Research Project 2016/–

Issued: 12/2016

Prediction of reported sleep quality based on physiological markers as measured with PSG

Mariam Zabihi

Philips Research Europe

Unclassified

Author's address: Mariam Zabihi
                    m.zabihi@philips.com

Title: Prediction of reported sleep quality based on physiological markers as measured with PSG

Author(s): Mariam Zabihi

Supervisor(s): Tim Weysen, Ad Denissen, Boris De Ruyter

Project: Sleep Research

Keywords: Research Project

Abstract: In sleep research, the relationship between objective and subjective scores of sleep quality is still an open question. In particular, sleep patterns change naturally during pregnancy. In this study, we analyze sixty three nocturnal sleep data from 39 women, pregnant in their second and third trimester, collected in an earlier study at Philips. The data consists of various polysomnographic measures as well as Groningen Sleep Quality Scale (GSQS) questionnaire evaluation. A combination of two sets of hand-crafted and EEG features directly extracted by convolutional neural network (CNN) has been used to find subjective relevant features in the data.

Conclusions: A LASSO analysis of the first set of hand-crafted sleep parameters yields to two predictors: N1 and sleep efficiency (SE) with 0.61 correlation to the subjective scores and three predictors for the second set of sleep parameters: the number of transition from REM to N2, SWS to N2 and N2 to N1 with 0.51 correlation value. The obtained results confirm previous findings. Moreover, CNN was able to predict the binary GSQS from raw EEG with 0.79 accuracy.

Contents

1 Introduction

1.1   Background

One of the interesting questions in sleep research is how the relationship between objective and subjective sleep scores can be addressed. Usually, objective sleep measurements are derived from the recorded body responses during sleep (e.g. EEG, heart rate, body position), while subjective measure is referred to the self-reported sleep experience [Rosipal et al., 2013; Argyropoulos et al., 2003; ÅKERSTEDT et al., 1994].

To understand the humans' sleep architecture, a typical hypnogram of a healthy adult is shown in figure 1.1[1](National Institutes of Health). The relevant terminology and definition of human sleep stages as recommended by [AASM and Iber, 2007] follows:

- Stage W (wakefulness) represents the waking state before or after subject experiences other sleep stages. Usually the corresponding EEG signals contain alpha rhythms (8-13 Hz).
- Stage N1 (NREM - non-rapid eye movement - 1 sleep) is a transition from wakefulness to sleep. The brain signal contains mixed frequency while the alpha rhythms amplitude has decreased.
- Stage N2 (NREM 2 sleep) mainly contains theta wave (4-7 Hz).
- Stage N3 (NREM 3 sleep) represents slow-wave sleep (SWS) and contains delta wave (0.5- 2 Hz). It is also known as deep sleep. This stage was previously considered to be two different stages: stage 3 and stage 4.
- Stage R (REM sleep) has the same frequency characteristics as stage W. It is assumed most of dreams occur in this stage.

The sleep architecture is the key to derive and analyze the objective measures of sleep [Rosipal et al., 2013].

Table 1.1 reviews several sleep studies that investigated various kinds of objective measures and subjective scores in multiple target groups. Some of these studies illustrate the relation between self-reported stress scores and objective sleep measurements. [Åkerstedt et al., 2007] reported chronic stress negatively affects the quality of sleep by increasing slow wave sleep (SWS) latency and wake time after sleep onset (WASO). [Hall et al., 2000] have shown that chronic stress increases the beta power while life acute stress

---

[1]https://science.education.nih.gov/supplements/nih3/sleep/guide/info-sleep.html
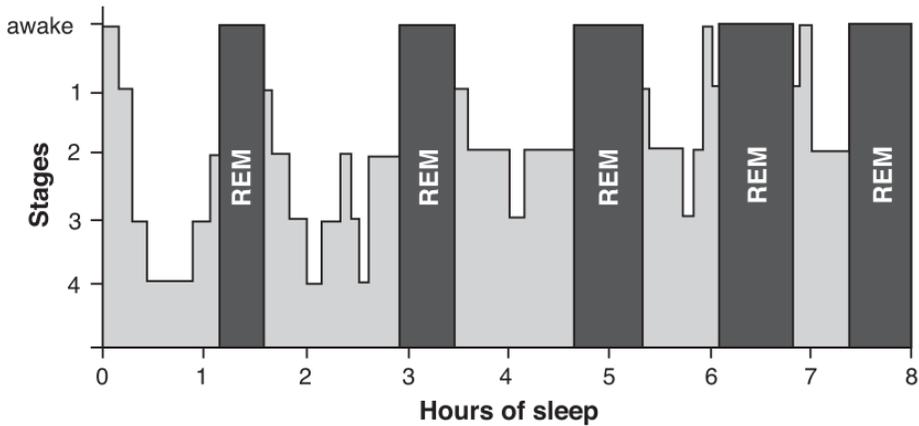
Figure 1.1: A typical hypnogram from a young, healthy adult. Light-gray areas represent NREM

attenuates the delta power. Furthermore, REM latency in the presence of acute stress decreased in healthy subjects but was unaffected in depressed subjects [Williamson et al., 1995].

The studies that specifically focused on subjective sleep quality scores reported various parameters as predictors.[Westerlund et al., 2016] indicates that stage N2 is negatively related to karolinska sleep questionnaire (KSQ) while SWS is positively correlated. Another Study by [Riedel and Lichstein, 1998] shows stage N1 as a negative indicator of the quality of sleep while SWS is positively associated with subjective sleep satisfaction in older adults with insomnia(OAWI). [Keklund and ÅKerstedt, 1997] shows sleep efficiency and SWS as the positive predictors of Karolinska Sleep Diary (KSD) scores. Stage N1 is introduced as a negative indicator of sleep quality in depressed target group in [Argyropoulos et al., 2003].

While it can be concluded that the relationship between subjective and objective quality is sensitive to type of questionnaires or subjects' demographic and medical conditions, most of these research have tried to predict this relationship by setting up a high-maintenance laboratory experiment by either having a control group or recording the data for a long period [Williamson et al., 1995; Hall et al., 2000; Kumar et al., 2012].

1.2   Problem description

The focus of the project is to recognize physiological markers from bodily responses during sleep which can explain and predict subjective quality of sleep scores. From a

machine learning perspective, the problem can be divided into two steps:

- Identification of a set of significant predictors of sleep quality
- Classification of the sleep data using the predictors.

Target group: In order to have a comprehensive model of subjective and objective sleep quality, the target group needs to be defined explicitly, since the sleep parameters are sensitive to additional subject specific characteristics such as age or gender [Kryger et al., 2011]. In this study, we focus on pregnant women. Despite the fact that 79% of pregnant women experience sleep disorders [Rezaei et al., 2013], they are one of the less-studied target groups.

Having sleep disturbances during pregnancy is associated with preterm labor and low birth weight and length, as well as increased risk of spontaneous abortion [Lobel et al., 2008; Mulder et al., 2002]. Moreover, it increases increases mother's likelihood of getting permanent sleep disorder later on in her life[Pien and Schwab, 2004]. Daily sleep monitoring could prevent reaching a threatening level of sleep disturbances. If there is enough evidence to support the hypothesis of the relationship between subjective and objective sleep scores and a predictive model can be conceived, a smart gadget can be developed to monitor sleep score. For example, a recent personal health product at Philips records the heart beats and can also perform it during sleep. Deploying the classifier on this product can provide the user with compelling information regarding the subjective sleep score [de Zambotti et al., 2016; Kumar et al., 2012].

1.2.1   Research Questions

Two main research questions will be addressed in this project:

1. Which sleep parameters/macro features during pregnancy are characteristics of sleep quality?
2. How to apply a representation learning algorithm on time series PSG data to classify the objective sleep quality?

1.3   Solution approach

In order to answer our research questions two sets of sleep parameters, here called macro-features, were extracted form polysomnographic (PSG) data and two different models have been achieved to predict the GSQS value. These models were used for classification task, later. In this study, macro-features refer to the hand-crafted sleep parameters while micro-features are the features that are extracted by the classifiers. A convolutional neural network (CNN) is applied on image representation of time series signals, to extract micro-feature. In the following chapter, the methods that we used will be explained. Chapter 3 details the results. In chapter 4, the obtained result will be discussed.

