Mediterranean Smart Cities Conference (MSCC 2024)
2-4 May, 2024, Martil, Morocco

# Classification of Indoor CO2 Levels: Exploring the Impact of Humidity, Temperature, and Occupancy on Air Quality Using Machine Learning Model

**DANI YASSINE**
University Hassan II
Casablanca
Laboratory: Modeling and Simulation of Intelligent Industrial Systems. (M2S2I)
Mohammedia, Morocco
yassinedani98@gmail.com
https://orcid.org/0009-0003-4908-5164

BELOUAGGADIA NAOUA
University Hassan II
Casablanca
Builders Lab, Builders Ecole d'ingénieur, 14600, Epron
Laboratory: Modeling and Simulation of Intelligent Industrial Systems. (M2S2I)
Mohammedia, Morocco
n.belouaggadia@gmail.com
https://orcid.org/0000-0001-9313-2974

JAMMOUKH MUSTAPHA
University of Hassan II
Casablanca
Centre Technique de plasturgie et de caoutchouc *(CTPC)*
Casablanca, Morocco
JAMOUKH@yahoo.fr
https://orcid.org/0000-0001-7020-8788

ENNADAFY HAMZA
Hassan II University
Casablanca
Signals, Distributed Systems and Artificial Intelligence laboratory (SSDIA)
Casablanca, Morrocco
hamza.ennadafy@gmail.com
https://orcid.org/0000-0001-8309-7271

*Abstract*— **This study investigates the influence of humidity, temperature, and occupancy on indoor CO2 levels using machine learning models. Utilizing a dataset sourced from open-access data, the analysis encompasses correlation matrix exploration, logistic regression modeling, and statistical comparisons between CO2 emissions and other variables. Results reveal strong correlations between humidity, temperature, and CO2 levels, with light and occupancy emerging as significant contributor. Logistic regression modeling achieves high accuracy (97.39%) in predicting CO2 emission classes. Descriptive statistics highlight the pronounced impact of variables on CO2 levels, with humidity and occupancy exhibiting notable influence. This research provides insights into factors influencing indoor CO2 levels and suggests practical implications for IAQ improvement strategies.**

***Keywords:*** *IAQ, CO2, Humidity, Temperature, Occupancy and Logistic Regression*

## I. Introduction

Indoor air quality (IAQ) is a critical aspect of our daily lives that often goes unnoticed. As we spend a significant portion of our time indoors, the quality of the air we breathe can have profound effects on our health, well-being, and productivity [15][16].

Poor IAQ can lead to a range of health issues, from minor irritations such as headaches and fatigue to more serious conditions like respiratory diseases and even cancer [17]. It is therefore crucial to monitor and manage IAQ to ensure a healthy and productive indoor environment.

One of the key indicators of IAQ is the concentration of carbon dioxide (CO2) [18]. High levels of CO2 can indicate poor ventilation and can lead to discomfort, reduced cognitive function, and in extreme cases, health risks.

However, CO2 levels are not the only factor affecting IAQ. Other environmental parameters, such as humidity and temperature, also play a significant role. Moreover, the occupancy level of a space can influences these parameters and, consequently, the IAQ.

The research problem addressed in this study is the need to comprehend the impact of humidity, temperature, and occupancy on indoor carbon dioxide (CO2) levels. While it is well-established that indoor air quality (IAQ) significantly influences human health and productivity, there remains a lack of comprehensive understanding regarding the specific contributions of these environmental factors to CO2 concentrations within enclosed spaces.

Humidity, temperature, and occupancy are known to influence indoor CO2 levels through various mechanisms [19].

For instance, inadequate ventilation coupled with high occupancy can lead to elevated CO2 concentrations as occupants exhale $CO_2$ during respiration. Additionally, temperature and humidity levels can affect ventilation rates and the release of $CO_2$ from building materials and indoor sources, further influencing IAQ.

By addressing this research problem, we aim to shed light on the complex interactions between humidity, temperature, occupancy, and indoor $CO_2$ levels. Understanding these dynamics is crucial for developing effective strategies to monitor and mitigate indoor air pollution, thereby promoting healthier indoor environments and enhancing overall human well-being and productivity.

## II. Literature Review

Accurate prediction of indoor $CO_2$ concentration is essential for maintaining optimal air quality and occupant well-being. Research by [5] showcased a highly precise five-steps-ahead prediction model, with an average difference of less than 17 ppm from actual $CO_2$ levels, highlighting its effectiveness in forecasting. Additionally, [6] emphasized the superior accuracy of the optimized Gaussian process regression model compared to other methods, enabling better capture of complex relationships for enhanced prediction. Similarly, [7] demonstrated the effectiveness of the Adaboost community algorithm in achieving high classification success rates, enhancing indoor environmental monitoring systems' accuracy and robustness.

