Fuzzy risk assessment system for indoor air quality and respiratory disease prevention

Éldman de Oliveira Nunes[1], Ariel Isaac Posada Barrera[2], Laura Margarita Rodríguez Peralta[2,3], Paulo Nazareno Maia Sampaio[4], Fabián Leonardo Cuesta Astudillo[2,5]
[1]Department of Computation, Universidade SENAI CIMATEC, Salvador, Brazil
[2]Department of Engineering, Faculty of Information Technology and Data Science, Universidad Popular Autónoma del Estado de Puebla (UPAEP), Puebla, Mexico
[3]Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), Puebla, México
[4]School of Engineering, Universidad de Lima (ULima), Lima, Peru
[5]GIFACT, Universidad Politécnica Salesiana (UPS), Cuenca, Ecuador

Article Info

Article history:

Received Oct 24, 2024
Revised Jun 19, 2025
Accepted Aug 6, 2025

Keywords:

Environmental pollutants
Fuzzy logic
Indoor air quality
Risk classification
Sick building syndrome

ABSTRACT

This study addresses the evaluation of indoor air quality, with a focus on mitigating respiratory diseases and sick building syndrome (SBS). Recognizing that different pollutants exhibit variable behavior depending on environmental factors and human activity, the objective was to develop a fuzzy logic-based classification system that integrates environmental variables such as temperature, relative humidity, and pollutant concentrations–particulate matter (PM10, PM2.5), carbon dioxide (CO_2), and total volatile organic compound (TVOC)-into a unified model. The method involved defining risk levels as low, moderate, high, and very high, and implementing 56 fuzzy rules to dynamically and accurately categorize these risks, based on measurements taken between 2022 and 2024 in the states of Morelos and Puebla under various relative humidity and temperature scenarios. The analysis of the results demonstrated robust system performance, with an overall accuracy of 94.08%, but also revealed challenges in distinguishing between adjacent risk classes. This research contributes to a better understanding of the complex impacts of air quality on health and reinforces efforts to mitigate respiratory problems and SBS in densely populated indoor environments.

Corresponding Author:

Éldman de Oliveira Nunes
Department of Computation, Universidade SENAI CIMATEC
Salvador, Brazil
Email: eldman.nunes@fieb.org.br

1. INTRODUCTION

The growing awareness of the impacts of indoor air quality on human health has intensified research in this area, particularly concerning the prevention of respiratory diseases and sick building syndrome (SBS). Building on previous studies, this research serves as a continuation of Portillo *et al.* [1], which identified variables such as carbon dioxide (CO_2), total volatile organic compound (TVOC), and particulate matter (PM2.5, and PM10) as pollutants that increase the risk of respiratory diseases in indoor spaces and contribute to SBS; Barrera *et al.* [2], which demonstrated a correlation between human activity and environmental variables in indoor spaces affecting pollutant concentration fluctuations; Barrera *et al.* [3], which evaluated the feasibility of forecasting pollutant changes based on human activity. Additionally, Barrera *et al.* [4] focused on forecasting CO_2 concentrations—used as the defining indicator of indoor air quality in that study—using a broader set of experimental data and machine learning models applied to a specific room.

CO_2 was identified as the most complex variable to predict due to its sensitivity to occupant behavior and ventilation patterns. Collectively, these studies highlight the critical role of indoor air quality in public health and emphasize the need for effective monitoring and management methods.

Mexico, with its diverse climatic and environmental conditions, provides a particularly relevant setting for studying the impacts of air quality. Research conducted along the Mexico-U.S. border in [5] and at Universidad Iberoamericana Torreón in [6] highlighted the importance of adaptive strategies for pollution management, underscoring the need for a more integrated and dynamic approach to air quality assessment. In this context, the present study introduces a fuzzy logic-based assessment system for predicting risks associated with indoor air quality. Utilizing data collected in Morelos and Puebla, the system integrates environmental variables such as temperature, relative humidity, and pollutant concentrations (CO_2, TVOC, PM2.5, and PM10) into a unified risk assessment model. Fuzzy logic was chosen for its capacity to manage uncertainty and imprecision in classifying environmental variables, offering a flexible approach for the continuous assessment of air quality and the identification of conditions that may affect occupants' health in indoor environments. The primary objective of this study is to develop and implement a fuzzy system that accurately classifies risk levels related to indoor air quality, thereby contributing to the prevention of respiratory diseases and SBS. This research distinguishes itself by applying a comprehensive fuzzy model that evaluates air quality based on multiple variables while considering the specific environmental factors and human activities influencing these pollutants.

The contribution of this work lies in the creation of a risk assessment model that enhances existing air quality monitoring practices. It underscores the importance of an integrated and adaptive approach to managing indoor air quality and preventing respiratory diseases, based on data from central Mexico. Furthermore, the study highlights the need for continuous adjustments to the membership functions and fuzzy rules to improve classification accuracy, particularly among adjacent risk categories.

