Journal of Architectural/Planning Research and Studies
(JARS) Vol.23(1).2026

https://doi.org/10.56261/jars.v23.275597

Plant Impact on Indoor Carbon Dioxide Concentration Using Ensemble Voting Prediction Models

Damrongsak Rinchumphu[1], Chinnapat Buachart[2]*, Warut Timprae[3], Sattaya Manokeaw[4], Pattaraporn Khuwuthyakorn[5], Ying-Chieh Chan[6], Worawut Kongwee[7]

[1, 2, 3, 4] Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand

[5] College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, Thailand

[6] Department of Civil Engineering, National Taiwan University, Taipei , Taiwan

[7] Integrated Science Program, Multidisciplinary and Interdisciplinary School, Chiang Mai University, Chiang Mai, Thailand

*Corresponding author email: chinapat@eng.cmu.ac.th

Received 21/12/2024 Revised 8/2/2025 Accepted 8/4/2025

Abstract

Indoor air quality, particularly carbon dioxide (CO_2) levels, is critical to occupants' health and comfort. This study developed predictive models for indoor CO_2 concentrations based on environmental variables, including light, temperature, humidity, and the presence of plants. Data collected from sensors within a controlled indoor environment were used to train predictive models using various techniques, including Artificial Neural Networks (ANN), k-Nearest Neighbors (k-NN), Random Forest, and Generalized Linear Models. Among standalone models, the ANN with a 70:30 train-test split yielded the best performance, achieving a root mean square error (RMSE) of 10.960, mean absolute error (MAE) of 7.300, and a coefficient of determination (R^2) of 0.640. The study further explored ensemble methods by combining ANN, k-NN, and Generalized Linear Models through soft voting. The optimal ensemble configuration—ANN and k-NN with a 90:10 split ratio—achieved an RMSE of 11.437, MAE of 8.153, and R^2 of 0.650, outperforming the standalone models. In addition, the results demonstrated that the presence of plants within a room reduced CO_2 levels under specific conditions (20-30°C and 200 lux), highlighting plants' potential to improve indoor air quality. This research suggests that ensemble models offer a viable solution for accurate indoor CO_2 prediction, with practical applications in indoor environmental management, especially when coupled with biophilic design elements such as indoor plants.

Keywords

Carbon dioxide; Indoor air quality; Plants; Soft voting classifier; Machine learning

1. Introduction

Maintaining good air quality in spaces regularly used by humans, especially indoor environments, is crucial to preventing potential health risks. Indoor plants commonly are used to reduce CO_2 concentration. While trees outside buildings help purify the air, indoor spaces often lack the number of plants necessary to perform this function. Plants play a vital role in reducing CO_2 through photosynthesis, which is influenced by environmental factors. The CO_2 synthesis by plants depends on these surrounding environmental conditions, meaning controlling these factors can help regulate CO_2 levels in a room effectively (Jirojwong et al., 2018).

Sensor technology is widely used for efficient data collection, providing valuable information. Factors such as air temperature, humidity, light intensity, and the species of trees present affect CO_2 levels (Candanedo & Feldheim 2016). These factors can serve as independent variables for predicting indoor CO_2 concentrations. Researchers have explored various forecasting models, but their effectiveness depends on the quality of the dataset. Manokeaw et al. (2022) developed an Artificial Neural Network (ANN) model to predict CO_2 levels in an office, producing reasonably accurate results. In our study, a voting classifier that combines ANN with other techniques is used to improve performance, as described by Imran et al. (2022). Our paper proposes an improved model using soft voting with ANN.

Significant progress has been made in predictive atmospheric quality modeling using various techniques. Among these, the Voting Classifier, which aggregates predictions from multiple models, has shown promise in ensemble prediction models for CO_2 concentration in office rooms. CO_2 is a naturally occurring greenhouse gas. In small quantities, it does not pose harm, and in fact is necessary for green plant growth, but excessive CO_2 levels can disrupt natural processes. Chemically, CO_2 consists of one carbon atom bonded to two oxygen atoms and plays an essential role in plant photosynthesis. Just as humans rely on oxygen, plants depend on CO_2 to thrive (Riham Jaber et al., 2017). Ambient CO_2 levels typically are around 400 ppm, but indoors, they can reach 1,000 ppm. It is critical to keep CO_2 levels below 1,500 ppm to avoid adverse human physiological effects such as excessive sweating, increased heart rate, and difficulty breathing. Effective ventilation systems help maintain safe CO_2 levels, protecting human health and promoting growth, especially in children (Health Canada, 2021).

In recent years, various machine learning techniques have been employed to predict indoor air quality parameters, including CO_2 concentration. Many studies have focused on standalone models such as Artificial Neural Networks (ANN), k-Nearest Neighbors (k-NN), and Random Forest (RF), but ensemble learning approaches have shown promising results in improving prediction accuracy. This study explores the effectiveness of soft voting ensemble models that combine ANN with other machine learning techniques to enhance the performance of CO_2 prediction.

The main goal of this study is to develop a robust predictive model for indoor CO_2 concentration using ensemble learning techniques. By leveraging real-time sensor data, the proposed model aims to provide an accurate estimation of CO_2 levels based on environmental variables such as temperature, humidity, light intensity, and the presence of indoor plants.

The advantages of this study include:

1. Improved CO_2 prediction accuracy through ensemble learning techniques, enabling better air quality monitoring in indoor environments.
2. Practical implications for smart building management, where predictive models can be integrated into automated ventilation control systems to maintain optimal air quality.
3. Demonstrating the potential of Snake Plants (*Sansevieria trifasciata*) in reducing indoor CO_2 concentration, supporting sustainable and biophilic design approaches in architecture.

By integrating machine learning with environmental science, and design, this research provides a foundation for data-driven strategies to enhance indoor air quality, benefiting occupants' health and well-being.

2. Background

In this section, the factors affecting the amount of CO_2 in ambient and room air are discussed. Subsequently, the machine learning models used to predict the amount of CO_2 are described.

2.1 Factors Affecting Carbon Dioxide

As discussed in the previous section, plants absorb CO_2 and produce O_2, in a process known as photosynthesis. Various factors, such as light, humidity, and temperature influence the processes essential for photosynthesis. The following sections will provide a detailed exploration and discussion of these influential factors.

