Domain Adaptation Using Large Scale Databases for Sleep Staging from a Novel In-Ear Sensor

Samuel Waters
swaters@bidmc.harvard.edu

Harvard University

Jonathan Berent
Nextsense, Inc

Prabhjyot Saini
Emory University

Gari Clifford
Emory University

Research Article

Keywords: Ear EEG, Wearable Medical Devices, Machine Learning, Transfer Learning, Automated Sleep Staging

Posted Date: February 22nd, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3965548/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: Competing interest reported. Dr. Gari Clifford is a paid consultant for NextSense, Inc., the sponsor of this research study. The terms of this arrangement have been reviewed and approved by Emory University in accordance with Emory University Policy 7.7, Policy for Investigators Holding a Financial Interest in Research. Jonathan Berent is the CEO of Nextsense, Inc. Prabhjyot Saini is an Employee of Nextsense, Inc. and received a salary for his work. The remaining authors have no conflicts of interest to declare.
Domain Adaptation Using Large Scale Databases for Sleep Staging from a Novel In-Ear Sensor

Samuel Waters\textsuperscript{1*}, Jonathan Berent\textsuperscript{2}, Prabhjyot Saini\textsuperscript{3}, Gari Clifford\textsuperscript{4}

\textsuperscript{1}\textsuperscript{*}Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, 330 Brookline Ave, Boston, 02215, MA, USA.
\textsuperscript{2}Neurology, Emory University, 12 Executive Park, 30329, GA, USA.
\textsuperscript{3}NextSense, Inc., Mountain View, CA, United States.
\textsuperscript{4}Biomedical Informatics at Emory University, Bioengineering at Georgia Tech, 101 Woodruff Circle, Atlanta, 30322, GA, United States.

*Corresponding author(s). E-mail(s): swaters@bidmc.harvard.edu;

Abstract

\textbf{Background:} Sleep disorders and poor sleep quality contribute to serious health problems and loss of quality of life, however there are few ways to reliably monitor the progression or treatment of sleep disorders for periods of more than a few nights. Long-term at-home monitoring of sleep quality is typically conducted via wrist-worn actigraphy sensors or subjective surveys, however such methods are far less reliable than full in-hospital polysomnography. Recent advances however have been made in using ear EEG as an alternative to scalp EEG for sleep staging, among other tasks. NextSense Inc. has developed a commercial EEG sensor which fits within the ear canal and which has been shown to be effective for seizure detection.

\textbf{Methods:} We developed and evaluated the effectiveness of a deep transfer learning-based automated sleep staging algorithm for use with the NextSense device. A deep neural network was pre-trained on 1100 in-hospital polysomnograms from the Wisconsin Sleep Cohort before fine-tuning on 8 recordings taken using the NextSense device. \textbf{Results:} Our approach achieved leave-one-subject-out cross-validation accuracy of 77.3\% and Cohen’s $\kappa$ of .647, beating the performance of multiple FDA-approved wearable scalp EEG sensors.

\textbf{Conclusion:} These findings validate the use of the NextSense earbuds for use in sleep staging and open the possibility for at-home sleep monitoring using methods which are both more reliable than wrist-worn sensors and less cumbersome than bulky scalp EEG sensors.
1 Introduction

1.1 Sleep Staging and Polysomnography

Sleep disorders such as sleep apnea, insomnia, and narcolepsy pose a substantial public health burden in the form of increased healthcare costs[1–3] and elevated risk of health problems[4–20]. Sleep disorders are most commonly diagnosed using polysomnography - the practice of bringing patients into a hospital setting to record their EEG, EMG, ECG, heart rate, and sometimes temperature, blood oxygenation, and respiratory rate for one full night. A trained human technician will then manually examine the entire recordings in 30 second epochs and manually determine which of several physiological states - wake, stage 1, stage 2, stage 3 or rapid-eye movement (REM) - according to rules developed by the American Academy for Sleep Medicine (AASM)[21]. After the recording has been scored, indicators of sleep health and pathology can be computed, such as total time asleep, percent of time asleep, and time to first REM epoch.

The need to bring subjects into a hospital setting makes in-hospital polysomnography infeasible for any form of long-term monitoring, and so long-term monitoring is typically conducted through less invasive methods such as sleep-diaries, surveys of sleep quality, or wrist-worn actigraphy sensors, which are far less accurate[22].

