The European Journal of Technical and Natural Sciences 2024, No 2
ISSN 2414-2352

PREMIER Publishing
ppublishing.org

Section 3. Physiology

DOI:10.29013/EJTNS-24-2-22-28

IDENTIFYING FACTORS RELATED TO SLEEP DISORDERS AMONG ADULTS IN NHIS 2022

Yiwen Zhou [1]

[1] Saint James School, MD

Cite: *Yiwen Zhou. (2024). Identifying Factors Related to Sleep Disorders Among Adults in Nhis 2022. European Journal of Technical and Natural Sciences 2024, No 2.* https://doi.org/10.29013/EJTNS-24-2-22-28

Abstract

Sleeping is a fundamental aspect of our daily lives. Sleep deprivation, such as dyssomnia, parasomnias, or insomnia, can have far-reaching implications. The lack of sleep can be due to plenty of reasons. Setting an application to predict people's potential risks of developing sleep disorders has become a pragmatic and intricate endeavor.

This paper discusses factors related to sleep disorders among adults using data from the National Health Interview Survey (NHIS) in 2022. Sleep deprivation, like insomnia, adversely affects mental and physical well-being. This study uses logistic regression along with regularization and cross-validation to analyze the data, considering variables like sex, medication usage, trouble staying asleep, feeling well-rested, and hours of sleep. The model performance is shown with a confusion matrix and a ROC curve analysis. This study also examines the correlation between variables and features' importance, highlighting variables like trouble staying asleep (SLPSTY_A) and feeling well-rested (SLPREST_A) as significant factors related to sleep disorders.

The conclusion is that variables such as sleep difficulties (SLPSTY_A) are significant, while waking up feeling well rested (SLPREST_A) is less critical. Overall, this article provides an overview of the research methods, findings, and factors related to adult sleep disorders.

Keywords: *Sleep disorder, NHIS, machine learning, logistic regression, correlation, ROC*

Introduction

Sleeping is a fundamental aspect of our daily lives. Two internal biological mechanisms govern sleep: circadian rhythm and sleep-wake homeostasis. The circadian rhythm is regulated by the biological clock and influenced by external factors such as ambient light, which triggers the release of hormones to regulate wakefulness (www.ninds.nih.gov/health-information/public-education/brain-basics/brain-basics-understanding-sleep./2023). Sleep-wake homeostasis is an alarm to signal the body when it is time to sleep and to determine the depth and intensity of sleep (URL: https://www.cdc.gov/nchs/nhis/quest_doc.htm/2022).

Every living organism, human, animal, and even plant needs to sleep. Beyond its physical benefits, sleep also provides significant mental advantages, reducing fatigue, increasing vitality, and aiding in healing.

Conversely, sleep deprivation, such as dyssomnia, parasomnias, or insomnia, can have far-reaching implications. These repercussions will involuntarily exacerbate lots of apprehensions and conjectures. The lack of sleep can be due to plenty of reasons. For instance, consistent hullabaloo, overwhelmed pressure, unhinged delusion, peevish temper, paranoid personality, or pensive mental health. Although there are some correspondent remedies to assist individuals in recovering from sleep disorders, most of the treatments are still tentative, equivocal, and inconsistent due to a limited understanding of the patient's unique circumstances. Setting an application to predict people's potential risks of developing sleep disorders has become a pragmatic and intricate endeavor.

The effects of sleep deprivation on N2-P3 component event-related potential waveforms associated with working memory were investigated through event-related potential (ERP) analysis (URL: https://www.cdc.gov/nchs/nhis/quest_doc.htm/2022). Sixteen healthy college students participated in working memory tasks and had their EEG data recorded before and after post-traumatic stress disorder. The study focused on the N2 and P3 parts of ERP related to working memory processes. The research emphasizes the adverse effects of sleep deprivation on working memory.

In particular, to fulfill this task, we pre-processed the dataset, built a logistic regression model, and tuned its hyper-parameters to find the most optimal parameters with superior predictive performance. Then the model was validated using different techniques and the coefficients were used to find the most important variables contributing to sleep disorder.

