

# Exploiting Air Quality Monitors to Perform Indoor Surveillance: Academic Setting

Prasenjit Karmakar

prasenjitkarmakar52282@gmail.com  
IIT Kharagpur, India

Swadhin Pradhan

swapradh@cisco.com  
Cisco Systems, USA

Sandip Chakraborty

sandipc@cse.iitkgp.ac.in  
IIT Kharagpur, India

## ABSTRACT

Due to public awareness and government regulations, low-cost air quality monitors are becoming ubiquitous in modern indoor spaces. These monitors primarily sense air pollutants to augment the end user's understanding of her indoors. Studies have shown that having access to one's air quality context reinforces the user's need to take necessary actions to improve the air over time. Hence, user's activities significantly influence the indoor air quality. Such correlation can be exploited to get hold of sensitive indoor activities from the side channel air quality fluctuations. In this study, we explore the odds of identifying eight different indoor activities (i.e., enter, exit, fan on, fan off, AC on, AC off, gathering, eating) in a research lab with the help of an in-house low-cost air quality monitoring platform named DALTON. Our extensive data collection and analysis over the three months shows 97.7% overall accuracy in our dataset.

## 1 INTRODUCTION

Air quality Monitors are becoming ubiquitous in modern indoor spaces due to government regulations and growing awareness among the general population. In 2023, this market was estimated to be US\$ 5006 million, which is expected to expand up to US\$ 11672 million within the next decade [9]. A typical air monitoring solution [1–3] provides the end user with an understanding of their pollution exposure. Such devices send data to cloud servers for storage and to offload computational overheads of analyzing long-term data rather than doing it on-device to maintain a low-power and portable form factor. Cloud storage and computing enable the development of online dashboards and mobile applications to visualize the overall pollution patterns [4], trigger alerts and notifications [16], derive countermeasures [17], etc., reinforcing the end user towards improving air quality for healthier indoors. However, sharing such indoor pollution data with a third party can be concerning due to the high correlation between the performed indoor activities and changes in pollution signature [14]. Therefore, the data can be used as a side channel to eavesdrop on indoor activities and carry out surveillance without the user's consent.

In the last decade, several studies have explored activity monitoring, and the literature can be grouped into two categories based on the utilized modality: (i) Direct Video and

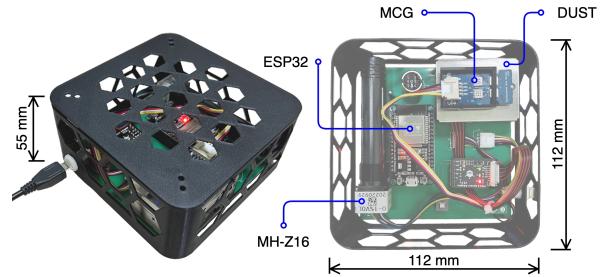
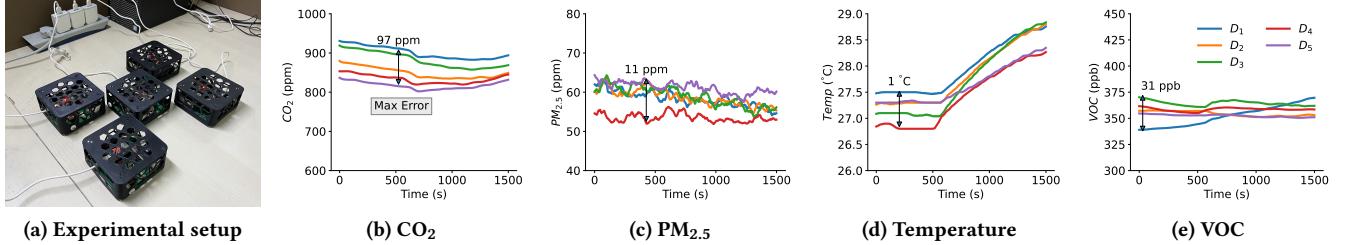


Figure 1: DALTON Sensing Module.

audio approaches, (ii) Side channel approaches like Wearables, mmWave, RFIDs, etc. Direct video [10] or audio [11] based approaches are privacy-invasive and usually require explicit permission from the end user to be operational. Moreover, users prefer to avoid capturing video or audio data in private spaces. In contrast, RF-based approaches are privacy-preserving and are largely being used in industrial or warehouse scenarios to track packages [15], monitor workers, etc. Several studies also explore mmWave in typical indoors to identify user activities in smart homes [13] and remote patient monitoring scenarios [8]. However, due to limited applications in general consumer spaces, RF-based approaches have yet to become part of our daily lives. Moreover, smart-watches are very effective in monitoring our daily activities [12] due to their wide adoption in recent years. Such wearables are incapable of continuous monitoring due to low battery life and users' discomfort for long-term usage.

