Unsupervised Classification of Spontaneous Action Potentials in Urinary Bladder

M. Padmakumar, B. Jacob, R. Venkatakrishnan, K. L. Brain, and R. Manchanda

Abstract—It is observed that the spontaneous action potentials (sAPs) observed in a urinary bladder smooth muscle cell are of different shapes. The biophysical mechanisms underlying this shape variety is not yet known. It is assumed that the syncytial properties of the smooth muscle tissue in urinary bladder affect the shape of the sAPs generated by the smooth muscle cell. Further investigation on the matter requires accurate identification of different types of sAPs observable from the detrusor smooth muscle cells. Since such a ground truth on the number of possible sAP classes is not available and the manual identification of the sAP classes from long intracellular recording is tedious and erroneous, it becomes necessary to use an unsupervised classification algorithm to classify the observed sAP signals. K-means clustering and hierarchical clustering algorithms (both agglomerative and divisive approaches) are some of such classical clustering algorithms available. Also considering the different ways in which the data can be presented (such as raw time domain data, Fourier transform, wavelet transform, and principal components), There are multiple approaches to do the signal classification. In this study, the clustering results of all these approaches are compared and the best performing methods are shortlisted. An internal measure called cluster balance was used to quantitatively evaluate the resulting clusters.

Index Terms—Action potential, cluster balance, pattern recognition, smooth muscles, unsupervised signal classification, urinary bladder.

I. INTRODUCTION

The neuromuscular system in our body is made of special cells called the excitable cells [1]. They maintain a potential difference between inside and outside of the cell. This difference in potential is caused by the semi permeable nature of the cell membrane which has ion selective pores called ion channels, where each ion channel is specific to a single ion [2]. The potential difference between the intracellular fluid (ICF) and the extracellular fluid (ECF) is called the resting membrane potential (RMP). The ion channels can be in an 'open' state when the ion movement through the channel is allowed, and otherwise in a closed state when no ion movement is possible. The way in which the ion channels

changes from one state to another, or the channel dynamics, are designed such that impulse-like signals called the action potentials (APs) could be generated by the excitable cells. The APs are the characteristic feature of excitable cells [1], [2].

The excitable cells are categorized in to two major groups - Nerve cells and Muscle cells. The nerve cells are specialized in decision making and information transfer. They transmit information – transformed into a series of APs – from one part of the body to other. Muscle cells receive input from the neurons, causing APs to get generated in them which in turn results in their contraction.

The muscle cells themselves could be classified into three subcategories [1]. They are (i) skeletal muscles, (ii) cardiac muscles, and (iii) Smooth muscles. The skeletal muscle cells are attached to the skeleton. They are innervated by motor neurons of central nervous system and could be controlled voluntarily. The individual skeletal muscle cells are electrically insulated from the neighboring cells and are controlled individually by a single motor neuron. The cardiac cells are muscle cells that are present in the heart. They form an electrically interconnected syncytium [3] so that the AP produced in one cardiac cell can travel to the neighboring cells. This electrical coupling is provided by special protein structures called the gap junctions. The third type of muscle cells, called the smooth muscle cells aid the involuntary movements of all the internal organs like the gut, urinary bladder, uterus, blood vessels, vas deferens, *et al*. They too make interconnected syncytium and are innervated by the sympathetic and parasympathetic nervous system [4], [5].

Smooth muscle cells are the smallest among the three muscle categories. Their width is about 2-6 μm which is around one tenth of that of the skeletal muscle cells [5]-[7]. The extent of electrical coupling between the smooth muscle cells in the syncytium varies with the organ. For example, the gap junctional coupling of uterus smooth muscle cells (at the initial stage of the pregnancy) are less compared to that of the vas deferens smooth muscle tissue. The smooth muscle syncytium is not homogeneous – the interconnected cells form bundles [7], [8]. The smooth muscle cells in a single bundle are well coupled to each other and the coupling across bundles is poor. The innervation style of the smooth muscle is also different from the skeletal muscles. The axons of the nerve cells pass through the intercellular space of the smooth muscle cells and occasionally form swellings along the length called varicosities [6]. These varicosities act as the neurotransmitter release points. A single axon can form multiple varicosities and thus excite multiple smooth muscle cells. Also, a single smooth muscle cell can receive inputs from multiple nerve cells [6]. All the above mentioned attributes of the smooth muscle tissue system – the small size,

the non-homogeneous electrical syncytium, and the many-to-many mapping of the nerve and muscle cells make the comprehension of the smooth muscle tissue very difficult. The details of the electrical and mechanical activities of the smooth muscle tissues are still under exploration [9].

