

EMPIRICAL RESEARCH

Open Access



Automatic classification of the physical surface in sound uroflowmetry using machine learning methods

Marcos Lazaro Alvarez^{1*} , Laura Arjona^{1†}, Miguel E. Iglesias Martínez^{2†} and Alfonso Bahillo^{3†}

Abstract

This work constitutes the first approach for automatically classifying the surface that the voiding flow impacts in non-invasive sound uroflowmetry tests using machine learning. Often, the voiding flow impacts the toilet walls (traditionally made of ceramic) instead of the water in the toilet. This may cause a reduction in the strength of the recorded audio signal, leading to a decrease in the amplitude of the extracted envelope. As a result, just from analysing the envelope, it is impossible to tell if that reduction in the envelope amplitude is due to a reduction in the voiding flow or an impact on the toilet wall. In this work, we study the classification of sound uroflowmetry data in male subjects depending on the surface that the urine impacts within the toilet: the three classes are water, ceramic and silence (where silence refers to an interruption of the voiding flow). We explore three frequency bands to study the feasibility of removing the human-speech band (below 8 kHz) to preserve user privacy. Regarding the classification task, three machine learning algorithms were evaluated: the support vector machine, random forest and k-nearest neighbours. These algorithms obtained accuracies of 96%, 99.46% and 99.05%, respectively. The algorithms were trained on a novel dataset consisting of audio signals recorded in four standard Spanish toilets. The dataset consists of 6481 1-s audio signals labelled as silence, voiding on ceramics and voiding on water. The obtained results represent a step forward in evaluating sound uroflowmetry tests without requiring patients to always aim the voiding flow at the water. We open the door for future studies that attempt to estimate the flow parameters and reconstruct the signal envelope based on the surface that the urine hits in the toilet.

Keywords Sound uroflowmetry, Machine learning, Automatic classification, Surface automatic classification, Acoustic voiding signals

1 Introduction

The growing interest in information and communication technologies is generating a paradigm shift in current health care systems, which are transitioning from face-to-face and reactive systems to remote and proactive systems. This has mutual benefits, as it provides advantages both for patients living in rural and hard-to-reach areas who have difficulty accessing such services and for healthcare providers, as it allows them to access up-to-date medical information and resources quickly and efficiently. As a result, the quality of medical care is improved and the associated costs are reduced.

[†]Marcos Lazaro Alvarez, Laura Arjona, Miguel E. Iglesias Martínez, and Alfonso Bahillo contributed equally to this work.

*Correspondence:

Marcos Lazaro Alvarez
alvarez.marcoslazaro@deusto.es

¹ Faculty of Engineering, University of Deusto, Bilbao 48007, Spain

² Grupo de Modelización Interdisciplinar, Instituto Universitario de Matemática Pura y Aplicada, Universitat Politècnica de València, Valencia 46022, Spain

³ Department of Signal Theory and Communications, Universidad de Valladolid, Valladolid 47011, Spain

One of the problems currently affecting the ageing population is lower urinary tract symptoms (LUTS). LUTS affect bladder storage, emptying and postvoiding, and they mostly affect the ageing male population and are caused by benign prostatic hyperplasia (BPH) [1]. LUTS lead to a decreased quality of life and a significant expenditure of health care resources [2].

It is estimated that more than 60% of the population of men over 60 years of age suffer from LUTS [3]. There is a non-invasive clinical test that is widely used to assess urinary tract function called uroflowmetry (UF) [4]. UF is used to provide objective evidence to evaluate the degree of prostate enlargement, overactive bladder, urinary incontinence and neurogenic bladder [5]. UF is performed with a uroflowmeter, a device that measures the bladder emptying rate as a function of time, the total volume voided and the duration of the process. With these values, urologists can obtain criteria to determine how well the urinary tract is functioning and thus obtain a diagnosis. A limitation associated with this test is that it generates situational stress in patients; this is known as shy bladder syndrome [6]. The patient is asked to void on demand in an unnatural environment, often with a very full bladder. This situation generates significant variability from test to test. As a result, it is recommended that more than one test should be performed, requiring several visits, which can be time-consuming and costly, to a clinic [7].

As an alternative that allows flow parameters to be measured remotely and in a natural environment, sound uroflowmetry (SU) has emerged; it attempts to estimate the flow parameters from the sound generated by the impact of urine on the water in the toilet. It has been shown that there is a good correlation between the flow parameters obtained by UF and those obtained from SU and the shapes of the visual flow traces [8, 9]. Multiple platforms have been developed to perform SU using various hardware configurations. These platforms make use of dedicated microphones [10] and use general-purpose devices such as smartphones [9, 11, 12], and recently, the first platform for performing SU using smart watches was developed and validated [13, 14].

One of the limitations associated with the SU test is that the person must aim the voiding flow at the water in the toilet at all times. If the voiding flow impacts the toilet walls (made of ceramic) instead of the water base, the sound units captured by the recording device decrease as a result of the change in the physical surface. This results in prediction errors in the flow and envelope parameters of the signal: the sound produced by the impact of the voiding flow on ceramic could be wrongly interpreted as a flow interruption or a decrease

in the flow rate. This limitation can become a serious problem if we consider the fact that the majority of the target population undergoing the test are elderly people. To address this limitation, in this work, we seek to explore SU audio signals from male subjects to extract patterns related to their characteristics in the frequency domain by using the fast Fourier transform (FFT) to identify the voiding flow impact surface. We develop a three-class classification algorithm with a high accuracy (ACC) for the detection of time intervals in which there is voiding against ceramic, voiding against water and silence in SU tests.

