

ARTICLE OPEN



Digital remote monitoring for screening and early detection of urinary tract infections

Alexander Capstick^{1,2}, Francesca Palermo^{1,2,7}, Kimberley Zakka^{3,4,7}, Nan Fletcher-Lloyd^{1,2}, Chloe Walsh^{1,2}, Tianyu Cui^{1,2}, Samaneh Kouchaki^{1,2,5}, Raphaella Jackson^{1,2}, Martin Tran^{1,2}, Michael Crone^{1,2}, Kirsten Jensen^{1,2}, Paul Freemont^{1,2}, Ravi Vaidyanathan^{1,2}, Magdalena Kolanko^{1,2}, Jessica True^{2,6}, Sarah Daniels^{1,2}, David Wingfield^{1,2}, CR&T Group*, Ramin Nilforooshan^{1,2,5,6} and Payam Barnaghi^{1,2,3,4}

Urinary Tract Infections (UTIs) are one of the most prevalent bacterial infections in older adults and a significant contributor to unplanned hospital admissions in People Living with Dementia (PLWD), with early detection being crucial due to the predicament of reporting symptoms and limited help-seeking behaviour. The most common diagnostic tool is urine sample analysis, which can be time-consuming and is only employed where UTI clinical suspicion exists. In this method development and proof-of-concept study, participants living with dementia were monitored via low-cost devices in the home that passively measure activity, sleep, and nocturnal physiology. Using 27828 person-days of remote monitoring data (from 117 participants), we engineered features representing symptoms used for diagnosing a UTI. We then evaluate explainable machine learning techniques in passively calculating UTI risk and perform stratification on scores to support clinical translation and allow control over the balance between alert rate and sensitivity and specificity. The proposed UTI algorithm achieves a sensitivity of 65.3% (95% Confidence Interval (CI) = 64.3–66.2) and specificity of 70.9% (68.6–73.1) when predicting UTIs on unseen participants and after risk stratification, a sensitivity of 74.7% (67.9–81.5) and specificity of 87.9% (85.0–90.9). In addition, feature importance methods reveal that the largest contributions to the predictions were bathroom visit statistics, night-time respiratory rate, and the number of previous UTI events, aligning with the literature. Our machine learning method alerts clinicians of UTI risk in subjects, enabling earlier detection and enhanced screening when considering treatment.

npj Digital Medicine (2024)7:11; <https://doi.org/10.1038/s41746-023-00995-5>

INTRODUCTION

Urinary Tract Infections (UTIs) are one of the most common bacterial infections in older adults, constituting around 25% of all infections^{1–5}. Clinical presentation ranges from self-limited illness to severe sepsis. UTIs account for ~9–31% of cases of severe sepsis which itself has an estimated mortality of 20–40%^{4,6–9}. To differentiate between asymptomatic bacteriuria and UTIs, clinicians rely on positive findings of bacteriuria and genitourinary symptoms. Diagnosis is further complicated by the presence of cognitive impairment or dementia since People Living with Dementia (PLWD) may find it challenging to report their symptoms, and this could result in further complications^{10–12}. As a result, acute infections might not be diagnosed until symptoms require hospitalisation¹³. In the United Kingdom, over 20% of hospital beds are occupied by PLWD, with 9% of these attributed to UTIs^{14–17}.

Currently, a urine sample test and acute changes in baseline cognition are used to diagnose UTIs in PLWD¹⁸. However, samples can be difficult to obtain due to urinary incontinence, cognitive impairment, sample contamination or previous use of antibiotics¹⁹ and are taken on suspicion of an infection, which may be delayed. Additionally, although they can be used as rapid detectors, dipstick tests have a high false positive rate for older adults and require action from the PLWD or their carer which limits their effectiveness for diagnosis^{3,20}. Highlighting UTI risk by identifying

early symptoms would allow for prompt diagnosis, improved health outcomes and effective allocation of healthcare resources.

Machine Learning (ML) offers opportunities for clinical diagnosis and decision-support and recent advances show promise for development of advanced predictive models that incorporate patient data to improve diagnostic performance. For UTI detection in PLWD, ML can improve diagnostic performance and timeliness. Several investigations have been conducted for UTI risk prediction on younger adult populations^{21,22}, which do not generalise to older adults. Existing methods developed for older adults also rely on typical symptoms as predictor variables, precluding their use in community-dwelling patients with dementia with atypical clinical manifestations and who may struggle expressing symptoms²³. In parallel, low-cost monitoring devices have been developed to offer complementary solutions to the typical diagnostic criteria^{24,25}. Rantz et al.²⁶ use activity data collected from in-home Passive Infra-Red (PIR) sensors to detect UTIs in older adults. However, their work is limited to 37 participants and does not utilise ML techniques. The study is also limited to the use of activity data and does not utilise physiological measurements. Work by Enshaeifar et al.¹⁷ employed an unsupervised approach to predict UTIs based on in-home sensors and physiological measurements, however their work showed insufficient diagnostic performance and required the participant to record their own physiology measurements twice a day.

¹Imperial College London, London, UK. ²UK Dementia Research Institute, Care Research and Technology Centre, London, UK. ³University College London, London, UK. ⁴Great Ormond Street Hospital NHS Foundation Trust, London, UK. ⁵University of Surrey, London, UK. ⁶Surrey and Borders Partnership NHS Foundation Trust, Leatherhead, UK. ⁷These authors contributed equally: Francesca Palermo, Kimberley Zakka. *A list of authors and their affiliations appears at the end of the paper.