[8] observed varying accuracy in occupancy estimation depending on the input variables used, noting lower accuracy with $CO_2$ concentration and differential pressure data in random forest and neural network models. Multilayer Perceptron showed strong performance in predicting indoor $CO_2$ levels [1]. Decision tree and hidden Markov models effectively predicted occupancy when incorporating occupancy data [2]. Integration of machine learning methods like Ridge regression, Decision Tree, Random Forest, and Multilayer Perceptron improved $CO_2$ concentration forecasting in offices [3]. Additionally, artificial neural networks have shown promise in predicting indoor air quality, including $CO_2$ levels [4].

Ensuring accurate prediction and classification of indoor environmental parameters is essential for effective indoor air quality management and occupant comfort. In [9], it was found that the highest accuracies were achieved by training Linear Discriminant Analysis, Classification and Regression Trees, and Random Forest models.

Moreover, [10] observed a positive correlation between outdoor temperature and window-opening probability. As outdoor temperatures rise, occupants are more likely to open windows to regulate indoor temperatures, thereby influencing indoor air quality dynamics.

In contrast, [11] proposed model structures that outperformed conventional linear autoregressive and

autoregressive exogenous models in estimating indoor $CO_2$ levels.

Furthermore, [12] developed a machine-learning algorithm capable of accurately estimating unmeasured variables using a limited number of sensors.

Similarly, [13] highlighted the potential of artificial intelligence in predicting indoor air quality to enhance building environments and public health.

Additionally, [14] identified the light sensor as the most significant variable in predicting occupancy status in office rooms.

## III. Methodology

### A. Data Overview

The dataset used in this study is sourced from an open-access database by Candanedo and Feldheim [9], focusing on indoor environmental parameters including temperature, humidity, light intensity, $CO_2$ concentration, and occupancy status. Preprocessing involved removing missing values, and the dataset was constrained to 3826 observations, it's essential to acknowledge potential limitations and biases inherent in the data. These could include sampling biases, measurement errors, or incomplete representation of certain environmental conditions.

Humidity, expressed either in relative terms or as a humidity ratio (the mass of water vapor to the mass of dry air, is indicative of the moisture content in the air. This can influence indoor $CO_2$ levels as higher humidity can lead to increased human respiration rate [23], thereby increasing $CO_2$ emissions. Temperature, measured in degrees Celsius, is not only a determinant of thermal comfort but can also indirectly affect $CO_2$ levels and the functioning of sensors. The comfort level determined by the temperature can influence human presence in a given indoor environment, which in turn impacts $CO_2$ emissions due to human respiration. Furthermore, temperature can influence the pressure of a gas due to the direct relationship between these two variables according to Gay-Lussac's law [21]. This can impact the performance and accuracy of $CO_2$ sensors [20].

Occupancy, represented as a binary variable, directly impacts $CO_2$ emissions. More occupants in a room typically result in higher $CO_2$ emissions due to human respiration [22].

Light intensity, quantified in Lux, is a dependent variable of occupancy. Higher light levels can potentially indicate greater occupancy. While light intensity does not directly influence $CO_2$ levels, it can impact other variables such as temperature. For instance, increased light intensity can raise the temperature of a room, which in turn can affect human comfort and presence, thereby influencing $CO_2$ emissions due to human respiration.

$CO_2$ concentration, measured in parts per million (ppm), is a critical indicator of indoor air quality. High $CO_2$ levels can indicate poor ventilation and air exchange, which can exacerbate the buildup of indoor pollutants.

By studying these variables in relation to $CO_2$ levels (Fig 2), we aim to understand their collective impact on indoor air quality dynamics, occupant comfort, and ventilation needs. This comprehensive analysis will aid in the development of strategies for optimizing indoor air quality and promoting occupant well-being in indoor environments.

The methodology employed in this study involved several key steps to analyze indoor $CO_2$ emissions levels and identify the most influential contributors. Initially, the dataset was utilized to calculate a correlation matrix, providing insights into the relationships between different environmental parameters. Subsequently, the dataset was divided into a training set (80%) and a testing set (20%) to facilitate model development and evaluation (Fig 1).

### B. *Correlation matrix*

The correlation matrix is a fundamental tool used to analyze relationships between variables in a dataset. It provides a comprehensive overview of the linear dependencies between pairs of variables, aiding in identifying patterns and associations (1).

The correlation matrix R is represented as:

$$R = \begin{pmatrix} 1 & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & 1 \end{pmatrix} \quad (1)$$

The formula to calculate the Pearson correlation coefficient (r) between two variables X and Y is given by:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \quad (2)$$

### C. *Performance Metrics*

The training set was used to train a logistic regression model, a widely-used classification algorithm suitable for binary classification tasks. The trained model was then tested on the remaining 20% of the dataset to assess its performance using various performance metrics, including accuracy (3), precision (4), recall (5), F1-score (6), and ROC AUC score (7). These metrics provided valuable insights into the model's ability to accurately classify CO2 emission levels.

