Patterns

CellPress
OPEN ACCESS

Preview

A journey toward artificial intelligence-assisted automated sleep scoring

Rui B. Chang[1,*]
[1]Department of Neuroscience, Department of Cellular and Molecular Physiology, Yale University School of Medicine, New Haven, CT 06520, USA
*Correspondence: rui.chang@yale.edu
https://doi.org/10.1016/j.patter.2021.100429

Sleep scoring is a tedious, time-consuming process that presents a huge challenge in clinics. Leveraging the state-of-the-art U-net architecture, Zhang et al. developed a deep learning algorithm to simultaneously annotate basic and pathologic sleep stages. This model can analyze a full-length sleep record in a few seconds with high accuracy.

Understanding human cognition—the way we think, learn, and grow—is a fundamental goal in modern neuroscience. While we are still at an early stage answering some key questions regarding natural intelligence, such as which algorithms are used by the mind, rapidly advancing artificial intelligence (AI) has already started changing our daily lives. Machine learning has shown remarkable potential in healthcare, facilitating speech recognition, clinical image analysis, and medical diagnosis. For example, there is a growing need for medical imaging diagnosis automation as it takes tremendous time and resource to train a human expert radiologist. Deep-learning AI architectures have been developed to analyze medical images of the brain, lungs, heart, breast, liver, skeletal muscle, and other visceral organs and identify abnormalities with impressively high accuracy,[1,2] some of which have already been used in clinics to assist disease diagnosis.

Another field that will benefit from machine learning is automated sleep scoring. Sleep is an essential physiological process, and many people are suffering from various sleep disorders. Abnormal sleep patterns are also indications of other psychiatric and neurological diseases.[3,4] Polysomnography (PSG) that monitors brain activity (electroencephalogram, EEG), muscle activity (electromyography, EMG), and eye movements (electrooculography, EOG) is the golden standard for sleep assessment. In clinics, annotating sleep stages from PSG recordings is still performed manually by well-trained human experts. This step is tedious and time consuming and generates significant inter-clinical and inter-operator variability. A number of deep-learning-based algorithms have been recently developed to overcome these challenges with the hope to make sleep scoring a fully automated process in clinics. Many of them can achieve reasonably high annotation accuracy of basic sleep stages including wake, rapid-eye-movement (REM), and non-rapid-eye-movement (NREM) in healthy individuals; however, their performance in patients with sleep disorders still needs to be improved.

In this issue of *Patterns*, Zhang and colleagues developed a deep learning algorithm with the goal of predicting not only basic sleep stages but also pathological stages.[5] They specifically focused on arousal and apnea, two common types of disease-related events. The model was developed based on a modified U-net architecture involving a one-dimensional (1D) convolutional neural network. The model was then trained, validated, and tested with thousands of whole-night PSG recordings from two datasets containing a mixture of both healthy individuals and patients with sleep disorders. The algorithm is able to predict most sleep stages including the two pathological ones with accuracy comparable to human experts. This paper thus represents a critical step toward developing machine learning based automatic sleep scoring for clinical uses.

For historical reasons, PSG recordings have been divided into 30-second epochs, and each epoch is annotated with the dominated sleep stage. However, sleep is a continuous process, and sleep stages gradually transition from one to another. Therefore, previous and following epochs need to be considered to better assess sleep stage transitions when manually scoring. Although this historical 30-second epoch standard was not established based on physiological reasons, it was adopted by the majority of machine learning algorithms. Interestingly, the possibility of using shorter epochs for improved temporal resolution in automated sleep scoring has been considered.[6] Instead of using segmented records, Zhang et al. take advantage of the U-net structure and use single entire-sequence-length sleep records as input. Their model clearly demonstrates an excellent performance using full-length records, providing an important groundwork for better integrating information at different resolutions and scoring sleep stages in a more continuous manner with higher temporal resolution and accuracy.

While including more recording channels would likely improve performance, the associated computational cost needs to be considered. Therefore, some algorithms are focused on using only EEG channels, or even a single EEG or EOG channel, for automated sleep scoring. Intriguingly, Zhang et al. shows that although non-EEG channels might not be necessary for predicting basic sleep stages, they are important for the prediction of disease-related stages such as arousal and apnea, demonstrating the importance of non-EEG channels in disease diagnosis. However, it is likely that a subset of these channels would be sufficient for prediction. Moreover, channels might be more primed for different disease-related events. Nailing down the minimum information needed for automated sleep scoring may facilitate its application in clinics in the future.

Furthermore, additional issues still need to be addressed for clinical applications. One major challenge is that PSG recorded at distinct clinic sites usually involves different channels/settings. Most algorithms were developed and trained using datasets from specific clinic sites. For example, the model presented by Zhang et al. requires re-training to account for different datasets to achieve satisfying performance. A possible solution for this problem was demonstrated in recent studies by training algorithms using datasets from multiple clinic sites.[7,8] Nevertheless, the work by Zhang et al. presents an exciting model for automated sleep scoring. An AI-assisted diagnostic revolution in sleep disorders is soon expected.

REFERENCES

1. Oren, O., Gersh, B.J., and Bhatt, D.L. (2020). Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. Lancet Digit Health *2*, e486–e488. https://doi.org/10.1016/S2589-7500(20)30160-6.
2. Yoon, H.J., Jeong, Y.J., Kang, H., Jeong, J.E., and Kang, D.Y. (2019). Medical Image Analysis Using Artificial Intelligence. Prog. Med. Phys. *30*, 49–58. https://doi.org/10.14316/pmp.2019.30.2.49.
3. Karna, B., and Gupta, V. (2021). Sleep Disorder (StatPearls Publishing). https://www.ncbi.nlm.nih.gov/books/NBK430685/.
4. Fiorillo, L., Puiatti, A., Papandrea, M., Ratti, P.L., Favaro, P., Roth, C., Bargiotas, P., Bassetti, C.L., and Faraci, F.D. (2019). Automated sleep scoring: A review of the latest approaches. Sleep Med. Rev. *48*, 101204. https://doi.org/10.1016/j.smrv.2019.07.007.
5. Zhang, H., Wang, X., Li, H., Mehendale, S., and Guan, Y. (2021). Auto-annotating sleep stages based on polysomnographic data. Patterns *3*, 100371.
6. Stephansen, J.B., Olesen, A.N., Olsen, M., Ambati, A., Leary, E.B., Moore, H.E., Carrillo, O., Lin, L., Han, F., Yan, H., et al. (2018). Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. Nat. Commun. *9*, 5229. https://doi.org/10.1038/s41467-018-07229-3.
7. Vallat, R., and Walker, M.P. (2021). An open-source, high-performance tool for automated sleep staging. eLife *10*, e70092. https://doi.org/10.7554/eLife.70092.
8. Perslev, M., Darkner, S., Kempfner, L., Nikolic, M., Jennum, P.J., and Igel, C. (2021). U-Sleep: resilient high-frequency sleep staging. NPJ Digit Med. *4*, 72. https://doi.org/10.1038/s41746-021-00440-5.