In the present work, the focus is on the smooth muscle cells in the urinary bladder. The urinary bladder is an organ which helps the mammals to store urine and micturate (or void) when needed. The layer of smooth muscle cells, called the detrusor, present in the bladder wall maintains the tone of the bladder during the storage phase and provides a coordinated contraction during the voiding phase [1]. The physiology of bladder is designed to match both these requirements. The detrusor tissue is made up of smooth muscle bundles which run in different directions [10]. The voiding process is initiated by the stimulation of the parasympathetic nerves. The varicosities present in the detrusor smooth muscle (DSM) tissue release neurotransmitter packets which depolarize the smooth muscle cells and thus generate action potentials in them, resulting in a coordinated contraction. During the storage phase, the varicosities spontaneously release neurotransmitter packets in a random manner [6]. These may cause APs (called the spontaneously generated action potentials, or spontaneous action potentials, or sAPs) in the smooth muscle cells. These sAPs, if generated, could travel to the nearby smooth muscle cells through the gap junctional coupling, resulting in the contraction of the corresponding smooth muscle bundle [9], [11]. These uncoordinated contractions of smooth muscle bundles caused by the spontaneous neurotransmitter release are believed to be the way by which the bladder maintains its tone during the storage phase.
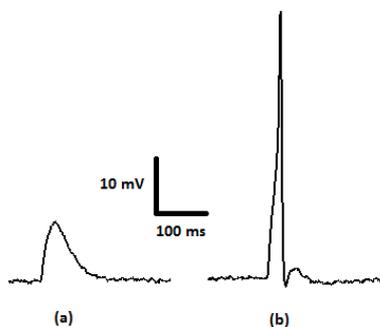
Fig. 1. Comparison between (a) an STD and (b) an sAP. Note the high amplitude and quick falling profile of the sAP.

All spontaneous neurotransmitter release from the varicosities cannot produce an AP in the smooth muscle cell [7]. If the intercellular gap junctional coupling is strong and if the neighboring cells are at the resting potential, the depolarizing ionic current initiated by the neurotransmitters get distributed to the neighboring smooth muscle cells via gap junctions, thus reducing the depolarization induced in the receiving smooth muscle cell [7], [9]. If the depolarization did not reach the threshold to produce an AP, the passive signals called the spontaneous transient depolarizations (STDs) are observed. The STDs are identified by its characteristic shape, which is commonly represented by an alpha function. A comparison of an AP and an STD is shown in Fig. 1. While the AP exhibits "all or none" property, whose amplitude is restricted in a narrow range, the STDs display graded amplitude which vary from near zero to tens of millivolts.

Apart from the sAPs and STDs, there are some other signals rarely observed in the DSM, the discussion about which are avoided in this paper.

It is interesting to note that the sAPs seen in the intracellular recording from a single DSM cell (DSMC) are not identical in shape [12]. The shapes of APs generated by different type of excitable cells may vary, but generally a single excitable cell, including the smooth muscle cells of other organs, always exhibit a signature AP shape. The reason behind this inconsistency is not well understood so far. However, it is assumed that the DSM syncytial properties such as non-homogeneous arrangement of smooth muscle cells, the neurotransmitter release profile of the varicosities, the size of the smooth muscle bundle, and / or the intensity of the gap junctional coupling, could modulate the sAP shape exhibited by the DSMC and thus induce the varied shape.

To explore the factors affecting the shape of the sAP, it is required to identify the different shapes that are observable from the DSMC recordings. There exist many clustering algorithms that are able to classify the signals into different clusters, such as *K*-means, hierarchical clustering methods, artificial neural networks, support vector machine, *et al.* [13]-[15]. But most of the clustering techniques need a training data set to optimize various parameters needed for the classification operation [14]. As we do not have a ground truth available for the groups present in the DSMC recordings, such a training data is not available. Thus ANN and SVM techniques cannot be used here. The *K*-means algorithm does not need a training data [16], but it requires information about the number of classes present in the data, which is not available as well. The hierarchical clustering methods (both agglomerative and divisive strategies) do not have these requirements and return a clustering tree that includes all the data points into one class on one end and all the data points as individual classes at the other [16]. The optimal clustering usually lies somewhere in between these two extremes, which need to be identified using some evaluation criteria.

The evaluation needed to identify the optimal clustering is based on the internal measurements on the clustered data, such as intracluster standard deviation and inter cluster spread. The Davies-Bouldin index and Dunn index are two widely used internal evaluation techniques [9], [17], [18]. But the values of those indices either monotonically increase or decrease with the number of clusters formed. Hence they cannot be used to find out the optimum number of clusters present in the data. Another such internal evaluation index is the clustering balance [13]. The clustering balance index is designed to assume a minimum value when the number of clusters formed is optimum. This minimum value obtained could also be used to compare the efficiency of different clustering algorithms as well. Hence the clustering balance is used also as the evaluation criteria for this work.

The efficiency of the clustering procedure could depend on the way the input data is represented [14], [15], [19]. In the present study, the input data for classification are the isolated signals sampled at a certain frequency, pre-processed so that the baseline shifted to zero, peaks aligned, and all the signals trimmed to equal length. These signals could be represented in other domains as well. For example, they could be transformed into frequency domain using Fourier transform,

to the wavelet coefficients using wavelet transform, or to the Eigen vectors using principal component analysis (PCA). It would be worthwhile to evaluate which of the data representation yield the best classification results.