Method

1 Data

The data used in this study is from the 2022 National Health Interview Survey (NHIS), a population-based survey established by the CDC to monitor the prevalence of the population health condition in the United States and evaluate the effects of illness on individuals (Effect of Sleep Deprivation on the Working Memory-Related N2-P3 Components of the Event-Related Potential Waveform). The survey includes family composition, tobacco use, alcohol, and drug usage, as well as other dietary behaviors and physical activity questions. The data is collected by a face-to-face survey of random household adults by asking demographic and related questions.

The dependent variable is coded as 'SLPFLL_A,' indicating how often the interviewee has trouble falling asleep in the past 30 days. We recorded the variable into a binary variable by categorizing those who have difficulty falling asleep as positive samples and the others as negative samples. The list of independent variables used in this study is shown below.

Table 1. *List of independent variables used in this study*

Name	Description	Response
SEX_A	The person's sex at birth	1: Male 2: Female
SLPMED_A	The frequency that people take any medication to help them fall asleep or stay asleep. Include both prescribed and over-the-counter medications.	1: Adults that are 18 and above
SLPSTY_A	The frequency people have trouble staying asleep	1: Adults that are 18 and above
SLPREST_A	The frequency people wake up feeling well-rested	1: Adults that are 18 and above
SLPHOURS_A	On average, the hours of sleep people get in a 24-hour period	1: Adults that are 18 and above

2 Statistical Method

2.1 Pre-processing

As most machine learning algorithms cannot deal with missing values, all the data points with lost entries or invalid responses to the dependent variable are excluded from training and testing.

One of the first steps of preprocessing is the feature standardization. The goal is to transform different variables into similar scales so that the machine learning model can treat them equally during the training process. In the formula shown below, *avg(x)* is the variable's mean value, and *std(x)* is the variable's standard deviation. Then each value *x* of the feature is replaced by y_i calculated as:

$$y_i = \frac{x - avg(x)}{std(x)}.$$

Finally, the dataset is split into the train and test sets: the train set (70%) is to help the model learn and the test dataset (30%) is for model performance evaluation and validation.

2.2 Logistic Regression

Logistic regression model was used to fit the current data and to predict future data. It's widely used in different fields, including machine learning, statistics, and social sciences.

The logistic regression model can be expressed with the formula:

$$ln\left(\frac{y}{1-y}\right) = w_0 + w_1x_1 + \ldots + w_mx_m$$

In the logistic regression, each feature x_i has its specific weight w_i, where w_0 is the intercept while w_1 through w_m are the coefficients of the independent variables.

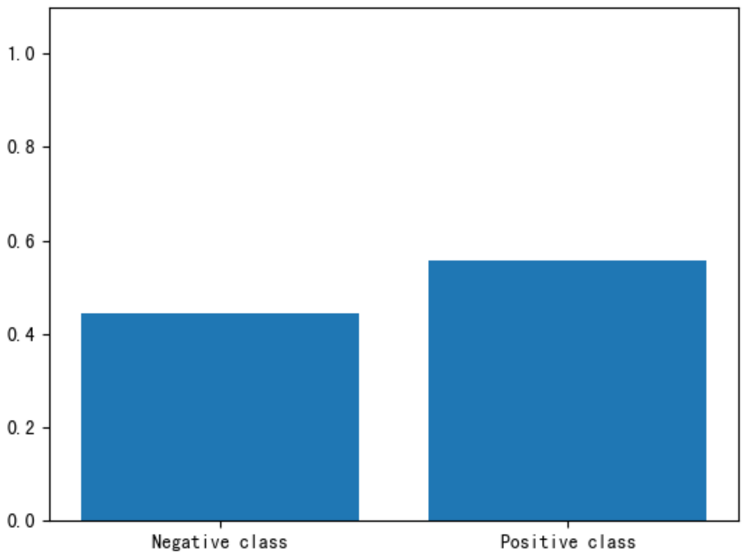
Our task is to find a set of parameters $w_0, \quad , w_m$ such that the loss function

$$l(y,u) = \| y - u \|_2^2$$

is minimized.