This article is structured as follows: section 2 reviews related works that provide the context and theoretical foundation for this research. Section 3 describes the data collected, outlines the methodology applied, and explains the construction and implementation of the fuzzy model. Section 4 presents the results and discusses their practical implications. Finally, section 5 concludes the study and offers directions for future research.

2. RELATED WORKS

Related studies on indoor air quality reveal a wide range of variables that are fundamental for the analysis and management of air quality. Research has highlighted the importance of specific pollutants such as PM2.5, PM10, CO_2, and TVOC, as well as environmental factors like relative humidity and temperature. In a study conducted by [7], the combination of PM2.5, PM10, carbon monoxide (CO), relative humidity, and temperature was found to provide a comprehensive view of air quality and its impact on health. Another study [8] expanded this analysis to include sulphur dioxide (SO_2), nitrogen dioxide (NO_2), and CO, emphasizing the importance of considering multiple pollutants for a complete assessment. Additionally, Ho *et al.* [9] focused on CO_2, PM2.5, TVOC, and relative humidity, suggesting that these variables are crucial for understanding the risks associated with indoor air quality. Other studies, such as [10], [11], confirmed the relevance of relative humidity and temperature, while expanding the analysis to include other pollutants like nitric oxide (NO), CO, and ozone (O_3). The breadth of these variables highlights the complexity of managing indoor air quality and underscores the need for integrated methods for effective assessment.

In the last decade, there has been an increasing application of fuzzy logic in the analysis of indoor air quality. Various studies have documented the effectiveness of fuzzy logic-based systems for monitoring and classifying air quality. Works such as those by [12]–[15] have demonstrated that fuzzy logic controllers are effective for dynamic monitoring and classification of indoor air quality. Liang *et al.* [16] introduced fuzzy assessment methods to evaluate the impact of chemical pollutants, showcasing the flexibility and adaptability of fuzzy logic in handling complex variables. Furthermore, the adaptive dynamic fuzzy inference system tree (ADFIST) model, integrated with IoT and fuzzy logic, as described in [17], underscores the application of fuzzy logic in more advanced environmental monitoring systems.

Monitoring techniques have also evolved, as exemplified in study [18], where a real-time monitoring system was implemented using a comprehensive framework. Other approaches, such as the technique for order preference by similarity to ideal solution (TOPSIS)-based method for assessing indoor environmental quality [19] and type-2 fuzzy controllers for automatic control [20], demonstrate the versatility of fuzzy logic across different control and monitoring contexts. Finally, the Mamdani-type fuzzy inference system [21] has been employed to monitor air quality and temperature, showcasing the robustness and adaptability of fuzzy logic in a wide range of applications. This literature review provides a solid foundation for the development of the fuzzy logic-based assessment model presented in this article, highlighting the relevance of an integrated and adaptive approach to managing indoor air quality and preventing respiratory diseases.

3. METHODOLOGY AND FUZZY MODEL IMPLEMENTATION

The dataset used in this analysis was obtained from the study presented in [3]. Data collection involved the use of the Databot and PMS5003 sensors, as described in [22], [23], respectively. The Databot sensor recorded environmental variables, including humidity, temperature, CO_2 levels, and TVOC, while the PMS5003 measured particulate matter concentrations, specifically PM2.5 and PM10. Additional data were captured using the IAQM-128W air quality monitor [24], which also included similar measurements.

In addition to environmental variables, metadata related to the type of ventilation system, door status, number of occupants, and room volume were collected across various locations in the cities of Puebla and Morelos, Mexico. Before analyzing the data, simultaneous comparisons were made between the readings from the Databot [22], PMS5003 [23], and the IAQM-128W air quality monitor [24] under different environmental conditions. These comparisons revealed no significant differences between the systems. The complete dataset comprised 74,826 distinct scenarios collected over periods ranging from 1 to 6 hours, utilizing both sensor technologies.

The dataset provides detailed observations on various environmental and occupancy metrics, with an emphasis on temperature, CO_2 levels, TVOC, and humidity. CO_2 concentrations, as described in [25], were measured in parts per million (ppm), ranging from 400 to 1,534 ppm, with an average of 513.37 ppm. TVOC levels, reported in parts per billion (ppb) [26], ranged from 0 to 2,184 ppb, with an average of 239.21 ppb. Humidity levels varied between 9.47% and 40.52%, with an average of 16.61%. The dataset also includes counts of occupants present, ranging from 0 to 33, with an average of 11.28. Measurements of particulate matter, specifically PM2.5 and PM10 [27], were documented in micrograms per cubic meter ($\mu g/m^3$), with averages of 9.42 $\mu g/m^3$ and 11.35 $\mu g/m^3$, respectively. Occupant density, door status, and ventilation conditions were recorded during measurements, monitoring the specific room where these metrics were recorded.