In this paper, a preliminary work to investigate on the most efficient automatic classification procedure suitable for the intracellular recorded signals from the DSMCs is presented. The data used for the study is the intracellularly recordings from mouse detrusor tissue. The distances between signals pairs were measured as Euclidian distance measure. The clustering results were evaluated using the cluster balance index. The effect of different data representations such as wavelet transformation and PCA were also studied.

II. METHOD

A. Data Acquisition

Mice of the BALB/c strain, of either sex, were sacrificed in accordance with the UK Animals (Scientific Procedures) Act 1986 and European Communities Council Directive 86/09/EEC. The urinary bladder was removed and the connective tissue surrounding the bladder removed, with the urothelium left intact. Tissue strips, which contained a few bundles of smooth muscle, 3 - 4 mm long and 1 - 2 mm wide, were cut. Strips were pinned out on a Sylgard-lined plate at the bottom of a shallow chamber (volume, approximately 1 ml), which was mounted on the stage of an upright microscope. Preparations were superfused with warmed (35°C) physiological saline solution (PSS) (composition, mM: NaCl, 120; KCl, 5.9; $MgCl_2$, 1.2; $CaCl_2$, 2.5; $NaHCO_3$, 15.5; NaH_2PO_4, 1.2 and glucose, 11.5; gassed with 95% O_2 and 5% CO_2) at a constant flow rate (100 ml/h), maintaining a pH of 7.2 - 7.3 [20].

Preparations, when pinned, were allowed to equilibrate for at least 30 min before electrophysiological recording. Individual bladder smooth muscle cells in muscle bundles were impaled with glass capillary microelectrodes, filled with 0.5 M KCl (tip resistance, 100-300 MΩ). Membrane potential changes were recorded using a high input impedance amplifier (Axoclamp-2B, Axon Instruments, Inc., Sunnyvale, CA, USA). Membrane potential changes were digitized using PowerLab/4SP (ADInstruments, Chalgrove, UK) at either 1 kHz or 4 kHz, and stored on computer for later analysis.

B. Data Pre-processing

The data files were resampled to 1 kHz sampling rate. In order to isolate the sAPs from the continuous intracellular recordings, the data file was accessed as segments of one-minute duration. The baseline, or the resting membrane potential (RMP), is determined in each of the segments and the activities with amplitude above a predefined threshold (set as 35 mV) from the RMP were identified and isolated. Those activities were manually verified to be sAPs and false detections were eliminated. The starting locations and span of identified sAPs were indexed in an associate file corresponding to the data file for later use. The associate files were prepared for all the data files used in this study.

To present the sAP signals to a classification algorithm, the segments of data file which contain the sAP signals - indexed in the associate files - were extracted first. These extracted sAP signals are peak aligned, and are trimmed to have equal lengths (or equal number of samples). The RMP is subtracted from all the signals so that the baselines of the signals are set to zero. After these pre-processing steps, the signals were stored in an array where each column represents the corresponding samples, and each row represents a sAP signal. This array is given as an input to the clustering algorithms.

C. Feature Extraction Techniques

The sAP signals digitized and recorded in time domain, could be transformed into different domains such as frequency domain (using Fourier transform), wavelet domain (using wavelet transform) or could be represented as eigenvectors (using PCA). The signal features, not very visible in time domain might get more prominence in other domains. Hence the signals in other domains were also used for the clustering operation. The details of the possible transformations are given below.

The wavelet allows the signals to be seen in a time-frequency domain. The time domain signals are converted to wavelet coefficients. A template wave segment (called mother wavelet) is used to obtain the wavelet coefficients. There are different mother wavelet functions are available, which are sometimes categorized into different families. It is generally assumed that the wavelet which has the maximum similarity with the signal of interest is preferred for the wavelet analysis. However due the variety of sAP signals, it is not possible to identify a single wavelet function which could be used for the classification application. As the discrete Meyer wavelet (dmey), the Daubechies family wavelets (db), Discrete Meyer wavelet and the symlet family wavelets (sym) were commonly used for biosignal applications, they were chosen to be used as the basis for the wavelet transformation incorporated to aid the signal classification operation. Out of these wavelet options, the best wavelets are to be shortlisted based on the quantitative evaluation of the clustering results.

Using principal component analysis (PCA) [19], it would be possible to convert the time domain data into a set of eigenvectors, the sum of which would represent the signals in a set of transformed axes such that the maximum signal variance is captured in the eigenvector corresponding to the highest eigenvalue. It had been shown that PCA based classification is useful for sound classification which has most of the important information stored below 500 Hz [21]. As the maximum frequency content of the intracellular recordings in the present study is below 500 HZ (as the sampling rate is 1 kHz), it is expected that the PCA would be helpful in the classification of the sAP signals as well.