✉email: alexander.capstick19@imperial.ac.uk; p.barnaghi@imperial.ac.uk

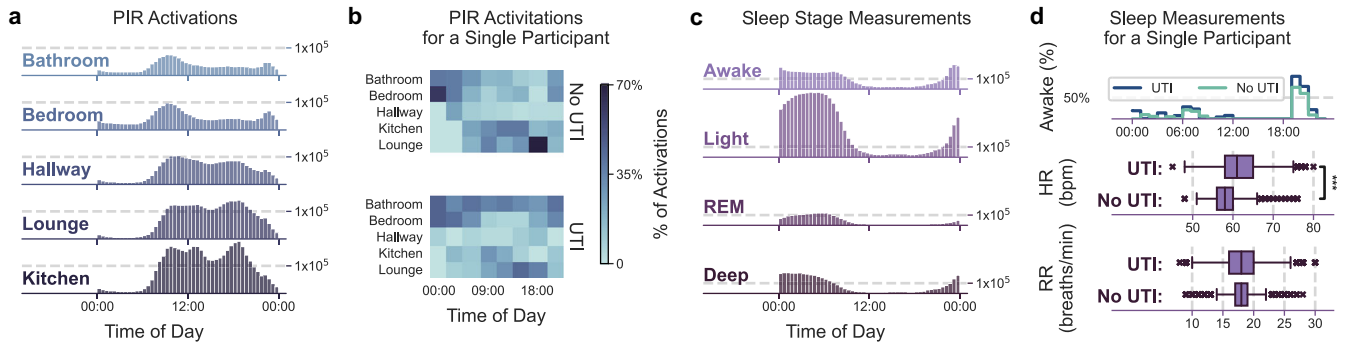


Fig. 1 **The dataset.** **a** All PIR activations measured and the corresponding time of day. **b** The average proportions of PIR activations in a time period for days containing a verified positive and negative UTI label for a selected participant. The difference in Bathroom activity is of note, showing a significant increase in use for those days corresponding to a positive UTI. **c** All sleep stage measurements within the dataset and the corresponding time of day. **d** The first of the graphs show the percentage of a given time window a single participant spent in bed and awake; the second and third shows the difference in heart rate and respiratory rate distributions for days labelled as positive and negative UTI for a single participant. The elements of the boxplots correspond to: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers. Here HR corresponds to heart rate and RR corresponds to respiratory rate.

This study presents a machine learning application to identifying the risk of UTI events in PLWD by analysing symptom-targeted features, engineered from continuous in-home activity and physiology data collected by low-cost and passive sensors (Fig. 1 presents an overview). Then, through optimisation and consultations with clinicians, we determine thresholds for the stratification of the risk scores to improve the algorithm's clinical applicability. The proposed approach has been evaluated in an observational clinical study consisting of 117 participants living with dementia within their own homes. We have worked closely with healthcare professionals to implement a reliable and non-intrusive UTI risk model. Our work will (1) aid clinicians in the early diagnosis of UTIs, and (2) enable a better understanding of in-home behaviour at the point of clinical decision-making. The use of high-resolution in-home observation and measurement data in conjunction with machine learning methods result in timely interventions that can have a significant impact on reducing preventable and unplanned hospital admissions in dementia patients. Such a tool allows for precise collection of urine samples for culture analysis, improved clinical outcomes, a reduction in the burden on healthcare services, and decreased antibiotic overuse and misuse in PLWD by reducing UTI detection time and providing practitioners with more complete pictures of their patients.

RESULTS

Model performance

We examined Logistic Regression (LR), Extreme Gradient Boosting Decision Trees (XGBoost)²⁷, Multilayer Perceptron (MLP), Self-Attention²⁸, Random Forest (RF)²⁹, and Naive Bayes (NB) in their effectiveness to predict UTI events and found the best-performing classification model was LR with L2 regularisation, acting over 3 consecutive days of data. Table 1 presents this model performance on the different data splits. Results from the other models are included in Supplementary Information Section 10. Analysis of model reliability and calibration can be seen in Supplementary Information Section 11.

Risk stratification

To improve flexibility of the model to varying clinical settings, we calculate stratified risk scores as discussed in Section: Stratification of Risk Scores for Clinical Reporting. Figure 2 shows the sensitivity and specificity that can be achieved on the validation set by stratifying the results. By varying the stratification thresholds, sensitivity and specificity can be balanced with the number of people given Green and Red alerts. In Supplementary Information Section 12.1, we present the performance variations when jointly

changing the Red and Green thresholds. Here, we select thresholds [0%, 30%], (30%, 80%), and (80%, 100%) (following interval notation) for groups Green, Amber, and Red respectively. Table 1 shows the results of grouping the risk predictions on the Red and Green groups.

Early detection

We evaluated the model's utility in correctly estimating the risk of UTIs prior to the recorded clinical urine tests. Figure 3 demonstrates specificity, sensitivity and the area under precision-recall curve for days prior to the recorded UTI events. This shows that 2 days prior to a sample test, our model achieved a sensitivity of 64.4 (95% CI = 61.1–67.8), specificity of 68.9 (95% CI = 66.8–71.0), and area under the precision-recall curve of 64.5 (95% CI = 63.0–66.0), and 4 days prior, a sensitivity of 64.4 (95% CI = 61.1–67.8), specificity of 71.9 (95% CI = 67.9–75.8), and area under the precision-recall curve of 65.4 (95% CI = 60.8–70.0).

