Computers in Biology and Medicine 193 (2025) 110337

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Comparison of three classifiers in detection of obstruction of the lower urinary tract using recorded sounds of voiding

Mario Jojoa-Acosta [a,*], Alfonso Bahillo [a], Laura Arjona [b], Rubén M. Lorenzo [a], Elba Canelón [c]

[a] Department of Signal Theory and Communications, Universidad de Valladolid, Valladolid, 47011, Spain
[b] Faculty of Engineering, University of Deusto, Av. Universidades, 24, 48007, Bilbao, Spain
[c] Hospital Universitario de Puerto Real, Puerto Real, 11510, Spain

ARTICLE INFO

Keywords:
Computer vision
Deep learning
Inception v3
Convolutional neural network
Scalogram
Wavelet
Low urinary tract symptoms

ABSTRACT

The aim of this research is to help health care professionals to automatically detect lower urinary tract disorders using sounds of voiding recorded at home. In total 93 patients were diagnosed as obstructed or non-obstructed in a hospital using traditional flow-metering technique. After they went to their houses to collect several micturition recordings (5–13 records per patient) by themselves using their Oppo smart watch. Our proposed method is based on the use of the wavelet scalogram to represent the collected sounds as images, which contains both time and frequency information. A deep learning model, the inception v3 convolutional neural network, is used to classify these recordings of the voiding into the categories of obstructed and non-obstructed. We compared the performance of our approach with classical techniques such as Support Vector Machine (SVM) and Multilayer Perceptron (MLP) using the envelope of the superposed sounds per patient as inputs. These recordings were obtained in home environments. The ground truth was built by physicians' labeling these sound recording. They used the gold standard uroflowmetry test, which gave them all the information to classify the patients as either obstructed or non-obstructed. The performance of the model in terms of the $F1$ score, accuracy, and area under the curve were 0.897, 0.891 and 0.901, respectively.

1. Introduction

Urethral obstructions are blockages in the urinary tract, which includes the kidneys, the bladder, the ureters, which carry urine from the kidneys to the bladder, and the urethra, which connects the bladder to the outside of the body [1]. Blockages may occur for many reasons, including gastrointestinal problems [2]. Such obstructions are more common in men, especially as they age and the prostate increases in size. Uroflowmetry is the most frequent and noninvasive physiological method to evaluate the obstruction of the lower urinary tract. Currently, this test is carried out at health care centers and involves the person urinating into a urinary flow meter. This process is unnatural and requires on-demand voiding, leading to significant test-to-test variability, as the situational stress of the patient affects the flowrate, giving unrepresentative results. Therefore, it is recommended to perform the uroflowmetry test more than once, which requires repeated, time-consuming, and expensive visits to the clinic. For this reason, we undertook the challenge to build an intelligent system able to automatically detect, at home, lower urinary tract obstructions, without the need to assist patients at a health center [3]. Therefore, the main objective of our proposed method is to classify patients with obstructed and non-obstructed lower urinary tracts [4]. To do this we decided to build an intelligent system based on computer vision [5]. The idea behind the use of computer vision is the trend to obtain more powerful models based on pretrained models published for fine tuning [6], trying to imitate human vision [7]. We concentrated our efforts on comparing our computer vision proposal with techniques that extract the information from the shape of the envelope of the recorded audios. This idea is based on the results of other authors, where they comment that the energy of the information is concentrated in the area under the curve of the amplitude of the recordings [8]. To perform a fair comparison, we selected the most appropriate models based on the literature. We will compare three different machine learning models performing the same task: Multilayer Perceptron [9], Support Vector Machine [10] and Inception v3 [11]. In addition, we used different techniques for the feature extraction to compare the performance of the models. We propose the use of decimated Hilbert Envelope [12], with different sizes of the feature space, and Scalograms [13], depending on

* Corresponding author.
E-mail address: mariofernando.jojoa@uva.es (M. Jojoa-Acosta).

https://doi.org/10.1016/j.compbiomed.2025.110337
Received 21 November 2024; Received in revised form 18 February 2025; Accepted 3 May 2025
Available online 23 May 2025

the specific algorithm to be used. Finally, it is important to observe the behavior of the models in unbalance data situations [14]. So, we wanted to observe the performance of the models using different coefficients of balancing data sets. US-49 is an unbalanced dataset with prevalent number of non-healthy sounds. At the opposite, US-93 has a prevalent number of healthy sounds. This idea helped us to the interpretation of the obtained metrics regarding the gap between amount of obstructed and non-obstructed collected sounds.

We use two datasets: the first is an unbalanced dataset called US-49 and the second is a balanced dataset called US-93. Both were collected for this research and are available on request.

The main opportunity for our contribution is the use of different overlapped audio recordings to obtain information about the phenomenon at different moments of the day, with the aim of improving the performance of the system and reducing the bias associated to the test-to-test variability, that could hide the real state of the urinary tract of the patient.

2. Related work

As the main objective of the present work is to classify the audio recordings into two categories, obstructed and non-obstructed patient, we present the most representative related work in this area.

In [15] the authors describe a comprehensive study of the sounds obtained from micturition to reproduce the traditional flow metric chart. They found that it is possible to use machine learning to predict voiding bladder volume values using acoustic signals. In addition, they demonstrate the use of smartwatches to perform diagnostics support in home environments. In our proposal, we obtained a model to directly classify the sounds in obstructed or non-obstructed depending on the phenomenon which produce them.

In [16] the authors evaluated the accuracy of an AI-based system in distinguishing between normal and abnormal urinary flows using voiding sound recordings. The study involved 233 male participants, including healthy individuals and those with lower urinary tract symptoms (LUTS). Machine learning algorithms, specifically a Gaussian Mixture Model – Universal Background (GMM-UBM) and a Long Short-Term Memory (LSTM) model, were trained on paired uroflowmetry and audio data. The results showed high classification accuracy, with the GMM-UBM model achieving 89.2% and the LSTM model achieving 91.1%. These findings suggest that AI-assisted audio-uroflowmetry could serve as a non-invasive screening tool for urinary flow abnormalities, particularly in settings lacking access to specialized urology services. The main difference with our work is that we used 5-folds cross-validation to increase the generalization capacity of the obtained model.

In [17], the authors trained a machine learning system to perform regression analysis of the uroflow chart, utilizing a multilayer neural network model with Mel Frequency Cepstral Coefficients (MFCC) for feature extraction. While we applied similar feature extraction techniques, our objective was distinct: we aimed to classify patients rather than replicate flow rates. In [8], researchers proposed using clinical urination sounds to reconstruct the uroflowmetric chart, leveraging time-frame calculations and energy concentration within those frames. This approach accommodates variable recording durations without losing critical information. However, their method's dependence on gold standard volume measurements for accurate reconstruction limits its diagnostic applicability.

