Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Prediction of sleep quality from smartphone sound recordings

Semester Thesis

Mohamed Ibrahim
Department of Information and Technology Management
Department of Electrical Engineering and Information Technology

2021

Advisor: Dr. Filipe Barata
Supervisor: Prof. Dr. Elgar Fleisch
Co-Supervisor: Prof. Dr. Roger Wattenhofer

January 10, 2021

Abstract

To predict the sleep quality, sound amplitudes recorded by smartphone microphones were used. The sleep quality was determined according to the PSQI score. The main focus was on Deep Learning models including LSTM. In addition, due to a slight correlation of the sleep w.r.t. to the sleep of the nights before, the data from previous days was fed in as a time series into the models. The results indicate, that sound amplitudes only might not have enough information to predict the sleep quality between subjects. Within subjects the results were moderate with a Matthews correlation coefficient of 0.2, F1 Score of 0.4 and an accuracy of 70%.

Acknowledgements

I would like to thank my Advisor Dr. Filipe Barata for the continuous support during the work and my fellow students for the interesting and inspiring discussions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Research in sleep quality has increased in the past 50 years, when people realized, that sleep quality can affect the person short term as well as long term, with numerous diseases resulting from it. This is well shown here [12] that the top 100 cited sleep medicine articles range from the 1970s till today. A lot of research has been conducted and it has been shown that sleep may cause negative effects in children like behavioural problems and in cognitive functioning, in adolescents it may cause psychological health and effect the school performance and in adults it may cause cognitive, memory and performance deficits as well as emotional distress just to mention a few things [9]. Therefore there is a high interest in identifying poor sleep quality in the first place, determine the factors causing it and trying to improve the sleep quality accordingly.

Early research gathered EEG signals during the night [14], where participants had to attach sensors during the sleep. With the increase of smartphone holders in the past 15 years, recent studies have shown, that 6 out of 10 people used their smartphone in their bed before going to sleep and put it near to their bed [4]. Therefore this and other works gathered data from the smartphone sensors and used it to predict the sleep quality, which is easier to conduct from a participants perspective. Only an app has to be installed and activated in the evening before going to sleep and stopped in the morning when waking up. Other ways to even reduce more the interaction of the person with the whole process is estimating the sleep duration with a best effort sleep (BES) model, which had an overall error in sleep duration of 40 minutes [2].

The drawbacks of using only the microphone of the smartphone might be that noise from other sleeping people in the same room might effect the measurement, particularly in the case when partners sleep in the same bed, makes the prediction even more difficult. Furthermore having a bed near a street or a train track will introduce non-periodical noise and this has to be filtered out. Additionally some people avoid having the smartphone near them during sleep because of the radiation, which makes this approach inapplicable for them.

1.1 Focus of this Work

The goal of this work is to investigate further and find out to what extend sleep quality can be inferred from only sound amplitudes collected by the smartphones microphone during the night.
This work focuses on classifying good vs poor sleep with sound amplitudes recorded during the sleep. Every day is labeled with a PSQI Score. Although the PSQI score is a subjective label, in this work we are trying to develop a Deep Learning model to predict the sleep quality.
Throughout the whole work, a between subject approach is pursued, which means, that training and testing

will be conducted on disjoint group of people. This will be the case in every section in chapter 4 unless otherwise stated. We explored different Deep Learning models and approaches to figure out what kind of conclusions one can derive from the data.

Chapter 2

Theoretical Background

This section will focus on three things. First, the sleep quality prediction using legacy methods like PSG and other wearable sensor based approaches. Second, other work done using smartphone sensors and lastly introducing the PSQI, the sleep quality score used

Polysomnography (PSG) is a sleep study where multiple sensors are attached to a person, usually performed in hospital or a sleep lab and is mostly used to identify sleep disorders and can therefore infer the sleep quality from that information. Although it is still used till today and gathers a lot of valuable information, it is not suitable for a wide range of use and not suited for a use at home at all at a regularly basis [5]. Since the 90s different other sensors were developed like the actigraphy sensors, which is worn like a watch on the wrist and can measure the activity as well as estimate the sleep time but is not able to measure additional parameters as brain activity (EEG) or heart rhythm (ECG). A study done in this area published in 2016 [13] used physical activity data from the day and evaluated the sleep quality of the night using the sleep efficiency, which is the ratio between the total sleep time and the total time spent in bed. At a threshold of 85%, they distinguished between poor and good sleepers. With this approach, a more objective method is used to assess the sleep quality. The prediction was done using Deep Learning models. Concurrent neural networks (CNN), multilayer perceptron (MLP) and long short-term memory (LSTM) have performed quite well in their work, particularly the LSTM got a sensitivity of 0.97 in detecting good sleepers and a specificity of 0.47 in detecting poor sleepers.

There has been some work done in the same area as mentioned before. Especially sleep quality prediction with PSQI scores using smartphone sensors or sensors in general [10]. The smartphone had to be carried during the day to measure the movement. The reason behind this, that previous work have shown, that physical activity does have a moderate effect on sleep quality [6]. The input data for their models were in particular the sound amplitudes, light intensity, sleep time estimates, and acceleration. In addition some device status information were used. The PSQI score was used as a label for the sleep quality of the night. The achieved accuracy was 81.48% with an F-score of 0.81 in detecting poor sleep quality [10].

The PSQI score is one of the most widely used metric to measure sleep quality of people [1], [3]. It assesses the sleep quality of a person over one month. It is standardized and used by many researchers and medical institutions. The questionnaire is filled out by the person after waking up and it takes 5 - 10 minutes to complete, making it an easy, quick and cheap way to analyze the sleep quality. The drawbacks can be, that it is rather subjective in comparison to other methods mentioned above.