3.1. Data description

The dataset used in this study comprises 100,275 rows and 12 columns. The columns in the database include a combination of numerical and categorical attributes. The numerical columns are: temperature, CO_2, TVOC, humidity, people density, people, PM2.5, PM10, normal ventilation, natural ventilation, and door open. The categorical column is risk_label, which represents the target class.

All data in the dataset is complete, with no null values, indicating the integrity of the records. The risk_label column, which is the target variable, contains four distinct classes with the following distribution: moderate, with 48,260 records (48.13% of the total); high, with 27,240 records (27.17%); low, with 17,318 records (17.27%); and very high, with 7,457 records (7.44%). This class distribution reveals an imbalance, with the moderate class being the most prevalent and the very high class being the least represented.

3.2. Feature selection

Feature selection is a crucial step in the machine learning model-building process. It directly influences the quality and efficiency of the model. Different approaches were utilized to select the most relevant attributes, each offering specific advantages in terms of accuracy, interpretability, and computational performance.

The filter approach is one of the simplest and fastest methods for feature selection, as it relies exclusively on the statistical properties of the data, independent of the predictive model. Among the methods used in this approach, analysis of variance (ANOVA) stands out, as it assesses the statistical significance of categorical variables in relation to the variable of interest. The chi-square test was also employed, being particularly useful for categorical variables, as it measures the independence between two variables. Pearson correlation was used to identify the linear relationship between continuous variables, allowing for the exclusion of redundant attributes that are strongly correlated.

The wrapper approach involves using a predictive model to evaluate the quality of subsets of attributes, which can provide higher accuracy but at a greater computational cost. In the context of this study, two main methods were applied. Forward feature selection iteratively adds attributes that improve the model's performance, starting with an empty set and inserting features until no significant improvements are observed. Recursive feature elimination (RFE) works in the opposite direction, iteratively removing the least important attributes until the most efficient subset for modeling is reached.

The embedded approach combines the benefits of the filter and wrapper approaches by integrating the feature selection process directly into the model training. In this study, decision trees were utilized as the main method in this approach. Decision trees automatically select the most important attributes during the model-building process, establishing a balance between predictive accuracy and model simplicity. This method not only performs classification but also ranks the attributes by their relevance, facilitating the interpretation and explanation of the model.

The comprehensive analysis of the attributes highlights the integrated importance of various features and the selection approaches utilized in this study. By employing methods such as ANOVA, chi-squared, and decision trees across different approaches—filter, wrapper, and embedded—we were able to effectively identify and prioritize the most relevant attributes. This rigorous methodology not only enhances the interpretability of the machine learning models developed but also ensures their efficiency in processing the dataset.

The analysis of the attribute selection methods revealed that TVOC consistently emerged as the most important attribute across all applied methods, reinforcing its high final importance. Additionally, the variables PM2.5, CO_2, and PM10 also showed significant relevance across various methods, confirming their importance in the selection process. The importance values assigned to each variable were normalized and combined to reflect the cumulative influence of each selection method. This approach enabled the creation of a final consolidated and ordered list, represented in Table 1, providing a comprehensive overview of the relative importance of the attributes. This proved useful for subsequent modeling and other analyses.

3.3. Fuzzy controller

The fuzzy risk assessment system for indoor air quality and respiratory disease prevention is illustrated in Figure 1, which highlights its main components: the fuzzification interface, the inference mechanism, the knowledge base, and the defuzzification interface. The process begins with the fuzzification stage, where linguistic variables, subjectively defined, are transformed into fuzzy sets. These fuzzy sets are then fed into the inference mechanism, which uses the knowledge base to activate the corresponding linguistic rules. For this system, the variables TVOC, CO_2, PM2.5, and PM10 were selected based on their importance as shown in Table 1, and the membership functions for these variables were modeled using a trapezoidal shape.

After inference, the system moves to the defuzzification phase, where the inferred fuzzy values are connected to the process to be controlled. In this step, the resulting fuzzy regions from the inference are converted into real values. The defuzzification method used is the centroid or center of gravity (COG), also known as the center of area (COA). This method calculates the center of the area under the resulting membership function curve, providing a value that represents the fuzzy system's output. In addition to fuzzification and defuzzification, the overall system involves data collection through sensors, information processing within the fuzzy system, and interaction with actuators that control processes based on the inferred output.

