PLOS ONE

RESEARCH ARTICLE

Use of machine learning to identify risk factors for insomnia

Alexander A. Huang[1☯]*, **Samuel Y. Huang**[2☯]

1 Northwestern University Feinberg School of Medicine, Chicago, IL, United States of America, **2** Virginia Commonwealth University School of Medicine, Richmond, VA, United States of America

☯ These authors contributed equally to this work.
* Alexander.huang@northwestern.edu

OPEN ACCESS

Citation: Huang AA, Huang SY (2023) Use of machine learning to identify risk factors for insomnia. PLoS ONE 18(4): e0282622. https://doi.org/10.1371/journal.pone.0282622

Editor: Zhe He, Florida State University, UNITED STATES

Received: October 6, 2022

Accepted: February 19, 2023

Published: April 12, 2023

Data Availability Statement: Data described in the manuscript are present at: https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?cycle=2017-2020. Code book and analytic code are available from https://github.com/huangs8/Use-of-Machine-Learning-to-Identify-Risk-Factors-for-Insomnia-.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Importance

Sleep is critical to a person's physical and mental health, but there are few studies systematically assessing risk factors for sleep disorders.

Objective

The objective of this study was to identify risk factors for a sleep disorder through machine-learning and assess this methodology.

Design, setting, and participants

A retrospective, cross-sectional cohort study using the publicly available National Health and Nutrition Examination Survey (NHANES) was conducted in patients who completed the demographic, dietary, exercise, and mental health questionnaire and had laboratory and physical exam data.

Methods

A physician diagnosis of insomnia was the outcome of this study. Univariate logistic models, with insomnia as the outcome, were used to identify covariates that were associated with insomnia. Covariates that had a $p<0.0001$ on univariate analysis were included within the final machine-learning model. The machine learning model XGBoost was used due to its prevalence within the literature as well as its increased predictive accuracy in healthcare prediction. Model covariates were ranked according to the cover statistic to identify risk factors for insomnia. Shapely Additive Explanations (SHAP) were utilized to visualize the relationship between these potential risk factors and insomnia.

Results

Of the 7,929 patients that met the inclusion criteria in this study, 4,055 (51% were female, 3,874 (49%) were male. The mean age was 49.2 (SD = 18.4), with 2,885 (36%) White patients, 2,144 (27%) Black patients, 1,639 (21%) Hispanic patients, and 1,261 (16%) patients of another race. The machine learning model had 64 out of a total of 684 features

that were found to be significant on univariate analysis (P<0.0001 used). These were fitted into the XGBoost model and an AUROC = 0.87, Sensitivity = 0.77, Specificity = 0.77 were observed. The top four highest ranked features by cover, a measure of the percentage contribution of the covariate to the overall model prediction, were the Patient Health Questionnaire depression survey (PHQ-9) (Cover = 31.1%), age (Cover = 7.54%), physician recommendation of exercise (Cover = 3.86%), weight (Cover = 2.99%), and waist circumference (Cover = 2.70%).

Conclusion

Machine learning models can effectively predict risk for a sleep disorder using demographic, laboratory, physical exam, and lifestyle covariates and identify key risk factors.

Introduction

Sleep is critical to a person's physical and mental health [1–6]. However, the prevalence of diagnosed sleep disorders among American patients has significantly increased over the past decade [1, 5, 7–10]. Sleep disorders are a broad categorization of disorders that encompass conditions that lead to difficulty falling asleep, poor sleep quality, early waking, circadian rhythm disorders, parasomnias, sleep-related movement disorders, and sleep-related breathing disorders [11–13]. This is particularly important as sleep disorders are a significant risk factor for diabetes, heart disease, obesity, and depression, leading to decreased quality of life and increased healthcare usage [14, 15]. Additionally, poor quality of sleep has been associated with decreased productivity at work and at school, increased stress, and decreased quality of life [16–19]. To combat the debilitating consequences of sleep disorders, a plethora of pharmacologic treatments have been introduced to the market and prescribed by physicians [20–26]. While medications have shown efficacy in decreasing sleep latency, significant side effects have been associated with these medications [27–32]. These include addiction, respiratory depression, decreased quality of sleep, and significant withdrawal symptoms when these medications are discontinued [21, 33–35]. Furthermore, due to the increasing prevalence of obstructive sleep apnea, continuous positive airway pressure (CPAP) machines are more regularly prescribed [27].

Despite recognition of sleep disorders as a strong contributor to increasing mortality and morbidity, little is known regarding specific risk factors that are strongly linked with increased probability of having sleep disorders. Given these limitations in the literature, we will leverage transparent machine-learning methods (Shapely Additive Explanations (SHAP) model explanations and model gain statistics) to identify pertinent risk-factors for sleep disorders and compute their relative contribution to model prediction of risk for sleep disorder; the NHANES 2017–2020 cohort, a large, nationally representative sample of US adults, will be used within this study.

Methods

A retrospective, cross-sectional cohort study using the publicly available National Health and Nutrition Examination Survey (NHANES) was conducted in patients who completed the demographic, dietary, exercise, and mental health questionnaire and had laboratory and physical exam data. The acquisition and analysis of the data within this study was approved by the

National Center for Health Statistics Ethics Review Board. Within this retrospective cohort, all data (medical records, survey information, demographic information) was fully anonymized before data analysis was carried out and all patients consented to their data being publicly available.

Dataset and cohort selection

The National Health and Nutrition Examination Survey (NHANES 2017–2020) is a program designed by the National Center for Health Statistics (NCHS), which has been leveraged to assess the health and nutritional status of the United States population. The NHANES dataset is a series of cross-sectional, complex, multi-stage surveys conducted by the Centers for Disease Control and Prevention (CDC) on a nationally representative cohort of the United States population to provide health, nutritional, and physical activity data. In the present study, we analyzed adult ($\geq$18 years old) patients in the NHANES dataset who completed the demographic, dietary, exercise, and mental health questionnaire and had laboratory and physical exam data.

Assessment of sleep disorder

The medical conditions file was used to identify patients with a sleep disorder. Participants were asked: "Have you ever told a doctor or other healthcare professional that you have trouble sleeping?" Participants who answered "Yes" to this question were considered to have a sleep disorder within this study.

Independent variable

Potential model covariates were identified within the demographics, dietary, physical examination, laboratory, and medical questionnaire datasets in NHANES. A total of 783 covariates were identified from the NHANES dataset. All covariates were extracted and merged with the sleep disorder indicator.

Model construction and statistical analysis

Univariate logistic models, with a sleep disorder as the outcome, were used to identify covariates that were associated with a sleep disorder. Covariates that had a $p<0.0001$ on univariate analysis were included within the final machine-learning model. Utilizing univariable logistic models to do an initial filter of the 700+ covariates that were within the dataset was used to ensure that all covariates used within the machine learning models were strong independent covariates. Furthermore, this initial filtering allowed for physician review of risk factors that were clinically relevant. After initial filtering, model importance statistics from machine-learning models were used to identify pertinent risk factors.

Four machine-learning methods were carried out: XGBoost, Random Forest (RF), Adaptive Boost (ADABoost), and Artificial Neural Network (ANN). All machine-learning models were constructed using 10-fold cross validation. Cross validation was applied to only the training set. A train:test (80:20) was used to compute the final set of model fit parameters. The model fit parameters used in this study were accuracy, F1, sensitivity, specificity, positive predictive value, negative predictive value, and AUROC (Area under the receiver operator characteristic curve).

A grid search of hyperparameters for the XGBoost, Random Forest, and Adaptive Boost methods was conducted. Trees were searched between 200 and 2000 at 100 tree increments, with the optimal number being 600 trees for all models. The artificial neural network was

comprised of an input layer with hidden layers and a scalar output layer. Additionally, the ReLu function at each hidden layer and a Sigmoid function at the output layer was used. The hyperparameters were determined by optimal accuracy across a grid search of 2–10 hidden layers, 128–1024 for hidden layer dimensions, and 64–512 for batch size. The hyperparameters that were most optimal were 4 hidden layers, 256 hidden layer dimensions, and 64 for the batch size.

The machine learning model XGBoost was used due to its prevalence within the literature as well as its increased predictive accuracy in healthcare prediction. Furthermore, XGBoost was chosen as the most optimal model based upon the mean AUROC: $(\text{AUROC}_{\text{XGBoost}} = 0.87 > \text{AUROC}_{\text{ADABoost}} = 0.84 > \text{AUROC}_{\text{ANN}} = 0.83 > \text{AUROC}_{\text{RF}} = 0.82, p<0.01)$.

Model feature importance statistics and SHAP visualization

Model covariates were ranked according to the Gain, Cover, and Frequency to identify risk factors for a sleep disorder. The Gain is the relative contribution of the feature within the model. The Cover is the number of observations related to this feature that were present. The Frequency is the percentage of times the feature occurs in the trees of the machine-learning model. The Gain statistic was chosen as the method to rank features based upon feature importance due to its ease of interpretation: the proportion the covariate contributed to the final prediction.

SHAP explanations were utilized to visualize the continuous covariates with the strongest relationship between the potential risk factors and a sleep disorder.

Results

Fig 1 shows of the 7,929 patients that met the inclusion criteria in this study, 4,055 (51% were female, 3,874 (49%) were male. The mean age was 49.2 (SD = 18.4), with 2,885 (36%) White patients, 2,144 (27%) Black patients, 1,639 (21%) Hispanic patients, and 1,261 (16%) patients of another race. A total of 2,302 (29%) of patients had a sleep disorder.