Table 1.1: A survey of subjective and objective sleep measurements

Study	Subject	Data	Aim
[Åkerstedt et al., 2007]	16 subjects	PSG, KSQ, 9-level rating scale[Wang et al., 2005]	Effects of stress on sleep
[Kumar et al., 2012]	50 subjects	Heartbeats, Subjective stress rating,	24-h stress monitoring
[Ross et al., 1994]	19 subjects (8 control, 11 test)	PSG	Posttraumatic stress disorder (PTSD) effects on sleep
[Hall et al., 2004]	59 subjects (28 control, 31 test)	PSG, Pittsburgh Sleep Quality Index (PSQI), Penn State Worry Questionnaire (PSWQ), Symptom Checklist-90R global distress, Visual Analogue Scale (VAS)	Sleep in presence of acute stress
[Williamson et al., 1995]	68 subjects (35 depressed)	PSG, Life Events Record (LER) [Coddington, 1972]	Acute stress effects on sleep
[Hall et al., 2000]	14 adults	EEG, PSQI, Hamilton Rating Scale for Depression (HRSD)	Chronic insomnia
[de Zambotti et al., 2016]	40 women approaching menopause (22 with insomnia)	Heartbeats, PSG, cortisol level, self-reported tension level Epworth Sleepiness Scale,(ESS),	Acute stress effects on sleep
[Facco et al., 2010]	189 pregnant women	Berlin questionnaire Women's health initiative insomnia rating scale, PSQI	Sleep disorders during pregnancy

Table 1.2: A survey of subjective and objective sleep measurements

Study	Subject	Data	Aim
[Westerlund et al., 2016]	33 subjects	PSG, KSQ	Subjective-objective sleep quality relationship
[Rosipal et al., 2013]	148 subjects	PSG, Self-reported sleep quality	Objective components for sleep quality
[Riedel and Lichstein, 1998]	47 subjects (59+ years old)	PSG Self-reported measurement	Subjective and objective sleep quality relationship in older adults
[Keklund and ÅKerstedt, 1997]	37 subjects	PSG, KSD	Subjective-objective sleep quality relationship
[Argyropoulos et al., 2003]	40 subjects	PSG, Self-reported measurement	Subjective-objective sleep quality relationship in depressed target group

2 Materials and Methods

2.1   Materials

2.1.1   Participants

Sleep data from 45 pregnant women (age: $21 - 38$, BMI in 2nd trimester: 23 ± 3, BMI in 3rd trimester: 24 ± 3) was collected from June to November 2015, in their second trimester as well as in their third trimester, using Alice PDx ambulatory EEG system (Philips Respironics, Monroeville)[1]. Data collection consisted of two nights of test sessions in participants' own homes.

By the end of data recording procedure, ten subjects quit the experiment (one subject in the first night and nine in the second night) due to increased discomfort during pregnancy and as such not able to participate. Details of data are available in A.1.2.

2.1.2   Data description

The physiological data consists of PSG recordings (EEG, ECG, EOG, EMG, respiratory effort, and body position). Subjects were also asked to measure their blood pressure three times with a few minutes in between and to fill out several subjective sleep and stress assessments (Pittsburgh Sleep Quality Index (PSQI) [Buysse et al., 1989], Groningen Sleep Quality Scale (GSQS) [Meijman et al., 1988], Perceived Stress Scale (PSS) [Cohen et al., 1983], Pregnancy-related Anxiety Questionnaire (PRAQ-R) [Huizink et al., 2004] and single-item stress symptoms measure).

The score distribution of the questionnaires and the correlation between each two of them are shown in figure 2.1. This figure shows that the distribution is not balanced in any of the questionnaires except for GSQS. In order to have a good model, we need to have uniformly distributed random scores and GSQS is the only subjective measurement that fits this definition. Therefore, it is considered to be the ground truth for the subjective sleep measurement.

GSQS is a measurement that indicates the subjective quality of sleep and contains 15 items. The maximum score of 14 indicates extremely poor sleep quality. In further

---

[1]http://www.usa.philips.com/healthcare/product/HC1043941/alice-pdx-portable-sleep-diagnostic-system
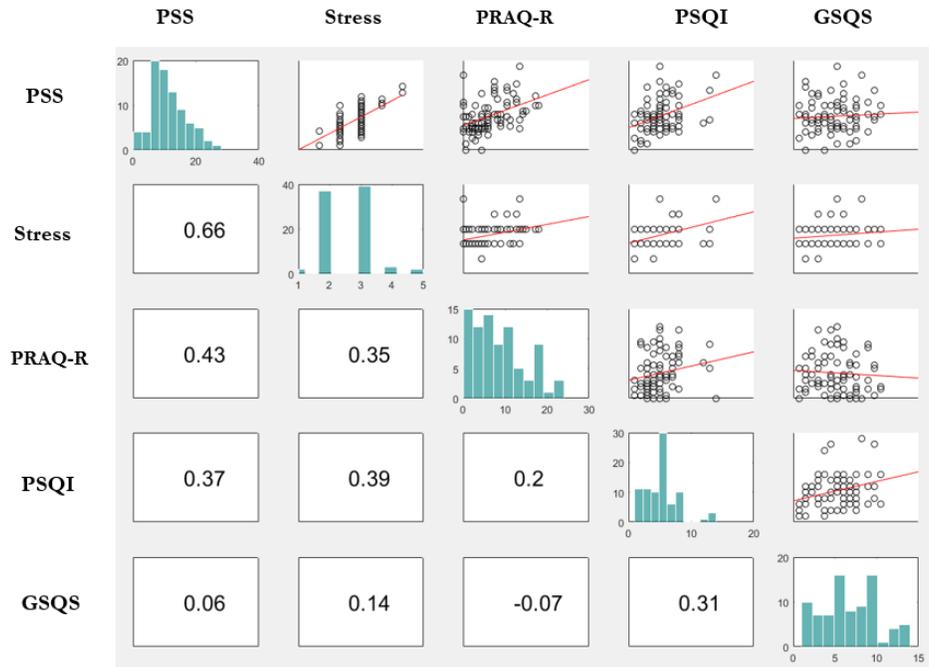
Figure 2.1: The relation between PSS, stress, PRAQ-R, PSQI, GSQS. Diagonal plots: score distributions, upper diagonal plots: scattered scores and lower diagonal plots: scores correlation.

analysis in this project, a total score of less than 8 demonstrates high quality of sleep and the score equal or greater than 8 represents poor quality of sleep[De Weerd et al., 2004; Kis et al., 2014].

Sleep stage annotation of PSG has been done visually by a sleep expert, using the American academy of sleep medicine (AASM) guideline.

Several samples are excluded from annotation process by the expert. Due to the availability of the sleep annotation and questionnaire scores, the number of samples has been narrowed down to 39 for the first night and 24 for the second night.

Furthermore, for frequency analysis in this study, only one EEG channel of PSG data (channel A1-C4) is analyzed. Details of PSG data are available in A.1.1.

2.2 Methods

The aim of the research is to extract those features from the collected PSG data that are necessary to obtain a predictive model of sleep quality during pregnancy. This model will help us distinguish between poor quality and high quality of sleep, later.

2.2.1   Sleep Parameters

For each subject, the macro-sleep features are calculated from the sleep stage annotation of the entire recorded signal. Each sleep stage-annotation refers to one of the events of sleep as introduced by [Rechtschaffen and Kales, 1968]. These events are the aforementioned sleep stages (awake, N1, N2, SWS, and REM) and arousal. The length of each annotated time-window excluding arousal is fixed at 30 seconds [AASM and Iber, 2007].

Finally, the physiological data is labeled with the results from the self-reported sleep quality score, GSQS.

The final dataset is an $N \times M$ matrix where N is the number of subjects, and M is the number of sleep parameters (i.e. features). Trimester is also considered as a sleep parameter. The corresponding labels (GSQS) form an $N \times 1$ vector. Figure 2.2 shows the setup of the feature matrix.
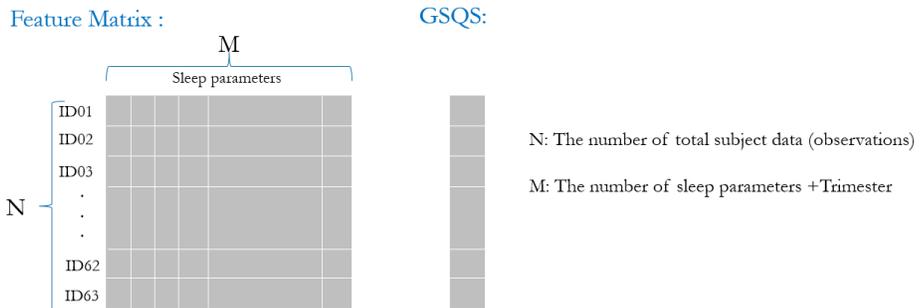
Figure 2.2: Feature-matrix

In this study, two different sets of features have been derived from the sleep stage-annotation. The first set of features contains well-known sleep parameters while the second one gives information of the number of transitions of each stage. The list of the parameters is shown in table 2.1.