2.1.1 Light intensity

Since plants need light to produce energy during photosynthesis, light is essential. Each plant has different light needs, and those with greener leaves often have a higher rate of photosynthesis than those with less green leaves or leaves of other colors because they have greater chlorophyll content.. For outdoor plants, the light energy source is shortwave radiation from the sun, while for indoor plants the light energy source may be from the sun, from grow light bulbs, or a combination of the two. Depending on how much light a plant absorbs, some plants naturally increase or decrease chlorophyll. The light needs of plants can be divided into three categories: those that need low light are those that are grown indoors and the majority of them are grown in rooms or offices with low light and cooler temperatures; those that need moderate light are those that tolerate sunlight and must be planted in a room or office building near a window or balcony where they can get some sunlight; and the last category is the plants that need high light or outdoor light. If the plant is in a darker area, it needs a grow light bulb to help it survive. However, the bulb should be placed a short distance from the plant to avoid drying the leaves. Due to plant adaptability, the light can be used both during the day and at night. Bulbs producing 2,400 candelas are the ideal light value for plants. It is recommended to turn on the light for short periods of time if there is a high illumination level; conversely, if there is a low illumination level, the light should remain on for a longer time (Dechachan, 2011; Manokeaw et al., 2022).

2.1.2 Temperature

One of the elements controlling plant photosynthesis is temperature. When the temperature climbs to 25 °C, the rate of photosynthesis increases. The rate of photosynthesis decreases with increasing temperature over this temperature limit. Time also plays a role in high temperatures; temperatures exceeding 40 °C cause disturbance. The 25-30 °C temperature range is comfortable for humans and is typical for indoor environments. As a result, indoor plants that photosynthesize in this range are appropriate for interior design (Rinchumphu et al., 2021). The interior and exterior design of a building need to be considered with respect to the suitability of temperature and comfort for plants and humans.

2.1.3 Water and humidity

Water (H_2O) is a necessary component of plant photosynthesis. It is delivered to the leaves via the phloem for photosynthesis. The soil and air humidity cause the leaves' stomata to open and close, resulting in CO_2 and O_2 diffusion. Because the stomata close during hydration, the rate of photosynthesis decreases to slow transpiration, resulting in a reduced capacity to take up CO_2. Furthermore, if terrestrial plants are in a flooded area or soils saturated with water, the roots are deprived of O_2, resulting in a decreased photosynthetic rate (Manokeaw et al., 2022). Low humidity or a dry atmosphere causes the stomata to close to prevent water loss and reduces the rate of photosynthesis since CO_2 diffusion into cells becomes limited. Furthermore, high humidity induces the stomata to open, increasing the diffusion efficiency of CO_2 into cells and hence increasing photosynthesis (Gubb et al., 2018).

2.1.4 Plants

This study collected two types of data: baseline measurements from an empty room and data with Snake Plants present to examine if plants help reduce CO_2 levels. The Snake Plant (Sansevieria trifasciata), specifically was chosen for its shade tolerance, effectiveness in air filtration, affordability, suitable size, and easy availability in the study area, as will be further described below.

The Snake Plant, scientifically named Sansevieria trifasciata and a member of the Agavaceae family, is known by various names. Sansevieria trifasciata is a long-lived herbaceous plant with rhizomes that spread along the soil surface, featuring distinctive, jointed, succulent leaves with tough or wavy edges in a variety of colors and patterns.

Adapted to arid environments, Snake Plants thrive in intense sunlight and cooler night temperatures. As a Crassulacean Acid Metabolism (CAM) plant, they open their stomata at night to minimize water loss, allowing them to absorb CO_2 for photosynthesis and capture moisture from the air. This adaptation helps them survive in dry conditions. Additionally, NASA has recognized Sansevieria trifasciata for its ability to absorb airborne pollutants, making it a highly effective plant for enhancing indoor air quality (Chiramongkolkan, 2008).

2.2 Forecasting Techniques

This research explores various forecasting techniques, including Artificial Neural Networks (ANN), Random Forest (RF), k-Nearest Neighbor k-NN, and Generalized Linear Models (GL), as a standalone model (ANN only) and also within a soft voting framework (ANN+other machine learning models), to identify the best approach for predicting indoor CO_2 concentrations . The details of each technique are described in the following sections (Imran et al., 2022).

2.2.1 Artificial Neural Network (ANN)

The Input Layer, Hidden Layer, and Output Layer comprise the three layers of ANN. The input layer initially receives data from numerous variables. The output layer gets input from experiences to analyze and locate the Hidden Layer, in which each variable is weighted to the outcome. The data are separated into two sets during this process: training and testing. For the most accurate model results, the steps in finding a hidden layer divide the data into 70% by 30% or 80% by 20% (Panyafong et al., 2020; Polat, 2012). As a result, Hidden Layers often do not have a single layer; instead, the number of layers and the node can change. Weighting data also mimics human decision-making (Boussabaine, 1996; Ranjan, 2019). The number of hidden layers and nodes necessary for the data to produce acceptable results will depend on the complexity of the research question at hand (Dechkamfoo et al., 2022).

2.2.2 Random Forest (RF)

RF is another form of the tree-based classification type. The principle of RF is based on the creation of various forms of decision trees; every tree has different model structures. After that, a vote for the best tree path is conducted. Yu et al. (2016) used the RF technique to predict the Air Quality Index (AQI) with an accuracy of 81.5 %. This level of accuracy is adequate to be used in a comparison with other forecasting techniques (Mahabub, 2020).

2.2.3 k-Nearest Neighbors (k-NN)

k-NN is a widely used forecasting technique based on classification and regression methods. Rui-jun et al. (2019) applied an improved k-NN forecasting technique to predict air quality, addressing issues related to the low accuracy and efficiency of conventional air quality meters. This approach achieved an accuracy of 94.53%, surpassing that of other established forecasting techniques (Cover & Hart, 1967).

2.2.4 Generalized Linear Model (GL)

The Generalized Linear Model (GL) forecasting technique is based on a linear model, where the dependent variable is assumed to have a linear relationship with independent variables through constant weight factors. Franklin et al. (2019) applied this approach to examine pregnant mothers' exposure to indoor air pollution and its impact on birth outcomes. They reported that higher exposure was associated with reductions in newborn birth weight and head circumference, as reflected in lower z-scores.

These forecasting techniques are useful in identifying the most effective approach for CO_2 prediction. This study aims to develop optimal forecasting models to estimate indoor CO_2 levels based on environmental factors, including light intensity, temperature, relative humidity, and the presence of plants. Conducted in Chiang Mai, Thailand, this study took place in 2022.