Fig 2 shows the comparison of different machine learning models that led to XGBoost being chosen for having the highest mean AUROC: $(\text{AUROC}_{\text{XGBoost}} = 0.87 > \text{AUROC}_{\text{ADABoost}} = 0.84 > \text{AUROC}_{\text{ANN}} = 0.83 > \text{AUROC}_{\text{RF}} = 0.82, p<0.01)$. The machine learning model had 64 out of a total of 684 features that were found to be significant on univariate analysis ($P<0.0001$ used). These were fitted into the XGBoost model and an AUROC = 0.87, Sensitivity = 0.77, Specificity = 0.78 were observed in Fig 3.

Fig 4 shows the top four highest ranked features by cover, a measure of the percentage contribution of the covariate to the overall model prediction, were the Patient Health Questionnaire depression survey (PHQ-9) (Cover = 31.1%), age (Cover = 7.54%), physician recommendation of exercise (Cover = 3.86%), weight (Cover = 2.99%), and waist circumference (Cover = 2.70%). Fig 5 shows the overall SHAP explanations for the covariates showing the PHQ-9, age, physician recommendation of exercise, weight, and waist circumference have the highest contribution to the model.

SHAP visualizations were conducted for the top four continuous covariates by model cover (Fig 6). We observed that increased PHQ-9 scores were strongly linked to the odds of a sleep disorder. Each increase in PHQ-9 score is associated with increased odds of a sleep disorder up to around a PHQ-9 score of 11, at which the odds of sleep disorder no longer increase with increased PHQ-9 score. Additionally, we observed a curvilinear relationship between weight and odds of a sleep disorder. There is no significant increase in odds of a sleep disorder with increasing weight for patients weighing under 80 kg, but after 80 kg, increased weight is associated with significantly increased odds of a sleep disorder. Furthermore, age was found to be a

Covariate	All Patients (N = 7929)	Patients with Sleep Disorders (N = 2302)	Patients without Sleep Disorders (N = 5627)	P-Value
Age: Mean (SD)	49.23 (18.35)	52.88 (16.73)	47.73 (18.77)	P<0.01
Gender Female: Count (Proportion)	4055 (0.51)	1300 (0.56)	2755 (0.49)	P<0.01
Gender Male: Count (Proportion)	3874 (0.49)	1002 (0.44)	2872 (0.51)	P<0.01
Race Other: Count (Proportion)	1261 (0.16)	284 (0.12)	977 (0.17)	P<0.01
Race White: Count (Proportion)	2885 (0.36)	1002 (0.44)	1883 (0.33)	P<0.01
Race Hispanic: Count (Proportion)	1639 (0.21)	427 (0.19)	1212 (0.22)	P<0.01
Race Black: Count (Proportion)	2144 (0.27)	589 (0.26)	1555 (0.28)	P<0.01
Income Poverty Ratio: Mean (SD)	2.6 (1.63)	2.58 (1.65)	2.62 (1.63)	P<0.01
Red blood cell count (million cells/uL) : Mean (SD)	4.73 (0.51)	4.68 (0.52)	4.74 (0.5)	P<0.01
Red cell distribution width (%): Mean (SD)	13.9 (1.39)	14.01 (1.38)	13.85 (1.38)	P<0.01
Cotinine, Serum (ng/mL) : Mean (SD)	57.44 (130.68)	69 (146.47)	52.65 (123.23)	P<0.01
Hydroxycotinine, Serum (ng/mL) : Mean (SD)	23 (62.39)	29.41 (75.42)	20.33 (55.89)	P<0.01
RBC folate (ng/mL) : Mean (SD)	514.51 (230.32)	541.57 (258.6)	503.66 (217.05)	P<0.01
Glycohemoglobin (%): Mean (SD)	5.82 (1.09)	5.93 (1.17)	5.78 (1.06)	P<0.01
HS C-Reactive Protein (mg/L) : Mean (SD)	4.09 (7.85)	5.02 (9.79)	3.69 (6.84)	P<0.01
Insulin (pmol/L) : Mean (SD)	89.85 (145.24)	109.97 (198.2)	81.2 (114.21)	P<0.01
Blood cadmium (ug/L) : Mean (SD)	0.46 (0.54)	0.51 (0.6)	0.44 (0.51)	P<0.01
Albumin, refrigerated serum (g/dL) : Mean (SD)	4.06 (0.35)	4.01 (0.35)	4.08 (0.35)	P<0.01
Alkaline Phosphatase (ALP) (IU/L) : Mean (SD)	78.23 (26.94)	80.63 (25.97)	77.23 (27.27)	P<0.01
Blood Urea Nitrogen (mg/dL) : Mean (SD)	14.86 (6.02)	15.42 (6.76)	14.63 (5.68)	P<0.01
Glucose, refrigerated serum (mg/dL) : Mean (SD)	101.39 (35.15)	104.83 (37.69)	99.95 (33.94)	P<0.01
Gamma Glutamyl Transferase (GGT) (IU/L) : Mean (SD)	31.31 (43.71)	35.3 (49.92)	29.65 (40.74)	P<0.01
Total Protein (g/dL) : Mean (SD)	7.15 (0.45)	7.1 (0.45)	7.17 (0.45)	P<0.01
N-acetyl-S-(n-propyl)-L-cysteine comt: Mean (SD)	169.52 (216.01)	202.78 (278.74)	155.6 (181.8)	P<0.01
BMXWT - Weight (kg) : Mean (SD)	84.02 (23.31)	88.92 (25.79)	82.01 (21.91)	P<0.01
BMXBMI - Body Mass Index (kg/m2) : Mean (SD)	30 (7.65)	31.86 (8.53)	29.24 (7.12)	P<0.01
BMXWAIST - Waist Circumference (cm) : Mean (SD)	100.65 (17.47)	105.28 (18.08)	98.78 (16.87)	P<0.01
SMQ020 - Smoked at least 100 cigarettes in life : Count (Proportion)	3261 (0.41)	1151 (0.5)	2110 (0.37)	P<0.01
SMQ690A - Used last 5 days - Cigarettes : Count (Proportion)	1376 (0.17)	477 (0.21)	899 (0.16)	P<0.01
SMQ856 - Last 7-d worked at job not at home? : Count (Proportion)	4263 (0.54)	1005 (0.44)	3258 (0.58)	P<0.01
MCQ160b - Ever told had congestive heart failure : Count (Proportion)	281 (0.04)	150 (0.07)	131 (0.02)	P<0.01
MCQ160c - Ever told you had coronary heart disease: Count (Proportion)	338 (0.04)	148 (0.06)	190 (0.03)	P<0.01
MCQ160d - Ever told you had angina/angina pectoris: Count (Proportion)	190 (0.02)	99 (0.04)	91 (0.02)	P<0.01
MCQ160e - Ever told you had heart attack: Count (Proportion)	351 (0.04)	149 (0.06)	202 (0.04)	P<0.01
MCQ160f - Ever told you had a stroke : Count (Proportion)	388 (0.05)	188 (0.08)	200 (0.04)	P<0.01
MCQ160m - Ever told you had thyroid problem: Count (Proportion)	908 (0.11)	383 (0.17)	525 (0.09)	P<0.01
MCQ160p - Ever told you had COPD, emphysema, ChB: Count (Proportion)	717 (0.09)	368 (0.16)	349 (0.06)	P<0.01
MCQ160l - Ever told you had any liver condition: Count (Proportion)	374 (0.05)	181 (0.08)	193 (0.03)	P<0.01
MCQ520 - Abdominal pain during past 12 months? : Count (Proportion)	1693 (0.21)	760 (0.33)	933 (0.17)	P<0.01
MCQ540 - Ever seen a DR about this pain: Count (Proportion)	1128 (0.14)	570 (0.25)	558 (0.1)	P<0.01
MCQ550 - Has DR ever said you have gallstones : Count (Proportion)	826 (0.1)	361 (0.16)	465 (0.08)	P<0.01
MCQ560 - Ever had gallbladder surgery? : Count (Proportion)	855 (0.11)	400 (0.17)	455 (0.08)	P<0.01
MCQ220 - Ever told you had cancer or malignancy: Count (Proportion)	814 (0.1)	313 (0.14)	501 (0.09)	P<0.01
MCQ300b - Close relative had asthma? : Count (Proportion)	2075 (0.26)	731 (0.32)	1344 (0.24)	P<0.01
MCQ300c - Close relative had diabetes? : Count (Proportion)	3653 (0.46)	1217 (0.53)	2436 (0.43)	P<0.01
MCQ300a - Close relative had heart attack? : Count (Proportion)	1011 (0.13)	435 (0.19)	576 (0.1)	P<0.01
MCQ366a - Doctor told you to control/lose weight: Count (Proportion)	2275 (0.29)	947 (0.41)	1328 (0.24)	P<0.01
MCQ366b - Doctor told you to exercise : Count (Proportion)	3330 (0.42)	1336 (0.58)	1994 (0.35)	P<0.01
MCQ366c - Doctor told you to reduce salt in diet : Count (Proportion)	2326 (0.29)	917 (0.4)	1409 (0.25)	P<0.01
MCQ366d - Doctor told you to reduce fat/calories: Count (Proportion)	2505 (0.32)	1013 (0.44)	1492 (0.27)	P<0.01
MCQ371a - Are you now controlling or losing weight : Count (Proportion)	5085 (0.64)	1605 (0.7)	3480 (0.62)	P<0.01
MCQ371c - Are you now reducing salt in diet: Count (Proportion)	4421 (0.56)	1378 (0.6)	3043 (0.54)	P<0.01
MCQ371d - Are you now reducing fat in diet: Count (Proportion)	4663 (0.59)	1488 (0.65)	3175 (0.56)	P<0.01
Used liquid diet : Count (Proportion)	279 (0.04)	123 (0.05)	156 (0.03)	P<0.01
Supplement lose weight: Count (Proportion)	233 (0.03)	101 (0.04)	132 (0.02)	P<0.01
Drank water lose weight : Count (Proportion)	2262 (0.29)	750 (0.33)	1512 (0.27)	P<0.01
Special diet lose weight : Count (Proportion)	289 (0.04)	124 (0.05)	165 (0.03)	P<0.01
Fewer carbs: Count (Proportion)	1336 (0.17)	491 (0.21)	845 (0.15)	P<0.01
Ate fruits veg: Count (Proportion)	2094 (0.26)	697 (0.3)	1397 (0.25)	P<0.01
Changed eating habits: Count (Proportion)	1681 (0.21)	584 (0.25)	1097 (0.19)	P<0.01
Ate less sugar: Count (Proportion)	1791 (0.23)	615 (0.27)	1176 (0.21)	P<0.01
Ate less junk food: Count (Proportion)	1897 (0.24)	645 (0.28)	1252 (0.22)	P<0.01
Weight loss surgery: Count (Proportion)	27 (0)	19 (0.01)	8 (0)	P<0.01
PHQ 9: Mean (SD)	3.18 (4.3)	5.44 (5.29)	2.25 (3.42)	P<0.01
Alcohol (gm) : Mean (SD)	122.25 (173.56)	137.48 (192.56)	115.71 (164.31)	P<0.01
Dietary fiber (gm) : Mean (SD)	16.17 (10.5)	15.2 (9.96)	16.56 (10.68)	P<0.01
Food folate (mcg) : Mean (SD)	206.71 (141.14)	194.26 (129.48)	211.8 (145.35)	P<0.01
Caffeine (mg): Mean (SD)	137.22 (202.64)	157.52 (232.69)	128.92 (188.37)	P<0.01
LBXTR Triglyceride (mg/dL) : Mean (SD)	103.06 (61.96)	112.24 (64.35)	99.15 (60.51)	P<0.01