In contrast, few studies [6, 7, 14] have explored correlations between indoor activities and air pollutants. For instance, [6] have identified cooking, smoking, and spraying activities. [14] has identified indoor meetings, walking in the corridor, cooking, window open, etc. To further explore the air quality modality in indoor surveillance, in this paper, we have identified eight activities (i.e., enter, exit, fan on, fan off, AC on, AC off, gathering, and eating) in an academic research lab from the data collected over three months period. The data is collected with an in-house sensing module named DALTON as shown in Figure 1. Firstly, we establish the influence of the activities on the air pollution signature of the indoor environment. After that, our extensive evaluation with simple off-the-shelf machine learning models



**Figure 2: Readings from five colocated *DALTON* devices indicate the variability across sensors made by the same vendor. The maximum error between two devices is within the error margin, as reported by the vendor.**

shows a maximum of 97.7% overall accuracy in classifying the considered activities.

## 2 EXPERIMENTAL SETUP

We have utilized our developed air quality monitoring platform named *DALTON* to collect pollution data in a research lab. We have deployed four sensing modules in four corner desks of the lab. Each *DALTON* module is lunchbox size (112 mm × 112 mm × 55 mm), equipped with multiple research-grade sensors that together measure the concentration of pollutants, such as *Particulate matter* (PM<sub>x</sub>), *Nitrogen dioxide* (NO<sub>2</sub>), *Ethanol* (C<sub>2</sub>H<sub>5</sub>OH), *Volatile organic compounds* (VOCs), *Carbon monoxide* (CO), *Carbon dioxide* (CO<sub>2</sub>), etc., with *Temperature* (T) and *Relative humidity* (RH). The device utilizes the ESP-WROOM-32 chip as the on-device processing unit that packs a dual-core Xtensa 32-bit LX6 MCU with WiFi 2.4GHz HT40 capabilities. The connectivity board is a two-layer printed circuit board (FR4 material). The outer shell of the module is a 3D printed (PLA+ material) hollow structure with honeycomb holes so that the air within the module is the same as outside, resulting in unbiased measurement of pollutants (at a sampling frequency of 1Hz).

Although the sensors are factory-calibrated, we have explicitly calibrated each sensor to ensure the correctness of the measurements. We have calibrated the PM<sub>2.5</sub>, Temperature, and Relative Humidity sensors using a reference Airthings device [2]. For the CO<sub>2</sub> readings, we have calibrated the MH-Z16 sensor to Zero point (400 ppm) and SPAN point (2000 ppm) as an initial step before the deployment. Further, we have turned on the self-calibration mode of the sensor so that it can judge the zero point intelligently and do the calibration automatically every 24 hours. The other measurements, such as NO<sub>2</sub>, C<sub>2</sub>H<sub>5</sub>OH, VOC, and CO are one-point calibrated before deployment and periodically cross-checked with a calibrated *DALTON* device during the data collection period. Figure 2 shows measurements from five colocated *DALTON* devices, validating acceptable variability across sensors made by the same vendor.

## 3 PILOT STUDY

In this section, we have conducted several pilot experiments to analyze the influence of indoor activities over the measurements of the air quality monitoring device. The observations are as follows:

