DOI: 10.15849/IJASCA.250330.17

Int. J. Advance Soft Compu. Appl, Vol. 17, No. 1, March 2025
Print ISSN: 2710-1274, Online ISSN: 2074-8523

An intelligent indoor air quality monitoring system

Sanaa Mouhim

InterDisciplinary Applied Research Laboratory - LIDRA, International University of Agadir - Universiapolis, Agadir, Morocco, ID:60268804.
e-mail: Sanaa.mouhim@e-polytechnique.ma

Abstract

The Internet of Things (IoT) is experiencing significant market growth thanks to advancements in electronic sensors, microprocessors, communication, and information media. With billions of devices interconnected , the potential for IoT applications is immense. This paper proposes an architecture for air quality monitoring systems based on IoT and Big Data technologies, providing tools, resources, and functionalities related to embedded systems, communication, processing, storage, data analysis, web, business, and security features. Furthermore, integrating IoT with Big Data enables the application of predictive models and machine learning algorithms to anticipate pollution fluctuations and proactively adjust ventilation systems. This predictive intelligence not only enhances occupant comfort and health by maintaining optimal indoor air quality but also improves energy efficiency by adjusting ventilation to meet actual needs.

Keywords-- air quality, big data, energy efficiency, intelligent systems, Internet of Things (IoT), predictive models.

1 Introduction

The research significance of developing an indoor air quality monitoring system using the Internet of Things and big data technologies lies in the health preservation of indoor occupants. The high level of outdoor and indoor pollutants, the long duration with which people stay indoors, and the negative correlation between ventilation and air conditioning usage result in direct impacts on indoor occupants' health and energy consumption. While other types of air pollution occur outside, indoor pollution has recently gained momentum because of the long urbanization process, densification of the population, increased amounts of building waste, unhealthy construction materials, established ventilation regulations, and famous exhaust systems that have been applied to buildings, contributing considerably to the deterioration of indoor air quality. Respiratory illnesses, including congestion, coughing, and dry skin, are caused by filthy air; this may be tied to the fact that humans spend 90% of their lives in indoor environments [1], [2].

To improve the quality of interior life, indoor air quality (IAQ) systems are devices or sets of devices designed to monitor, analyze and improve the air quality inside buildings. These systems are increasingly used in homes, offices, schools and other indoor environments to ensure healthy air and prevent health problems. To improve air quality

Received 7 January 2025; Accepted 20 March 2025

in indoor environments, several metrics are used, such as the CO2 concentration, measurement of fine airborne particles, detection of toxic gases such as CO or O3, measurement of temperature and humidity to monitor thermal comfort, and prevention of mold growth [3]. Therefore, real-time monitoring is needed to achieve timely detection and correction of pollution episodes to improve IAQ, and data analysis is essential for identifying the causes and their effective solutions. In contrast, the IoT allows devices to communicate and share data, whereby it has revolutionized domestic infrastructure, making our homes more energy-efficient, functional, and safe [4]. The information given by IoT devices, once processed by big data systems, enables the identification of trends and forecasts of changes in air quality and the automatic triggering of actions such as ventilation on/off or alerts. This, of course, requires special attention to data security and scalability to enable smooth integration and secure usage in different environments.

The main question we are trying to answer in this paper is how IoT, artificial intelligence models and big data-based systems can be used for predicting fluctuations in indoor air quality while optimizing energy efficiency in smart buildings. Systems can then proactively readjust heating, ventilation, and air conditioning (HVAC) systems, reducing energy consumption while maintaining optimal conditions for occupant health. This approach not only improves the energy sustainability of intelligent buildings but also enables more reactive and autonomous management of air control systems, transforming the way buildings respond to internal environmental changes.

We start our analysis with a literature review of QAI monitoring systems using IoT technologies and big data platforms. We explore the architectures proposed, the challenges faced and the solutions adopted to enable effective monitoring. This review also highlights systems and sensors employing air samples, as well as analytical tools for improving the indoor air quality. We then propose the predictive models that can be integrated into the system in each specific case as well as the system hardware and software architecture.

2 Literature review

The article “Development of an IoT-Based Indoor Air Quality Monitoring Platform” [4] describes the development of Smart-Air, an IoT platform for real-time indoor air quality monitoring that uses sensors to measure various pollutants. Data are analyzed and visualized via an AWS-based cloud server, with alerts in the event of moderate or poor air quality. With respect to the integration of machine learning models, this article proposes the use of machine learning techniques to improve the accuracy of air quality forecasts.

The article “Sensing Data Fusion for Enhanced Indoor Air Quality Monitoring” [5] focuses on the development of an air quality management system designed for smart buildings that integrates multi sensor data to enhance indoor air quality (IAQ) monitoring. The authors propose an approach that merges the indoor air quality index (IAQI) with humidex to create an enhanced indoor air quality index (EIAQI). This system uses a network of waspmote sensors to measure various indoor air pollutants in real time. Real-time monitoring is utilized through the deployment of a network of waspmote sensors that continuously measure various indoor air pollutants and environmental parameters. The real-time data collected from the sensors are processed via an extended fractional-order Kalman filter (EFKF).

The article "A comprehensive review on indoor air quality monitoring" [6] provides an extensive overview of the current state of indoor air quality (IAQ) monitoring systems. Several machine learning models are mentioned as potential tools for predicting future air quality conditions. Attention has focused mainly on long short-term memory (LSTM) and gated recurrent unit (GRU) architectures, which are recurrent neural network (RNN) architectures that are notable for their ability to analyze patterns in historical air quality data and make reliable predictions about future conditions.

The article "Building an indoor air quality monitoring system based on the architecture of the Internet of Things" [7] presents the development of a comprehensive indoor air quality monitoring system that utilizes IoT technology. This study incorporates fuzzy logic control mechanisms to manage the indoor environment on the basis of the collected data. Unfortunately it does not delve into predictive analytics or forecast future air quality conditions.

The article titled "An Improvement Strategy for Indoor Air Quality Monitoring Systems" [8] introduces two main decision-making algorithms that consider measurement uncertainty, allowing for more reliable assessments and interventions in indoor environments: the utility cost test algorithm, which evaluates decisions under conditions of uncertainty by considering the potential consequences of those decisions, referred to as utilities, and the fixed risk algorithm, which is designed to manage the risk associated with exceeding thresholds.

The article "Establishment of Smart Living Environment Control System" [9] proposes several key technologies for the establishment of a smart living environment control system, such as the IoT, sensing components, cloud computing and mobile technologies. The article does not explicitly mention specific AI tools used for prediction within the smart living environment control system. However, it does refer to the integration of various advanced technologies, such as big data and cloud computing, which are often associated with AI applications.