Chapter 3

Materials and Methods

3.1 Data

Data was collected from participants, who have got a phone, which was placed by them next to their bed every night. The most useful and complete data was from the microphone, which recorded the sound amplitudes. These were recorded continuously during the whole night and afterwards 2 features, the Root Mean Square (RMS) and the standard deviation, were extracted for every 10 minutes window.

After they woke up, they filled out the questionnaire to gather the PSQI score. The total score ranges from 0 to 21 points, each of the 7 components can contribute between 0 and 3 points. A lower score should reflect a better sleep while good sleepers could be classified with a PSQI score less or equal than 5 and poor sleepers with a PSQI score above 5.

The data had to be cleaned in advance. In particular sleep hours of less than 2h were dismissed as well as nights with no sleep data. Furthermore if questions were not completely answered, which leads to a loss of data regarding the sleep quality of that night, were also disregarded. In addition only people with 20 or more nights were taken into consideration. Outliers with more than 3 standard deviations were removed.

The final data we worked with consists of 72 participants, the number of days varies across the participants. In figure 3.1 the number of participants with x days in the data is shown.

The PSQI score of each night will be used to distinguish between good and poor sleepers. After using the threshold of 5 as mentioned above we get the following distribution in figure 3.2. We have a ratio of 3:1 of good to poor sleep.

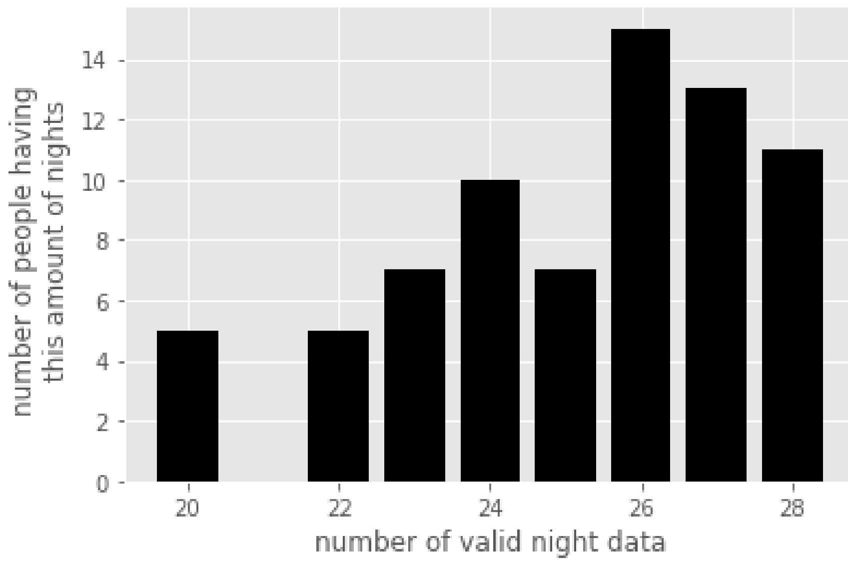


Figure 3.1: number of people having x nights of data in the data set

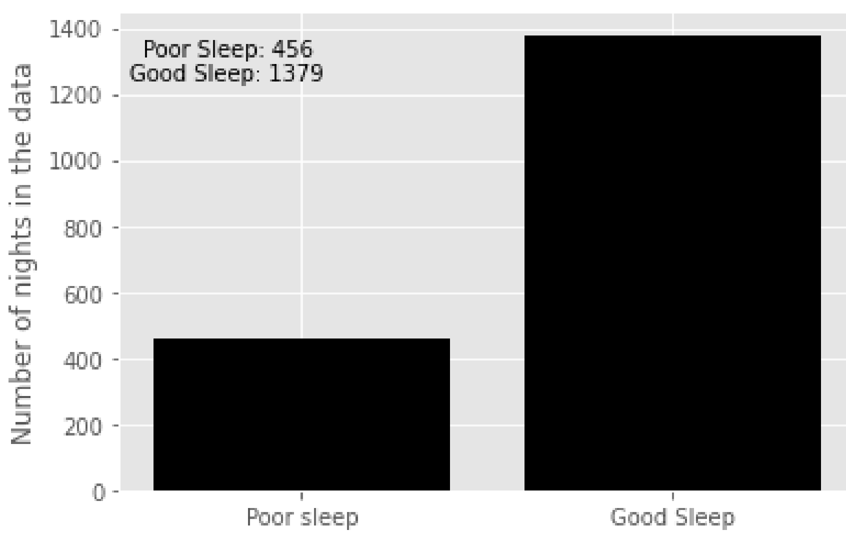
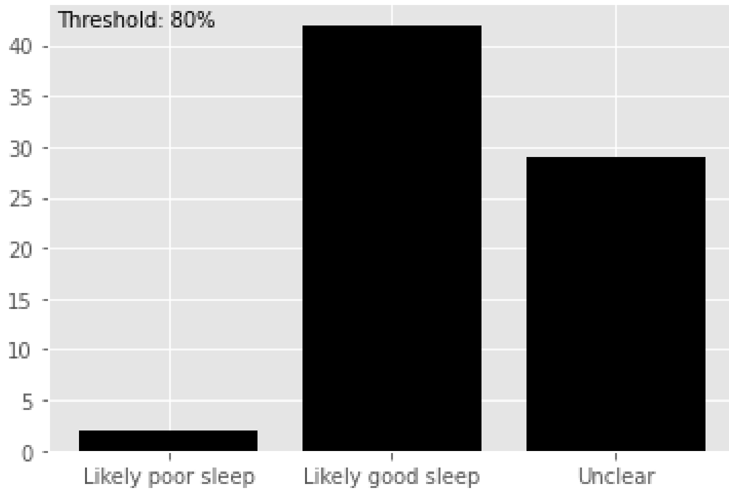
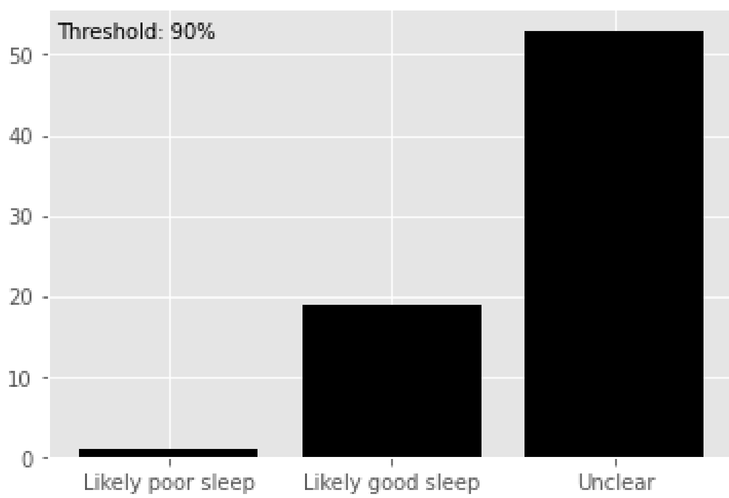