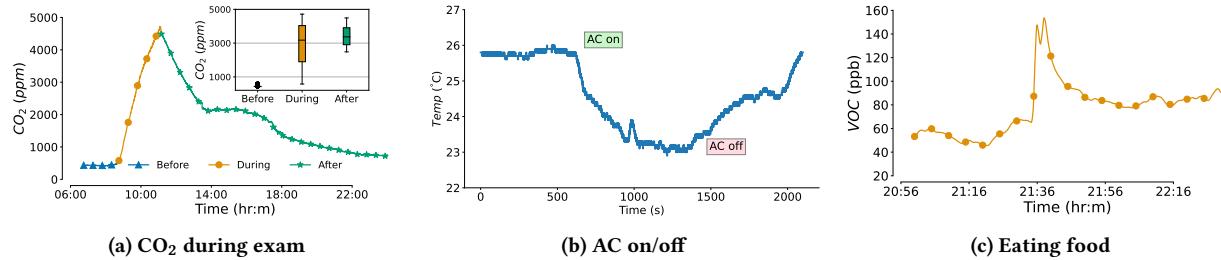
### 3.1 Indoor Gatherings

To measure the impact of large indoor gatherings in indoor spaces, we collected data from a classroom during mid-semester exams at the university. Before conducting the experiments, we surveyed several classrooms and identified an ideal one equipped with split AC, and all the windows are therefore closed. The exam was scheduled for two hours, from 9:00 am to 11:00 am in the morning. In the early morning hours before the exam, we ensured that the CO<sub>2</sub> concentration was at the expected level (close to 400 ppm) to understand the pollution footprint of 40 students gathering in the classroom.

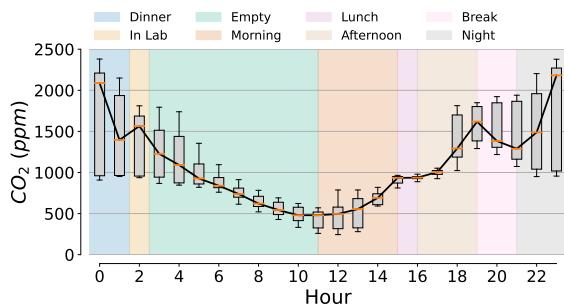
As the students started arriving at the venue at 8:45 am, the CO<sub>2</sub> accumulated as the windows were closed for effective air-conditioning. Notably, the split ACs circulate the airflow within the room, rather than pulling air from outside, to ensure effective air-conditioning with minimal energy cost [5]. However, this made the pollutants accumulate, which the students could not sense; instead, they felt comfortable with the cool breeze of the airflow. The pollutants reached peak levels (almost 5000 ppm) at the last 10 minutes of the exam. Further, we observe from the floating figure in Figure 3a that the pollutants remain trapped in that space for a long time even after all the students leave the classroom. Therefore, pollutants due to consecutive indoor gathering can add up and result in long-term accumulation of CO in indoors.

### 3.2 Air Conditioning

The air monitors are mostly equipped with temperature and humidity sensors. Air conditioning systems directly impact the temperature. Therefore, when the AC is turned on AC, the temperature goes down and vice versa. Similar observations



**Figure 3: Air monitor's measurements due to different activities - (a) Accumulation of CO<sub>2</sub> when students enter the classroom and drop upon exit during an exam, (b) Temperature change with AC on/off, (c) Eating food.**



**Figure 4: Variation of CO<sub>2</sub> concentration with indoor activities throughout the day.**

can be seen in Figure 3b, where the temperature goes down from 26 ° C to 23 ° C. The temperature starts rising as soon as the air conditioning is off.

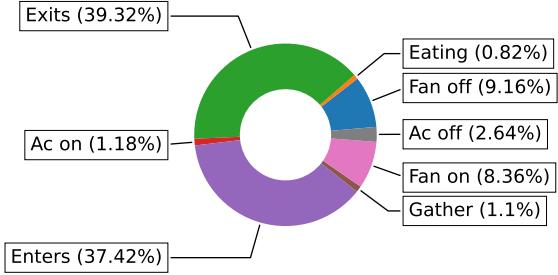
### 3.3 Eating Food

Figure 3c shows the sudden increase in volatile organic compounds (VOC) concentration due to eating fruits near a DALTON module. We can observe elevated levels of VOC for the duration of eating activity (10 minutes). The pollutant starts normalizing as soon as the activity ends and the table is cleaned. During the pilot experiments, we observed that food scraps can act as a long-term pollution source until they are removed from the indoor space.

