Detecting Sleep Apnea from raw physiological signals by Dreem

Challenge data Report

Représentations Parcimonieuses

MVA, 2021

NAIT SAADA Thiziri
Télécom Paris
thiziri.naitsaada@telecom-paris.fr

TORREZ PÉREZ Claudia
Télécom Paris
claudia.torresperez@telecom-paris.fr

Abstract

Lack of training data and need of hand labelling by experts are crucial challenges in fields such as medicine or genetics where experiments to generate them can have a high cost and be time consuming. To deal with this issue, Neural Networks are powerful tools to classify samples following a distribution that is underlying by the training data. Since their introduction by Lecun et al. *[7], the research community have opened up Neural Networks to a very large bunch of possible applications, from classification of images, videos to time series.*

In the context of our challenge, Dreem, a French start-up, has collected data coming from eight different psychological channel signals. The objective is to infer whether or not the subject is going through an episode of apnea, given these eight inputs. This task reveals to be extremely challenging for different reasons. The first one is that these data come from sensors, that might not be as precise and robust as we would wish. While sleeping, the subject is likely to move, and the sensor to be affected by these changes. As we noticed, several sample signals were noisy, which is why we performed a careful denoising task. A second challenge is to find a suitable criterion to minimize while performing the training of our model when dealing with highly imbalanced data as sleep apnea is rare compared to normal sleep.

In this paper we propose a model combining the feature extraction and classification abilities of a Convolutional Neural Network, with a more interpretable Graph Signal Classifier that makes use of the domain knowledge of the specific changes of behavior of EEG signals during sleep apnea events.

1. Introduction

Sleep Apnea is one of the most common breathing-related sleep disorders, that concerns equally men and women. It is also known to be a disorder that is commonly under-diagnosed whereas it can result in dangerous unhealthy behaviours when coupled with risks of heart failure. Episodes of sleep can be characterized by repetitions of some patterns in physiological signals, that can be affected and/or deteriorated by an apnea. On the one hand, detecting a period of Sleep Apnea from physiological signal recordings requires a sufficient expertise in order to be able to localize abrupt changes of the signals from the its original trend. Note that even for experts, the diagnosis task can end up being particularly difficult when considering that these small repeated patterns in sleep signals is strongly dependent on the patient that is under study. Some estimate it can take around an hour for an expert to manually detect these periods of sleep apnea. On the other hand, the recent surge of interest in Machine Learning Models has given rise to many applications, one of them being healthcare. Tools from Computer Vision has gradually tend to be used in other fields to perform automatic tasks, such as detection in medical Time Series. This turns out to be very effective when a sufficient amount of data is provided.

1.1. Motivation

In this challenge, Dreem, a French neurotechnology startup, has collected several physiological signals from sleep episodes, where periods of sleep apnea and non apnea have been manually labelled. Each record is composed of 90 seconds of 8 physiological signals corresponding to, respectively : Abdominal contraction, Respiratory Airflow from the subject, Cardiac activity, Record Thoracic contraction, Snoring indicator, O_2 saturation of the blood and two EEG (electroencephalogram signals) derivation signals. Each signal is sampled at 100 Hz. The goal of this study is to present methods that can be used in order to automatically detect sleep apnea events within a signal of 90 seconds, given these 8 corresponding input signals. If this task is performed correctly, one could easily imagine the saving cost of such diagnoses. By making them more affordable and efficient, such methods could help detect the numerous

undetected cases.

1.2. Related work

Many investigations have been pursued to try to detect apnea from psychological signals based on Deep Learning. Among them, DeepSleepNet [11] and its recent extension TinySleepNet [12] are examples of models based on Deep Learning tools, that try to capture meaningful features of the input data from stacking multiple convolutional layers. Another way to go has been explored by adapting the famous Computer Vision model U-Net [9] to time series data, so-called U-Time [8]. Many other works have designed their own architecture model, which is the result of computational expensive hyperparameters tuning. In [4], which has been published by Dreem, the presented model, DOSED (Dreem One Shot Event Detector) relies on a loss that is designed to suit better the tackled problem of predicting events inside a 90 seconds signal, which consists in finding the center and duration of each sleep apnea period.

