Expert Systems With Applications 126 (2019) 187–199

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Closed-loop deep brain stimulation based on a stream-clustering system

C. Camara[a,b,d,*], K. Warwick[c], R. Bruña[b], T. Aziz[d], E. Pereda[b,e]

[a] Department of Computer Science, Carlos III University of Madrid, Madrid, Spain
[b] Centre for Biomedical Technology, Technical University of Madrid, Madrid, Spain
[c] Vice Chancellors Office, Coventry University, Coventry, United Kingdom
[d] Oxford Functional Neurosurgery, University of Oxford, Oxford, United Kingdom
[e] Department of Industrial Engineering & IUNE, Universidad de La Laguna, Tenerife, Spain

ARTICLE INFO

Article history:
Received 21 October 2018
Revised 14 February 2019
Accepted 17 February 2019
Available online 18 February 2019

Keywords:
Clustering
Data stream mining (DSM)
Expert system
Deep brain stimulation (DBS)
Parkinson's disease (PD)
Neural engineering

ABSTRACT

Idiopathic Parkinsons disease (PD) is currently the second most important neurodegenerative disease in incidence. Deep brain stimulation (DBS) constitutes a successful and necessary therapy; however, the continuous stimulation it provides can be associated with multiple side effects. DBS uses an implanted pulse generator that delivers, through a set of electrodes, electrical stimulation to the target area, normally the Sub Thalamic Nucleus. Recently, Closed-loop DBS has emerged as a promising new strategy, where the device stimulates only when necessary, thereby reducing any adverse effects. Here, we present a Closed-loop DBS system for PD, which is able to recognize, with 100% accuracy, when the patient is going to enter into the tremor phase, thus allowing the device to stimulate only in such cases. The expert system has been designed and implemented within the data stream mining paradigm, suitable for our scenario since it can cope with continuous data of a theoretical infinite length and with a certain variability, which uses the synchronization among the neural population within the Sub Thalamic Nucleus as the continuous data stream input to the system.

1. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder. It is expected to grow to pandemic proportions by 2040, surpassing Alzheimer's disease (Dorsey & Bloem, 2018). Nonetheless, it remains an idiopathic disease in about 95% of cases (Farrer, 2006), although it is known that it is caused by the degeneration of dopaminergic neurons of the Substantia Nigra Compacta (SNC).

The loss of neurons in this brain area produces an imbalance between the direct and indirect pathways, with the resultant prevalence of the indirect pathway. This disequilibrium is responsible for the symptoms of the disease, which include tremor of the limbs at rest (the so-called resting tremor (RT)), muscle rigidity, inability to initiate precise movements (akinesia) and slow motion, especially in complex voluntary movements (bradykinesia). PD is a disease that encompasses different subtypes. In this study we work with patients suffering from *benign tremulous parkinsonism*. The symptomatology present in this patients includes: 1) A prominent RT, being this the main symptom; 2) symptoms not related to tremor remain mild; 3) majority absence of gait disorder; 4) absence of disability apart from tremor (Josephs, Matsumoto, & Ahlskog, 2006). RT makes the patient transit between different movement states: non-tremorous resting state (*NT*), in which the patient does not experience symptomatology; the tremor state (*T*) itself, in which the patient experiences RT, and a third state called Tremor Onset (*TO*), which is supposed to hold the key to understanding the transition between *NT* and *T*.

The time-stamp indicating the beginning of the tremor episode was determined by a movement disorder specialist, and tremor-onset was detected through the EMG signal as described in 3.2.2. Data labels have not been used in the proposed system, but they have served for the previous verification of the existence of such tremor states in the used data.

To alleviate these symptoms, the first option is usually a pharmacological treatment, with or without dopaminergic effects. However, these patients have a tremor with high resistance to medication, even at the highest tolerable doses of levodopa (Josephs et al., 2006). In addition, the use of levodopa leads to

* Corresponding author at: Centre for Biomedical Technology, Technical University of Madrid, Madrid, Spain.
E-mail addresses: carmen.camara@ctb.upm.es (C. Camara), k.warwick@reading.ac.uk (K. Warwick), ricardo.bruna@ctb.upm.es (R. Bruña), tipu.aziz@nds.ox.ac.uk (T. Aziz), eperdepa@ull.edu.es (E. Pereda).

https://doi.org/10.1016/j.eswa.2019.02.024
0957-4174/

dyskinesias (LID), in which the patient suffers from involuntary movements that may ultimately be worse than the original PDs symptomatology.

A second line of treatment in such cases, is Deep Brain Stimulation (DBS). Treatment with DBS can be administered in conjunction with levodopa if necessary, for the rest of the symptoms, if present, including bradykinesia and rigidity. DBS consists of the surgical implantation of a neuro-stimulator, an implantable medical device (IMD) that uses an implanted pulse generator (IPG) to deliver electrical current through a set of electrodes to the surgical target, normally the SubThalamic Nucleus (STN), restoring its normal functioning. In the context of cardiac illnesses, pacemakers have the ability to adapt the stimulation to perform event-response in real time. However, presently, neurostimulators, once implanted, provide a continuous stimulation, which may induce adverse effects such as paresthesia, psychiatric or cognitive malfunction and even an increased risk of suicide (Benabid, Chabardes, Mitrofanis, & Pollak, 2009; Sugiyama, 2015). Moreover, current DBS treatment requires that the battery must be changed on average every 3–5 years (Medtronic, 2018), although in practice it occurs more frequently than this to prevent the deterioration of the treatment. Real-time adaptive (closed-loop) DBS systems represent a better strategy, in which the IMD stimulates only when necessary, on demand (Camara et al., 2015a; 2015c; Little et al., 2015; Priori, Foffani, Rossi, & Marceglia, 2013) thereby reducing the adverse effects. Yet the implementation of such a strategy requires the knowledge of what features of the STN activity change when (or ideally, shortly before) the symptoms appear.

Disentangling the behaviour of the STN in the aforementioned states (*NT, TO* and *T*) is the key to understanding the nature of the tremor. Furthermore, studying the temporal dynamics of such activity could be useful as well in designing better DBS treatments, even tailoring them to the individual concerned, as a further step towards personalised medical care. However, this task is exceptionally difficult, since it requires recordings that capture the neural activity of patients while transitioning from *NT* to *T*, in a natural and spontaneous way, in the absence of medication. Furthermore, local field potentials recorded from the STN (STN-LFP) in humans are generally acquired in the peri-operative period, after the surgical implantation of the electrodes, but prior to the final internalisation of the neurostimulator, and are not usually accessible afterwards. These two restrictions make this kind of recordings exceptional. In this work, we analyse four of them from our dataset, which fulfil these properties. They allow us to characterise the dynamical transitions, by identifying the different states. Especially interesting is the fact of being able to detect the TO state, and distinguish it from T, as it is the first step in preventing or suppressing the tremor by means of closed-loop DBS, before the tremor actually starts up for the patient.

2. Neural oscillations versus synchronization

LFPs provide one of the best sources of information with regard to brain activity, since they enable us to observe the original signal generated within deep structures without the need to apply a mathematical algorithm of source reconstruction.

One of the interesting features we can explore using the LFP is the neural oscillations, whose deviation from the typical patterns of healthy brain activity is often an indication of a pathological condition. Such deviations may not be restricted to the modification of, e.g., the spectral power of the individual LFPs in the frequency domain, but may include alterations in the communication between neural populations, that is, their functional and effective connectivity (Friston, 1994). Indeed, the relation between local synchronization, as reflected by the spectral power at certain frequency bands, and connectivity among neurons remains an open question. Some dopaminergic studies have suggested that the pathological performance of the STN, in which neurons oscillate synchronously with high amplitude, involves also an increase in spatial synchronization. When neurons trigger at the same time, there is no temporal delay, which is necessary for the communication between neurons, that is, functional and/or effective connectivity (Hohlefeld et al., 2013). Thus, in these studies it has been observed that during ON medication periods, the pathological spatial synchronization disappears and connectivity between STN neurons is increased.

2.1. LFP-STN neural oscillations

There are many studies on local synchronization. It is well known that when neurons oscillate synchronously in the basal ganglia, mainly in the subthalamus, it entails dysfunctional motor states in PD patients. This has been observed at the single cell (Lourens et al., 2013) and at the LFP level (Schnitzler & Gross, 2005; Weinberger, Hutchison, & Dostrovsky, 2009). Most of these studies found excessive neuronal synchronization in the beta frequency band (between 12 and 30 Hz), which is linked to bradikinesia and rigidity (Brown & Williams, 2005; Weinberger et al., 2009). Additionally, this beta synchronization is linearly related to the degree of levodopa administered (Kühn, Kupsch, Schneider, & Brown, 2006) and to the treatment of PD using DBS (Brown et al., 2004; Meissner et al., 2005; Wingeier et al., 2006).

However, the influence that beta band synchronization has on the occurrence of the resting tremor is still an open question. Indeed, some authors claim that there is no relationship between these events (Weinberger et al., 2006), while others maintain the contrary (Levy, Hutchison, Lozano, & Dostrovsky, 2000). Additionally, it has been observed that there is no causal link between the dopamine deficit in the striatum and the severity of the tremor (Deuschl et al., 2000).

There is also controversy as to whether this beta band synchronization is generated within the STN or if it merely reflects the overall synchronization in the basal ganglia circuit (Bevan, Atherton, & Baufreton, 2006; Schnitzler & Gross, 2005). In this regard there are in-depth studies that conclude that these oscillatory patterns, at least in the beta band, are generated within the STN (Kühn et al., 2005).

2.2. LFP-STN synchronization

There are not many studies on functional connectivity in STN-LFP, whose analysis we believe is essential not only to understand the functioning of the basal ganglia itself but also to improve the treatments of some extended neurological diseases, such as PD (Benabid et al., 2009). It has been observed that, as in the case of neural oscillations, connectivity in the beta band is modulated by levodopa (Hohlefeld et al., 2013). Yet to the best of our knowledge, there is no published research that studies synchronization within the STN during both tremorous and atremorous states.