For this purpose, this work makes use of a set of machine learning (ML) algorithms from mixed voiding event data obtained during this study. This work hopes to provide an essential step forward to improve the performance of SU tests and contribute to increasing their reliability by removing the requirement that patients always target water in SU tests; instead, this method detects the physical environment and acts accordingly.

The paper is organised as follows: Section 2 briefly reviews the state of the art for audio feature extraction and classification using ML; Section 3 presents the materials and methods proposed in this research, where the specific characteristics of the dataset, feature selection and the theoretical foundations of the ML algorithms used in the classification process are described; Section 4 shows the results obtained from the proposed methodology; and, finally, Section 5 provides some concluding remarks.

2 Related work

2.1 Feature extraction in audio signals

Feature extraction is the process of identifying the distinctive properties of a signal [15], which are subsequently used as inputs for classification methods. Features can be extracted from signals in one of three domains: the frequency domain, the time domain and the time-frequency domain. In the frequency domain, spectral components obtained using the FFT [16], mel spectrograms [17] and mel frequency cepstral coefficients (MFCCs) [18] are conventionally used. In the time domain, several statistics have been used to characterise the discriminant information, such as the zero crossing rate (ZCR) and kurtosis [19]. Finally, in [20, 21], novel approaches for the computational analysis of auditory scenes using time-frequency representations and discriminative content extraction are performed.

Within all domains, the frequency domain includes a wide variety of representations [22], and MFCCs have been used extensively with both classical and deep learning approaches to obtain a high ACC [23].

2.2 Audio signal classification using ML

Audio classification has become a focus of attention in audio processing and pattern recognition research. It is difficult to find an optimal classifier and to select the optimal features from several features extracted from an audio fragment. Several methods have been proposed; they range from traditional signal processing techniques to more recent techniques using deep learning approaches. In [24, 25], support vector machine (SVM)-based classifiers were proposed for audio signal classification. Other works made use of the SVM and random forest (RF), and a comparison of the behaviour of both classifiers showed that better results are obtained using RF [23, 26].

With the advent of deep learning, more advanced techniques have been developed that can learn sound tagging tasks exceptionally well; they have become the standard in mobile and embedded applications [27]. These techniques include the convolutional neural network (CNN), recurrent neural network (RNN) and their variants, such as convolutional recurrent neural networks (CRNNs). In [28], an extensive study investigating CNN sets for audio classification is carried out, and in [29], a study using an RNN to classify environmental sound signals is carried out; very satisfactory results were obtained in both cases.

In summary, automatic audio classification is an active area of research, and there have been significant advances in both traditional and deep learning-based approaches. In this paper, we develop a classification algorithm to determine the surface in SU tests to classify when voiding against water or against ceramic is occurring or when there is silence (absence of voiding). To the best of our knowledge, there are no previous works that use ML for surface classification in SU tests. As a result, there are no datasets of voiding sounds that include the three sound labels, so we have created a dataset of labelled sounds that was used to train our ML algorithms.

3 Materials and methods

3.1 Dataset description

For the classification task, we have created, from real voiding event audio recordings that have been segmented into 1-s chunks, a dataset of 6481 1-s audio clips recorded with a professional microphone, the Ultramic384. This device has a highly sensitive audio sensor that allows a sampling rate (SR) of 384 kHz, allowing the study of a wide frequency spectrum. All the voiding audio clips recorded were obtained from 15 male subjects voiding in a standing up position.

The audio recordings were carried out in four Spanish domestic bathrooms, where the height of the toilet water from the floor was approximately 15 cm. The recording

device was placed above the water tank of the toilet, with an approximate height above the floor of 90 cm. The audio clips are composed of three classes: voiding against ceramic (ceramic class), voiding against water (water class) and silence (silence class). The ceramic, water and silence classes represent 32.5 %, 34 % and 33.5 % of the total recordings, respectively. The experimental procedures conform to the provisions of the Declaration of Helsinki (as revised in Edinburgh in 2000). Table 1 shows the proportions of the audio clips recorded in each of the bathrooms according to the class and the dimensions of each of the bathrooms. The procedure for the collection of the audio recordings of each class is detailed below:

- **Ceramic class:** It is composed of 2108 1-s audio clips that correspond to 104 voiding events of 15 different subjects who were aiming at the toilet wall. Subsequently, we took the time intervals of the recordings in which only ceramics surface sounds were present based on the validation of the participants and fragmented the audio recordings into 1-s frames.
- **Water class:** It is composed of 2203 1-s audio clips corresponding to 96 audio recordings of 12 subjects aiming at the water base. The audio recordings were fragmented using the same procedure that was used for the ceramic class.
- **Silence class:** This class does not represent a physical surface as such but is associated with an interruption of the voiding flow. It is composed of 2170 silent audio recordings made while a person was present in the bathroom, with the objective of recording the characteristics of the breathing process when there is an absence of voiding.