In contrast, [18] employed microscopy images of bladder cells for deep learning-based urinary tract infection detection, underlining the importance of remote sensing for early diagnosis. Similarly, [19] combined patient demographic data with six machine learning algorithms to predict lower urinary tract infections, demonstrating the potential of these methods in assisting early and precise diagnoses. A broader perspective was provided in [20], where the authors reviewed machine learning applications in healthcare, emphasizing that blending diverse data sources can significantly enhance system performance in analyzing urodynamics and detecting urinary symptoms.

The work of [21] used acoustic signals to classify voiding patterns through a long short-term memory (LSTM) model, achieving high accuracy in categorizing patients using a one-versus-all strategy. This approach aligns with our goal of classifying patients into distinct groups, such as normal, lower urinary tract symptoms (LUTS), or obstruction. However, their study lacked rigorous validation methods, such as cross-validation, to assess the model's generalizability. Similarly, [22] used deep learning algorithms to classify urination flow as normal or abnormal, achieving promising results. Our proposed model demonstrates superior classification performance, as evidenced by a higher area under the curve (AUC) in distinguishing obstructed vs. non-obstructed patients. However, direct comparison was hindered by the unavailability of their dataset.

While previous studies have demonstrated the potential of machine learning and deep learning techniques in healthcare, particularly for analyzing uroflowmetry charts, acoustic signals, and demographic data, they present several limitations. Many approaches, such as [21,22], focus on patient classification but often lack robust validation strategies to ensure generalizability. Others, like [8], rely on external gold standards, limiting their practical applicability in real-world scenarios. Furthermore, approaches such as [17,19] leverage alternative data types (e.g., microscopy images and demographic data) but do not utilize sound recordings as a primary diagnostic tool, leaving a gap in leveraging this non-invasive, cost-effective modality.

Our work addresses these gaps by introducing a novel framework for classifying lower urinary tract obstruction using recorded voiding sounds. Unlike prior studies, our approach emphasizes the use of rigorous validation techniques, such as cross-validation, to ensure the model's reliability and generalization across diverse datasets. Additionally, by comparing the performance of three classifiers, we offer a comprehensive analysis of their effectiveness for this task, setting a new benchmark for classification accuracy. Our focus on classifying patients as obstructed versus non-obstructed, position our work to contribute not only as a practical diagnostic tool but also as a validated step forward in the integration of acoustic signal processing and computer vision for medical applications.

3. Materials and method

We recruited 106 male volunteers between the ages of 18 and 85 (44 without urological comorbidities aged around 20, and 62 with urological comorbidities aged around 70), who agreed to participate in the study. All the sound recording both obstructed and non-obstructed were collected at home. The diagnosis (obstructed or non-obstructed) of all the volunteers was done at the hospital based on the result of the standard uroflowmetry collected there. All the data were collected following the same procedure approved by the Valladolid East Health Area Medicine Research Ethics Committee on 27 July 2023 with reference PI-GR-23-3275. The dataset US-49 was collected first and involved most of the volunteers with urological comorbidities. The data added was collected in a second round and involved most of the volunteers without urological comorbidities. Normally the audio data collection is considered as a stochastic process. However, the duration of the micturition is distributed following these parameters: 87 s in average and 37 s of standard deviation. The proposed solution has two stages: the first is to record and store the sounds of the urination at home of adult men using an Oppo smartwatch. With those sounds coming from patients previously classified as obstructed or non-obstructed, we trained an artificial intelligence model to detect abnormalities in the visual representation of these recordings. The second is to perform an automatic classification of new sounds using the trained model. Our approach is based on the idea that the obstruction phenomena is not completely deterministic, and it is affected by external sources. One patient could have more than one sound; sometimes with normal

Table 1
Characteristics of the dataset of audio recordings.

	US-49	Data added	US-93
Patients	62	44	106
Labeled patients	49	44	93
Obstructed patients	37	0	37
Non-obstructed patients	12	44	56
Undetermined Patients	13	0	13
Total audio files	869	227	1096
Sample frequency	16 000	16 000	16 000

characteristics and other ones with urinary track obstruction features. This situation is common in this kind of disease, where one obstructed sound is needed to conduce to an obstructed diagnostic. In addition, paruresis disorder could generate hospital examination procedures to obtain wrong diagnosis, since the patient cannot miction normally, simulating an obstruction. To mitigate this situation, we decided to use sounds obtained in comfortable and known environments for the patient, who collects between five to thirteen sounds in an observation window of three days. To preserve a holistic approach of the phenomena, we need to superpose all the sounds to conserve all information as possible before to be analyzed by the machine learning model.

An Oppo smartwatch was used to collect voiding sounds at home. It was configured to capture audio data at a sample rate of 16,000 samples per second. A smartwatch, worn by each participant on his wrist, provided a hands-free and minimally intrusive way to gather audio samples in real time. The smartwatch was programmed to automatically record the sounds of voiding, and the data were stored securely for later analysis. This approach allowed for consistent capture of high-quality sound with minimal interference, aiding in the accurate evaluation of the acoustic patterns associated with urination.

3.1. Data collection

The collection of the dataset was led and performed by the Hospital Puerto Real (Spain). All the volunteers signed an informed consent describing the procedure, which is not invasive, and in compliance with the European Fundamental Rights of citizens, and fully respecting ethical aspects. The data were processed fairly for the specified purposes and based on the consent of the person concerned. The volunteers include patients between 18–85 years of age with lower urinary tract symptoms who had been prescribed flowmetry from health care professionals and diagnosed as obstructed. Afterwards, we balanced the dataset with volunteers between 18–85 years of age without urological comorbidities (non-obstructed), who volunteered to have a flowmetry performed on them at the hospital. The dataset consists of a set of sound-based voidings per volunteer taken at home. All volunteers were asked to urinate against the water of the toilet bowl for three consecutive days after a uroflowmetry performed at the hospital. On average, 10 recordings per volunteer were collected. The final number depends on each case, as some voids were removed for different reasons such as high background noise, or voids performed out of the home in wall mounted toilets. The voiding events were recorded using an Oppo smartwatch and the UroSound platform as motivated and described in [23]. Table 1 shows the main characteristics of the datasets used for our research work. After the whole of this process, we obtained two datasets: US-49, with unbalanced data, which was collected at the beginning of our research, US-93, which was balanced by adding more audio recordings from volunteers without urological comorbidities. Their main characteristics are shown in Table 1.