Besides, a drawback of many studies is that they consider only the beta band. Yet in (Marceglia et al., 2006; Priori et al., 2004), the authors found different behaviours at the lower (<20Hz) and upper (>20Hz) beta bands. This result suggests that the oscillations take place only in the lower beta band, and what is observed in the upper beta is just a “contamination” of these activities. In order to be as thorough as possible, we estimated the synchronization levels during tremor and atremorous episodes, not only in the lower ([12–20] Hz) and upper ([20–30] Hz) beta bands, but also in the tremor ([3–7] Hz), alfa ([8–12] Hz) and gamma ([30–45] Hz) bands.

Given the heterogeneity across subjects, and to be certain that the results are valid, we performed the analysis per subject. This

allows us not only to deal with inter-subject heterogeneity, but also to observe, in a subject-specific way, what happens to the synchronization level in the STN before, during, and after the appearance of the tremor.

The purpose of this paper is twofold. Firstly, we want to characterise synchronization in the STN, and study which of them are more informative differentiating the distinct states. To this end we apply both functional and effective connectivity methods. Moreover, and given that it remains an open question, we want to know which frequency bands show changes in the synchronization with the appearance of the tremor. Secondly, we face the design of a data stream mining system which is able to identify the dynamical states in which the patient transits, to recognise and adapt itself to concept drifts, to perform in real time and to do it whilst employing a limited amount of resources (memory).

3. Signal preprocessing

3.1. Patients dataset

The dataset used in this study consists of four recordings from parkinsonian patients diagnosed with tremor-dominant PD, who underwent surgery for the implantation of a neurostimulator (DBS treatment) at the John Radcliffe Hospital in Oxford, UK. The successful implantation of the DBS electrode lead within the STN was verified with postoperative magnetic resonance. The local research ethics committee of the Oxfordshire Health Authority approved the recordings and informed consent was obtained from the patients involved.

Neurophysiological data was acquired by employing a Medtronic DBS Lead Model 3387, which contains four electrodes spaced 1.5 mm apart. This allows for three different contact pair (electrode) configurations ($0+1, 1+2,$ and $2+3$) to be simultaneously recorded in a bipolar fashion with one contact used as reference.

Thus, each record consists of three channels of LFP, collected through the electrodes located in the basal ganglia, specifically the STN. LFPs capture the summarised electrical activity of the neuron population between each contact. Recordings were performed during a peri-operative observation period in which the depth electrodes were already implanted but were not stimulating. As a result, data was accessible for recording. The original dataset is composed of 33 files, but we used in this study only those four records capturing the neural activity of patients while transitioning from *NT* to *T*, in a natural and spontaneous way, in the absence of medication. This ratio (4 records out of 33) reflects the difficulty in obtaining recordings such as those used in this research.

3.2. Data preparation

3.2.1. Signal preprocessing

LFP data was two-pass filtered into tremor, alfa, lower-beta, upper-beta and gamma bands using band-pass filtering with a 500 order FIR filter designed with Hamming window, and using 2 seconds of real data as padding. The movement artefacts around 1Hz and line noise (in Europe, 50 Hz) were excluded after this filtering step.

Some of the connectivity methods work with the analytic signal. To this end the Hilbert transform of the filtered LFP data was performed, which can be expressed as $A(x) = x + iH(x)$, where x is the original signal, $H(x)$ is its Hilbert transform, and i is the imaginary unit.

Finally, the data was segmented in windows containing 10 cycles at the central frequency of the band, with an overlap of 50%. The level of overlapping is motivated for capture with high temporal resolution the rich dynamic behaviour that STN-LFP has.

The EMG signal in tremor is made up of bursts, whose peaks are at a frequency of 30 Hz and above. Consequently, EMG data were high-pass filtered above 30 Hz using a two-pass procedure with a 500 order FIR filter designed using a Hamming window. The EMG signal was then rectified using the Hilbert envelope. This rectified signal is filtered again a FIR 2–45 Hz filter. With this, the tremor in Hilbert's rectified signal is already at low frequencies.

3.2.2. Tremor state labelling

Each recording holds a single tremor-onset event. A clinical specialist on movement disorders marked the time in which the tremor starts. Tremor onset was calculated based on the amplitude of filtered and rectified EMG signal, following the same procedure used previously with this data (Bakstein et al., 2012). The magnitude of the EMG time series value was checked against a threshold of 3 times the mean of the EMG amplitude in the first 5s of the recording (which contain atremorous data). If a peak of high tremor frequency activity is detected at any point in time, the average of the following 5sg (time enough to cover any period of tremor-onset) of data is calculated to confirm the tremor-onset detection. Specifying a threshold would also detect peaks of tremor activity, however short magnitude spikes may trigger an incorrect detection.

A single time of tremor-onset was calculated for each recording. This mark, together with the one determined by the clinical specialist divide the recording into the 3 tremor states of which it is composed: NT(atremorous state), TO(tremor-onset state) and T(tremor state).

4. Connectivity

The analysis of synchronized activity is a relatively novel approach to employ when observing the functioning of the brain. There are several mathematical methods to estimate connectivity. For instance, functional connectivity (FC) methods quantify the statistical dependence between temporal series without giving any information about the directionality. On the other hand, effective connectivity (EC) methods rely on the assumption that, when studying brain connectivity, true interactions between two neural sources appear with a certain time delay, the time in which the information is travelling (Nolte et al., 2004). These are able to detect this directionality.

In this work, we applied two classical linear methods (Coherence and Amplitude Envelope Correlation), two information-theory based methods (Mutual Information and Phase Transfer Entropy), a phase synchronization method (Phase-Locking Value) and a high-order spectra method (Bicoherence). All of them are FC measures, except the phase transfer entropy, which belongs to EC methods. These measures have been applied after filtering the data to the bands of interest: Tremor, Alfa, Lower Beta, Upper Beta and Gamma bands separately. By applying this battery of measures, we intended to thoroughly characterise both linear and nonlinear connectivity within the LFP of the STN in order to have a detailed description of the pattern of FC/EC within the STN. Henceforth, we briefly describe these indexes.

4.1. Classical methods

4.1.1. Coherence

The coherence function estimates the linear correlation between two signals x and y as a function of the frequency. It is defined as the cross-spectrum $C_{xy}(f)$ normalised by the product of the individual power spectral densities (auto spectrum) of each signal,

$C_{xx}(f)$ and $C_{yy}(f)$ (Pereda, Quiroga, & Bhattacharya, 2005):

$$\Gamma_{xy}(f) = \frac{C_{xy}(f)}{\sqrt{C_{xx}(f)C_{yy}(f)}}; 0 \leqslant |\Gamma_{xy}(f)| \leqslant 1 \quad (1)$$

We used the Welch's averaged, modified periodogram method to estimate the auto and cross spectra, since we are dealing with finite amount of data.

4.1.2. Amplitude envelope correlation

Amplitude envelope correlation (AEC) is an index developed to detect signal coupling without phase coherence. It is able to detect synchronization in a less precise coupling of signals than coherence, since the envelope of band-limited signals does not change as rapidly as the signals themselves (Bruns, Eckhorn, Jokeit, & Ebner, 2000).

Given two signals x and y, AEC is the mean correlation of their envelopes. We calculate AEC as:

$$AEC = corr(abs(x_h), abs(y_h)) \quad (2)$$

where x_h and y_h are the Hilbert analytical signals of x and y respectively.

4.2. PS methods phase - locking value

Phase synchronization indexes (PS) are used to investigate if the phases of two oscillators are coupled, even if their amplitudes may not be. If two systems present PS, it means that the difference of their phase over time remains bounded.

The preferred index to observe this phenomenon in neuroscience is arguably the Phase-Locking Value (PLV), which measures the relative phase difference between x and y and estimates how it is distributed over the unit circle (Lachaux et al., 1999). It is defined as:

$$PLV_{x,y} = T^{-1}\left|\sum_{T} e^{i[\phi_x(t)-\phi_y(t)]}\right|; 0 \leqslant PLV_{x,y} \leqslant 1 \quad (3)$$

where ϕ_x and ϕ_y are the phases of x and y, extracted from the Hilbert analytical signal, respectively, wrapped to the interval [0, 2π). Thus, we work with the *cyclic relative phase*, avoiding the phase slips of 2π that may be present in the signal. We get a value of one when the phase difference across windows is the same, and a value of zero when it varies randomly.

4.3. Information theory based

4.3.1. Mutual information

Mutual Information (MI) is one of the most used indexes of interdependence based on information theory. MI draws from Shannon's concept of entropy (Shannon & Weaver, 1949), which can be regarded as the amount of information a variable holds. Therefore, it is also a measure of its uncertainty.

Given a random signal x, with a probability distribution $p(x) = P\{X = x\}$, $x \epsilon X$, its entropy is defined as:

$$H(X) = -\sum_{x\in X} p(x)\log p(x) \quad (4)$$

Joint Entropy is the extension of the concept of entropy to two variables. The joint entropy for a couple of discrete random X and Y signals is:

$$H(X,Y) = -\sum_{x\in X} -\sum_{y\in Y} p(x,y)\log p(x,y) \quad (5)$$

The joint entropy will always be less than or equal to the sum of the individual entropies: $H(X,Y) \leq H(X) + H(Y)$; equality being met only if they are independent variables.

Henceforth, for a pair of random variables, we can estimate the amount of information they share, which is the concept of MI:

$$MI = \sum_{x\in X}\sum_{y\in Y} p(x,y)log\frac{p(x,y)}{p(x)p(y)} \quad (6)$$

where $p(x, y)$ is the joint probability. If $MI = 0$, then the variables are independent. Otherwise (i.e., if $MI > 0$), there exists some degree of interdependence between both signals (Pereda et al., 2005).

