

UC Berkeley

Replication/Extension Papers 2025 - 2026

Title

Predicting Smoking Relapse Using Sleep Metrics: A Machine Learning Replication and Extension

Permalink

<https://escholarship.org/uc/item/933722dz>

Journal

Undergraduate Lab at Berkeley

Authors

Krishnan, Anika

Miller, Isaac

Kulkarnic, Soham

et al.

Publication Date

2025-04-01

Data Availability

Associated data will be made available after this publication is published.

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Predicting Smoking Relapse Using Sleep Metrics: A Machine Learning Replication and
Extension

Anika Krishnan^a

Isaac Miller^b, Soham Sudhir Kulkarni^c, Josh Jarek Calen^d, Darian Liu^e

Undergraduate Lab at Berkeley, Cognitive Science and Psychology Division

University of California, Berkeley

June 29, 2026

^a Research Lead, University of California, Berkeley

^b Research Student, University of California, Berkeley

^c Research Student, University of California, Berkeley

^d Research Student, University of California, Berkeley

^e Research Student, University of California, Berkeley

Abstract

From sleep quality to self-efficacy, smoking cessation attempts are affected by several factors. Determining which combination of attributes is most impactful at highlighting relapse risk can be difficult. A study from the University of Bristol attempted to do so by forming prediction models that would send messages based on ecological momentary assessment (EMA) and actigraphy data to prevent smoking relapse. We were able to partially validate the goals of this study by using a Random Forest machine learning model, utilizing leave-one-out cross-validation (LOOCV). Because the original temporal EMA and actigraphy data were not publicly available, we could not fully replicate the study's Chronic Sleep Approach. Instead, our replication relied on available interview responses and total sum scores of variables such as sleep and relapse risk. We then extended the project by applying additional machine learning models, including Support Vector Machines (SVM), Logistic Regression, Elastic Net, and XGBoost, to determine whether sleep-related and smoking-related variables could predict outcomes such as quit attempts, sleep duration, sleep quality, and perceived relapse risk. Our results suggested that the sleep-related variables, in particular, the Pittsburgh Sleep Quality Index component scores, were useful predictors of cessation. However, some models with very high accuracy appeared to be affected by uneven data distribution and class imbalances, weighted input columns, or overfitting. Overall, we were able to partially support the original study's analytical and computational approach, even without access to longitudinal data. These findings suggest that sleep metrics may be valuable in predicting smoking relapse risk, though future studies should use larger datasets with repeated daily or weekly measurements to make stronger conclusions.

Keywords: EMA, JITAI, smoking cessation, machine learning, sleep

Introduction

Relapse is common in cessation attempts of all addictive behaviors. Smoking, one of the most widespread forms of substance use, is no exception. Countless smokers attempt cessation but struggle to maintain abstinence. Because relapse can be influenced by several behavioral, psychological, and physiological factors, researchers have developed different methods of data collection to better understand when a smoker may be at risk for relapse.

A notable method is the Ecological Momentary Assessment (EMA), which allows researchers to collect information from participants about their current experiences, emotions, and behaviors at set intervals throughout the day. EMAs provide immense benefits as they can be used to characterize individuals to pinpoint how a factor in their mood influences their smoking decisions. For example, smokers who continue to view themselves as smokers, compared to those who have disassociated themselves with that label, may be more likely to relapse (Callaghan et al., 2021). Therefore, evaluating EMA responses can be useful for understanding a smoker's behavior during a quit attempt.

The paper we are replicating selected an EMA method that functioned through a stopwatch. This method was designed to reduce disruption to participants' daily routines, making it less burdensome than approaches that rely on smartphones or in-person communication (Haynes et al., 2018). As the University of Bristol paper notes, actigraphy data is additionally imperative. Actigraphy can track indicators of sleep quality, such as sleep efficiency, which may provide insight into a participant's risk of relapse. Prior research has indicated the relationship

between sleep and pain, highlighting the unidirectionality of the linkage. Specifically, while sleep quality directly impacts cessation success (i.e., the worse one sleeps, the harder it is to avoid relapse), one's relapse risk is indirectly affected by pain intensity — pain only worsens sleep (Thai et al., 2025). Sleep's vitality in preventing relapse also stems from withdrawal intensification that occurs due to poor sleep, due to disturbing one's circadian rhythm (Davey et al., 2017). Impairments to attention, mood, reward processes, and stress regulation are additional sleep deprivation problems. Smoking cravings may also heighten after a poor night's sleep.