### 3.4 Occupancy Patterns

An overall daily pattern in CO<sub>2</sub> variation for the collected data from the research lab is shown in Figure 4. We observe that indoor activities and occupancy influence the overall CO<sub>2</sub> levels. As shown in the figure, CO<sub>2</sub> concentration keeps rising when the lab is occupied, and during the dinner, lunch, and break hours, it falls slightly due to less occupancy. The lab members usually come to the lab at around 10 am in the morning. The CO<sub>2</sub> starts accumulating till 2:00 pm in the noon when the members go to lunch. However, the CO<sub>2</sub>

remains at a similar concentration due to less ventilation. The pollutant increases during the afternoon and evening hours due to maximum occupancy before the evening break. Again, during the night hours, the CO<sub>2</sub> accumulates until the dinner break. In summary, indoor pollutants such as CO<sub>2</sub> are significantly influenced by indoor activities and occupancy patterns.



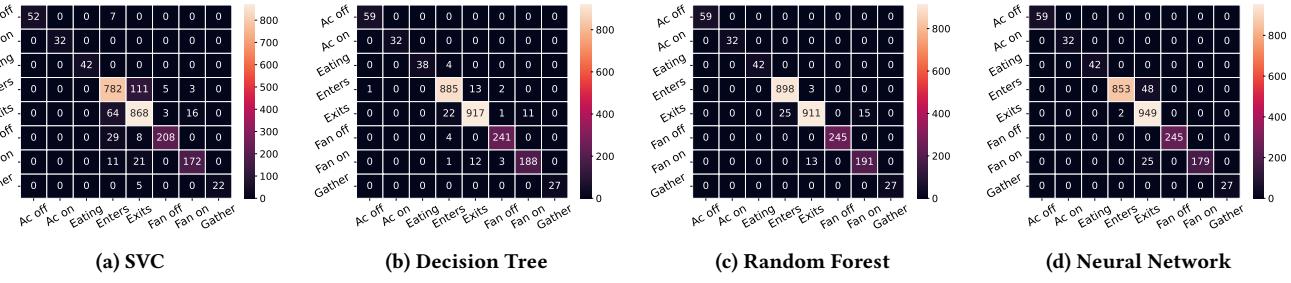
**Figure 5: Class Distribution in the collected Dataset.**

## 4 DATASET & FEATURE ENGINEERING

We have deployed the DALTON air quality monitors over three months in four corner desks of the lab. In total, seven volunteers annotated their activities in the lab throughout the entire data collection process. For example, the annotations are collected when someone exits or enters the lab, the fan is turned on or off, the AC is turned on or off, people gather for discussion, and someone eats food in the lab. Figure 5 shows the class distribution of the collected dataset. We can observe that most annotations comprise people entering or exiting the lab, followed by the fan on or off activity. However, AC on-and-off events are relatively less frequent. Lastly, indoor gathering and eating in the lab is very infrequent.

### 4.1 Features

We compute several statistical features over a sliding window of duration  $\tau$  to capture each pollutant's abrupt changes and



**Figure 6: Testing confusion matrix - (a) SVC with polynomial kernel, (b) Decision Tree with max depth 30, (c) Random Forest with max estimators 50 and depth 10, (d) Neural Network with three 64 neuron hidden layer.**

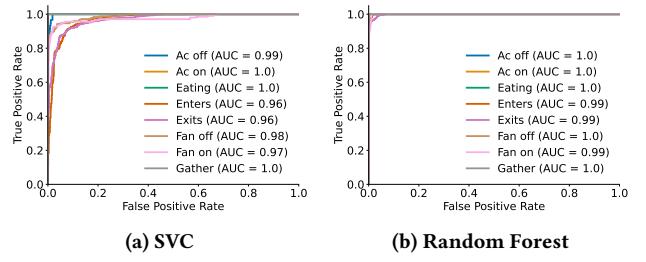
long and short-term accumulation. The features calculated for each pollutant are as follows:

