CET CHEMICAL ENGINEERING TRANSACTIONS

VOL. 127, 2026

A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it

Guest Editors: Laura Capelli, Marzio Invernizzi, Selena Sironi

ISBN 979-12-81206-28-1; **ISSN** 2283-9216

DOI: 10.3303/CET26127010

Real-Time Environmental Identification for Personal Odour and VOC Monitoring using Portable Multi-Sensor Data

Feiling Pan*, James A. Covington

School of Engineering, University of Warwick, Coventry, UK
Feiling.Pan@warwick.ac.uk

Volatile organic compounds (VOCs) and odours vary considerably across urban micro-environments because of differences in emission sources, occupancy, and ventilation. Fixed-site monitoring does not always reflect the changing exposures individuals encounter during everyday travel and activity. In addition, the nature and concentration of VOCs can differ considerably between micro-environments. Therefore, real-time identification of the surrounding environment can help interpret the potential risk associated with elevated VOC levels. This study evaluated an automated method for environmental recognition using the PONG4 portable multi-sensor device. A labelled dataset was generated from real-time monitoring across 15 urban micro-environments, including transport, commercial, campus, and community settings. Five machine-learning algorithms were assessed using 5-fold cross-validation. Random Forest and XGBoost achieved the highest classification accuracy, both reaching 81.1%, and outperformed the single decision tree model, which achieved 66.6%. Feature-importance analysis showed that PM4, CO_2, and relative humidity contributed most strongly to environmental classification, with importance weights of 0.271, 0.130, and 0.094, respectively. In addition, t-SNE visualisation showed partial clustering across several micro-environment categories. These findings suggest that portable multi-sensor data, combined with machine learning, can support automated environmental identification and may be useful for context-aware VOC and odour exposure assessment.

1. Introduction

Volatile organic compounds (VOCs) and odours in urban micro-environments can vary markedly over time and location because of differences in traffic emissions, commercial activity, building materials, and indoor-outdoor ventilation (Dutta et al., 2016). Conventional air-quality monitoring still relies mainly on fixed stations. Although these are useful for measuring regional background conditions, they are less effective for capturing the changing environments that individuals encounter as they move between locations and transport modes (Bagkis et al., 2025). This limits accurate assessment of personal exposure in micro-environments (Li et al., 2024), as pollutant composition and intensity differ between settings, such as transport systems and commercial spaces (Maung et al., 2022). Therefore, reliable identification of the surrounding environment is important for interpreting odour and VOC measurements (Sekar et al., 2023). This would provide a useful context to potential sources of VOCs based on the specific environment and therefore would make the use of low-cost VOC sensors, for personal monitoring, more effective. However, this relies on being able to rapidly identify the specific environment to interpret the VOC readings. To achieve this identification, we propose to measure a wide range of environmental factors, within a micro-environment, and then apply machine-learning to the data for environment classification. Machine-learning methods have previously been shown to perform well with complex environmental datasets (Nasir et al., 2022), which supports their use here. Improved identification of the surrounding micro-environment may help link pollutant patterns to specific settings and support more accurate assessment of personal exposure. In this study, we used an in-house personal monitor, PONG, to collect multi-sensor environmental data during daily activities. The units used here were the fourth-generation version of the platform, PONG4, which extends the previously published PONG3 design (Pan & Covington, 2024) by monitoring a wider range of environmental parameters. PONG4 was used to capture environmental responses across different urban settings. These data were combined with on-site scene labels to create a structured dataset covering a range

Paper Received: 28 March 2026; Revised: 20 May 2026; Accepted: 5 June 2026
Please cite this article as: Pan F., Covington J.A., 2026, Real-Time Environmental Identification for Personal Odour and VOC Monitoring using Portable Multi-Sensor Data, Chemical Engineering Transactions, 127, 55-60 DOI:10.3303/CET26127010

of locations, including shopping areas, transport environments, and dining spaces. Machine-learning algorithms were then applied to extract features from the sensor signals and classify environmental scenarios. This framework was used to explore the potential of portable sensing and data-driven analysis for automated environmental recognition and individual-level exposure assessment.

