Computational Phenotyping in Polysomnography:

Using Interpretable Physiology-Based Machine Learning Models to Predict Health Outcomes

Chris Fernandez[1,2] Sam Rusk [1,2] Nick Glattard[1,2] Mehdi Shokoueinejad.[1,3]
[1] Department of Population Health Sciences, University of Wisconsin-Madison
[2] EnsoData Research, EnsoData Inc.
[3] Department of Biomedical Engineering, University of Wisconsin-Madison

Introduction & Motivation

- Machine learning models have grown in popularity for the analyzing sleep and 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
- Approach to predict adverse health outcomes based on:
  - Common clinical variables
  - Interpretable physiological features
  - Provide clear explanation as to why each prediction is made

PSG data offers a window into the dynamic multivariate human physiological state, trajectory, and health

- Each of the 3.2 billion DNA base pairs in a human genome can be encoded by two bits—800 megabytes for the entire genome
- Sequence of nucleotides comprising DNA is relatively static... while environment within each cell that five trillion copies of DNA sit in is highly variable
- Genome sequence may not tell exposure to toxic water, how badly injured in a fall, how a recent surgery or change in medication affected health, healthier this versus last year
- By some estimates, your physiological state at any point in time contains roughly 10^{18} (a million trillion) times more information than resides in your genetic code

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 PSG biomarkers is key step to improving OSA diagnostic tests and therapies
- Investigate basic science of sleep pathophysiological contributions to health risk

Study Sample

- Data obtained with IRB approval from the National Sleep Research Resource (NSRR)
- Sleep Heart Health Study (SHHS1), a multi-cohort longitudinal study with 11 institutions
- PSG dataset over 300 GB with a cross-sectional analyses of adults (N = 5,803), ages 39-90 (M ± SD = 63.2 ± 11.2 years)

PSG Characteristics

- Compumedics P-Series Sleep Monitoring System used to collect unattended Type II PSG signal data:
  - C3/A2 and C4/A1 EEGs,125 Hz
  - Right and left EOG, 50 Hz
  - Submental EMG, 125 Hz
  - Airflow by nasal-oral thermocouple, 10 Hz
  - Abdominal inductive plethysmography bands,10 Hz
  - Finger-tip pulse oximetry,1 Hz
  - ECG,125 Hz for most SHHS-1 studies
  - Body position by mercury gauge sensor
  - Ambient light by recording garment light sensor

Experimental Methods: Features

- 1,541 interpretable physiological and clinical features computationally derived from the dataset
- Used to predict 8 outcome variables including all-cause mortality, stroke, CHD, and CVD
- These features included:
  - 435 Clinical Observation variables
    - Included cigarette packs per year, blood pressure, cholesterol, others understood to contribute to outcomes
  - 1170 PSG variables
    - Including sleep architecture, AHI, respiratory indices, SpO2 trends, arousal and PLMS indices, and event characteristics

