Machine Learning for Automated Bladder Event Classification from Single-Channel Vesical Pressure Recordings

V. Abbaraju[1,2], K. Lewis[3] and S.J.A. Majerus[1,2]

[1]Department of Electrical, Computer and Systems Engineering, Case Western Reserve University, Cleveland, OH
[2]Advanced Platform Technology Center, Louis Stokes Cleveland Veterans Affairs Medical Center, Cleveland, OH
[3]Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, OH
{vxa112, sjm18}@case.edu, lewisk9@ccf.org

***Abstract*— Analyzing urodynamic study (UDS) tracings can be prone to error in the presence of artifacts and subjective due to lack of standardization in clinical UDS interpretation. As such, the diagnosis of patients undergoing UDS would greatly benefit from a standardized, automated method to assist clinicians in interpreting UDS tracings. In this work, we evaluated a machine learning framework for automatically classifying bladder events from single-channel vesical pressure recordings (P_{VES}) ($N = 60$) into 4 possible classes: abdominal event (i.e., cough or Valsalva), voiding contraction, detrusor overactivity (DO) and no event. Wavelet multiresolution analysis of P_{VES} was used to extract time-frequency localized wavelet coefficient vectors which were segmented into 0.8 second segments with 55 statistical features per segment. Feature selection was subsequently applied for three classifier architectures: a k-nearest classifier (KNN), an artificial neural network classifier (ANN) and a support vector machine classifier (SVM). Each classifier was trained and evaluated using five-fold cross validation, from which we derived the sensitivity, specificity, F1 score and AUC for all four classes and the overall classification accuracy for each classifier. The KNN, ANN and SVM classifiers labeled 7,861 0.8 second P_{VES} segments with 91.5%, 90.8% and 82.4% accuracy, respectively. We have thus proposed the first framework for automated multi-event bladder classification using single-channel UDS data.**

Keywords*— *urodynamics, wavelet signal processing, bladder event classification

I. Introduction

Multi-channel urodynamics (UDS) is a diagnostic procedure that provides a functional assessment of lower urinary tract function. UDS provides qualitative and quantitative analyses of bladder relaxation and contractile behavior during retrograde bladder filling and voiding [1]. This information allows physicians to diagnose lower urinary tract dysfunction, including detrusor overactivity (DO), detrusor underactivity and urinary incontinence, among other diagnoses. Although UDS allows for precise calculation of several parameters with well-established guidelines, the interpretation of UDS traces can be subjective and variable. This is in part due to the presence of artifacts and technician-related errors [2] and may also be due to lack of standardization in clinical urodynamic interpretation [3]. An automated, standardized methodology for the detection and classification of bladder events in UDS tracings may allow physicians to focus on meaningful periods in the study, speed up interpretation, prevent errors, and improve consistency when analyzing UDS tracings.

Typically, a multi-channel UDS study is conducted with two inserted pressure-sensing catheters: an intraurethral catheter which measures vesical pressure (P_{VES}), and a vaginal or rectal catheter which measures abdominal pressure (P_{ABD}). The detrusor pressure is calculated as the difference between the two:

$$P_{DET} = P_{VES} - P_{ABD} \quad (1)$$

P_{DET} is evaluated by a clinician in the context of other pressure and flow tracings to interpret UDS tracings [1]. However, using two sensors for recordings—especially during ambulatory urodynamics, which allows for longer periods of data recording—is a source of physical discomfort in patients due to the insertion of the catheters [4].

Thus, there arises a need for a standardized, automated methodology for the interpretation of ambulatory UDS tracings using single-channel pressure data acquired using only an intravesical sensor. Prior work focused on automated binary DO detection using pattern recognition, modelling instances of DO as wave-shapes and making predictions based on the measured similarity of an unlabeled event to a bank of aggregated DO wave-shapes [5]. Newer methods have built on automating DO classification by applying machine learning, extracting time-domain and frequency-domain features from windowed P_{DET} data for binary classification of DO from already-detected, yet unlabeled, events [6]. While such techniques rely on both sensors (P_{VES} and P_{ABD}) for event detection and classification, detecting bladder contractions and pressure artifacts from single-channel data (P_{VES}) using wavelet multiresolution analysis and adaptive thresholding has also been investigated [7].