1.2 Ear EEG

The demand for more effective methods of long-term monitoring of sleep quality has led to calls for more research on automated sleep staging using at-home wearable sensors[23]. A number of recent approaches have focused on sensors which fit within the ear. Long-term ambulatory scalp EEG requires uncomfortable and unwieldy form factors which are prone to poor signal quality and detachment of the leads[24], whereas electrodes which fit within the ear like ear buds provide a more comfortable, more user-friendly, and less cumbersome form factor. The reduced bone mass between the ear canal and brain also allows for a strong EEG signal. Ear EEG has been shown to be similar in content and quality to nearby scalp electrodes[25] and is capable of detecting seizures[26, 27], alpha rhythms during eye closure[25, 28, 29], and auditory brain responses[30].

Considerable research on the use of ear EEG for sleep staging has been conducted using both cEEGrid[31–34] and within-ear[35–40] EEG. Most approaches have used statistical learning models[25, 31, 32, 34–39, 41], although several have also used deep-learning mechanisms[33, 40].

1.3 NextSense Ear EEG System

NextSense Inc. has developed an ear-EEG system (Figure 1) which has been shown to be capable of seizure detection[42, 43]. The device consists of an enclosure housing
the electronics including the analog-to-digital converter (ADC), micro-SD card reader, lithium-ion battery, opto-sync mechanism for synchronizing with clinical systems, and additional microprocessors and components for communication and other related functions. Two custom-molded earbuds are connected to the enclosure via a micro-HDMI cable. Each earbud has been 3D printed with biocompatible resin (Carbon, SIL30) from a 3D scan of the subject’s ear using the United Sciences eFit Scanner (United Sciences, Atlanta, Georgia, USA) and the Secret Ear Designer software (Cyflex AG, Zürich, Switzerland) for modeling. Earbuds are coated with a proprietary conductive polymer on two locations in each ear. Silver rivets are threaded through the earbud to bridge the conductivity between the earbud and the electronics module.

2 Problem Statement

Automated sleep staging using wearable sensors can facilitate long-term monitoring of sleep pathology. The goal of this research is to evaluate the effectiveness of the NextSense Ear-EEG system for automated sleep staging.

3 Methods

3.1 Dataset

One full-night recording was taken on each of 8 healthy adult subjects. Subjects were recruited via flyers and enrolled at Emory University Sleep Clinic. Inclusion criteria included healthy adults ages 18-60 years of age living within 20 miles of the Emory
Sleep Center in Atlanta, Georgia. Additional criteria included normal body mass index (BMI) ($\geq 18.5$ and $\leq 28.0$ kg/m$^2$) and subjects did not require regular sleep aids or wake promoting medications (including some over the counter cold/allergy medications). All data collection was approved by the Emory Institutional Review Board under IRB protocol number 00003148 on 09/20/21.

Subjects were 75% male and were 37.5% White, 25% hispanic, 12.5% Black, 12.5% Asian, and 12.5% Middle Eastern. Ground truth sleep scores were obtained from a vote of several human sleep staging technicians using simultaneous full hospital PSG, including 10-20 montage scalp EEG, EMG, and ECG. Ages ranged from 20–44 (Figure 2). Subjects had been wearing the NextSense earbuds for hours prior to going to sleep in order collect data for separate experiments, so for our analysis we included only the 15 minutes of wakeful activity prior to initially falling asleep in both the training and test data.

A consensus mechanism was used to obtain the ground-truth sleep scores in order to minimize the effect of inter-rater differences in scoring. Scoring of the dataset was conducted according to AASM standards[21] by three independent sleep technicians. Each epoch was labeled N1, N2, N3, REM, or Wake. The annotators were then evaluated based on the number of times they had correctly scored compared to consensus. The consensus of the three scores was used as the ground truth, and the top annotator was designated as the tie-breaking vote when there was no consensus for a given epoch (Figure 3).