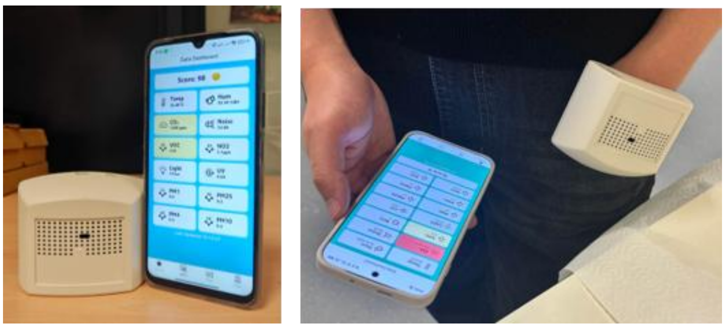
2. Methodology

PONG4 was developed as the fourth-generation version of the Personalized envirONmental quality monitorinG system. It follows the previously reported PONG3 platform (Pan & Covington, 2024), which measured ten environmental and air-quality parameters but did not include particulate matter sensing. The addition of PM measurement was a change in PONG4, alongside an increase in sensor outputs. The device records physical conditions, particulate matter, VOC-related responses, and inorganic gases, giving a broader measurement set for comparing urban micro-environments and interpreting odour and VOC readings. PONG4 uses a microcontroller with an Xtensa LX7 dual-core processor running at 240 MHz. Sensor readings are collected every 5 s, equivalent to 0.2 Hz, and are sent to the PONG-APP by Bluetooth Low Energy. Power is supplied by two 10440 rechargeable lithium-ion batteries, each rated at 3.7 V and 320 mAh, giving about 6 h of operation. The electronics are housed in a ventilated ERGO-CASE enclosure measuring 96 mm × 80 mm × 40 mm. Table 1 lists the sensors, specifications, outputs, and reference values used for context.

Table 1: PONG4 sensors

Parameter	Part number	Manufacturer	Accuracy	Output	Risk Range
Light	LTR390-UV	Lite-On Inc.	±10%	lux	>50000 lux
UV	LTR390-UV	Lite-On Inc.	±20%	Index	>8 index
Noise	CMA-4544PF-W	CUI Devices	±2dB	dB	>120 dB
Temperature	ENS210	ScioSense	±0.2°C	°C	>35°C
Humidity	ENS210	ScioSense	±2.5%	%r.h.	>70%r.h.
VOC	ENS161	ScioSense	±15%	ppb	>660ppb
CO2	SEN66	ScioSense	±100	ppm	>5000 ppm
PM	SEN66	Sensirion AG.	±10	μg/m³	>35 μg/m³
NOx	SEN66	Sensirion AG.	±15 index	Index	>300 Index
VOC	SEN66	Sensirion AG.	±15 index	Index	>400 Index
NO2	NO2-D4	Alphasense	±15%	ppb	>0.053 ppm
SO2	SO2-D4	Alphasense	±15%	ppb	>0.2 ppm
CO	CO-D4	Alphasense	±15%	ppb	>9pm

The two VOC entries in Table 1 refer to different sensor responses. The ENS161 gives an estimated VOC concentration in ppb, whereas the SEN66 gives a VOC index based on relative changes in mixed VOC exposure. These outputs were treated as complementary features, not duplicate measurements. During use, PONG4 communicates with the PONG-APP through Bluetooth Low Energy. The app records sensor values and GPS positions every 5 s, saves the data locally as CSV files, and can synchronise data to cloud storage. It also provides real-time feedback by comparing selected readings with fixed comfort, warning, and risk thresholds, (risk range defined in table 1). These thresholds are used for user alerts and colour-coded route maps, while the exported raw data are used for machine-learning analysis. Future versions could allow more user-defined application-specific thresholds. Figure 1 shows the PONG4 unit and mobile-app interface.



(a) *PONG4 using on table* (b) *PONG4 using in life*

Figure 1: PONG4 use example with unit and phone app

PONG4 was used to collect labelled measurements across 15 urban micro-environments in Coventry, UK, and Hefei, China, between January and February 2026. The final dataset contained 1,367 labelled samples. Each sample consisted of a synchronised set of sensor readings and a corresponding micro-environment label recorded during sampling. The monitored environments included transport settings, commercial indoor spaces, campus areas, and community environments. Raw CSV files exported from the PONG-APP were screened before model training. Initial sensor start-up periods were removed, missing values were handled by [insert method], and continuous sensor variables were standardised using Z-score normalisation. To avoid data leakage, normalisation parameters were calculated from the training data only and then applied to the validation or test data. Scene labels were matched to sensor timestamps to assign each record to a micro-environment class. Five machine-learning algorithms were compared for micro-environment recognition: decision tree, random forest, support vector machine, k-nearest neighbour, and XGBoost. Model performance was evaluated using stratified 5-fold cross-validation. Accuracy, precision, recall, and macro-F1 score were used to evaluate performance, as macro-F1 is more informative when class sizes are uneven.

