Better Air, Better Health:
Creating an indoor Air Quality Monitoring and Predictive System

MARIA JOSE MOLINA-CONTRERAS, PhD
MARIA JOSE MOLINA-CONTRERAS, PhD

PhD in Plant Molecular Biology

Data Scientist

DIY, Microcontrollers & IoT

@MJ_BIO
@mjmolina
mjmolinacontreras
introduction with career path
motivation to develop this project
Factors Contributing to Poor Indoor Quality

- Occupants
- Humidity and moisture
- Ventilation design
- Pollutants sources
- Odors and gases
Poor Indoor quality reasons
Is CO2 an Indoor Pollutant? Direct Effects of Low-to-Moderate CO2 Concentrations on Human Decision-Making Performance

Low Level Carbon Dioxide Indoors—A Pollution Indicator or a Pollutant? A Health-Based Perspective

How indoor environmental quality affects occupants’ cognitive functions: A systematic review
scientific publications about CO2 in human health
**CO2: Ranges & Health**

- **Outdoor Environment**
  - CO2 ppm (part per million): <350
  - Indicates a healthy environment.

- **Indoor with good air exchange**
  - CO2 ppm: 350-1000
  - Poor Air Quality (Feel sleepy)

- **Indoor with poor air exchange**
  - CO2 ppm: 1000-5000
  - Very Poor Air Quality: not recommended to be >8 hours
    - Symptoms: Downsiness, Headaches, Loss of attention, Nausea, Others

- **Extreme bad Air Quality**
  - CO2 ppm: >5000
  - Permanent damages

Adapted from: https://www.airthings.com/
co2 ranges and health
project structure
Sensors devices

PMSA003I Air Quality Breakout (Adafruit)
Particles: PM1.0, PM2.5 and PM10.0 concentration in both standard & environmental units

Data collection

Air Pollutants

SCD-30 Sensirion (Adafruit)
Co2: range 400-10000 ppm
Speaker notes

sensor devices models and data collection
Monitoring system: First idea (QT py)
Monitoring system: particles sensor devices running circuitpython
Monitoring system: Second idea

MicroPython
Speaker notes

monitoring system: co2 sensor running micropython
What’s next?
Data collection
devices for data collection
Data collection & Monitoring system

SCD-30
CO2
PMSA0031I
Air Quality Breakout

Raspberry Pi 4
Cables

Raspberry Pi Zero W
Project Structure

Monitoring

TCP Socket in Python

Sensor1

Sensor2

Data collection
Speaker notes

Project structure with monitoring system
Monitoring system: Data

- **Co2**
- **Humidity**
- **°C**
monitoring system data co2, humidity
Monitoring system: Data
monitoring system for co2 in my apartment.
Project Structure

Data collection

Predictive

Machine Learning Predictions

Ventilation activation
project structure: predictive system
Values

Time

Predictive system: Data type
predictive system time series introduction
How can we use this information to predict values?
how can we use this information to predict values? plot with CO2 values in times.
A Practical and Adaptive Approach to Predicting Indoor CO₂

Giacomo Segala ¹,²,³,* ②, Roberto Doriguzzi-Corin ² ③, Claudio Peroni ¹, Tommaso Gazzini ¹ and Domenico Siracusa ² ③

Appl. Sci. 2021, 11, 10771

Table 1. Summary of the SoA.