In addition, there have also been some works that tackle the sleep apnea detection problem by using common signal processing and machine learning tools. [2] focuses only on the respiratory signals as a way to detect Obstrusive Sleep Apnea Events. They developed a voice detection tool to distinguish silences in respiratory signals from normal breathing. On the other hand, [13] and [10] used only EEG signals whose behaviour is harder to interpret when looking at the raw signal captured by the PSG, but which is essential to diagnose sleep apnea. Both of them extracted the EEG subband signals using a bandpass filter, and computed, for each of them, statistical and signal features such as energy and entropy. [13] then created a dataset with those features that was fed to an SVM and an ANN in order to classify signals as normal sleep or apena events. KNN was also used for that same matter. [10] also used SVM and KNNs to classify sleep signals, but the only features that were used were the inter-band energy ratio. However, both of these papers focus on the classification of entire signals instead of the detection of sleep apnea events of variable lengths inside a signal.

1.3. Methodology

The aim of this presentation is to design a model that makes the most of psychological experts knowledge and investigations. Indeed, we regret the lack of interpretability of the deep learning previous approaches, excepted for DOSED, which has tried to adapt their approach to their setting. In section 2, we will present the processing of the data that has been performed. Section 3 presents two models we have developed in order to perform the detection task. The first one, based on Deep Learning tools, tries to capture the meaningful features of the data when the second one tries to take benefit from expertise knowledge on EEG. The final model is a combination of both. We will present our methodology and general implementation. Section 4 shows the results we obtained from this model.

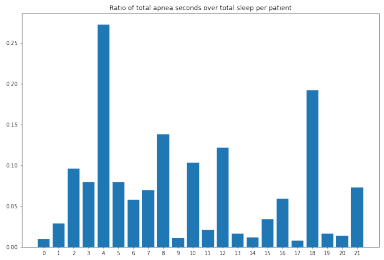


Figure 1: The classification task is highly imbalanced

2. Features extraction

2.1. Exploring the data

The first part of our investigation was to explore and have a better understanding of the input data, as well as the apnea events.

2.1.1 Distribution of apnea events

As mentioned previously, this task is extremely difficult because of the subject-dependency of the feature signals as well as the distribution of apnea events. This enforces us to treat each patient individually in the exploration of the data but we should keep in mind that our final objective is to catch meaningful features that discriminates apnea from nonapnea for all the patients. In 1, one can see first that for barely all patients, the ratio of duration of apnea over the duration of total sleep is extremely low, meaning that the classification is highly imbalanced between labels corresponding to an apnea compared to normal sleep.

On the other hand, in 2, we highlight the complexity of the problem which is to find a general automatic detector on any patient whereas his associated apnea distribution is specific to this subject. Therefore, the idea is not to try to recognize some general patterns shared by all patients going through apnea episodes but to try to catch a difference in the behavior of the input signals when doing an apnea. This change of behaviour is believed to be shared among any patient, hence legitimating this challenge and this idea of automatic detection for any patient.

2.1.2 Statistical features of the inputs

Our first intuition has been to see whether it could be possible to recover the class of the input given the statistical features of each of the eight signals. There seems not to be any clear and simple enough relation between these statistical features (among the mean, median, variance, maximum,

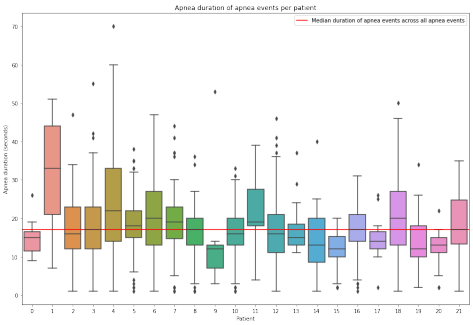


Figure 2: Distribution of apnea events duration over patients is subject-dependent.

minimum) and the class the signal is associated to to make such a feature extraction. Moreover, too many information would be lost if considering statistical features instead of the eight signals themselves. Please see the boxplots in **??** to see the statistical differences computed on each of the eight input signals of normal sleep and apnea events.

2.2. Denoising the data

Before the signals were used to train any models, it was important to preprocess them with the goal to remove the irrefutable noise that was present in them, with the most substancial source of noise being the captors. Since the eight signals come from captors placed in different parts of the subject's body, who is prone to moving through the night, the signals can be polluted with the captors' reactions to sudden movements and changes of position which may result in outliers and signal change points. The signals are also polluted with the unavoidable noise that accompanies each captor and that appears through the different stages of data storage, procession and transmission.
In order to recover as much of the original signals as possible, we decided to use Singular Spectrum Analysis (SSA) on each signal separately. This method consists on computing the Singular Value Decomposition of each signal's x trajectory matrix X on a window of length N_w with an overlap of $N_o = N_w - 1$. The trajectory matrix $X \in \mathbb{R}^{N_w \times (N-N_w+1}$ contains on each row i the signal's samples from i to $N - N_w + i$: $(x[i] \dots x[N - N_w + i])$, with the resulting column j being composed of the signal's samples from j to $N_w + j$: $(x[j] \dots x[N_w + j])^T$. The second step of SSA is to analyze the singular value distribution and form groups of singular values that correspond to similar phenomenon as to decompose the trajectory according to these groups. SSA is useful to denoise noisy signals where the signal-to-noise ratio is reasonable, meaning that the pure signal is dominant when compared to the noise. Thus, the pure signal is associated to the trajectory matrix' largest singular values while the noise corresponds to the smallest ones.

