Automatic Classification of Audio Uroflowmetry with a Smartwatch

Girish Narayanswamy[1], Laura Arjona[2], Luis E. Díez[2], Alfonso Bahillo[2], Shwetak Patel[3]

Abstract— **Prior work has shown the classification of voiding dysfunctions from uroflowmeter data using machine learning. We present the use of smartwatch audio, collected through the UroSound platform, in order to automatically classify voiding signals as normal or abnormal, using classical machine learning techniques. We train several classification models using classical machine learning and report a maximal test accuracy of 86.16% using an ensemble method classifier.**

Clinical relevance— **This classification task has the potential to be part of an essential toolkit for urology telemedicine. It is especially useful in areas that lack proper medical infrastructure but still host ubiquitous audio capture devices such as smartphones and smartwatches.**

I. INTRODUCTION

The rapid ageing of society, combined with the expected reduction in the working population, threatens the sustainability of health systems. Furthermore, the current COVID19 pandemic continues to aggravate this situation, especially for the elderly, and for those in less developed / rural regions that lack proper medical infrastructure. Consequently, new strategies are needed to transition from in-person and reactive health care systems to remote and proactive systems that focus on patient welfare through continuous and non-intrusive care.

A problem frequently associated with ageing is voiding dysfunction. This highly prevalent issue has a major impact on the quality of life for a large number of individuals (more than 60% of men over 60 years of age) [1]. Remote assessment of people with voiding dysfunction may allow for the capture of unique voiding features, thus facilitating the prompt diagnosis of a disease.

Recent works have demonstrated the feasibility of using mobile devices, such as smartphones [2], [3], [4], [5], [6] and smartwatches [7], in a home environment, to characterize urinary flow patterns by capturing the sound generated when the urine stream hits the water in a toilet bowl. This test is known as audio uroflowmetry. The scientific literature shows that the use of machine learning techniques to estimate flow parameters from urinary voiding sounds is viable, with improved accuracy anticipated as the algorithm is continuously refined with additional training samples [8]. Therefore, the quantity and quality of the audio samples are of paramount importance in order to train and validate these algorithms. Unfortunately, obtaining such samples is non-trivial and most existing works train and validate their methods from sounds obtained from a standard uroflowmeter device, instead of a water-based toilet bowl, whose acoustic characteristics are different [9], [10]. Additionally, information provided by the urine flow parameters is not always very significant, as there is often great variability among patients.

It is usually more meaningful to classify whether the flow envelope corresponds to a normal or abnormal flow, or to evaluate the envelope evolution over time [11]. According to [11], the association between flow shapes and underlying pathologies can be thoroughly researched, and abnormal flow shapes are carefully associated with underlying pathologies. This classification task, which can be performed by an experienced urologist listening to the urinary flow audio, can be automated thanks to the use of machine learning techniques. This functionality has the potential to be part of an essential toolkit for urology telemedicine, especially in under-served areas and in communities that lack proper medical infrastructure. The advantages of using a smartwatch compared to a smartphone to perform audio uroflowmetry are clear: as a wearable device, smartwatches can be worn continuously, facilitating the capture of voiding behavioural data in real time, in natural environments, with multiple repeated measurements.

The primary objective of this work is to evaluate the accuracy of machine learning techniques to support the automatic identification of pathologies associated with the urinary tract and distinguish between normal (healthy) and abnormal (unhealthy) voiding signals. These algorithms are trained and validated with a dataset obtained from a cohort of patients and are run in a cloud-based backend.

II. RELATED WORK

The state of the art in machine learning applied to uroflowmetry can be divided into two main categories. The first includes works that focus on accurately estimating the flow rate and voided volume in standard units (ml/s and ml, respectively) from sound-based uroflowmetries [12], [3], [8]. The results demonstrate that it is a rather complex problem, as an accurate calibration is required to extract the data in standard units from audio signals.

The second category includes works that use machine learning to automatically classify the voiding shape as normal or abnormal, both from standard uroflowmetry test data [10], [13] and from sound-based uroflowmetry [14].

