Journal of Physics:
Conference Series



PAPER • OPEN ACCESS

# Advancing indoor environmental monitoring: A machine learning approach for post-data collection activity validation

To cite this article: H Routledge *et al* 2025 *J. Phys.: Conf. Ser.* **3140** 112018

# Advancing indoor environmental monitoring: A machine learning approach for post-data collection activity validation

**H Routledge, R Tang and S Stamp**

Institute for Environmental Design and Engineering, University College London, UK

E-mail: `hannah.routledge.22@ucl.ac.uk`

**Abstract.** Occupant behaviour is a significant driver of air quality in indoor spaces, but due to the innate variability of these activities, they can be complex sources. This makes it difficult to assess their direct health impact. Validation methods currently used to research occupant activities are also often disruptive and vulnerable to inaccuracies, making research in this area difficult. This pilot study aims to establish whether it is possible to develop a new method to conduct activity research by identifying activities post-data collection. Household environmental data was collected in 3 houses over 3 months, alongside participants completing an activity diary. This data was used to identify an activity's 'signature' and develop generalised and individual activity identification models using machine learning (ML). The generalised activity model achieved 55-64% overall accuracy when tested on training data, and approximately 30% when applied to non-training data. These results indicated that a single model was not the best approach to accurately identify events. Using individual specialised models increased accuracy significantly, achieving 72-92% when applied to the training dataset, and 57-79% when applied to the non-training dataset. This suggests that the method is viable, and it is possible to accurately identify various activities using household environmental data.

## 1. Introduction

Air pollution is currently the 2nd highest risk factor for death globally [1] and accounts for millions of premature deaths each year [2]. A large proportion of these reported deaths can be attributed to household air pollution, and whilst most occur in low or middle income countries [2], poor indoor air quality (IAQ) is still of significant concern in high-income countries.

In high-income countries, people will spend up to 90% of their time in indoor environments [3, 4]. Due to this, the behaviour of the occupants within the spaces can be a significant driver for IAQ changes. Due to the innate variability in many of these activities, they can be complex sources of pollutants, making it difficult to understand their direct impact on our health.

There is increasing importance in understanding these pollutant sources as buildings become more air tight to improve energy efficiency and reduce energy wastage. Sometimes there can be an unintended effect of worsening the IAQ in these spaces [5, 6, 7], often due to the ventilation systems either not being used or not working as they should. Internal sources may then have a greater impact on the IAQ [8] as it will take longer for pollutant concentrations to decay. This results in occupants being exposed to a potentially higher dose of pollutants for an extended period of time, causing a higher health risk [9]. These factors further emphasise the importance of understanding dynamics of indoor pollutants, their sources, and how they may affect us, as this will allow for suitable changes to be made to preserve both immediate and long-term health.

To research occupant behaviour, observations of real world environments are often conducted. These tend to use validation methods (such as activity diaries) that can be disruptive for participants, as well as being susceptible to human error [10], resulting in potentially unreliable data. This makes research to identify indoor pollutant sources more difficult, which consequentially causes a delay in understanding how we can make changes to improve health.

The aim of this pilot study was to begin development of a method to identify occupant activities post-data collection, and assess whether it would be viable to use on a larger scale. This was completed by collecting indoor environmental data and using this, train a machine learning model to recognise the signatures of these activities. Newer low cost sensors were also used as they provide higher time resolution multi-sensor air quality data when compared to more traditional methods. This method should be more reliable and less disruptive to participants, making it easier to conduct future research into the impact of these behaviours on health, and how we can make changes to our actions and environment to make these activities safer.

## 2. Method

### *2.1. Data Collection*

Indoor environmental data for this study was collected via a natural experiment in 3 houses over 3 months, where sensors were placed in the kitchen of each home. The sensors used were: Eltek AQ211 loggers [11] for all air quality measurements, HOBO UX120-006M [12] with a split-core AC current sensor [13] to measure current at the electricity meter, and HOBO U12-012 [14] for measuring light intensity.

Alongside the environmental data collection, participants were asked to keep a paper activity diary of their oven, hob, microwave, and kettle use. Whilst activity diaries are susceptible to human error, it was decided that this would still be suitable for activity validation in this proof-of-concept pilot. Whilst some events maybe missed or mislabelled, overall the viability of the method can still be established before further refinement. Diagrams of the kitchen arrangement and sensor placement for each of the observed homes can be seen in Figure 1.