3.2 Feature selection

The first step in audio classification is to select the best procedure for characterising each audio sample in the dataset. First, we perform a spectral analysis in the entire frequency band recorded by our specialised microphone (0–192 kHz) to determine where the components that

Table 1 Proportion of audio clips of each class recorded in each bathroom

Bathroom (length-width-height)	Ceramic class (%)	Water class (%)	Silence class (%)
1 (404-175-271 cm)	48.1	23.2	49.4
2 (300-102-255 cm)	14.7	43.1	38.8
3 (230-157-261 cm)	11.3	11.9	11.8
4 (240-150-230 cm)	25.9	21.8	0

provide the most information in the classification process are located. For this purpose, we extract 1000 linear-binned FFT samples for each 1-s audio clip, where the frequency range (0–192 kHz) is divided into 1000 equally spaced intervals, and for each interval, we sum the absolute values of the amplitudes of the components present in each interval, finally obtaining a vector with 1000 values that characterises each audio clip. Then, we perform supervised feature selection and classification using RF and build a model using a Gini impurity-based metric [30]. By using the Gini impurity to measure the quality of our split criterion, we can quantify the weighted impurity of each feature in the tree, indicating its importance. Figure 1 shows the predictive power of each frequency component based on Gini impurities for the ceramic, water and silence audio clips in our dataset. This figure shows that the bins around 1 kHz, 17 kHz and 30 kHz contain the greatest predictive power for the task of distinguishing among the three classes. To develop the ML models, we selected the band from 0 to 22.05 kHz because it is the frequency band captured by the vast majority of commercial microphones ($SR = 44.1$ kHz). This represents a compromise between the model performance and the cost and availability of the microphone being used.

For the study of the 0–22.05 kHz band, we extract a 20-linear-bin FFT. Next, to visualise the degree of separability between the three classes, we apply the dimensionality reduction technique t-distributed stochastic neighbour embedding (t-SNE) [31], which converts similarities between data points into joint probabilities. The results are shown in Fig. 2; they demonstrate a high degree of separability between the three classes.

3.3 Sound classification model

In this subsection, we build three supervised ML algorithms to classify the physical void impact surface in an SU test. We have selected three models for our study:

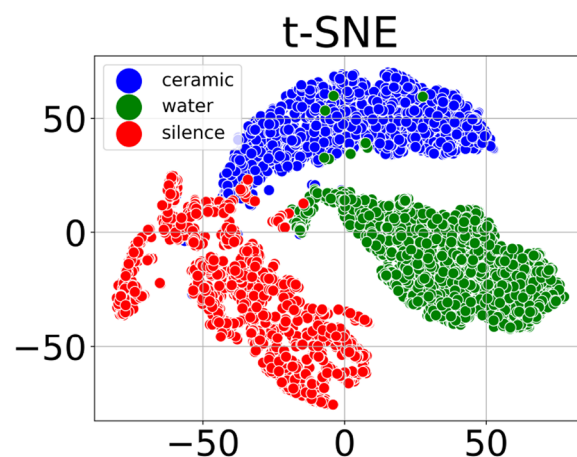


Fig. 2 t-SNE plot that shows that the ceramic (blue), water (green) and silent (red) classes can be distinguished well

an SVM, an RF and a k-nearest neighbours (k-NN) classifier. We applied the stratified k-fold cross-validation method with $k = 10$ to divide our data into training and testing sets for each of the algorithms used. This validation method provides a robust and reliable estimate of a model's performance on unseen data and ensures that each split maintains a class distribution similar to that of the original dataset. These models have been selected because our dataset is too small to apply deep learning techniques. Below, we detail why we chose each model:

- SVM: It is a supervised learning algorithm used mostly for classification purposes. This algorithm is easy to use and will provide the best output, even when it is tested on limited-size training datasets [23]. The only data-dependent step is the choice of the kernel and the corresponding feature space [32]. In our case, we have used the polynomial kernel since it generally performs better in classifying high-

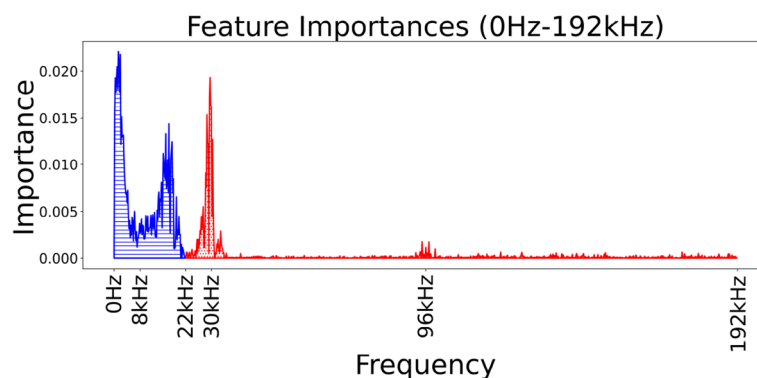


Fig. 1 Predictive power (importance) of each frequency component in the classification task with three classes: ceramic, water and silence. The frequency band selected in our algorithms is shown in blue. The importance is calculated using the Gini impurity with a random forest model

dimensional data when the data are not linearly separable, which is the case for the data in this paper.

- **RF:** This method is used for popular ML tasks related to regression and classification in any domain of interest. RF works by constructing an outsized quantity of decision trees. Random decision forests prevent decision trees from overfitting the training data [23]. For the selection of the number of estimators, a parameter that indicates the number of trees in the forest, we have experimentally tested different values and selected a value of 10 trees.
- **k-NN:** It is one of the simplest and most common classifiers, yet it can compete with the most complex classifiers in the literature [33]. k-NN is based on the idea of clustering data of the same nature. In other words, objects of the same category should be closer in terms of distance [34]. The core of this classifier depends mainly on measuring the distance or similarity between the tested examples and the training examples. To use the classifier, it is necessary to determine the number of neighbours; in our case, it is three.

