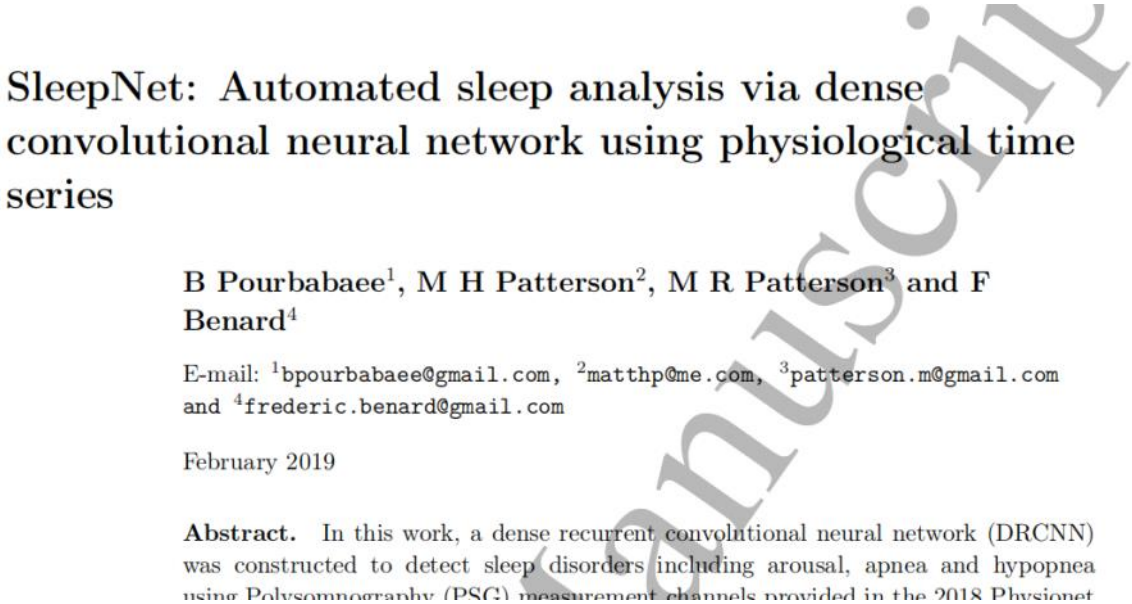
Healthcare Presentation

Monday, March 11, 2024 2:38 PM



SleepNet: Automated sleep analysis via dense convolutional neural network using physiological time series

B Pourbabaee[1], M H Patterson[2], M R Patterson[3] and F Benard[4]

E-mail: [1]bpourbabaee@gmail.com, [2]matthp@me.com, [3]patterson.m@gmail.com and [4]frederic.benard@gmail.com

February 2019

Abstract. In this work, a dense recurrent convolutional neural network (DRCNN) was constructed to detect sleep disorders including arousal, apnea and hypopnea using Polysomnography (PSG) measurement channels provided in the 2018 Physionet

Introduction:

The introduction of the research paper "SleepNet: automated sleep analysis via dense convolutional neural network using physiological time series" highlights the critical role of sleep quality in maintaining good health and its linkage with various negative health outcomes, including depression, obesity, and cardiovascular diseases. It emphasizes the prevalence of sleep disorders like apnea and hypopnea, which significantly impact sleep quality and have been extensively studied. The paper notes the high reliability of detecting these events using Polysomnography (PSG), a standard method for investigating sleep quality through multiple physiological signals.

Furthermore, the introduction addresses less studied sleep disruptions, such as non-apnea/hypopnea arousals, which can be caused by various factors like respiratory effort, teeth grinding, and muscle jerks, among others. The paper underlines the challenges and high costs associated with manual detection of such arousals, advocating for an automated detection method to facilitate research and improve health outcomes.

The paper reviews past efforts in automated arousal detection, including frequency analysis of EEG channels and various machine learning approaches, noting the advantages of deep learning in handling large datasets and reducing the need for manual feature engineering. The research aims to assess the accuracy of deep learning in detecting non-apnea/hypopnea arousals, acknowledging the potential for larger scale studies to better understand the health risks associated with arousal frequency.

Methodology:

The methodology section of the paper introduces a sophisticated approach to sleep analysis using a dense recurrent convolutional neural network (CNN), specifically tailored for detecting arousal regions alongside apnea/hypopnea and sleep/wake intervals within Polysomnography (PSG) data. This network is an enhancement of the DenseNet architecture, featuring multiple Dense Convolutional Units (DCUs). These units are designed such that each convolutional layer within a unit is directly connected to every other layer, ensuring maximal flow of information and facilitating feature learning from the physiological time series data.