3. Result

3.1 Sensor feature distributions across micro-environments

PONG4 measurements were collected across 15 urban micro-environments in Coventry, UK, and Hefei, Anhui Province, China, between January and February 2026. The dataset contained 1,367 labelled samples. Figure 2 compares the distributions of four representative sensor features: PM4, VOC index, CO_2, and relative humidity, which are the top influential variables in this study. These variables were selected because they showed clear variation across several micro-environments and are relevant to likely differences in particle sources, occupancy, ventilation, and indoor material emissions.

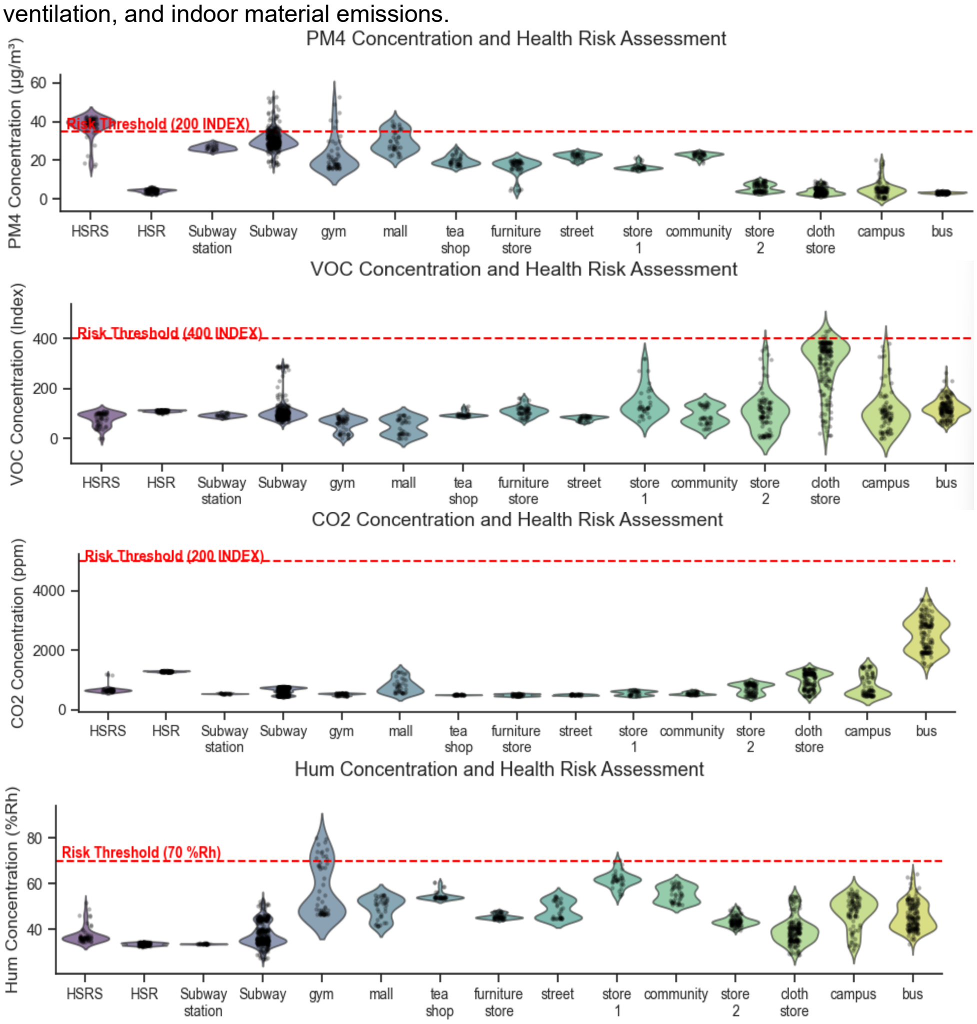


Figure 2: PM4, VOC index, CO_2 and relative humidity across micro-environments. Dashed lines indicate app-based reference or alert thresholds used for contextual interpretation.

