Neuroinformatics

Management and Harmonization of Neuronal Data Recordings for Research: The Case of WebBioBank

--Manuscript Draft--

Manuscript Number:	NEIN-D-21-00115
Full Title:	Management and Harmonization of Neuronal Data Recordings for Research: The Case of WebBioBank
Article Type:	Original Article
Keywords:	Parkinson's disease; deep brain stimulation; local field potentials; biosignals; data management; signal biobank
Corresponding Author:	AmirAli Farokhniaee, Ph.D. University College Dublin - National University of Ireland: University College Dublin IRELAND
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University College Dublin - National University of Ireland: University College Dublin
Corresponding Author's Secondary Institution:	
First Author:	AmirAli Farokhniaee, Ph.D.
First Author Secondary Information:	
Order of Authors:	AmirAli Farokhniaee, Ph.D. Lorenzo Rossi Madeleine Lowery Mattia Arlotti Sara Marceglia
Order of Authors Secondary Information:	
Funding Information:	Science Foundation Ireland (19/IFA/7419) — Dr. AmirAli Farokhniaee
Abstract:	WebBioBank (WBB) is a web-based platform previously developed to collect and share anonymous patients' biosignals with neurodegenerative disorders such as Parkinson's disease (PD) undergoing clinical treatments and deep brain stimulation (DBS). WBB signals were mainly local field potentials (LFPs) collected in over 15 years' experience, with heterogeneous systems and in different experimental settings. The goal of this study was therefore to categorize and harmonize WBB biosignals to fully exploit its research potential. As a first example of harmonized data use, we evaluated the distribution of beta band rhythm (13-30 Hz) power of subthalamic nucleus local field potentials (STN LFPs) over age, biological sex and clinical status. We summarized the recordings information in a metadata file and developed Matlab scripts to format and organize the biosignals. 953 signals from 98 PD and Tourette patients were categorized based on signal types such as LFPs, personal information such as age, and clinical status such as on-off medication and stimulation. The old .txt files of signals were modified and their .json format were created. Additionally, signal gains and units were estimated if missing. To obtain an overall view of the harmonized WBB potential, we reported the signals distributions based on age and sex of the patients as well as their medication and stimulation status and beta band power. The algorithms developed in this study can be used to organize any kind of large-scale signal biobank. The future plan is to embed online signal processing tools on WBB for scientific analysis and clinical investigations.

Management and Harmonization of Neuronal Data Recordings for Research: The Case of WebBioBank

AmirAli Farokhniaee[1*], Lorenzo Rossi[2], Madeleine Lowery[1], Mattia Arlotti[2], and Sara Marceglia[3]

[1] *School of Electrical & Electronic Engineering, University College Dublin, Ireland*
[2] *Newronika S.p.A., Milan, Italy*
[3] *Dipartimento di Ingegneria e Architettura, Università degli Studi di Trieste, Trieste, Italy*

***Abstract*—** WebBioBank (WBB) is a web-based platform previously developed to collect and share anonymous patients' biosignals with neurodegenerative disorders such as Parkinson's disease (PD) undergoing clinical treatments and deep brain stimulation (DBS). WBB signals were mainly local field potentials (LFPs) collected in over 15 years' experience, with heterogeneous systems and in different experimental settings. The goal of this study was therefore to categorize and harmonize WBB biosignals to fully exploit its research potential. As a first example of harmonized data use, we evaluated the distribution of beta band rhythm (13-30 Hz) power of subthalamic nucleus local field potentials (STN LFPs) over age, biological sex and clinical status. We summarized the recordings information in a metadata file and developed Matlab scripts to format and organize the biosignals. 953 signals from 98 PD and Tourette patients were categorized based on signal types such as LFPs, personal information such as age, and clinical status such as on-off medication and stimulation. The old .txt files of signals were modified and their .json format were created. Additionally, signal gains and units were estimated if missing. To obtain an overall view of the harmonized WBB potential, we reported the signals distributions based on age and sex of the patients as well as their medication and stimulation status and beta band power. The algorithms developed in this study can be used to organize any kind of large-scale signal biobank. The future plan is to embed online signal processing tools on WBB for scientific analysis and clinical investigations.

Keywords: Parkinson's disease, deep brain stimulation, local field potentials, biosignals, data management, signal biobank.

I. INTRODUCTION

WebBioBank (WBB) is a web-based platform previously developed by Newronika S.p.A (Rossi et al., 2014) and collects biosignals recorded from patients with neurodegenerative diseases in different conditions and from different brain\body areas. WBB is a multi-centre web-based databank that combines clinical data collection (electronic health records, EHRs) with signal processing and analysis tools (Rossi et al., 2014) . This is of extreme importance in pathophysiological studies, where the relationship between clinical data and recorded signals is crucial for uncovering the underlying neurophysiological mechanisms such as in deep brain stimulation (DBS) used for Parkinson's disease (PD).

The WBB system architecture and language are explained in detail in (Rossi et al., 2014). The signals are in the .txt format since the LFP recordings are stored often according to the EDF (European Data Format) standard and also because of its common usage between all biosignal acquisition systems. Nevertheless, these signals have been stored in the WBB server over more than 15 years of experiments in a range of text file formats (more than different 17 formats), with little consistency over their associated clinical information. In this aspect, the first objective of this study was to harmonize all recorded biosignals on WBB, by (1) producing a unique format and style, with unique ID, (2) gathering all the sparse signals information (including patients' data) in a metadata sheet that can be retrieved by calling the unique ID, (3) creating the corresponding .json files, and (4) estimating the signals gain and unit in case of missing. The second objective of this study was to categorize the harmonized dataset stored in the new style based on two classes of signal type and patients, described in section II. This comes with a study of signals' and patients' features distributions and at the end, a report on the variation of beta band power (beta power, in short) over age, sex and medication-stimulation groups.

WBB biosignals consist of local field potentials (LFPs), electroencephalographical (EEG),

electromyographical (EMG) and electrocardiography (ECG) recordings from patients with PD and Tourette syndrome gathered over years from both male and female patients with different age ranges. The patients underwent DBS implantation surgery, by which the LFPs were recorded in different settings such as off/on DBS, off/on medication, rest/move tasks etc. (Rossi et al., 2007; Rossi et al., 2008; Cortese at al., 2015; Arlotti et al., 2018).

In this paper we explain the methods and algorithms that we utilized to organize and categorize the biosignals data set on WBB. These algorithms can be extended and utilized for any type of large-scale bank of biosignals. In addition, we evaluated the signals distributions based on age and sex of the patients as well as their medication and stimulation status, which provides an overall view of the data stored in WBB. Finally, we reported the distribution of the beta power in STN LFPs in different age and sex groups of the patients and also in different combinations of medication and stimulation states, as a proof of concept of the harmonized biosignals' potential.