The University of Bristol study applied machine learning methods to actigraphy and EMA data, as well as interview responses collected from each participant. Their goal was to develop algorithms that could predict when an individual was at risk of imminently smoking. Through the same stopwatch that conducted the EMAs, a just-in-time adaptive intervention (JITAI) message would be sent to the participant to discourage smoking. In this study, the JITAI was a message sent to subjects through the stopwatch at moments when the participant was predicted to be at higher risk.

Our utilization of machine learning originated from a desire to discover patterns within the data collected by the original researchers. However, we were unable to fully replicate the original study's conclusions based on participants' daily collected data because the temporal EMA and actigraphy data were not publicly available. Specifically, we were unable to reproduce their Chronic Sleep Approach (CSA) methodology as they did not provide chronological data on each participant throughout the experimental time period. Instead, we looked for trends within the available survey data to partially validate the original study's approach. For the replication

portion of our investigation, we selected features based on qualitative interview questions and the study's weighted scoring system, especially variables related to sleep and relapse risk. We used a Random Forest model with leave-one-out cross-validation to examine whether the available data could support the original study's emphasis on sleep and behavioral risk factors. For the extension portion, we evaluated how well additional machine learning models could predict selected sleep-related and smoking-related outcomes from the available variables. Because the University of Bristol study focused on how actigraphy-related data could speak to smokers' behavioral decisions during a cessation period, our target variables centered around participant sleep measurements, smoking behaviors, quit attempts, and relapse risk. Incorporating additional models allowed us to further assess the relationship between a smoker's behavioral profile and sleep measurements, while also recognizing the limitations of working without temporal participant data.

Methods

2.1 Dataset Cleaning and Preparation

Before beginning our analysis, we cleaned the dataset to ensure our models would not run into incomplete data. We first dropped null and invalid columns, then reviewed the remaining variables to determine which columns could be used as input features and target variables. After cleaning the dataset, we ran exploratory data analysis (EDA) to better understand the available variables and choose useful weighted data columns for modeling.

Because the original University of Bristol study relied on temporal ecological momentary assessment (EMA) and actigraphy data, we originally aimed to replicate their Chronic Sleep

Approach (CSA) modeling. This approach would have used metrics from each subject's previous night of sleep to predict whether the subject was at risk of relapse and whether a just-in-time adaptive intervention (JITAI) should be sent. However, the published dataset did not include the longitudinal data necessary to reconstruct day-to-day changes in participant behavior. Without access to this chronological data, we were limited in our ability to make strong conclusions about how daily variation in sleep and other variables affected smoking attitudes or relapse risk.

2.2 Replication Methods

For the replication portion of the study, we used a Random Forest model to examine whether the available survey and summary-score variables could partially support the original study's analytical approach. After completing EDA, we selected motivation to quit as the predicted variable and used the remaining columns as input variables. We focused on the variables related to perceived relapse risk, sleep measures, and interview responses because these variables were most closely connected to the original study's emphasis on smoking behavior and sleep-related relapse risk.

We used GridSearchCV to train and tune the Random Forest model. GridSearch helped identify the hyperparameter combination that best balanced model performance across accuracy, recall, and precision. The selected model used low values for the minimum samples per split and the minimum samples per leaf, suggesting that the Random Forest model needed to fit closely to the small dataset in order to maintain predictive performance. However, because this also increased the risk of overfitting, we used leave-one-out cross-validation (LOOCV) to evaluate the model more carefully. LOOCV allowed each participant to be used once as the test case while

the remaining participants were used for training. This helped us assess whether the model was able to produce consistent predictions across the dataset.

2.3 Extension Methods

Building upon our initial exploratory data analysis and Random Forest replication model, we expanded our methodology to include additional machine learning models. For the extension portion of the study, we used XGBoost, Support Vector Machine (SVM), Logistic Regression, and Elastic Net models to better understand the predictive weight of our features. This extension allowed us to examine whether the patterns found in the Random Forest model remained consistent across other modeling approaches.