Accuracy: The proportion of correctly classified instances among the total number of instances.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (3)$$

Precision: The proportion of true positive predictions among all positive predictions made by the model.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (4)$$

Recall (Sensitivity): The proportion of true positive predictions among all actual positive instances in the dataset.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (5)$$

F1-score: The harmonic mean of precision and recall, providing a balanced measure of a model's performance.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

ROC AUC Score: The area under the Receiver Operating Characteristic (ROC) curve, indicating the model's ability to distinguish between positive and negative classes across different thresholds.

$$\text{ROC AUC Score} = \int_0^1 \text{TPR}\left(\text{FPR}^{-1}(t)\right)\, dt \quad (7)$$

### D. *Logistic regression*

Logistic regression is a popular statistical method used for binary classification tasks [27], such as predicting the probability of a binary outcome based on one or more predictor variables.

The selection of Logistic Regression was motivated by its suitability for binary classification tasks, which aligns with the nature of our research problem. As we aimed to predict binary outcomes related to indoor CO2 levels, Logistic Regression emerged as a natural choice due to its ability to model the probability of a binary outcome based on one or more predictor variables. Additionally [29], Logistic Regression provides interpretable results in terms of odds ratios, making it easier to understand the influence of each predictor variable on the outcome. Given these advantages and the simplicity of the model, Logistic Regression was deemed appropriate for our analysis.

The logistic function is defined as:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (8)$$

where z is a linear combination of predictor variables.

$$z = \beta_0 + \beta_1 . x_1 + \beta_2 . x_2 + \cdots + \beta_n . x_n \quad (9)$$

The logistic regression model predicts the probability of the positive class using this function and classifies observations based on a threshold (typically 0.5) (10). Mathematically, the logistic regression model is represented as:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}} \quad (10)$$

Furthermore, a statistical analysis was conducted on the features of each class (low emission and high emission) using the training data. By comparing the statistical characteristics of each feature between the two classes, insights were gained into the features that most significantly influenced indoor CO2 emissions levels. This analysis was complemented by a comparison with the correlation matrix, allowing for the identification of the most influential contributors to indoor CO2 levels.

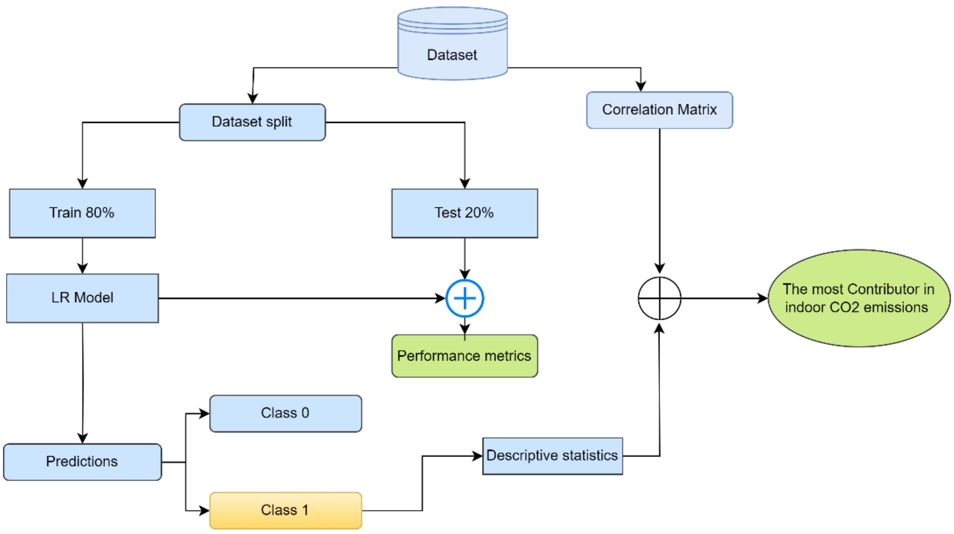


Fig 1: The followed Methodology for determining Indoor CO2 emissions most contributor