In Table 1, we present a comparison of several studies on air quality monitoring systems, analyzing various aspects, such as the metric used, energy efficiency, sampling rate, storage, compliance with standards and the predictive capability of the models.

Table 1 Comparison of studies on air quality monitoring systems on the basis of metrics, energy efficiency, sampling rate, storage, norms and prediction algorithms

article	Metrics used	Energy Efficiency	Sampling rate	Storage	Standards compliance	Prediction
[4]	Carbon dioxide (CO_2), Volatile organic compounds (VOCs), Aerosols, Temperature and humidity	✓	Not specified	Amazon Web Services	Indoor Air Quality Control Act (Korea)	Not invoked
[5]	Indoor Air Quality Index (IAQI), Humidex, Carbon monoxide (CO), Carbon dioxide (CO_2), Ammonia (NH_3), Hydrogen (H_2), Hydrogen sulfide (H_2S), Volatile organic compounds (VOCs) such as ethanol (C_2H_6O) and toluene (C_7H_8), Suspended particulates	Not explicitly invoked	Not specified	Not specified	Not specified	fractional extended Kalman filter (FEKF), used to manage uncertainties, noise and missing measurements that can affect prediction performance
[6]	Particulate Matter (PM), Carbon Dioxide (CO_2),	✓	Not specified	Web Servers	Not specified	Long Short-Term Memory (LSTM)

	Volatile Organic Compounds (VOC), Temperature and Relative Humidity (RH), Air Quality Indices (AQI), Ozone (O3)					and Gated Recurrent Unit (GRU). anticipate indoor pollution levels based on historical data.
[7]	Particulate Matter (PM2.5), Carbon Dioxide (CO2), Carbon Monoxide (CO), Air Quality Index (AQI)	Not explicitly invoked	30 seconds.	Terminal computer	ASHRAE[1]	vector Kalman filter, to predict pollutant concentrations in indoor air
[8]	Carbon dioxide (CO2), Total volatile organic compounds (TVOCs), Fine particulate matter (PM2.5)	✓	Not specified	Cloud	Not specified	Not specified
[9]	PM2.5 and PM10, Harmful gases, Air Quality Indices	Not explicitly invoked	Not specified	Google Cloud Drive,	ISO 7730 et ISO 10551	Not invoked
[10]	Air Quality Indices (AQI), Carbon monoxide (CO), Methane, Liquefied Petroleum Gas (LPG), Smoke, Temperature and Humidity, Meteorological Data	Not explicitly invoked	Not specified	ThingSpeak	Not specified	Long-term memory neural network (LSTM) to account for temporal variations in pollution levels

As shown in Table 1, intelligent indoor air quality systems still have diverse approaches and methodologies. While a few studies have matured enough to use methods such as AI and predictive models, others still focus only on pollutant measurement alone while neglecting other factors (e.g., energy efficiency and standard compliance). An analysis of the key points is as follows:

- **Metrics used:** Most items measure air quality indices, including fine particulate matter (PM2.5 and PM10), carbon dioxide (CO2), carbon monoxide (CO), volatile organic compounds (VOCs), ozone (O3), and other indicators such as humidity and temperature. These measurements are essential for assessing indoor and outdoor air quality. Some articles include additional measurements, such as liquefied petroleum gas (LPG), methane, ammonia, and suspended particulates. This shows a variety of approaches for detecting pollution.
- **Energy efficiency:** Few articles explicitly address energy efficiency. Only a few seem to take this into account, which could indicate a lack of emphasis on this aspect or that energy efficiency is an indirect priority in ventilation systems.
- **Sampling** rate: The sampling rate was specified in only one article (30 s) and was not used in the other studies. However, the sampling rate is important, as it can influence system responsiveness and predictive accuracy. The absence of this information makes it difficult to assess the real-time performance of systems.
- **Data storage:** Storage methods vary, with some studies using the cloud (Google Cloud Drive, ThingSpeak, Amazon Web Services), whereas others do not specify the

[1] American Society of Heating, Refrigerating and Air-Conditioning Engineers

storage tool used. The use of the cloud suggests easy access to data and the ability to process large volumes, although this raises questions of security and latency.

- **Compliance with standards:** A few articles mention compliance with air quality standards. As compliance is rarely explicit in the majority of articles, this shows an effort that must be made to comply with standards. Compliance is essential to guarantee the reliability and conformity of systems in regulated environments.

- **Prediction:** Predictive techniques vary, with some papers using Kalman filters to improve predictions by accounting for noise and uncertainties, whereas others use long-memory neural networks (LSTMs) to process temporal data. Several articles do not include predictive capabilities. Machine learning techniques such as LSTM and gated recurrent unit (GRU) networks tend toward sophisticated predictions based on time series models.

On the basis of these comparisons, in the following section, we propose an analysis of indoor air quality standards and a comparison of big data technologies in terms of real-time data processing. We also focus on integrating IoT monitoring systems with machine learning and deep learning networks for reliable prediction decisions.

3 Technical study of the various components of an intelligent indoor air quality monitoring system

3.1 Indoor air quality standards

Indoor air quality standards are essential for ensuring indoor air quality and energy efficiency. Here are some of the best-known standards, with a description and the main metrics they use:

- **ASHRAE 62.1 and 62.2** (United States) [11]: Published by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), this standard is widely adopted for establishing ventilation and indoor air quality requirements. ASHRAE 62.1 applies to commercial and institutional buildings, whereas ASHRAE 62.2 applies to residential dwellings. It imposes minimum air exchange rates to ensure good indoor air quality.

- **ISO 17772-1** (International) [12]: This international standard sets energy performance and indoor air quality criteria for ventilation systems in nonresidential buildings. It covers energy efficiency and thermal comfort while ensuring adequate levels of ventilation.

- **EN 13779** (Europe) [13]: European standard EN 13779 specifies ventilation and air-conditioning requirements for nonresidential buildings. It includes indoor air performance criteria on the basis of pollution levels and contaminant sources. It is used to define air quality classes and thermal comfort in indoor spaces.

- **WELL Building Standard** (International) [14]: Although not solely focused on ventilation, this standard for occupant health and well-being in buildings pays particular attention to the indoor air quality. This highlights best practices in ventilation and air filtration for commercial and residential buildings.

- **LEED** (leadership in energy and environmental design): [15]: This is a sustainable building certification that includes indoor air quality criteria. Although not specifically a ventilation standard, it imposes requirements for ventilation systems to reduce pollutants and maximize energy efficiency.