4.3.2. Phase transfer entropy

Phase Transfer Entropy (PhTE) quantifies the directed interaction between two time series by applying the concept of transfer entropy to their phases (Lobier, Siebenhühner, Palva, & Palva, 2014). It has the advantage, when compared with other directed methods such as Granger or Dynamic Causal Modelling, to be model free (not making assumptions about the system) and works with only one parameter, thereby reducing the possibility of erroneous results due to the election of parameters. PhTE is defined as:

$$PhTE_{x\to y} = H\left(\phi_y(t), \phi_y(t')\right) + H\left(\phi_y(t'), \phi_y(t')\right) - H\left(\phi_y(t')\right) - H\left(\phi_y(t), \theta_y(t'), \phi_x(t')\right) \quad (7)$$

where $\phi_x(t')$ and $\phi_y(t')$ are the past states at time point $t = t - \delta$: $\phi_y(t') = \phi_y(t-\delta)$, $\phi_x(t') = \phi_x(t-\delta)$, and $H(\bullet)$ is the entropy and the joint entropy as defined previously in Eqs. (4) and (5).

4.4. High - order spectra methods - bicoherence

Power spectrum analysis is usually performed via Fourier Transform (FT), of the second order statistic of the signal. But, unfortunately, this measure loses information about the phase relationships between frequency components, a fact that has been linked with impaired functions in the brain (Wong, Clifton, & Tarassenko, 2009).

The bispectrum is a two dimensional version of the FT based on the third order cumulant of the signal. It is defined as:

$$B(f_1, f_2) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} R(m,n)e^{-j2\pi f_1 m}e^{-j2\pi f_2 n} \quad (8)$$

where $R(m, n)$ is the third order cumulant as a function of the lags m and n, and f_1 and f_2 are the frequencies in study.

The bispectrum can be used to investigate if the signals at f_1, f_2 and $f_1 + f_2$ are synchronized, which would mean that the oscillation at $f_1 + f_2$ is due to the nonlinear relationship between both signals. Bicoherence (BICOH) can be calculated as a normalised version of the bispectrum (Mendel, 1991):

$$BICOH(f_1, f_2) = \frac{B(f_1, f_2)}{\sqrt{P(f_1)P(f_2)P(f_1+f_2)}} \quad (9)$$

where $P(f)$ is the power spectrum at frequency f. For incoherent signals this measure tends to zero.

5. Performance evaluation of the synchronization indexes

The entity-relation diagram of Fig. 1 illustrates the dimensions of information in our dataset. We have information on what is happening within the STN between each pair of contacts, at each frequency band, and from the perspective of each connectivity index. This constitutes a lot of information, but certainly not all of it is useful in identifying the movement states. We therefore need to reduce the dimensionality of the data, ignoring in the subsequent steps the irrelevant attributes. This step is important for several reasons: Firstly, to improve the performance of our task, since subjoining irrelevant features is self-defeating for ML algorithms. Secondly to reduce the computational requirements of the

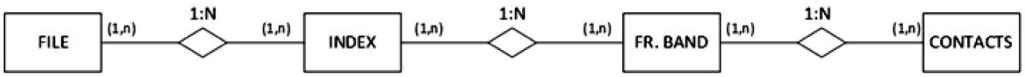

Fig. 1. E-R model. The E-R diagram represents the relevant entities of our dataset and their interrelationships.

Table 1
Features selection results using a wrapper method with a K-NN algorithm.

File	Selected bands (Accuracy)	Selected Indexes (Accuracy)
1	TB: 96.56%	PhTE: 93.78% BiCOH: 88.15% MI: 87.02%
	LB: 91.43%	BiCOH: 81.31% PhTE: 75.57% MI: 75.95%
2	TB: 98.35%	PhTE: 88.75% BiCOH: 86.89% MI: 82.10% PLV: 75.50%
	LB: 92.99%	MI: 80.09% BiCOH: 78.86% PhTE: 72.05% PLV: 70.95%
3	TB: 93.13%	BiCOH: 76.13% MI: 66.63% PhTE: 61.50%
	LB: 87.15%	BiCOH : 76.13% MI: 66.63% PhTE : 61.50%
4	TB: 93.15%	BiCOH: 79.42% MI: 64.39%
	LB: 85.87%	BiCOH: 67.99% MI: 57.31%

system, since fewer features need to be calculated by the IPG in real time. Finally, it allows us to improve the explicability of our results, thereby gaining insight into the FC mechanisms within the STN associated to the appearance of the tremor.

To this end, we could simply perform a manual selection of the attributes, by removing or adding them one by one and testing the results, but this is a nave and arduous process. Methods from ML are however very useful. In this work we used a wrapper method with a KNN classifier with $k = 5$ performing a ten-fold cross validation for the attribute selection problem. KNN is a lazy and non-parametric algorithm, which is highly sensitive to irrelevant attributes. Although wrapper methods are computationally more expensive than filter methods, they render better performance. Feature selection is calculated here in order to unveil which synchronization indexes and frequency bands are more informative. This process will not be carried out by the IMD, and it is not part of the proposed clustering system.

We applied the feature selection procedure in two steps. Firstly, we studied the preferred frequency bands, those that produce the highest accuracy. Subsequently we performed a second round of feature selection in these bands. We have summarised the selected bands and features in Table 1.

The preferred bands in all cases were the tremor (TB) and the lower-beta (LB) bands. The selected indexes per each band were the same in all the cases, except for the phase indexes, PLV and PhTE which are not always selected per all the recordings. However, the degree of accuracy varied across files, so we cannot conclude at this point that only one index is valid, but we would choose three for further analysis.

Before applying a feature selection procedure we had results from six synchronization measures, each of them calculated per each of the four evaluated frequency bands. After that, we reduced the dimensionality, choosing only two frequency bands (TB and LB) and only three indexes, PhTE, MI and BiCOH. Hereinafter we continue our experimentation only with these selected features; only these will be evaluated in our system.

6. Data stream mining

6.1. Data stream mining

Data mining refers to the set of technologies to handle larger datasets to find patterns, trends or rules and explain data behaviour (Witten, Frank, Hall, & Pal, 2016). In classical data mining techniques, a model is built during the so-called training phase, in order to make future predictions, recognising the class of a given sample not seen by the system before - the so-called testing phase. These technologies have consolidated due to the huge amount of data, which is collected and handled on a day to day basis. Indeed, this is a trend that continues growing at a fast pace in different areas, especially in the healthcare context (Miljkovic et al., 2016).

However, this approach does not always fit well in real-time analytic scenarios, in which it is important to analyze the data, extract the relevant features and take decisions in real time. In other words, while data mining can handle well considerable quantities of data, it does not consider the continuous supply of data. Models cannot be updated when new information arrives. This implies that they are not able to self-adapt in response to the observation of new samples, and the complete training process has to be repeated. Furthermore, the length of the data feed is much larger than the storage for instance, our scenario in which we have a neural signal monitored during the entire life of an individual.

Contrary to traditional data mining, data stream mining (DSM) involves a set of algorithms, emerged as a paradigm to address the continuous data problem and changes in the behaviour of the stream. A data stream can be considered a signal with time as an independent variable, which needs to be processed online, as each sample arrives. In this way, data samples of such signals will feed the model sequentially, training it in a continuous way. In such systems, we do not distinguish between the training and the testing phase, as both tasks are continuously updating the model online. This way of functioning is very interesting, since it allows the system to adapt itself to changes in the data, as we will discuss hereinafter.

There are, though, two main algorithmic challenges in this scenario. First, the data stream has infinite length and arrives fast, and therefore it is necessary to extract information from it in real time. Second, the data may be evolving, experiencing shifts (non-stationary data such as neurological signals) and models must adapt themselves when there are changes in the data (Bifet, Gavaldà, Holmes, & Pfahringer, 2018).

The core assumption is that training samples arrive very fast and should be processed and discarded to make room for new samples, thus being processed one time only. More specifically, DSM presents a set of different requirements (Bifet et al., 2018):

Uniqueness: Each sample must be processed before a new one arrives, and it has to be done only once, without the retrieval of any previous samples being possible. Since our source of data is the STN, our data stream will be endless. This fact makes this requirement critical for our purpose.

Limitation of resources: The hardware that will analyse the data, the IPG, as any other IMD, has restricted capabilities of

energy, storage, and computing power (Camara, Peris-Lopez, & Tapiador, 2015b), principally due to its implantation. This limitation of memory is one of the main motivations behind using data streams, as memory can be overloaded when too much data is stored in it.

Limitation of time: DSM algorithms should scale linearly with the number of samples to work in a limited amount of time. The algorithm should be able to process the data at the speed of the stream. In the proposed system, the algorithm has to test and train at the speed of the neurostimulator register and pre-process a time window, with the minimum plausible delay to detect the shift in the tremor state as soon it appears.

Evolving time series: DMS works with non-stochastic time series, which evolve over time. The changes in the data are called *concept drifts*, and our DSM algorithm has to able to detect them, adapting the output.

Readiness: The algorithm must be able to commence working after a learning period as small as possible.

Immediacy: An algorithm should be ready to produce the best model at any time regardless of the number of processed samples.

6.2. Stream clustering

The objective of our system is to detect the tremor through the data stream recorded from the electrodes. In this scenario, the labels of the new samples are unknown to the system; in fact, its purpose is to create them. We are therefore in an unsupervised learning scenario. The objective of stream clustering is to cluster the samples of the time series according to the speed at which the data stream is generated, updating the clustering with each newly arriving sample, and within the constraints presented in the previous section.

As with classical clustering techniques, the main goal of stream clustering is grouping the instances into clusters according to their commonalities, so that instances within each cluster are similar to each other while instances from different clusters are distinct. Obviously, the success in this task is closely related to the quality of our features, in our case the value of the synchronization indexes. This is one of the reasons why a feature selection was applied prior to the analysis with DSM.