Here, we present a machine learning framework for classifying four classes of bladder events from single-channel pressure recordings (Fig. 1). To our knowledge, this is the first example of multi-event classification from single-channel UDS pressure data. Section II describes the methodology for annotating and pre-processing the dataset, statistical feature extraction using wavelet multiresolution analysis, classification using different machine learning classifiers, and classifier evaluation and tuning. Section III discusses the performance of the classifiers and potential performance improvements for a real-time hardware implementation of the framework.

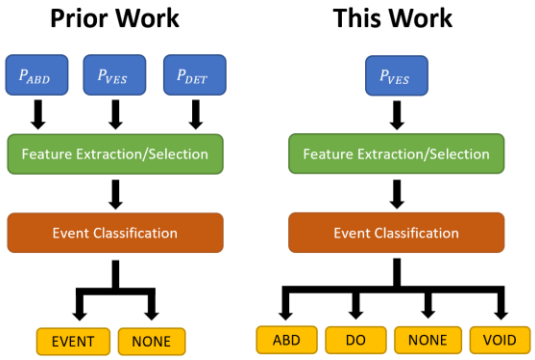

Fig. 1. While prior work focused on single-event detection or binary classification from multiple pressure channels (left), we focused on four-class bladder event classification using single-channel vesical pressure recordings (right).

II. Methodology

A. Dataset Preparation

60 UDS tracings sampled at 10 Hz were obtained from 34 human subjects with overactive bladder or neurogenic urinary incontinence through urodynamic testing conducted at the Louis Stokes Cleveland Department of Veteran Affairs Medical Center and the Cleveland Clinic. Anonymized data were harvested from previous studies and were IRB exempt. Each tracing included P_{VES}, P_{ABD}, P_{DET}, volume and flow information and was analyzed by a blinded urologist to annotate abdominal events (ABD), voiding contractions (VOID), and detrusor overactivity (DO) (Fig. 2). Noisy P_{VES} segments, typically found towards the start or end of a tracing and caused by catheter motion, were discarded from each tracing.

B. Wavelet-Based Feature Extraction

DO events and voiding contractions were usually represented as low-frequency rises and falls, while abdominal artifacts contained high-frequency spikes in pressure. While the fast Fourier transform (FFT) and short-time Fourier transform (STFT) are typically used to determine the frequency content of a time-varying signal, they suffer from the trade-off between time and frequency resolution. The discrete wavelet transform (DWT) was therefore used for improved temporal localization of frequency content in non-stationary signals. Previously, the Daubechies 4 wavelet was shown to be useful for extracting time-frequency localized information from P_{VES} data [7].

The five-level discrete wavelet transform (DWT) using the Daubechies 4 wavelet was applied on the entire P_{VES} signal in each tracing, resulting in five sets of detail and approximation coefficients for each tracing. Each set of coefficients was subsequently upsampled by a 2^n factor using spline interpolation, where n denotes the transform level, to retain the original time scale of the P_{VES} tracing. The result was 10 wavelet coefficient vectors per P_{VES} tracing (Fig. 3). Each P_{VES} tracing was split into non-overlapping segments of 0.8 seconds in length (8 P_{VES} samples), and each P_{VES} segment was then labeled with one of four possible class labels, depending on which event was present in the segment: 'ABD', 'VOID,' 'DO' or 'NONE' (if no event was present).

In prior work in classifying time-varying physiological signals such as ECG and EEG using machine learning, statistical properties of wavelet coefficients such as the mean, maximum, Shannon entropy, power, etc. were used as features in classifiers [8-9]. Adopting this