Fig 1. Demographic information and disease characteristics. Descriptive statistics for demographic characteristics and all covariates within the machine learning model, stratified by whether patients had a sleep disorder. Covariates with SMQ or MCQ labeled in front of it were asked the question written; responses were numeric (integer number) for SMQ and binary (yes, no) for MCQ. Abbreviations: DR = Doctor.

https://doi.org/10.1371/journal.pone.0282622.g001

significant risk factor for a sleep disorder, with odds of a sleep disorder increasing between age 20 until age 60, at which point there does not appear to be an increase in sleep disorder with increasing age. Lastly, there is a strong relationship between waist circumference and a sleep disorder. There is no significant increase in odds of a sleep disorder with increasing waist circumference until after 100cm, at which there is a significant increase in odds of a sleep disorder with increasing waist circumference.

	Metrics	Minimum	5th Percentile	25th Percentile	Median	75th Percentile	95th Percentile	Maximum	Mean	Standard Deviation	Range
XGBoost	Accuracy	0.657	0.709	0.734	0.771	0.796	0.808	0.894	0.762	0.019	0.193
	F1	0.655	0.739	0.750	0.772	0.798	0.823	0.875	0.766	0.006	0.185
	Sensitivity	0.675	0.727	0.770	0.774	0.791	0.847	0.887	0.768	0.023	0.195
	Specificity	0.565	0.693	0.743	0.762	0.797	0.823	0.936	0.782	0.025	0.333
	Positive Predictive Value	0.668	0.731	0.761	0.785	0.819	0.863	0.941	0.806	0.025	0.246
	Negative Predictive Value	0.544	0.640	0.720	0.719	0.774	0.809	0.913	0.737	0.034	0.331
	AUROC	0.755	0.800	0.833	0.840	0.861	0.905	0.918	0.865	0.010	0.173
Random Forest	Metrics	Minimum	5th Percentile	25th Percentile	Median	75th Percentile	95th Percentile	Maximum	Mean	Standard Deviation	Range
	Accuracy	0.645	0.715	0.736	0.744	0.789	0.804	0.879	0.766	0.020	0.215
	F1	0.682	0.717	0.748	0.749	0.793	0.813	0.859	0.780	0.001	0.158
	Sensitivity	0.645	0.734	0.764	0.763	0.786	0.817	0.857	0.784	0.006	0.191
	Specificity	0.575	0.700	0.737	0.757	0.797	0.845	0.923	0.740	0.010	0.332
	Positive Predictive Value	0.648	0.709	0.768	0.806	0.805	0.860	0.944	0.790	0.033	0.260
	Negative Predictive Value	0.532	0.657	0.711	0.703	0.732	0.806	0.896	0.722	0.027	0.323
	AUROC	0.726	0.789	0.844	0.824	0.867	0.861	0.891	0.818	0.006	0.151
Artificial Neural Network	Metrics	Minimum	5th Percentile	25th Percentile	Median	75th Percentile	95th Percentile	Maximum	Mean	Standard Deviation	Range
	Accuracy	0.659	0.705	0.743	0.773	0.785	0.806	0.848	0.773	0.011	0.191
	F1	0.671	0.733	0.736	0.762	0.782	0.816	0.869	0.770	0.017	0.198
	Sensitivity	0.649	0.715	0.745	0.784	0.799	0.832	0.854	0.793	0.014	0.197
	Specificity	0.559	0.669	0.750	0.762	0.765	0.816	0.902	0.753	0.018	0.323
	Positive Predictive Value	0.635	0.730	0.750	0.781	0.816	0.826	0.905	0.791	0.013	0.258
	Negative Predictive Value	0.536	0.654	0.706	0.717	0.735	0.813	0.895	0.725	0.015	0.338
	AUROC	0.727	0.786	0.800	0.848	0.869	0.884	0.920	0.831	0.011	0.155
Adaptive Boosting	Metrics	Minimum	5th Percentile	25th Percentile	Median	75th Percentile	95th Percentile	Maximum	Mean	Standard Deviation	Range
	Accuracy	0.663	0.725	0.743	0.755	0.765	0.798	0.875	0.744	0.014	0.188
	F1	0.657	0.729	0.731	0.739	0.771	0.791	0.883	0.743	0.005	0.189
	Sensitivity	0.640	0.719	0.773	0.780	0.776	0.828	0.889	0.796	0.007	0.194
	Specificity	0.563	0.678	0.720	0.740	0.765	0.850	0.930	0.754	0.010	0.324
	Positive Predictive Value	0.665	0.722	0.769	0.791	0.840	0.832	0.929	0.800	0.040	0.260
	Negative Predictive Value	0.549	0.656	0.676	0.748	0.758	0.803	0.886	0.748	0.012	0.329
	AUROC	0.724	0.812	0.820	0.837	0.848	0.858	0.904	0.839	0.014	0.140

Fig 2. Comparison of different machine learning models. Comparison of four machine learning models (XGBoost, Random Forest, Artificial Neural Network, Adaptive Boosting) using the model statistics computed from the 20% test set: Accuracy, F1, Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, and AUROC.

https://doi.org/10.1371/journal.pone.0282622.g002

Fig 7A show SHAP explanations for the XGBoost model and show a positive relationship between each gm of alcohol use and odds of a sleep disorder. Likewise Fig 7B show a positive relationship between each mg of caffeine and odds of a sleep disorder.

Discussion

In this retrospective, cross sectional cohort of United States adults, a machine learning model utilizing demographic, laboratory, physical examination, and lifestyle questionnaire data had strong predictive accuracy (AUROC = 0.87). The greatest predictors for a sleep disorder included depression (PHQ-9), weight, age, and waist circumference.

Prior studies have accurately predicted the presence of sleep disorders using machine-learning methods from a variety of datasets using numerous machine-learning methods [36–38]. Short-term insomnia detection was conducted using a single-channel sleep Electrooculography. Furthermore, natural language processing on 18,901 tweets was conducted to find correlations between words related to insomnia and negative health information [38–40]. Furthermore, a comparative study of 15 machine learning algorithms identified 14 main factors for the prediction of insomnia, identifying that vision problems, mobility problems, and sleep disorders were significantly related to insomnia [38, 39]. These studies highlight the utility of machine learning models in identifying patients at risk for sleep disorders. What our study adds to the literature is a large dataset (N = 7,929) and a diverse wealth of potential covariates (700+ covariates) to study how lifestyle, diet, demographic, and medical covariates are able to predict insomnia.

