Is the diagnosis of sleep disorders and tendencies towards psychopathological traits possible through the detection and analysis of sleep spindles?

By

Agathe de Vulpian

A thesis submitted to
Ecole Polytechnique
for the degree of
BACHELOR OF SCIENCE

Data Science and Mining team (DaSciM)
Laboratoire Informatique de l'X (LIX)
Ecole Polytechnique
January - March 2020

ABSTRACT

Sleep spindles are characteristic electroencephalogram (EEG) signatures of stages 2 and 3 of non-rapid eye movement sleep. Here we detect sleep spindle in the goal of using their characteristics (duration, frequency, density) to discover if it is possible to use them to diagnose sleep disorders or psychopathological traits. In complement we use the answers to the second version of the Minnesota Multiphasic Personality Inventory (MMPI-2) which gives an insight on the patients' mental health and personality traits.

We reach the conclusion that using sleep spindles - detected in the way we describe in this paper - cannot help in the diagnosis of Paradoxical Insomnia.

Moreover we show that, upon trying to predict tendencies towards psychopathological traits using the spindle data collected, a difference between non-psychotic traits (such as Depression or Somatisation) and psychotic traits (such as Paranoia and Schizophrenia) arises. Indeed, in the first case spindles are unable to predict accurately if a patient has a high or low tendency towards a non-psychotic trait. However, in the second case the spindles allow to predict accurately that a patient has low tendency towards a psychotic trait. While the detection of low tendencies could help physicians in eliminating these psychotic traits from the diagnosis and could therefore prove to be a useful result, the inability to predict high tendencies for all types of psychopathological traits offers a positive finding for patient privacy.

KEY WORDS

Sleep spindles, MMPI-2, Psychopathological traits, Paradoxical Insomnia , EEG, (Non)Psychotic traits

ÉCOLE POLYTECHNIQUE
UNIVERSITÉ PARIS-SACLAY

Statement of Academic Integrity Regarding Plagiarism

I, the undersigned de Vulpian Agathe hereby certify on my honor that:

1. The results presented in this report are the product of my own work.
2. I am the original creator of this report.
3. I have not used sources or results from third parties without clearly stating thus and referencing them according to the recommended rules for providing bibliographic information.

I hereby declare that this work contains no plagiarized material.

Date : 12.04.2020 .

Signature :  .

Table of Contents

Chapter 1

Introduction

Sleep spindles are brain waves that occur during sleep. In the last decade, sleep spindles have attracted increasing attention due to the the many intriguing relationships between spindles and various diseases for instance Parkinson or Alzheimer and sleep spindles and cognitive faculties such as memory consolidation, intelligence, dream recall and sleep preservation. Nonetheless, a methodological wall has impeded the study of sleep spindles as their detection is of utmost difficulty as we will see. Moreover the unavailability of large, standardized, high-quality databases has also greatly restricted the findings on this topic.

The goal of this thesis is to discover if the detection and analysis of sleep spindles can improve the diagnosis of sleep disorders and the diagnosis of tendencies towards a given psychopathological trait. Prior studies have focused on trying to characterise spindles in order to understand their genetic architecture [1]. Studies have focused on characterising sleep spindles in more precise population groups such as patients suffering from major depressive disorder (MDD) [2], mentally retarded children [3] or Schizophrenia patients and their non affected relatives [4]. But studies have yet to discover to what extent it is possible to use sleep spindles to diagnose particular sleep disorders such as Paradoxical Insomnia or psychological conditions such as Hypochondria.

This study was made possibles through the rare access to medical data regrouping both the work of a sleep expert and of a psychiatrist.

1.1 Sleep

Sleep is composed of 4 phases, the first three fit in to the Non Rapid Eye movement (NREM) category and are denoted by N1, N2 and N3. The last is the REM phase which stands for rapid eye movement. A sleep cycle is the progression through the various stages of NREM sleep to REM sleep before beginning the progression again with NREM sleep. Typically, a person begins a sleep cycle every 90-120 minutes resulting in four to six cycles per sleep time. (see figure 1.1)

REM is the time when the most vivid dreams occur. The rapid eye movements that occur can be seen as sharp, rapid movements. Brain waves during REM sleep are considered to be of low amplitude and mixed frequency consistent with higher activity than that seen in Stages 2 and 3. Each phase is characterised by different brain waves. Our study focuses on sleep spindles which occur in phases N2 and N3. A detailed explanation of sleep spindles is given in chapter 2.

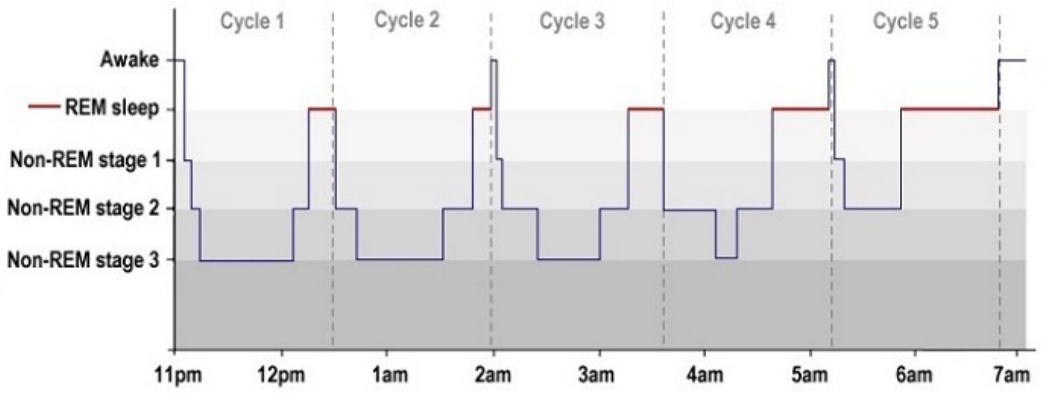


Figure 1.1 Sleep cycles broken down into the 4 phases

1.2 Motivations

According the the World Health Organisation, approximately one in four people in US have a mental illness at some point in their life. It would therefore be a key contribution to facilitate the diagnosis of such disorders.

Regarding sleep disorders, we choose to focus our study on insomnia. Indeed 30% of adults in the U.S. suffer from Insomnia making it the most common sleep disorder. More precisely, we explore Paradoxical Insomnia, otherwise know as Sleep Misperception as much like most pshyopathological traits the diagnosis of this sleep disorder relies greatly on the patient's perception and is thus subjective. Meaning that the diagnosis is often very difficult and inexact and would benefit from improvement.

We explore the possibility of the existence of a bio marker (sleep spindles) which could facilitate the diagnosis of such disorders and traits. We choose sleep spindles firstly as they are characteristic of our sleep and secondly because abundant studies have shown undeniable links between sleep disorders and psychophathological traits such as Depression, Anxiety, Post-traumatic Stress, Schizophrenia ,Bipolarism, Eating Disorders, Borderline Personality Disorder and others [23] showing that our sleep mechanism deeply affects our mental health.

1.3 Method

We first focus on detecting sleep spindles in the patients' Electroencephalogram (EEG). To do so in the most accurate manner we detect them using three different algorithms. Then we create a spindle data set by extracting the density, average duration, average frequency and average number of oscillation within all sleep spindles in stages 2 and 3 of non-rapid eye movement sleep.

Next using this data set we attempt to predict Paradoxical Insomnia (PI). PI is a sleep disorder characterised by a complaint of severe insomnia disproportional to the presence of objective sleep

disturbance. Put in simpler words the patient believes he does not sleep well (not long enough or wakes up many times) while in reality he does. Through visualisation of the data set and using classification techniques we come to the conclusion that the spindle data cannot properly predict Paradoxical Insomnia and can thus not help in it's diagnosis.

Lastly we focus on analysing if the spindle data is able to predict a patient's tendency towards a psychopathological trait. This was done using the scales of the MMPI-2 questionnaire (detailed below) all patients used in the study had answered. We focused on 5 psychopathological traits: Hypochondria, Depression, Somatization, Paranoia and Schizophrenia.