In addition, elastic-net regularization was applied to constrain model complexity and prevent the model from over-fitting with an L-1 ratio being 0.5. The optimal regularization strength is found through grid search and 5-fold cross-validation. Cross-validation is a statistical technique utilized to measure the performance of machine learning models. It is frequently employed in practical machine learning to assess and pick a suitable model for a specific task. This is due to its simplicity in comprehension and implementation, leading to skill evaluations (Brownlee, Jason. 2020). The Holdout method represents the most straightforward form of cross-validation. The dataset is divided into two sets: the training and test sets. Function approximators only construct functions from the training set. Subsequently, the task of the function approximator is to predict the output value for the data in the test set. The errors it generates are aggregated as described earlier, resulting in the mean absolute error for the test set. This metric is then used to evaluate the performance of the model (Brownlee, Jason. 2020). Meanwhile, as the distribution of the negative and positive classes is approximately equal, we gave the positive class the same weight as the negative class in the training process.

Figure 1. *Percentage of the negative class and positive class in the data set*



2.3 Model Validation

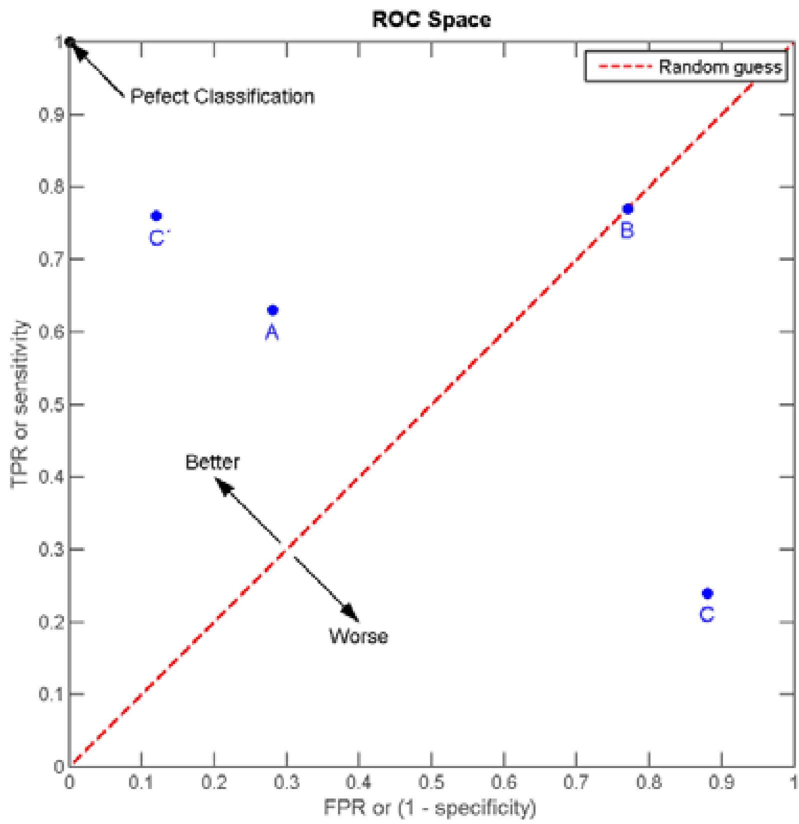
A confusion matrix is a table that visualizes the performance of an algorithm. Figure 2 is an example of the confusion matrix.

Figure 2. *Confusion matrix example*

		Predicted condition	
	Total population = P + N	Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

A receiver operating characteristic curve, or ROC curve, is a plot that visualizes a binary classifier's predicting power. A sample ROC plot is shown in Figure 3. Area Under Curve (AUC) can find the area under the ROC curve which can make model comparison easier.