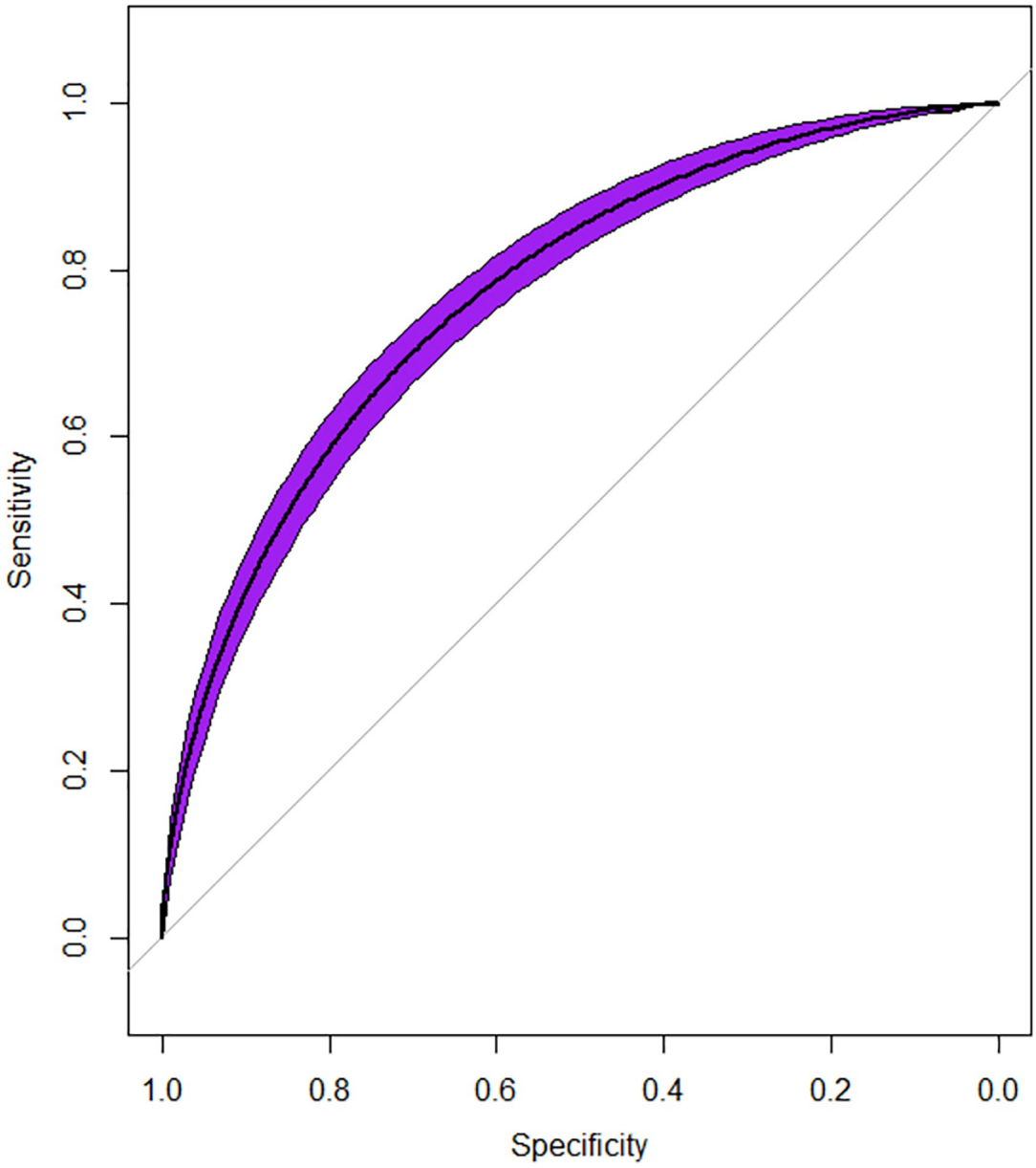

Fig 3. Receiver operator characteristic curve and model statistics. The Receiver operating characteristic curve for the machine-learning model predicting a sleep disorder. AUROC = 0.87.

https://doi.org/10.1371/journal.pone.0282622.g003

The visualizations completed for the top four continuous covariates were concordant with current literature: there is strong epidemiological evidence that sleep problems are heavily linked with depression. Multiple papers have found difficulty falling asleep and decreased hours of sleep with increased depression [41–53]. Additionally, depression has bene linked to lower quality sleep and increased day time exhaustion [31, 34, 46, 54, 55]. There is also strong literature evidence for the link between weight and sleep disorders [4]. There is epidemiological evidence for the relationship between increased age and increased sleep disorders, older age has been associated with increased sleep latency, decreased time spent in rapid eye movement (REM) sleep and stage-3 sleep, and increased frequency of waking up during the night

Feature	Gain	Cover	Frequency
PHQ 9	0.309	0.197	0.061
Age	0.075	0.094	0.061
MCQ366b - Doctor told you to exercise	0.039	0.048	0.019
BMXWT - Weight (kg)	0.030	0.034	0.038
BMXWAIST - Waist Circumference (cm)	0.027	0.021	0.040
Alcohol (gm)	0.025	0.027	0.041
Blood cadmium (ug/L)	0.025	0.031	0.041
Food folate (mcg)	0.024	0.018	0.038
Caffeine (mg)	0.023	0.021	0.034
RBC folate (ng/mL)	0.023	0.021	0.036
MCQ540 - Ever seen a DR about this pain	0.022	0.038	0.017
Red blood cell count (million cells/uL)	0.021	0.020	0.034
Dietary fiber (gm)	0.020	0.012	0.032
N-acetyl-S-(n-propyl)-L-cysteine comt	0.019	0.020	0.029
LBXTR Triglyceride (mg/dL)	0.019	0.018	0.030
Glucose, refrigerated serum (mg/dL)	0.018	0.015	0.030
Insulin (pmol/L)	0.017	0.013	0.028
HS C-Reactive Protein (mg/L)	0.017	0.016	0.031
Red cell distribution width (%)	0.017	0.028	0.028
BMXBMI - Body Mass Index (kg/m^2)	0.015	0.010	0.024
Alkaline Phosphatase (ALP) (IU/L)	0.015	0.011	0.028
Total Protein (g/dL)	0.015	0.016	0.024
Gamma Glutamyl Transferase (GGT) (IU/L)	0.014	0.012	0.025
Glycohemoglobin (%)	0.013	0.015	0.023
Blood Urea Nitrogen (mg/dL)	0.012	0.009	0.020
Cotinine, Serum (ng/mL)	0.011	0.009	0.019
Hydroxycotinine, Serum (ng/mL)	0.011	0.010	0.015
SMQ856 - Last 7-d worked at job not at home?	0.011	0.020	0.009
Albumin, refrigerated serum (g/dL)	0.010	0.008	0.017
MCQ366a - Doctor told you to control/lose weight	0.010	0.011	0.008
MCQ520 - Abdominal pain during past 12 months?	0.008	0.008	0.009
MCQ300a - Close relative had heart attack?	0.008	0.016	0.010
MCQ160p - Ever told you had COPD, emphysema, ChB	0.008	0.016	0.008
MCQ366c - Doctor told you to reduce salt in diet	0.007	0.011	0.007
MCQ160b - Ever told had congestive heart failure	0.006	0.018	0.008
MCQ560 - Ever had gallbladder surgery?	0.006	0.013	0.007
MCQ300b - Close relative had asthma?	0.006	0.018	0.008
MCQ160m - Ever told you had thyroid problem	0.005	0.008	0.005
Fewer carbs	0.005	0.008	0.005
Changed eating habits	0.004	0.009	0.005
SMQ020 - Smoked at least 100 cigarettes in life	0.004	0.007	0.006
MCQ220 - Ever told you had cancer or malignancy	0.003	0.004	0.004
Used liquid diet	0.003	0.011	0.004
MCQ366d - Doctor told you to reduce fat/calories	0.002	0.002	0.003
MCQ371a - Are you now controlling or losing weight	0.002	0.003	0.004
Ate less junk food	0.002	0.001	0.003
MCQ160l - Ever told you had any liver condition	0.002	0.007	0.003
MCQ300c - Close relative had diabetes?	0.002	0.001	0.003
MCQ160f - Ever told you had a stroke	0.002	0.005	0.002
MCQ371c - Are you now reducing salt in diet	0.001	0.001	0.002
Ate fruits veg	0.001	0.001	0.002
MCQ160e - Ever told you had heart attack	0.001	0.001	0.002
Special diet lose weight	0.001	0.001	0.001
Drank water lose weight	0.001	0.000	0.002
Gender	0.001	0.000	0.002
Ate less sugar	0.001	0.001	0.001
Supplement lose weight	0.001	0.002	0.001
MCQ371d - Are you now reducing fat in diet	0.001	0.000	0.001
SMQ690A - Used last 5 days - Cigarettes	0.000	0.000	0.001
MCQ510f - Liver condition: Other liver disease	0.000	0.001	0.000
MCQ550 - Has DR ever said you have gallstones	0.000	0.000	0.001
MCQ160d - Ever told you had angina/angina pectoris	0.000	0.000	0.000
MCQ160c - Ever told you had coronary heart disease	0.000	0.000	0.000
Weight loss surgery	0.000	0.001	0.000

Fig 4. Model gain statistics. The Gain, Cover, and Frequency of all covariates within the XGBoost model. The Gain represents the relative contribution of the feature to the model and is the most important metric of model importance within this study. Covariates ordered according to the Gain statistic.