A notable aspect of the proposed network is the integration of a bidirectional Long Short-Term Memory (LSTM) layer. This LSTM layer is augmented with a residual skip connection and additional convolutional layers to adapt the LSTM's output to the required shape for further processing. This design choice aims to capture both the temporal dynamics and complex feature interactions in the physiological signals indicative of sleep events.

To assist in the training and decision-making processes, the authors introduce a remapping mechanism. This mechanism simplifies the network's output interpretation, making it more straightforward to compute the probabilities of different sleep events and the associated losses during training. Furthermore, the methodology leverages a multi-task learning framework, employing auxiliary tasks such as distinguishing between apnea-hypopnea/normal breathing and sleep/wake states. This multi-task approach is intended to share learned representations across related tasks, thereby enhancing the model's overall generalization capability and improving its performance on the primary task of arousal detection.

Labeling:
In the context of this research, each sleep event (like arousal, apnea/hypopnea, normal breathing, etc.) is assigned a specific label within a vector to categorize the type of event occurring at any given time during sleep. For example:

- Target arousal detection task: The labels could be something like 1 for target arousal events, -1 for non-target arousal events (which could include apnea/hypopnea or wakefulness), and 0 for normal, undisturbed sleep. Target arousals in the context of this research paper refer to sleep disruptions or disturbances that are distinct from apnea and hypopnea events.
- Apnea-hypopnea/normal detection task: Labels might be 1 for any apnea-hypopnea events and 0 for normal breathing.
- Sleep/wake detection task: Labels could be 1 for various sleep stages (like REM, NREM1, NREM2, NREM3) and 0 for wakefulness, with -1 possibly used for undefined stages.

Bins and Remapping Mechanism:
Given the labeling system, the researchers then organize these labels into "bins" to manage the output of the neural network more effectively. Each bin represents a combination of labels corresponding to the different tasks (e.g., arousal detection, apnea-hypopnea detection, sleep/wake detection).

However, not all combinations of labels might occur in the actual data, leading to some bins being "empty." To manage this, a "remapping" mechanism is used, where:

- Non-empty bins: These are bins with data that correspond to actual, observed combinations of sleep events. These are directly used in the model's output layer to compute probabilities for each event combination.
- Empty bins: These bins don't correspond to observed data and are effectively ignored in the analysis, reducing complexity.
- Remapping: Some bins are "remapped" to simplify the model further. For instance, if a bin initially represents a combination of labels that is either very rare or doesn't make logical sense (like having an arousal event during wakefulness), it might be remapped to a more common or logical bin. This process helps in simplifying the network's output layer, making it easier to interpret the results and reducing the computational burden.

Materials and Pre-processing:

The materials and preprocessing section details the dataset and the steps taken to prepare the physiological signals for analysis. The dataset comprises PSG data from 1,985 subjects, recorded at the MGH sleep laboratory for sleep disorder diagnosis purposes. The data is split into two sets; the first is used for training, validation, and testing of the model, while the second serves as a blind test set for evaluating the models' performance in the challenge context.

For the analysis, the authors exclude the electrocardiogram (ECG) signal, focusing on 12 other channels relevant to sleep scoring. The preprocessing routine starts with the application of an anti-aliasing finite impulse response (FIR) filter to all channels, aimed at reducing noise and preventing aliasing. This is followed by downsampling the signals to 50 Hz and removing the DC bias to center the signals around zero.

A critical step in the preprocessing is the normalization of the signals, conducted on an 18-minute moving window for each channel. This normalization involves subtracting the mean and dividing by the root-mean-square (RMS) value of the signal within the window, employing fast Fourier transform (FFT) convolution for efficiency. The choice of an 18-minute window ensures a 90% overlap between consecutive baseline windows, maintaining the integrity of significant breathing variations. Oxygen saturation (SaO2) measurements are handled differently, being scaled to fall within a specific range to prevent overwhelming the neural network with large values. This meticulous preprocessing is designed to standardize the physiological signals, making them conducive for analysis by the dense recurrent CNN

NREM Sleep: This is divided into three stages, NREM1, NREM2, and NREM3, each progressively deeper than the last.

- NREM1: This is the lightest stage of sleep, often considered the transition phase between wakefulness and sleep. It's characterized by slow eye movements, reduced muscle activity, and the ability to be easily awakened. People in this stage might experience sudden muscle contractions known as hypnic jerks.

- NREM2: This stage accounts for the majority of our sleep time. It is deeper than NREM1 and features specific patterns on an EEG, including sleep spindles and K-complexes. There are no eye movements, and heart rate and breathing slow down.