PM4 concentrations were highest in high-speed railway stations, gyms, malls, and subway carriages, while campus and bus environments generally showed lower PM4 levels. These differences may reflect particle resuspension, passenger movement, mechanical wear, and differences in ventilation. The VOC index was elevated in several indoor commercial environments, particularly clothing stores and retail spaces, which may reflect emissions from textiles, furnishing materials, cleaning products, or fragranced products. CO_2 concentrations were generally below the app-based alert threshold, although bus environments showed relatively higher values, probably because of higher occupant density and more limited air exchange. Relative humidity also varied between environments, especially between indoor commercial spaces, transport settings, and outdoor or semi-outdoor locations. Figure 3 shows examples of the environmental quality mapping function in five settings: Coventry city, a gym, Hefei urban area, Hefei subway, and the China high-speed railway. The mapped environmental score varied spatially within and between these environments, indicating that PONG4 can capture changes in environmental conditions during movement. These maps are intended to visualise spatial variation in the measured sensor outputs and should be interpreted as comparative environmental indicators rather than regulatory exposure assessments.

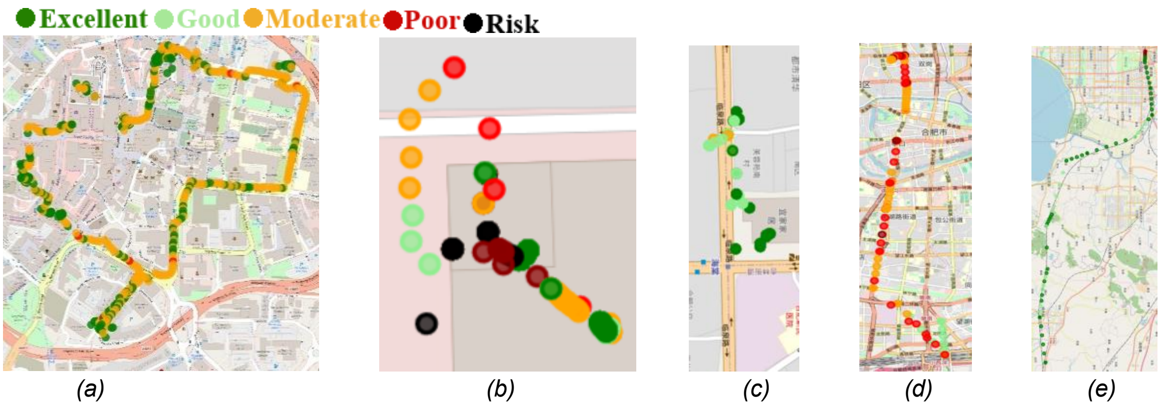


Figure 3: Environmental quality distribution across five example settings: (a) Coventry city, (b) gym, (c) Hefei urban area, (d) Hefei subway, and (e) China high-speed railway.

3.2 Micro-environment classification performance

To evaluate whether the multi-sensor data could distinguish between micro-environments, five machine-learning algorithms were compared: decision tree, random forest, support vector machine, k-nearest neighbour, and XGBoost. Random forest and XGBoost achieved the highest classification accuracy, both reaching 81.1%. This suggests that ensemble methods were better able to capture non-linear relationships in the multi-sensor dataset than the other models tested. SVM and KNN achieved accuracies of 74.8% and 74.0%, respectively, while the single decision tree had the lowest accuracy at 66.6%. The weaker performance of the decision tree suggests that a single-tree structure was less effective at representing the variability of the environmental sensor signals.

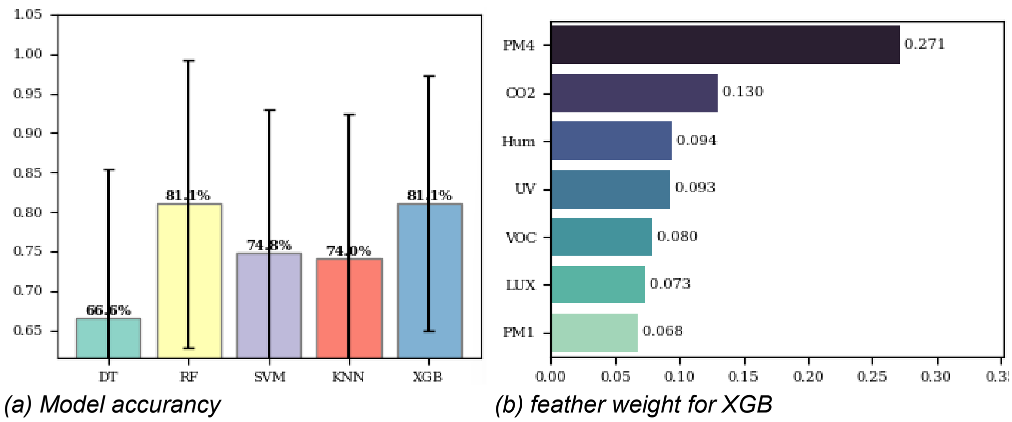


Figure 4: (a)Classification accuracy and XGBoost feature importance for the five machine-learning models. DT: decision tree; RF: random forest; SVM: support vector machine; KNN: k-nearest neighbour; XGB: extreme gradient boosting. (b)Feather weight for XGB