Figure 3 shows a graphical pipeline diagram of the proposed methodology. Our input data are the SU audio signals. First, the audio signal is segmented into 1-s frames. Next, the FFT is applied to each of the frames to process the data in the frequency domain, and 20 linear bins are extracted. These bins are the input features of the classification algorithms. Finally, the algorithm outputs the classification results: the signal is predicted to be in the ceramic class, water class or silence class.

4 Results and discussion

We next evaluate the three different ML algorithms using three different frequency bands. The first band, 0–22.05 kHz, covers the entire band available for the vast majority of commercial recording devices (SR = 44.1 kHz); this includes devices integrated into smartphones and

smartwatches and dedicated devices. The second one corresponds to 0–8 kHz, which includes only information within the human speech band. Finally, the third one from 8 to 22.05 kHz is selected to evaluate the algorithms for the case in which it is necessary to preserve the users' privacy by eliminating human speech components.

For each of the three bands, we used 20 linear-binned FFT features. We used stratified 10-fold validation to ensure that each fold of the dataset is class-balanced across labels. For each model, we report the following performance metrics: the F1-score, ACC, standard deviation (SD), false positive rate (FPR) and false negative rate (FNR). Figure 4 shows the confusion matrices for each of the three models in the three frequency bands analysed. Table 2 shows the results obtained. This table shows that similar results are obtained for the three models, with the values of the ACC and F1-score ranging from 89.38 to 99.46% for the three models across the three frequency bands. Overall, the RF model presents the best performance results for each frequency band for the task of classifying the physical surface in SU tests. Furthermore, we can safely remove the human speech frequency band and consider the range 8–22.0 kHz, since the RF model maintains a high ACC (93.29%) and F1-score (93.30%). We believe that the removal of human speech could be a requirement for some users who want privacy in their SU test.

These positive results reinforce the decision in this work to consider frequencies below 22.05 kHz, eliminating the need for specialised microphones. This demonstrates that the surface can be classified accurately in SU tests using commercial recording devices. Therefore, it is not necessary to use specialised and expensive recording equipment with sample rates above 44.1 kHz.

4.1 Surface classification in mixed-surface SU audio clips

Next, we need to validate our models for the typical voiding event in which, within the same voiding event, the urine impacts both the water and the ceramic surface.

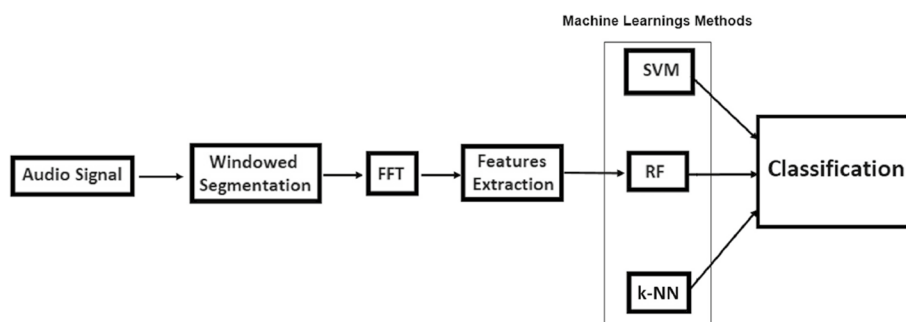


Fig. 3 Diagram showing the pipeline of the proposed methodology

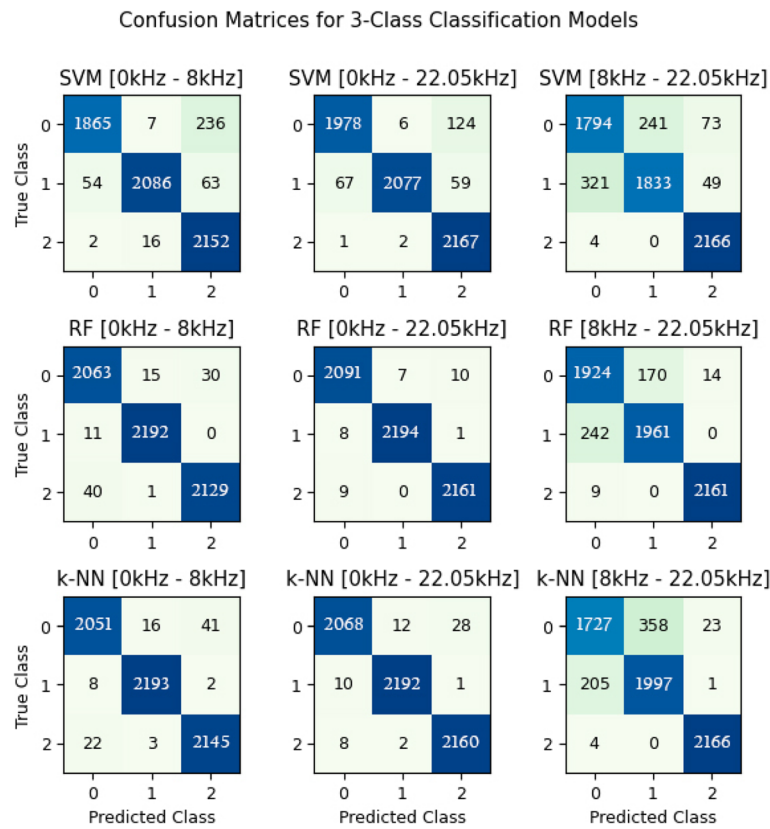


Fig. 4 Confusion matrices for three-class classification models: ceramic class (0), water class (1) and silence class (2)