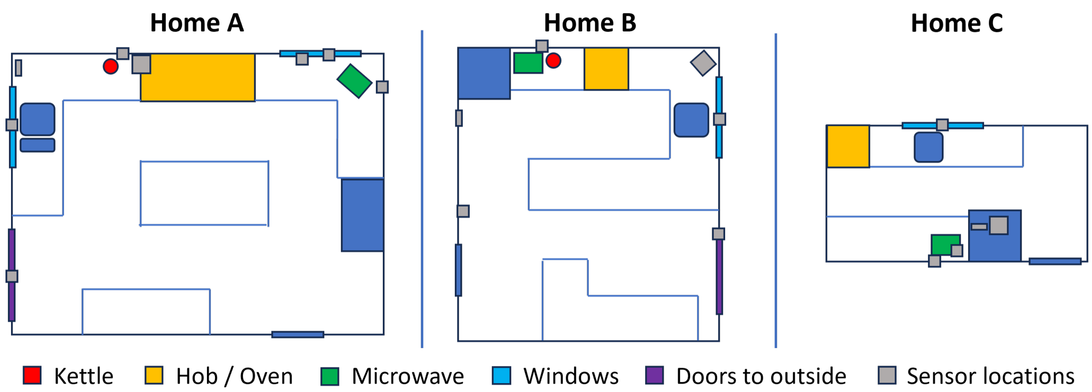


**Figure 1.** Diagrams of the experimental set-up in each of the monitored homes.

The selection of activities observed were chosen by estimating how much data would be needed for the data analysis, and then dividing it over the amount of participating houses. This indicated the number of times an activity needed to be performed over the pilot period to provide enough data, which then further gave an idea of what kinds of activities could be selected.

According to Smolic [15], the number of examples of an activity needed to be able to use machine learning is equal to 10 times the number of columns in the dataset. As the data collected had, on average, 13 columns, this meant that at least 130 examples of each activity was needed. Dividing this over 3 houses meant that activities needed to be performed approximately 43 times per house. As the pilot took place over 3 months, this meant that the activities chosen needed to be ones that would likely be performed at least once every other day. From this, use of the hob, oven, kettle and microwave were chosen. Whilst other activities may be considered more significant for health impacts, choosing ones that are more frequent gave the ability to test the method, and then it could later be developed further to include other activities.

Further benefits from having chosen these specific activities would be to test the limitations of the data analysis method being applied. As the microwave and kettle would have similar effects on their environment, they were likely to have a similar signature and thus would test the method in seeing if it is able to identify the differences between similar activities.

Indoor environmental data was collected using a range of sensors, including an air quality sensor measuring temperature, relative humidity , CO2, CO, NO2, PM (1, 2.5 & 10) and tVOCs, as well as electricity data using a current clamp on the electricity meter, and light intensity (lux). Examples of the signatures seen in the data collected can be seen in Figure 2.