Figure 4 shows that Random Forest and XGBoost achieved the highest classification accuracy, both reaching 81.1%, while the decision tree model showed the lowest accuracy at 66.6%. This indicates that ensemble models were better able to capture non-linear relationships within the multi-sensor dataset. The XGBoost feature-importance analysis showed that PM4 was the most influential variable, with an importance weight of 0.271. CO_2 and relative humidity were also important contributors, with weights of 0.130 and 0.094, respectively. These results suggest that differences in particulate levels, occupancy-related ventilation, and moisture conditions were important for distinguishing between micro-environments. VOC signals, light intensity, and UV also contributed to classification, indicating that the model used a combination of air-quality and physical environmental features rather than relying on a single pollutant. To further examine whether the sensor profiles formed distinguishable groups, t-SNE was used to project the high-dimensional sensor dataset into two dimensions. Figure 5 shows the resulting projection for the 15 labelled micro-environments. The t-SNE coordinates do not represent physical units, rather, they show the relative similarity between samples based on the measured sensor features.

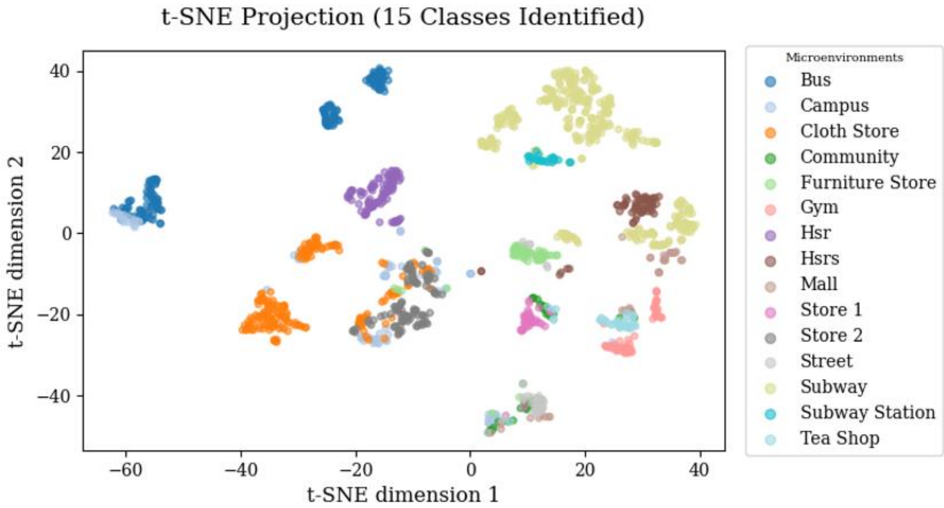


Figure 5: t-SNE projection of PONG4 sensor features across 15 urban micro-environments

The t-SNE projection shows partial clustering of several micro-environment categories, indicating that the multi-sensor measurements captured differences between some environments. High-speed railway and campus samples formed relatively distinct groups, while several commercial and public environments showed greater overlap. This overlap is expected because some indoor environments may share similar ventilation conditions, occupancy patterns, or emission sources. The partial separation observed in Figure 5 is consistent with the classification results in Figure 4, where the best-performing models achieved good but not perfect accuracy. To examine class-specific performance, a normalised confusion matrix was generated for the best-performing classification model, shown in Figure 6.

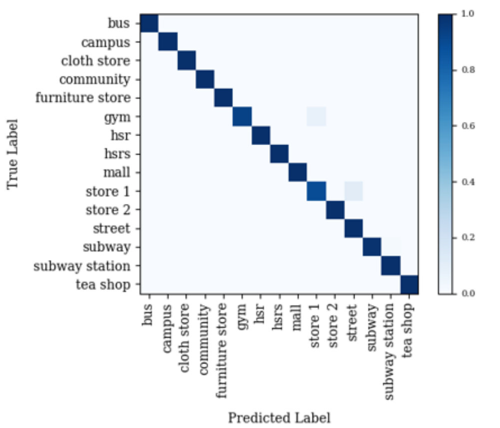


Figure 6: Confusion Matrix for Micro-environment Recognition

This suggests that the multi-sensor feature set generated by the PONG4 platform provided useful discriminatory information for scene recognition. Some misclassification was still present, indicating partial overlap between environments with similar characteristics. Overall, these results support the use of portable multi-sensor data for automated recognition of urban micro-environments and suggest potential value for context-aware exposure monitoring.