D. Classification Techniques

One of the most frequently used clustering technique is the vector quantization using the K-means algorithm. In this algorithm, it is assumed that the number of clusters present in the data is known and is given as the value of K. The algorithm then randomly chooses K data points as the cluster seeds. All the data points are then separated into K groups based on the smallest Euclidean distance from the seeds. The centroid of each of the groups thus formed are calculated and called cluster centers. All the data points are then regrouped, keeping the cluster centers as new seeds. This procedure is iterated until the cluster centers are stabilized and remain the

same for two consecutive iterations. The drawback of this method is that the final cluster result depends on the allotment of initial seeds. Another limitation, specific to our application, is that the number of sAP signal classes present in a data pool is unknown. Hence it would be required to run the K-means clustering procedure repeatedly with value of K ranging from 1 to 20 (It is assumed that the number of sAP classes observed in a single cell recording is less than 20). The quantification of the clustering result corresponding to each value of K is evaluated and the best result is assumed as the optimal classification.

The requirement of the knowledge of the number of clusters demanded by the centroid models such as k-means clustering mentioned above is not present in the connectivity models such as the hierarchical clustering algorithm. This makes them more suitable for our current requirement. In the divisive approach, all the signals present in the input data pool are considered to be the part of one single cluster to begin with. This cluster then divided into two so that the average distance between the signals belonging to the same cluster is minimized. Each of these new clusters formed are then further divided into two each following the same procedure. This procedure is continued until the number of clusters formed is equal to the number of signals present in the data pool, or in other words all clusters formed have only one member in it. The optimal number of clusters formed lies between the starting state and the ending state, which is identified either by visual inspection or by quantitative evaluation of the clustering result at after each stage.

The other approach to hierarchical clustering, called the agglomerative approach, follows the above procedure in reverse direction. Here, every signal is initialized as separate clusters. Then signals with smallest distance between them are combined to make a new bigger cluster. This process is continued until all the signals end up as a part of one single cluster. The Euclidean distance is used as a measure of distance between two signals. However the distance between two clusters, called the cluster linkage, could be evaluated in multiple ways. The following are the commonly used linkages in agglomerative clustering:

Single Linkage: The distance between two clusters is calculated as the closest distance between two clusters.

Complete Linkage: The distance between the two farthest points in the two clusters.

Average Linkage: The average of the distances measured between all pair of signals taken between two clusters.

Weighted Linkage: The weighted linkage is similar to the average linkage but the distances are also based on the number of elements in the cluster and this is important when the clusters that are being compared are not of equal sizes.

Ward's Linkage: In this method, the increment in the total intra-cluster variance after merging the two clusters together is taken as the linkage between those clusters. The cluster pair with minimum linkage, *i.e.*, the pair when combined causes the minimum increment in intra-cluster variance is combined in the next level of agglomerative clustering.

The clustering techniques mentioned above, including the different possibilities of the cluster linkage, are the available options for the unsupervised classification for the sAP signals. To find out the optimum number of clusters present in the data pool, and to quantify the quality of clusters produced by different approaches, a robust index need to be defined. We propose to use the cluster balance as such an internal measure of cluster quality. The cluster balance index is described below in detail.

E. Evaluation of Classification Results

For developing an unsupervised Classification algorithm it is necessary to quantitatively evaluate different classification techniques to help determine the best algorithm by iterating and getting the best combination using the evaluation technique. External evaluation is unsuitable for this study because there are no training data available with the ground truth to evaluate the efficiency of the clustering techniques (Fig. 2).

In the case of internal evaluation Davies–Bouldin index and Dunn index have been commonly used in literature [22]. Davies–Bouldin index is ratio of intra-cluster distances and inter-cluster distances [17]. When the cluster has been formed with the minimum ratio, it is said to have the best classification. However this method would be useful only when the number of clusters in a data pool is predefined. As the number of clusters is increased, the ratio keeps on decreasing monotonously and hence makes it impossible to determine the optimum number of clusters present in the data pool. A similar situation exists in the case of Dunn index which is a ratio of inter-cluster distance and intra-cluster distance. Here the value keeps on increasing monotonously [18].

The clustering balance as proposed in [10] offers a technique that quantifies the balance between the intra-cluster variance and the inter-cluster variance of the data. The intra cluster variance is to be minimum and the inter cluster variance should be maximum. As the number of clusters increased, the intra cluster variance decreases monotonously and the inter cluster variance increases monotonously. Mathematically, the intra cluster variance (or intra cluster error sum of squares, Δ) is defined as

$$\Delta = \sum_{j=1}^{k} \sum_{i=1}^{n_j} e(q_i^j, q_0^j) \tag{1}$$

where q_i^j and q_0^j are the ith data point and centroid of the jth cluster respectively. k represents the number of clusters and the value of nj represents the number of data points present in the jth cluster. Function $e(\mathbf{a}, \mathbf{b})$ represents the measurement of Euclidean distance between the vectors $\mathbf{a}$ and $\mathbf{b}$. The inter-cluster error sum of squares Γ is defined as

$$\Gamma = \sum_{j=1}^{k} e(q_0^j, q_0) \tag{2}$$

where q_0 is the centroid of the entire data pool. The cluster balance index (ε) is then designed as the weighted sum of the intra cluster error sum and the inter cluster error sum as given below.

$$\varepsilon = \Delta(\alpha) + (1-\alpha)\Gamma \tag{3}$$

where α is a constant which vary between 0 and 1. IN this paper, the value of α is kept as 0.5, giving equal weightage to Δ and Γ. Cluster balance value is high when the number of clusters is very low, owing to the high value of intra cluster

error sum. If the number of clusters is too large, the inter cluster error sum becomes very large and influence the ε value. When the number of clusters is optimum, a balance is reached and the value of ε attains a minimum value.
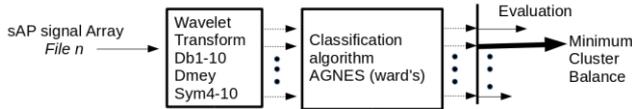
Fig. 2. Schematic diagram explaining the procedure for the selection of the best performing wavelet functions for classification procedure.

