Predicting Depression Risk Amongst Those with Sleep Disorder Using Machine Learning

Alice Yu
Gillings School of Global Public Health
Department of Biostatistics

April 28 2023

Abstract

Depression continues to widely effect public health, and has a strong link to sleep disorders. We aimed to predict depression risk among those who suffer from a sleep disorder. CART and random forest were the predictive models used to predict depression outcome of the participants classified as having a sleep disorder using demographic, depression screener, and sleep disorder datasets from NHANES 2017-March 2020 pre-pandemic data. The metrics used to evaluate model performance are receiving operating characteristic (ROC) curves, overall area under the curve (AUC), precision-recall AUC (PR-AUC), and accuracy. CARTs ROC curve curves more towards the top-left corner than random forest, which mirrors closer to a perfect ROC curve. Further, CART has a higher overall AUC of 0.71 and PR-AUC of 0.271, and the accuracies between CART and RF are practically the same. This suggests that CART may be a better performing model to predict depression risk.

Introduction

The increasing trend of the prevalence of depression throughout populations has made this mental health illness an alarming public health concern.[1] Moreover, a very strong association exists between depression and sleep.

The National Institute of Mental Health Epidemiologic Catchment Area conducted a study 20 years ago regarding sleep disturbances and psychiatric disorders. The study emphasized the statement aforementioned about association. It found that sleep disturbances and disorders, which were self reported, such as insomnia are typically the first onset symptom for risk of depression. Depression emerged in 14% of participants suffering from insomnia and 9.9% from hyperinsomnia either at the first interview or at the follow-up interview that occurred one year later.[2]

This risk has since increased. In a more recent study to assess association of sleep factors and clinically relevant depression (CRD) using NHANES (2007-2014) data, logistic regression was used on three models with different characteristics of participants to analyze this link. They found those with a sleep disorder were 2.03, 2.03, and 3.04 times more likely to have CRD than those without a sleep disorder, respectively.[3] Additionally, the results from a longitudinal study looking at the association of sleep-related breathing disorder (SRBD) and depression found participants with minimal SRBD had increased odds of depression risk by 1.6-fold compared to no SRBD. Similarly, moderate or worse SRBD increased by 2.6-fold. Thus, there is a positive correlation: as SRBD level increases, the risk for depression increases as well.[4] Aside from being a main symptom of depression, it is proven as one of the few risk factors for suicide.[5] Even if all symptoms except sleep problems are resolved, significant increase in risk of relaspe or recurrence remains.[6]

The relationship between sleep disorders and depression is well-established, however, prediction of depression risk within the population suffering from sleep disorders has yet to be investigated. To address this, we assessed the efficacy of different machine learning models on predicting risk of depression. We also applied feature importance to narrow down which covariates have a larger effect on the model.

Methods

Statistical Analysis

Descriptive statistics were used to characterize the sleep disorder sample by depression status (no depression vs depression). Categorical characteristics

were presented as percentages and continuous by mean and standard deviation. Correlations by depression outcome were analyzed to view relationships between sleep disorder and depression variables and check for multicollinearity.

5-fold cross validation (CV) for two machine learning classification models: CART without pruning (CART) and random forest (RF) were used to predict depression risk. These models were used because continuous and categorical covariates were used to predict a binary outcome. Due to imbalanced data, 6:1 (no depression:depression) ratio, pruning, which reduces size of decision trees by removing redundant or non-critical tree sections, was not used. Sample weights, provided in the NHANES demographic dataset, were used when fitting the model on the train dataset. The dependent variable for the model was depression outcome. The independent variables include gender (RIAGENDR), age (RIDAGEYR), race/Hispanic origin with Asian category (RIDRETH3), education level (DMDEDUC2), marital status (DMDMARTZ), ratio of family income to poverty (INDFMPIR), sleep hours during weekdays/workdays (SLD012), sleep hours during weekends (SLD013), snoring frequency (SLQ030), snort or stop breathing frequency (SLQ040), whether told doctor trouble sleeping (SLQ050), sleep time on weekdays/workdays in hours (slq300_hours), wake time on weekdays/workdays in hours (slq310_hours), wake time on weekends in hours (slq330_hours), sleep time on weekends (slq320_hours), and sleep debt.

