Automatic Diagnosis of Voiding Dysfunction from Sound Signal

Petr Hurtík*, Michal Burda*, Jan Krhut†‡, Peter Zvara‡, Libor Luňáček†‡
*Institute for Research and Applications of Fuzzy Modeling, NSC IT4Innovations, University of Ostrava, Czech Republic
†Department of Urology, University hospital, Ostrava, Czech Republic
‡Department of Surgical Studies, University of Ostrava, Czech Republic

***Abstract*—The aim of this paper is to present the results of an experiment towards Sonouroflowmetry, a novel approach for recognition of potential voiding dysfunctions based on machine learning classification of sound records that are obtained while a patient urinates into water in a toilet bowl. Such approach could enable a diagnosis of the voiding dysfunctions via a mobile device. We provide a comparison of 69 state-of-the-art classification methods.**

I. Introduction

Voiding dysfunction is highly prevalent and has a major impact on the quality of life of a large proportion of men. *Uroflowmetry* (UF) is a widely used non-invasive instrument for evaluation of bladder emptying. The measuring procedure involves the person to urinate into the uroflowmeter, often at a pre-determined time and at specified procedure area. This process is unnatural as it requires "on-demand" voiding often with either low or very high bladder filling, which leads to significant variability that should be mitigated by repeating the measures, which is costly and time consuming.

In this paper, we present an experiment that was performed towards the development of *Sonouroflowmetry* (SUF) [1], a novel approach for recognition of potential voiding dysfunctions based on capturing the sound generated when urine stream is hitting the water level in the toilet bowl. The idea is to enable the patients to perform the measurements at their home and at the time that is suitable for them. For that purpose, e.g. a cell phone with special sound recording application (or other small, cheap, and user-friendly sound recording device) may be used. Whereas the first paper, [1], gives overview from the medical point of view, this work is aimed to the mathematical and information science point of view.

The sound recognition is not a new problem. There exist a lot of real-life applications such as recognition of actually played song [2], speech recognition [3] in cell phones etc. The problem is very often solved by using neural networks [4]. Paper [4] also shows a comparison of existing methods and focuses on environmental sound recognition (such as footsteps, glass breaking etc). Their main idea is similar to ours, i.e. within the supervised machine learning methodology, we extract features from the sound signal and train the classification model. The most problematic part is naturally to develop the right set of features that correspond to the domain knowledge. Whereas the paper [4] focuses mainly on frequencies of a sound function, we focus on features reflecting medical expert's knowledge.

In this paper, we address several objectives. The first one is to explore possibilities of recording a sound using cell phones. The second one is to define characteristic values describing function that reflects expert decision. The third goal is to define an own learning algorithm and compare it with state-of-the-art algorithms. The last one (and the most important) is to analyze the possibilities of making diagnoses of a potential voiding dysfunction at home using a user-friendly non-invasive device such as a cell phone or other mobile device.

The structure of this paper is as follows: in Section II, the method of data collection and the structure of source dataset is discussed. We describe the data preparation phase, features extraction, and classification in Section III. In Section IV, the comparison of 69 state-of-the-art classification methods is presented. We discuss the obtained results and comment the former unsuccessful approaches in Section V. The conclusion and some proposals for future research are drawn in Section VI.

II. Data

In order to evaluate the possibility to recognize the voiding dysfunction based on a sound signal captured during urinating, the urologists prepared a simulation environment that allows repeated experiments under identical conditions. A water stream of constant height was directed into a container with water of constant column level. A microphone was put into a constant distance and sound recording was performed with predefined constant settings. Within the sound recording, a uroflowmeter was measuring the water flow. Figure 1 shows the simulation environment schematically.