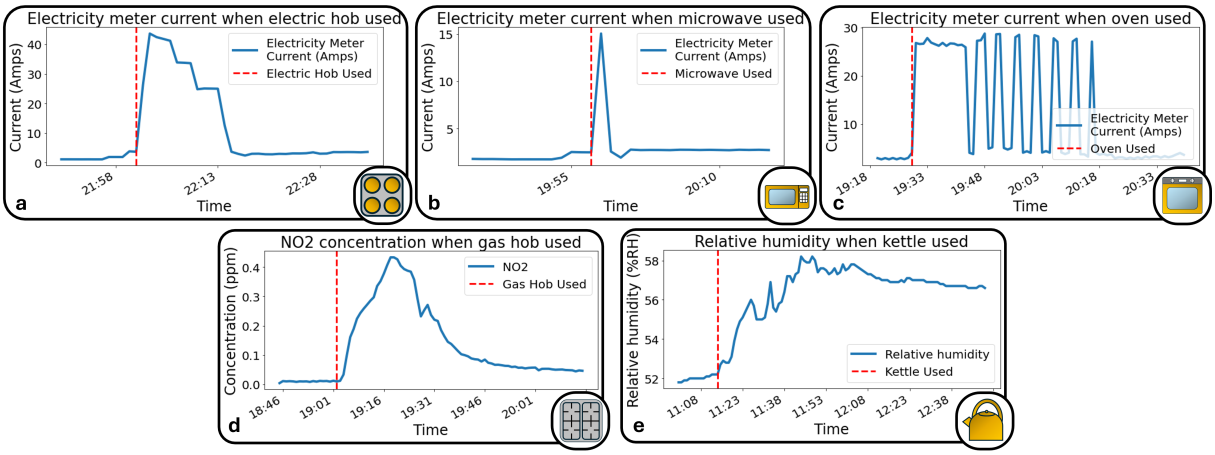


**Figure 2.** Graphs showing partial signatures of each activity. **a)** Current when electric hob used. **b)** Current when microwave used. **c)** Current when oven used. **d)** NO2 when gas hob used. **e)** Relative humidity when kettle used.

### *2.2. Data Analysis*

Logistic regression (LR) and long short-term memory (LSTM) were both individually used to build the identification models, doing this gave an indication of how different methods may perform. LR was chosen as it is simple to apply and showed how ML may perform with the format and skew of the dataset. LSTM was chosen as it performs well with time series data, which is important in this case as each data point is not independent from previous ones.

Due to the nature of the data collected, there was a large skew within the dataset towards non-activities, as a large portion of the time nothing will be happening (e.g. at night). This meant that the data needed to be sampled to reduce the amount of non-activities before using for training. This was done by counting the number of events occurring in the dataset, and randomly sampling approximately the same number of non-events, keeping the dataset balanced.

Most of the measured variables were fed to the model in their raw form, these include: temperature (plus 30 minutes of lagged readings), relative humidity (plus 30 minutes of lagged readings), CO, NO2, PM (1, 2.5 & 10), tVOCs, and electricity meter current. Light intensity and CO2 were excluded as they were found to have no and negative impact respectively.

Due to the activity diary used not providing durations of each event, this had to be assumed within the dataset. It was decided that microwave and kettle durations would be set at 5 minutes, and gas hob, electric hob and oven durations would be set at 30 minutes. In the dataset, each activity was defined to be 'true' for the full duration due to the lagged nature of IAQ signatures.

The classification labels given to the models to sort data into were: electric hob, gas hob, oven, microwave, kettle, multiclass (multiple activities occurring simultaneously), and no activity.

Each model was trained on 80% of the data from House A and House B combined, and then tested using the remaining 20% and the entire dataset from House C. Doing this gave the model the ability to train on the largest number of variables and activities, while also giving an indication of the generalisability of the model. Model performance is assessed using accuracy, which is calculated by the number of correct predictions over total predictions.

Initially, both types of ML method (LR & LSTM) were used to create single models that could identify every activity (generalised), then further developed into individual (specialised) models for each activity. This was done to see how identification accuracy would be affected.

## 3. Results

### *3.1. Generalised Activity Model*

The identification accuracies for each activity over both machine learning methods and datasets when using the single model can be seen in Table 1.

**Table 1.** Generalised Activity Model Accuracy.

| | Logistic Regression | | Long Short-Term Memory (LSTM) | |
|---|---|---|---|---|
| | Training Data (%) | House C (%) | Training Data (%) | House C (%) |
| Electric Hob | 66 | 0 | 74 | 0 |
| Gas Hob | 62 | - | 76 | - |
| Oven | 32 | 0 | 49 | 6 |
| Microwave | 12 | 0 | 20 | 0 |
| Kettle | 45 | - | 54 | - |
| Multiclass | 22 | 4 | 31 | 11 |
| No Activity | 67 | 60 | 71 | 61 |
| Overall | 55 | 29 | 64 | 27 |

It can be seen that, while the LSTM model does perform slightly better overall on the training data than LR, there are clear weaknesses within the identification for both models. Neither of the methods perform well identifying the individual activities, with around half of the accuracies reporting to be less than 50%. This is further highlighted when looking at accuracy of microwave identification where neither model was able to achieve more than 20% accuracy.

When applying the trained model to the data collected from House C, there are no classification accuracy results for either kettle or gas hob as this home did not have those appliances. Neither model performed well on this data, having low to zero accuracy for all classifications other than no activity. This indicates that the models at this stage are not generalisable and are unable to be accurately applied to other homes.