Figure 3. *A sample ROC plot*

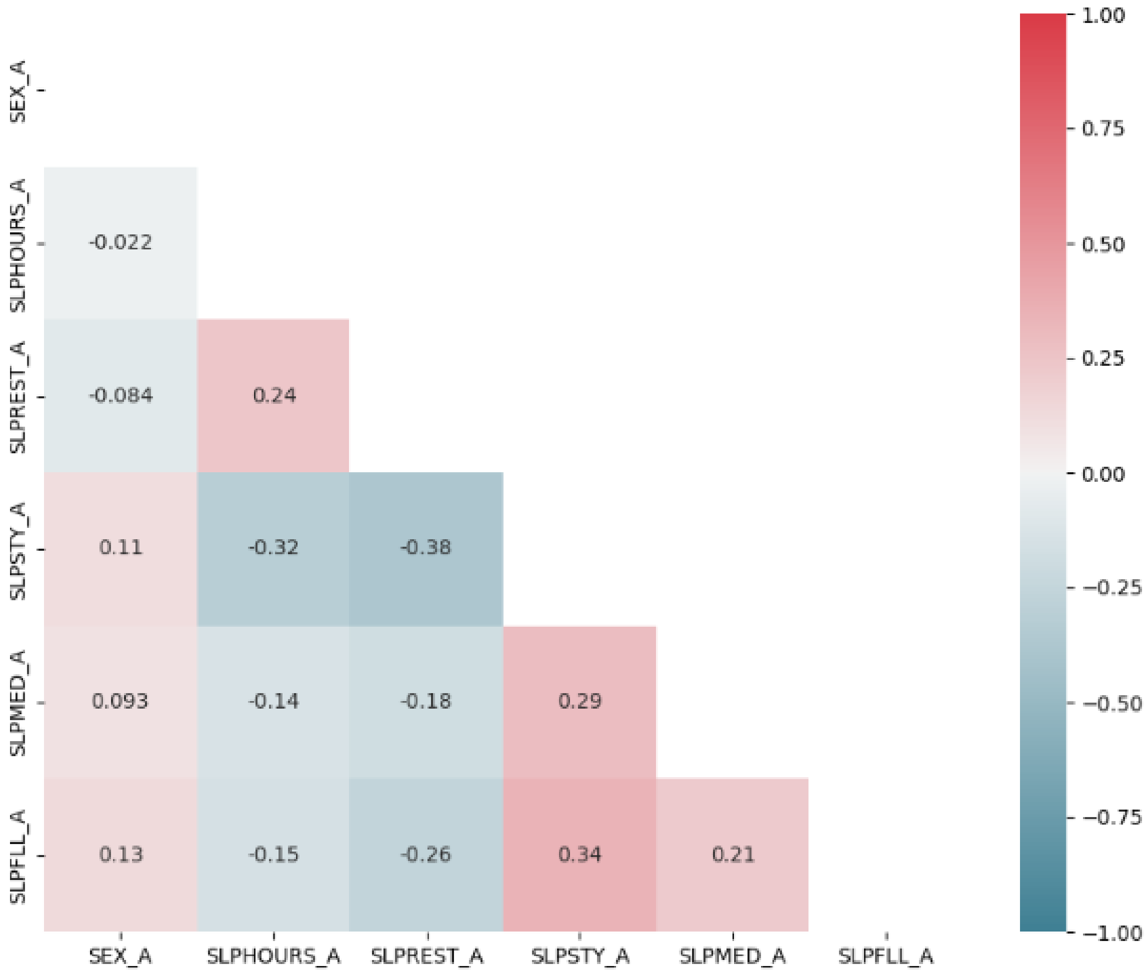


Results

3.1 Chorogram

Figure 4 is a chorogram that visualizes the correlation coefficients between variables. The dependent variable SLPFLL_A has a 0.34 positive correlation with the independent variable SLPSTY_A. It means insomnia is relatively incredibly related to having trouble staying asleep. The dependent variable SLPREST_A has a – 0.38 negative correlation with the dependent variable SLPSTY_A. It means waking up feeling well-rested or not is oppositely related to having trouble staying asleep.

Figure 4. *Correlation among variables*



3.2 Confusion matrix and ROC curve

Figure 5 is the confusion matrix of the trained logistic regression model. As shown in Figure 5, we can see that the logistic regression model has a relatively high (89.4%) true positive rate.