The number of parameters or predictors is relatively large in comparison with the number of available observations and it is plausible that a model with all predictors faces over-fitting. Hence, first, the features should be reduced to a reasonable number. In order to find the parameters that are relevant to quality of sleep, Least Absolute Shrinkage and Selection Operator (LASSO) regression has been applied to the feature matrix. A simple linear regression method minimizes the least square error to find the optimum coefficients. LASSO uses same methodology by adding a regularization parameter λ to the error function. By choosing the right value of λ, the irrelevant coefficients to data tend toward zero and the parameters that are useful for explaining the data remain.

After obtaining the robust model of the predictors, the regression problem can be shifted to a classification problem. By setting a threshold on GSQS, the labels become

Table 2.1: The list of sleep parameters in the two feature set

First Feature Set	Second Feature Set
Total Sleep Time (TST)	N1-W
Total Stage1 / TST (N1 index)	N1-N2
Total Stage2 / TST (N2 index)	N1-SWS
Total SWS / TST (SWS index)	N1-REM
Total REM/ TST (REM index)	N2-W
Total Arousal/ TST (Arousal index)	N2-N1
REM Latency	N2-SWS
Wakeness After Sleep Onset (WASO)	N2-REM
Total Sleep Period (TSP)	SWS-W
Wakefulness/ TST (W index)	SWS-N1
TST/TSP (Sleep Efficiency)	SWS-N2
N1/N2	SWS-REM
	REM-W
	REM-N1
	REM-N2

binary representing two classes of poor quality and high quality of sleep.

The Z-scored dataset is divided as follows: 80 percent train set and 20 percent test set. Since we have a small sample set, the split process has been performed in a 1000-iteration loop. Namely, we have performed 1000 random splits of the data into training and test set. Additionally, the trained model's parameters were obtained using 10-fold cross-validation on the training set. To evaluate the performance of the classification model Cohen's kappa coefficient is used [Cohen, 1968]. The new setup of the data set has been shown in figure 2.3.

2.2.2   EEG features

Dealing with EEG time series signals in this scope is not trivial. Besides the common challenge on EEG analysis, like low signal to noise ratio (SNR) [Stober et al., 2015], the difficulties of this specific dataset stem from two challenges. First, the number of observations is limited. Second, not all parts of the time series signal contain the features relevant to sleep quality and if so, it is not clear what kind of features are expected to be extracted. To overcome the addressed issues, we enlarge the number of observation, first, and then we "clean" the dataset by the following approaches.

To augment the number of samples, we define a new setup of dataset which is shown in figure 2.4. The dimension of the new dataset is $WN \times L$ where W is the sum of the number of windows for each signal, and L is the length of one epoch. Considering 30-second epoch, according to AASM 2007 standard criteria, and 200 Hz sampling rate, L
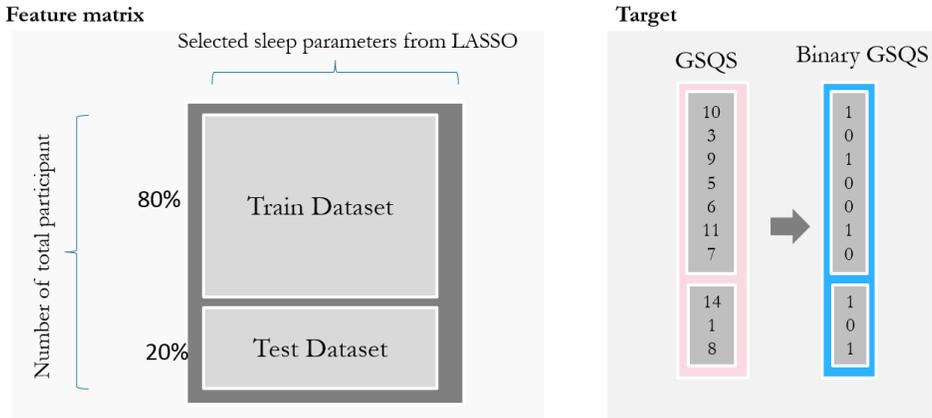
Figure 2.3: Feature-matrix setup for classification task

contains 6000 time samples. Each epoch gets the label corresponding to binary GSQS. Since the sleep stage is known, it is also possible to investigate the quality of sleep for each stage, separately.
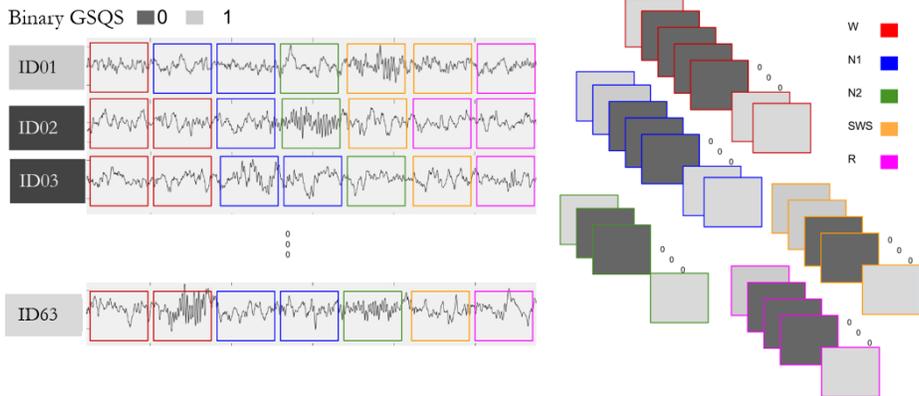
Figure 2.4: A rough sketch of PSG annotation in each stage-window corresponding with the subjective sleep score reported by binary GSQS

As a first step of cleaning the dataset, the time series signal for each subject is divided

to four quarters. Each quarter roughly represents one cycle of sleep. The frequency features of a specific quarter are extracted by applying Fast Fourier Transformation (FFT) to the corresponding epochs. Six frequency ranges are introduced to represent frequency features: delta [0-3.9 Hz], theta [3.9-7.02 HZ], alpha [7.02-11.7 Hz], sigma [11.7-14.04 Hz], beta1 [14.04-21.84 Hz] and beta2 [21.84-30.03 Hz] [Bastien et al., 2003]. The average of the power spectrum of each frequency range is considered as a parameter. The same feature matrix setups, explained in figure 2.2 and figure 2.3, are used to find the quarter of sleep which contains the most information about the quality of sleep. Note that the number of parameters ends up to six here ($M = 6$). The frequency analysis provides the approximate location of relevant features. This information later will be used to feed the classifier.

Having a rich and almost clean dataset, now it is possible to apply more complex classifiers on data. One of the powerful algorithms of feature learning of high dimensional data is Deep Neural Networks (DNNs). In particular, given the success of CNNs in visual processing tasks, it is the first choice for providing the flexible map from high-dimensional observations to the task-relevant features [Finn et al., 2015; Koutnik et al., 2014].

In order to use CNN, the dataset needs to be represented as an image dataset. Converting one dimension signals to image is a common approach in speech recognition tasks [Abdel-Hamid et al., 2012, 2014]. The following steps explain how one single epoch can be converted to an image:

1. Defining three frequency ranges R[0-10 Hz], G[10-20 Hz], B[20-30 Hz].
2. Applying FFT on Hamming window of size $l = 1200$ and overlap of 1150 in order to calculate power density.
3. Removing power density of frequencies greater than 30 Hz.
4. Mapping R, G, and B frequencies' power densities to gray-scale values [0, 255].
5. Moving the Hamming widow to the next point (which is the next 50th time sample) and repeating the steps 2-5 till the end of the epoch.
6. Concatenating R, G, and B matrices as RGB image.

The steps are shown in figure 2.5.

The size of the obtained image is 100×100. The image size can be modified depending on the structure of the network. It is also possible to feed the network with one gray-scale image contains 0-30 Hz frequency range, instead of an RGB image.