Our method consisted on computing the SVD of each signal, looking at the singular values' distribution and then reconstructing the pure signal by taking only the vectors that are associated to the highest singular values. For a window of length 14, we found that in all of the cases, the first singular value was enough to represent the original signal yet in some dimensions the second singular value's contribution was substantial too. However, the remaining 12 values only composed the remaining noise in the signal. Figures 3 and 4 show the singular values associated to each vector on a given channel of the entire training set.

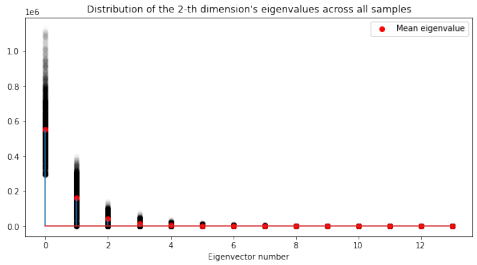


Figure 3: Mean singular values of the SVD of each signal on the third channel (Cardiac Activity) of the data set.

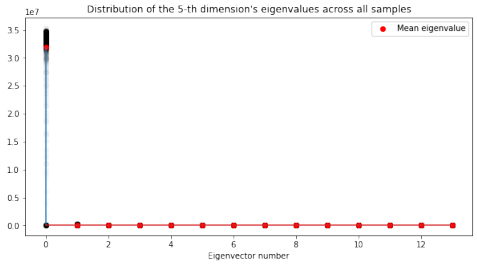


Figure 4: Mean singular values of the SVD of each signal on the sixth channel (O2 saturation of the blood) of the data set.

Figure 5 shows an example of the signals we obtain when using SSA as a denoising tool. The denoised signal appears smoother, with fewer high frequency variations. The denoising steps becomes essential when exploring the data statistics, as some essential statistical features, such as the mean, are not invariant to outliers.

2.3. EEG Analysis

The input data can be divided into two parts :

- Abdominal contraction, Respiratory Airflow from the subject, Cardiac activity, Record Thoracic contraction, Snoring indicator, O_2 saturation of the blood
- The two EEG derivations : EEG1 and EEG2

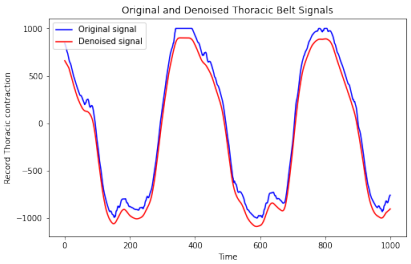


Figure 5: Zoom into a Record Thoracic contraction signal and its denoised version obtained with SSA. The denoised signal is shifted downwards to make the difference between both signals clearer.

We decided to perform a special processing to the EEG signals, that are known from experts to behave distinctly depending on the state of the subject (apnea or non apnea). Indeed, during sleep, low frequency bands in EEG signals contain a higher percentage of energy when compared to higher frequency bands, however, this behavior changes to its opposite during sleep apnea events, caused by the receptors' reaction to lack of O_2 and the alarm they send to the brain to start breathing. Strictly speaking, from a neurologist expert point of view, it makes sense to try to detect apnea based on these two features, on which we will have a very particular look at.

2.3.1 Spectrogram and Mel-Spectrogram of EEG

In 6, 7, we can see that neither the FFT transform coefficients nor the continuous wavelet transform coefficients seem to be directly and obviously correlated with a simple function to the output classification. Thus, we decided to explore further reading on the topic.

