Sanitary sanity: Evaluating privacy preserving machine learning methods for post-occupancy evaluation

B. HIGGS, B. DOHERTY, N. GARDNER, M. HAEUSLER
BVN & UNSW, Sydney, Australia
BaptisteHiggs@gmail.com, ben@notionparallax.co.uk, n.gardner@unsw.edu.au, m.haeusler@unsw.edu.au

Abstract. Traditional post-occupancy evaluation (POE) of building performance has typically privileged physical building attributes over human behavioural data. This is due to a lack of capability and is especially the case for private spaces such as Sanitary Facilities (SFs). A privacy-preserving sensor-based system using Machine Learning (ML) was previously developed, however it was limited to basic body position classification. Yet, SF usage behaviour can be significantly more complex. This research accordingly builds on the aforementioned work to expand behavioural classifications using a sensor-based ML system. Specifically, the case study uses a GridEYE thermal sensor array, which is trained on a cubicle location within a workplace SF. A variety of ML algorithms are then evaluated on their behaviour-classifying ability. A detailed analysis of behaviour-classification performance is then provided. A system with greater fidelity is thus demonstrated, albeit hampered by imprecise behaviour definitions. Regardless, this contributes to the capability of the broader field of research that is investigating Evidence Based Design (EBD) by extending the ability to examine human behaviour, especially in private spaces. This further contributes to the growing body of work surrounding the provision of SFs.

Keywords. EBD; Data; Internet of Things; Machine Learning; Post Occupancy Evaluation;

1. Introduction

"Although many architects profess interest in post-occupancy performance, only a handful have taken action to derive more reliable and sophisticated ways to gather data" (Shapiro 2019). Current standards such as the BCA (the Building Code of Australia, the document relevant to the jurisdiction of the case study) are built on minimal if any data (Doherty et al. 2019). Mathematical queueing models are put forward as the correct way of assessing these issues, with little practical evidence to back these suggestions up. This is especially demonstratable in SFs, as high privacy requirements result in difficulty in gathering data for EBD. Further, architectural practices generally disregard high fidelity human behavioural data gathering techniques in lieu of physical characteristics of the building such as temperature and volume (Li et al. 2018). The experience of users in a building is assumed to be correlated with physical variables such as these, at least with regard to the gathering of data for further EBD.

There are few existing methods to collect high fidelity behavioural post occupancy information. Computer vision techniques that have been successful in high fidelity classification cannot be used, especially in private spaces, due to infringements on privacy (NSW Government 2005).

This research addresses these issues by evaluating the potential fidelity of a sensor and ML based privacy preserving behavioural classification system. This uses a SF cubicle as a context for the case study, as in Australia it has the greatest necessary privacy provision. The case study specifically deals with the Australian context, although similar concerns are evident internationally. This system develops the ability for architectural practice to firstly understand the usage of SFs and thus improve SF provision, while also contributing to the greater field of higher fidelity in data collection for EBD.

Collecting meaningful data, and thus enabling EBD, allows for design to more appropriately address the real needs of the individuals that inhabit the space. Given that this system is scalable, findings can be more easily applied to a wider variety of spaces than otherwise, even privacy-sensitive spaces, such as a "change room, toilet facility or shower or other bathing facility at a workplace" (NSW Government 2005, p.7).

2. Research Aims

The overarching aim of this research is to develop the capability for greater Post Occupancy Evaluation (POE) fidelity, especially in private spaces. This is especially with the goal of aiding EBD practices, however utilising the technique for EBD is outside the scope of this research. The case study of a SF is chosen as it represents potentially the most private space, and as such the strictest privacy concerns when developing a data-gathering system.

Further, this research also aims to understand the specific behaviours within a SF cubicle context that are able to be classified, both in an attempt to contribute to the field of SF provisions, but also to extrapolate information about characteristics of issues that could be apparent in a broader range of private spaces.

3. Research Questions

Given the aims of the research outlined above, alongside the current state of scholarly research on the topic, the questions that will be addressed by this research are:

What degree of fidelity can be reached when using a combination of sensor technology and ML for a privacy-preserving, data gathering system to measure human behaviour in buildings?

More specifically, how can workplace SFs usage be quantified using such a system?

4. Methodology

When conducting research, there are a variety of approaches a researcher may take. This is true not only for the method used – "a mode of procedure, especially an orderly or systematic mode", but also for the methodology – "a branch of logic dealing with the logical principles underlying the organisation of the various special sciences, and the conduct of scientific inquiry" (Macquarie Dictionary 2017). This research adopted the Action Research (AR) methodology, a term first coined in the mid-20th century by its pioneer, Lewin (1946). Since then the methodology has been built out significantly, with several different types, tools, and implementations now in use (O'Brien 2001). Compared to other methodologies (although not exclusively), AR is learning by doing. It is actively addressing a practical concern of an immediate situation, while also trying to advance scholarly knowledge (O'Brien 2001). Due to this, it may seem similar to problem solving in a consulting context. However, there are some key differences–it has scientific motivations, it is committed to producing knowledge for this wider scientific community, it has an involved rather than objective approach, it founds its recommendations on theoretical frameworks, and it has an experimental model of deriving situational understanding (Lippitt & Lippitt 1978). Although the specific outworking of AR takes many different forms, it is commonly understood to be a cyclical, iterative process of change. A regularly referenced series of stages is described by Baskerville (1999), represented in Figure 1.