- **BREEAM** (Building Research Establishment Environmental Assessment Method) [16]: This is another environmental certification system for buildings. Specific ventilation performance is needed to optimize air quality and occupant comfort while minimizing the carbon footprint.

- **WHO** (World Health Organization) [17]: Although the WHO does not provide specific standards for indoor air, it does provide global guidelines for air pollutants that can be applied indoors.

The metrics used by these standards to ensure good air quality are as follows:

- **CO_2** (ppm): Indicator of ventilation quality; high levels may indicate poor ventilation.
- **PM2**.5/PM10 (µg/m³): Fine particles of different sizes, with PM2.5 being more hazardous to health because of its ability to penetrate deep into the lungs.
- **VOCs** (µg/m³): volatile organic compounds (VOCs) that can be harmful and can be produced from building materials, cleaning products, etc.
- **Formaldehyde (HCHO):** Hazardous chemical often found in furniture and building materials; carcinogenic.
- **Ozone (O_3):** Ozone can enter buildings from outside or be generated by internal appliances.
- **Carbon** monoxide (CO): Toxic gas from combustion; high levels indicate a significant health risk.

In Morocco, the indoor air quality standards used may vary according to specific project requirements and local regulations. However, ASHRAE standards 62.1 and 62.2, as well as ISO 17772-1, are often applied in construction and ventilation projects, even at the international level, as they are recognized worldwide and widely adopted in various countries, including Morocco, for certain quality projects. In Table 2, we present the values we used in the rest of our study.

Table 2 The recommended level used for different metrics for the rest of our study

Metric	Recommended Threshold	Explanation/Importance
CO_2 (ppm)	≤ 1000	Acceptable level to maintain proper air exchange and prevent drowsiness or discomfort.
PM2.5 (µg/m³)	≤ 25	Protects occupants from fine particles that can penetrate deep into the lungs.
PM10 (µg/m³)	≤ 50	Limits larger particles that can irritate the respiratory tract.
VOCs (µg/m³)	≤ 300	Reduces chemical pollutants emitted by furniture, paints, and cleaning products.
Formaldehyde (HCHO, µg/m³)	≤ 100	Limits exposure to this carcinogenic pollutant commonly found in furniture and materials.
Carbon Monoxide (CO, ppm)	≤ 9	Protects against this toxic gas from combustion appliances (heaters, kitchens).
Temperature (°C)	20°C - 26°C	Thermal comfort range for occupants, depending on the seasons in Morocco.
Relative Humidity (%)	30% - 60%	Prevents issues of overly dry or humid air, reducing risks of mold or irritation.
ACH (Air Changes per Hour)	≥ 0.5	Ensures sufficient air exchange to avoid pollutant buildup.
Ozone (O_3, ppb)	≤ 51	Reduces respiratory irritation caused by high ozone levels.

3.2 Real-Time Data Processing

Big data technologies play a crucial role in intelligent ventilation systems, particularly when integrated with the Internet of Things (IoT). In such a system, IoT sensors continuously collect environmental data (such as temperature, humidity, and air quality) and transmit this information in real time. Owing to big data technologies, these massive volumes of data can be processed, analyzed and stored efficiently, enabling complex insights to be extracted and trends in air quality to be detected.

In the literature, several tools are available for real-time data processing. In Table 3, we compare these tools on the basis of several key factors, such as latency, type of real-time processing, data persistence, ecosystem integration and supported languages.

Table 3 Comparison of real-time data processing tools

Tool	Latency	Streaming type	Persistence	Integrated ecosystem	Supported langages
Spark Streaming	Medium (micro batch)	Micro batch	Yes	Excellent (Spark)	Java, Scala, Python
Apache Flink	Low	True streaming	Yes	Moderate	Java, Scala
Apache Storm	Low	True streaming	No	Low	Java, Python
Kafka Streams	Low	True streaming (Kafka)	Yes	Excellent (Kafka)	Java
Google Dataflow	Variable	True streaming/batch	Yes	Excellent (Google)	Java, Python
Azure Stream Analytics	Average	True streaming (SQL)	Limited	Excellent (Azure)	SQL
Amazon Kinesis	low	True streaming	No	Excellent (AWS)	Java, Python

Depending on the complexity of our system and its needs in terms of latency, persistent state and integration, we may use either of the following solutions:

- Apache Flink is perfectly suited for applications that demand real-time processing on the basis of a persistent state and have minimal latency.
- Integration with Kafka: Kafka Streams is a great option if we already manage the flow of data with Kafka.
- Spark ecosystem: Spark Streaming is good for batch analysis and machine learning applications.

3.3 Prediction Models

For an intelligent ventilation system, a predictive model can be designed to improve air quality and optimize energy efficiency by anticipating various factors and adjusting system parameters accordingly. To achieve these objectives, we propose below the key points that can be improved and the models best suited to each case.

3.3.1 Air Quality Improvement

- CO_2 concentration predictions: predict CO_2 levels on the basis of room occupancy. The model can anticipate a rise in CO_2 during high occupancy hours and adjust ventilation accordingly to maintain safe levels.

- PM2.5 and PM10 particle prediction: Environmental data are used to predict the infiltration of fine particles, especially in urban areas or during pollution episodes. The system could then trigger more intense air filtration cycles.

- Concentration of volatile organic compounds (VOCs): Predicts the accumulation of pollutants from furniture, cleaning products or cooking, and adjusts ventilation to effectively disperse them.

- Prediction of outdoor air quality: from meteorological data and outdoor pollution, the model could avoid using outdoor air during periods of high pollution and instead focus on indoor air recycling.

3.3.2 Energy optimization

- Predicted ventilation requirements based on occupancy: By learning occupant occupancy patterns, the model can anticipate times of high or low occupancy and adjust ventilation intensity to avoid overuse.

- Temperature and humidity prediction: On the basis of weather forecasts and comfort preferences, the system can adjust ventilation to balance temperature and humidity, reducing the load on heating and cooling systems.

- Optimum ventilation times: Predict when the outdoor air quality is best (e.g., early morning) and adjust the system to use outdoor air only during these times, reducing the need for filtration and air conditioning.

- Energy load forecasting: Based on outdoor temperature forecasts and equipment usage patterns, the model can anticipate periods of high energy consumption and adapt ventilation to avoid peaks, thus improving overall energy efficiency.

For each prediction use case in an intelligent ventilation system, several machine learning models may be particularly suitable, depending on the nature of the data, the complexity of the problem, and the frequency of the desired predictions.

3.3.3 Suitable models for improving air quality

- **CO_2 concentration prediction:**

✓ Time series models such as ARIMA or SARIMA can be used if the CO_2 data follow a cyclical trend (such as regular room occupancy).