Feature importance

The most important features influencing predictions were identified using SHapley Additive exPlanations (SHAP)³⁰, a method for producing explainable predictions and calculating contributions from individual features to risk scores. The results of this, on the test set, can be seen in Fig. 4a and reveal that the number of previous confirmed UTI events, the standard deviation of the nocturnal respiratory rate, the nocturnal average heart rate, and the number of nocturnal awake states were positively correlated with a higher risk score. We can also breakdown single predictions to understand contributions to a risk score, as show in Fig. 4b. Further examples can be seen in Supplementary Information Section 13.

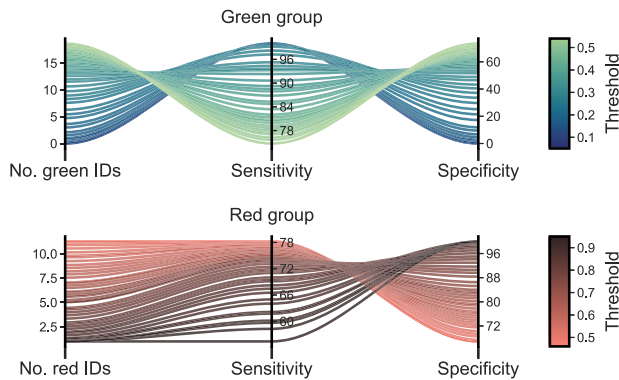
Frequency of generated alerts

To understand the requirements of our model in a clinical setting, we calculated the risk groups of each day of data between 2022/10/05 and 2022/12/01, for each of the PLWD in our dataset. We find that on average, each of the PLWD will receive 0.25 Green alerts, 0.69 Amber alerts and 0.06 Red alerts each day. In Supplementary Information Section 14, we visualise how this risk score varies over time and in Supplementary Information Section 15, we present the model performance on subsets of the sensors. Additionally, in Supplementary Information Section 16, we compare the performance between those with recurrent and non-recurrent UTIs and in Supplementary Information Section 17, we compare the results between Male and Female participants.

Table 1. Mean (95% CI) % of sensitivity, specificity, and area under the precision-recall curve of the UTI prediction model on the different data splits with 10 bootstrap repeats.

Before risk Stratification				
		Sensitivity	Specificity	AUC Precision-Recall
Date	Validation	67.3 (65.9–68.6)	69.1 (66.9–71.4)	67.7 (66.4–69.0)
Date	Test	54.5 (52.7–56.4)	73.0 (71.2–74.8)	54.4 (53.4–55.4)
Date-ID	Validation	87.7 (85.2–90.2)	66.0 (64.4–67.7)	78.3 (76.8–79.7)
Date-ID	Test	65.2 (64.3–66.2)	70.9 (68.6–73.1)	63.5 (61.8–65.2)
After Risk Stratification				
		Sensitivity	Specificity	Precision
Date	Validation	86.6 (80.9–92.3)	94.5 (91.7–97.3)	87.3 (82.2–92.4)
Date	Test	69.0 (64.4–73.5)	94.1 (92.0–96.2)	81.9 (75.5–88.2)
Date-ID	Validation	98.3 (95.5–101.1)	90.0 (85.5–94.5)	81.7 (74.4–89.1)
Date-ID	Test	74.7 (67.9–81.5)	87.9 (85.0–90.9)	77.0 (71.9–82.1)

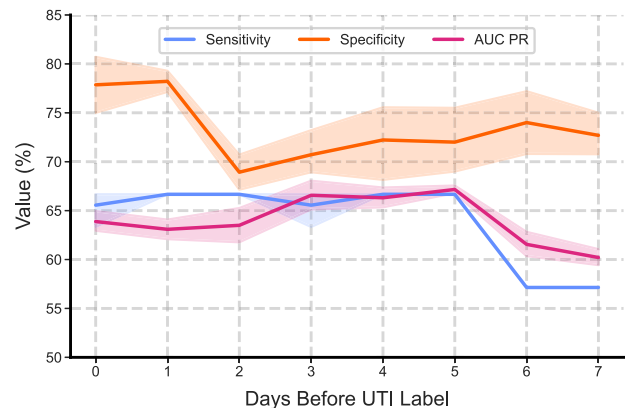
These results are reported before and after risk stratification is performed.

**Fig. 2 Sensitivity and specificity with different thresholds.** The variations in sensitivity and specificity that can be achieved on the validation set by changing the thresholds for defining Green and Red groups. The line colours represent different threshold values. Sensitivity and specificity are calculated on the data in the validation set corresponding to Red and Green alerts. Here, to calculate the metrics on the Green group, the Red threshold was set at >50%, and when calculating the metrics on the Red group, the Green threshold was set at ≤50%. This figure shows the average results from the “Date Split” and “Date-ID Split”.

DISCUSSION

The 2020 report of the Lancet Commission on dementia prevention, treatment, and care emphasises the significance of individualised interventions to address complex medical problems in dementia, which result in unnecessary hospital admissions, accelerated functional decline, and decreased quality of life³¹. An area of priority development is infection prevention and timely detection and treatment³². By conducting preliminary experiments into early identification of possible UTIs in remote healthcare settings, we hope to contribute to directly addressing this priority by investigating more individualised, predictive, and preventative healthcare.