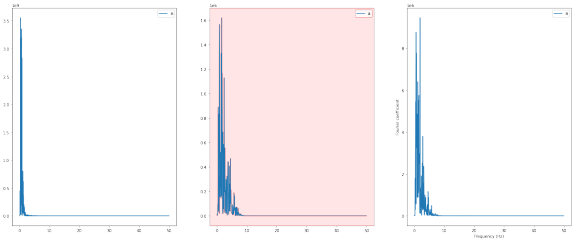


Figure 6: Spectrogram computed on windows of 5 seconds for three successive periods of non apnea(white background), apnea (pink) and non apnea (white)

2.3.2 Proposed feature for EEG : inter band energy

The electroencephalogram (EEG) is a recording of the electrical activity of the brain from the scalp. The recorded

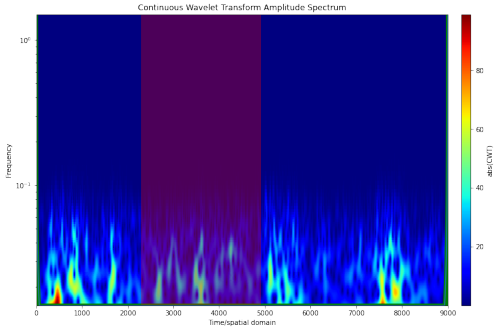


Figure 7: Mel spectrogram for three successive periods of non apnea (white background), apnea (pink) and non apnea (white)

Figure 8: Example of EEG1 signal

waveforms reflect the cortical electrical activity. Many studies have shown that the response of this kind of signals to a certain filter can reflect whether the subject is doing an apnea or not. Indeed, each sleep phase corresponds to the emergence of typical frequencies in the brain. As a result, EEG signal is partitioned into five frequency bands as explained in [3]:

- Delta waves : 0-4 Hz
- Theta waves : 4-7 Hz
- Alpha waves : 8-12 Hz
- Sigma waves : 12-16 Hz
- Beta waves : 16-40 Hz

Defining the energy of the p-th band of a signal x as :

$$E_p = \sum_{n=1}^{N} x_p[n]^2$$

and the ratio $p - q$ as the ratio of the energy of the p-th band over the energy of the q-th band :

$$R_{p,q} = \frac{E_p}{E_q}$$

in [10] the authors proposed an inter-band energy ratio feature defined as the ratio of the energy of the signal response to these different filters, meaning that the signal is encoded by these new features : $(R_{i,j} \forall i \neq j \in \delta, \theta, \alpha, \sigma, \beta)$. More precisely, the authors observed that only a few of these ratio were sufficient to detect sleep apnea : $R_{\delta,\theta}, R_{\delta,\alpha}, R_{\delta,\sigma}, R_{\delta,\beta}, R_{\theta,\beta}$.

We found quite interesting to rely on such expertise on EEG to build a classification model. More precisely, we

want here to make the most of this observation to build a model which is consistent with this approach. To do so, we decided to build two models in parallel, one based on the whole input data and a second one based on graphs that rely on these new introduced features of the EEG signals. Further details will be given in next section.

3. Models

As we believe that there is not any simple relation between these features and the classification output, we decided to implement a Convolutional Neural Network that will try to capture the meaningful features of the whole bunch of eight signals and relate them to their class, as it seems to be encoded by a complex function. On top of that, we would like to help the algorithm in its task by completing its predictions with a classification that will be performed based on EEGs. In Machine Learning, such combination of learning is called inductive bias. The final proposed model is an ensembling model between the deep learning classifier and the one based on graphs. The final prediction would result in a weighted average of both prediction models, where the weight λ should be tuned :

$$p^{\textit{final}} = \lambda p^{CNN} + (1-\lambda) p^{\textit{graphs on EEG}}$$

3.1. Convolutional Neural Network

The architecture we chose is inspired from U-Time ([8]), with a deep architecture of 8 convolutional layers, whose increasing and then decreasing width makes it look symmetric. We used Max Pooling layers as well. All the details can be found in 1. The first layer is a linear layer that enables us to downsample the input data with a weighted average, whose weights are also learnt. The input data have shape $(N, 8, 90, 100)$, where N denotes the batch size and the output shape is $(N, 2, 90)$, corresponding to the probabilities of belonging to each of the two classes (apnea or non apnea). Taking the argument maximum of these two probabilities gives us the predicted class. We can consider that each input window of 90 seconds is treated as an image, but the difference with traditional Computer Vision architecture here is that each pixel (second) of the image (window) is assigned to a class. Somehow, it looks very similar to Object Localization: apnea events corresponding to the object that needs to be detected in the image. This has motivated our choice of architecture. To prevent from overfitting, we used Dropout, and we chose Adam optimizer with a weight decay to perform the minimization of the loss. The choice of the loss is discussed in the following section.

