Shwetha, H.R., Sharath, S.M., Guruprasad, B., Rudraswamy, S.B., 2022. MEMS based metal oxide semiconductor carbon dioxide gas sensor. Micro and Nano Engineering 16, 100156. https://doi.org/10.1016/j.mne.2022.100156.

Singh, O.P., Malarvili, M., 2018. Review of infra red carbon-dioxide sensors and Cpnogram features for developing asthma-manufacturing device. J. Clin. Diagn. Res. 12 (10), 1–6. https://doi.org/10.7860/JCDR/2018/35870.12099.

Stankova, M., Vilanova, X., Calderer, J., Llobet, E., Ivanov, P., Gràcia, I., Cané, Correig, X., 2004. Detection of SO_2 and H_2S in CO_2 stream by means of WO_3-based micro-hotplate sensors. Sensor. Actuator. B 102, 219–225.

TE Connectivity, 2020. Sensors for Corrosive Environment.

Thekkethil, S.R., Roy, K.E.R., Thomas, R.J., Neumann, H., Ramalingam, R., 2016. Mathematical model for a novel cryogenic flow sensor using fibre bragg gratings. Procedia Technology 24, 477–484.

Tsui, L.-K., Benavidez, A., Palanisamy, P., Evans, L., Garzon, F., 2017. Quantitative decoding of the response a ceramic mixed potential sensor array for engine emissions control and diagnostics. Sensor. Actuator. B 249, 673–684.

Vilanova, X., Llobet, E., Brezmes, J., Calavia, R., Sanchez, X., Correig, X., 2003. A multisensor system for monitoring the quality of carbon dioxide in the beverage industry. In: Paper Presented at the 12th International Conference on Solid State Sensors, Actuators and Microsystems, Boston, MA,USA, June 8-12, 2003. https://doi.org/10.1109/SENSOR.2003.1217023.

Walsh, P.T., Gant, S.E., Dowker, K.P., Batt, R., 2011. Response of electrochemical oxygen sensors to inert gas–air and carbon dioxide–air mixtures: measurements and mathematical modelling. J. Hazard Mater. 186, 190–196.

Wang, J., Lian, S., Lei, B., Li, B., Lei, S., 2022. Co-training neural network-based infrared sensor array for natural gas monitoring. Sensors and Actuators: A. Physical 335, 113392. https://doi.org/10.1016/j.sna.2022.113392.

Xu, Y., Pu, H., Li, Y., Wang, H., 2023. Flow pattern identification for gas-oil two-phase flow based on a virtual capacitance tomography sensor and numerical simulation. Flow Meas. Instrum. 92, 102376. https://doi.org/10.1016/j.flowmeasinst.2023.102376.

Yao, P., Yu, Z., Zhang, Y., Xu, T., 2023. Application of machine learning in carbon capture and storage: an in-depth insight from the perspective of geoscience. Fuel 333, 1–13. https://doi.org/10.1016/j.fuel.2022.126296.

Yoro, K.O., Daramola, M.O., 2020. CO_2 emission sources, greenhouse gases, and the global warming effect. In: Rahimpour, M.R., Farsi, M., Makarem, M.A. (Eds.), Advances in Carbon Capture Methods, Technologies and Applications. Woodhead Publishing, Cambridge, pp. 3–28. https://doi.org/10.1016/B978-0-12-819657-1.00001-3.

Zaki, S.E., Basyooni, M.A., Shaban, M., Rabia, M., Eker, Y.R., Attia, G.F., Yilmaz, M., Ahmed, A.M., 2019. Role of oxygen vacancies in vanadium oxide and oxygen functional groups in graphene oxide for room temperature CO_2 gas sensors. Sensor Actuator Phys. 294, 17–24. https://doi.org/10.1016/j.sna.2019.04.037.

Zhang, W., Xu, T., Li, Y., 2011. Modeling of fate and transport of co-injection of H_2S with CO_2 in deep saline formations. J. Geophys. Res. 116, B02202. https://doi.org/10.1029/2010JB007652.

Zhang, R., Xia, J., Ahmed, I., Gough, A., Armstrong, I., Upadhyay, A., Fu, Y., Enemali, G., Lengden, M., Johnstone, W., Wright, P., Ozanyan, K., Pourkashanian, M., McCann, H., Liu, C., 2023. A fast sensor for non-intrusive measurement of concentration and temperature in turbine exhaust. Sensors and Actuators: B. Chemical 396, 134500. https://doi.org/10.1016/j.snb.2023.134500.