Evaluation of Electronic Medical Record Artificial Intelligence Screening Tools for Undiagnosed OSA

ensodata research

Sam Rusk, BS[1], Yoav Nygate, MS[1], Fred Turkington, BS[1],
Chris Fernandez, MS[1], Nick Glattard, MS[1], Nathaniel F. Watson, MD, MSc[2]
[1]EnsoData Research, EnsoData, Madison, WI, USA
[2]Department of Neurology, University of Washington School of Medicine, Seattle, WA, USA

Introduction

Background

The STOP-BANG is a concise, simple and widely adopted obstructive sleep apnea (OSA) screening tool. However, it has limited predictive ability and is susceptible to subjective reporting bias.

Artificial Intelligence (AI) methodologies can be utilized together with existing data in electronic medical records (EMRs) to create new screening tools to increase diagnostic sensitivity and facilitate discovery of preclinical OSA phenotypes.

Previous work has shown that EMR and demographic variables have predictive power to various AHI severities. We look to investigate this relationship over two independent datasets with reference to the STOPBANG survey, the clinical standard for OSA screening.

As OSA severity is best defined by the continuous AHI value, we report AI model performance over three disease severity thresholds 5, 15, and 30. These different OSA severity sub-group analysis allows us to investigate EMR based variable predictiveness over different OSA severities. Additionally, we report the top-5 random forest GINI feature importance's for each of the models to approximate which features have stronger predictive performance to the output OSA severity target.

Methodology

The study comprised two independent retrospective sleep study datasets: 1) Type III HSATS (N=5583) and, 2) Type I polysomnograms (N=1037). The first dataset was the Sleep Heart Health Study (SHHS) collected by the National Health Lung & Blood Institute (NHLBI). Of the 5583 sleep studies included in the analysis, x were male and y were female. The second dataset (Dataset2) was collected from a private sleep center in the United States as part of standard clinical practice. Of the 1037 sleep studies included in the analysis, x were male and y were female.

Each contained raw sleep study waveforms, manually scored sleep events (respiratory, arousal, sleep staging), and standard report indices (apnea-hypopnea index; AHI, arousal index).

The first dataset contained 90 EMR based metadata variables and the second dataset contained 54 EMR based metadata variables. Two random forest models were trained to detect OSA diagnostic thresholds (AHI> 5, AHI>15, and AHI>30) over two different screening models: STOP-BANG and Common Clinical Data Set (CCDS)-OSA (all metadata variables simulating EMR CCDS standard). Additionally, as a reference for current OSA screening standard of care, the output scores from STOP-BANG (STOP-BANG score) and Epworth Sleepiness Scale (ESS score) were analyzed using the same methodologies. STOP-BANG outputs a value between 0 and 8 and ESS outputs a value between 0 and 24. We simply divide each questionnaire score by the maximum value to achieve a value between 0 and 1 as an input to ROC analysis.

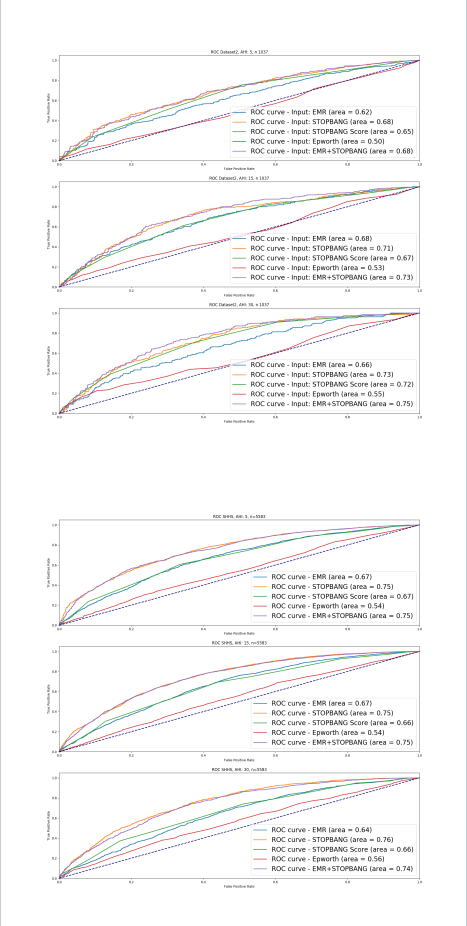
Receiver Operator Characteristic curve (ROC) analysis was performed for all 4 input types: STOP-BANG score, ESS score, STOP-BANG random forest, and EMR random forest. Each input type was analyzed over each AHI severity threshold 5, 15, and 30. We report ROC Area Under the Curve (AUC) as the primary metric for OSA severity predictive power.