Table 1. Consolidated importance of attributes across selection methods

Attribute	ANOVA	Chi-squared	Pearson	Forward	RFE	Decision tree	Integrated importance
TVOC	0.83	0.95	0.25	0.63	0.75	0.75	0.69
PM2.5	0.02	0.01	0.12	0.07	0.25	0.25	0.12
PM10	0.02	0.01	0.16	0.06	0.00	0.00	0.04
CO2	0.06	0.04	0.08	0.04	0.01	0.01	0.04
Temperature	0.01	0.00	0.15	0.05	0.00	0.00	0.03
People density	0.01	0.00	0.10	0.04	0.00	0.00	0.02
Humidity	0.02	0.00	0.01	0.08	0.00	0.00	0.02
People	0.00	0.00	0.07	0.02	0.00	0.00	0.02
Door open	0.01	0.00	0.06	0.00	0.00	0.00	0.01
Natural ventilation	0.02	0.00	0.02	0.01	0.00	0.00	0.01
Normal ventilation	0.00	0.00	0.02	0.00	0.00	0.00	0.00
Sum	1.0	1.0	1.0	1.0	1.0	1.0	1.0

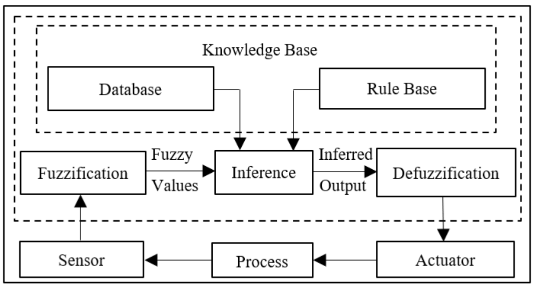

Figure 1. Diagram of the fuzzy risk assessment system for indoor air quality and respiratory disease prevention

3.4. Membership functions

The membership functions for the variables TVOC, CO_2, PM2.5, PM10, and risk were defined using trapezoidal functions. A 15% overlap was applied to provide a smooth and representative transition between different levels of membership. This approach ensures clarity in the classification of membership levels.

For the TVOC variable, the range is defined from 0 to 5,500 ppb. The membership functions are as follows: 'Low' covers the range [0, 0, 64, 74] ppb, where values below 64 ppb are considered safe for indoor environments. The 'Moderate' category is represented by [42, 65, 220, 243] ppb, indicating that prolonged exposure may cause discomfort. The 'High' function spans [215, 221, 600, 610] ppb, signaling an increased health risk, especially for sensitive individuals. Finally, 'Very high' is defined as [590, 601, 5,500, 5,510] ppb, where levels are potentially dangerous and require immediate intervention.

For CO_2 concentration, the range is 250 to 5,000 ppm. The 'Low' function is established for [250, 275, 538, 625] ppm, with values below 275 ppm indicating good ventilation. The 'Moderate' category covers [275, 625, 875, 1,050] ppm, where levels up to 625 ppm are acceptable but may cause discomfort for some individuals. 'High' spans [625, 1,050, 1,500, 2,100] ppm, suggesting poor ventilation and an increased risk of fatigue. Lastly, 'Very high' is defined as [1,050, 2,100, 3,750, 5,000] ppm, where levels can cause headaches and drowsiness, potentially leading to asphyxiation at higher concentrations.

For PM2.5, with a range of 0 to 500 µg/m³, the membership functions are defined as follows: 'Low' for [0, 0, 12.75, 15] µg/m³, where concentrations below 12 µg/m³ are safe. 'Moderate' encompasses [12.75, 15, 34, 40] µg/m³, representing acceptable air quality with risks for sensitive individuals. The 'High' function is defined for [34, 40, 51, 60] µg/m³, indicating unhealthy air quality for sensitive groups. Finally, 'Very high' is defined as [51, 60, 250, 500] µg/m³, representing levels dangerous to public health.

For PM10, with a range of 0 to 999 µg/m³, the membership functions are configured as follows: 'Low' for [0, 0, 51, 60] µg/m³, where concentrations below 51 µg/m³ are considered safe. 'Moderate' encompasses [51, 60, 144.5, 170] µg/m³, representing acceptable air quality with some risks for sensitive individuals. The 'High' function spans [144.5, 170, 229.5, 270] µg/m³, indicating unhealthy air quality for sensitive groups. Finally, 'Very high' covers [229.5, 270, 600, 999] µg/m³, characterizing levels dangerous to public health.

Lastly, for the risk variable, with a range of 0% to 100%, the membership functions are defined as follows: 'Low' for [0, 0, 21.25, 25%], where the risk of contamination or health issues is low. The 'Moderate' function covers [21.25, 25, 46.75, 55%], indicating a moderate risk with the possibility of discomfort. 'High' is defined for [46.75, 55, 76.25, 85%], representing a high risk of adverse health effects, especially for sensitive groups. Finally, 'Very high' is defined as [76.25, 85, 90, 100%], where the risk is very high and requires immediate action to mitigate exposure.