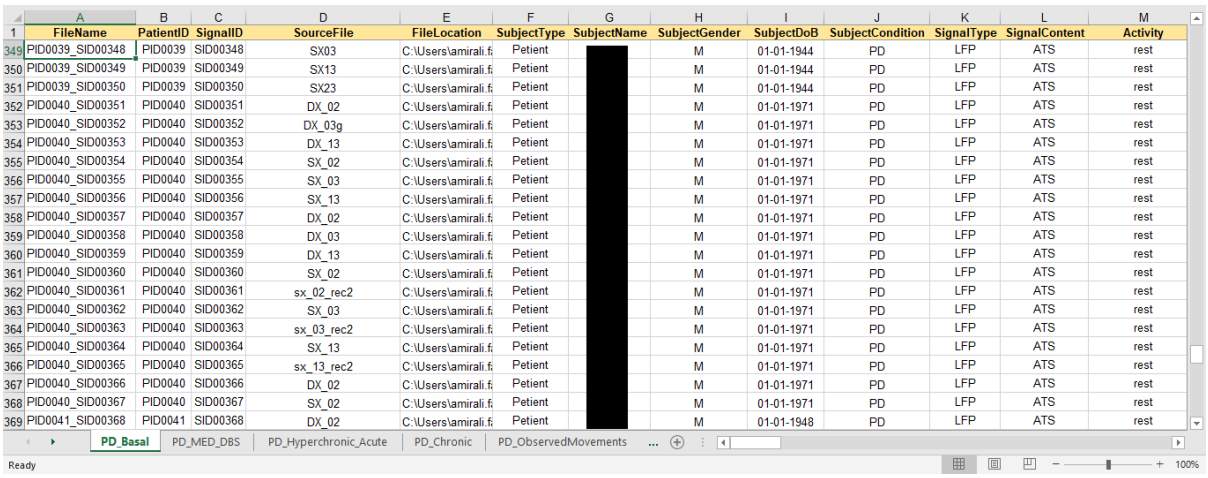
II. Materials and Methods

This section explains the methods including the Matlab script codes that are used to clean up and categorize the data on Newronika WBB data base. The codes and their corresponding descriptions are available at a github repository: https://github.com/aafarokh/WBB_NWK. We have converted the existing text files for the signals in diverse formats (old files) to unique text file format (new files), explained in section II. B, below. The codes can be used to clean up (organize) and group any kind of biological and brain recording data bank, where big amount of data is stored and their manual selection and edit would be a cumbersome task.

A. Metadata file

The first step in managing large data sets can be the creation of sheets that include a summary of each recording information. A unique pateint ID (PID) is given to each patient. Each signal

is labeled by a signal ID (SID) that is unique. At the end, each old file is named by combining PID and SID so that the new file name is in the format of PID*abcd*_SID*vwxyz* where small letters are integers. The information of each recording and subject is gathered in a metadata file as a sheet, such as in Microsoft Office Excel format. Each row presents one of our signals with the corresponding filename PID*abcd*_SID*vwxyz* and the columns gather the overall information of the subject and the recorded signal. The columns in the metadata file are explained in Table 1. Part of the metadata file is shown in Figure 1.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	FileName	PatientID	SignalID	SourceFile	FileLocation	SubjectType	SubjectName	SubjectGender	SubjectDoB	SubjectCondition	SignalType	SignalContent	Activity
349	PID0039_SID00348	PID0039	SID00348	SX03	C:\Users\amirali.fa	Petient	[redacted]	M	01-01-1944	PD	LFP	ATS	rest
350	PID0039_SID00349	PID0039	SID00349	SX13	C:\Users\amirali.fa	Petient		M	01-01-1944	PD	LFP	ATS	rest
351	PID0039_SID00350	PID0039	SID00350	SX23	C:\Users\amirali.fa	Petient		M	01-01-1944	PD	LFP	ATS	rest
352	PID0040_SID00351	PID0040	SID00351	DX_02	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
353	PID0040_SID00352	PID0040	SID00352	DX_03g	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
354	PID0040_SID00353	PID0040	SID00353	DX_13	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
355	PID0040_SID00354	PID0040	SID00354	SX_02	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
356	PID0040_SID00355	PID0040	SID00355	SX_03	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
357	PID0040_SID00356	PID0040	SID00356	SX_13	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
358	PID0040_SID00357	PID0040	SID00357	DX_02	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
359	PID0040_SID00358	PID0040	SID00358	DX_03	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
360	PID0040_SID00359	PID0040	SID00359	DX_13	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
361	PID0040_SID00360	PID0040	SID00360	SX_02	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
362	PID0040_SID00361	PID0040	SID00361	sx_02_rec2	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
363	PID0040_SID00362	PID0040	SID00362	SX_03	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
364	PID0040_SID00363	PID0040	SID00363	sx_03_rec2	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
365	PID0040_SID00364	PID0040	SID00364	SX_13	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
366	PID0040_SID00365	PID0040	SID00365	sx_13_rec2	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
367	PID0040_SID00366	PID0040	SID00366	DX_02	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
368	PID0040_SID00367	PID0040	SID00367	SX_02	C:\Users\amirali.fa	Petient		M	01-01-1971	PD	LFP	ATS	rest
369	PID0041_SID00368	PID0041	SID00368	DX_02	C:\Users\amirali.fa	Petient		M	01-01-1948	PD	LFP	ATS	rest

Figure 1. A section of metadata file. The names of the patients are hidden.

Table 1. Metadata file columns and their properties.

Column	Description	Value format
FileName	The new file name	PID*abcd*_SID*vwxyz*; *a,b,c,d v, w, x, y* and *z* are integers.
PatientID	The unique ID for each patient	PID*abcd; a,b,c,d* are integers
SignalID	The unique ID of each signal	SID*vwxyz; v, w, x, y, z* are integers
SourceFile	The name of the old file	arbitrary
FileLocation	The stored location of the old file	
SubjectType	Type of the subject in the study	Patient or Animal
SubjectName	Name of the subject	arbitrary
SubjectGender	The gender of the subject	M/F
SubjectDoB	Subject date of birth	dd-mm-yyyy
SubjectCondition	The subject's pathology	PD, TS, OCD+TC, etc.
SignalType	The type of signal	LFP, EEG, EMG, ECG, etc.
SignalContent	Voltage trace or spectral content	ATS etc.
BodySite	The recording site on the subject body	STN, GPi, etc.
Channel	The recording channels	R02, L13, etc.
Medication	Medication state	on/off
Stimulation	Stimulation state	on/off
SignalUnit	Unit of the recorded signal	V
Number_Column_Source File	This column is used for the data conversion code and mentions the	