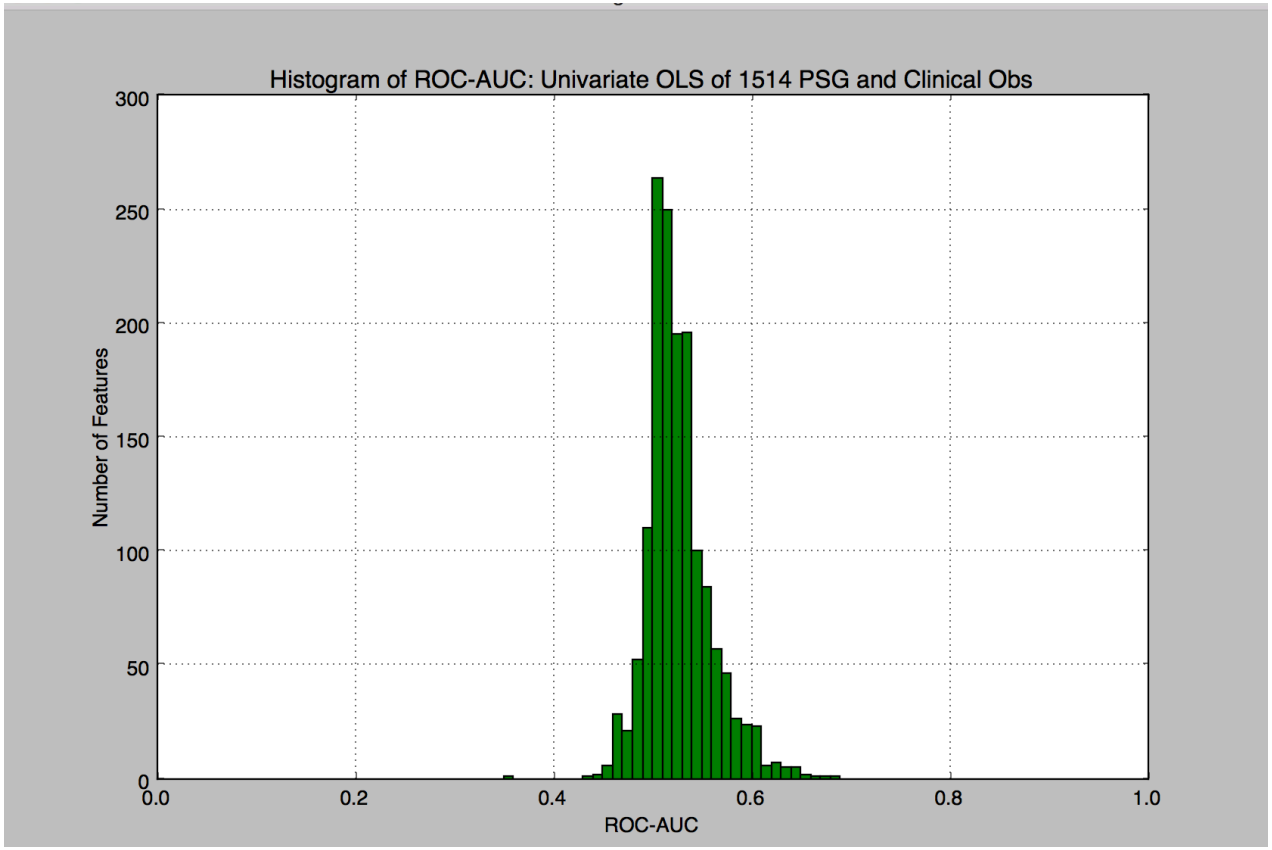
Experimental Methods: Models

- Machine learning models were trained, optimized, and evaluated
  - N=1306 subjects used for training, 5-fold CV gridsearch hyperparameter opitimizaiton utilized on training set
  - N=4497 subjects “held out” for final validation testing results
- Aim to model relationship between interpretable features and health outcomes
- Utilized several methods:
  - Ordinary Least Squares
  - Random Forest
  - Deep MLP, Kernel SVM, Naïve Bayes, KNN, Gaussian process, QDA, LASSO, Logistic Regression, AdaBoost

Endpoint 1: Statistical Analysis of Predicative Value of Individual Features

- Ordinary Least Squares analysis utilized to analyze each of the 1,541 interpretable features individually
- 5-year all-cause mortality was selected as the health outcome of interest to be predicted to focus analysis
- Receiver Operating Characteristic (ROC) analysis used to calculate the TPR and NPR at varied thresholds
- Predicative value evaluated compared to a random chance predictor using the ROC-AUC measure

Endpoint 1: Distribution of Feature ROC-AUC



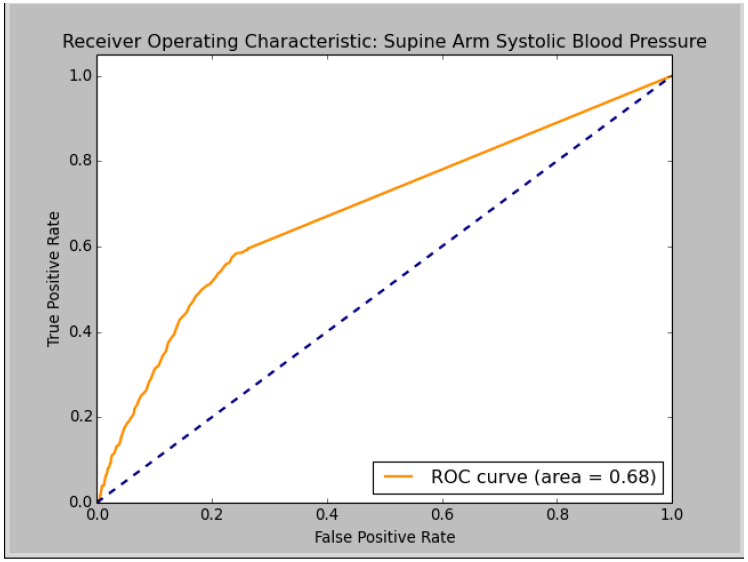
Statistical analysis of demonstrated 83% (1276/1541) of features held predictive value utilizing the basic univariate OLS models