We present a machine learning pipeline for continuous UTI risk screening via analysis of passively collected in-home activity and physiology data. We considered several models and found that LR acting over 3 days attained the top performance (sensitivity of 65.2% (95% CI = 64.3–66.2) and 54.5% (95% CI = 52.7–56.4), and specificity of 70.9% (95% CI = 68.6–73.1) and 73.0% (95% CI = 71.2–74.8) on “Date-ID Split” and “Date Split” respectively).

**Fig. 3 Performance at days prior to UTI label.** The performance of the model when tasked with analysing the risk of a UTI from the test set, at different numbers of days prior to the verified recorded date. The error bands represent the 95% confidence interval (1000 bootstrap samples) of the mean of the values from the 10 bootstrap repeats.

The performance was higher on “Date-ID Split” than “Date Split”, which we hypothesise is due to some PLWD who have opposing labels in the training and testing data. In this case, in “Date Split”, the model might over-fit to the training data from a PLWD. However, in “Date-ID Split” all data seen by the model during testing is from participants not appearing in training. The ratios of positive to negatives in the test sets of the “Date Split” and the “Date-ID Split” are 0.31 and 0.32 respectively and 0.47 and 0.47 for the training set of the “Date Split” and the “Date-ID Split” respectively.

Through stratification, risk scores were transformed into more accessible groups, allowing for the flexible management of actionable alerts within a time period. Following this, the performance on the Green and Red groups were significantly improved, achieving a sensitivity of 74.7% (95% CI = 67.9–81.5) and 69.0% (95% CI = 64.4–73.5), and specificity of 87.9% (95% CI = 85.0–90.9) and 94.1% (95% CI = 92.0–96.2) on the “Date-ID Split” and “Date Split” respectively.

SHAP analysis then highlighted the features most strongly predictive of the risk score. Our analysis shows that an increase in the number of previously confirmed UTI events was associated

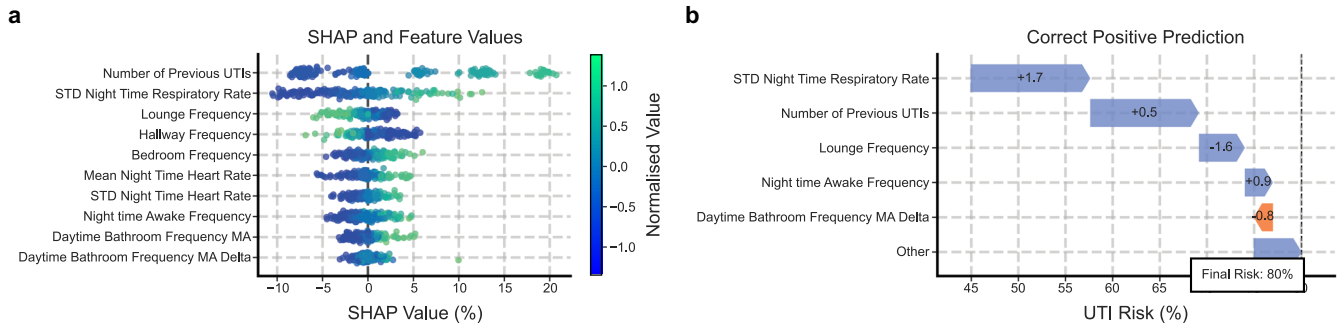


Fig. 4 SHAP values. **a** The feature importance for the top 10 most important features, as calculated by SHAP on the test set, and their corresponding feature values. The colour represents the normalised feature value, whilst the position in the x-axis represents the contribution that value made to the prediction. “MA” refers to the moving average, whilst “Delta” refers to the percentage change in the value from the previous day. These values are calculated on the test set of the “Date-ID Split”. **b** The breakdown of a single prediction shows how each feature contributed to a correct prediction of a positive UTI. Here, the values on the arrows correspond to the normalised feature value in units of standard deviations away from the mean.

with a positive UTI prediction, agreeing with the literature³³. We also highlighted the frequency of the lounge and hallway activations as negatively correlated with risk score, whilst the bedroom frequency was positively correlated. We postulate that this results from participants spending more time in bed due to interrupted sleep, or due to the effects of comorbidities. Third, increases in the standard deviation of the night time respiratory rate and the night time average heart rate were correlated with a higher risk of a UTI. Nocturnal respiratory rate has been linked to stress, reflects physiologic and pathophysiologic determinants, and has been suggested as a biomarker for impending hospitalisation^{34–38}. Increased nocturnal awake occurrences were associated with a higher UTI risk, suggesting PLWD with UTIs were having more disturbed nights of sleep; in agreement with the literature³⁹. This could additionally explain why increased standard deviation of the night time respiratory rate and the night time average heart rate were correlated with a higher risk of a UTI. Considering the clinical manifestations of UTIs in older adults, our feature importance results agree with the current understanding of UTIs in PLWD.