	column in the old file where the new file is constructed on	
SignalFs	Signal sampling frequency (Hz)	Integer
SignalGain	Gain of the signal	Integer
EstimatedGain	The estimated gain in case the gain was not reported in the old file and corresponding data base	Integer
SignalBandWitdth	Bandwidth of the recorded signal	Integer-Integer
Notch	If Notch filter was used on the original signal or not	on/off
RecordingData	Date of the signal recording	dd-mm-yyyy
RecordingHour	Hour of the signal recording	hh:mm:ss
SignalDuration_s	Duration of the signal (s)	Real
Comments	Any comments	String
LinkPaper	Mentions the paper where the data has been used/analyzed	Hyperlink
BackupFolder	The folder where the backup of the old signal exists	Storage location
StimFreq	Stimulation frequency (Hz)	Integer
StimAmp	Stimulation amplitude (mA)	Real
StimPulseDuration_us	Stimulation pulse duration (μs)	Real

B. Converting old files to the new ones

The objective of this data management is to read the old files in the data base, which contain different format of saving the signals and convert them to a one-column .txt file (new files). All signal files are in .txt format, either the old or the new versions. The files that contain different signals from different subjects in the old format could have different rows and columns with various headers depending on how it was saved originally after the experiment or clinical trial. The first column contains the time in most of the cases and the other rows contain the signals. The signals could be different types of the same subject, such as LFP, EMG or different channels of one LFP recording. So far, we have identified 17 types of recording formats on WBB data base. The new files contain just one column where the first row indicates the sampling frequency (*Fs*) in Hz, and the rest of the rows are the signal values at each time sample.

C. Estimate the signals gain, evaluating the signal unit and duration

In some cases, the gain of the signal i.e., the value that the recording device has used to record the signal was not known. However, as an example, we know that dividing the value of the signal by gain for LFPs should be in microvolts (μV) order of magnitude. Also, in some cases

it can be guessed by other existing values of the same subject (e.g., similar recordings). Hence, our algorithm reads the metadata file and wherever that the gain was not given, tries to estimate a gain based on the SignalType and writes it to the metadata sheet and save. We compute the unit of each signal summarized in the metadata file based on the already given/estimated signal gain. The result is written in the SignalUnit column of the metadata file for each corresponding signal and is saved. The duration of each signal in the metadata file is also calculated and its value was added up to calculate the total recording time per sheet and overall.

D. Create IEEE JSON from the new files

For web applications, it is extremely handy to merge every signal with the corresponding information from the metadata file in one single file in IEEE JSON (.json) format. An algorithm is written that merges these information and saves each new signal in .json format in parallel with .txt file that already exists. This algorithm helps researchers and web developers along with bioengineers to save any signal in .txt format along with its information as a .json file using Matlab.

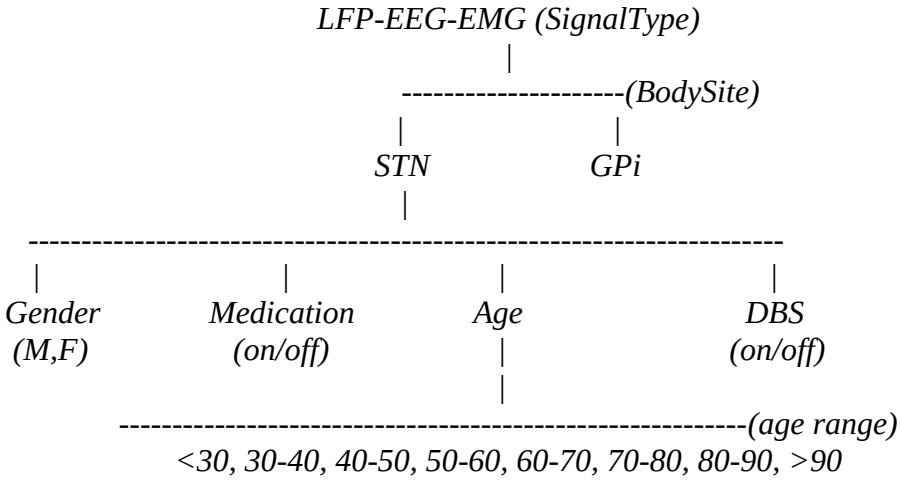
E. Clean up the data set

There were some files in the WBB with no given *Fs* and since this property is essential for any kind of data analysis, one might want to keep only the signals with given *Fs*. The other essential property of the signal that is required for data analysis is the SignalType, which indicates if it is LFP, EEG or any other type of biosignal. In this regard, we read the metadata file, identify those files that have at least both *Fs* and SignalType reported and save them in a new folder called Clean. The corresponding .json files will be saved in the new folder as well.

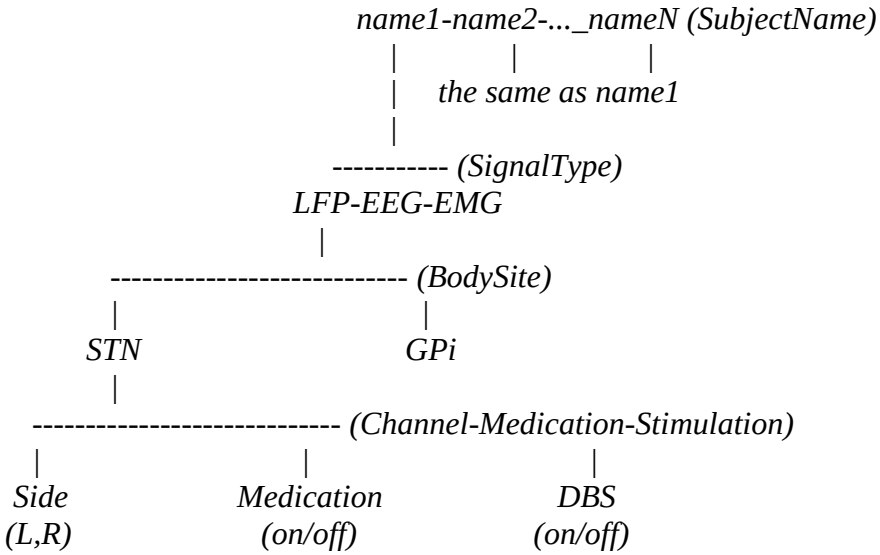
F. Data Identification & Categorization

We have categorized the signals on WBB based on two distinct classes.