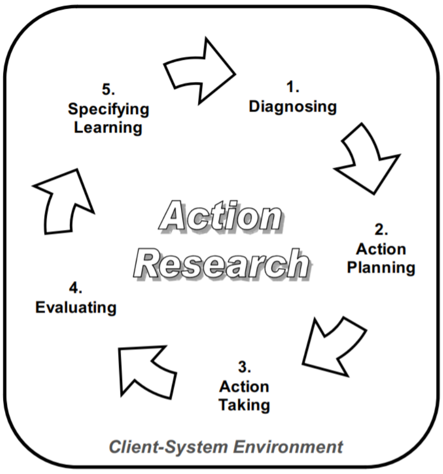

Figure 1. AR stage diagram (Baskerville 1999, p.16, Azhar et al. 2010, p.89)

AR can be applied to a variety of contexts, including the context of construction and engineering. Noting this context, Azhar et al. conclude "it can be recommended that the steps and phases of canonical AR be followed with little modification other than to adapt to the context of the individual research projects" (2010, p.97). With this in mind, AR may be applied to this work. This research adopts the case study of an Australian architectural practice's SFs, with the intent of developing a method for privately classifying behaviour using sensors that may be applied elsewhere. A variety of methods may be used, the GridEYE thermal sensor array was chosen due to its high degree of fidelity without compromising privacy, as well as its use in similar studies with relative success (Berry & Park 2017, Doherty et al. 2019). Using ML techniques to classify between the data points was chosen due to its ability to recognise and accurately distinguish complex patterns. The development of the specific ML models will be where the iterative cycle of the AR methodology is implemented most visibly–the models are trained, results analysed, problems with the results specified, and then the models trained again. SF cubicles were chosen as the context due to the required level of high privacy, thus producing a method that could be employed in many other contexts of less than or equal to levels of privacy.

5. Literature review

SFs have had great significance in most people's day to day lives throughout history (Kira 1976, pp.5–6). Despite this, not much has been written about the design and use of these spaces prior to Alexander Kira's 1966 (later revised in 1976) seminal work, "The Bathroom". This work was the culmination of 17 years of research, and described SF needs and wants qualitatively–only later did Kira collate quantitative data on their usage from various sources (Kira 1976, Kira 1994). Others, some of whom Kira cited, performed time-based studies on SF usage using stopwatches and human observation (Reid & Novak 1975, Anthony & Dufresne 2007, Rawls 1988). In particular, Rawls (1988) asked participants to fill out a survey after using the SFs, which detailed the behaviours they self-professed to. In each of these studies there were gaps of knowledge; 34.7% of SF users' behaviours in Rawls' study were unaccounted for (Rawls 1988, p.81), Reid and Novak only studied urinal usage (Reid & Novak 1975, p.265), and Anthony and Dufresne amalgamated various other pieces of data rather than collect their own (Anthony & Dufresne 2007, pp.272–274).

Given only these data sources, the academic consensus is that current public SF provision is inequitable (Greed 2003, Edwards & McKie 1996, Rawls 1988, Anthony & Dufresne 2007, Molotch & Norén 2010, Banks 1991). Greed suggests that this is to the extent of being unlawful under European Union & Equal Opportunities law (Greed 2003, p.8). Citing Rawls and Kira among others, Anthony and Dufresne alongside Greed agree that female SF provision equity translates to at least double the amount of fixtures (Greed 2003, Anthony & Dufresne 2007, p.272). Despite this, often the opposite is true in England due to the impact of laws enacted less than a century ago (Greed 2019, p.909). Greed (2016, pp.1, 5) suggests that public SF policy "is one of the last frontiers of gender inequality", causing great harm including how "50% of girls in Africa do not continue with school because of lack of toilets".

Understanding the behaviours that occur in SFs is an important step towards closing this inequity of provision. Scholars on the topic recognise that there are an inexhaustible variety of behaviours that occur in these facilities, however mostly from anecdotal evidence (Kira 1976, p.156, Molotch & Norén 2010, p.9). Molotch and Norén extend this, noting that defining the activities that take place in SFs and thus inform their designs turn such people into "moral entrepreneurs", feeling obliged to "assume responsibility for others' actions" (Molotch & Norén 2010, p.8, Becker 2008, p.147). This is especially true when it comes to behaviours such as sex, smoking, and violent crime among many others (ibid). Users often self-select specific SFs to suit their behaviours, one characteristic noted as often being important being privacy. Studying anonymous homosexual intercourse in the 1960s, Humphreys noted "the most active [SFs for homosexual intercourse] studied were all isolated" (1970, p.31). The survey results from Rawls' study suggest however that a minority of these behaviours are common in SFs, with over 80% of the self-reported behaviours of respondents being categorised within the behaviours "Wash Hands, Urinate, Check Appearance, Straighten Clothes, Comb/Brush Hair, Straighten Tie, Talk" (1988, pp.110–118). However, this study has the disadvantage of only accounting for participants that had the willingness and time to participate in the study, and those that did respond subjectively reporting their behaviours. There is little evidence to suggest from the survey of literature that a study has overcome these barriers in monitoring SF behaviours quantitatively using sensors.

According to a recent practice review, built environment firms that use Post Occupancy Evaluation (POE) largely lack quantitative sensor-based behavioural analysis (Li et al. 2018). Over 80% of surveyed POE reports use the most popular subjective method, "Occupant Survey", whereas just over 40% used the most popular Indoor Environmental Quality (IEQ) passive method, "Thermal" (Figure 2). There were no recordings of behavioural analysis.