3. Research Methodology

3.1 Data collection

The data collection process was conducted in a controlled indoor environment to investigate the impact of plants on CO_2 concentration levels (Figure 1). The experimental setup was conducted in a closed room of 25.0 m^2 with a ceiling height of 3.0 m. The room contained a single window (1.5m × 1.2m), which remained closed during data collection to minimize external air exchange. The experiment was conducted both during the day (08:00–17:00) and night (18:00–06:00) for 7 days to analyze the impact of light conditions on CO_2 concentrations. A total of five Snake Plants (*Sansevieria trifasciata*) with an average height of 0.75 m were placed centrally in the room. No human presence was allowed during data collection to isolate the effect of plants on CO_2 reduction. The ventilation system was turned off to ensure CO_2 variations were solely due to plant activity and environmental factors.

To gather the necessary environmental data, four sensor modules were installed within the room to measure temperature (°C), relative humidity (%), light intensity (lux), and CO_2 concentration (ppm) in real-time. The sensors were calibrated before deployment to ensure accurate data collection. Data was recorded every 10 minutes over a designated experimental period and the measurements were transmitted to a cloud-based storage system (Google Cloud Platform) for further analysis (Timprae, 2021).

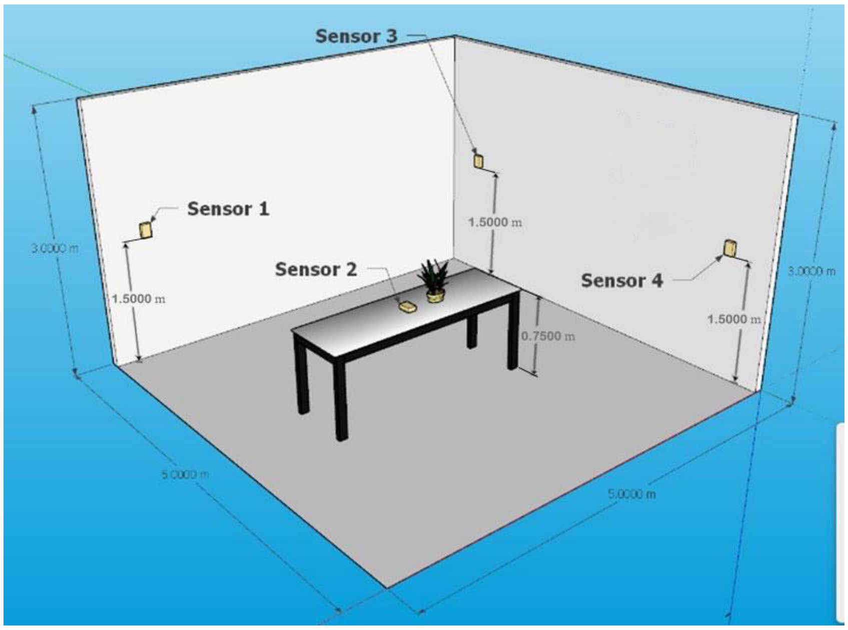

Figure 1. Data collection via sensors in the closed room.

The data collected from sensors included temperature, humidity, light intensity, and CO_2 concentration. These variables served as input features for training machine learning models, where the CO_2 concentration was the target variable to be predicted. The dataset consists of 2,196 records, which underwent preprocessing steps, including normalization and handling missing values before being used in model training. The models were trained and validated using different train-test split ratios (70:30, 80:20, and 90:10), and their performance was evaluated using RMSE, MAE, and R^2.

In this experiment, low-code computer software was used for the forecasting models. There is much low-code software available, one of which is RapidMiner Studio, a free-to-use program with limited functionality. However, the functions in the free version are sufficient for analyzing primary data that were reviewed as accurate and precise (Chaysiri & Ngauv, 2020; László & Ghous, 2020). RapidMiner Studio is a tool for data analysis that can be used to analyze a wide range of data, whether it be statistical analysis, data correlation, or forecast modeling, in order to predict the future, including sales, customer usage, or other dependent variables of interest (Sitthikankun et al., 2021). RapidMiner Studio also has been applied by many users to create forecast models. Geetha & Nasira (2014) used RapidMiner Studio to build forecast models with an accuracy of 81.78% for weather forecasts, proving that the models are sufficient and may be used for forecasts of different weather conditions. Additionally, Çelik and Başarır (2017) used RapidMiner Studio and ANN to predict the prices of precious metals like gold, silver, platinum, and palladium.

3.2 Selection of the machine learning model

The process starts with importing data from Excel and selecting the relevant columns. Next, the role of each variable is defined, categorizing them as integers, real numbers, or labels. The dataset is then divided into training and testing subsets. In this study, three different train-test split ratios were used: 70:30, 80:20, and 90:10. A higher training proportion (such as 90:10) allows the model to learn from more data but may lead to lower generalizability, while a lower training proportion (such as 70:30) reserves more data for testing, which helps assess the model's performance on unseen samples. To determine the most effective approach for CO_2 concentration prediction, both standalone and ensemble machine learning models were tested.

- Standalone ANN Model: The Artificial Neural Network (ANN) model was used independently as a baseline. ANN consists of multiple interconnected layers, where each node processes input values using weighted connections and activation functions. The model learns through backpropagation and adjusts weights to minimize errors.
- Soft Voting Ensemble Models: To improve prediction accuracy, soft voting models were implemented, combining ANN with other machine learning techniques:
    - ANN+GL (Generalized Linear Model): Combines ANN's non-linearity with the GL's statistical regression approach.
    - ANN+RF (Random Forest): Uses multiple decision trees to enhance ANN's predictive performance and robustness.
    - ANN+k-NN (k-Nearest Neighbors): Incorporates k-NN's instance-based learning with ANN's adaptive feature learning.

The forecasting equations also were reconstructed using Eqs. (1), (2) and (3) (Imran et al. 2022):

$$ML_{1p}=ML_{1}(data) \quad (1)$$

$$ML_{2p}=ML_{2}(data) \quad (2)$$

$$ML_{1}+ML_{2}=\frac{1}{2}(ML_{1p}+ML_{2p}) \quad (3)$$

In Eq. (1) and (2), ML_{1p} and ML_{2p} represent the predicted values generated by the ML_1 and ML_2 machine learning models. According to Eq. (3), the prediction is calculated by averaging the result from ML_1 and ML_2. After the ensemble vote, ML_1+ML_2 models are obtained and then tested against the test data. The trial and error of different split ratio values was conducted to obtain the best model results. In this experiment, the split ratio was divided into three types: 90:10, 80:20, and 70:30, which are the most commonly used and have been found to yield the best model results (Polat 2012). The modeling flow is presented in Figure 2.