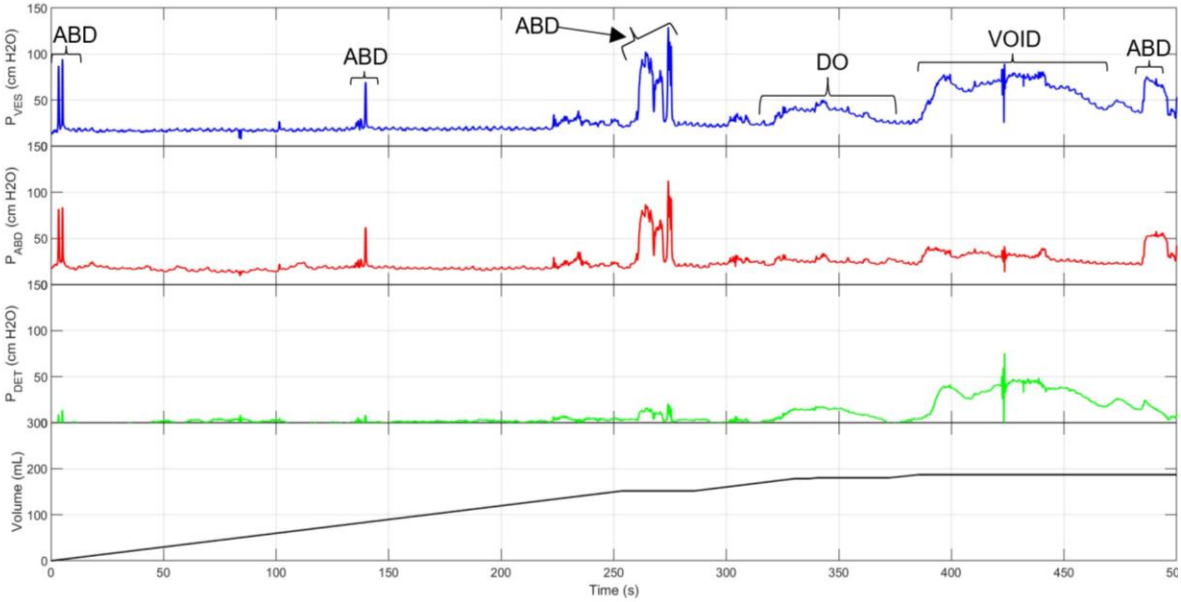

Figure 2. Typical UDS tracing with annotations. The annotations were performed using information from all three pressure channels: P_{VES}, P_{ABD} and P_{DET}. Our work focused on extracting features and making predictions solely from P_{VES}.

approach for vesical pressure, for each P_{VES} segment, the maximum, mean absolute value (MAV), median and Shannon entropy from each of the 10 wavelet coefficient vectors was computed, resulting in 40 features for each P_{VES} segment. Furthermore, for each P_{VES} segment, the cross-correlation between approximation coefficients and corresponding detail coefficients in each DWT level was calculated, resulting in five cross-correlation vectors for each P_{VES} segment. The maximum, mean and median from each cross-correlation vector were also computed, resulting in 15 additional features for each P_{VES} segment.

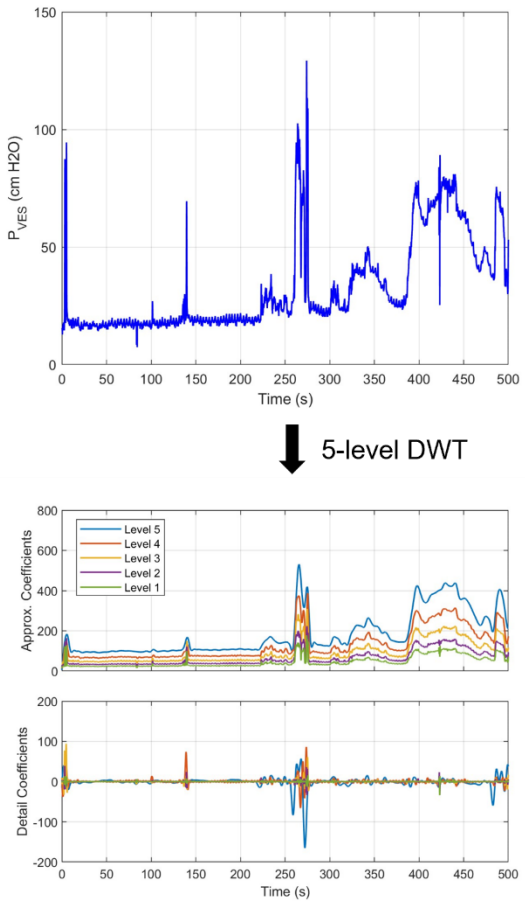

Fig. 3. Application of 5-level DWT on P_{VES} tracing, resulting in 10 total wavelet coefficient vectors. The approximation and detail coefficients localized frequency content in P_{VES}.