3x1 Linear Layer ReLU (channels=8, padding=1)
Max Pooling layer
3x1 Convolution Layer ReLU (channels=8, padding=1)
Max Pooling layer
3x1 Convolution Layer ReLU (channels=16, padding=1)
Max Pooling layer
3x1 Convolution Layer ReLU (channels=32, padding=1)
Max Pooling layer
3x1 Convolution Layer ReLU (channels=64, padding=1)
Max Pooling layer
3x1 Convolution Layer ReLU (channels=128, padding=1)
Max Pooling layer
3x1 Convolution Layer ReLU (channels=64, padding=1)
3x1 Convolution Layer ReLU (channels=32, padding=1)
Max Pooling layer
3x1 Convolution Layer ReLU (channels=16, padding=1)
1x1 Convolution Layer ReLU (channels=2, padding=0)

Table 1: CNN architecture 5-16-32-64-128-64-32-16

3.2. Choice of the loss

3.2.1 Weighted Cross Entropy Loss

We naturally chose the Cross Entropy Loss as a criterion for this binary classification task. However, recalling that the two classes are highly imbalanced, we corrected it artificially by strongly penalizing the errors made on the rare class. More precisely, we computed the ratio of positive samples over negative samples and we corrected the loss with this weight as shown below :

$$\mathcal{L}_{CE}(y, p^{pred}) := -(y \log(p^{pred}) + (1-y)\log(1-p^{pred}))$$
$$\mathcal{L}_{CE}^{weighted}(y, p^{pred}) := -(\alpha_1 y \log(p^{pred}) + \alpha_2 (1-y)\log(1-p^{pred}))$$
$$\text{where } \alpha_1 = \frac{|\textit{negative samples}|}{|\textit{positive samples}|}$$

3.2.2 Differentiable F1 score

In addition to this, we used another loss which is directly derived from the F1 score. Recall the F1 score is non differentiable and computes a score between the binary outputs and the true class. Instead of considering binary outputs (which is essentially the cause of the non differentiability of the loss), we consider the probabilities. For instance, if the ground truth is 1 and the model prediction is 0.3, we calculate it as 0.3 true positive and 0.7 false negative. If the ground truth is 0 and the model prediction is 0.4, we calculate it as 0.6 true negative and 0.4 false positive. This yields to a differentiable loss on which gradient descent can be performed and that we believe to be best suited to maximize the final objective, which is the F1 score, than the cross entropy.

3.3. Graphs on EEG

The idea behind the separation of the 8 feature signals into two groups (the EEG apart) is that it leads us to consider a multivariate series and not a multi-modal one as it was before this split. Indeed, now that we have *homogeneous* data, a classification task based on graphs and smoothness makes sense. Note that it would not have made any sense to build a graph with nodes corresponding to the 8 initial channels (Abdominal contraction, Respiratory Airflow from the subject, Cardiac activity, Record Thoracic contraction, Snoring indicator, O_2 saturation of the blood, EEG1 and EEG2) as the notion of smoothness on it would need to be redefined.

Thus, to overcome this multi-modal issue, we decided only to consider an EEG classifier based on graphs. The nodes of this graph are set to be the 5 most interesting inter-band energy defined previously in section 2.3.2. In the end, we have 5 nodes for each EEG, so a total of 10 nodes in the graph. Therefore, each signal can be fully described (regarding its 2 EEGs only) with this graph and is said to be a graph signal. Many works have been investigated on classification of EEG signals based on graphs and our work gets its inspiration from some of them. For example, [1] performs graph signals classification in the same spirit. To define the edges between these 10 nodes, we derived two methods that have been developed in the literature. The first one, ad hoc, is based on the computation of the correlation between these nodes. If this correlation exceeds a certain threshold (in absolute value), the edge between both nodes is set to this value, otherwise it is set to 0. We also tried to find the graph's Laplacian by solving a convex optimisation, following a method described in [5].

Two graphs are built, each corresponding to class apnea and non apnea. Then, to perform classification of a new signal, we compare its smoothness on both graphs. A graph signal is said to be smooth on a graph if it fits well the graph in terms of its Laplacian. In other words, there is a homogeneity or a continuity in each node's values. On the other hand, a graph signal is not considered smooth on a graph if such homogeneity is no longer present. This would mean, for example, that connected nodes have highly different values.

As we discussed earlier, one can observe inter-band energy behavior changes on EEG signals during the apnea events and the transitions that enter and exit it, with higher frequency bands having a bigger energy proportion during apnea events. Thus, we expected the normal sleep graphs and apnea graphs to have structures. We predict the class whose learnt graph minimizes the smoothness of the new incoming signal. This process is done considering a sliding window of size 500 on the signal of size 9000, this choice being made as to capture the features on a reasonable duration. The overlap of this window is set arbitrarily between 10 and 100, with the goal to capture also the features of the transition between normal sleep and an apnea event, and viceversa. We believe that if these inter band energy ratios are different enough between the apnea and non apnea classes, their underlying structure and thus the topology of their associated graph should differ from a class to another.