Table 2 Evaluation of models by frequency range in terms of the classification ACC, F1-score, SD, FPR and FNR

	Frequency range	ACC (%)	F1-score (%)	SD (%)	FPR (%)	FNR (%)
SVM	[0–8 kHz]	94.17	94.18	0.82	2.92	5.89
SVM	[0–22.05 kHz]	96.00	96.01	0.65	2.00	4.01
SVM	[8–22.05 kHz]	89.38	89.27	1.15	5.30	10.63
RF	[0–8 kHz]	98.50	98.50	0.36	0.75	1.51
RF	[0–22.05 kHz]	99.46	99.46	0.32	0.27	0.54
RF	[8–22.05 kHz]	93.29	93.30	0.72	3.35	6.71
k-NN	[0–8 kHz]	98.58	98.63	0.44	0.71	1.44
k-NN	[0–22.05 kHz]	99.05	99.05	0.33	0.47	0.95
k-NN	[8–22.05 kHz]	90.88	90.82	1.36	4.57	9.20

We collected 15 voiding events in two bathrooms corresponding to bathrooms 2 and 3 in Table 1. The audio recordings for these tests were not used in the training phases of our models. The participants were asked to aim the voiding flow at the toilet ceramic and water within the same voiding event. Table 3 summarises the characteristics of the voiding forms performed.

During the tests, there were time intervals, especially at the end of some tests, in which the flow gradually

Table 3 Voiding characteristics

Signal	Physical surface combinations	Repetitions
1	Silence-water-ceramic-water-silence	2
2	Silence-ceramic-water-ceramic-silence	2
3	Silence-water-ceramic-silence	3
4	Silence-ceramic-water-silence	2
5	Silence-water-silence	2
6	Silence-ceramic-silence	4

decreased until it became a dribble. We considered this indeterminate and did not take it into account in the evaluation of the algorithm (see Fig. 5, where the indeterminate time is marked with grey dots). This is because it was impossible for the volunteers who performed the test to determine accurately whether these seconds corresponded to voiding against ceramic or water. It is important to note that this time interval contains a mixture of dribbling against water and ceramic.

These intervals generate some uncertainty in the classification task but become somewhat meaningless if we consider that, according to urologists' criteria, the final seconds of the voiding event do not provide relevant information for screening or diagnosis.

In the 15 audio recordings processed, 700 s were analysed, corresponding to 258, 222 and 220 s of the ceramic, water and silence classes, respectively. To evaluate the automatic classification of the impact surface, we used the RF classifier with the features extracted for the 0–20.05 kHz band. We selected this configuration because it provided the best overall classification results. Additionally, most commercial recording devices allow recording in this band, which facilitates its implementation.

Figures 5, 6 and 7 show the results obtained by the algorithm for three selected voiding events. Red, blue and green represent the silence, ceramic and water classes, respectively, for each 1-s interval. The circles represent the ground truth, while the diamonds represent

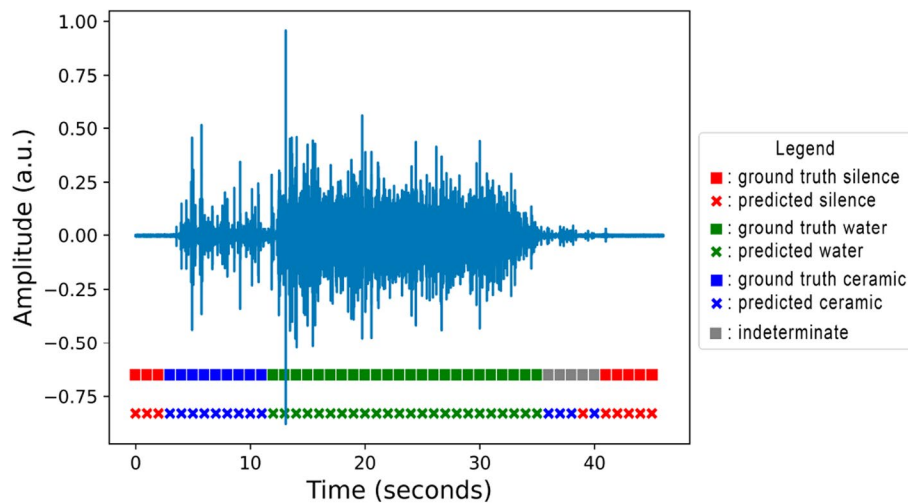


Fig. 5 Results for signal four, repetition one (see Table 3)

