Urinary Flow Estimation through Sound–based Uroflowmetry and Machine Learning

Doctoral Thesis by

Marcos Lazaro Alvarez Arteaga

within the doctoral programme
Engineering for the Information Society and Sustainable Development

Supervised by:
Dr. Alfonso Bahillo
Dr. Laura Arjona

Faculty of Engineering
University of Deusto

A thesis submitted for the degree of
Doctor of Engineering

Bilbao, September 2025

Urinary Flow Estimation through Sound–based Uroflowmetry and Machine Learning

Doctoral Thesis by

Marcos Lazaro Alvarez Arteaga

Supervised by:
Dr. Alfonso Bahillo
Dr. Laura Arjona

ALVAREZ ARTEAGA, MARCOS-LAZARO (FIRMA)

Firmado digitalmente por ALVAREZ ARTEAGA, MARCOS-LAZARO (FIRMA)
Nombre de reconocimiento (DN): c=ES, serialNumber=60470363M, sn=ALVAREZ, givenName=MARCOS-LAZARO, cn=ALVAREZ ARTEAGA, MARCOS-LAZARO (FIRMA)
Fecha: 2025.09.12 17:23:06 +02'00'

BAHILLO MARTINEZ ALFONSO - 71926682R

Digitally signed by BAHILLO MARTINEZ ALFONSO - 71926682R
Date: 2025.09.15 09:15:42 +02'00'

Firmado por ARJONA AGUILERA, LAURA (FIRMA) el día 14/09/2025 con un certificado emitido por AC DNIE 004

Doctoral Candidate	Thesis Supervisor	Thesis Supervisor

Faculty of Engineering
University of Deusto

A thesis submitted for the degree of
Doctor of Engineering

Bilbao, September 2025

Dedication

Esta tesis doctoral está especialmente dedicada, A MIS PADRES!!!, mis queridos padres, Mariluz de la Caridad y Pablo Nivaldo, su amor inquebrantable y fortaleza han guiado mi vida y mis estudios. A mi esposa Marely y a mi hijo Fabio Lazaro, cuyo amor, paciencia y alegría dan sentido y fuerza a todo lo que hago. A mi hermano, Dario, cuya presencia y apoyo siempre me han recordado que nunca camino solo. Y en cariñosa memoria de mis abuelos, Pablo e Isabel María, cuya bondad y valores continúan inspirándome, incluso en su ausencia. A toda mi familia, tías, primos y a todos quienes han contribuido a mi formación desde mis primeros días y me han apoyado a lo largo de este camino. Gracias por ser parte de esta historia.

Declaration

I hereby declare that, with the exception of explicit references to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification at this or any other university. Except as mentioned in the text and Acknowledgements, this dissertation is entirely my own work.

Marcos Lazaro Alvarez Arteaga
September 2025

Acknowledgments

La culminación de esta tesis doctoral no habría sido posible sin el apoyo, el aliento y la guía invaluables de numerosas personas e instituciones.

En primer lugar, deseo expresar mi profunda gratitud a mis queridos padres, Mariluz de la Caridad y Pablo Nivaldo, por su amor incondicional, su confianza y su fe en mis capacidades, que han sido el pilar y ejemplo fundamental de mi formación personal y profesional.

A mi esposa Marely y a mi hijo Fabio Lazaro, cuyo amor, paciencia y alegría han sido mi mayor inspiración y motivo de esfuerzo durante estos años.

A mi hermano, Dario, por caminar a mi lado como hermano y amigo, siempre con un apoyo incondicional. A toda mi familia extendida y amigos, quienes con su apoyo constante me han dado la fuerza necesaria para superar cada desafío.

Quiero expresar mi más profundo agradecimiento a mis directores de tesis, Alfonso Bahillo y Laura Arjona, por su apoyo constante, sus comentarios perspicaces y su motivación continua a lo largo de todo este proceso. Su mentoría ha sido fundamental para mi crecimiento como investigador independiente, alentándome siempre a ampliar los límites de mi curiosidad y mi compromiso con el aprendizaje.

También agradezco al Ministerio de Ciencia, Innovación y Universidades (MICIU) y a la Agencia Estatal de Investigación (AEI) por financiar este trabajo a través de la "Ayuda para contratos predoctorales" de 2020 (ref. PRE2020-095612). Este apoyo ha sido determinante para impulsar mi carrera académica.

Extiendo mi más sincero agradecimiento al grupo de investigación Deusto Smart Mobility y a Deusto-Tech por proporcionar un entorno de investigación estimulante y colaborativo, así como por ofrecer los recursos necesarios para completar este trabajo. Igualmente, estoy en deuda con mis colegas y amigos, cuya amistad, apoyo y entusiasmo hicieron de este camino una experiencia mucho más gratificante y enriquecedora.

Quisiera expresar mi especial agradecimiento al Prof. Alípio M. Jorge y a la Prof. Elsa Ferreira Gomes por hacer posible mi estancia de investigación internacional en INESC TEC. Esta experiencia ha sido increíblemente valiosa tanto para mi desarrollo profesional como personal. A todos MUCHAS GRACIAS!!!

Marcos Lazaro Alvarez Arteaga
Bilbao, September 2025

Abstract

Sound-based uroflowmetry (SU) offers a non-invasive alternative to traditional clinical uroflowmetry (UF) by analyzing the acoustic signals generated during urination. Leveraging consumer-grade audio devices and machine learning (ML) techniques, SU enables a low-cost and remote evaluation of the lower urinary tract function (LUTS). This thesis investigates the feasibility, performance and robustness of SU through four interrelated studies, each addressing a key aspect of urinary flow estimation using sound and artificial intelligence (AI). The methodology is grounded in a novel dataset comprising both real and synthetic voiding sound recordings, paired with diverse ML models.

The first study is motivated by the versatility of SU compared to UF, as it enables frequent and discreet assessments in natural environments like the patient's home. This allows for a more comprehensive representation of individual voiding patterns over time, overcoming the situational bias and variability typically associated with single in-clinic measurements. We recruited 50 male volunteers aged 18–60. After excluding three recordings due to noise or uncertainty in stream direction, 47 valid test cases were retained for analysis. The study compares SU and UF in estimating the urinary flow curve and voided volume (VV), taking as reference a commercial and medically certified Minze uroflowmeter. Audio signals were simultaneously recorded using three different devices: an Ultramic384 (UM) microphone, a Mi A1 smartphone (Phone), and an Oppo smartwatch (Watch). These signals were then segmented and used to train ML models. Mean absolute errors (MAE) for flow rate estimation were 2.6, 2.5 and 2.9 ml/s, with R^2 values of 84%, 83% and 79%, respectively. For voiding flow and VV, Lin's concordance coefficients reached 0.9 and 0.85, respectively. These results confirm that SU enables reliable estimation of key uroflowmetry parameters using low-cost, portable devices.

The second study addresses one of the key limitations in SU: the difficulty of obtaining real, labeled datasets due to ethical, technical and logistical constraints. Additionally, clinical data are often imbalanced with respect to urinary flow rates. To overcome these challenges, a simulated dataset of voiding sounds was created, covering flow rates from 1 to 50 ml/s in 1 ml/s increments, generated using a precision peristaltic pump and recorded with the same three devices used in the first study under controlled conditions. This dataset enables the training of highly generalizable mod-

els and addresses the lack of labeled data in SU, promoting reproducibility in future research.

The third study stems from the need to assess the generalization capacity of ML models, that is, their ability to perform well on real-world data after being trained on synthetic datasets. To this end, random forest (RF), gradient boosting, support vector machines and convolutional neural networks were compared in both regression and classification tasks. Among them, the RF model achieved the best performance on the synthetic dataset and was therefore selected for partial retraining using the real SU recordings from the first study. After this retraining, the RF model achieved MAE below 2.5 ml/s and quadratic weighted kappa scores above 0.86, validating the synthetic-to-real transfer approach and demonstrating its practical feasibility.

The fourth study addresses a critical challenge in SU, driven by the practical difficulty many individuals particularly elderly patients experience in maintaining a continuous urinary stream aimed at the toilet water. This often leads to the stream impacting other surfaces, such as ceramic, which alters the acoustic characteristics of the signal and may cause errors in the estimation of flow parameters during SU tests. The main objective of this study was to develop a system capable of automatically detecting whether the urinary stream impacts a ceramic surface, thereby enhancing the reliability of subsequent estimation models. To address this, a supervised ML classifier was developed to distinguish between three scenarios: water impact, ceramic impact and absence of voiding (silence). A supervised ML classifier was developed to distinguish these three scenarios using frequency-domain features. Furthermore, an analysis was performed after removing audible frequencies below 8 kHz to assess performance in privacy-sensitive applications. The classifier achieved an accuracy of up to 99.46%, enhancing the reliability of downstream flow estimation models.

Overall, this thesis establishes a robust framework for SU using ML, addressing signal variability, demonstrating accuracy across devices, introducing a reference synthetic dataset and incorporating an analysis of acoustic privacy to ensure suitability in sensitive home environments. The findings support the development of affordable and privacy-preserving home-based urinary care systems. Despite current limitations related to devices heterogeneity and acoustic variability, this work contributes meaningfully to remotely support the diagnose of LUTS and digital health innovation.

Resumen

La sonouroflujometría (SU) representa una alternativa no invasiva a la uroflujometría clínica tradicional (UF), al analizar las señales acústicas generadas durante la micción. Mediante el uso de dispositivos de audio de consumo y técnicas de aprendizaje automático (ML), SU permite una evaluación remota y de bajo coste de los síntomas del tracto urinario inferior (LUTS). Esta tesis analiza la viabilidad, el rendimiento y la robustez de SU a través de cuatro estudios interrelacionados, cada uno abordando un aspecto clave en la estimación del flujo urinario utilizando sonido e inteligencia artificial (IA). La metodología se fundamenta en un conjunto de datos novedoso que combina grabaciones reales y sintéticas de sonidos de micción, junto con diversos modelos de ML.

El primer estudio parte de la versatilidad de la SU frente a la uroflujometría tradicional (UF), ya que permite realizar evaluaciones frecuentes y discretas en entornos naturales como el hogar del paciente. Esto facilita una representación más completa de los patrones miccionales individuales a lo largo del tiempo, superando el sesgo situacional y la variabilidad típicamente asociada a las mediciones puntuales realizadas en entornos clínicos. Se reclutaron 50 voluntarios varones de entre 18 y 60 años. Tras excluir tres grabaciones por ruido o incertidumbre en la dirección del chorro, se analizaron 47 casos válidos. El estudio compara la SU y la UF en la estimación de la curva de flujo urinario y el volumen miccional (VV), utilizando como referencia el uroflujómtro comercial y medicalmente certificado, Minze. Las señales de audio fueron grabadas simultáneamente mediante tres dispositivos distintos: un micrófono Ultramic384 (UM), un teléfono inteligente Mi A1 (Phone) y un reloj inteligente Oppo (Watch). Posteriormente, estas señales fueron segmentadas y utilizadas para entrenar modelos de ML. Los errores absolutos medios (MAE) para la estimación del flujo fueron de 2,6, 2,5 y 2,9 ml/s, con valores de R^2 del 84 %, 83 % y 79 %, respectivamente. En cuanto al flujo miccional y el VV, los coeficientes de concordancia de Lin alcanzaron 0,9 y 0,85 respectivamente. Estos resultados confirman que la SU permite estimar de forma fiable los parámetros clave de la UF utilizando dispositivos portátiles y de bajo coste.

El segundo estudio responde a una de las principales limitaciones actuales de la SU: la dificultad para obtener conjuntos de datos reales etiquetados, debido a restricciones éticas, técnicas y logísticas. Además, los datos clínicos reales suelen estar desequilibra-

dos en cuanto a los rangos de flujo urinario. Para superar estos obstáculos, se diseñó un conjunto de datos simulado de sonidos de micción con flujos que van de 1 a 50 ml/s en incrementos de 1 ml/s, generado mediante una bomba peristáltica de precisión y grabado con los tres dispositivos usados en el primer estudio bajo condiciones controladas. Este conjunto de datos permite entrenar modelos altamente generalizables y cubre la carencia de datos etiquetados en SU, fomentando así la reproducibilidad en futuras investigaciones.

El tercer estudio parte de la necesidad de evaluar la capacidad de generalización de los modelos de ML, es decir, su rendimiento cuando se aplican a datos reales tras haber sido entrenados con datos sintéticos. Para ello, se compararon distintos enfoques (random forest (RF), gradient boosting, máquinas de vectores soporte y redes neuronales convolucionales) en tareas de regresión y clasificación. Entre ellos, el modelo de RF fue el que obtuvo el mejor rendimiento sobre el conjunto de datos sintético y, por ello, fue seleccionado para ser reentrenado parcialmente con las grabaciones reales de SU utilizadas en el primer estudio. Tras su reentrenamiento parcial con datos reales, los modelos alcanzaron errores absolutos medios por debajo de 2,5 ml/s y valores de kappa ponderado cuadrático superiores a 0,86, validando así el enfoque de transferencia de datos sintéticos a reales.

El cuarto estudio aborda un desafío crítico en la SU, motivado por la dificultad práctica que muchas personas, especialmente los pacientes de edad avanzada, enfrentan para mantener un flujo urinario continuo dirigido hacia el agua del inodoro. Esta limitación conlleva con frecuencia a que el chorro impacte sobre otras superficies, como la cerámica, lo que modifica las propiedades acústicas de la señal y puede inducir errores en la estimación de los parámetros de flujo a partir de las pruebas SU. El objetivo principal de este estudio fue desarrollar un sistema capaz de detectar automáticamente si la micción se produce sobre una superficie cerámica, a fin de mejorar la fiabilidad de los modelos de estimación posteriores. Se presenta un clasificador basado en ML, entrenado para distinguir entre estos tres escenarios utilizando características frecuenciales. Además, se realizó un análisis eliminando las frecuencias audibles por debajo de 8 kHz, con el objetivo de evaluar el rendimiento del sistema en contextos sensibles a la privacidad. El clasificador alcanzó una precisión de hasta el 99,46 %, lo que mejora la fiabilidad de los modelos posteriores de estimación de flujo.

En conjunto, esta tesis establece un marco sólido para la SU basada en ML, abordando la variabilidad de las señales, demostrando precisión en distintos dispositivos, proponiendo un conjunto de datos sintético de referencia e incorporando un análisis de la privacidad acústica para garantizar su aplicabilidad en entornos domésticos sensibles. Los resultados respaldan el desarrollo de sistemas de monitorización miccional domiciliaria asequibles y respetuosos con la privacidad. A pesar de las limitaciones actuales en cuanto a heterogeneidad de dispositivos y variabilidad acústica, este trabajo contribuye significativamente a apoyar de forma remota el diagnóstico de los LUTS y a la innovación en salud digital.

Contents

List of Figures

List of Tables

Abbreviations

AI artificial intelligence. 4, 6, 11, 13–15, 18, 20, 24, 36, 37, 46, 47, 74, 79, 81

AUC area under the receiver operating characteristic curve. 18, 19

BPH benign prostatic hyperplasia. 2

CCC Lin's concordance correlation coefficient. 6, 17, 19, 20, 23, 47

CNN convolutional neural network. 9, 47

DL deep Learning. 15, 17–19

FFT fast fourier transform. 7, 9, 23, 62, 78

FT flow time. 2, 4, 12, 14–17, 19

GB gradient boosting. 15, 19, 47

GBR gradient boosting regressor. 6, 9, 23

ICC intraclass correlation coefficient. 16, 19

k-NN k-nearest neighbours. 7, 15, 19, 62, 78

LUTS lower urinary tract symptoms. 2, 3, 6, 12, 16

MAE mean absolute error. 6, 10, 19, 20, 23, 47, 74

MFCC mel-frequency cepstral coefficients. 9, 15, 19, 37, 80

ML machine learning. 4, 5, 7–11, 15, 19, 22, 47, 61, 62, 74–80

MSE mean squared error. 10

PCC Pearson Correlation Coefficient. 15, 17

Phone Mi A1 smartphone. 6, 19, 23, 37, 47, 80

Qave average flow rate. 2, 4, 8, 14–17, 19, 74, 76

Qmax maximum flow rate. 2, 4, 8, 12, 14–17, 19, 74, 76

QWK quadratic weighted kappa. 10

RF random forest. 7–9, 19, 47, 62, 78

RFR random forest regressor. 6, 23, 47

RMSE root mean square error. 16, 17, 19

SMAPE symmetric mean absolute percentage error. 17

SR sampling rate. 13, 37

SU sound-based uroflowmetry. xi, xii, 3–24, 36, 46, 47, 61–63, 74–78, 80, 81

SVM support vector machine. 7, 9, 19, 62, 78

SVR support vector regressor. 6, 23, 47

t-SNE t-distributed stochastic neighbour embedding. 8

TQmax time to maximum flow. 2, 4, 16, 17

UF traditional uroflowmetry. xi, 2–4, 6, 9, 10, 12–17, 22, 23, 77

UM Ultramic384. 6, 19, 23, 37, 47, 80

VV voided volume. 2, 4, 6, 8, 10, 12, 14–20, 22, 23, 74, 76, 77

Watch Oppo smartwatch. 6, 19, 23, 37, 47, 80

ZCR zero-crossing rate. 9

"A journey of a thousand miles begins with a single step."

— Lao Tzu

Chapter 1

Introduction

1.1 Context and motivation

One of the most prevalent health issues associated with population aging is lower urinary tract symptoms (LUTS), affecting over 2.3 billion individuals worldwide, which accounts for 45.8% of the global population [3]. LUTS encompass conditions related to obstructions in the urinary tract, including the kidneys, bladder, ureters (which carry urine from the kidneys to the bladder) and the urethra (which connects the bladder to the outside of the body) [4]. Urethral obstructions can occur for a variety of reasons, including gastrointestinal problems [5]. These obstructions are more common in men, especially as they age and the prostate enlarges due to benign prostatic hyperplasia (BPH) [6]. It is estimated that more than 60% of men over forty years old experience LUTS [2]. Such dysfunctions can significantly compromise patients' quality of life, causing persistent discomfort, psychological distress and considerable clinical impact [7].

To diagnose and monitor LUTS, a non-invasive and physiological test known as traditional uroflowmetry (UF) is commonly used. It has been shown to provide objective evidence regarding prostate enlargement, overactive bladder, urinary incontinence, or neurogenic bladder [8]. This test is currently performed in clinical settings, where patients are asked to urinate into a funnel that collects urine in a container equipped with a uroflowmeter, which measures several parameters associated with the voiding process. These parameters help assess bladder emptying function and include voided volume (VV), flow time (FT), average flow rate (Qave), maximum flow rate (Qmax) and time to maximum flow (TQmax) [2]. Figure 1.1 graphically illustrates the meaning of each of these urinary flow parameters.

Despite its clinical utility, UF has certain limitations. Being an in-clinic procedure, patients must void in an unfamiliar environment and on demand, often with a bladder that is either underfilled or overfilled. This introduces considerable test-to-test variability, as the stress of the clinical setting may influence flow rate and lead to

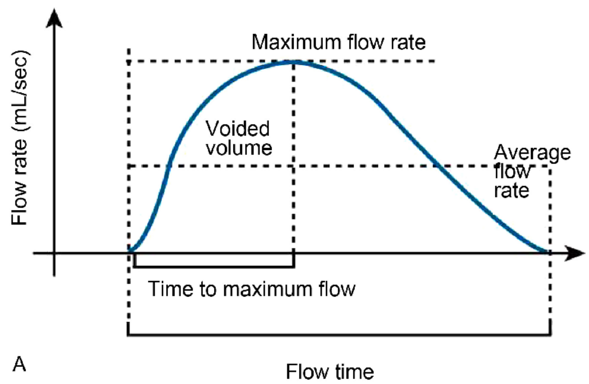


Figure 1.1: Main parameters in the UF test (source: [1])

non-representative outcomes [9]. Therefore, it is recommended that UF be repeated multiple times, requiring costly and time-consuming clinical visits. This challenge is expected to worsen in the coming decades, as the aging global population will lead to a significant increase in the number of men requiring medical diagnosis for LUTS.

Conducting UF tests in home settings has the potential to improve data collection on voiding patterns, thereby supporting clinical decision-making [10]. This need has led to the emergence of portable dedicated uroflowmeters. However, such devices have not been widely adopted in routine clinical practice due to their high cost and complexity, which pose a barrier particularly for older patients [11].

As a remote and proactive alternative to overcome the limitations of UF, sound-based uroflowmetry (SU) has emerged. SU characterizes flow patterns by capturing the sound produced when the urine stream impacts the toilet water. This approach offers a non-invasive, portable and potentially automatable solution for LUTS monitoring, enabling more accessible, home-based and discreet evaluations without the need for specialized clinical equipment. The scientific literature has shown good correlation between UF and SU in terms of signal shape and several key urinary flow parameters [12, 13].

Unlike UF, which requires dedicated electronic equipment, patient presence in clinical environments and technical supervision, SU can be implemented using consumer devices such as smartphones [12], smartwatches [14–16], or standalone microphones. This enables repeated testing in a natural environment like the home, reducing clinical setting bias, lowering healthcare costs and enhancing adherence to therapy monitoring (see Figure 1.2).

Moreover, SU aligns with current trends in personalized medicine and remote monitoring, where continuous and passive acquisition of biometric data is valued without compromising user privacy. Despite its demonstrated potential, current SU research still faces significant limitations. Existing studies have been conducted with small sample sizes, lack standardized voiding environments and do not provide public or reproducible datasets. Furthermore, few studies have addressed the complete signal

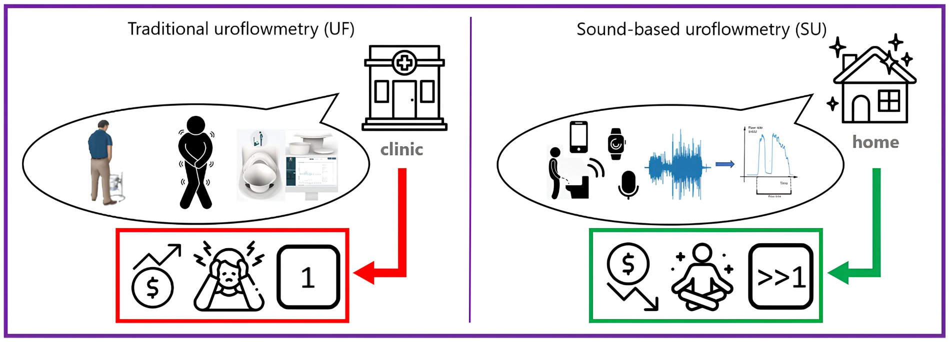


Figure 1.2: Comparison between UF and SU. Key differences are shown, highlighting SU's advantages: lower cost, testing in natural environments and the ability to perform more frequent assessments.

analysis pipeline from acoustic environment classification to accurate urinary flow estimation. The absence of robust solutions that function in uncontrolled conditions using consumer-grade devices limits its widespread clinical application.

In this context, this thesis proposes a comprehensive machine learning (ML)-based approach for estimating urinary parameters from acoustic signals. The work is structured as a compendium of publications, each addressing a different component of the problem. Through these contributions, the thesis aims to advance the state of the art in SU, offering scientifically sound and technically viable solutions for integration into home-based monitoring systems thus contributing to more accessible, personalized and patient-centered urological care.

1.2 Hypothesis and scientific contribution

Thesis Hypothesis:
The combined use of ML techniques and acoustic signals captured using consumer-grade microphones enables an accurate estimation of five key voiding parameters (VV, Qmax, FT, TQmax and Qave) under realistic conditions, thereby facilitating the development of accessible, non-invasive and privacy-preserving home-based urinary monitoring systems.

This thesis is based on the application of artificial intelligence (AI), specifically ML techniques, to the processing of audio signals captured during urination. Using recordings obtained via high-fidelity microphones, smartphones and smartwatches, models were developed to classify the physical surface impacted by the urine stream, to estimate urinary parameters and to analyze the shape of the acoustic envelope, an interpretable representation of the signal that can assist clinicians in identifying typical voiding patterns (e.g., normal, prostatic, fluctuating, intermittent and plateau [8, 17]).

The proposed approach also incorporates the use of synthetic data generated with precision peristaltic pumps to train models under controlled conditions, which are subsequently validated in real-world scenarios. This combination of simulated and real data constitutes a methodological innovation that helps overcome the scarcity of annotated clinical data. Additionally, the approach includes the analysis of model performance under privacy-preserving conditions by excluding audible frequency bands particularly those associated with human speech thereby enabling applications in sensitive home environments.

1.3 Structure and included publications

This thesis follows a compendium format, with all publications focused on the use of ML applied to SU for urinary flow analysis. Figure 1.3 provides a graphical overview of the methodological workflow followed throughout this research.

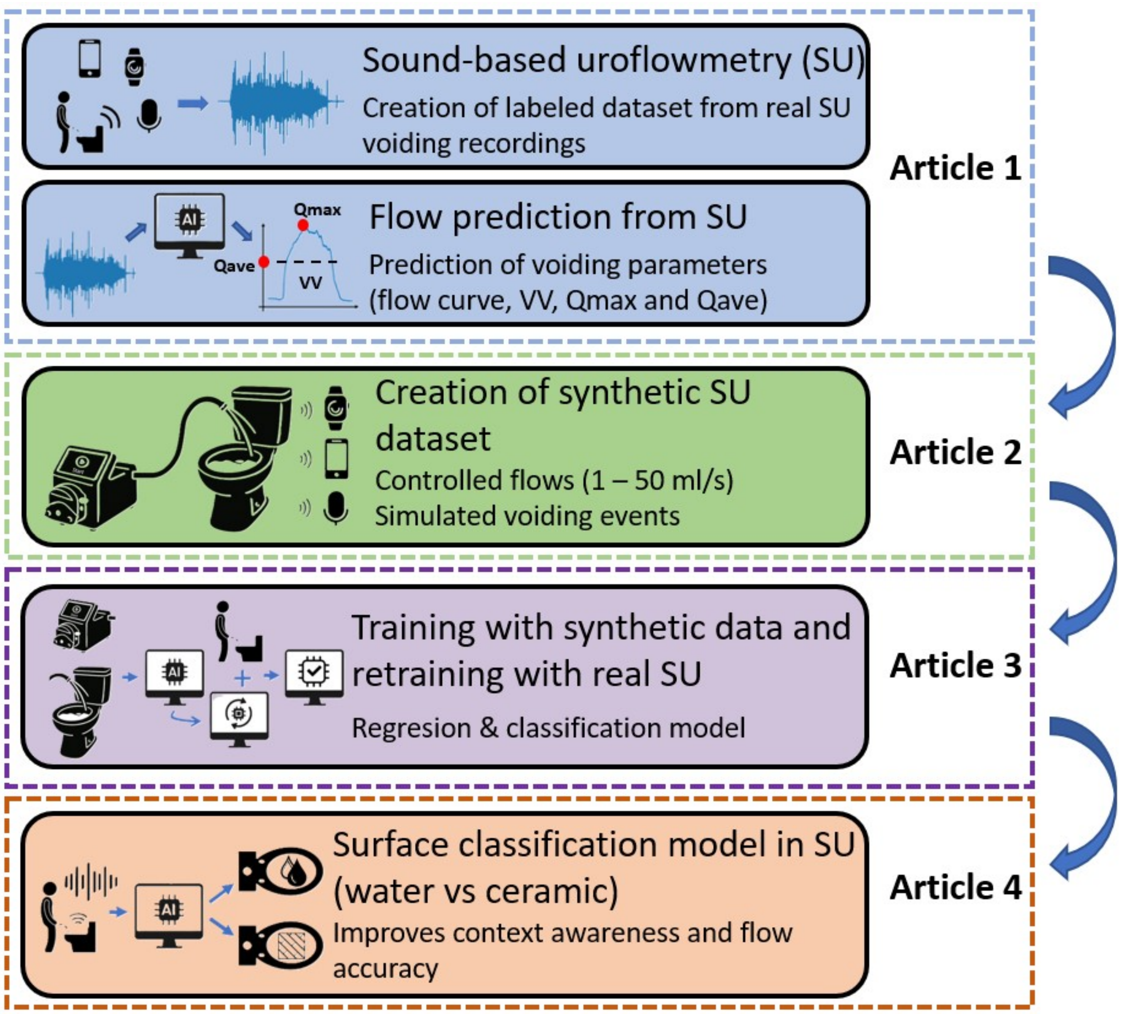


Figure 1.3: Graphical overview of the methodological workflow of the thesis.

The following sections describe each of the included articles in detail:

- **Chapter 2 – Article 1:** This chapter presents a system for predicting urinary flow curves and VV based on acoustic analysis of urination. As part of this study, a new dataset of real SU recordings was created using 50 male volunteers aged 18–60, who urinated into a commercial and medically certified Minze uroflowmeter [18] while the sound was simultaneously recorded using three different devices: a professional microphone Ultramic384 (UM), a Mi A1 smartphone (Phone) and an Oppo smartwatch (Watch). After excluding three recordings due to noise or uncertainty in stream direction, 47 valid test cases were retained for analysis. The audio signals were processed to extract spectral features and train regression models such as random forest regressor (RFR), gradient boosting regressor (GBR) and support vector regressor (SVR). Experimental results, evaluated against a uroflowmeter Minze, showed a mean absolute error (MAE) below 2.9 ml/s for flow prediction, with Lin's concordance correlation coefficient (CCC) above 0.85 for VV and above 0.90 for flow rate. The 0–8 kHz frequency band was found to contain over 80% of the relevant information.

  This work outperforms previous studies and demonstrates the feasibility of performing SU tests at home using everyday devices, promoting a more accessible, continuous and user-friendly approach for monitoring LUTS. Comparisons with conventional methods (UF) showed a high correlation, identifying the Watch as the best alternative in terms of convenience, portability, accuracy and cost [19].

- **Chapter 3 – Article 2:** This chapter presents a synthetic dataset of simulated voiding events with precise flow control using a peristaltic pump in a real toilet environment. Simultaneous audio recordings were collected using three representative devices: a professional microphone (UM), a Phone and a Watch, covering flow rates from 1 to 50 ml/s in 1 ml/s increments.

  This work addresses a critical need in the SU field: the lack of public, standardized and labeled datasets, which has thus far hindered the development and validation of AI algorithms in this domain. The generated dataset aims to establish a shared research foundation, enabling supervised training and reproducible evaluation of flow estimation models. Additionally, the controlled conditions make this dataset a robust preliminary resource for validating models prior to using real-world signals.

  The article also confirms the technical validity of the procedure, detailing the equipment, recording methodology and verification of actual flow rates delivered by the pump. This contribution is thus significant for the scientific community working on accessible, home-based and automated technologies for lower urinary tract evaluation [20].

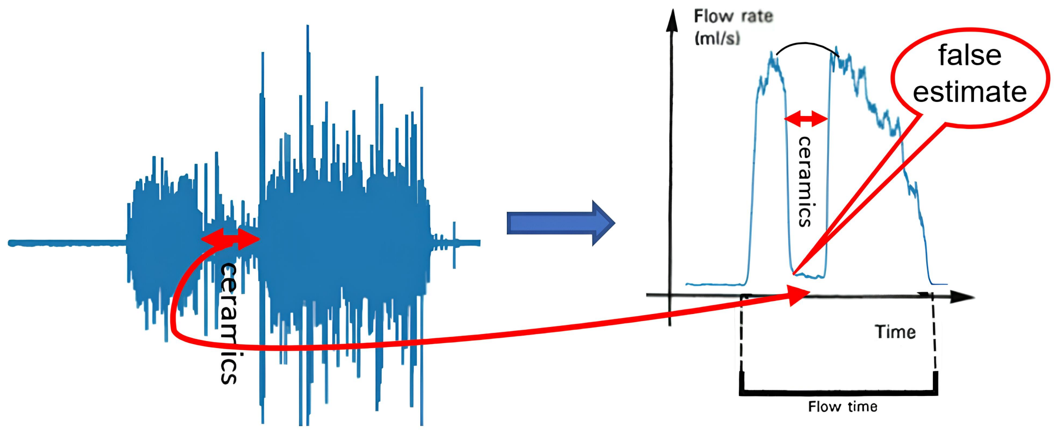


Figure 1.4: Effect of impact surface change in SU. The left panel shows a sound signal from urination impacting a ceramic surface; the right panel shows the resulting error in flow estimation due to this surface change.

- **Chapter 4 – Article 3:** This chapter evaluates the feasibility of training ML models (for both classification and regression tasks) using synthetic data generated from simulated voiding events under flow control. To assess the models in real-world conditions and evaluate the effectiveness of synthetic data training, the best-performing models were subsequently retrained and validated using a dataset of real voiding sound recordings. This strategy helps circumvent ethical and logistical challenges related to real data collection in early development stages, while also providing a reproducible and controlled foundation for model development. Moreover, it highlights that synthetic data can serve as a valuable pretraining resource, improving model robustness when applied to real clinical signals and setting a precedent for effective simulation-based development of SU tools [21].

- **Chapter 5 – Article 4:** This chapter presents an ML-based model for classifying the surface where the urine stream impacts during SU tests, distinguishing between ceramic and water, as well as absence of urination (silence). This classification is critical since acoustic characteristics vary significantly depending on the impact surface and a decrease in sound envelope amplitude may result either from an actual flow reduction or a surface change (e.g., from water to ceramic) (see Figure 1.4). This phenomenon can lead to misinterpretation of voiding parameters if the acoustic context is unknown [22].

  The proposed system uses ML algorithms such as support vector machine (SVM), random forest (RF) and k-nearest neighbours (k-NN), trained on a novel dataset created as part of this work, consisting of 6481 one-second clips recorded in domestic toilets with real subjects. Feature extraction was performed primarily in the frequency domain using fast fourier transform (FFT), linear band analysis and dimensionality reduction via t-distributed stochastic neighbour embedding (t-SNE) to assess class separability.