Figure 3.2: Ratio of good sleep to poor sleep data

Taking a look at the each person individually, the probability of being more likely to sleep either good or poor was of interest. The outcome was when setting the threshold at 80%, which means, that if 80% of the nights of this person were classified as good sleep, then we could say, that this person is more likely to sleep good. The same can be done for a poor sleeper. There might be some people, who fulfill neither of those conditions, they were labeled as unclear. This assessment was also made for a threshold of 90% and is shown in figure 3.3. We can see that until 80% more than half of the participants sleep good in at least 80% of the nights recorded, which changes dramatically when increasing the threshold to 90% where the likely good sleepers decrease by more than 50%.



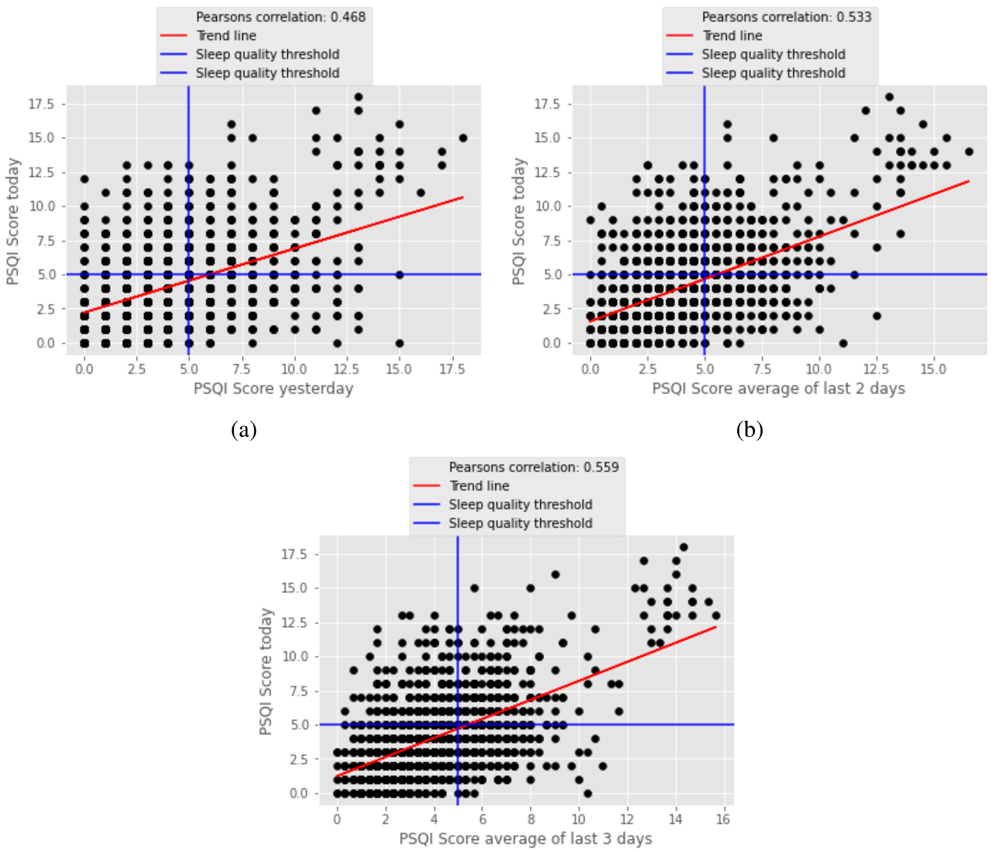
(a)



(b)

Figure 3.3: Number of participants who are more likely to sleep poor, good or none of both with different thresholds (a) 80% (b) 90%

Another thing that was investigated was the relationship of the sleep quality of the last 1, 2 or 3 days in contrast to today's sleep quality across all participants. The outcome was kind of expected, since there is a significant amount of people tending to sleep either good or poor in general. The Pearson correlation coefficient as well as the the trend line is shown in figure 3.4.



(a)

(b)

(c)

Figure 3.4: PSQI Score of sleep today vs
(a) PSQI score of the sleep of yesterday
(b) Average PSQI score of the last 2 days
(c) Average PSQI score of the last 3 days

The first and last hour of the sleep might yield some information about the sleep quality as well. People experiencing poor sleep tend to take longer to fall asleep as well as take more time to leave the bed in the morning. This was explored with the RMS of the sound amplitudes of the first and last hour. The spearman correlation of the RMS of the first hour and the last hour compared to the sleep quality is shown in the following table 3.1. Additionally in figure 3.5 the scatter plot of RMS of the sound amplitude of the first hour vs the last hour is shown. The labels are colored and the densities or shown on the sides of the graph. From the table 3.1 and the figure 3.5 it is obvious that there is slight to no correlation between the sound amplitudes of the first respectively the last hour of the sleep with the sleep quality. There is a tiny more significant correlation of the last hour to the sleep quality in comparison to the first hour.

