

UC San Diego

Family Medicine Research Symposium 2025

Title

Development of an Automated Machine Learning Algorithm for Screening of OSA in Primary Care Settings

Permalink

<https://escholarship.org/uc/item/1kj0h0hb>

Authors

Mandvi, Ammar

Shen, Alice

Chow, Christopher

Publication Date

2026-09-17

Peer reviewed

Purpose

Obstructive sleep apnea (OSA) is a highly prevalent disease and a major risk factor for stroke and all-cause mortality. The gold standard for OSA diagnostic testing is a sleep study, which can be an expensive test subject to insurance denials. As such, screening questionnaires for OSA, such as STOP-BANG, have been developed with high sensitivity but variable specificity. They are also burdensome to apply in primary care settings due to time limitations and questionnaire fatigue among patients and physicians. We propose to develop an algorithm built into electronic health records (EHRs) to screen for OSA from commonly available EHR data to automate screening for patients who should be referred for sleep study. Methods: We curated an anonymized dataset of over 6000 UCSD Health patients who received sleep studies between the years of 2014 to 2023 with AHI scores. We trained a machine learning (ML) model on a set of both clinical and demographic features including race, sex, age, lipid results, tobacco usage, and active medications list, to predict which patients would have an AHI count greater than five, indicating the presence of sleep apnea. We evaluated different learning algorithms, including deep neural networks and gradient-boosting decision trees. Results: We were able to predict the presence of sleep apnea in the above context with an AUC-PR of 0.78. This is competitive with the STOP-BANG questionnaire and is significant because it is an automated screening test that does not require additional questionnaires or patient interviews. We performed fairness analyses along categorical features such as race and zip code that may be correlated with socioeconomic status and found that the proposed algorithm provides fair net benefit across the categories within this data set. Conclusions: The importance of sensitive and specific screening tests is widely recognized, especially when diagnostic testing is costly or invasive. However, improving the performance of questionnaire-based screening tests typically involves adding more questions, limiting the feasibility of their use during time-constrained visits. Our findings suggest that ML algorithms can be trained and deployed on EHR data to create robust, automated screening tools for OSA that alleviate the questionnaire and interview burden on patients and providers. This promising technology may be applied in screening for other diseases commonly seen in primary care settings.