Here we use an ad-hoc algorithm for DSM, *ClusterGenerator*, since it ensures an evaluation unbiased by possible incorrect outcomes of stream clustering algorithms (Kremer et al., 2011). We briefly describe it now.

The algorithm forms and continuously maintains microclusters, a technique used in other stream-clustering approaches. A microcluster is a compact representation of the data distribution, from which we can derive the mean and the standard deviation. It is represented as a cluster feature tuple $CF = (n, LS, SS)$, where the data stored in it are the number of points within the microcluster n, their linear sum LS, and their squared sum SS. Microclusters are representatives for sets of similar data points.

Stream clustering algorithms usually alternate between two phases, commonly called online and offline phases. In the online phase, microclusters are created. Working with microclusters is a key point, since in this way it is not necessary to access the past samples of the stream, i.e each sample will be accessed only once, which satisfies the uniqueness condition. During the offline phase time is not critical, and is typically used for the user to make an analysis of the existing clusters at each temporal moment, stored in the so-called *snapshots*. This phase typically requires more than one pass to the stored data (to the microclusters) and is a very useful phase in big data analytics environments. Cluster Generator adopts only the online phase, which assures us a lower computation requirement, which is a very important property in our domain. In addition, the offline phase usually reduces the number of microclusters. However, in the use scenario of the presented system, the number of microclusters is small and determined by the clinical problem, so that offline clustering is omitted.

Cluster generator is a partitioning-based clustering algorithm. The cluster boundaries are determined as described in (Gärtner, 1999; Welzl, 1991), that suggest creating spherical clusters of the smallest possible radius that contains all the instances of the cluster. The pseudocode of this algorithm is presented in Algorithm 1 and described herein below. For further details, the source code of the used algorithm is publicly available at A.Bifet (2018).

Algorithm 1 Smallest enclosing disks.

```
minidisk(P,S)
    D := MD(∅,S)
    if |S| = d+1
        return D
    end
    while i≤ n
        if P_i ∉ D
            D := minidisk (P −{p_i}, S ∪ {p_i})
        end
    end
    return D
end
```

Given a set of points $P = \{p_1, \ldots, p_n\} \subseteq \mathbb{R}^d$, let D($P$) denote the closest disk of smallest radius that contains P. D(P) exists and it is unique. For $P, S \subseteq \mathbb{R}^d$, $P \cap S = \emptyset$, let MD(P, S) be the smallest disk containing P, with all the points in S on its boundary. So we have MD(P, $\emptyset$)=D(P), and MD($\emptyset$, S) to be the smallest disk containing all the points of S on the boundary. $D(P)$ is determined by at least tree points on its boundary. i.e there is a subset S of P on the boundary of $D(P)$ such that $S \leq 3$ and $D(P) = D(S)$; so if $D(P - \{p\}) \neq D(P)$, then $p \in S$, and p lies on the boundary of $D(P)$ (Welzl, 1991).

minidisk is a recursive algorithm that computes $D(P, S)$ incrementally by adding the points in P successively while maintaining the smallest enclosing disk. When the algorithm calls minidisk(P,∅), all sets S that resulting from recursive calls are affinely independent (Gärtner, 1999)

Summarizing, the reasons why we opted for DSM are as follows:

- Neural activity is a source of non-linear, non-stationary data. Thus, it calls for a system able to adapt itself, learning in a continuous way.
- IMDs are devices with restricted capabilities of energy, storage, and computing power. Any proposed measure to operate imbibed in them should be as efficient as possible. DSM methods are thought to obtain the maximum accuracy with minimum time and memory use (Bifet et al., 2018).
- It has been applied to other problems dealing with sensor data and the Internet of Things, due to its capacity to monitor processes and improve their quality (Bifet et al., 2018; Gaber, Zaslavsky, & Krishnaswamy, 2005)
- The main motivation for choosing a stream-clustering technique, instead of a classification one, is determined by the context of application of the system: a clinical environment during a neurosurgery procedure. In this scenario, the best of systems that could be designed, is one that would be able to recognise the different movement states autonomously with a high level of accuracy. This is due to the short time available in the clinical environment for a possible training of the system. Note that the system must be prepared for full operation during the perioperative period. For this reason, the main objective is to design a system capable of separating the samples that arrive in real time, without the need of prior training or data labelling.

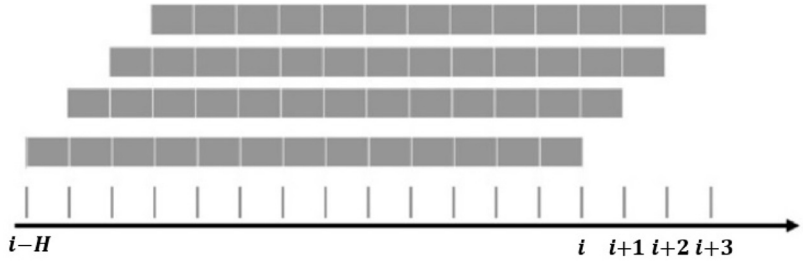

Fig. 2. Example of sliding window strategy.

With a stream-clustering system, we only need to expose the patient to the different movement states, so that the clusters are open. This process is faster and more efficient than collecting long periods of subthalamic signal, then labelling them and training a classification system. This fact is key to the usability of the system in a real clinical environment.

7. System description

One of the relevant aspects of the system is how data is treated. In a real-time application, in which neural samples arrive continuously to the system in a non-predefined order, an on-line analysis was used to evaluate LFP streams following a sliding window strategy, in which the size of the window was fixed and the buffer kept the newest instances. Similar to the first-in, first-out data structures (Gama, 2010), whenever a new instance was inserted into the window, another instance $i - H$ was forgotten, where H represents the window size (see Fig. 2).

In clinical application, the system starts up during the peri-operative period. The IPG starts registering the STN-LFP signals, which are divided into 10-cycles windows. After that, the synchronization index is calculated over each LFP window and per each contact pair, obtaining a set of coefficients:

$$S = I(LFP) = \{I(LFP)_{w=1}, I(LFP)_{w=2}, \cdots, (LFP)_{w=n}\} \tag{10}$$

where $I(LFP)_{w=i}$ is:

$$I(LFP)_{w=i} = S_{w(i)} = I_{w=i}^{c_1-c_2}, I_{w=i}^{c_1-c_3}, I_{w=i}^{c_2-c_3} \tag{11}$$

where w is the number of the observed window, c_x and c_y are the channel pairs between which the synchronization is calculated, I is the synchronization measure employed and S represent the result of the synchronization per each window, i.e our datastream. Finally, the LFP stream S is sent to the clustering algorithm.

According to Algorithm 2, the system works in the following way. Firstly, during the **set-up phase** the necessary clusters are initialised inside the system. For that, the system observes the activity of the STN for some time, during which it is necessary for the patient to experience the different movement states. During this time the signals on each electrode are collected, pre-processed and the synchronization measure is extracted. The output of the synchronization index determines autonomously which cluster the sample will go to. The process finishes when three clusters (NT, TO and T) are created, one per movement condition.

Once the clusters are initialised, the system enters into the **operation phase**, whose core is very similar to the set-up phase. The main difference is that in this phase the clusters are already created. The system should not open new clusters, but merely evaluate the arriving samples, decide the cluster each sample should go to and update the microcluster of the elected cluster after the instance is included. We will later discuss the evaluation of this process. The operation phase of the system is depicted in Fig. 3.

8. Experimental analysis

8.1. Clustering-streams results

8.1.1. Evaluation measures

Before presenting the results, we describe the evaluation measures. We employed external measures, since the ground truth is available against which to compare the clustering result.

Algorithm 2 Set-up and operation phases.

```
procedure Set-Up Phase (during peri-operative period)
    while num (tremor states seen) < 3
        capture STN records
        pre-process dataraw
        get synchronization index
        if sample belongs to existing cluster
            add sample to cluster
            update microcluster
        else
            open new cluster
        end
    end
end

procedure Operation Phase (after the neuroestimulator is
 internalized)
    capture STN records
    pre-process dataraw
    get synchronization index
    add sample to cluster
    update microcluster
end
```

Accuracy (ACC): Fractions of instances assigned to their correct cluster.

$$acc = \frac{\sum_{i=1}^{N_w} TP}{Nw} \tag{12}$$

where TP means a true positive (an instance correctly identified), and N_w is the total number of windows in the record.

Cluster Mapping Measure (CMM): CMM is an ad-hoc measure for stream clustering. Similar to accuracy, it quantifies how a given clustering is different from the ground truth, but in this case taking into account the type of errors that can occur during stream clustering, called faults. These faults can be derived from missed, misplaced or noisy points. The estimated CMM ranges between 0 and 1. The lower the faults, so the closer CMM is to 1 (Kremer et al., 2011).