✓ LSTM (long short-term memory) or gated recurrent units (GRUs), which are recurrent neural networks (RNNs) adapted to time series, can handle more complex temporal dependencies, especially when historical CO_2 data show irregular variations.

- PM2.5 and PM10 particle forecasting

✓ Random forest regressor or gradient boosting: To predict particle levels as a function of multiple variables (temperature, humidity, and outdoor air quality), these models handle tabular data and nonlinear interactions well.

✓ LSTM: This method is used for long-term forecasting, especially if the PM2.5 and PM10 levels follow temporal patterns influenced by factors such as seasonal pollution.

- Prediction of VOCs and other chemical pollutants

✓ XGBoost or random forest: These tree-based algorithms are powerful for forecasting chemical pollutants, taking into account the environmental characteristics and potential sources of pollution in buildings.

✓ Linear regression: For simple data or with few influential variables, linear regression may suffice, offering a lighter model.

- Predicting outdoor air quality

✓ SARIMA for cyclical weather data.
✓ Facebook Prophet: Useful for forecasting daily and weekly time series with seasonal patterns, incorporating days of the week, holidays and seasonal effects.

3.3.4 Suitable models for energy optimization

- **Ventilation requirements according to occupancy**

✓ K-nearest neighbors (KNNs) for detecting occupancy patterns based on similar periods in the past.
✓ Classification by artificial neural networks (ANNs) if occupancy data are based on multiple sensors (such as motion detectors and CO_2 sensors).
✓ Clustering models such as K-means are used to detect occupancy patterns and adjust operations accordingly.

- **Temperature and humidity prediction**

✓ SARIMA for seasonal and regular temperature forecasts.
✓ Random forest regressors or gradient boosting machines (GBMs) manage nonlinear dependencies between meteorological variables and the indoor environment.

- **Optimum ventilation times**

✓ LSTM to predict periods of best outdoor air quality.
✓ Tree-based time series models such as XGBoost and random forest are based on occupancy and weather forecasts.

- **Energy load forecasting**

✓ Multivariable linear regression or Ridge/LASSO regression is used if energy load data are influenced mainly by a combination of linear factors.
✓ Deep learning with recurrent neural networks (LSTMs) for complex temporal dependencies in energy consumption and temperature data.

In Table 4, we summarize the machine learning models suitable for each use case in an intelligent ventilation system.

Table 4 Machine learning models suitable for each use case in an intelligent indoor air quality monitoring system

Use Case	Objective	Recommended Machine Learning Models
Prediction of CO_2 concentration	Anticipate variations to adjust ventilation	ARIMA/SARIMA , LSTM, GRU
Prediction of PM2.5 and PM10 particles	Monitor and adjust air filtration	Random Forest Regressor, Gradient Boosting, LSTM
Prediction of VOCs and chemical pollutants	Detect and reduce pollution sources	XGBoost, Random Forest, Linear Regression
Prediction of outdoor air quality	Optimize ventilation based on outdoor conditions	SARIMA, Prophet, LSTM

Ventilation needs based on occupancy	Adjust ventilation according to occupancy	KNN, Artificial Neural Networks (ANN), K-Means
Prediction of temperature and humidity	Maintain optimal thermal comfort	SARIMA, Random Forest Regressor, Gradient Boosting
Optimal hours for ventilation	Determine the best times for ventilation	LSTM, XGBoost,Random Forest
Prediction of energy load	Optimize energy consumption	Multivariable Linear Regression, LSTM

For temporal forecasting and continuous prediction (such as indoor air quality prediction), time series models and recurrent neural networks (RNNs) are particularly well suited. On the other hand, for energy and occupancy optimization tasks, classification and regression models (such as random forest or gradient boosting) are often more effective, especially when combined with an analysis of occupancy patterns and energy loads. Depending on resources and system complexity, hybrid models (combining several models) can also be used for the best results.

4 The proposed architecture for an indoor air quality monitoring system based on IA, IOT and big data technologies

As shown in Figure 1, the architecture of the proposed system consists of 5 main layers: the data collection layer, transmission layer, analysis and decision-making layer, storage layer and presentation layer.

- Sensors and Actuator Layers (Data Collection): This layer contains physical devices, IoT sensors that measure environmental parameters (temperature, humidity, CO2, etc.) and actuators. In addition to internal sensors, our system uses external sensors to collect environmental data. To ensure smooth interaction, we use the MQTT communication protocol to orchestrate communication between the various components of an IoT system, including a Raspberry Pi, ESP modules and sensors/actuators. In this context, the Raspberry Pi generally acts as an MQTT broker, centralizing exchanges. Sensors connected to ESPs periodically publish data such as temperature, humidity, or CO_2 levels on specific topics. Actuators, such as fans or air conditioners, subscribe to these topics and react according to the messages received. The ESP uses Wi-Fi communication to connect to the Raspberry Pi and transmit or receive messages. Transmission and Flow Management Layer (Kafka): This layer collects, distributes and manages data in real time via Kafka. It receives data flows from the sensors and distributes them to the other layers for processing. In effect, sensor data are published in Kafka topics in real time via MQTT. Kafka distributes messages to consumers, i.e., the other layers for processing and decision-making. Kafka connectors are used to archive processed data or important events. These historical data are retrieved for retrospective analysis or to improve optimization models.

- Analysis and Decision-Making Layer (Data Processing and Machine Learning): This layer analyzes data in real time, applying machine learning algorithms to predict trends and make decisions on the activation of IoT devices (fans, air conditioners, etc.). Kafka Streams is used for real-time stream processing, whereas machine learning models are used for prediction. To this end, a Kafka topic has been configured to receive continuous predictions, which are then used by the optimization algorithms to adjust the ventilation and air conditioning systems. The decision-making system takes data provided by internal and external sensors into account when alternating between air-conditioning and ventilation.

- Storage layer: This layer is responsible for storing, processing and managing data from IoT sensors and ventilation devices. It stores historical data for retrospective analysis, reporting and long-term performance management. This layer also retrieves data for training machine learning models.

Presentation layer: This layer is responsible for interaction with the end user. A dashboard provides a real-time view of air quality, temperature and other environmental parameters. The web or mobile application allows the user to activate and deactivate ventilation and adjust parameters. An alert system provides notification of the air quality status and recommended actions.