For models attempting to predict the motivation to quit ladder score (MOT_QUIT_LADDER), we explicitly removed the “PRE” and “ACP” columns. These variables were removed because they served as expected outputs and were highly correlated to the target variable, so their inclusion could have caused collinearity. To evaluate whether our findings generalized beyond motivation to quit, we selected several behavioral and sleep-related target variables:

SB_Q4: Whether the participant engaged in a quit attempt within the last 12 months;

PSQI_Component_3_Score_SD: Participant sleep duration;

PSQI_Component_5_Score_SDIST: Participant sleep quality;

PSRQ_RISK_SCORE: Perceived risk total score.

For feature evaluation, we utilized Elastic Net to extract absolute coefficient values and generate top-feature importance rankings across different targets. Logistic Regression and XGBoost models were also evaluated using LOOCV so that the models could be tested as thoroughly as possible despite the small dataset size. Across these extension models, we compared performance using metrics such as accuracy, F1 score, and ROC-AUC when applicable. This allowed us to examine not only whether a model achieved high accuracy, but also whether its predictions were reliable.

Results

3.1 Exploratory Data Analysis

We carried out EDA within the initial replication stages of our experimentation. One notable result was a scatterplot illustrating the relationship between total perceived barriers to smoking cessation and predicted relapse risk, measured by PSRQ score (Fig. 1). Pearson's correlation coefficient of 0.525 indicates a moderate positive association between the two variables. As participants' barrier score rose, their perceived relapse risk scores also tended to increase.

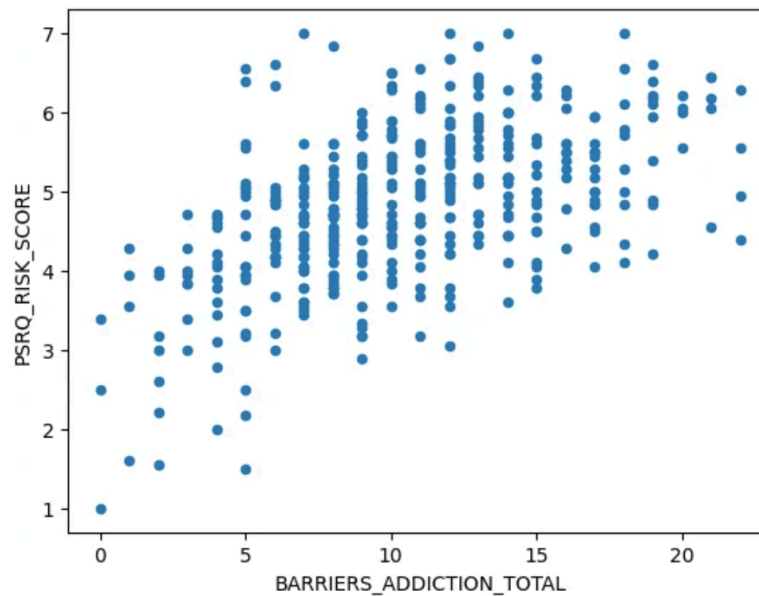


Fig. 1. Barriers to Preventing Relapse Mapped with Relapse Risk

3.2 Random Forest Replication Model

After completing EDA, we began predictive modeling utilizing an RF model. For the replication portion, we selected a participant's total perceived risk score and the duration of their last relapse event as input variables to predict motivation to quit. This model produced nearly 99% accuracy, though this result was interpreted cautiously because of the small dataset and possible overfitting. This initial result motivated us to further examine how sleep-related variables and smoking-related behaviors were associated with cessation outcomes.

3.3 Extension Models

The preliminary EDA and Random Forest modeling guided the extension portion of our project. In this stage, we narrowed our focus to sleep-related and smoking-related variables and

tested additional models, including Support Vector Machine (SVM), Logistic Regression, Elastic Net, and XGBoost. These models were used to evaluate whether the available variables could predict quit attempts, sleep duration, sleep quality, and perceived relapse risk.

Feature	Random Forest	SVM (Elastic Net)	LogReg (LOOCV)	XGBoost (LOOCV)	Feature Importance Distribution
SB 4	0.6747	0.6988	0.6626	0.7160	even
PSQI 3 Score	0.7469	0.6746	0.8107	0.9976	skewed
PSQI 5-Sum Score	0.5783	0.7711	0.6481	0.5922	even

Fig. 2. Top Features Tested With Most Accurate Variation of Classifier Models

3.4 Quit Attempts in the Last 12 Months

SB_Q4, which measured whether a participant engaged in a quit attempt within the last 12 months, performed well as a target variable across models. The SVM model achieved an accuracy of 69.88% and an F1 score of 0.6991, with a relatively even distribution of predictions. The Logistic Regression model with leave-one-out cross-validation (LOOCV) produced a slightly lower accuracy of 66.26%. The XGBoost model with LOOCV achieved the strongest performance for this target, reaching 71.60% accuracy and a ROC-AUC of 0.7567.