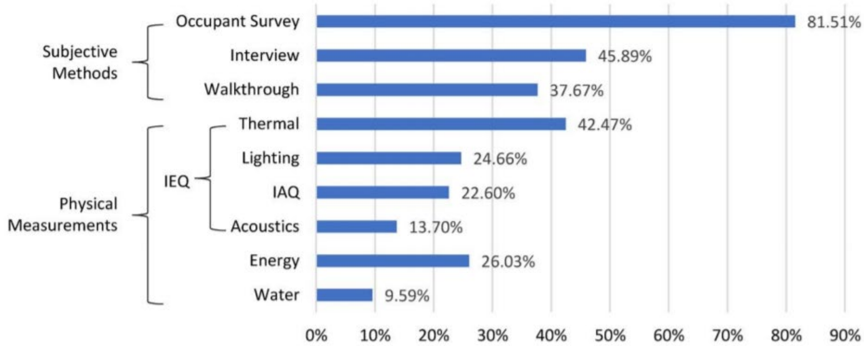

Figure 2. Types of POE used in public reports since 2010 (Li et al. 2018, p.19).

There have been a few notable studies that attempt to uncover behaviour through passive data gathering (Herkel et al. 2008, Sailer et al. 2013, Spinney et al. 2015, Wang & Shao 2018, Berry & Park 2017). Herkel et al recorded user behaviour regarding the opening and closing of windows relative to weather conditions using a bespoke sensor setup which recorded window angle and temperature among other variables (Herkel et al. 2008, p.590). Sailer et al. (2013, p.13) used RFID tags on participants in a workspace environment to track movement behaviour, comparing the results of automated and manual data collation, noting that they each have their own strengths and weaknesses. Spinney et al (2015, p.17) also use RFID tags alongside sitting time & physical activity sensors to demonstrate a system that has a greater fidelity of behavioural information, although they do not recommend RFID systems for future use due to their low spatial resolution. Wang and Shao (2018, pp.22–23) use Wi-Fi instead, noting higher spatial resolution, enabling them to classify behaviour based on location, although only inferentially based on location and duration of stay. All of these locating methods require a user to have a device with them, unlike Berry and Park (2017) who employ GridEye thermopile array sensors to track participant movement in a space. Berry and Park's method is a more widely applicable system while still ensuring privacy unlike popular computer vision classification techniques (Brunetti et al. 2018). The cost of privacy however is that this technique had little accuracy, with predefined walking paths largely deviating from the system's predictions (Berry & Park 2017, pp.142–143). As a potential solution to this, Han (2012) and Yang (2014) among others have found success in combining several lower resolution datasets into one ML classification model. This model however was applied to building occupancy rather than occupant behavioural analysis. None of these aforementioned methods have been applied to a SF context.

The recent article by the authors' is the most pertinent piece of work to this study, using sensor-based technology to quantify presence and behaviour in a workplace SF (Doherty et al. 2019). A three-class model was highly successful in this study for the classification of basic body position classification, showing promise for the expansion into a greater list of classifiable states. This study however does not collect any data on cubicle usage, but rather collects data only on urinal usage, which is demonstrated by Rawls to be a small subset of SF usage (Rawls 1988, pp.132–135). With regard to fidelity of behaviour, the study demonstrates a system capable of only classifying between sitting and standing in a cubicle, which can be used to infer further information about potential behaviours, but with limited detail (Doherty et al. 2019). As discussed above, there are a plethora of behaviours that happen in SFs that cannot be accounted for in the system built in that study. The focus of this study will be to determine the level of behavioural fidelity that can be reached using these sensor-based ML techniques, specifically in a workplace SF.

6. Case Study

This research explores the possibility of using a combination of sensor technology and ML to distinguish between SF behaviours. In doing so, the system aims to further the fields of EBD in private spaces. In order to ensure that the privacy of individuals is maintained in such a system, a sensor that is unable to infringe on privacy was chosen – the GridEYE thermal array sensor. It produces 64 temperatures per reading, an 8x8 grid dependant on the position of the sensor.

6. 1 METHOD

6.1.1 Develop

Initially, a housing for the electronics (GridEYE, Raspberry Pi, and wiring) was developed to support them in their appropriate position. To see more details on the development of this underlying system, see Doherty et al (2019).

Further, the software layer that allows for the training of the ML models had to be developed. The system used includes two main components:

Sensor-capturer: This script interfaces with the GridEYE, collecting temperature information via the I2C protocol. It translates the information into a useful format for transmission (Celsius), writing continuously to a file. The script uses metadata specified in currentMeta.csv when it writes.

Behaviour-informer: This script runs through the specified behaviours, updating the currentMeta.csv files as it runs. This is the script that the operator interacts with during the experiment.

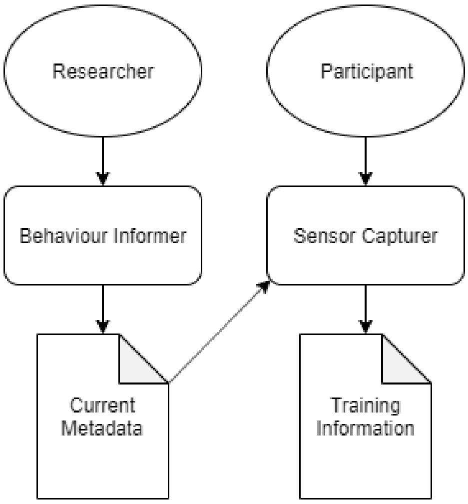

Figure 3. Script architecture diagram.