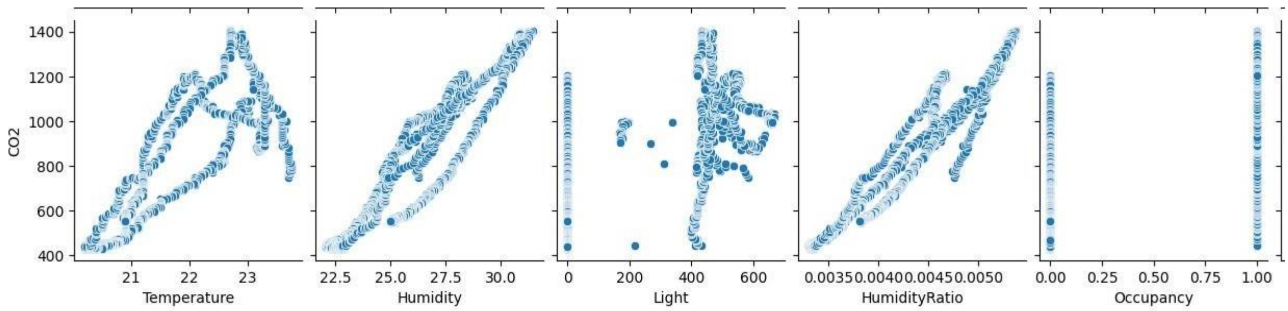


Fig 2: Visualizing CO2 variation with other features

## IV. Results

### A. *Correlation results*

In the results section, we commence with the analysis of the correlation matrix, which unveils the interplay between various variables. Through this examination, we discern the strength and direction of relationships, providing crucial insights into potential patterns within the dataset.

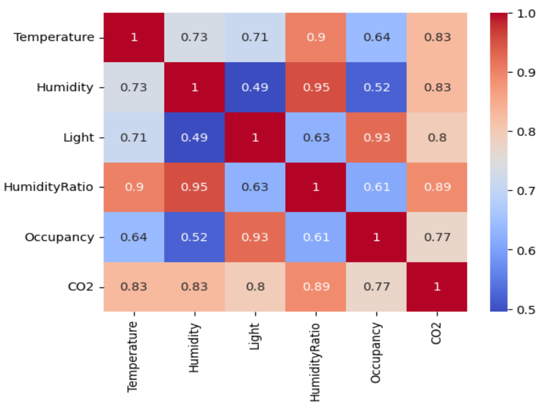


Fig 3: Correlation Matrix

The correlation matrix highlights significant relationships in the dataset. Humidity and humidity ratio show a strong correlation of 0.95, reflecting their inherent link. Occupancy and light correlate closely at 0.93, consistent with sensor functionality. However, determining CO2 emission contributors is complex due to strong correlations with multiple variables. Correlations surpassing 0.5 denote substantial interdependence, emphasizing dataset complexity. Visualizing CO2 emissions and other variables can offer clearer insights, aiding pattern recognition and facilitating comprehensive data interpretation.

To illustrate the variation of CO2 levels over time in minutes, we can create a line plot where the x-axis represents time in minutes and the y-axis represents CO2 levels in ppm (Fig 4). This visualization will provide a clear depiction of how CO2 concentrations fluctuate throughout the observation period, offering insights into temporal patterns and trends.

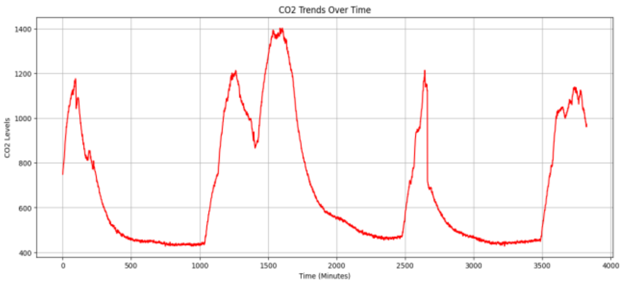


Fig 4: Indoor CO2 trends in (ppm) over time (Minutes)

After applying the logistic regression model, we obtained two distinct classes as illustrated in the figure below (Fig 5). Each data point is color-coded to represent its assigned class, facilitating a visual understanding of the model's classification outcome.

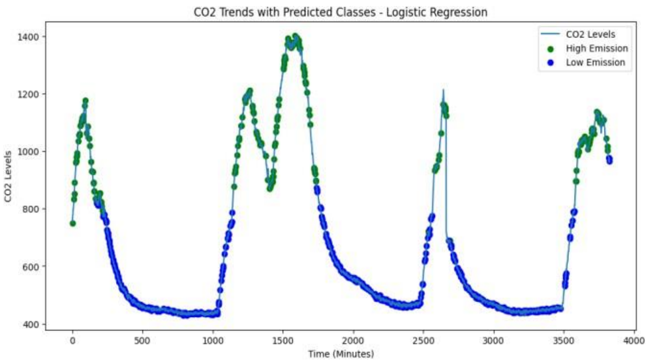


Fig 5: Classification result

The result yielded the following distribution: Class 0 -2643 instances, and Class 1 - 1183 instances.