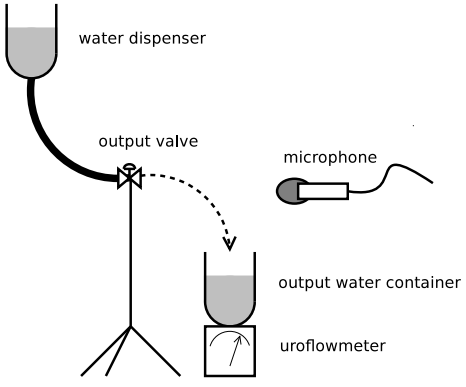

Fig. 1. Scheme of the simulation environment.

978-1-4799-7560-0/15 $31.00 © 2015 IEEE
DOI 10.1109/SSCI.2015.190

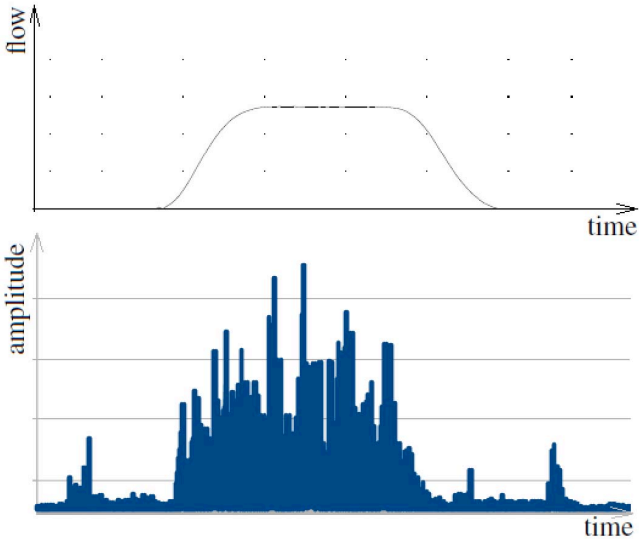

Fig. 2. The result of the UF measure is the graph of the flow-rate (y-axis) in time (x-axis) – top. The result of the SUF measure is the sampled sound volume (y-axis) in time (x-axis) – bottom.

The expert urologists simulated the flow of healthy (group A) and diseased men (group B) by changing the flow rate by obstructing or releasing the tube of water output. The simultaneous UF measures provided them an instant feedback of a quality of the simulation and allowed later a comparison with the SUF method being developed.

A total of 208 simultaneous UF and SUF recordings were included into the final analysis. Group A consisted of 114 recordings representing physiological voiding patterns. Group B consisted of 94 recordings representing the most frequent subtypes of pathological voiding patterns, namely: urethral stricture ($n = 27$), bladder outlet obstruction due to benign prostatic enlargement ($n = 29$), detrusor-sphincter dyssynergia ($n = 19$), and detrusor acontractility ($n = 19$).

Roughly speaking, the healthy patients are expected to have relatively high and stable stream of voiding, whereas pathological patients exhibit variable stream with lower average flow. Let us remark, whereas such differences can be easily seen and recognized using uroflowmetry, the behavior of sound is different, distorted by several artifacts and the recognition of pathological patterns is not so straightforward.

III. Methods

The whole process of machine learning consisted in the following parts (see also Fig. 4). The recorded sound signal in the WAV format was pre-processed and transformed into a time-series vector of integers that correspond to sampled amplitudes. From them, several numeric features were extracted so that the set of recorded source sound signals could be transformed into a numeric matrix, where rows correspond to measures (i.e. sound records) and columns contain the values of the extracted features. Also a column of correct classification to healthy/diseased was appended to the matrix.

Such feature matrix was repeatedly split into a training and testing part. The training part was used for machine learning to create a classification model. The testing part was used to evaluate the model performance. With respect to the size of the original dataset, we have decided to use the *leave-one-out* scheme (a.k.a. *jackknife*) for partitioning of features matrix into a training and testing part. In each step, a single row was considered as testing, while the rest was used to create the model. The resulting model was evaluated against testing row. Such process was repeated for each row, and the results were averaged.