8.1.2. Experimental analysis

In order to follow this section, we first introduce the concepts of weight and horizon (Kremer et al., 2011):

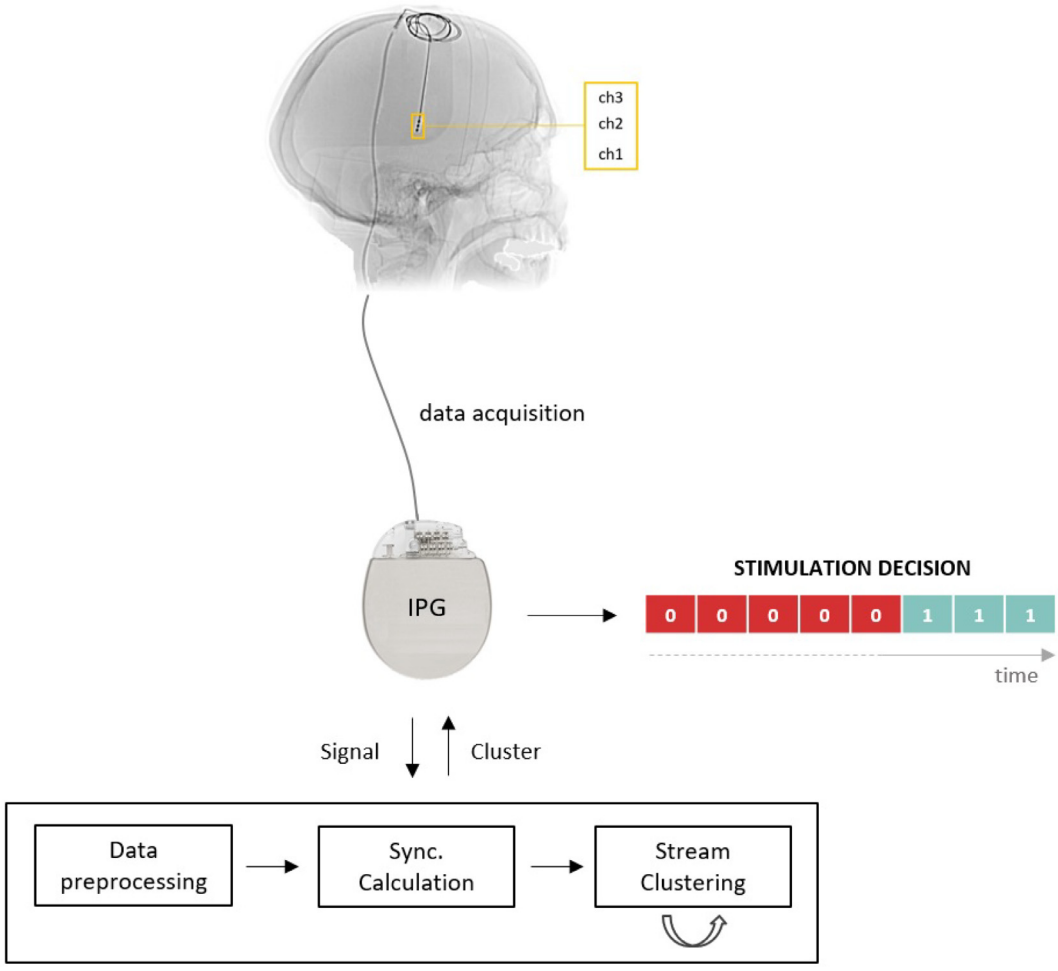

Fig. 3. Stream clustering - Closed-loop DBS system model.

Weight. Consider t_i (the time in which the sample $S_{w(i)}$ arrived to the system) and t_0 (the present time), with $t_i < t_0$. The weight of $S_{w(i)}$ is the decay function: $weight(S_{w(i)}) = 2^{-\lambda(t_0 - t_i)}$, where λ is a parameter that controls the ageing of the function. In this case, $1/\lambda$ is the half-life of $S_{w(i)}$. In our case, λ is set to zero, so we do not consider decay in our stream.

Horizon. Since this paradigm works with real-time and infinite data, stream clustering techniques have to forget past samples. To this end, only a subset of the recent samples of the stream S is considered at a given time. The horizon (H) for a stream S and a defined threshold value of ξ, is defined as $H = \{S_{w(i)} \epsilon S \mid W(S_{w(i)}) \xi\}$.

As stated previously, the records capture the neural activity of patients while transitioning from NT to T. Thus, the sequence order the recordings is $NT \rightarrow TO \rightarrow T$, as shown in Fig. 4.a. If we evaluate the system over these records, we can test if the system can adapt to concept drifts, detecting the three states, and thus opening the three clusters. However, this would be similar to training the system, without testing it later (Although in stream clustering we do not talk of *training* and *testing*, we will refer to it this way for the sake of clarity), since with this experimentation it is not possible to test immediacy and readiness, essential features to indicate if the system will work properly in real-time and real environment states. We need to *test* the system once the clusters are opened. To this end, as stated in Algorithm 2, we firstly expose the system to the tremor states (set-up phase), to later *test* states (operation-phase) to see if it is able to identify correctly the arriving instances, by adding them to the correct cluster.

To this end, the files were reordered, positioning a subsection of windows of each state at the beginning of the file, as shown in Fig. 4b and 4c. The number of selected windows is the 50% of the smallest section in the file. We used this subsection as a *training* period, in which the system was exposed to all the tremor states. Accordingly, a cluster of each type is opened. The horizon in this case was fixed to the number of windows within this subsection ($3x$ in the example of the figure).

It is important to remark here that this period is just to simulate the functioning of the system in a real environment. It is not a typical training phase itself, since it is the system itself that opens the clusters based on the instances it receives, the label information is not provided. The sequence order of the files to carry out the experimental analysis is therefore $NT \rightarrow TO \rightarrow T \rightarrow NT \rightarrow TO \rightarrow T$.

At this point, the issue of forgetfulness becomes important. To assess the influence of whether maintaining the opened clusters or not, we designed two possible configurations for the operation phase: OP-1, corresponding with Fig. 4c, in which we provoked the system to forget some of the clusters by fixing the parameter H low enough; and OP-2, corresponding with Fig. 4b, where the opened clusters are not forgotten. The results for both configurations are presented in Table 2.

In the OP-1, the system reaches the so-called point of forgetfulness. An example is depicted in Fig. 5, in which the cluster T has been forgotten. Under this situation, if the arriving points are not strongly separable from the existing ones, they wouldn't be assigned to their proper class. We have tested this, and have found

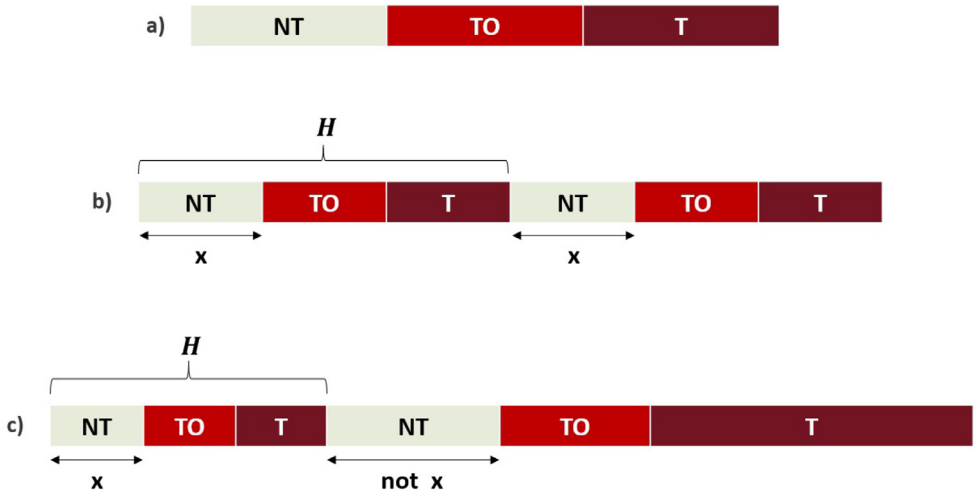

Fig. 4. a) Original order of the states in the recordings. b) States re-ordered to test OP-2: in which initialised clusters during peri-operative period will be maintain in the system memory. We simulate that by making the section of the operation phase as long as sections of set-up phase c) States re-ordered to test OP-1: in which initialised clusters during peri-operative period are not maintained in the system memory. We simulate that by making the sections of the operation phase longer than sections of set-up phase.

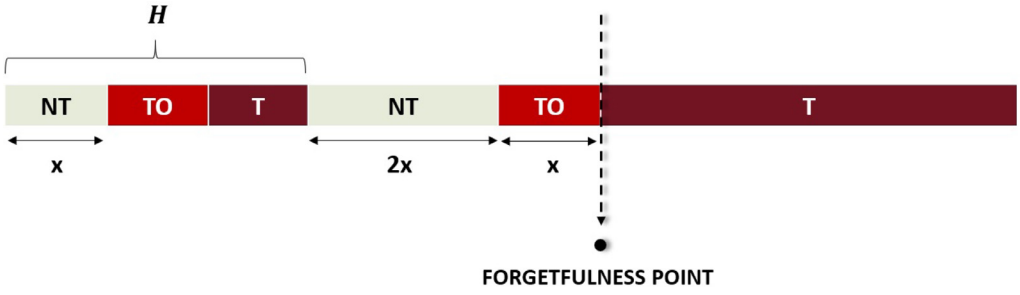

Fig. 5. Forgetfulness Point. By fixing *H* short enough we simulate the case in which at some point a cluster is forgotten by system. In the so-called *forgetfulness point*, the system forgets *T* cluster.

Table 2
Results of stream clustering for configurations OP-1 in which clusters are not maintained in the system and OP-2 in which the system takes advantage of the peri-operative period initializing the clusters in the system.

File	Band	Feature	OP-1		OP-2	
			ACC	CMM	ACC	CMM
1	**TB**	PhTE	66.89	0.6729	100	0.7614
		MI	66.89	0.6852	100	0.7837
		BICOH	66.89	0.66	100	0.7730
	LB	PhTE	73.87	0.7977	100	0.8318
		MI	73.87	0.8815	100	0.9298
		BICOH	73.87	0.852	100	0.9183
2	**TB**	PhTE	73.51	0.7284	100	0.7961
		MI	73.51	0.7089	100	0.7589
		BICOH	73.51	0.7049	100	0.7653
	LB	PhTE	51.73	0.8704	100	0.93
		MI	51.73	0.8595	100	0.8907
		BICOH	51.73	0.8447	100	0.8628
3	**TB**	PhTE	50.81	0.6063	100	0.9201
		MI	50.81	0.6250	100	0.9071
		BICOH	50.81	0.6144	100	0.9665
	LB	PhTE	51.73	0.6914	100	0.9880
		MI	51.73	0.6670	100	0.9835
		BICOH	51.73	0.6660	100	0.9812
4	**TB**	PhTE	84.47	0.8228	100	0.9144
		MI	84.47	0.8228	100	0.9483
		BICOH	84.47	0.8899	100	0.9264
	LB	PhTE	77.90	0.5972	100	0.9741
		MI	77.90	0.5641	100	0.9870
		BICOH	77.90	0.9768	100	0.9810

that this error never happened when comparing a sample of *TO* or *T* with a sample of *NT*. However, we have found that it is possible for the system to get confused between *TO* and *T* instances if one of these clusters is forgotten.