This study contains a few limitations that would also allow for future research directions. Whilst this work was conducted using readily available and low-cost sensors (with preliminary analysis of sensor importance presented in Supplementary Information Section 15), further directions of work could improve the understanding of the balance between the cost and complexity of deployment and UTI risk prediction performance. The deployed PIR sensors allow us to collect data at low cost, but they do not allow for the distinction between data generated by the person of study and other members of the house. Further work could explore methods of passively collecting personalised data. We found that the sleep mat (which does collect personalised data) significantly improved the analysis performance (Supplementary Information Section 15). In Supplementary Information Section 11, we discuss the model reliability and calibration and find that our model overestimates UTI risk, likely because of the data imbalance in the training set. This motivates the applied risk stratification which allows the monitoring team to balance the sensitivity and specificity with the number of generated alerts; however, when deploying this system, work should be done to understand the trade-off between false positives and false negatives to ensure that the risk groups are well-calibrated. Finally, whilst this work focused on an important section of the population (People Living with Dementia), it would be helpful to apply these techniques in a larger cohort study or one containing older adults in community living environments such as care homes, assisted living, or skilled nursing facilities.

Our feasibility study was conducted within real-world in-home settings on data collected in (near) real-time using off-the-shelf and low-cost sensory technologies and engineered, clinically meaningful, features for predicting UTIs determined by clinicians, urine sample analysis, and a clinical monitoring team. We provide preliminary evidence for the use of such an operation and model which, with further testing, could prove to reduce delays in detecting UTIs in PLWD, and potentially reduce the number of avoidable hospital admissions when used to support clinicians with care. The proposed approach can be scaled rapidly and enable human-in-the-loop decision support by taking advantage of technological advancements, cloud computing, and machine learning. Moreover, risk stratification allows for model calibration to improve patient outcomes and care delivery whilst balancing the cost associated with testing for UTIs. Within an ongoing study or in production, the group thresholds can be modified over time to account for care team resources. SHAP analysis will enable the presentation of explainable results (such as in Supplementary Information Section 13), allowing clinicians to explore why the UTI algorithm has made a given prediction. Additional future work will involve continuing to investigate our in-home monitoring systems’ effects on clinical outcomes, as well as patients’ quality of life.

When deployed, our model will be continually trained on new data as collected. To ensure the performance consistently meets a minimum standard, we will routinely evaluate the model on a test set and track its performance. Feature importance will also be monitored to confirm the algorithm is producing clinically founded results. This will enable rapid debugging of errors and maintain a high level of quality in predictions.

METHODS

Study design and population

This study was performed in collaboration with Imperial College London and Surrey and Borders Partnership NHS Trust. Participants were recruited from the following: (1) health and social care partners within the primary care network and community NHS trusts, (2) urgent and acute care services within the NHS, (3) social services who oversee sheltered and extra care sheltered housing schemes, (4) NHS Community Mental Health Teams for older adults (CMHT-OP), and (5) specialist memory services at Surrey and Borders Partnership NHS Foundation Trust. All participants provided written informed consent. Capacity to consent was assessed according to Good Clinical Practice, as detailed in the Research Governance Framework for Health and Social Care (Department of Health 2005) and the Mental Capacity Act 2005. Participants were provided with a Participant Information Sheet

(PIS) that includes information on how the study used their personal data collected in accordance with the GDPR requirements. If the participant was deemed to lack capacity, a personal or professional consultee was sought to provide written consent to the study. Additionally, capacity of both the participant and study partner is assessed at each research visit. Research staff conducting the assessment have completed the NIHR GCP training and Valid Informed Consent training. If a participant is deemed to lack capacity but is willing to take part in the research, a personal consultee is sought in the first instance to sign a declaration of consent. If no personal consultee can be found, a professional consultee, such as a key worker, is sought. This process is included in the study protocol and ethical panel approval is obtained.

Eligible study participants included adults >50 years with a clinically ascertained diagnosis of dementia or mild cognitive impairment and current or previous treatment at a psychiatric unit. Participants lacking capacity for informed consent were required to have a partner or caregiver who had known them for at least 6 months and was able to attend research assessments with them. Exclusion criteria were as follows: (1) patients receiving treatment for terminal illness (2) presence of severe mental health conditions including depression, anxiety, psychosis, and agitation (3) presence of active suicidal thoughts. In total, 117 participants were selected for participation using the above-mentioned recruitment process.

The cohort characteristics can be seen in Table 2 and a patient disposition is available in Supplementary Information Section 1.

Data collection and definition of outcome

Demographic data was collected during the baseline assessment, whilst psychometric scales were used to collect various physical and cognitive data during regular visits. In-home observation and measurement data was obtained using low-cost off-the-shelf monitoring technologies, including PIR sensors (for measuring activity) and sleep monitoring devices. Figure 1 presents cohort-wide sleep and activity activations, and differences in sleep and activity for a participant with both UTI positive and negative days. PIR sensors can detect motion within 9 metres and with a maximum angle of 45° and the sleep mat device can monitor breathing rates, heart rates, and sleep states. For an illustration of the layout of sensors see Supplementary Information Section 2.

Urine samples were collected from several enrolled participants to be labelled by clinicians. Additionally, a baseline algorithm developed in our previous work⁴⁰ suggested patients to the study monitoring team to check for additional symptoms of UTIs and arrange a sample collection and refer to the GP if needed. Once samples were collected, a urine sample analysis was performed and the results sent to clinicians, who with information from the monitoring team, determine a UTI. In total, we have 258 labelled urine samples from 64 participants, of which 81 were confirmed positive UTI cases. If a single day has been labelled, we assume the preceding and proceeding 3 days would also be labelled the same (see Supplementary Information Section 6). This extends the number of labelled days of data to 1752, consisting of 534 positives and 1218 negatives. For our experimentation, we used data collected between 2021/06/28 and 2022/12/01. The models were trained to predict whether a participant had a UTI on a given day (24 h time window). The distribution of labels can be found in Supplementary Information Section 3.