1.4 Results

It is not possible to predict Paradoxical Insomnia using spindle data - detected in the manner described in this paper.
Moreover, in using spindle data top predict psychopathological traits, 2 groups appeared those of the non-psychotic traits and psychotic traits. Psychotic disorders are severe mental disorders that cause abnormal thinking and perceptions. It was found that for the non-psychotic traits it is not possible to predict a high or low tendency. However for psychotic traits the use of spindles is unable to predict a high tendency towards them but can accurately predict a low tendency. Allowing to eliminate with certainty the possibility that the patients suffers from these psychotic traits.
Having only access to the data for 267 patients limits the importance of these results. However it is important to note that the spindle data we detect is in accordance with the results found in studies characterising sleep spindles [1] (for instance a higher density of spindles in N2, higher density in younger patients) thus the methods we use to detect the sleep spindles and create the data set are reliable and can be reused for a larger scale study.

1.5 The data

We use the data collected by Olivier Pallanca through the course of several years in the frame of his work as a psychiatrist and neurophysiologist at Centre d'investigation et de traitement de l'insomnie at the Pitié Salpêtrière hospital in Paris. For each patient we have an overnight polysomnographic recording from which we extract specifically the EEG recording for spindle analysis. Polysomnography includes Electroencephalography (EEG) measuring brain activity, Electromyography (EMG) measuring muscle activity or skeletal muscle activation, Electrooculography (EOG) measuring eye movements and Electrocardiography (ECG) measuring heart rhythm. EOG and EMG are necessary to score sleep phases but only EEG is needed for the spindle anal-

ysis. Moreover we have a binary label indicating if he/she suffers from paradoxical insomnia and the answers the second version of the Minnesota Multiphasic Personality Inventory (MMPI-2).

1.5.1 Electroencephalogram (EEG)

The EEG recording uses a frequency of 256Hz. 8 electrodes are placed on the patient and the signals are recorded between 2 electrodes, their names and positioning can be found in figure 1.2. Our study focuses on 6 channels:

- $A_1 - C_4$
- $A_2 - C_3$
- $A_1 - F_4$
- $A_2 - F_3$
- $A_1 - O_2$
- $A_2 - O_1$

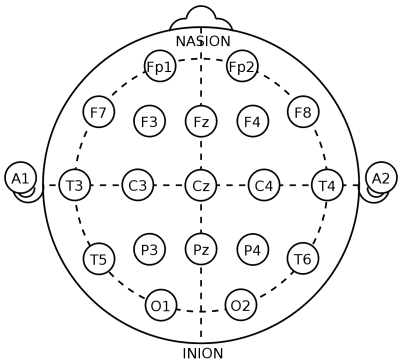


Figure 1.2 Scheme of electrode positioning on the head based on the International 10–20 system (image taken from wikipedia)

We apply to the data a notch filter of 50Hz to cancel any potential interference created by the electric plugs in the patient's room. The recordings last 6-14 hours but we only focus on certain parts of the night.

1.5.2 Minnesota Multiphasic Personality Inventory (MMPI)

MMPI is a test that assesses personality traits and psychopathology. It is primarily intended to test people who are suspected of having mental health or other clinical issues.
We use the MMPI-2 created in 1989 which replaced the original MMPI created in the 1940s. The test is composed of 500+ questions designed with 10 clinical scales which assess 10 major categories of abnormal human behaviour (see figure 1 in the appendix for a detailed description), and four validity scales, which assess the person's general test-taking attitude and whether they answered the items on the test in a truthful and accurate manner. This test can only be administered by health professionals. It is important to note that the test does not diagnose mental illnesses but rather gives a scale indicating to which extent the patient has a tendency toward certain psychopathological traits. In this study we use the scales given by the MMPI-2 after having processes the true/false answers of the patient.

Chapter 2

Sleep spindles detection

The investigation of sleep spindles rests heavily on our ability to reliably and consistently identify spindle patterns from background EEG activity. A task involving many obstacles, including: a fuzzy definition of spindles, low inter-expert agreement on their scoring, lack of consensus on standard techniques for their automated detection and inconsistencies in the methods used to evaluate the performance of automated detectors. In this section the goal is to find and understand the algorithms which allow to detect to the best of our ability the sleep spindles in phases N2 and N3. We encountered all of the above mentioned challenges and have overcome them at best to give coherent results. We explain in great depth our detection methods and how we built the data set as this is the main difficulty in handling sleep spindles.

2.1 Sleep spindles

Analysis of EEG signals during wake is of utmost difficulty as the abundance of waves creates overlapping signals which we are unable to distinguish and analyse. At night whilst the body is at rest, the signals are less agitated and we are thus able to conduct studies.
In the EEG recording we then can identify a variety of brainwaves that are characterised by their frequency and shape. Alpha waves between $8-12$Hz, Delta waves $0.5-4$Hz, Sigma waves $11-16$ Hz which constitute the sleep spindles. Lastly K-complexes which are high-amplitude biphasic waves composed of an initial negative sharp wave followed by a slow wave. (see figure plot sleep spindles)

Sleep spindles are one of the landmarks and defining characteristics of N2 sleep, which is the sleep stage in which you spend the more time during the night (approx. 45 %). According to the formal definition by the American Academy of Sleep Medicine (AASM), a sleep spindle is "a train of distinct 11–16 Hz waves, predominant over central EEG derivations and lasting more than 0.5 s". Sleep spindles are sudden bursts of oscillatory brain activity generated in the reticular nucleus of the thalamus. These brainwaves are called sleep spindles because of how they look when printed out on an EEG reading as can be seen in figure 2.1.

Sleep spindles were first studied by Gibbs and Gibbs in the 1960 [16] [17], and have been of interest since. Although the functional role of sleep spindles remains unclear, there is evidence they are involved in a number of domains including maintaining disconnection from the external

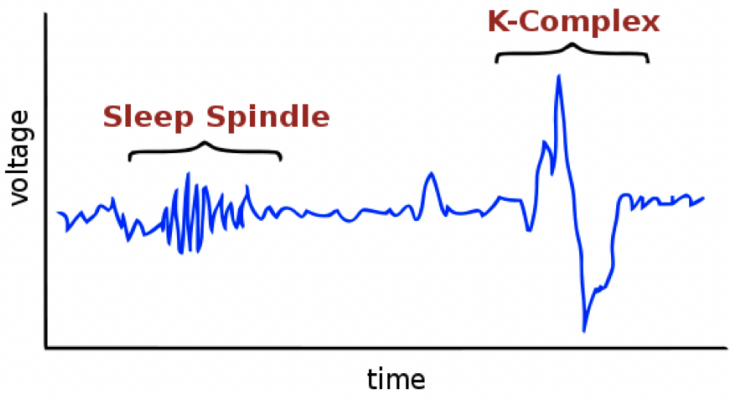


Figure 2.1 Plot of sleep spindles and k-complex

environment during sleep [7], sleep-dependent memory consolidation [9] and cortical development [8]. Many studies have been conducted to better characterise the sleep spindles and in particular to understand the differences age and gender may entail on the spindle characteristics. It was found that spindle number, density and duration are lower in the elderly compared to the young adults [10], a result which is confirmed with the patients of this study.

2.2 Different detection methods

2.2.1 Using Wavelet convolution

A wavelet is a wave-like oscillation with an amplitude that begins at zero, increases, and then decreases back to zero. also known as a wave packet in physics. Morlet's wavelet has the particularity of being designed to have the optimal properties for detecting spindles-like activity as it has the shape of a sleep spindle. (see figure 2.2). This method convolves the Morlet's wavelet with the EEG signal. If a real spindle is present in the EEG signal, it is going to be multiplied by the spindle-like wavelet, thus resulting in a very high amplitude signal. Applying a threshold to the resulting amplitude allows to detect the spindles. Methods such as RAY2015 [20] and UCSD - University of California, San Diego (unpublished) use convolution.

2.2.2 Using detection thresholds

A second more common and more precise way of detecting sleep spindles is through the use of thresholds. The goal of this is to progressively identify by elimination which parts of the EEG signal compose sleep spindles. Many methods use thresholds, for instance YASA [13], Moelle 2011 [18], FAAST2 [19] each make use of thresholds based on the value of the root mean square (RMS). They follow a different detailed protocol but with the same structure.