Actually, this would not suppose a problem in the particular case of our system, since both clusters (*TO-cluster* and *T-cluster*) will produce the same output, since the decision as to whether to apply stimulating pulses or not in both cases would be positive. In fact, we could perfectly have merged these two states. However, we decided to maintain them separately to test the power of the synchronization measures and our system in distinguishing each individual state.

One might think that increasing the value of the horizon could fix the problem, however: 1) it would imply more memory use, a fact we want to avoid; and 2) using a large value of *H*, the cluster tails become longer, increasing the probability of overlapping clustering (Kremer et al., 2011).

The solution could be to open the clusters and maintain them in memory, possibly as a background task. Remember that we would not maintain all the information and points of the clusters, just their *CFs*. To this end, we designed experiment 2, corresponding to Fig. 4b, in which we wanted to simulate the behaviour of the system when opened clusters are not forgotten. As can be observed in the results for configuration 2 in Table 2, the accuracy improves significantly, reaching 100% in all the observed cases. This is because when maintaining the learned clusters in the system, the confusion between *TO* and *T* does not occur. Note that keeping the information about the opened clusters is not equivalent to

Table 3
Accuracy results comparison between stream and classical clustering.

File	Band	Feature Subset	Stream-Clustering	K-means++	Density-based (Canopy)
1	**TB**	PhTE	100	46.82	45.17
		MI	100	44.37	52.98
		BiCOH	100	30.46	34.43
		PhTE + MI + BiCOH	-	59.60	50.99
	LB	PhTE	100	58.62	51.54
		MI	100	57.89	58.07
		BiCOH	100	54.44	56.3
		PhTE + MI + BiCOH	-	48.27	60.25
2	**TB**	PhTE	100	45.69	46.35
		MI	100	54.96	53.64
		BiCOH	100	45.35	54.3
		PhTE + MI + BiCOH	-	50.34	55.63
	LB	PhTE	100	57.35	60.60
		MI	100	54.80	50.81
		BiCOH	100	59.53	62.79
		PhTE + MI + BiCOH	-	43.38	55.35
3	**TB**	PhTE	100	61.3	59.67
		MI	100	47.73	39.61
		BiCOH	100	55.78	54.15
		PhTE + MI + BiCOH	-	48.92	48.64
	LB	PhTE	100	63.43	53.83
		MI	100	62.54	45.06
		BiCOH	100	63.59	55.24
		PhTE + MI + BiCOH	-	36.85	52.5
4	**TB**	PhTE	100	52.01	50.31
		MI	100	54.03	51.55
		BiCOH	100	52.01	51.08
		PhTE + MI + BiCOH	-	48.76	43.78
	LB	PhTE	100	64.39	61.14
		MI	100	58.82	53.89
		BiCOH	100	61.46	55.8
		PhTE + MI + BiCOH	-	37.84	57.38

the classical approximation of training and testing. This is later addressed in the discussion section.

The obtained results in OP-2 show a very good level of immediacy and readiness, since we were able to detect all changes of states (our concept drifts), adapting the output.

To control stochasticity levels in the outputs due to the online modification of the clusters, the determination of the rest of the system's hyperparameters has been done by fine tuning. The percentage of change in *radius increase, radius decrease* and *cluster addition* has been set to 0.5, 0.5 and 0.3 respectively. Higher values lead to a loss of accuracy of the system. Finally, in our system we do not allow the removal of created clusters or the joining of two clusters, since, as previously stated, the system achieves best performance when the necessary clusters are opened and maintained in the system. In this respect, it has been verified that the rate of 0.3 in the *cluster addition* hyperparameter is sufficient to open only the necessary clusters. i.e, no more than the three necessary clusters are opened.

8.2. Comparing DSM with a traditional clustering approach

In this section we present results of accuracy for two classical and popular clustering approaches for the sake of comparing them with stream clustering. Both chosen because they carry relatively little computation: i) K-means++ (Arthur & Vassilvitskii, 2007) using 3 clusters and ii)a density-based clustering using a canopy algorithm (McCallum, Nigam, & Ungar, 2000), with 3 clusters; a periodic pruning rate of 0 to avoid deleting open clusters (as in stream clustering); The tight distance T2 and loss distance T1 have been set individually for each record: T2 has been set based on the feature (synchronization index) standard deviation as $SD = 0.5 * SD/(max - \min)$; and $T1 = T2 * 1.25$.

We perform the analysis by considering the features both alone and combined, to improve the results of the classical clustering techniques. However, as can be seen in Table 3, the results of classical techniques are far from the performance obtained with DSM.

The results confirm our initial hypothesis which is why we opted for DSM: neural activity constitutes a source of non-linear and non-stationary data. For that reason a system able to adapt itself, learning in a continuous way would benefit the performance, increasing significantly the global accuracy and using the IMD resources much more effectively.

9. Discussion

9.1. Heterogeneity of LFP connectivity across patients

Previous studies on LFP-STN show that it presents a high degree of heterogeneity across patients (Levy et al., 2000; McNeely et al., 2011; Whitmer et al., 2012). This suggests that several connectivity patterns could exist, which have yet to be identified (Hohlefeld et al., 2013).

We have not delved deeply into this question, but the applied feature selection procedure reveals some information about it. Some synchronization measures are more stable across patients, while others are more specific, not showing changes in all cases, as is the case of PLV and PhTE.

Despite the inter-subject variability, the connectivity patterns revealed from some indexes perfectly detect the change between tremor states in all the cases. This is exactly the reason why they have been selected as features, while others cannot properly distinguish among states, and were discarded. Certainly, more studies in this direction are necessary.

9.2. Preferred frequency bands

As mentioned before, previous studies have found that local synchronization in the beta band is linked with bradykinesia and

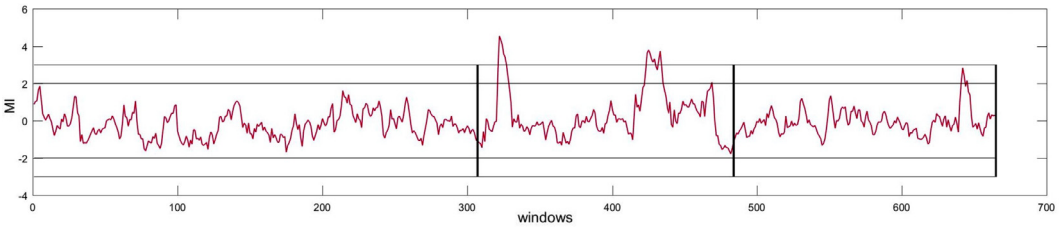

Fig. 6. Dynamic Analysis of LFP-STN measured with MI over the tremor band. The left and right black vertical line represents the transition from *NT* to *TO* and from *TO* to *T*, respectively. The horizontal lines represent the 2 and 3 standard deviation thresholds for statistical significance.

rigidity, but not with the tremor. Our results show that the connectivity fluctuations when tremor appears are more appreciable on the tremor and the lower beta bands. We have observed, as in other studies that not all the beta band is involved at the same level in PD symptomatology (Marceglia et al., 2006; Priori et al., 2004).

9.3. Application of LFP connectivity to closed-loop deep brain stimulation

Presently, PD has no cure. Therefore, the treatments are aimed at combating the associated symptomatology. The first option is in most cases is treatment based on levodopa. However, it can lead to numerous complications, and with the advance of the disease, some patients have to undergo surgery to change from a pharmacological treatment to neurostimulation via an implantable medical device called a neurostimulator (Perlmutter & Mink, 2006).

As commented in the introduction, both ECG signals and pathological events in cardiac diseases are well known; the first pacemaker dates back to 1958, while the first HFS-STN device (high frequency stimulation of the STN) did not appear until 1993 (Benabid et al., 1994; Benazzouz, Gross, Féger, Boraud, & Bioulac, 1993). Moreover, we know less about neural oscillations, because the signal is less accessible and more complex than ECG. Maybe for these reasons, neurostimulators and pacemakers do not work in the same way. More research work is needed concerning LFP connectivity and its relation to DBS (Benabid et al., 2009). In this sense, the main contribution of this paper is to find that some connectivity measures are able to distinguish with high accuracy between tremorous and atremorous states directly from LFP-STN activity, employing a stream clustering system. This approximation is appropriate for the closed-loop DBS problem since:

1. It does not need any other measure. It only requires as input the LFP signal that can be recorded by the IPG.
2. The system has been tested in immediacy and readiness parameters, showing that it is able to detect the change between states in real time, with no delay and with 100% accuracy.

Of course, as mentioned previously, these good results are not only the outcome of the stream clustering system but of the connectivity indexes employed as meters of LFP activity. In fact, if we perform an analysis into the dynamic of such measurements we can see how they reflect the drifts between states and that the mean level of synchronization varies across states. Fig. 6 shows an example of dynamics measured by MI index calculated over the tremor band.

As a last observation, as depicted in Fig. 6, the physical symptoms appear in the T period. Thus, since we are able to separate *TO* instances from *T* instances we could decide whether to stimulate only in the *T* period or from the beginning of *TO*. We could even prepare the IPG once the *TO* instances arrive to the clustering, and launch the stimulation when the *T* instances appears. All the possibilities will be perfectly possible with rigorous accuracy in the presented system.