Automatically identifying the impact surface overcomes a major limitation of traditional SU, which typically requires patients to direct the stream exclusively toward the toilet water to ensure reliable results. This requirement is unnatural and often unfeasible for elderly or mobility-impaired individuals, who represent a large portion of the affected population.

The results show classification rates above 99% with the RF model and 93.29% accuracy even when excluding frequency bands below 8 kHz (typically associated with human speech), thereby also contributing to user privacy. This work lays the foundation for estimating voiding parameters by surface type, eliminating physical constraints on patient behavior and enabling non-invasive, home-based urinary tract assessment.

1.4 General and specific objectives

General Objective

To develop a non-invasive and accessible system for estimating key voiding parameters (including the urinary flow curve, VV, Qmax and Qave) based on the analysis of SU signals using ML techniques, suitable for home environments and to validate the system with both real clinical recordings and synthetically generated data under controlled conditions.

Specific Objectives:

1. To create a labeled dataset of real SU recordings from urination events, using synchronized measurements obtained with a commercial and medically certified Minze uroflowmeter as ground truth.

2. To develop and validate regression models for estimating key urodynamic parameters (including the urinary flow curve, Qmax, Qave and VV) from SU signals captured using accessible devices such as smartphones or smartwatches.

3. To create a synthetic dataset of voiding events with precise annotations and flow control to support reproducible research in SU.

4. To train ML models on synthetic data and evaluate their generalization capacity on real recordings from clinical settings, demonstrating their applicability in real-world contexts.

5. To train ML models capable of automatically classifying the voiding surface (water or ceramic) or the absence of urination (silence), in order to mitigate flow estimation errors caused by acoustic context changes.

6. To analyze the impact of using frequency bands above 8 kHz for urinary flow estimation as a privacy-preserving strategy without compromising analytical quality.

1.5 Common methodology

This section presents the general methodological framework supporting the various studies included in this doctoral thesis. Although each experimental chapter has its own objectives, specific procedures and contextual particularities, all share a coherent methodological approach that ensures homogeneity, comparability and reproducibility of the results.

The thesis is structured around four complementary investigations that, while employing different experimental setups and data sources, follow a common methodological cycle adapted to the characteristics of SU. This methodology is organized into the following key stages:

- **Literature review:** A comprehensive analysis was conducted of existing literature related to UF and SU, as well as recent applications of ML techniques to biomedical signal processing. This review helped establish the working hypothesis, identify scientific gaps and guide the design of subsequent studies.

- **Experimental design and data acquisition:** Each study defined specific protocols for acquiring acoustic signals related to voiding events, using various types of recording devices (professional microphone, smartphone and smartwatch). In the third study, a synthetic dataset with flow control was developed to facilitate reproducible model training.

- **Preprocessing and feature extraction:** The acoustic signals were segmented and obtained under controlled conditions to minimize background noise and ensure signal clarity. Relevant features were then extracted in spectral, statistical and cepstral domains (FFT, zero-crossing rate (ZCR), variance, mel-frequency cepstral coefficients (MFCC), among others), tailored to the objectives of each chapter.

- **Model design and implementation:** ML models were developed for classification and regression tasks, including algorithms such as SVM, RF, GBR and convolutional neural network (CNN). Model selection depended on the nature of the problem and the complexity of the available data.

- **Evaluation and result analysis:** Models were validated using standard metrics (MAE, mean squared error (MSE), R^2, Accuracy, quadratic weighted kappa (QWK)), with particular attention to cross-validation for assessing generalization capacity and preventing overfitting. Results were compared with prior state-of-the-art studies and across different recording and segmentation configurations.

- **Discussion and cross-validation:** The results were critically analyzed to assess the robustness of the proposed hypotheses, identifying strengths, limitations and areas for improvement. Particular emphasis was placed on evaluating the generalization capacity of models trained on synthetic data to real recordings and the feasibility of implementation in home-based settings.

This modular and replicable approach enabled a systematic exploration of various aspects of the non-invasive urinary flow estimation problem, laying the groundwork for its future integration into accessible and privacy-conscious clinical solutions.

1.6 Thesis structure

This doctoral thesis adopts a compendium format, where the main scientific results are presented as articles published in indexed international journals. Each chapter contributes in a structured manner to the research narrative, reflecting the evolution of the proposals, experiments and validations carried out throughout the doctoral work.

The overall structure of the thesis is as follows:

- **Chapter 1 – Introduction:** This chapter presents the general context of the research, the motivation behind the study, the general and specific objectives, the thesis hypothesis and the common methodology adopted across the different studies. It also introduces the structure of the thesis and the included publications.

- **Chapter 2 – State of the Art:** This chapter offers a detailed review of the literature related to SU and ML techniques applied to biomedical signal analysis. It examines prior advances, limitations of traditional methods and current opportunities that justify the chosen research approach.

- **Chapter 3 – Flow prediction in sound-based uroflowmetry:** This chapter describes a system for estimating key voiding parameters such as urinary flow and VV, using acoustic recordings captured from three types of consumer devices: professional microphone, smartphone and smartwatch. The performance of the proposed models is compared with commercial and medical certified uroflowmeters and their correlation with UF tests is analyzed.

- **Chapter 4 – Annotated dataset of simulated voiding sound for urine flow estimation:** This chapter presents the design and creation of a synthetic dataset of simulated voiding events under controlled flow conditions. The dataset is intended to support supervised research in SU, promoting the development of standardized labeled data for training and validating algorithms.

- **Chapter 5 – Leveraging synthetic data to develop a machine learning model for voiding flow rate prediction from audio signals:** This chapter explores the generalization capacity of ML models trained exclusively on synthetic data, evaluating their applicability in real-world conditions. The results validate the utility of the generated dataset and its potential to reduce the need for large volumes of real clinical data in sensitive contexts.

- **Chapter 6 – Automatic classification of the physical surface in sound uroflowmetry using machine learning methods:** This chapter addresses the problem of automatically classifying the voiding impact surface (water or ceramic), as well as the absence of voiding (silence), using ML techniques. Its relevance is justified by the need to improve urinary flow estimation accuracy and reduce restrictive requirements in home-based SU testing.

- **Chapter 7 – Conclusions and future work:** The final chapter synthesizes the main findings and contributions of the thesis. It assesses the extent to which the objectives were met, discusses the limitations encountered during the research and proposes possible future research directions focused on improving SU through AI and deploying these systems in real clinical settings.

This research was conducted with the approval of the Valladolid East Health Area Medicine Research Ethics Committee on 27 July 2023, under reference PI-GR-23-3275 (minutes number 16/2023), at the Hospital Clínico Universitario de Valladolid. The Ethics Committee complies with Good Clinical Practice (GCP) standards (CPMP/ICH/135/95). All participants provided written informed consent for the use of both conventional and SU data.

Chapter 2

Literature Review

2.1 Introduction

This chapter provides a review of the state of the art in literature that explores the use of SU for estimating key urinary parameters (flow and VV) employed by urologists to evaluate LUTS. A normal urinary flow pattern for male patients typically follows a bell-shaped curve (see Figure 1.1), where Qmax exceeds 15 ml/s and is this value is reached within the first 5 seconds of voiding. The rest of the curve varies depending on VV and FT. A Qmax below 10 ml/s is considered abnormal. In female patients, Qmax values typically range between 20 and 36 ml/s [8, 23–25]. Our work focuses on flow parameter prediction in male patients who void while standing. Women were excluded from our study due to notable differences in voiding habits, such as voiding position (sitting vs. standing), which affect the impact surface in the toilet and thereby alter the acoustic profile of the voiding process.

In addition to Qmax and VV, the shape of the flow curve in UF can indicate potential abnormalities in the lower urinary tract. However, a complete urodynamic evaluation is needed for a definitive diagnosis. Urinary flow is influenced by factors such as detrusor muscle contractility, urethral resistance, bladder volume at voiding onset and testing conditions [24]. Five typical flow curve patterns have been identified to guide clinical interpretation: normal, prostatic, fluctuating, intermittent and plateau [8, 17]. Each pattern reflects a distinct pathophysiological mechanism. For example, a fluctuating flow shows an irregular curve with multiple peaks, typically due to detrusor-sphincter dyssynergia or bladder instability. Intermittent flow appears as alternating pauses and resumptions in voiding, potentially linked to interrupted detrusor contractions or effortful voiding. Figure 2.1 illustrates examples of abnormal flow patterns.

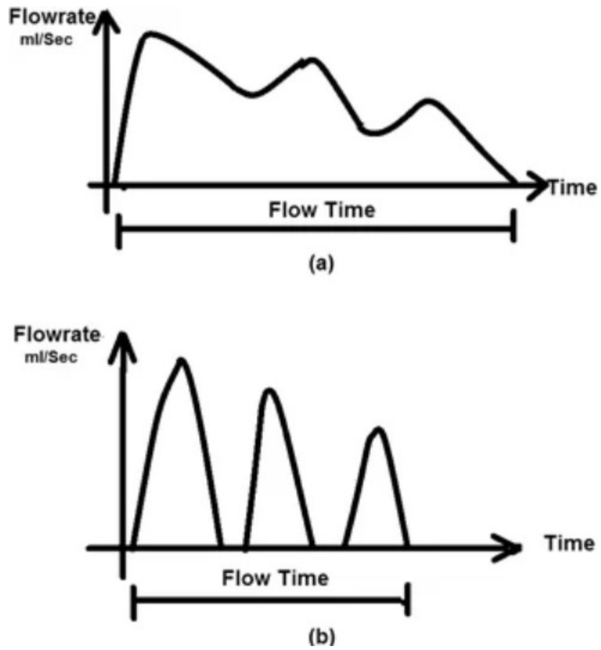


Figure 2.1: Abnormal flow curve sketches, (a) fluctuating and (b) intermittent (source: [2])

Advancements in information technology and the rapid development of AI applied to acoustic signal processing have opened new opportunities to transform SU into a clinically viable tool. These innovations address critical technical challenges, such as accurate flow estimation in uncontrolled environments and improving robustness to ambient variability. Within this context, the present thesis significantly contributes to the SU field by developing methods and resources aimed at clinical integration, enabling more accessible, continuous and patient-centered assessments.

2.2 Fundamentals of SU and early developments

SU analyzes the acoustic signal generated when the urine stream impacts the toilet bowl to estimate urinary flow characteristics. This emerging technique offers a non-invasive and accessible alternative to UF, particularly in telemedicine contexts.

The origins of SU trace back to the work of [26], who proposed an approach based on the acoustic characterization of urinary flow in controlled simulations. Inspired by studies estimating rainfall rates from the sound of water droplets impacting surfaces, the authors explored extracting urodynamic information from the acoustic signal generated during voiding. They designed a system using a syringe pump with a fixed flow rate to simulate urine. A digital microphone with a sampling rate (SR) of 44 kHz recorded the splashing sound of urine hitting a transparent surface. The recordings were spectrally analyzed using time-frequency and wavelet transforms. This experimental work did not progress to patient testing and the authors concluded that only general correlations could be drawn between acoustic signatures and flow parameters.

In [27], one of the first clinical validations of SU was conducted by recording the voiding stream's sound and comparing it with UF. A microphone placed near the toilet captured the sound during spontaneous voiding. The acoustic signal's amplitude envelope and temporal characteristics were analyzed and compared with UF data. The authors reported good correlation in the general curve shape and parameters such as VV and FT, though Qmax estimation was less accurate. The study's limitations included the lack of quantitative error metrics and reliance on visual analysis, but it supported SU's initial clinical feasibility.

[28] assessed SU's validity using simultaneous acoustic recordings and UF. The study involved 25 healthy men, with a mobile phone placed near the toilet to record the urine stream sound. Signal analysis focused on time-domain sound intensity via the amplitude envelope to estimate voiding parameters. Results showed high correlations for FT (0.87), moderate for VV (0.68) and low for Qmax (0.38). The use of simple linear correlation techniques limited overall accuracy.

A more recent study by [29] proposed a mobile SU system using smartphone recorded voiding sounds in real-life settings. The study evaluated 97 male and female patients with urinary symptoms, comparing SU-derived parameters with UF results. High correlation coefficients were reported for VV and Qave in both genders and acceptable correlation for Qmax. Despite promising results, the study lacked detailed signal processing methods and parameter estimation techniques, limiting reproducibility and generalizability.

In summary, SU is a relatively recent research area in urodynamics, with limited but growing contributions exploring its theoretical and clinical applicability. Early studies have validated its feasibility as a non-invasive technique but often relied on basic signal analysis and lacked methodological transparency, hindering reproducibility. However, advancements in acoustic signal processing and AI, particularly in audio-based regression and classification, are enabling a new generation of more robust and interpretable SU studies. Rather than aiming to replace conventional UF, SU complements it by enabling non-invasive, frequent and decentralized monitoring of voiding events. This facilitates a more comprehensive representation of individual urinary patterns over time, mitigating the situational variability and clinical constraints of isolated in-office UF tests. Consequently, SU emerges as a promising tool for digital health applications and long-term urological monitoring in home-based settings.

2.3 Application of AI in SU

In recent years, there has been a growing interest in applying AI techniques to the analysis of acoustic signals generated during urination, as a non-invasive means of estimating relevant urodynamic parameters. This section analyzes various representative studies that integrate ML and deep Learning (DL) models in the context of SU, with a special focus on the methodologies employed, sample sizes, evaluation metrics and main limitations.

One of the first studies to explore the feasibility of AI to estimate urodynamic parameters from voiding sounds was conducted by [16], who developed a ML model to predict urinary flow characteristics from acoustic recordings obtained with a smartphone during urination. The study included 25 healthy men, recording a total of 52 voiding sessions. The acoustic signals were segmented into 0.1-second frames to extract MFCC, which served as input to an ensemble model based on k-NN and gradient boosting (GB). This model was trained with 35 sessions and validated with the remaining 17. In the testing phase, significant correlations were achieved between the model's predicted values and those obtained by UF: VV (r=0.83), FT (r=0.96), Qave (r=0.70) and Qmax (r=0.69), all with p<0.01. The highlighted limitations include the small sample size, the exclusion of women, the use of a single smartphone model and potential acoustic variability in real-life conditions. Despite these, the study demonstrates the potential of this technology as a portable, non-invasive and accessible tool for remote urological monitoring.

The subsequent study [30] investigated the prediction of variable water jet flows impacting a free surface using sounds generated during liquid-liquid interaction as the source of information. This work provides evidence that acoustic flow prediction is feasible even outside the urological context and lays useful foundations for extrapolation to clinical scenarios such as SU. A total of 51 flow episodes were collected, simultaneously recording the sound with a smartphone and the actual flow using a high-precision gravimetric system. Two approaches were proposed and compared for estimating the flow trajectory: one based on ML using MFCC features extracted from the sounds and another based on acoustic parameters related to spectral energy. The ML model was trained with 70% of the data (over 10,000 audio frames) and validated with the remaining 30%. Volume prediction showed a high correlation with the actual value (r=0.83). The estimated flow trajectories were evaluated using Euclidean and Fréchet distances, demonstrating comparable performance between the two methods, although the ML approach also enabled estimation of absolute volumes. The study's limitations include the need for large training datasets in the ML approach and the restriction of the second method to relative estimates.

A representative study is [15], which evaluated a smartphone-based SU system by comparing it with UF in a cohort of 112 participants (66 men and 46 women). The system records the urination sound and applies prediction models to estimate parameters such as Qmax, Qave and VV. The analysis showed strong correlations between both methods: in men, Pearson Correlation Coefficient (PCC) values were $r = 0.88$, $r = 0.91$ and $r = 0.95$ for Qmax, Qave and VV, respectively; in women, values were $r = 0.78$, $r = 0.93$ and $r = 0.96$. Although Qmax showed statistically significant differences compared to UF, the authors highlight that the acoustic system provides comparable results with the added advantage of enabling home use, which could facilitate urological follow-up without relying on expensive clinical equipment or controlled environments. However, the study does not describe the specific mapping between audio features and flow parameters, nor does it detail the methods used to construct the prediction models. Furthermore, the use of correlation coefficients and mean comparisons implies an association between ground truth and predicted values, but does not offer a proper assessment of model accuracy (such as error metrics or validation procedures) limiting the interpretability and reproducibility of the results.

In [31], an SU technique based on the acoustic energy of the sound generated during urination, captured using a smartphone, was proposed to estimate the urinary flow curve over time. The method was applied to a cohort of 44 healthy men whose voiding was recorded while urinating into a cup with water placed over a conventional uroflowmeter. Based on the sound energy, segmented into time frames of 0.2 to 1 second, a flow curve was generated by scaling the energy values by the ratio of VV to the area under the energy curve. The model's accuracy was high, achieving a correlation of 0.993 and an root mean square error (RMSE) of 2.37 ml/s for 1-second frame duration. However, a major limitation of the approach is that it requires accurate knowledge of the voided VV to perform the sound-to-flow transformation, which restricts its fully automated applicability. Despite these limitations, this study provides a systematic and reproducible methodology that lays the groundwork for the development of more accessible portable urodynamic evaluation solutions.

In [32] conducted a prospective comparative study to assess the feasibility and accuracy of a mobile SU App in male pediatric patients with LUTS. The study included 16 boys with a median age of 9 years, whose voids were recorded using a machine-learning-based app installed on an iPhone. Acoustic recordings of the urine stream were processed to estimate parameters such as Qmax, Qave, FT and VV, which were compared with those obtained using UF. Strong correlations were observed between both methods for Qmax (intraclass correlation coefficient (ICC)=0.755, $p=0.005$), FT (ICC=0.974, $p<0.001$) and VV (ICC=0.930, $p<0.001$), but not for Qave (ICC=0.442, $p=0.135$). Visually, the flow curves generated by both techniques showed good agreement. Main limitations include the small sample size and validation with only one specific smartphone model in a controlled environment. Furthermore, although volume estimation is performed indirectly from the area under the curve, the model relies on a standardized environment that may not be easily replicated in home settings. Nevertheless, this work represents a valuable contribution to extending the clinical use of SU in pediatric populations, highlighting its potential as a complementary tool for home monitoring of voiding dysfunctions.

[33] conducted a clinical validation of the use of a SU App (TeleSono-UroFlow) in adult men, comparing its results with those of UF in an outpatient setting. The study included 44 male patients treated at a urology clinic and three healthy volunteers to assess intra-individual repeatability. Parameters such as Qmax, TQmax, FT

and the shape of the flow curve were compared. Results showed low correlation for Qmax in clinical patients ($r = 0.12$), moderate for TQmax ($r = 0.46$) and very high for FT ($r = 0.91$), with symmetric mean absolute percentage error (SMAPE) of 32.6%, representing 67.4% accuracy in flow shape. In healthy volunteers, Qmax repeatability reached a correlation coefficient of $r = 0.72$, with 82.3% accuracy in curve shape. Study limitations include a high exclusion rate (50% of recordings) due to technical issues, acoustic variability from the environment (microphone position, room acoustics, type of receptacle) and low sensitivity for detecting variations in Qmax, a key parameter for diagnosing urinary tract obstruction. Despite these limitations, the study suggests that SU may be useful for assessing temporal parameters and monitoring voiding patterns in out-of-hospital settings, which is especially relevant in the context of SU.

In [34] conducted a prospective study to validate the accuracy of a mobile app for acoustic VV estimation (proudP®), comparing it to conventional estimation using bladder ultrasound before and after voiding. The study included 49 hospitalized male patients, collecting a total of 245 voiding events. Acoustic volume estimation was based on analyzing the sound generated when urinating directly into the toilet bowl water, while reference estimation was obtained through three-dimensional bladder ultrasound scans. Results showed significant correlation between the two techniques (PCC=0.811, $p < 0.001$), with an average difference of 16 ml and an RMSE of 75.9 ml. Linear regression yielded a determination coefficient of $R^2 = 0.61$, indicating that a substantial proportion of the variability in acoustic estimation can be explained by the reference method. However, key limitations include inter-individual variability in accuracy, influenced by factors such as toilet type, water volume and temperature, patient posture and possible acoustic interference. Moreover, the study was limited to hospitalized patients, which may restrict generalization to home or rural contexts where acoustic conditions and technological access may vary considerably. Despite these limitations, the study demonstrates the viability of acoustic VV measurement as a non-invasive and automatable alternative, helping to reduce the psychological burden of UF.

In [35] developed and validated a DL-based system, Audioflow, to predict uroflowmetry parameters and detect abnormal urinary flow patterns from urination sounds recorded with a smartphone. In an open prospective study, 534 men aged 21 to 80 were recruited at Singapore General Hospital, yielding 331 valid recordings after excluding distorted recordings or those with VV below 150 ml. The algorithm, comprising two deep neural network models, was trained to estimate Qmax, Qave and VV, as well as to classify flows as normal or abnormal based on the assessment of three expert urologists. Correlations between the algorithm's predictions and UF were high according to CCC: 0.77 for Qmax, 0.85 for Qave and 0.84 for VV. Although these values are slightly lower than those reported in previous studies using PCC, the authors argue that CCC provides a more rigorous measure of absolute agreement, not just linear correlation as PCC does. Additionally, the system's ability to identify abnormal flows was notable, achieving an area under the receiver operating characteristic curve (AUC) of 0.892, sensitivity of 87.3% and specificity of 77.5%. Main study limitations include validation only in men, being conducted at a single site and model sensitivity to the acoustic environment (toilet type, materials, patient posture). Nevertheless, this work represents a significant advance in SU, demonstrating that a DL-based approach can offer diagnostic precision comparable to that of human experts and applicability in remote monitoring contexts.

Overall, the reviewed studies demonstrate that using AI techniques applied to voiding acoustic signals allows estimation of urodynamic parameters with acceptable accuracy. These results reflect the clinical potential of SU as a non-invasive and portable tool. However, flow estimation remains a challenge due to its high physiological variability and the sensitivity of sound to external factors such as surface type, patient posture, or acoustic environment.

Despite these advances, common limitations are identified in most studies:

- **Lack of standardization** in recording protocols, devices used and environmental acoustic conditions.
- **Small sample sizes** and limited population representativeness (e.g., studies focused exclusively on men or a single clinical setting).
- **Low methodological transparency**, with little description of algorithms used and training parameters, making replicability difficult.
- **Dependence on auxiliary variables** such as VV, which are not always available in unsupervised contexts.
- **Lack of external or multicenter validations** to assess model robustness under diverse conditions.
- **Absence of public and open-access databases** of voiding acoustic signals, hindering objective comparison between studies, limiting reproducibility and slowing scientific progress in SU.

These limitations highlight the need for more robust, reproducible methodological proposals adapted to real-world clinical and home environments. Table 2.1 provides a comparative summary of the most representative studies in SU, highlighting their main methodological characteristics, evaluation metrics and limitations. The last row summarizes the contributions of this doctoral thesis in relation to the state of the art.

Table 2.1: Comparative summary of key studies in SU

Study	Participants	Devices	Methods	Metrics / Results	Limitations
[16]	25 men (52 sessions)	Phone	MFCC + k-NN/GB ensemble	VV (r^1=0.83), FT (r=0.96), Qave (r=0.70), Qmax (r=0.69)	Small sample, no women, single device, basic metrics
[30]	51 episodes	Phone + gravimetric	MFCC + ML / spectral energy	VV (r=0.83); Euclidean/Fréchet distance	Non-clinical setting, synthetic liquid stream
[15]	112 (66 male/46 female)	Phone	Audio features (unspecified) + ML	Qmax (r=0.88 male/ r=0.78 female), Qave, VV (0.91–0.96)	No method detail, lacks accuracy/error metrics
[31]	44 men	Phone	Energy scaling of sound waveform	RMSE[2]: 2.37 ml/s, r=0.993	Requires known VV, limited automation
[32]	16 boys	Phone	ML model (not detailed)	ICC[3]> 0.75 (Qmax, FT, VV)	Small sample, no generalization, manual environment
[33]	47 men	Phone	Amplitude-based metrics	FT: r=0.91, Qmax: r=0.12 (low)	50% exclusion rate, noisy environment
[34]	49 men (245 events)	Phone + 3D ultrasound	Sound amplitude vs. bladder scan	r=0.81, RMSE = 75.9 ml	Hospital-only, low generalizability
[35]	534 men (331 valid)	Phone	DL	CCC[4] : 0.77–0.85; AUC[5] = 0.89 (abnormal flows)	Single site, men only, acoustic sensitivity, low interpretability (no feature-level analysis)
This Thesis	**Real: 50 men (47 valid x 3 devices); Synthetic: 50 events x 3 devices**	**UM, Watch, Phone**	**ML: RF, SVM, k-NN, DL; impact surface classification; synthetic-to-real validation**	**MAE[6] < 2.6 ml/s (Flow); CCC: Flow (0.9), VV(0.85); Accuracy > 99% (surface type)**	**Male-only, standing posture, recorded quiet environment, limited real-world generalization**

[1] **r**: denotes the Pearson Correlation Coefficient (PCC)
[2] **RMSE**: Root Mean Squared Error
[3] **ICC**: Intraclass Correlation Coefficient
[4] **CCC**: Lin's Concordance Correlation Coefficient
[5] **AUC**: Area Under the ROC Curve
[6] **MAE**: Mean Absolute Error

2.4 Differential contributions of this doctoral thesis

This thesis systematically addresses the main deficiencies identified in the literature on SU, offering innovative solutions from both technical and methodological perspectives. In particular, the following advances stand out:

- **Incorporation of heterogeneous devices** (dedicated microphone, mobile phone, smartwatch) in the data collection process, allowing the evaluation of portability and applicability of different recording devices in SU testing.

- **Creation of a novel dataset of real SU recordings**, obtained from 50 male volunteers using a certified Minze uroflowmeter under controlled conditions. A total of 47 validated test cases were retained and the dataset is available to researchers upon request.

- **Application of supervised regression and classification techniques**, selected and optimized through quantitative validations (MAE, CCC, weighted kappa), surpassing the merely correlational approach of previous studies.
- **Creation and publication of a public and standardized dataset** of simulated voiding sounds, contributing to scientific reproducibility and facilitating objective comparison between AI models.
- **Cross-validation between synthetic and real data**, demonstrating that models pre-trained on simulations can effectively generalize when applied to clinical signals.
- **Automatic classification of the acoustic environment** (water or ceramic), enabling predictive models to adapt to different conditions and eliminating the need to aim specifically at the water, an unrealistic constraint for elderly patients.
- **Analysis of high-frequency bands > 8 kHz** to preserve the patient privacy in sensitive environments, without compromising acoustic prediction accuracy. The 0–8 kHz range encompasses the majority of human speech content [36], so discarding it minimizes the risk of unintentionally recording intelligible verbal information.

Taken together, these contributions not only extend the state of the art in the field of SU, but also provide a clear path toward clinical implementation by combining technical robustness, generalization capability and suitability for real-world settings.

2.5 Conclusion

The review presented in this chapter confirms that SU has gradually evolved in recent years, driven by advances in AI and acoustic processing. While the literature has validated the general feasibility of estimating voiding parameters such as VV and urinary flow, significant limitations remain, including the dependence on ideal acoustic conditions, lack of public datasets and limited standardization of recording and evaluation techniques. Additionally, most studies have focused on male participants voiding while standing, with limited exploration of female subjects or seated voiding, which restricts the generalizability of current findings.

In this context, this doctoral thesis offers a comprehensive and pioneering contribution to the state of the art in SU. Robust models are developed for estimating urodynamic parameters with competitive error metrics and clinically meaningful agreements. The design, validation and creation of a standardized open dataset lay the groundwork for a more cohesive and accelerated research community. Finally, a hybrid workflow (synthetic–real) is validated, enabling the training of versatile models capable of generalizing to real-world conditions, even with low-cost technological resources.

It also addresses fundamental issues that have so far been scarcely explored, such as the automatic classification of the voiding environment (water or ceramic), which removes the assumption of idealized conditions (e.g., urination only into water) and enables the deployment of SU systems in realistic, everyday scenarios.

Overall, these advances bring SU closer to actual clinical and home use, offering a non-invasive, accurate and accessible alternative for the functional evaluation of the lower urinary tract, especially valuable in aging populations or those with limited access to specialized care centers.

Chapter 3

Flow prediction in sound-based uroflowmetry

3.1 Introduction

This chapter presents the first article published as part of the compendium of this doctoral thesis. The study addresses the automatic estimation of urinary flow and VV from acoustic signals recorded during voiding, using ML regression algorithms. The approach is based on SU tests and compares its results with a UF device, the Minze uroflowmeter [18], evaluating its accuracy and feasibility for home applications.

3.2 Reference

Alvarez, M. L., Arjona, L., Jojoa-Acosta, M., & Bahillo, A. (2025). *Flow prediction in sound-based uroflowmetry*. **Scientific Reports**, *15*(1), 643. https://doi.org/10.1038/s41598-024-84978-w

3.3 Status

Table 3.1 shows the publication status of the paper presented in this chapter.

3.4 Summary and contributions

The article presents a methodology to estimate urinary flow and VV based on the analysis of real acoustic signals recorded during voiding. As part of this work, a new dataset of real SU recordings was created, involving the participation of fifty healthy males (18–60 years) recruited specifically for this investigation. Each participant voided into a clinically calibrated conventional Minze uroflowmeter [18], which recorded the flow

Publication status	Published
Journal	Scientific Reports
Publisher	Nature Portfolio (Springer Nature)
Year	2025
JCR Impact Factor (2024)	3.9
JCR Quartile	Q1 in Multidisciplinary Sciences
DOI	https://doi.org/10.1038/s41598-024-84978-w
Citations (as of September 2025)	4 (according to Scopus)

Table 3.1: Publication details for Paper 1

while the sound was simultaneously captured using three different recording devices (UM, Phone and Watch).

Several regression algorithms (RFR, GBR and SVR) were evaluated, along with acoustic feature extraction techniques based on FFT segmented into 1000 ms blocks. It was observed that the 0–8 kHz frequency band contained 83% of the useful information, being sufficient for accurately estimating the flow.

The main results indicate that:

- A novel dataset of real SU recordings was created from 50 healthy male volunteers, with synchronized ground truth provided by a certified Minze uroflowmeter. After excluding noisy or uncertain cases, 47 validated test recordings were retained for analysis. This dataset supports reproducibility and is available to researchers upon request.
- The MAE was 2.86 ml/s using the Watch, with a CCC of 0.9 for flow and 0.85 for VV.
- The Watch was considered the most suitable device due to its comfort and potential for automated home testing, although devices like the UM and Watch showed slightly lower MAE rates.
- The feasibility of applying SU using accessible consumer-grade devices under a non-invasive approach was demonstrated.

3.4.1 Key contributions

- Creation of a validated dataset of 47 real SU recordings with synchronized ground truth from a certified clinical uroflowmeter.
- Direct comparison between SU and UF with cross-validation on 47 valid test cases, selected from an initial cohort of 50 participants after excluding recordings affected by noise or surface uncertainty.

- Evaluation of the impact of time segmentation and audio frequency on model accuracy.
- Replicable methodological proposal including experimental design, signal processing and model training.

3.4.2 Limitations noted by the authors

- Environmental variability (toilet material, acoustic conditions, etc.) was not evaluated.
- The study was limited to male participants, preventing immediate generalization to female populations.
- The absence of public databases with SU sounds in the literature remains a major limitation for the development and benchmarking of AI-based algorithms. This, combined with the high variability in environments and recording devices, hinders the reproducibility of results by other researchers.

3.5 Full article

The final version of the article is included below:

scientific reports

Check for updates

OPEN

Flow prediction in sound-based uroflowmetry

Marcos Lazaro Alvarez[1,3]✉, Laura Arjona[1,3], Mario Jojoa-Acosta[2,3] & Alfonso Bahillo[2,3]

Sound-based uroflowmetry (SU) offers a non-invasive alternative to traditional uroflowmetry (UF) for evaluating lower urinary tract dysfunctions, enabling home-based testing and reducing the need for clinic visits. This study compares SU and UF in estimating urine flow rate and voided volume in 50 male volunteers (aged 18–60), with UF results from a Minze uroflowmeter as the reference standard. Audio signals recorded during voiding were segmented and machine learning algorithms (gradient boosting, random forest, and support vector machine) estimated flow parameters from three devices: Ultramic384k, Mi A1 smartphone, and Oppo smartwatch. The mean absolute error for flow rate estimation were 2.6, 2.5 and 2.9 ml/s, with R^2 values of 84%, 83%, and 79%, respectively. Analysis of the Ultramic384k's frequency range showed that the 0–8 kHz band contained 83% of significant components, suggesting higher sampling frequencies are unnecessary. A 1000 ms segment size was optimal for balancing computational efficiency and accuracy. Lin's concordance coefficients for urine flow and voided volume using the smartwatch (0–8 kHz, 1000 ms) were 0.9 and 0.85, respectively, demonstrating that SU is a reliable, cost-effective alternative to UF for estimating key uroflowmetry parameters, with added patient convenience.

