UNIVERSITY OF WEST BOHEMIA

# Automatic detection of sleep spindles by deep learning methods

Jan Rychlik
rychlikj@students.zcu.cz

In Pilsen 30.09.2021

# Introduction

Scoring sleeping EEG data and identifying sleeping spindles is a necessary part of analyzing human sleep. However, detection of sleep spindles in large datasets by human experts is time-consuming and costly. Machine learning algorithms to detect spindles are cost-efficient and reproducible. Also, the annotations using machine learning methods achieve a positive outcome.

A master thesis focused on the annotation of sleep spindles originated within the neuroinformatics research group at the University of West Bohemia. Neuroinformatics research group deals with the processing of biosignals, especially electroencephalography (EEG) signals and event-related potentials (ERPs)**.** The group has been active since 2008 and has about twenty members. The master thesis will be processed by a student of the medical informatics study program, Jan Rychlík, and will mainly focus on sleep data processing. The supervisor of the master thesis is associate professor Roman Mouček and the reviewer will be Lukáš Vařeka, Ph.D.

# State of the art

The article ‘Massive online data annotation, crowdsourcing to generate high-quality sleep spindle annotations from EEG’ [1] mainly describes the MASS (Montreal Archive of Sleep Studies) dataset and its origin. Datasets were scored/annotated by experts and non-experts; the machine learning methods with positive output to scoring sleep spindles are explained.

A group of scientists at FH Aachen University of Applied Sciences used deep learning methods to identify and classify three phases of animal (mice) sleep (Wake, REM, Non-REM)[2].

Technol Health Care [4] used the LSTM and CNN neural networks to annotate three phases of human sleep (Wake, REM, Non-REM) on Sleep-EDF dataset. The achieved accuracy was 93.47% using the Fpz-Cz electroencephalogram channel.

The previous articles described the use of neural networks for three phases of sleep only. **We plan to use a neural network to identify sleep spindles themselves.**

CNN networks are commonly used to process EEG data. They have achieved great success, for example, in the identification of epileptic patients.[6]

In general, LSTM and CNN networks have proven to be suitable for the analysis of EEG signals.[7]

## Sleep datasets available

### Sleep-EDF

The sleep-EDF database is an open-source dataset and contains 197 whole-night PolySomnoGraphic sleep recordings, containing EEG, EOG, chin EMG, and event markers. Some records also contain the values of respiration and body temperature. Corresponding hypnograms (sleep patterns) were manually scored by well-trained technicians according to the Rechtschaffen and Kales manual, and are also available. [5]

### Mass-dataset

The Montreal Archive of Sleep Studies (MASS) is an open-access and collaborative database of laboratory-based polysomnography (PSG) recordings O'Reilly, C., et al. (2014) J Seep Res, 23(6):628-635. Its goal is to provide a standard and easily accessible source of data for benchmarking the various systems developed to help the automation of sleep analysis. It also provides a readily available source of data for fast validation of experimental results and for exploratory analyses. Finally, it is a shared resource that can be used to foster large-scale collaborations in sleep study.

### Mice dataset

The dataset consists of polysomnographic recordings of 18 mice with a total recording duration of 52 days (corresponding to 454,301 windows (segments) of 10 seconds each, see Table 1). Each mouse was recorded for 3 days, except for one mouse whose recording spanned 1 day because it unexpectedly passed away. The data were acquired during a study at the University of Tübingen which tested the influence of dietary variations on sleep. This dataset is introduced only as an example on analyzing sleep EEG data by neural network. Because the data comes from mices, it can't be used in our experiment.

# Methods

The following preprocessing and processing methods were used on the mentioned datasets.

## Sleep-EDF

Long Short-Term Memory (LSTM), convolutional neural networks (CNNs), and their combination were used for Sleep-EDF dataset processing.
LSTM is a kind of time recurrent neural network. LSTM can process and predict important events with long intervals and delays in time series. LSTM includes the input gates, the output gates, and the forget gates.

CNN is frequently used to recognize two-dimensional images but it was found successful also for EEG patterns processing. It combines the local perception, weight sharing, and down sampling. Local perception allows hidden units not to be full connections. Weight sharing enables different signals to share a convolution kernel.

The proposed CNN-LSTM algorithm contains 7 layers to realize automatic sleep staging. Four layers of 1D CNN and three layers of LSTM were used.

## Mass-dataset

Mostly preprocessing methods were used for mass-dataset.

| |
|---|
| Band-pass filters (11–15 Hz) the EEG signal to compute its envelope. An upper (8 x mean) and lower (2 × mean) thresholds are used to detect spindles.<br>*Code available on github.com/swarby/SpindleAlgorithms_NatMeth_2014* |
| Band-pass filters (11.3–15.7 Hz) the EEG signal to compute the root mean squared (RMS) on sliding windows (100 ms length with a step of 50 ms), and then applies a threshold (1.5 × STD).<br>*Code available on github.com/swarby/SpindleAlgorithms_NatMeth_2014* |

