Computational Phenotyping In Polysomnography: Using Interpretable Physiology-Based Machine Learning Models to Predict Health Outcomes

Chris R. Fernandez[1,2], Samuel J. Rusk[1,2], Nick J. Glattard[1,2], and Mehdi Shokoueinejad[2,3]
University of Wisconsin School of Medicine and Public Health - Department of Population Health Sciences[1],
EnsoData Research[2], University of Wisconsin - Department of Biomedical Engineering[3]

Introduction & Motivation

- Machine learning models have grown in popularity for the analyzing PSG data…but the practical utility of many are disadvantaged by significant lack of interpretability
- Clinically, it can be challenging to understand what determinant health factors are incorporated into predictive models to estimate the likelihood of key health outcomes
- In contrast, we utilize a Computational Phenotyping approach to predict adverse health outcomes based on:
  - Common clinical variables
  - Interpretable physiological features
  - Providing a clear explanation as to why each estimation is made.

Computational Phenotyping Background

- Goal is to develop methods to model and predict thousands of phenotypes in order to:
  - advance biomedical science
  - and improve human health
- Identification of polysomnographic (PSG) biomarkers is a key step towards improving OSA diagnostic tests and therapies
- Investigate sleep pathophysiological contributions to health risk

Study Sample and PSG Characteristics

The data was obtained with IRB approval from the National Sleep Research Resource (NSRR) in [4]. The present analysis uses Sleep Heart Health Study (SHHS1), a multi-cohort longitudinal study with 11 institutions focused on sleep disordered breathing and cardiovascular outcomes.

The Polysomnography (PSG) dataset, totaling over 300 GB, is composed of a cross-sectional analyses of adults (N = 5,803), ages 39-90 (M ± SD = 63.2 ± 11.2 years), who completed an at-home PSG while participating in the Sleep Heart Health Study.

Compumedics P-Series Sleep Monitoring System was used to collect unattended Type II Polysomnography signal data:

- C3/A2 and C4/A1 EEGs, sampled at 125 Hz
- Right and left electrooculograms, sampled at 50 Hz
- A bipolar submental electromyogram, sampled at 125 Hz
- Thoracic and abdominal excursions, recorded by inductive plethysmography bands and sampled at 10 Hz
- Airflow detected by a nasal-oral thermocouple, 10 Hz
- Finger-tip pulse oximetry, sampled at 1 Hz
- ECG from a bipolar lead, sampled at 125 Hz for most SHHS-1 studies and 250 Hz for SHHS-2 studies
- Heart rate derived from the ECG and sampled at 1 Hz
- Body position, using a mercury gauge sensor
- Ambient light, by a light sensor secured to the recording garment

Univariate PSG Variable and Clinical Observation Feature Analysis

Experimental Methods

In total, 1,541 interpretable physiological and clinical features were computationally derived from the dataset and used to predict 8 outcome variables including all-cause mortality, stroke, CHD, and CVD. These features included:

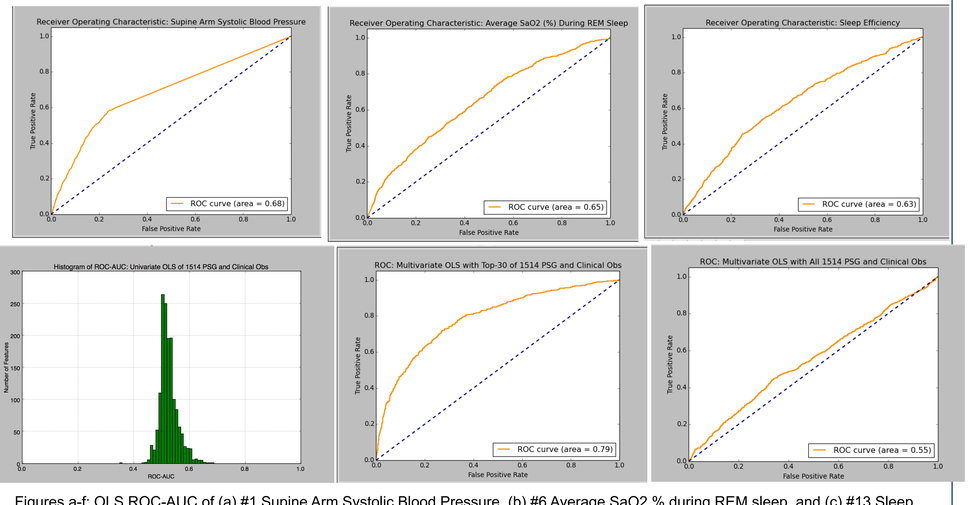
- 435 Clinical Observation variables including cigarette packs per year, blood pressure, cholesterol, and others that are well understood to contribute to outcomes analyzed
- 1170 PSG variables including sleep architecture indices, apnea and other respiratory indices, measures of spo2, arousal indices, PLMS indices, and event characteristics

Machine learning techniques including Ordinary Least Squares, Random Forests, and Deep Neural Networks were trained, optimized, and evaluated to model the relationship between the interpretable features and health outcomes.