Keywords Acoustic voiding signals, Flow prediction, Machine learning, Sound-based uroflowmetry

The rapid development of information and communication technologies is transforming healthcare systems, becoming remote and more proactive. This evolution brings benefits to both patients, by facilitating their access to medical services, and healthcare providers, who can obtain updated information and resources more quickly and efficiently. As a result, the quality of healthcare improves and associated costs decrease. Lower urinary tract symptoms (LUTS) are a problem affecting over 1900 million people worldwide[1], leading to a decreased quality of life and significant healthcare resource expenditure[2]. The most widely used test to detect possible LUTS issues is UF[3]. UF is a non-invasive test based on estimating urine flow as a function of time, voided volume (VV), and the duration of the voiding process. This test is carried out in a clinic, where the patient must urinate into a uroflowmeter device. However, the accuracy of UF can be limited by various factors, such as an insufficient VV (which should be > 150 ml)[4] or the situational stress experienced by the patient due to the unnatural voiding environment. This process is often performed on demand, frequently with a low or very high bladder filling, rather than urinating when the patient feels physiologically ready[5]. Additionally, flow rates vary significantly throughout the day[6], and a single test may not be representative of a patient's regular daily voiding patterns. This leads to non-reproducible flow measurements[7], resulting in biased contributions from this test toward diagnosis[8]. Therefore, multiple tests are recommended, leading to numerous prolonged and costly visits for both the clinic and the patient[1]. To solve the challenges associated with UF, SU emerges to enable monitoring the void in a natural and comfortable environment for the patient, such as at home. This test seeks to estimate the flow patterns from the sound generated by urine hitting the water surface in the toilet bowl. Previous studies have shown a good correlation between the sound-derived flow parameters and those generated by UF[8,9]. To collect the audio signal for the SU tests, various hardware devices have been used in the literature. Smartphones are the most commonly used devices for performing SU tests due to their great versatility[10–12]. Other works have used professional microphones[13] and smartwatches[14,15] for the same purpose.

Recent studies have confirmed the potential of SU in estimating flow parameters, with strong correlations between SU and UF tests. In[8], a moderate correlation with conventional uroflowmetry was reported, with a Pearson's correlation coefficient of 0.38, 0.57, and 0.68 for maximum flow rate (Qmax), average flow rate (Qave), and VV, respectively. In[10], the authors reported strong correlations between Qmax, Qave, and VV of 0.88, 0.91, and 0.95 among men, and 0.78, 0.93, and 0.96 among women, respectively, for flow curve patterns obtained with their SU method and those obtained with a conventional UF device. However, the methods used to estimate flow parameters were not fully described and cannot be replicated. In[16], the estimation of sound flow parameters was

[1]Faculty of Engineering, University of Deusto, 48007 Bilbao, Spain. [2]Department of Signal Theory and Communications, Universidad de Valladolid, 47002 Valladolid, Spain. [3]Marcos Lazaro Alvarez, Laura Arjona, Mario Jojoa-Acosta, and Alfonso Bahillo have contributed equally to this work. ✉email: alvarez.marcoslazaro@deusto.es

analysed. However, the methodology requires one to know the VV to obtain accurate estimates,which seems not feasible if we want to perform these tests at home. Finally, an analysis of the correlation of voiding characteristics was performed in[12] using deep learning on SU audio recordings. As an evaluation metric, the Lin concordance coefficient was used instead of the Pearson correlation coefficient applied in[10]. Although Lin's concordance coefficients were 0.77, 0.85, and 0.84 for Qmax, Qave, and VV respectively, lower than the corresponding Pearson coefficients obtained in[10], the authors argue that Lin's concordance coefficient is more appropriate as an evaluation metric since it describes the degree of agreement between two measurements rather than the linear relationship given by Pearson.

The main objective of this research is to evaluate and estimate urine flow from SU tests using state-of-the-art machine learning (ML) regression algorithms. The comparative evaluation study considers different frequency bands in the range (0–96 kHz) and recording devices such as a Dodotronic Ultramic384, an Oppo smartwatch, and a Xiaomi Mi A1 smartphone. Furthermore, we used a professional Minze uroflowmeter[17] which allows us to obtain the actual flow for each time instant of the recorded audio signal. The Minze uroflowmeter, which serves as the ground truth, is factory calibrated and was purchased in November 2022. This device is a certified medical instrument that is compliant with the ISO 13485:2016 standard, ensuring highly accurate flow rate measurements. The data collected from this flowmeter was utilized to train the ML model.

The paper is organised as follows: "Related work" section briefly reviews the state of the art in audio feature extraction and flow prediction from SU audios using ML; "Patients and methods" section presents the materials and methods proposed in this research, describing the study design and the population, the characteristics of the datasets, and the procedures and theoretical foundations followed in the analysis of flow prediction in SU tests using different recording devices; "Results" section shows the results obtained from the proposed methodology; and finally, "Conclusions" section provides some concluding remarks.

Related work

Feature extraction in audio signals

There are multiple techniques in the literature to extract features from audio signals to be used in artificial intelligence models. Feature extraction techniques span several domains. The features of the temporal, frequency, cepstral, wavelet, and time-frequency domains have been explored for different types of audio signals, including speech, music, and environmental sounds[18,19]. The integration of modern ML algorithms with audio signal processing techniques has led to significant advances in audio classification tasks[18]. Studies have investigated the performance of different deep learning models using various audio features, such as the Mel Spectrogram and mel-frequency cepstral coefficients (MFCCs), both independently and in combination with ensembles[20]. The choice of features depends on both the dataset and the model, and feature combinations are generally restricted to those that perform well individually[19].

In our research, we used the N linear-binned FFT of the audio signal to be analyzed, where N is an integer. The audio signal is processed using the Fast Fourier Transform (FFT) which is organised into a specific number of linearly distributed segments (or "bins"). Compared to other audio feature extraction techniques, such as cepstral analysis or MFCCs, N linear-binned FFT can be more intuitive and straightforward in its implementation. Furthermore, when combined with traditional ML regression methods, it offers a robust and efficient solution to extract and use audio features in our flow estimation analysis in SU tests with different recording devices.

ML algorithms in SU

To estimate the voiding flow from SU tests, many algorithms have been implemented in the literature. Finding the optimal model and selecting the most suitable features from various audio segments can be a complex process. The proposed methods range from traditional signal processing techniques to more recent approaches utilizing deep learning. In[11], ML was used with k-nearest neighbours and gradient boosting, and the model was trained using the MFCCs. In[12], a deep neural network (DNN) with three hidden layers was used, taking spectral centroids, chroma vectors, and mel-frequency cepstral coefficients (MFCCs) as input features[21,22]. In[23], an long short-term memory (LSTM) model was used for a time series prediction, with loudness representing the magnitude of the urinary sound and roughness representing the signal change pattern as inputs.

Patients and methods

Study design and population

We recruited 50 male volunteers without urological comorbidities, between the ages of 18 and 60, who agreed to participate in the study. The average height of the participants was found to be 175 cm, with a standard deviation of 4 cm. All participants provided their informed consent in writing for the use of conventional and acoustic uroflowmetry data. This study was approved by the Valladolid East Health Area Medicine Research Ethics Committee on 27 July 2023 with reference PI-GR-23-3275 (minutes number 16/2023). The Ethics Committee mentioned above complies with the GCP standards (CPMP/ICH/135/95).

Procedures

The test consisted of urinating in a Minze uroflowmeter basin, while three different microphones recorded the sound. The Minze uroflowmeter basin, made of plastic, had been prefilled with 400 ml of water at the bottom of the container to ensure that the sound generated by the urination was against the water, simulating the conditions of a real toilet bowl. The Minze uroflowmeter has a resolution of 1 ml/s, with an accuracy of ± 2.5 ml/s for flow and ± 30 ml for VV[17]. The sampling rate is 10 Hz and records the flow curve, VV, Qmax, Qave, time to maximum flow and voiding time. Figure 1 shows an example of the data provided by Minze software during a test.

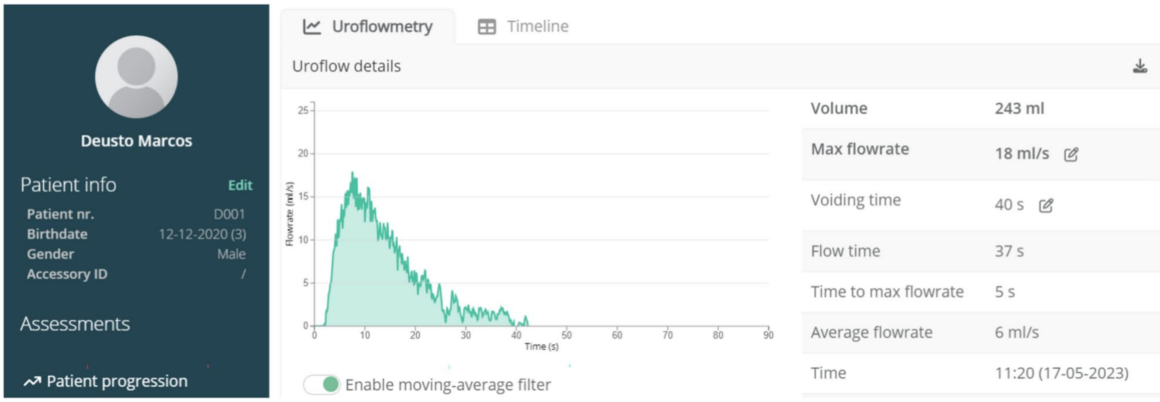


Fig. 1. Example of data recorded by the Minze uroflowmeter software during a test.

All participants were given instructions immediately before the test, asking them to aim at the toilet water as accurately as possible. After completing the test, they were asked if they had been able to meet this requirement, and all affirmative responses were subsequently validated by reviewing the audio recordings. Finally, recordings containing background noise, or those where there was uncertainty regarding time intervals in which urine may have hit the walls of the basin, were excluded from the analysis.

The entire voiding process on the Minze uroflowmeter was recorded using three sound recording devices:

- *Ultramic384 (UM)*: a high-quality microphone (FG23629 microphone sensor from Knowles), used in studies to record void events SU[14] with a sampling rate (SR) of 384 kHz, allowing the study of a wide frequency spectrum. For our tests, we used a SR of 192 kHz, because tests showed that there was no information above the 96 kHz band. This device is not intuitive to use as it requires additional hardware components to operate. Besides, it is not as versatile compared to smartwatches and smartphones (the device has no other use beyond audio recording).
- *Mi A1 smartphone (Phone)*: it integrates a medium-quality microphone (SPU0410LR5H-QB microphone sensor from Knowles), with a SR of 44.1 kHz, and it has also been used in similar studies. This device is intuitive to use, although not as versatile during urination as a smartwatch[24].
- *Oppo smartwatch (Watch)*: it integrates a medium-quality microphone (chipset details not publicly disclosed), validated for use in SU applications[14], with a SR of 44.1 kHz. It is the most intuitive and versatile recording device for the urination process.During the test, the Watch and the UM were placed opposite the user at a height of 80 cm above the basin water level, that is the average height of a toilet cistern. The test participants must wear the smartwatch on the wrist, pointing its microphone to the basin during the recording. Figure 2 shows the laboratory environment set up in the bathroom where the tests were performed. The bathroom had dimensions of 404 cm (length) × 175 cm (width) × 271 cm (height), with ceramic tiles covering the floor and walls, and a plasterboard ceiling.

Dataset description

For each performed test, we obtained three audio files in WAV format, corresponding to each of the recording devices. For each test, the corresponding uroflowmetry curve provided by Minze software, containing the flow information for each time unit (100 ms resolution), was matched to these audio files.

After excluding the test that presented the issues mentioned in "Patients and methods" section, we obtained a total of 47 valid tests. For each test, we obtained the corresponding flow values with a frequency of 10 Hz for further analysis.

Correlation between SU signal envelopes and Minze measurements

To study the sound signals' waveform and compare it with the flow graphs provided by Minze software, we obtained the sound envelope from each one of the recorded audios following the methodology proposed by[14]. We then evaluated the linear relationship between the flow envelope given by Minze software and the sound signal envelope of the three recording devices. We used the Pearson correlation coefficient to analyze the 47 audio signals recorded by each of the devices. We selected Pearson's correlation coefficient because we are interested in the linear relationship rather than the level of similarity. Figure 3 shows the result of evaluating the envelope correlation using a box plot. It can be seen that the interquartile range (IQR) for the three devices is very close to the median and mean and is it relatively narrow. This suggests that the correlations are consistently high and that there is little variability in the correlation between the waveform provided by Minze and the envelope of the sound signal for each device, supporting the idea of estimating flow from sound in SU tests. Figure 4 shows a random selection of eight tests in which a very high correlation is demonstrated, further supporting the idea of estimating flow from sound using SU tests.

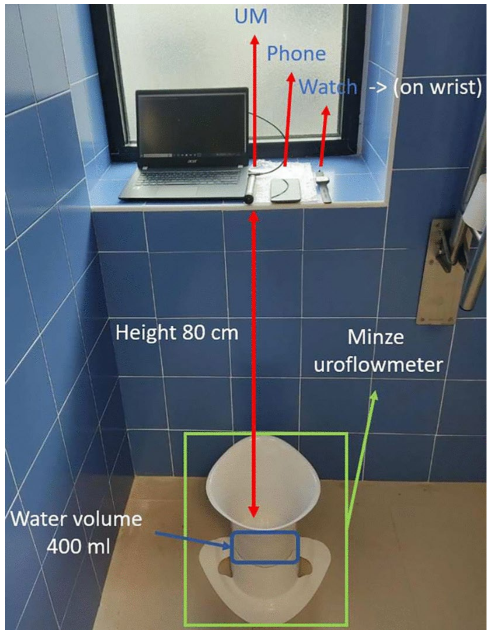


Fig. 2. Laboratory data collection scenario showing the Minze uroflowmeter with a water volume of 400 ml in the basin and the three recording devices: UM, Phone, and Watch, along with their respective heights.

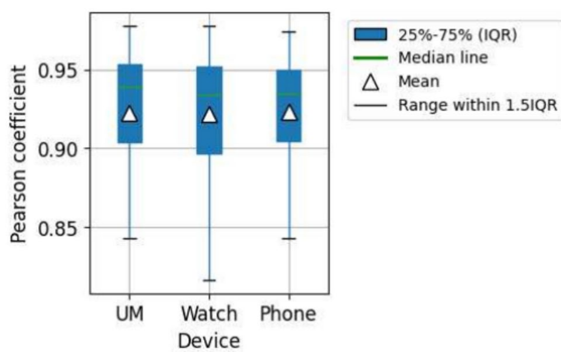


Fig. 3. Results of evaluating Pearson's correlation for the three devices to compare the envelop extraction.

General research diagram

Figure 5 shows a graphical diagram of the proposed methodology for analyzing SU audios to predict the voiding flow rate. Our input data are the SU audio signals in WAV format recorded with the three devices (see "Patients and methods" section). Each audio is associated with its corresponding flow data point provided by the Minze software. To analyse the audio signals, the first step was to synchronize all the signals with the data given by Minze. The audio signals were then split into 100-ms segments and each segment was labelled with the flow value provided by the Minze software. We selected a segment size of 100 ms because the Minze software provides flow data points sampled at 10 Hz. As a result of the audio segmentation, we obtained 13,060 100-millisecond-long segments of SU labelled with their corresponding flow value.

To estimate the flow from sound, we used traditional ML regression algorithms. The input features for training the algorithms were frequency domain features. We utilized the N linear-binned fast fourier transform (FFT), where the frequency range is divided into N equally spaced intervals. For each interval, we sum the absolute values of the amplitudes of the components present in each interval, obtaining a vector with N values that characterize each audio segment. Linear-binned FFT is a widely used technique to extract the most relevant features of the frequency spectrum of an audio signal and has demonstrated effectiveness for audio analysis. For our experiments, we selected N = 20, a value chosen based on experimentation with various values of N.

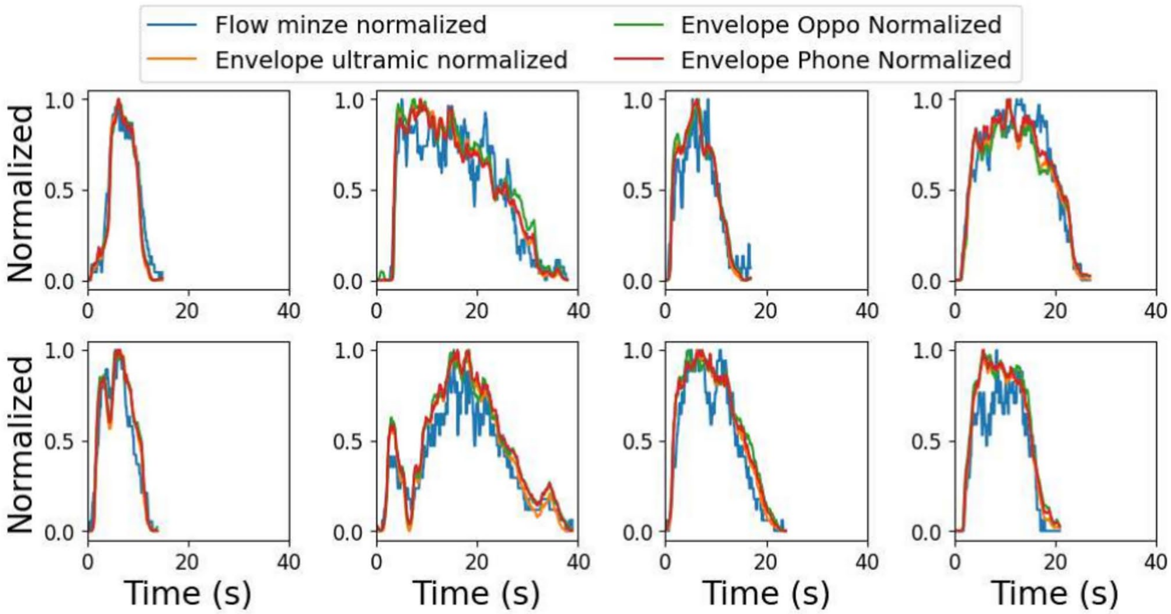


Fig. 4. Comparison of Minze flow data with the sound envelope of the signal given by UM, Watch, and Phone, for a selection of eight randomly selected signals.

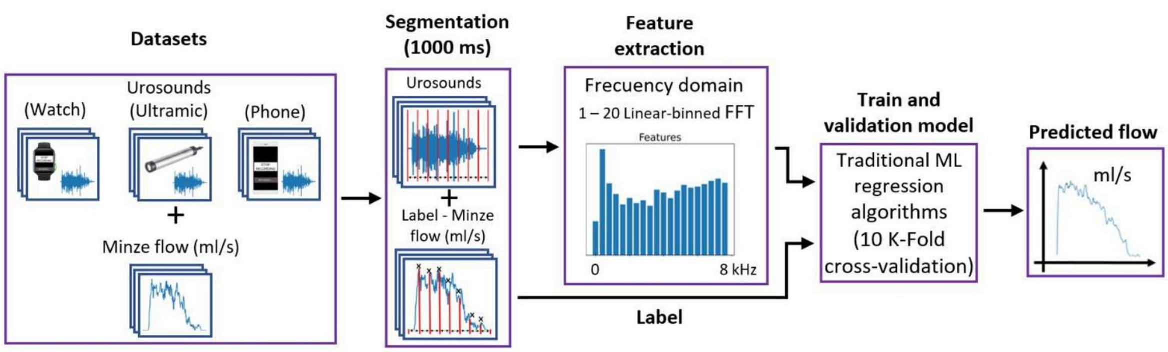


Fig. 5. Diagram showing the pipeline of the proposed methodology, consisting of 4 main steps: data extraction, audios segmentation, feature extraction, and finally model training and validation.

The performance of the regression models was evaluated using tenfold cross-validation. In this process, the dataset is divided into 10 subsets; in each iteration, one subset is used as the test set, while the other nine are used for training. The model's performance is assessed on the test set for each fold, and the final results are reported as the average error across all tenfolds. This ensures that all data is used for both training and testing, but never simultaneously, providing an unbiased estimate of the model's predictive capability.

To predict the flow rate of the labelled audio segments, we performed an analysis using three ML regression algorithms, with the aim to evaluate which one performs best in the flow prediction task. Below are the details of the algorithms used:

- *Random forest regressor (RF)*: is widely used in various fields such as healthcare, where the prediction of continuous values is essential. It is a versatile and powerful algorithm known for its effectiveness in a variety of regression tasks.
- *Support vector regressor (SVR)*: is suitable for tasks where capturing non-linear relationships is essential, and it can be a powerful tool in various regression applications, including time-series analysis.
- *Gradient boosting regressor (GBR)*: is widely used in practice for tasks such as time-series forecasting and various regression applications where accurate predictions are essential.

Figure 6 shows the evaluation results of different regression algorithms in terms of their MAE for each one of the recording devices, with a segment duration of 100 ms. The MAE is defined as the average of the absolute differences between the predicted and true flow values, computed across all labeled audio segments, regardless of the flow magnitude. To train our ML algorithms, we used tenfold cross-validation due to its multiple advantages in terms of model performance evaluation. We used a value of N = 13,060, which corresponds to the number of labelled 100-ms audio segments. It can be observed that, in general, the three algorithms show similar results, with RF performing slightly better.

For our remaining analysis, we used RF because it presents a compromise between accuracy and implementation complexity.

Feature selection

The first step to train our regression model is to select the best procedure for characterizing each audio sample in the dataset. First, we performed an analysis across the entire frequency band recorded by the professional UM microphone (0–96 kHz) to identify which frequency components have the most influence on the flow prediction. For this purpose, we extracted 1000 linear-binned FFT samples for each 100-ms audio segment. As a result, we obtained a vector with 1000 values that characterize each audio segment. We trained a supervised RF algorithm using the 1000 linear-binned FFT samples for each labelled audio segment with the corresponding flow label provided by Minze. Subsequently, from the trained RF model, we obtained a Gini impurity-based metric[25] to measure the quality of our split criteria. This metric allows us to quantify the weighted impurity of each feature in the tree, indicating its relative importance in the model. Furthermore, the Gini impurity metric provides us with an effective way to interpret the model and understand which frequencies are most influential in the predictions.

Figure 7 shows that the frequency components that contribute the most information to the prediction of flow from sound are in the lower band of the signal spectrum. The upper part of the spectrum above 8 kHz is not relevant for our application. For the remaining analyses, we used the (0–8 kHz) band, which contains 83% of the frequency components that contribute the most information to the model, and is the band in which the majority of commercial devices record.

Analysis of different audio segment sizes

Once we analyzed in "Patients and methods" section the frequency bands that have a higher influence on the model prediction results, we then performed an analysis to study the influence of the audio segment size taking the 0–8 kHz frequency band. The Minze uroflowmeter used has a SR of 10 Hz, so the maximum resolution we can achieve for the audio signals is a segment duration of 100 ms.

For each recording device, we analyzed eight different audio segment sizes: 100, 200, 500, 800, 1000, 1100, 1200, and 1500 ms. For segments longer than 100 ms, we took the maximum flow value provided by Minze for the corresponding timestamp as the flow label. We selected this value because, according to urologists, the most valuable information in a uroflowmetry test is the maximum flow reached by the patient at any given moment. For this study, we used a RF model taking as input features the 20 linear-binned FFT samples for each audio segment.

Figure 8 shows the evaluation results in terms of MAE. Across all devices, the segment size that yielded the lowest MAE, indicating the best performance, was 1000 ms. The audio recordings from the Phone and UM showed similar and lower errors compared to those from the Watch device.

Despite the Watch exhibiting slightly higher errors than the other two devices (on average 0.37 ml/s higher than the phone and 0.30 ml/s higher than the UM), it was selected for further analysis. This decision was driven

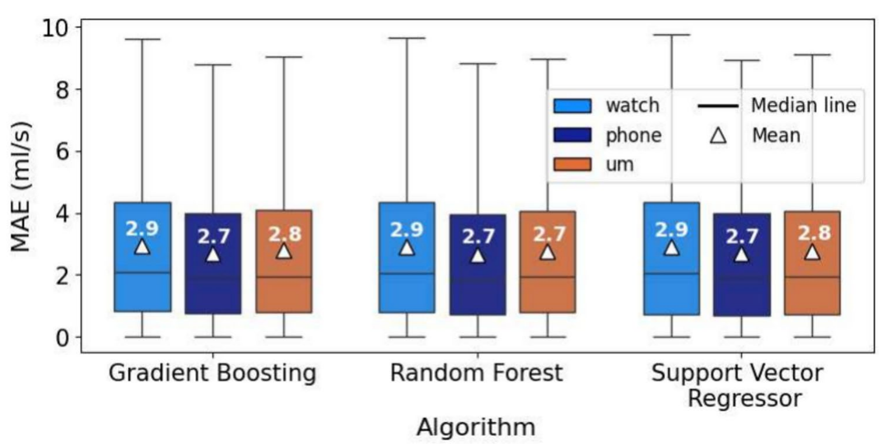


Fig. 6. Evaluation results of the three regression algorithms for each recording device, in terms of the MAE, measured in ml/s. The MAE is the mean of the absolute differences between the predicted and actual flow values, calculated across all labeled audio segments.

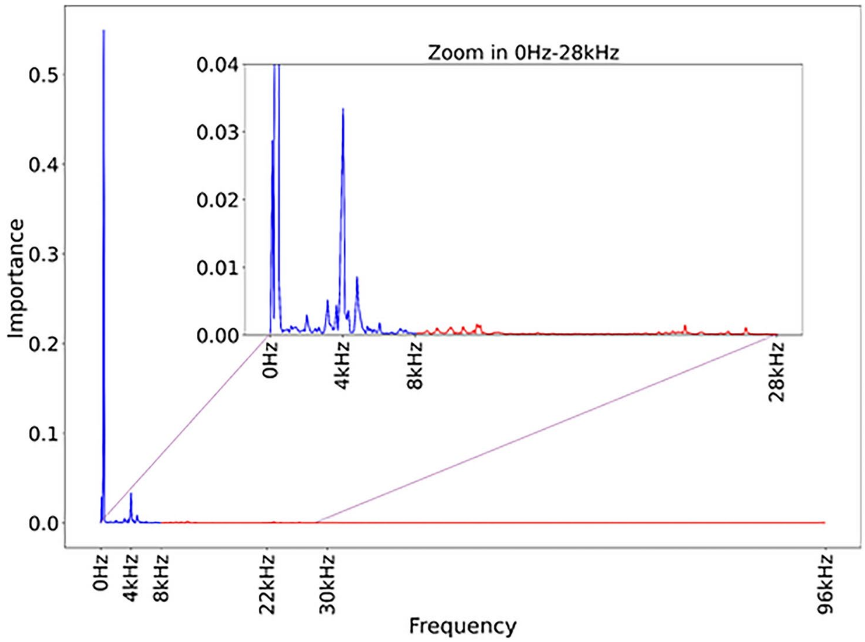


Fig. 7. Predictive power (importance) of each frequency component in the flow prediction task from SU. The frequency band selected in our algorithms is shown in blue, showing the highest values of importance. Importance is calculated using the Gini impurity metric with an RF model.

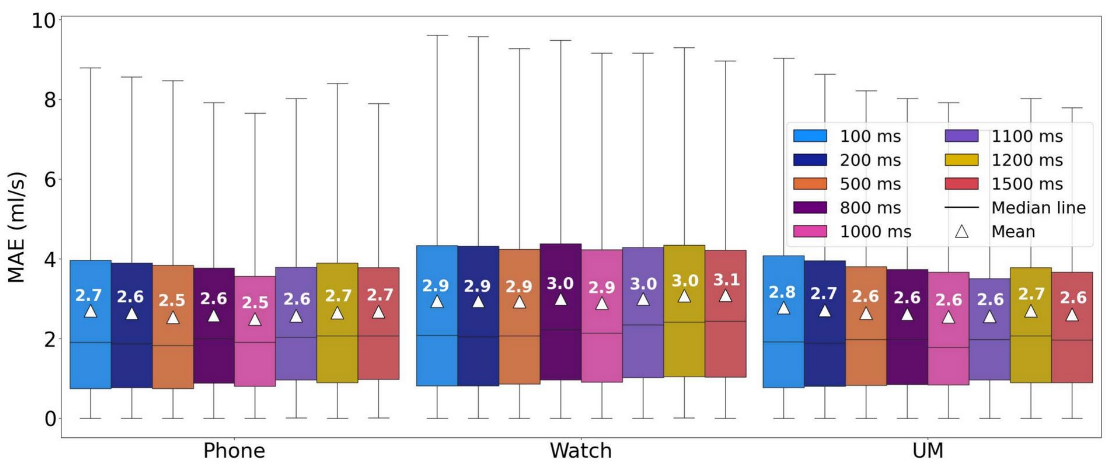


Fig. 8. Analysis of the MAE for the RF prediction model, comparing different audio segment sizes (ms) and the three different recording devices. The MAE is the mean of the absolute differences between the predicted and actual flow values, calculated across all labeled audio segments.

by Watch's versatility and ease of use within the SU context, which offsets the slight increase in error. Figure 8 showed that the error increase of the Watch compared to the other devices is not significant.

Frequency analysis

Once we have determined the audio segments size that yields the best prediction results, we then analyse the frequency range within the (0–96 kHz) band that also obtains the best flow estimation results. This study is performed using the audios recorded with the UM with a audio segment size of 1000 ms. We train a RF model utilising the 20 linear-binned FFT as input features.

To carry out this analysis, we considered the following points:

- We eliminated the low-frequency noise bands (0–250 Hz) and (0–1200 Hz). These frequency ranges have been adopted from related works[14,16], respectively.
- To study the flow prediction results in environments where the user privacy preservation is required, we considered the removal of three frequency bands: (0–4 kHz) and (0–8 kHz) that correspond to the conversational band, and (0–16 kHz) that corresponds to the human audible range.
- An analysis was performed within the (0–8 kHz) band because it contained the most information, as shown in the analysis performed in "Patients and methods" section. In addition, this frequency range is used by the vast majority of commercial devices. This selection represents a compromise between the performance of the model and the cost and availability of the microphone.

Figure 9 shows the results of evaluating the RF model for 18 different frequency bands, from 0 to 96 kHz. The best results were obtained for the (0–8 kHz) and (0–5 kHz) bands, demonstrating that most of the prediction power for the flow prediction task is found in the lower end of the spectrum. It can be observed that when the speech band (0–4 kHz) is removed, the error increases. However, it could be an alternative, despite being less accurate, for applications where preserving privacy is crucial.

Results

Flow estimation with selected parameters

The voiding flow is the variable that provides the most valuable information for urologists from a flowmetry test. To predict the flow using audio signals, in this section we train a new RF model using the parameter settings selected in the analyses performed in the previous sections.

We trained a RF model using audio signals recorded from the Watch, segmented into 1000-ms segments, within the band of 0–8 kHz. As features, we used the 20 linear-binned FFT. To mitigate the effect of overfitting, we used k-fold cross-validation, as it provides a robust and reliable estimate of a model's performance on unseen data. For our algorithms, we selected k = 10 to achieve a reasonable balance between bias and variance in the performance estimate. Figure 10 presents the results obtained with a MAE of 2.86 ml/s. We can consider that the error obtained by our algorithm is significantly low, since our reference device, the Minze uroflowmeter, according to the manual introduces a base error of ± 2.5 ml/s. To assess the correlation between the predicted and actual values, we applied Lin's concordance correlation coefficient for each 1000-ms segment, yielding a value of 0.9. This high coefficient suggests a strong positive relationship, indicating that, as voiding flow increases, the model's predictions also increase in a manner consistent with the field measurements.

Voiding volume estimation with selected parameters

Another important parameter for urologists to evaluate the urinary tract is the VV. In our dataset, we also have the VV associated with each of the audio recordings, obtained from the Minze uroflowmeter. To estimate the VV for each test, we sum the estimated flows corresponding to each audio signal from the SU tests, as obtained in "Results" section. We evaluated Lin's concordance correlation coefficient between the predicted and actual

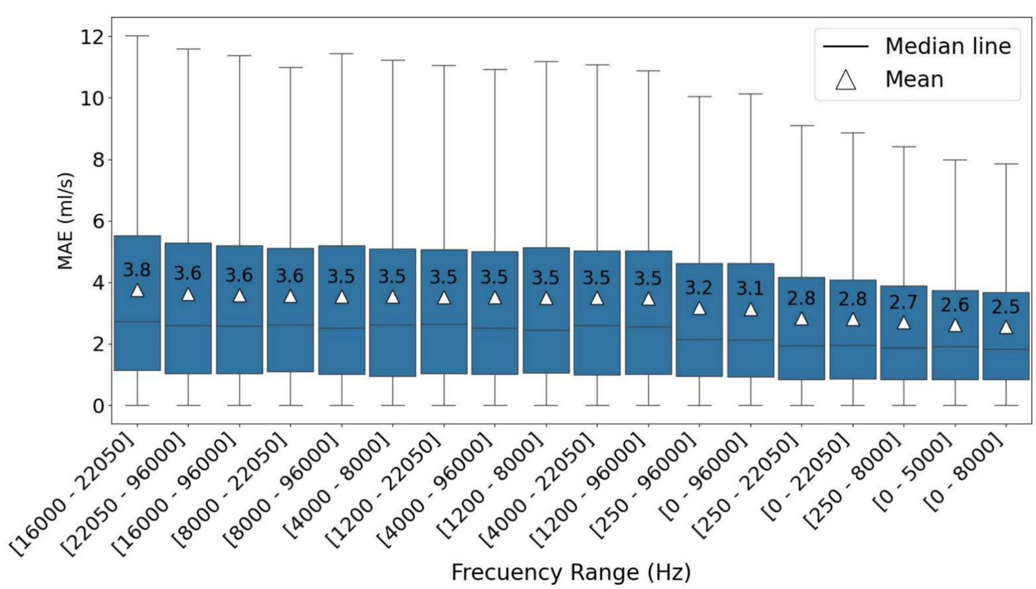


Fig. 9. Analysis of the RF model MAE value for different frequency bands, using the UM microphone. The MAE is the mean of the absolute differences between the predicted and actual flow values, calculated across all labeled audio segments.