- *Class-1* hierarchy collects signals to different "Types":



- *Class-2* hierarchy collects signals to different "Patients":



G. The cleaning pipeline and data organization routine

In this section we summarize the data management routine that we have utilized on WBB dataset. The following matlab files can be found on the github page (https://github.com/aafarokh/WBB_NWK).

1- Enter the info in the metadata file as complete as possible by referring to the old format of the files and sparse information. **a)** FileName: Add as many as the

columns of the old file dictates. The old text file contains different columns. The person must open the old file and check how many of these columns are related to recorded signal at different channels. **b)** If the recording date of the signal is missing, write down 00:00:00 in RecordingDate column. **c)** The most important factors to keep the signals and provide analysis on them are sampling frequency (Fs) and type of the signal (SignalType). **d)** By checking the names, the person needs to make sure if a subject name is not repeated in previous rows and sheets, otherwise the same PID should be used. **e)** The columns called EstimatedGain, SignalDuration and SignalUnit should be left blank as they will be filled out later automatically by the Matlab scripts. **f)** If stimulation is on, try to provide info for the last three columns of the sheets which are the stimulation frequency, amplitude and pulse width.

2- Run *FileAdaptor.m* to create the new files. This code should be run once the FileName columns in the metadata file and their corresponding Fs are fixed.

3- Run *Gainestimator.m* to estimate the missing gains in the metadata file.

4- Run *SignalUnit_calc.m* to evaluate signals units.

5- Run *SignalDuration_calc.m* to evaluate the duration of the signals in seconds.

6- Run *JsonAdaptor.m* to create corresponding json files for each new file. Note that this step can be run only if the previous steps were run already. Also note that this step can be skipped in this routine without a problem to the whole concept.

7- Run *CleanUp.m* to place the files with given Fs and SignalType in a subdirectory for further spectral and statistical analysis.

8- Run *AutomatedDataCheck.m* to make a folder for each sheet of the metadata file called Diagrams and plots the signal as a time series, computes its spectrogram and Power Spectral Density (PSD) and plot them in a figure specific to the signal.

9- Run *Signal_identifier_classifier.m* to classify signals under classes 1 and 2.

10- Run *SingalsStatistics.m* to evaluate different statistical aspects of WBB such as pie-plots distributions and box plots.

H. Analysis example: calculation of the total and normalized beta power

To better understand the potential use of this harmonized dataset, we considered only the LFP signals recorded from the STN of parkinsonian patients and calculated the power of the beta band rhythm (13-30 Hz). Beta band rhythm of STN LFPs is known to be correlated with motor impairments in PD (Little and Brown 2012) and is used as a biomarker for control strategies in DBS (Arlotti et al., 2018; Bronte-Stewart et al., 2020) at the optimum levels (Blumenfeld et al., 2015; Farokhniaee and Lowery 2021).

According to the literature, beta power was defined and calculated as the area under the PSD curve from 13 to 30 Hz, Eq. 1:

$$Beta\ Power = \int_{13}^{30} PSD(f)\ df \qquad \text{(Eq. 1)}$$

where *f* is the frequency in Hz. The normalized beta power was defined as the ratio of the beta power with respect to the total power from 1 to 125 Hz as stated in Eq. 2:

$$Normalized\ Beta\ Power = Beta\ power\ /\ \int_{1}^{125} PSD(f)\ df \qquad \text{(Eq. 2)}$$

The choice of 125 Hz for the higher limit of the total power was to exclude any stimulation artefact that induces high power at 130 Hz if DBS was on.

III. RESULTS

953 signals from 98 patients, of which 89 are with PD, were identified and organized based on personal information and clinical status such as on/off medication and stimulation. This is equivalent to more than 98 hours of recordings from different patients including their LFP, EEG and EMG signals. The old .txt files of the signals were modified and their .JSON format were created. Additionally, signal gains were estimated if missing. We reported some information of the data stored in WBB based on diversity of sex, medication and stimulation

status, types of the signals and the distribution of beta power within different subgroups as below.

A. Sex Statistics

The total number of male and female patients, and the breakdown in signals by biological sex, in WBB are presented in Figure 2 A and B respectively.

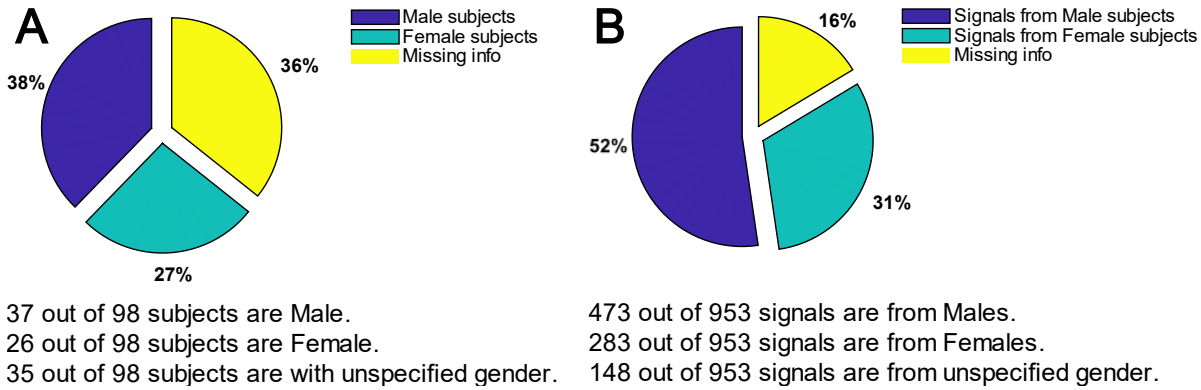


Figure 2. **A)** Patients' biological sex distribution pie diagram calculated over 98 patients on WBB. **B)** Signals' sex distribution shows how many of the 953 signals are recorded from male or female subjects.

B. Signals Type

We evaluated the number of signals in the data bank which are either LFP, EEG, EMG or ECG, presented in Figure 3.

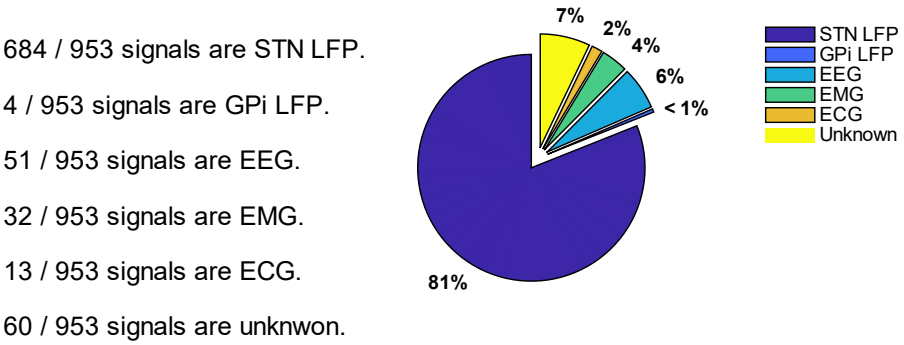


Figure 3. Different types of signals (SignalType) distributed in WBB. The signals can be LFPs from STN or GPi, EEG, EMG and ECG.