The generated images often contain similar features at different horizontal locations. Hence, the data augmentation is only possible by mirroring the image. The size of the input kernel or window is chosen relatively large to ignore the temporal properties of the signal while the stride is two. Max-pooling layer is used after each convolutional layer to reduce the size and make the network robust in spite of the small changes in image features. After investigating several network structures, a structure inspired by Lenet net[LeCun et al., 1998] is chosen.
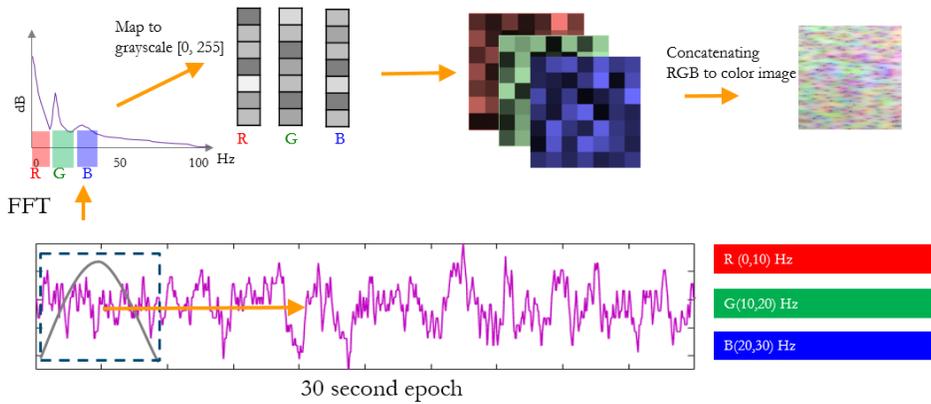
Figure 2.5: Single epoch representation as an image

Figure 2.6 shows that the structure has been used as a classifier in the following analysis on EEG signals. The network has two convolutional layers and two fully connected final layers. To ovoid over-fitting, the dropout is also used to randomly remove some of the nodes during training with probability of $p = 0.5$. Rectified Linear Units (ReLu) is used as an activation function to boost the nonlinear properties of the decision function. Caffe has been used as a deep network framework and the rest of analysis is done by MATLAB.

In order to have a baseline for evaluating the network performance, Support Vector Machine (SVM) is applied on the frequency features of dataset.
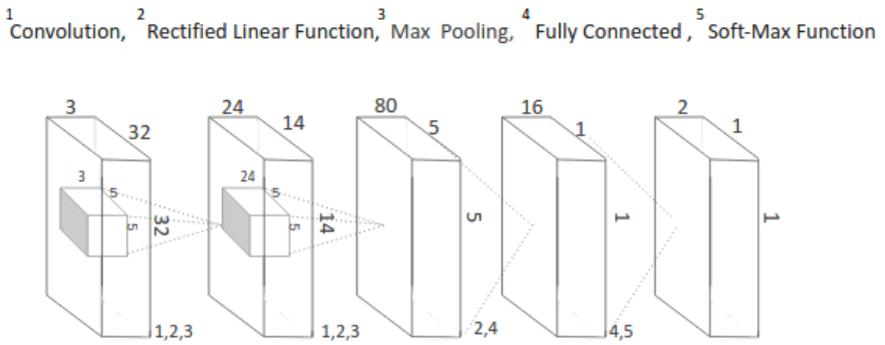
Figure 2.6: The architecture of the convolutional neural network

3 Results

3.1   Subjective sleep measurement

Figure 3.1 shows how GSQS values change through second trimester to third trimester for each subject. We cannot conclude that the trimester of pregnancy has a specific impact on sleep quality, as we did not find a pattern in changing subjective sleep quality from second trimester to third.
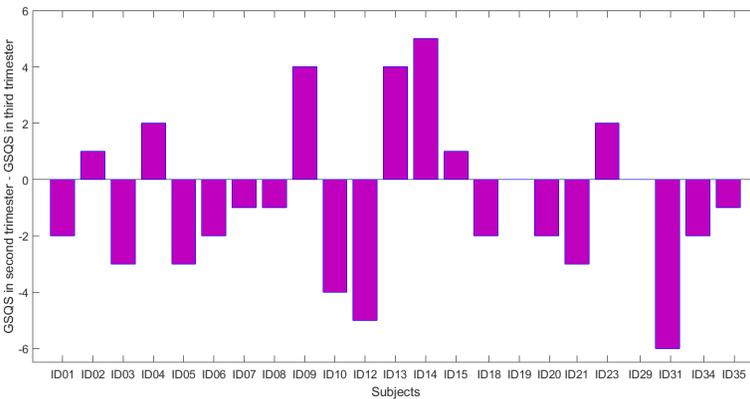
Figure 3.1: The change of GSQS through second trimester to third for each participate.

3.2   Objective sleep measurement

3.2.1   Macro-features

The first set of features that have been analyzed is a set of well-known sleep parameters. The properties and the correlation of a single sleep parameter and GSQS are shown in

table 3.1. The trimester is also considered as a feature to investigate if it does have any impact on the final model.

Table 3.1: Sleep parameters statistics

Sleep Parameters	GSQS < 8	GSQS ≥ 8	Correlation
TST (min)	439 ± 41	399 ± 89	**-0.29**[*]
N1 index	0.1 ± .03	0.14 ± 0.04	**0.51**[**]
N2 index	0.46 ± 0.06	0.47 ± 0.07	0.00
SWS index	0.21 ± 0.06	0.2.0 ± 0.07	-0.07
REM index	0.23 ± 0.04	0.20 ± 0.04	**-0.36**[*]
Arousal index	0.04 ± 0.01	0.05 ± 0.02	**0.30**[*]
REM latency (min)	95 ± 31	102 ±0.59	0.10
WASO (min)	70 ± 91	61 ± 50	-0.07
TSP (min)	535 ± 52	532± 115	-0.02
W index	0.06 ±0.04	0.11 ± 0.07	**0.33**[*]
TST/TSP	0.82 ± 0.05	0.75 ±0.07	**-0.51**[**]
N1/N2	0.21 ± 0.06	0.30 ±0.14	**0.48**[**]

$*p < 0.02$
$**p < 0.0001$

Feature selection is performed by using a LASSO reduction model. As mentioned earlier in Methods chapter, the optimum value of λ leads to find the coefficients of the parameters that are supported by data. The trace plot of Mean Square Error (MSE) for different values of λ is shown in figure 3.2. The MSEs of 10-fold cross-validated data are calculated for each specific λ value. The minimum standard deviation of MSEs indicates a robust and optimum λ. The corresponding coefficients to the robust λ are introduced as the predictors. Note that these results are obtained after 100 iterations. Figure 3.2 shows the average MSE of 100 of 10-fold cross-validations of data.

By finding the optimum λ, irrelevant parameters are removed from the model and features space is narrowed down from thirteen to two parameters which are N1 index and sleep efficiency (SE). The obtained Pearson correlation is **r=0.61**, $p < 10^{-7}$. The models show that the amount of N1 and SE correlates with the reported sleep quality in the morning such that the individuals who reported poor sleep have shown more N1 sleep and a lower SE.

A set of linear classifiers is applied to the binary GSQS-labeled model. The average of achieved accuracy and kappa index of each classifier after 1000 iterations are reported in table 3.2.

Figure 3.3 shows the scattered values of N1 and SE and the separability of the samples. For a better visualization, red dots represent the poor quality of sleep while the blue
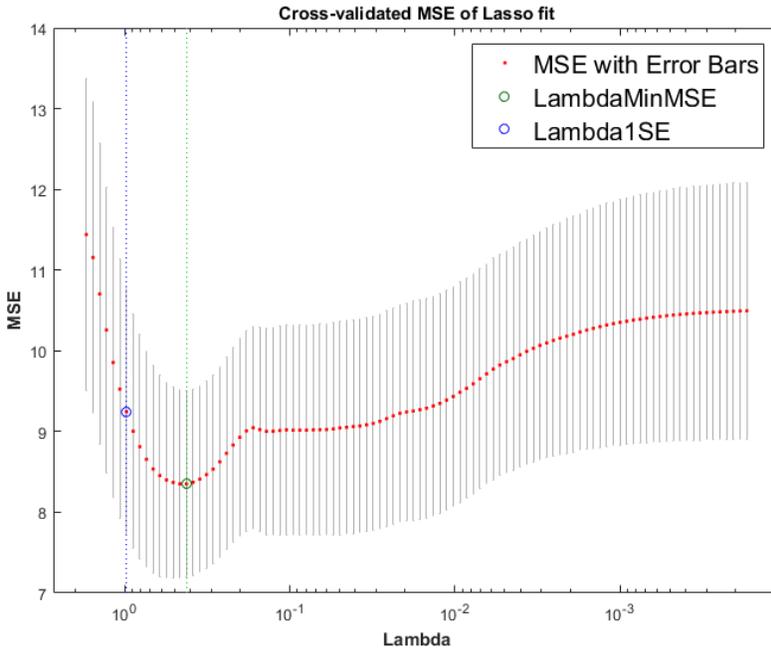
Figure 3.2: The trace plot of standard deviation of MSE against the Lambda regularization parameter for the first feature set

ones indicate high quality of sleep.