[1]Girish Narayanswamy is with the Department of Electrical and Computer Engineering, University of Washington, Seattle WA 98195, USA girishvn@uw.edu

[2]Laura Arjona, Luis E. Díez, and Alfonso Bahillo are with the Faculty of Engineering, University of Deusto, Bilbao, 48007, Spain {laura.arjona, luis.enrique.diez, alfonso.bahillo}@deusto.es

[3]Shwetak Patel is with the Paul G. Allen School of Computer Science and Engineering and the Department of Electrical and Computer Engineering, University of Washington, Seattle WA 98195, USA shwetak@cs.washington.edu

To the best of our knowledge, there is no prior work in classifying the voiding shape, extracted from sound-based uroflowmetry tests, performed with a smartwatch. These tests consist of recording the impact of the voiding event with the toilet water through the use of a smartwatch.

III. Methods

This section presents the hardware specifications of the platform used to record voiding events and then introduces the clinical study performed to collect audio uroflowmetry tests from 14 volunteer patients. We then discuss the dataset, data augmentation, feature extraction, and machine learning model selection and training. Fig. 1 presents an overview of the procedure, which is explained in detail in this section.

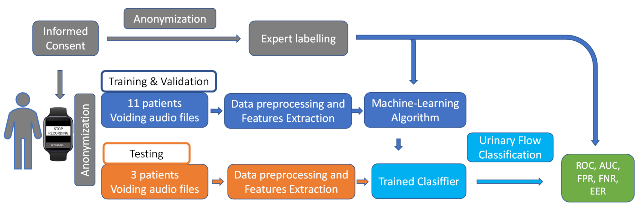

Fig. 1. Statistical analysis procedure: from the smartwatch voiding event audio recording, to the machine learning model evaluation.

A. Hardware Specifications

To collect the audio uroflowmetry tests, we used the UroSound application (open-sourced for Android devices) [15], the first application to perform audio uroflowmetry tests with a smartwatch by recording the sound produced when the voiding flow impacts the toilet water. The audio uroflowmetry data collected with this app achieves a good correlation between acoustic and standard uroflowmetry with respect to the voiding shape [7]. In particular, we use the UroSound app installed on an Oppo Smartwatch (a commercially available smartwatch). UroSound is configured to record audio with a sampling rate (fs) of 16 kHz, a bit depth of 16 bits/sample, and in an uncompressed WAV format.

B. Clinical Data Collection

We devised and conducted a clinical study with 14 volunteer patients from two pelvic floor health clinics located in Spain. The experimental procedures described below conform to the provisions of the Declaration of Helsinki (as revised in Edinburgh 2000). We have published a public GitHub repository containing the dataset, where each audio is named with the time and date of the recording [15]. The collection procedure is as follows:

1) The patient is informed of the project, the conditions of the tests, and then signs an informed consent form.
2) The patient is asked to take a standard uroflowmetry test in the clinic, using a uroflowmeter.
3) The patient receives a smartwatch from the physician that they wear and use for three consecutive days to record the sound of their urination.
4) At the end of this period, the patient returns the smartwatch to the clinic.
5) Personnel from the clinic send all the audios to a central server using the same smartwatch app. The audios are anonymously associated with a patient through an identification (ID) value.
6) All audios are deleted from the smartwatch. The smartwatch is then disinfected, fully-charged, and ready for the next patient.

C. Dataset

The dataset is comprised of 153 voiding audios, from Oppo smartwatches, across 14 study participants. We performed expert labelling according to the urologist author of [7]. The labels were assigned based on audio playback and visual analysis of the extracted envelopes, as explained in Section II.F. Across all audios, 47 are labeled as **Abnormal** flows, while the remaining 106 are labeled as **Normal** flows. A normal flow curve is smooth without any rapid changes in amplitude. As the shape of the flow curve is determined by the kinetics of the contractions of the detrusor, a smooth muscle, it does not, in general, show rapid variations [16]. An abnormal voiding event is visualised as a non-bell shaped flow-audio signal envelope. Abnormal flows account for approximately 30% of all recordings in the dataset. This information is outlined in Table I, where **0** represents **Normal** flows, and **1** represents **Abnormal** flows. This table also presents the distribution of trials (number of audio recordings) for each of the patients, as well as the number of trials assigned to each one of the two labels. For each patient the number of trials was highly dependant on their engagement in the study and ranged from 3 to 16.