	Correlation coefficient	**Significance test**
First hour	0.034	0.129
Last hour	0.072	0.001

Table 3.1: Spearman's Rank Correlation Coefficient and significance of the RMS of the first and last hour in correlation with the sleep quality

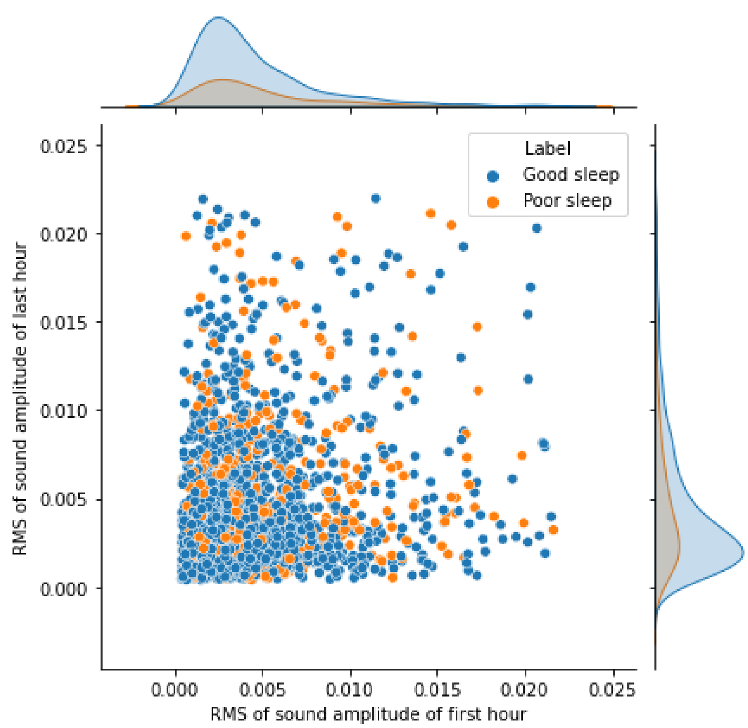


Figure 3.5: Scatter plot of the RMS of the sound amplitude of the first vs the last hour and each point coloured with its label

Decreasing the number of features to the most significant and contributing features is essential to have a rather less complicated model and less noise during training. A famous approach here is extracting the most contributing features using Principal Component Analysis (PCA). In general, a Singular Value Decomposition (SVD) is made and the largest eigenvalues, which are more likely to contribute and contain the most information, are extracted and used for training. In this case 68 principal components came out to explain at least 95% of the variance. The Spearman's rank correlation coefficient is 0.13 for the first principal component and 0.12 for the second principal component, both with a high confidence.

3.2 Hardware and Software

For the training, the Euler cluster of the ETH was used. The whole work was done in python. The framework used for the Machine Learing models is keras, which uses tensorflow as backend. A more detailed list of the packages used are listed in the Appendix A.

3.3 Model

The main model used from chapter 4.3 to chapter 4.5 is a recurrent neural network with 2 LSTM layers and one Dense layer at the end. For the different approaches the parameters were tuned. For the optimizer RMSprop was used. A visualization of the model is shown in figure 3.6. The input layer dimensions may vary from chapter to chapter. As an example an input layer dimension of 7x168 is shown. While the number of features of each day is constant at 168 except in chapter 4.5, the number of days, in this example 7, varies, since for different problem settings, other window day sizes performed better than others.

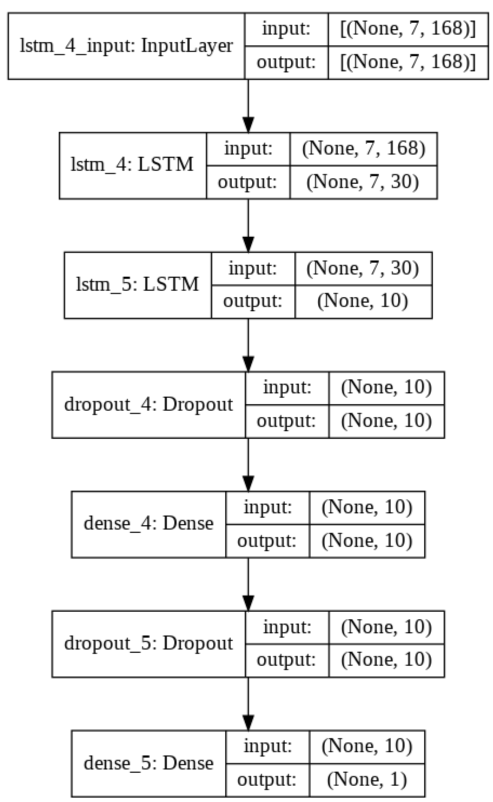


Figure 3.6: RNN Model with the first input layer having a dimension of 7x168, where 168 are the features from the sound amplitudes and 7 depicts the number of days fed into the model.

Chapter 4

Experiments and Results

4.1 Metrics

To be able to evaluate the results and compare them to each other, appropriate metrics had to be chosen. Since the dataset is imbalanced (3:1), this had to be particularly taken into account.

The Matthews correlation coefficient (MCC) has been first introduced in 1975 and has been since then proven its worth [7]. In its simplest form it is used for binary classification tasks especially with medical data where the data imbalance is daily present. It takes the true and false positives and negatives into account and is normalized and ranges therefore from -1 to 1. The greater the absolute value of coefficient the higher the correlation, while a a positive number indicates a positive correlation and a negative number indicates a negative correlation.