- NREM3: Often referred to as deep sleep or slow-wave sleep, NREM3 is the deepest stage of NREM sleep. It's characterized by delta waves on an EEG, slow breathing, and reduced heart rate and muscle activity. Waking someone from this stage can be difficult, and if awakened, a person might feel disoriented for a few minutes.

REM Sleep: This stage is known for rapid eye movements, increased brain activity, and vivid dreams. It's characterized by paralysis of most voluntary muscles (REM atonia), preventing people from acting out their dreams. REM sleep typically begins around 90 minutes after falling asleep and recurs about every 90 minutes, getting longer later in the night.

The cycle of NREM and REM sleep repeats itself about 4-6 times throughout a typical night, with variations in duration and intensity as the night progresses. Each stage plays a crucial role in various brain and body restoration functions, memory consolidation, and emotional regulation

Sleep Disorder Detector Model:

The Sleep Disorder Detector Model section of the paper elaborates on the design and functioning of the proposed Dense Recurrent Convolutional Neural Network (DRCNN) aimed at identifying sleep-related disorders, focusing on arousal detection, apnea/hypopnea identification, and sleep/wake classification. This section delves into the DRCNN's architecture, highlighting its multi-task learning approach to leverage various sleep event annotations for enhanced network generalization.

DRCNN Network Structure
The DRCNN architecture is optimized for efficiency, handling full-night recording data downsampled to 50 Hz to reduce computational demands. It comprises multiple blocks, including Dense Convolutional Units (DCU1 and DCU2) and a Long Short-Term Memory (LSTM) layer, each serving distinct functions in processing the input signals. Initial layers (DCU1s) feature max-pooling to downsample signals, facilitating the handling of 50 Hz input data at a reduced resolution of 1 Hz. Subsequent DCU2 blocks incorporate depthwise separable convolutions and scaled exponential linear unit (SELU) activations, enriched with weight, position-wise, and stochastic batch normalization techniques to enhance the network's ability to learn complex features from the data. Dilated convolutions expand the receptive field within DCU2s, with dilation rates varying across the network to capture a broad range of signal features. The LSTM layer, equipped with a residual skip connection, processes temporal dependencies in the data, with additional convolutions adapting LSTM outputs for further analysis. The network employs weight normalization and hyperbolic tangent (tanh) activation for stability during training.

Learning Mechanism
The multi-task learning framework is a cornerstone of the DRCNN's design, allowing the network to leverage correlations between different sleep events to improve overall performance and generalization. This approach uses ground truth vectors representing various sleep conditions (target arousal, non-target arousal, normal, apnea/hypopnea, and wake states) to define multiple tasks within the network. An innovative bin remapping mechanism simplifies the network's output layer, enabling the computation of joint probabilities for different sleep conditions. This facilitates the handling of complex multi-label scenarios inherent in sleep data.

Training the DRCNN involves balancing the contributions of the primary task (target arousal detection) and auxiliary tasks (apnea-hypopnea/normal and sleep/wake detection) using a weighted cross-entropy loss function. The optimization process, utilizing the Adam method, focuses on iteratively improving the model based on performance metrics on a validation set, employing a checkpointing mechanism to save the best-performing model configurations.

The network's efficacy is assessed using area under the precision-recall curve (AUPRC) and area under the receiver operating characteristic curve (AUROC) metrics, with performance enhancements achieved through ensemble modeling based on multiple training/validation data folds. The ensemble approach averages predictions from multiple DRCNN instances trained on different data subsets, aiming to boost the model's robustness and predictive accuracy.

This sophisticated model design, incorporating advanced neural network techniques and a multi-task learning strategy, underscores the potential of deep learning in transforming sleep disorder detection and analysis, offering insights into the complex interplay of physiological signals during sleep.

Empirical Results: GitHub Results: https://github.com/matthp/Physionet2018_Challenge_Submission

The empirical results section outlines the performance of the Dense Recurrent Convolutional Neural Network (DRCNN) in analyzing sleep disorders using Polysomnography (PSG) data. The DRCNN was trained on PSG data, excluding the ECG signal, across four folds to ensure robustness and reliability. Each fold included distinct sets of training, validation, and testing records, ensuring comprehensive evaluation.

The network's effectiveness was gauged using the Area Under the Precision-Recall Curve (AUPRC) and the Area Under the Receiver Operating Characteristic Curve (AUROC) across different tasks: sleep/wake detection, target arousal detection, and apnea-hypopnea/normal detection. The performance metrics were averaged across the four folds to present a unified overview of the model's capabilities.