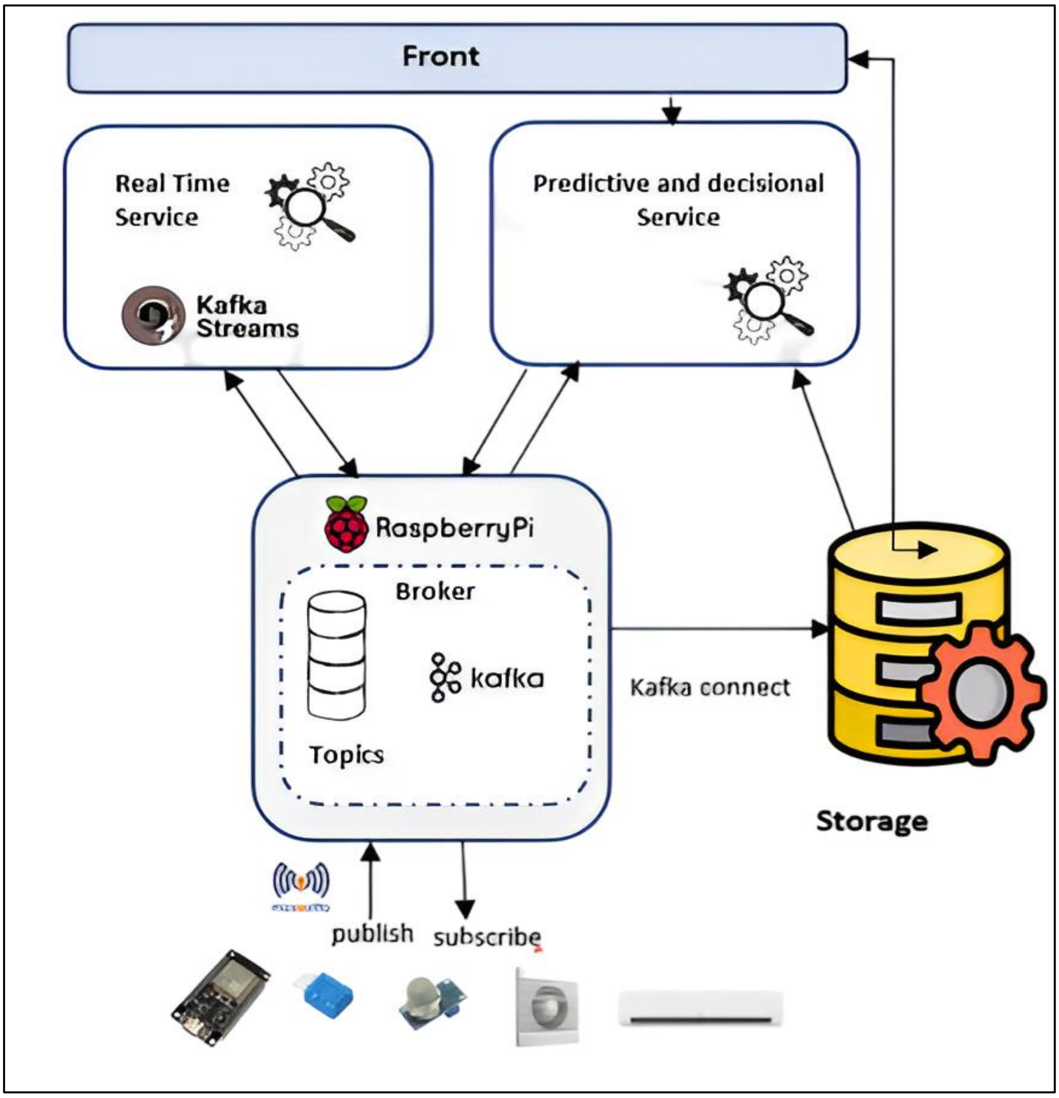

Figure 1 Architecture proposed for the intelligent indoor air quality monitoring system

This modular architecture enables greater scalability and flexibility for future additions, such as integrating other IoT devices or enhancing the prediction model.

5 Realization

Our test environment is a studio-style dome equipped with connected objects (IoT) and artificial intelligence (AI) at the heart of the UNIVERSIAPOLIS campus (Figure 2).

Figure 2 Test environment

The project combines real-time monitoring and control of connected objects through a platform, as well as prototypes designed at the dawn of modern technology and machine learning for the optimization and automation of daily domestic life. This dome is composed of a bedroom, a dining room, a living room, a hall and a kitchen. We equipped all the rooms with temperature, humidity and CO2 sensors. Figure 3 shows the locations of the temperature and CO2 sensors in the smart home.

Figure 3 Location of temperature and CO2 sensors in the smart home

As our system alternates between ventilation and air conditioning, we equipped our smart home with an air conditioner, as shown in Figure 4.

Figure 4 Air conditioner deployed in Smart Home

5.1 choice of components

The choice of components is crucial to ensure that the system operates smoothly. The sensors and actuators selected were chosen for their accuracy and reliability. The table below shows the components used in the project, together with their main characteristics and their functions in the system.

Table 5 List of electronic components used in our intelligent indoor air quality monitoring system

COMPONENT	TYPE	CHARACTERISTICS	FUNCTIONALITY IN THE SYSTEM
Raspberry Pi 3B+	Single-board computer	64-bit quad-core, 1 GB of LPDDR2 SDRAM, MicroSD card	Central controller of the data collected from sensors
Arduino Mega	Microcontroller	Microcontroller	Managing sensors and actuators
5 x MQ135	Air quality sensor	Measures harmful gases (CO2, NH3, etc.)	Monitoring air quality in each room and outdoors
MQ2	Gas sensor	Detects gases (VOCs, smoke)	Identifying dangerous gas levels in the kitchen
5 x DHT22	Temperature and humidity sensor	Temperature range: -40 to 80°C; humidity: 0 to 100%	Monitoring temperature and humidity in each room
ACS712	Current sensor	Range: 5A, 20A, or 30A	Measuring system energy consumption
4 x Fans	Actuators	12 V DC	Ensuring air circulation in the rooms
IR Module	Actuator	Infrared remote control	Remotely controlling air conditioning
ESP32	Microcontroller	Built-in Wi-Fi, Bluetooth	Transmit collected data or receive orders
HLK-PM01	Power supply module	AC-DC conversion	Converts AC power to low voltage for microcontrollers and other components
Relays	Switching devices	-	Remotely controlling equipment

Our IoT system uses an Arduino board to manage the sensors and actuators. Temperature, humidity and gas (such as butane) sensors are connected to the Arduino to collect environmental data. Actuators include fans connected to a relay module. The air conditioner is controlled via infrared with a Smart Air Conditioner control system deployed in the smart Home shown in figure 5.