TABLE I
Dataset Audios by Participant

User	Trials	# 0	# 1	User	Trials	# 0	# 1
A2	12	9	3	B5	12	7	5
A3	14	5	9	B7	14	9	5
A4	15	8	7	B8	7	3	4
A5	9	8	1	B9	8	4	4
A6	12	12	0	B10	3	2	1
A7	15	11	4	B11	9	9	0
B3	16	12	4	B12	7	7	0

This table shows that most patients present at least one trial for each of the two label types. This demonstrates the variability among patients in terms of their voiding parameters as mentioned in the Introduction section, and highlights the relevance of performing uroflowmetry tests more than once. UroSound facilitates this process by using a smartwatch as opposed to a standard uroflowmeter device.

D. Data Augmentation

Due to the relatively small size of the dataset, audio data augmentation was used in order to increase the number of samples for both training and testing. Each audio file was augmented to produce 3 additional audio samples. This amounted to a total of 612 voiding audio samples. These augmentations are outlined in Table II.

TABLE II
AUGMENTATIONS FOR VOIDING AUDIOS

Augmentation Type	Description
N/A (Original)	Original audio with no augmentation
Quite	Original audio with gain of -5dB
Loud	Original audio with gain of 5dB
Gaussian Noise	Noise added to original audio with SNR of 5dB

E. Data Preprocessing

The dataset audios are fed through a preprocessing pipeline that extracts the envelope of the voiding signal. This pipeline is broken down into the following stages.

1) **Raw Audio:** The raw audio is read in along with sampling rate fs.
2) **Lowpass Filter:** A lowpass filter, with cutoff frequency of $\frac{3fs}{10}$ is applied to remove high frequency noise.
3) **Hampel Filter:** A Hampel filter is applied to remove and replace outliers in a 256 sample window using the median.
4) **Envelope Detection:** The voiding signal envelope is extracted through the use of a Hilbert filter with a window of 128 samples.
5) **Moving Median Smoothing:** A sliding moving-median window, of fs samples, is applied to smooth-out envelope noise.

This pipeline, along with example voiding signals and corresponding envelopes are illustrated in Fig. 2.

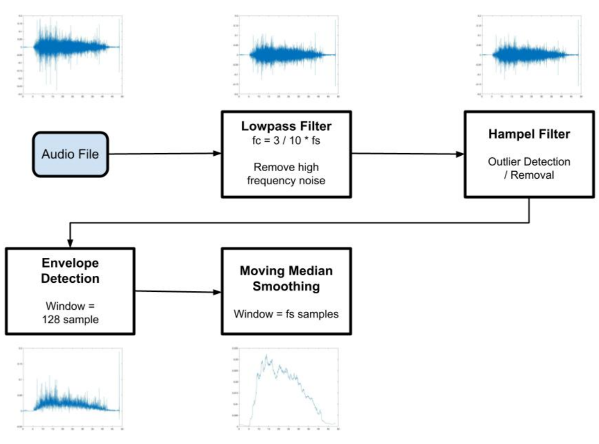

Fig. 2. Audio Preprocessing and Envelope Extraction.

F. Feature Definitions and Feature Extraction

Once the audio signals are preprocessed, a number of features are derived from the voiding signal envelope in order to train a classical machine learning model. Fig. 3 illustrates many of the defined features extracted from the audio envelope, while Fig. 4 shows an example of extracted features from an actual voiding event audio sample.