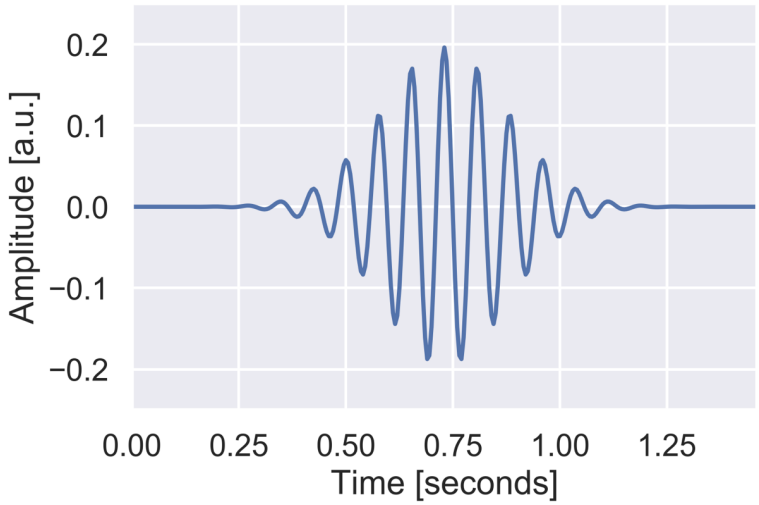


Figure 2.2 Morlet's wavelet

1. Detect signals within the sigma frequency range (11-16 Hz).
2. Compute the Root mean square (RMS) of the detected signals using an adjustable window size and step.
3. Compute the RMS threshold: $RMS_{thresh} = RMS_{mean} + 1.5 * RMS_{std}$
4. Spindles are detected whenever $RMS > RMS_{thresh}$
5. Only the spindles lasting between 0.5s and 3s are retained.

The core structure may include some more complex or detailed steps depending of the algorithm. For instance step 1 can either be achieved by applying a simple band pass filter. Or, by making use of a Short Term Fourier Transform (STFT) to detect whenever the signal has a relative power in the sigma frequency range ≥ 0.2 to ensure that the increase in sigma power is actually specific to the sigma frequency range and not just due to a global increase in power (e.g. caused by artefacts).

2.3 Our detection

2.3.1 Choosing the algorithms

The wavelet convolution was not accurate enough as the wavelet did not permit to detect a satisfying proportion of the "real" spindles since spindles are not as regular in shape, frequency and oscillations as we would have hoped. We then considered creating multiple different wavelets with slightly different characteristics to cover a wider range of spindle shape in the aim of detecting more.

However this method was deemed too complicated and we decided to turn towards other methods.

When looking into the methods using thresholds we brought our attention to Yet Another Spindle Algorithm (YASA) published in 2018 [13]. YASA applies three different thresholds: relative σ power to detect signals within the sigma frequency range, correlation to detect spindles that are visible on the raw EEG signal by requiring a high correlation between raw EEG signal and the filtered sigma burst. And lastly root mean square threshold to detect increase of energy in the EEG signal. This algorithm is of particular interest as it not only offers a very precise detection but also outputs many interesting parameters concerning the spindles for instance the peak frequency of the spindle or the number of oscillations in the spindle.

In a will to offer the most precise study we decided to use multiple detection methods. The domain expert directed me to a paper he had written in 2018 [14], which bench marked multiple sleep spindles algorithms by comparing their precision, recall and f-score. The study compared seven published algorithms: Moelle11, Fe07, Nir11, Ray15, FASST, mar13, La18 and 2 unpublished algorithms, the UCSD algorithm (University of California based on wavelet analysis) and CONCORDIA. It concluded that Moelle 2011 gave the highest accordance with the ground truth detected by the sleep expert. Thus I include Moelle 2011 in this study (abbreviated to Mo). Moelle 2011 follows the basic structure described above while offering the possibility to adjust all parameters (highcut and lowcut for the band pass filter, window size and step, min duration and max duration for the spindle).

Lastly, the sleep expert had previously used the application BrainRT to conduct other studies on his data. This application detects the different brain waves and in particular the sleep spindles. This spindle detection was added to the study and called in Violet (related to the colors used by the application BrainRT). The detection uses thresholds on the amplitude of the signal rather than the frequency thus offers a complementary view with respect to the 2 previous methods.

2.3.2 Tuning hyper parameters

The domain expert detected spindles on one file: EDFMANUELX4. This detection is extremely time consuming as the sleep recording is over 8 hours long while spindles last between 0.5 and 2 seconds which explains we only had one such annotated file. The domains expert's detection is used as the ground truth to tune the parameters of the three methods used.

2.3.3 Merging spindles

The initial data is composed of 6 channels. Moelle 2011 and Violet detect spindle on each channel separately and did not include a way of merging these results. Indeed, by analysing these detection we noticed many spindles overlapped in time, thus computing the spindle density per 30 seconds by simply summing the spindles found on each channel would have been error prone. Instead we decided on a merge rule presented in figure 2.3 where the red line represents the duration of the merged spindle. The merge rule decided was the one used by YASA and was added to the code of Moelle 2011 and Violet so that all 3 methods could be in accordance. Note, we use density per 30 seconds for historical reasons. Indeed before the use of computers, the scoring of sleep was done on paper. One paper corresponded to 30 seconds, thus 30 seconds is considered as an epoch for sleep scoring.

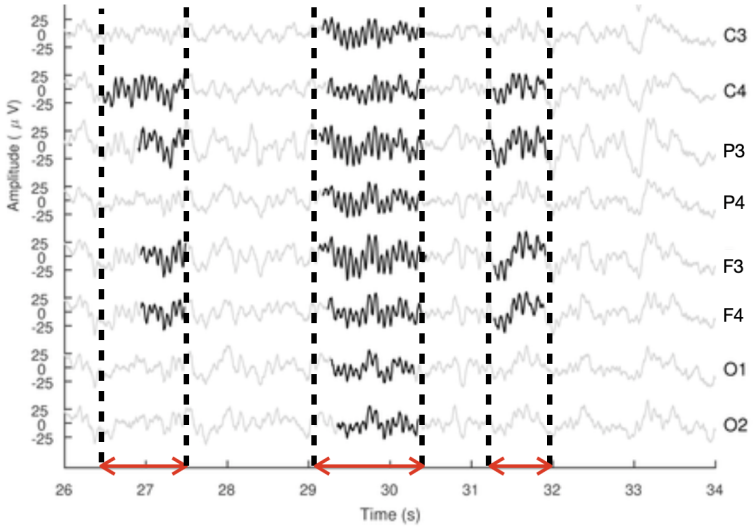


Figure 2.3 Spindle merging rule

Chapter 3

Creating the data set

There exists major challenges of dealing with real world data sets. For instance having to rely on experts to understand the data, deciding how to deal with missing values or having to arrange data in common file format to be use by machine learning methods. Here, we had for each patient a 3GB EEG file, the goal is to extract the important features relating to sleep spindles (density, duration, frequency) to obtain tabular data that we could input into machine learning models.

3.1 File organisation

For each patient we had two csv files :

1. The EEG recording which consists of over 16 000 000 points for each of the 6 channels.This csv had been previously extracted from the original edf using the python module mne [5].
2. An annotation file evaluated by a sleep expert identifying the sleep phases.

Thus for each patient 2 files were necessary. Simply having these files, in the right directory, under the correct format, both with the matching name was a hassle. Many files were missing, or had not been exported correctly. Moreover, the filenames had been entered by hand by the medical staff at the hospital and thus featured many typos. We had to identify these and correct them in order to be able to use all the files. The difficulties encountered emphasised the need of standardizing recordings in hospitals so as to make the data usable for Machine Learning tasks.

3.2 Feature Extraction

3.2.1 In numbers

- The study has 267 patients aged 18 to 76 of whom 56% are Female.
- In total we have 746.4 hours of N2 sleep, that is on average 10140 seconds per individual corresponding to 169 minutes per patient
- In total we have 247.4 hours of N3 sleep, that is on average 3298 seconds per individual corresponding to 55 minutes per patient.