These scripts must be used together during the running of the experiment. The data-capturer should be run in the background before starting the behaviour-informer script. These components together will capture the information (timestamped heatmaps with associated behaviours) necessary to train the ML models to classify behaviours based on heatmaps.

6.1.2 Recruit

The recruitment process for individuals to help train the system is essential to ensure low bias in the system’s predictions. The recruitment process should involve an equitable cross section of those that may be classified by the system.

1. Individually ask potential participants if they are free at the time of the study and are able and willing to participate.
2. When selecting people to ask, ensure that participants are diverse. Aim to choose individuals who have different body types and sizes to

those that have already volunteered. The final mix should have no one characteristic dominating larger than 70% of the individuals (e.g. 7 men and no women would be inadequate).

3. For those that are willing to participate, provide a copy of the method, alongside detailing the practical requirements of the study in conversation.
4. Provide the specific time and location of the study, allocating provisions for transport.

6.1.3 Set Up

The cubicle must be set up to ensure the experiment can continue without discrepancies between different participants, which would lead to inaccurate classifications. For example, if some of the participants didn't have a pen, the action of writing into a book would be different to those that did.

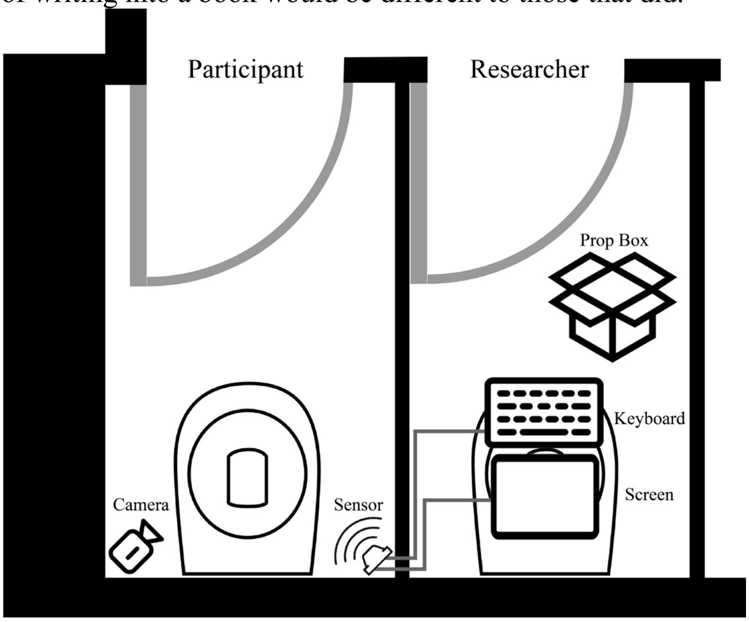

Figure 4. Cubicles layout.

1. Place signs in front of the two cubicles as well as outside the bathroom, informing participants as well as others in a similar vicinity of the experiment taking place. The signs in front of the cubicles in use should display that they are 'temporarily out of service'.
2. Collect the following items in a box in the researcher's cubicle:
   - Lighter & metal rod (to simulate smoking)
   - Eyeglasses

- Contacts case
- Book for reading
- Book for writing
- Pen
- Beer bottle (filled with appropriately chilled water)
- Tampon box
- Jumper/Jacket
- Water bag the size of a baby, heated to body temperature

3. Place the portable screen, keyboard, and portable battery in the researcher's cubicle. Plug the screen into the portable battery, powering it. Plug the screen and keyboard into the sensor
4. Using double sided tape, attach the roof-to-sensor connection module to a location that is square with the corner and allows the sensor to be placed so that it fits tightly with the corner.
5. Using extension cables if necessary, plug the power, keyboard and screen into the sensor. Use duct tape to secure the cables together while keeping them out of the way in the corner, also providing support to relieve weight.
6. Power on the sensor, ensuring that the system works, script included. After this, shut down the system and unplug the screen to conserve battery power.
7. Attach the camera to the ceiling of the cubicle, in the adjacent corner to the sensor. Connect the camera app to ensure that the viewing angle is appropriate. Turn off the camera after finalising its position.

6.1.4 Train System

This is the part that participants will be involved with. Uniformity in how the system is trained and how the participants are interacted with is important.

1. When participants arrive at the location, power on the system.
2. Introduce them to the items for the exercise. Ensure that they are not introduced to how they could act out the given behaviours, rather without actions explain that they may use certain items for certain behaviours.
3. Show the participants the setup of both cubicles, introducing them to what they as well as the researcher will be doing. Ensure that the items are in the researcher's cubicle.
4. Turn on the camera to ensure the entire process can be verified post-experiment if necessary.
5. Lead them into their cubicle. Ensure there is another researcher outside the bathroom to answer any questions non-participating individuals may have during the process.

6. Move into the researcher's cubicle and start the system. Inform the participant that the process has begun.
7. Note the behaviour that is listed as about to be recorded by the script, telling the participant to act out the behaviour. If an item is required, pass it underneath the barrier between the cubicles to the participant.

Figure 5. Footage of participant 'reading' from camera.