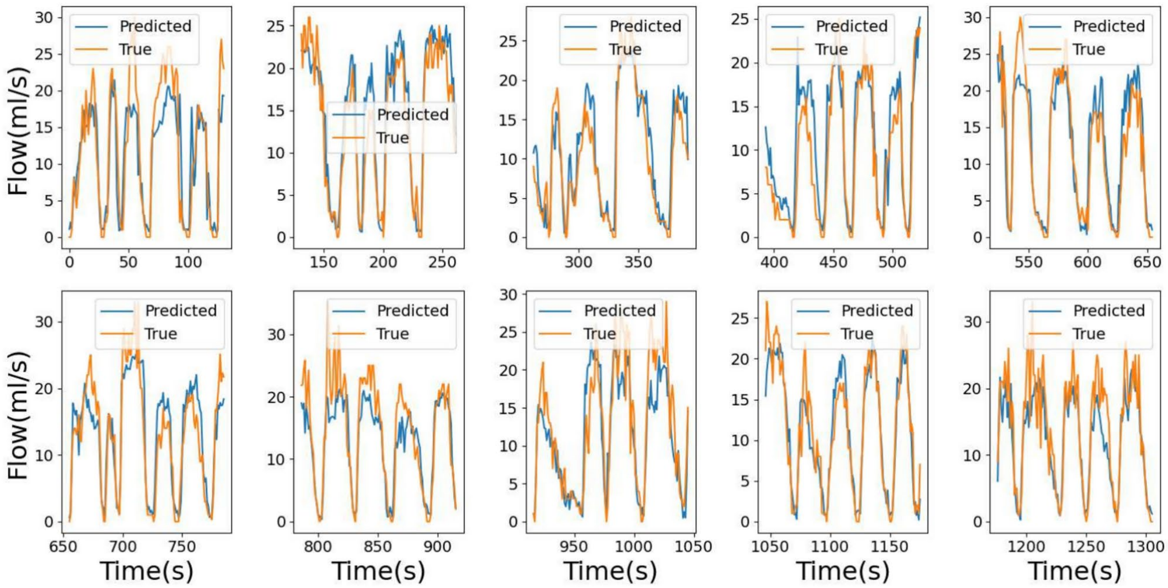


Fig. 10. Comparison of the UF (orange) and SF (blue) flow curves. To obtain the SF curves, we selected the Watch audios, and the model selected was a RF with 20 linear-binned FFT as input features, taking segment sizes of 1000 ms.

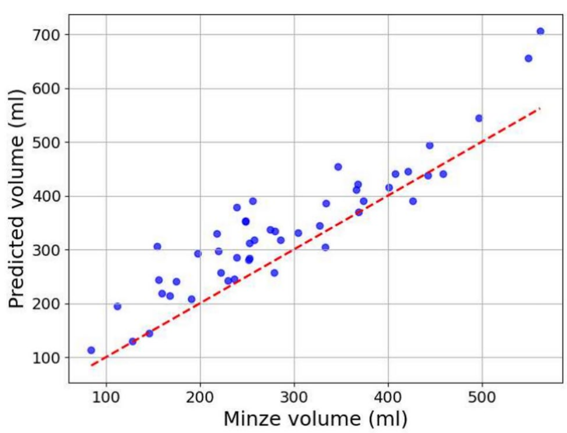


Fig. 11. Comparison between the VV given by Minze flowmeter and the predicted volume, calculated as the sum of the estimated flows corresponding to each audio signal from the SU tests.

volume values and obtained a value of 0.85. Figure 11 shows the result of the comparison between the estimated and actual volumes provided by Minze uroflowmeter.

Conclusions

The comparative results to conduct SU tests using three different recording devices showed similar performance for the Watch and UM, and slightly lower performance for the Watch, for the task of estimating the flow rate. However, the choice of the Watch is based on its ease of use and versatility for conducting UF tests at home. Also, the estimation error difference across the three devices is not significant. Moreover, smartwatches, being wearable objects unlike the other microphones, are more suitable for these types of tests because:

- They are very comfortable for the user as they do not interfere with the act of voiding, can be activated automatically without requiring any action from the patient (especially useful for individuals with low digital skills, such as the elderly and children).

- Since they have a fixed position on the user's body, they allow for sound recording from a consistent distance from where the sound is produced.
- They can be used continuously, making it possible to maintain a voiding diary that measures multiple flows at different times of the day and night, which could be considerably more useful and objective for determining any pathophysiology.

These characteristics position the Watch as a more versatile and user-friendly device for extended home monitoring compared to handheld alternatives such as the Watch, while providing significantly high accuracy to estimate the flow rate.

Contrarily, the UM microphone is not a wearable device, requires additional data collection equipment, is not user-friendly, is bulky, and is not widely distributed, which could limit its use in SU. Its use is only justified if one seeks to obtain information in frequencies above 22 kHz (ultrasound applications), which is particularly beneficial for applications where it is necessary to preserve privacy by eliminating the audible human band.

Limitations

- *Environmental variability* The generalization of the algorithm across different environments has not been evaluated. Variations in environment, as well as differences in toilet material, size, and shape, could potentially affect the acoustics of urination event. To address these variations, additional data collection in diverse settings is necessary to refine and calibrate the algorithm for different conditions.
- *Gender differences* Our study primarily enrolled male participants due to the differences in voiding habits between men and women. Factors such as standing versus sitting position influence how the urine stream impacts the toilet bowl, thereby altering the acoustics of the urination process. As part of future work, we intend to collect a new dataset that includes female volunteers, aiming to enhance the model's applicability across diverse patient demographics.

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

Received: 21 October 2024; Accepted: 30 December 2024
Published online: 03 January 2025

References
1. Alothmany, N. et al. Critical review of uroflowmetry methods. *J. Med. Biol. Eng.* **38**, 685–696 (2018).
2. Arjona, M. F. & Sanz, I. P. Hiperplasia benigna de próstata: Una afección de elevada prevalencia en el paciente de edad avanzada. *Rev. Española Geriatr. Gerontol.* **43**, 44–51 (2008).
3. Sorel, M. R., Reitsma, H. J., Rosier, P. F., Bosch, R. J. & de Kort, L. M. Uroflowmetry in healthy women: A systematic review. *Neurourol. Urodyn.* **36**, 953–959 (2017).
4. Rodrigues Pessoa, R. & Kim, F. *Urodynamics and Voiding Dysfunction* 7th edn, 452–454 (Abernathy's Surgical Secrets, 2018).
5. Kuoch, K. L., Meyer, D., Austin, D. W. & Knowles, S. R. Classification and differentiation of bladder and bowel related anxieties: A socio-cognitive exploration. *Curr. Psychol.* **40**, 4004–4011 (2021).
6. Golomb, J., Lindner, A., Siegel, Y. & Korczak, D. Variability and circadian changes in home uroflowmetry in patients with benign prostatic hyperplasia compared to normal controls. *J. Urol.* **147**, 1044–1047 (1992).
7. Sonke, G. S. et al. A method for estimating within-patient variability in maximal urinary flow rate adjusted for voided volume. *Urology* **59**, 368–372 (2002).
8. Krhut, J. et al. Comparison between uroflowmetry and sonouroflowmetry in recording of urinary flow in healthy men. *Int. J. Urol.* **22**, 761–765 (2015).
9. GÄrtner, M. et al. Evaluation of voiding parameters in healthy women using sound analysis. *LUTS Lower Urin. Tract Symptoms* **10**, 12–16 (2018).
10. Lee, Y. J., Kim, M. M., Song, S. H. & Lee, S. A novel mobile acoustic uroflowmetry: Comparison with contemporary uroflowmetry. *Int. Neurourol. J.* **25**, 150 (2021).
11. Aslim, E. J. et al. Pilot study for the comparison of machine-learning augmented audio-uroflowmetry with standard uroflowmetry in healthy men. *BMJ Innov.* **6**, bmjinnov-2019-000382 (2020).
12. Lee, H. J. et al. Development and validation of a deep learning system for sound-based prediction of urinary flow. *Eur. Urol. Focus* **9**, 209–215 (2023).
13. Alvarez, M. L., Arjona, L., Iglesias Martínez, M. E. & Bahillo, A. Automatic classification of the physical surface in sound uroflowmetry using machine learning methods. *EURASIP J. Audio Speech Music Process.* **2024**, 12 (2024).
14. Arjona, L., Díez, L. E., Bahillo, A. & Arruza-Echevarría, A. Urosound: A smartwatch-based platform to perform non-intrusive sound-based uroflowmetry. *IEEE J. Biomed. Health Inform.* **27**, 2166–2177 (2022).
15. Narayanswamy, G., Arjona, L., Díez, L. E., Bahillo, A. & Patel, S. Automatic classification of audio uroflowmetry with a smartwatch. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* 4325–4329 (IEEE, 2022).
16. El Helou, E. et al. Mobile sonouroflowmetry using voiding sound and volume. *Sci. Rep.* **11**, 11250 (2021).
17. Homepage - Minze Health, accessed on 28 October 2024; https://minzehealth.com/ (2024).
18. Sharma, G., Umapathy, K. & Krishnan, S. Trends in audio signal feature extraction methods. *Appl. Acoust.* **158**, 107020 (2020).
19. Alías, F., Socoró, J. C. & Sevillano, X. A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. *Appl. Sci.* **6**, 143 (2016).
20. Turab, M., Kumar, T., Bendechache, M. & Saber, T. *Investigating multi-feature selection and ensembling for audio classification*. arXiv preprint arXiv:2206.07511 (2022).
21. Piczak, K. J. Environmental sound classification with convolutional neural networks. In *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)* 1–6 (IEEE, 2015).
22. Balamurali, B. et al. Deep neural network-based respiratory pathology classification using cough sounds. *Sensors* **21**, 5555 (2021).
23. Jin, J., et al. Prediction of urinal flowrate by using urinary acoustic signals based on LSTM neural network. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* 5807–5810, vol. 261 (Institute of Noise Control Engineering, 2020).

24. Castillo, Y., Cámara, M. A., Blanco-Almazán, D. & Jané, R. Characterization of microphones for snoring and breathing events analysis in mhealth. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 1547–1550 (IEEE, 2017).
25. Iravantchi, Y., Ahuja, K., Goel, M., Harrison, C. & Sample, A. Privacymic: Utilizing inaudible frequencies for privacy preserving daily activity recognition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* 1–13 (2021).

Acknowledgements
This research was supported by the Spanish Ministry of Science and Innovation under SWALU project (ref. CPP2022-010045) and 'Ayuda para contratos predoctorales 2020 (ref. PRE2020-095612)' funded by MICIU/AEI /10.13039/501100011033 and co-financed by FSE invierte en tu futuro. Additionally, partial support was provided by the Ministry under the Aginplace project (ref. PID2023-146254OB-C41 and ref. PID2023-146254OA-C44).

Author contributions
M.L.A., A.B., and L.A. conceived the study. M.L.A. and L.A. designed the experiments, while M.J.-A. developed the theoretical framework. M.L.A. performed the computations and, together with A.B., contributed to sample preparation and data collection. All authors participated in the analysis and interpretation of the results, and contributed to the writing and revision of the manuscript. M.L.A. drafted the manuscript, incorporating feedback from all authors.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to M.L.A.

Reprints and permissions information is available at www.nature.com/reprints.

Chapter 4

Annotated dataset of simulated voiding sound for urine flow estimation

4.1 Introduction

This chapter presents the second article published as part of the compendium of this thesis. The study describes the creation and validation of a dataset of realistically simulated labeled urinary flow sounds in a real bathroom, aimed at supporting research on urinary flow estimation using AI applied to SU. This work addresses one of the main limitations in the SU literature: the lack of public and standardized datasets that enable the development and objective comparison of prediction models for SU testing.

4.2 Reference

Alvarez, M. L., Arjona, L., Bahillo, A., & Bernardo-Seisdedos, G. (2025). *Annotated dataset of simulated voiding sound for urine flow estimation*. **Scientific Data**, *12*(1), 1–7. https://doi.org/10.1038/s41597-025-05358-1

Dataset available at: https://doi.org/10.6084/m9.figshare.27606642.v1 [37]

4.3 Status

Table 4.1 shows the publication status of the paper presented in this chapter.

Publication status	Published
Journal	Scientific Data
Publisher	Nature Portfolio (Springer Nature)
Year	2025
JCR Impact Factor (2024)	6.4
JCR Quartile	Q1 in Multidisciplinary Sciences
DOI	https://doi.org/10.1038/s41597-025-05358-1
Citations (as of September 2025)	1 (according to Scopus)

Table 4.1: Publication details for Paper 2

4.4 Summary and contributions

This article documents the creation of a dataset of emulated voiding sounds labeled with urinary flow values, recorded in a controlled environment using three different devices: a professional microphone UM, a smartphone Phone and a smartwatch Watch. The signals were generated using a calibrated precision peristaltic pump that generated flow rates ranging from 1 to 50 ml/s in 1 ml/s increments, reproducing realistic physiological conditions.

The dataset consists of 60-second WAV recordings for each flow value, along with metadata including device, SR and flow rate. Silent samples were also recorded. The dataset is publicly available via the Figshare repository [37].

4.4.1 Key contributions

- First public and labeled dataset of simulated voiding sounds available to the scientific community.
- Multi-device recording, promoting standardization and comparison across AI models.
- Python scripts for preprocessing, feature extraction (MFCC), segmentation and supervised model training.
- Technical validation of the generated flow using system calibration and relative error assessment.

4.4.2 Limitations noted by the authors

- Recordings were performed in a controlled environment simulating male voiding in a standing position.

- No anatomical variations (e.g., female subjects) or real-world acoustic contexts with uncontrolled background noise were included.
- The outlet height of the peristaltic pump was fixed at 85 cm to simulate the average male standing voiding posture , which may limit acoustic variability compared to natural anatomical differences or other postures.
- The outlet height of the peristaltic pump was fixed at 85 cm to approximate the urine stream exit point in standing adult males, which may limit acoustic variability compared to natural anatomical differences or alternative voiding positions.

4.5 Full article

The final version of the article is included below:

www.nature.com/scientificdata

scientific data

OPEN

DATA DESCRIPTOR

Check for updates

Annotated dataset of simulated voiding sound for urine flow estimation

Marcos Lazaro Alvarez[1]✉, Laura Arjona[1], Alfonso Bahillo[2] & Ganeko Bernardo-Seisdedos[3]

Sound-based uroflowmetry is a non-invasive test emerging as an alternative to standard uroflowmetry, estimating voiding characteristics from the sound generated by urine striking water in a toilet bowl. The lack of labeled flow sound datasets limits research for developing supervised AI algorithms. This work presents a dataset of simulated urinary flow sound recordings at flow rates from 1 to 50 ml/s, in increments of 1 ml/s, against water in a real toilet bowl. Flow generation employed an L600-1F precision peristaltic pump, with simultaneous recordings from three devices: high-quality Ultramic384k microphone, Mi A1 smartphone and Oppo smartwatch. Water was expelled through a 6 mm diameter nozzle (simulating the urethra) from a variable height of 73 to 86 cm, mimicking adult urination. The dataset provides 60-seconds labeled, constant-flow audio recordings (WAV format). This resource is intended to support research on sound-based urinary flow estimation by developing and validating supervised artificial intelligence algorithms.

Background & Summary

The growing development of artificial intelligence (AI) is transforming healthcare systems, steering them towards a more proactive and remote approach, which promises to redefine healthcare at multiple levels. This technological evolution allows healthcare providers to anticipate diseases through predictive analysis based on massive and real-time health data, significantly improving the early detection of pathologies, personalization of treatments and management of chronic diseases[1]. At the same time, for patients, AI systems are facilitating more convenient and continuous access to healthcare, eliminating the barrier of distance and improving treatment adherence through remote monitoring and automated reminders[2].

One medical test significantly benefiting from AI is sound-based uroflowmetry (SU). This innovative technique seeks to estimate urinary flow patterns during bladder emptying based on the sound generated by urine striking the water surface in a toilet bowl. SU emerges as a remote and proactive alternative to uroflowmetry (UF), a standard clinical test performed by urologists to detect issues associated with urinary tract symptoms (LUTS), such as obstructions or voiding dysfunctions.

UF, while effective, is conducted in a clinical setting where the patient must urinate into a uroflowmeter, a device that measures critical parameters such as urinary flow rate, voided volume and the times involved in the voiding process[3]. However, the effectiveness of UF can be affected by contextual factors, such as the stress or discomfort patients may experience when undergoing the test in an unfamiliar or unnatural environment, potentially altering their normal voiding patterns[4]. Moreover, just one uroflowmetry test could not be representative enough of how the patient usually voids. Therefore, SU improves patient adherence allowing home based interventions and reduces the variability of the results increasing the number of tests.

The main challenge in developing AI-based research related to SU relies on the absence of public datasets with labeled flows. This issue poses a significant challenge for researchers and limits the advancement of associated clinical applications. Currently, there are multiple works in the literature that address the estimation of flow parameters in SU[5–9], but most researchers create their own databases independently, leading to significant variations in terms of experimental design, recording devices used, flow parameters and characteristics of the subjects under study. This lack of standardization makes it difficult to compare studies, impedes the replicability

[1]Faculty of Engineering, University of Deusto, Bilbao, 48007, Spain. [2]Department of Signal Theory and Communications, Universidad de Valladolid, Valladolid, 47011, Spain. [3]Department of Medicine, Faculty of Health Sciences, University of Deusto, Bilbao, 48007, Spain. ✉e-mail: alvarez.marcoslazaro@deusto.es

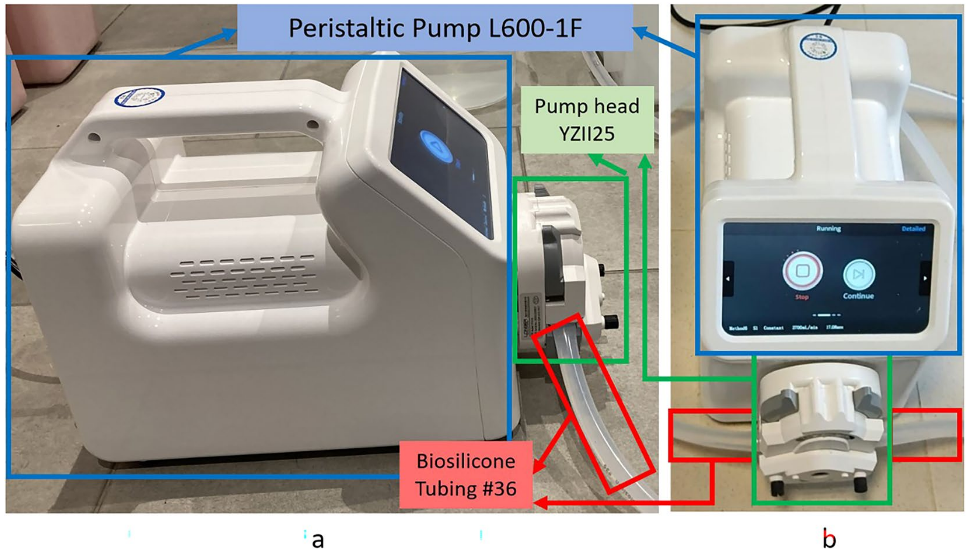


Fig. 1 Precise laboratory peristaltic pump (L600-1F) 0,16 μl/min-3000 ml/min: (**a**) side view and (**b**) front view.

of experiments and creates inconsistencies in AI based algorithms applied to SU. Moreover, these datasets are not public.

Among the recording devices used in different studies, there is a wide variety, including smartphones[8–10], dedicated microphones[11,12] and smartwatches[13,14], which produce acoustic data with very different characteristics, further widening the gap between technological developments and clinical applications. The heterogeneity in the recording and data processing protocols limits the ability to develop robust and generalizable algorithms that can be validated under various clinical conditions.

The availability of labelled datasets is essential for training AI models capable of accurately predicting urinary flows and detecting pathological conditions earlier. In addition, the SU combined to these AI-based models can be used to create a digital voiding diary that measures multiple flows at different times of the day and night, which could be significantly more useful and objective for determining any pathology. Without well-curated public datasets, the possibility of creating models that can be applied to different devices and clinical scenarios is compromised, limiting the potential impact of SU as an accessible and reliable diagnostic tool. Creating a standardized, public, multi-device dataset, even if based on synthetic flows, could be a key step toward democratizing this technology and its effective integration into clinical practice.

In this work, we have undertaken the task of creating a synthetic flow dataset ranging from 1 to 50 ml/s in increments of 1 ml/s against water in a real toilet, using a L600-1F precision peristaltic pump and recording with three devices: high-quality Ultramic384k microphone, Mi A1 smartphone and Oppo smartwatch. The water was expelled through a 6-mm diameter nozzle that simulates the average external urethral meatus of the adult male[15], from a variable height of 73-86 cm. This clean dataset, without any other sound but the simulated urine striking the water of a real toilet bowl, could be used as the basement over which voids are simulated in real environments adding background noise before training AI-based models.In the following sections, we will provide a detailed analysis of the steps and methods used to produce the data.

Methods

Flow generating device. For the generation of the dataset, we used a L600-1F precision peristaltic pump[16], which generates flows in the range of 0.16 μl/min – 3000 ml/min (2.67 nl/s – 50 ml/s), depending on the head and tube diameter used. For our application, we aimed to generate flows between 1 and 50 ml/s, as this is the flow range observed in uroflowmetry tests according to the International Continence Society (ICS)[17], corresponding to flows of 60 ml/min – 3000 ml/min. In our case, we used the YZII25 pump head with 9.5 mm inner diameter x 2.4 mm wall thickness biosilicone tubing (code #36), which covers a flow range of 160 μl/min – 3000 ml/min (see Fig. 1). The length of the biosilicone tubing was 15 m and we used this length to place the peristaltic pump in another room outside the bathroom where the microphones were recording, to isolate the environment from the noise produced by the pump. We had to calibrate the pump using a graduated cylinder to guarantee the flow values supplied by the peristaltic pump, as it will be shown in Section 3, validating the accuracy of the flows.

Table 1 shows the specifications provided by the manufacturer for the L600-1F pump model.

Audio recording devices. Building on the diversity of recording devices reported in previous SU studies, we selected three representative types to reflect common practices and support methodological consistency. Smartphones are widely used due to their availability, integrated microphones and ease of use[6,8,9]. Dedicated microphones are used in studies that require high-fidelity or privacy-focused acoustic data[11,12]. Smartwatches, on the other hand, enable hands-free, body-fixed recording and are increasingly adopted for long-term, home-based

Parameter	L600-1F
Speed	0.1 rpm – 600 rpm CW/CCW
Speed resolution	0.1 rpm (speed $<$ 100 rpm), 1 rpm (speed $>$ 100 rpm)
Flow rate	0.16 μl/min–3000 ml/min
Dispensing volume	100 μl–9999 l ($\pm$2% accuracy with calibration)

Table 1. Specifications for L600-1F Pump Model.

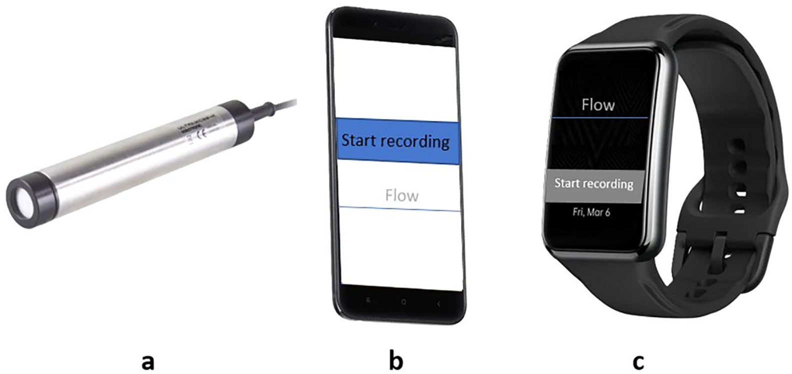


Fig. 2 Audio recording devices used: (**a**) UM, (**b**) Phone and (**c**) Watch.

monitoring[13,14]. In particular, the selection of the smartwatch model used in our dataset is supported by the comparative evaluation presented in[13], which assessed multiple smartwatch devices and justified the use of the Oppo smartwatch for its recording performance and suitability for non-intrusive SU testing.

Accordingly, our dataset includes one representative device from each of these categories (smartphone, smartwatch and professional microphone). This selection is intended to capture a broad range of realistic acoustic scenarios while supporting reproducibility and cross-device algorithm development. The selected devices are shown in Fig. 2 and described below:

- Ultramic384K (UM): a high-quality external USB microphone (Ultramic384k by Dodotronic)[18], incorporating a Knowles FG23629 MEMS sensor. It is capable of a maximum maximum sampling rate (SR) of 384 kHz, allowing for detailed analysis of a wide frequency spectrum. In our tests, we used a SR of 192 kHz, sufficient to analyze up to 96 kHz, while minimizing data volume. The microphone was connected to a laptop where recordings were triggered using a custom Python script with pre-set parameters.
- Mi A1 smartphone (Phone)Phone: a Xiaomi Mi A1 smartphone equipped with an integrated Knowles SPU0410LR5H-QB MEMS microphone, commonly found in mobile consumer devices. The microphone operated at a SR of 48 kHz, capturing frequencies up to 24 kHz. Smartphones have been extensively used in previous SU studies due to their accessibility, built-in microphones and ease of software integration. In our setup, recordings were managed using a custom Android app with fixed parameters.
- Oppo smartwatch (Watch): an Oppo smartwatch integrating a medium-quality embedded microphone (exact chipset not publicly disclosed), validated in prior SU studies[13]. The watch recorded audio at a SR of 44.1 kHz and was selected for its practicality during voiding, offering a fixed recording position on the wrist without interfering with the act. Recording was initiated using a companion Android mobile app.

Experimental Setup. Figure 3 shows the position of the recording devices in the bathroom where the flow recording tests were conducted. The heights from the ground for the UM, Phone and Watch were 84, 95 and 86 cm, respectively. The selection criteria for the height and position of the UM and Phone were based on the average position and height of a toilet tank. For the Watch, the position and height simulated the average wrist height from the ground for a person standing with relaxed arms by their sides.

The height of the water outlet nozzle was set at a reference height of 85 cm, which approximates the typical height of the urine stream exit in standing adult males. While the average male hip height is around 90 cm[19], the anatomical location of the urethral meatus is slightly lower, justifying the use of 85 cm as a realistic reference point for simulating voiding conditions. To ensure that the stream consistently impacted the water surface in the toilet bowl across all flow conditions, the nozzle height was adjusted between 73 and 86 cm depending on flow intensity. At lower flows, a more horizontal jet trajectory was required and the inclination of the outlet tube was also modified accordingly. These variations remained within realistic anatomical limits and were not expected to significantly affect the acoustic properties of the recorded sound.

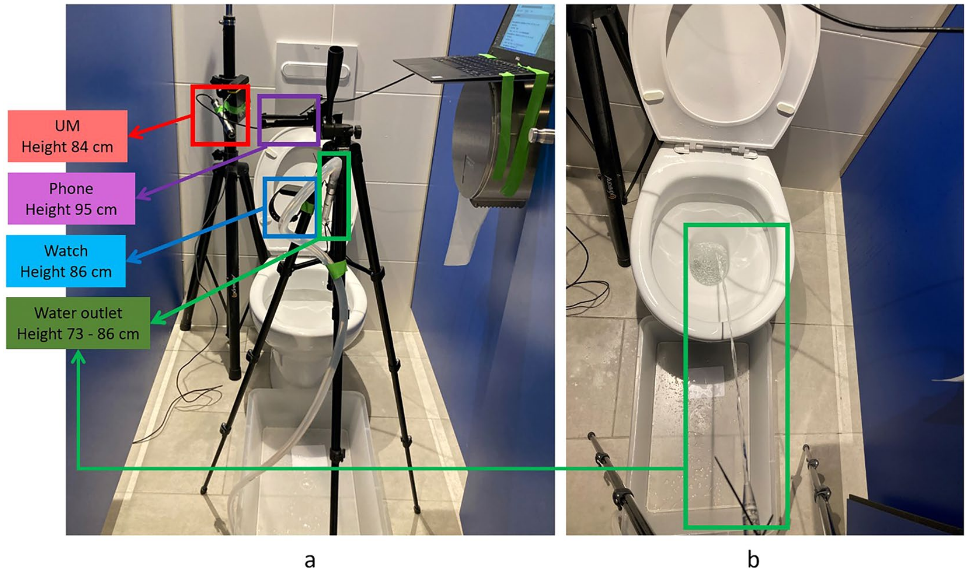


Fig. 3 Recording stage: (**a**) front view and (**b**) top view.

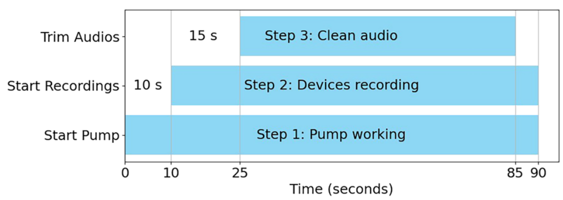


Fig. 4 Process Flow Diagram.

Data Acquisition. Our goal is to obtain labeled audio samples of constant flows with a duration of 60 seconds for each recording device. The data were collected under controlled conditions to minimize background noise and external interference. The pump was programmed for each of the flows from 1 – 50 ml/s in 1 ml/s increments, with a duration of 90 seconds. Once the pump was activated (Fig. 4, Step 1: Pump working), the recording process for each microphone was started activating the UM, Phone and Watch. The recording start process was as follows: for the UM, the "Enter" key was pressed on the laptop keyboard; for the Phone and Watch, a touch button was pressed in their respective Android based mobile apps. All devices were configured to record for 80 seconds once initiated (Fig. 4, Step 2: Devices recordings). The file name format for the recordings of the different flow audios for each device is as follows: "[device]_f_[flow].wav" (where *device* is UM, Phone, or Watch and *flow* is the flow between 1 and 50 ml/s). Once all the audio files were obtained with their respective flow labels for each device, they were trimmed to 60 seconds of pure urination sound with their corresponding flow. For this, we trimmed the first 15 seconds of each audio to remove the initial noise related to the start of recording and the last 5 seconds, finally obtaining labeled 60-seconds audio files (Fig. 4, Step 3: Clean audio). Figure 4 shows a diagram of the progress followed from step 1 to step 3 to obtain the audio recordings from the dataset. The trimmed audio "[device]_f_[flow]_[duration]s.wav," where the number of seconds in the audio is added under the tag *duration*. Figure 5 shows the flowchart for the audio collection procedure.

We have also included 30 seconds of total silence recording for the three devices in the same environment where all the flow data were recorded, labeled with a flow value of 0.

Data Records

The audio dataset described in this work has been deposited in the Figshare repository[20] (https://doi.org/10.6084/m9.figshare.27606642) and is organized as follows:

- Three folders named `Ultramic_1min`, `Phone_1min` and `Oppo_1min`, corresponding to the recording devices UM, Phone and Watch, respectively.
- Within each folder, there are 51 audio files in WAV format for each flow rate (1-50 ml/s, in increments of 1 ml/s), as well as an additional file representing 30 seconds of silence.

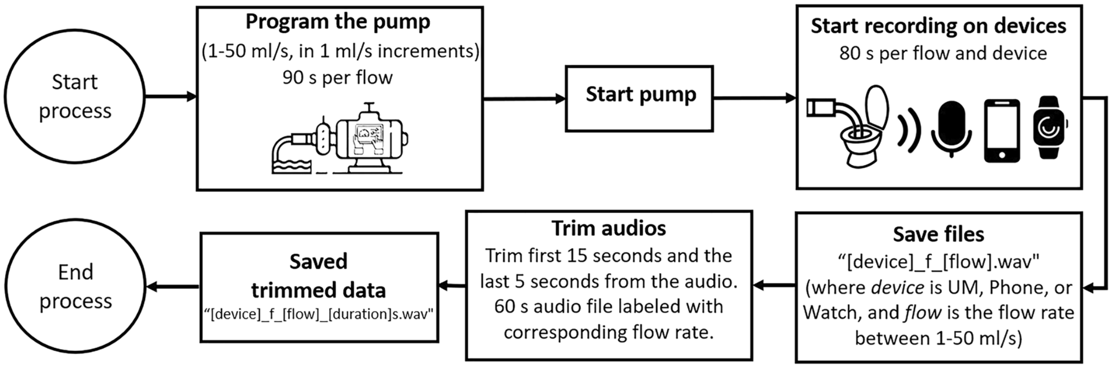


Fig. 5 Flowchart describing the audio collection procedure.

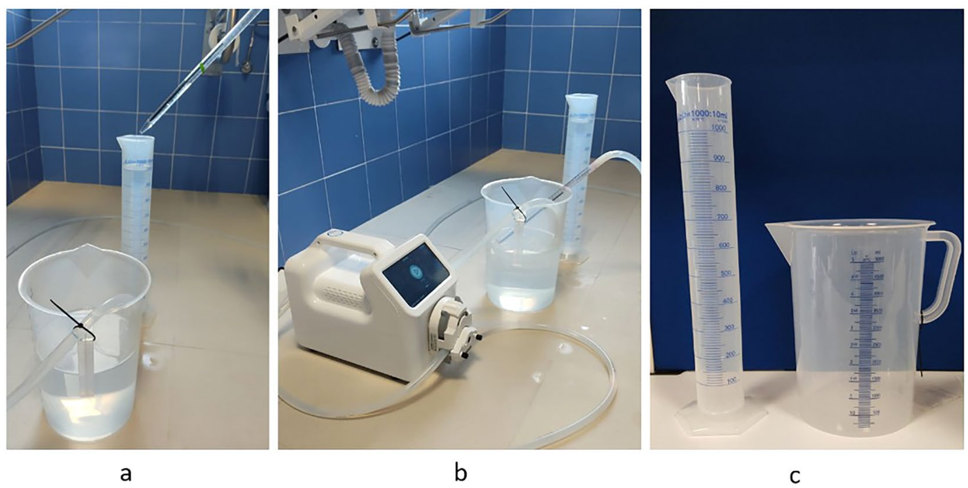


Fig. 6 Flow validation scenario with the peristaltic pump: (**a,b**) calibration test setup and (**c**) graduated cylinder.