4. Results

4.1. Final Performance

Please find in 10 the training and validation loss decrease of our model while training. We can see in **??** that it is correlated to the increasing of the objective, the F1 score, which confirms our intuition of loss choice. The dropout is set to 0.2, Adam optimizer's parameters are chosen such that the learning rate equals to $1e^{-}3$ and the weight decay to $1e^{-}3$. Batch size is set to 64.

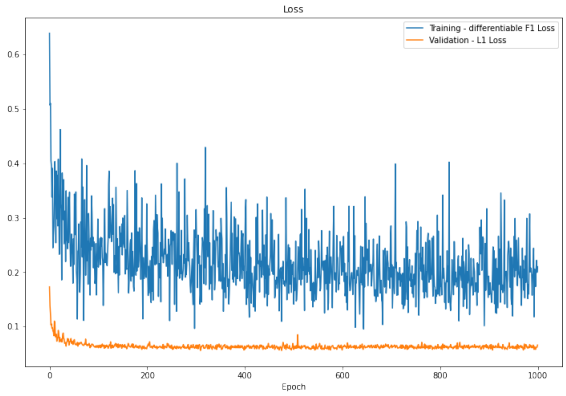


Figure 9: Losses during training

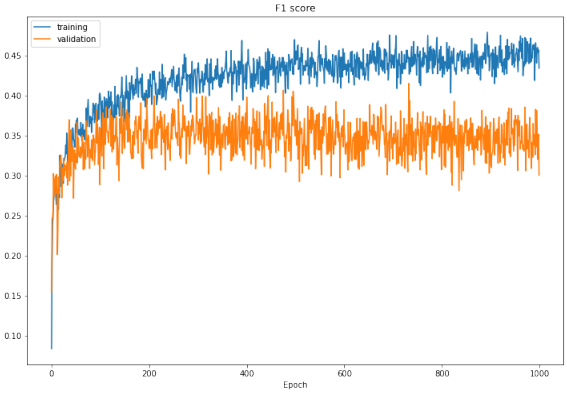


Figure 10: F1 score during training

The performance of our final CNN model are reported in 2.

Regarding the model based on graphs that should classify signals from their EEG channels, the results given are not really satisfying. Still, performing ensembling methods

Training set	Validation set	Testing set
0.444	0.415	0.282

Table 2: F1 score of our model

between both models seems to alienate the performance of the first classifier. Therefore, we decided not to retain this approach as part of our final model. We try to analyze the reasons of this unsuccessful classification from the graph model in the following section.

4.2. Discussion

The new introduced features inter band energy extracted from EEGs are said to contain significant features in the detection of apnea episodes. However, we have realized that on this particular data we were given, it was not the case as shown in 13, **??**. Therefore, the built graphs were not really discriminating signals of sleep apnea and normal sleep. In reality, both graphs from sleep apnea and normal sleep classes were barely identical. So does the notion of smoothness. Further work in that direction could be interesting. Especially because of the arise of efficiency of Graphs Neural Networks, combining these two tools could help improve our model performance. Further investigation could be performed on the use of Hidden Markov Models (HMM) to estimate the apnea event from the EEG signals. This model is commonly used in time series prediction (see [6] for instance). In addition, a final post processing step could be considered in order to enforce the predictions to be consistent with the real duration of an apnea episode. As shown in 2, its median duration is around 18 seconds. Therefore considering transition probabilities between each state, given a current state and coupled with the common duration of an apnea, would be of interest. Strictly speaking, we noticed that the label data were concatenation of blocks of 0 or 1 (with a mean size of 18 for blocks of 1). There is a kind of continuity between them. Thus, these transition probabilities would encourage the predictions to be consistent with realistic apnea disorder.

5. Conclusion

We have decided to design in parallel two models and to combine their predictions. The first one, based on Deep Learning, has enabled us to capture meaningful characteristics from a multi-modal time series. To feed the second one, some homogeneity in the inputs were required to use legitimately tools from graph theory. To do so, we followed trained sleep expertise that highlights the importance of EEG signals in the sleep apnea detection task. We introduced an inductive bias by extracting these features from the signals and we tried to perform classification based on it. Unfortunately, this last method has not revealed to be satisfying as we realized that the data were not really consistent with experts recent studies on the topic and the introduced features revealed not to be very suited to the problem.