Receiving operating characteristic (ROC) curve and overall area under the curve (AUC), accuracy, and precision-recall AUC (PR-AUC) are the metrics to evaluate the models. Note that due to data imbalance, accuracy alone is not sufficient enough to evaluate model performance. Thus, other metrics were used as well. Additionally, average feature importance scores for RF were calculated to see which predictors have the largest effect on a model. R version 4.0.3 (R Core Team 2020) was used to perform all statistical analyses.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where

True Positive (TP): participant predicted to have depression and actually has depression

True Negative (TN): participant predicted to not have depression and actually doesn't have depression

False Positive (FP): participant predicted to have depression, but doesn't actually have depression

False Negative (FN): participant predicted to not have depression, but actually has depression

Results

Data Source

National Health and Nutrition Examination Surveys (NHANES) 2017-March 2020 pre-pandemic data was used. The sample scheme is based on a complex, multistage, probability design and has a sample size (n = 8965) that consists of adults and children in the United States.

The analysis dataset was compiled by merging three different datasets: demographic, sleep disorder, mental health - depression screener, where the latter two are located under the questionnaire data section, by respondent sequence number.[7] Of the sample size, 4459 observations had at least one missing value. If we were to drop those with missing values, we would lose about half of our observations. This results in losing relevant information and bias. Thus, we imputed the missing categorical variables with mode and continuous with mean. The final dataset used for analysis was subsetted to include only participants with a sleep disorder.

Classification of Depression

The depression screener listed 10 different questions, which only 9 were taken into account, omitting the question "Difficulty these problems have caused." For each, there were four choices: "not at all," "several days," "more than half the days," and "nearly every day." The responses were given a point value ranging from 0 to 3. Then, a depression score was computed by summing the numbers. Following the depression screening instrument, Patient Health Questionnaire (PHQ-9) classification, participants were categorized under the following depression statuses: minimal (0-4), mild (5-9), moderate

(10-14), moderately severe (15-19), and severe (20-27).[8] These classifications were dichotomoized into two outcomes: 0 or 1, denoting no depression, defined as those with minimal or mild (0-9), and depression, defined as those with moderate, moderately severe or severe (10-27), respectively.

Classification of Sleep Disorder

Three aspects were used to classify those with sleep disorder: sleep duration during weekdays, sleep debt, and whether participant has told doctor about sleeping trouble (binary, yes or no). Sleep time was divided into three categories: insufficient (<6 hours per day), normal (6 - 8 hours per day), excessive (>8 hours per day).[9] Only insufficient and excessive were considered. Sleep debt was calculated by the difference of sleep duration during the weekends and weekdays. If sleep debt was ≥2 hours, then participants were categorized as having weekday sleep debt; participants with <2 hours were categorized as not having weekday sleep debt.[10]

Statistical Analysis Results

Table 1 shows baseline characteristics of 4004 participants who suffer from a sleep disorder. For no depression and depression groups, females are more prevalent with a proportion of 53% and 64% respectively. The mean age in years of depression, 49.09 (SE 17.27) is slightly higher than no depression, 48.91 (SE 17.55). The difference in mean depression scores is 11.33. The mean sleep debt of no depression 1.61 (SE 1.93) is about 2 times greater than the depression participants 0.82 (SE 1.85) (**Table 1**).

Table 1. Demographic and sleep disorder characteristics of participants from NHANES 2017-March 2020 pre-pandemic data by depression outcome

	No depression N = 3,440	Depression N = 564
Gender		
Female	1,807 (53%)	362 (64%)
Male	1,633 (47%)	202 (36%)
Age (years)	48.91 (17.55)	49.09 (17.27)

[1] n (%); Mean (SD)