More details concerning the spindle data may be found in the appendix in figure 2

3.2.2 Working with phase segments

For each patient we first set out to obtain the 9 features seen in figure 3.1 over their entire N2 and N3 sleep. This turned out to have 2 issues. First of all this meant we had quite a small data, as we were limited to 534 observations (2 per patient, one for N2, one for N3 sleep with 267 patients) . Secondly this hid a lot of information as we were averaging values over multiple N2 or N3 segments. Indeed during the night, a patient enters multiple times in N2 or N3 sleep, these N2 and N3 segments which can last anywhere from 30 to 3000 seconds.
Thus we decided to not only extract the features over all of N2 or N3 sleep, but also to extract them for each segment. This allowed to have many more samples as there are on average more than 10 N2 sleep segments per patient. With this, we added one feature to the data set: the duration in seconds of each segment (seg dur). It is important to note that the overall density over N2 or N3 is not the average of the segment densities since the density is measured with respect to the duration of each segment, hence it is a weighted average.

File	mo dens	mo av dur	violet dens	violet av dur	seg dur	yasa dens	yasa av dur	yasa av freq	yasa av osci nb
ALAL591004_1	1.581395	0.886432	0.976744	0.659722	1290.0	0.046512	0.656250	12.966834	7.750000
ALAL591004_4	2.000000	0.879232	0.750000	0.629296	360.0	0.083333	0.752604	13.809815	8.833333
ALAL591004_7	2.016667	0.811435	1.900000	0.632812	1800.0	0.016667	0.546875	13.530694	7.200000
ALAL591004_8	1.781250	0.865063	1.750000	0.634364	960.0	0.062500	0.761719	12.895143	9.625000
ALAL591004_9	2.200000	0.892116	1.880000	0.648011	750.0	0.120000	0.713728	13.196670	8.642857
\LAL591004_12	2.266667	0.754969	1.783333	0.685571	1800.0	0.050000	0.605078	13.317776	7.500000
ALAL591004	2.030405	0.818519	1.496622	0.638580	8880.0	0.040541	0.701282	13.218796	8.255119

Figure 3.1 Feature extraction for patient ALAL591004 in phase N2. Files with underscore correspond to N2 segments, missing segments were dropped due to missing values.

3.3 Checking for errors

Errors can occur due to an error in the code or a misunderstanding of the data. It is therefore of extreme importance to check the coherence of the results obtained. Units for instance can be a source of mistake as the duration and time in the original data are given in microseconds while the data we extract uses seconds.

A key method to spot errors is through visualisation, this is indeed how an error was found in the code extracting the features using the Violet algorithm. We plotted the densities in decreasing order, this gave the plot in figure 3.2. Clearly this led to notice that some densities were much too high. Indeed, the densities should be no bigger than 10 spindles per 30 seconds. Moreover, the

densities should be more or less the same amongst all segments while here we see huge disparities (standard deviation amongst N2 segments of patient ELSO850228 - first patient to the left of the x axis in figure 3.2 - was of of 317,62)

This error was due to a misunderstanding of the annotation file, which caused some N2 segments to last the entirety of the recording rather than the found duration. The code was fixed and we obtained the results in 3.3 which are indeed much more coherent.

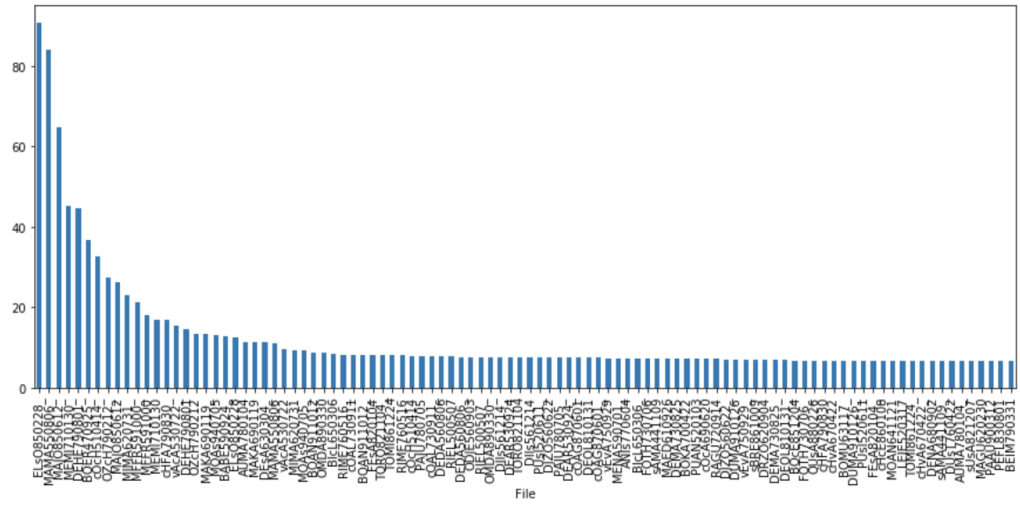


Figure 3.2 Densities found by Violet in decreasing order, in y axis the density of spindles per 30 seconds, in x axis N2 segments with their file name

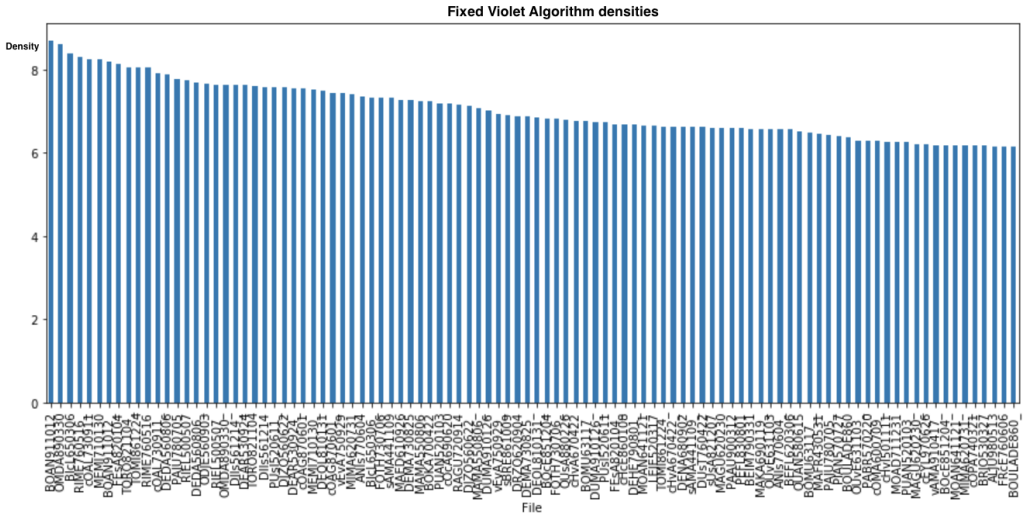


Figure 3.3 Densities found by fixed Violet, in y axis the density of spindles per 30 seconds, in x axis N2 segments with their file name

Chapter 4

Sleep Disorder: Paradoxical Insomnia

Patients who have this relatively uncommon form of insomnia have a tendency to underestimate sleep duration and overestimate wakefulness relative to polysomnographic measures. This misconception often hides another mental disorder such as depression or anxiety. The diagnosis of Paradoxical Insomnia (PI) allows to uncover and treat these mental disorders. Using machine learning would accelerate and reduce the cost of the diagnosis.

We show that it is not possible to predict PI using the spindle data collected. Moreover, using the MMPI-2 (detailed in chapter 5) data we are able to give a slightly more accurate diagnosis, but including the spindle data to the latter prediction does not improve its accuracy confirming the fact that the spindle characteristics - extracted as we did in this paper - cannot improve the diagnosis of Paradoxical insomnia.

4.1 Background

4.1.1 Types of insomnia

Paradoxical insomnia, otherwise known as sleep state misperception, is the reporting of severe insomnia without objective evidence of sleep disturbance or significant impairment of daytime function. Note that there exists a rarer type of PI where the patient believes he sleeps well whilst the quality of his sleep is less than good. However, there are extremely few cases like this since in this case the patient does not come to see a sleep expert. Thus we do not consider it in the study. Another more common type of insomnia is Psychophysiological insomnia. A patient who suffers from this becomes anxious that they are not going to get enough sleep and frets about how a lack of sleep is going to affect their next day.