The clustering balance index is mainly designed to obtain the optimum number of clusters present in the data. However, it can be used to quantitatively evaluate the quality of the clusters. The lower value of the cluster balance would indicate the better clustering, irrespective of the number of clusters. This principle is used to compare the efficiency of different clustering algorithms for a common data pool.

F. Shortlisting the Optimum Clustering Procedure

If there are *N* feature-selection methods and *M* clustering methods available, there would be *N×M* ways in which the signal classification be done for each of the data files. Hence to minimize the number of options in hand, the low performing options are to be eliminated.

There is a large number of wavelet options (db1 to db20, sym1 to sym20, and dmey) available for the feature selection operation. The performance of each of the wavelet function is studied by fixing all other parameters like the classification algorithm and the input data file. The best performing wavelets would be shortlisted for further analysis and the rest are discarded.

The options available for classification algorithms are K-means, divisive hierarchical clustering, agglomerative hierarchical clustering with five possible linkage options-single, complete, average, weighted, and Ward's. Out of these seven available options, the approaches which perform consistently bad in classifying sAPs are to be ignored as well. This is done by fixing the input data format as the time domain signals, and executing the available classification algorithms on each of the data file. The algorithm(s) which performs consistently poor for all data files are rejected.

After eliminating the less performing options in the feature selection stage and the classification stage, the performance of different combinations of the feature selection and clustering methods are studied. Combinations that consistently perform worse for all the data files are rejected, and the combinations that give good clustering balance value for majority of the data files are retained. These shortlisted combinations of feature detection and classification algorithms were used for unsupervised signal classification of sAP signals recorded intracellularly from detrusor smooth muscle cells.

III. RESULTS

The data files in which the intracellular recordings were stored were resampled to 1 kHz and inspected. Each of the data file represented recording taken from a single tissue preparation. It was observed that some of the long-duration data files contained recordings from more than one cells. In such cases, the sections of data which represented the recording from individual cell were isolated and treated as separate files, called cell recordings. 11 such cell recordings were chosen for the classification study presented in this paper. In this paper, these cell recordings are addressed as File 1, File 2 ... and File 11 respectively for convenience. The sAP signals present in each of the cell recordings were identified using a threshold of 35 mV above the baseline (see Section methods) and their locations were indexed in corresponding associate files. The sAP signals indexed in the associate files were manually verified and the erroneous detections, if present ($< 2\%$) were eliminated.

The time domain sAP signal array for the classification purpose was prepared by extracting the data segments where sAP signals were present, peak aligning them, subtracting the baseline (RMP) value from the signals, and fixing the duration of the sAP signals so that the entire time span of all the sAP signals were contained. The pre-processed signals were then arranged in an array such that the rows represented individual signals. This signal array was then given to the feature-extraction stage, the output of which is then given as an input to the classification algorithm.

A. Shortlisting the Wavelet Functions for Feature Detection

The wavelet functions under considerations were Daubechies (db1-db10), Symlets (sym4-sym10), and discrete Meyer (dmey). These wavelet functions, one after another, used to take wavelet transforms of the signals in the sAP signal array and the transformed signals were given as input to the agglomerative clustering procedure with Ward's linkage. The resulting clusters were evaluated using cluster balance index. For each cell recording (or file), the wavelet functions were arranged in the ascending order of the cluster balance value. The wavelet functions consistently placed in the first 4 places are db1, dmey, db2, and sym8 respectively. Two best performing wavelet functions - db1 and dmey - were shortlisted for feature selection application and the rest of the wavelets were discarded.

B. Shortlisting the Clustering Techniques

As described in the Section II, the sAP signal array, without any transformations, were given as the input to different classification algorithms and their performance were measured using the cluster balance index. As the K-means clustering algorithm followed a centroid based model which is different from the connective model followed by the hierarchical algorithms, the evaluation for the former was done separately (Fig. 3).
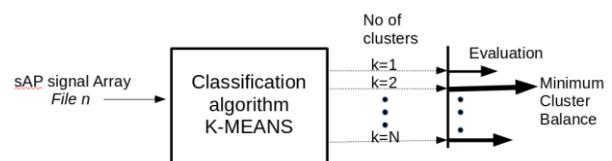
Fig. 3. Schematic diagram showing the evaluation of optimal cluster number using *K*-means algorithm.