A. Pre-process and Transformation

During the pre-process phase, the records with artifacts were removed from the experiment. By artifact we mean a sound that is not caused by the water flow - e.g. cough, phone ringing etc. The sound signal was trimmed from quiet parts in the beginning and in the end of the record. We used two methods of artifact removal. The first one is manual - an expert processes each sound graph and if an artifact occurs, the particular measure is removed. The second approach detects the biggest continuous part of non-zero sound amplitude automatically and trims everything else.

The records were transformed to a sequence of non-negative integers by taking the absolute values of the samples, i.e. each record was treated as a vector $\mathbf{a}$ of sequence $\mathbf{a} = (|a_k|)_{k=1}^{\ell}$ where ℓ is the length of the signal.

See Figure 2 for a comparison of the flow-rate measure performed by the UF device and the corresponding sound record.

B. Features Extraction

Based on the analysis of the results from UF and the experts' knowledge (see [5]), we have defined the following features, which we believe capture the important effects that are capable of the desired classification, and which were computed from the source sound signal sequences. In total, we have defined 27 features denoted with f_i, for $i \in \{1, \ldots, 27\}$. Figure 3 illustrates some of the features extracted from the sound signal.

Let $\mathbf{a} = (a_1, \ldots, a_\ell)$ be a vector representing the absolute values of the sound signal.

Let $\mathbf{s}_\tau$ be a vector of values from vector $\mathbf{a}$ without values that are lower than τ, i.e. $\mathbf{s}_\tau = (a_k \in \mathbf{a} \mid a_k > \tau)$.

Let $\mathbf{p} = (p_1, \ldots, p_\ell)$ be such permutation of vector $\mathbf{a}$ that $p_k \leq p_{k+1}$, for $k \in \{1, \ldots, \ell - 1\}$.

Let $\mathbf{g}_\delta = (g_1, \ldots, g_\ell)$ be a δ-gradient vector computed from $\mathbf{a}$ as follows (for $d \in \{1, \ldots, \ell\}$):

$$g_d = \max\{a_k \mid k \in \{d-\delta, \ldots, d+\delta\} \cap \{1, \ldots, \ell\}\} - \min\{a_k \mid k \in \{d-\delta, \ldots, d+\delta\} \cap \{1, \ldots, \ell\}\}.$$

Note that the proposed gradient is not a standard way of how the gradient is usually computed e.g. in computer graphics (see [6] for more details).

The final list of extracted features is as follows:

- The arithmetic mean of non-zero values:

$$f_1 = \frac{1}{|\mathbf{s}_0|} \sum_{s \in \mathbf{s}_0} s.$$

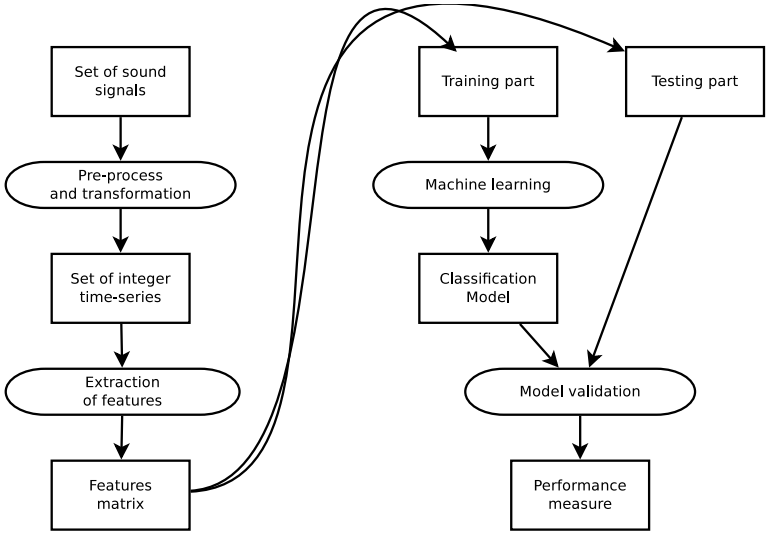

Fig. 4. Flow-chart of the whole sound signal processing and machine learning, as performed within our experiment. Processes are depicted as ovals, artifacts (i.e. data sets, models, results) as rectangles. The right part of the diagram was repeated multiple times, accordingly to the leave-one-out (jackknife) scheme of validation.