8. Start the recording process for that activity after having told the participant to act it out. Train for a total of 20 seconds – although it is more important to ensure that the data recorded is of the participant acting out that behaviour, so stop if they indicate they are about to stop. Once the recording process is stopped, inform the participant it has. Ask the participant to return any items necessary.
9. Repeat steps 6 & 7 for the remainder of the activities that are listed by the system. Ensure that the participant always feels capable and willing to act out the listed behaviour.
10. Tell the participant that the process has finished. Turn off the camera, shut down the system, and unplug the screen to conserve power.
11. Ask the participant to meet you outside. Debrief the experience, asking if they had any difficulty during the process. If there are any noted difficulties in completing the task, take note and remove the data for that behaviour for that individual.
12. Repeat steps 1-11 for at least 6 participants.

6.1.5 Remove

Immediately after the participants have finished, remove the system to safeguard privacy infringements from occurring.

1. After collecting all the data, remove the system from the cubicles. Ensure that the camera, sensor, screen, keyboard and wiring are no longer in the cubicles.
2. Remove all the signs (in front of the cubicles as well as outside the bathroom) that were placed.
3. Ensure that are no traces left by the experiment in the SF.

6.1.6 Extract

Extracting the useful results from the trained ML models provides a platform to analyse the usefulness of using the techniques. Specifically, confusion matrices (a record of the distinguishability of each activity between each other activity for each model) is a useful metric.

1. Start the raspberry pi that collected the data, ensuring that it has an internet connection.
2. Start your personal computer for data analysis, ensuring that it is connected to the same network as the raspberry pi.
3. Use a Secure Copy Protocol (SCP) to transfer the recorded files from the raspberry pi to your data analysis computer.
4. Ensure Python 3.6 is installed on this computer, and that SKLearn, Matplotlib, Tensorflow, and Pandas are installed as libraries.
5. Using Tensorflow, develop a ConvLSTM model that accepts an 8x8 array of temperatures.
6. Using an 85% training to 15% validation split, train the model with the temperature data given.
7. Using the same split of data, use SKLearn to develop the K-Nearest Neighbours (KNN), Support Vector Machine (SVM), and Logistic Regression (LR) models. These models are explained further in the results (6.2.2) section.
8. Observe the results of each of these models, iteratively tuning hyperparameters such as Gamma and Kernel to tune the model until it obtains the highest validation accuracy.
9. Visualise these results using matplotlib to gain further insight into the nature of the relationships between the results.

6. 2 RESULTS

6.2.1 Initial Test

Initially, a list of 7 behaviours was used: sitting, browsing phone, take phone call, standing, straighten clothes, comb brush hair, leave. Six individuals acted these behaviours out, which trained the ConvLSTM model. The result was a 99.8% validation accuracy between each of the behaviours. This result confirmed that these 7 behaviours could easily be distinguished between, and so further behaviours were added.

6.2.2 Further Results

From the literature review, a total of 51 potential behaviours were noted, however, not all could be attempted due to various logistical reasons (e.g. 'bombing' the SF would be difficult to replicate safely and cheaply). Eventually, a list of 27 behaviours was used: breast feeding, urinate, defecate, change clothes, put-in/take-out contacts, clean glasses, take medicine, smoke, talk, change diaper, adjust jewellery/scarf, change pad/tampon, use drugs, nap, read, vandalise, take phone call, drink alcohol, hide, write notes, have solace, cover seat with toilet paper, squat on toilet, eat food, exercise, deal drugs, spy.

The results of using these behaviours for ML model training are varied and have lower accuracies (as would be expected) relative to the initial results that used fewer behaviours. The highest validation accuracies (VAs) achieved are recorded here:

Table 1. Support Vector Machine (SVM) accuracies. Different kernels are different methods for attempting to split up the data. Gamma/Degree are the 'strengths' used for classification (Degree applies to "Kernel: Poly", Gamma applies to the others).

Gamma	**Degree**	**Kernel: RBF VA**	**Kernel: Sigmoid VA**	**Kernel: Poly VA**	**Kernel: Linear VA**
0.001	**2**	31%	4%	48%	49%
0.1	**3**	80%	4%	50%	51%
10	**4**	4%	4%	63%	48%

Table 2. K-Nearest Neighbours (KNN) Accuracies. K is the amount of known closest results the algorithm looks at, before averaging the results to find the correct classification.

K	**1**	**10**	**100**	**1000**
VA	77%	78%	63%	26%

Table 3. Logistic Regression (LR) Accuracies. This algorithm develops a function that gives a probability of each outcome being correct, and chooses the option with the highest.

Classification Strength	**0.001**	**0.1**	**1**	**100**
VA	27%	30%	28%	32%

Table 4. Convolutional Long Short-Term Memory (ConvLSTM) Accuracies. The batch size is how many time-readings previous to the chosen reading are looked at when deciding the classification. Using previous readings is unique to this model, which uses its 'memory' to help it understand what is happening. It also uses 'convolutions', a way to find significant features in each model, simplifying the two-dimensional data.

Batch Size	**2**	**6**	**10**	**14**
VA	4%	4%	4%	4%

The best result achieved by any ML model developed is 80% by the SVM with a Radial Basis Function (RBF) kernel and a 0.1 gamma. A confusion matrix identifies some of the areas that contribute to its downfall. A perfectly performing confusion matrix will have 100% for the right/down diagonal, and 0% for all others. Higher percentages in these other areas indicate behaviours that are mistaken for each other. Figure 6 displays how most activities are being classified correctly, but not all.