- (i) **Maximum and Minimum (max, min):** These reflect the highest and lowest levels of pollutants recorded indoors. High maximum values could indicate very poor air quality. Very low minimum values suggest good ventilation.
- (ii) **Standard Deviation (std):** This measures how much the pollutant levels fluctuate. A high standard deviation indicates large swings in pollutant levels due to inconsistent ventilation or sporadic pollutant sources (like gathering).
- (iii) **Rate of Change (roc):** This measures how quickly pollutant levels rise or fall. Rapid increases might occur if there's a sudden release of pollutants (e.g., eating food), while rapid decreases might indicate effective ventilation. We consider the rate of change of pollutants for both raising ( $roc_{raise}$ ) and falling ( $roc_{fall}$ ) edges.
- (iv) **Peak Count (peak<sub>c</sub>):** This is the number of times pollutant levels exceed a certain unsafe threshold. Multiple peaks could indicate recurring sources of pollutant-generating activities or inconsistent ventilation.
- (v) **Peak Duration (peak<sub>Δ</sub>):** This measures the total time that pollutant levels were above a certain unsafe threshold. Longer spikes in pollutant concentration could mean a long-duration activity or may point to issues with ventilation or persistent pollutant sources.
- (vi) **Long Stay ( $Δ_{exc}$ ):** This represents the duration of moderate pollution levels above the safe threshold. Extended periods of moderate pollution indicate poor air quality for prolonged periods. It also suggests inadequate ventilation or persistent sources of pollutants.

$$\mathcal{F} = \bigcup_{d \in \mathcal{D}} \{f(p) : \forall f, p \in \psi \times \pi\} \quad (1)$$

Let the set of sensing modules  $\mathcal{D} = \{d_i | i = 1, 2, \dots, N\}$ , where  $N$  is the number of modules, and the set of pollutants  $\pi = \{CO_2, VOC, PM_{2.5}, PM_{10}, T, H\}$ . The set of computed functions over  $\tau^1$  minute sliding window of each pollutant

<sup>1</sup>We took  $\tau = 10$  minutes as per empirical observations.



**Figure 7: AUC-ROC Curves - (a) SVC with polynomial kernel, (b) Random Forest with max estimators 50 and depth 10.**

$\psi = \{min, max, std, roc_{raise}, roc_{fall}, peak_c, peak_\Delta, \Delta_{exc}\}$ . Therefore, the set features, including all deployed devices, are shown in Equation 1. These features are used to train simple off-the-shelf ML models in the next section.

## 5 EVALUATION

In this section, we evaluated our setup with multiple off-the-shelf machine-learning models. We have kept the models lightweight, considering the efficiency of the system. Figure 6 shows confusion matrices of four models that performs with above 80% F1-score during our testing. The SVC shows 87.9% F1-score with polynomial kernel. The best-performing model, random forest, shows 97.7% testing F1-score. The respective AUC-ROC curves are shown in Figure 7.

From the confusion matrix shown in Figure 6c, we can observe that the random forest classifier is facing confusion with class pairs such as (enter, exit), (exit, fan on). The primary reason is the lab protocol, which insists the members turn on the fans while leaving the lab. Moreover, multiple members simultaneously enter and exit the lab, leading to confusion in the machine learning model.

Finally, Table 1 summarizes the detailed evaluation of 70-30 random split and 5-fold cross-validation experiments

**Table 1: Performance of the machine learning models in 70-30 random split and 5-fold cross-validation experiments in the collected dataset.**