4.1.2 Link between Insomnia, Depression and Anxiety

To illustrate the importance of being able to diagnose properly paradoxical insomnia we use the study [15] conducted on 63 patients with Paradoxical Insomnia (PI), 63 patients with Psychophysiological insomnia (PsyI) and 63 normal sleepers (NS) from southwest China. The authors were able to extract the following conclusions: both Psychophysiological insomnia and Paradoxical in-

somnia patients had significantly higher anxiety and depression than NS. Moreover, compared to PsyI patient, PI patients had slightly lower anxiety and significantly higher depression.

4.1.3 Diagnosing Paradoxical Insomnia

PI diagnosis is composed of 2 measures evaluating the time slept by the patient. The first measure is evaluated by the patient himself using a sleep log he completes for the duration of 1 week. The second is evaluated by the sleep expert's analysis of the polysomnographic and actimetric data. Indeed to proceed to the diagnosis the patient must undergo one night of polysomnographic measurements and wear an activity monitor for a week. Thus the first is subjective while the second is objective. When the difference in the time slept per night given by the 2 measures is above one hour, the patient is considered to have paradoxical insomnia.

4.1.4 Utility of using machine learning to predict Paradoxical Insomnia

The utility of using machine learning would be to be able to diagnose PI while cutting down on all this costs. Both time and money. It is important to understand that the goal here is not to predict the second factor, indeed we attempt at predicting the composition of the 2 factors. Distinguishing this type of insomnia from the others is important as the treatment will be different.

4.2 Paradoxical Insomnia (PI) Label

For each of our 267 patients, the domain expert had proceeded to the diagnose PI in the traditional manner explained above. We were thus able to add a binary PI target label to our data set for each of our 267 patients. Our data set contains exactly 50% of each label.

- 1: patient suffers from PI
- 0: patient does not suffer from PI

4.3 Predicting PI using spindle data

4.3.1 Missing Values

YASA records significantly less spindles than Moelle11 and Violet. This causes to question how missing values should be handled. Through many tests we found that the best results are obtained:

- by dropping the missing values rather than filling them with the value 0.
- by merging the detection from all three algorithms to have a data set with 9 features rather than simply using each algorithm separately to avoid having missing values.

4.3.2 Visualization

We first set out to visualise our data in the goal of seeing if any particular clusters or outliers could be detected. As our data set is composed of 9 features naturally we used dimension reduction techniques. Recall the 9 features: Mo density (/30s), Mo average duration(s), Violet density (/30s), Violet average duration(s), YASA density (/30s), YASA average duration(s), YASA average frequency (Hz), YASA average number of oscillations and Segment duration(s).
Firstly we employed principal component analysis, the results are found in 4.1.

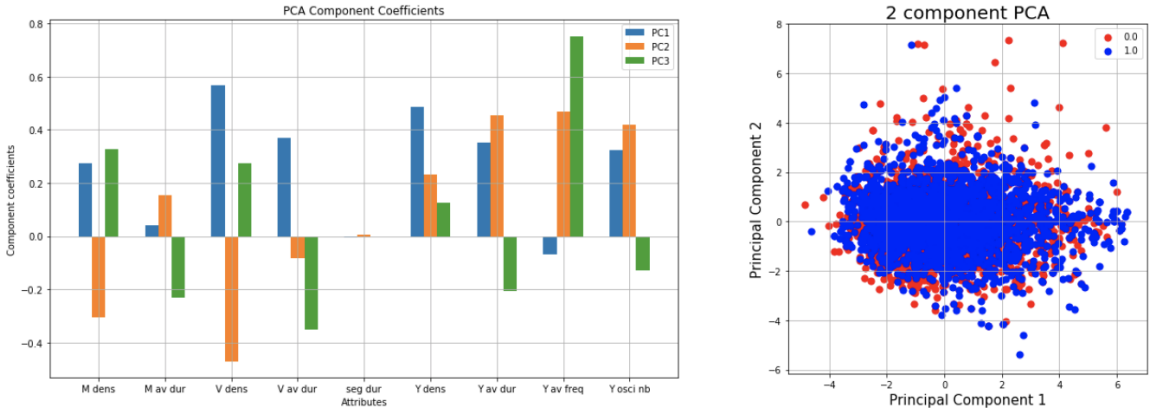


Figure 4.1 Principal Component Analysis, data used: 6874 rows ×9 columns

On the left of figure 4.1 we may see that the feature "seg dur" has little to no importance. This feature was included in the aim of serving as a weight to give more importance to longer segments clearly the machine leaerning algorithm did not use it as such. As seen on the right of figure 4.1, PCA is not enough to have a clear visualisation of the data, this motivated the choice of using t-SNE. t-SNE is a method that presents many advantages with respect to PCA mainly that unlike PCA, t-SNE is not limited to linear projections, which makes it suited to all sorts of dataset.

The left plot of figure 4.2 gives the same conclusion as the PCA: the data cannot be grouped into target classes after dimension reduction. However working on a more restricted data set (6874 × 2) focusing only on the spindle detection given by Moelle 2011 we were able to obtain clearer groups. (right of figure 4.2).

Naturally we thought to cluster these points to see if the clusters had any particular characteristics. We use Spectral clustering, agglomerative clustering and K-means clustering. K-means gave the most satisfying results. However when looking at the characteristics of each cluster (right of figure 4.3) there are no difference in the density or duration of the spindles.

As this type of visualisation and clustering was inconclusive we brought our attention to logistic regression to have another type of visualisation. Using keras we built a neural net whose second

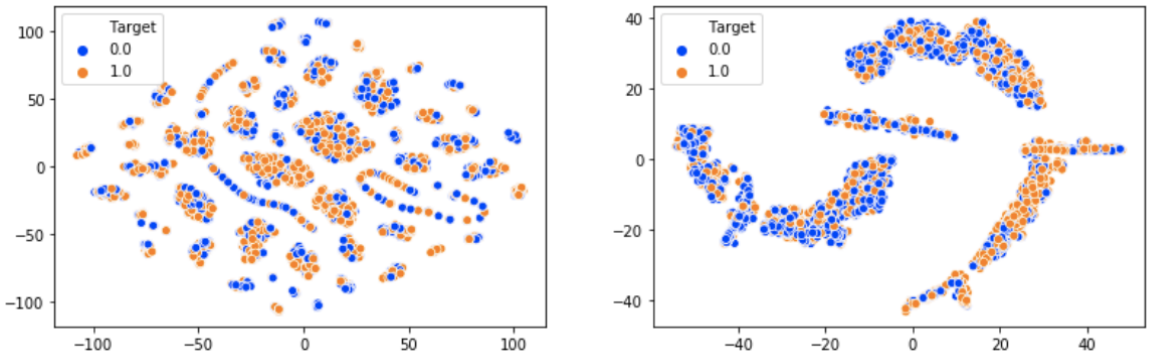
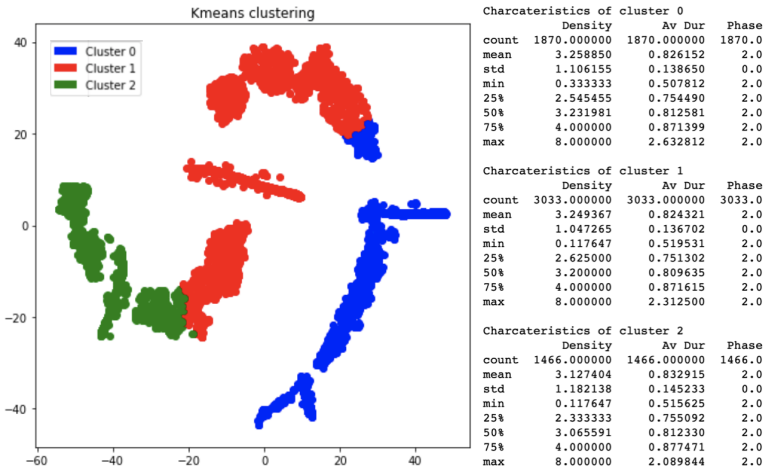


Figure 4.2 TSNE (left) for the entire data set with perplexity=20, (right) for data only from Moelle2011 with perplexity=150