Even though we had limited computational resources and so finding best tuned parameters was already difficult, what we found the most difficult in this task was the lack of interpretability of the features that were retained by the Neural Network to perform the classification. Because we wanted our final model to make the most of expertise knowledge, we extracted from EEG signals some features that are believed to discriminate both classes and added this inductive bias using a graph classifier model. A last point that needed to be addressed on is the multi-modal aspect of the time series, which prevented us from building a graph directly on all the channels but only on the EEG signals in order to preserve the sense of smoothness, as defined in graph theory.

References

[1] H. B. Ahmed, D. Dare, and A. O. Boudraa. Graph signals classification using total variation and graph energy informations. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 667–671, 2017. 6

[2] L. Almazaydeh, K. Elleithy, M. Faezipour, and A. Abushakra. Apnea detection based on respiratory signal classification. *Procedia Computer Science*, 21:310–316, 10 2013. 2

[3] M. R. Azim, S. A. Haque, M. S. Amin, and T. Latif. Analysis of eeg and emg signals for detection of sleep disordered breathing events. In *International Conference on Electrical Computer Engineering (ICECE 2010)*, pages 646–649, 2010. 4

[4] S. Chambon, V. Thorey, P. J. Arnal, E. Mignot, and A. Gramfort. Dosed: a deep learning approach to detect multiple sleep micro-events in eeg signal. *arXiv preprint arXiv:1812.04079*, 2018. 2

[5] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst. Learning laplacian matrix in smooth graph signal representations, 2016. 6

[6] M. Hanif, F. Sami, M. Hyder, and M. Ch. Hidden markov model for time series prediction. *Journal of Asian Scientific Research*, 7:196–205, 01 2017. 7

[7] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. 1

[8] M. Perslev, M. H. Jensen, S. Darkner, P. J. Jennum, and C. Igel. U-time: A fully convolutional network for time series segmentation applied to sleep staging, 2019. 2, 5

[9] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. 2

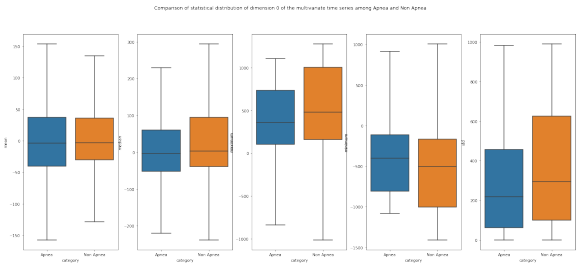
[10] S. Saha, A. Bhattacharjee, and S. A. Fattah. Automatic detection of sleep apnea events based on inter-band energy ratio obtained from multi-band eeg signal. *Healthcare Technology Letters*, 6, 04 2019. 2, 4

[11] A. Supratak, H. Dong, C. Wu, and Y. Guo. Deepsleepnet: a model for automatic sleep stage scoring based on raw single-channel eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(11):1998–2008, Nov 2017. 2

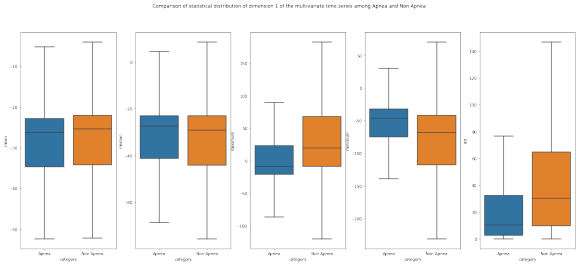
[12] A. Supratak and Y. Guo. Tinysleepnet: An efficient deep learning model for sleep stage scoring based on raw single-channel eeg. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 641–644, 2020. 2

[13] V. Vimala, K. Ramar, and M. Ettappan. An intelligent sleep apnea classification system based on eeg signals. *Journal of Medical Systems*, 43, 01 2019. 2

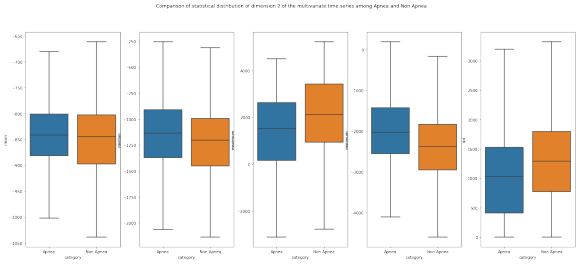
A. Statistical features of the input data