## B. *Model Evaluation:*

TABLE 1: EVALUATION METRICS FOR LR MODEL

| **Performance metrics** | **Value** |
|---|---|
| ***Accuracy*** | 97.39% |
| ***Precision*** | 96.90% |
| ***Recall*** | 94.40% |
| ***F1-score*** | 95.63% |
| ***ROC AUC Score*** | 96.54% |

The performance metrics of the logistic regression model showcase its effectiveness in classifying indoor CO2 levels (Table 1). With an accuracy of 97.39%, the model demonstrates a high level of overall correctness in its predictions. The precision score of 96.90% indicates the model's capability to correctly identify high CO2. Despite this, the recall score of 94.40% suggests a slight tendency to miss some instances of high CO2 levels. However, the F1-score of 95.63%, which harmonizes precision and recall, reflects a robust balance between these two metrics. Lastly, the ROC AUC score of 96.54% further corroborates the model's discriminative power in distinguishing between high and low CO2 levels. These metrics collectively affirm the logistic regression model's efficacy in accurately classifying indoor air quality based on key environmental parameters.

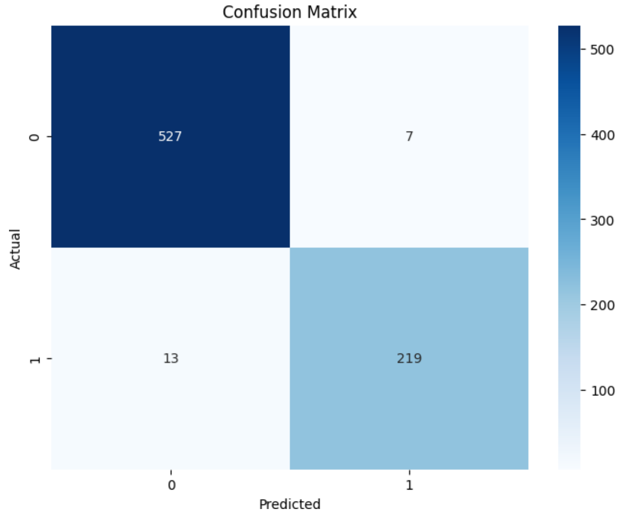


Fig 6: Confusion Matrix

The confusion matrix (Fig 6) offers a granular insight into the model's predictive accuracy by delineating true positives, true negatives, false positives, and false negatives. With only 7 false positives and 13 false negatives, the model demonstrates robust classification performance, as gleaned from the 20% testing split of the dataset.

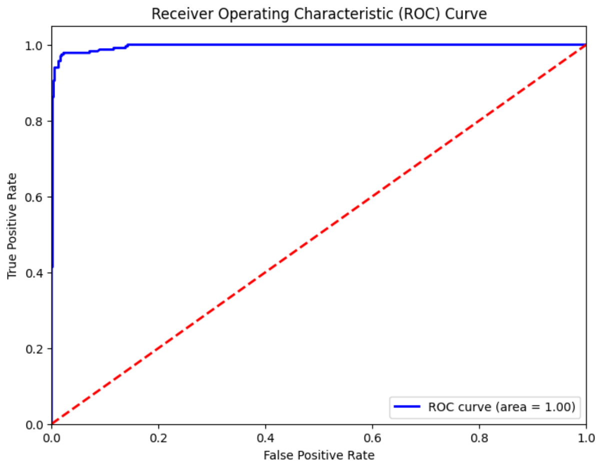


Fig 7: Receiver operating characteristic curve

An ROC AUC (7) score of 0.96 indicates that the model has a high discriminatory power, with a large area under the ROC curve. This implies that the model performs well in distinguishing between the positive and negative classes, with a high TPR and low FPR across various threshold settings (Fig 7).