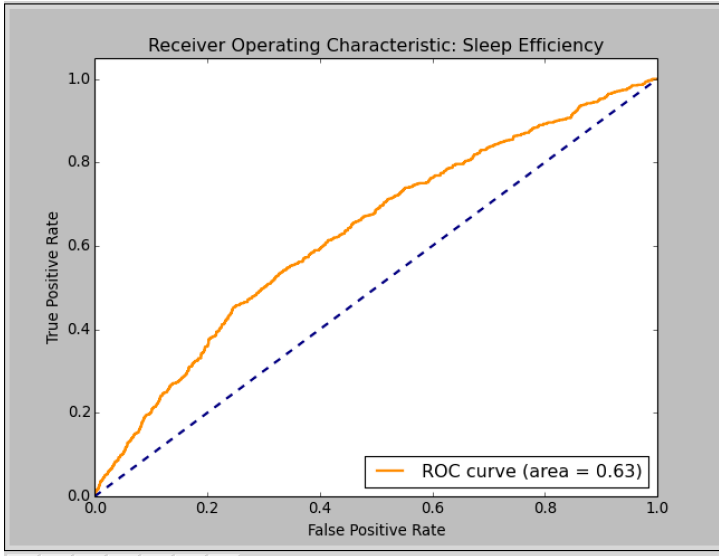
Endpoint 1: Feature Predictive Utility Ranking

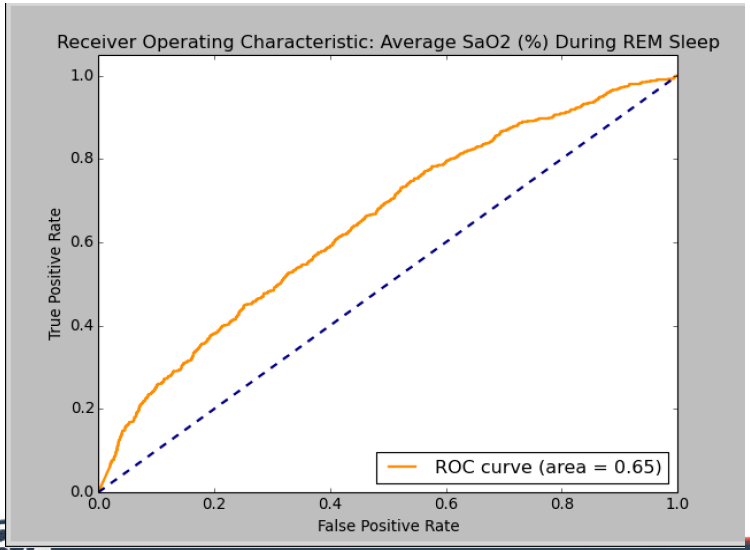
Table of Top-30 PSG variable and Clinical observation features ranked by ROC-AUC:

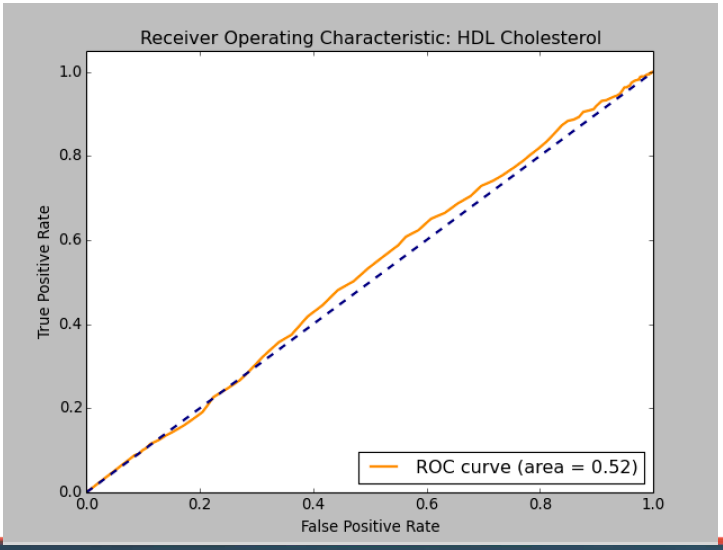
ROC-AUC	Top-30 Feature Definition
0.68	Supine arm systolic blood pressure
0.67	Forced Expiratory Volume in One Second at SHHS1
0.67	Supine ankle systolic blood pressure
0.66	Physical Functioning Standardized Score
0.66	Physical Functioning Raw Score
0.65	Average SaO2 % during REM sleep
0.65	Quality of Life (SHHS1): General health
0.65	Average SaO2 in REM sleep
0.64	SF-36 Calculated (SHHS1): Physical Component Scale Standardized Score
0.64	Systolic BP: reading 3 of 3 (SHHS1)
0.64	Systolic BP: reading 1 of 3 (SHHS1)
0.64	Systolic BP: reading 2 of 3 (SHHS1)
0.63	Any Anti-Hypertensive Medication (SHHS1)
0.63	PSG Report (SHHS2): Sleep Efficiency
0.63	Hypertension (SHHS1)
0.63	Minutes spent in REM sleep
0.63	Time in REM sleep (SHHS1)
0.63	Ventricular rate
0.63	Quality of Life (SHHS1): Health is excellent
0.62	Quality of Life (SHHS1): Health limits walking more than a mile
0.62	Average SaO2 in non-REM sleep
0.62	Wake After Sleep Onset
0.62	Average Systolic BP (SHHS1)
0.61	Has SHHS1 Adverse Event form
0.61	NREM power density at 14.0 Hertz
0.61	NREM power density at 13.5 Hertz
0.61	Percent of sleep time SaO2 is below 95%
0.61	Quality of Life (SHHS1): Health limits moderate activities
0.61	Has ECG data (SHHS1)
0.61	Maximum SaO2 during REM sleep