3.2. Methods

This section describes the method employed to obtain the results and conclusions of the study. The block diagram in Fig. 1 shows the sequence of steps followed.

3.2.1. Pre-processing

This stage is necessary to eliminate possible biases associated with the acquisition of the signals under study. We normalized the amplitude and performed a zero padding. In addition, the audios were filtered using a low pass 30 order FIR filter with a rejection band in 8 kHz. The use of this filter is to prevent aliasing. The designed filter uses a Hamming windowing method to guarantee a linear phase for our application. In order to eliminate possible bias associated with the amplitude, we normalized the signal. Eq. (1) shows the normalization of the amplitude of the audio signal.

$$x(t) = \frac{x(t) - X_{min}}{X_{max} - X_{min}} \tag{1}$$

3.2.2. Curation of the sounds

As the volunteers in this study have more than one audio to characterize their potential voiding dysfunction, we organized the audios with the aim of obtaining one audio array per volunteer. To do this, we organized all the audios one by one in an $m \times n$ matrix, where m is the number of audios associated to each volunteer and n is the maximum length of the associated audios. Since audios have different lengths, we used zero padding to get an equal number of samples per row.

3.3. Overlapping sounds

The matrix described in Eq. (2) contains for each register (row vector) the normalized audio record obtained in a different urination process.

$$X(t) \rightarrow \begin{bmatrix} x_{11}(t) & \cdots & \cdots & x_{1n}(t) \\ \vdots & \ddots & \ddots & \vdots \\ x_{m1}(t) & \cdots & \cdots & x_{mn}(t) \end{bmatrix} \tag{2}$$

Next, as we want to conserve all the possible information contained implicitly in all voiding events, our perspective is based on the idea that urination is a stochastic process, and each voiding is a realization of that stochastic process. Based on this principle, we can evaluate the stationary properties of the sound process, finding that the process is highly non-stationary. Eq. (3) shows the condition used to affirm this premise.

$$R_{XX}(t_1, t_2) \neq R_{XX}(t_1 - t_2, 0) \tag{3}$$

To conserve all the information associated to the studied phenomena, we decided to overlap all the rows of the sound matrix to obtain a register that better represents the characteristics of the patient's urination with the arithmetic sum of each other one. This overlapping is conducive to conserving all the information related to the frequency and shape components of each sound. Eq. (4) shows the method of overlapping the sounds.

$$\begin{bmatrix} x_{11}(t) & \cdots & x_{1n}(t) \\ \vdots & \ddots & \vdots \\ x_{m1}(t) & \cdots & x_{mn}(t) \end{bmatrix} \rightarrow X_s(t) = \sum_{i=1}^{m} \vec{x}_i \tag{4}$$

It is important to highlight that this overlapped final signal was normalized again to mitigate the effects related to the amplitude of the signals and the number of signals available per volunteer.

3.4. Feature extraction

In the design of a classification system, different algorithms can be chosen. Nowadays, the use of deep learning has spread significantly, because there is no need to explicitly build a feature extraction stage of the studied signals. These features are the key values or elements to be considered when determining whether a sample belongs to one of the given classes. However, in this work, we present explicit feature extraction techniques that were combined with automatic classification and deep learning algorithms [24]. We applied three different techniques to extract the features from the audio signals: time-domain envelope, and scalogram.

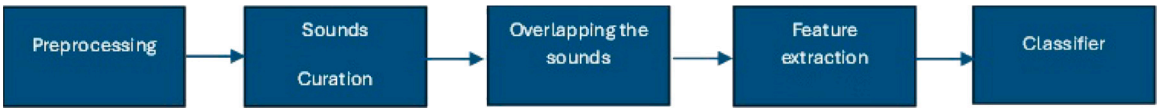

Fig. 1. Block diagram of the method.

3.4.1. Features extraction based on envelope

Intuitively, we can see that the shape of the amplitude of the audio signal contains the information with which to determine whether or not there is an obstruction in the lower urinary tract. When we say shape, we are talking about the shape of the envelope of the overlapped register. It is important to remember that $X_s(t)$ is the arithmetic sum of all sounds recorded by one volunteer. Consequently, this signal has information of the whole urination process of a single volunteer over three consecutive days. To calculate the envelope of the overlapped audios, we used the Hilbert transform [25]. This mathematical tool uses the analytical signal in the domain of the complex numbers as the basis for the calculation of the envelope, since its norm corresponds to each of the points of the envelope of the input signal. Eq. (5) shows the analytical signal formed by the Hilbert transform.

$$\begin{aligned} H_s(t) &= Ht\{X_s(t)\} \\ S(t) &= X_s(t) + H_s(t) \\ envelope\{X_s(t)\} &= |S(t)| \end{aligned} \tag{5}$$

where $S(t)$ is the analytic signal and Ht is the Hilbert transformation.

This envelope should have the same length, in terms of the number of samples, for all input signals (audio recordings). The requirement to use the same number of inputs for the proposed classification algorithms is mandatory. To achieve this, we proposed to decimate dynamically the input signal. In this way, we achieved an equal number of samples in the feature space. Fig. 2 shows an example of the envelopes calculated for an obstructed volunteer and a non-obstructed volunteer. For our research work, we selected two values to fix the dimensionality of the space. We used the values 100 and 1000 as the two possible dimensions of the feature spaces. With this constraint, we executed experiments to search for differences using these different numbers of features. The values selected correspond to the second and third powers of ten.

3.4.2. Feature extraction based on the scalogram with the continuous wavelet transform

The inception v3 model was selected because it is a well-known model. In addition, this model is widely used in industry environments, since it is available for free. This condition will lead us to perform in the future, one explainability study, needed in all areas, but mainly in healthcare applications. In addition, this algorithm blended with scalograms representing the sounds as inputs, has a high performance in sound classification task. We built graphical representations of the configured overlapped signal to feed a fine-tuned computer vision algorithm, the inception v3. We used scalograms based on the wavelet transform to get a 2D scale – time representation of the audio signal [26]. This graph contains the behavior of the spectral components of all the superposed audios versus time. The representation is the feature map extracted from each patient with the whole of the information of the phenomena of urination over three consecutive days. This time-scale representation offers some advantages, since it allows us to see the changes in the frequency over time. To carry out this process, we calculated the continuous wavelet transform with the Morse standard wavelet [27]. Fig. 3 shows an example of one scalogram obtained for an obstructed volunteer and another for a non-obstructed volunteer. As we can observe, the energy is distributed in different ways depending on the category of the sound. In an obstructed volunteer, the image has more lobes than for the non-obstructed one. This led us to infer that the information is contained in the distribution of the energy in its frequency-time dependence.