While evaluating the *K*-Means classification algorithm it was observed that every iteration of the algorithm on the same signal data resulted in different results, including the

optimum number of clusters. This was caused by the random initialization of the cluster seeds which during iterations would then influence the final result. An attempt was made to tackle this problem by using the multiple runs of *K*-means algorithm on the same data file and finalize the clustering result with the minimum cluster balance index as the best result obtainable from the *K*-means procedure. A file was run 60 times through the *K*-Means clustering algorithm, and the optimum number of clusters obtained are shown in the Table I. It could be observed that the method do not conclusively suggest a single number for the clusters present in the data pool. Also considering the huge amount of computational requirement to follow this procedure, it is concluded that the *K*-means algorithm cannot be used for the purpose of unsupervised sAP signal classification.

TABLE I: OCCURRENCE OF OPTIMAL CLUSTER NUMBER IN 60 ITERATIONS OF *K*-MEANS ALGORITHM ON FILE 1

# of clusters	16	9	8	7	6	5	3
count	1	5	13	16	20	3	2

Combined evaluation procedure was followed for shortlisting the set of clustering methods from AGNES with different linkages (single, complete, average, weighted, and Ward's) and DIANA. The signal data was passed on to the clustering algorithm and each method was ranked as per its

performance which was determined by the ascending order of the Cluster Balance value. This process was repeated on all the eleven files used for the study and the results are shown in Table II. From the table, it could be observed that the AGNES procedure with the 'single' linkage consistently performed poor for all the data files analysed. Other methods exhibit mixed performances for different files. Hence the AGNES technique with 'single' linkage was removed (along with the *K*-means clustering) from the study and the rest were accepted.

TABLE II: PERFORMANCE EVALUATION OF HIERARCHICAL CLUSTERING ALGORITHMS (AGNES AND DIANA) FOR EACH OF THE DATA FILES. 'D' STANDS FOR DIANA. THE ENTRIES OTHER THAN DIANA REPRESENT DIFFERENT LINKAGES USED IN AGNES CLUSTERING

File	Best	→	→	→	→	Worst
File1	war.	com.	avg.	avg.	D	sin.
File2	war.	com.	wei.	D	avg.	sin.
File3	war.	com.	D	avg.	wei.	sin.
File4	war.	wei.	com.	D	avg.	sin.
File5	war.	D	com.	wei.	avg.	sin.
File6	war.	com.	D	avg.	avg.	sin.
File7	war.	com.	D	wei.	avg.	sin.
File8	war.	com.	D	wei.	avg.	sin.
File9	war.	com.	D	wei.	avg.	sin.
File10	com.	ward	wei.	D	avg.	sin.
File11	war.	wei.	com.	D	avg.	sin.
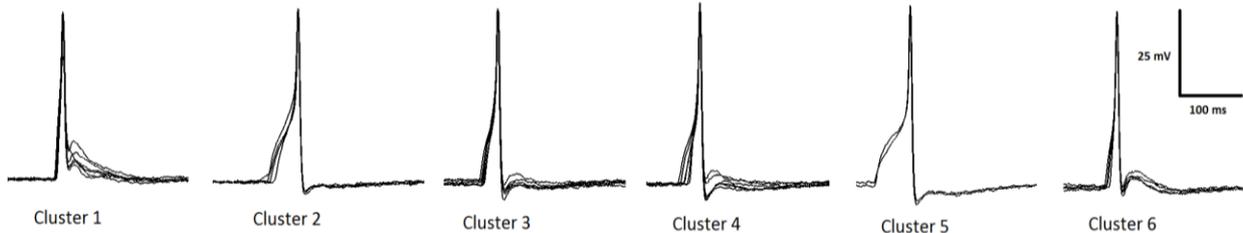
Fig. 4. The clustering result obtained for File 1 using the proposed combination of wavelet (db1) feature extraction and AGNES-ward clustering algorithm.

C. The Best Combination of the Feature and the Clustering Technique

Five types of data representation techniques - (i) time domain, (ii) Fourier transform, (iii) wavelet transform - db1, (iv) wavelet transform - dmey, and (v) PCA and five types of classification algorithms - (i - iv) AGNES with linkages 'complete', 'average', 'weighted', and Ward's, and (v) DIANA were shortlisted after eliminating the low performing options. It means that there exists 25 ways in which the clustering operation can be carried out for every data file. In order to identify the right combination of data representation and clustering algorithm which gives rise to the best clustering result all the combinations were tried out on each of the data files. The clustering balance was obtained for every clustering results thus generated. Based on the value of the clustering balance, each of the combinations were ranked from best to worst. The combination which gave the best clustering result for each of the data file is tabulated in Table III. From the table, it could be observed that the best unsupervised classification of sAP signals were obtained using AGNES with Ward's or complete linkage when the data was presented in time domain or wavelet domain (db1).

As an example of the output obtained from the procedure, the classification of sAP signals in the File 1 is given in Fig.