Figure 5 Control board deployed for air conditioner remote control

As mentioned above, we use the Kafka broker to ensure a continuous, reactive data flow and Kafka streams for real-time data processing. For the Kafka broker, we create 4 topics used to organize messages received from the various sensors and another topic to record data from the predictive system (Figure 6).

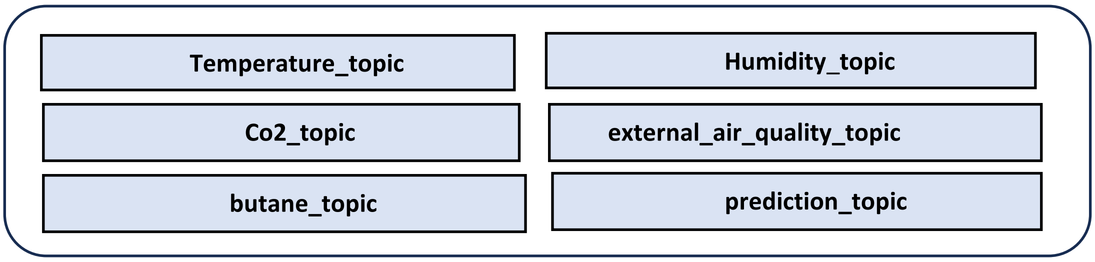

Figure 6 Kafka Topics

These data are used by the decision-support system to switch between ventilation and air conditioning to ensure optimum energy efficiency and occupant comfort. In Table 6, we propose detailed scenarios for choosing between ventilation and air-conditioning.

Table 6 Detailed scenarios proposed for choosing between ventilation and air-conditioning

Scenario	**Internal Conditions**	**External Conditions**	**System Action**	**Justification**
Moderate temperature and air quality (ventilation)	• Temperature: 20–26°C • CO_2: good • PM2.5: good	• Temperature: 20–26°C • CO_2: good • PM2.5: good	Ventilation activated	Ideal conditions for renewing indoor air with external air.
High internal and external temperature (air conditioning)	• Temperature: >28 • CO_2 <1000 ppm • PM2.5: < 12 µg/m³	• Temperature: >30 • CO_2< 700 ppm • PM2.5 < 12 µg/m³	Air conditioning activated	Ventilation is ineffective due to external heat; air conditioning is needed.
Good temperature, poor internal air quality (ventilation)	• Temperature: 20–26°C • CO_2: 1000–2000 ppm • PM2.5: 12–35 µg/m³	• Temperature: 20–26°C • CO_2 <1000 • PM2.5 < 12 µg/m³	Ventilation activated	Ventilation to reduce CO_2 levels and internal particles.
Cold external and warm internal air (isolation)	• Temperature: 20–26°C • CO_2 <1000 ppm • PM2.5: 12–35 µg/m³	• Temperature: <20 • CO_2 <1000 ppm • PM2.5: 12–35 µg/m³	No action (isolation)	Maintain indoor conditions without ventilation or air conditioning.
Poor external air quality (air conditioning with filtration)	• Temperature: >28°C • CO_2<1000 ppm • PM2.5 >14 µg/m³	• Temperature: 25°C • CO_2 >1500 ppm • PM2.5>40 µg/m³	Air conditioning activated with advanced filtration	Avoid introducing polluted air while maintaining a comfortable temperature.
High indoor humidity,	• Temperature: 22°C • Humidity: 80%	• Temperature: 20°C - Humidity: 60%	Ventilation activated with humidity control	Reduce excessive indoor humidity

moderate external conditions				using moderate external air.
Heatwave outside, good indoor air quality	• Temperature: > 28C • CO_2: <1000 ppm • PM2.5: 12–35 µg/m³	• - Temperature: 40°C - CO_2<1000 ppm - PM2.5: 12–35 µg/m³	Eco-mode air conditioning	High external temperature requires minimal air conditioning to maintain comfort.
Sandstorm outside (closed system, filtration)	• Temperature: 20-26°C • CO_2<1000 ppm • PM2.5: 12–35 µg/m³	• - Temperature: 30°C - CO_2:> 2000 ppm - PM2.5: >100 µg/m³	Air conditioning activated with advanced filtration	Prevent polluted air from entering while maintaining acceptable indoor air quality.

5.2 System programming logic

System programming is based on well-defined logic, which enables sensors and actuators to be managed according to the data collected. The key steps we used in the programming logic of our indoor air quality system are as follows:

- **Initialization**: The main role of the initialization stage is to prepare the system to collect, process and react to data in real time. This step configures the foundations for the entire treatment process, defining data sources, communication channels and tools for processing information. Kafka Streams is initialized with KafkaStreams.start(), which prepares the system to consume real-time data streams from sensors and external sources and to create real-time processing pipelines to make decisions on the basis of the latest data.

Initialize Kafka topics for sensor data

```
temperature_topic = "temperature_topic"    //Topic for indoor temperature data
humidity_topic = "humidity_topic"          //Topic for indoor humidity data
co2_topic = "co2_topic"                    //Topic for CO2 levels
external_air_quality_topic = "external_air_quality_topic"//Topic for external air quality data
prediction_topic = "prediction_topic"      //Topic for predictions
butane_sensor_topic = "butane_sensor_topic" //Topic for butane gas detection data
//Start the Kafka Streams for real-time data processing
stream = KafkaStreams.start()
```

- **Collect** and **Aggregate Sensor Data:** This stage begins by collecting data from various sensors deployed in the environment (e.g., temperature, humidity, CO2, air quality, butane gas sensors). Once the data have been collected, the aggregation stage involves combining several data streams from different sources. They are then filtered, reduced and transformed to make them usable by the control system.

Consume streams of data from various sensors

```
temperature_stream = stream.consume(temperature_topic)      //Temperature data
humidity_stream = stream.consume(humidity_topic)            //Humidity data
co2_stream = stream.consume(co2_topic)                      //CO2 data
```

```
external_air_quality_stream = stream.consume(external_air_quality_topic)//External air
quality data
butane_stream = stream.consume(butane_sensor_topic)        //Butane sensor data

//Aggregate all sensor data into a unified stream
sensor_data_stream = stream.join(
   temperature_stream, humidity_stream, co2_stream, external_air_quality_stream
)
//Applying prediction logic to the aggregated data
predicted_data_stream = sensor_data_stream.map(data => {
 //Call a prediction function or model to predict air quality and temperature
   prediction = PredictAirQualityAndTemperature(data)
   return data.merge(prediction)//Combine input data with prediction
})
//Send prediction results to the prediction topic
predict_data_stream.foreach(prediction => {
   stream.produce(prediction_topic, prediction)//Publish predictions for further processing
})
```

- **Safety Check for Butane Detection:** The butane sensor stream is continuously monitored. If butane gas is detected, the system immediately activates an alarm and shuts down all ventilation and air conditioning systems for safety. If no butane is detected, the system follows its normal logic.