TABLE 2: DESCRIPTIVE STATISTICS FOR THE FEATURES IN CLASS 0 AND 1

| | Statistics | Class 0 | Class 1 |
|---|---|---|---|
| **Temperature** | ***Mean*** | 20,96 | 22,55 |
| | ***Std*** | 0,52 | 0,64 |
| | ***Min*** | 20,2 | 21,2 |
| | ***25%*** | 20,65 | 22,1 |
| | ***50%*** | 20,89 | 22,6 |
| | ***75%*** | 21,2 | 23 |
| | ***Max*** | 23,76 | 24,41 |
| **Humidity** | ***Mean*** | 24,36 | 27,58 |
| | ***Std*** | 1,47 | 1,79 |
| | ***Min*** | 22,1 | 24,39 |
| | ***25%*** | 22,89 | 26,2 |
| | ***50%*** | 24,5 | 27,1 |
| | ***75%*** | 25,245 | 28,5 |
| | ***Max*** | 27,79 | 31,47 |
| **Light** | ***Mean*** | 51,48 | 458,43 |
| | ***Std*** | 140,06 | 177,35 |
| | ***Min*** | 0 | 0 |
| | ***25%*** | 0 | 438 |
| | ***50%*** | 0 | 464 |
| | ***75%*** | 0 | 535,69 |
| | ***Max*** | 585,2 | 1697,25 |
| **Humidity Ratio** | ***Mean*** | 0,0037 | 0,0046 |
| | ***Std*** | 0,00033 | 0,00036 |
| | ***Min*** | 0,0033 | 0,0038 |
| | ***25%*** | 0,0034 | 0,0044 |
| | ***50%*** | 0,0037 | 0,0045 |
| | ***75%*** | 0,0039 | 0,0049 |
| | ***Max*** | 0,0048 | 0,0054 |
| **Occupancy** | ***Mean*** | 0,10 | 0,84 |
| | ***Std*** | 0,31 | 0,35 |
| | ***Min*** | 0 | 0 |
| | ***25%*** | 0 | 1 |
| | ***50%*** | 0 | 1 |
| | ***75%*** | 0 | 1 |
| | ***Max*** | 1 | 1 |

In the descriptive statistics (Table 2) for each variable in classes 0 and 1, it becomes evident that all statistical values are higher in class 1. This observation reinforces the strong correlation of CO2 with other variables.

To assess the influence of each variable, we calculate the Relative Mean Difference of each variable between classes 0 and 1.

$$\mathrm{RMD} = \left(\frac{\text{Class 1 mean} - \text{Class 0 mean}}{\text{Class 0 mean}}\right) \times 100\% \qquad (11)$$

Variables with higher percentage values indicate higher influence in determining CO2 levels, as they exhibit greater disparity between the two classes.

TABLE 3: RELATIVE MEAN DIFFERENCE OF THE FEATURES

| Variables | *RMD* |
|---|---|
| ***Temperature*** | 7,56% |
| ***Humidity*** | 13,20% |
| ***Light*** | 790,40% |
| ***Humidity Ratio*** | 24,76% |
| ***Occupancy*** | 680,78% |

In conclusion, our analysis reveals that occupancy and light have the most significant impact on indoor CO2 levels (Table 3). Despite their high correlation of 0.93 (Fig 3), indicating a strong relationship, we must exclude light as a contributor due to its direct association with occupancy [28]. However, our findings underscore that the majority of indoor CO2 emissions originate from occupiers, highlighting the pivotal role of human presence in indoor air quality dynamics.

## V. DISCUSSION:

Our study highlights the significance of addressing elevated CO2 levels in indoor environments, especially as observed in class 1, where the minimum CO2 level reaches about 750 ppm. These findings exceed the recommended threshold of 1,000 ppm for healthy indoor spaces, suggesting a pressing need for intervention to mitigate potential health risks. By adhering to established guidelines categorizing CO2 levels based on ppm concentrations, our analysis underscores the importance of maintaining indoor CO2 levels within the range of 350 to 1000 ppm for optimal indoor air quality [24]. Proactive measures, such as enhancing ventilation and promoting occupant awareness, are essential to achieve this goal and create healthier indoor environments. Leveraging ongoing research and technological advancements in IAQ monitoring further supports efforts to improve indoor air quality and enhance overall well-being.

In our findings, humidity levels below 30%, particularly observed in class 0, may pose health concerns as they fall below the recommended range of 30-50% for indoor environments. This ideal range helps maintain balanced moisture levels, promoting a healthy indoor environment. Conversely, elevated humidity levels above 50% can lead to health risks and structural damage, while low humidity below 30% may result in dryness of the skin, lips, and eyes, along with potential respiratory discomfort over time [25].

TABLE 4: RANGES RECOMMENDED OF INDOOR CO2 AND HUMIDITY

| Factors | Ranges | Description |
|---|---|---|
| **Indoor CO2 level** | *350-450 ppm* | Ideal CO2 levels for well-ventilated outdoor air [24]. |
| | *450-600 ppm* | Excellent indoor air quality with ample ventilation [24]. |
| | *600-1000 ppm* | Acceptable levels for most indoor spaces but may indicate a need for increased ventilation [24]. |
| | *Above 1000 ppm* | Elevated CO2 levels that may lead to health concerns if sustained [24]. |
| **Humidity** | *30-50%* | This is the ideal range for indoor humidity [25]. |
| | *Above 50%* | High indoor humidity levels can pose serious health risks [25]. |
| | *Below 30%* | Low indoor humidity can cause dry skin, lips, and eyes [25]. |