The precision recall curve (PR Curve) and its Area under the curve (PR AUC) metric are also well suited for imbalanced datasets in comparison to the widely used ROC [11]. It makes it possible to extract more information related to the performance in an imbalanced dataset.

Another metric used in medical environments are the sensitivity and specificity. While the sensitivity measures the percentage of poor sleepers whom have been correctly identified across all poor sleepers, the specificity measures the same but for the good sleepers. This will help evaluating the performance of the models in predicting good and poor sleepers separately.

In addition to the above metrics the accuracy, a dummy classifier, predicting always the majority class and the F1 Score were taken into account

4.2 Oversampling and Normalization

In order to provide enough data for both classes for the training, oversampling was performed with the adaptive synthetic sampling algorithm (ADASYN) on the minority class, which are the poor sleepers in our case to reduce overfitting. The validation and the test set were left unchanged.
The data was also normalized in advance to enable the model to take all features provided into account.

4.3 Classification model

The first approach was to treat each night as an independent sample and train with 80% and test on 20% of the data to evaluate the prediction performance. For this different model architectures were examined, but the results were, as expected from the data analysis, not astonishing. Here is a quick overview of the results of a RandomForest shown in table 4.1. The model tends to overfit to the data although certain measures were taken into account from the previous sections and the MCC indicates in this case almost no correlation between the predicted labels and the true labels.

MCC	PR AUC	Sensitivity	Specificity	Accuracy	Dummy classifier	F1 Score
0.01	0.34	0.27	0.74	0.65	0.74	0.23

Table 4.1: Results of training a RandomForest

4.4 Time Series Classification

As shown in chapter 3.1 the sleep quality of today might correlate to a certain extend with the sleep of the days before. This raised the idea to use a time series of data of previous days to predict the sleep quality of today. In addition to the between subject approach, a within subject approach was examined where the training data was split in a way, such that 70% of the data of each person was used for training, 15% for validation and 15% for testing. With this approach, the model will have seen data from each person. In figure 4.1 the prediction with the time series is visualized.

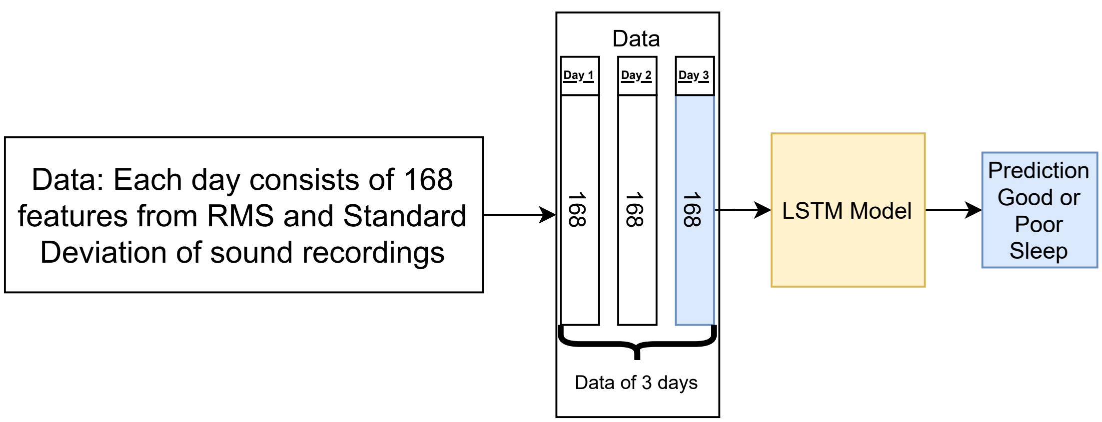
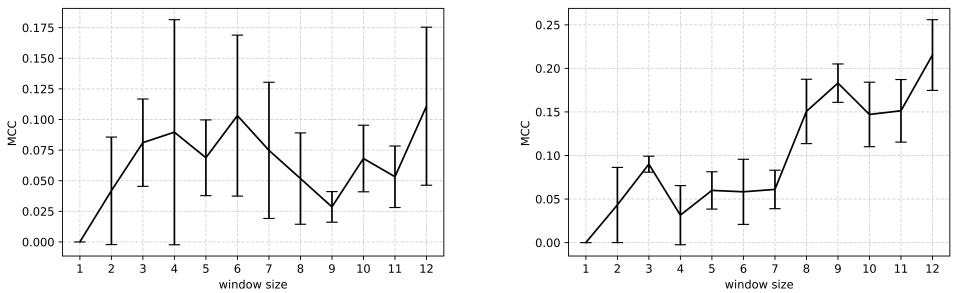


Figure 4.1: Example of a time series with 3 days

These number of days in the past can be variable. In figure 4.1 an example for 3 days is shown. We have explored different number of days and the result is shown in figure 4.2(a). Due to a high standard deviation in the cross validation it is hard to tell which window size performs the best. Therefore the following results in table 4.2 were conducted with a window size of 6, although other window sizes could have been chosen.

In contrast to the figure 4.2(a) the same was conducted within subjects as explained above. The results are displayed in figure 4.2(b).



(a) (b)

Figure 4.2: Matthews correlation coefficient w.r.t. to number of window size of the days
(a) between subjects, (b) within subjects

	MCC	PR AUC	Sensitivity	Specificity	Accuracy	Dummy classifier	F1 Score
between subjects	0.05	0.34	0.24	0.77	0.63	0.76	0.24
within subjects	0.20	0.47	0.36	0.82	0.70	0.76	0.40

Table 4.2: Results of training and predicting between subjects and within subjects

4.5 Time Series Classification with label Feedfoward

To investigate if there is a boost in performance, we decided to add another feature to the time series, which is the respective label of the night. In the first part we used the trained model from above to predict the first night of the person and afterwards a new model was trained, which had a larger input shape due to the additional feature. A visualization of the workflow is shown in figure 4.3.