Although some research on wearable sleep staging has involved removing epochs with higher levels of noise or artifacts from the dataset, this research is intended as an evaluation of how the NextSense device and sleep staging algorithm perform in a

![Subject Age Histogram](image_url)
3.2 Model Architecture

We used a convolutional neural network architecture which we previously found to be effective for transfer learning on sleep staging tasks[44]. The architecture consists of 4 convolutional layers followed by a single dense layer, with each convolutional layer being followed by a ReLU, max-pooling, and batchnorm layer (Fig 4). The ADAM adaptive learning rate was used as the optimizer, and hyperparameters were tuned using Bayesian hyperparameter tuning.

real-world setting where such noise and artifacts may be present, and so we have not removed any samples from the dataset.
3.3 Pre-processing

The NextSense device contains electrodes on the ear canal and within the cymba of the ear. For our analysis we used the across ear-canals channel, which contains a stronger EEG signal than the cymba channels or same-ear channels. As with the architecture, we used the pre-processing steps which we found effective in our previous sleep staging paper\[44\]. EEG for both the source and target datasets was downsampled to 100Hz using polyphase resampling with zero-phase lowpass FIR anti-aliasing filter, after which the short-time Fourier transforms of 30-second epochs were taken and used as input for the neural network.

3.4 Transfer Learning Procedure

EEG recorded from the ear canal differs from EEG collected using scalp electrodes used in hospital PSG in terms of the number of channels, noise content and artifacts (Figure 5), and so a sleep staging algorithm trained on scalp EEG would be less effective on ear EEG. Because the target dataset had only a small number of subjects, a transfer learning procedure was used to boost performance. The Wisconsin Sleep Cohort (WSC)\[45, 46\] is a publicly available polysomnography dataset containing over 1100 subjects who are either healthy or have sleep disordered breathing, with EEG at the C3-A2 channel recorded at 100 Hz. The WSC dataset was used as a source dataset for pre-training. Early stopping was conducted by randomly withholding 10% of the training set for use in evaluating the model during training. To reduce bias due to class imbalance, samples from both the source and target datasets were weighted so that all classes would have the same total weight.

The machine learning task was to classify 30-second epochs as wake, REM, stage 1, stage 2, or stage 3. The model was pre-trained on the WSC dataset before freezing all neural network layers except for the five closest layers to the output. The unfrozen layers were then re-trained on the NextSense recordings. Performance was assessed using leave-one-subject-out cross-validation.
Table 1: Leave-one-subject-out Cross-validation results of trained neural network on data collected using NextSense earbuds.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (%)</td>
<td>77.3 ± 8.0</td>
</tr>
<tr>
<td>Precision (%)</td>
<td>60.8 ± 8.8</td>
</tr>
<tr>
<td>Recall (%)</td>
<td>57.7 ± 10.2</td>
</tr>
<tr>
<td>Macro F1</td>
<td>53.0 ± 10.6</td>
</tr>
<tr>
<td>Cohen’s κ</td>
<td>.647 ± .112</td>
</tr>
</tbody>
</table>

Table 2: Confusion matrix of trained neural network on data collected using NextSense earbuds.

<table>
<thead>
<tr>
<th>True Label</th>
<th>Wake</th>
<th>REM</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wake</td>
<td>94.8</td>
<td>0.8</td>
<td>1.3</td>
<td>3.0</td>
<td>0.0</td>
</tr>
<tr>
<td>REM</td>
<td>16.0</td>
<td>46.2</td>
<td>7.3</td>
<td>30.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Stage 1</td>
<td>30.8</td>
<td>12.2</td>
<td>24.9</td>
<td>31.4</td>
<td>0.8</td>
</tr>
<tr>
<td>Stage 2</td>
<td>5.6</td>
<td>9.3</td>
<td>4.5</td>
<td>71.9</td>
<td>8.7</td>
</tr>
<tr>
<td>Stage 3</td>
<td>6.8</td>
<td>0.1</td>
<td>0.0</td>
<td>42.6</td>
<td>50.5</td>
</tr>
</tbody>
</table>

4 Results

Automated sleep staging using the trained neural network achieved a leave-one-subject-out cross-validation accuracy of 77.3% and Cohen’s κ of .647 (Table 1). The neural network was particularly effective for identifying wake and stage 2 epochs but was prone to misclassifying REM and stage 3 samples as stage 2 (Table 2).