Charcateristics of cluster 0

	Density	Av Dur	Phase
count	1870.000000	1870.000000	1870.0
mean	3.258850	0.826152	2.0
std	1.106155	0.138650	0.0
min	0.333333	0.507812	2.0
25%	2.545455	0.754490	2.0
50%	3.231981	0.812581	2.0
75%	4.000000	0.871399	2.0
max	8.000000	2.632812	2.0

Charcateristics of cluster 1

	Density	Av Dur	Phase
count	3033.000000	3033.000000	3033.0
mean	3.249367	0.824321	2.0
std	1.047265	0.136702	0.0
min	0.117647	0.519531	2.0
25%	2.625000	0.751302	2.0
50%	3.200000	0.809635	2.0
75%	4.000000	0.871615	2.0
max	8.000000	2.312500	2.0

Charcateristics of cluster 2

	Density	Av Dur	Phase
count	1466.000000	1466.000000	1466.0
mean	3.127404	0.832915	2.0
std	1.182138	0.145233	0.0
min	0.117647	0.515625	2.0
25%	2.333333	0.755092	2.0
50%	3.065591	0.812330	2.0
75%	4.000000	0.877471	2.0
max	8.000000	2.089844	2.0

Figure 4.3 K-mean clustering of figure 4.2 (right) with n=3 clusters and the description of each cluster

to last layer was composed of 3 nodes. We trained this net then upon testing instead of taking the probability given by the last layer we took the values of the nodes of the second to last layer which we plotted in a 3D graph. However since logistic regression did not yield good result for predicting PI the 3D graph was unable to give more information on the data. (these graphs may be seen in the appendix in figure 3).
Hence we conclude that it is not possible to obtain more information through visualisation and dimension reduction.

4.3.3 Classification

Multiple classification methods were tested: Random Forest (RF), K- Nearest Neighbours (KNN), Support Vector Classifier (SVC) and Logistic Regression (LR).

Hyper parameters were adjusted and multiple cross validation techniques were tested. How-

ever, no test gave a mean accuracy of above 52%, indeed this result is clearly illustrated in the Cumulative gains curve of figure 4.4. This curve is an evaluation curve that assesses the performance of the model for a binary problem and compares the results with the random pick. We may see that the orange and blue curve barely lift off from the black baseline proving the prediction is no better than random. Hence, the sleep spindle data - detected in the manner described in this paper - cannot predict accurately Paradoxical Insomnia.

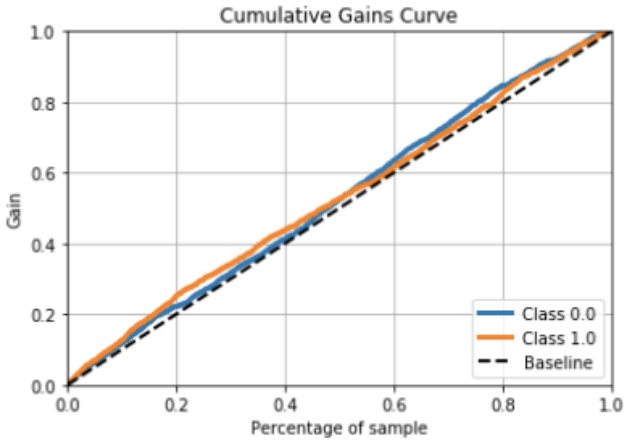


Figure 4.4 Cumulative Gain Curve, Random Forest Classifier using spindle data to predict Paradoxical Insomnia

Going further we attempted to see if associating the spindle data to another data set could give a more accurate diagnosis. We thus used the MMPI-2 data (described in the following chapter) to predict Paradoxical Insomnia. This gave an accuracy of above 50%. However when adding the spindle data to the MMPI-2 data, the accuracy was not improved. Proving detecting spindle - in the manner described in this paper - cannot be useful for the diagnosis of PI.

4.4 Conclusion

Then we reach the same conclusion as in [11] which stats that sleep spindles characteristics are not predictive of sleep misperception. We had not included this paper in initial hypothesis as due to the small number of patients (approx 50), the authors claimed the need to replicate the study to be able to confirm these findings. Here we confirm these findings.

Chapter 5

Predicting Tendencies Toward Psychopathological Traits

Using the patient's answers to the MMPI-2, the questionnaire creates scales evaluating certain psychopathological tendencies for each patient. We focus our attention on trying to see if these are reflected in the patient's sleep spindles. To uncover if sleep spindles could be a bio marker used for the diagnosis of particular psychopathological traits, we use the spindle data to predict particular scales of the MMPI-2 data.

We show that in all cases spindle data does not allow to predict the patient's tendency towards a psychopathological trait. This offers positive results concerning the patient's privacy. However we also show that for more severe psychoses such as Paranoia the spindle data does allow to accurately eliminate the possibility of the psychosis from the diagnosis. This result can help the doctor in refining his diagnosis and reach a more exact diagnosis more rapidly.

5.1 Diagnosing psychopathological traits

Psychopathology is the scientific study of mental disorders. We use the terms psychophathological traits rather than mental illnesses as in the medical community the fact of diagnosing mental illnesses through the use of physiological data and computers is not yet accepted. Indeed today, the diagnosis of psychopathological traits relies on patient reports and clinician observations, rather than on objective biological tests. This is what constitutes the difficulty of their diagnosis. There exist no test as for a same disorder patients may express and report different symptoms thus no generalisation can be made. Without a proper and complete diagnosis the treatment proposed cannot be entirely effective and thus the help provided to the patient cannot be fully beneficial.

The goal of this chapter is to try and discover if through the analysis of physiological data (sleep spindles) some psychopathological traits may arise. Should this hypothesis be validated, the outcome would not be that patients could simply be diagnosed by computers; on the contrary any finding would aim at assisting doctors and facilitating, accelerating and rendering more precise their diagnosis so as to better orient the patient for a more suited treatment.

5.2 MMPI-2 features

The domain expert gave us access to the interpreted scores of the MMPI-2. Meaning that the features used are scales that are determined from the answers of the 500+ true/false questions. This data consists of 85 un-correlated columns, and can be one of the followings:

- an information concerning the patient (Ex: age, gender)
- a score from 0-100 assessing the patient's tendency towards a psychopathological trait.(Ex: Hypochondria, Extraversion, Depression, Anxiety).
- a score from 0-100 assessing a personality trait in relation with a psychopathological trait (Ex: the statement I am scared of dying in my sleep is linked to Anxiety)
- a score from 0-100 assessing how truthfully the patient is taking the test (Ex: Lying).

5.3 Definitions

We give a short definition of the psychopathological traits explored in the classification. We divide in 2 groups psychotic and non-psychotic traits. This distinction is chosen as it changes fundamentally the medical treatment.
Psychotic disorders are severe mental disorders that cause abnormal thinking and perceptions. People with psychoses loose touch with reality. Two of the main symptoms are delusions and hallucinations. Delusions are false beliefs, such as thinking that someone is plotting against you or that the TV is sending you secret messages. Hallucinations are false perceptions, such as hearing, seeing, or feeling something that is not there.
In non-psychotic traits we find two main categories: Anxiety and Depression.

Non-psychotic traits:
• **Hypochondria** is the fact of being abnormally anxious about your health.
• **Depression** is a mood disorder that causes a persistent feeling of sadness and loss of interest.
• **Somatisation** is the manifestation of psychological distress by the presentation of physical symptoms. Somatisation is an expression of anxiety.

Psychotic traits:
• **Paranoia** is a thought process heavily influenced by anxiety or fear, often to the point of irrationality and delusion. It is characterized by delusions of persecution, unwarranted jealousy, or exaggerated self-importance.
• **Schizophrenia** is a condition in which people interpret reality abnormally.

5.4 Classification

The psychopathological traits are presented as a score from 0-100. This score is based on the answers to the questions of the MMPI-2. To these scores we apply a cutoff of 65. Indeed a scale is clinically significant if it has a score of above 65, this cutoff is a common practice for papers using the MMPI-2 [21], [22].
This allows to have a binary problem (label 1 for scores above 65, 0 otherwise), however this cutoff does not in all cases create a class balance thus the F1 score is the preferred measurement to interpret the results.