These membership functions were adjusted to ensure smooth transitions between categories. This enables accurate and effective assessment of pollution and risk levels. The assessments are conducted in monitored environments.

3.5. Fuzzy rules

The fuzzy rules defined for the risk assessment system incorporate a comprehensive set of conditions. These conditions combine the levels of TVOC, CO_2, PM2.5, and PM10 to determine the overall level of risk. These rules are critical for modeling the fuzzy logic system, enabling a detailed assessment of environmental conditions and their impact on health.

For the "Low" risk class, 8 rules were established. These rules are triggered when all parameters—TVOC, CO_2, PM2.5, and PM10—are at low levels, ensuring that the overall risk is assessed as low. This set includes combinations of different levels for PM2.5 and PM10, while TVOC and CO_2 remain low.

For the "Moderate" risk class, 15 rules were defined. These rules cover a combination of conditions where at least one parameter is at moderate levels, while the others may be at low or moderate levels. They account for variations in the concentrations of TVOC, CO_2, PM2.5, and PM10, reflecting a moderate risk in environments where some factors may begin to cause discomfort.

The "High" class is represented by 18 rules. These rules are activated in situations where at least one parameter is at a high level and others may be at low or moderate levels. The combination of high levels of TVOC, CO_2, PM2.5, and PM10 may indicate significant health risks, necessitating special attention to conditions affecting air quality.

Finally, for the "Very high" risk class, 15 rules were specified. These rules apply when at least one parameter is very high, reflecting the most critical health condition. They cover scenarios where TVOC, CO_2, PM2.5, and PM10 are at dangerous levels, indicating the need for immediate intervention to protect health.

In total, the fuzzy logic system consists of 56 rules. These rules are distributed among the different risk classes. This ensures a detailed and adaptable approach to air quality assessment and management.

4. RESULTS AND DISCUSSION

The analysis of the results obtained from the fuzzy logic-based classification system reveals a robust overall performance, with an accuracy of 94.08%. However, a detailed analysis of the metrics by class, both in absolute and percentage terms, is essential to better understand the magnitude of correct and incorrect predictions and identify possible areas for improvement. Table 2 presents the confusion matrix in percentages. Table 3 presents the classification report, which includes the precision, recall, and F1-score metrics by class, as well as the corresponding support values.

Table 2. Confusion matrix in percentages

Real risk	Predicted risk				
	Low (%)	Moderate (%)	High (%)	Very high (%)	Support (%)
Low	84.00	16.00	0.00	0.00	84.00
Moderate	0.10	98.00	1.90	0.00	0.10
High	0.00	6.30	92.20	1.50	0.00
Very high	0.00	0.00	1.00	99.00	0.00

Table 3. Classification report: precision, recall, F1-score, and support by class

Class	Precision	Recall	F1-score	Support
Low	1.00	0.84	0.91	17,318
Moderate	0.91	0.98	0.95	48,260
High	0.96	0.92	0.94	27,240
Very high	0.95	0.99	0.97	7,457
Accuracy		0.94		100,275
Macro avg	0.96	0.93	0.94	100,275
Weighted avg	0.94	0.94	0.94	100,275

Class low: the low class exhibited a precision of 1.00, indicating that all 14,544 instances classified as low by the system indeed belonged to this class. This result is extremely positive, as it demonstrates that the system made no incorrect predictions for this specific category. However, the recall was 84%, suggesting that out of the 17,318 actual instances of the low class, the system correctly identified 14,544 but failed to capture the remaining 2,774 instances, which were classified as moderate. While 2,774 may seem like a high number of errors, it is essential to contextualize this value concerning the total number of samples in the low class. These errors represent approximately 16% of the actual low instances. Therefore, although there is some confusion between the low and moderate classes, the system still manages to correctly identify the majority of low instances. The perfect precision, combined with a recall of 84%, results in an F1-score of 0.91, reflecting overall good performance, although there is room for improvement in distinguishing between low and moderate.

Class moderate: the moderate class contains the largest number of samples, with a total of 48,260 actual instances. The fuzzy system successfully identified 47,289 of these instances, resulting in a recall of 98%. However, there were 971 instances that were incorrectly classified, with 39 confused with low and 932 with high. While the confusion with low is minimal, the confusion with high is more significant. With 932 moderate instances classified as high, the system demonstrated some difficulty in differentiating these two classes, which corresponds to approximately 2% of the actual moderate instances. Nevertheless, the precision of 0.91, combined with the high recall, leads to an F1-score of 0.95. Indicating that despite the confusions, the system performs robustly in correctly classifying the majority of moderate instances.