Table 3.2: The performance of listed classifiers on model (N1, SE)

	Accuracy	Kappa	Chance
Logistic Regression	0.76	0.51	0.60
Support Vector Machine (SVM)	0.77	0.53	0.60
Naive Bayes	0.80	0.56	0.60

The second set of sleep parameters that has been analyzed is a set of sleep-stage transition numbers. Each parameter indicates the number of transition from a specific stage to another for each subject. Table 3.3 shows the correlation of a single stage transition (from the current stage to the new stage) and GSQS. LASSO regression was able to reduce the number of parameters (including trimester) from sixteen to three with the same setup that
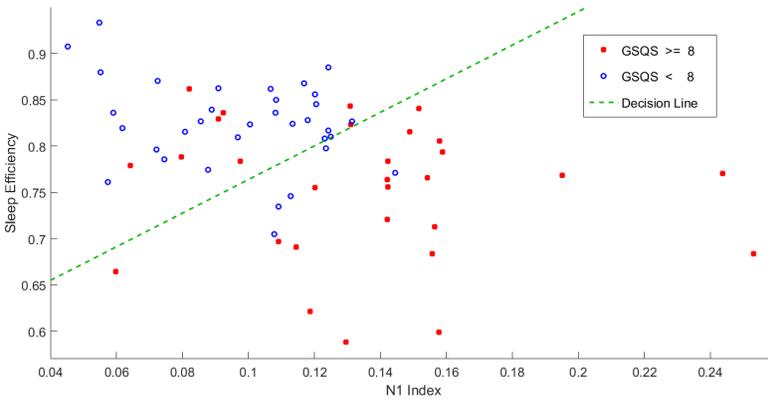
Figure 3.3: The predictor model (N1,SE) for the first set of feature set

has been explained for the first feature set (100 iterations and 10-fold cross-validation). The obtained model contains three predictors: REM-N2, SWS-N2, and N2-N1. The correlation of the predicted scores and GSQS is **r=0.53** with $p < 10^{-5}$.

Table 3.3: Sleep-stage transitions statistics

Sleep Parameters	GSQS < 8	GSQS ≥ 8	Correlation
N1-W	4.94 ± 3.77	6.24 ± 4.02	0.17
N1-N2	33.94 ± 12.26	40.83 ± 16.04	0.24
N1-SWS	0.06 ± 0.34	0.03 ± 0.19	-0.04
N1-REM	3.09 ± 1.76	2.31 ± 2.95	-0.16
N2-W	5.15 ± 2.27	4.97 ± 2.63	-0.04
N2-N1	25.03 ± 12.93	31.52 ± 14.35	0.23
N2-SWS	25.41 ± 7.92	19.38 ± 10.21	**-0.32**
N2-REM	4.82 ± 1.78	4.34 ± 1.65	-0.14
SWS-W	0.62 ± 0.85	0.55± 0.63	-0.04
SWS-N1	0.94 ±1.01	0.79 ± 0.86	-0.08
SWS-N2	23.76 ± 7.70	17.79 ±9.97	**-0.32**
SWS-R	0.18 ± 0.39	0.28 ±0.53	0.11
R-W	1.47 ± 1.24	1.52 ±1.35	0.02
R-N1	4.35 ± 2.50	4.34 ±3.88	0.00
R-N2	2.29 ± 1.85	1.14 ±1.13	**-0.35**

The same set of linear classifiers has been applied to binary GSQS-labeled model.
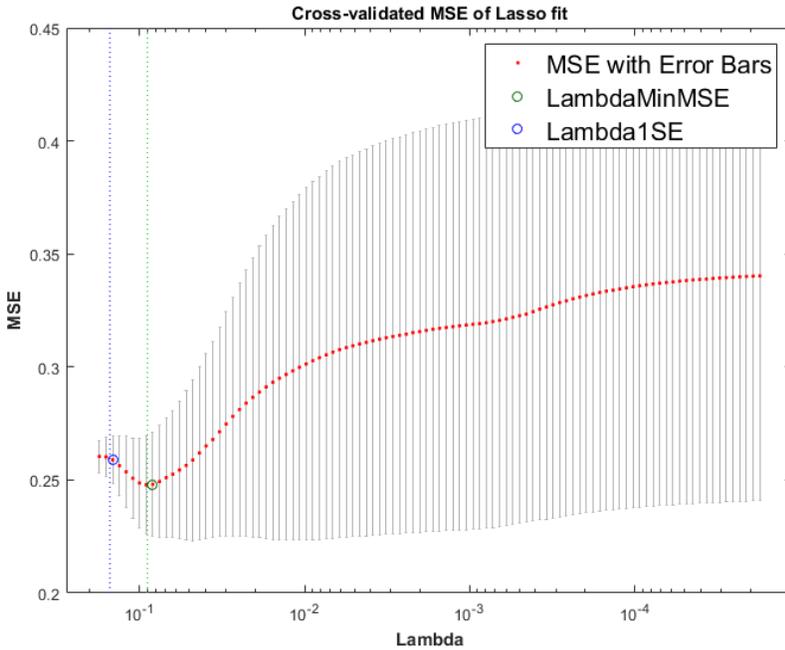
Figure 3.4: The trace plot of standard deviation of MSE against the Lambda regularization parameter for the second feature set

The average performance has been shown in table 3.4.

Figure 3.5 shows the scattered selected parameters by LASSO. The number of transitions from REM to N2 and SWS to N2 are positively correlated with GSQS while the number of transitions from N2 to N1 has a negative correlation with GSQS.

Table 3.4: The performance of listed classifiers on model (REM-N2, SWS-N2, N2-N1)

	Accuracy	Kappa	Chance
Logistic Regression	0.71	0.42	.60
Support Vector Machine (SVM)	0.70	0.39	0.60
Naive Bayes	0.72	0.44	0.60
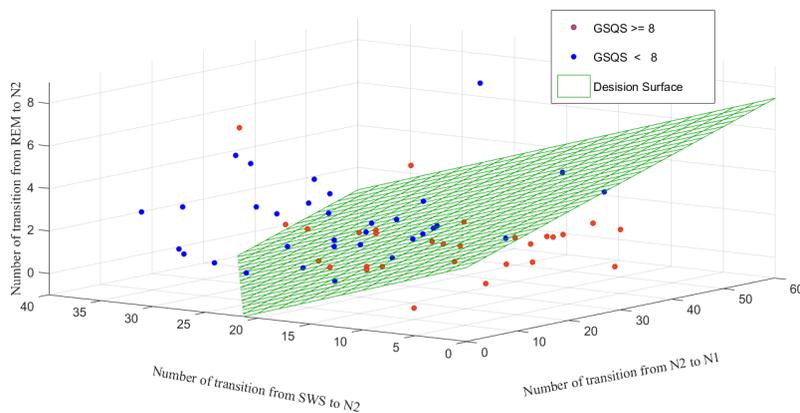
Figure 3.5: The predictor model(REM-N2, SWS-N2, N2-N1) for the second set of feature set

3.2.2   Micro-features

As it is mentioned in Methods section, in order to find the most informative part of sleep regrading the frequency features, we analyse the average power spectrum of the six frequency ranges (delta [0-3.9 Hz], theta [3.9-7.02 HZ], alpha [7.02-11.7 Hz], sigma [11.7-14.04 Hz], beta1 [14.04-21.84 Hz] and beta2 [21.84-30.03 Hz]) for each sample. Each specific average power spectrum is considered as a predictor. The estimated sleep quality is obtained by applying regression analysis on the six power spectrum values. The correlation of the predicted sleep quality score and GSQS for each quarter and for each stage has been shown in table 3.5.

Table 3.5: The average frequency features analysis for each quarter and each stage of sleep

	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
N1	0.10	0.17	0.13	0.09	0.10
N2	0.13	0.29	0.21	0.20	0.18
N3	0.10	0.26	0.16	0.15	0.15
R	0.10	0.33	0.20	0.19	0.13
W	0.41	0.43	0.32	0.34	**0.42**
ALL	0.23	**0.36**	0.21	0.20	0.25

The average frequency features analysis for each quarter reports that the second quarter ($r = 0.36$) and wakefulness ($r = 0.42$) have the highest correlation values among the sleep quarters and sleep stages, respectively.