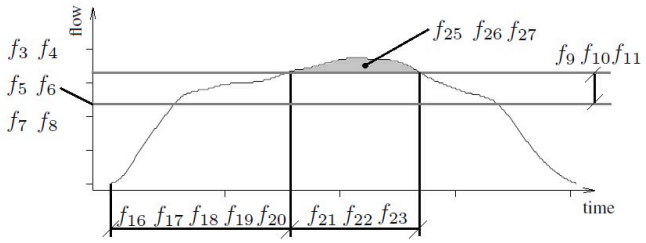

Fig. 3. The illustration of how the features extracted from SUF relate to the flow-rate function captured by UF.

This feature corresponds to the area under the curve of the sound signal, which approximates the volume of the voided urine.

- The arithmetic mean of the sequence vector $\mathbf{a}$:

$$f_2 = \frac{1}{\ell}\sum_{k=1}^{\ell} a_k.$$

In comparison with the previous case, here also the silent parts of the signal are taken into consideration.

- Percentiles of sound amplitude that approximates a maximum flow rate. The aim of the percentile is to mitigate the effect of a noise:

$$f_3 = p_{1\ell}, \quad f_4 = p_{0.95\ell}, \quad f_5 = p_{0.85\ell},$$

$$f_6 = p_{0.8\ell}, \quad f_7 = p_{0.75\ell}.$$

- Median:

$$f_8 = p_{0.5\ell}.$$

- A difference between 90% and 40% percentiles, which serves as a measure of variability because accordingly to the experts, the high variability often relates to pathological patterns. The 90 and 40 percentages were selected empirically.

$$f_9 = p_{0.9\ell} - p_{0.4\ell}.$$

Also the following variants were used:

$$f_{10} = p_{1\ell} - p_{0.9\ell}.$$

$$f_{11} = p_{1\ell} - p_{0.7\ell}.$$

- The arithmetic mean of the 1, 10, 100, and 1000-gradient:

$$f_{12} = \frac{1}{\ell}\sum_{g\in\mathbf{g}_1} g; \quad f_{13} = \frac{1}{\ell}\sum_{g\in\mathbf{g}_{10}} g.$$

$$f_{14} = \frac{1}{\ell}\sum_{g\in\mathbf{g}_{100}} g; \quad f_{15} = \frac{1}{\ell}\sum_{g\in\mathbf{g}_{1000}} g.$$

- Time to reach the 50%, 60%, 70%, 80%, and 90% percentile:

$$f_{16} = \min\{k \mid a_k > p_{0.5\ell}\};$$

$$f_{17} = \min\{k \mid a_k > p_{0.6\ell}\};$$

$$f_{18} = \min\{k \mid a_k > p_{0.7\ell}\};$$

$$f_{19} = \min\{k \mid a_k > p_{0.8\ell}\};$$

$$f_{20} = \min\{k \mid a_k > p_{0.9\ell}\}.$$

Accordingly to the experts, a healthy person should reach the maximum flow rate of voiding faster than a person with a dysfunction.

- Thresholded time:

$$f_{21} = |\mathbf{s}_{100}|; \quad f_{22} = |\mathbf{s}_{200}|; \quad f_{23} = |\mathbf{s}_{300}|.$$

- The product of the arithmetic mean and the arithmetic mean of non-zero values:

$$f_{24} = f_1 \cdot f_2$$

- Sums of thresholded values:

$$f_{25} = \sum_{s \in \mathbf{s}_{100}} s; \quad f_{26} = \sum_{s \in \mathbf{s}_{200}} s; \quad f_{27} = \sum_{s \in \mathbf{s}_{300}} s.$$

Let us remark that we have also tried various features derived from the frequency domain of the sound records. However, no significant influence on the resulting classification was observed. We expect the frequency filters and features based on frequencies to be useful later when developing automatic filters of sound artifacts and noise.