In figure 4.3 the feed forward is initiated with the original model with an input shape of 168 to predict the label and pass it to the second model which has an input shape of 169. In this example it has been shown for a window size of 3. Here we have also explored different window sizes and the best result was given at a window size of 9. The results are summarized in the table 4.3.

MCC	PR AUC	Sensitivity	Specificity	Accuracy	Dummy classifier	F1 Score
0.14	0.50	0.21	0.79	0.37	0.77	0.32

Table 4.3: Results of feedforwarding the label of each day

In contrast to the above methodology, where a lot of days of data have to sacrificed when the window size increases until we are able to use the rest of the data to predict the sleep quality, a similar approach has been tried out, where the Feedfoward of the label only was done for the label of the last day. The setup is very similar and is shown in figure 4.4.

Since these approaches do increase the complexity of the whole system and make conclusions hard under certain circumstances, we pursued a direct comparison with a similar setup, where the first label was not predicted anymore by a model but rather we fed in the true label of the last day. This way the yellow model in figure 4.4 became obsolete and only the purple model was needed to predict the sleep quality. The

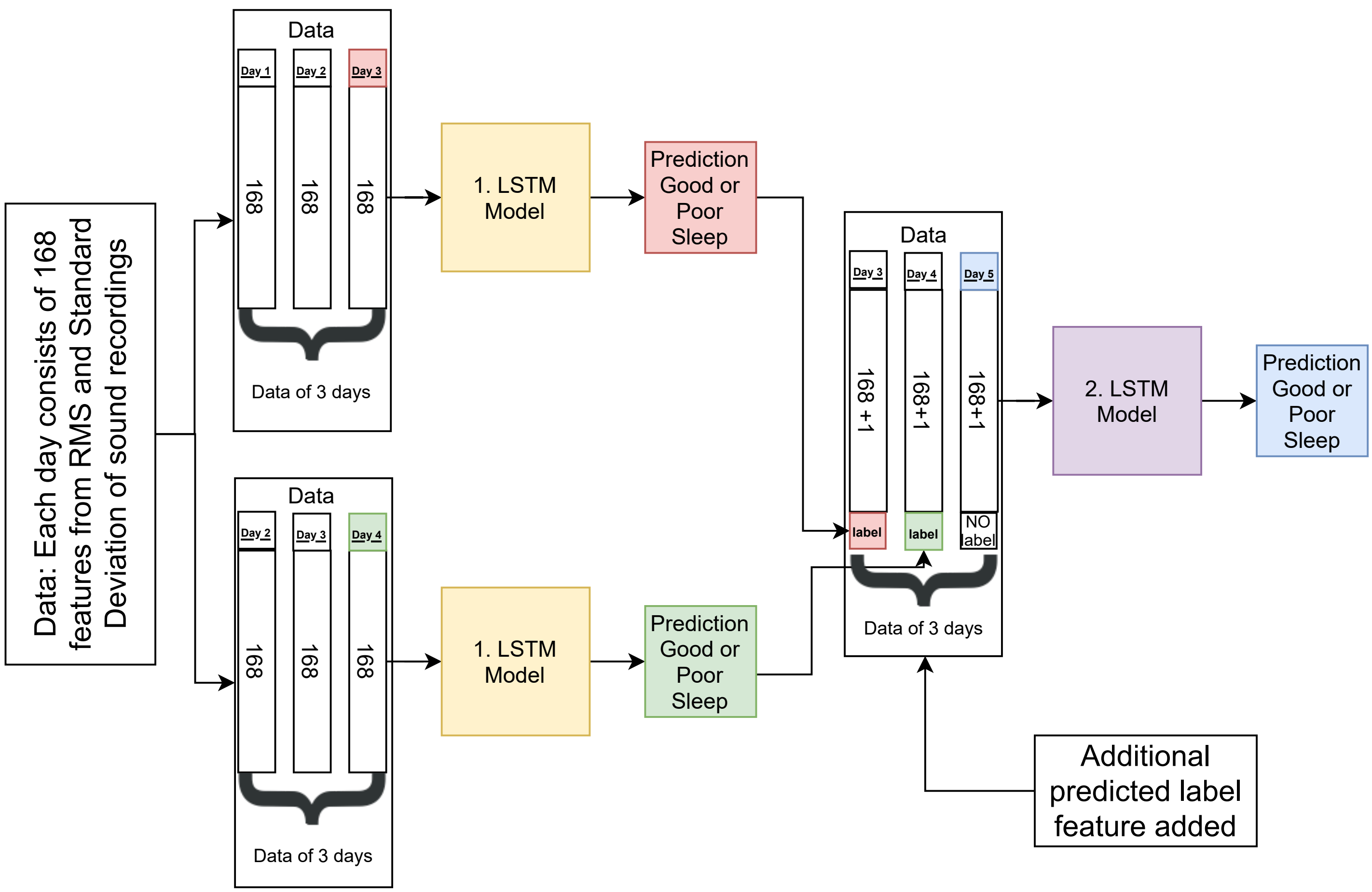


Figure 4.3: Example of a time series with 3 days, feeding every night its predicted label to use model 2 to predict the sleep quality with the additional feature