	No depression N = 3,440	Depression N = 564
Race/Hispanic origin with NH Asian		
Mexican American	441 (13%)	64 (11%)
Non-Hispanic Asian	311 (9.0%)	18 (3.2%)
Non-Hispanic Black	966 (28%)	140 (25%)
Non-Hispanic White	1,142 (33%)	216 (38%)
Other Hispanic	401 (12%)	77 (14%)
Other Race - Including Multi-Racial	179 (5.2%)	49 (8.7%)
Ratio of family income to poverty	2.59 (1.51)	2.03 (1.36)
Marital status		
Married/Living with partner	2,054 (60%)	258 (46%)
Never married	637 (19%)	128 (23%)
Widowed/Divorced/Separated	749 (22%)	178 (32%)
Education level (Adults 20+)		
9-11th grade (Includes 12th grade, no diploma)	376 (11%)	78 (14%)
College graduate or above	741 (22%)	62 (11%)
High school graduate/GED or equivalent	784 (23%)	151 (27%)
Less than 9th grade	255 (7.4%)	55 (9.8%)
Some college or AA degree	1,284 (37%)	218 (39%)
Told doctor trouble sleeping	2,021 (59%)	500 (89%)
Frequency of feeling overly sleepy during day		
Almost always: 16-30 times a month	324 (9.4%)	162 (29%)
Never	459 (13%)	35 (6.2%)
Often: 5-15 times a month	721 (21%)	186 (33%)
Rarely: 1 time a month	752 (22%)	46 (8.2%)
Sometimes: 2-4 times a month	1,184 (34%)	135 (24%)
Wake time on weekends	8.13 (2.15)	8.19 (2.64)
Sleep Debt (sleep time on weekends - weekdays)	1.61 (1.93)	0.82 (1.85)
Sleep hours (weekends)	8.78 (1.95)	8.07 (2.23)
Sleep hours (weekdays/workdays)	7.17 (1.73)	7.24 (2.02)
Wake time on weekdays/workdays	6.59 (2.57)	7.16 (2.78)
Sleep time on weekdays/workdays	15.62 (9.54)	12.67 (10.05)
Sleep time on weekends	13.53 (10.40)	10.72 (10.09)

[1] n (%); Mean (SD)

	No depression N = 3,440	Depression N = 564
Frequency of snorting or stop breathing		
Frequently: 5 or more nights a week	247 (7.2%)	74 (13%)
Never	2,533 (74%)	355 (63%)
Occasionally: 3-4 nights a week	236 (6.9%)	48 (8.5%)
Rarely: 1-2 Nights a week	424 (12%)	87 (15%)
Snore Amount		
Frequently: 5 or more nights a week	1,249 (36%)	255 (45%)
Never	814 (24%)	117 (21%)
Occasionally: 3-4 nights a week	592 (17%)	94 (17%)
Rarely: 1-2 Nights a week	785 (23%)	98 (17%)

[1] n (%); Mean (SD)

There are no high correlations between variables pertaining to sleep disorder and depression, i.e., there is little to no association. The overall correlations are weak in comparison to the correlations by depression outcome (**Figure 1**). Since there are no high correlations, multicollinearity is not an issue, and thus all variables of interest were used in the models.