The second quarter of sleep draws out from each participant's EEG signal and is converted to images. The image dataset feeds to CNN, later. The performance of the network for 15219 input images (75% train dataset and 25% test dataset) is shown in figure 3.6

The network is able to predict poor sleep quality and high sleep quality directly from EEG signals with 78% accuracy. In this analysis, the chance is 54.5% (dataset is not exactly balanced). In order to have a baseline, SVM is applied to EEG epochs with the same setup of the train and test dataset. The features are the average power densities of six ranges of frequencies introduced previously in Methods section. The obtained accuracy of SVM is around 57% which is slightly above the chance level.
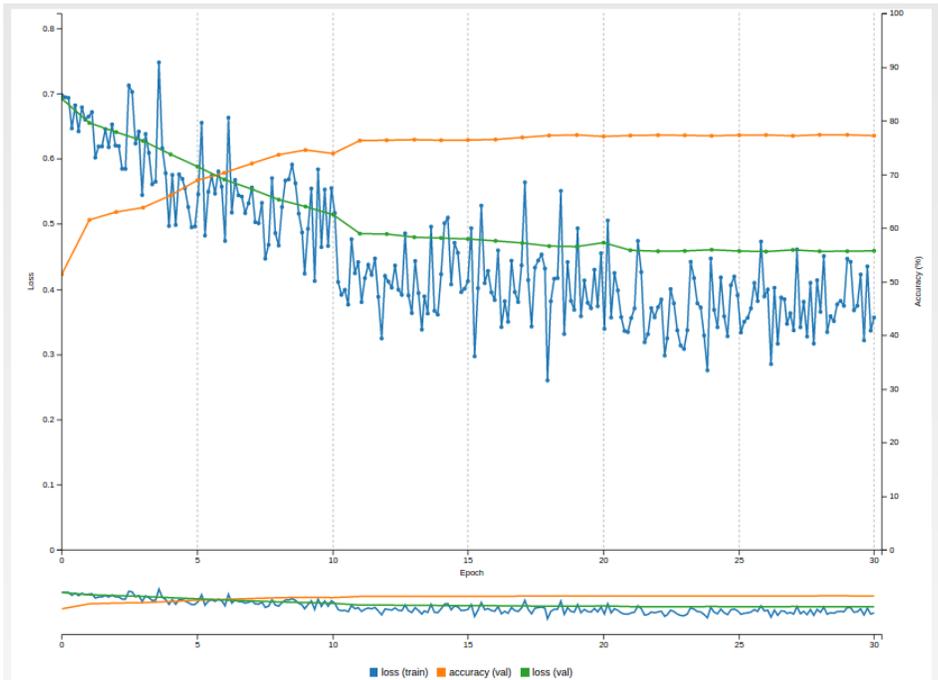
Figure 3.6: The test loss, training loss and accuracy of the network for 30 network-epoch

4 Discussion

This study investigated the relationship between subjective and objective sleep scores. The outcome of applying LASSO regression on the first set of sleep parameters indicated that N1 index and sleep efficiency are associated with self-reported sleep quality in GSQS. Particularly, the high amount of N1 index is a strong predictor of low quality of sleep. N1 index indicates how much of total sleep time is spent in transition from wakefulness to sleep. Hence, it is notable how this parameter plays a role in sleep quality score. In the other hand, the high amount of the sleep efficiency is seemed to be evident to predict the high quality of sleep.

Having a larger number of subjects can be helpful to set a certain N1 index as a threshold for recognizing low quality of sleep, immediately. For example in this particular dataset, 31% of the samples can be recognized as poor quality of sleep with probability of $P = 1$ by setting a threshold for N1 index > 0.127. In other words, if the subject spends more than 12.7% of sleep in stage N1, she suffers from poor quality of sleep.

In support of our findings, [Riedel and Lichstein, 1998; Keklund and ÅKerstedt, 1997; Argyropoulos et al., 2003] reported sleep efficiency and stage N1 as the sleep quality predictors. Although, as it is mentioned before, the effective parameters are strongly related to the subject's profile (age, gender, health condition). Therefore, there is a variety of predictive models reported in different studies.

The result obtained from the number of stage transitions set shows that the number of transitions from stage R and SWS to N2 are positively, and the number of transition from stage N2 to N1 are negatively, correlated with sleep quality scores. It is interesting to observe that the content of sleep, specifically SWS and REM, are not important for predicting the subjective score while their number of transitions to N2 might be useful to make a predictive model. [ÅKERSTEDT et al., 1994].

One of the important drawbacks of PSG analysis is the limited number of observations. Initially, the PSG data from 39 and 24 subjects were available for their second and third trimesters, respectively. One way of analysis the data was splitting the trimesters and doing the analysis separately. The result of this approach is available in A.2.2. Alternatively, by combining these two trimesters, not only the the number of samples is increased but also it is possible to investigate the effect of the trimester on sleep quality score by including it as a predictor. One might argue that the data from a certain sub-

ject appears in both test and training sets (24 subjects share the data in both trimesters), hence the classifier's outcome might be disturbed. In general, this seems to be a valid point, although the result has shown no evidence that the second or third trimester can be important to predict the score in this study. This leads to an interesting remark that the relationship between objective and subjective sleep quality measures is independent from the trimester.

Regarding to micro-feature analysis, the result of EEG frequency analysis indicates that the second quarter of sleep carries the most frequency information to predict the sleep score. The result of CNN shows that it is possible to classify the sleep data in two classes of poor and high sleep quality by having several epochs of EEG. The structure of the network can be improved using sets of horizontal and vertical rectangular kernels instead of square kernels. Applying deconvolutional network on the trained network, it is possible to observe the most important frequency features.

Apart from the common constraints in sleep study (e.g. the limited number of observations), unequally distributed subjective scores have an important effect on the identifying of the sleep quality predictors. This might not be an issue as long as the samples are divided into two groups of high and poor sleep quality. The nature of the experiment can be improved by including broader types of participants in regard to age, gender and medical conditions. Precisely, the reported result in the present study is interpretable for pregnant target group, only. It will be interesting to test the predictive models in a larger group of subjects with a different profile.

5 Conclusion

N1 index and sleep efficiency are correlated with the reported sleep quality in the morning. Individuals that reported poor sleep quality show more N1 index and a lower sleep efficiency.

Low number of transitions from REM and SWS to N2 in association with high number of transitions from N2 to N1 indicate poor quality of sleep.

The most distinguishable frequency features relevant to sleep quality can be identified mostly in the second quarter of sleep.

The quality of sleep has been predicted by CNN with 78 percentage accuracy while the baseline accuracy obtained by SVM is about 57 percentage.

References

AASM and Conrad Iber. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*. American Academy of Sleep Medicine, 2007.

Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, and Gerald Penn. Applying convolutional neural networks concepts to hybrid nn-hmm model for speech recognition. In *2012 IEEE international conference on Acoustics, speech and signal processing (ICASSP)*, pages 4277–4280. IEEE, 2012.

Ossama Abdel-Hamid, Abdel-Rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu. Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on audio, speech, and language processing*, 22(10):1533–1545, 2014.

TORBJÖRN ÅKERSTEDT, KEN Hume, David Minors, and Jim Waterhouse. The meaning of good sleep: a longitudinal study of polysomnography and subjective sleep quality. *Journal of Sleep Research*, 3(3):152–158, 1994.

Torbjörn Åkerstedt, Göran Kecklund, and John Axelsson. Impaired sleep after bedtime stress and worries. *Biological psychology*, 76(3):170–173, 2007.

Spilios V Argyropoulos, Jane A Hicks, Jon R Nash, Caroline J Bell, Ann S Rich, David J Nutt, and Susan J Wilson. Correlation of subjective and objective sleep measurements at different stages of the treatment of depression. *Psychiatry Research*, 120(2):179–190, 2003.

Célyne H Bastien, Mélanie LeBlanc, Julie Carrier, and Charles M Morin. Sleep eeg power spectra, insomnia, and chronic use of benzodiazepines. *SLEEP-NEW YORK THEN WESTCHESTER-*, 26(3):313–317, 2003.

Daniel J Buysse, Charles F Reynolds, Timothy H Monk, Susan R Berman, and David J Kupfer. The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry research*, 28(2):193–213, 1989.

R Dean Coddington. The significance of life events as etiologic factors in the diseases of childrenii a study of a normal population. *Journal of psychosomatic research*, 16(3): 205–213, 1972.

Jacob Cohen. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4):213, 1968.

Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. A global measure of perceived stress. *Journal of health and social behavior*, pages 385–396, 1983.

Al De Weerd, Sanne De Haas, Andreas Otte, Dorothée Kasteleijn-Nolst Trenité, Gerard Van Erp, Adam Cohen, Marieke De Kam, and Joop Van Gerven. Subjective sleep disturbance in patients with partial epilepsy: A questionnaire-based study on prevalence and impact on quality of life. *Epilepsia*, 45(11):1397–1404, 2004.

Massimiliano de Zambotti, David Sugarbaker, John Trinder, Ian M Colrain, and Fiona C Baker. Acute stress alters autonomic modulation during sleep in women approaching menopause. *Psychoneuroendocrinology*, 66:1–10, 2016.

Francesca L Facco, Jamie Kramer, Kim H Ho, Phyllis C Zee, and William A Grobman. Sleep disturbances in pregnancy. *Obstetrics & Gynecology*, 115(1):77–83, 2010.