### *3.2. Specialised Activity Models*

When moving to specialised models for each activity as opposed to a generalised model, the 'multiclass' classification was dropped. The 'multiclass' classification was considered too ambiguous as it had no differentiation between which activities were occurring at the same time, and it was likely that there would be a large amount of variation. This would mean that defining clear decision boundaries would be difficult, causing inaccuracies within the model.

The 'no activity' classification is also not reported in the table as it differed for each model and wouldn't be one conclusive result as in the generalised model.

The classification accuracy using the specialised individual models can be seen in Table 2.

**Table 2.** Specialised Activity Model Accuracy.

| | Logistic Regression | | Long Short-Term Memory (LSTM) | |
|---|---|---|---|---|
| | Training Data (%) | House C (%) | Training Data (%) | House C (%) |
| Electric Hob | 87 | 79 | 90 | 78 |
| Gas Hob | 92 | - | 91 | - |
| Oven | 86 | 60 | 77 | 57 |
| Microwave | 82 | 75 | 77 | 64 |
| Kettle | 72 | - | 81 | - |

When comparing tables 1 and 2, a significant improvement in the identification accuracy for every activity can be seen for both datasets and model architectures.

Looking at the accuracy when applied to the training data, identification accuracies for each activity increase by 21-70% when using LR, and 15-57% for LSTM. The largest increase for both of these models was seen in microwave identification.

The largest change seen when comparing the the generalised and specialised models is in the results when applied to the House C dataset, going from being unable to identify the events, to identifying them with reasonable accuracy. It can be seen that there is a 60-79% and 51-78% increase for LR and LSTM respectively, with the largest increase for both being the electric hob.

## 4. Discussion

The results of this pilot study show that when using specialised individual activity models, it is possible to accurately identify selected activities from a dataset post-data collection. This suggests that this method, with further development, could potentially be used going forward to further research into the direct impact of various behaviours on our health.

A large increase in accuracy can be seen when the model development method changes from using a single model for every activity together, to single models each focussed on their own activity. There are several reasons for why this may be observed.

Firstly, having an individual model for each activity means the model will not be influenced by the skew between how many times each activity occurs. Whilst the data was sampled to reduce the skew of events to non-events, this does not affect the difference between the number of times each activity occurred. This means that the model may be more biased towards identifying activities that are more prevalent within the dataset, reducing the identification accuracy for other events. An example of this can be seen for microwave in Table 1. This activity had fewer examples within the training dataset, in addition to being a short duration activity, potentially making the general model less likely to identify it.

Having more models also means that each model can focus on specific features of an activity, rather than balancing between different classifications. This means it develops clearer decision boundaries as only one class is indicated from the rest, rather than deciding between 7.

Individual models should additionally allow for better handling of ambiguous cases. As the generalised model requires the event to be a single classification, and since the ‘multiclass’ label does not have a clear definition, this may cause confusion when multiple labels are technically true at the same time. As the individual models do not rely on each other for their prediction, they may cope better when multiple events occur over the same time period.

There are several limitations both in the design of this pilot and nature of the activities which may hinder the performance of the model, these may account for some of the performance deterioration between the training data and House C.

Machine learning models generally perform better when given more training data, this is because it allows for better understanding of variation between examples of events. The sample size used for this study is small when compared to what would normally be expected, and while it may be enough data to test the viability of the method, it may mean that generalisability is impacted. Having fewer homes over a short period of time could mean the model is less exposed to any behavioural changes that may occur due to occupant diversity (e.g. cultural cooking styles, different house occupancy, etc.) or seasonal changes (window opening, cooking styles, etc.). Increasing sample size in further research could improve both accuracy and generalisability, as it will allow for the model to be exposed to more variability of what an activity may look like in different situations. This has been implemented in further work building on from this pilot.

Paper activity diaries were used as the activity validation method during the data collection. This method is often considered more unreliable due to relying on participants self-reporting their actions, which has many opportunities to introduce errors (e.g. forgetting to report, reporting incorrect timestamps, etc.). Further work being undertaken to build on this pilot relies less on self-reporting where possible, such as including sensors that measure plug load for appliance usage, this should increase the reliability of validation by reducing the impact of human error.