Results

CCDS-OSA ROC-AUC exceeded STOP-BANG score ROC-AUC for both sleep study collections for all AHI severity thresholds. Additionally, STOP-BANG random forest ROC-AUC exceeded STOP-BANG score ROC-AUC for both sleep study collections for all AHI severity thresholds. Further, STOP-BANG random forest marginally exceeded CCDS-OSA ROC-AUC in SHHS dataset for AHI>5, AHI>15, and AHI>30 and in Dataset2 for AHI>5. Additionally, ESS was observed to have low predictive performance ROC-AUC < 0.56 over each dataset at all three AHI severities.

Additionally, we analyzed the Gini feature importance ranking of the trained CCDS-OSA model and the STOP-BANG random forest model to evaluate which variables showed highest predictive value of OSA.

The ranking revealed the top 5 features were the five physiologic based STOP-Bang parameters, followed by EMR based physiologic measurements such as HDL, triglycerides, systolic BP, and disease conditions such as diabetes, hypertension, and depression.



SHHS STOP-BANG		SHHS EMR		Dataset2 STOP-BANG		Dataset2 EMR	
Feature	**GINI**	**Feature**	**GINI**	**Feature**	**GINI**	**Feature**	**GINI**
Neck	0.250	HDL Cholesterol	0.072	BMI	0.291	Epworth Score	0.115
Age	0.200	Triglyceride Cholesterol	0.059	Neck	0.239	Breath Hold Sleep	0.056
BMI	0.179	Systolic Blood Pressure	0.052	Age	0.233	Hypertension	0.036
Gender	0.119	Cholesterol	0.052	Gender	0.094	Awakening Gasp Breath	0.032
Blood Pressure	0.086	Estrogen Supplement	0.049	Stop Breathing	0.056	Allergies	0.029

Conclusion

This study shows that while STOP-BANG contains data critical to OSA screening, a variety of other EMR-based parameters can improve performance of OSA detection. Even further, leveraging non-linear relationships between the 8 STOP-BANG survey questions showed increased ROC-AUC performance suggesting that even the most standardly collected clinical variables can benefit from utilizing basic AI methodologies.

AI-based EMR screening can provide a critical tool for more systematic and accurate screening of undiagnosed sleep apnea.

Nationwide standards facilitating patient EMR data interoperable health information exchange, particularly the United States Core Data for Interoperability (USCDI CCDS), holds promise to foster broad clinical and research opportunities.

Resulting data sharing will allow application of AI screening tools at the population health scale with ubiquitous, existing EMR data to improve population sleep health.

Future Work

- Test for repeatability with CCDS dataset collected from real world health system EMR archive
- Test against other primary diagnostic indices. For example ODI 3%, ODI 4%, TST, ect.
- With many datasets, explore STOP-BANG ROC-AUC to investigate clinical pathway deviations from site to site

References

- Kim YJ, Jeon JS, Cho SE, Kim KG, Kang SG. Prediction Models for Obstructive Sleep Apnea in Korean Adults Using Machine Learning Techniques. Diagnostics (Basel). 2021 Mar 30;11(4):612. doi: 10.3390/diagnostics11040612. PMID: 33808100; PMCID: PMC8066462.