Chelsea Finn, Xin Yu Tan, Yan Duan, Trevor Darrell, Sergey Levine, and Pieter Abbeel. Learning visual feature spaces for robotic manipulation with deep spatial autoencoders. *arXiv preprint arXiv:1509.06113*, 2015.

Martica Hall, Daniel J Buysse, Peter D Nowell, Eric A Nofzinger, Patricia Houck, Charles F Reynolds III, and David J Kupfer. Symptoms of stress and depression as correlates of sleep in primary insomnia. *Psychosomatic medicine*, 62(2):227–230, 2000.

Martica Hall, Raymond Vasko, Daniel Buysse, Hernando Ombao, Qingxia Chen, J David Cashmere, David Kupfer, and Julian F Thayer. Acute stress affects heart rate variability during sleep. *Psychosomatic medicine*, 66(1):56–62, 2004.

Anja C Huizink, Eduard JH Mulder, Pascale G Robles de Medina, Gerard HA Visser, and Jan K Buitelaar. Is pregnancy anxiety a distinctive syndrome? *Early human development*, 79(2):81–91, 2004.

Göran Keklund and Torbjörn ÅKerstedt. Objective components of individual differences in subjective sleep quality. *Journal of sleep research*, 6(4):217–220, 1997.

Anna Kis, Sára Szakadát, Péter Simor, Ferenc Gombos, Klára Horváth, and Róbert Bódizs. Objective and subjective components of the first-night effect in young nightmare sufferers and healthy participants. *Behavioral sleep medicine*, 12(6):469–480, 2014.

Jan Koutnik, Jürgen Schmidhuber, and Faustino Gomez. Online evolution of deep convolutional network for vision-based reinforcement learning. In *From Animals to Animats 13*, pages 260–269. Springer, 2014.

Meir H. Kryger, Thomas Roth, and William Dement. *Principles and Practice of Sleep Medicine, 5th Edition*. Elsevier, 2011. ISBN 978-1-4377-0731-1.

Mohit Kumar, Sebastian Neubert, Sabine Behrendt, Annika Rieger, Matthias Weippert, Norbert Stoll, Kerstin Thurow, and Regina Stoll. Stress monitoring based on stochastic fuzzy analysis of heartbeat intervals. *Fuzzy Systems, IEEE Transactions on*, 20(4): 746–759, 2012.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Marci Lobel, Dolores Lacey Cannella, Jennifer E Graham, Carla DeVincent, Jayne Schneider, and Bruce A Meyer. Pregnancy-specific stress, prenatal health behaviors, and birth outcomes. *Health Psychology*, 27(5):604, 2008.

TF Meijman, AH de Vries-Griever, G De Vries, and R Kampman. The evaluation of the groningen sleep quality scale. *Groningen: Heymans Bulletin (HB 88-13-EX)*, 2006, 1988.

EJH Mulder, PG Robles De Medina, AC Huizink, BRH Van den Bergh, JK Buitelaar, and GHA Visser. Prenatal maternal stress: effects on pregnancy and the (unborn) child. *Early human development*, 70(1):3–14, 2002.

Grace W Pien and Richard J Schwab. Sleep disorders during pregnancy. *Sleep*, 27: 1405–1417, 2004.

Allan Rechtschaffen and Anthony Kales. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. 1968.

Elham Rezaei, Zahra Behboodi Moghadam, and Khadijeh Saraylu. Quality of life in pregnant women with sleep disorder. *Journal of family & reproductive health*, 7(2):87, 2013.

Brant W Riedel and Kenneth L Lichstein. Objective sleep measures and subjective sleep satisfaction: how do older adults with insomnia define a good night's sleep? *Psychology and aging*, 13(1):159, 1998.

Roman Rosipal, Achim Lewandowski, and Georg Dorffner. In search of objective components for sleep quality indexing in normal sleep. *Biological psychology*, 94(1):210–220, 2013.

Richard J Ross, William A Ball, David F Dinges, Nancy B Kribbs, Adrian R Morrison, Steven M Silver, and Francis D Mulvaney. Rapid eye movement sleep disturbance in posttraumatic stress disorder. *Biological psychiatry*, 35(3):195–202, 1994.

Sebastian Stober, Avtial Sternin, Adrian M Owen, and Jessica A Grahn. Deep feature learning for eeg recordings. *arXiv preprint arXiv:1511.04306*, 2015.

Jiongjiong Wang, Hengyi Rao, Gabriel S Wetmore, Patricia M Furlan, Marc Korczykowski, David F Dinges, and John A Detre. Perfusion functional mri reveals cerebral blood flow pattern under psychological stress. *Proceedings of the National Academy of Sciences of the United States of America*, 102(49):17804–17809, 2005.

Anna Westerlund, Ylva Trolle Lagerros, Göran Kecklund, John Axelsson, and Torbjörn Åkerstedt. Relationships between questionnaire ratings of sleep quality and polysomnography in healthy adults. *Behavioral sleep medicine*, 14(2):185–199, 2016.

Douglas E Williamson, Ronald E Dahl, Boris Birmaher, Raymond R Goetz, Beverly Nelson, and Neal D Ryan. Stressful life events and eeg sleep in depressed and normal control adolescents. *Biological Psychiatry*, 37(12):859–865, 1995.

A Appendix

A.1   Materials

A.1.1   PSG data properties

Table A.1: The properties of PSG data

Channel Name	Sampling Rate	Channel's Number
ECG II	200 Hz	Ch 1
Effort THO	100 Hz	Ch 2
Effort ABD	100 Hz	Ch 3
EMG chin	200 Hz	Ch 4
EOG-R	200 Hz	Ch 5
EEG O2- A1	200 Hz	Ch 6
EEG A1- A2	200 Hz	Ch 7
EEG F4- A1	200 Hz	Ch 8
EEG C4- A1	200 Hz	Ch 9
EEG C3-A2	200 Hz	Ch 10
L-EOC	200 Hz	Ch 11
Body	1 Hz	Ch 12
ECG I	1 KHz	Ch 13

A.1.2   Missing PSG data

The list of available PSG data, GSQS and corresponding sleep annotation is shown in table A.2-A.4.

Table A.2: PSG Data subject 1-16

Subject	Trimester	PSG	Annotated PSG	Questionnaire
ID01	2	✓	✓	✓
	3	✓	✓	✓
ID02	2	✓	✓	✓
	3	✓	✓	✓
ID03	2	✓	✓	✓
	3	✓	✓	✓
ID04	2	✓	✓	✓
	3	✓	✓	✓
ID05	2	✓	✓	✓
	3	✓	✓	✓
ID06	2	✓	✓	✓
	3	✓	✓	✓
ID07	2	✓	✓	✓
	3	✓	✓	✓
ID08	2	✓	✓	✓
	3	✓	✓	✓
ID09	2	✓	✓	✓
	3	✓	✓	✓
ID10	2	✓	✓	✓
	3	✓	✓	✓
ID11	2	✓	✓	✓
	3	-	-	-
ID12	2	✓	✓	✓
	3	✓	✓	✓
ID13	2	✓	✓	✓
	3	✓	✓	✓
ID14	2	✓	✓	✓
	3	✓	✓	✓
ID15	2	✓	✓	✓
	3	✓	✓	✓
ID16	2	✓	✓	✓
	3	✓	-	✓

Table A.3: PSG Data subject 17-32

Subject	Trimester	PSG	Annotated PSG	Questionnaire
ID17	2	✓	✓	✓
	3	-	-	-
ID18	2	✓	✓	✓
	3	✓	✓	✓
ID19	2	✓	✓	✓
	3	✓	✓	✓
ID20	2	✓	✓	✓
	3	✓	✓	✓
ID21	2	✓	✓	✓
	3	✓	✓	✓
ID22	2	✓	✓	✓
	3	✓	-	✓
ID23	2	✓	✓	✓
	3	✓	✓	✓
ID24	2	✓	✓	✓
	3	✓	-	✓
ID25	2	✓	✓	✓
	3	✓	-	✓
ID26	2	✓	✓	✓
	3	✓	-	-
ID27	2	✓	✓	✓
	3	-	-	-
ID28	2	-	-	-
	3	-	-	-
ID29	2	✓	✓	✓
	3	✓	✓	✓
ID30	2	✓	✓	✓
	3	✓	-	✓
ID31	2	✓	✓	✓
	3	✓	✓	✓
ID32	2	✓	-	-
	3	✓	-	✓

