ElderNet : Automated Electroencephalography Sleep Stage Scoring for Elderly Patient Populations

Michael Silvernagel, Spyridon Baltsavias, Abhijeet Phatak [mpsilver,sbaltsav,aphatak @stanford.edu]
Department of Electrical Engineering, Stanford University

Introduction (Problem statement)

Importance of sleep monitoring

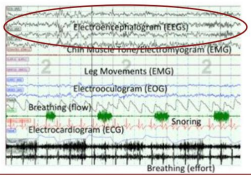
- Poor sleep quality associated with obesity, diabetes, high blood pressure, stroke, and cardiovascular disease [1]
- Under-recognized health issue: affects ~50 to 70 million Americans [2]
- Over age 60: disturbed sleep, 40-50%; chronic insomnia, 28%

Polysomnography (PSG)

- Gold standard for sleep monitoring
- Issues:
  - Requires 11 different sensors
  - Complicated setup
  - Scored by human technician

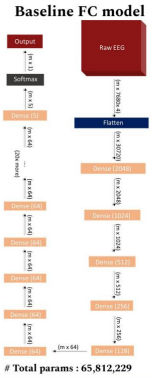
Automated scoring on EEG

- Subset of data - improves comfort, reduces complexity, removes need for human scoring
- Deep learning shows promise [3,4]
- Efficacy on older populations?
  - Reduction in deep sleep → greater class imbalance [5]
  - Worse signal-to-noise



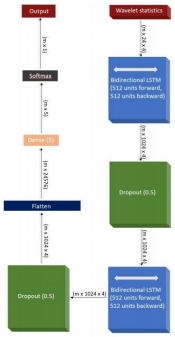
Methods (Models / Data Analysis)

Baseline FC model



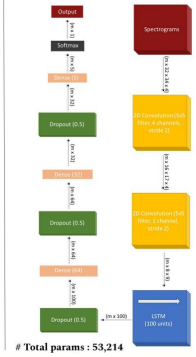
Total params : 65,812,229

Wavelet coef. + BiLSTM



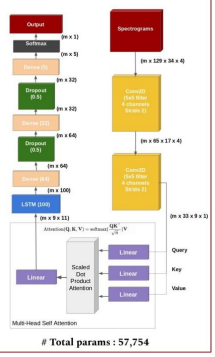
Total params : 8,536,069

Spectrogram+CNN+LSTM



Total params : 53,214

Spectrogram + CNN + Attention + LSTM



Total params : 57,754

Future Work

- Compare performance with *different/larger* elderly patient datasets
- Continue architecture search to *close the gap* between model and human error
- Try to make the models more *interpretable*
- Use *knowledge distillation* for compact deployable models in wearables

Acknowledgements

Thanks to Dr. Makoto Kawai for providing the dataset and labels, and our project mentor Aarti Bagul for her helpful guidance and advice. We also thank Amazon AWS for providing the necessary computational resources for our project.

Approach (Data + Preprocessing)

EEG Dataset

- 85 patients (67.2±5.7 years)
- 8 hours scored on 30 second intervals → 83,193 valid examples of labeled data
- 4 EEG ch.: C3-A2; C4-A1; O1-A2; O2-A1
- Train-dev- test split: 80%-10%-10%

Challenges

#1: Class imbalance: Need for weighted loss function

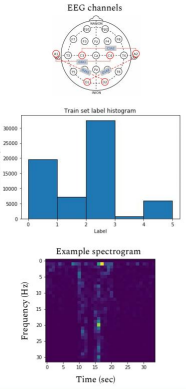
#2: Complex time-freq. relationships: Preprocess to reduce model complexity

Wavelet coefficient statistics

- Capture time-frequency features on frequency bands of interest
- Calculate RMS, variance, skew kurtosis level 3 - 8 detail coefficients

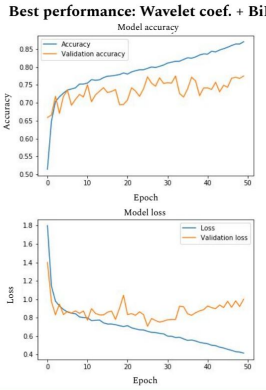
Spectrograms

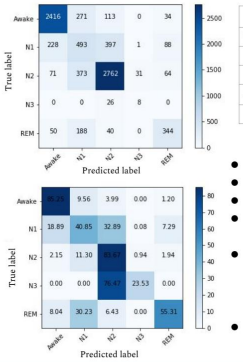
- Produce image representing signal frequencies and their amplitude over time
- Limit frequency span to 0-32 Hz



