Using Machine Learning to Diagnose Narcoleptic Patients

Leslie Yao1,3, Caroline Wang2,3, Justin Lee3, Dr. Marianne Bezaire3

Evergreen Valley High School, San Jose, CA1, Westford Academy, Westford, MA2, Boston University, Boston, MA3

INTRODUCTION

- Narcolepsy is a sleep disorder characterized by excessive daytime drowsiness that affects around 200,000 people in the U.S. annually and currently has no cure
- Symptoms of narcolepsy are easily misdiagnosed with other disorders, and it can take over 10 years for patients to be diagnosed correctly and given treatment
- Patients are diagnosed by recording different types of signals monitored throughout their sleep, such as a polysomnogram (PSG) and the multiple sleep latency test (MSLT)
  - Electrophysiological monitoring methods include electroencephalogram (EEG), electrooculography (EOG), and electromyography (EMG)
- Hypnodensity graphs are a probability distribution of sleep stages over time, which provide more information about the progression of sleep stages than hypograms
- Hypograms are manually produced by sleep scorers who analyze the patient’s signals during sleep
- Our study aims to use machine learning to accurately diagnose patients with type 1 narcolepsy (T1N), also called narcolepsy with cataplexy, from polysomnographic data
- In T1N, the sleep-inducing neurotransmitter hypocretin is unable to be produced, causing abnormal sleep patterns
- Research has showed a strong association of T1N with the HLA-DQB1 gene, which is a part of a group of genes called the HLA complex that helps the immune system distinguish between body-produced proteins and foreign invaders, suggesting that narcolepsy is an autoimmune disorder

METHODS

- Resources:
  - Python 3, GPyTensor 1.2.0, and TensorFlow 1.14.0, EDFBrowser
  - Dataset: Stanford Sleep Cohort, from stanfordmedicine.app.box.com, containing polysomnographic sleep recording data
  - Used 28 control patient files, 5 other hypersomnia patient files, and 6 type 1 narcolepsy patient files
- Machine Learning:
  - Code based on software provided by GitHub
  - Used convolutional neural networks (CNNs), a class of deep learning algorithms
  - Efficient for using image input data to generate an output (predicted diagnosis)
- EDFBrowser was used to reveal the channels in each patient’s data allowing the adjustment of the program’s channel indices to correspond with each electrophysiological monitoring method
- We used our program to generate hypnodensity graphs of 15 s epochs over an 8 hour sleep period from the polysomnographic data
- Using 16 training models, we analyzed each graph and obtained a computer-generated diagnosis of the patient as well as a sleep score
  - A score of -0.03 to 1 indicated type 1 narcolepsy, while a score of -1 to -0.03 indicated other (no narcolepsy/other sleep disorder)
  - Accuracy of the algorithm was measured using a 2-class confusion matrix

RESULTS

- The overall accuracy 0.8718 represents the total proportion of correct predictions, which is high given the limited number of training data
- However, the recall is proportionally calculated to the number of data files for each individual category of either control or narcoleptic patients and as a result, the low recall of narcoleptic patients alone is masked by the weight of the control patients’ recall
  - Essentially, the model is accurate at diagnosing patients without T1N but inaccurate at diagnosing patients with T1N
- The precision indicates that when the model presents a positive diagnosis, the patient is only truly narcoleptic about 67% of the time
  - This presents itself as a concerning matter since if one is diagnosed with T1N, it is relatively likely to give a false positive
- In order to represent both the precision and recall in one score, the harmonic mean of the two values is calculated as the F1-score
  - Five of the patient data files were patients with other hypersomnia conditions, all five tested negative for T1N
  - The model can accurately differentiate T1N from other hypersomnia conditions

DISCUSSION/CONCLUSIONS

- Limitations:
  - Channel indices vary across different institutions, and across different sleep technicians from each institution
  - Our data was skewed: 15.38% of patient samples were narcoleptic patients although in reality, less than 1% have narcolepsy
- Weighted averages of measurements are not proportional to population, meaning the weight of the truly narcoleptic patients counts more than it should when performing analysis of data
  - The narcolepsy sleep files consisted of more fragmented sleeping patterns, resulting in lower performance and accuracy
  - Given the time constraints, only a limited number of patients were able to be tested
  - This model can only diagnose a subject with T1N and is unable to test for any other sleep conditions
- Our Algorithm Can Be Improved By:
  - Using more models to train, resulting in a higher accuracy and lower chance of overfitting
  - Analyzing more channels from polysomnographic data (such as EKG signals) to compile more information and generate a more accurate diagnosis
- Practical Application in the Real World:
  - With increased accuracy and recall for narcoleptic patients, this algorithm could become an integral part of helping technicians diagnose T1N in hospitals, along with using Multiple Sleep Latency Tests, measured hypocretin levels, and DNA analysis
  - The end goal of using machine learning to diagnose sleep disorders is to create an overarching algorithm that is able to accurately differentiate between sleep disorders and diagnose a patient upon analysis of his/her sleep recording

REFERENCES


ACKNOWLEDGEMENTS

We would like to thank the teaching assistants Johan Martinez-Fuentez and Sophia Li for helping us troubleshoot code and our parents Rita Jin, Tim Yao, Yan Wang, and Chun You for funding our research opportunity at BU RISE. Special thanks to Hyatt Moore, Eileen Leary, and Emmanuel Mignott for providing us with validation diagnoses and responding to our inquiries.