C. Medication and Stimulation Statistics

Figure 4 shows the number of all signal types that are recorded during different medication and/or stimulation states.

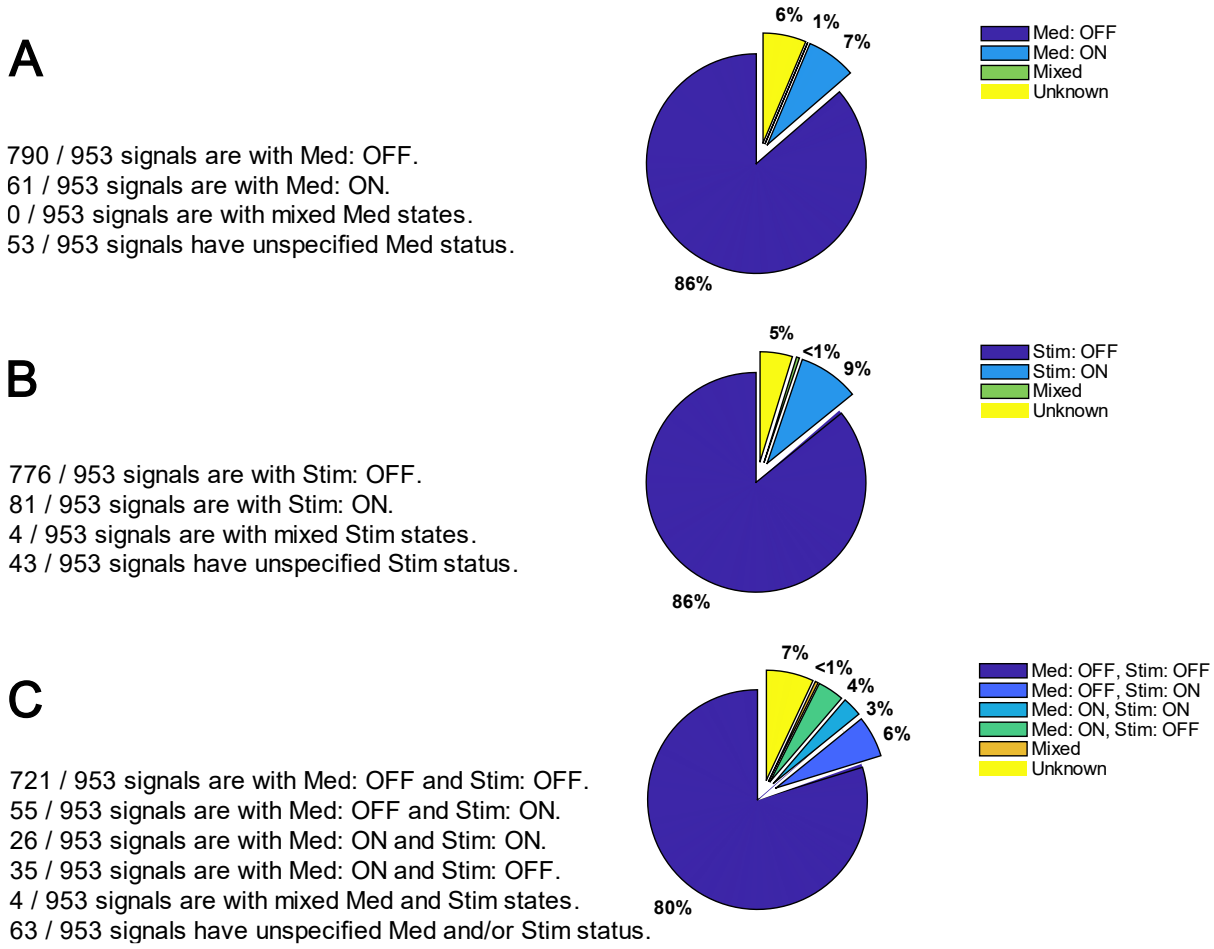
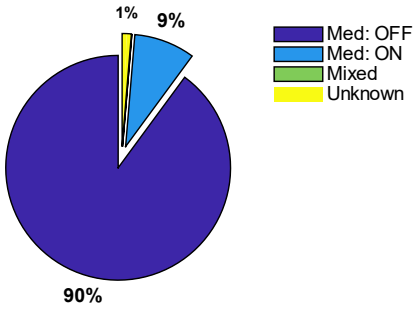


Figure 4. The distribution of all signals on WBB based on their **A)** medication status only, **B)** stimulation status only and **C)** both medication and stimulation states.

Figure 5 illustrates the same statistics as in Figure 4 but for STN LFPs only.

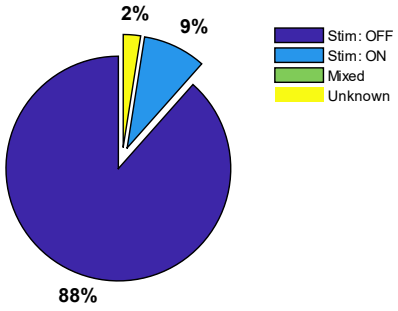
A

615 / 684 signals are with Med: OFF.
60 / 684 signals are with Med: ON.
0 / 684 signals are with mixed Med states.
9 / 684 signals have unspecified Med status.



B

605 / 684 signals are with Stim: OFF.
62 / 684 signals are with Stim: ON.
0 / 684 signals are with mixed Stim states.
17 / 684 signals have unspecified Stim status.



C

571 / 684 STN LFPs are with Med: OFF and Stim: OFF.
36 / 684 STN LFPs are with Med: OFF and Stim: ON.
26 / 684 STN LFPs are with Med: ON and Stim: ON.
34 / 684 STN LFPs are with Med: ON and Stim: OFF.
0 / 684 STN LFPs are with mixed Med and Stim states.
17 / 684 STN LFPs have unspecified Med and/or Stim status.

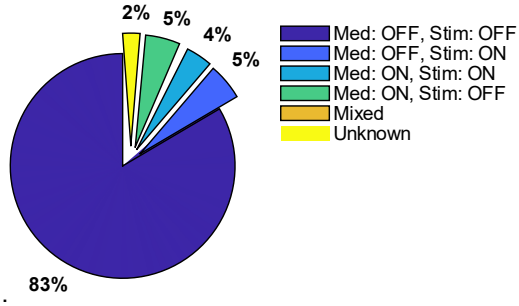


Figure 5. The distribution of STN LFPs on WBB based on their **A)** medication status only, **B)** stimulation status only and **C)** both medication and stimulation states.