Results and Discussion

Best performance: Wavelet coef. + BiLSTM





Comparison Table

	M1	M2	M3	M4
Train acc	0.51	0.86	0.81	0.78
Val acc	0.51	0.75	0.70	0.70
Test acc	0.41	0.75	0.72	0.72
# params	65.8 M	8.5 M	53.2 k	57.8 k
epochs	100	50	100	100

Observations

- Test accuracy: 75%
- Approaching human error: ~80%
- Model overfits at ~25 epochs
- High accuracy for majority classes (Awake, N2)
- Weighted loss function helps with class imbalance, but classification on minority classes (N1, N3, REM) can be improved
- Wavelet compression allows for training of bigger models

References

[1] M.H. Kryger, T. Roth, and W.C. Dement. Principles and Practice of Sleep Medicine E-Book. Elsevier Health Sciences, 2017.

[2] S Fulda and H Schulz. "Cognitive dysfunction in sleep disorders". In: Sleep medicine reviews 5.6 (2001), pp. 423–445.

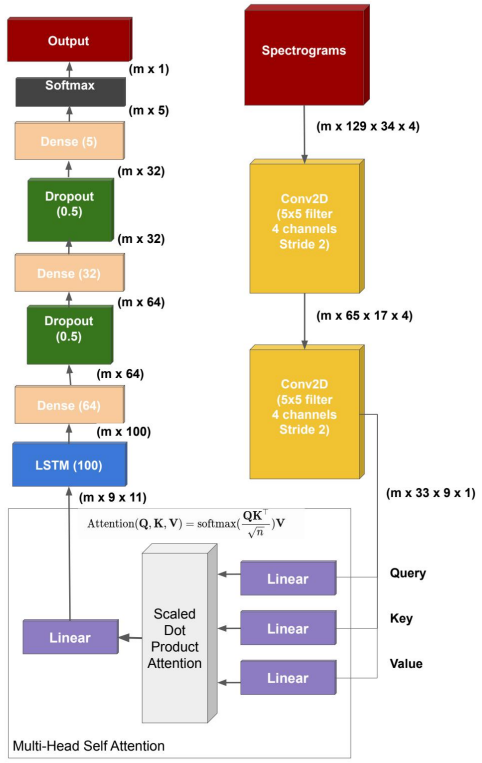
[3] Akara Supratak et al. "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG". In: IEEE Transactions on Neural Systems and Rehabilitation Engineering 25.11 (2017), pp. 1998–2008.

[4] Alexander et al. Malafeev. "Automatic Human Sleep Stage Scoring Using Deep Neural Networks". In: Frontiers in Neuroscience 12.781 (2018).

[5] Mirjam Munch. "EEG sleep spectra in older adults across all circadian phases during NREM sleep". In: Sleep 33.3 (2010), pp. 389–401.

[6] Ashish Vaswani et al. "Attention is all you need". In: Advances in neural information processing systems . 2017, pp. 5998–6008.

[7] Giulio Ruffini et al. "Deep learning with EEG spectrograms in rapid eye movement behavior disorder". In: bioRxiv (2018)


Output
(m x 1)
Softmax
(m x 5)
Dense (5)
(m x 32)
Dropout (0.5)
(m x 32)
Dense (32)
(m x 64)
Dropout (0.5)
(m x 64)
Dense (64)
(m x 100)
LSTM (100)
(m x 9 x 11)
Spectrograms
(m x 129 x 34 x 4)
Conv2D (5x5 filter 4 channels Stride 2)
(m x 65 x 17 x 4)
Conv2D (5x5 filter 4 channels Stride 2)
(m x 33 x 9 x 1)
$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{n}})\mathbf{V}$
Linear
Linear
Linear
Linear
Query
Key
Value
Scaled Dot Product Attention
Multi-Head Self Attention

(m x 65 x 17 x 4)