Model	Parameters	70-30 Random Split						5-Fold Cross-validation			
		Training (Weighted)			Testing (Weighted)			Accuracy (Mean)		Accuracy (Std)	
		F1-score	Precision	Recall	F1-score	Precision	Recall	Train	Test	Train	Test
SVM	Linear kernel	0.831	0.837	0.833	0.817	0.827	0.821	0.495	0.497	0.0156	0.0176
	Polynomial kernel	<b>0.892</b>	<b>0.894</b>	<b>0.892</b>	<b>0.879</b>	<b>0.881</b>	<b>0.879</b>	<b>0.543</b>	<b>0.542</b>	<b>0.0041</b>	<b>0.0075</b>
	RBF kernel	0.815	0.836	0.82	0.798	0.82	0.804	0.53	0.53	0.0036	0.01
Naive Bayes	Gaussian	0.403	0.727	0.399	0.399	0.717	0.39	0.424	0.419	0.0105	0.0047
Decision Tree	Max depth 10	0.949	0.95	0.949	0.924	0.924	0.924	0.967	0.951	0.0025	0.0051
	Max depth 20	0.992	0.992	0.992	0.975	0.975	0.976	0.992	0.974	0.0008	0.0039
	Max depth 30	<b>0.992</b>	<b>0.992</b>	<b>0.992</b>	<b>0.976</b>	<b>0.976</b>	<b>0.976</b>	<b>0.992</b>	<b>0.975</b>	<b>0.0007</b>	<b>0.0039</b>
	Max depth 40	0.992	0.992	0.992	0.976	0.976	0.976	0.992	0.975	0.0007	0.0039
k-Nearest Neighbour	Neighbour 10	<b>0.981</b>	<b>0.981</b>	<b>0.981</b>	<b>0.975</b>	<b>0.975</b>	<b>0.975</b>	<b>0.985</b>	<b>0.979</b>	<b>0.0008</b>	<b>0.0022</b>
	Neighbour 20	0.972	0.972	0.972	0.967	0.967	0.967	0.983	0.979	0.0002	0.0024
	Neighbour 30	0.959	0.96	0.96	0.956	0.956	0.957	0.979	0.976	0.0015	0.0044
	Neighbour 40	0.946	0.946	0.947	0.947	0.947	0.947	0.975	0.972	0.0021	0.0036
Logistic Regression	-	0.791	0.801	0.796	0.764	0.777	0.77	0.581	0.577	0.0088	0.0145
Random Forest	Max estimator 30										
	Max depth 10	0.988	0.988	0.988	0.979	0.979	0.979	0.989	0.977	0.0005	0.0021
	Max estimator 50	<b>0.988</b>	<b>0.988</b>	<b>0.988</b>	<b>0.977</b>	<b>0.977</b>	<b>0.977</b>	<b>0.989</b>	<b>0.979</b>	<b>0.0006</b>	<b>0.0049</b>
	Max depth 10	0.99	0.99	0.99	0.979	0.979	0.979	0.989	0.977	0.0012	0.0037
Neural Network	Hidden [64, 64]	0.981	0.981	0.981	0.973	0.973	0.973	0.925	0.92	0.0229	0.0165
	Hidden [64, 64, 64]	<b>0.982</b>	<b>0.983</b>	<b>0.982</b>	<b>0.978</b>	<b>0.979</b>	<b>0.978</b>	<b>0.947</b>	<b>0.943</b>	<b>0.0104</b>	<b>0.0135</b>
	Hidden [128, 128]	0.981	0.982	0.981	0.979	0.979	0.979	0.912	0.91	0.0116	0.0114
	Hidden [128, 128, 128]	0.982	0.982	0.982	0.978	0.978	0.978	0.95	0.946	0.0174	0.0203

across seven machine learning models with varying parameters. According to the table, the neural network with three 64-neuron hidden layers performs best in the random split experiment. However, the random forest shows the highest accuracy in the 5-fold cross-validation and very similar performance in the random split experiments. Moreover, the random forest is more computationally efficient and light-weight. The best parameters for each of the models are shown in bold font in the table.

## 6 CONCLUSION

In this paper, we explored potential side-channel applications of pollutant measurements from ubiquitous air monitoring solutions to identify indoor activities. Therefore, sharing pollution data with third parties may cause privacy concerns, as with such capabilities, one can carry out indoor surveillance without the end user's consent. In this work, we have collected pollution data annotated with eight indoor activities (i.e., enter, exit, fan on, fan off, AC on, AC off, gathering, eating) in a research lab over three months. Our analysis highlights that indoor pollutants are greatly influenced by the activities being performed. We can predict the underlying activity from the pollution data with 97.7% F1-score using a simple light-weight random forest model.