D. Age Statistics

The age range distribution of the patients is shown in Figure 6.

2 / 98 patients are below 30 y.o.
3 / 98 patients are between 30-40 y.o.
9 / 98 patients are between 40-50 y.o.
7 / 98 patients are between 50-60 y.o.
13 / 98 patients are between 60-70 y.o.
2 / 98 patients are between 70-80 y.o.
0 / 98 patients are between 80-90 y.o.
1 / 98 patients is above 90 y.o.
61 out of 98 subjects are with unspecified age.

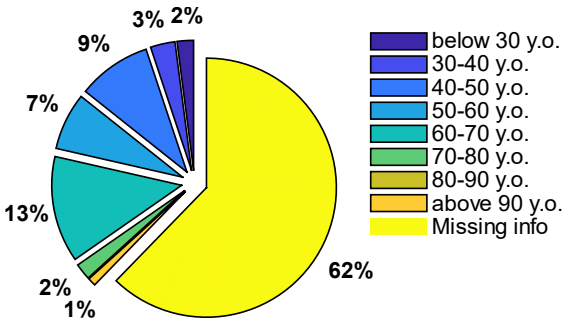


Figure 6. Number of patients in different age groups.

We also counted the number of STN LFPs that are recorded from different age ranges of patients and the results are shown in Figure 7.

9 / 684 STN LFPs are recorded from patients below 30 y.o.
13 / 684 STN LFPs are recorded from patients between 30-40 y.o.
133 / 684 STN LFPs are recorded from patients between 40-50 y.o.
51 / 684 STN LFPs are recorded from patients between 50-60 y.o.
113 / 684 STN LFPs are recorded from patients between 60-70 y.o.
19 / 684 STN LFPs are recorded from patietns between 70-80 y.o.
0 / 684 STN LFPs is recorded from patients between 80-90 y.o.
5 / 684 STN LFPs are recorded from patients above 90 y.o.
341 out of 684 STN LFPs are with unspecified age.

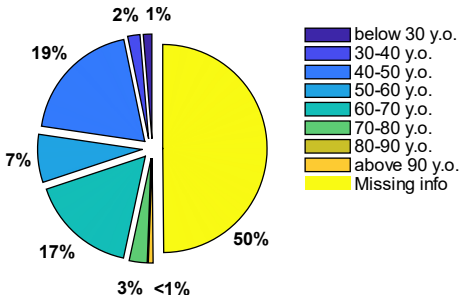


Figure 7. Age distribution of the STN LFPs in WBB.

E. Beta Power Distributions

Figures 8 to 10 show the distribution of the beta power and normalized beta power over STN LFPs obtained from different age range, sex and medication and stimulation status. The data harmonization process enabled us to obtain the beta power distribution across different clinical variables and different patient's conditions (such as medication off/on or stimulation off/on).

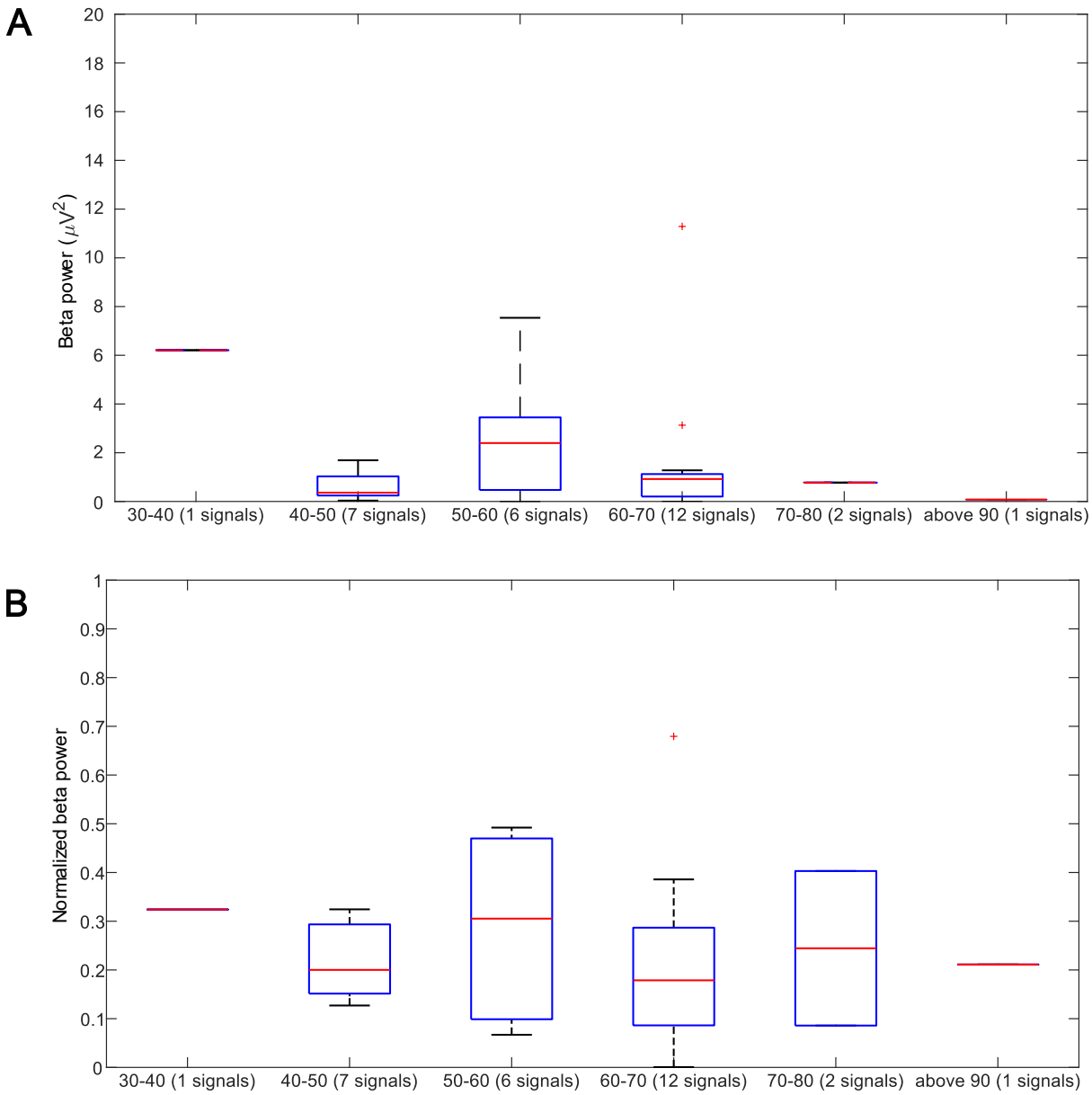


Figure 8. The distribution of STN LFPs **A)** beta power and **B)** normalized beta power over different age ranges. Medication and stimulation are off.