Statistical Analysis of Predicative Value of Individual Features

Ordinary Least Squares (OLS) Regression analysis was utilized to analyze each of the 1,541 interpretable physiological and clinical features individually. To focus analysis, 5-year all-cause mortality was selected as the health outcome of interested to be predicted.

Receiver Operating Characteristic (ROC) analysis was utilized to calculate the True Positive Rate (TPR) and True Negative Rate (NPR) at varied thresholds. Predicative value was calculated by the ROC Area Under the Curve (ROC-AUC) metric compared to random chase. The statistical analysis of demonstrated that 83% (1276/1541) of the features held predictive value utilizing the basic univariate OLS models.

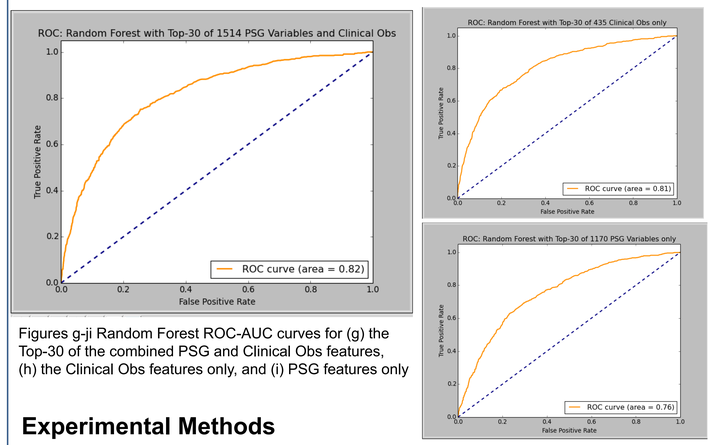


Figures a-f: OLS ROC-AUC of (a) #1 Supine Arm Systolic Blood Pressure, (b) #6 Average SaO2 % during REM sleep, and (c) #13 Sleep Efficiency. (d) Histogram ROC-AUC distribution of all 1541 features, and Multivariate OLS AUC for (e) the Top-30 and (f) all features

References

[1] Caffo, Brian, et al. "A novel approach to prediction of mild obstructive sleep disordered breathing in a population-based sample: the Sleep Heart Health Study." Sleep 33.12 (2010): 1641.
[2] Davis, Jesse, and Mark Goadrich. "The relationship between Precision-Recall and ROC curves." Proceedings of the 23rd international conference on Machine learning. ACM, 2006.
[3] Dean, D. A., Goldberger, A. L., Mueller, R., Kim, M., Rueschman, M., Mobley, D., Sahoo, S. S., Jayapandian, C. P., Cui, L., Morrical, M. G., Surovec, S., Zhang, G. Q., & Redline, S. (2016). Scaling Up Scientific Discovery in Sleep Medicine: The National Sleep Research Resource. Sleep, 5, 1151–1164

Multivariate Health Outcome Prediction Performance



Figures g-ji Random Forest ROC-AUC curves for (g) the Top-30 of the combined PSG and Clinical Obs features, (h) the Clinical Obs features only, and (i) PSG features only

Experimental Methods

Random Forests were selected as the primary multivariate tool for this analysis based on empirical and theoretical factors:

- Robust to noisy, missing, and unbalanced data
- Leverages ensemble learning and bootstrap statistics
- Superior ROC and PRC characteristics in our optimizations
- Produces interpretable feature importance's consistent with univariate OLS based approach but with improved accuracy

Table 1: Multivariate Model Comparison for Predicting All-Cause 5-Year Mortality

N = 5,803 subjects	ROC-AUC	Accuracy	Precision	Recall	Support
Random Forest: PSG and Clinical Obs	0.82	77.4%	86%	78%	4497
Random Forest: Clinical Obs only	0.81	75.1%	85%	75%	4497
OLS: PSG and Clinical Obs	0.79	72.9%	85%	73%	4497
Deep MLP: PSG and Clinical Obs	0.78	77.9%	84%	78%	4497
Random Forest: PSG only	0.76	70.3%	84%	70%	4497

Discussion & Conclusions

- Statistical analysis of features demonstrates that many sleep indices (e.g. average SaO2, SE, REM time) are of equal or in some cases greater predictive value than common Clinical Observations (e.g. blood pressure, HDL, tobacco/year)
- Simple models such as Ordinary Least Squares can be used to statistically analyze the predictive value of individual physiological factors in relation to critical health outcomes
- Multivariate OLS performs well compared to complicated methods given a valuable subset of physiological variables
- Random Forests are robust to common variations in data, provide interpretable outputs, and leading AUC, PRC, Accuracy
- Sleep phenotypes provide targets for diagnostics and treatment

Acknowledgements

Feedback from Prof. Rob Nowak, Prof. Jerome Dempsey, Prof. Paul Peppard, Dr. Mihaela Teodorescu, Dr. Thomas Penzel, Brady Riedner, Dr. Giulio Tononi NIH Center for Predictive Computational Phenotyping, others. Email: chris@ensodata.io