Monitor the butane sensor stream for safety alerts

```
butane_alert_stream = stream.consume(butane_sensor_topic)
//Check for butane gas detection in real time
butane_alert_stream.foreach(alert => {
   if (alert.butane_detected == true) {
//Trigger safety protocol if butane is detected
      ActivateAlarm()                  //Sound the alarm
      DeactivateAllSystems()              //Turn off all ventilation and air conditioning
systems
   } else {
    //If no butane is detected, proceed with normal system operation
      ProcessPredictions()
   }
})
```

- **Decision Making Based on Predictions**: Predictions from the prediction_topic are processed to make real-time decisions. On the basis of air quality and temperature predictions, the system can activate or deactivate ventilation or air conditioning. The system handles eight scenarios, as described in the decision-making table (Table 6).

Make decisions on the basis of predictions

```
function ProcessPredictions() {
//Consume prediction data from the prediction topic
   predictions_stream = stream.consume(prediction_topic)

//Make decisions on the basis of predictions
   predictions_stream.foreach(data => {
```

```
      if (data.air_quality == "good" AND data.temperature == "normal") {
       //Scenario 1: Optimal air quality and temperature
         SetMode("standby")                //No action needed
      }
      else if (data.air_quality == "poor" AND data.temperature == "normal") {
       //Scenario 2: Poor air quality
         Activate ventilation()            //Purify indoor air
      }
      else if (data.temperature == "high") {
         if (data.air_quality == "good") {
          //Scenario 3: High temperature, good air quality
            ActivateCooling()             //Turn on air conditioning
         } else {
          //Scenario 4: High temperature, poor air quality
            Activate cooling and ventilation()//Cool and purify the air
         }
      }
      else if (data.air_quality == "average" AND data.temperature == "normal") {
       //Scenario 5: Average air quality
         MaintainVentilation("open")       //Keep ventilation running
      }
      else if (data.temperature == "low") {
       //Scenario 6: Low temperature
         DeactivateCooling()               //Turn off air conditioning
         ActivateVentilationIfNecessary()   //Ventilate if required
      }
      else if (data.air_quality == "poor" AND data.temperature == "normal") {
       //Scenario 7: Poor air quality
         MaintainVentilation("closed")      //Prevent external pollution
      }
   })
}
```

- **Action handlers:** Each function handles a specific action for the system (e.g., turning on/off ventilation or air conditioning). The alarm function and system deactivation are critical for safety in the case of butane detection.

Function Handlers

```
function ActivateAlarm() {
   SendCommand("alarm", "ON")             //Send a command to turn on the alarm
   LogEvent("Butane detected! Alarm activated.")
}
function DeactivateAllSystems() {
   SendCommand("ventilation", "OFF")       //Turn off ventilation
   SendCommand("air_conditioning", "OFF")  //Turn off air conditioning
   LogEvent("All systems deactivated for safety.")
}
function SetMode(mode) {
   if (mode == "standby") {
      SendCommand("ventilation", "OFF")   //Put system in standby
      SendCommand("air_conditioning", "OFF")
   }
}
function ActivateVentilation() {
```

```
    SendCommand("ventilation", "ON")        //Start ventilation
}
function ActivateCooling() {
    SendCommand("air_conditioning", "ON")   //Start air conditioning
}
function ActivateCoolingAndVentilation() {
    SendCommand("air_conditioning", "ON")   //Start both cooling and ventilation
    SendCommand("ventilation", "ON")
}
function MaintainVentilation(state) {
    if (state == "open") {
        SendCommand("ventilation", "ON")    //Keep ventilation running
    } else if (state == "closed") {
        SendCommand("ventilation", "OFF")   //Stop ventilation
    }
}
function ActivateVentilationIfNecessary() {
    SendCommand("ventilation", "ON")        //Activate ventilation only if needed
}
function DeactivateCooling() {

    SendCommand("air_conditioning", "OFF")  //Turn off air conditioning
```

6 Discussion

The indoor air quality system we propose in this study combines the most innovative technologies in terms of connected objects, data analysis and real-time processing. The strengths of the proposed solution, as well as its limitations for possible improvement, are analyzed below.

Comfort and energy efficiency: To optimize energy consumption while maintaining adequate thermal comfort, we alternated between fan and air conditioning based on real-time data (temperature and air quality). In addition, the use of forecasts (data-based predictions) allows us to adjust actions before conditions become uncomfortable.

Improved air quality: deployed sensors monitor indoor and outdoor air quality to activate ventilation or shut down systems, ensuring a healthy environment. The management of pollutants (such as CO2 or particles) prevents respiratory problems, especially in urban or industrial environments.

Increased safety owing to gas detection: The detection of gases such as butane activates an alarm and deactivates all other systems to prevent the risk of explosion or fire. This approach prioritizes occupant safety since the butane sensor takes priority over all other operations.

Connectivity and remote management: With ESP8266, the system can be monitored and controlled remotely via an application or cloud platform, offering great flexibility and continuous monitoring.

Scalability and customization: Using the Kafka and Kafka Streams, the system can easily handle large volumes of data and expand to include more sensors or actuators.

The decision system considers both the data received from the various sensors and the predictions based on the data. With this in mind, we propose a set of scenarios with a focus on energy efficiency and occupant comfort.

With regard to limitations for future improvement, we cite the following:

Dependence on connectivity: If the Wi-Fi network is unstable, or in the event of an ESP8266 failure, the system could be partially or completely disabled, compromising its reliability.

System complexity: Coordination between multiple sensors, actuators and a real-time processing system such as Kafka makes the system complex to configure, maintain and troubleshoot.

Sensor error risks: Incorrect calibration or malfunction of sensors (gas, temperature, humidity) could lead to incorrect actions, such as disabling a fan when butane gas is detected.

Data and communication security: Data collected and commands must be encrypted and use sufficient security measures, as the system may be vulnerable to cyberattacks or malicious interception.

Machine learning-based prediction: In our study, we proposed an in-depth analysis of the machine learning models that may be most appropriate in different scenarios. This first version of the system does not integrate the prediction part since the development of the best model requires real data collected from our test environment. The integration of prediction into our indoor air quality system is the subject of progress.

7 Conclusion

IoT technologies integrated with data processing algorithms in indoor air quality monitoring systems represent a significant leap toward the intelligent management of domestic and industrial environments. This article focuses on a system proposal that integrates a modular architecture based on sensors and actuators with a real-time data processing infrastructure using Kafka Streams. The ability to monitor and control critical parameters such as temperature, humidity, CO_2, and external pollutants with this approach while ensuring predictive decision-making is readily possible.