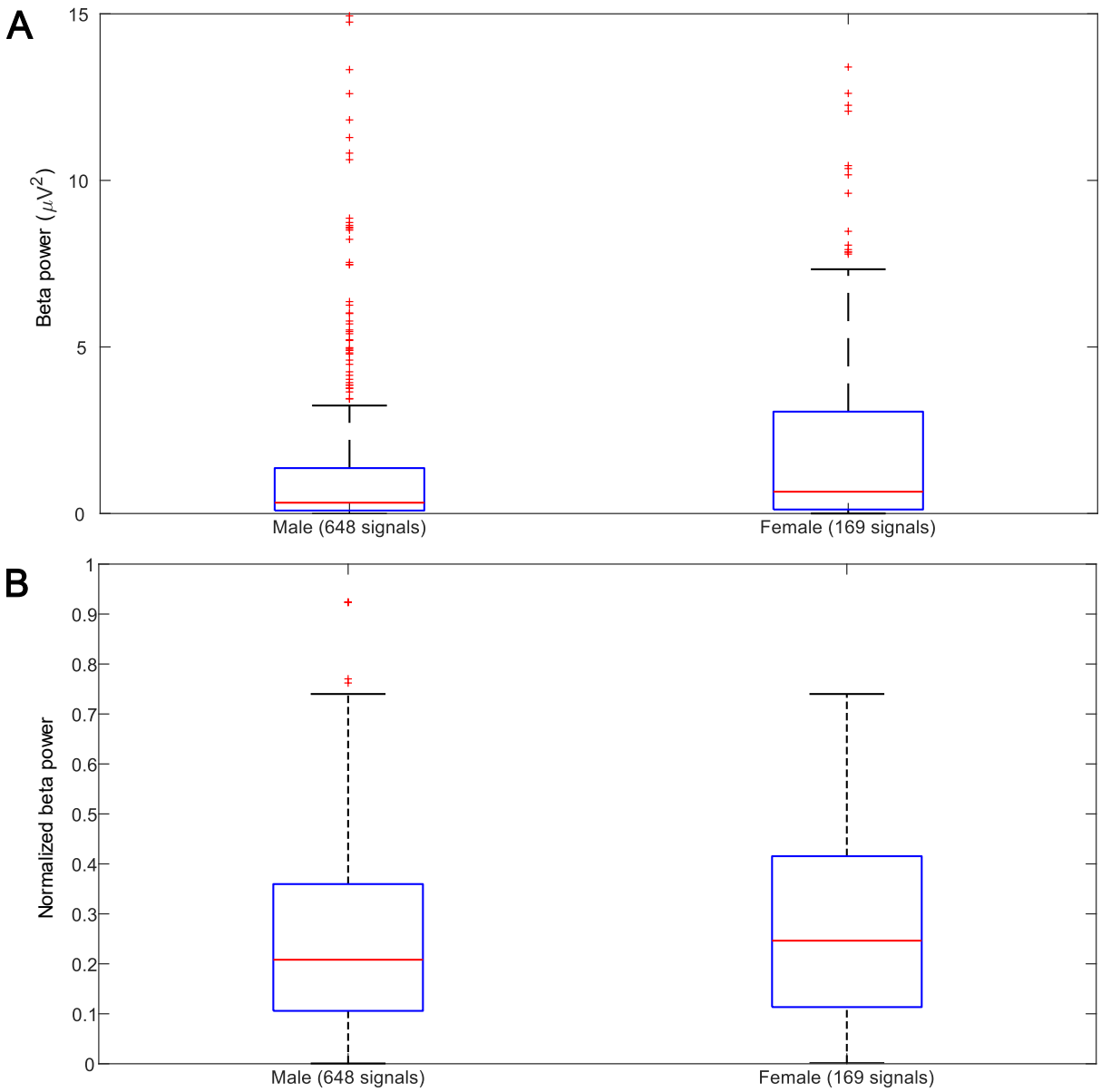


Figure 9. The distribution of STN LFPs **A)** beta power and **B)** normalized beta power for male and female groups during off medication and off stimulation.

