CALIFORNIA SCIENCE & ENGINEERING FAIR
2018 PROJECT SUMMARY

Name(s)
Lan Jiang

Project Number
38768

Project Title

Defining a New Diagnostic Paradigm in Primary Central Nervous System Hypersomnias through Statistical Machine Learning

Abstract

Objectives/Goals

The multiple sleep latency test (MSLT) is the current gold standard for diagnosing primary central nervous system hypersomnias. While existing thresholds for defining a positive MSLT are sufficient for diagnosing narcolepsy type 1, the arbitrary nature of the thresholds result in both the mischaracterization of over 28% of hypersomnias on initial testing and the poor differentiation of other hypersomnias, thereby negatively impacting treatment efficiency. The study objective was to determine whether better differentiation of primary central nervous system hypersomnias - narcolepsy type 1, narcolepsy type 2, idiopathic hypersomnia - is possible, by incorporating data from preceding polysomnograms and defining new thresholds for the MSLT.

Methods/Materials

Cases from the world's largest hypersomnia database at the Stanford Narcolepsy Center were combined with a control population derived from the Wisconsin Sleep Cohort. Five machine-learning models - stepwise multinomial logistic regression, decision trees, random forests, gradient boosting machine, and recursive partitioning and regression trees - were developed to address the unique multinomial categorization problem. Transparent, reproducible, and comparable methods were then created to adjust for confounders and extract information from the machine learning "black box" to elucidate the mechanisms of each algorithm and thus improve clinical interpretability.

Results

For classification accuracies in the validation set, stepwise multinomial logistic regression performed the best (0.95 vs 0.83-0.88 for other models) and was the only model that had consistently strong category-specific accuracies. In addition to expected MSLT features, new features of interest from the preceding polysomnogram (e.g. total sleep time, N2 percent) greatly improved the ability to differentiate hypersomnias.

Conclusions/Discussion

By incorporating existing clinical information at different thresholds, all models perform excellently at categorization (well above the 25% accuracy expected for chance, with 4 categories) and significantly above current MSLT accuracies. By integrating additional elements from the diagnostic work-up, these results provide doctors with ways to improve the diagnosis and treatment of their patients without needing to reverse their fundamental clinical practice, and deliver great value to researchers hoping to better identify these disorders for investigation.

Summary Statement

I significantly improved the diagnostic accuracy of sleeping disorder hypersomnias by incorporating existing clinical information at novel thresholds, thereby elucidating disease patterns not readily apparent in common clinical practice.

Help Received

Dr. Logan Schneider of Stanford University provided access to data and other laboratory resources and guidance on my model formulation and analyses interpretations. Dr. Emmanuel Mignot of Stanford University provided me with the facilities to conduct my research.