Figure 5. *Confusion matrix*

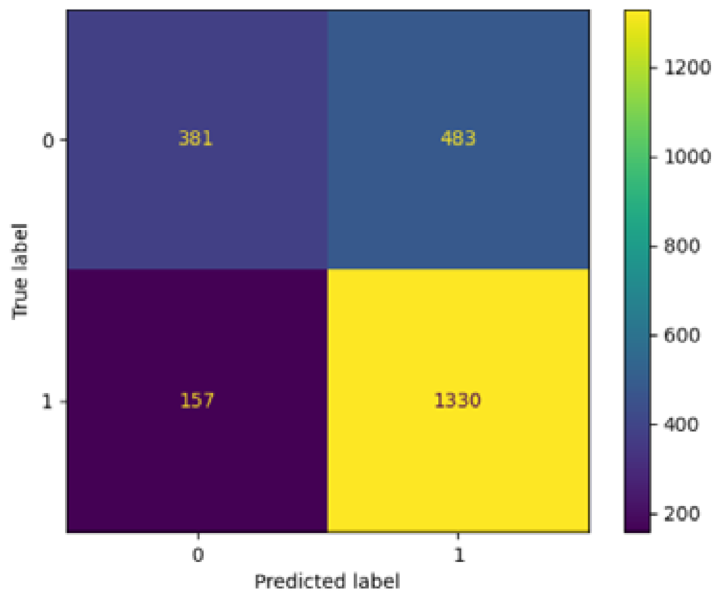


Figure 6 displays the ROC curve for the trained logistic regression model. It can be concluded that the model has results much better than random guessing (the diagonal in the ROC curve), and the AUROC score is 0.75.

Figure 6. *The ROC curve for the logistic regression model*

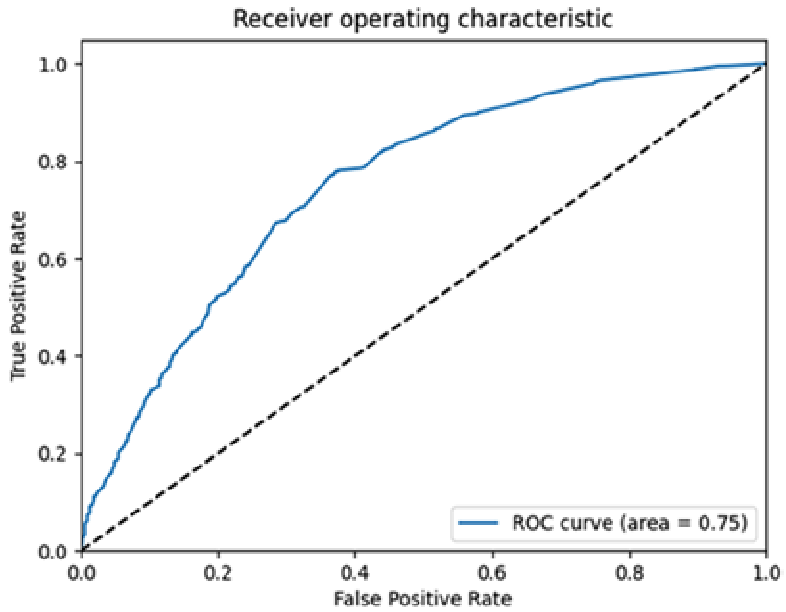
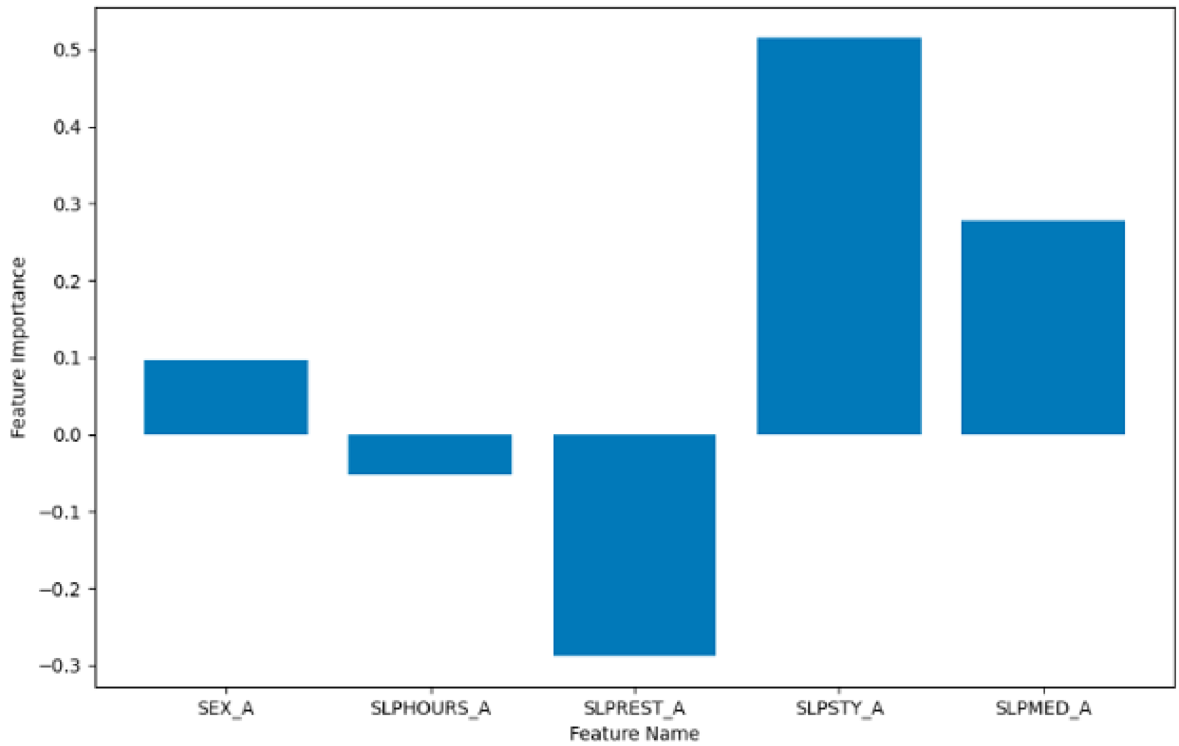