4. Conclusions

Sources of VOCs and odours strongly depend on the type of micro-environment. Therefore, rapid identification of the surrounding micro-environment can help interpret VOC exposure risk when using low-cost VOC sensors. This study evaluated an automated approach for urban environmental identification by combining the PONG4 portable multi-sensor platform with machine-learning models. Data collected across 15 micro-environment types showed that Random Forest and XGBoost achieved the highest classification accuracy, both reaching 81.1%, indicating that ensemble methods were better able to distinguish between complex environmental profiles than the other models tested. Feature-importance analysis showed that PM4, CO_2, and relative humidity were among the most informative variables for environmental classification, while t-SNE visualisation indicated partial clustering across several micro-environment categories. However, the limited geographical range and short sampling period may restrict model generalisability across cities, seasons, building types, and ventilation conditions. Overall, the results suggest that portable multi-sensor devices can support automated recognition of urban environments and may be useful for context-aware VOC and odour exposure assessment. Future work will expand the geographical and temporal range of the dataset and investigate how VOC sensor responses can be related to specific VOC and odour sources within these micro-environments.

References

Bagkis E., Hassani A., Schneider P., DeSouza P., Shetty S., Kassandros T., Khan J., 2025, Evolving trends in application of low-cost air quality sensor networks: challenges and future directions, npj Climate and Atmospheric Science, 8(1), 335.

Dutta T., Kim K.H., Uchimiya M., Kumar P., Das S., Bhattacharya S.S., Szulejko J., 2016, The micro-environmental impact of volatile organic compound emissions from large-scale assemblies of people in a confined space, Environmental research, 151, 304–312.

Kollapudi P., Alghamdi S., Veeraiah N., Alotaibi Y., Thotakura S., Alsufyani A., 2022, A New Method for Scene Classification from the Remote Sensing Images, Computers, Materials & Continua, 72(1), 311–325.

Li L., Xu Z., Zhou S., 2024, Research progress in micro-scale environmental monitoring, exposure and individual response in Chinese literature, Transactions in Urban Data, Science, and Technology, 3(3-4), 121–135.

Li X., Yi S., Cundy A. B., Chen W., 2022, Sustainable decision-making for contaminated site risk management: A decision tree model using machine learning algorithms, Journal of Cleaner production, 371, 133612.

Liu C., Wu H., Li J., Liu X., Yan S., Wang A., 2025, Hyperspectral image cross-scene classification combined domain adaptation with XGBoost, International Conference on Measurement, Communication, and Virtual Reality, 13634, 124–132.

Maung T. Z., Bishop J. E., Holt E., Turner A. M., Pfrang C., 2022, Indoor air pollution and the health of vulnerable groups: a systematic review focused on particulate matter (PM), volatile organic compounds (VOCs), International journal of environmental research and public health, 19(14), 8752.

Meng L., van den Elzen S., Pezzotti N., Vilanova A., 2023, Class-constrained t-SNE: combining data features and class probabilities, IEEE Transactions on Visualization and Computer Graphics, 30(1), 164–174.

Nasir N., Kansal A., Alshaltone O., Barneih F., Sameer M., Shanableh A., Al-Shamma'a A., 2022, Water quality classification using machine learning algorithms, Journal of Water Process Engineering, 48, 102920.

Pan F., Covington J. A., 2024, A portable personalized environmental quality monitoring system (PONG) version 3, IEEE Sensors Journal, 24(14), 22926–22933.

Sathyanarayanan S., Tantri B. R., 2024, Confusion matrix-based performance evaluation metrics, African Journal of Biomedical Research, 27(4S), 4023–4031.

Sekar A., Varghese G. K., Varma R., 2023, Chemical Fingerprinting of Volatile Organic Compounds—A Forensic Tool to Apportion Pollution Sources in Industrial Micro-Environments, Asian Conference on Indoor Environmental Quality, 27–39.

Thapa A., Horanont T., Neupane B., Aryal J., 2023, Deep learning for remote sensing image scene classification: A review and meta-analysis, Remote Sensing, 15(19), 4804.

Wang W., Xing L., Ren P., Jiang Y., Wang G., Liu B., 2023, Subspace prototype learning for few-shot remote sensing scene classification, Signal Processing, 208, 108976.