3.4.3. Class separability analysis

In classification or class detection applications, it is necessary to evaluate how separable the datasets are with respect to their labels. This is very common in medical applications, where the objective is to determine if a sample input to the system belongs to an obstructed or non-obstructed volunteer. For the present case, two labels were used: obstructed and non-obstructed. For this research work, the Calinski–Harabaz index (CHI) [28], a clustering index, is used. It returns a measure of the distances between the possible clusters existing within the dataset. A greater distance between centroids indicates a greater ability of the datasets to be separated with a supervised or unsupervised algorithm. Eq. (6) shows the mathematic formula for the CHI.

$$\begin{aligned} B &= \sum_{q \in k} n_q (c_q - c_E)(c_q - c_E)^T \\ W &= \sum_{q \in k} \sum_{x \in q group} (x - c_q)(x - c_q)^T \\ CHI &= \frac{B}{W} \frac{n_E - k}{k - 1} \end{aligned} \tag{6}$$

where k is the number of groups, n_q is the number of points in the group k, c_q is the center of group q, n_E is the number of data points, and c_E is the center of all the points. We calculated this score only for time envelope features. Based on our observations, we concluded that the CHI does not directly show the separability property in a high dimensional feature space, since it is highly dependent on the number of proposed clusters.

3.5. Classifier

Different metrics, such as the $F1$ score, accuracy, and the AUC, were used to measure the performance of the classifiers, blended with 5-fold cross-validation. We split the total dataset in k (5) sets randomly conformed, and in the followings steps we used one for validation and the rest to train the model. Only in the first step, the dataset was randomized. Finally, we calculated the mean and the standard deviation of the calculated metric. Multilayer Perceptron (MLP), Support Vector Machine (SVM), and Inception V3 are widely used machine learning models with distinct architectures and applications. Multilayer Perceptron (MLP) is a feedforward neural network that learns complex patterns using multiple layers and backpropagation. Support Vector Machine (SVM) is a powerful algorithm that finds an optimal hyperplane for classification and regression, especially effective for high-dimensional data. Inception V3 is a deep convolutional neural network (CNN) designed for image recognition, utilizing inception modules to capture multi-scale features efficiently

3.5.1. Multilayer Perceptron (MLP)

One of the key strengths of MLPs is their ability to learn and represent complex non-linear relationships in the data through activation functions applied at each node. Common activation functions include rectified linear unit (ReLU), introducing non-linearity to the model, enabling it to capture intricate patterns. Despite their powerful abilities, MLPs require careful tuning of hyperparameters, such as the number of hidden layers, number of neurons per layer, and the learning rate, as well as regularization techniques, to prevent overfitting and ensure good generalization to new data.

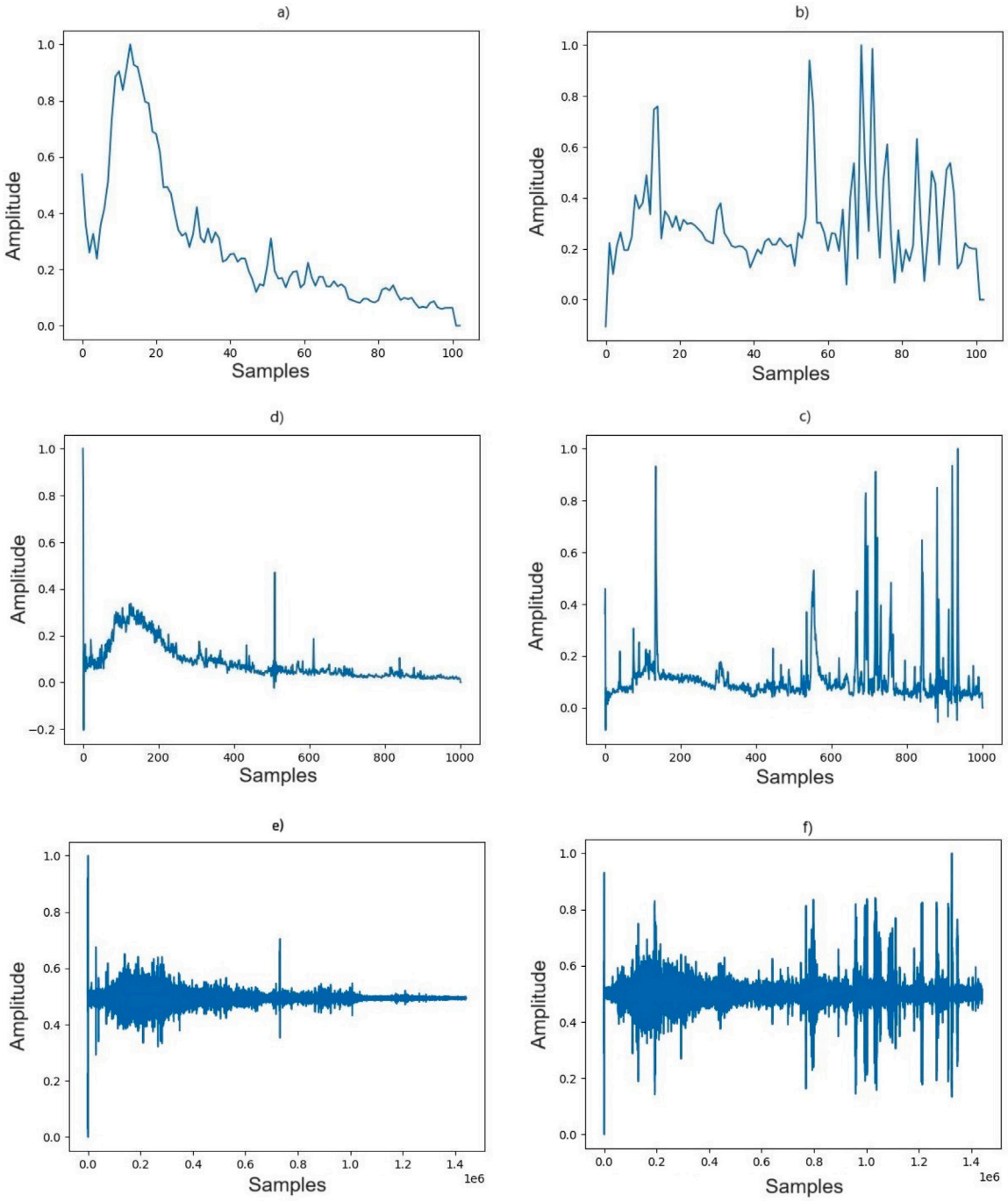

Fig. 2. (a) Envelope of a non-obstructed volunteer decimated and truncated to 100 samples. (b) Envelope of an obstructed volunteer decimated and truncated to 100 samples. (c) Envelope of an obstructed volunteer decimated and truncated to 1000 samples. (d) Envelope of a non-obstructed volunteer decimated and truncated to 1000 samples. (e) Original signal of a non-obstructed volunteer more than 1.4 million samples. (f) Original signal of a obstructed volunteer more than 1.4 million samples. For more information, see https://github.com/mario42004/urosound.