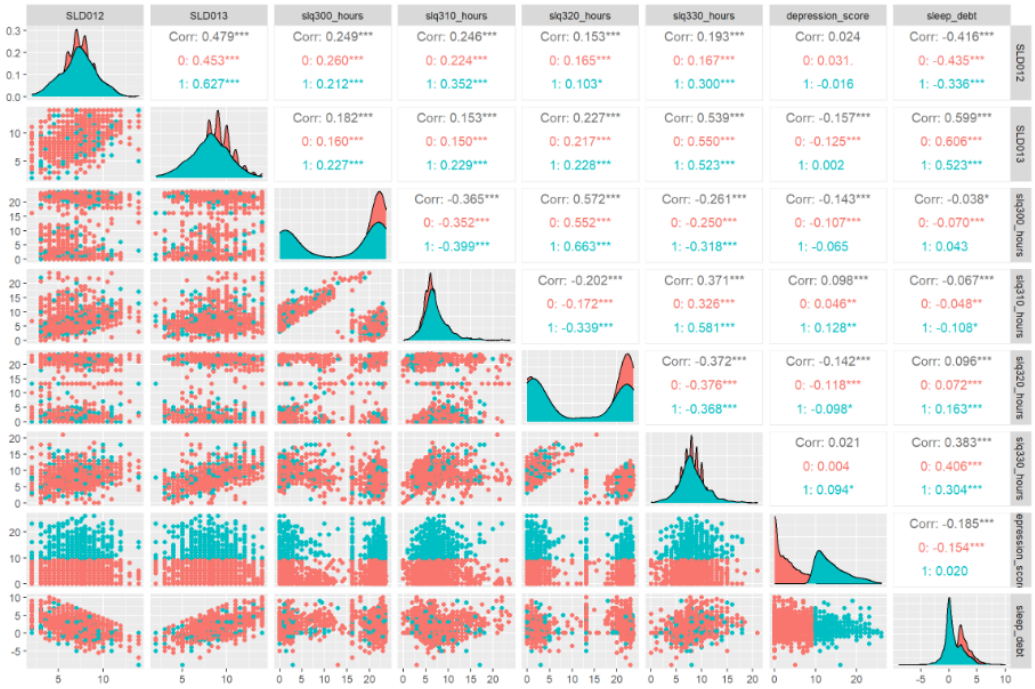

Figure 1. Overall correlations between sleep disorder and depression variables and correlations by depression outcome. The correlations in red represent no depression outcome, and blue represents depression outcome.

5-fold CV was used to fit models (CART and RF) with sample weights (WTMECPRP) on the train set, which were then predicted on the test set. This resulted in 5 models for each fold which were used to construct a ROC curve and calculate overall AUC. This was done by combining all predictions and responses from each fold and then graphing the ROC curve and computing the AUC of the overall model. In general, an ideal ROC curve is when the classifiers curve closer to the top-left corner. **Figure 2** shows that the CART ROC curve curves more closely to the top-left corner and the RF ROC curve is essentially a 45-degree diagonal; suggesting CART ROC resembles closer to a perfect ROC curve.

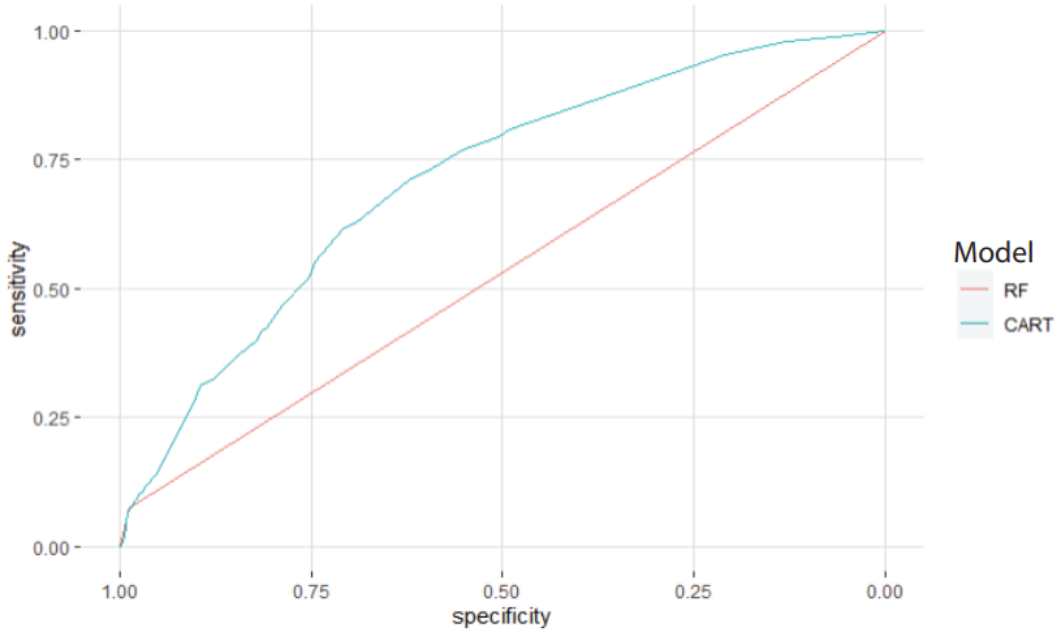

Figure 2. Overall ROC curves by classification model. The curve in red denotes RF model and blue denotes CART model.