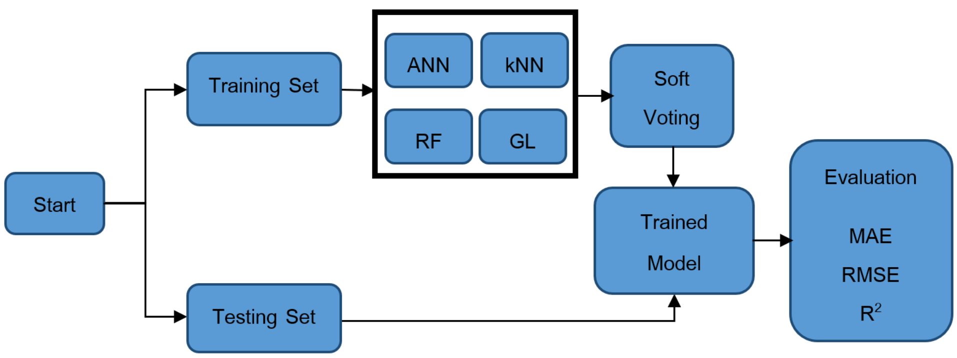

Figure 2. Modeling structure.

3.3 Measurement of model accuracy

After training, the model's performance was evaluated on the testing set using three error metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). These metrics were chosen to assess different aspects of model accuracy, allowing for a comprehensive evaluation of the predictive capabilities (Imran et al., 2022).

3.3.1 Root mean square error (RMSE)

Root mean square error (RMSE) measures the square root of the average squared difference between the predicted and actual CO_2 values. It penalizes larger errors more than smaller ones, making it useful for detecting significant deviations in predictions. A lower RMSE value indicates better model accuracy. The RMSE is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{4}$$

where n is the number of observations for the test dataset, $\hat{y}_i$ is the predicted CO_2 value, and y_i is the measured value for the i-th observation.

3.3.2 Mean absolute error (MAE)

Mean absolute error (MAE), represents the average absolute difference between predicted and measured values. Unlike RMSE, MAE does not square the errors, making it less sensitive to outliers. A lower MAE value signifies a higher degree of prediction accuracy. The formula for MAE is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{5}$$

3.3.3 Coefficient of determination (R2)

Coefficient of determination (R^2) indicates how well the independent variables explain the variance in the dependent variable (CO_2 concentration). An R^2 value closer to 1 suggests that the model can accurately predict the CO_2 levels, whereas a lower R^2 value indicates a weaker model performance (Mai et al., 2021). The formula for R^2 is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2} \tag{6}$$

where $\bar{y}_i$ is the mean (average) of the measured values, y_i.

These metrics provide a robust evaluation framework: RMSE and MAE quantify the magnitude of prediction errors, while R^2 assesses the explanatory power of the model.

4. Results and Discussion

This section presents the CO_2 concentration values measured from sensors and analyzes the effectiveness of machine learning models in predicting indoor CO_2 levels. Figures 3 and 4 illustrate the relationships between CO_2 concentration, temperature, and relative humidity, while Table 1 quantifies these relationships using correlation coefficients. Table 3 compares the accuracy of different machine learning models used for CO_2 prediction.

4.1 CO_2 concentration analysis

The measured CO_2 concentration values were analyzed under different environmental conditions, including variations in temperature and relative humidity. Figure 3 presents the average CO_2 concentration recorded at different temperature levels. The data were obtained from real-time sensor recordings and analyzed to determine the correlation between CO_2 concentration and temperature changes. The trend suggests that as temperature increases, CO_2 levels tend to decrease slightly, which may be attributed to enhanced photosynthetic activity by the Snake Plants.

Similarly, Figure 4 illustrates the relationship between CO_2 concentration and relative humidity. The data indicate that higher relative humidity levels are associated with lower CO_2 concentrations, potentially due to improved stomatal opening in plants, which facilitates CO_2 exchange.

The correlation analysis in Table 1 further quantifies these relationships, revealing a weak negative correlation between temperature and CO_2 concentration (-0.269) and a moderate positive correlation between relative humidity and CO_2 concentration (0.513). These findings confirm that both temperature and humidity play a role in influencing indoor CO_2 levels.

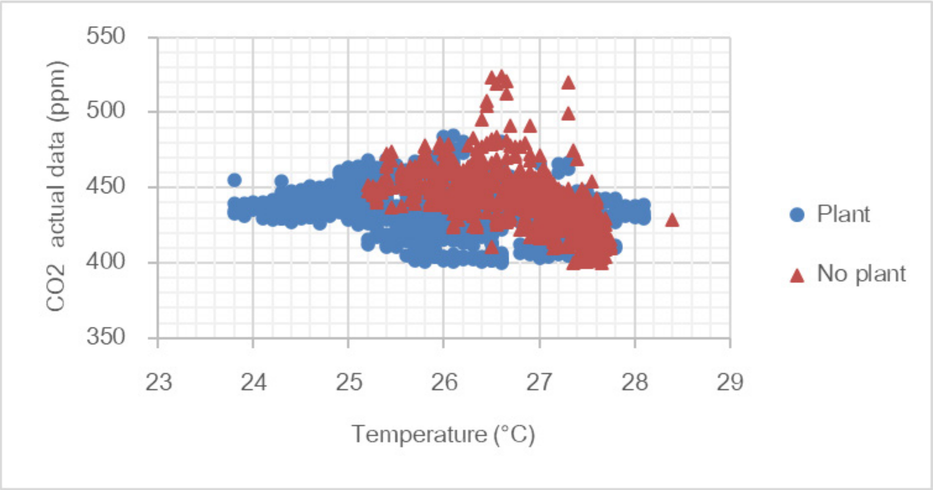

Figure 3. The average value of measured CO_2 versus temperature.