9.4. Maintain opened clusters strategy

As stated in Section 8, the system is conceived to be initialised during the peri-operative period. To this end, the system observes the activity of the STN during the necessary time to watch the patient transiting between the movement states, and thus it is able to open the required clusters

We have tested several experiments to observe the effect that forgetting the initialised clusters has in classifying the subsequent instances, and subsequently in the accuracy of the system. The best results were though obtained when we maintained the opened clusters in the system, since the possibility of confusing an arriving instance to a similar one belonging to another cluster reduced drastically. Nonetheless, due to the conclusions obtained from experimentation, we strongly recommend to initialise and maintain the clusters opened in the system, to avoid misclassification.

As previously stated, this approximation is not equivalent to the classical approximation of training and testing for two main reasons:

1. In classical clustering, the clusters centroids are set during the training phase and remain constant afterwards. When we need to know the class of an instance, we evaluate to which cluster it belongs.
   In stream clustering, we update the cluster structure after the inclusion of the new sample: The centroids of clusters can be re-positioned, and the radius of the cluster can increase or decrease, as illustrated in Fig. 7.
2. In classical clustering all the information about the clusters is maintained in memory, whilst DSM algorithms operate using a limited amount of memory. To this end the stream clustering algorithm stores the so-called *CFs*, presented in section five, and the structure to maintain only that strictly necessary to operate. This information is updated each step, evolving at the time and form STN-LFP activity does.

9.5. Stochasticity of the system

There are two components in the system that can be stochastic in nature:

1. The inputs. We could find two sources of stochasticity in the inputs:
   - The noise present in the data. To reduce the effects of possible noise in the data, recordings have been pre-processed as presented in Section 3.2.
   - The peak frequency of Parkinsonian tremor may not always be at the same frequency varying slightly, not only across subjects, but also in different windows of observation of the same subject. In order to be sure we capture the tremor peak, two approaches can be taken:
     i) Identify the peak frequency between the 3–7 Hz band, denoted with *fp*, and filter the signal between $(fp-0.5)$ and $(fp+0.5)$. Note here that the peak frequency in PD have an approximate bandwidth of 1Hz. However, the

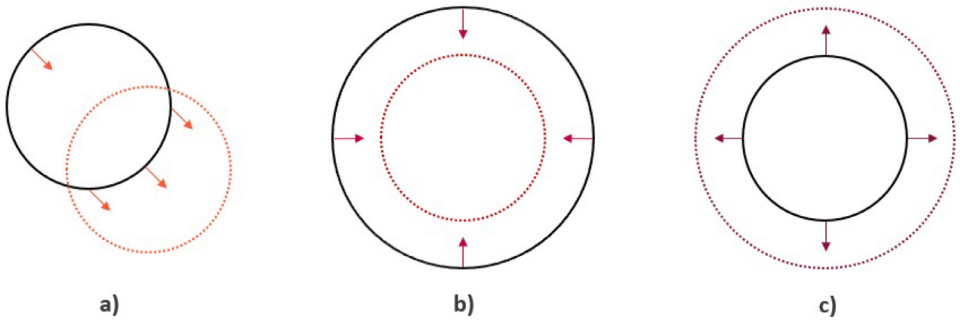

Fig. 7. Ways of updating clusters in stream clustering technique. a) cluster re-positioned. b) radius decrease and c) radius increase. The direction of the arrows indicates in which direction the cluster evolves.

tremor peak is not always exactly at the same frequency so that, to be sure to capture the peak, we should perform an ad-hoc detection and filtering process for each signal window. This approach would be very expensive, given the restricted capabilities of energy, and computing power of the IMDs, which would limit the application of our system.

ii) Use the entire tremor [3–7] band. In this respect, the results show that the selection of the whole tremor band gives a high accuracy, probably because, being the tremor a narrow peak, most of the power in the band is concentrated surrounding this peak, acting like a natural narrow-pass filter. We checked the data and, in fact, up to 78.7% of the power of the band falls withing 1 Hz of the tremor peak. A recent paper (Bruña, Maestú, & Pereda, 2018) showed that, when using Hilbert transform (as in PhTE), the existence of a clear frequency peak is enough to ensure that the phase is extracted correctly, and it is not needed to use a narrower band. This is probably also true for MI and BiCOH.

2. The cluster movements: To control stochasticity levels in the outputs due to an incorrect update of the clusters, the selection of the systems hyperparameters that control such updates has been determined by fine tuning, as stated in Section 8.1

10. Conclusions and future work

Health-care expert systems meet the challenge of processing of physiological signals to help diagnose, identify symptoms, improve treatments to ultimately ameliorate the quality of life of patients (Nancy, Khanna, & Kannan, 2017; Parisi, RaviChandran, & Manaog, 2018; Shi, Boudouh, & Grunder, 2017; Wu et al., 2010). An example of these devices are the IMDs, which, implanted within the body, treat medical conditions, monitor the state or improve the functioning of some body part. These devices deal with infinite and non-stationary signals such as LFP. The supported system needs the intelligence to adapt itself to changes and provide the most appropriate treatment in each moment. In this demanding scenario, DSM emerges as a promising technique to deal with this sort of restrictions. To the best of our knowledge, none of the existing solutions uses LFP signals as data streams. Therefore, we take advantage of the full potential of DSM and have designed a closed-loop DBS system using LFP streams.

In the first part of the work, we studied whether the patterns of LFP connectivity within the STN change when the motor symptoms of PD emerge. As suggested by the results, the tremor onset implies a change in connectivity in some frequency bands, which can be used to improve the DBS systems currently employed. Our results also show the potential of STN-LFP synchronization streams for closed-loop DBS purposes. In fact, the behavior of the clustering, which is the core of the system, is remarkable, achieving an accuracy of 100% in all cases. The system has demonstrated that it is able to detect concept drifts in the data by clustering correctly the arriving instances with a high level of immediacy.

The results are promising since, to the best of our knowledge, this is the first study attaining such levels of accuracy. Nevertheless, as a future work, it would be ideal to reproduce them in new datasets with more patients, when available. We hope this contribution can serve as a seed to future work that explores the use of STN-LFP synchronization for closed-loop DBS. In addition, and given the accuracy levels achieved here, it would also be interesting to study if data stream mining algorithms respond well in other on-demand stimulation scenarios or in other actuating devices in the medical environment.

Finally, we would like to mention that there is another line of research in closed-loop DBS that proposes the use of systems able to adapt in real time the parameters of the stimulation (the frequency, duration and amplitude of a square-wave pulse train) (Feng, Greenwald, Rabitz, Shea-Brown, & Kosut, 2007; Rosin et al., 2011). The objective of such systems is to modulate the waveform of the stimulation, which presently in practice must be hand-tuned by the clinician during the visit of the patient to the hospital periodically every 3–12 months (Deuschl et al., 2006). For their part, the aim of our system is to detect when the patient needs the stimulation itself, to reduce the side effects induced by chronic DBS and to make more efficient use of the IPGs battery. In our opinion, the combination of both kinds of approaches will constitute the complete solution for an intelligent DBS system, able to adapt the stimulation parameters by itself, and also capable of start up and shut down itself as required by the changing dynamics of the STN in real time.

Credit authorship contribution statement

C. Camara: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. **K. Warwick:** Supervision, Writing - original draft, Writing - review & editing. **R. Bruña:** Methodology, Writing - original draft, Writing - review & editing. **T. Aziz:** Resources, Writing - review & editing. **E. Pereda:** Supervision, Writing - original draft, Writing - review & editing.

Acknowledgments

We would like to thank the anonymous reviewers whose contributions have helped to improve the manuscript.

References

Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms* (pp. 1027–1035). Society for Industrial and Applied Mathematics.

Bifet A. (2018). Stream clustering - github.

Bakstein, E., Burgess, J., Warwick, K., Ruiz, V., Aziz, T., & Stein, J. (2012). Parkinsonian tremor identification with multiple local field potential feature classification. *Journal of Neuroscience Methods, 209*(2), 320–330.

Benabid, A., Pollak, P., Gross, C., Hoffmann, D., Benazzouz, A., Gao, D., et al. (1994). Acute and long-term effects of subthalamic nucleus stimulation in parkinson's disease. *Stereotactic and Functional Neurosurgery, 62*(1–4), 76–84.

Benabid, A. L., Chabardes, S., Mitrofanis, J., & Pollak, P. (2009). Deep brain stimulation of the subthalamic nucleus for the treatment of Parkinson's disease. *The Lancet Neurology, 8*(1), 67–81.

Benazzouz, A., Gross, C., Féger, J., Boraud, T., & Bioulac, B. (1993). Reversal of rigidity and improvement in motor performance by subthalamic high-frequency stimulation in MPTP-treated monkeys. *European Journal of Neuroscience, 5*(4), 382–389.

Bevan, M. D., Atherton, J. F., & Baufreton, J. (2006). Cellular principles underlying normal and pathological activity in the subthalamic nucleus. *Current Opinion in Neurobiology, 16*(6), 621–628.

Bifet, A., Gavaldà, R., Holmes, G., & Pfahringer, B. (2018). *Machine learning for data streams: With practical examples in MOA*. MIT Press.

Brown, P., Mazzone, P., Oliviero, A., Altibrandi, M. G., Pilato, F., Tonali, P. A., & Di Lazzaro, V. (2004). Effects of stimulation of the subthalamic area on oscillatory pallidal activity in Parkinson's disease. *Experimental Neurology, 188*(2), 480–490.

Brown, P., & Williams, D. (2005). Basal ganglia local field potential activity: character and functional significance in the human. *Clinical Neurophysiology, 116*(11), 2510–2519.

Bruña, R., Maestú, F., & Pereda, E. (2018). Phase locking value revisited: Teaching new tricks to an old dog. *Journal of Neural Engineering, 15*(5), 056011.

Bruns, A., Eckhorn, R., Jokeit, H., & Ebner, A. (2000). Amplitude envelope correlation detects coupling among incoherent brain signals. *Neuroreport, 11*(7), 1509–1514.

Camara, C., Isasi, P., Warwick, K., Ruiz, V., Aziz, T., Stein, J., & Bakštein, E. (2015a). Resting tremor classification and detection in Parkinson's disease patients. *Biomedical Signal Processing and Control, 16*, 88–97.

