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].

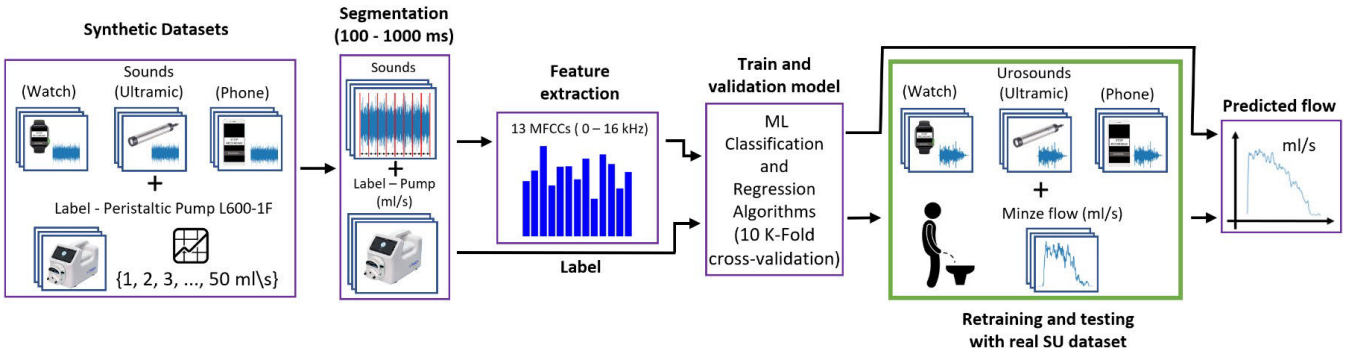


FIGURE 1. Diagram showing the pipeline of the proposed methodology, consisting of 4 main steps: data extraction, audio segmentation, feature extraction and finally model training and validation.

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.

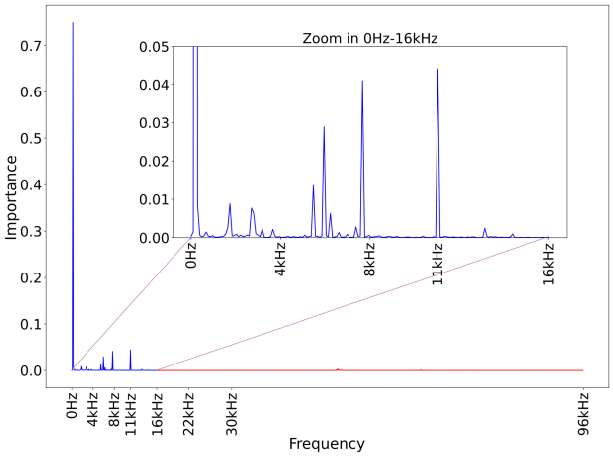
Figure 2 (b) illustrates the relative feature importance of each frequency component in the urinary flow estimation process using RFC. Similar to the regression-based approach, the results indicate that the most relevant frequencies are predominantly located in the lower spectrum, specifically below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

E. SELECTED FEATURES FOR THIS STUDY

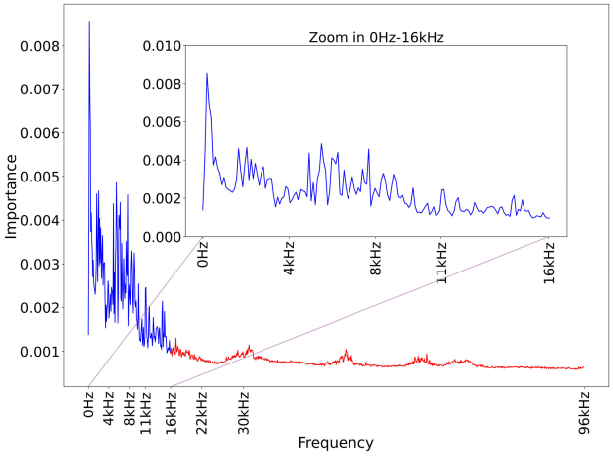
For the analysis of the audio signals, we evaluated various extraction features, including MFCC, ZCR and Chroma, each providing complementary information about the signal:

- 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].