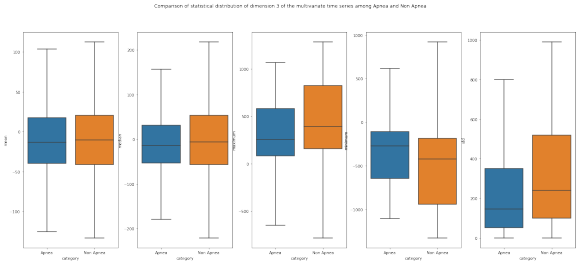
(a) First channel : Abdominal contraction



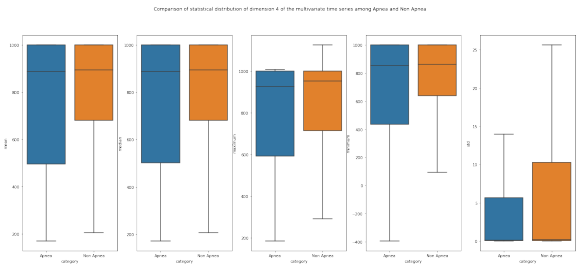
(b) Second channel : Respiratory Airflow from the subject



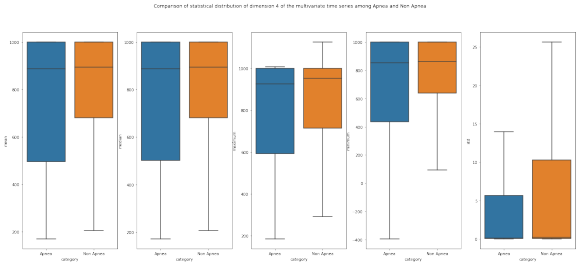
(c) Third channel : Cardiac activity



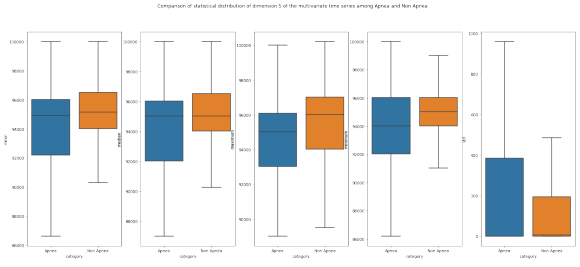
(d) Fourth channel : Record Thoracic contraction



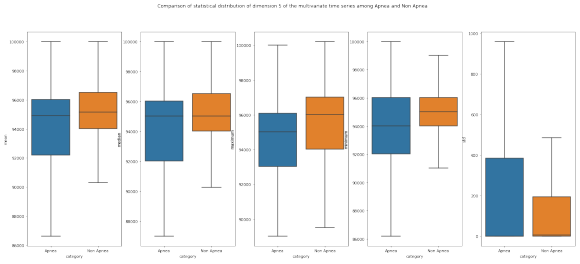
(a) Fifth channel : Snoring indicator



(b) Sixth channel : O2 saturation of the blood



(c) Seventh channel : EEG1 derivation



(d) Eighth channel : EEG2 derivation

Figure 11: Some statistics from each of the eight channel composing the multi-modal input time series. From left to right : mean, median, maximum, minimum, standard deviation.

B. Statistical features of the inter band energy extracted from the dataset

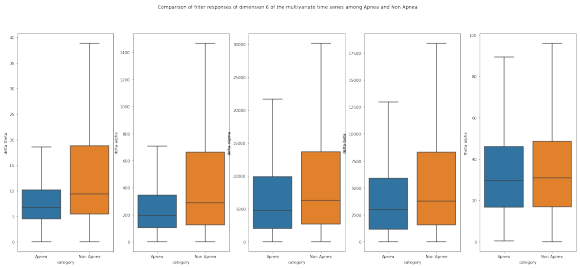


Figure 12: EEG 1 - Distribution of the 5 most important inter band energy features.
On this dataset, they seem not to be discriminating both classes

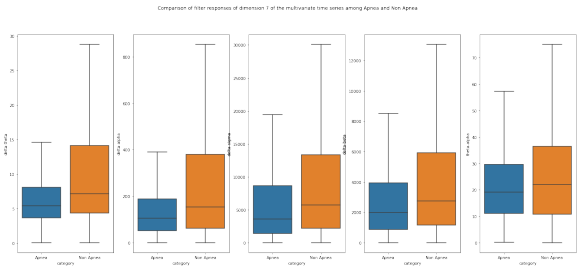


Figure 13: EEG 2 - Distribution of the 5 most important inter band energy features.
On this dataset, they seem not to be discriminating both classes