Table A.4: PSG Data subject 33-45

Subject	Trimester	PSG	Annotated PSG	Questionnaire
ID33	2	✓	✓	✓
	3	-	-	✓
ID34	2	✓	✓	✓
	3	✓	✓	✓
ID35	2	✓	✓	✓
	3	✓	✓	✓
ID36	2	✓	✓	✓
	3	-	-	✓
ID37	2	✓	✓	✓
	3	✓	-	✓
ID38	2	✓	✓	✓
	3	✓	-	✓
ID39	2	✓	-	-
	3	✓	-	✓
ID40	2	✓	✓	✓
	3	-	-	-
ID41	2	✓	✓	✓
	3	-	-	✓
ID42	2	✓	✓	✓
	3	-	-	✓
ID43	2	✓	-	-
	3	✓	-	✓
ID44	2	✓	✓	✓
	3	-	-	-
ID45	2	✓	-	-
	3	-	-	-

A.1.3   GSQS

1. I had a deep sleep last night
2. I feel that I slept poorly last night
3. It took me more than half an hour to fall asleep last night
4. I woke up several times last night
5. I felt tired after waking up this morning
6. I feel that I didn't get enough sleep last night
7. I got up in the middle of the night
8. I felt rested after waking up this morning
9. I feel that I only had a couple of hours' sleep last night
10. I feel that I slept well last night
11. I didn't sleep a wink last night
12. I didn't have trouble falling asleep last night
13. After I woke up last night, I had trouble falling asleep again
14. I tossed and turned all night last night
15. I didn't get more than 5 hours' sleep last night

The first question does not count for the total score.
One point if answer is 'true': questions 2, 3, 4, 5, 6, 7, 9, 11, 13, 14, 15.
One point if answer is 'false': questions 8, 10, 12.
Maximum score 14 points.

A.2   Result

A.2.1   Objective and Questionnaire's scores

Figures A.1-A.4 show the correlation between a single parameter and the questionnaire.
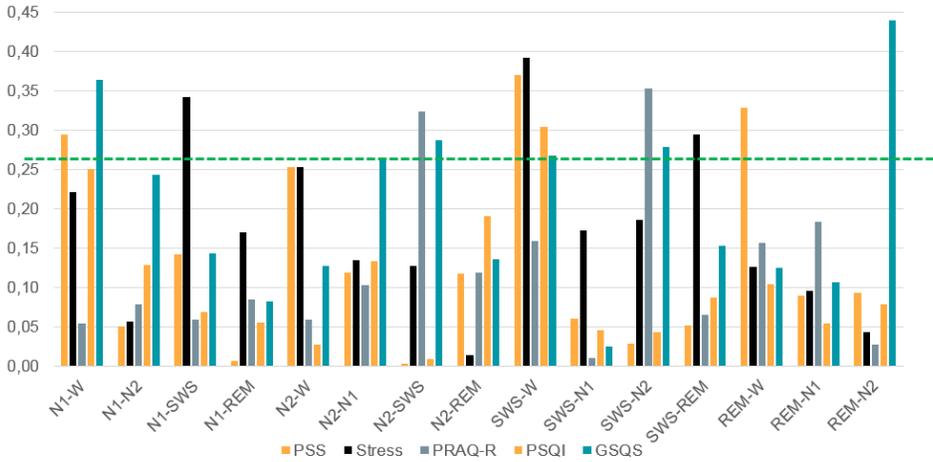
Figure A.1: Correlation of a single sleep transition parameter and questionnaires' scores in 2nd trimester
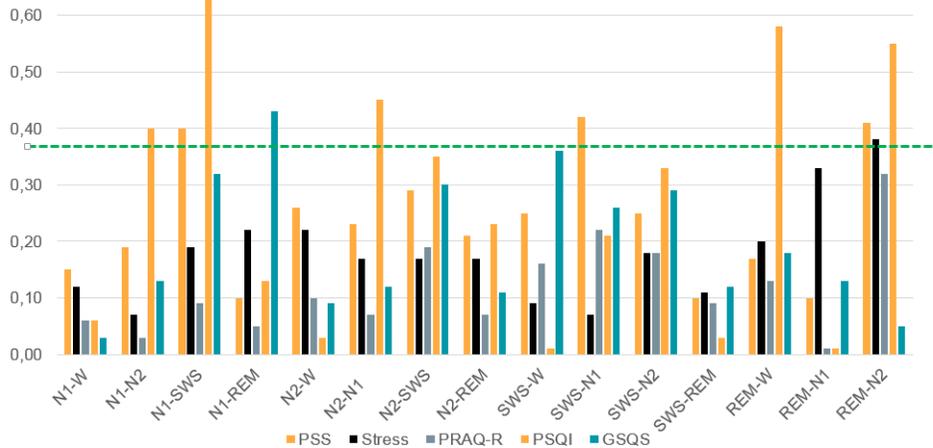
Figure A.2: Correlation of a single sleep transition parameter and questionnaires' scores in 3rd trimester
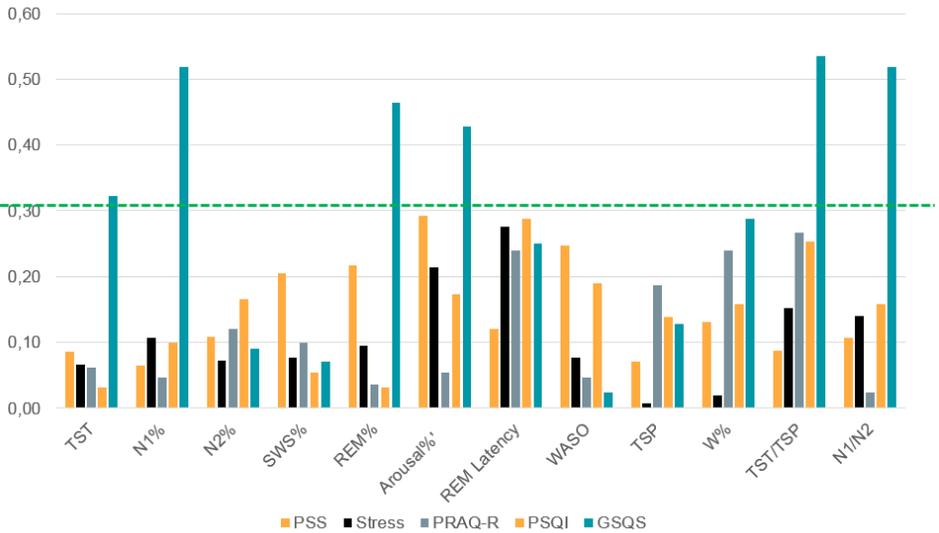
Figure A.3: Correlation of a single sleep parameter and questionnaires' scores in 2nd trimester
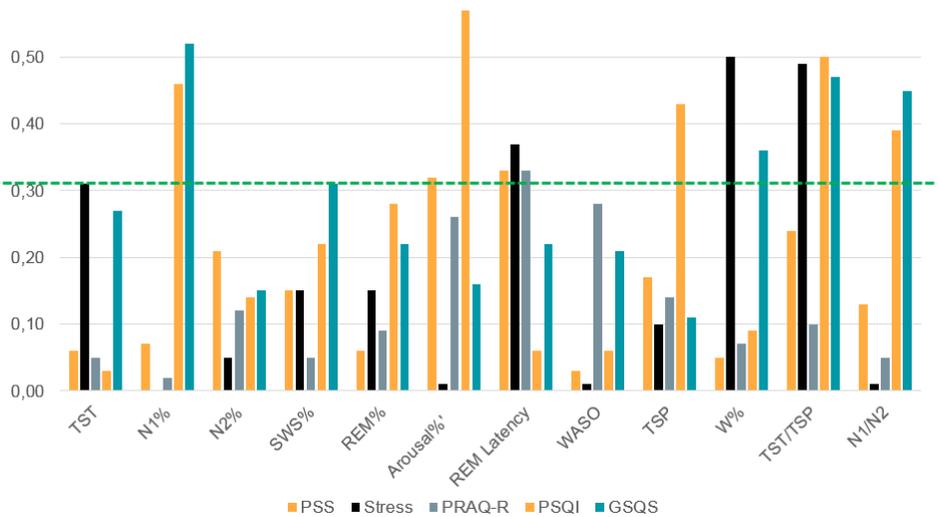
Figure A.4: Correlation of a single sleep parameter and questionnaires' scores in 3rd trimester

A.2.2   Individual trimester analysis

Table A.5: Result of predictive model (N1, SE) for 2nd and 3rd trimester

	Correlation with GSQS	Navie Bayes	SVM	Logistic Regression
2nd Trimester	0.63	0.78	0.79	0.76
3rd Trimester	0.60	0.72	0.73	0.71

Table A.6: Result of predictive model (REM-N2, SWS-N2, N2-N1) for 2nd and 3rd trimester

	Correlation with GSQS	Navie Bayes	SVM	Logistic Regression
2nd Trimester	0.57	0.78	0.72	0.76
3rd Trimester	0.32	0.47	0.55	0.53