Endpoint 1: Univariate Feature ROC Analysis





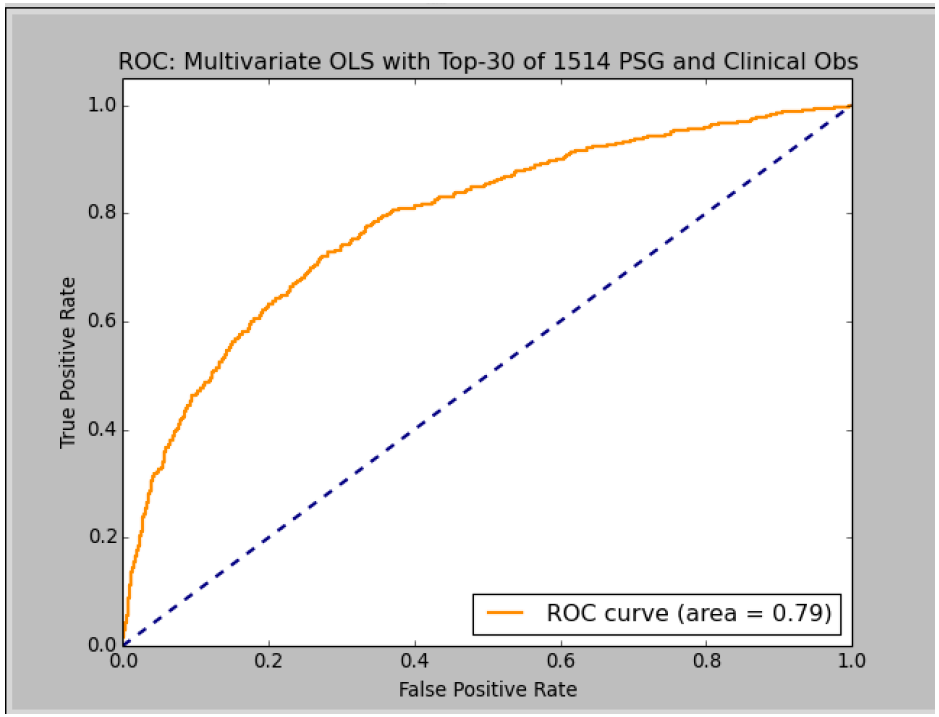


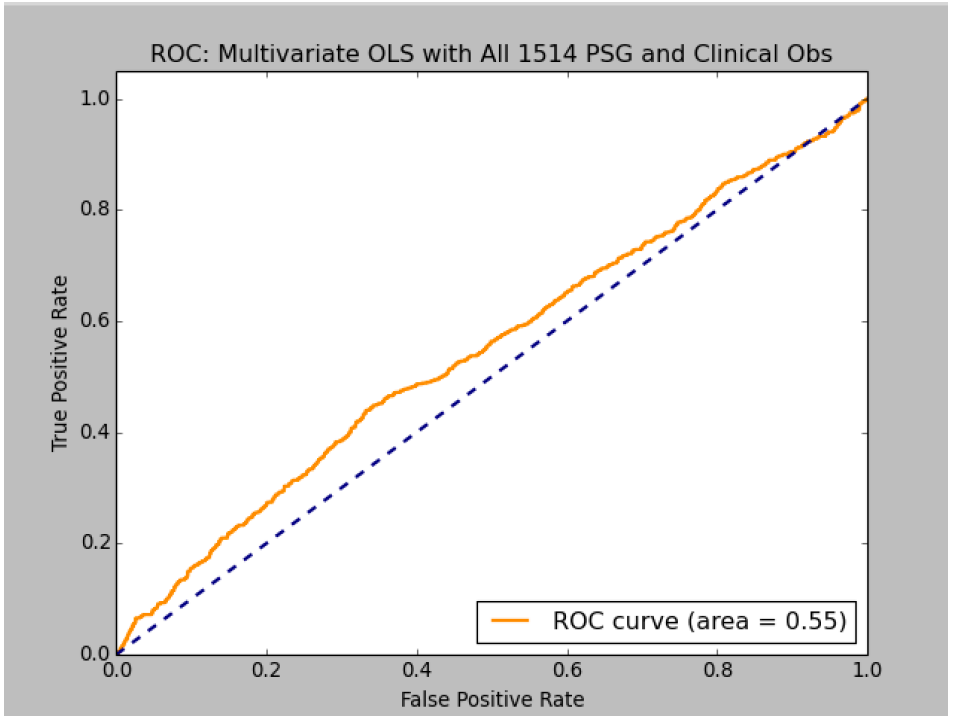


Endpoint 2: Statistical Analysis of Multivariate Health Outcome Prediction Performance

- Human physiology and disease are multivariate systems, we live in a multivariate world
- Aim is improve prediction performance for health outcomes by using multiple feature inputs
- Want to take advantage of uncorrelated feature interactions with multivariate modeling approach
  - Example: (BP and SE) or (HDL and SpO2)

Endpoint 2: Multivariate Feature ROC Analysis





Multivariate OLS trained with Top-30 features from univariate OLS predictive utility analysis outperforms All-1514 multivariate OLS

Endpoint 2: Multivariate Model Selection

Random Forests were selected as the primary multivariate tool by empirical and theoretical factors:

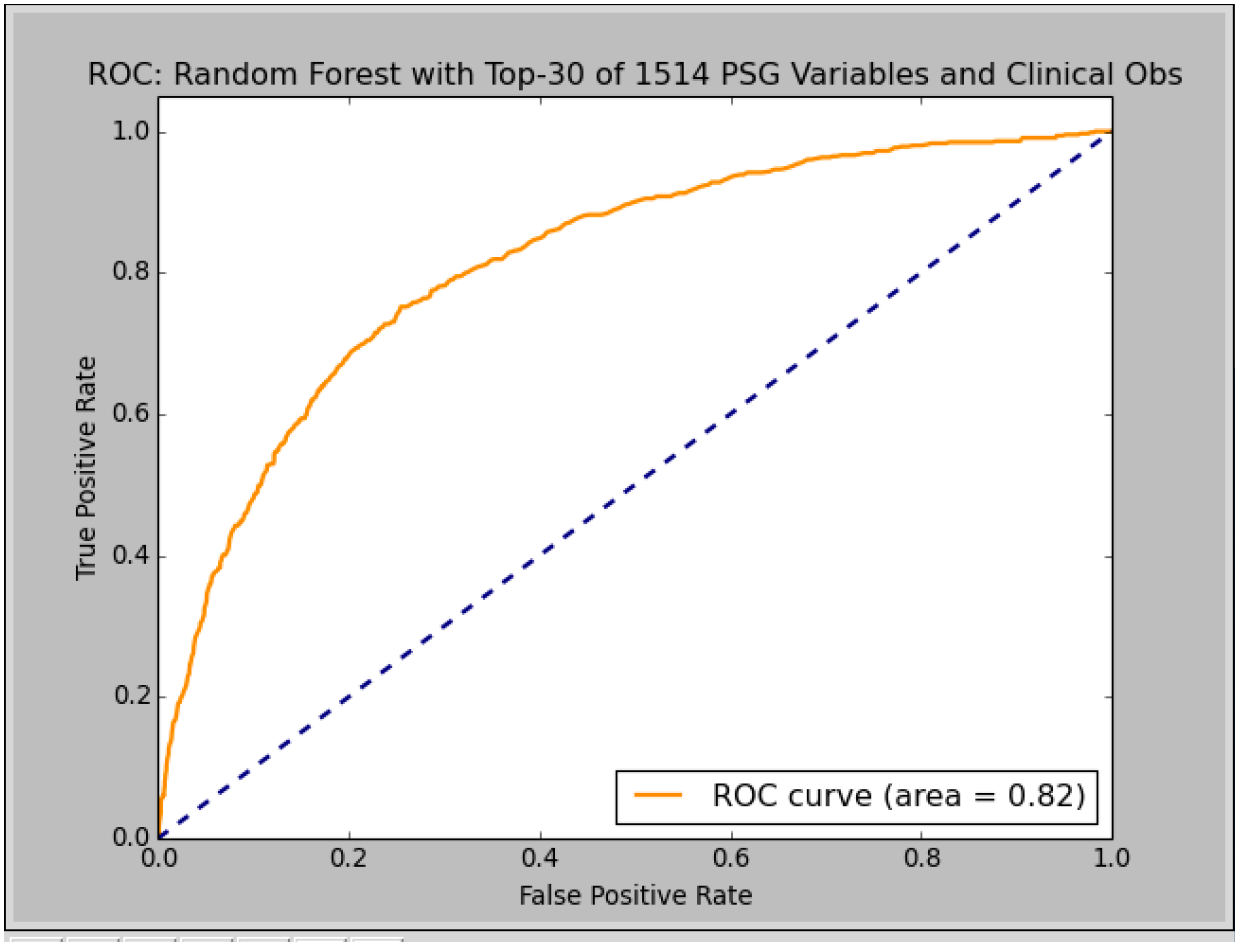
- Robust to noisy, missing, and unbalanced data
- Ensemble learning and bootstrap statistics
- Superior ROC and PRC characteristics in our optimizations versus other methods
- Produces interpretable feature importance's consistent with univariate OLS based approach but with improved accuracy

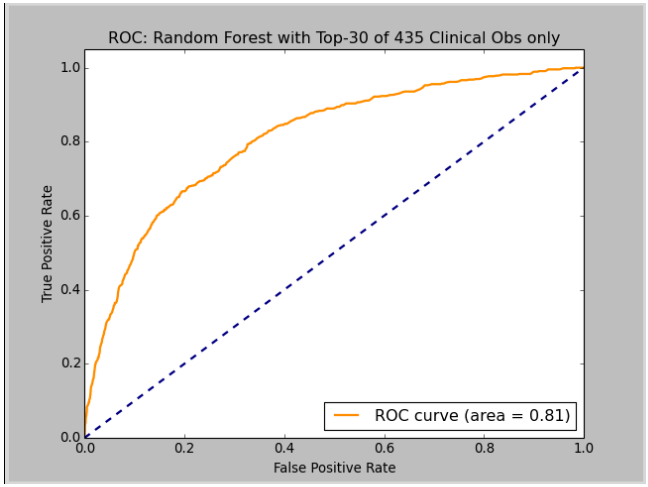
Endpoint 2: Feature Predictive Utility Ranking