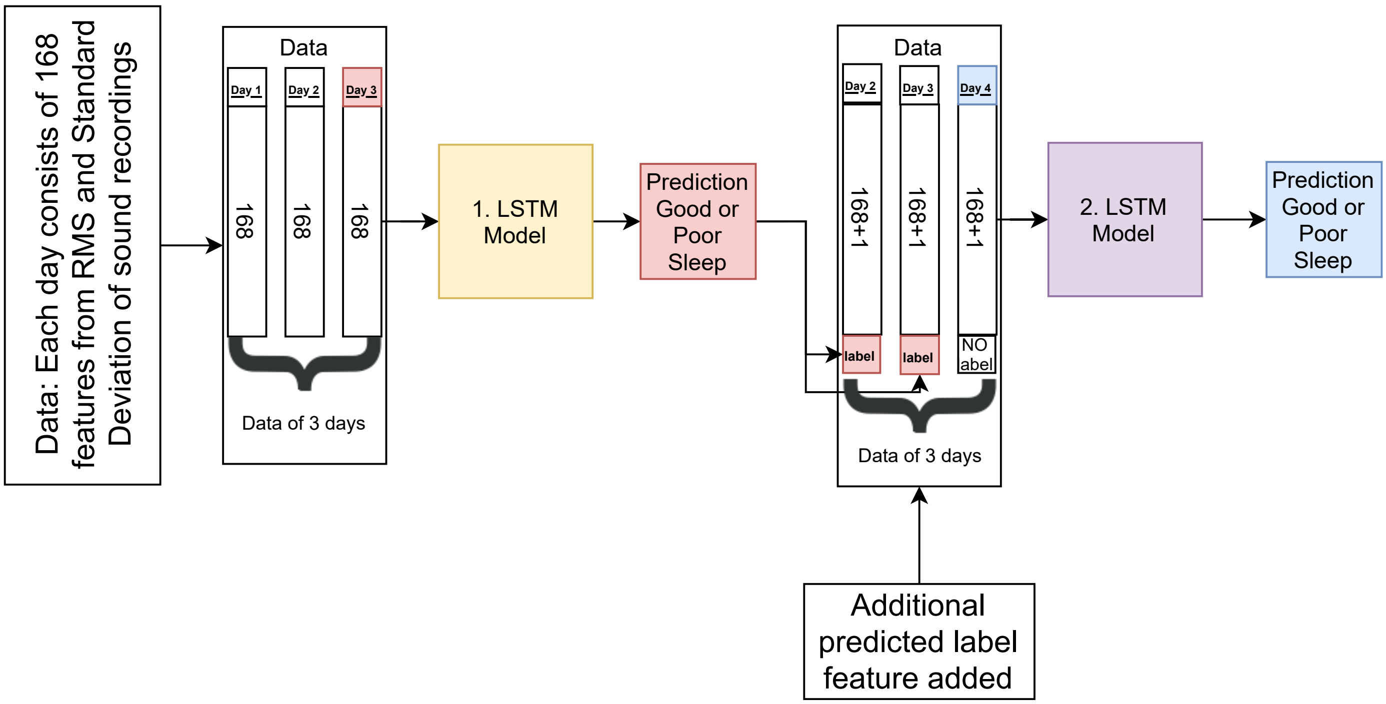


Figure 4.4: Example of a time series with 3 days with a feeding the predicted label of only the last night to use model 2 to predict the sleep quality with the additional feature

new label predicted was fed into the prediction of the next day and so on. The results of the both approaches with a window size of 3 are shown in the following table 4.4.

	MCC	PR AUC	Sensitivity	Specificity	Accuracy	Dummy classifier	F1 Score
Label prediction	0.07	0.40	0.21	0.74	0.46	0.76	0.28
True label	0.09	0.43	0.24	0.78	0.52	0.77	0.30

Table 4.4: Results of the model with predicting and feedforwarding the label and the model when the true label was feedforwarded

4.6 New approach: Predicting sleep quality change

The results with the approach of predicting sleep with hard decision boundary of a global PSQI Score of 5 delivered kind of limited results. A new approach was thus investigated where the objective is if one can predict how the sleep quality changes from day to day. In this case we have a multi class prediction, where the sleep can get better, poorer or stays the same. According to research conducted on the global PSQI score a threshold value of 3 was reached [8]. This means if the global PSQI score of a person does not change by more than 3 points from the day before, one can say, that sleep remained more or less the same or rather did not change significantly. The downside of this can be that if the sleep of a person was very poor at the beginning and becomes slowly better over time, this will not be captured by this model design. Nonetheless we will be investigating the model, if it is able to capture significant changes in the sleep quality.

The model used here is also an 2 LSTM layer and 1 Dense Layer, and parameter tuning was also performed to adjust to the new model approach. The results were conducted for a threshold value of 2,3 and 4. The results are summarized in table 4.5. The sensitivity and specificity include 3 entries, where the first entry represents the value for the first class ”no significant sleep quality change”, the second entry represents the second class ”sleep gets significantly better” and the third entry represents the third class ”sleep gets significantly poorer”.

	MCC	**PR AUC**	**Sensitivity**	**Specificity**	**Accuracy**	**Dummy classifier**	**F1 Score**
TH: 2	0.05	0.47	[0.39, 0.26, 0.44]	[0.69, 0.74, 0.63]	0.37	0.52	0.37
TH: 3	0.04	0.42	[0.4, 0.26, 0.25]	[0.59, 0.67, 0.71]	0.36	0.71	0.39
TH: 4	0.04	0.43	[0.35, 0.35, 0.25]	[0.64, 0.67, 0.68]	0.34	0.84	0.41

Table 4.5: Results of the 3 classification model with different threshold values

Chapter 5

Discussion & Conclusion

We have seen by now different ways in tackling the research question. From chapter 4.3 to chapter 4.5 different approaches were investigated in predicting the sleep quality of a person. This was introduced with a simple RandomForest classifier which had a relatively low MCC of 0.006 and an F1 Score of 0.23. The time series summarized in table 4.2 performed similarly to the RandomForest. Afterwards the label of each day in the time series was included (table 4.3) which brought a significant improvement in MCC and F1 Score, but had a relatively low accuracy.