Data pre-processing and feature selection

In addition to sensor readings, we performed feature engineering inspired by well-known symptoms of UTIs such as incontinence, urgency and increased frequency of urination, and behavioural changes (<https://www.nhs.uk/conditions/urinary-tract-infections->

Table 2. Characteristics of the study cohort.

Characteristic	Entire cohort		Labelled cohort	
	No.	%	No.	%
Sex				
Female	54	46	27	42
Male	63	54	37	58
Birth year				
1920-1930	13	11	5	8
1930-1940	47	40	27	42
1940-1950	41	35	26	41
1950-1960	13	11	5	8
1960-1970	2	2	1	2
1970-1980	1	1	0	0
Ethnicity				
White	95	81	60	94
Asian	8	7	3	5
Black/African/Caribbean	3	3	0	0
Mixed/Multiple Groups	1	1	0	0
N/A	10	9	1	2
Household				
Lives Alone	45	38	16	25
Lives with Partner	60	47	73	73
N/A	12	10	1	2
Primary diagnosis				
Alzheimer's	61	52	39	61
Vascular Dementia	10	9	5	8
Parkinson's	5	4	2	3
Other and Mixed	40	34	18	28
N/A	1	1	0	0

Some participants requested not to share their information outside the study and correspond to the Not Available information.

utis/) to allow clinical interpretability and improve model performance and generalisability.

Raw features were: (1) frequency of bathroom, bedroom, hallway, kitchen, lounge activations; (2) mean and standard deviation of nocturnal heart rate and respiratory rate; (3) nocturnal awake occurrences. Engineered features were: (4) bathroom day and nocturnal frequencies, moving average, and percentage change; (5) mean and standard deviation of the movement time from any location within the house to bathroom; (6) daily entropy in PIR sensor activation; (7) number of previous UTIs to date. More information on the features selected can be found in Supplementary Information Section 4.

Data collection occurred outside controlled environments using in-home devices so missing measurements inevitably occurred. To limit the effects of incomplete data⁴¹, we imputed missing values based on strategies depending on the given features (see Supplementary Information Section 5 for more information).

Analysis platform

All analyses were performed on a secure computing environment at Imperial College London using Python version 3.9. The Pandas⁴², Numpy⁴³, Scikit-Learn⁴⁴, and Pytorch⁴⁵ packages formed much of our pipeline.

Methodology

To ensure generalisability, we evaluated our work in two different ways.

Firstly, the dataset was split temporally into training and testing subsets in an 80:20 ratio. The data collected from 2021/06/28 to 2022/10/05 represented 80% ($n = 1394$ days in total, from $p = 54$ participants) of the dataset, whilst the data between 2022/10/05 and 2022/12/01 represented 20% ($n = 358$, $p = 39$).

This formed the first analysis, evaluating the model at making predictions on future data from the same cohort as it was trained on. We will refer to this experimental setting as "Date Split".

In the second analysis, we used a leave-one-out cross-validation strategy⁴⁶. Here, data was split in the same way as in the first evaluation method. Then, training and testing of our machine learning models was performed using a leave-one-out strategy on data from each of the PLWD. This way, we are able to test the model performance on data from participants outside of the cohort it has been trained on. We will refer to this experimental setting as "Date-ID Split".

During model development and whilst optimising model parameters, validation sets were produced by splitting the training data on the date 2022/09/11. All experiments were performed multiple times, with each run using a bootstrap sample⁴⁶ of the training set to ensure reproducibility. See Supplementary Information Section 8 for a visualisation of this evaluation.

We used sensitivity, specificity, and area under the precision-recall curve to measure model performance (for definitions of metrics, please see Supplementary Information Section 7).

Model development

We tested Logistic Regression (LR), Extreme Gradient Boosting Decision Tree (XGBoost)²⁷, Multilayer Perceptron (MLP), Self-Attention²⁸, Random Forest (RF)²⁹ and Naive Bayes (NB) models at predicting the risk of UTIs. Hyper-parameters were tuned using Bayesian optimisation on train-validation splits, with the model producing the highest area under the precision-recall curve (on validation data) selected for the final analysis. The number of days of data used as input to the model was jointly tested, ranging from 1 day to 7 days. Supplementary Information Section 9 contains information on the decisions made regarding each step of the UTI model pipeline.

Stratification of risk scores for clinical reporting

Risk scores from the model are stratified into three groups, used to inform clinical decisions in a concise way and provide precise control over the number of actionable alerts. Outputs are split into the groups Green, Amber, and Red; referring to minimal, medium, and high risk of a UTI respectively. By varying these thresholds, we can balance levels of sensitivity and specificity for the different groups with the number of alerts. This allows our process to be flexible to different clinical scenarios and resources.

Within this work, the optimal thresholds used in our analysis of results were calculated using the algorithm's predictions on the data collected between 2022/09/11 and 2022/10/05 (validation data), and with feedback from a clinical team. More information on the risk stratification is included in Supplementary Information Section 12.