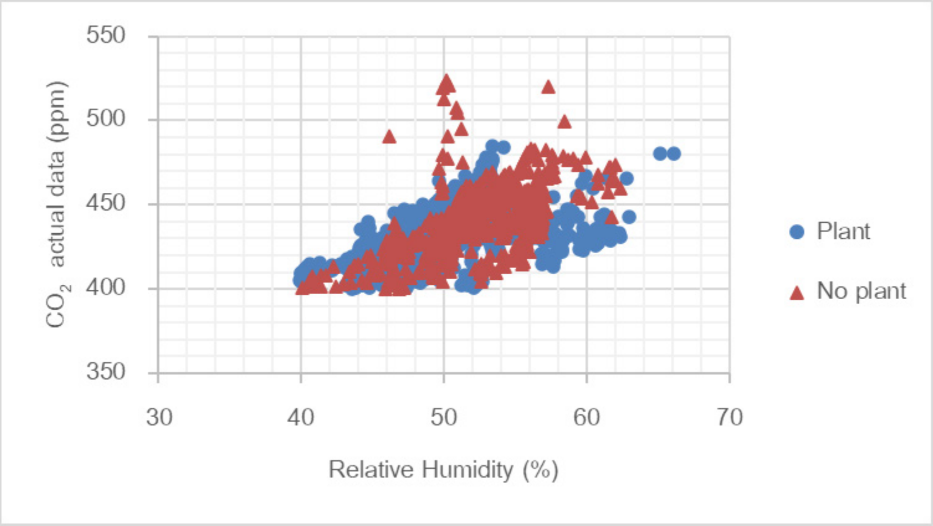

Figure 4. The average value of measured CO_2 versus relative humidity.

CO_2 concentration in an empty room without plants, in Figures 3 and 4 illustrate the relationships between CO_2 concentration, temperature, and relative humidity under different conditions, including when the room is empty without plants. When no plants are present, CO_2 concentration primarily is influenced by ambient temperature and air exchange. In an unventilated space, CO_2 tends to accumulate over time due to human respiration (if occupants are present) or minimal air circulation. However, as temperature increases, CO_2 levels exhibit a slight decline. This can be attributed to the expansion of air at higher temperatures, leading to increased air movement and slight diffusion of CO_2 away from the measurement zone. Additionally, at higher temperatures, buoyancy-driven air currents may enhance ventilation, resulting in gradual CO_2 reduction (Ateş & Khameneh, 2023). When the plants are absent, the relationship between CO_2 and humidity is largely dictated by room ventilation and external air infiltration. Higher humidity levels often coincide with higher CO_2 concentrations, particularly in enclosed environments where moisture retention correlates with limited air exchange. The presence of humidity may reduce CO_2 diffusion efficiency, leading to localized CO_2 buildup. Conversely, when humidity decreases, the room's air tends to be drier, which can facilitate better CO_2 dispersion and slight reductions in measured concentration (Jiang et al., 2024).

These trends indicate that in an empty room without plants, CO_2 accumulation primarily is affected by air circulation, temperature-driven buoyancy effects, and humidity-induced air retention. In contrast, when plants are present, their photosynthetic activity contributes to a reduction in CO2 concentration, altering these relationships.

Table 1. Correlation Coefficients between Temperature, Relative Humidity, and CO2 ($p < 0.001$).

Variables	**Temperature (°C)**	**Relative humidity (%)**	**CO_2 (ppm)**
Temperature (°C)	1		
Relative humidity (%)	–0.155	1	
CO_2 (ppm)	–0.269	0.513	1

4.2 Model performance evaluation

To assess the accuracy of CO_2 prediction models, different machine learning approaches were tested. Figure 5 shows the architecture of ANN to predict the concentration of CO2 (Manokeaw et al., 2022). The details of ANN used in this study are listed in Table 2. The activation functions used in the hidden layers were a sigmoid function while a linear function was used in the output layer. Table 3 summarizes the RMSE, MAE, and R^2 values for each model, providing a comparative analysis of their performance. Among standalone models, the ANN model with a 70:30 train-test split achieved the best results, with an RMSE of 14.003, MAE of 10.869, and R^2 of 0.397.

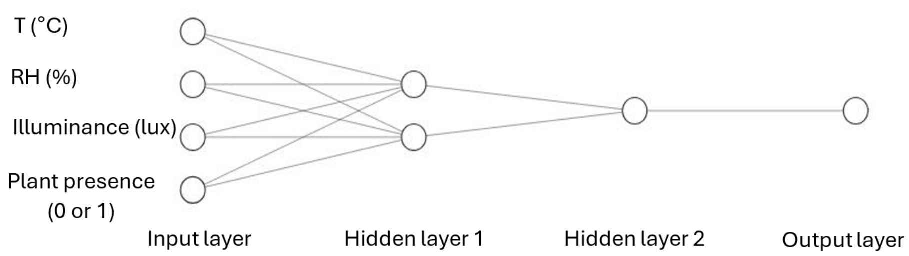

Figure 5. The architecture of ANN model used in this study (Plant presence, 0 = No; 1 = yes).

Table 2. Details of ANN Input, Values of Weights and Bias[1].

Layer	Node	Activation Function	Weights	Bias
Hidden 1	1	Sigmoid	T (°C): 1.474, RH (%): -2.482, lux: 0.829, Plant Presence: -0.435	-1.354
	2	Sigmoid	T (°C): 0.251, RH (%): -0.453, lux: -0.576, Plant Presence: 0.999	0.039
Hidden 2	1	Sigmoid	Node 1: -1.998, Node 2: -0.995	-0.216
Output	1	Linear (Regression)	Node 1: 1.444	0.049 (Threshold)

[1] Abbreviations: T (°C) = Temperature in degrees Celsius, RH (%) = Relative Humidity in percentage, lux = Illuminance in lux (unit of light intensity) and Plant Presence = Presence of plant (0 = No, 1 = Yes).

However, ensemble models improved prediction accuracy. The best-performing model, ANN+k-NN (90:10 split), achieved an RMSE of 11.437, MAE of 8.153, and R^2 of 0.650, demonstrating notable improvement over standalone models.

Table 3. Comparison of RMSE, MAE, and R2 for CO2 Standalone and Ensemble Models.