Thus, 55 features were computed for each 8-sample P_{VES} segment in each of the 60 P_{VES} tracings (Fig. 4). The N_i x 55 feature vectors from each P_{VES} tracing, where N_i denotes the number of P_{VES} segments in the i^{th} P_{VES} tracing, were combined to form an initial single 53,496 x 55 feature matrix.

C. Dimensionality Reduction through Feature Selection

Feature selection was applied to reduce the training time and overfitting of the classifiers [8-9]. The most relevant features from the feature matrix were identified using the ReliefF algorithm, which assigns weights to features using a nearest-neighbors approach [10]. ReliefF rewards features with large differences for observations of different classes and penalizes features with large differences for observations of the same class [10]. This technique identified the m highest-ranked features, where m denotes the number of features chosen for each classifier architecture.

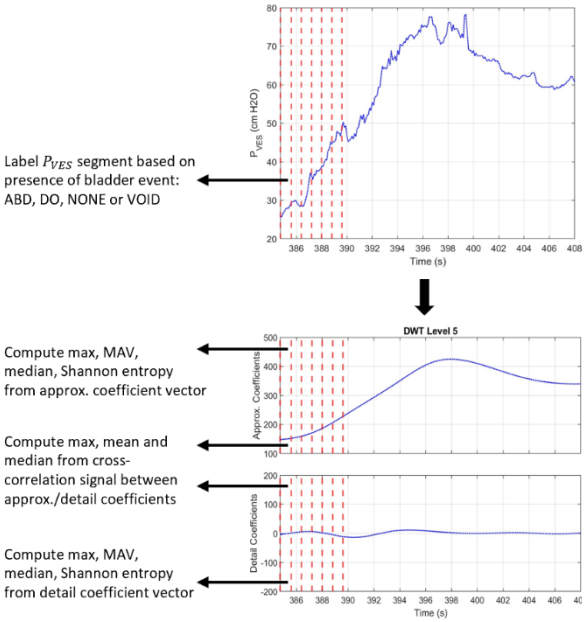

Fig. 4. Example of segmentation of P_{VES} and statistical feature extraction from fifth level of the DWT. This process was performed for all five DWT levels, resulting in 55 features for each P_{VES} segment.

D. Classifier Selection and Training

The majority (~78%) of the initial dataset was made up of observations belonging to the 'NONE' class, followed by ~15% of observations belonging to 'VOID.' To balance the dataset and limit bias during training and cross-validation, observations from each of the four classes were randomly chosen such that each class represented ~25% of the observations. Using this approach, the final dataset comprised of 7,861 total observations: 1,861 'ABD' events, 2,000 'DO' events, 2,000 'NONE' events, and 2,000 'VOID' events.

Three initial classifier models for the machine learning framework were evaluated: 1) a k-nearest neighbors classifier (KNN) with $k = 1$; 2) an artificial feed-forward neural network classifier (ANN) with a 10-neuron hidden layer activated using the ReLU function; 3) a support vector machine classifier with the radial basis function kernel (SVM-RBF), $C = 1$ and $\gamma = 1$. The initial hyperparameters for the ANN and SVM-RBF classifiers were chosen based on prior work in physiological signal classification using DWT-based features [8-9]. The k value for the KNN classifier was chosen based on initial testing with the 7,861 x 55 feature matrix, which showed

that using a single neighbor resulted in strong classification accuracy. Each classifier was trained and evaluated using 5-fold cross validation.