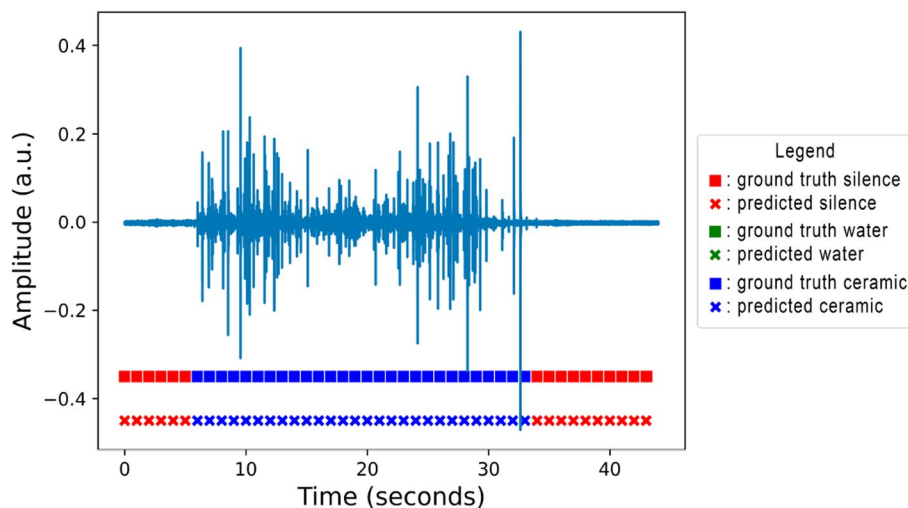


Fig. 6 Results for signal six, repetition two (see Table 3)

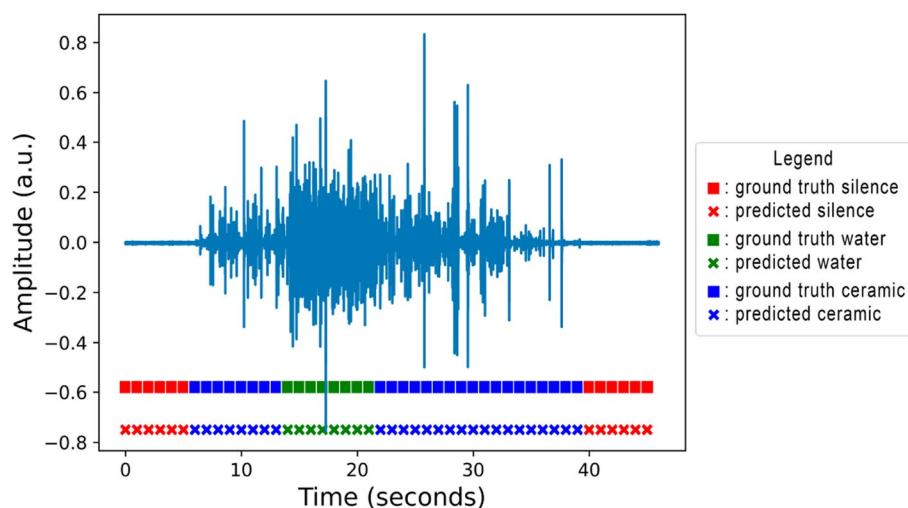


Fig. 7 Results for signal two, repetition one (see Table 3)

the inference made by the RF algorithm. By comparing the ground truth and the output of the RF model, we obtained a classification ACC of 98.17 %.

5 Conclusions

This work addresses the problem of the automatic classification of the physical voiding flow impact surface in SU tests. One of the SU requirements is that the voiding flow must always impact the water in the bowl of the toilet. However, in a real-world scenario, the voiding flow impacts the toilet wall often. This requirement represents a constraint, especially for elderly people and children. If this requirement is not met, the estimation of the flow parameters will be negatively affected.

We built a dataset of 6481 1-s audio clips labelled as silent (no voiding), ceramic (voiding against ceramic) and water (voiding against water) to train three automatic classification models. Three algorithms were trained to automatically evaluate the classification of the surface in three frequency bands within the 0–22.05 kHz commercial band: the SVM, RF and k-NN. The results show that the RF classifier using the FFT-based features in the frequency range of 0–22.05 kHz obtains a classification ACC of 99.46 % for distinguishing among voiding events against ceramic or water and silence (absence of voiding flow). Furthermore, we can safely remove the human speech frequency band and consider the range 8–22.05 kHz, since the RF model maintains a high ACC (93.29%) and F1-score (93.30%).

Next, we collected data from 15 real SU tests performed by three male subjects in three different bathrooms. The subjects were instructed to change the impact surface during the voiding event. We validated the positive inference performance of the model for differentiating among

the three surfaces. With this work, we open the door for new studies that will allow the analysis of the voiding flow and the extraction of the envelope parameters as a function of the surface that the urine impacts. The results will allow SU tests to be performed without the existing limitation of always targeting the water in the toilet.

5.1 Future work

For future work, our goal is to study the estimation of the voiding parameters (flow rate and volume) as a function of the surface that the voiding flow impacts (water or ceramic) and to be able to eliminate the requirement in current SU tests to always aim at the water in the toilet bowl. Additionally, we will analyse the reconstruction of the signal envelope in the time intervals in which the voiding flow impacts a ceramic surface, as if it had impacted water. This will allow us to automatically classify the voiding patterns according to the four existing patterns in the literature, normal, intermittent, fluctuating and plateau, which each represent a set of underlying dysfunctions, regardless of the voiding impact surface.

Abbreviations

LUTS	Lower urinary tract symptoms
BPH	Benign prostatic hyperplasia
UF	Uroflowmetry
SU	Sound uroflowmetry
FFT	Fast Fourier transform
ML	Machine learning
MFCCs	Mel frequency cepstral coefficients
SVM	Support vector machine
RF	Random forest
CNN	Convolutional neural network
RNN	Recurrent neural network
SR	Sampling rate
t-SNE	t-distributed stochastic neighbour embedding
k-NN	K-nearest neighbours

ACC	Accuracy
SD	Standard deviation
FPR	False positive rate
FNR	False negative rate

Authors' contributions

Marcos Lazaro Alvarez, Laura Arjona, Miguel E. Iglesias Martínez and Alfonso Bahillo contributed to the conception and design of the study. Marcos Lazaro Alvarez and Laura Arjona organised the database and developed the software. Marcos Lazaro Alvarez and Miguel E. Iglesias Martínez wrote the first draft of the manuscript. Marcos Lazaro Alvarez, Laura Arjona, Miguel E. Iglesias Martínez and Alfonso Bahillo wrote sections of the manuscript. All authors contributed to manuscript revision and read and approved the submitted version.