3.5.2. Support Vector Machine (SVM)

The versatility of SVM lies in its ability to handle both linear and non-linear classification problems using kernel tricks. The choice of the kernel function significantly impacts the classifier's performance and adaptability to different types of data distributions. Additionally, SVM includes regularization parameters (C) to balance the trade-off between achieving a low training error and maintaining a large margin. This helps in controlling overfitting and ensuring better generalization on unseen data. SVMs are particularly effective in scenarios with a clear margin of separation.

3.5.3. Inception v3 convolution network

Inception v3, a deep convolutional neural network, is an enhancement of the original inception architecture (GoogLeNet) designed to improve efficiency and accuracy in image classification tasks [11]. The model incorporates several innovative techniques, including factorized convolutions, which decompose larger convolutions into smaller, more manageable pieces, and asymmetric convolutions, which split a single convolution into two operations. These strategies significantly reduce the computational complexity and the number of parameters, leading to faster training times and less overfitting. Additionally, inception v3 introduces the use of "label smoothing", which prevents the network

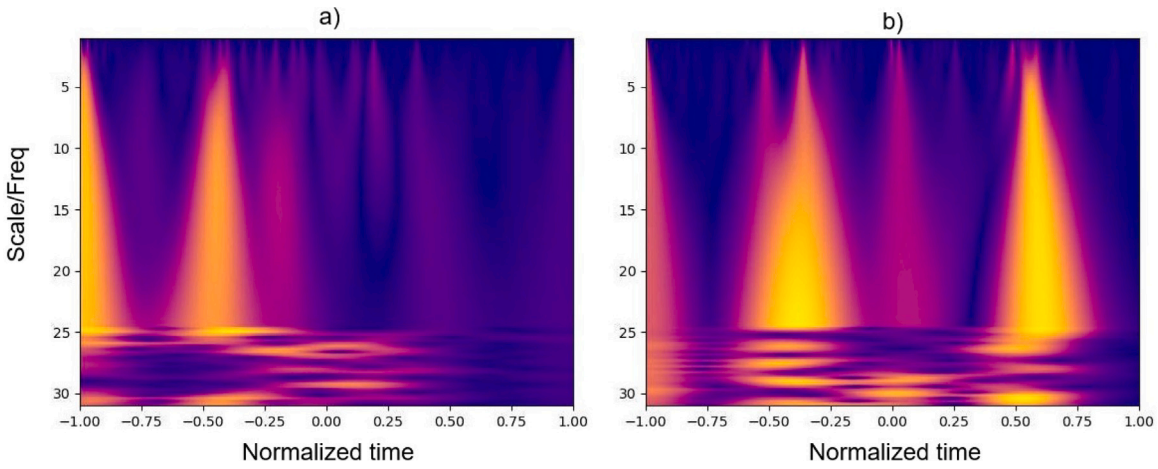

Fig. 3. (a) Wavelet scalogram for non-obstructed overlapped signal. (b) Wavelet scalogram for obstructed overlapped signal.

from becoming overly confident in its predictions, thereby improving its ability to generalize. The architecture of inception v3 comprises multiple inception modules, each containing parallel convolutional layers with filters of varying sizes. These modules allow the network to capture multi-scale features and learn spatial hierarchies effectively. The network also integrates auxiliary classifiers at intermediate layers to promote gradient flow and prevent vanishing gradients during backpropagation. Another key component is the use of batch normalization, which standardizes the inputs to each layer, accelerating the convergence and improving the performance. Overall, inception v3's carefully designed architecture balances depth and computational efficiency, making it a powerful model for image recognition tasks. Transfer learning is a widely-used technique to save computational cost in terms of training models with a huge number of parameters. It is important to highlight that we fine-tuned the inception v3 ImageNet pretrained model for our specific application. To do this, we adjusted the combination of hyperparameters to find which one led to the best performance on classification tasks.

3.5.4. Adjusting the hyperparameters

In all cases, we applied the grid search method to find the best combination of hyperparameters for each model. Table 2 shows the optimal configuration for each one of them. Hyperparameters are crucial in defining the behavior of a classifier, as they control aspects of the learning process and the model's ability. They have a significant impact on the classifier's performance and generalization ability. Tuning these hyperparameters can improve the model's accuracy, robustness, and efficiency, making hyperparameter optimization a critical step in developing an effective classifier. Through grid search, we predefined a set of values of the hyperparameter and they were systematically explored to determine the optimal settings for the model. This approach ensured that we considered a wide range of possible combinations, providing a comprehensive view of the hyperparameter landscape. By fine-tuning these values, we aimed to balance the model's complexity with its ability to generalize to new data, avoiding problems like overfitting or underfitting. Additionally, the optimization process was guided by cross-validation, which further enhanced the model's reliability by evaluating its performance across multiple data splits.

In neural networks, particularly fully connected architectures such as MLP, the number of hidden layers and neurons in each layer are critical hyperparameters. These parameters define the model's ability to learn complex patterns. However, increasing these numbers can also lead to overfitting if not carefully controlled. Other important hyperparameters include the activation function, which determines the output per each layer to back-propagate the error in each iteration.

For the SVM, key hyperparameters included the penalty parameter C and the type of kernel. The parameter C controls the trade-off between achieving a low error on the training data and a low margin, impacting the model's tolerance to misclassified samples. The choice of kernel (e.g. linear, polynomial, RBF) is also essential, as it determines how the data is transformed in higher dimensions, which can significantly impact the accuracy of the classification.

For the inception v3 model, a deep convolutional neural network architecture, fine-tuning often involves adjusting layer-specific learning rates and batch sizes to optimize the efficiency and accuracy of the training. Additionally, modifying the number of epochs is crucial, as it controls the number of times the entire dataset is passed through the network. Because of the complexity of inception v3, parameters such as weight decay and batch normalization momentum can also play a role in stabilizing the training and achieving faster convergence.

4. Results

The following tables present our results. Table 3 presents the separability index, for the classes of obstructed and non-obstructed, of the feature space conformed by the envelope.