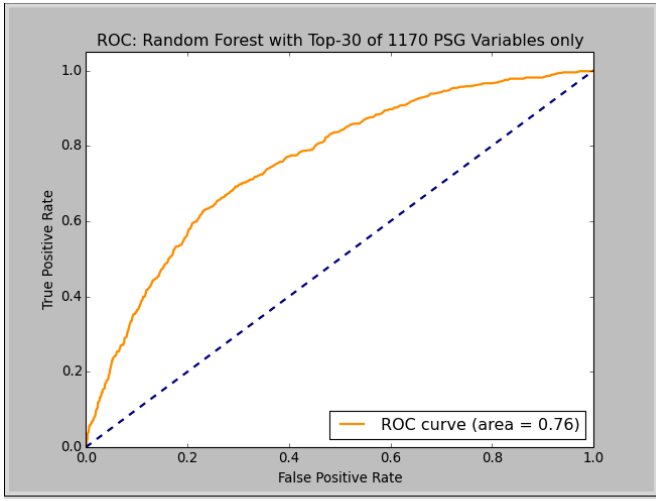
Table of Top-30 PSG variable and Clinical observation features ranked by Gini Importance:

Gini Importance (Mean Decrease Impurity)	Feature Definition
0.067	Supine arm systolic blood pressure
0.044	Forced Expiratory Volume in One Second at SHHS1
0.034	Has ECG data (SHHS1)
0.014	Ventricular rate
0.014	PSG Report (SHHS2): Sleep Efficiency
0.010	Quality of Life (SHHS1): General health
0.008	Cigarette pack-years (SHHS1)
0.008	Percent of sleep time SaO2 is below 95%
0.007	HDL cholesterol
0.006	SF-36 Calculated (SHHS1): Physical Functioning Standardized Score
0.006	Number of days since the baseline PSG until collected: ECG (SHHS1)
0.006	SF-36 Calculated (SHHS1): Physical Component Scale Standardized Score
0.005	Minimum Heart Rate (REM, Other, all oxygen desaturations)
0.005	Has SHHS1 Quality of Life form
0.005	SF-36 Calculated (SHHS1): Physical Functioning Raw Score
0.005	Forced Vital Capacity at SHHS1
0.004	Wake After Sleep Onset
0.004	Systolic BP: reading 3 of 3 (SHHS1)
0.004	Average Systolic BP (SHHS1)
0.004	Cholesterol
0.004	Minimum HR with arousal (REM, Other, 3% oxygen desaturation)
0.004	Triglycerides
0.004	Neck Circumference (SHHS1)
0.004	Sleep Time
0.004	Sleep onset time
0.003	Gender
0.003	Ankle-arm BP Index (SHHS1)
0.003	Number of oxygen desaturation with at least 2% oxygen desaturation
0.003	REM Latency II - excluding wake
0.003	Sleep time used in calculations

Endpoint 2: PSG-only, Obs-only, and Combined Random Forest analysis







Top-30 Random Forest with combined PSG and Clinical Obs data outperforms all other models including PSG-only and Obs-only

Endpoint 2: Statistical Analysis of Multivariate Health Outcome Prediction Performance

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

- Statistical analysis of features shows that sleep indices (e.g. average SaO2, SE, REM time) are of equal or sometimes greater predictive value than common clinical observations (e.g. blood pressure, HDL, tobacco/year)
- Simple models such as OLS can be used to statistically analyze the predictive utility of individual physiological factors in relation to critical health outcomes
- Multivariate OLS performs well compared to state-of-the-art methods given a valuable subset of physiological variables
- Random Forests are robust to common variations in data, provide interpretable outputs, and leading ROC-AUC, PRC, and accuracy performance in this study

Conclusion

- Computational Phenotyping provides a framework for analysis and discovery of predictive phenotypes, biomarkers, and interactions
- Applying this approach to PSG data offers a promising method to identify targets for new diagnostics and therapeutics
- Opportunity to advance basic science of sleep, better understand relationship to other psychiatric, neurological, and cardiopulmonary conditions

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