4. By visual inspection it could be qualitatively rated that the resulting clusters were indeed satisfactory.

TABLE III: OPTIMAL COMBINATION OF FEATURE EXTRACTION AND CLUSTERING TECHNIQUE FOR EACH DATA FILE

File Name	Feature extraction	Cluster Method
File 1	Wavelet - db1	AGNES-ward
File 2	Time domain	AGNES-ward
File 3	Wavelet - db1	AGNES-ward
File 4	Time domain	AGNES-ward
File 5	Wavelet - db1	AGNES-ward
File 6	Time domain	AGNES-ward
File 7	Time domain	AGNES-ward
File 8	Time domain	AGNES-ward
File 9	Time domain	AGNES-ward
File 10	Time domain	AGNES-complete
File 11	Wavelet - db1	AGNES-ward

IV. DISCUSSION

The study presented here shortlists a handful of approaches that could be taken for the unsupervised classification of intracellularly recorded sAP signals from a mouse detrusor. It

is to be noted that different data files preferred different approaches to get the maximum clustering efficiency. This could be due to the fact that the properties of sAP signals present in different data files are different in nature, and the efficiency of a clustering technique depends on the properties such as type of signals, the extent of dissimilarity between the clusters, *et al.* This means that for the unsupervised classification of sAP signals in a new data file, the 4 different clustering approaches shortlisted in the study had to be carried out, and the best among the clustering results thus produced is to be finalized.

The efficiency of the clustering algorithm could be further improved by combining the signal features of multiple domains. For example, the time domain signal and its wavelets transform could be combined with appropriate weights to form a composite-feature vectors which could help the classification algorithm to discriminate between the different clusters better. The aspects of the features in other domains such as the frequency spectrum, PCA components *et al.* could also be included in the composite vector. The biophysical details about the signals such as the amplitude, time course *et al.* could also be used.

The major challenge faced in the study was the evaluation of the clusters produced. The most efficient way to evaluate a clustering procedure is to test it against a known result. Unfortunately, such a ground truth is currently not available for the signals of interest. Hence it was required to restrict ourselves to the internal measures of cluster evaluation. The cluster balance index, on which the results of the proposed work depend heavily, is not a widely used mea-sure for the quantitative evaluation of the clustering result. However the index gave satisfactory results in the present work.

The unsupervised classification of the spontaneous action potentials would be very helpful in exploring the syncytial arrangement, the nature of innervation and the electrical properties of the detrusor smooth muscle tissue. The different signal groups obtained from the clustering results could be used to prepare a comprehensive list of the templates corresponding to various signal types observable from the detrusor tissue of a mouse bladder. As these templates are related to the biophysical environment in which the cell is present, they could be used to predict the biophysical environment of the cell from which the recording is taken. Such investigations would be carried forward in the future.

V. CONCLUSION

It could be concluded that AGNES classification technique with Ward's linkage is most efficient for the unsupervised classification for the spontaneous APs of the detrusor smooth muscle cells. The preprocessing of the signals using wavelet (db1 or Haar) transform increases the clustering efficiency in some of the data files. The entity called clustering balance could be used to estimate the optimal number of clusters and also as a criterion for the internal evaluation of the clustering algorithm.

ACKNOWLEDGEMENTS

REFERENCES

[1] A. C. Guyton and J. E. Hall, *Textbook of Medical Physiology*, ED-10, Saunders, 2000.

[2] R. Plonsey and R. C. Barr, *Bioelectricity: A Quantitative Approach,* Springer Science & Business Media, 2007.

[3] L. N. Katz, *Electrocardiography,* Lea & Febiger, 1946.

[4] J. N. Langley and H. K. Anderson, "On the Innervation of the pelvic and adjoining viscera: Part I. The lower portion of the intestine," *The Journal of Physiology*, vol. 18, no. 1-2, pp. 67–105, 1895.

[5] M. R. Bennett, "Structure and electrical properties of the autonomic neuromuscular junction," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, pp. 25-34, 1973.

[6] R. Manchanda, "Membrane current and potential change during neurotransmission in smooth muscle," *Current Science*, vol. 69, no. 2, pp. 140–150, 1895.

[7] M. R. Bennett, *Autonomic Neuromuscular Transmission*, CUP Archive, 1972, vol. 30.

[8] M. Koyanagi and M. Miyoshi, "Three-dimensional arrangement of muscle bundles in the outer layers of rodentia vasa deferentia," *Tissue and Cell*, vol. 32, no. 5, pp. 351-357, 2000.

[9] S. Appukuttan, K. L. Brain, and R. Manchanda, "A computational model of urinary bladder smooth muscle syncytium," *Journal of Computational Neuroscience*, vol. 38, no. 1, pp. 167-187, 2015.

[10] M. Hayase, H. Hashitani, K. Kohri, and H. Suzuki, "Role of K$^+$ channels in regulating spontaneous activity in detrusor smooth muscle in situ in the mouse bladder," *The Journal of Urology*, vol. 181, no. 5, pp. 2355-2365, 2009.

[11] Y. Abe and T. Tomita, "Cable properties of smooth muscle," *The Journal of Physiology*, vol. 196, no. 1, pp. 87-100, 1968.

[12] M. Padmakumar, K. Bhuvaneshwari, and R. Manchanda, "Classification and analysis of electrical signals in urinary bladder smooth muscle using a modified vector quantization technique," in *Proc. 2012 International Conference on Signal Processing and Communications*, 2012.