https://doi.org/10.1371/journal.pone.0282622.g004

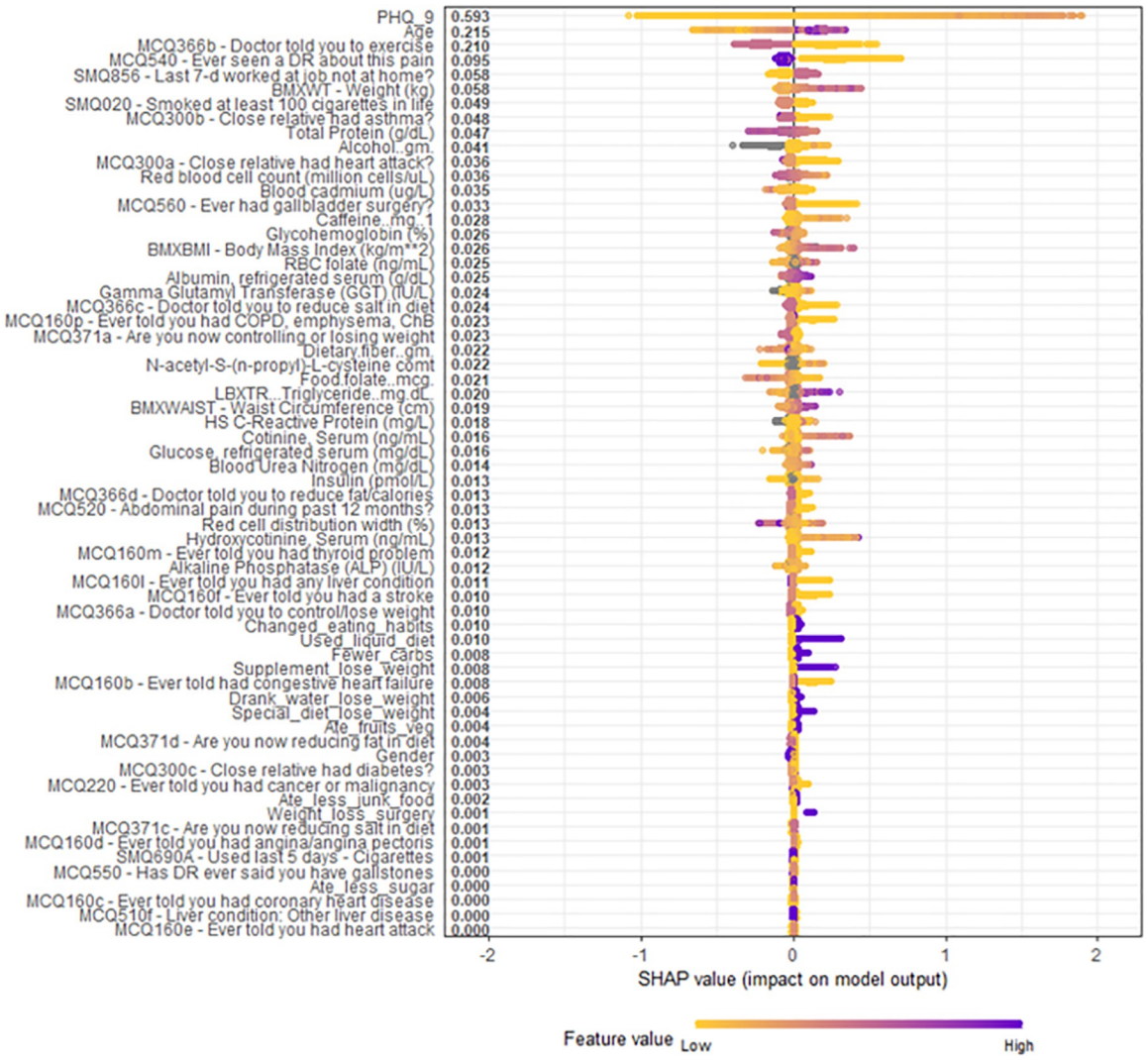

Fig 5. Overall SHAP explanations. SHAP explanations, purple color representing higher values of the covariate while yellow representing lower values of the covariate. X-axis is the change in log-odds for a sleep disorder. Covariates ordered according to the Gain statistic. Covariates with SMQ or MCQ labeled in front of it were asked the question written; responses were numeric (integer number) for SMQ and binary (yes, no) for MCQ. Abbreviations: DR = Doctor.

https://doi.org/10.1371/journal.pone.0282622.g005

[56–62]. Furthermore, increased caffeine usage has been found to be linked with difficulty falling asleep, decreased time falling asleep, and decreased quality of sleep [63]. Additionally, increased alcohol is associated with sleep disorders, leading to decreased sleep latency and potential physiologic need for alcohol as a depressant to allow for sleep in multiple patients [64–66].

Since visualizations for risk factors match literature relationships, we have increased confidence that the machine learning model is able to capture the actual physiological relationships of these covariates. These transparent machine-learning tools allow for increased confidence that these algorithms are picking up true signal within these covariates to predict the presence of a sleep disorder rather than just replicating potential biases stemming from systemic data-quality errors that are present within the dataset. Additionally, these SHAP visualizations allow us to interpret that the increase predictive power of these machine-learning methods is

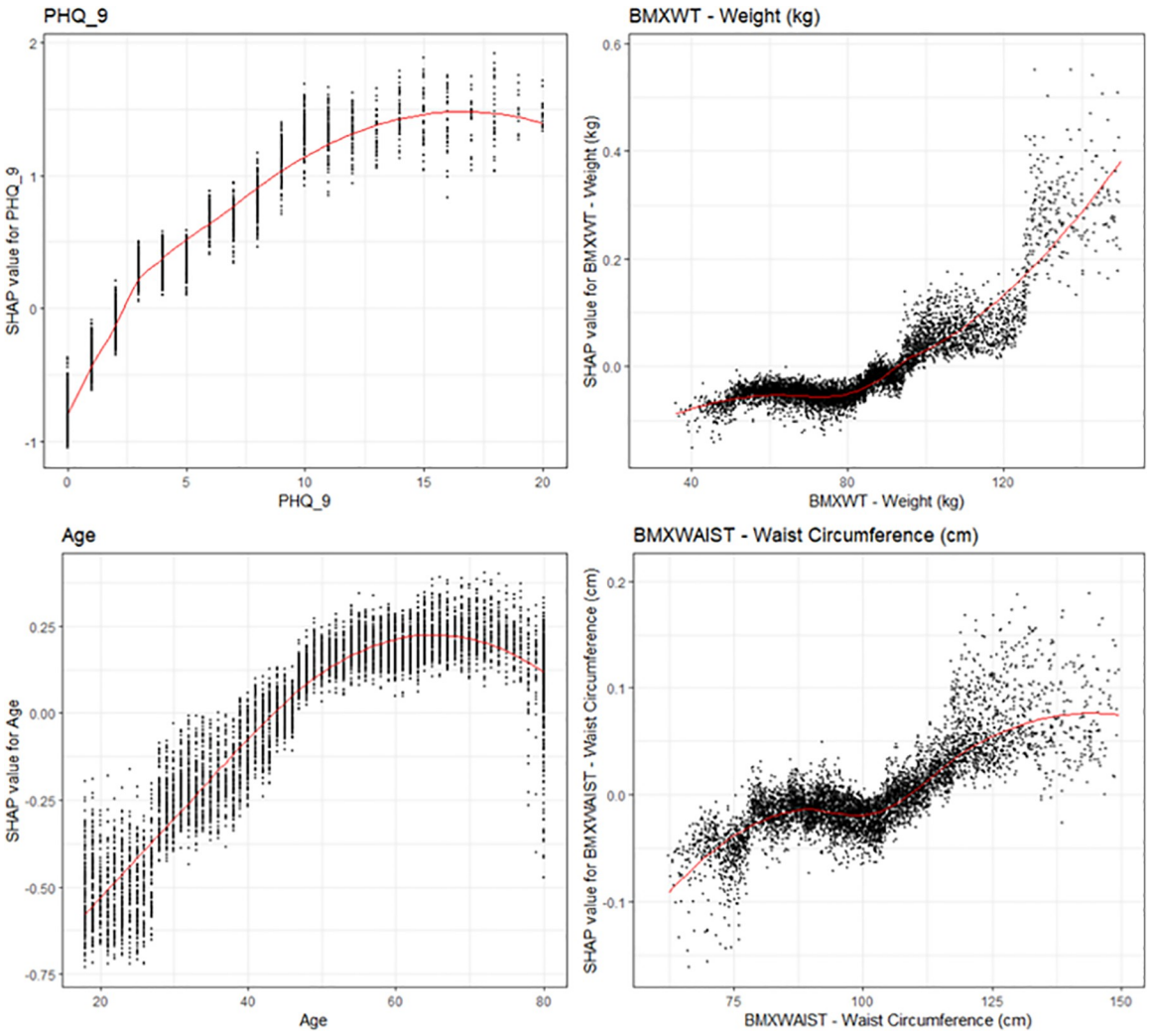

Fig 6. SHAP explanations for the Top 4 continuous covariates. SHAP explanations, covariate value on the x-axis, change in log-odds on the y-axis, red line represents the relationship between the covariate and log-odds for CAD, each black dot represents an observation. Covariates: top left–PHQ-9, top right–Body weight, bottom left–patient age, bottom right–waist circumference.

https://doi.org/10.1371/journal.pone.0282622.g006

associated with the ability for these non-parametric methods to more accurately capture the non-linear interactive relationship between the covariates, rather than just over-fitting the model to get increased accuracy.

The greatest strength of this algorithmic method for identification of the covariates is the ability to search through hundreds of covariates systematically without relying upon judgment form the researcher, which may be muddled by potential personal biases. This method also allows for the ranking of the relative importance of each of these covariates through the cover statistic, which allows us to obtain the relative contribution to the prediction each covariate has and thus infer from there an estimate for the relative contribution to true risk for a sleep disorder that each patient has. Another strength is that after these covariates are selected and the model built, SHAP visualizations can be used to make sure that each of the covariate either matches current literature understandings of the covariate's association with a sleep disorder or in the case of a discrepancy, allow researchers to validate the plausibility of this feature and then evaluate for potential errors in data-quality.