- Each file includes metadata specifying the flow rate, recording device, SR and duration in seconds, serving as annotations for supervised learning tasks. No additional annotations or manual classifications were applied beyond these flow rate labels.

Technical Validation

To validate the quality of the dataset, we conducted multiple trials to ensure consistency in the recordings. The validation of the flow rate supplied by the peristaltic pump was performed using a graduated cylinder of 1000 ml volume with a resolution ± 10 ml (see Fig. 6). The results obtained are presented in Table 2. It contains the programmed flow rates, the programmed operation time, the expected volume in ml and the real volume emptied by the pump, as it comes from the factory.

The validation results confirm that the peristaltic pump has a high degree of accuracy in most of the programmed flow rates, with minimal differences between the expected and actual emptied volumes. For higher flow rates, such as 50 ml/s to 25 ml/s, the relative error percentages are quite low, remaining below 2.1%. However, for lower flow rates, specifically between 20 ml/s and 5 ml/s, the relative errors increase significantly, reaching 10%. These deviations can be attributed to the use of a 15-meter hose, as small pressure losses and internal friction have a greater impact on reduced flow rates. At higher flows, the pump can better compensate for these losses, maintaining a more stable flow.

To reduce errors associated with low flow rates, we calibrated the pump for low flows from 5 to 20 ml/s in increments of 5 ml/s. Table 2 shows the results for the real calibrated volume and their corresponding calibrated absolute errors from 5 to 20 ml/s. In conclusion, the results are consistent and suitable for the purpose of this study, providing a solid foundation for creating a dataset with accurately labelled flows.

The use of the UM microphone provided high-resolution sound data, while recordings from the Phone and Watch represent more accessible alternatives that may be more commonly available for practical applications.

Programmed		Expected	Real volume		Relative error (%)	
Flow	Time	volume	Uncalibrated	Calibrated	Uncalibrated	Calibrated
(ml/s)	(s)	(ml)	(ml)	(ml)		
50	20	1000	1000	—	0.00	—
45	22	990	1000	—	1.01	—
40	25	1000	990	—	1.00	—
35	28	980	960	—	2.04	—
30	33	990	970	—	2.02	—
25	40	1000	990	—	1.00	—
20	50	1000	950	990	5.00	1.00
15	60	900	830	900	7.78	0.00
10	60	600	560	590	6.67	1.67
5	60	300	270	295	10.00	1.00

Table 2. Validation of the peristaltic pump with times in seconds, expected volumes and actual emptied volumes.

Usage Notes

Researchers can use this dataset to develop and evaluate AI-based models for estimating urinary flow rates based on sound recordings. Possible applications include training regression models to predict continuous flow rates from acoustic features (e.g., Melfrequency cepstral coefficients (MFCC)), or classification models to categorize flow profiles into clinically relevant groups (e.g., low, normal, high). The labeled structure of the recordings supports supervised learning, model validation and benchmarking in a reproducible manner.

To support reproducibility and practical application of the dataset, a sample Python notebook compatible with Google Colab has been provided as supplementary material in the dataset repository (`sound_dataset_processing.ipynb`)[20]. The script implements the recommended preprocessing steps, including background noise addition to simulate real environments, noise reduction via high-pass filtering, audio segmentation, MFCC-based feature extraction, normalization across devices to address microphone variability and supervised modeling with 10-fold cross-validation. Additionally, the high sampling rate of the UM allows for a detailed analysis in the frequency domain, which could provide deeper insights into the acoustic properties associated with different flow rates.

As a limitation, it should be noted that the dataset was created under controlled conditions simulating a standing adult male with relaxed arms, at a fixed height of approximately 85 cm. Variations in toilet structure or anatomical factors (including gender differences) and voiding posture (sitting or standing) may influence the acoustic characteristics of real voiding events. Researchers should take these factors into account when generalizing findings to clinical settings.

Code availability

The Python notebook `sound_dataset_processing.ipynb` has been developed to support the preprocessing and modeling of the dataset. The notebook is available in the dataset repository[20] and demonstrates preprocessing, feature extraction, normalization across devices and supervised modeling in a reproducible manner.

Received: 12 November 2024; Accepted: 5 June 2025;
Published online: 13 June 2025

References

1. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine* **25**, 44–56 (2019).
2. Krittanawong, C., Zhang, H., Wang, Z., Aydar, M. & Kitai, T. Artificial intelligence in precision cardiovascular medicine. *Journal of the American College of Cardiology* **69**, 2657–2664 (2017).
3. Fowler, C. J., Griffiths, D. & De Groat, W. C. The neural control of micturition. *Nature Reviews Neuroscience* **9**, 453–466 (2008).
4. Kuoch, K. L., Meyer, D., Austin, D. W. & Knowles, S. R. Classification and differentiation of bladder and bowel related anxieties: A socio-cognitive exploration. *Current Psychology* **40**, 4004–4011 (2021).
5. Alvarez, M. L., Arjona, L., Jojoa-Acosta, M. & Bahillo, A. Flow prediction in sound-based uroflowmetry. *Scientific Reports* **15**, 643 (2025).
6. Lee, H. J. *et al*. Development and validation of a deep learning system for sound-based prediction of urinary flow. *European Urology Focus* **9**, 209–215 (2023).
7. Lee, D. G. *et al*. A prospective comparative study of mobile acoustic uroflowmetry and conventional uroflowmetry. *International Neurourology Journal* **25**, 355 (2021).
8. Lee, Y. J., Kim, M. M., Song, S. H. & Lee, S. A novel mobile acoustic uroflowmetry: comparison with contemporary uroflowmetry. *International Neurourology Journal* **25**, 150 (2021).
9. Aslim, E. J. *et al*. Pilot study for the comparison of machine-learning augmented audio-uroflowmetry with standard uroflowmetry in healthy men. *BMJ Innovations* 6 (2020).
10. Comiter, C. V & Belotserkovsky, E. A novel mobile uroflowmetry application for assessing lower urinary tract symptoms. *Neurourology and Urodynamics*, S56–S57 (2019).
11. Alvarez, M. L., Arjona, L., Iglesias Martínez, M. E. & Bahillo, A. Automatic classification of the physical surface in sound uroflowmetry using machine learning methods. *EURASIP Journal on Audio, Speech and Music Processing* **2024**, 12 (2024).
12. Hurtík, P., Burda, M., Krhut, J., Zvara, P & Lunácek, L. Automatic diagnosis of voiding dysfunction from sound signal. In: *2015 IEEE Symposium Series on Computational Intelligence*, 1331–1336 (2015).

13. Arjona, L., Diez, L. E., Martinez, A. B. & Echevarria, A. A. Urosound: A smartwatch-based platform to perform non-intrusive sound-based uroflowmetry. *IEEE Journal of Biomedical and Health Informatics*. (2022).
14. Narayanswamy, G., Arjona, L., Díez, L. E., Bahillo, A. & Patel, S. Automatic classification of audio uroflowmetry with a smartwatch. In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 4325–4329 (2022).
15. Standring, S. Bladder, prostate and urethra. In: *Gray's Anatomy: The Anatomical Basis of Clinical Practice* **40**, 1250–1252 (2008).
16. FluidControlSolutions. Precise laboratory peristaltic pump. https://www.fluidcontrol24.com/systems/laboratory-peristaltic-pumps.html.
17. Schäfer, W. *et al*. Good urodynamic practices: uroflowmetry, filling cystometry and pressure-flow studies. *Neurourology and Urodynamics* **21**, 261–274 (2002).
18. Dodotronic. Ultramic 384K BLE. https://www.dodotronic.com/product/ultramic-384k-ble/.
19. Pheasant, S. *Bodyspace: Anthropometry, Ergonomics and Design*. Taylor & Francis, Bristol, PA (1988).
20. Alvarez, M. L., Arjona, L., Bahillo, A. & Bernardo-Seisdedos, G. Annotated dataset of simulated voiding sound for urine flow estimation. *Figshare*. https://doi.org/10.6084/m9.figshare.27606642 (2025).

Acknowledgements

This research was supported by the Spanish Ministry of Science and Innovation under SWALU project (ref. CPP2022-010045) and 'Ayuda para contratos predoctorales 2020 (ref. PRE2020-095612)' funded by MICIU/AEI /10.13039/501100011033 and co-financed by FSE invierte en tu futuro. Additionally, partial support was provided by the Ministry under the Aginplace project (ref. PID2023-146254OB-C41 and ref. PID2023-146254OA-C44) financed by MICIU/AEI/10.13039/501100011033 and FEDER, UE.

Author contributions

Marcos Lazaro, Laura and Alfonso conceived the study. Marcos Lazaro, Laura and Ganeko designed the experiments and contributed to sample preparation and data collection. All authors participated in the analysis and interpretation of the results and contributed to the writing and revision of the manuscript. Marcos Lazaro drafted the manuscript, incorporating feedback from all authors.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential competing of interest.

Additional information

Correspondence and requests for materials should be addressed to M.L.A.

Reprints and permissions information is available at www.nature.com/reprints.

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

Chapter 5

Leveraging synthetic data to develop a machine learning model for voiding flow rate prediction from audio signals

5.1 Introduction

This chapter presents the third article published as part of the compendium of this doctoral thesis. The study addresses a comprehensive strategy for training and validating AI models in the context of SU. The proposed methodology is based on training supervised algorithms using synthetic data generated under controlled conditions [37] and subsequently validating them with real SU recordings, aiming to ensure better system generalization and robustness.

5.2 Reference

Alvarez, M. L., Bahillo, A., Arjona, L., Nogueira, D. M., Gomes, E. F., & Jorge, A. M. (2025). *Leveraging synthetic data to develop a machine learning model for voiding flow rate prediction from audio signals*. **IEEE Access**, *13*, 127240–127251. https://doi.org/10.1109/ACCESS.2025.3590626

5.3 Status

Table 5.1 shows the publication status of the paper presented in this chapter.

Publication status	Published
Journal	IEEE Access
Publisher	IEEE Xplore
Year	2025
JCR Impact Factor (2024)	3.6
JCR Quartile	Q2 in Computer Science, Information Systems
DOI	https://doi.org/10.1109/ACCESS.2025.3590626
Citations (as of September 2025)	0 (according to Scopus)

Table 5.1: Publication details for Paper 3

5.4 Summary and contributions

The article proposes a two-stage methodology to improve the accuracy of urinary flow prediction from acoustic signals. In the first stage, various ML models (RF, GB, SVR and CNN) are trained and compared using the simulated SU data generated in the study from Chapter 4. In the second stage, the best models are retrained using real voiding recordings from 47 male volunteers, obtained from the study described in Chapter 3, allowing for performance validation under real-world conditions.

Experimental results indicate that models pretrained with synthetic flow data achieve effective adaptation when fine-tuned with real SU signals. In particular, evaluation of the RFR model with optimized hyperparameters showed a reduction in MAE of 0.7, 0.3 and 0.3 ml/s for the Watch, Phone and UM devices respectively, compared to the results reported in Chapter 3 for the same dataset. Additionally, MAE values below 2.5 ml/s were achieved for flow prediction, with CCC above 0.86. These findings confirm the viability of using synthetic data as a pretraining strategy to enhance flow estimation performance in real scenarios.

5.4.1 Key contributions

- Proposal of a dual training and validation scheme to improve generalization in real-world environments.
- Systematic comparison between AI models trained solely on real data and models pretrained on synthetic data and subsequently re-trained on real data.
- Demonstration that combining synthetic and real data can overcome the limitations of clinical data availability.

5.4.2 Limitations noted by the authors

- Validation is limited to healthy male participants voiding while standing; women and patients with urological conditions were not included.
- Generalization to uncontrolled home environments still requires further study.
- The model does not yet account for variability in acoustic settings such as different toilet types or environmental noise.

5.5 Full article

The final version of the article is included below:

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 July 2025, accepted 16 July 2025, date of publication 18 July 2025, date of current version 24 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3590626

RESEARCH ARTICLE

Leveraging Synthetic Data to Develop a Machine Learning Model for Voiding Flow Rate Prediction From Audio Signals

MARCOS LAZARO ALVAREZ [1], ALFONSO BAHILLO [2], LAURA ARJONA [1], DIOGO MARCELO NOGUEIRA [3,4], ELSA FERREIRA GOMES [4,5], AND ALÍPIO M. JORGE [3,4]

[1] Faculty of Engineering, University of Deusto, Bilbao, 48007 Bizkaia, Spain
[2] Department of Signal Theory and Communications, Universidad de Valladolid, 47002 Valladolid, Spain
[3] Department of Computer Science, Faculdade de Ciências da Universidade do Porto, 4169-007 Porto, Portugal
[4] INESC TEC-Institute for Systems and Computer Engineering, Technology, and Science, Campus da FEUP, 4200-465 Porto, Portugal
[5] Instituto Superior de Engenharia do Porto, 4249-015 Porto, Portugal

Corresponding author: Marcos Lazaro Alvarez (alvarez.marcoslazaro@deusto.es)

This work was supported in part by the Spanish Ministry of Science, Innovation and Universities (MICIU) through the SWALU Project under Grant CPP2022-010045; in part by the 2020 "Ayuda para contratos predoctorales," funded by MICIU and the State Research Agency Agencia Estatal de Investigación (AEI), 10.13039/501100011033, and co-financed by the European Social Fund Fondo Social Europeo (FSE) under the slogan "FSE invierte en tu futuro," under Grant PRE2020-095612; in part by the Basque Government through the Hazitek Program under the BATHMIC Project, Grant ZL-2024/00481; and in part by the Ministry through the Aginplace Project, funded by MICIU, AEI (10.13039/501100011033), and the European Union (UE) through the European Regional Development Fund Fondo Europeo de Desarrollo Regional (FEDER), under Grants PID2023-146254OB-C41 and PID2023-146254OA-C44.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Valladolid East Health Area Medicine Research Ethics Committee on 27 July 2023 (reference PI-GR-23-3275, minutes number 16/2023), and the Committee complies with GCP standards (CPMP/ICH/135/95).

ABSTRACT Sound-based uroflowmetry (SU) is a non-invasive technique emerging as an alternative to traditional uroflowmetry (UF) to calculate the voiding flow rate based on the sound generated by the urine impacting the water in a toilet, enabling remote monitoring and reducing the patient burden and clinical costs. This study trains four different machine learning (ML) models (random forest, gradient boosting, support vector machine and convolutional neural network) using both regression and classification approaches to predict and categorize the voiding flow rate from sound events. The models were trained with a dataset that contains sounds from synthetic void events generated with a high precision peristaltic pump and a traditional toilet. Sound was simultaneously recorded with three devices: Ultramic384k, Mi A1 smartphone and Oppo Smartwatch. To extract the audio features, our analysis showed that segmenting the audio signals into 1000 ms segments with frequencies up to 16 kHz provided the best results. Results show that random forest achieved the best performance in both regression and classification tasks, with a mean absolute error (MAE) of 0.9, 0.7 and 0.9 ml/s and quadratic weighted kappa (QWK) of 0.99, 1.0 and 1.0 for the three devices. To evaluate the models in a real environment and assess the effectiveness of training with synthetic data, the best-performing models were retrained and validated using a real voiding sounds dataset. The results reported an MAE below 2.5 ml/s and a QWK above 0.86 for regression and classification tasks, respectively.

INDEX TERMS Machine learning, non-invasive voiding monitoring, sound-based uroflowmetry, sound voiding signals, voiding flow estimation.

I. INTRODUCTION

The rapid development of information and communication technologies (ICT) is transforming healthcare systems by making them more proactive and remote. This transformation offers significant benefits, such as improving the quality of healthcare services and reducing associated costs. Patients gain broader access to medical services, while healthcare providers can leverage real-time information and resources to optimize treatment strategies. Among the various areas

The associate editor coordinating the review of this manuscript and approving it for publication was Yongqiang Cheng.

influenced by these advancements, the application of machine learning (ML) has proven to be revolutionary, enabling predictive analytics, personalized treatments and remote monitoring systems that have significantly improved healthcare and treatment adherence [1], [2].

One of the healthcare fields where these technologies have shown great potential is urology, particularly in uroflowmetry (UF) testing. Lower urinary tract symptoms (LUTS) affect more than 1.9 billion people worldwide, causing a significant reduction in quality of life and increasing the burden on healthcare systems [3], [4]. The standard test for diagnosing LUTS is UF, a non-invasive procedure that evaluates parameters such as maximum flow rate (Qmax), average flow rate (Qave), voided volume (VV) and voiding time. However, UF is typically conducted in clinical settings, where patients must urinate into a uroflowmeter. This process often causes stress or discomfort, altering natural voiding patterns and potentially introducing significant variability in the results [5], [6]. Furthermore, a single UF test may not be sufficiently representative of a patient's habitual voiding patterns. Sound-based uroflowmetry (SU) has emerged as a promising alternative, allowing patients to perform voiding tests in the comfort of their homes. SU estimates urinary flow from the sound produced by urine impacting the water surface in a toilet, providing a more natural and non-invasive means of assessing LUTS [7], [8]. This approach has shown strong correlations with conventional UF results, achieving Pearson correlation coefficients of up to 0.95 for key parameters such as Qmax and VV [9], [10]. Therefore, SU improves patient adherence by enabling home-based interventions and reduces result variability by increasing the number of tests.

Another challenge to the widespread adoption of SU is the heterogeneity of recording devices. Previous studies have employed a variety of equipment, including professional microphones [11], smartphones [9], [12] and smartwatches [13], [14], each producing acoustic data with different characteristics. This variability complicates the development of generalizable ML models and limits their clinical applicability. Despite its potential, developing robust SU systems faces challenges due to the lack of publicly available datasets with labeled flow rates, leading to inconsistencies in experimental designs, recording protocols and model implementations across studies [12], [15].

To overcome these limitations, we developed a synthetic dataset of voiding flow sounds recorded under controlled conditions. The decision to use synthetic data was driven by several factors:

- the absence of publicly available, labeled datasets for urinary sound analysis,
- the ethical and logistical constraints of collecting large-scale real-world recordings from human subjects, and
- the need to generate a balanced and reproducible dataset suitable for training and benchmarking ML models.

Synthetic data offer the advantage of consistent labeling and controlled variability across recording devices, which are essential for developing generalizable models. Furthermore, such a dataset provides a standardized experimental framework that facilitates reproducibility and comparison across future studies.

To address these challenges, this study analyzes a synthetic and balanced void flow dataset generated through controlled simulations with three recording devices: Ultramic384 (UM), Mi A1 smartphone (Phone) and Oppo smartwatch (Watch). The synthetic dataset covers flow rates from 1 to 50 ml/s, a range that encompasses the full spectrum of male voiding flows according to UF studies [16]. These data were captured in a realistic environment using a high-precision peristaltic pump and a real ceramic toilet with water at the bottom, ensuring controlled and representative flow conditions. This balanced dataset, recorded in a noise-free environment where only the sound of urine impacting the water surface was present, serves as a robust foundation for training ML models while offering the potential to simulate real-world voiding scenarios by incorporating background noise.

To assess the real-world suitability of models trained on synthetic data, we re-trained and evaluated the best-performing model using the real SU voiding dataset from [17], which includes natural voiding events recorded under controlled conditions with the same devices. This process allowed us to validate the model's ability to predict flow rates from real SU signals, demonstrating its effectiveness in practical scenarios. Notably, this re-trained model achieved improved performance compared to the results reported in [17], further supporting the benefits of pre-training with synthetic data. A detailed analysis of these improvements is presented in Section IV.

Furthermore, we conducted a comprehensive feature analysis to identify the most relevant acoustic characteristics for flow estimation, offering insights into key factors that influence model performance. The study also compares the performance of regression and classification models across the three recording devices, further exploring the feasibility of sound-based urinary flow estimation. Lastly, a privacy-preserving analysis was performed to assess the impact of removing frequency bands containing identifiable speech information, ensuring the system remains effective while safeguarding user privacy.

Additionally, the availability of a public and balanced synthetic dataset enables fair benchmarking across different research efforts in urinary flow prediction. By training models on a shared synthetic baseline, researchers can objectively compare algorithm performance under controlled conditions. The most promising models from this evaluation can then be re-trained on real-world SU signals to adapt to environmental and behavioral variability, ensuring both reproducibility and practical relevance.

While synthetic datasets offer a standardized and reproducible environment for initial training and benchmarking, they may not fully capture the acoustic complexity of real-world conditions. This introduces a domain gap between synthetic and real urination audio. To bridge this gap, we retrain the top-performing models on real SU recordings, allowing them to adapt to natural variability in device acoustics and user behavior. Future work may further explore domain adaptation or transfer learning to enhance model robustness across environments.

The paper is organized as follows: Section II briefly reviews the state of the art in audio feature extraction and flow prediction from SU audio signals using ML models; Section III presents the materials and methods proposed in this research, describing the study design, dataset characteristics and the procedures and theoretical foundations followed in analyzing flow prediction in SU tests using different recording devices; Section IV presents the results obtained from the proposed methodology; and finally, Section V provides some concluding remarks.

II. RELATED WORK

A. ARTIFICIAL INTELLIGENCE (AI) IN SOUND ANALYSIS

AI has played a crucial role in analyzing audio signals, enabling both classification and regression tasks while offering innovative solutions for predicting continuous parameters or discriminating between classes based on acoustic features. Regression models are particularly useful for estimating continuous variables such as amplitude, frequency, or flow rates, whereas classification models are used to assign discrete labels to data, such as identifying flow rate ranges or detecting specific acoustic events.

Techniques such as random forest regressor (RFR), an ensemble learning method that constructs multiple decision trees and averages their predictions, support vector regressor (SVR), which maps features into higher dimensions to model complex nonlinear relationships, gradient boosting regressor (GBR), a technique that sequentially builds models to correct the errors of prior ones and convolutional neural network (CNN), deep learning architectures effective for learning patterns from time-frequency representations, have demonstrated effectiveness in capturing both continuous relationships in regression tasks and class patterns in classification tasks. For instance, in regression, recent studies have shown that combining traditional acoustic features, such as mel-frequency cepstral coefficients (MFCC)—which model the human auditory system's perception of sound frequency—with ML models can enhance the estimation of continuous parameters, including urinary flow rate and environmental sound intensity [10], [12]. Furthermore, supervised ML models, such as k-nearest neighbors (KNN) and GBR, have been successfully applied to estimate flow rate from audio signals [18]. These models, trained using extracted MFCC, effectively capture spectral characteristics relevant to flow estimation.

For classification tasks, deep learning architectures such as CNN and Multilayer Perceptrons (MLP) have shown success in classifying environmental sounds and segmenting acoustic parameters [19], [20], [21]. The integration of advanced ML techniques in audio analysis has enabled accurate predictions and robust label assignments, opening new applications in sound-based flow estimation, environmental monitoring and diagnosis based on acoustic parameters. Moreover, the use of acoustic features such as MFCC, zero-crossing rate (ZCR)—which measures how frequently the signal waveform crosses the zero amplitude axis—and chroma features (Chroma) provides complementary information that enhances the robustness of ML models in noisy and diverse environments, supporting both regression and classification tasks [22], [23]. These feature sets play a crucial role in improving model generalization, making AI-based approaches more reliable for practical applications.

B. SOUND FEATURE EXTRACTION TECHNIQUES

Feature extraction in audio signal processing is essential for various applications, including speech analysis, music classification and environmental sound detection. The complexity of audio signals, characterized by non-stationarities and discontinuities, requires robust extraction techniques capable of effectively capturing relevant signal characteristics. These methods can be broadly categorized into temporal, spectral, cepstral and time-frequency approaches. Temporal features analyze waveform properties, while spectral techniques focus on frequency content.

Studies such as [24] highlight that MFCC are among the most widely used techniques for feature extraction in audio-based applications. MFCC are designed to model how humans perceive frequency variations, making them effective for capturing key spectral characteristics in various acoustic environments. Their robustness in representing complex sound patterns has been well established across different domains. In the context of SU, [12] demonstrated that MFCC effectively capture spectral features relevant to voiding sounds, enabling accurate estimation of urinary flow parameters using ML techniques. The study underscores that MFCC, by providing a perceptually motivated frequency representation, enhance the ability to model the acoustic properties of urinary flow, reinforcing their suitability for UF applications. Furthermore, MFCC have been applied in other fluid dynamics studies; for instance, [18] employed MFCC to estimate the flow rate of liquid jets impacting a water surface, demonstrating their capability in characterizing acoustic signatures associated with fluid motion.

Another widely used technique in audio analysis is linearly binned fast fourier transform (FFT), which converts time-domain signals into their frequency-domain representation. This method provides a more uniform frequency resolution across the spectrum, making it suitable for applications beyond human speech. This technique has been applied in the domain of SU as an input feature for prediction and classification models [11], [25], further validating its utility in bioacoustic analysis.

An emerging alternative is the continuous wavelet transform (CWT), which offers multiresolution analysis capabilities ideal for non-stationary signals. Scalograms derived from CWT have shown to improve the models performance in certain acoustic recognition tasks using CNN, particularly when capturing localized time-frequency variations [26]. However, due to their high computational cost and the high dimensionality of the resulting representations, they were not adopted in our pipeline, which was designed to remain computationally efficient and to reduce the risk of overfitting given the limited size of our labeled datasets. In fact, preliminary tests with Mel-spectrograms and CNN yielded suboptimal performance due to overfitting, even after multiple architecture and training refinements.

C. FLOW RATE ESTIMATION WITH SOUND

A promising strategy for improving voiding flow estimation from sound involves pretraining models on synthetic data and subsequently retraining them on real-world recordings. This approach leverages the controlled nature of synthetic datasets to develop robust initial models, which can then be fine-tuned using real data to enhance generalization and practical applicability [27].

Despite the growing interest in using ML for voiding flow estimation based on sound [9], [10], [12], [15], [17], [28], [29], a fundamental limitation in the existing literature is the lack of standardized voiding flow datasets and the wide variability in evaluation metrics. This inconsistency hinders direct comparisons between studies and limits their real-world applicability. Previous research has utilized different datasets of voiding sounds, each recorded under unique conditions, employing distinct preprocessing steps and model evaluation criteria. Consequently, cross-study comparisons become inconsistent, non-equitable and difficult to generalize.

To address this limitation, it is essential to use a standardized and publicly accessible dataset that allows for fair comparisons between different algorithms. A synthetic voiding flow dataset recorded under controlled real-world conditions offers a valuable foundation for this purpose, as it provides a common benchmark while enabling the creation of diverse testing environments by systematically introducing variations such as background noise and recording conditions. This structured approach ensures that models trained on synthetic data can be optimized under uniform conditions before being retrained and tested on real voiding events, ultimately enhancing their reliability and applicability in practical scenarios.

III. MATERIALS AND METHODS

A. GENERAL DIAGRAM OF THE RESEARCH

The proposed methodology for analyzing flow estimation from the synthetic dataset is illustrated in Figure 1. The flow generation system was based on a L600-1F precision peristaltic pump [30], capable of producing flows within a 0.16 μl/min – 3000 ml/min (2.67 nl/s – 50 ml/s) range, depending on the selected tubing and pump head. For this study, we focused on simulating flow rates between 1 and 50 ml/s, as these values correspond to those typically observed in UF assessments [16]. This setup allowed us to generate controlled and reproducible flow conditions, ensuring accurate ground-truth labels for model training and validation.

To capture the synthetic voiding flow audio recordings, which lasted 60 seconds each, we employed three different recording devices with distinct frequency response ranges and sampling rate (SR):

- UM: A high-fidelity microphone designed for capturing a broad frequency spectrum, enabling the analysis of both audible and ultrasonic components. The selected device was configured with a SR of 192 kHz, allowing spectral analysis up to 96 kHz. The recordings were managed through a USB connection to a laptop, with a Python-based script controlling the capture parameters [31].
- Phone: A smartphone microphone with standard recording capabilities, configured with a 48 kHz SR, capturing frequencies in the 0–24 kHz range. Given the accessibility and widespread use of smartphones in previous SU studies, we developed a dedicated Android application to enable standardized audio recording with predefined settings.
- Watch: A wearable device microphone, validated for use in SU applications [13], featuring a 44.1 kHz SR. This device was chosen due to its non-intrusive nature, allowing for audio collection from a fixed position without interfering with the voiding process. A custom-built Android application was used to initiate recordings seamlessly, ensuring consistent data capture across different sessions.

Next, the audio signals are splitted into segments of equal durations ranging from 100 ms to 1000 ms to evaluate how segment size affects flow estimation accuracy. For each segments, features related to the 13 MFCC are extracted and used as inputs for the ML models in the flow estimation task. Two approaches for flow estimation will be evaluated: one using regression models and the other using multi-class classification models. After identifying the best model configurations for flow prediction and classification using a synthetic audio dataset, we re-trained and evaluated these models using a real SU test dataset [17]. The retraining process employed the same models configuration that achieved the best results on the synthetic dataset. Finally, the results obtained for the flow estimation are presented.

B. SYNTHETIC DATASET DESCRIPTION

We used a synthetic voiding event dataset to develop and train our ML models for flow rate prediction. This dataset, fully described and validated in [32], consists of labelled audio recordings of constant water streams generated by a peristaltic pump and captured using three different recording devices: a high-quality microphone UM, a Phone and a Watch. Each audio sample corresponds to a specific flow rate ranging from 1 to 50 ml/s, in increments of 1 ml/s. The recordings were performed under controlled laboratory conditions to ensure high acoustic quality and minimal background noise. The complete dataset is publicly available at [33].

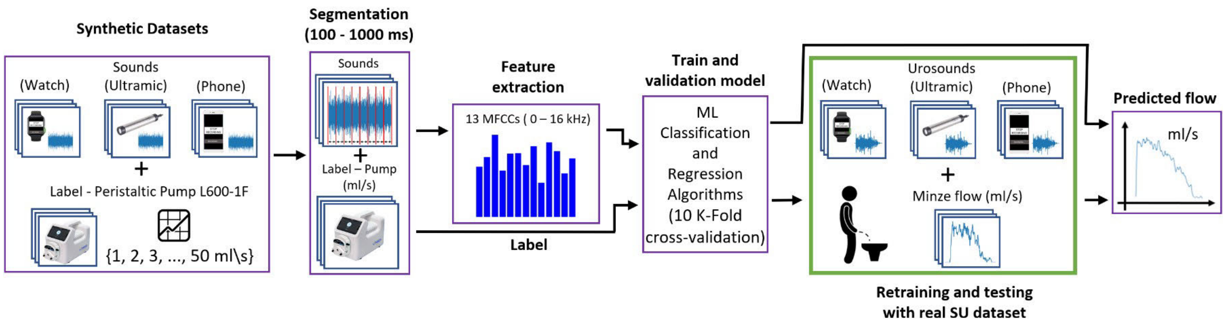


FIGURE 1. Diagram showing the pipeline of the proposed methodology, consisting of 4 main steps: data extraction, audio segmentation, feature extraction and finally model training and validation.

On the one hand, the density of human urine typically ranges between 1.005 and 1.030 g/cm^3, depending on the concentration of solutes such as urea, salts and other substances [34]. On the other hand, pure water density is 1 g/cm^3 a 4°C, that is, between 0.5% and 3.0% less dense than human urine. Consequently, we assume this difference to be negligible.

Each audio file in the dataset follows a standardized naming format: "`[device]_f_[flow]_[duration]s`", where *device* specifies the recording hardware (UM, Phone, or Watch), *flow* represents the corresponding voiding flow rate in ml/s and *duration* indicates the length of the audio segment in seconds. For each recording device, the final dataset contains 60-second-long audio clips. These clips were obtained after trimming the first 15 s and the last 5 s from the original recording to ensure clean samples and remove artifacts.

This structured and labelled dataset enables the training and evaluation of ML models under reproducible and balanced conditions. Its public availability also facilitates benchmarking and promotes transparency in future research.

C. REAL DATASET DESCRIPTION

We have used a second dataset consisting of real voiding events, to retrain the best models that were designed with the synthetic dataset. This dataset consists of 47 real voiding events recorded in a controlled environment, as described in [17]. This study received approval from the Valladolid East Health Area Medicine Research Ethics Committee on 27 July 2023 (reference PI-GR-23-3275, minutes number 16/2023) and the Committee complies with GCP standards (CPMP/ICH/135/95).

Voiding flow rates were measured using a medical uroflowmeter from Minze [35], which provides a 10 Hz resolution and an accuracy of ±2.5 ml/s. Simultaneously, three sound recording devices (UM, Phone and Watch) captured the corresponding audio signals. To ensure consistent sound generation, the Minze uroflowmeter basin was pre-filled with 400 ml of water, simulating a real toilet environment where the primary acoustic source is the impact of urine against the water surface. Testing was carried out in a tiled bathroom with controlled acoustics to minimize background noise. Participants were instructed to direct their urine stream precisely at the water and compliance was verified through audio analysis. Recordings containing extraneous noise or uncertainties regarding the location of the urine impact were excluded. Each trial produced three synchronized WAV audio files (one per recording device), along with the corresponding UF data, enabling a comprehensive evaluation of SU.

We acknowledge that the limited size and variability of the real-world dataset may restrict the generalizability of the models. This limitation is primarily due to the logistical difficulty of collecting synchronized voiding sound and flow measurements in clinical settings. Additionally, real voiding events exhibit spontaneous and unbalanced distributions across flow rates, which makes it challenging to build a uniformly distributed dataset. Currently, no public dataset exists for SU with labeled flow values, which further limits reproducibility and comparability across studies. To address these issues, our team is collaborating with multiple clinical institutions to progressively expand the dataset under diverse acoustic and physiological conditions.