C. The State-of-the-Art Machine Learning Techniques

The feature matrix $F = [f_{j,i}]$ is created for $j = 1, \ldots, n$, $i = 1, \ldots, 27+1$, where i-th column corresponds to the feature f_i and the 28th column encodes the target classification to healthy/diseased, as provided by the experts, and j-th row corresponds to the j-th record of total $n = 208$.

We have tested several well known techniques that are available in the *caret* package [7] of the R statistical environment [8] – see Table I. We believe that the selection of methods nicely covers the state-of-the-art at the field of machine learners.

D. Our Learning Method

We have developed also a learning algorithm based on the idea of a binary decision tree [9]. The algorithm was originally used for visual recognition of jewelry stone defects [10]. The main idea is similar to decision trees and it is as follows: at each node of a tree, a decision is made which divides an input set into two disjoint sets. However, the selection of a feature to be used in a node as well as a construction of a separation condition in a node differs to the classical decision trees. For details, see [10]. In comparison with the original approach presented in [10], the following changes were made: the input function is one dimensional instead of two dimensional; new semantic based features were used (see section III-B); detection and removing extreme values in a learning phase was designed.

The most significant benefit of our approach is its simplicity that leads to very low computational cost, which makes our classification algorithm ideal for execution on mobile devices. For the purpose of classification problem described in this paper, we have further improved the algorithm by extending the classification condition in each node by introducing more thresholds to separate the extreme feature values. However, the mathematical background is out of the aim of the paper. In Table I of the results, our algorithm is denoted as "PROPOSED".

IV. Results

As can be seen from Table I, the majority of tested methods classified correctly more than 90 % of cases. The best accuracy (0,9471) was achieved by Bagged AdaBoost (*AdaBoostM1*) [11], Model Averaged Neural Network (*avNNet*) [12], and *glmnet* [13]. Such good quality of classification serve us as an empirical evidence of correctly selected features.

The proposed algorithm demonstrated one of the best accuracy rates in the task of jewelry stone defects recognition (see [10]). In the task of classification of sound records described in this paper, the accuracy of our improved algorithm is 91.35 %.

TABLE I. Classification Accuracy as Obtained from Various Classification Methods

method	accuracy	method	accuracy
AdaBoost.M1	0.9471	Boruta	0.9183
avNNet	0.9471	bstLs	0.9183
glmnet	0.9471	bstSm	0.9183
fda	0.9423	ctree	0.9183
spls	0.9423	FRBCS.CHI	0.9183
svmPoly	0.9423	FRBCS.W	0.9183
kernelpls	0.9375	partDSA	0.9183
nnet	0.9375	rFerns	0.9183
pls	0.9375	rpart2	0.9183
plsRglm	0.9375	**PROPOSED**	0.9135
simpls	0.9375	AdaBag	0.9135
widekernelpls	0.9375	bagFDAGCV	0.9135
bstTree	0.9327	glm	0.9135
LogitBoost	0.9327	parRF	0.9135
sda	0.9327	RFlda	0.9135
bagEarth	0.9279	RSimca	0.9135
CSimca	0.9279	blackboost	0.9087
lda	0.9279	gamLoess	0.9038
Mlda	0.9279	hdda	0.9038
pda	0.9279	lssvmRadial	0.9038
rocc	0.9279	C5.0Rules	0.8990
RRFglobal	0.9279	treebag	0.8990
svmRadialCost	0.9279	pam	0.8942
svmRadial	0.9279	PenalizedLDA	0.8846
svmRadialWeights	0.9279	C5.0Tree	0.8798
bagEarthGCV	0.9231	SLAVE	0.8750
bagFDA	0.9231	knn	0.8413
C5.0Cost	0.9231	lvq	0.8365
C5.0	0.9231	rbfDDA	0.8077
gcvEarth	0.9231	ada	0.8039
kknn	0.9231	GFS.GCCL	0.7596
multinom	0.9231	mlpWeightDecay	0.5913
nodeHarvest	0.9231	rbf	0.5481
rf	0.9231	oblique.tree	0.4519
RRF	0.9231	protoclass	0.4519

V. Discussion

This paper describes a successful approach for classification of SUF sound records. The results look promising, indeed, the way to obtain them was quite complicated. In this section, we would like to describe some previous attempts and dead-ends.