Including ROC curve, overall accuracy, AUC, and PR-AUC were computed to evaluate model performance. To calculate overall accuracy, the sum of correct predictions was divided by the sum of total number of predictions from the 5 models. The difference between RF and CART accuracy is 0.0001, which is a negligible amount. Since the accuracies are so similar, overall accuracy cannot distinguish which model performed better. An overall AUC and PR-AUC of 1 indicates a perfect predictive model, i.e., a higher value indicates better performance. Due to its higher overall AUC of 0.71, CART shows better performance in distinguishing between depression outcome. In comparison, the RF overall AUC is very close to 0.5, suggesting no discrimination between outcomes, i.e., bad predictive ability and is no better than randomly guessing. Similarly, CART has a higher overall PR-AUC than RF (**Table 2**). These results suggest that CART may be a better predictive model for depression outcome.

Table 2. Overall evaluation metrics of classification models

	Accuracy	AUC	PR-AUC
Random Forest	0.859	0.531	0.189
CART	0.858	0.71	0.271

RF feature importance describes which variables are more relevant and lead to a better model. The higher the importance score, the larger the effect the variable has on the predictive model. The top 3 predictors are, in decreasing average importance score, SLQ050, SLQ120, slq330_hours. Following these 3, sleep debt is close to wake time on weekends. SLQ050 has the largest effect since it has the highest score. DMDEDUC2 has the lowest average variable importance score, and thus, has the least effect on depression prediction (**Figure 3**).

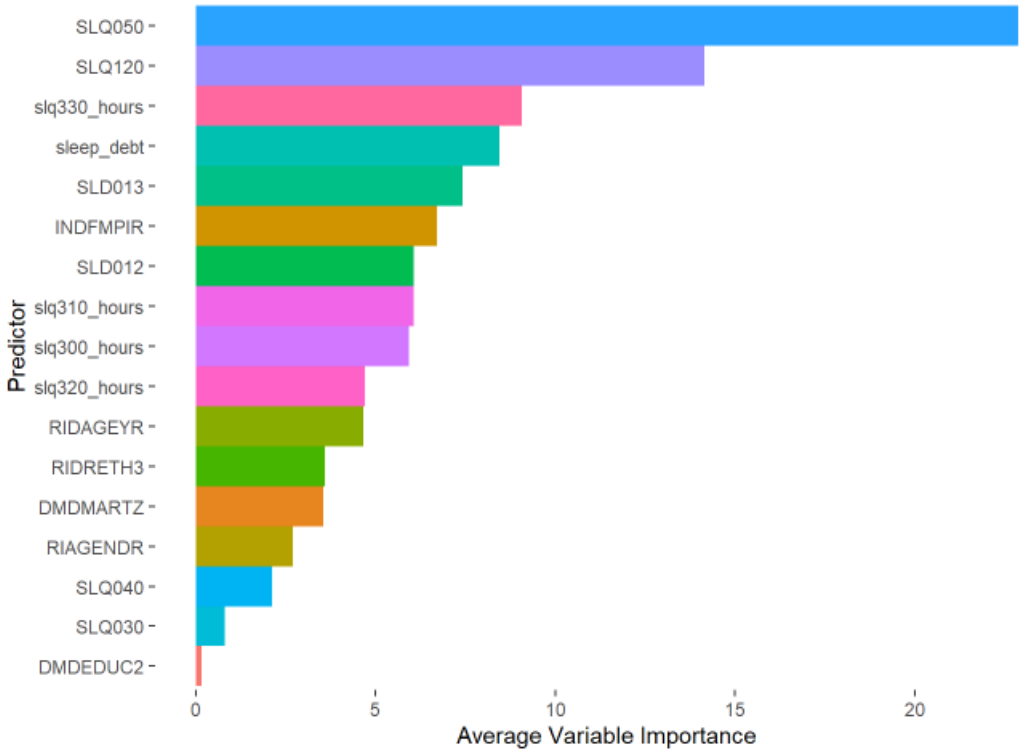

Figure 3. Average feature importance scores for RF of covariates, in descending score, to show the effect on predicting depression outcome

Discussion

CART and RF were used to predict depression risk on participants that have a sleep disorder from NHANES 2017-March 2020 pre-pandemic data. Overall, CART seems to perform better in predicting depression outcome than RF. CART has a ROC curve closer to an ideal one, higher overall AUC and PR-AUC. The overall accuracies for RF and CART are essentially the same. The PR-AUCs are lower than AUCs due to the fact that PR analysis does not take into account TN and gives more weight to the minority class, depression; whereas ROC takes into account all four: TP, TN, FP, FN.