Split ratio	Model	RMSE	MAE	R^2
90:10	ANN	15.050	11.049	0.349
	ANN+k-NN	11.437	8.153	0.650
	ANN+RF	13.205	9.715	0.513
	ANN+GL	15.384	11.647	0.319
80:20	ANN	14.031	10.779	0.376
	ANN+k-NN	11.194	8.238	0.602
	ANN+RF	12.483	9.529	0.505
	ANN+GL	14.260	11.056	0.354
70:30	ANN	14.003	10.869	0.397
	ANN+k-NN	11.218	8.432	0.616
	ANN+RF	12.508	9.590	0.519
	ANN+GL	14.089	11.064	0.386

Figure 6 compares the percentage error among three models: the best soft voting model (ANN+k-NN), the standalone ANN model, and the least effective model (ANN+GL). The figure shows that the ANN+k-NN model consistently maintains errors within a 10% range, whereas the other two models exhibit larger deviations, exceeding 15% in some cases. This confirms the robustness of the ANN+k-NN model in predicting CO_2 concentrations.

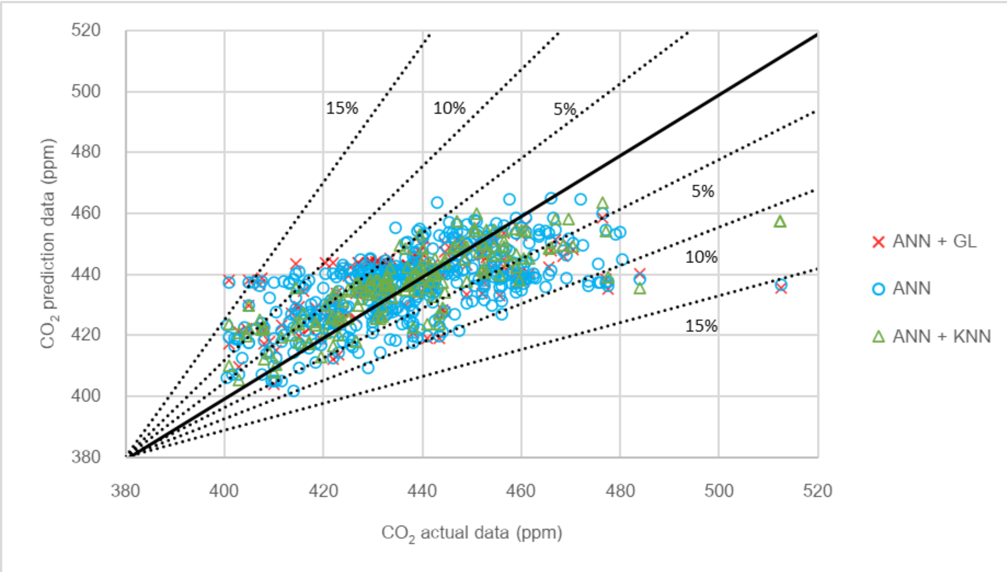

Figure 6. Comparison of percentage error among the best model (ANN+k-NN), the standalone ANN model, and the least effective model (ANN+GL).

4.3 Impact of Indoor Plants on CO_2 Reduction

To investigate the effect of indoor plants on CO_2 concentration, comparisons were made between rooms with and without Snake Plants. Figure 7 shows that the presence of Snake Plants led to a reduction in CO_2 levels, particularly within the 20–30°C temperature range under a light intensity of 200 lux. The greatest reduction was observed at approximately 28.5°C, suggesting that this temperature range is optimal for CO_2 absorption by the plants.

These findings support the hypothesis that indoor plants can effectively reduce CO_2 concentration under specific environmental conditions. However, at temperatures exceeding 30°C, the CO_2 reduction rate declined, indicating that extremely high temperatures may limit plant mitigation efficiency.

4.4 Limitations, Assumptions and Sensitivity Analysis

While the results provide valuable insights into indoor CO_2 dynamics, some limitations and assumptions must be considered:

1. The dataset was collected in a controlled environment with no human presence, which may not fully reflect real-world indoor conditions where human respiration contributes to CO_2 levels.
2. Only one plant species (Snake Plant) was tested. Future studies should explore different plant types to determine their relative effectiveness in CO_2 absorption.
3. The experiment was conducted in a room with a single closed window (1.5m × 1.2m), minimizing external air exchange. If the room had multiple windows or active ventilation, CO_2 levels might fluctuate differently.
4. The impact of plant size and quantity was not fully explored. If a larger number of Snake Plants was placed in the room, CO_2 uptake might have been greater. Conversely, if the plant size were smaller, the CO_2 reduction effect might be diminished.

Additionally, if some assumptions change, such as using a room without windows or altering the number of plants, the CO_2 reduction dynamics would likely differ. For example:

1. If the room had no windows and limited air circulation, the CO_2 reduction effect from plants might become more pronounced, as there would be less external air exchange.
2. If the number of Snake Plants were doubled, CO_2 absorption might increase proportionally, but only up to a certain threshold, beyond which the effect could plateau due to environmental limitations.
3. Conversely, if fewer plants were used, the CO_2 reduction effect would be lower, possibly resulting in higher CO_2 levels in the room atmosphere over time.

Despite these limitations, this study demonstrates the feasibility of using ensemble machine learning models for accurate CO_2 prediction and highlights the practical benefits of incorporating indoor plants for air quality management.

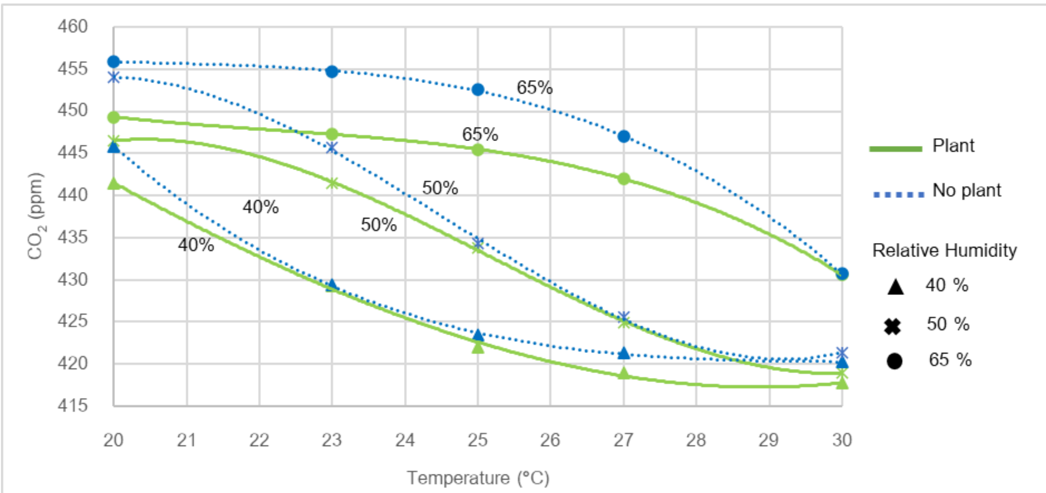

Figure 7. Comparison of CO_2 forecasts from the ANN+k-NN model between plant and non-plant conditions.