Funding

This research was supported by the Spanish Ministry of Science and Innovation under the Peace of Mind project (ref. PID2019-105470RB-C31). Miguel E. Iglesias Martínez's work was supported by the postdoctoral research scholarship 'Ayudas para la recualificación del sistema universitario español 2021-2023. Modalidad: Margarita Salas', UPV, Ministerio de Universidades, Plan de Recuperación, Transformación y Resiliencia, Spain. It was funded by the European Union-Next Generation EU.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare no competing interests.

Received: 12 September 2023 Accepted: 29 January 2024

Published online: 16 February 2024

References

- B. Chughtai, J.C. Forde, D.D.M. Thomas, L. Laor, T. Hossack, H.H. Woo, A.E. Te, S.A. Kaplan, Benign prostatic hyperplasia. *Nat. Rev. Dis. Prim.* **2**(1), 1–15 (2016)
- M.F. Arjona, I.P. Sanz, Hiperplasia benigna de próstata: una afección de elevada prevalencia en el paciente de edad avanzada. *Rev. Esp. Geriátria Gerontol.* **43**(1), 44–51 (2008)
- J.C. Santos, C.E. Smet, Prevalencia de síntomas del tracto urinario inferior de llenado en pacientes varones que acuden a consulta de urología en España. la urgencia urinaria como predictor de calidad de vida. *Actas Urol. Esp.* **40**(10), 621–627 (2016)
- M.R. Sorel, H.J. Reitsma, P.F. Rosier, R.J. Bosch, L.M. de Kort, Uroflowmetry in healthy women: A systematic review. *Neurourol. Urodyn.* **36**(4), 953–959 (2017)
- W. Schäfer, P. Abrams, L. Liao, A. Mattiasson, F. Pesce, A. Spangberg, A.M. Sterling, N.R. Zinner, P.V. Kerrebroeck, Good urodynamic practices: Uroflowmetry, filling cystometry, and pressure-flow studies. *Neurourol. Urodyn. Off. J. Int. Continence Soc.* **21**(3), 261–274 (2002)
- K.L. Kuoch, D. Meyer, D.W. Austin, S.R. Knowles, Classification and differentiation of bladder and bowel related anxieties: A socio-cognitive exploration. *Curr. Psychol.* **40**, 4004–4011 (2021)
- N. Alothmany, H. Mosli, M. Shokouinejad, R. Alkashgari, M. Chiang, J.G. Webster, Critical review of uroflowmetry methods. *J. Med. Biol. Eng.* **38**, 685–696 (2018)
- D.G. Lee, J. Gerber, V. Bhatia, N. Janzen, P.F. Austin, C.J. Koh, S.H. Song, A prospective comparative study of mobile acoustic uroflowmetry and conventional uroflowmetry. *Int. Neurourol. J.* **25**(4), 355 (2021)
- Y.J. Lee, M.M. Kim, S.H. Song, S. Lee, A novel mobile acoustic uroflowmetry: Comparison with contemporary uroflowmetry. *Int. Neurourol. J.* **25**(2), 150 (2021)
- P. Hurtík, M. Burda, J. Krhut, P. Zvara, L. Lunáček, Automatic diagnosis of voiding dysfunction from sound signal. 2015 IEEE symposium series on computational intelligence. (Cape Town, 2015), p. 1331–1336. <https://doi.org/10.1109/SSCI.2015.190>
- E.J. Aslim, B. Balamurali, Y.S.L. Ng, T.L.C. Kuo, K.S. Lim, J.S. Chen, J.M. Chen, L.G. Ng, Pilot study for the comparison of machine-learning augmented audio-uroflowmetry with standard uroflowmetry in healthy men. *BMJ. Innovations.* **6**, bmjinnov-2019 (2020), <https://doi.org/10.1136/bmjinnov-2019-000382>
- C.V. Comiter, E. Belotserkovsky, in *Neurourology and Urodynamics*, vol. 38. A novel mobile uroflowmetry application for assessing lower urinary tract symptoms (Philadelphia, 2019), pp. 556–557, www.ics.org/2018/abstract/175
- L. Arjona, L. E. Diez, A. Bahillo, A. Arruza-Echevarría, UroSound: A smartwatch-based platform to perform non-intrusive sound-based uroflowmetry. *IEEE J. Biomed. Health Infor.* **27**(5), 2166–2177 (2023), <https://doi.org/10.1109/JBHI.2022.3140590>
- G. Narayanswamy, L. Arjona, L. E. Diez, A. Bahillo, S. Patel, Automatic classification of audio uroflowmetry with a smartwatch. 2022 44th annual international conference of the IEEE Engineering in Medicine & Biology Society (EMBC). (Glasgow, United Kingdom, 2022) p. 4325–4329, <https://doi.org/10.1109/EMBC48229.2022.9871611>
- L. Bobrowski, T. Łukaszyk, Feature selection based on relaxed linear separability. *Biocybernetics Biomed. Eng.* **29**(2), 43–59 (2009)
- K. Moreland, E. Angel, The FFT on a GPU. In *Proceedings of the ACM SIGGRAPH/EUROGRAPHICS conference on graphics hardware (HWWW '03)*. (Eurographics Association, Goslar, DEU, 2003), p1 12–119
- Q. Zhou, J. Shan, W. Ding, C. Wang, S. Yuan, F. Sun, H. Li, B. Fang, Cough recognition based on mel-spectrogram and convolutional neural network. *Front. Robot. AI* **8**, 580080 (2021)
- N. Sato, Y. Obuchi, Emotion recognition using mel-frequency cepstral coefficients. *Inf. Media Technol.* **2**(3), 835–848 (2007)
- Wang, W. (Ed.), Machine audition: principles, algorithms and systems: principles, algorithms and systems. (IGI Global, 2010), [https://books.google.es/books?hl=es&lr=&id=WWWAQAQBAJ&oi=fnd&pg=PR1&dq=Principles,+Algorithms+and+Systems+\(IGI+Global,+2010\)&ots=TJfQaDCnjg&sig=hWE2xEdXHlGf53nmIPYuo2Fo#v=onepage&q=Principles%2C%20Algorithms%20and%20Systems%20\(Igi%20Global%2C%202010\)&f=false](https://books.google.es/books?hl=es&lr=&id=WWWAQAQBAJ&oi=fnd&pg=PR1&dq=Principles,+Algorithms+and+Systems+(IGI+Global,+2010)&ots=TJfQaDCnjg&sig=hWE2xEdXHlGf53nmIPYuo2Fo#v=onepage&q=Principles%2C%20Algorithms%20and%20Systems%20(Igi%20Global%2C%202010)&f=false)
- J. Ye, T. Kobayashi, N. Toyama, H. Tsuda, M. Murakawa, Acoustic scene classification using efficient summary statistics and multiple spectro-temporal descriptor fusion. *Appl. Sci.* **8**(8), 1363 (2018)
- 2004 IEEE International conference on acoustics. Speech, and signal processing. 2004 IEEE International Conference on acoustics, speech, and signal processing. (Montreal, 2004), <https://doi.org/10.1109/ICASSP2004.1326738>
- L.L. Wyse, Audio Spectrogram Representations for Processing with Convolutional Neural Networks. *Proceedings of the First International Workshop on Deep Learning and Music joint with IJCNN*. Anchorage, US. **1**(1), 37–41 (2017), <https://doi.org/10.48550/arXiv.1706.09559>
- B. Vimal, M. Surya, Darshan, V. S. Sridhar, A. Ashok, MFCC Based audio classification using machine learning. 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). (Kharagpur, 2021), p. 1–4, <https://doi.org/10.1109/ICCCNT51525.2021.9579881>
- P. Dhanalakshmi, S. Palanivel, V. Ramalingam, Classification of audio signals using SVM and RBFNN. *Exp. Syst. Appl.* **36**(3), 6069–6075 (2009)
- F. Rong, Audio classification method based on machine learning. 2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). (Changsha, 2016) p. 81–84, <https://doi.org/10.1109/ICITBS.2016.98>
- M. R. Ansari, S. A. Tumpa, J. A. F. Raya, M. N. Murshed, comparison between support vector machine and random forest for audio classification. 2021 International conference on Electronics, Communications and Information Technology (ICECIT), (Khulna, 2021), p. 1–4, <https://doi.org/10.1109/ICECIT54077.2021.9641152>
- A. Subasi, M. Radhwan, R. Kurdi, K. Khateeb, IoT based mobile healthcare system for human activity recognition, 2018 15th Learning and Technology Conference (L&T), (Jeddah, 2018), p. 29–34, <https://doi.org/10.1109/LT.2018.8368507>