| |
|---|
| Band-pass filters (11–15 Hz) the EEG signal to compute the RMS on sliding windows (25 ms length with a step of 25 ms), and then applies a threshold (95th percentile).<br>*Code available on github.com/swarby/SpindleAlgorithms_NatMeth_2014* |
| Transforms the EEG signal into continuous Morlet wavelets to compute the moving average on sliding windows (0.1 sec length), and then applies the threshold (4.5 × mean).<br>*Code available on github.com/swarby/SpindleAlgorithms_NatMeth_2014* |
| Computes the absolute (Mean Square) sigma (11–16 Hz) power, the relative sigma power with Power Spectral Analysis, the covariance and correlation between sigma filtered and the unfiltered EEG signal on sliding windows (0.3 sec length with a step of 0.1 sec). It then detects a spindle if the 4 features extracted from EEG exceed their respective threshold (1.25 $\mu V^2$, 1.6 × STD, 1.3 × STD and 69%).<br>*Code available on github.com/swarby/A7_LacourseSpindleDetector* |
| Decomposes the EEG signal into 3 components: Direct Current, oscillation around 13.5 Hz and other frequency components (0.3–30 Hz). Spindles are detected from the oscillation around 13.5 Hz with an upper (2.33 × STD) and lower (0.1 × STD) thresholds applied in sliding windows (60 sec length).<br>*Code available on github.com/stuartfogel/detect_spindles* |
| Decomposes the EEG signal into 3 components: transient (t), low-frequency (lf) and oscillations (s). *S* are represented sparsely with Short Time Fourier Transform (1.28 sec length with a step of 0.32 sec). It then detects spindles by thresholding (c1 = 0.03) the Teager-Kaiser energy operator (energy smooth) of *s* band-pass filtered (11.5–15.5 Hz). Parameters initialization: lambda0 = 0.6, lambda1 = 7, lambda2 = 8.5, mu = 0.5 and c1 = 0.03.<br>*Code available on github.com/aparek/detok* |

### Mice dataset

A simple neural network was used to annotate the dataset, which contained two hidden layers of 113 and 96 neurons.

# Solution

Successful annotations of data and classification results described above encouraged us to use preprocessing and processing methods including deep learning algorithms to identify sleep spindles in the sleep datasets available.

LSTM (Long Short-Term Memory) neural networks and convolutional neural networks (CNN) neural networks have been chosen as the best candidates to identify sleep spindles mainly due to the following reasons. LSTM neural networks are recurrent neural networks developed to work with sequences, signals, etc. This type of network is thus suitable for processing time series. CNN is a class of artificial neural networks, originally and most commonly applied to analyze visual imagery. However, CNN networks have also proven to be suitable to analyze EEG signals as described in State of the art.

The neural networks mentioned above will be trained on a data set annotated in 2014 during the project: Montreal Archive of Sleep Studies: an open-access resource for instrument benchmarking and exploratory research [3].

Since, generally, men have fewer spindles than women and the frequency of spindles also decreases with age, these parameters will be used as inputs to the neural networks mentioned above; therefore complete data will help to achieve more reasonable classification results(Biosignals and sleep stages, Subset-specific annotations, Open-access descriptors, Restricted-access descriptors).

The training of neural networks will follow a standard procedure using cross-validation. The data set will be divided into a training and test data set.

## Data

Annotated data from the MASS group [8] presented in the research paper: Massive online data annotation, crowdsourcing to generate high quality sleep spindle annotations from EEG [1] will be used as a dataset. The data will be handled in accordance with the code of ethics and meet the licensing conditions of the MASS group [8].

# Success criteria

The project aims to verify if it is possible to annotate sleep spindles by using automated machine/deep learning methods, specifically the neural network described above.

The project outcomes will be supplemented by a written master thesis describing the state of the art and opportunities, methods and results when using deep learning methods (neural networks) for identification of sleep spindles in EEG data. The common metrics will be used when evaluating the results. The results will be compared to the results of State of the art processing methods. The content of the diploma thesis will be finally used as a base for a conference or journal paper.

# References


[1] Lacourse, K., Yetton, B., Mednick, S. *et al.* Massive online data annotation, crowdsourcing to generate high quality sleep spindle annotations from EEG data. *Sci Data* 7, 190 (2020). https://doi.org/10.1038/s41597-020-0533-4

[2] Grieger, N., Schwabedal, J.T.C., Wendel, S. *et al.* Automated scoring of pre-REM sleep in mice with deep learning. *Sci Rep* 11, 12245 (2021). https://doi.org/10.1038/s41598-021-91286-0

[3] Christian O'Reilly, Nadia Gosselin, Julie Carrier, Tore Nielsen, 09 June 2014 https://doi.org/10.1111/jsr.12169

[4] Zhao D, Jiang R, Feng M, Yang J, Wang Y, Hou X, Wang X. A deep learning algorithm based on 1D CNN-LSTM for automatic sleep staging. Technol Health Care. 2021 Jun 25. DOI: 10.3233/THC-212847. Epub ahead of print. PMID: 34180436.

[5]Niklas Grieger, Justus T. C. Schwabedal, Stefanie Wendel, Yvonne Ritze & Stephan Bialonski. Automated scoring of pre-REM sleep in mice with deep learning

https://doi.org/10.13026/C2X676

[6] Wei, X., Zhou, L., Chen, Z. *et al.* Automatic seizure detection using three-dimensional CNN based on multi-channel EEG. *BMC Med Inform Decis Mak* 18, 111 (2018). https://doi.org/10.1186/s12911-018-0693-8

[7] Li, Gen & Lee, Chang & Jung, Jason & Youn, Young Chul & Camacho, David. (2019). Deep learning for EEG data analytics: A survey. Concurrency and Computation: Practice and Experience. 32. 10.1002/cpe.5199.

[8] MASS group http://ceams-carsm.ca/en/mass/