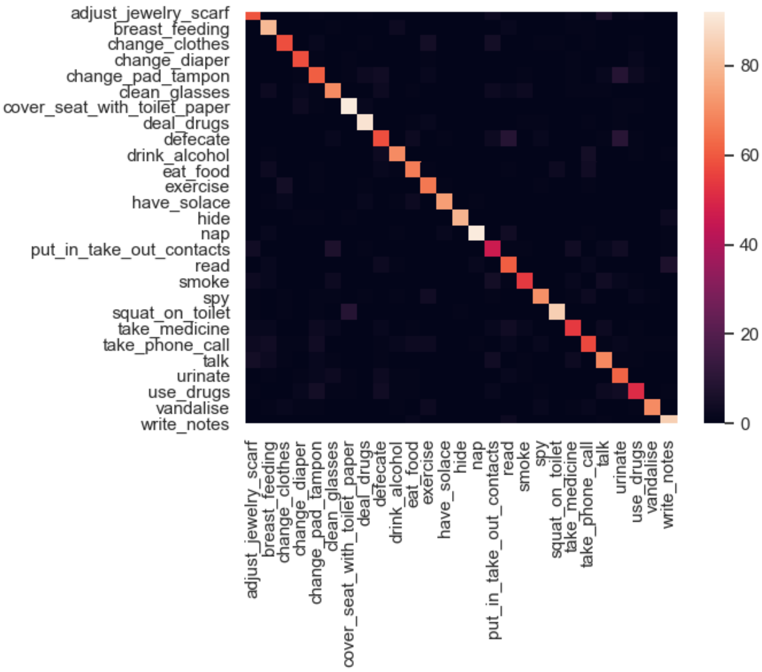

Figure 6. KNN Accuracies (Actual activities vertical, what is inferred by the model horizontal)

The following ranks behaviour combinations that have a confusion rate of over 3% from the same model:

1. Defecate & urinate: 6.5%
2. Adjust jewellery/scarf & talk: 6.0%
3. Defecate & read: 6.0%
4. Change pad/tampon & urinate: 5.5%
5. Change clothes & exercise: 5.0%
6. Clean glasses & put-in/take-out contacts: 5.0%
7. Read & write notes: 5.0%
8. Squat on toilet & cover seat with toilet paper: 4.5%
9. Change pad/tampon & use drugs: 4.0%
10. Put-in/take-out contacts & adjust jewellery/scarf: 3.5%
11. Take phone call & drink alcohol: 3.5%
12. Take phone call & eat food: 3.5%

Some activities, such as defecation, urination, and reading are highly present in wrong inferences. The following ranks how often activities are being mistakenly identified:

1. Urinate
2. Read
3. Change pad/tampon
4. Put-in/take-out contacts
5. Defecate
6. Breast feeding
7. Talk
8. Take phone call
9. Clean glasses
10. Exercise

The two behaviour lists noted previously (mistaken identifications and mistaken pairs) having similar behaviours listed suggests that certain behaviours are more easily misassociated than others. This trend can be confirmed when viewing other ML models:

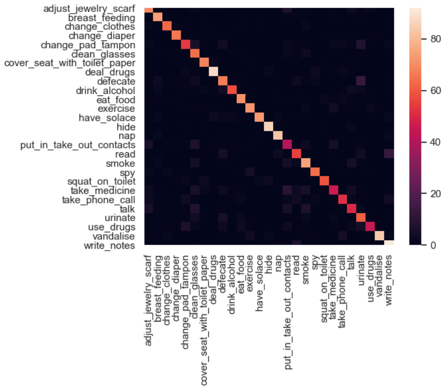

Figure 7. KNN (K: 10)

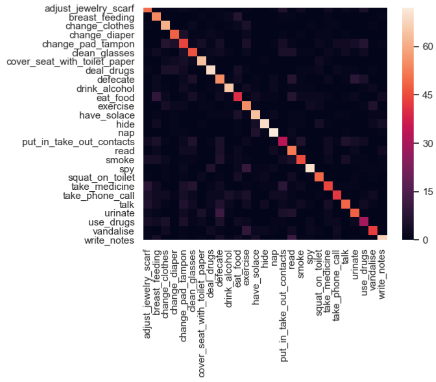

Figure 8. SVM (K: Poly, D: 4)

Table 5. A clearer rank of the misassociated behaviours of the models in Figure 7 & 8

	KNN (K: 10)	SVM (K: Poly, D: 4)
1.	Clean glasses	Defecate
2.	Put-in/take-out glasses	Change pad/tampon
3.	Urinate	Read
4.	Smoke	Exercise
5.	Defecate	Breast feeding

Although there is some variance, there is a trend towards certain activities being more difficult to classify. This can be confirmed by a pair contrast, which used the ConvLSTM model. The pair contrast shows the highest validation accuracy achieved for each pair of behaviours:

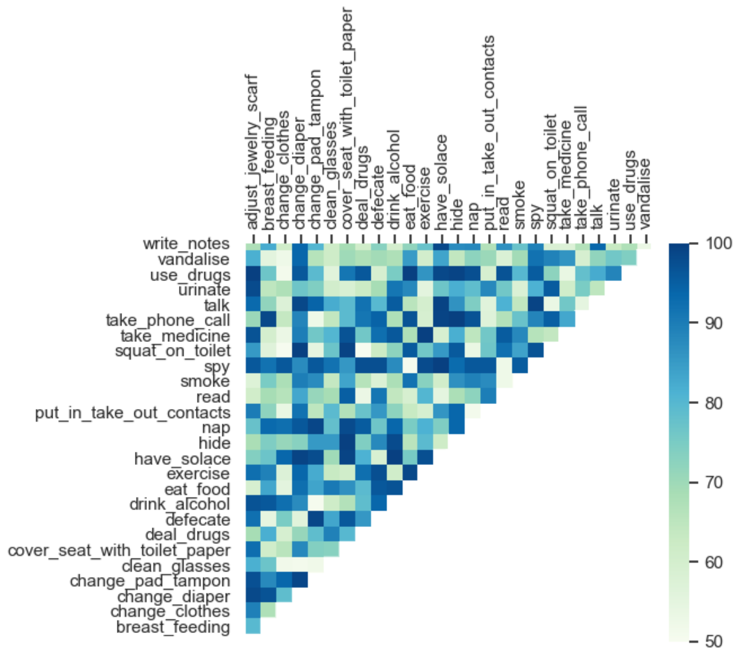

Figure 9. ConvLSTM Pair Contrast. Shows the accuracy percentage of correctly distinguishing between each pair of behaviours.

For example, ‘read’ has quite low distinguishability, as was seen by the other pieces of data, whereas ‘spy’ has quite a high distinguishability. However, there are notable anomalies between this analysis and the others. For example, the behaviour ‘vandalise’ appeared to have a low distinguishability but did not appear to be highly confused with other behaviours in previous figures. This suggests that this behaviour was highly confused with others, but instead of being inferred, it often was inferred to be other behaviours. This is potentially due to the high variability of interpretation for ‘vandalise’. A similar trend can be seen in Figure 6 in the ‘change clothes’ row & column.

7. Discussion (evaluation and significance)

The overarching aim of this research was to evaluate the potential fidelity of a sensor and ML based system for the distinction of behaviours in private spaces. There was also a secondary aim of understanding the particular behaviours that could be identified within a SF cubicle, contributing to the field of SF provision.

With regard to this second aim, minor trends within the data could be identified. Behaviours such as ‘read’, ‘defecate’, ‘urinate’, ‘vandalise’, and ‘change clothes’ were noted to be difficult to distinguish. However, there were two classes of suggested distinguishability–similarity to other behaviours (e.g. defecate & urinate), and variability in interpretation (e.g. vandalise). Overall, the ability of the more successful ML models to accurately distinguish between behaviours was high, with an overall 80% accuracy rate. As this interprets frames of a continuous stream, just one second of data at 5fps should increase the confidence of a consistent reading to well over 99.9% accuracy.

Having said this, there are definite limitations to the research. The list of behaviours selected is in no way comprehensive, although it aims to cover a broad spectrum of behaviours. A fully comprehensive behaviour identifying system may potentially be impossible, as it would be impossible to pre-train a system with infinite classifications, although good design decisions would value such a system. A limited list of inferences could be made from the developed system, albeit better than previously developed systems.

Regarding the primary aim of evaluating the potential fidelity of the system, it has been established that the fidelity of the developed system is greater than previously established systems, although improvements can still be made, and such a system should merely be the starting point for further research into the field. Such research should address further questions on the particular ML models suited to classifying such minimal data, as well as how hardware could be improved to suit the classification of behaviours without infringing privacy. As previously noted, a conglomeration of different sensors and ML models may be suited to solving this issue more effectively (Han et al. 2012, Yang et al. 2014). Further, the definition of what is privacy infringing and what is privacy preserving (especially with reference to cameras) is a difficult question that is paramount to this field, and should be addressed both in a scholarly and legal context.

8. Conclusion

This research study has developed the fidelity of a privacy preserving sensor-based ML system. It has iteratively improved an ML algorithm for a previously optimised sensor placed within a SF cubicle. The research demonstrates that a wide range of activities are possible to distinguish between in a SF setting, albeit with limitations with respect to activities that have limited physical differences, and activities that display a wide range of user interpretations. This contributes to the field of EBD, by empowering the techniques, as well as further aiding the growing discourse on SF provision.

The initial question regarding the degree of behavioural fidelity that can be reached when using a combination of sensor technology and ML while preserving privacy (using a SF as a case study) has been answered, albeit with further investigation being necessary. There will always be further hardware and ML improvements to further investigate. Further, the study includes only a partially comprehensive list of cubicle behaviours–a completely exhaustive list would be ideal, but is potentially impossible. Pre-training a model to classify into set behaviours may be an improper method of capturing cubicle behaviours entirely due to the biases inherent in not allowing for new behaviours.

Regardless, there are many further directions this research points towards taking both in the context of SF provision and EBD data capture. Primarily, this outcome gives rise to the opportunity for closer to holistic data capture of SF behaviours, which would allow for further SF EBD. The narrow application of this method should also be broadened to other spatial scenarios to understand the true potential use of this technique in private places. Overall, the research provides a useful method for high-fidelity privacy preserving sensor-based EBD applications using ML, but should be treated as a step towards further research in a potentially highly impactful field.

Acknowledgements

Thank you to BVN Architecture for their support and allowing us to use their SFs for this research. Thank you to Narridh Khean for his aid in developing the ConvLSTM. Thank you to Aiden Ray & Ishaan Varshney for their expertise in developing the initial physical system. And of course, thank you to the many friends and colleagues who volunteered their time to this research.