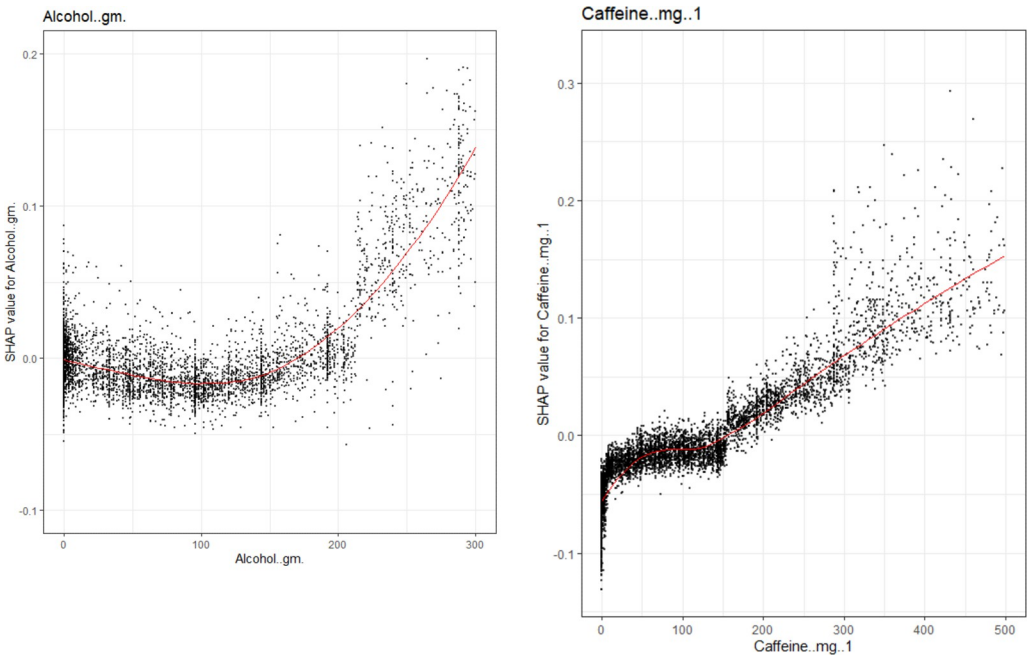

Fig 7. **a**: Covariates of interest to evaluate sensibility of the model. SHAP explanations for the relationship between Alcohol and odds of a sleep disorder. Covariate value on the x-axis, change in log-odds on the y-axis, red line represents the relationship between the covariate and log-odds for a sleep disorder, each black dot represents an observation. **b:** SHAP explanations for the relationship between Caffeine intake and odds of a sleep disorder. Covariate value on the x-axis, change in log-odds on the y-axis, red line represents the relationship between the covariate and log-odds for a sleep disorder, each black dot represents an observation.

https://doi.org/10.1371/journal.pone.0282622.g007

A potential weakness to this machine-learning analysis is the necessity of the retrospective nature of this cohort. The covariates that were selected within this study will be better at predicting risk for a sleep disorder for this cohort than for other cohorts. However, this was limited by the use of training: testing sets to be able to minimize the errors that come with overfitting. Furthermore, visualizations of SHAP allow researchers to test for physiologic plausibility of each of these covariates and allows for effective analysis by researchers of whether these effects are due to true signal or if they are just noise that may be contributing to a type-1 error.

Given the analysis of the strengths and weaknesses of these methods, we argue that use of machine-learning methods can be an effective first step in the identification of risk-factors that can then be further selected by clinicians based upon the specific clinical presentation.

Limitations

This study has several strengths and weaknesses. We utilized the NHANES dataset, which is a retrospective cohort, carrying the limitations of retrospective studies. However, this study allows for the selection of a large cohort, evaluation of data quality, and due to the publicly available nature of the cohort, allows for increased replication and follow-up studies based upon the same cohort. Furthermore, the cohort relied on surveys to obtain the outcome of interest (a sleep disorder requiring medical attention) as well as the dietary and lifestyle information. More accurate measurements may have been achieved with prospective studies with automated measurement of foods. However, self-reported survey information allows for the volume of participants to be included within this study. Another weakness was the voluntary nature of this cohort, with participants choosing to opt into the study instead of being

randomly selected. This may artificially select a different cohort that may significantly differ from the population. However, our analysis found a demographically diverse population, so these results may still be generalizable to other cohorts.

Conclusion

Machine learning models can effectively predict risk for a sleep disorder using demographic, laboratory, physical exam, and lifestyle covariates and identify key risk factors. Depression, age, weight, and waist circumference were the strongest predictors of sleep disorder.

Author Contributions

Conceptualization: Alexander A. Huang, Samuel Y. Huang.

Data curation: Alexander A. Huang.

Formal analysis: Alexander A. Huang, Samuel Y. Huang.

Funding acquisition: Samuel Y. Huang.

Investigation: Alexander A. Huang.

Methodology: Alexander A. Huang.

Project administration: Alexander A. Huang.

Resources: Alexander A. Huang.

Software: Alexander A. Huang, Samuel Y. Huang.

Supervision: Alexander A. Huang.

Validation: Alexander A. Huang, Samuel Y. Huang.

Visualization: Alexander A. Huang.

Writing – original draft: Alexander A. Huang.

Writing – review & editing: Alexander A. Huang, Samuel Y. Huang.

References

1. Wang J, Ren X. Association Between Sleep Duration and Sleep Disorder Data from the National Health and Nutrition Examination Survey and Stroke Among Adults in the United States. Med Sci Monit. 2022; 28:e936384. Epub 20220703. https://doi.org/10.12659/MSM.936384 PMID: 35780293; PubMed Central PMCID: PMC9261468.
2. Wolf C, Wolf S, Weiss M, Nino G. Children's Environmental Health in the Digital Era: Understanding Early Screen Exposure as a Preventable Risk Factor for Obesity and Sleep Disorders. Children (Basel). 2018; 5(2). Epub 20180223. https://doi.org/10.3390/children5020031 PMID: 29473855; PubMed Central PMCID: PMC5836000.
3. Guo Q, Xie W, Peng R, Ma Y, Chong F, Wang Y, et al. A Dose-Response Relationship Between Sleep Duration and Stroke According to Nonhealth Status in Central China: A Population-based Epidemiology Survey. J Stroke Cerebrovasc Dis. 2019; 28(7):1841–52. Epub 20190507. https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.04.016 PMID: 31076320.
4. Piepoli MF. Editor's Presentation Benefit of healthy lifestyle on cardiovascular risk factor control: Focus on body weight, exercise and sleep quality. Eur J Prev Cardiol. 2019; 26(12):1235–8. https://doi.org/10.1177/2047487319861847 PMID: 31298111.
5. Carriedo-Diez B, Tosoratto-Venturi JL, Canton-Manzano C, Wanden-Berghe C, Sanz-Valero J. The Effects of the Exogenous Melatonin on Shift Work Sleep Disorder in Health Personnel: A Systematic Review. Int J Environ Res Public Health. 2022; 19(16). Epub 20220817. https://doi.org/10.3390/ijerph191610199 PMID: 36011832; PubMed Central PMCID: PMC9408537.
6. Sivertsen B, Hysing M, Harvey AG, Petrie KJ. The Epidemiology of Insomnia and Sleep Duration Across Mental and Physical Health: The SHoT Study. Front Psychol. 2021; 12:662572. Epub

20210614. https://doi.org/10.3389/fpsyg.2021.662572 PMID: 34194368; PubMed Central PMCID: PMC8236531.