[13] Y. Jung, H. Park, D. Du, and B. L. Drake, "A decision criterion for the optimal number of clusters in hierarchical clustering," *Journal of Global Optimization*, vol. 25, no. 1, pp. 91–111, 2003.

[14] V. Estivill-Castro, "Why so many clustering algorithms: a position paper," *ACM SIGKDD Explorations Newsletter,* vol. 4, no. 1, pp. 65-75, 2002.

[15] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, "Unsupervised spike detection and sorting with wavelets and super paramagnetic clustering," *Neural Computation,* vol. 16, no. 8, pp. 1661-1687, 2004.

[16] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, vol. 344, 2009.

[17] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, pp. 224–227, 1979.

[18] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," pp. 32-57, 1973.

[19] D. A. Adamos, E. K. Kosmidis, and G. Theophilidis, "Performance evaluation of PCA-based spike sorting algorithms," *Computer Methods and Programs in Biomedicine*, vol. 91, no. 3, pp. 232-244, 2008

[20] E. Meng, J. S. Young, and A. F. Brading, "Spontaneous activity of mouse detrusor smooth muscle and the effects of the urothelium," *Neurourology and Urodynamics*, vol. 27, no. 1, pp. 79–87, 2008.

[21] Y. Sun and J. N. Daigle, "A pca-based vehicle classification system in wireless sensor networks," in *Proc. Wireless Communications and Networking Conference*, 2006, vol. 4, pp. 2193–2198.

[22] S. Ray and R. H. Turi, "Determination of number of clusters in *k*-means clustering and application in colour image segmentation," in *Proc. 4th International Conference on Advances in Pattern Recognition and Digital Techniques*, India, 1999, pp. 137–143.

Mithun Padmakumar has completed his master in biomedical engineering from IIT Bombay in 2010 and bachelor in electrical and electronic engineering from College of Engineering Trivandrum, Kerala, India, in 2008.

He is currently pursuing his Ph.D. degree in the Department of Biosciences and Bioengineering, Indian Institute of Technology (IIT) Bombay. He has attended the IBRO school on computational neuroscience in Hyderabad, and worked as a research associate in the Signal Processing and Instrumentation Lab, IIT Bombay where he developed algorithms for suppressing the motion artifacts and EMG noise from the ambulatory ECG recordings. His area of interest includes biomedical signal processing and pattern recognition.

Binil Jacob has completed his master in biomedical engineering from IIT Bombay in 2015 and bachelor in electrical and electronic engineering from College of Engineering Kidangoor, Kerala in 2012.

He is currently the co-founder and chief technology officer of a medical device startup called Medprime Technologies Pvt Ltd, Mumbai. He has a strong belief in the open source culture and takes keen interest in areas of embedded systems and signal processing.

R. Venkatakrishnan has completed his master in biomedical engineering from IIT Bombay in 2013 and bachelor in industrial biotechnology from sASTRA University Thanjavur, TamilNadu, India, in 2011.

He is a teacher for India alumnus (2013-2015) and is currently a software programmer specializing in web technologies at GQuotient Pvt. Ltd., Chennai, Tamil Nadu.

Keith L. Brain graduated with a Ph.D. in physiology in 1998 and received MB/BS (Hons.) in 1999 from the University of Sydney, Australia. There, in collaboration with Max Bennett at the University of Sydney, he developed a confocal microscopy technique for measuring calcium changes in autonomic nerve terminals.

He worked as a medical officer at the Royal Prince Alfred Hospital in Sydney, while he was awarded a Nuffield Medical Fellowship and joined the Department of Pharmacology, Oxford, in 2001. He is now working as a senior lecturer at the University of Birmingham. He is also a lecturer in Pharmacology at Christ Church College, Oxford.

Apart from the Nuffield Medical Fellowship, he was awarded a Welcome Trust Career Development Fellowship (2005), and University Research Lectureship (2007). He is the International Secretary of the International Society for Autonomic Neuroscience.

Rohit Manchanda received the B.A. degree (with honors) and the M.A. degree in physiological sciences from the University of Oxford, Oxford, U.K. He received his Ph.D. degree from Oxford University in 1989, working on electrophysiology of autonomic neurotransmission in the Department of Pharmacology.

He served as a pool officer in the Council of Scientific and Industrial Research (CSIR), All India Institute of Medical Sciences (AIIMS), New Delhi. Presently he is a Professor and heads the Department of Biosciences and Bioengineering at the Indian Institute of Technology, Mumbai, India. His research interests include electrophysiolology and biophysics of neurotransmission, electrical signaling and information processing in excitable cells and tissues, and in the application of signal processing techniques to synaptic potentials.

He has been awarded Career Award for Young Teachers, All India Council for Technical Education, (AICTE), 1997-2000, BOYSCAST Fellowship, Department of Science & Technology, 1998-1999, University of Oxford, U. K., Award for Excellence in Teaching, IIT Bombay, 2006 and 2010, Annual Felicitation Award, Society for Cancer Research & Communication, 2007, and Tibor Jones South Asia Prize, 2012.