Class high: the high class, with a total of 27,240 actual instances, was correctly identified by the system in 25,122 cases, resulting in a recall of 92%. However, there was a notable confusion with the moderate and very high classes, with 1,718 instances of high being classified as moderate and 400 instances classified as very high. In percentage terms, these errors represent approximately 6.3% of the actual high instances confused with moderate and 1.5% confused with very high. These results indicate that while the precision of 0.96 is high. Suggesting that the system makes good predictions when indicating high, the recall of 92% shows that there is a small but significant degree of confusion. This confusion may be attributed to the proximity of the characteristics that define moderate and high, or high and very high, indicating a potential need for adjustments in the membership functions for these classes.

Class very high: the very high class was the smallest in terms of samples, with 7,457 actual instances. The system performed exceptionally well, with a recall of 99%, correctly identifying 7,382 of these instances. Only 75 instances were confused with high, representing approximately 1% of the actual very high instances. The precision of 0.95, coupled with a recall of 99%, results in an F1-score of 0.97. Highlighting that the fuzzy system is particularly well-suited to identifying the very high class. The minimal confusion with high suggests that for the majority of cases, the system can distinguish very high risks with high accuracy.

5. CONCLUSION

The analysis of the results obtained from the fuzzy logic-based classification system reveals a robust overall performance, with an accuracy of 94.08%. The system demonstrated satisfactory results, particularly in terms of high precision and recall for most classes. However, the greatest difficulty lies in distinguishing between adjacent classes, such as low and moderate or moderate and high, where the overlap in the membership functions leads to confusion. Although the magnitude of these errors, expressed in absolute terms, may seem significant, the relative percentages indicate that the system still correctly classifies the majority of instances within each class. These results demonstrate that the fuzzy system is capable of making accurate and reliable classifications in most cases. However, to further enhance performance, it is essential to refine the membership functions and fuzzy rules. Such improvements could provide better differentiation between adjacent risk levels, thereby increasing the effectiveness of the system in precise classification and risk-based decision-making. The contributions of this work include the validation of the use of fuzzy logic in risk classification systems, demonstrating its ability to handle uncertainties and variations in data. Furthermore, the study offers a solid foundation for future research, which could focus on exploring advanced techniques for optimizing membership functions and the application of hybrid methods that integrate other machine learning models with fuzzy logic. It is also suggested to conduct tests on different datasets and application scenarios to validate the system's versatility and robustness, ensuring its adaptability in various contexts.

ACKNOWLEDGMENTS

Not applicable.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Éldman de Oliveira Nunes	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			
Ariel Isaac Posada Barrera			✓	✓	✓	✓	✓	✓		✓	✓			
Laura Margarita Rodríguez Peralta	✓					✓	✓			✓		✓	✓	✓
Paulo Nazareno Maia Sampaio	✓					✓				✓		✓		✓
Fabián Leonardo Cuesta Astudillo			✓			✓	✓	✓		✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

Not applicable.

ETHICAL APPROVAL

Not applicable.

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author [AIPB] on request.

REFERENCES

[1] J. D. Portillo, L. M. R. Peralta, P. N. M. Sampaio, and E. D. O. Nunes, "Determining environmental indicators related to the propagation of contagious diseases and health issues: a systematic literature review," in *2023 18th Iberian Conference on Information Systems and Technologies (CISTI)*, 2023, pp. 1–5, doi: 10.23919/CISTI58278.2023.10212010.

[2] A. I. P. Barrera, L. M. R. Peralta, É. de O. Nunes, and P. N. M. Sampaio, "Influence of indoor conditions on sick building syndrome: a data-driven investigation," in *Information Technology and Systems (ICITS 2024)*, 2024, pp. 46–57, doi: 10.1007/978-3-031-54235-0_5.

[3] A. I. P. Barrera, L. M. R. Peralta, É. D. O. Nunes, P. N. M. Sampaio, and F. L. C. Astudillo, "Improving public health policies with indoor air quality predictive models," in *2023 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2023, pp. 221–226, doi: 10.1109/CSCI62032.2023.00040.

[4] A. I. P. Barrera, L. M. R. Peralta, É. de O. Nunes, P. N. M. Sampaio, and F. L. C. Astudillo, "Beyond one room: comprehensive predictive analysis of CO2 in indoor air quali," in *2024 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 2024, pp. 1–6, doi: 10.1109/LA-CCI62337.2024.10814744.

[5] A. M. -Rangel, J. Baek, T. Roh, X. Xu, and G. Carrillo, "Assessing impact of household intervention on indoor air quality and health of children with asthma in the US-Mexico border: a pilot study," *Journal of Environmental and Public Health*, vol. 2020, pp. 1–9, 2020, doi: 10.1155/2020/6042146.