Key findings from the empirical evaluation include:

- For the **target arousal detection task**, the ensemble model achieved an AUROC of **0.931** and an AUPRC of **0.543** on the test set, indicating a significant capability in identifying arousal events from PSG data.
- In the **apnea-hypopnea/normal detection task**, the ensemble model's performance was robust with an AUROC of **0.965** and an AUPRC of **0.783**, showcasing high accuracy in differentiating between apneic events and normal breathing.
- For the **sleep/wake detection task**, the ensemble model demonstrated high efficiency with an AUROC of **0.960** and an AUPRC of **0.832,** effectively distinguishing sleep stages.

Additionally, the DRCNN model was evaluated for its ability to estimate the severity of apnea-hypopnea conditions using standard sleep medicine metrics: sleep efficiency (SE), arousal index (AI), and apnea-hypopnea index (AHI). The model's predictions closely mirrored actual clinical assessments, with mean absolute errors (MAE) for these metrics within acceptable ranges, indicating the model's practical utility in clinical settings.

The confusion matrices for apnea-hypopnea severity grade estimation provided further insights into the model's diagnostic accuracy, with the ensemble approach showing a balanced classification across different severity grades of sleep apnea.

Overall, the DRCNN demonstrated promising results in automated sleep analysis, with the ensemble model outperforming individual models in most tasks. These findings underscore the potential of advanced neural network architectures in enhancing sleep disorder diagnostics and treatment planning.

Discussion:

The discussion section reflects on the DRCNN's performance in detecting sleep disorders, emphasizing the ensemble model's enhanced capability in identifying target arousal, apnea-hypopnea, and sleep/wake intervals. The DRCNN's performance, particularly in target arousal detection with an AUPRC of 0.543, is commendable and stands out among entries from the 2018 Physionet challenge, except for one model that achieved a higher AUPRC through hyper-parameter search outside the official challenge timeline.

The multi-task learning approach not only bolstered the detection of arousal regions but also facilitated the accurate identification of other sleep disorders, evidenced by the model's reasonable estimations of sleep efficiency (SE), arousal index (AI), and apnea-hypopnea index (AHI). These estimations suggest the DRCNN could be instrumental in generating automated sleep monitoring reports with minimal errors.

The model's tendency to overestimate the severity of apnea-hypopnea cases is highlighted as preferable to underestimations or missed detections, as false positives are easier for technicians to correct. Despite the model's primary focus on arousal detection, it also shows satisfactory performance in detecting apnea/hypopnea events and sleep/wake states, although the study acknowledges that dedicated models or a shift in focus might yield even better results for these specific tasks.

Comparisons with other studies indicate varying degrees of success in using machine learning and deep learning for sleep disorder detection, with some models achieving high accuracy, specificity, and sensitivity in similar tasks. However, the discussion also points out the need for a more direct comparison to human performance, given the variability in human annotations and the limitation of the dataset being annotated by a single scorer.

The single-site nature of the dataset poses a significant limitation to the generalizability of the model, emphasizing the need for additional data from various sites to ensure broader applicability. Moreover, the reliance on a specific set of PSG channels may restrict the model's use with different equipment or in settings where these channels are not available.

The discrepancy noted between the AUPRC and AUROC metrics suggests a trade-off between sensitivity and precision, indicating a challenge in maintaining high precision in a dataset where normal cases significantly outnumber abnormal ones. This observation underscores the complexities of sleep disorder detection and the need for continued refinement of machine learning models to improve their clinical utility.

Conclusion:

In the conclusion, the paper summarizes the development and evaluation of a modified dense convolutional neural network, designed to diagnose sleep disorders such as arousal, apnea, and hypopnea from 12 PSG channels provided in the 2018 Physionet challenge dataset. This network integrates multiple convolutional and LSTM blocks and employs a multi-task learning approach with hard parameter sharing to leverage correlated task information, enhancing the model's generalization capabilities.

The study trained and evaluated four DRCNN models on diverse subsets of the data, culminating in an ensemble model that averages the predictions of these models. The ensemble approach proved superior, outperforming individual model strategies. Notably, the ensemble model achieved an AUPRC of 0.54 on the challenge's blind test set, securing the first-place position during the official stage of the Physionet challenge.

Additionally, an ablation study was conducted to assess the impact of various model components and configurations on performance, including activation functions, normalization techniques, the presence of LSTM layers, signal normalization methods, residual mappings, weight normalization, the inclusion of auxiliary tasks, and dilation rates in the network's convolutional units. Results from this study, presented in the paper's appendix, highlight the importance of these elements in achieving optimal model performance and offer insights into potential areas for further improvement and refinement.