## VI. CONCLUSION:

In conclusion, our study elucidates the critical importance of maintaining optimal indoor air quality (IAQ) by addressing elevated CO2 levels. With the observed minimum CO2 level in class 1 exceeding the recommended threshold for healthy indoor spaces, it becomes evident that immediate action is necessary to mitigate potential health risks associated with high CO2 concentrations. By adhering to established guidelines and implementing proactive measures such as enhancing ventilation and promoting occupant awareness [26], we can strive towards creating healthier indoor environments conducive to well-being and productivity. Furthermore, leveraging ongoing research and technological advancements in IAQ monitoring will continue to play a pivotal role in supporting efforts to improve IAQ and create safer, more comfortable indoor spaces for all occupants. Ultimately, our findings underscore the importance of prioritizing IAQ management as a fundamental aspect of building design and operation, with the potential to significantly impact the health and well-being of individuals in indoor environments.

## ACKNOWLEDGMENT

I extend my heartfelt appreciation to the University Hassan II for its invaluable resources and conducive research environment, which greatly facilitated the completion of this study. I am particularly grateful to my diligent supervisors for their unwavering guidance and support, instrumental in bringing this work to fruition.

## REFERENCES

[1] S. Taheri and A. Razban, “Learning-based CO2 concentration prediction: Application to indoor air quality control using demand-controlled ventilation,” Building and Environment, vol. 205, p. 108164, Nov. 2021, doi: https://doi.org/10.1016/j.buildenv.2021.108164

[2] S. H. Ryu and H. J. Moon, “Development of an occupancy prediction model using indoor environmental data based on machine learning techniques,” Building and Environment, vol. 107, pp. 1–9, Oct. 2016, doi: https://doi.org/10.1016/j.buildenv.2016.06.039

[3] J. Kallio, J. Tervonen, P. Räsänen, R. Mäkynen, J. Koivusaari, and J. Peltola, “Forecasting office indoor CO2 concentration using machine learning with a one-year dataset,” Building and Environment, vol. 187, p. 107409, Jan. 2021, doi: https://doi.org/10.1016/j.buildenv.2020.107409

[4] W. Wei, O. Ramalho, L. Malingre, S. Sivanantham, J. C. Little, and C. Mandin, “Machine learning and statistical models for predicting indoor air quality,” Indoor Air, vol. 29, no. 5, pp. 704–726, Jul. 2019, doi: https://doi.org/10.1111/ina.12580

[5] B. Khazaei, A. Shiehbeigi, and A. R. Haji Molla Ali Kani, “Modeling indoor air carbon dioxide concentration using artificial neural network,” International Journal of Environmental Science and Technology, vol. 16, no. 2, pp. 729–736, Jan. 2018, doi: https://doi.org/10.1007/s13762-018-1642-x

[6] N. R. Kapoor et al., “Machine Learning-Based CO2 Prediction for Office Room: A Pilot Study,” Wireless Communications and Mobile Computing, vol. 2022, pp. 1–16, Mar. 2022, doi: https://doi.org/10.1155/2022/9404807

[7] T. PalabaS and K. Eroğlu, “Occupancy detection from temperature, humidity, light, CO2 and humidity ratio measurements using machine learning techniques,” IEEE Xplore, May 01, 2018. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8404198 (accessed Dec. 02, 2022).

[8] J. Kim, J.-I. Bang, A.-S. Choi, J. Mun, and M. Sung, “Estimation of Occupancy Using IoT Sensors and a Carbon Dioxide-Based Machine Learning Model with Ventilation System and Differential Pressure Data,” vol. 23, no. 2, pp. 585–585, Jan. 2023, doi: https://doi.org/10.3390/s23020585

[9] L. M. Candanedo and V. Feldheim, “Accurate occupancy detection of an office room from light, temperature, humidity and CO 2 measurements using statistical learning models,” Energy and Buildings, vol. 112, pp. 28–39, Jan. 2016, doi: https://doi.org/10.1016/j.enbuild.2015.11.071

[10] X. Dai, J. Liu, and X. Zhang, “A review of studies applying machine learning models to predict occupancy and window-opening behaviours in smart buildings,” Energy and Buildings, vol. 223, p. 110159, Sep. 2020, doi: https://doi.org/10.1016/j.enbuild.2020.110159

[11] B. Vergauwen, Oscar Mauricio Agudelo, Raj Thilak Rajan, F. Pasveer, and Bart De Moor, “Data-driven modeling techniques for indoor CO2 estimation,” Oct. 2017, doi: https://doi.org/10.1109/icsens.2017.8234156

[12] A. Kaligambe, G. Fujita, and T. Keisuke, “Estimation of Unmeasured Room Temperature, Relative Humidity, and CO2 Concentrations for a Smart Building Using Machine Learning and Exploratory Data Analysis,” Energies, vol. 15, no. 12, p. 4213, Jun. 2022, doi: https://doi.org/10.3390/en15124213