7. Shapiro CM, Dement WC. ABC of sleep disorders. Impact and epidemiology of sleep disorders. BMJ. 1993; 306(6892):1604–7. https://doi.org/10.1136/bmj.306.6892.1604 PMID: 8329929; PubMed Central PMCID: PMC1678030.
8. Trosman I, Ivanenko A. Classification and Epidemiology of Sleep Disorders in Children and Adolescents. Child Adolesc Psychiatr Clin N Am. 2021; 30(1):47–64. Epub 20201020. https://doi.org/10.1016/j.chc.2020.08.002 PMID: 33223068.
9. Bliwise DL. Epidemiology of Age-Dependence in Sleep Disordered Breathing (Sdb) in Old Age: The Bay Area Sleep Cohort (Basc). Sleep Med Clin. 2009; 4(1):57–64. https://doi.org/10.1016/j.jsmc.2008.11.004 PMID: 20161180; PubMed Central PMCID: PMC2726445.
10. Kerkhof GA. Epidemiology of sleep and sleep disorders in The Netherlands. Sleep Med. 2017; 30:229–39. Epub 20161105. https://doi.org/10.1016/j.sleep.2016.09.015 PMID: 28215254.
11. Ancoli-Israel S. Epidemiology of sleep disorders. Clin Geriatr Med. 1989; 5(2):347–62. PMID: 2665916.
12. Partinen M. Epidemiology of sleep disorders. Handb Clin Neurol. 2011; 98:275–314. https://doi.org/10.1016/B978-0-444-52006-7.00018-6 PMID: 21056193.
13. Shamsadini A, Amizadeh M, Motamed S, Rajabi F. Epidemiology of Sleep Disorders Among Adolescents in Kerman, Iran, in 2019. Soc Work Public Health. 2022:1–6. Epub 20220613. https://doi.org/10.1080/19371918.2022.2089310 PMID: 35695466.
14. Pluzhnikov MS, Blotskii AA. [Epidemiology and laser correction of chronic diseases observed in snoring and obstructive sleep apnea]. Vestn Otorinolaringol. 2002;(3):12–5. PMID: 12227021.
15. Rainer MA, Palmer PH, Xie B. Sleep Duration and Chronic Disease Among Older Native Hawaiians or Other Pacific Islanders and Asians: Analysis of the Behavioral Risk Factor Surveillance System. J Racial Ethn Health Disparities. 2022. Epub 20220915. https://doi.org/10.1007/s40615-022-01409-0 PMID: 36109435.
16. You Y, Zhong X, Liu G, Yang Z. Automatic sleep stage classification: A light and efficient deep neural network model based on time, frequency and fractional Fourier transform domain features. Artif Intell Med. 2022; 127:102279. Epub 20220309. https://doi.org/10.1016/j.artmed.2022.102279 PMID: 35430040.
17. Jeong E, Cha KS, Shin HR, Kim EY, Jun JS, Kim TJ, et al. Corrigendum to "Alerting network alteration in isolated rapid eye movement sleep behavior disorder patients with mild cognitive impairment" [Sleep Med. 89 (2022) 10–18/SLEEP-D-21-00679]. Sleep Med. 2022;98:87–8. Epub 20220705. https://doi.org/10.1016/j.sleep.2022.06.019 PMID: 35803115.
18. Shah A, Turkistani A, Luenam K, Yaqub S, Ananias P, Jose AM, et al. Is Shift Work Sleep Disorder a Risk Factor for Metabolic Syndrome and Its Components? A Systematic Review of Cross-Sectional Studies. Metab Syndr Relat Disord. 2022; 20(1):1–10. Epub 20211008. https://doi.org/10.1089/met.2021.0070 PMID: 34637354.
19. Vinstrup J, Jakobsen MD, Andersen LL. Poor Sleep Is a Risk Factor for Low-Back Pain among Healthcare Workers: Prospective Cohort Study. Int J Environ Res Public Health. 2020; 17(3). Epub 20200205. https://doi.org/10.3390/ijerph17030996 PMID: 32033339; PubMed Central PMCID: PMC7036951.
20. Johnson EO, Roehrs T, Roth T, Breslau N. Epidemiology of alcohol and medication as aids to sleep in early adulthood. Sleep. 1998; 21(2):178–86. https://doi.org/10.1093/sleep/21.2.178 PMID: 9542801.
21. Merino D, Gerard AO, Van Obberghen EK, Ben Othman N, Ettore E, Giordana B, et al. Medications as a Trigger of Sleep-Related Eating Disorder: A Disproportionality Analysis. J Clin Med. 2022; 11(13). Epub 20220704. https://doi.org/10.3390/jcm11133890 PMID: 35807172; PubMed Central PMCID: PMC9267629.
22. Bergman E, Loyttyniemi E, Myllyntausta S, Rautava P, Korhonen PE. Temporal changes in self-reported sleep quality, sleep duration and sleep medication use in relation to temporal changes in quality of life and work ability over a 1-year period among Finnish municipal employees. J Sleep Res. 2022: e13605. Epub 20220415. https://doi.org/10.1111/jsr.13605 PMID: 35429092.
23. Forthun I, Waage S, Pallesen S, Moen BE, Bjorvatn B. A shift to something better? A longitudinal study of work schedule and prescribed sleep medication use in nurses. Occup Environ Med. 2022. Epub 20220620. https://doi.org/10.1136/oemed-2022-108251
24. Lambek R, Thomsen PH, Sonuga-Barke EJS, Jennum P, Sorensen AV. The Association between Sleep Problems and Neuropsychological Deficits in Medication-naive Children with ADHD. Behav Sleep Med. 2022; 20(4):429–41. Epub 20210603. https://doi.org/10.1080/15402002.2021.1931222 PMID: 34081546.
25. Leite FMC, Xavier JCS, Silva RP, Wandekoken KD, Tavares FL, Amorim MHC. Prevalence and factors associated with the use of sleep-inducing medication among women receiving Primary Health Care: a

cross-sectional study in Vitoria, Espirito Santo, Brazil, 2014. Epidemiol Serv Saude. 2022; 31(1): e2021347. Epub 20220420. https://doi.org/10.1590/S1679-49742022000100016 PMID: 35475997.

26. Masse M, Henry H, Cuvelier E, Pincon C, Pavy M, Beeuwsaert A, et al. Sleep Medication in Older Adults: Identifying the Need for Support by a Community Pharmacist. Healthcare (Basel). 2022; 10(1). Epub 20220113. https://doi.org/10.3390/healthcare10010147 PMID: 35052310; PubMed Central PMCID: PMC8775744.

27. Celik Y, Balcan B, Peker Y. CPAP Intervention as an Add-On Treatment to Lipid-Lowering Medication in Coronary Artery Disease Patients with Obstructive Sleep Apnea in the RICCADSA Trial. J Clin Med. 2022; 11(1). Epub 20220105. https://doi.org/10.3390/jcm11010273 PMID: 35012012; PubMed Central PMCID: PMC8745784.

28. Darchia N, Campbell IG, Basishvili T, Eliozishvili M, Tchintcharauli T, Oniani N, et al. Longitudinal assessment of NREM sleep EEG in typically developing and medication-free ADHD adolescents: first year results. Sleep Med. 2021; 80:171–5. Epub 20210201. https://doi.org/10.1016/j.sleep.2021.01.052 PMID: 33601229.

29. Forthun I, Waage S, Pallesen S, Moen BE, Bjorvatn B. Sleep medication and melatonin use among Norwegian nurses—A cross-sectional study. Nurs Open. 2022; 9(1):233–44. Epub 20210917. https://doi.org/10.1002/nop2.1057 PMID: 34534412; PubMed Central PMCID: PMC8685790.

30. Full KM, Pusalavidyasagar S, Palta P, Sullivan K, Shin JI, Gottesman RF, et al. Associations of late-life sleep medication use with incident dementia in the Atherosclerosis Risk in Communities (ARIC) Study. J Gerontol A Biol Sci Med Sci. 2022. Epub 20220414. https://doi.org/10.1093/gerona/glac088 PMID: 35421897.

31. Gong YC, Chen X, Song QT, Gan Y, Zhang B, Li BS, et al. A randomized placebo-controlled study: Phellodendron Bawei tablets combined with standard management can improve storage symptoms, sleep quality, and medication compliance in patients with benign prostatic hyperplasia compared to placebo with standard management. Transl Androl Urol. 2021; 10(8):3423–31. https://doi.org/10.21037/tau-21-588 PMID: 34532267; PubMed Central PMCID: PMC8421819.

32. Hou J, Zhang R. Clinical Analysis on the Effects of Tandospirone Citrate Assisted by Drawing Therapy on Medication Compliance and Sleep Quality in Patients with Anxiety Disorders. Emerg Med Int. 2022; 2022:9295627. Epub 20220830. https://doi.org/10.1155/2022/9295627 PMID: 36081955; PubMed Central PMCID: PMC9448592.

33. Miettinen T, Sverloff J, Lappalainen OP, Linton SJ, Sipila K, Kalso E. Sleep problems in pain patients entering tertiary pain care: the role of pain-related anxiety, medication use, self-reported diseases, and sleep disorders. Pain. 2022; 163(7):e812–e20. Epub 20210923. https://doi.org/10.1097/j.pain.0000000000002497 PMID: 34561395; PubMed Central PMCID: PMC9199106.

34. Mozer CL, Bhagat PH, Seward SA, Mason NR, Anderson SL, Byron M, et al. Optimizing Oral Medication Schedules for Inpatient Sleep: A Quality Improvement Intervention. Hosp Pediatr. 2021; 11(4):327–33. Epub 20210317. https://doi.org/10.1542/hpeds.2020-002261 PMID: 33731336; PubMed Central PMCID: PMC8006200.

35. Yamamoto H, Nakagawa E, Kita Y, Kaga Y, Inagaki M. Effect of anti-attention-deficit hyperactivity disorder (ADHD) medication on clinical seizures and sleep EEG: A retrospective study of Japanese children with ADHD. Neuropsychopharmacol Rep. 2021; 41(4):511–21. Epub 20211020. https://doi.org/10.1002/npr2.12215 PMID: 34668641; PubMed Central PMCID: PMC8698674.

36. Borhani F, Shafiepour Motlagh M, Ehsani AH, Rashidi Y, Maddah S, Mousavi SM. On the predictability of short-lived particulate matter around a cement plant in Kerman, Iran: machine learning analysis. Int J Environ Sci Technol (Tehran). 2022:1–14. Epub 20221109. https://doi.org/10.1007/s13762-022-04645-3 PMID: 36405244; PubMed Central PMCID: PMC9643923.

37. Peng J, Yang K, Tian H, Lin Y, Hou M, Gao Y, et al. The mechanisms of Qizhu Tangshen formula in the treatment of diabetic kidney disease: Network pharmacology, machine learning, molecular docking and experimental assessment. Phytomedicine. 2022; 108:154525. Epub 20221106. https://doi.org/10.1016/j.phymed.2022.154525 PMID: 36413925.

38. Ahuja R. Comparative study of various machine learning algorithms for prediction of insomnia. Research Anthology on Machine Learning Techniques, Methods, and Applications. 2022;IGI Global:776–99.

39. Zulfiker M, Kabir N., Biswas A. Chakraborty A., P. Predicting Insomnia Using Multilayer Stacked Ensemble Model. In International Conference on Advances in Computing and Data Sciences 2021.

40. Siddiqui MM, Srivastava G, Saeed SH. Diagnosis of insomnia sleep disorder using short time frequency analysis of PSD approach applied on EEG signal using channel ROC-LOC. Sleep Sci. 2016; 9(3):186–91. Epub 20160721. https://doi.org/10.1016/j.slsci.2016.07.002 PMID: 28123658; PubMed Central PMCID: PMC5241612.

41. Chen RF, Cai Y, Zhu ZH, Hou WL, Chen P, Wang J, et al. Sleep disorder as a clinical risk factor of major depression: associated with cognitive impairment. Asian J Psychiatr. 2022; 76:103228. Epub 20220809. https://doi.org/10.1016/j.ajp.2022.103228 PMID: 35973338.