Through feature importance in RF, five predictors related to sleep disorder had the largest effect on predicting depression risk. This supports previous

studies where they found symptoms of sleep disorders being strongly associated with depression risk and relapse in depression.[11]

However, these results should be interpreted with caution as there are limitations to the study. Due to lack of data for the minority class, the results are biased towards the majority class, no depression. This leads to inaccurate prediction models and ending up with a model that may not be the best. Further, the dataset is not an accurate representation of the population that have a sleep disorder, meaning poor generalizability. Also, sleep time was self-reported in the survey. Since an objective way to measure it was not used, this can lead to response bias.

This study extends existing findings on predicting depression risk and opens up for further investigation concerning the sleep disorder population. More studies should be performed to improve generalizability and accuracy of results. The intention is to be able to decrease depression risk and catch it early on so that sleep disordered individuals receive help before symptoms worsen.

Data and Code Availability

Data is publicly available on CDC NCHS site under the NHANES section. Code is available on github at the following link: `https://github.com/alicewyu/biostat_ms_thesis`

References

[1] D. Moreno-Agostino, Y.-T. Wu, C. Daskalopoulou, M. T. Hasan, M. Huisman, and M. Prina, "Global trends in the prevalence and incidence of depression:a systematic review and meta-analysis," *Journal of Affective Disorders*, vol. 281, pp. 235–243, 2021.

[2] D. E. Ford and D. B. Kamerow, "Epidemiologic Study of Sleep Disturbances and Psychiatric Disorders: An Opportunity for Prevention?," *JAMA*, vol. 262, pp. 1479–1484, 09 1989.

[3] L. Chunnan, S. Shaomei, and L. Wannian, "The association between sleep and depressive symptoms in us adults: data from the nhanes (2007–2014)," *Epidemiology and Psychiatric Sciences*, vol. 31, p. e63, 2022.

[4] P. E. Peppard, M. Szklo-Coxe, K. M. Hla, and T. Young, "Longitudinal Association of Sleep-Related Breathing Disorder and Depression," *Archives of Internal Medicine*, vol. 166, pp. 1709–1715, 09 2006.

[5] M. Y. Agargun, H. Kara, and M. Solmaz, "Sleep disturbances and suicidal behavior in patients with major depression," *Journal of Clinical Psychiatry*, vol. 58, no. 6, pp. 249–251, 1997.

[6] D. Nutt, S. Wilson, and L. Paterson, "Sleep disorders as core symptoms of depression," *Dialogues in Clinical Neuroscience*, vol. 10, no. 3, pp. 329–336, 2008. PMID: 18979946.

[7] C. for Disease Control and P. (CDC), "National health and nutrition examination survey data,"

[8] K. Kroenke, R. L. Spitzer, and J. B. Williams, "The phq-9: validity of a brief depression severity measure," *Journal of general internal medicine*, vol. 16, no. 9, pp. 606–613, 2001.

[9] W. Li, D. Chen, W. Ruan, Y. Peng, Z. Lu, and D. Wang, "Associations of depression, sleep disorder with total and cause-specific mortality: A prospective cohort study," *Journal of Affective Disorders*, vol. 298, pp. 134–141, 2022.

[10] T. Cabeza de Baca, K. L. Chayama, S. Redline, N. Slopen, F. Matsushita, A. A. Prather, D. R. Williams, J. E. Buring, A. M. Zaslavsky, and M. A. Albert, "Sleep debt: the impact of weekday sleep deprivation on cardiovascular health in older women," *Sleep*, vol. 42, no. 10, p. zsz149, 2019.

[11] D. Riemann, "Does effective management of sleep disorders reduce depressive symptoms and the risk of depression?," *Drugs*, vol. 69, pp. 43–64, 2009.