References

Anthony, KH & Dufresne, M 2007, 'Potty Parity in Perspective: Gender and Family Issues in Planning and Designing Public Restrooms', Journal of Planning Literature, vol. 21, no. 3, pp. 267–294.

Azhar, S, Ahmad, I, & Sein, MK 2010, 'Action Research as a Proactive Research Method for Construction Engineering and Management', Journal of Construction Engineering and Management, vol. 136, no. 1, pp. 87–98.

Banks, TL 1991, 'Toilets as a Feminist Issue: A True Story', Berkeley Women's Law Journal, vol. 6, no. 1, pp. 263–289.

Baskerville, RL 1999, 'Investigating Information Systems with Action Research', Communications of the Association for Information Systems, vol. 2, viewed 18 October 2019, <https://aisel.aisnet.org/cais/vol2/iss1/19>.

Becker, HS 2008, Outsiders, Simon and Schuster.

Berry, J & Park, KS 2017,'A Passive System for Quantifying Indoor Space Utilization', in, Disciplines & Disruption: Proceedings of the 37th Annual Conference of the Association for Computer Aided Design in Architecture (ACADIA), Cambridge, MA, pp.138–145.

Brunetti, A et al. 2018, 'Computer vision and deep learning techniques for pedestrian detection and tracking: A survey', Neurocomputing, vol. 300, pp. 17–33.

Doherty, B et al. 2019, 'Location intelligence: Developing an indoor-localisation system for quantifying presence and behaviour in office environments', Architectural Science Review.

Edwards, J & McKie, L 1996, 'Women's Public Toilets: A Serious Issue for the Body Politic', European Journal of Women's Studies, vol. 3, no. 3, pp. 215–230.

Greed, C 2003, Inclusive Urban Design: Public Toilets, 1st edition, Taylor & Francis, Amsterdam.

Greed, C 2016, ‘Taking women’s bodily functions into account in urban planning and policy: Public toilets and menstruation’, Town Planning Review, vol. 87, no. 5, pp. 505–24.

Greed, C 2019, ‘Join the queue: Including women’s toilet needs in public space’, The Sociological Review, vol. 67, no. 4, pp. 908–926.

Han, Z, Gao, R, & Fan, Z 2012, ‘Occupancy and indoor environment quality sensing for smart buildings’, 2012 IEEE I2MTC - International Instrumentation and Measurement Technology Conference, Proceedings.

Herkel, S, Knapp, U, & Pfafferott, J 2008, ‘Towards a model of user behaviour regarding the manual control of windows in office buildings’, Building and Environment, vol. 43, no. 4, pp. 588–600.

Humphreys, L 1970, ‘Tearoom trade. Impersonal sex in public places’, Beitrage zur Sexualforschung, vol. 54, no. 0, pp. 1–138.

Kira, A 1976, The Bathroom, New and expanded edition, The Viking Press, New York, N.Y.

Kira, A 1994, ‘Letter from Alexander Kira to Susan Cunningham’.

Lewin, K 1946, ‘Action Research and Minority Problems’, Journal of Social Issues, vol. 2, no. 4, pp. 34–46.

Li, P, Froese, TM, & Brager, G 2018, ‘Post-occupancy evaluation: State-of-the-art analysis and state-of-the-practice review’, Building and Environment, vol. 133, pp. 187–202.

Lippitt, G & Lippitt, R 1978, The consulting process in action, University Associates.

Macquarie Dictionary 2017, Macquarie Dictionary, Seventh, Macquarie Dictionary Publishers, Sydney.

Molotch, H & Norén, L (eds.) 2010, Toilet: Public Restrooms and the Politics of Sharing, NYU Press.

NSW Government 2005, Workplace Surveillance Act 2005.

O’Brien, R 2001, Overview of Action Research Methodology, <http://www.web.ca/~robrien/papers/arfinal.html>.

Rawls, SK 1988, Restroom usage in selected public buildings and facilities: A comparison of females and malesPhD Thesis, Virginia Polytechnic Institute and State University, <https://vtechworks.lib.vt.edu/handle/10919/53598>.

Reid, E & Novak, P 1975, ‘Personal space: An unobtrusive measures study’, Bulletin of the Psychonomic Society, vol. 5, no. 3, pp. 265–266.

Sailer, K, Pachilova, R, & Brown, C 2013,‘“Human Versus Machine”: Testing validity and insights of manual and automated data gathering methods in complex buildings’, in YO Kim, HT Park, & KW Seo (eds.), 9th International Space Syntax Symposium.

Shapiro, GF 2019, Honorable Mention: Post Occupancy Data Device (PODD), a Hands-Free Way to Monitor Projects | Architect Magazine, viewed 29 November 2019, <https://www.architectmagazine.com/awards/r-d-awards/honorable-mention-post-occupancy-data-device-podd-a-hands-free-way-to-monitor-projects_o>.

Spinney, R et al. 2015, ‘Indoor Tracking to Understand Physical Activity and Sedentary Behaviour: Exploratory Study in UK Office Buildings’, PLOS ONE, vol. 10, no. 5, p. e0127688.

Wang, Y & Shao, L 2018, ‘Understanding occupancy and user behaviour through Wi-Fi-based indoor positioning’, Building Research & Information, vol. 46, no. 7, pp. 725–737.

Yang, Z et al. 2014, ‘A systematic approach to occupancy modeling in ambient sensor-rich buildings’, SIMULATION, vol. 90, no. 8, pp. 960–977.