	Hypochondria	Depression	Somatisation	Paranoia	Schizophrenia
Mean F1 score for label 0	0.478	0.657	0.514	0.818	0.822
Mean F1 score for label 1	0.614	0.528	0.557	0.226	0.139

Figure 5.1 Mean F1 scores over 10 folds for the prediction of 5 psychopathological traits

5.5 Interpretation and Conclusion

We can separate the interpretation of figure 5.1 into two groups. First the non-psychotic traits: Hypochondria, Depression and Somatisation. Through the F1 scores for labels 0 and 1 revolving around 0.5, we may conclude that sleep spindles cannot help to draw conclusions on these psychopathological traits. In truth this is a positive results as it respects the patient's privacy. This preserves the importance of seeing a mental health expert.

Second, the psychotic traits: Paranoia and Schizophrenia which are characterised low F1 scores for label 1, but particularly high F1 scores for label 0. The particularly low F1 scores for label 1 signify that the sleep spindles cannot accurately predict if a patient suffers from these psychotic traits. However, the high F1 scores for label 0 illustrate that the sleep spindles are able to accurately eliminate the possibility of the patient suffering from these psychoses. This is a very interesting result as it could help domain experts in their diagnosis as they could eliminate these possibilities to better focus their diagnosis. It is interesting to note that we can interpret differently for psychotic and non-psychotic traits, this leads to question whether the link between the psychopathological traits and the characteristics of the sleep spindles is proportional to the severity of the trait. Indeed psychotic traits are considered to be more severe diagnosis.

Chapter 6

Limitations and Further Works

6.1 Limitations

This study is conducted on a relatively small data set composed of 267 patients. Moreover these patients had consulted the domain expert for sleep related issues. Hence, these patients do not represent a broad enough population. The study should be conducted on patients from a wider background to be properly generalised. Moreover, not all parameters were taken into account for instance patient's medical treatments were not recorded whereas they may affect sleep spindles.

The spindle detection has its limits as the three detection algorithms used gave drastically different measures for the density and duration of the spindles. However they conserved the order, meaning that if patient A had a higher density than patient B evaluated by Moelle 2011 then YASA would also show a higher density of spindles for A than B. Then these values can be used within the study but any given value may not be extracted and interpreted. For instance stating that female patients have an average spindle density of 3.5 is false but saying that female patients have a higher spindle density than men may be true.

The third and most impacting limitation is the coronavirus outbreak. Indeed, our work was interrupted. Thus we were not able to add more patients to the study and chapter 5 is not as detailed as we would have wished.

6.2 Further works

Further works would include adding more patients to the study, possibly from different clinics. Moreover it could be interesting to focus on detecting the spindles on each channel individually rather than as a whole to see if we can deduce further conclusion by knowing which part of the brain generated the spindles.

Lastly the difficulties in creating the data set shows the complexity of using recorded medical data for machine learning tasks. In the coming decades huge progress must be done towards standardising data recording and making a link between those who record the data in the medical

facilities and those who exploit them so as to be able to use medical data in machine learning. Indeed applying ML to the medical field requires collecting clean, complete, usable data, rare access to such data limits the progress ML can bring to the medical field. It is for this reason that the medical discipline in which machine learning has brought the most is imaging as there are standard procedures for recording images.

Bibliography

[1] Purcell, S. M., et al. "Characterizing sleep spindles in 11,630 individuals from the National Sleep Research Resource." *Nature communications* 8.1 (2017): 1-16.

[2] Plante, D. T., et al. "Topographic and sex-related differences in sleep spindles in major depressive disorder: a high-density EEG investigation." *Journal of affective disorders* 146.1 (2013): 120-125.

[3] Shibagaki, Masamitsu, Sigehiro Kiyono, and Kazuyoshi Watanabe. "Spindle evolution in normal and mentally retarded children: a review." *Sleep 5.1* (1982): 47-57.

[4] D'Agostino, Armando, et al. "Sleep endophenotypes of schizophrenia: slow waves and sleep spindles in unaffected first-degree relatives." *npj Schizophrenia* 4.1 (2018): 1-8.

[5] A. Gramfort, M. Luessi, E. Larson, D. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, M. Hämäläinen (2013) *MEG and EEG data analysis with MNE-Python, Frontiers in Neuroscience* Volume 7, ISSN 1662-453X, [DOI]

[6] Hathaway, Starke Rosecrans, and John Charnley McKinley. "Minnesota Multiphasic Personality Inventory; Manual, revised." (1951).

[7] Steriade, Mircea. *Neuronal substrates of sleep and epilepsy.* Cambridge University Press, 2003.

[8] Khazipov, Rustem, et al. "Early motor activity drives spindle bursts in the developing somatosensory cortex." *Nature* 432.7018 (2004): 758-761.

[9] Fogel, Stuart M., and Carlyle T. Smith. "The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation." *Neuroscience Biobehavioral Reviews* 35.5 (2011): 1154-1165.

[10] Crowley, Kate, et al. "The effects of normal aging on sleep spindle and K-complex production." *Clinical neurophysiology* 113.10 (2002): 1615-1622.

[11] Normand, Marie-Pier, Patrick St-Hilaire, and Célyne H. Bastien. "Sleep spindles characteristics in insomnia sufferers and their relationship with sleep misperception." *Neural plasticity* 2016 (2016)

[12] NO Dylan Buckley (December 2019) *What Is The MMPI Test And What Does It Say About You?.*

[13] Vallat, R. "YASA (yet another spindle algorithm): A fast and open-source sleep spindles and slow-waves detection toolbox." *Sleep Medicine* 64 (2019): S396.

[14] Pallanca, Olivier, Sammy Khalife, and Jesse Read. "Detection of sleep spindles in NREM 2 sleep stages: Preliminary study benchmarking of algorithms." 2018 *IEEE International Conference on Bioinformatics and Biomedicine (BIBM).* IEEE, 2018.

[15] Liao, Jingping, Shichao Zhu, and Xiaolin Li. "anxiety and depression in paradoxical insomnia: a case–control study." *Neuropsychiatric disease and treatment* 14 (2018): 231.

[16] Gibbs, Erna L., and Frederic A. Gibbs. "Extreme spindles: correlation of electroencephalographic sleep pattern with mental retardation." *Science* 138.3545 (1962): 1106-1107.

[17] Adams, C. L., E. L. Gibbs, and F. A. Gibbs. "Asynchronism of Electrical Activity of Frontal Lobes During Sleep: A Late Sequel of Frontal Lobotomy." *AMA Archives of Neurology Psychiatry* 77.3 (1957): 237-242.

[18] Mölle, Matthias, et al. "Fast and slow spindles during the sleep slow oscillation: disparate coalescence and engagement in memory processing." *Sleep* 34.10 (2011): 1411-1421.

[19] Leclercq, Yves, et al. "fMRI artefact rejection and sleep scoring toolbox." *Computational intelligence and neuroscience* 2011 (2011).

[20] Ray, Laura, et al. "Expert and crowd-sourced validation of an individualized sleep spindle detection method employing complex demodulation and individualized normalization." *Frontiers in human neuroscience* 9 (2015): 507

[21] Keller, Laura Sue. *Assessment of chronic pain patients with the MMPI-2*. Vol. 2. U of Minnesota Press, 1991.

[22] Greene, Roger L. The MMPI-2: *An interpretive manual*. Allyn Bacon, 2000.

[23] American Sleep Disorders Association. "The international classification of sleep disorders." *Diagnostic and coding manual* (1990): 1990.

[24] Villani, Cédric, et al. "Donner un sens à l'intelligence artificielle." (2018).