42. Chen X, Jiang F, Yang Q, Zhang P, Zhu H, Liu C, et al. Bilateral repetitive transcranial magnetic stimulation ameliorated sleep disorder and hypothalamic-pituitary-adrenal axis dysfunction in subjects with major depression. Front Psychiatry. 2022; 13:951595. Epub 20220825. https://doi.org/10.3389/fpsyt.2022.951595 PMID: 36090377; PubMed Central PMCID: PMC9452697.

43. Gardner W, Fuchs F, Durieux L, Bourgin P, Coenen VA, Dobrossy M, et al. Slow Wave Sleep Deficits in the Flinders Sensitive Line Rodent Model of Depression: Effects of Medial Forebrain Bundle Deep-Brain Stimulation. Neuroscience. 2022; 498:31–49. Epub 20220622. https://doi.org/10.1016/j.neuroscience.2022.06.023 PMID: 35750113.

44. Hoyos CM, Bartlett DJ, Phillips CL. Is Obstructive Sleep Apnea a Risk Factor for Depression in Coronary Artery Disease? Ann Am Thorac Soc. 2019; 16(1):49–50. https://doi.org/10.1513/AnnalsATS.201810-728ED PMID: 30592447.

45. Huan Y, Mujun X, Xin L, Ping Z, Limei F, Aming L, et al. Short Sleep Duration as a Risk Factor for Depression, Anxiety and Fatigue in Patients with Leukemia. Neuropsychiatr Dis Treat. 2022; 18:1573–82. Epub 20220729. https://doi.org/10.2147/NDT.S362229 PMID: 35937713; PubMed Central PMCID: PMC9346604.

46. Kazmi N, Wallen GR, Yang L, Alkhatib J, Schwandt ML, Feng D, et al. An exploratory study of pro-inflammatory cytokines in individuals with alcohol use disorder: MCP-1 and IL-8 associated with alcohol consumption, sleep quality, anxiety, depression, and liver biomarkers. Front Psychiatry. 2022; 13:931280. Epub 20220811. https://doi.org/10.3389/fpsyt.2022.931280 PMID: 36032219; PubMed Central PMCID: PMC9405018.

47. Li M, Cui J, Xu B, Wei Y, Fu C, Lv X, et al. Sleep Disturbances and Depression Are Co-morbid Conditions: Insights From Animal Models, Especially Non-human Primate Model. Front Psychiatry. 2021; 12:827541. Epub 20220125. https://doi.org/10.3389/fpsyt.2021.827541 PMID: 35145441; PubMed Central PMCID: PMC8821160.

48. Mungo A, Hein M, Lanquart JP, Loas G. [Atypical depression as a risk factor for obstructive sleep apnea syndrome in young adults]. Encephale. 2022; 48(2):171–8. Epub 20210604. https://doi.org/10.1016/j.encep.2021.02.018 PMID: 34092378.

49. Rosenstrom T, Jokela M, Puttonen S, Hintsanen M, Pulkki-Raback L, Viikari JS, et al. Pairwise measures of causal direction in the epidemiology of sleep problems and depression. PLoS One. 2012; 7(11):e50841. Epub 20121130. https://doi.org/10.1371/journal.pone.0050841 PMID: 23226400; PubMed Central PMCID: PMC3511346.

50. Ryu D, Jee HJ, Kim SY, Hwang SH, Pil GB, Jung YS. Luteolin-7-O-Glucuronide Improves Depression-like and Stress Coping Behaviors in Sleep Deprivation Stress Model by Activation of the BDNF Signaling. Nutrients. 2022; 14(16). Epub 20220812. https://doi.org/10.3390/nu14163314 PMID: 36014820; PubMed Central PMCID: PMC9412559.

51. Stubbs B, Koyanagi A, Thompson T, Veronese N, Carvalho AF, Solomi M, et al. The epidemiology of back pain and its relationship with depression, psychosis, anxiety, sleep disturbances, and stress sensitivity: Data from 43 low- and middle-income countries. Gen Hosp Psychiatry. 2016; 43:63–70. Epub 20160930. https://doi.org/10.1016/j.genhosppsych.2016.09.008 PMID: 27796261.

52. Sumi Y, Masuda F, Kadotani H, Ozeki Y. The prevalence of depression in isolated/idiopathic rapid eye movement sleep behavior disorder: A systematic review and meta-analysis. Sleep Med Rev. 2022; 65:101684. Epub 20220809. https://doi.org/10.1016/j.smrv.2022.101684 PMID: 36150254.

53. Wang D, Zhao J, Zhai S, Ye H, Bu L, Fan F. Does sleep disturbance predicts posttraumatic stress disorder and depression among college students during COVID-19 lockdown? A longitudinal survey. Front Public Health. 2022; 10:986934. Epub 20220913. https://doi.org/10.3389/fpubh.2022.986934 PMID: 36176529; PubMed Central PMCID: PMC9514232.

54. Alamir YA, Zullig KJ, Kristjansson AL, Wen S, Misra R, Montgomery-Downs H. A theoretical model of college students' sleep quality and health-related quality of life. J Behav Med. 2022. Epub 20220813. https://doi.org/10.1007/s10865-022-00348-9 PMID: 35962152.

55. Prajsuchanai T, Tanphaichitr A, Hosiri T, Ungkanont K, Banhiran W, Vathanophas V, et al. Prevalence of high-risk for obstructive sleep apnea in attention deficit hyperactivity disorder children referred to psychiatry clinic and impact on quality of life. Front Psychiatry. 2022; 13:926153. Epub 20220722. https://doi.org/10.3389/fpsyt.2022.926153 PMID: 35935414; PubMed Central PMCID: PMC9353399.

56. Costa e Silva JA, Chase M, Sartorius N, Roth T. Special report from a symposium held by the World Health Organization and the World Federation of Sleep Research Societies: an overview of insomnias and related disorders—recognition, epidemiology, and rational management. Sleep. 1996; 19(5):412–6. https://doi.org/10.1093/sleep/19.5.412 PMID: 8843532.

57. Lemma S, Patel SV, Tarekegn YA, Tadesse MG, Berhane Y, Gelaye B, et al. The Epidemiology of Sleep Quality, Sleep Patterns, Consumption of Caffeinated Beverages, and Khat Use among Ethiopian College Students. Sleep Disord. 2012; 2012:583510. Epub 20121125. https://doi.org/10.1155/2012/583510 PMID: 23710363; PubMed Central PMCID: PMC3581089.

58. Matricciani L, Fraysse F, Grobler AC, Muller J, Wake M, Olds T. Sleep: population epidemiology and concordance in Australian children aged 11–12 years and their parents. BMJ Open. 2019; 9(Suppl 3):127–35. Epub 20190704. https://doi.org/10.1136/bmjopen-2017-020895 PMID: 31273023; PubMed Central PMCID: PMC6624061.

59. Salman LA, Shulman R, Cohen JB. Obstructive Sleep Apnea, Hypertension, and Cardiovascular Risk: Epidemiology, Pathophysiology, and Management. Curr Cardiol Rep. 2020; 22(2):6. Epub 20200118. https://doi.org/10.1007/s11886-020-1257-y PMID: 31955254.

60. Su S, Li X, Xu Y, McCall WV, Wang X. Epidemiology of accelerometer-based sleep parameters in US school-aged children and adults: NHANES 2011–2014. Sci Rep. 2022; 12(1):7680. Epub 20220510. https://doi.org/10.1038/s41598-022-11848-8 PMID: 35538108; PubMed Central PMCID: PMC9090869.

61. Urtnasan E, Park JU, Joo EY, Lee KJ. Deep Convolutional Recurrent Model for Automatic Scoring Sleep Stages Based on Single-Lead ECG Signal. Diagnostics (Basel). 2022; 12(5). Epub 20220515. https://doi.org/10.3390/diagnostics12051235 PMID: 35626390; PubMed Central PMCID: PMC9140070.

62. Vidigal TA, Brasil EL, Ferreira MN, Mello-Fujita LL, Moreira GA, Drager LF, et al. Proposed management model for the use of telemonitoring of adherence to positive airway pressure equipment—position paper of the Brazilian Association of Sleep Medicine—ABMS. Sleep Sci. 2021; 14(Spec 1):31–40. https://doi.org/10.5935/1984-0063.20200086 PMID: 34917271; PubMed Central PMCID: PMC8663734.

63. Ota M, Maki Y, Xu LY, Makino T. Prolonging effects of Valeriana fauriei root extract on pentobarbital-induced sleep in caffeine-induced insomnia model mice and the pharmacokinetics of its active ingredients under conditions of glycerol fatty acid ester as emulsifiers. J Ethnopharmacol. 2022; 298:115625. Epub 20220813. https://doi.org/10.1016/j.jep.2022.115625 PMID: 35970315.

64. Bhatt SP, Guleria R, Vikram NK, Gupta AK. Non-alcoholic fatty liver disease is an independent risk factor for inflammation in obstructive sleep apnea syndrome in obese Asian Indians. Sleep Breath. 2019; 23(1):171–8. Epub 20180722. https://doi.org/10.1007/s11325-018-1678-7 PMID: 30032465.

65. Piekarski D, Sullivan EV, Pfefferbaum A, Zahr NM. Poor subjective sleep predicts compromised quality of life but not cognitive impairment in abstinent individuals with Alcohol Use Disorder. Alcohol. 2022; 103:37–43. Epub 20220721. https://doi.org/10.1016/j.alcohol.2022.07.001 PMID: 35870739.

66. Yang S, Guo X, Liu W, Li Y, Liu Y. Alcohol as an independent risk factor for obstructive sleep apnea. Ir J Med Sci. 2022; 191(3):1325–30. Epub 20210610. https://doi.org/10.1007/s11845-021-02671-7 PMID: 34110582; PubMed Central PMCID: PMC9135842.