28. L. Nanni, G. Maguolo, S. Brahnem, M. Paci, An ensemble of convolutional neural networks for audio classification. *Appl. Sci.* **11**(13), 5796 (2021)
29. M. Scarpiniti, D. Comminiello, A. Uncini, Y.-C. Lee, Deep recurrent neural networks for audio classification in construction sites. 2020 28th European Signal Processing Conference (EUSIPCO). (Amsterdam, 2021), p. 810–814, <https://doi.org/10.23919/Eusipco47968.2020.9287802>
30. Y. Iravantchi, K. Ahuja, M. Goel, C. Harrison, A. Sample, privacyMic: Utilizing inaudible frequencies for privacy preserving daily activity recognition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21) Article 198*. (Association for Computing Machinery, New York, 2021), p. 1–13. <https://doi.org/10.1145/3411764.3445169>
31. L. van der Maaten, G. Hinton, Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.* **9**(2579–2605), 9 (2008)
32. A. Mammone, M. Turchi, N. Cristianini, Support vector machines. *Wiley Interdiscip. Rev. Comput. Stat.* **1**(3), 283–289 (2009)
33. H.A. Abu Alfeilat, A.B. Hassanat, O. Lasassmeh, A.S. Tarawneh, M.B. Alhasanat, H.S. Eyal Salman, V.S. Prasath, Effects of distance measure choice on k-nearest neighbor classifier performance: A review. *Big Data* **7**(4), 221–248 (2019)
34. C.H. Chen, W.T. Huang, T.H. Tan, C.C. Chang, Y.J. Chang, Using k-nearest neighbor classification to diagnose abnormal lung sounds. *Sensors* **15**(6), 13132–13158 (2015)

Publisher's Note

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