A. Approach 1 – A Proprietary Sound Recording Software on a Mobile Device

The result of UF is a function u of flow-rate on time (see Figure 2). From that, the physician can see the process of voiding of the patient in order to formulate a diagnosis. The original idea was to develop SUF to get results similar to UF. In other words, having a sound record s, which is a function of volume of sound on time, we wanted to find a mapping $s \to s'$ such that $\mathrm{corr}(u, s')$ is maximized (or equal to 1, ideally).

For that purpose, a proprietary sound recording application on a mobile device was used to obtain a set of measures, where the UF measure was performed simultaneously with the SUF recordings.

Accordingly to the experts, the most important characteristics of the UF graph are: maximum voiding flow-rate ($Qmax$);

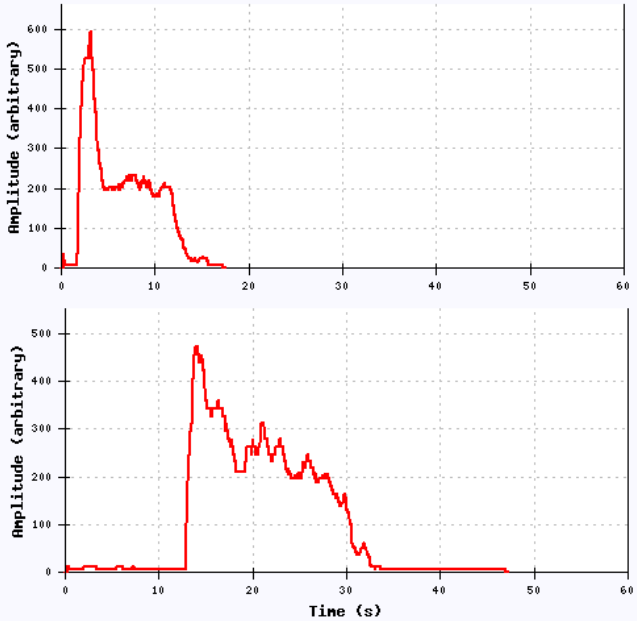

Fig. 6. Cell phone automatic volume normalization problem: whereas the top graph reaches higher value than lower one, UF says the maximal value is opposite. Moreover, consider the peak at the beginning of the recording that we suppose is caused by the Automatic Gain Control too.

the total voiding time ($Time$); and the total voided volume (Vol), which can be represented with the area under the curve in the UF graph (see Figure 2). Therefore, the comparison between UF and SUF was performed as a correlation on these three characteristics. Figure 5 shows the obtained correlations.

We tried to develop the mapping $s \rightarrow s'$ using various convolutions and non-linear filters. As depicted in Figure 5, our mapping improved the correlation of $Time$ (from 0.56 to 0.91), but on the other hand, the correlation of $Qmax$ or Vol remained almost the same or got even worse. The reason of that problem was discovered in the Android OS that automatically normalizes the sound volume in order to make the record audible and pleasant to the humans. Such feature is called the *Automatic Gain Control*. However, it is undesirable for SUF measures. See Figure 6 for visual demonstration of that problem.

B. Approach 2 – Developing a Sound Recording Software for Android OS

Based on the findings discussed above, we developed an own Android application for sound recording with automatic volume control turned off in the Android SDK. However, probably because of some bug in the SDK, the cell phone still insisted to control the volume automatically. Therefore, we postponed the development of the mobile device version until the problems with automatic gain control get resolved.

C. Approach 3 – Developing a Sound Recording Software on the Personal Computer

Finally, a sound recording software was developed for a personal computer running the MS Windows OS that can use an integrated or external microphone. Unfortunately, although the sound recording quality improved significantly, the correlations between SUF and UF did not improve as expected.