[13] J. Saini, M. Dutta, and G. Marques, “Indoor air quality prediction systems for smart environments: A systematic review,” Journal of Ambient Intelligence and Smart Environments, vol. 12, no. 5, pp. 433–453, Sep. 2020, doi: https://doi.org/10.3233/ais-200574

[14] A. Mirugwe, "Accurate Occupancy Detection of an Office Room From Light, Temperature, Humidity and CO2 Measurements Using Statistical Learning Models," SSRN Electronic Journal, 2020, doi: https://doi.org/10.2139/ssrn.3686755

[15] J. Saini, M. Dutta, and G. Marques, "A comprehensive review on indoor air quality monitoring systems for enhanced public health," Sustainable Environment Research, vol. 30, no. 1, Jan. 2020, doi: https://doi.org/10.1186/s42834-020-0047-y

[16] US EPA,OAR, "Publications about Indoor Air Quality | US EPA," US EPA, Jul. 30, 2014. https://www.epa.gov/indoor-air-quality-iaq/publications-about-indoor-air-quality (accessed Jan. 16, 2020).

[17] G. D. Nielsen, S. T. Larsen, and P. Wolkoff, "Re-evaluation of the WHO (2010) formaldehyde indoor air quality guideline for cancer risk assessment," Archives of Toxicology, vol. 91, no. 1, pp. 35–61, May 2016, doi: https://doi.org/10.1007/s00204-016-1733-8

[18] M. J. Rueda López, G. Guyot, B. Golly, M. Ondarts, F. Wurtz, and E. Gonze, "Relevance of CO2-based IAQ indicators: Feedback from long-term monitoring of three nearly zero-energy houses," Journal of Building Engineering, vol. 44, p. 103350, Dec. 2021, doi: https://doi.org/10.1016/j.jobe.2021.103350

[19] Y. Liu, X. Ni, Y. Wu, and W. Zhang, "Study on effect of temperature and humidity on the CO2 concentration measurement," IOP Conference Series: Earth and Environmental Science, vol. 81, p. 012083, Aug. 2017, doi: https://doi.org/10.1088/1755-1315/81/1/012083

[20] A. Samad, D. R. Obando Nuñez, G. C. Solis Castillo, B. Laquai, and U. Vogt, "Effect of Relative Humidity and Air Temperature on the Results Obtained from Low-Cost Gas Sensors for Ambient Air Quality Measurements," Sensors, vol.20, no. 18, p. 5175, Sep.2020, doi: https://doi.org/10.3390/s20185175

[21] "Gay-Lussac's Law," University College Cork. https://www.ucc.ie/en/physics/study/undergraduate/thelaboratories/firstyearphysicslabs/gay-lussacslaw/

[22] A. Franco and F. Leccese, "Measurement of CO2 concentration for occupancy estimation in educational buildings with energy efficiency purposes," Journal of Building Engineering, vol. 32, p. 101714, Nov. 2020, doi: https://doi.org/10.1016/j.jobe.2020.101714

[23] J. C. Guevara, E. M. Yahia, R. M. Beaudry, and L. Cedeño, "Modeling the influence of temperature and relative humidity on respiration rate of prickly pear cactus cladodes," Postharvest Biology and Technology, vol. 41, no. 3, pp. 260–265, Sep. 2006, doi: https://doi.org/10.1016/j.postharvbio.2006.04.012

[24] G. Rubin, "Acceptable Levels of CO2 in Your Home: What Is Safe?," YourIAQ, Oct. 02, 2023. https://youriaq.com/acceptable-levels-of-co2-in-home/

[25] "What Is The Ideal Indoor Humidity Level For Your House," IAQ.Works, May 20, 2021. https://iaq.works/humidity/indoor-humidity-level-why-is-the-40-60-range-ideal/

[26] X. Lu, Z. Pang, Y. Fu, and Z. O'Neill, "The nexus of the indoor CO2 concentration and ventilation demands underlying CO2-based demand-controlled ventilation in commercial buildings: A critical review," Building and Environment, vol. 218, p. 109116, Jun. 2022, doi: https://doi.org/10.1016/j.buildenv.2022.109116

[27] D. W, S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression. New York, Etc.: John Wiley and Sons, Cop, 2013.

[28] Z. Luna, "Feature Selection in Machine Learning: Correlation Matrix | Univariate Testing | RFECV," Geek Culture, Aug. 10, 2021. https://medium.com/geekculture/feature-selection-in-machine-learning-correlation-matrix-univariate-testing-rfecv-1186168fac12

[29] M. P. LaValley, "Logistic Regression," Circulation, vol. 117, no. 18, pp. 2395–2399, May 2008, doi: https://doi.org/10.1161/circulationaha.106.682658