Table 1 summarizes the key characteristics of both the synthetic and real-world datasets, including their type, number of samples, flow rate range, typical sample duration and recording devices used.

TABLE 1. Summary of the datasets used in this study, including recording conditions and devices.

Dataset	Subject	# Samples	Flow Range (ml/s)	Duration per Sample (s)	Devices
Synthetic	Controlled (Pump)	150 audio clips (50 events × 3 devices)	1–50	60	UM, Phone, Watch
Real	Human subjects	141 audio clips (47 events × 3 devices)	1–35	Varied [12–48]	UM, Phone, Watch

D. FEATURE EXTRACTION FOR SYNTHETIC DATASET

In this section, we describe the frequency analysis conducted to identify the spectral components that contribute the most to voiding flow estimation, addressing the problem from both a regression and classification perspective. Our objective is to determine the frequency bands with the highest predictive contribution to voiding flow estimation. The predictive power is measured by supervised model-based feature importance scores. In particular, we used the RFR for regression tasks and the random forest classifier (RFC) for classification tasks. RFC is a tree-based ensemble learning method that aggregates predictions from multiple decision trees to improve classification accuracy. We calculated the mean squared error (MSE) for the regression models and Gini impurity for the classification models. For the two types of models, we analyze the entire spectrum captured by the specialized microphone UM (0–96 kHz). To achieve this, we extract 1000 linearly binned FFT features for each 100 ms labeled audio segment, dividing the full spectrum into 1000 equally spaced frequency bins. This choice allows for a fine-grained representation of the spectral content, ensuring that all frequency components are adequately captured while maintaining computational efficiency. By using a linear binning approach, we preserve the resolution across the entire frequency range, avoiding biases that could arise from non-uniform binning. Within each bin, we sum the absolute values of the amplitudes of the present frequency components, ultimately obtaining a feature vector of 1000 values per segment. This approach enables us to systematically assess the contribution of different frequency bands to voiding flow estimation, facilitating a robust analysis of their relative feature importance.

1) FEATURE EXTRACTION FOR REGRESSION MODELS

First, we perform a supervised feature selection using RFR from scikit-learn [36], leveraging its ability to estimate feature importance based on the reduction in variance in the target variable [37]. The feature importance scores are computed by measuring the decrease in MSE when a particular frequency band is used to split the data. This approach enables us to quantify the contribution of each frequency bin to the regression task, identifying the spectral regions that influence the most to the voiding flow estimation. Feature importance scores were computed using a RFR, based on the average reduction in MSE across all decision trees. This analysis was used to evaluate the predictive contribution of each frequency bin to the regression task, independent from the final model used for flow estimation.

Figure 2(a) illustrates the predictive significance of different frequency components, highlighting that the most relevant information is primarily concentrated in the lower-frequency bands, specifically below 16 kHz.

The results obtained from the evaluation of the regression models align closely with those reported in [17], where the real UF flow signals were analysed using the same feature extraction methodology. In that study, 1000 linearly binned FFT features were extracted from 100 ms audio segments recorded with UM and feature importance was assessed using identical techniques.

The strong correspondence between our synthetic and real datasets suggests that the acoustic characteristics of the synthetic flow dataset exhibit patterns comparable to those observed in real UF tests. This reinforces the feasibility of using a synthetic dataset to train ML models for voiding flow estimation, providing a controlled and reproducible framework for algorithm development.

By leveraging a standardized synthetic dataset, researchers can systematically benchmark and compare different models under uniform conditions before applying them to real voiding events. This structured approach facilitates the optimization of model selection, ensuring that the most effective algorithms are identified, refined and validated before their deployment in real-world applications.

2) FEATURE EXTRACTION FOR CLASSIFICATION MODELS

Next we evaluate features importance using a classification-based approach, where audio segments are labeled with integer values between 1 and 50 ml/s. In our case, we do not consider flows greater than 50 ml/s, as these values are not typically observed in male UF [16]. Feature importance is evaluated based on the Gini impurity metric [38], which quantifies the weighted impurity of each frequency band and provides insight into its relative contribution to classification tasks. This metric is particularly useful in assessing how well a frequency band can separate different voiding flow categories.

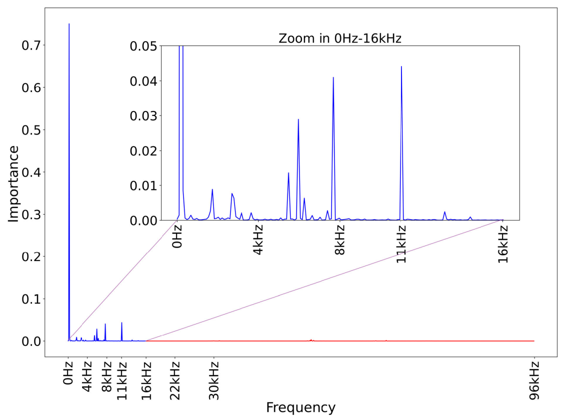
Figure 2 (b) illustrates the relative feature importance of each frequency component in the urinary flow estimation process using RFC. Similar to the regression-based approach, the results indicate that the most relevant frequencies are predominantly located in the lower spectrum, specifically below 16 kHz, further reinforcing the critical role of low-frequency components in flow estimation.

E. SELECTED FEATURES FOR THIS STUDY

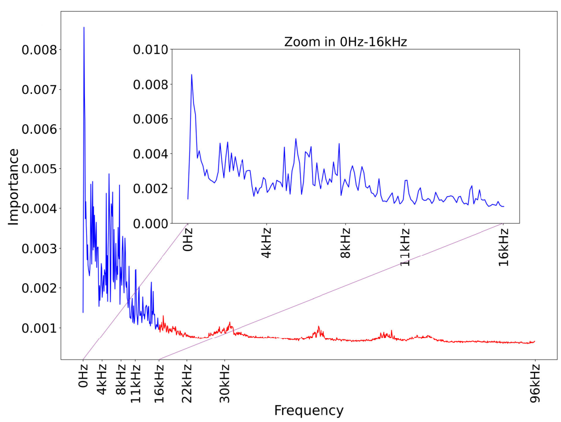
For the analysis of the audio signals, we evaluated various extraction features, including MFCC, ZCR and Chroma, each providing complementary information about the signal:

- MFCC: Capture the spectral and timbral structure of the sound, providing a compact representation of frequency characteristics relevant to voiding flow estimation.
- ZCR: Represent temporal information by quantifying the signal's granularity and its rhythmic elements.
- Chroma: Describe harmonic and tonal information, which may contribute to distinguishing different flow characteristics.

We trained a RFR regression model using different combinations of these features in order to identify the best representation for our signals. The results, presented in Figure 3, show that incorporating ZCR and 12 Chroma alongside the 13 MFCC did not yield better results than using MFCC alone. Given that MFCC provide equivalent performance with fewer features, the MFCC features are selected for the remaining analysis of this work to reduce model complexity while maintaining the predictive accuracy.



(a) Regression-based feature importance using MSE with the RFR



(b) Classification-based feature importance using Gini Impurity with the RFC

FIGURE 2. Relative feature importance of frequency components for voiding flow estimation using (a) regression (RFR) and (b) classification (RFC) approaches. Both methods reveal that the most relevant spectral regions (highlighted in blue) are concentrated in the lower frequency range, specifically below 16 kHz.

IV. RESULTS AND DISCUSSION

A. FLOW PREDICTION MODELS WITH SYNTHETIC DATASET

1) ANALYSIS OF REGRESSION MODELS

To select the model to be used in flow prediction, we performed an analysis of several regression models, including:

- RFR: A widely utilized model in numerous domains, including healthcare, known for its robustness and effectiveness in tackling complex regression problems with high-dimensional data [36], [37].
- SVR: A regression model capable of capturing complex non-linear relationships between acoustic features and voiding flow rates, making it suitable for flow estimation tasks [12].
- GBR: A robust ensemble learning method that incrementally refines predictions by correcting errors from previous iterations, making it well-suited for capturing complex patterns in voiding flow estimation [10], [39].

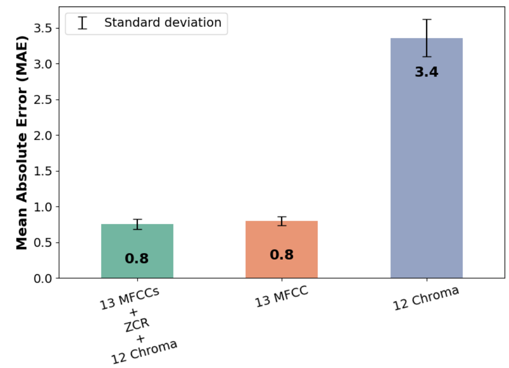


FIGURE 3. Evaluation of RFR performance with different combinations of audio feature inputs.

- CNN: A deep learning architecture particularly effective in processing grid-like data, such as images or time-series. CNN are commonly used in tasks such as image classification, speech recognition and natural language processing due to their ability to learn hierarchical feature representations [40]. In this context, CNN are fed with MFCC images, which represent the time-frequency structure of the audio signal, enabling the network to extract relevant patterns from the spectral content [11], [41].

All models were trained using a segment size of 1000 ms for each device, considering the frequency band from 0 to 16 kHz, where the most relevant information for flow prediction is concentrated. For the RFR, SVR and GBR models, hyperparameters were optimized using GridSearchCV with 10-fold cross-validation (K=10). The optimization included key parameters such as the number of estimators, tree depth and minimum samples per split for RFR; the number of estimators, learning rate and tree depth for GBR; and the regularization parameter, kernel coefficient and epsilon for SVR. In the case of the CNN, Keras Tuner was employed to explore the optimal combination of hyperparameters, including the number of filters in the convolutional layers, the dropout rate in various layers, the number of units in the dense layer and the application of L2 and L1 regularization in the dense layer.

Although CNN are commonly applied in audio analysis due to their ability to extract hierarchical features from spectrogram representations, their performance in this study was limited. The CNN models were trained on the synthetic dataset, which, while well-structured, is relatively small. This constraint led to signs of overfitting during training and poorer generalization compared to ensemble-based models such as Random Forest. Future work may address this limitation by applying data augmentation techniques or exploring more regularized CNN architectures.

The results obtained are shown in Figure 4. It can be observed that the RFR model achieved the best results in the flow estimation task. Therefore, for subsequent analyses, we will use RFR as the base model.

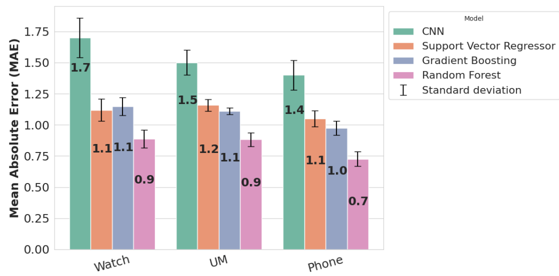


FIGURE 4. **Evaluation results of the four regression models for each recording device, in terms of the mean absolute error (MAE), measured in ml/s.**

a: ANALYSIS OF THE AUDIO SEGMENT DURATION

Once the RFR model was identified as the best-performing model in terms of regression metrics for each device, its flow prediction performance was further evaluated using different segment sizes within the 0–16 kHz frequency band. For this analysis, the audio signals were segmented into segments of 100, 200, 500 and 1000 ms. Figure 5 presents the results of the evaluation of the different segment sizes. It can be observed that as the segment duration increases, the MAE error decreases. Segments with durations longer than 1000 ms were not included, as increasing the segment size could compromise the system's resolution, which refers to its ability to estimate flow accurately over small time intervals.

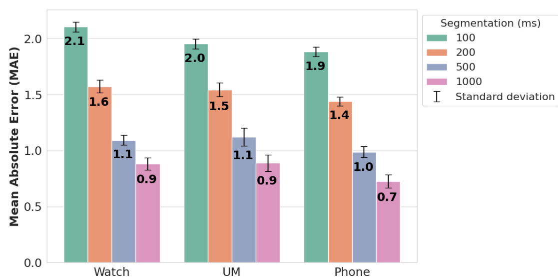


FIGURE 5. **Analysis of the MAE for the RFR prediction model, comparing different audio segment sizes (ms) and the three different recording devices.**

2) ANALYSIS OF CLASSIFICATION MODELS

Another way to approach flow prediction is as a multiclass classification problem, where audio segments are labeled with integer values between 1 and 50 ml/s. We do not consider flows greater than 50 ml/s, as these values are not observed in male UF [16].

Therefore, we trained a RFC model using the MFCC coefficients from the 0–16 kHz frequency band, extracted from 1000 ms audio segments as input features. For validation, we applied 10-fold cross-validation (K=10) and optimized the hyperparameters using GridSearchCV. The optimization included key parameters such as the number of estimators, tree depth and minimum samples per split for RFC.

To evaluate the model's performance, we used the quadratic weighted kappa (QWK) metric, in addition to accuracy. The QWK is a metric particularly useful in problems where the classes have an inherent order, such as in our case, where the flows range from 1 to 50 ml/s [42]. Unlike accuracy, which only measures the proportion of correct predictions, the QWK differentially penalizes errors based on the distance between the prediction and the true class. This is crucial in this problem, as a prediction error of 1 ml/s is much less significant than an error of 10 ml/s.

Figure 6 shows the evaluation results of the RFC model for each device. Although the accuracy metric provides an overall view of the classifier's performance, the QWK provides a more precise measure of how well the model maintains the ordinal relationship between the classes. The values of 1 obtained in the QWK for each device indicate that the model performs well and not only classifies correctly but also makes predictions that are close to the true value when it errs, which is key for the system's reliability in flow estimation.

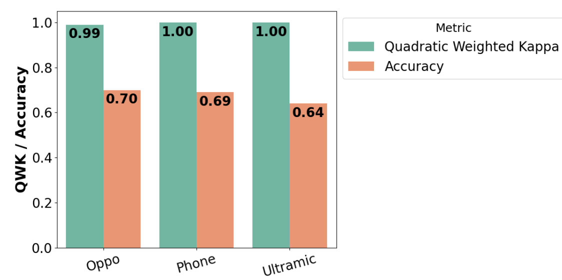


FIGURE 6. **Evaluation of the RFC model for each device in terms of QWK and accuracy.**

The following figure 7 presents the confusion matrix for the UM device, where it can be observed that the errors are close to the actual value when the system makes a mistake.

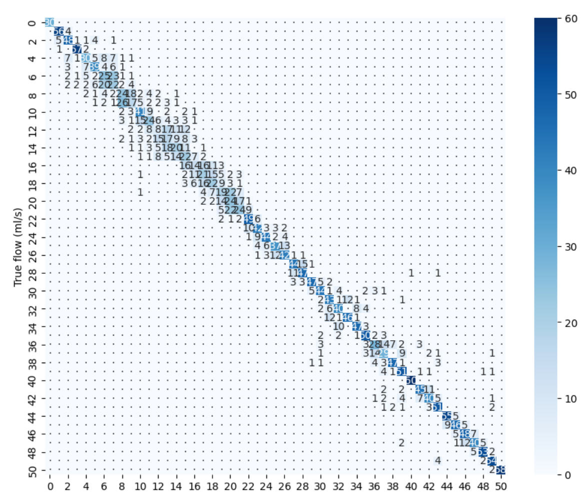


FIGURE 7. **Confusion matrix for the RFC model with UM audio recordings. The model was trained and evaluated using the synthetic dataset, considering a frequency range of 0–16 kHz and a segment size of 1000 ms.**

We also trained a CNN optimized for the flow classification task using Keras Tuner as the hyperparameter optimizer. The results were not better than those obtained using RFC model, as we obtained accuracy metrics of 0.49, 0.52 and 0.47 and QWK values of 0.80, 0.81 and 0.83 for the Watch, Phone and UM devices, respectively.

The results obtained from the evaluation of various regression and classification models for flow prediction over the synthetic dataset showed that using MFCC as input features with a segment size of 1000 ms and using RFR for regression and RFC for classification provided the best results in terms of MAE ± standard deviation (SD), accuracy and QWK. Table 2 presents the results obtained for the regression model and classification model that achieved the best results for the Watch, Phone and UM devices.

TABLE 2. Performance of RFR and RFC models on the synthetic dataset using MFCC features.

Device	RFR	RFC	
	MAE ± SD	Accuracy	QWK
Watch	0.9 ± 0.7	0.70	0.99
Phone	0.7 ± 0.06	0.69	1.00
UM	0.9 ± 0.06	0.64	1.00

It is also shown that for the flow prediction task, the use of ultrasound frequencies (20 - 96 kHz) is not necessary, as the most relevant information is found in the frequency band below 16 kHz, which explains why the results using the three recording devices are similar. Additionally, this validates the use of Watch as a viable and comfortable alternative for SU tests. The use of the Watch has advantages over the other devices in terms of ease of use and versatility, as it does not require patient intervention, making it particularly useful for individuals with limited digital experience, such as children and the elderly. Furthermore, its fixed position on the body ensures a constant recording distance and enables continuous monitoring of voiding events throughout the day, facilitating a more accurate analysis of potential alterations in the voiding pattern.

B. FLOW PREDICTION MODELS WITH REAL DATASET

After obtaining the best configurations of the models for flow prediction and classification trained on a synthetic audio dataset, we retrained and evaluated these models on real voiding dataset from [17]. For this evaluation, we used the same configurations and models that achieved the best results in flow prediction for the synthetic dataset. Table 3 presents the results of assessing the RFR model on the real labeled flow data for each device.

The evaluation results of the RFR model with the best hyperparameter configurations show a reduction in MAE of 0.7, 0.3 and 0.3 ml/s for the Watch, Phone and UM, respectively, compared to the results obtained in [17] for the same dataset.

The synthetic dataset, apart from containing a balanced number of samples for each flow rate value, it does not include possible human noises or artifacts that could affect the results. These two situations can explain the improved results.

Compared to previous SU studies, our approach provides reproducible and open methodologies, avoiding the dependency on additional parameters such as voided volume. While studies such as [9] and [10] report high correlation coefficients between SU and UF parameters, their flow estimation pipelines are not fully described or reproducible. In contrast, our RFR model achieves an MAE below 2.5 ml/s on real recordings, an improvement over prior work, such as [10], where only Lin's concordance coefficients were reported (0.77–0.85) and no direct error metrics were provided. Our method also does not require prior flow or volume knowledge, making it suitable for practical deployment and fair benchmarking.

For the classification model using RFC, the results are shown in Table 3. It can be observed that the model has low accuracy, but the QWK values greater than 0.90 show that the classification errors are close to the real values. However, it was observed that for the flow prediction task, regression models are more suitable.

TABLE 3. Evaluation results of the RFR and RFC models on real voiding events using MFCC features.

Device	RFR			RFC	
	MAE ± SD	MSE ± SD	R^2	Accuracy	QWK
Watch	2.20 ± 0.20	2.92 ± 0.22	0.88	0.21	0.93
Phone	2.18 ± 0.21	2.89 ± 0.26	0.88	0.19	0.92
UM	2.29 ± 0.15	3.08 ± 0.18	0.86	0.20	0.91

1) ANALYSIS OF PRIVACY-PRESERVING ENVIRONMENTS

After validating our models on real SU audio within the 0–16 kHz frequency range, we conducted an analysis of their performance in scenarios where user privacy must be preserved. Specifically, we examined the impact of removing frequency bands that contain identifiable speech information to assess the feasibility of privacy-preserving flow estimation.

To this end, we evaluated the performance of the flow prediction and classification models under two conditions, aiming at preserving user privacy by removing frequency bands containing identifiable speech information.

1) Removing the human conversational band 0–8 kHz, retaining only the 8–16 kHz range.
2) Removing only the primary human speech band 0–3.4 kHz, retaining the 3.4–16 kHz range.

The first approach ensures that most speech content is removed, preserving only higher-frequency components that may still contribute to flow estimation. The second approach retains additional spectral information beyond conversational speech while still filtering out the lower-frequency components associated with voice intelligibility.

These band selections are supported by well established literature: 3.4 kHz is the upper bound of narrowband speech in telecommunication standards, while 8 kHz encompasses the broader conversational speech spectrum [43]. The first approach ensures that nearly all intelligible content is excluded, while the second retains additional spectral information beyond the critical speech region.

Importantly, we did not filter or modify the original audio waveform. Instead, we employed a frequency-limited feature extraction approach using `librosa.feature.mfcc` from the Librosa library [44], with parameter settings `fmin=3400` or `8000` and `fmax=16000`. This configuration computes MFCC coefficients solely from the selected frequency bands, effectively excluding all content within the removed speech regions. This method ensures that only non-speech-related frequency components are used as input features, preserving user privacy while maintaining model integrity.

To assess the effect of these frequency restrictions, we trained and validated the models using the 3.4–16 kHz and 8–16 kHz bands, applying MFCC with a frame duration of 1000 ms. Table 4 presents the comparative performance of the RFR and RFC models when operating in the 8–16 kHz and 3.4–16 kHz frequency bands.

TABLE 4. Performance of flow prediction models on real voiding events using MFCC in privacy-preserving environments with a segment size of 1000 ms.

Device	RFR			RFC	
	MAE ± SD	MSE ± SD	R^2	Accuracy	QWK
8–16 kHz					
Watch	2.98 ± 0.26	3.96 ± 0.33	0.77	0.18	0.85
Phone	2.95 ± 0.23	3.98 ± 0.33	0.77	0.16	0.85
UM	2.88 ± 0.25	4.04 ± 0.30	0.76	0.18	0.84
3.4–16 kHz					
Watch	2.89 ± 0.25	3.90 ± 0.26	0.78	0.18	0.88
Phone	2.88 ± 0.27	3.85 ± 0.32	0.79	0.17	0.86
UM	2.91 ± 0.15	4.07 ± 0.03	0.76	0.19	0.84

The results indicate that removing lower-frequency bands (either 0–8 kHz or 0–3.4 kHz) leads to an increase in prediction error, confirming that most predictive information is concentrated in the lower part of the spectrum. However, despite this reduction in accuracy, the models maintained a reasonable level of performance, suggesting that flow prediction remains feasible even when privacy-preserving measures are applied. Between the two approaches, filtering only the 0–3.4 kHz band resulted in a slightly better performance compared to removing the entire 0–8 kHz band.

These findings highlight a potential trade-off between accuracy and privacy. While eliminating the conversational band (0–3.4 or 0–8 kHz) reduces the model's effectiveness, it offers a viable alternative for applications where protecting voice information is a priority. Further optimization, such as refining feature extraction methods or leveraging alternative frequency-based representations, could help mitigate the loss of accuracy while preserving privacy.

V. CONCLUSION

In conclusion, the use of a balanced synthetic void flow dataset recorded with three different devices has proven to be an effective strategy for identifying ML models capable of improving the existing models for real void flow prediction. These models can then be retrained to predict flow in real SU signals. This approach not only offers a practical solution in the absence of publicly available and balanced datasets for voiding flow sounds but also enhances the adaptability of models to different recording devices.

Furthermore, the availability of this dataset enables researchers to systematically evaluate the performance of various algorithmic approaches and make objective comparisons with previous studies. By providing a standardized reference framework, it facilitates future research in this field.

Once the most effective models are identified using the synthetic dataset, retraining them on real SU recordings allows for adaptation to natural acoustic conditions. This hybrid approach maximizes generalization by leveraging the consistency of synthetic training while tuning to the variability of real-world environments. Such a workflow supports the development of robust models with clinical applicability.

Experimental results indicate that models trained on synthetic flow data effectively adapted when retrained with real SU signals. Specifically, the evaluation of the RFR model with the best hyperparameter configurations demonstrated a reduction in MAE of 0.7, 0.3 and 0.3 ml/s for the Watch, Phone and UM, respectively, compared to the results reported in [17] for the same dataset. These findings confirm the feasibility of leveraging synthetic data to improve flow estimation performance in real-world scenarios.

Additionally, the synthetic dataset creates opportunities to explore the impact of environmental noise on estimation accuracy. Since recordings were conducted in a controlled environment with minimal background noise, future experiments could introduce varying levels of noise to assess how estimation errors fluctuate across different acoustic scenarios. Such analyses would contribute to the development of new strategies that enhance models robustness in different environment conditions.

The ability of these models to generalize is a crucial factor in establishing SU as a clinically viable and validated alternative to UF. Therefore, integrating synthetic flow data with real-world signals and assessing their performance across diverse acoustic environments represents a critical step toward advancing this technology for clinical applications.

Finally, although our system involves audio acquisition via mobile or wearable devices, the primary processing is designed to occur in the cloud. The random forest models, exported in open neural network exchange (ONNX, an open format for interoperable machine learning models) format, have compact sizes (approximately 8.9 MB) and demonstrate efficient inference (approximately 0.2 ms per 1-second segment) when evaluated on a server-like environment (Intel Xeon @ 2.20 GHz). This hybrid architecture ensures minimal computational demand on edge devices, which act only as audio acquisition and transmission units, enabling real-time applications without requiring local inference capabilities. This consideration aligns with recent advances in computational efficiency in AI [45].

AUTHOR CONTRIBUTIONS STATEMENT

Marcos Lazaro Alvarez: Writing–review and editing, Writing–original draft, Validation, Methodology, Investigation, Conceptualization. Alfonso Bahillo: Writing–review and editing, Project administration, Conceptualization. Laura Arjona and Diogo Marcelo Nogueira: Writing–review and editing, Software, Methodology, Formal analysis, Conceptualization. Elsa Ferreira Gomes and Alípio M. Jorge: Writing–review and editing, Supervision, Methodology, Investigation.

ADDITIONAL INFORMATION

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AVAILABILITY OF DATA AND MATERIALS

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

REFERENCES

[1] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nature Med.*, vol. 25, no. 1, pp. 44–56, Jan. 2019.

[2] C. Krittanawong, H. Zhang, Z. Wang, M. Aydar, and T. Kitai, "Artificial intelligence in precision cardiovascular medicine," *J. Amer. College Cardiology*, vol. 69, no. 21, pp. 2657–2664, 2017.

[3] N. Alothmany, H. Mosli, M. Shokoueinejad, R. Alkashgari, M. Chiang, and J. G. Webster, "Critical review of uroflowmetry methods," *J. Med. Biol. Eng.*, vol. 38, no. 5, pp. 685–696, Oct. 2018.

[4] M. Fernández Arjona and I. Pereira Sanz, "Hiperplasia benigna de próstata: Una afección de elevada prevalencia en el paciente de edad avanzada," *Revista Española de Geriatría y Gerontología*, vol. 43, no. 1, pp. 44–51, Jan. 2008.

[5] K. L. J. Kuoch, D. Meyer, D. W. Austin, and S. R. Knowles, "Classification and differentiation of bladder and bowel related anxieties: A socio-cognitive exploration," *Current Psychol.*, vol. 40, no. 8, pp. 4004–4011, Aug. 2021.

[6] G. S. Sonke, C. Robertson, A. L. M. Verbeek, W. P. J. Witjes, J. J. M. C. H. de la Rosette, and L. A. Kiemeney, "A method for estimating within-patient variability in maximal urinary flow rate adjusted for voided volume," *Urology*, vol. 59, no. 3, pp. 368–372, Mar. 2002.

[7] J. Krhut, M. Gärtner, R. Sýkora, P. Hurtík, M. Burda, L. Luňáček, K. Zvarová, and P. Zvara, "Comparison between uroflowmetry and sonouroflowmetry in recording of urinary flow in healthy men," *Int. J. Urol.*, vol. 22, no. 8, pp. 761–765, Aug. 2015.

[8] M. GÄrtner, J. Krhut, P. Hurtik, M. Burda, K. Zvarova, and P. Zvara, "Evaluation of voiding parameters in healthy women using sound analysis," *LUTS, Lower Urinary Tract Symptoms*, vol. 10, no. 1, pp. 12–16, Jan. 2018.

[9] Y. J. Lee, M. M. Kim, S. H. Song, and S. Lee, "A novel mobile acoustic uroflowmetry: Comparison with contemporary uroflowmetry," *Int. Neurourology J.*, vol. 25, no. 2, pp. 150–156, Jun. 2021.

[10] H. J. Lee, E. Aslim, B. Balamurali, L. Y. S. Ng, T. Kuo, K. S. Lim, C. Lin, C. J. Clarke, P. Privasarshinee, J. Jer-Ming, and L. G. Ng, "Development and validation of a deep learning system for sound-based prediction of urinary flow," *Eur. Urol.*, vol. 81, p. S1250, Feb. 2022.

[11] M. L. Alvarez, L. Arjona, M. E. Iglesias Martínez, and A. Bahillo, "Automatic classification of the physical surface in sound uroflowmetry using machine learning methods," *EURASIP J. Audio, Speech, Music Process.*, vol. 2024, no. 1, p. 12, Feb. 2024.

[12] 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 Innov.*, vol. 6, no. 4, pp. 199–203, Oct. 2020.

[13] L. Arjona, L. E. Díez, A. Bahillo, and A. Arruza-Echevarría, "UroSound: A smartwatch-based platform to perform non-intrusive sound-based uroflowmetry," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 5, pp. 2166–2177, May 2023.

[14] G. Narayanswamy, L. Arjona, L. E. Díez, A. Bahillo, and S. Patel, "Automatic classification of audio uroflowmetry with a smartwatch," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2022, pp. 4325–4329.

[15] E. El Helou, J. Naba, K. Youssef, G. Mjaess, G. Sleilaty, and S. Helou, "Mobile sonouroflowmetry using voiding sound and volume," *Sci. Rep.*, vol. 11, no. 1, p. 11250, May 2021.

[16] W. Schäfer, P. Abrams, L. Liao, A. Mattiasson, F. Pesce, A. Spangberg, A. M. Sterling, N. R. Zinner, and P. V. Kerrebroeck, "Good urodynamic practices: Uroflowmetry, filling cystometry, and pressure-flow studies," *Neurourol. Urodynamics, Off. J. Int. Continence Soc.*, vol. 21, no. 3, pp. 261–274, 2002.

[17] M. L. Alvarez, L. Arjona, M. Jojoa-Acosta, and A. Bahillo, "Flow prediction in sound-based uroflowmetry," *Sci. Rep.*, vol. 15, no. 1, p. 643, Jan. 2025.

[18] B. T. Balamurali, E. J. Aslim, Y. S. L. Ng, T. L. 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," in *Proc. Int. Conf. Signal Process. Commun. (SPCOM)*, Jul. 2020, pp. 1–5.

[19] M. Arumugam and M. Kaliappan, "An efficient approach for segmentation, feature extraction and classification of audio signals," *Circuits Syst.*, vol. 7, no. 4, pp. 255–279, 2016.

[20] D. M. Nogueira, C. A. Ferreira, E. F. Gomes, and A. M. Jorge, "Classifying heart sounds using images of motifs, MFCC and temporal features," *J. Med. Syst.*, vol. 43, no. 6, p. 168, Jun. 2019.

[21] Z. Zhang, S. Xu, S. Zhang, T. Qiao, and S. Cao, "Attention based convolutional recurrent neural network for environmental sound classification," *Neurocomputing*, vol. 453, pp. 896–903, Sep. 2021.

[22] F. Alías, J. Socoró, and X. Sevillano, "A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds," *Appl. Sci.*, vol. 6, no. 5, p. 143, May 2016.

[23] F. Korzeniowski and G. Widmer, "Feature learning for chord recognition: The deep chroma extractor," 2016, *arXiv:1612.05065*.

[24] U. Ayvaz, H. Gürüler, F. Khan, N. Ahmed, T. Whangbo, and A. A. Bobomirzaevich, "Automatic speaker recognition using mel-frequency cepstral coefficients through machine learning," *Comput., Mater. Continua*, vol. 71, no. 3, pp. 5511–5521, 2022.

[25] L. Arjona, Y. Iravantchi, A. Sample, M. L. Alvarez, A. Bahillo, and E. Canalon, "Privacy-preserving automatic collection of acoustic voiding events," in *Proc. 45th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2023, pp. 1–4.

[26] D. T. Phan, "Comparison performance of spectrogram and scalogram as input of acoustic recognition task," in *Proc. Future Inf. Commun. Conf.*, 2025, pp. 660–673.

[27] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Boochoon, and S. Birchfield, "Training deep networks with synthetic data: Bridging the reality gap by domain randomization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 1082–10828.

[28] M. Dawidek, R. Singla, L. Spooner, L. Ho, and C. Nguan, "Clinical validation of an audio-based uroflowmetry app in adult males," *Can. Urolog. Assoc. J.*, vol. 16, no. 3, p. 120, Jul. 2021.

[29] D.-G. Lee, J. Gerber, V. Bhatia, N. Janzen, P. F. Austin, C. J. Koh, and S. H. Song, "A prospective comparative study of mobile acoustic uroflowmetry and conventional uroflowmetry," *Int. Neurourol. J.*, vol. 25, no. 4, pp. 355–363, Dec. 2021.

[30] FluidControlSolutions. *Precise Laboratory Peristaltic Pump*. Accessed: Feb. 17, 2025. [Online]. Available: https://www.fluidcontrol24.com/systems/laboratory-peristaltic-pumps.html

[31] Dodotronic. *Ultramic 384K BLE*. Accessed: Jan. 17, 2025. [Online]. Available: https://www.dodotronic.com/product/ultramic-384k-ble/

[32] M. L. Alvarez, L. Arjona, A. Bahillo, and G. Bernardo-Seisdedos, "Annotated dataset of simulated voiding sound for urine flow estimation," *Sci. Data*, vol. 12, no. 1, pp. 1–7, Jun. 2025.