The features described in Table III are automatically derived from the voiding signal envelopes as follows:

TABLE III
UROFLOWMETRY CLASSIFICATION FEATURES

Feature	Variable	Description
Voiding Time	T	Start to end time interval of voiding
Time to Maximal Flow	T_{max}	Time interval from voiding start to time of maximal flow
Maximal Flow Rate	Q_{max}	Maximal value of processed flow envelope
Average Flow Rate	Q_{avg}	Average value of processed flow envelope
Interruptions	$Ints$	Number of times urine flow falls to or below a value of $max(B_{noise}, 20\%Q_{max})$
Fluctuations	$Flucts$	Number of times urine flow peaks with a prominence of at least 20% of Q_{max}
Background Noise	B_{noise}	Minimum of averaged first and last seconds of envelope signal

1) **Background Noise** (B_{noise}): Let x_{avg1} represent the average value of the first second of the envelope signal. Let x_{avg2} represent the average value of the last second of the envelope signal. $B_{noise} = min(x_{avg1}, x_{avg2})$.
2) **Maximal Flow Rate** (Q_{max}): The highest value reached by the envelope signal, while ignoring the first and last second of the signal (these are ignored due to mic edge effects).
3) **Voiding time** (T): Let T_{start} be the time at which the envelope signal exceeds B_{noise} before reaching a value of at least 20% of Q_{max} for the first time. Let T_{end} be the time at which the envelope signal falls to or below B_{noise} after passing the final value of 20% of Q_{max}. $T = T_{end} - T_{start}$.
4) **Time to Maximal Flow** (T_{max}): The time interval from T_{start} to the first time occurrence of maximal flow.
5) **Average Flow Rate** (Q_{avg}): The sum of all envelope values, from T_{start} to T_{end} divided by T.
6) **Fluctuations** ($Flucts$): The number of times urine flow peaks with a prominence of at least 20% of Q_{max}.
7) **Interruptions** ($Ints$): The number of times urine flow drops to or below a value of B_{noise} or 20% of Q_{max}.

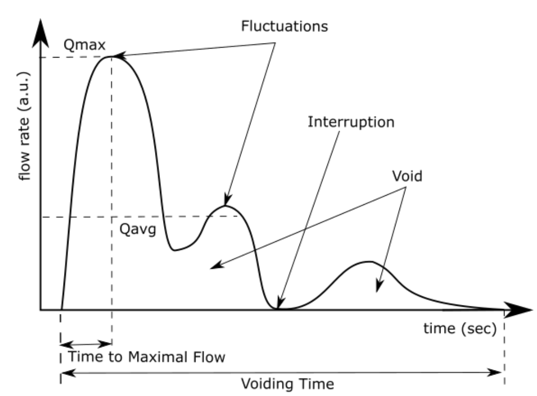

Fig. 3. Feature definitions based on the voiding event audio envelope. Y-axis represents the flow rate in arbitrary units (a.u.).

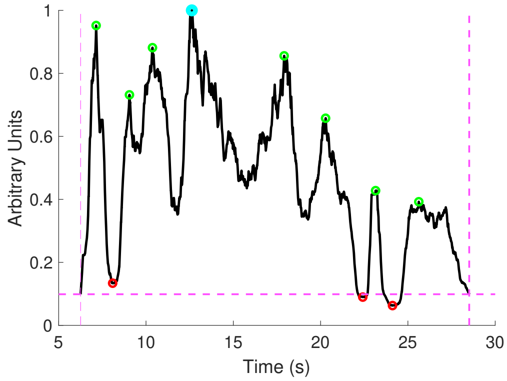

Fig. 4. Example of extracted envelope and features from an audio file from patient A7 (November 11th 2021 at 2:09am). This envelope corresponds to an Abnormal flow (label 1). Blue circle denotes Q_{max}. Green circles denote fluctuations. Red circles denote interruptions. Horizontal dotted line denotes B_{noise}. Vertical dotted lines denote T_{start} and T_{end}.