E. Classifier Tuning and Performance Metrics

Each classifier was evaluated on overall classification accuracy (# of correctly predicted P_{VES} segment labels). Throughout the framework development and testing process, the hyperparameters and the number of features for each classifier were adjusted to improve performance (Fig. 5). For instance, ReliefF was applied to select the 12 most relevant features for the KNN and SVM-RBF classifiers; testing showed that there was no increase in classification accuracy when more than 12 features were used for these two classifiers. Furthermore, setting $\gamma = 0.94$ improved performance of the SVM-RBF classifier. For the ANN classifier, adding a second hidden layer, increasing the number of neurons in each hidden layer to 100, and including all 55 features substantially increased classification accuracy. The final architectures for all three classifiers that were tested are shown in Table I.

TABLE I. FINAL CLASSIFIER ARCHITECTURES

Classifier	Number of Features	Hyperparameters
KNN	12	$k = 1$
ANN	55	Hidden layers: 2 x 100 neurons/layer Activation: ReLU
SVM-RBF	12	$\gamma = 0.94$

Furthermore, for each classifier, the specificity, sensitivity and F1 score for each class was calculated as

$$Sensitivity\ (Recall) = \frac{TN}{TP + FP} \quad (2)$$

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$F1 = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (5)$$

where TP, FP, TN and FN represented the true positives, false positives, true negatives, and false negatives, respectively. Furthermore, for each classifier, the receiver operating characteristic (ROC) and the area under the ROC curve (AUC) were calculated.

III. RESULTS AND DISCUSSION

The overall accuracy and the per-class performance metrics for the three machine learning classifiers were calculated (Fig. 6 and Table I). Although KNN classifier performed the best in terms of overall classification accuracy, the ANN classifier's performance was comparable to the KNN in terms of sensitivity, specificity and F1 scores for the four classes. Furthermore, the ANN classifier achieved the best ROC

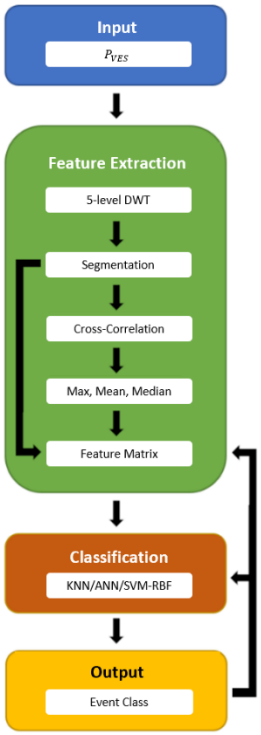

Fig. 5. Flow of bladder event classification framework with underlying steps from P_{VES} input to event class output. During classifier testing and tuning, feedback from the output was used to iterate on feature selection and classifier hyperparameters to optimize classification performance.

AUC for all four classes. The high sensitivity of the KNN classifier for 'DO' suggests that it most effectively detected these events in this class, but the ANN generalized more effectively to all four classes of events.

Despite achieving upwards of 82% overall classification accuracy across all 3 classifiers, there was room for better performance. In this study, each 0.8 second-long P_{VES} segment was treated as an independent event; the full length of bladder events, which can range from 5 seconds for an abdominal artifact to 100 seconds for a voiding contraction, was not considered. This approach had its advantages, as the three classifiers labeled short time segments without relying on memory. However, using temporal information based on previous segments would improve performance. For instance, recurrent neural networks (RNNs), and specialized RNNs, such as the long short-term memory (LSTM) can combat the vanishing gradient prevalent in neural networks with large hidden layers. For example, LSTMs have achieved promising results in ECG signal classification [11].

While the classifiers performed well at classifying the different types of bladder events ('ABD,' 'DO,' and 'VOID'), they were not as adept at distinguishing these events from 'NONE' segments, i.e., noise.