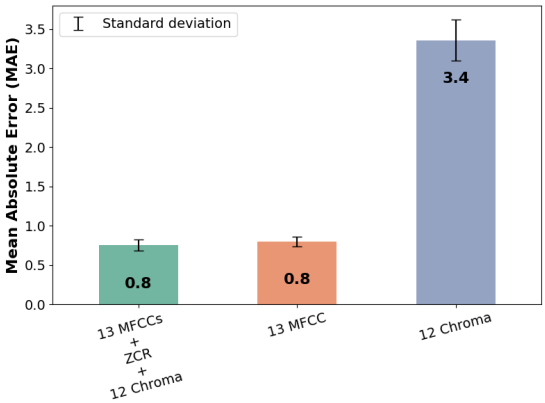


FIGURE 3. Evaluation of RFR performance with different combinations of audio feature inputs.

- 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.

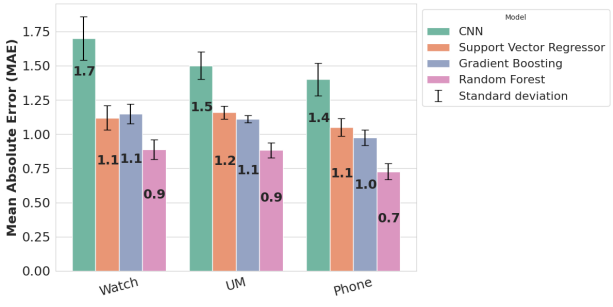


FIGURE 4. Evaluation results of the four regression models for each recording device, in terms of the mean absolute error (MAE), measured in ml/s.

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.

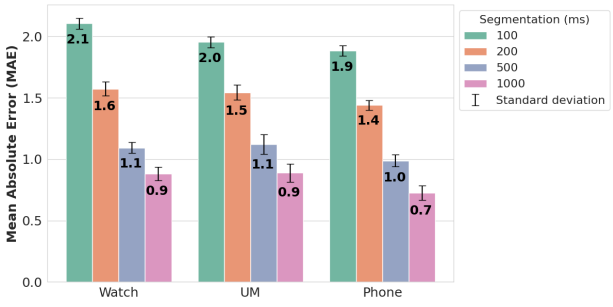


FIGURE 5. Analysis of the MAE for the RFR prediction model, comparing different audio segment sizes (ms) and the three different recording devices.

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.

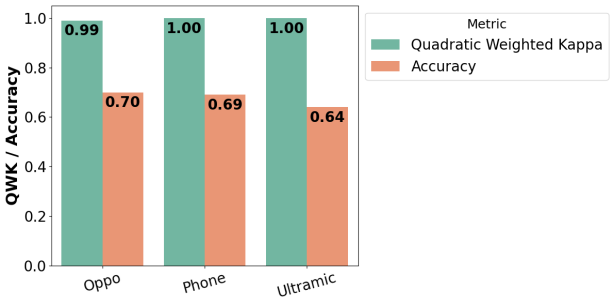


FIGURE 6. Evaluation of the RFC model for each device in terms of QWK and accuracy.

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.

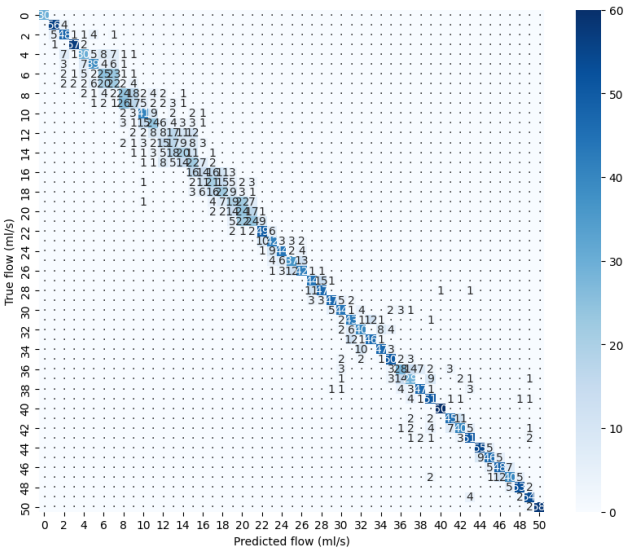


FIGURE 7. Confusion matrix for the RFC model with UM audio recordings. The model was trained and evaluated using the synthetic dataset, considering a frequency range of 0–16 kHz and a segment size of 1000 ms.

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 $\pm$ 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 $\pm$ 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 $\pm$ SD	MSE $\pm$ 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.