APPENDIX

Scale	Items	Description
Hysteria	60	There are five aspects monitored: shyness, cynicism, neuroticism, poor physical health, and headaches.
Psychopathic Deviate	50	Looks at the type of social maladjustment the individual might have, as well as whether they have strong pleasant experiences. It will also take a look at complaints or problems the individual has with family and other authority figures, as well as their level of boredom, social alienation, and self-alienation.
Hypochondriasis	32	Evaluates vague or nonspecific complaints that pertain to the functioning of the body. These are focused on the abdomen and back and tend to persist despite having negative medical tests
Depression	57	Measures the level of clinical depression within the tested individual. This is done by looking at the individual's morale, a lack of hope in the future, and any level of dissatisfaction that an individual may have with their life in general.
Masculinity/Femininity	56	Looks at activity versus passivity, personal sensitivity, interests, hobbies, and aesthetic preferences. In general, it looks at the rigidity of the individual's conformity to stereotypes for masculine or feminine activities and roles.
Paranoia	40	It gauges interpersonal sensitivity as well as suspiciousness and even the morality of self-righteousness that an individual has.
Psychasthenia	48	Tests the inability of an individual to resist a specific type of action or thought even if they may be maladaptive. It looks at abnormal fears, difficulty in concentration, guilty feelings and thoughts or self-criticisms in its view of obsessive-compulsive behaviors.
Schizophrenia	78	Evaluates strange thoughts and perceptions, poor relationships with the family, problems concentrating, problems with impulse control, social alienation, disturbing questions on self-worth and self-identity and even lack of interests and sexual problems.
Hypomania	46	Looks at more mild degrees of elated but unstable moods, psychomotor excitement, flighty thoughts and ideas and overactivity in general, including irritability and egocentricity.
Social Introversion	69	Looks at how introverted or extroverted the individual is. It looks at limited social skills, preference for being alone versus with others and whether the individual does well with a group of friends or not.

Figure 1 Description of the 10 scales in the MMPI which assess 10 major categories of abnormal human behaviour. Items indicate the number of question relating to this scale. [12]

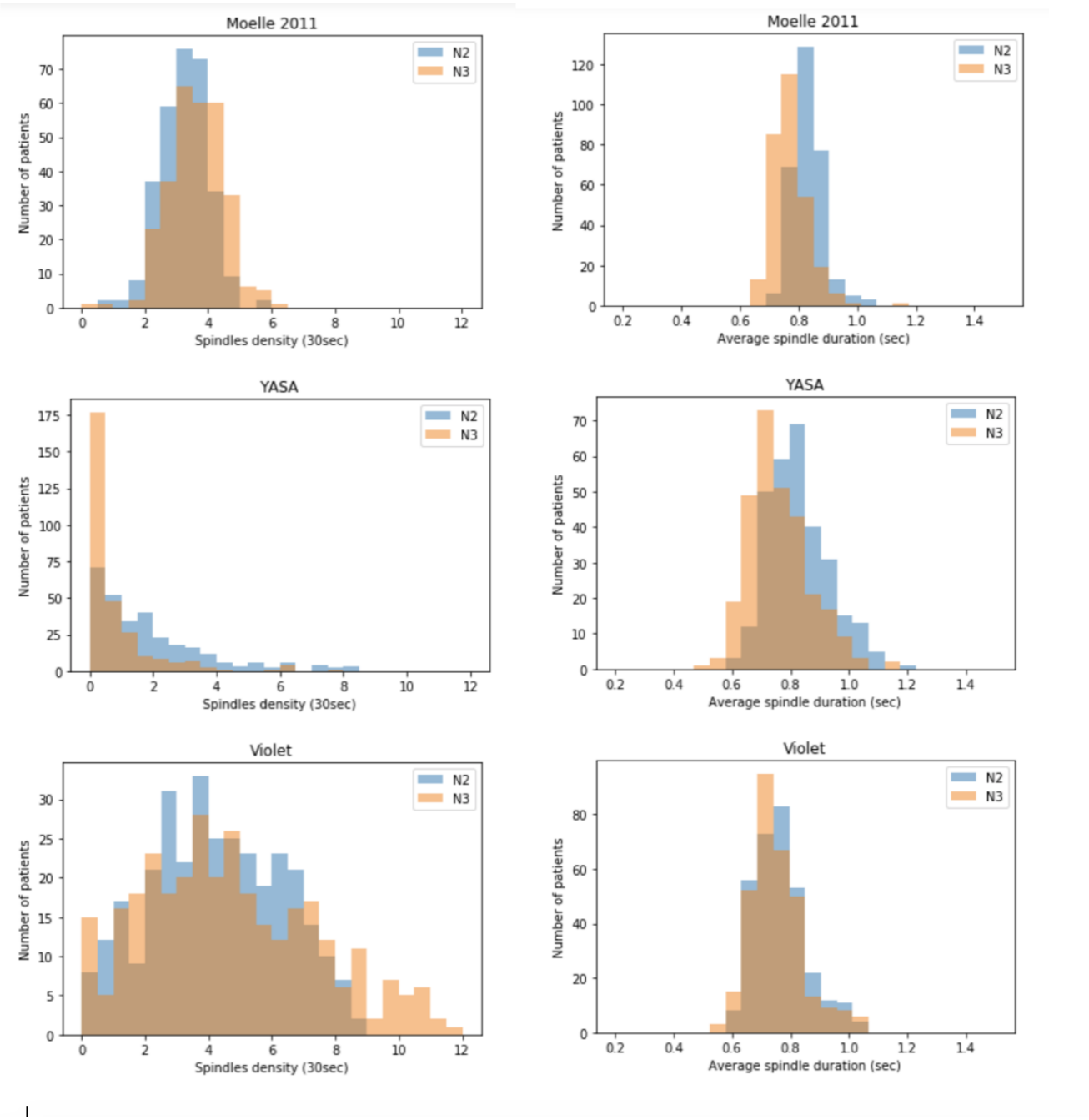


Figure 2 Distribution of spindles density and average duration for the algorithms YASA, Moelle2011 and Violet the 267 patients

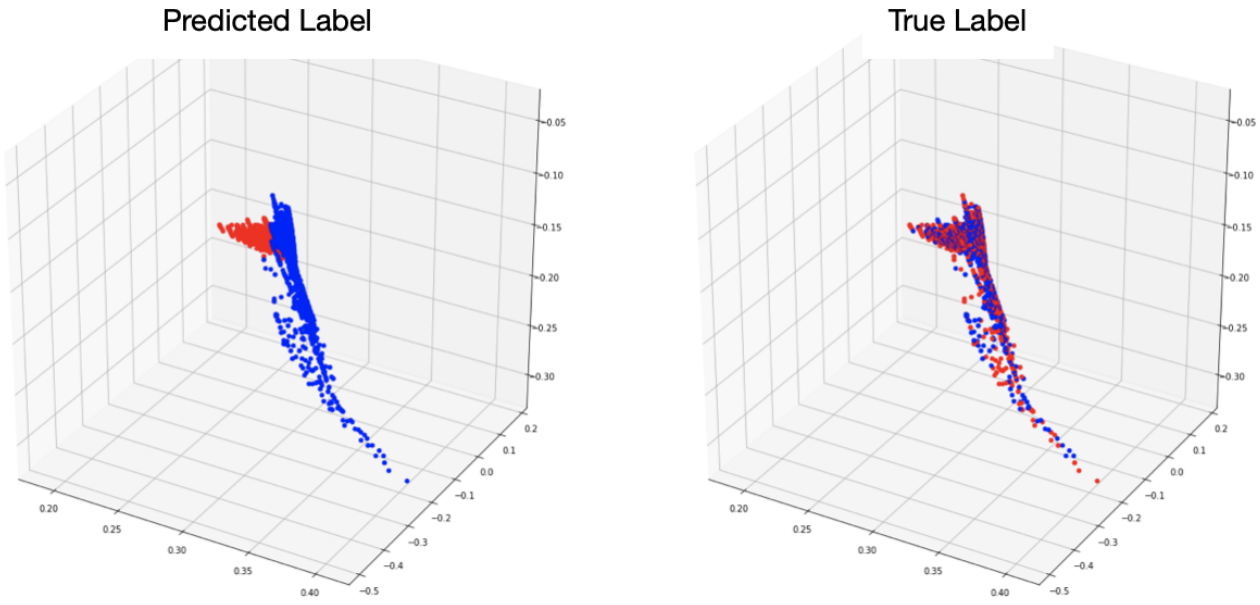


Figure 3 3D plot of the values of the three nodes on the second to last layer of the neural network for Logistic Regression. Left is the plotted points colored with the label predicted by the Logistic Regressor, on the right the points are colored with their true label.