As we can observe, the CHI improves if we reduce the number of samples of the envelope from 1000 to 100. This is particularly true in our study. The calculation of the index is based on the selection of the ideal number of clusters in a group of data. So, the number of clusters is dependent on the properties of the feature space: in our case, the size of this space. We also observe an improvement of the performance of the classical machine learning classification algorithms if the CHI is greater.

Secondly, we formed another feature space using the scalogram with the wavelet transform. Since the amount of data was limited, a 5-fold cross-validation was carried out. Table 4 shows the values of the metrics obtained in this way for each model for the two datasets used.

In bold, we show the best results obtained for the proposed task. Inception v3 model with the scalogram feature extraction technique performed better than the other classical machine learning models with their own feature extraction methods. This led us to conclude that it is possible to use computer vision algorithms to classify patients using audio recordings converted to images, achieving high performance. This could open the door to future work, where the challenge is the construction of a general-purpose artificial intelligence model to be applied to health care tasks, simplifying the architecture of the system with multiple classes and nature of inputs. In addition, we contribute to the state of the art with a model which has reached an $F1$ score, accuracy, and area under the curve with values 0.897, 0.891 and 0.901,

Table 2
Hyperparameters used for each model. Values in [] are the grid search ranges.

Multilayer Perceptron					
Dataset features	Number of layers	Activation function	Neurons per layer	Output Neurons	Optimizer
US-49 – Decimated to 100	3 [1–10]	ReLU - ReLU - ReLU	100,50,20	1 tansig	SGD
US-49 – Decimated to 1000	3 [1–10]	ReLU - ReLU - ReLU	1000,500,200	1 tansig	SGD
US-93 – Decimated to 100	3 [1–10]	ReLU - ReLU - ReLU	100,40,20	1 tansig	SGD
US-93 – Decimated to 1000	3 [1–10]	ReLU - ReLU - ReLU	1000,300,200	1 tansig	SGD

Support Vector Machine			
Dataset features	Normalized Gamma	Kernel	C
US-49 – Decimated to 100	0.8 [0.01–1]	RBF	1 [1-10]
US-49 – Decimated to 1000	0.8 [0.01–1]	RBF	1 [1-10]
US-93 – Decimated to 100	0.9 [0.01–1]	RBF	1.2 [1-10]
US-93 – Decimated to 1000	0.9 [0.01–1]	RBF	1.2 [1-10]

Inception v3				
Dataset features	Batch Size	Epochs	Learning rate step decay	Optimizer
US-49 – Scalogram	4 [2–16]	Early detection - 8	0.1 [0.1,0.01,0.001]	AdaGrad
US-93 – Scalogram	8 [2–16]	Early detection - 5	0.1 [0.1,0.01,0.001]	SGD

Table 3
Calinski–Harabaz Index (CHI) for 100 and 1000 samples decimated envelopes.

Number of samples	US-49 CHI	US-93 CHI
100	0.437	**0.769**
1000	0.414	0.678

respectively, for the obstructed vs non-obstructed classification task, using only a reduced amount of audio recordings of voiding from 93 male participants.

5. Discussion

The use of voiding sound recordings, analyzed using advanced signal processing techniques and deep learning, has proven to be an effective and non-invasive tool for the early detection of lower urinary tract obstructions [29]. This can reduce the need for invasive tests and frequent clinical visits, improving the patient experience and optimizing the use of healthcare resources. The ability of the scalogram-based model to accurately classify voiding patterns as obstructed or non-obstructed suggests that this technique could be integrated into clinical practice to personalize patient treatment. This allows continuous home-based monitoring, enabling adjustments of treatment based on the patient's real-time progress without the need for frequent in-person consultations.

The traditional technology of urinary flow meters offers a highly reliable method of detecting obstruction of the lower urinary tract [30]. However, this method has some disadvantages since it is necessary to employ it in a hospital environment and with the assistance of a medical doctor. Our proposed method mitigates this disadvantage, since we process audio recordings obtained from patients in home environments without any professional supervision.

The frequency resolution of the recording device could be a key issue, since in our method all the information related to the disease phenomenon needs to be captured. In [15] the authors mentioned that the spectrum of recording audio is lower than 20,000 Hertz, so achieving this technical feature is mandatory to apply our approach.

The use of images that represent the data helped us extract features in domains that have not yet been explored [31]. For instance, the spectrum of the recordings of the voiding contains information about the phenomena in their scale-time-dependent components. This scalogram contains all the features of audio composition that can describe all the urination process itself. The computer vision techniques offered us high accuracy in terms of performance, since the pretrained structures used were built with a high density of trainable parameters, such as the inception v3 network. All complex structures used in our work are available to be used with general transfer learning applications. This idea is conducive to generalizing computer vision structures for general tasks such as health care audio classifications.

6. Conclusion

The use of a feature extraction technique based on the spectrogram is better compared to the technique based on the decimated Hilbert envelope. These results suggest that the shape of the audio wave contains noisy information probably related to another phenomenon such as voice or external noise. Since the phenomenon is non-stationary, it is necessary to use the time information blended with the frequency information as well. The changes in frequency at the time domain gives a fingerprint of the sound, since the scalogram contains the whole of the information of the studied signals. This feature space, based on the scalogram, improved the separability just by itself.

The use of sounds to detect urinary tract disorders in home environments is a milestone because we are opening the door to future applications using the same approach as with other sound sources in human beings. This breakthrough could pave the way for non-invasive, cost-effective diagnostic tools, making it easier for individuals to monitor their health from the comfort of their own homes. By analyzing specific sounds generated during bodily functions, such technologies can potentially identify early signs of health issues that might otherwise go unnoticed until symptoms become severe. Furthermore, as the technology improves, it could be adapted to detect a wide range of conditions, offering personalized healthcare solutions and reducing the burden on traditional medical facilities.

The main contribution of our work is the use of a set of micturition sounds recordings got in home environments using a commercial smart watch to detect obstruction and non-obstruction in low urinary track system. This approach helps healthcare professionals and healthcare providers to carry out an early detection of this disorder without the need to attend to the patient directly in hospital or health centers.

The main limitation of our work is the number of sounds used in our study. It is true that we are getting a precedent in the obstruction detection applying machine learning, however, to build more reliable detection model is mandatory to use a greater number of recordings per patient, and in the same way, a higher number of participants should be involved in our study.

Finally, it could be a good idea to explore new computer algorithms such as vision transformers. It is necessary to use this technique to detect specific diseases in the patient and not only to assign the patient to the class of obstructed or non-obstructed. This objective would be reached when more labeled data becomes available for this purpose.