<table>
<thead>
<tr>
<th>Related Work</th>
<th>Dataset Size</th>
<th>Input Variables</th>
<th>AI Architecture</th>
<th>Automated Model Update</th>
</tr>
</thead>
<tbody>
<tr>
<td>This Work</td>
<td>Adaptive (Max 30 days)</td>
<td>Temperature, Humidity, CO₂</td>
<td>1D CNN</td>
<td>Yes</td>
</tr>
<tr>
<td>Vanus et al. [7]</td>
<td>One year</td>
<td>Temperature, humidity, time, date</td>
<td>Random Forest</td>
<td>No</td>
</tr>
<tr>
<td>Khorram et al. [8]</td>
<td>242 days</td>
<td>CO₂, weekday, hour, minute</td>
<td>ANN</td>
<td>No</td>
</tr>
<tr>
<td>Ahn et al. [9]</td>
<td>Six months</td>
<td>Fine dust, light amount, VOC, CO₂, temperature and humidity</td>
<td>GRU</td>
<td>No</td>
</tr>
<tr>
<td>Kallio et al. [10]</td>
<td>One year</td>
<td>CO₂, PIR, temperature and humidity</td>
<td>Ridge, Decision Tree, Random Forest, MLP</td>
<td>No</td>
</tr>
<tr>
<td>Sharma et al. [11]</td>
<td>One week</td>
<td>Indoor NO₂, wind speed, wind direction, number of student</td>
<td>LSTM</td>
<td>No</td>
</tr>
<tr>
<td>Putra et al. [12]</td>
<td>One week</td>
<td>CO₂</td>
<td>ANN</td>
<td>No</td>
</tr>
<tr>
<td>Khazaei et al. [13]</td>
<td>One week</td>
<td>CO₂, humidity, temperature</td>
<td>MLP</td>
<td>No</td>
</tr>
<tr>
<td>Skön et al. [14]</td>
<td>Six months</td>
<td>Temperature, humidity</td>
<td>MLP</td>
<td>No</td>
</tr>
</tbody>
</table>
different options practical predicting indoor co2 that we could use to predict co2
Monitored variables:

- Carbon dioxide concentration [ppm]
- Humidity [%]
- Temperature [°C]
- Particles [ppm]
- Activity level (occupancy of the room)

Dataset creation: The data is not being randomly shuffled before splitting

Normalization: It is a common way of doing this scaling (only be computed using the training data)
Predictive system: variables, dataset creation, normalization
Windowing: Based on window of consecutive samples from the data.

N = number of days or hours
windowing system to predict our co2 values.
Neural Network:

1-dimensional convolutional neural network (CNN)

Illustration from "Python for Finance Cookbook"
Predictive system: Methodology

Analyzing and extracting features along the time-dimensional axis of the input data.

```
1 model_cnn = Sequential()
2 model_cnn.add(Conv1D(filters=64, kernel_size=3,
3    activation='relu',
4    input_shape=(WINDOW_SIZE,features))
5 model_cnn.add(MaxPooling1D(pool_size=2))
6 model_cnn.add(Flatten())
7 model_cnn.add(Dense(50, activation='relu'))
8 model_cnn.add(Dense(1))
```

Input environmental variables (i.e., temperature, humidity, CO2...) covering windows with 10 minutes input.

It aims to learn the most valuable information.

Predicts the CO2 value for the next hour.
CNNs are less computationally demanding, making them suitable for limited-power and resource-constrained devices.

Neural Network Architecture:

1-dimensional convolutional neural network (CNN)

Input layer: input environmental variables (i.e., temperature, humidity, CO2...) covering windows with 10 minutes input.

1D Convolutional Layer: It is devoted to analyzing and extracting features along the time-dimensional axis of the input data.

Max Pooling layer. It aims to learn most valuable information from extracted feature vectors by applying a subsampling operation to the output matrix from the CNN layer.

Flatten layer. It reshapes the input matrix to provide a one-dimensional feature vector which can be used to make predictions by the output layer.

Output layer. This linear fully connected layer, whose output consists of a single neuron, predicts the CO2 value for the next quarter hour.
• Root Mean Squared Error (RMSE) as the performance metric.
• **Hyper parameters**: Different hyper parameter experiments were performed.
• **Windowing**:
  I picked some random days (not included in the training dataset) and predict the values. I averaged all the results to get overall performance per window.
metrics and results from the experiments
Predictive system: Results
Results from prediction
Sensor1
Sensor2
Data collection

Project Structure

Monitoring

TCP Socket
in python

Predictive

Machine Learning Predictions
Ventilation activation
project structure schema complete
How everything looks together?
How everything looks together? Sensors
Predictive system:

- 900 ppm
- 24.5 °C
- 30%
- Prediction: 1000
Predictive system: Next

Work in Progress

- System implementation using microcontrollers (ESP32)
- Use TensorFlow Lite for ML in microcontrollers
- Include particles information in the modelling stage
- Model and system optimization
How will you improve your home?