Camara, C., Peris-Lopez, P., & Tapiador, J. E. (2015b). Security and privacy issues in implantable medical devices: A comprehensive survey. *Journal of Biomedical Informatics, 55*, 272–289.

Camara, C., Warwick, K., Bruña, R., Aziz, T., Del Pozo, F., & Maestú, F. (2015c). A fuzzy inference system for closed-loop deep brain stimulation in Parkinsons disease. *Journal of Medical Systems, 39*(11), 155.

Deuschl, G., Herzog, J., Kleiner-Fisman, G., Kubu, C., Lozano, A. M., Lyons, K. E., et al. (2006). Deep brain stimulation: Postoperative issues. *Movement Disorders, 21*(14), S219–S237.

Deuschl, G., Raethjen, J., Baron, R., Lindemann, M., Wilms, H., & Krack, P. (2000). The pathophysiology of Parkinsonian tremor: A review. *Journal of neurology, 247*(5), V33–V48.

Dorsey, E. R., & Bloem, B. R. (2018). The Parkinson pandemica call to action. *JAMA Neurology, 75*(1), 9–10.

Farrer, M. J. (2006). Genetics of Parkinson disease: Paradigm shifts and future prospects. *Nature Reviews Genetics, 7*(4), 306–318.

Feng, X.-j., Greenwald, B., Rabitz, H., Shea-Brown, E., & Kosut, R. (2007). Toward closed-loop optimization of deep brain stimulation for Parkinson's disease: Concepts and lessons from a computational model. *Journal of Neural Engineering, 4*(2), L14–21.

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping, 2*(1–2), 56–78.

Gaber, M. M., Zaslavsky, A., & Krishnaswamy, S. (2005). Mining data streams: A review. *ACM Sigmod Record, 34*(2), 18–26.

Gama, J. (2010). *Knowledge discovery from data streams*. CRC Press.

Gärtner, B. (1999). Fast and robust smallest enclosing balls. In *European symposium on algorithms* (pp. 325–338). Springer.

Hohlefeld, F., Huchzermeyer, C., Huebl, J., Schneider, G.-H., Nolte, G., Brücke, C., et al. (2013). Functional and effective connectivity in subthalamic local field potential recordings of patients with Parkinsons disease. *Neuroscience, 250*, 320–332.

Josephs, K. A., Matsumoto, J. Y., & Ahlskog, J. E. (2006). Benign tremulous Parkinsonism. *Archives of Neurology, 63*(3), 354–357.

Kremer, H., Kranen, P., Jansen, T., Seidl, T., Bifet, A., Holmes, G., & Pfahringer, B. (2011). An effective evaluation measure for clustering on evolving data streams. In *Proceedings of the 17th acm sigkdd international conference on knowledge discovery and data mining* (pp. 868–876). ACM.

Kühn, A. A., Kupsch, A., Schneider, G.-H., & Brown, P. (2006). Reduction in subthalamic 8–35 hz oscillatory activity correlates with clinical improvement in Parkinson's disease. *European Journal of Neuroscience, 23*(7), 1956–1960.

Kühn, A. A., Trottenberg, T., Kivi, A., Kupsch, A., Schneider, G.-H., & Brown, P. (2005). The relationship between local field potential and neuronal discharge in the subthalamic nucleus of patients with Parkinson's disease. *Experimental Neurology, 194*(1), 212–220.

Lachaux, J.-P., Rodriguez, E., Martinerie, J., Varela, F. J., et al. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping, 8*(4), 194–208.

Levy, R., Hutchison, W. D., Lozano, A. M., & Dostrovsky, J. O. (2000). High-frequency synchronization of neuronal activity in the subthalamic nucleus of Parkinsonian patients with limb tremor. *Journal of Neuroscience, 20*(20), 7766–7775.

Little, S., Beudel, M., Zrinzo, L., Foltynie, T., Limousin, P., Hariz, M., et al. (2015). Bilateral adaptive deep brain stimulation is effective in Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, jnnp–2015.

Lobier, M., Siebenhühner, F., Palva, S., & Palva, J. M. (2014). Phase transfer entropy: A novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions. *Neuroimage, 85*, 853–872.

Lourens, M. A. J., Meijer, H. G. E., Contarino, M., Van Den Munckhof, P., Schuurman, P., van Gils, S. A., & Bour, L. (2013). Functional neuronal activity and connectivity within the subthalamic nucleus in Parkinsons disease. *Clinical Neurophysiology, 124*(5), 967–981.

Marceglia, S., Foffani, G., Bianchi, A., Baselli, G., Tamma, F., Egidi, M., & Priori, A. (2006). Dopamine-dependent non-linear correlation between subthalamic rhythms in Parkinson's disease. *The Journal of Physiology, 571*(3), 579–591.

McCallum, A., Nigam, K., & Ungar, L. H. (2000). Efficient clustering of high-dimensional data sets with application to reference matching. In *Proceedings of the sixth ACM SIGKDDinternational conference on knowledge discovery and data mining* (pp. 169–178). ACM.

McNeely, M., Hershey, T., Campbell, M., Tabbal, S., Karimi, M., Hartlein, J., et al. (2011). Effects of deep brain stimulation of dorsal versus ventral subthalamic nucleus regions on gait and balance in Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 82*(11), 1250–1255.

Medtronic (2018). Replacing your device.

Meissner, W., Leblois, A., Hansel, D., Bioulac, B., Gross, C. E., Benazzouz, A., & Boraud, T. (2005). Subthalamic high frequency stimulation resets subthalamic firing and reduces abnormal oscillations. *Brain, 128*(10), 2372–2382.

Mendel, J. M. (1991). Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications. *Proceedings of the IEEE, 79*(3), 278–305.

Miljkovic, D., Aleksovski, D., Podpečan, V., Lavrač, N., Malle, B., & Holzinger, A. (2016). Machine learning and data mining methods for managing Parkinsons disease. In *Machine learning for health informatics* (pp. 209–220). Springer.

Nancy, J. Y., Khanna, N. H., & Kannan, A. (2017). A bio-statistical mining approach for classifying multivariate clinical time series data observed at irregular intervals. *Expert Systems with Applications, 78*, 283–300.

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., & Hallett, M. (2004). Identifying true brain interaction from eeg data using the imaginary part of coherency. *Clinical Neurophysiology, 115*(10), 2292–2307.

Parisi, L., RaviChandran, N., & Manaog, M. L. (2018). Feature-driven machine learning to improve early diagnosis of Parkinson's disease. *Expert Systems with Applications, 110*, 182–190.

Pereda, E., Quiroga, R. Q., & Bhattacharya, J. (2005). Nonlinear multivariate analysis of neurophysiological signals. *Progress in Neurobiology, 77*(1–2), 1–37.

Perlmutter, J. S., & Mink, J. W. (2006). Deep brain stimulation. *Annual Review of Neuroscience, 29*, 229–257.

Priori, A., Foffani, G., Pesenti, A., Tamma, F., Bianchi, A., Pellegrini, M., et al. (2004). Rhythm-specific pharmacological modulation of subthalamic activity in parkinson's disease. *Experimental Neurology, 189*(2), 369–379.

Priori, A., Foffani, G., Rossi, L., & Marceglia, S. (2013). Adaptive deep brain stimulation (ADBS) controlled by local field potential oscillations. *Experimental Neurology, 245*, 77–86.

Rosin, B., Slovik, M., Mitelman, R., Rivlin-Etzion, M., Haber, S. N., Israel, Z., et al. (2011). Closed-loop deep brain stimulation is superior in ameliorating parkinsonism. *Neuron, 72*(2), 370–384.

Schnitzler, A., & Gross, J. (2005). Normal and pathological oscillatory communication in the brain. *Nature reviews neuroscience, 6*(4), 285.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of information*. University of Illinois Press.

Shi, Y., Boudouh, T., & Grunder, O. (2017). A hybrid genetic algorithm for a home health care routing problem with time window and fuzzy demand. *Expert Systems with Applications, 72*, 160–176.

Sugiyama, K. (2015). Complications of deep brain stimulation. In *Deep brain stimulation for neurological disorders* (pp. 195–206). Springer.

Weinberger, M., Hutchison, W. D., & Dostrovsky, J. O. (2009). Pathological subthalamic nucleus oscillations in PD: can they be the cause of bradykinesia and akinesia? *Experimental Neurology, 219*(1), 58–61.

Weinberger, M., Mahant, N., Hutchison, W. D., Lozano, A. M., Moro, E., Hodaie, M., et al. (2006). Beta oscillatory activity in the subthalamic nucleus and its relation to dopaminergic response in parkinson's disease. *Journal of Neurophysiology, 96*(6), 3248–3256.

Welzl, E. (1991). Smallest enclosing disks (balls and ellipsoids). In *New results and new trends in computer science* (pp. 359–370). Springer.

Whitmer, D., De Solages, C., Hill, B. C., Yu, H., Henderson, J. M., & Bronte-Stewart, H. (2012). High frequency deep brain stimulation attenuates subthalamic and cortical rhythms in Parkinson's disease. *Frontiers in Human Neuroscience, 6*, 155.

Wingeier, B., Tcheng, T., Koop, M. M., Hill, B. C., Heit, G., & Bronte-Stewart, H. M. (2006). Intra-operative STN DBS attenuates the prominent beta rhythm in the STN in Parkinson's disease. *Experimental Neurology, 197*(1), 244–251.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Wong, D., Clifton, D., & Tarassenko, L. (2009). An introduction to the bispectrum for EEG analysis. In *Postgraduate conference in biomedical engineering & medical physics* (pp. 61–62).

Wu, D., Warwick, K., Ma, Z., Burgess, J. G., Pan, S., & Aziz, T. Z. (2010). Prediction of Parkinsons disease tremor onset using radial basis function neural networks. *Expert Systems with Applications, 37*(4), 2923–2928.