By emphasizing energy efficiency, the system optimizes the operation of fans and air conditioners to reduce energy consumption while sustaining a healthy, comfortable indoor environment. The use of recognized standards, such as MQTT for communication, ensures interoperability, reliability, and security for the data exchanged. In addition, the inclusion of a gas sensor for butane leak detection provides an essential layer of safety by automatically disabling those systems in the presence of possible danger.

However, several limitations have been reported, such as dependence on stable network connectivity, risks from sensor errors, and difficult maintenance within complex environments. These constraints provide further opportunities for development: increasing sensor redundancy, local autonomy in the case of network failure, and the incorporation of machine learning algorithms to make more accurate predictions.

Finally, the predictions that can be made by the system regarding air quality, weather conditions, or the lifetime of equipment seal its place not only as a reactive device but also as a proactive device. This further underlines the potential of such a system to meet the rising demands for sustainable and intelligent management of the indoor air quality. It therefore constitutes a very valuable contribution to the quest for health, safety, and energy efficiency across diverse environments.

ACKNOWLEDGMENTS

The author acknowledges support from the International University of Agadir, particularly the Polytechnic School. The author is thankful to the editor and anonymous reviewers for their constructive suggestions and comments.

References

[1] Tran VV, Park D, Lee YC. Indoor Air Pollution, Related Human Diseases, and Recent Trends in the Control and Improvement of Indoor Air Quality. Int J Environ Res Public Health. 2020 Apr 23;17(8):2927. doi: 10.3390/ijerph17082927. PMID: 32340311; PMCID: PMC7215772.

[2] Belova, Anna & Dagli, Rahul & Economu, Nicole & Hartley, Seth & Holder, Chris & Hubbard, Heidi & Justice, Michele & Lettes, Sarah & Raymer, Paul & Silva, Raquel. (2022). Literature Review on the Impacts of Residential Combustion Final Report Literature Review on the Impacts of Residential Combustion Final Report.

[3] M. Benammar, A. Abdaoui, S. Ahmad, F. Touati, and A. Kadri, “A Modular IoT Platform for Real-Time Indoor Air Quality Monitoring,” *Sensors*, vol. 18, no. 2, p. 581, Feb. 2018, doi: 10.3390/s18020581.

[4] J. Jo, B. Jo, J. Kim, S. Kim, and W. Han, “Development of an IoT-Based Indoor Air Quality Monitoring Platform,” *Journal of Sensors*, vol. 2020, no. 1, p. 8749764, 2020, doi: 10.1155/2020/8749764.

[5] Q. P. Ha, S. Metia, and M. D. Phung, “Sensing Data Fusion for Enhanced Indoor Air Quality Monitoring,” *IEEE Sensors J.*, vol. 20, no. 8, pp. 4430–4441, Apr. 2020, doi: 10.1109/JSEN.2020.2964396.

[6] J. Saini, M. Dutta, and G. Marques, “Indoor Air Quality Monitoring Systems Based on Internet of Things: A Systematic Review,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 14, Art. no. 14, Jan. 2020, doi: 10.3390/ijerph17144942.

[7] Sung, WT., Hsiao, SJ. Building an indoor air quality monitoring system based on the architecture of the Internet of Things. *J Wireless Com Network* **2021**, 153 (2021). https://doi.org/10.1186/s13638-021-02030-1

[8] C. De Capua, G. Fulco, M. Lugarà, and F. Ruffa, “An Improvement Strategy for Indoor Air Quality Monitoring Systems,” *Sensors*, vol. 23, no. 8, Art. no. 8, Jan. 2023, doi: 10.3390/s23083999.

[9] W.-L. Hsu *et al.*, “Establishment of Smart Living Environment Control System,” *Sensors and Materials*, vol. 32, no. 1, p. 183, Jan. 2020, doi: 10.18494/SAM.2020.2581.

[10] “Shanmugaraja et al. - 2021 - Analysis of air quality using IoT with machine lea.pdf.”Available: https://iopscience.iop.org/article/10.1088/1742-6596/1916/1/012188/pdf

[11] “standard-62.2-fact-sheet.pdf.” Accessed: Nov. 21, 2024. [Online]. Available: https://www.ashrae.org/file%20library/about/government%20affairs/advocacy%20toolkit/virtual%20packet/standard-62.2-fact-sheet.pdf

[12] “ISO-17772-1-2017.pdf.” Accessed: Nov. 21, 2024. [Online]. Available: https://cdn.standards.iteh.ai/samples/60498/7e804a657a2947d4a2b4ad5b9f0cc06a/ISO-17772-1-2017.pdf

[13] M. Brandt, “Calcul des débits de ventilation en non-résidentiel,” 2019.

[14] “well-building-standard-10-26-15-web_lr-french.pdf.” Accessed: Nov. 21, 2024. [Online]. Available: https://a.storyblok.com/f/52232/x/6e652913a1/well-building-standard-10-26-15-web_lr-french.pdf

[15] “LEED_Certification_Guidebook_March_2011.pdf.” Accessed: Nov. 21, 2024. [Online].Available: https://dgs.dc.gov/sites/default/files/dc/sites/dgs/page_content/attachments/LEED_Certification_Guidebook_March_2011.pdf

[16] "BREEAM_In-Use_Standard.pdf." Accessed: Nov. 21, 2024. [Online]. Available: https://tools.breeam.com/filelibrary/BREEAM%20In%20Use/BREEAM_In-Use_Standard.pdf

[17] "Guidelines for air quality", Published by the World Health Organization, Geneva Cluster of Sustainable Development and Healthy Environment (SDE), Department of Protection of the Human Environment (PHE), Occupational and Environmental Health Programme (OEH). Accessed: Nov. 21, 2024. [Online]. Available: https://www.miteco.gob.es/content/dam/miteco/es/calidad-y-evaluacion-ambiental/temas/atmosfera-y-calidad-del-aire/guidelinesforairquality-2000_tcm30-188065.pdf

Notes on contributors

Mouhim Sanaa is an Assistant professor at Universiapolis since 2014. She received her Ph.D. in Artificial intelligence in 2014 from the Ibn Zohr University. In her research, Sanaa has focused on adapting complex algorithms such as intelligent systems and optimized neural networks for resource-constrained embedded devices . As head of the AI Lab in the UNIVERSIAPOLIS innovation garden, her work focuses on the development of intelligent systems based on context modeling in smart home environments.