Ethics approval

The study received ethical approval from the London-Surrey Borders Research Ethics Committee; TIHM 1.5 REC: 19/LO/0102. The study is registered with National Institute for Health and Care Research (NIHR) in the United Kingdom under Integrated Research Application System (IRAS) registration number 257561.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

CODE AVAILABILITY

The code used in this study will be made available by the corresponding author upon reasonable request.

Received: 18 August 2023; Accepted: 11 December 2023;

Published online: 13 January 2024

REFERENCES

1. Foxman, B. & Brown, P. Epidemiology of urinary tract infections. *Infect. Dis. Clin. North Am.* **17**, 227–241 (2003).
2. Linhares, I., Raposo, T., Rodrigues, A. & Almeida, A. Frequency and antimicrobial resistance patterns of bacteria implicated in community urinary tract infections: a ten-year surveillance study (2000–2009). *BMC Infect. Dis.* **13**, 19 (2013).
3. Mody, L. & Juthani-Mehta, M. Urinary tract infections in older women: a clinical review. *JAMA* **311**, 844 (2014).
4. Gharbi, M. et al. Antibiotic management of urinary tract infection in elderly patients in primary care and its association with bloodstream infections and all cause mortality: population based cohort study. *BMJ* **364**, E23001 (2019).
5. Godbole, G. P., Cerruto, N. & Chavada, R. Principles of assessment and management of urinary tract infections in older adults. *J. Pharm. Pract. Res.* **50**, 276–83 (2020).
6. Wagenlehner, F. M. et al. Diagnosis and management for urosepsis: Items in urosepsis. *Int. J. Urol.* **20**, 963–70 (2013).
7. Tal, S. et al. Profile and prognosis of febrile elderly patients with bacteremic urinary tract infection. *J. Infect.* **50**, 296–305 (2005).
8. Martin, G. S., Mannino, D. M. & Moss, M. The effect of age on the development and outcome of adult sepsis*. *Crit. Care Med.* **34**, 15–21 (2006).
9. Levy, M. M. et al. Outcomes of the surviving sepsis campaign in intensive care units in the USA and Europe: a prospective cohort study. *Lancet Infect. Dis.* **12**, 919–924 (2012).
10. Arvanitakis, Z., Shah, R. C. & Bennett, D. A. Diagnosis and management of Dementia: review. *JAMA* **322**, 1589 (2019).
11. Cooper, C. et al. Inequalities in receipt of mental and physical healthcare in people with dementia in the UK. *Age Ageing* **52**, afw208v1 (2016).
12. Callahan, C. M. et al. Errors in self-reports of health services use. *Alzheimer Dis. Assoc. Disord.* **29**, 75–81 (2015).
13. Toot, S., Devine, M., Akporobaro, A. & Orrell, M. Causes of hospital admission for people with dementia: a systematic review and meta-analysis. *J. Am. Med. Direct. Assoc.* **14**, 463–70 (2013).
14. Sampson, E. L., Blanchard, M. R., Jones, L., Tookman, A. & King, M. Dementia in the acute hospital: prospective cohort study of prevalence and mortality. *Br. J. Psychiat.* **195**, 61–66 (2009).
15. Rao, A., Suliman, A., Vuik, S., Aylin, P. & Darzi, A. Outcomes of dementia: systematic review and meta-analysis of hospital administrative database studies. *Archiv. Gerontol. Geriatr.* **66**, 198–204 (2016).
16. Mukadam, N. & Sampson, E. L. A systematic review of the prevalence, associations and outcomes of dementia in older general hospital inpatients. *Int. Psychogeriatr.* **23**, 344–355 (2011).
17. Enshaeifar, S. et al. Machine learning methods for detecting urinary tract infection and analysing daily living activities in people with dementia. *PLoS One* **14**, e0209909 (2019).
18. Krinitski, D., Kasina, R., Klöppel, S. & Lenouvel, E. Associations of delirium with urinary tract infections and asymptomatic bacteriuria in adults aged 65 and older: a systematic review and meta-analysis. *J. Am. Geriatr. Soc.* **69**, 3312–23 (2021).
19. Woodford, H. J. & George, J. Diagnosis and management of urinary infections in older people. *Clin. Med.* **11**, 80–3 (2011).
20. Deville, W.L., Yzermans, J.C., Duijn, N.P., Bezemer, D. et al. The urine dipstick test useful to rule out infections. A meta-analysis of the accuracy. *BMC Urol.* <https://doi.org/10.1186/1471-2490-4-4> (2004).
21. Little, P. et al. Developing clinical rules to predict urinary tract infection in primary care settings: sensitivity and specificity of near patient tests (dipsticks) and clinical scores. *Br. J. Gener. Pract.* **56**, 606–12 (2006).