These extracted features are based off those used by prior work [13] in an algorithmic uroflowmetry classification approach. It is important to note that the features described by this prior work are derived from data produced from a standard uroflowmeter, a device used to precisely measure urine flow. This work, on the other hand, derives similar features from smartwatch audio. For this reason flow parameters such as Q_{max} are given in arbitrary units as opposed to ml/s. Additionally, the audio signal level is dependant on a number of factors including, but not limited to, room acoustics, background noise, and proximity to signal source. For these reasons, we include the Q_{avg} and B_{noise} features to help the model normalize across the uncontrollable environmental factors.

G. Model Selection

Prior work [13] utilizes regression forest models to obtain high prediction accuracy for diagnosis of abnormal voiding from uroflowmeter data. We choose a number of similar ensemble based models. We train three classification models from classical machine learning: an ensemble learning model, and two random forest models. These models were selected experimentally, as they presented high overall classification accuracy. The performance of these models is compared and contrasted in a later section.

- **Ensemble Method:** An ensemble classification method as described by the LogitBoost algorithm [17].
- **Random Forest 1:** A group of 250 bagged classification trees, using 2 features, at random, for decision splits, as described by Breiman's random forest algorithm [18].
- **Random Forest 2:** A group of 250 bagged classification trees, using all 7 features for decision splits.

H. Model Training

Of the 14 participants 11 are used for training, while 3 are held out and comprise the test set. Both testing and training sets are randomly constructed to maintain approximately 30% abnormal flow audios, to reflect the split of data in the full dataset. Data from participant B10 is always used to train due to the relatively low number of associated audios. Additionally, data from participant A3 is also always used to train due to the relatively high proportion of abnormal flow audios. We then perform 3-fold cross-validation with training data comprised from 11 participants, and testing data comprised from the remaining 3 patients.

IV. RESULTS AND DISCUSSION

This section presents the results of training and evaluating the three machine learning models presented in the previous section. The receiver operating characteristic curve (ROC) along with the area under the curve (AUC) are presented in Fig. 5. The ROC curve plots the true positive rate (TPR), that provides information about what proportion of the class 1 got correctly classified, and the false positive rate (FPR), that provides information of what proportion of the class 0 got incorrectly classified. This figure shows that the ensemble method achieves the highest AUC with a value of 0.89. This means that this model has the highest measure of separability: the best model prediction of Normal (or 0) shapes as Normal, and Abnormal shapes (or 1) as Abnormal.

The equal error rate (EER) along with false positive rate (FPR) and false negative rate (FNR) curves are shown in Fig. 6. EER represents the best threshold to choose as it is the point where FPR and FNR are equal. A lower EER is considered better. The three models present an EER value of close to 0.2. The exact values are shown in Table IV. This table presents the evaluation results by model.

TABLE IV
RESULTS BY MODEL. BOLD REPRESENTS THE BEST RESULT

Model	FPR	TPR	EER	AUC	ACC
Ensemble Method	6.67%	67.86%	**21.21**%	**0.8919**	**86.16**%
Regression Forest	**3.42**%	55.56%	22.24%	0.8754	85.62%
Random Forest	9.12%	**68.10**%	23.12%	0.8476	84.03%

As shown in Table IV and Fig. 5, the ensemble method model ($AUC = 0.89$) slightly outperforms both tree ensemble models. Fig. 6 further demonstrates that the ensemble model better balances specificity (1-FPR) and sensitivity (TPR) to produce a lower equal error rate (EER) than either of the two random forest models. This ensemble model achieves a high test accuracy ($ACC = 86.16\%$) on voiding data from patients not present in the training set.

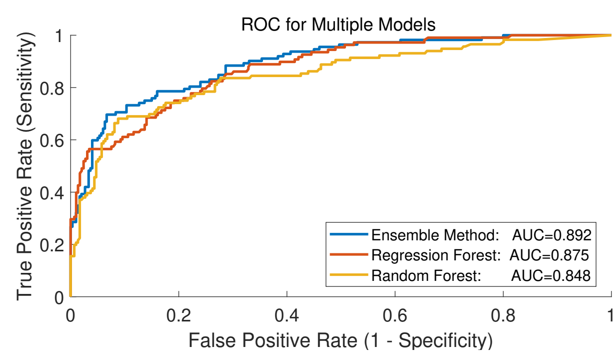

Fig. 5. Model evaluation in terms of ROC and AUC.