Figure 7. *Feature importance*



3.3 Feature Importance

On the feature importance graph (figure 7), SLPSTY_A has the biggest feature importance of 0.5. On the other hand, SLPREST_A shows minor feature importance of approximately –0.3.

Discussion

This paper discusses factors related to sleep disorders among adults using data from the National Health Interview Survey (NHIS) in 2022. Sleep deprivation, like insomnia, adversely affects mental and physical well-being. The study uses logistic regression to analyze the data, considering variables like sex, medication usage, trouble staying asleep, feeling well-rested, and hours of sleep. Logistic regression is applied with regularization and cross-validation. The results include a confusion matrix and ROC curve analysis. It shows the model's performance regarding true positives, false positives, and true negatives. The ROC curve indicates the model's ability to discriminate between cases and non-cases. The study also examines the correlation between variables and features' importance, highlighting variables like trouble staying asleep (SLPSTY_A) and feeling well-rested (SLPREST_A) as significant factors related to sleep disorders.

The study concluded by discussing the importance of features, emphasizing that variables such as sleep difficulties (SLPSTY_A) are significant, while waking up feeling well rested (SLPREST_A) is less critical. Overall, this article provides an overview of the research methods, findings, and factors related to adult sleep disorders.

References

"Brain Basics: Understanding Sleep." National Institute of Neurological Disorders and Stroke, www.ninds.nih.gov/health-information/public-education/brain-basics/brain-basics-understanding-sleep. Accessed: 17 July, 2023.

Centers for Disease Control and Prevention. (2022, March 20). NHIS – Questionnaires. URL: https://www.cdc.gov/nchs/nhis/quest_doc.htm.

Effect of Sleep Deprivation on the Working Memory-Related N2-P3 Components of the Event-Related Potential Waveform. URL: https://www.frontiersin.org/articles/10.3389/fnins.2020.00469/full

Brownlee, Jason. "A Gentle Introduction to K-Fold Cross-Validation". MachineLearningMastery.Com, 2 Aug. 2020. URL: https://machinelearningmastery.com/k-fold-cross-validation

submitted 02.02.2024;
accepted for publication 15.02.2024;
published 30.04.2024

Contact: yiq213@outlook.com