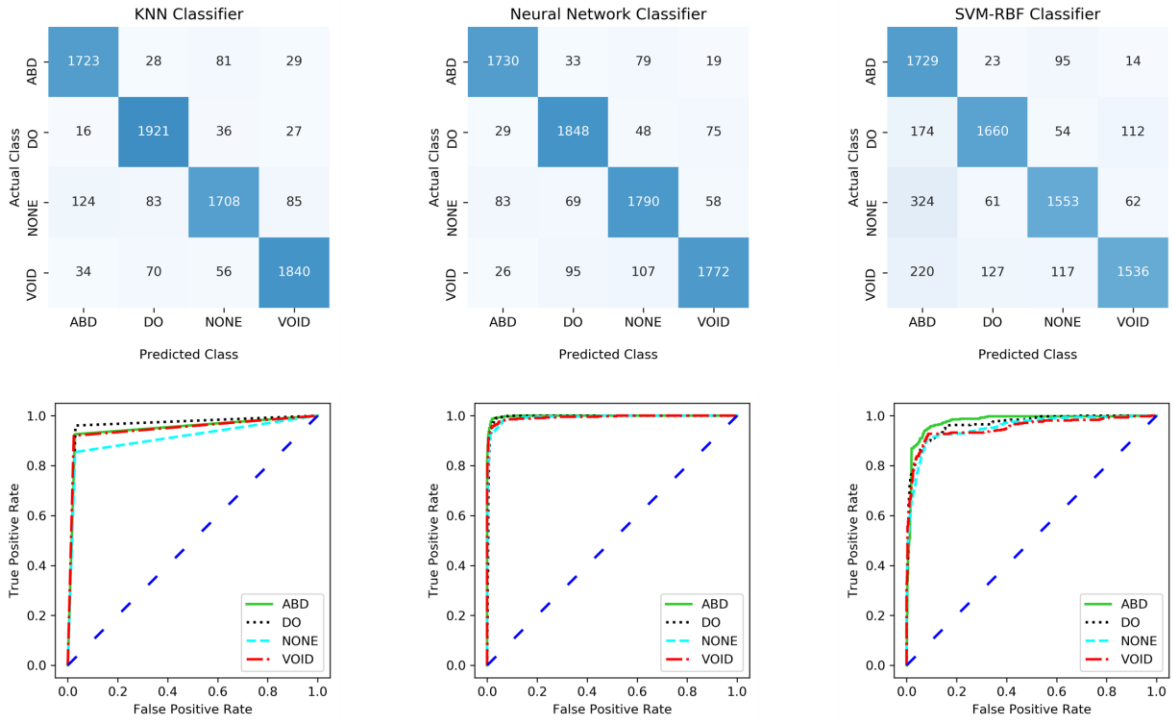

Fig. 6. Summary of prediction results using five-fold cross validation of all three classifiers. The confusion matrices (top) show the true positives, false positives, true negatives, and false negatives per class for each classifier. The ROC curves (bottom) indicate the trade-off between sensitivity and specificity per class for each classifier.

TABLE II. SUMMARY OF RESULTS FOR MULTI-CLASS CLASSIFICATION

Classifier	Event	Sensitivity	Specificity	Precision	Recall	F1	ROC AUC	Accuracy
KNN	ABD	92.58%	97.10%	0.91	0.93	0.92	0.95 ± 0.00	91.49%
	DO	96.05%	96.91%	0.91	0.96	0.94	0.96 ± 0.00	
	NONE	85.40%	97.05%	0.91	0.85	0.88	0.91 ± 0.01	
	VOID	92.00%	97.59%	0.93	0.92	0.92	0.95 ± 0.01	
ANN	ABD	92.96%	97.70%	0.93	0.93	0.93	0.99 ± 0.01	90.83%
	DO	92.40%	96.64%	0.90	0.92	0.91	0.99 ± 0.01	
	NONE	89.50%	96.01%	0.88	0.90	0.89	0.98 ± 0.02	
	VOID	88.60%	97.41%	0.92	0.89	0.90	0.99 ± 0.01	
SVM-RBF	ABD	92.91%	88.03%	0.71	0.93	0.80	0.97 ± 0.00	82.41%
	DO	83.00%	96.40%	0.89	0.83	0.86	0.97 ± 0.00	
	NONE	77.65%	95.46%	0.85	0.78	0.81	0.95 ± 0.01	
	VOID	76.80%	96.79%	0.89	0.77	0.82	0.95 ± 0.01	