D. Approach 4 – Searching the Mapping with Differential Evolution

Until now, we tried to define the mapping $s \rightarrow s'$ manually with the aid of an expert knowledge. In our subsequent experiment, we searched the mapping with the help of differential evolution. A function that processed the raw sound signal was encoded into the genes of the differential evolution and we optimized the fitness function maximizing the correspondence of SUF and UF. As the task was very computationally demanding, we have run it on Anselm HPC in Ostrava.

Unfortunately, we have failed to find a suitable solution as it seems that the sound of voiding does not reflect ideally the flow rate. This unsuccessful research led us to reformulating the objective and not to try to obtain UF results from SUF, but to directly classify dysfunctions from SUF, as described in this paper, which finally led to success.

VI. Conclusion

As can be seen from the reported results, the SUF method of voiding dysfunction classification seems very promising. However, this experiment is still too much laboratory-centric and is far away from being directly applicable to the medical practice. We see many challenges that must be tackled before reaching the applicability in practice.

For instance, the real-life sound recordings may be encumbered with the bias caused by:

- varying sound conditions in a room where the measure takes place;
- sound jamming, artifacts;
- different quality of devices performing the recording;
- different height of urine stream per patient;
- different size and shape of toilet bowl, different amount of water inside;
- different distance or position of recording device microphone, etc.

We tried to avoid all such sources of bias in this preliminary study by setting constant laboratory conditions.

For the readers who are interested, the CSV file with data used for classification was uploaded to `graphicwg.irafm.osu.cz/storage/suf.csv`.
If you are interested, you can test your machine learning algorithm and send us your results.

We have also tested our own classification algorithm that was successfully used in the stone defect recognition task [10]. Although it does not achieve the highest success rate in the task described above, its proven benefit is again a very low computational cost.

Future work may address classification of the recorded sounds not only to healthy/diseased, but also to the type of a disease or dysfunction. We plan to define more features,