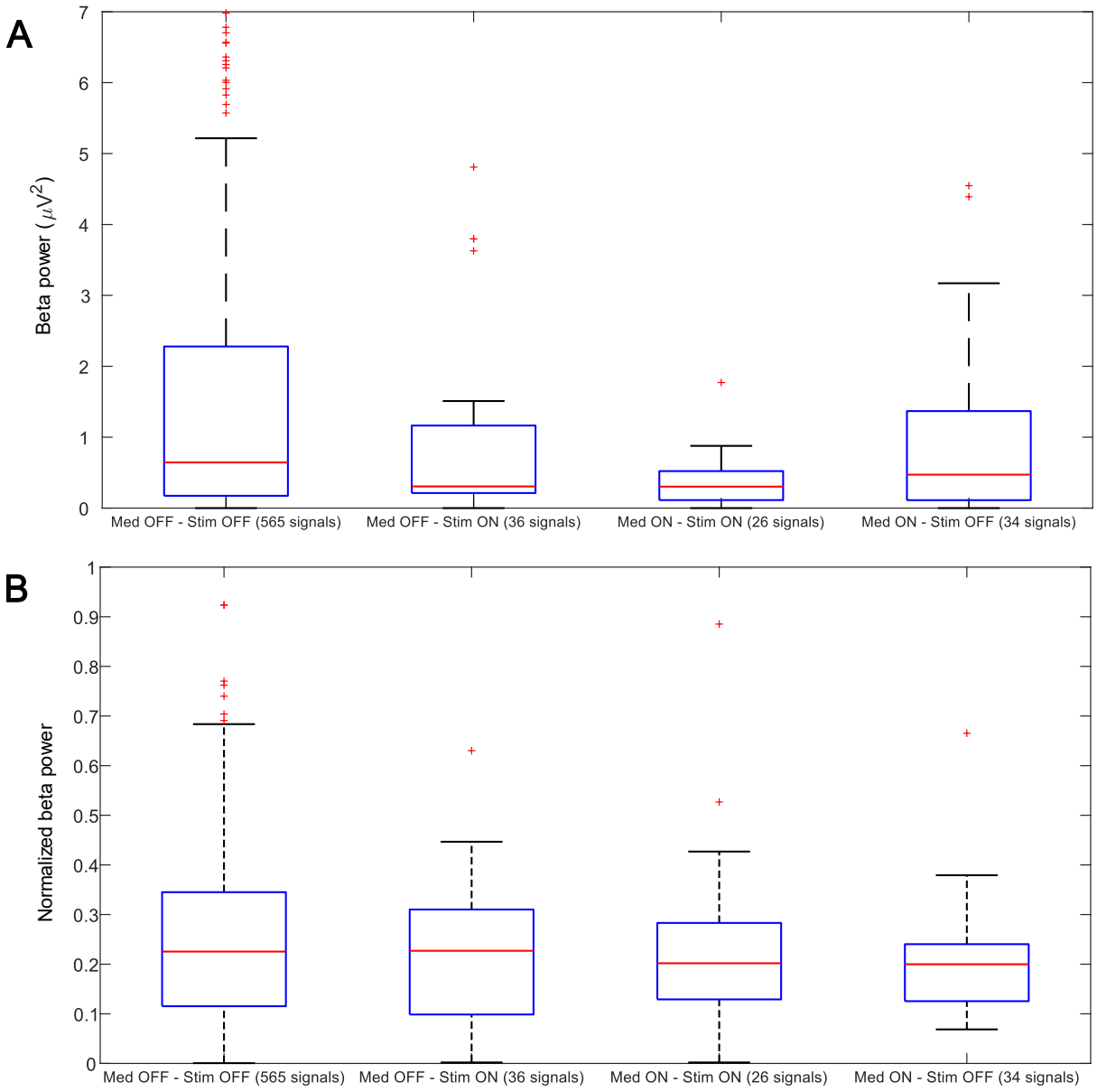


Figure 10. The distribution of **A)** beta power and **B)** normalized beta power for LFPs recorded from STN during different combinations of medication and stimulation states.

IV. Discussion and Concluding Remarks

In this paper we presented an update on WBB data base and explained the algorithms used to clean up and organize the large amount of data, which is useful for biomedical informatics community to utilize for any other large data base of biosignals. We also reported an overview of the available biosignals in WBB data base including the distribution of types of the signals, patients' age range and sex plus their medication and stimulation status. Missing information in pie charts were mostly because of inconclusive or missing clinical notes.

The neuro-biosignals collected in WBB represent the direct neuronal response to many conditions, including pharmacological or neuromodulation treatments, motor tasks, action observation experiments, cognitive and behavioral stimuli in several neurological and neuropsychiatric disorders, particularly PD.

We organized and grouped the large dataset of WBB coming from multiple experiments over different protocols, which provides more information than smaller datasets collected from single experiments. However, data harmonization is essential to increase our knowledge on neural signals and their relationship with pathologies. Here, as a proof of concept, we considered a biomarker of PD such as beta power and by data harmonization, were able to evaluate its distribution over sex, age range and medication/stimulation status groups.

The developers of WBB design the platform such that the authorized users, researchers and clinicians would be able to view the signal time series and the power spectral density that is computed for each arbitrary minute of the signal. The appearance, tabs, data tables and case report forms are getting modified. The questionnaires, amount of medication and clinical information of the patient will be implemented in the patient's information section (tab). The device tab would include the name and properties of the medical devices paired with the patient and used in the DBS electrode implantation such as Medtronic. WBB developers are in the process of implementing signal pre-processing and processing algorithms to the platform such

as filtering in different rhythms, power spectral densities, spectrograms, and other signal features. This is a step beyond state of the art of such platforms that exist to date, which gives the authorized users the ability to extract important and basic features of the signals known to biomedical engineers, clinicians and researchers right on the web, without a priori advanced knowledge of coding computational algorithms.

DECLARATIONS

Funding: This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 19/IFA/7419. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Conflicts of interests: The authors have no conflicts of interests.

Availability of Data and Material: The data would be available after publication for authorized researchers.

Code availability: All the Matlab scripts developed for this study can be found at: https://github.com/aafarokh/WBB_NWK.

Authors' contributions: LR, SM, AF: Conceptualization; AF, SM, MA: Data curation; AF, ML: Formal analysis; AF, ML: Funding acquisition; AF, SM, LR: Methodology; LR, MA: Project administration; LR: Resources; LR, SM, ML: Supervision; AF, SM, ML: Validation; AF: Visualization; AF: Writing - original draft; AF, ML, SM: Writing - review & editing.

Ethics approval: N/A

Consent to participate and publication: All authors consent to participate in this research and the publication.

REFERENCES

Arlotti M, Marceglia S, Foffani G, Volkmann J, Lozano AM, Moro E, et al. Eight-hours adaptive deep brain stimulation in patients with Parkinson disease. Neurology. 2018;90(11):e971–6.

Blumenfeld Z, Velisar A, Koop MM, Hill BC, Shreve LA, Quinn EJ, et al. Sixty hertz neurostimulation amplifies subthalamic neural synchrony in parkinson's disease. PLoS One. 2015;10(3):1–12.

Bronte-stewart HM, Petrucci MN, Day JJO, Afzal MF, Parker JE, Kehnemouyi YM, et al. Perspective: Evolution of Control Variables and Policies for Closed-Loop Deep Brain Stimulation for Parkinson' s Disease Using Bidirectional Deep-Brain-Computer Interfaces. 2020;14(August):1–8.

Cortese F, Rosa M, Arlotti M, Cogiamanian F, Ardolino G, Rampini P, et al. Abnormal local field potentials precede clinical complications after DBS surgery for Parkinson's disease: A case report. Clin Neurophysiol. 2015;126(5):1056–8.

Farokhniaee A, Lowery M. Cortical network effects of subthalamic deep brain stimulation in a thalamo-cortical microcircuit model. J Neural Eng. 2021;

Little S, Brown P. What brain signals are suitable for feedback control of deep brain stimulation in Parkinson's disease? Ann N Y Acad Sci 2012;1265:9–24.

Rossi E, Rosa M, Rossi L, Priori A, Marceglia S. WebBioBank: A new platform for integrating clinical forms and shared neurosignal analyses to support multi-centre studies in Parkinson's Disease. J Biomed Inform 2014;52:92–104. http://dx.doi.org/10.1016/j.jbi.2014.08.014

Rossi L, Foffani G, Marceglia S, Bracchi F, Barbieri S, Priori A. An electronic device for artefact suppression in human local field potential recordings during deep brain stimulation. J Neural Eng. 2007;4(2):96–106.

Rossi L, Marceglia S, Foffani G, Cogiamanian F, Tamma F, Rampini P, et al. Subthalamic local field potential oscillations during ongoing deep brain stimulation in Parkinson 's disease. 2008;76:512–21.