TABLE III. SUMMARY OF RELATED WORKS

Work	Year	Channels	Feature Extraction	Classification	Classes	Findings
R. Karam et al. [7]	2016	1	5-level DWT	Adaptive thresholding	2	97% true positive rate
H.-H. S. Wang et al. [5]	2020	3	Wave-shape manifold model	Dynamic time warping	2	81.35% accuracy, 0.84 ROC AUC
K.T. Hobbs et al. [6]	2021	3	Windowed time/freq. analysis	SVM-RBF	2	0.91 ROC AUC
Proposed	2022	1	5-level DWT	KNN, ANN, SVM-RBF	4	91.49%, 90.83%, 82.41% accuracies, ROC AUCs ranging 0.91-0.99

Future work could include a two-stage filtering process: 1) a "coarse" filter that separates useful P_{VES} data from noise and detects the onset of a bladder event; 2) a "fine" classifier that uses wavelet-based features and temporal information from both current/past states to classify the event of the current P_{VES} segment being processed. Coarse detection was previously demonstrated through adaptive thresholding and adaptive sampling techniques [7, 11]. The fine classification step would rely on a machine learning classifier as presented in this work.

IV. Conclusion

A novel supervised machine learning framework for predicting UDS tracing segments from single-channel P_{VES} data was demonstrated. Three machine learning classifiers, using statistical features extracted using the DWT, were capable of classifying P_{VES} segments at accuracies ranging from ~82% for an SVM-based classifier to ~91% for a KNN-based classifier. This work showed the efficacy of a machine learning approach for automated bladder event classification using only a single catheter line; this may prove useful for aiding in UDS interpretation and in reducing the number of catheters required for UDS in some patients. Future work will include a two-step event detection/classification process and implementing the framework in hardware as a real-time algorithm.

References

[1] M. J. Drake, S. K. Doumouchtsis, H. Hashim, and A. Gammie, "Fundamentals of urodynamic practice, based on International Continence Society good urodynamic practices recommendations," *Neurourology and Urodynamics*, vol. 37, no. S6, pp. S50–S60, 2018.

[2] J. Melgaard and N. J. M. Rijkhoff, "Detecting Urinary Bladder Contractions: Methods and Devices," *JST*, vol. 04, no. 04, pp. 165–176, 2014.

[3] T. L. Frenkl *et al.*, "Variability of urodynamic parameters in patients with overactive bladder," *Neurourology and Urodynamics*, vol. 30, no. 8, pp. 1565–1569, 2011.

[4] A. M. Suskind *et al.*, "Patient perceptions of physical and emotional discomfort related to urodynamic testing; a questionnaire-based study in men and women with and without neurologic conditions," *Urology*, vol. 85, no. 3, pp. 547–551, Mar. 2015.

[5] H.-H. S. Wang, D. Cahill, J. Panagides, C. P. Nelson, H.-T. Wu, and C. Estrada, "Pattern recognition algorithm to identify detrusor overactivity on urodynamics," *Neurourology and Urodynamics*, vol. 40, no. 1, pp. 428–434, 2021.

[6] K. T. Hobbs *et al.*, "Machine Learning for Urodynamic Detection of Detrusor Overactivity," *Urology*, vol. 159, pp. 247–254, Jan. 2022.

[7] R. Karam *et al.*, "Real-Time Classification of Bladder Events for Effective Diagnosis and Treatment of Urinary Incontinence," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 721–729, Apr. 2016.

[8] A. J. Chashmi and M. C. Amirani, "An Efficient and Automatic ECG Arrhythmia Diagnosis System using DWT and HOS Features and Entropy- Based Feature Selection Procedure," *J Electr Bioimpedance*, vol. 10, no. 1, pp. 47–54, Aug. 2019.

[9] A. al-Qerem, F. Kharbat, S. Nashwan, S. Ashraf, and khairi blaou, "General model for best feature extraction of EEG using discrete wavelet transform wavelet family and differential evolution," *International Journal of Distributed Sensor Networks*, vol. 16, no. 3, p. 1550147720911009, Mar. 2020.

[10] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *Journal of Biomedical Informatics*, vol. 85, pp. 189–203, Sep. 2018.

[11] S. Saadatnejad, M. Oveisi, and M. Hashemi, "LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 515–523, Feb. 2020.