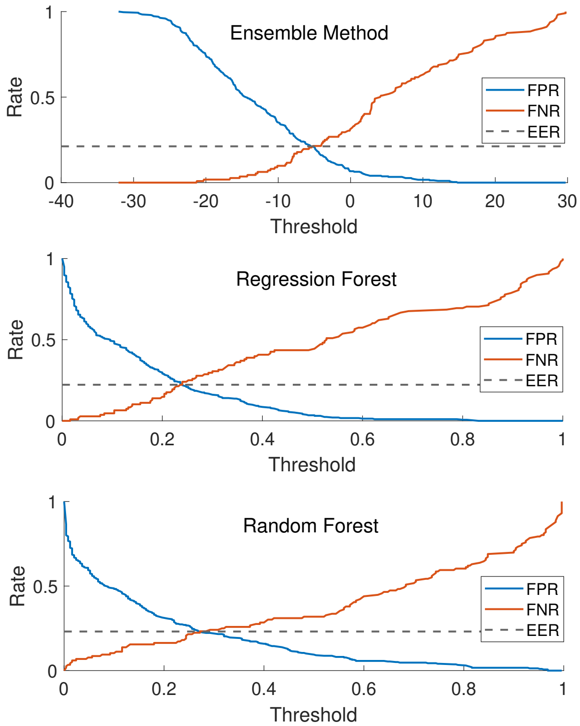

Fig. 6. Models evaluation in terms of FPR, FNR, and EER.

V. CONCLUSIONS AND FUTURE WORK

In this work we present the use of classical machine learning to classify normal and abnormal voiding flows from smartwatch audio. Using 3-fold cross validation we achieve a 21.21% equal error rate and an 86.16% classification accuracy.

The current state of the art in the field of audio uroflowetry is promising, and there is ongoing work to improve its reliability in real-world conditions. However, further development of this technology and standardization are required for both the recording devices and the flow estimation algorithms.

We are currently working on extending our clinical study to collect significantly more audio uroflowmetry tests. We hope to explore deep-learning techniques on our dataset and compare and contrast these results with those of the traditional machine learning methods. We believe that a deep-learning approach may be able to better generalize over environmental noise, and models such as convolutional neural networks may be able better diagnose abnormal voiding flows.

ACKNOWLEDGMENT

This research was supported by the Spanish Ministry of Science and Innovation under the PeaceOfMind project (ref. PID2019-105470RB-C31). A special thank you to all the volunteers / patients that participated in our clinical study.

REFERENCES

[1] J. Cambronero Santos and C. Errando 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 Urológicas Españolas*, vol. 40, no. 10, pp. 621–627, Dec. 2016, publisher: Elsevier.

[2] D. Simon, E. Margolis, H. Rana, M. Blizzard, A. Jung, and A. Lee, “Comparing a Mobile Acoustic Uroflowmetry Device to Standard Uroflowmetry, a Validation Study,” in *ics.org*, Nov. 2020, publisher: International Continence Society. [Online]. Available: https://www.ics.org/2020/abstract/183

[3] E. J. Aslim, B. B. T, Y. S. L. Ng, T. L. C. Kuo, K. S. Lim, J. S. Chen, J.-M. Chen, and L. G. Ng, “Pilot study for the comparison of machine-learning augmented audio-uroflowmetry with standard uroflowmetry in healthy men,” *BMJ Innovations*, pp. bmjinnov–2019–000 382, Jul. 2020, publisher: BMJ Specialist Journals Section: Artificial intelligence. [Online]. Available: https://innovations.bmj.com/content/early/2020/07/14/bmjinnov-2019-000382