[6] J. A. M. -Villalba, M. Y. L. -Porres, A. M. G. -Villa, and S. J. A. -Guerrero, "Microbiological and total particles suspended indoor air quality at the Universidad Iberoamericana Torreón," *Journal of Enviromental Sciences and Natural Resources*, vol. 6, no. 18, pp. 20–34, 2020, doi: 10.35429/JESN.2020.18.6.20.34.

[7] M. S. H. Sassi and L. C. Fourati, "Deep learning and augmented reality for IoT-based air quality monitoring and prediction system," in *2021 International Symposium on Networks, Computers and Communications (ISNCC)*, 2021, pp. 1–6, doi: 10.1109/ISNCC52172.2021.9615639.

[8] M. L. Jyothi and R. S. Shanmugasundaram, "Role of nano-sensors towards CO2 concentrations in an indoor classroom environment to improve occupational health," in *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 2022, pp. 1–7, doi: 10.1109/ICDCECE53908.2022.9792733.

[9] Y.-H. Ho, P.-E. Li, L.-J. Chen, and Y.-L. Liu, "Indoor air quality monitoring system for proactive control of respiratory infectious diseases: poster abstract," in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, 2020, pp. 693–694, doi: 10.1145/3384419.3430456.

[10] M. L. Hoang, M. Carratù, V. Paciello, and A. Pietrosanto, "Body temperature—indoor condition monitor and activity recognition by MEMS accelerometer based on IoT-alert system for people in quarantine due to COVID-19," *Sensors*, vol. 21, no. 7, 2021, doi: 10.3390/s21072313.

[11] C. Spandonidis, S. Tsantilas, F. Giannopoulos, C. Giordamlis, I. Zyrichidou, and P. Syropoulou, "Design and development of a new cost-effective internet of things sensor platform for air quality measurements," *Journal of Engineering Science and Technology Review*, vol. 13, no. 6, pp. 81–91, 2020, doi: 10.25103/jestr.136.12.

[12] B. W. Dionova, M. N. Mohammed, S. Al-Zubaidi, and E. Yusuf, "Environment indoor air quality assessment using fuzzy inference system," *ICT Express*, vol. 6, no. 3, pp. 185–194, 2020, doi: 10.1016/j.icte.2020.05.007.

[13] J. Saini, M. Dutta, and G. Marques, "ADFIST: adaptive dynamic fuzzy inference system tree driven by optimized knowledge base for indoor air quality assessment," *Sensors*, vol. 22, no. 3, 2022, doi: 10.3390/s22031008.

[14] M. Zareb, B. Bakhti, Y. Bouzid, C. E. Batista, I. Ternifi, and M. Abdenour, "An intelligent IoT fuzzy based approach for automated indoor air quality monitoring," in *2021 29th Mediterranean Conference on Control and Automation (MED)*, 2021, pp. 770–775, doi: 10.1109/MED51440.2021.9480313.

[15] R. B. Caldo, "FLC-based indoor air quality assessment for ASHRAE standard conformance," in *Industrial Engineering, Management Science and Applications 2015*, 2015, pp. 711–718, doi: 10.1007/978-3-662-47200-2_75.

[16] X. Liang, Z. Li, H. Zhang, and X. Hong, "Study of the characteristics and comprehensive fuzzy assessment of indoor air chemical contamination in public buildings," *Frontiers in Public Health*, vol. 9, 2021, doi: 10.3389/fpubh.2021.579299.

[17] J. Saini, M. Dutta, and G. Marques, "Modeling indoor PM2.5 using adaptive dynamic fuzzy inference system tree (ADFIST) on internet of things-based sensor network data," *Internet of Things*, vol. 20, 2022, doi: 10.1016/j.iot.2022.100628.

[18] A. Yasin, J. Delaney, C.-T. Cheng, and T. Y. Pang, "The design and implementation of an IoT sensor-based indoor air quality monitoring system using off-the-shelf devices," *Applied Sciences*, vol. 12, no. 19, p. 9450, 2022, doi: 10.3390/app12199450.

[19] F. Lolli, A. M. Coruzzolo, and E. Balugani, "The indoor environmental quality: a TOPSIS-based approach with indirect elicitation of criteria weights," *Safety Science*, vol. 148, 2022, doi: 10.1016/j.ssci.2021.105652.

[20] I. Naskar and A. K. Pal, "Automatic indoor air quality control of COVID-19 patient facilities using type-2 fuzzy controller," *Journal of Physics: Conference Series*, vol. 1937, no. 1, 2021, doi: 10.1088/1742-6596/1937/1/012035.

[21] M. Tennakoon, "A fuzzy inference system prototype for indoor air and temperature quality monitoring and hazard detection," *Journal of Environmental Informatics*, vol. 16, no. 2, pp. 70–79, 2010, doi: 10.3808/jei.201000179.