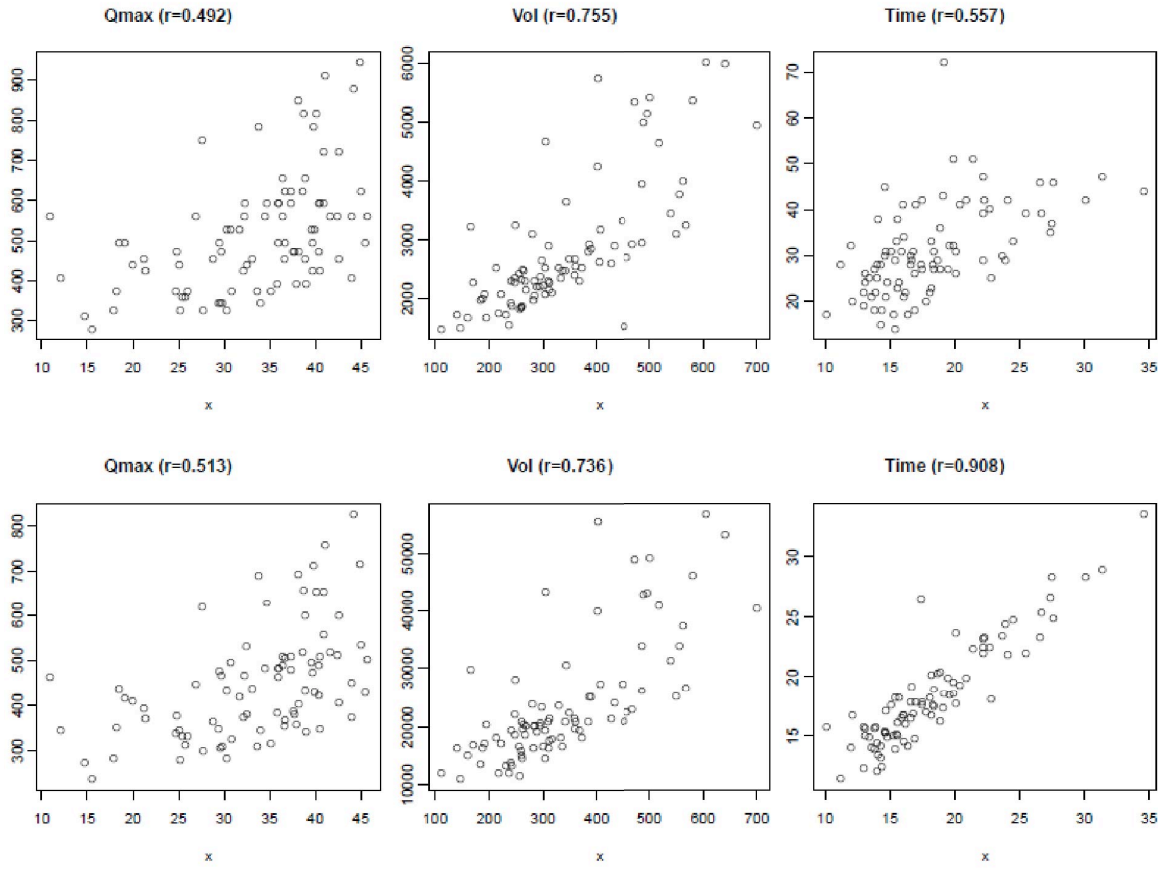

Fig. 5. Correlations between UF (x-axis) and SUF (y-axis) with respect to $Qmax$, Vol, and $Time$. Top: original data (s); Bottom: transformed data (s').

employ some advanced feature selection technique, or perform classification with some ensemble of classifiers. The main objective is to improve sound recognition pre-processing and artifact removal.

ACKNOWLEDGMENT

This work was supported by the European Regional Development Fund in the project of IT4Innovations Center of Excellence (CZ.1.05/1.1.00/02.0070, VP6) and by the IT4I XS project number LQ1602.

REFERENCES

[1] J. Krhut, M. Gärtner, R. Sỳkora, P. Hurtík, M. Burda, L. Luňáček, K. Zvarová, and P. Zvara, "Comparison between uroflowmetry and sonouroflowmetry in recording of urinary flow in healthy men," *International Journal of Urology*, 2015.

[2] A. Wang, "The shazam music recognition service," *Communications of the ACM*, vol. 49, no. 8, pp. 44–48, 2006.

[3] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 1, pp. 30–42, 2012.

[4] M. Cowling and R. Sitte, "Comparison of techniques for environmental sound recognition," *Pattern recognition letters*, vol. 24, no. 15, pp. 2895–2907, 2003.

[5] A. J. Wein and E. S. Rovner, "Pathophysiology and classification of voiding dysfunction," in *Voiding Dysfunction*. Springer, 2000, pp. 3–24.

[6] N. Madrid and P. Hurtik, "Lane departure warning for mobile devices based on a fuzzy representation of images," *Fuzzy Sets and Systems*, 2015, (accepted for publickation).

[7] M. Kuhn, *caret: Classification and Regression Training*, 2014, R package version 6.0-37. [Online]. Available: http://CRAN.R-project.org/package=caret

[8] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2015. [Online]. Available: http://www.R-project.org/

[9] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.

[10] P. Hurtik, M. Burda, and I. Perfilieva, "An image recognition approach to classification of jewelry stone defects," in *IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), 2013 Joint*. IEEE, 2013, pp. 727–732.

[11] E. Alfaro, M. Gámez, and N. García, "adabag: An R package for classification with boosting and bagging," *Journal of Statistical Software*, vol. 54, no. 2, pp. 1–35, 2013. [Online]. Available: http://www.jstatsoft.org/v54/i02/

[12] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, 4th ed. New York: Springer, 2002, iSBN 0-387-95457-0. [Online]. Available: http://www.stats.ox.ac.uk/pub/MASS4

[13] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010. [Online]. Available: http://www.jstatsoft.org/v33/i01/