[33] M. L. Alvarez, L. Arjona, A. Bahillo, and G. Bernardo, "Annotated dataset of simulated voiding sound for urine flow estimation," *Figshare*, vol. 13, Jun. 2025, doi: 10.6084/m9.figshare.27606642.v1.

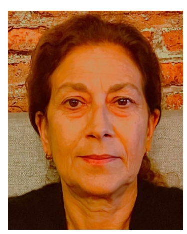
[34] M. Pradella, R. M. Dorizzi, F. Rigolin, and B. E. Statland, ''Relative density of urine: Methods and clinical significance,'' *CRC Crit. Rev. Clin. Lab. Sci.*, vol. 26, no. 3, pp. 195–242, Jan. 1988.
[35] (2024). *Homepage-Minze Health*. Accessed: Oct. 28, 2024. [Online]. Available: https://minzehealth.com/
[36] J. Hao and T. K. Ho, ''Machine learning made easy: A review of scikit-learn package in Python programming language,'' *J. Educ. Behav. Statist.*, vol. 44, no. 3, pp. 348–361, Jun. 2019.
[37] A. Cutler, D. R. Cutler, and J. R. Stevens, ''Random forests,'' *Mach. Learn.*, vol. 45, pp. 157–175, Jan. 2012.
[38] Y. Iravantchi, K. Ahuja, M. Goel, C. Harrison, and A. Sample, ''PrivacyMic: Utilizing inaudible frequencies for privacy preserving daily activity recognition,'' in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2021, pp. 1–13.
[39] J. H. Friedman, ''Greedy function approximation: A gradient boosting machine,'' *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
[40] Y. LeCun, Y. Bengio, and G. Hinton, ''Deep learning,'' *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[41] L. Deng and D. Yu, ''Deep learning: Methods and applications,'' *Found. Trends Signal Process.*, vol. 7, nos. 3–4, pp. 197–387, 2014.
[42] S. Vanbelle, ''A new interpretation of the weighted Kappa coefficients,'' *Psychometrika*, vol. 81, no. 2, pp. 399–410, Jun. 2016.
[43] W.-S. Hsu, ''Robust bandwidth extension of narrowband speech,'' Dept. Elect. Comput. Eng., McGill Univ., Montreal, QC, Canada, Tech. Rep., 2004.
[44] B. McFee, C. Raffel, D. Liang, D. P. W. Ellis, M. McVicar, E. Battenberg, and O. Nieto, ''Librosa: Audio and music signal analysis in Python,'' in *Proc. 14th Python Sci. Conf. (SciPy)*, K. Huff and J. Bergstra, Eds., Austin, TX, USA, Jul. 2015, pp. 18–25.
[45] D. T. Phan, T. A. Huynh, V. T. Pham, C. M. Tran, V. T. Mai, and N. Q. Tran, ''Optimal scalogram for computational complexity reduction in acoustic recognition using deep learning,'' 2025, *arXiv:2505.13017*.

MARCOS LAZARO ALVAREZ received the bachelor's degree in telecommunications and electronics engineering from the University of Pinar del Río, Cuba. He is currently pursuing the Ph.D. degree with the Faculty of Engineering, University of Deusto, Spain. Since 2016, he has been a Professor and a Researcher with UPR, focusing on software development and data science. In January 2022, he joined the Deusto Smart Mobility Research Group, University of Deusto, as a Researcher. His research interests include ML and deep learning applications in healthcare, particularly in sound-based medical diagnostics and biomedical signal processing.

ALFONSO BAHILLO received the degree in telecommunications engineering, in 2006, and the Ph.D. degree from the University of Valladolid, Spain, in 2010, and the PMP Certification with PMI, in 2014. From 2006 to 2010, he joined CEDETEL as a Research Engineer. From 2006 to 2011, he was an Assistant Professor with the University of Valladolid. From 2010 to 2012, he was with LUCE Innovative Technologies as the Product Owner. From 2013 to 2017, he held a postdoctoral position with the University of Deusto. From 2017 to 2020, he was the Director of DeustoTech-Fundación Deusto, University of Deusto. Currently, he is an Associate Professor with Universidad de Valladolid. He has worked (leading some of them) in more than 35 regional, national, and international research projects and contracts. He has co-authored more than 40 research articles published in international journals, more than 45 communications in international conferences, and three national patents. His research interests include local and global positioning techniques, ambient assisted living, biomedical and health informatics, and smart cities.

LAURA ARJONA received the B.Sc. degree in telecommunications engineering from the University of Granada, the M.Sc. degree from UNED University, and the Ph.D. degree from the University of Deusto. She is currently an Assistant Professor with the Faculty of Engineering, University of Deusto. She was awarded two different postdoctoral fellowships. Firstly, she was awarded the Washington Research Foundation Innovation Postdoctoral Fellow in Neuroengineering, where she developed a wireless communication protocol for backscattered-based neural implants working at the Paul. G. Allen School of Computer Science and Engineering, University of Washington, Seattle. Next, she was awarded a Juan de La Cierva Postdoctoral Research Fellowship, working on developing non-invasive digital health applications for ambient assisted living, in particular, to help diagnose lower urinary tract symptoms with sound-based uroflowmetries. Her research interests include wireless, mobile, and remote sensing to promote health and wellbeing, and embedded systems.

DIOGO MARCELO NOGUEIRA received the B.Sc. degree in biomedical engineering from Universidade de Trás-os-Montes e Alto Douro, and the M.Sc. degree in medical physics from the Faculty of Sciences, University of Porto, Portugal, where he is currently pursuing the Ph.D. degree in computer science. He has been a Researcher with INESC TEC, since 2012. From 2012 to 2016, he developed his activity at the Center for Applied Photonics. Since 2016, he has belonged to the Laboratory of Artificial Intelligence and Decision Support (LIAAD). His research interests include data mining and machine learning, particularly in health data. From 2018 to 2021, he was with the Data Intelligence Service, Centro Hospitalar Universitário São João, Porto. Since 2018, he has been a Visiting Professor with the Informatics Engineering Department, Institute of Engineering, Polytechnic Institute of Porto, teaching courses related to data analysis, artificial intelligence, machine learning, and programming.



ELSA FERREIRA GOMES received the bachelor's, master's, and Ph.D. degrees from the University of Porto, in 1990, 1999, and 2006, respectively. She is currently a Coordinating with ISEP, Polytechnic of Porto, and a Senior Researcher with the Artificial Intelligence and Decision Support Laboratory, INESC TEC. Her research interests include the application of machine learning and deep learning techniques to sound analysis and biomedical and healthcare applications.

ALÍPIO M. JORGE is currently pursuing the Ph.D. degree in computer science with the University of Porto. He is a Full Professor with the Department of Computer Science, Faculdade de Ciências da Universidade do Porto (FCUP) and is a Coordinator with LIAAD, Artificial Intelligence and Decision Support Laboratory, INESC TEC. He leads research projects in the areas of natural language processing, machine learning, data science, recommender systems, and artificial intelligence. He has co-chaired international conferences (ECML/PKDD 2005, 2015, and 2025, DSAA 2022, and Discovery Science 2009), workshops, and seminars. He coordinated the Portuguese strategy for AI, in 2019.

Chapter 6

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

6.1 Introduction

This chapter presents the fourth article published as part of the compendium of this doctoral thesis. The study addresses the automatic classification of the physical surface where the urinary stream impacts (water or ceramic), as well as the estimation of absence of voiding (silence), using ML techniques applied to acoustic signals collected in SU tests.

The motivation for this work arises from the recognition that the acoustic characteristics of the urinary stream depend significantly on the type of surface it impacts. In conventional practice, it is assumed that voiding impacts the toilet water, but in reality—especially among elderly patients—this is not always the case. When the stream hits ceramic, the acoustic properties of the signal are altered, introducing errors in the estimation of urodynamic parameters. This study proposes a novel method to automatically identify the type of surface, aiming to improve system robustness and eliminate the need to require patients to aim specifically at the water during SU tests.

6.2 Reference

Álvarez, M. L., Arjona, L., Iglesias Martínez, M. E., & Bahillo, A. (2024). *Automatic classification of the physical surface in sound uroflowmetry using machine learning methods*. **EURASIP Journal on Audio, Speech and Music Processing**, 2024(1), Article 12. https://doi.org/10.1186/s13636-024-00278-0

6.3 Status

Table 6.1 shows the publication status of the paper presented in this chapter.

Publication status	Published
Journal	EURASIP Journal on Audio, Speech, and Music Processing
Publisher	Springer
Year	2024
JCR Impact Factor (2024)	1.9
JCR Quartile	Q2 in Acoustics
DOI	https://doi.org/10.1186/s13636-024-00332-y
Citations (as of September 2025)	3 (according to Scopus)

Table 6.1: Publication details for Paper 4

6.4 Summary and contributions

This article introduces the first ML-based approach to identify the type of impact surface (water or ceramic) in SU tests conducted on male patients. A dataset of 6481 one-second audio clips was created by recording the voiding sounds of 15 male participants in four different bathrooms.

Supervised models (SVM, k-NN, RF) were trained on spectral features (20-linear-bin FFT) extracted from the audio signals. Three frequency bands were evaluated: full band (0–22.05 kHz), low frequency (0–8 kHz) and high frequency (8–22.05 kHz). The latter is particularly relevant, as it excludes the human-audible range and addresses the impact on results in applications where privacy must be preserved during SU testing.

Among the models studied, the RF model performed best:

- **Overall accuracy:** 99.46% (full band (0–22.05 kHz)).
- **F1-score:** 99.46%.
- **Accuracy in high-frequency band (8–22.05 kHz):** 93.29%, enabling applications where user privacy must be preserved without capturing information within the human voice range.

Additionally, a validation was conducted using 15 additional mixed voiding events not used during training. Each event contained segments with ceramic, water and silence (absence of voiding) within the same episode. The results achieved 98.17% accuracy, demonstrating strong generalization capability.

6.4.1 Key contributions

- First automatic classifier of surface type (water or ceramic) in SU, based on real human voiding data.
- High-accuracy results even in high-frequency bands, supporting implementations where user privacy is a concern.
- Key contribution toward more autonomous, robust and real-world-applicable SU systems.

6.4.2 Limitations noted by the authors

While the proposed classification model demonstrates strong performance under controlled conditions and generalizes well to realistic mixed recordings, several limitations remain.

- First, the dataset used for training and validation was recorded in a laboratory setting with minimal environmental noise, which may not fully reflect real-life home environments.
- Second, the model is currently limited to classifying three predefined classes (water, ceramic and silence) and does not yet consider edge cases such as simultaneous mixed surfaces or background noise interference.

Future work should include testing the model in a wider variety of home-like acoustic environments, including different toilet types, bathroom acoustics and user behaviors. It would also be valuable to explore data augmentation strategies and noise-robust features to further enhance generalization. Finally, integrating surface classification directly into the pipeline for real-time flow estimation could improve the overall robustness of SU systems.

6.5 Full article

The final version of the article is included below:

Alvarez *et al.*
EURASIP Journal on Audio, Speech, and Music Processing (2024) 2024:12
https://doi.org/10.1186/s13636-024-00332-y

EURASIP Journal on Audio, Speech, and Music Processing

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.

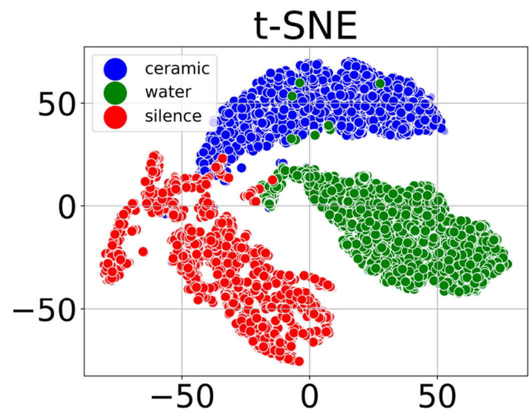


Fig. 2 t-SNE plot that shows that the ceramic (blue), water (green) and silent (red) classes can be distinguished well

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: 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-

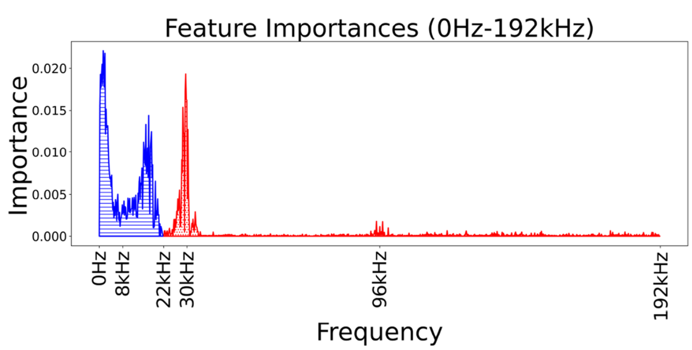


Fig. 1 Predictive power (importance) of each frequency component in the classification task with three classes: ceramic, water and silence. The frequency band selected in our algorithms is shown in blue. The importance is calculated using the Gini impurity with a random forest model

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.

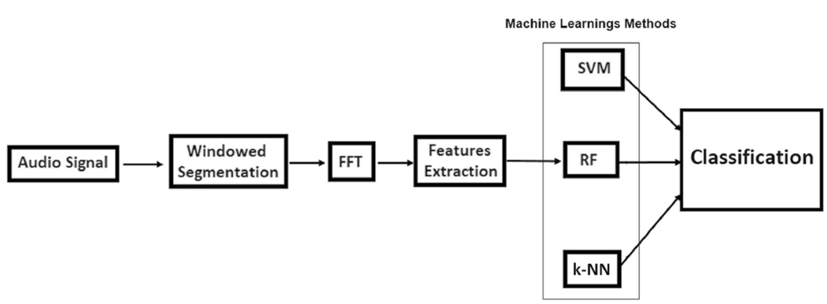


Fig. 3 Diagram showing the pipeline of the proposed methodology

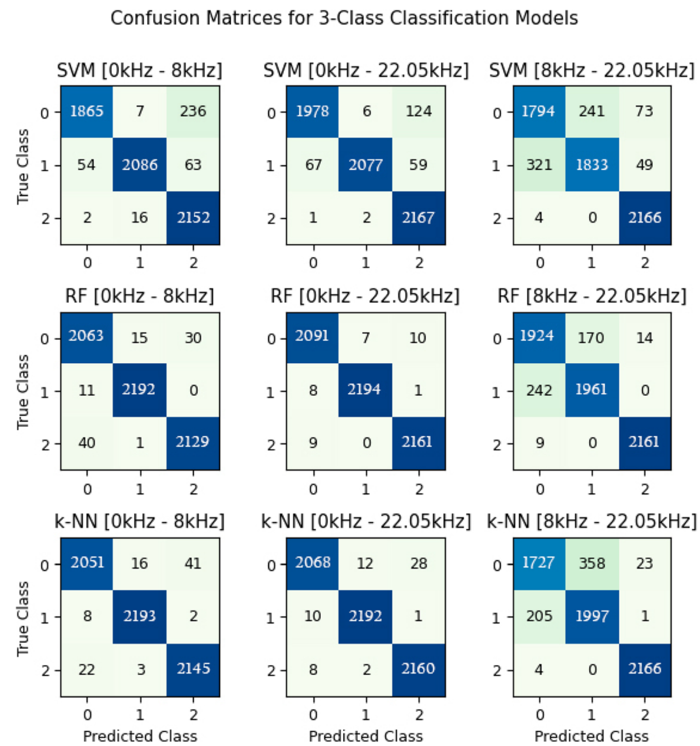


Fig. 4 Confusion matrices for three-class classification models: ceramic class (0), water class (1) and silence class (2)

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

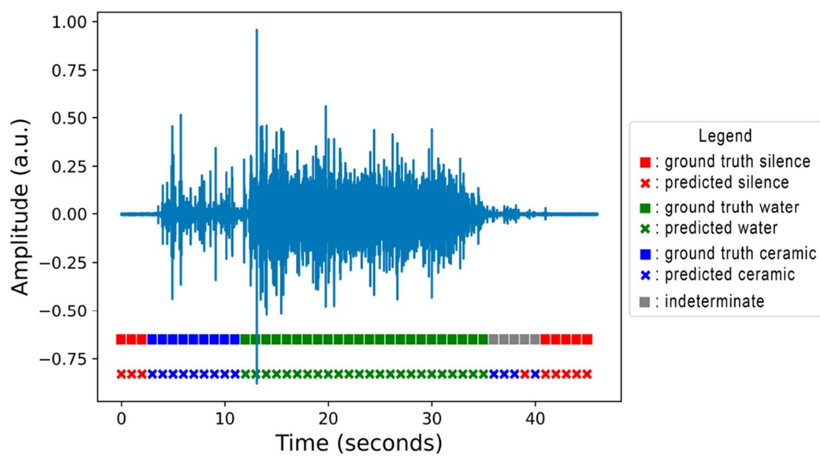


Fig. 5 Results for signal four, repetition one (see Table 3)

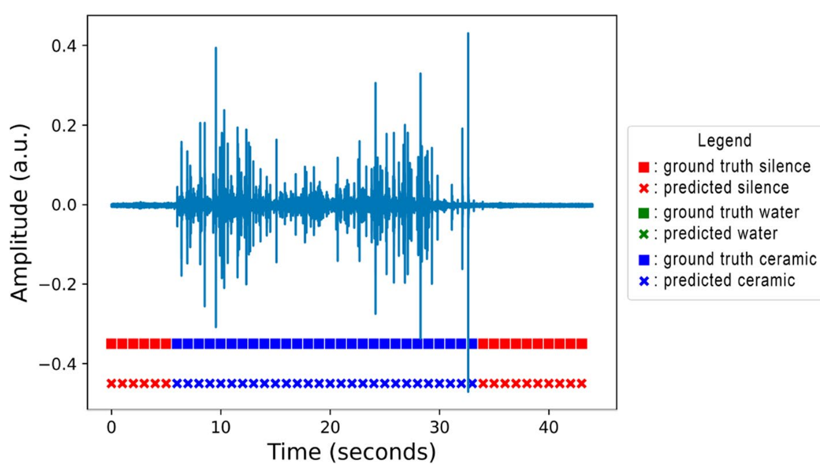


Fig. 6 Results for signal six, repetition two (see Table 3)

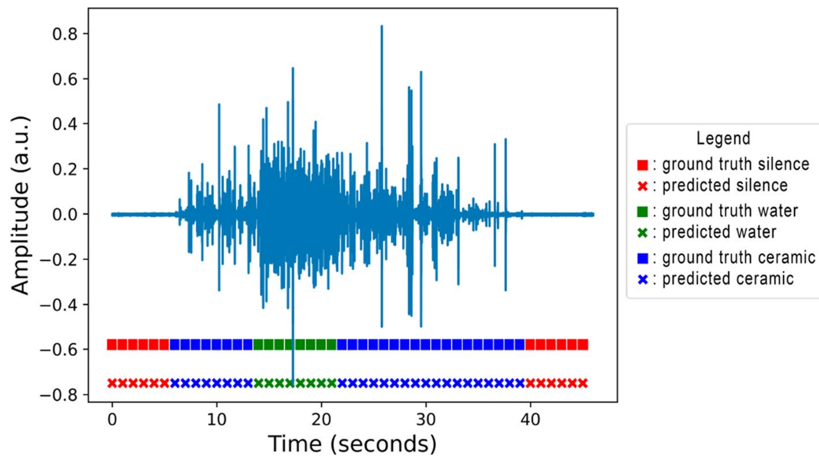


Fig. 7 Results for signal two, repetition one (see Table 3)

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
1. 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)
2. M.F. Arjona, I.P. Sanz, Hiperplasia benigna de próstata: una afección de elevada prevalencia en el paciente de edad avanzada. Rev. Esp. Geriatría Gerontol. **43**(1), 44–51 (2008)
3. 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)
4. 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)
5. 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)
6. 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)
7. N. Alothmany, H. Mosli, M. Shokoueinejad, R. Alkashgari, M. Chiang, J.G. Webster, Critical review of uroflowmetry methods. J. Med. Biol. Eng. **38**, 685–696 (2018)
8. 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)
9. 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)
10. P. Hurtík, M. Burda, J. Krhut, P. Zvara, L. Lunácek, 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
11. 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
12. 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. S56–S57, www.ics.org/2018/abstract/175
13. L. Arjona, L. E. Díez, 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
14. G. Narayanswamy, L. Arjona, L. E. Díez, 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
15. L. Bobrowski, T. Łukaszuk, Feature selection based on relaxed linear separability. Biocybernetics Biomed. Eng. **29**(2), 43–59 (2009)
16. K. Moreland, E. Angel. The FFT on a GPU. *In Proceedings of the ACM SIGGRAPH/EUROGRAPHICS conference on graphics hardware (HWWS '03)*. (Eurographics Association, Goslar, DEU, 2003), p112–119
17. 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)
18. N. Sato, Y. Obuchi, Emotion recognition using mel-frequency cepstral coefficients. Inf. Media Technol. **2**(3), 835–848 (2007)
19. 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=WWWAQAAQBAJ&oi=fnd&pg=PR1&dq=Principles,+Algorithms+and+Systems+(IGI+Global,+2010)&ots=TjFQaDCnjg&sig=hWE2xEdXHIGfq53nmlPIYuo2Fo#v=onepage&q=Principles%2C%20Algorithms%20and%20Systems%20(IGI%20Global%2C%202010)&f=false
20. 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)
21. 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/ICASSP.2004.1326738
22. 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
23. 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
24. P. Dhanalakshmi, S. Palanivel, V. Ramalingam, Classification of audio signals using SVM and RBFNN. Exp. Syst. Appl. **36**(3), 6069–6075 (2009)
25. 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
26. 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
27. 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. Brahnam, 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

Chapter 7

Conclusions

This chapter presents a synthesis of the main findings and contributions achieved throughout the development of this doctoral thesis. Across the various chapters, the potential of SU and AI methods for the non-invasive estimation of urodynamic parameters, as well as the classification of voiding patterns in real and simulated environments, has been demonstrated. In addition, two novel datasets of both real and synthetic voiding sounds were created, providing a valuable resource for training and validating ML models in this domain. Under the paradigm of SU, it has been shown that urinary flow can be captured, processed and interpreted using consumer-grade devices such as smartphones, smartwatches and commercial microphones. This represents a step forward in the telemonitoring of urodynamic parameters from a natural environment like the patient's home, making it more accessible and reducing long wait times and costs at clinical centers.

The results obtained across the different studies included in this thesis allow us to conclude that:

- Clinically relevant parameters such as VV, Qmax and Qave have been successfully estimated from acoustic signals captured under real-world conditions, achieving MAE below 2.5 ml/s and high concordance coefficients.

- A comprehensive dataset of real SU recordings has been created using synchronized measurements with a certified uroflowmeter. This resource is available upon request and supports future research and model validation.

- A synthetic SU dataset has been created under controlled conditions using a precision peristaltic pump. This publicly available resource promotes reproducibility and enables objective model comparisons in the field.

- A hybrid modeling strategy has been validated, in which models pretrained on synthetic data and then retrained on real data have achieved performance comparable to models trained exclusively on real data.

- A system for automatic classification of the voiding impact surface (water or ceramic) has been created, significantly improving the robustness of predictive models and removing unrealistic clinical constraints such as requiring users to aim only at the water surface.
- The use of frequency bands above 8 kHz has been shown to preserve user privacy while maintaining model performance, supporting ethical and practical deployment in home environments.

This work lays the foundation for the development of clinical solutions that are accessible, autonomous and privacy-conscious, aligned with the principles of personalized medicine and telehealth monitoring. These contributions have been formalized through multiple publications in indexed journals and international conferences, consolidating the academic impact of this research.

7.1 General conclusions

The results achieved throughout this doctoral thesis have successfully fulfilled the general objective, which focused on the development of a non-invasive and accessible system for estimating voiding parameters from acoustic signals using ML techniques in both simulated and real environments. To accomplish this goal, five specific objectives were systematically addressed, each translated into a set of studies or experimental developments presented in different chapters of this thesis.

Table 7.1 provides a schematic summary of the correspondence between the specific objectives presented in the introduction and the chapters in which the related work was carried out. This overview helps to visualize how each component of the theoretical-methodological framework materialized into concrete contributions—from acoustic environment classification and voiding parameter estimation to privacy analysis and cross-validation with synthetic data. These contributions consolidate the feasibility of SU as a robust, reproducible and privacy-respecting alternative, with potential for home and clinical applications.

Table 7.1: Correspondence between specific objectives and related work.

SO	Description of the Specific Objective	Corresponding Work or Chapter
1	Create a labeled dataset of real SU recordings from urination events, using synchronized measurements obtained with a commercial and medically certified Minze uroflowmeter as ground truth.	Chapter 3: Dataset generation and validation using real recordings.
2	Develop regression models to estimate Qmax, Qave and VV from acoustic signals recorded with smartphones/smartwatches.	Chapters 3 and 5: Estimation of urodynamic parameters using accessible devices.
3	Create and publish a synthetic dataset with accurate annotations and controlled flow rates.	Chapter 4: Generation of synthetic dataset and controlled labeling.
4	Evaluate the generalization capability of models trained with synthetic data when applied to real recordings.	Chapter 5: Model transfer and synthetic-to-real cross-validation.
5	Design ML models to classify the voiding surface (water or ceramic), reducing errors due to impact surface variability.	Chapter 6: Classification of voiding impact surface in SU using ML.
6	Analyze the impact of >8 kHz frequency bands on flow estimation, with a focus on privacy.	Chapters 6, 3 and 5: Acoustic privacy study and high-frequency band analysis.

7.2 Scientific contributions

This doctoral thesis has been developed within the framework of a predoctoral grant awarded under the 2020 call for predoctoral contracts (FPI2020) (ref. PRE2020-095612). Throughout the research process, a total of four scientific articles have been published in indexed journals, including two in Q1 journals with impact factors of 3.9 and 6.4 and two in Q2 journals with impact factors of 1.9 and 3.6. Additionally, one conference paper was presented at the 45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE EMBC 2023), held in Sydney, Australia, contributing to the international dissemination of the results.

The research was further enriched by an international mobility stay at the INESC TEC – Institute for Systems and Computer Engineering, Technology and Science in Porto, Portugal, fostering collaboration and knowledge exchange at the international level. Additionally, a national research stay was carried out at the Artificial Intelligence and Knowledge Engineering (AIKE) research group of the University of Murcia, Spain, contributing to the development of a novel dataset of simulated urination sounds impacting ceramic surfaces and supporting the advancement of methodologies for urinary flow estimation.

The academic journey also included four formal seminar presentations aligned with the PhD progress milestones, along with participation in eight external seminars and twelve specialized doctoral courses. Moreover, participation in scientific events has been recognized through two official certificates. Table 7.2 provides a detailed summary of all these scientific contributions.

S.No	Title	Number	Description
1	PhD grant	1	FPI 2020 (ref. PRE2020-095612)
2	Journal articles	4	IF 1.9 (Q2), 3.9 (Q1), 6.4 (Q1) and 3.6 (Q2)
3	Conference articles	1	IEEE EMBC 2023, Sydney, Australia
4	International mobility	1	INESC TEC, Porto, Portugal
5	National mobility	1	AIKE research group, Murcia, Spain
6	Seminars presented	4	One per year aligned with PhD progress
7	Seminars attended	8	Related to doctoral research
8	Study courses	12	Specialized PhD training courses
9	Certificates	1	Conference presentation certificates

Table 7.2: Scientific contribution summary.

7.2.1 Published articles in indexed journals

This doctoral thesis has led to the publication of four scientific articles in peer-reviewed journals, each aligned with the core objectives of this research.

The **first article**, published in a Q1 journal in the Multidisciplinary Sciences category with an impact factor of 3.9, involved the creation of a real SU dataset (47 validated cases) using a certified uroflowmeter and multiple recording devices. It proposes a ML system for estimating urinary flow and VV from acoustic signals, achieving high concordance with UF. Further details can be found in Table 3.1.

The **second article**, published in a Q1 journal in the Multidisciplinary Sciences category with an impact factor of 6.4, focuses on the design, simulation and public release of a synthetic micturition dataset with precise flow control and annotation. This resource contributes to reproducible research in SU. The corresponding publication is described in Table 4.1.

The **third article**, published in a Q2 journal in Computer Science, Information Systems category with an impact factor of 3.6, explores model generalization through validation on real clinical data after training on synthetic recordings. This work confirms the practical transferability of ML models across environments and is summarized in Table 5.1.

The **fourth article**, published in a Q2 journal in Acoustics category with an impact factor of 1.9, introduces a methodology for detecting the micturition surface (water or ceramic) using supervised classification models. This work validates the feasibility of acoustic environment compensation for flow parameter estimation, as presented in Table 6.1.

Each of these publications corresponds to a dedicated chapter in the thesis and collectively validates the robustness, applicability and reproducibility of the proposed system for non-invasive urinary flow estimation.

7.2.2 Conference presentation: IEEE EMBC 2023

Table 7.3 summarizes the contribution to the international conference derived from the early stages of this doctoral research. The presented paper introduced a novel method for classifying the physical surface of urination (water or ceramic) using sound-based analysis. This task addresses one of the main challenges in SU systems: the acoustic distortion produced when the urine stream impacts different materials within the toilet bowl. Unlike the approach presented in Chapter 6, this early study did not rely on ML techniques. Instead, it employed classical signal processing methods such as the FFT, variance and kurtosis, and was validated using a limited dataset comprising 24 experimental recordings. This foundational work enabled the development of more advanced methodologies later in the PhD, including the study published in [22], which leveraged ML models (RF, SVM and k-NN) to classify three surface types (water, ceramic and silence) based on a substantially larger dataset of over 6,000 one-second audio segments collected in four different bathroom environments. This evolution reflects the methodological transition from traditional acoustic feature engineering to automated, data-driven classification under realistic conditions. The conference study directly contributed to Objective 1 of the thesis and is detailed in the IEEE EMBC 2023 proceedings, as shown in Table 7.3.

Conference	45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE EMBC 2023)
Date & Location	July 25–28, 2023: Sydney, Australia
Publisher	IEEE
Indexing	IEEE Xplore, Scopus
DOI	https://doi.org/10.1109/EMBC40787.2023.10340174
Title	Classification of the Physical Surface in Sound-Based Uroflowmetry
Authors	Marcos L. Alvarez, Miguel E. Iglesias Martínez, Laura Arjona, Alfonso Bahillo

Table 7.3: Conference publication details (IEEE EMBC 2023).

7.2.3 International mobility: INESC TEC

As part of the academic training during this doctoral thesis, an international research stay was conducted at INESC TEC – Institute for Systems and Computer Engineering, Technology and Science, located in Porto, Portugal. The stay took place from January 16 to April 16, 2025 and was aimed at strengthening scientific collaboration and advancing research activities in the field of sound-based non-invasive uroflowmetry.

During this three-month period, the research focused on the development of AI models for estimating urinary flow from audio recordings. The project, entitled *"Leveraging synthetic data to develop a ML model for voiding flow rate prediction from audio signals"*, formed the foundation for Chapter 5.

The main contributions of the stay included the segmentation and preprocessing of synthetic audio signals, extraction of acoustic features, model training and validation and subsequent transfer learning using real-world urination recordings. Moreover, a scientific article based on this research was prepared and submitted to a peer-reviewed international journal, demonstrating the scientific value of the work carried out during the stay.

This international experience not only contributed directly to achieving Objective 4 of the thesis—validating synthetic data models against real clinical data—but also reinforced institutional collaboration between the University of Deusto and INESC TEC, promoting knowledge exchange and international cooperation in biomedical signal processing and AI.

7.2.4 National mobility: AIKE research group

As part of the research activities developed during this doctoral thesis, a research mobility was carried out at the AIKE research group of the University of Murcia, from July 29 to August 29, 2025. This collaboration provided an excellent opportunity to advance research efforts in SU.

The work conducted during this mobility focused on the development of a novel dataset and methodologies for urinary flow estimation when the urine stream impacts ceramic toilet surfaces. The main activities included:

- Design and implementation of experimental protocols to generate synthetic urination sounds under controlled flow conditions.

- Acquisition of multi-device audio recordings (UM, Phone, Watch) in a real bathroom environment.

- Signal segmentation, preprocessing and extraction of acoustic features using MFCC.

- Preliminary training and evaluation of ML models for flow estimation.

- Drafting and preparation of a scientific article derived from this work.

This mobility represents a significant step toward the creation of robust and generalizable datasets for SU.

7.3 Study limitations

- The real clinical dataset consisted of 47 validated voiding recordings from healthy male participants aged 18–60, all in a standing position. The models have not yet been evaluated on female subjects or sitting voiding positions, which may present different acoustic characteristics.

- Recordings were conducted under relatively quiet, semi-controlled conditions, minimizing background noise and ensuring consistent microphone placement. In fully uncontrolled home environments, additional factors such as variable ambient noise, microphone distance and echo could affect signal quality and classification performance.

- Although the sample size was sufficient for exploratory analysis and initial validation, expanding the dataset to at least 300–350 recordings across diverse users (including different sexes, age groups and clinical backgrounds) would be necessary to confirm the robustness and generalizability of the proposed models.

7.4 Future research directions

- Expand multicenter clinical trials to include female, pediatric and urologically diverse populations to improve system generalization.

- Incorporate self-learning and continuous adaptation models to refine system performance based on individual users.
- Explore sound anonymization techniques and federated learning approaches to enhance privacy and data security.
- Develop a fully integrated mobile application including audio recording, real-time analysis and clinical feedback with user-friendly interfaces for patients and healthcare professionals.

Overall, this thesis demonstrates that combining SU signals with AI techniques holds significant potential to transform urological clinical practice by providing non-invasive, accessible and accurate diagnostic tools.