Table 4
5-fold mean and standard deviation performance of the algorithms with different techniques of feature extraction.

Dataset	Feature extraction technique	Classification model	$F1$-Score	Accuracy	AUC
US-49	Decimated to 100	MLP	$(\overline{0.655}, 0.092)$	$(\overline{0.655}, 0.092)$	$(\overline{0.554}, 0.086)$
US-49	Decimated to 1000	MLP	$(\overline{0.531}, 0.104)$	$(\overline{0.531}, 0.017)$	$(\overline{0.491}, 0.095)$
US-49	Decimated to 100	SVM	$(\overline{0.765}, 0.036)$	$(\overline{0.755}, 0.035)$	$(\overline{0.659}, 0.049)$
US-49	Decimated to 1000	SVM	$(\overline{0.755}, 0.029)$	$(\overline{0.756}, 0.034)$	$(0.408, 0.036)$
US-49	Scalogram	Inception v3	$(\overline{\mathbf{0.856}}, 0.106)$	$(\overline{\mathbf{0.759}}, 0.134)$	$(\overline{\mathbf{0.743}}, 0.112)$
US-93	Decimated to 100	MLP	$(\overline{0.791}, 0.045)$	$(\overline{0.801}, 0.048)$	$(\overline{0.682}, 0.052)$
US-93	Decimated to 1000	MLP	$(\overline{0.765}, 0.034)$	$(\overline{0.764}, 0.022)$	$(\overline{0.772}, 0.011)$
US-93	Decimated to 100	SVM	$(\overline{0.823}, 0.002)$	$(\overline{0.817}, 0.012)$	$(\overline{0.831}, 0.009)$
US-93	Decimated to 1000	SVM	$(\overline{0.773}, 0.008)$	$(\overline{0.778}, 0.012)$	$(\overline{0.802}, 0.016)$
US-93	Scalogram	Inception v3	$(\overline{\mathbf{0.897}}, 0.011)$	$(\overline{\mathbf{0.891}}, 0.024)$	$(\overline{\mathbf{0.901}}, 0.009)$

CRediT authorship contribution statement

Mario Jojoa-Acosta: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alfonso Bahillo:** Project administration, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Laura Arjona:** Writing – review & editing, Writing – original draft, Validation, Resources, Conceptualization. **Rubén M. Lorenzo:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis. **Elba Canelón:** Visualization, Validation, Project administration, Funding acquisition, Data curation, Conceptualization.

Ethical approval

This study was approved by the Medicine Research Ethics Committee of the Valladolid East Health Area on July 27th, 2023, under the reference PI-GR-23-3275 (minutes number 16/2023). The above-mentioned Ethics Committee complies with GCP standards (CPMP/ICH/135/95).

Funding

This research was partially supported by the Spanish Ministry of Science and Innovation under the Aginplace (ref. PID2023-146254OB-C41) and Swalu (ref. CPP2022-010045) projects.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Special thanks to the participants who were engaged in the collection of the dataset, and to Sebastián Armijos, CEO of Process Med company (https://www.e-processmed.com), who provided invaluable support through resources, digital health expertise, and active participation throughout the work.

Data availability

The data that support the findings of this study are available on reasonable previous request to the corresponding author, Mario Jojoa-Acosta at mariofernando.jojoa@uva.es.

References

[1] J.M. Hollingsworth, T.J. Wilt, Lower urinary tract symptoms in men, BMJ 349 (2014) g4474, http://dx.doi.org/10.1136/bmj.g4474, URL https://www.bmj.com/content/349/bmj.g4474.

[2] Evaluation of a Gastrointestinal Symptoms Questionnaire | Digestive Diseases and Sciences, URL https://link.springer.com/article/10.1007/s10620-006-9120-6.

[3] E. Crigger, K. Reinbold, C. Hanson, A. Kao, K. Blake, M. Irons, Trustworthy augmented intelligence in health care, J. Med. Syst. 46 (2) (2022) 12, http://dx.doi.org/10.1007/s10916-021-01790-z.

[4] K.K.Y. Chew, M. Kas, P. Mancuso, Urinary tract obstruction secondary to fungal balls: A systematic review, Soc. Int. D' Urol. J. 5 (3) (2024) 227–236, http://dx.doi.org/10.3390/siuj5030034, Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, URL https://www.mdpi.com/2563-6499/5/3/34.

[5] B.X. Yu, J. Chang, H. Wang, L. Liu, S. Wang, Z. Wang, J. Lin, L. Xie, H. Li, Z. Lin, Q. Tian, C.W. Chen, Visual Tuning, ACM Comput. Surv. (2024) http://dx.doi.org/10.1145/3657632, Just Accepted.

[6] J.S. Bowers, G. Malhotra, M. Dujmović, M.L. Montero, C. Tsvetkov, V. Biscione, G. Puebla, F. Adolfi, J.E. Hummel, R.F. Heaton, B.D. Evans, J. Mitchell, R. Blything, Deep problems with neural network models of human vision, Behav. Brain Sci. 46 (2023) e385, http://dx.doi.org/10.1017/S0140525X22002813, URL https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/.

[7] A. Parvaiz, M.A. Khalid, R. Zafar, H. Ameer, M. Ali, M.M. Fraz, Vision Transformers in medical computer vision—A contemplative retrospection, Eng. Appl. Artif. Intell. 122 (2023) 106126, http://dx.doi.org/10.1016/j.engappai.2023.106126, URL https://www.sciencedirect.com/science/article/pii/S095219762300310X.

[8] Mobile sonouroflowmetry using voiding sound and volume | Scientific Reports, URL https://www.nature.com/articles/s41598-021-90659-9.

[9] L.B. Almeida, Multilayer perceptrons, in: Handbook of Neural Computation, CRC Press, 1996.

[10] M. Tanveer, T. Rajani, R. Rastogi, Y.H. Shao, M.A. Ganaie, Comprehensive review on twin support vector machines, Ann. Oper. Res. (2022) http://dx.doi.org/10.1007/s10479-022-04575-w.

[11] M. Mujahid, F. Rustam, R. Álvarez, J. Luis Vidal Mazón, I.d.l.T. Díez, I. Ashraf, Pneumonia classification from X-ray images with inception-V3 and convolutional neural network, Diagnostics 12 (5) (2022) 1280, http://dx.doi.org/10.3390/diagnostics12051280, Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, URL https://www.mdpi.com/2075-4418/12/5/1280.

[12] M.F. Jojoa Acosta, Comparación de tres sistemas de clasificación de señales fonocardiográficas basados en distintas técnicas de extracción de características, 2018, Accepted: 2019-11-01T13:53:14Z Publisher: Universidad del Cauca, URL http://repositorio.unicauca.edu.co:8080/xmlui/handle/123456789/1316.