5 Discussion

We were able to achieve strong automated sleep staging performance using the NextSense device, close to what state-of-the-art approaches have achieved on in-hospital polysomnography[47–61] and close to or better than approaches using wearable scalp EEG sensors[62–64]. Furthermore, results were on par with the 79.1% accuracy and κ of .704 achieved in our previous work using the same neural network and transfer learning method on a wearable scalp EEG sensor[44]. Results were also consistent with other approaches using ear-EEG, which have achieved Cohen’s κ ranging from .61–.767[35, 36, 38–41]; and using cEEGrid, which have achieved Cohen’s κ ranging from .20–.762[31–34]. The performance on the NextSense device however is particularly noteworthy given that it only made use of ear-EEG and did not require cumbersome additional sensors for EMG and EOG as other approaches do. Performance is also impressive given the small quantity of data available for training.

The NextSense device along with our work opens up options for user-friendly and accurate long-term at-home monitoring of sleep pathologies in a way which is more comfortable and less cumbersome than bulky headbands or approaches requiring additional EMG or EOG sensors, yet which is far more reliable than non-EEG based approaches to sleep monitoring such as actigraphy, sleep diaries or subjective surveys.
of sleep quality. The approach also does not require any wet electrodes or disposable parts, making it even more feasible for long-term monitoring.

Several shortcomings of this study include the small sample size along with the fact that all subjects were healthy, however data collection on subjects with sleep pathologies is currently underway. Some future directions include training and testing on patients with sleep disorders, as well as potentially improving results further using open-source state-of-the-art neural networks. Data collection using soft-electrodes which fit any subject without the need for custom-fitting is also currently underway.

6 Conclusion

The NextSense ear-EEG sensor is capable of achieving strong automated sleep staging performance, on par with results achievable using scalp EEG. The high performance makes it feasible for use in comfortable and user-friendly long-term monitoring of sleep quality and sleep disorder treatment. More research is needed to evaluate performance on larger cohorts of test subjects and on sleep disorder patients.

Abbreviations

EEG - Electroencephalography, EMG - Electromyography, ECG - Electrocardiogram, AASM - American Academy for Sleep Medicine, REM - Rapid eye movement, BMI - Body mass index, WSC - Wisconsin Sleep Cohort, ADC - Analog-to-digital converter

Declarations

Ethics Approval and Consent to Participate

The experiment was approved by the Internal Review Board of Emory University (approval number 00001737). All subjects signed a written informed consent.

Availability of Data and Materials

Data cannot be posted publicly due to patient privacy concerns, but can be made available privately on request. To obtain the dataset, please email the corresponding author (Samuel Waters) at swaters@bidmc.harvard.edu.

Competing Interests & Disclosures

Dr. Gari Clifford is a paid consultant for NextSense, Inc., the sponsor of this research study. The terms of this arrangement have been reviewed and approved by Emory University in accordance with Emory University Policy 7.7, Policy for Investigators Holding a Financial Interest in Research. Jonathan Berent is the CEO of NextSense, Inc. Prabhjyot Saini is an Employee of NextSense, Inc. and received a salary for his work. The remaining authors have no conflicts of interest to declare.
Funding
The Cognitive Empowerment Program is supported by a generous investment from the James M. Cox Foundation and Cox Enterprises, Inc., in support of Emory’s Brain Health Center and Georgia Institute of Technology. We would also like to thank the staff of the Emory Sleep Center and Biomedical Informatics departments for their assistance in data acquisition and computing resources.

Authors’ Contributions
S.W. designed, trained, and evaluated the machine learning algorithms, and wrote most of the paper.
J.B. is the CEO of NextSense. He wrote the sections of the paper relating to the NextSense technology and the consensus labeling for scoring.
P.S. was involved in designing, planning and carrying out the study at the Emory Sleep Center; collecting and validating raw and summarized study data.
G.C. was the PI of the project. He provided mentorship to S.W. and provided feedback for algorithm development.
All authors reviewed the manuscript and provided feedback.

References


[17] Lim ASP, Yu L, Kowgier M, Schneider JA, Buchman AS, Bennett DA. Modification of the relationship of the apolipoprotein E ε4 allele to the risk of Alzheimer disease and neurofibrillary tangle density by sleep. JAMA neurology.