5. Conclusions

This study developed a machine learning model to predict CO_2 concentrations based on data collected from an empty room using four sensors to measure relative humidity, temperature, light intensity, and CO_2 levels. Standalone ANN models and soft voting models (ANN+GL, ANN+RF, and ANN+k-NN) were tested across split ratios of 70:30, 80:20, and 90:10, with their accuracy evaluated using RMSE, MAE, and R^2 metrics.

The key findings are as follows:

1. Soft voting improved the R^2 efficiency of the standalone ANN model. The standalone ANN models yielded R^2 values of 0.349, 0.376, and 0.397 for split ratios of 70:30, 80:20, and 90:10, respectively. The best-performing soft voting model, ANN+k-NN, achieved R^2 values of 0.650, 0.602, and 0.616 for these split ratios, demonstrating notable improvement over the standalone ANN.
2. A plot of measured versus predicted CO_2 data shows that the ANN+k-NN model maintained a percentage error within 15%, with most errors being less than 10%. This indicates that ensemble learning techniques can enhance prediction accuracy and stability.
3. Using the ANN+k-NN model to forecast CO_2 levels in rooms with and without Snake Plants revealed that the presence of Snake Plants effectively reduced CO_2 in the temperature range of 20–30°C, under light intensity of 200 lux. The CO_2 reduction was most pronounced at temperatures around 28.5°C.

5.1 Benefits of Snake Plants for Indoor CO_2 Reduction

The results highlight the potentially important role of Snake Plants in improving indoor air quality by reducing CO_2 levels. The key benefits include:

- Snake Plants actively absorb CO_2 during the night through Crassulacean Acid Metabolism (CAM) photosynthesis, making them particularly effective in enclosed spaces with minimal ventilation.
- The ability of Snake Plants to survive in low light conditions makes them ideal for indoor environments such as offices, classrooms, and residential spaces where CO_2 levels can accumulate.
- By lowering CO_2concentration, Snake Plants may contribute to improved cognitive function, enhanced productivity, and overall occupant well-being, particularly in high-occupancy areas.

5.2 Benefits of the Machine Learning Models

The application of machine learning models in this study demonstrates their potential in predicting and managing indoor CO_2 levels. The key benefits include:

- The ANN+k-NN model provided a more accurate and robust prediction of CO_2 concentration compared to standalone models, highlighting the advantage of ensemble learning.
- These models can be integrated into smart indoor air quality management systems, allowing for real-time monitoring and optimization of ventilation strategies based on predicted CO_2 trends. Moreover, they offer promising applications when embedded within Building Information Modeling (BIM) and digital twin frameworks for enhanced building performance analysis and decision-making.
- The ability to predict CO_2 fluctuations based on environmental factors (temperature, humidity, and light) enables proactive interventions, such as adjusting HVAC (Heating, Ventilation, and Air Conditioning) settings or recommending indoor plant placements for air purification.

These findings suggest that a combination of machine learning models and biophilic design strategies, such as incorporating Snake Plants in indoor environments, can enhance indoor air quality management. Future research should explore the impact of different plant species and environmental conditions to further refine predictive models for CO_2 optimization in various indoor settings.

6. Practical Applications and Future Work

The findings from this study provide valuable insights into indoor air quality management and predictive modeling for CO_2 concentration. The developed machine learning models and experimental results have multiple real-world applications, particularly in smart buildings, environmental monitoring, and health-focused indoor spaces.

6.1 Applications in Smart Buildings

The CO_2 prediction models developed in this study can be integrated into smart building management systems to optimize ventilation strategies. By predicting indoor CO_2 concentration based on temperature, humidity, and light levels, automated HVAC systems can adjust airflow and fresh air intake in real-time to maintain air quality.

For example:

- In office buildings, where high occupancy levels contribute to CO_2 buildup, predictive models can help dynamically control air circulation to enhance worker productivity.
- In schools and universities, CO_2 prediction models can be used to regulate classroom ventilation, ensuring that students and teachers have access to fresh air to improve cognitive performance.
- In residential smart homes, AI-driven automation can use CO_2 forecasts to activate air purifiers, open windows, or adjust plant positioning to optimize indoor air quality.

6.2 Environmental and Health Benefits

Reducing indoor CO_2 improves human well-being. Prolonged exposure to elevated CO_2 levels (above 1,000 ppm) can cause headaches, drowsiness, and decreased cognitive function. The ability to predict and manage CO_2 levels ensures healthier indoor environments, benefiting long-term health.

Indoor plants as natural air purifiers. The study demonstrates that Snake Plants can effectively reduce CO_2 concentration, particularly at temperatures around 28.5°C. This supports the implementation of biophilic design strategies in urban architecture, where plants are integrated into indoor spaces to enhance air quality. Energy efficiency improvements. By using machine learning models to predict CO_2 fluctuations, HVAC systems can be operated more efficiently, reducing unnecessary energy consumption while maintaining optimal air quality.

6.3 Future Research Directions

While this study demonstrates the effectiveness of ensemble machine learning models in CO_2 prediction, further research can extend its applications in the following ways:

1. Expanding plant species analysis: Future work should investigate how different plant types contribute to CO_2 absorption under varying environmental conditions.
2. Incorporating real-time human activity data: Adding human occupancy levels as a factor can improve model accuracy in predicting CO_2 fluctuations in dynamic indoor environments.
3. Deploying IoT-based smart monitoring systems: Integrating AI models with IoT sensors can enable real-time CO_2 tracking and automated decision-making for ventilation control.

By combining advanced machine learning techniques with sustainable indoor environmental design, this research contributes to the development of intelligent air quality management solutions that enhance human health and energy efficiency.

Acknowledgements

The authors thank support staff from the City Research and Development Center, Faculty of Engineering, Chiang Mai University.

Funding

This research has received funding support from the NSRF via the Program Management Unit for Human Resources and Institutional Development, Research and Innovation [grant number B16F640189].

Conflicts of Interest

The authors declare no conflict of interest.