7.5 Funding declaration

This research has been primarily supported through the 2020 call for predoctoral contracts (*FPI 2020*) under the grant *Ayuda para contratos predoctorales 2020* (ref. PRE2020-095612), funded by the Spanish Ministry of Science and Innovation (MICI-U/AEI/10.13039/501100011033) and co-financed by the European Social Fund (FSE) under the program *"FSE invierte en tu futuro."*. Additionally, this work has received funding from the AGINPLACE project (PID2023-146254OB-C41 and PID2023-146254OA-C44), financed by MICIU/AEI/10.13039/501100011033 and by FEDER, European Union.

The research was conducted under the umbrella of the Deusto Smart Mobility research group at the University of Deusto. The author would also like to acknowledge the institutional support received from the Basque Government.

This thesis reflects only the author's views and the funding agencies are not responsible for any use that may be made of the information it contains.

Appendix A

Appendices

A.1 Conference Paper (IEEE EMBC 2023)

The following pages include the full version of the conference paper presented at IEEE EMBC 2023, titled:

"Classification of the Physical Surface in Sound-based Uroflowmetry"

2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340174

Classification of the Physical Surface in Sound-based Uroflowmetry

Marcos L. Alvarez[1], Miguel E. Iglesias Martínez[2], Laura Arjona[1] and Alfonso Bahillo[3]

***Abstract*— This work constitutes a first approach to automatically classify the urination medium for non-invasive sound based uroflowmetry tests. Often the voiding flow impacts the toilet wall (often made of ceramic) instead of the water. This causes a reduction in the amplitude of the recorded audio signal, and thus a reduction in the amplitude of the extracted envelope. Analysing the envelope alone, it is not possible to tell accurately if the reduction in the amplitude is due to a low voiding flow or an impact on the toilet walls. In this work, we carry out a study on the classification of sound uroflowmetry data depending on the medium where the urine impacts within the toilet: water or ceramic. In the analysis, a classification algorithm is proposed to identify the physical medium automatically based on the urination acoustics. The classification algorithm takes as input the frequency spectrum, the variance, and the kurtosis of the audio signal corresponding to a voiding event.**

***Clinical relevance*— Sound uroflowmetry has a strong correlation with the standard uroflowmetry. It is useful for the non-invasive detection of pathologies associated with the urinary tract as a support tool for information processing and screening. It consists of a characterization of the urinary flow patterns by capturing the sound generated when the urine stream impacts the water in the toilet. Identifying the medium which originates the sound is of paramount importance to better interpret the sound uroflowmetry.**

I. INTRODUCTION

One of the problems that is currently affecting the aging population is lower urinary tract symptoms (LUTS). LUTS are those affecting bladder storage and emptying and postvoiding, usually caused by an enlargement of the prostate. They can lead to a decrease in quality of life and a significant expenditure of health care resources [1]. It is estimated that more than 60% of the population of men over sixty years suffer from LUTS [2]. LUTS have a non-invasive and physiological test to assess lower urinary tract obstruction, the uroflowmetry (UF). This procedure needs to be performed in a clinic in a face-to-face setting, which creates situational stress for patients causing significant test-to-test variability [3]. Therefore, it is recommended to repeat the test more than once, which generates several, time-consuming and costly visits to the clinic. As a remote and proactive alternative to solve the limitations of UF, sound uroflowmetry (SU) has emerged, which characterizes flow patterns by capturing the sound generated when the urine stream reaches the water in the toilet bowl. The scientific literature has shown a good correlation between UF and SU in terms of the shape of the visual trace and some of the parameters obtained from the voided flow [4], [5]. In current works analyzing SU, dedicated microphones have been used as recording devices [6]. General-porpuse devices have also been used, including smartphones [5], [7], [8] and smartwatches [4], [9]. One of the main limitations of SU present in current systems is that it requires that the person points the entire urine flow towards the water base in the toilet bowl during the entire test (avoiding the toilet wall). If the urine hits the toilet wall instead of the water, it can be affect considerably the acoustic characteristics of the test making current SU systems invalid [4]. This is specially important in some European countries, where the water level is lower than the average in some other countries such as the United States of America. The requirement that patient must point to the water of the toilet bowl during the complete duration of the voiding event can be invalidate current SU tests.

The main objective of this work is to be able to classify the intervals of urination against ceramic and water, and thus to be able to automatically detect the medium where urination is performed and make a better assessment of the acoustic characteristics of the SU test.

II. RELATED WORK

Due to the non-existence of databases of voiding sounds in which voiding versus water and ceramic are present, we have generated a novel database of sounds. We have developed a classifier that employs summary statistics as efficient and flexible features. Currently, several works have been proposed that make use of a wide variety of algorithm for processing the acoustic signals in SU.

Classical sound labeling systems focus on predicting the content of an audio from the analysis of Fast Fourier Transforms (FFT) and Mel Frequency Cepstral Coefficients (MFCC), then the prediction is made by the use of machine learning (ML) algorithms for classification [10]. In [11] the spectrogram on logarithmic scale is proposed, in [12], it is used the gammatone cepstral coefficients (GTCC) and in [13] the MFCC. Several statistics have also been used to characterize the discriminant information in both time and frequency domains, such as the use of the zero crossing rate

[1]Marcos L. Alvarez, Laura Arjona are with the Faculty of Engineering, University of Deusto, Bilbao, 48007, Spain {alvarez.marcoslazaro, laura.arjona}@deusto.es

[2]Miguel E. Iglesias Martínez is with Grupo de Modelización interdisciplinar, from the Instituto Universitario de Matemática Pura y Aplicada, Universitat Politècnica de València, E-46022 Valencia, Spain and with Grupo de Ingeniería Física, Escuela de Ingeniería Aeronáutica y del Espacio, Universidad de Vigo, Edif. Manuel Martínez Risco, 32004 Ourense. Spain. miigmar@upv.es, miguel.iglesias.martinez@uvigo.gal

[3]Alfonso Bahillo is with the Department of Signal Theory and Communications, University of Valladolid, 47011, Spain alfonso.bahillo@uva.es

(ZCR) and the spectral kurtosis [14]. Another method uses the mean and standard deviations to convert time-frequency matrix representations into a compact feature vector [12]. Additionally, in [15] a novel approach to computational analysis of auditory scenes using time-frequency representations and discriminative content extraction is performed.

In our work, making use of the spectral characteristics with the FFT and statistical measures of kurtosis and variance of the audio signals, we have implemented an algorithm for classifying the medium for the SU tests. We discard the use of ML and deep learning models due to the limited size of our dataset.

III. MATERIAL AND METHODS

The tests were carried out in a typical European home bathroom, with dimensions of 2.5m x 1.5m from October 9th to 23th, 2022. For this first study, real voiding events were collected from a healthy volunteer (in a standing position), and in future works, we will expand the number of people, including both healthy people and patients. The goal is to identify the medium where the void impacts (ceramic or water). It is known that loudness is proportional to flow [5]. When the void impacts against the toilet wall, the loudness decreases and can cause uncertainty in the subsequent processing of the acoustic signal because it can be interpreted as a decrease in flow when in fact it is not. The section IV presents a novel algorithm to identify the medium and solve the mentioned uncertainty. The experimental procedures conform to the provisions of the Declaration of Helsinki (as revised in Edinburgh 2000).

During the same voiding event, the volunteer person pointed to the walls of the toilet and to the water. The Ultramic384K, a high sensitive digital audio microphone [16] was used to capture the acoustic energy, with a sample rate (SR) of 384kHz. It was located in a fixed position on the toilet tank (about 84 - 105 cm), allowing us to have no variability in terms of distance from one test to another. The intervals duration of voiding, ceramic, and silence between voiding events is approximately 3 seconds. The recording time for each sequence was 50 seconds. Table I summarizes the data set characteristics carried out in the experimental test. We collected 1171 seconds of audio, of which approximately 490, 362 and 319 seconds are silence, water and ceramic respectively.

TABLE I: DATA SET CHARACTERISTIC

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

IV. PROPOSED ALGORITHM

The process of automatic detection of the medium (water or ceramic) can be organized in 3 steps, presented next: IV-A Threshold detection, IV-B Outlier removal, and IV-C Validation.

A. Step 1: Threshold detection

Firstly, the initial 5 seconds of the recording are taken to obtain the background noise N_b threshold which is subtracted from the signal in the frequency domain [10]. As a quantitative indicator for the detection of ceramic and water, we use the Fourier Transform absolute value with a rectangular window of the original signal. The time series audio signal is analyzed in windows of 0.6-seconds-long with an overlap of 0.3%. The duration of the window and the overlap values were chosen experimentally from the results obtained using linear regression of audio recordings with syringes simulating constant flow and validating the values on a standard uroflowmeter.

To determine the medium where the void impacts (ceramic or water) we establish the following thresholds (silence, that is, no presence of void is also identified):

$$\frac{N_b}{A_f} \leq ceramic < 1 \quad | \quad water > 1 \quad | \quad silence < \frac{N_b}{A_f}$$

where the A_f factor is an adjustment factor to detect the lower limit of urination against ceramic. The value of A_f=3.2 was set experimentally from 8 audios . Based on these thresholds, we classify each signal frame (0.6-seconds-long) of the audio, obtaining a matrix of labels of silence, water or ceramic as appropriate. Fig. 4a shows the result.

B. Step 2: Outlier removal

We apply an algorithm to eliminate outliers that may exist in the signal in the previous step. The proposed algorithm takes the matrix of labels (silence, ceramic and water) obtained from the previous step and verifies the groups of isolated labels (outliers) that correspond to a duration of less than 0.85s (taking into account that each label corresponds to the classification of an audio segment of 0.6 seconds). If there are groups of outliers less than or equal to 2, these will change its value and will update their value with the classification value that had the signal in the previous frame. Fig. 4b shows the result.

C. Step 3: Validation Process

Once the ceramic, water and silence intervals have been obtained, the last step is validation. According to Step 1, there may be intervals where ceramic is identified when there really is voiding against water in which the voiding flow was slowly decreasing (or a voiding practically with drips) until it falls on the ceramic threshold. To solve this, we use a second filter based on the kurtosis (K) of these doubt interval. K is a statistical measure used to describe the degree to which scores cluster in the tails or the peak of a frequency distribution. To demonstrate the above, we evaluated K for 17 audio clips of different duration, between 2 and 24 seconds. 9 of them were voiding against water, and 8 voiding against ceramic. Fig. 1 shows the obtained results.

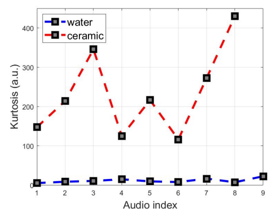


Fig. 1: Kurtosis of the ceramic and water audio signals intervals, respectively

This figure shows notable differences in K. Fig. 4c shows the result.

Table II shows the proposed K threshold values in this procedure. Additionally, if after applying the filter using K

TABLE II: K and MVAS threshold values

K value	MVAS values	Clasification label
$k \leq 28$	MVAS ≥ 15	water
$k \geq 35$	MVAS ≤ 8	ceramic
$28 < k < 35$	$8 <$ MVAS < 35	doubt

in the ceramic intervals the result is in the area of doubt, it is proposed to apply the mean of the variance of the absolute value of the spectrum (MVAS) in the intervals. Table II shows the MVAS threshold values proposed. These were calculated experimentally using the same audio clips used for K. Fig. 2 shows the obtained results.

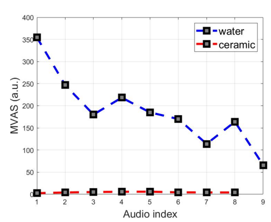


Fig. 2: MVAS of the ceramic and water audio signals intervals, respectively

Finally, if after applying the kurtusos and the MVAS filters uncertainty remains to identify the medium, it is proposed to calculate a probability percentage based on the two previous results (*K and MVAS threshold*), that will indicate whether or not the null hypothesis is discarded regarding the classification of the physical medium. In case of the result presents a probability higher or equal to 0.5, the classification result is accepted as valid. Fig. 3 shows the block diagram of the proposed methodology.

V. Results

This section presents the obtained results from the experiment for all the acoustic voiding signals used. We used

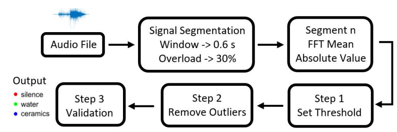
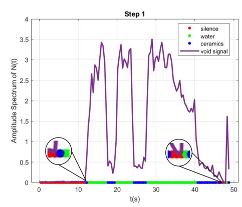
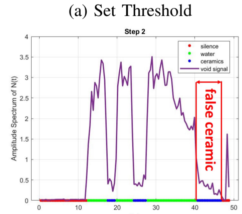


Fig. 3: Block diagram of the proposed algorithm.

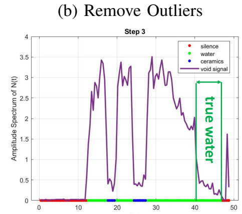
a frequency of SR = 384kHz, but tests were performed only using the signal information contained in the 0 - 44 kHz band, obtaining similar results. Therefore, devices with lower SR could be used for classification. Figures 4a, 4b and 4c show the obtained results using the proposed algorithm corresponding to steps 1, 2, 3. As an example, we take a signal that follows the voiding pattern corresponding to audio 1 repetition 1 of Table I. The sections labeled as red belong to silent zones, those labeled as green correspond to void against water and those labeled as blue, to void against ceramic.



(a) Set Threshold



(b) Remove Outliers



(c) Validation with kurtosis and MVAS

Fig. 4: Results for the signal 1 repetition 1, Table I

Fig. 4 shows that the classification of the physical medium is achieved according to the amplitude curve obtained once the proposed algorithm has been applied. Although there may be areas of doubt, an example of these is in Fig. 4a when urination is ending and the flow is decreasing, the algorithm classifies that last time interval as ceramic when it was truly urination against water, in the validation process these errors can be solved using the proposed K and MVAS validation as shown in fig. 4c. The correlation values obtained between the real interval and the interval classified by the proposed method for silence, water and ceramics of each one of the 24 audios were 0.8644, 0.9481 and 0.8960 respectively.

Finally, 1171 seconds of audio were analyzed, of which approximately 490, 362 and 319 were silence, water and ceramic respectively. Table III shows the mean relative error of the 24 audios samples for silence, water and ceramics intervals respectively. Urination against ceramics proved to be the worst performing case due to its acoustic similarity to the end of an urination event when the flow is very low or dripping on the water is present. Table III shows the results of the mean relative error on the real values and those obtained by the algorithm for the times of silence, water and ceramics of the 24 audios that were analyzed.

Among the interval prediction errors for the case of ceramics and water, 58.3% of the intervals are located in the final 4-6 seconds of the voiding events. Although these errors are slightly higher compared to most of the void duration, they lose a little meaning if one takes into account that according to urologists, the final seconds of the voiding event do not provide relevant information for screening.

TABLE III: Mean Relative error in Function of the Physical Surface

Physical medium	**Mean relative error (%)**
Silence	9.1
Water	12.1
Ceramics	16.5

VI. CONCLUSIONS

This work is a first approach in the study of the automatic classification of the physical medium in sound based uroflowmetry. Current SU systems have the limitation that patients must point the void towards the toilet water during the test. This is a limitation of current SU systems since most patients with problems related to the lower urinary tract are elderly and this requirement can rise rejections. This study will allow the development of SU systems for home environments that overcome the limitation mentioned. This paper has shown the theoretical and practical basis for the evaluation of a methodology to automatically classify the physical medium in acoustic uroflowmetry tests. A relative classification error and correlations values were obtained as a quantitative measure of the algorithm efficiency, which represents an improvement in acoustic uroflowmetry diagnostic tests. We are working on a more extensive study for future work, adding a more significant number of patients and performing the classification process using ML to generalize further and improve the results obtained.

ACKNOWLEDGMENT

This research was supported by the Spanish Ministry of Science and Innovation under the Peace Of Mind project (ref. PID2019-105470RB-C31) and ”Ayudas para contratos predoctorales para la formación de doctores 2020” (ref. PRE2020-095612). Miguel E. Iglesias Martínez's work was supported by the postdoctoral research scholarship “Ayuda para potenciar la investigación postdoctoral (PAID-PD-22)” from Universitat Politècnica de València.

REFERENCES

[1] M. F. Arjona and I. P. Sanz, “Hiperplasia benigna de próstata: una afección de elevada prevalencia en el paciente de edad avanzada,” *Revista Española de Geriatría y Gerontología*, vol. 43, no. 1, pp. 44–51, 2008.

[2] J. C. Santos and 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 Urológicas Españolas*, vol. 40, no. 10, pp. 621–627, 2016.

[3] N. Alothmany, H. Mosli, M. Shokoueinejad, R. Alkashgari, M. Chiang, and J. G. Webster, “Critical review of uroflowmetry methods,” *Journal of Medical and Biological Engineering*, vol. 38, no. 5, pp. 685–696, 2018.

[4] L. Arjona, L. E. Diez, A. B. Martinez, and A. A. Echevarria, “Urosound: A smartwatch-based platform to perform non-intrusive sound-based uroflowmetry,” *IEEE Journal of Biomedical and Health Informatics*, 2022.

[5] Y. J. Lee, M. M. Kim, S. H. Song, and S. Lee, “A novel mobile acoustic uroflowmetry: comparison with contemporary uroflowmetry,” *International Neurourology Journal*, vol. 25, no. 2, p. 150, 2021.

[6] P. Hurtík, M. Burda, J. Krhut, P. Zvara, and L. Lunácek, “Automatic diagnosis of voiding dysfunction from sound signal,” in *2015 IEEE Symposium Series on Computational Intelligence*, 2015, pp. 1331–1336.

[7] E. J. Aslim, B. Balamurali, and Y. S. L. e. a. Ng, “Pilot study for the comparison of machine-learning augmented audio-uroflowmetry with standard uroflowmetry in healthy men,” *BMJ Innovations*, vol. 6, no. 4, 2020.

[8] C. V. Comiter and E. Belotserkovsky, “A novel mobile uroflowmetry application for assessling lower urinary tract symptoms,” in *NEUROUROLOGY AND URODYNAMICS*, vol. 38. WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2019, pp. S56–S57.

[9] G. Narayanswamy, L. Arjona, L. E. Díez, A. Bahillo, and S. Patel, “Automatic classification of audio uroflowmetry with a smartwatch,” in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022, pp. 4325–4329.

[10] Y. Iravantchi, K. Ahuja, M. Goel, C. Harrison, and A. P. Sample, “Privacymic: Utilizing inaudible frequencies for privacy preserving daily activity recognition,” *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021.

[11] S. Chu, S. Narayanan, and C.-C. J. Kuo, “Environmental sound recognition with time–frequency audio features,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1142–1158, 2009.

[12] D. Barchiesi, D. Giannoulis, D. Stowell, and M. D. Plumbley, “Acoustic scene classification: Classifying environments from the sounds they produce,” *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 16–34, 2015.

[13] J. T. Geiger, B. Schuller, and G. Rigoll, “Recognising acoustic scenes with large-scale audio feature extraction and svm,” *IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events*, 2013.

[14] W. Wang, *Machine Audition: Principles, Algorithms and Systems: Principles, Algorithms and Systems*. IGI Global, 2010.

[15] J. Ye, T. Kobayashi, N. Toyama, H. Tsuda, and M. Murakawa, “Acoustic scene classification using efficient summary statistics and multiple spectro-temporal descriptor fusion,” *Applied Sciences*, vol. 8, no. 8, p. 1363, 2018.

[16] “Dodotronic.” [Online]. Available: https://www.dodotronic.com/

A.2 Conference Certificate (IEEE EMBC 2023)

The following page shows the official certificate of participation in the 45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE EMBC 2023), held in Sydney, Australia.

IEEE EMBC 2023

24-27 July 2023 • ICC Sydney

Mr Marcos Lazaro Alvarez Arteaga

has attended

45th Annual International Conference on the IEEE Engineering in Medicine and Biology Society

Nigel Lovell

Nigel Lovell

Sally McArthur

IEEE EMBC 2023 Conference Chairs

A.3 International Mobility Certificate

The following page shows the certificate of the research stay conducted at INESC TEC – Institute for Systems and Computer Engineering, Technology and Science, Porto, Portugal, from January 16 to April 16, 2025, as part of the doctoral training.

CERTIFICATE OF RESEARCH STAY

To whom it may concern

This is to certify that Mr. Marcos Lazaro Alvarez Arteaga, PhD student at the University of Deusto – Faculty of Engineering (Bilbao, Spain), has successfully completed a research stay at INESC TEC – Institute for Systems and Computer Engineering, Technology and Science, in Porto (Portugal), from **January 16, 2025 to April 16, 2025**.

During this period, Mr. Alvarez actively participated in the research activities of the host institution, contributing to the development of artificial intelligence models for non-invasive urinary flow estimation using sound-based analysis.
His work was carried out within the scope of the project entitled:

"Leveraging Synthetic Data to Develop a Machine Learning Model for Voiding Flow Rate Prediction from Audio Signals."

This study focuses on training and evaluating machine learning algorithms using synthetic urination sound data generated with a high-precision peristaltic pump and a conventional toilet setup. These models were designed to estimate voiding flow rate from audio features, particularly MFCCs, and were later retrained and validated using real voiding sound data.

Mr. Alvarez's activities during the research stay included:

- Generation of synthetic datasets under controlled experimental conditions.
- Signal segmentation and extraction of acoustic features.
- Training and evaluation of machine learning models.
- Validation of model performance with real voiding data.
- Preparation and submission of a scientific article derived from this work, which has been submitted to an international journal and is currently under peer review.

These contributions were essential for advancing research in sound-based uroflowmetry and reinforce the collaboration between the University of Deusto and INESC TEC. Porto, April 16, 2025

Prof. Alípio Jorge
Supervisor
Department of Computer Science
Faculdade de Ciências
Universidade do Porto
Coordinator of LIAAD / INESC TEC
Portugal
amjorge@fc.up.pt.pt

Prof. Elsa Gomes
Supervisor
Departamento de Engenharia Informática
Instituto Superior de Engenharia do Porto
Senior Researcher
LIAAD / INESC TEC
Portugal
efg@isep.ipp.pt

A.4 National Mobility Certificate

The following page shows the certificate of the research mobility carried out at the AIKE Research Group (Artificial Intelligence and Knowledge Engineering), Department of Information and Communication Engineering, University of Murcia (Spain), from July 29 to August 29, 2025, as part of the doctoral training.

UNIVERSIDAD DE MURCIA

Departamento
de Ingeniería de la Información
y las Comunicaciones

CERTIFICADO

D. José Tomás Palma Méndez, Profesor del Departamento de Ingeniería de la Información y las Comunicaciones de la Universidad de Murcia,

CERTIFICA:

Que D. Marcos Lazaro Alvarez Arteaga, doctorando de la Universidad de Deusto (Facultad de Ingeniería), ha realizado una estancia de investigación en este Departamento desde el 29 de julio hasta el 29 de agosto de 2025.

Durante dicho periodo, el doctorando ha trabajado en el ámbito de la Uroflujometría basada en sonido, participando en el diseño experimental, adquisición y procesamiento de señales acústicas de micciones simuladas contra superficies cerámicas de inodoros.

El trabajo realizado forma parte de un artículo científico titulado:

"Annotated dataset of simulated voiding sounds impacting ceramic surfaces for urine flow estimation", que se encuentra en preparación para su envío a la revista Scientific Data.

Y para que conste a los efectos oportunos, se expide el presente certificado en Murcia, a 29 de agosto de 2025.

Atentamente,

PALMA MENDEZ JOSE TOMAS - 31248453D

Firmado digitalmente por PALMA MENDEZ JOSE TOMAS - 31248453D
Fecha: 2025.08.29 14:13:29 +02'00'

José Tomás Palma Méndez
Investigador Principal Grupo de Investigación AIKE
Departamento de Ingeniería de la Información y las Comunicaciones
Universidad de Murcia
Email: jtpalma@um.es
Tel: +34 868 884631
Fax: +34 868 884151

Departamento de Ingeniería de la Información y las Comunicaciones
UGAD de Informática
Facultad de Informática
Campus de Espinardo
30100 Murcia
ESPAÑA

Tlf.: 868 88 1775
http://www.um.es/web/diic
https://www.um.es/web/ugad/unidades/informatica

Bibliography

[1] W Scott McDougal, Alan J Wein, Louis R Kavoussi, Alan W Partin, and Craig A Peters. *Campbell-Walsh urology 11th edition review E-book*. Elsevier Health Sciences, 2015.

[2] Nazeeh Alothmany, Hisham Mosli, Mehdi Shokoueinejad, Rayan Alkashgari, Michelle Chiang, and John G Webster. Critical review of uroflowmetry methods. *Journal of Medical and Biological Engineering*, 38:685–696, 2018.

[3] Amy Y Zhang and Xinyi Xu. Prevalence, burden, and treatment of lower urinary tract symptoms in men aged 50 and older: a systematic review of the literature. *SAGE Open Nursing*, 4:2377960818811773, 2018.

[4] John M Hollingsworth and Timothy J Wilt. Lower urinary tract symptoms in men. *Bmj*, 349, 2014.

[5] HJ Bovenschen, MJR Janssen, MGH Van Oijen, RJF Laheij, LGM Van Rossum, and JBMJ Jansen. Evaluation of a gastrointestinal symptoms questionnaire. *Digestive diseases and sciences*, 51:1509–1515, 2006.

[6] Bilal Chughtai, James C Forde, Dominique Dana Marie Thomas, Leanna Laor, Tania Hossack, Henry H Woo, Alexis E Te, and Steven A Kaplan. Benign prostatic hyperplasia. *Nature reviews Disease primers*, 2(1):1–15, 2016.

[7] Manuel Fernández Arjona and Ignacio Pereira Sanz. Hiperplasia benigna de próstata: una afección de elevada prevalencia en el paciente de edad avanzada. *Revista Española de Geriatría y Gerontología*, 43(1):44–51, 2008.

[8] Werner Schäfer, Paul Abrams, Limin Liao, Anders Mattiasson, Francesco Pesce, Anders Spangberg, Arthur M Sterling, Norman R Zinner, and Philip van Kerrebroeck. Good urodynamic practices: uroflowmetry, filling cystometry, and pressure-flow studies. *Neurourology and Urodynamics: Official Journal of the International Continence Society*, 21(3):261–274, 2002.

[9] Sanjay Sinha. The use of uroflowmetry as a diagnostic test. *Current Urology Reports*, 25(5):99–107, 2024.

[10] Zoe S Gan and Stephen A Zderic. Current state and future considerations for home uroflowmetry. *Nature Reviews Urology*, 20(9):515–516, 2023.

[11] Laura Arjona, Sergio Hernández, Girish Narayanswamy, Alfonso Bahillo, and Shwetak Patel. Autonomous collection of voiding events for sound uroflowmetries with machine learning. *Biomedical Signal Processing and Control*, 105:107556, 2025.

[12] Laura Arjona, Luis E Díez, Alfonso Bahillo, and Antonio Arruza-Echevarría. Urosound: A smartwatch-based platform to perform non-intrusive sound-based uroflowmetry. *IEEE Journal of Biomedical and Health Informatics*, 27(5):2166–2177, 2022.

[13] Craig V Comiter and Edward Belotserkovsky. A novel mobile uroflowmetry application for assessling lower urinary tract symptoms. In *Neurourology and Urodynamics*, volume 38, pages S56–S57. WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2019.

[14] Marcel GÄrtner, Jan Krhut, Petr Hurtik, Michal Burda, Katarina Zvarova, and Peter Zvara. Evaluation of voiding parameters in healthy women using sound analysis. *LUTS: Lower Urinary Tract Symptoms*, 10(1):12–16, 2018.

[15] Young Ju Lee, Michelle M Kim, Sang Hun Song, and Sangchul Lee. A novel mobile acoustic uroflowmetry: comparison with contemporary uroflowmetry. *International Neurourology Journal*, 25(2):150, 2021.

[16] Edwin Jonathan Aslim, BT Balamurali, Yun Shu Lynn Ng, Tricia Li Chuen Kuo, Kheng Sit Lim, Jacob Shihan Chen, Jer-Ming Chen, and Lay Guat Ng. Pilot study for the comparison of machine-learning augmented audio-uroflowmetry with standard uroflowmetry in healthy men. *BMJ Innovations*, 6(4), 2020.

[17] W Schäfer, H Rübben, R Noppeney, and F-J Deutz. Obstructed and unobstructed prostatic obstruction: a plea for urodynamic objectivation of bladder outflow obstruction in benign prostatic hyperplasia. *World Journal of Urology*, 6:198–203, 1989.

[18] Homepage - Minze Health. https://minzehealth.com/, 2024. Accessed on 2024-10-28.

[19] Marcos Lazaro Alvarez, Laura Arjona, Mario Jojoa-Acosta, and Alfosno Bahillo. Flow prediction in sound-based uroflowmetry. *Scientific Reports*, 15(1):643, 2025.

[20] Marcos Lazaro Alvarez, Laura Arjona, Alfonso Bahillo, and Ganeko Bernardo-Seisdedos. Annotated dataset of simulated voiding sound for urine flow estimation. *Scientific Data*, 12(1):1–7, 2025.

[21] Marcos Lazaro Alvarez, Alfonso Bahillo, Laura Arjona, Diogo Marcelo Nogueira, Elsa Ferreira Gomes, and Alípio M. Jorge. Leveraging synthetic data to develop a machine learning model for voiding flow rate prediction from audio signals. *IEEE Access*, 13:127240–127251, 2025.

[22] Marcos Lazaro Alvarez, Laura Arjona, Miguel E Iglesias Martínez, and Alfonso Bahillo. Automatic classification of the physical surface in sound uroflowmetry using machine learning methods. *EURASIP Journal on Audio, Speech, and Music Processing*, 2024(1):12, 2024.

[23] Paul Abrams and Paul Abrams. *Urodynamic techniques*. Springer, 1997.

[24] BG Urbonavičius and P Kaškonas. Urodynamic measurement techniques: A review. *Measurement*, 90:64–73, 2016.

[25] Eric Chieh-Lung Chou, Pei-Yu Yang, Wen-Hui Hsueh, Chao-Hsiang Chang, and Nai-Hsin Meng. Urinating in the standing position: a feasible alternative for elderly women with knee osteoarthritis. *The Journal of urology*, 186(3):949–953, 2011.

[26] Darren Hitt, Katarina Zvarova, and Peter Zvara. Urinary flow measurements via acoustic signatures with application to telemedicine. In *39th AIAA Fluid Dynamics Conference*, page 4300, 2009.

[27] Katarina Zvarova, Michal Ursiny, Timothy Giebink, Kathleen Liang, Jerry G Blaivas, and Peter Zvara. Recording urinary flow and lower urinary tract symptoms using sonouroflowmetry. *Canadian Journal of Urology*, 18(3):5689, 2011.

[28] Jan Krhut, Marcel Gärtner, Radek Sỳkora, Petr Hurtík, Michal Burda, Libor Luňáček, Katarína Zvarová, and Peter Zvara. Comparison between uroflowmetry and sonouroflowmetry in recording of urinary flow in healthy men. *International Journal of Urology*, 22(8):761–765, 2015.

[29] Ju L Young, S Jeeyoung, J Jiyoung, S Min-Ho, C Hansol, and L Sangchul. A novel mobile acoustic uroflowmetry: Comparison of uroflowmetry and mobile acoustic uroflowmetry. *Neurourology and Urodynamics*, 37:S147–S148, 2018.

[30] BT Balamurali, Edwin Jonathan Aslim, Yun Shu Lynn Ng, Tricia Li Chuen Kuo, Jacob Shihang Chen, Dorien Herremans, Lay Guat Ng, and Jer-Ming Chen. Acoustic prediction of flowrate: varying liquid jet stream onto a free surface. In *2020 International Conference on Signal Processing and Communications (SPCOM)*, pages 1–5. IEEE, 2020.

[31] Elie El Helou, Joy Naba, Karim Youssef, Georges Mjaess, Ghassan Sleilaty, and Samar Helou. Mobile sonouroflowmetry using voiding sound and volume. *Scientific Reports*, 11(1):11250, 2021.

[32] Dong-Gi Lee, Jonathan Gerber, Vinaya Bhatia, Nicolette Janzen, Paul F Austin, Chester J Koh, and Sang Hoon Song. A prospective comparative study of mobile acoustic uroflowmetry and conventional uroflowmetry. *International Neurourology Journal*, 25(4):355, 2021.

[33] Mark T Dawidek, Rohit Singla, Lucie Spooner, Louisa Ho, and Christopher Nguan. Clinical validation of an audio-based uroflowmetry application in adult males. *Canadian Urological Association Journal*, 16(3):E120, 2021.

[34] Hwanik Kim, Changhee Ye, Gyoohwan Jung, Hoyoung Ryu, Jeong Woo Lee, Seong Jin Jeong, and Sangchul Lee. Validation of acoustic voided volume measure: a pilot prospective study. *World Journal of Urology*, 41(2):509–514, 2023.

[35] Han Jie Lee, Edwin Jonathan Aslim, BT Balamurali, Lynn Yun Shu Ng, Tricia Li Chuen Kuo, Cindy Ming Ying Lin, Christopher Johann Clarke, Prachee Priyadarshinee, Jer-Ming Chen, and Lay Guat Ng. Development and validation of a deep learning system for sound-based prediction of urinary flow. *European Urology Focus*, 9(1):209–215, 2023.

[36] Wei-shou Hsu. *Robust bandwidth extension of narrowband speech*. PhD thesis, McGill University, 2004.

[37] Marcos Lazaro Alvarez, Laura Arjona, Alfonso Bahillo, and Ganeko Bernardo. Annotated dataset of simulated voiding sound for urine flow estimation. figshare, June 2025. Accessed: 2025-06-17.