[22] Databot, "The future of data powered learning transform your STEAM programs today with databot!," *Databot*. 2023. Accessed: May 29, 2025. [Online]. Available: https://databot.us.com/

[23] Plantower, "The plantower PMS5003 and PMS7003 air quality sensor experiment," *Aqicn*. 2017. Accessed: May 29, 2025. [Online]. Available: http://aqicn.org/sensor/pms5003-7003/

[24] INKBIRD, "Wi-Fi 8-in-1 air quality monitor IAQM-128W," *EU INKBIRD*, [Online]. Available: https://eu.inkbird.com/products/wi-fi-air-monitor-iaqm-128w

[25] J. E. Janssen, T. J. Hill, J. E. Woods, and E. A. B. Maldonado, "Ventilation for control of indoor air quality: a case study," *Environment International*, vol. 8, no. 1–6, pp. 487–496, 1982, doi: 10.1016/0160-4120(82)90067-8.

[26] Y. Tsuchiya, "Volatile organic compounds in indoor air," *Chemosphere*, vol. 17, no. 1, pp. 79–82, 1988, doi: 10.1016/0045-6535(88)90046-X.

[27] J. J. Quackenboss, M. D. Lebowitz, and C. D. Crutchfield, "Indoor-outdoor relationships for particulate matter: exposure classifications and health effects," *Environment International*, vol. 15, no. 1–6, pp. 353–360, 1989, doi: 10.1016/0160-4120(89)90048-2.

BIOGRAPHIES OF AUTHORS

Éldman de Oliveira Nunes has a Ph.D. and M.Sc. degree in Computer Science from the University of Fluminense Federal (UFF), Brazil. He is Assistant Professor in the Lato Sensu Postgraduate Program-BigData/AI at SENAI/CIMATEC. His research interests include artificial intelligence, data science, computer vision, and image processing. He can be contacted at email: eldman.nunes@fieb.org.br.

Ariel Isaac Posada Barrera is currently pursuing a Ph.D. in Information Technology at UPAEP, Puebla. He completed an academic exchange program at Universidad del Rosario. He holds a degree in Engineering in Computational Technologies from Tecnológico de Monterrey and a Master's degree in Information Technology Management from Universidad Tecmilenio. He has professional experience in software engineering and development. His certifications include Digital Forensics Essentials (DFE) v1 from EC-Council, Data Science and Data Engineering from Data Science Dojo, and Senior Data Scientist from Tecnológico de Monterrey. Currently, he is involved in research projects on cybersecurity and indoor air quality analysis. He can be contacted at email: arielisaac.posada@upaep.edu.mx.

Laura Margarita Rodríguez Peralta holds a bachelor's degree in Computer Science from Benemérita Universidad Autónoma de Puebla (BUAP) and a Master's in Computer Science from Universidad de las Américas Puebla (UDLAP), Mexico. She obtained her DEA in Telecommunications and Systems from Joseph Fourier University, France and her Ph.D. in Computer Science and Telecommunications from Institut National Polytechnique de Toulouse (INPT), France. She has been a professor and researcher at the University of Madeira (Portugal), the University of Salvador (Brazil), and UPAEP (Mexico), where she directed the Faculty of Information Technologies. She has published extensively and participates in research projects, focusing on IoT, cybersecurity, and computer networks. She can be contacted at email: lauramargarita.rodriguez01@upaep.mx.

Paulo Nazareno Maia Sampaio completed his Ph.D. in Informatics and Telecommunications at Université Paul Sabatier (Toulouse III), with his doctoral research conducted at the Laboratory of Analysis and Architecture of Systems (LAAS-CNRS) in Toulouse, France. He has participated in several research and development projects, with the results of his work published in international conferences and scientific journals. His research and development activities include proposing solutions for sustainable cities and communities through the design and optimization of communication networks, the design and implementation of multimedia systems and virtual/augmented reality solutions, as well as the study of cybersecurity issues. He can be contacted at email: pmaia@ulima.edu.pe.

Fabián Leonardo Cuesta Astudillo is a Computer Engineer and Commercial Engineer, he is currently a Ph.D. student in Information Technology and Electronic Business at the Universidad Popular Autónoma del Estado de Puebla, he has postgraduate studies in: Master in Artificial Intelligence, Master in Neuromarketing, Master in Integral Auditing, Master in Computer Science with mention in distributed applications, Master in Business Management, Higher Diploma in Socio-Educational Research, actively participates in research projects in artificial intelligence at international level together with experts from Brazil and Mexico, actively participates in the Salesian Polytechnic University in the research group of biomedical engineering and in the research group in finance, auditing, accounting and taxation with scientific articles related to neuroscience and artificial intelligence applied to the field of engineering and business. He can be contacted at email: fabian.cuesta@upaep.edu.mx.