[13] J. Gelpud, S. Castillo, M. Jojoa, B. Garcia-Zapirain, W. Achicanoy, D. Rodrigo, Deep learning for heart sounds classification using scalograms and automatic segmentation of PCG signals, in: I. Rojas, G. Joya, A. Català (Eds.), Advances in Computational Intelligence, Springer International Publishing, Cham, 2021, pp. 583–596, http://dx.doi.org/10.1007/978-3-030-85030-2_48.

[14] Machine Learning and Bias in Medical Imaging: Opportunities and Challenges | Circulation: Cardiovascular Imaging, URL https://www.ahajournals.org/doi/abs/10.1161/CIRCIMAGING.123.015495.

[15] M.L. Alvarez, L. Arjona, M. Jojoa-Acosta, A. Bahillo, Flow prediction in sound-based uroflowmetry, Sci. Rep. 15 (1) (2025) 643, http://dx.doi.org/10.1038/s41598-024-84978-w, Publisher: Nature Publishing Group, URL https://www.nature.com/articles/s41598-024-84978-w.

[16] E. Aslim, B. B T, L. Ng, T. Kuo, J. Chen, J.-M. Chen, L. Ng, Can machine-learning (ML) augmented audio-uroflowmetry distinguish between normal and abnormal flows from voiding sounds? 19, 2020, e2141, http://dx.doi.org/10.1016/S2666-1683(20)34049-0,

[17] R.A. Taylor, C.L. Moore, K.-H. Cheung, C. Brandt, Predicting urinary tract infections in the emergency department with machine learning, PLOS ONE 13 (3) (2018) e0194085, http://dx.doi.org/10.1371/journal.pone.0194085, Publisher: Public Library of Science, URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0194085.

[18] N. Liou, T. De, A. Urbanski, C. Chieng, Q. Kong, A.L. David, R. Khasriya, A. Yakimovich, H. Horsley, A clinical microscopy dataset to develop a deep learning diagnostic test for urinary tract infection, Sci. Data 11 (1) (2024) 155, http://dx.doi.org/10.1038/s41597-024-02975-0, Publisher: Nature Publishing Group, URL https://www.nature.com/articles/s41597-024-02975-0.

[19] J.M. Knorr, G.T. Werneburg, Machine learning and artificial intelligence to improve interpretation of urodynamics, Curr. Bl. Dysfunct Rep. 19 (1) (2024) 44–53, http://dx.doi.org/10.1007/s11884-023-00734-2.

[20] S. Ellahham, N. Ellahham, M.C.E. Simsekler, Application of artificial intelligence in the health care safety context: Opportunities and challenges, Am. J. Med. Qual. 35 (4) (2020) 341–348, http://dx.doi.org/10.1177/1062860619878515, Publisher: SAGE Publications Inc.

[21] J. Jin, Y. Chung, W. Kim, Y. Heo, J. Jeon, J. Hoh, J. Park, J. Jo, Classification of bladder emptying patterns by LSTM neural network trained using acoustic signatures, Sensors 21 (16) (2021) 5328, http://dx.doi.org/10.3390/s21165328, Number: 16 Publisher: Multidisciplinary Digital Publishing Institute, URL https://www.mdpi.com/1424-8220/21/16/5328.

[22] H.J. Lee, E.J. Aslim, B.T. Balamurali, L.Y.S. Ng, T.L.C. Kuo, C.M.Y. Lin, C.J. Clarke, P. Priyadarshinee, J.-M. Chen, L.G. Ng, Development and validation of a deep learning system for sound-based prediction of urinary flow, Eur. Urol. Focus. 9 (1) (2023) 209–215, http://dx.doi.org/10.1016/j.euf.2022.06.011, URL https://www.sciencedirect.com/science/article/pii/S2405456922001407.

[23] 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. Heal. Informatics 27 (5) (2023) 2166–2177, http://dx.doi.org/10.1109/JBHI.2022.3140590, Conference Name: IEEE Journal of Biomedical and Health Informatics, URL https://ieeexplore.ieee.org/abstract/document/9670631.

[24] D.-M. Huang, J. Huang, K. Qiao, N.-S. Zhong, H.-Z. Lu, W.-J. Wang, Deep learning-based lung sound analysis for intelligent stethoscope, Military Med. Res. 10 (1) (2023) 44, http://dx.doi.org/10.1186/s40779-023-00479-3.

[25] Hilbert Envelope Extraction from Real Discrete Finite Signals Considering the Nonlocality of Hilbert Transform | IEEE Conference Publication | IEEE Xplore, URL https://ieeexplore.ieee.org/abstract/document/9213286.

[26] M.F. Siddique, Z. Ahmad, J.-M. Kim, Pipeline leak diagnosis based on leak-augmented scalograms and deep learning, Eng. Appl. Comput. Fluid Mech. (2023) Publisher: Taylor & Francis, URL https://www.tandfonline.com/doi/abs/10.1080/19942060.2023.2225577.

[27] E.A. Martinez-Ríos, R. Bustamante-Bello, S. Navarro-Tuch, H. Perez-Meana, Applications of the generalized morse wavelets: A review, IEEE Access 11 (2023) 667–688, http://dx.doi.org/10.1109/ACCESS.2022.3232729, Conference Name: IEEE Access, URL https://ieeexplore.ieee.org/abstract/document/9999638.

[28] X. Wang, Y. Xu, An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index, IOP Conf. Ser.: Mater. Sci. Eng. 569 (5) (2019) 052024, http://dx.doi.org/10.1088/1757-899X/569/5/052024, Publisher: IOP Publishing.

[29] Journal of Medical Internet Research - A Smart Diaper System Using Bluetooth and Smartphones to Automatically Detect Urination and Volume of Voiding: Prospective Observational Pilot Study in an Acute Care Hospital, URL https://www.jmir.org/2021/7/e29979/.

[30] Critical Review of Uroflowmetry Methods | Journal of Medical and Biological Engineering, URL https://link.springer.com/article/10.1007/s40846-018-0375-0.

[31] T. Gupta, A. Kamath, A. Kembhavi, D. Hoiem, Towards general purpose vision systems: An end-to-end task-agnostic vision-language architecture, 2022, pp. 16399–16409, URL https://openaccess.thecvf.com/content/CVPR2022/html/Gupta_Towards_General_Purpose_Vision_Systems_An_End-to-End_Task-Agnostic_Vision-Language_Architecture_CVPR_2022_paper.html.