22. McIsaac, W. J., Moineddin, R. & Ross, S. Validation of a decision aid to assist physicians in reducing unnecessary antibiotic drug use for acute cystitis. *Arch. Intern. Med.* **167**, 2201–6 (2007).
23. Yourman, L. C., Kent, T. J., Israni, J. S., Ko, K. J. & Lesser, A. Association of dementia diagnosis with urinary tract infection in the emergency department. *J. Am. College Emerg. Phys. Open* **1**, 1291–6 (2020).
24. Majumder, S. et al. Smart homes for elderly healthcare—recent advances and research challenges. *Sensors* **17**, 2496 (2017).
25. Turjamaa, R., Pehkonen, A. & Kangasniemi, M. How smart homes are used to support older people: an integrative review. *Int. J. Old. People Nurs.* **14**, e12260 (2019).
26. Rantz, M.J. et al. Using sensor networks to detect urinary tract infections in older adults. In *Proceedings of IEEE International Conference on e-Health Networking Applications and Services (Healthcom)* 142–9 (IEEE, 2011).
27. Chen, T., Guestrin, C. et al. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–945 (ACM, 2016).
28. Vaswani, A. et al. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* 6000–6010 (Curran Associates, 2017).
29. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
30. Lundberg, S. M. et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2**, 749 (2018).
31. Livingston, G. et al. Dementia prevention, intervention, and care: 2020 report of the Lancet commission. *Lancet* **396**, 413–46 (2020).
32. Pickett, J. et al. A roadmap to advance dementia research in prevention, diagnosis, intervention, and care by 2025. *Int. J. Geriatr. Psychiat.* **33**, 900–6 (2018).
33. Aydin, A., Ahmed, K., Zaman, I., Khan, M. S. & Dasgupta, P. Recurrent urinary tract infections in women. *Int. Urogynecol. J.* **26**, 795–804 (2014).
34. Harrington, N. et al. Nocturnal respiratory rate dynamics enable early recognition of impending hospitalizations. *medRxiv* <https://doi.org/10.1101/2022.03.10.22272238> (2022).
35. Seymour, C. W. et al. Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* **315**, 762 (2016).
36. Smith, G. B., Prytherch, D. R., Meredith, P., Schmidt, P. E. & Featherstone, P. I. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation* **84**, 465–470 (2013).
37. Subbe, C. P. Validation of a modified early warning score in medical admissions. *QJM* **94**, 521–526 (2001).
38. Tobin, M. J. et al. Breathing patterns. *Chest* **84**, 202–205 (1983).
39. Chartier-Kastler, E. & Davidson, K. evaluation of quality of life and quality of sleep in clinical practice. *Eur. Urol. Suppl.* **6**, 576–84 (2007).
40. Li, H. et al. Deep representation for connected health: semi-supervised learning for analysing the risk of urinary tract infections in people with dementia. *arXiv* <https://arxiv.org/abs/2011.13916> (2022).
41. Aste, M., Boninsegna, M., Freno, A. & Trentin, E. Techniques for dealing with incomplete data: a tutorial and survey. *Patt. Anal. Appl.* **18**, 1–29 (2014).
42. McKinney, W. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* 2nd edn, Vol. 3 (eds Walt, V.S. & Millman, J.) Ch 51–56 (Springer, 2010).
43. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–62 (2020).
44. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–30 (2011).
45. Paszke, A. et al. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems* 8024–3566 (Curran Associates, 2019).
46. Hastie, T., Tibshirani, R., Friedman, J. et al. *The Elements of Statistical Learning* (Springer, 2009).

CR&T GROUP

Alexander Capstick ^{1,2}✉, Francesca Palermo^{1,2,7}, Nan Fletcher-Lloyd ^{1,2}, Chloe Walsh ^{1,2}, Tianyu Cui^{1,2}, Samaneh Kouchaki ^{1,2,5}, Raphaella Jackson^{1,2}, Martin Tran^{1,2}, Michael Crone ^{1,2}, Kirsten Jensen^{1,2}, Paul Freemont ^{1,2}, Ravi Vaidyanathan ^{1,2}, Magdalena Kolanko^{1,2}, Jessica True^{2,6}, Sarah Daniels ^{1,2}, David Wingfield^{1,2}, Ramin Nilforooshan ^{1,2,5,6} and Payam Barnaghi ^{1,2,3,4}✉

A full list of members and their affiliations appears in the Supplementary Information.

ACKNOWLEDGEMENTS

This study is funded by the UK Dementia Research Institute (UKDRI) Care Research and Technology Centre funded by the Medical Research Council (MRC), Alzheimer's Research UK, Alzheimer's Society (grant number: UKDRI-7002), and the UKRI Engineering and Physical Sciences Research Council (EPSRC) PROTECT Project (grant number: EP/W031892/1). Infrastructure support for this research was provided by the NIHR Imperial Biomedical Research Centre (BRC) and the UKRI Medical Research Council (MRC). The funders were not involved in the study design, data collection, data analysis or writing the manuscript.

AUTHOR CONTRIBUTIONS

AC, FP: Conceptualisation, Methodology, Software, Formal analysis, Investigation, Data Processing, Writing—Original Draft, Review and Editing, Visualisation; KZ, NFL: Writing—Original Draft, Review and Editing; CW: Writing—Original Draft, Review and Editing, Data Collection; TC: Methodology, Writing—Review and Editing; SK: Methodology, Writing—Original Draft, Review and Editing; RJ, MT, MC, KJ, RV, MK, SD: Reviewing, Data Collection; PF: Reviewing, Data Collection, Funding Acquisition; JT: Data Collection; DW: Methodology, Data Collection; RN: Clinical Study Lead, Conceptualisation, Data Collection, Writing—Review and Editing, Funding Acquisition; PB: Conceptualisation, Methodology, Writing—Original Draft, Review and Editing, Supervision, Funding Acquisition; FP and KZ contributed equally to this work.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-023-00995-5>.

Correspondence and requests for materials should be addressed to Alexander Capstick or Payam Barnaghi.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024