A large number of different environmental variables were measured and used to train the models, this may mean that the models become less applicable to other studies and datasets due to the large amount of additional information needed. This may cause the models to become redundant as they are not widely accessible, even if they work accurately. To deal with this, further research could be undertaken to see how the performance of the models changes when certain variables are added or taken away. This will allow for a better understanding of how the models may be used within the wider context of the research area.

## 5. Conclusion

Overall, the results of the pilot study show that it is possible for a machine learning model to be trained to accurately identify selected activities from a dataset post-data collection. Using a single generalised activity model produced between 55-64% overall accuracy for the portion of training data, and 27-29% overall when applied to House C, depending on whether logistic regression or long short-term memory was used. These results can be significantly improved by changing from a single model to individual specialised activity models. By making this change, the identification accuracies can be improved to 72-92% and 57-79% for the training and House C data respectively, again dependent on the method used.

The success of this method suggests it could be used as a more accurate and less disruptive method of completing occupant behaviour studies within residential settings. This could help further research into how these activities may affect health and identify targeted interventions.

Work to develop this method further is currently being completed to continue to improve the model identification accuracy, including observing additional activities, more reliable validation, and adding further case studies to the training dataset.

## References

[1] Health Effects Institute. *State of Global Air 2024. Special Report.* Boston, MA: Health Effects Institute; 2024.

[2] World Health Organization. Household air pollution. World Health Organization: Geneva, Switzerland;. Available from: https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health.

[3] Hussein T, Paasonen P, Kulmala M. Activity pattern of a selected group of school occupants and their family members in Helsinki-Finland. *The Science of the total environment.* 2012;**425**:289-92.

[4] Matz CJ, Stieb DM, Davis K, Egyed M, Rose A, Chou B, et al. Effects of age, season, gender and urban-rural status on time-activity: CanadianHuman Activity Pattern Survey 2 (CHAPS 2). *International journal of environmental research and public health.* 2014;**11**(2):2108-24.

[5] Gupta R, Kapsali M. Empirical assessment of indoor air quality and overheating in low-carbon social housing dwellings in England, UK. *Advances in Building Energy Research.* 2016;**10**(1):46-68.

[6] McGill G, Oyedele LO, McAllister K. An investigation of indoor air quality, thermal comfort and sick building syndrome symptoms in UK energy efficient homes. *Smart and Sustainable Built Environment.* 2015;**4**(3):329-48.

[7] Vardoulakis S, Dimitroulopoulou C, Thornes J, Lai KM, Taylor J, Myers I, et al. Impact of climate change on the domestic indoor environment and associated health risks in the UK. *Environment international.* 2015;**85**:299-313.

[8] Energy Saving Trust. Energy efficient ventilation in dwellings – a guide for specifiers; 2006. Available from: https://www.beama.org.uk/static/uploaded/0926d783-01b5-4cf1-9d53dc0198e2d0e1.pdf.

[9] Dimitroulopoulou C. Ventilation in European dwellings: A review. *Building and Environment.* 2012;**47**:109-25.

[10] Chatzidiakou L, Krause A, Kellaway M, Han Y, Li Y, Martin E, et al. Automated classification of time-activity-location patterns for improved estimation of personal exposure to air pollution. *Environmental health : a global access science source.* 2022;**21**(1):125.

[11] Eltek Ltd. Eltek GenII AQ211 Manual;. Available from: https://www.manualslib.com/manual/3291499/Eltek-Genii-Aq211.html#manual.

[12] Onset Computer Corporation. Onset HOBO Manual;. Available from: https://www.manualslib.com/manual/1300587/Onset-Hobo.html#product-UX120-006M.

[13] Onset Computer Corporation. Split-core AC current Sensor;. Available from: https://www.onsetcomp.com/products/sensors/ctv-x?srsltid=AfmBOoqEdMNjHZHTX6YQZ5MfnbOSXo26AjTxB1Su9gJN5zF-Q3crBJ50.

[14] Onset Computer Corporation. Hobo U12-012 Manual;. Available from: https://www.manualslib.com/manual/3147901/Hobo-U12-012.html#manual.

[15] Smolic H. How Much Data is Needed for Machine Learning?; 2024. Available from: https://graphite-note.com/how-much-data-is-needed-for-machine-learning.