## REFERENCES

- [1] Prana Air. 2023. Parnaair Sensible+ Air Quality Monitor. <https://www.parnaair.com/air-quality-monitor/sensible-plus-air-monitor/>.
- [2] Airthings. 2023. Airthings. <https://www.airthings.com/>.
- [3] Amazon. 2022. AirKnight 9-in-1 Indoor Air Quality Monitor. <https://www.amazon.in/AIRKNIGHT-Portable-Formaldehyde-Detector-Monitoring/dp/B09LY1QBDM>.
- [4] Justas Brazauskas, Rohit Verma, Vadim Safronov, Matthew Danish, Jorge Merino, Xiang Xie, Ian Lewis, and Richard Mortier. 2021. Data Management for Building Information Modelling in a Real-Time Adaptive City Platform. *arXiv preprint arXiv:2103.04924* (2021).
- [5] Nan Ding, Fudan Liu, Feng Pang, Jingyu Su, Lianyu Yan, and Xi Meng. 2023. Integrating the living wall with the split air conditioner towards indoor heating environment improvement in winter. *Case Studies in Thermal Engineering* 47 (2023), 103061.
- [6] Biyi Fang, Qiumin Xu, Taiwoo Park, and Mi Zhang. 2016. AirSense: an intelligent home-based sensing system for indoor air quality analytics. In *ACM UbiComp*. 109–119.
- [7] Ennio Gambi, Giulia Temperini, Rossana Galassi, Linda Senigagliesi, and Adelmo De Santis. 2020. ADL recognition through machine learning algorithms on IoT air quality sensor dataset. *IEEE Sensors Journal* 20, 22 (2020), 13562–13570.
- [8] Zhanjun Hao, Hao Yan, Xiaochao Dang, Zhongyu Ma, Wenzhe Ke, and Peng Jin. 2023. RMVS: Remote Monitoring of Vital Signs with mm-Wave Radar. In *Proceedings of the 1st ACM Workshop on AI Empowered Mobile and Wireless Sensing* (Sydney, NSW, Australia) (*MORSE '22*). Association for Computing Machinery, New York, NY, USA, 1–6. <https://doi.org/10.1145/3556558.3558578>
- [9] Future Market Insights. 2023. Indoor Air Quality Monitor Market Outlook (2023 to 2033). <https://www.futuremarketinsights.com/reports/indoor-air-quality-monitor-market>.
- [10] Muhammad Attique Khan, Kashif Javed, Sajid Ali Khan, Tanzila Saba, Usman Habib, Junaid Ali Khan, and Aaqif Afzaal Abbasi. 2024. Human action recognition using fusion of multiview and deep features: an application to video surveillance. *Multimedia tools and applications* 83, 5 (2024), 14885–14911.
- [11] Gierad Laput, Karan Ahuja, Mayank Goel, and Chris Harrison. 2018. Ubicoustics: Plug-and-Play Acoustic Activity Recognition. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA) (*SUIT '18*). Association for Computing Machinery, New York, NY, USA, 1–10. <https://doi.org/10.1145/3210727.3210750>

- Technology* (Berlin, Germany) (*UIST '18*). Association for Computing Machinery, New York, NY, USA, 213–224. <https://doi.org/10.1145/3242587.3242609>
- [12] Kavous Salehzadeh Niksirat, Lev Velykoivanenko, Noé Zufferey, Mauro Cherubini, Kévin Huguenin, and Mathias Humbert. 2024. Wearable Activity Trackers: A Survey on Utility, Privacy, and Security. *ACM Comput. Surv.* 56, 7, Article 183 (apr 2024), 40 pages. <https://doi.org/10.1145/3645091>
  - [13] Argha Sen, Anirban Das, Swadhin Pradhan, and Sandip Chakraborty. 2023. Continuous Multi-user Activity Tracking via Room-Scale mmWave Sensing. *arXiv preprint arXiv:2309.11957* (2023).
  - [14] Rohit Verma, Justas Brazauskas, Vadim Safronov, Matthew Danish, Ian Lewis, and Richard Mortier. 2021. RACER: Real-Time Automated Complex Event Recognition in Smart Environments. In *29th ACM SIGSPATIAL GIS*. 634–637.
  - [15] Chenshu Wu, Feng Zhang, Beibei Wang, and K. J. Ray Liu. 2020. mSense: Towards Mobile Material Sensing with a Single Millimeter-Wave Radio. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 106 (sep 2020), 20 pages. <https://doi.org/10.1145/3411822>
  - [16] Sailin Zhong, Hamed S Alavi, and Denis Lalanne. 2020. Hilo-wear: Exploring wearable interaction with indoor air quality forecast. In *Extended Abstracts of the 2020 ACM CHI*. 1–8.
  - [17] Hao-Cheng Zhu, Chen Ren, and Shi-Jie Cao. 2022. Dynamic sensing and control system using artificial intelligent techniques for non-uniform indoor environment. *Building and Environment* 226 (2022), 109702.