Data Availability Statement

Data may be obtained from the corresponding author upon reasonable written request.

Use of Generative Artificial Intelligence (AI) and AI-Assisted Technologies

ChatGPT-4.0 was used solely to improve the clarity and readability of the language during manuscript preparation, without contributing to the scientific content.

Author Contributions

D.R., C.B.: Conceptualization; W.T.: Software; S.M., W.K.: Data curation; P.K., Y.C.C.: Validation; D.R.: Writing – original draft; C.B.: Writing – review & editing.

References

Ateş, E., & Khameneh, E. T. (2023). Effects of the number of people, temperature, relative humidity, and CO2 parameters on indoor air quality in higher education institution classrooms. *Edelweiss Applied Science and Technology, 7*(2), 164-181.

Boussabaine, A. H. (1996). The use of artificial neural networks in construction management: A review. *Construction Management and Economics, 14*(5), 427-436.

Candanedo, L. M., & Feldheim, V. (2016). Accurate occupancy detection of an office room from light, temperature, humidity and CO_2 measurements using statistical learning models. *Energy and Buildings, 112*, 28-39.

Çelik, U., & Başarır, Ç. (2017). The prediction of precious metal prices via artificial neural network by using RapidMiner. *Alphanumeric Journal, 5*(1), 45-54.

Chaysiri, R., & Ngauv, C. (2000). Prediction of closing stock prices using the artificial neural network in the Market for Alternative Investment (MAI) of the Stock Exchange of Thailand (SET). *International symposium on integrated uncertainty in knowledge modelling and decision making* (pp.335-345). Springer.

Chiramongkolkan, U. (2008). *Saint George's sword* (5th ed.). Amarin Printing and Publishing.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions On Information Theory, 13*(1), 21-27.

Dechachan, R. (2011). *Guide to planting ornamental plants in office building*. Thaiqualitybooks.

Dechkamfoo, C., Sitthikankun, S., Kridakorn Na Ayutthaya, T., Manokeaw, S., Timprae, W., Tepweerakun, S., Tengtrairat, N., Aryupong, C., Jitsangiam, P., & Rinchumphu, D. (2022). Impact of rainfall-induced landslide susceptibility risk on mountain roadside in northern Thailand. *Infrastructures, 7*(2), 17.

Franklin, P., Tan, M., Hemy, N., & Hall, G. L. (2019). Maternal exposure to indoor air pollution and birth outcomes. *International Journal of Environmental Research and Public Health, 16*(8), 1364.

Geetha, A., & Nasira, G. (2014). Artificial neural networks' application in weather forecasting–using RapidMiner. *International Journal of Computational Intelligence and Informatics, 4*(3), 177-182.

Gubb, C., Blanusa, T., Griffiths, A., & Pfrang, C. (2018). Can houseplants improve indoor air quality by removing CO_2 and increasing relative humidity?. *Air Quality, Atmosphere & Health, 11*(10), 1191-1201.

Health Canada. (2021). *Residental indoor air quality guidelines: Carbon dioxide.*

Imran, H., Ibrahim, M., Al-Shoukry, S., Rustam, F., & Ashraf, I. (2022). Latest concrete materials dataset and ensemble prediction model for concrete compressive strength containing RCA and GGBFS materials. *Construction and Building Materials, 325*, 126525.

Jiang, J., Irga, P., Coe, R., & Gibbons, P. (2024). Effects of indoor plants on CO2 concentration, indoor air temperature and relative humidity in office buildings. *PLoS One, 19*(7), e0305956.

Jirojwong, P., Sangkanchanavanich, C., Sreshthaputra, A., Pongsuwan, S., & Choruengwiwat, J. (2018). Comparative study of air exchange rates in a high-rise residence by using both tracer gas dilution and fan pressurization methods. *Journal of Architectural/Planning Research and Studies (JARS), 15*(1), 127-134.

László, K., & Ghous, H. (2020). Efficiency comparison of Python and RapidMiner. Multidiszciplináris Tudományok, 10(3), 212-220.

Mahabub, A. (2020). A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers. *SN Applied Sciences, 2*(4), 1-9.

Mai, H.-V. T., Nguyen, T.-A., Ly, H.-B., & Tran, V. Q. (2021). Prediction compressive strength of concrete containing GGBFS using random forest model. *Advances in Civil Engineering, 2021.*

Manokeaw, S., Nim-Anutsonkun, T., Chaiya, T., Timprae, W., & Rinchumphu, D. (2022). Assessment of CO_2 reduction potential of indoor plants using artificial neural network in classrooms. *Journal of Hunan University Natural Sciences, 49*(5).

Panyafong, A., Neamsorn, N., & Chaichana, C. (2020). Heat load estimation using Artificial Neural Network. *Energy Reports, 6*, 742-747.

Polat, G. (2012). ANN approach to determine cost contingency in international construction project. *Journal of Applied Management and Investments, 1*(2), 195-201.

Ranjan, C. (2019). *Rules-of-thumb for building a Neural Networ*k. Retrieved August 22, 2023, from https://towardsdatascience.com/17-rules-of-thumb-for-building-a-neural-network-93356f9930af

Riham Jaber, A., Dejan, M., & Marcella, U. (2017). The effect of indoor temperature and CO2 levels on cognitive performance of adult females in a university building in Saudi Arabia. *Energy Procedia, 122*, 451-456.

Rinchumphu, D., Phichetkunbodee, N., Pomsurin, N., Sundaranaga, C., Tepweerakun, S., & Chaichana, C. (2021). Outdoor thermal comfort improvement of campus public space. *Advances in Technology Innovation, 6*(2), 128.

Rui-jun, Y., Dan-feng, D., & Feng, Y. (2019). Application of improved KNN algorithm in air quality assessment. *Proceedings of the 2019 3rd High Performance Computing and Cluster Technologies Conference* (pp.108-112).

Sitthikankun, S., Rinchumphu, D., Buachart, C., & Pacharawongsakda, E. (2021). Construction cost estimation for government building using artificial neural network technique. *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies, 12*(6), 1-12.

Timprae, W. (2021). *Effective environmental factor to photosynthetic active radiation in greenhouse planting* [Master's thesis, Chiang Mai University].

Yu, R., Yang, Y., Yang, L., Han, G., & Move, O. A. (2016). RAQ–A random forest approach for predicting air quality in urban sensing systems. *Sensors, 16*(1), 86.