Feedforwarding only the label of the last day as in table 4.4 suffered a little bit in MCC and the F1 Score, although these minor differences still lie in the standard deviation. Using the predicted label and feedforward it to the next model performed just slightly worse than feeding the true label in, which might indicate, that the prediction was quite good. But this stays in contrast to the results from before in table 4.2, that the model prediction was poor. One reason might be, that the model did not value the new feature as much as we expected.

In table 4.2 the same model was trained but the first one with a between subjects and the other one with a within subject approach. The differences were quite significant, indicating that if the model has seen some data of each person, it might be able to perform better. Additionally from figure 4.2 one may derive that the greater the time series was, the better it performed.

The last experiment in chapter 4.6, where the objective was changed compared to the other models, the results were comparable to what we have seen by now. In all three cases with different thresholds the results were close to each other. Therefore varying the threshold did not seem to change the result.

In all tasks similarities have been observed. The specificity, which in our case was how well the good sleepers were identified was always above 0.7, while the sensitivity, which describes how well poor sleepers were identified, was between 0.2 and 0.3. This is in alignment with other research [10], that predicting poor sleepers was the challenging task.

In a nutshell, objectively the performance was relatively poor from all models compared to related work introduced in the beginning with a similar dummy classifier performance and a model performance with an F score of 0.81 and an accuracy of 81.48% [10]. One should note that additional data was fed into the model e.g. the acceleration data during the day. From other research it is not clear which features dominated the prediction of the models. But one may conclude that either the model selection or the approach was not suitable. Another reason might be, that only sound amplitudes do not have enough information to be able to predict the sleep quality.

If the classification results can be improved in the between subject case, a general model can be trained and an app could be developed, where the trained model could be deployed. People will get an information of their sleep quality of their day and can track it on a daily basis to try to improve their sleep in the next following days if needed.

If the classification in the within subject came ou t to be more promising, building the application will be more challenging. Training will be required from the first couple of days of each participant in addition to the PSQI questionnaire which has to be filled out everyday to obtain the score.

Further research could try different approaches, especially in the model selection or in collecting different types of data and investigate how much does every feature contribute to the prediction and what combination of features results in a good prediction. Furthermore a prediction into the future, which means, predicting the sleep quality of the next following days, could be pursued. Last but not least a similar approach to transfer learning could be explored. A general model could be trained with many participants data, and afterwards the same model could be retrained with individual data, potentially decreasing the amount of data required for training.

Appendix A

The First Appendix

The main packages used were the following, a more detailed overview is given in the code:

- keras, tensorflow, sklearn, scipy, statistics, numpy, seaborn

Bibliography

[1] Daniel J Buysse, Charles F Reynolds III, Timothy H Monk, Carolyn C Hoch, Amy L Yeager, and David J Kupfer. Quantification of subjective sleep quality in healthy elderly men and women using the pittsburgh sleep quality index (psqi). *Sleep*, 14(4):331–338, 1991.

[2] Zhenyu Chen, Mu Lin, Fanglin Chen, Nicholas D Lane, Giuseppe Cardone, Rui Wang, Tianxing Li, Yiqiang Chen, Tanzeem Choudhury, and Andrew T Campbell. Unobtrusive sleep monitoring using smartphones. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, pages 145–152. IEEE, 2013.

[3] Yuriko Doi, Masumi Minowa, Makoto Uchiyama, Masako Okawa, Keiko Kim, Kayo Shibui, and Yuichi Kamei. Psychometric assessment of subjective sleep quality using the japanese version of the pittsburgh sleep quality index (psqi-j) in psychiatric disordered and control subjects. *Psychiatry research*, 97(2-3):165–172, 2000.

[4] Liese Exelmans and Jan Van den Bulck. Bedtime mobile phone use and sleep in adults. *Social Science & Medicine*, 148:93–101, 2016.

[5] Max Hirshkowitz. The history of polysomnography: tool of scientific discovery. In *Sleep medicine*, pages 91–100. Springer, 2015.

[6] M Alexandra Kredlow, Michelle C Capozzoli, Bridget A Hearon, Amanda W Calkins, and Michael W Otto. The effects of physical activity on sleep: a meta-analytic review. *Journal of behavioral medicine*, 38(3):427–449, 2015.

[7] B.W. Matthews. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 405(2):442 – 451, 1975.

[8] Lucy M McDonnell, Lauren Hogg, Lynn McDonnell, and Patrick White. Pulmonary rehabilitation and sleep quality: a before and after controlled study of patients with chronic obstructive pulmonary disease. *NPJ Primary Care Respiratory Medicine*, 24(1):1–5, 2014.

[9] Goran Medic, Micheline Wille, and Michiel EH Hemels. Short-and long-term health consequences of sleep disruption. *Nature and science of sleep*, 9:151, 2017.

[10] Jun-Ki Min, Afsaneh Doryab, Jason Wiese, Shahriyar Amini, John Zimmerman, and Jason I Hong. Toss'n'turn: smartphone as sleep and sleep quality detector. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 477–486, 2014.

[11] Brice Ozenne, Fabien Subtil, and Delphine Maucort-Boulch. The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *Journal of clinical epidemiology*, 68(8):855–859, 2015.

[12] Bong Soo Park, Il Hwan Kim, and Kang Min Park. Top 100 cited articles on sleep medicine: A bibliometric analysis. *European Neurology*, 83(1):111–120, 2020.

[13] Aarti Sathyanarayana, Shafiq Joty, Luis Fernandez-Luque, Ferda Ofli, Jaideep Srivastava, Ahmed Elmagarmid, Teresa Arora, and Shahrad Taheri. Sleep quality prediction from wearable data using deep learning. *JMIR mHealth and uHealth*, 4(4):e125, 2016.

[14] Wei Zhao, Xinxi Wang, and Ye Wang. Automated sleep quality measurement using eeg signal: first step towards a domain specific music recommendation system. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1079–1082, 2010.