[4] Y. J. Lee, M. M. Kim, S. H. Song, and S. Lee, “A Novel Mobile Acoustic Uroflowmetry: Comparison With Contemporary Uroflowmetry,” *Journal of the Korean Continence Society*, Jan. 2021, publisher: Korean Continence Society. [Online]. Available: http://www.einj.org/journal/view.php?doi=10.5213/inj.2040250.125

[5] M. Gärtner, J. Krhut, P. Hurtik, M. Burda, K. Zvarova, and P. Zvara, “Evaluation of Voiding Parameters in Healthy Women Using Sound Analysis: Sonouroflowmetry in Women,” *LUTS: Lower Urinary Tract Symptoms*, vol. 10, no. 1, pp. 12–16, Jun. 2016. [Online]. Available: http://doi.wiley.com/10.1111/luts.12134

[6] C. V. Comiter and E. Belotserkovsky, “A novel mobile uroflowmetry application for assessing low urinary tract symptoms,” in *Neurourology and Urodynamics*, Aug. 2018, publisher: International Continence Society. [Online]. Available: https://www.ics.org/2018/abstract/175

[7] L. Arjona, L. Enrique Diez, A. Bahillo Martinez, and A. Arruza Echevarria, “Urosound: A smartwatch-based platform to perform non-intrusive sound-based uroflowmetry,” *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2022.

[8] E. Aslim, B. T B, Y. Ng, T. Kuo, J. Chen, J. Chen, and L. Ng, “A novel machine-learning augmented audio-uroflowmetry – comparison with standard uroflowmetry,” in *European Urology Supplements*, vol. 18, Mar. 2019, p. e1773. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1569905619312849

[9] J. S. de Haan, “Standardised, automatic description of uroflowmetry curve shapes,” February 2020. [Online]. Available: http://essay.utwente.nl/80668/

[10] Song Sang Hun, choo minsoo, Young Ryu Ho, and Lee Sangchul, “Interpretation of uroflowmetrogram: application with artificial intelligence,” *The Journal of Urology*, vol. 203, Apr. 2020.

[11] R. Li, Q. Zhu, M. Nibouche, and A. Gammie, “Urine flow rate curve shapes and their descriptors,” *Neurourology and Urodynamics*, vol. 37, no. 8, pp. 2938–2944, 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/nau.23743

[12] B. B. T, E. J. Aslim, Y. S. L. Ng, T. Li, C. Kuo, J. S. Chen, D. Herremans, L. G. Ng, and J.-M. Chen, “Acoustic prediction of flowrate: varying liquid jet stream onto a free surface,” 2020.

[13] S. Baas, “A machine learning approach to the automatic classification of female uroflowmetry measurements,” June 2016. [Online]. Available: http://essay.utwente.nl/70570/

[14] E. Aslim, B. Balamurali, Y. Ng, T. Kuo, J. Chen, J. Chen, and L. Ng, “Pt311 - can machine-learning (ml) augmented audio-uroflowmetry distinguish between normal and abnormal flows from voiding sounds?” *European Urology Open Science*, vol. 19, p. e2141, 2020, eAU20 Virtual Congress and Theme Week. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666168320340490

[15] 2021. [Online]. Available: https://github.com/DeustoTech/UroSound

[16] W. Schäfer, P. Abrams, L. Liao, A. Mattiasson, F. Pesce, A. Spangberg, A. M. Sterling, N. R. Zinner, P. van Kerrebroeck, and International Continence Society, “Good urodynamic practices: uroflowmetry, filling cystometry, and pressure-flow studies,” *Neurourology and Urodynamics*, vol. 21, no. 3, pp. 261–274, 2002.

[17] J. Friedman, T. Hastie, and R. Tibshirani, “Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors),” *The Annals of Statistics*, vol. 28, no. 2, pp. 337 – 407, 2000. [Online]. Available: https://doi.org/10.1214/aos/1016218223

[18] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct 2001. [Online]. Available: https://doi.org/10.1023/A:1010933404324