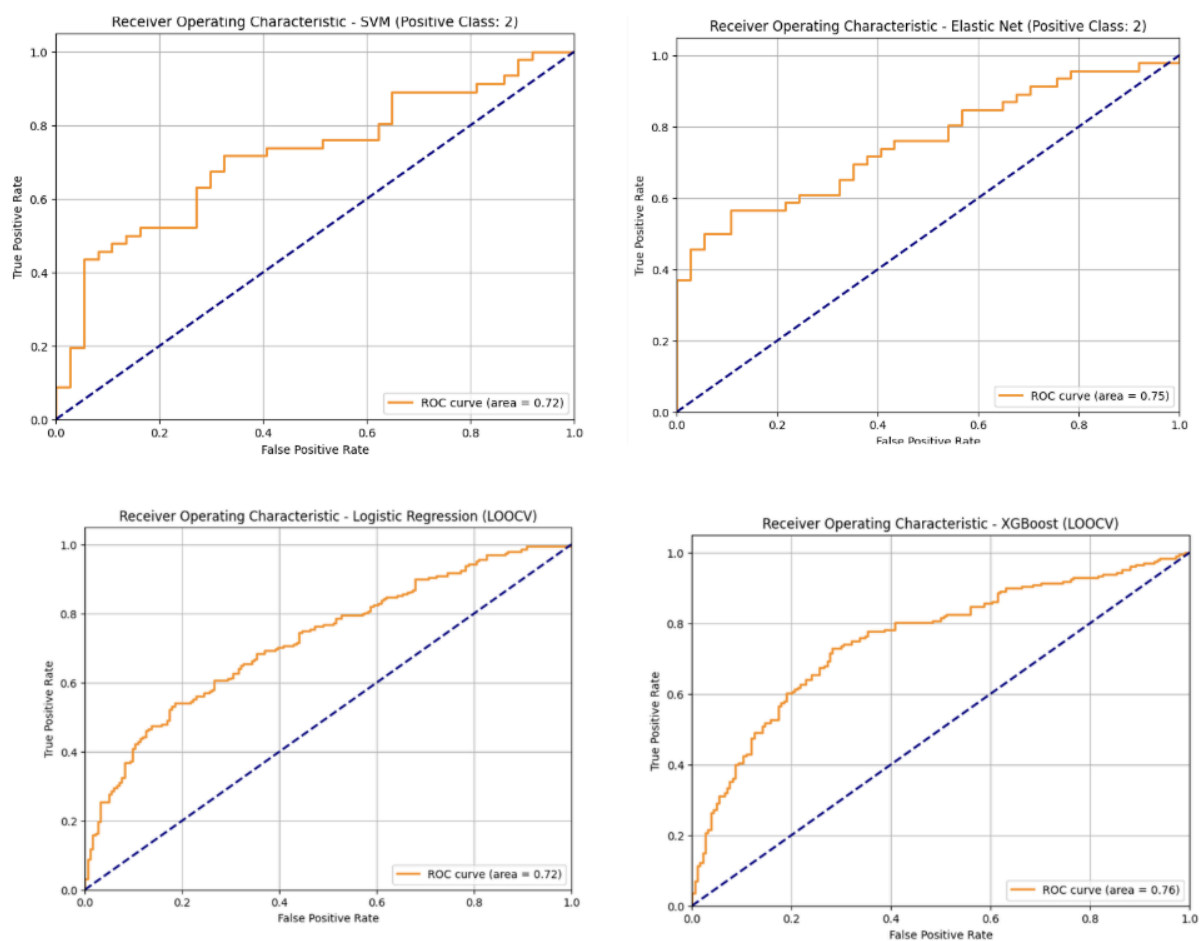


Fig. 3. ROC-AUC Curves for SVM, Elastic Net, Logistic Regression, and XGBoost utilizing SB_Q4

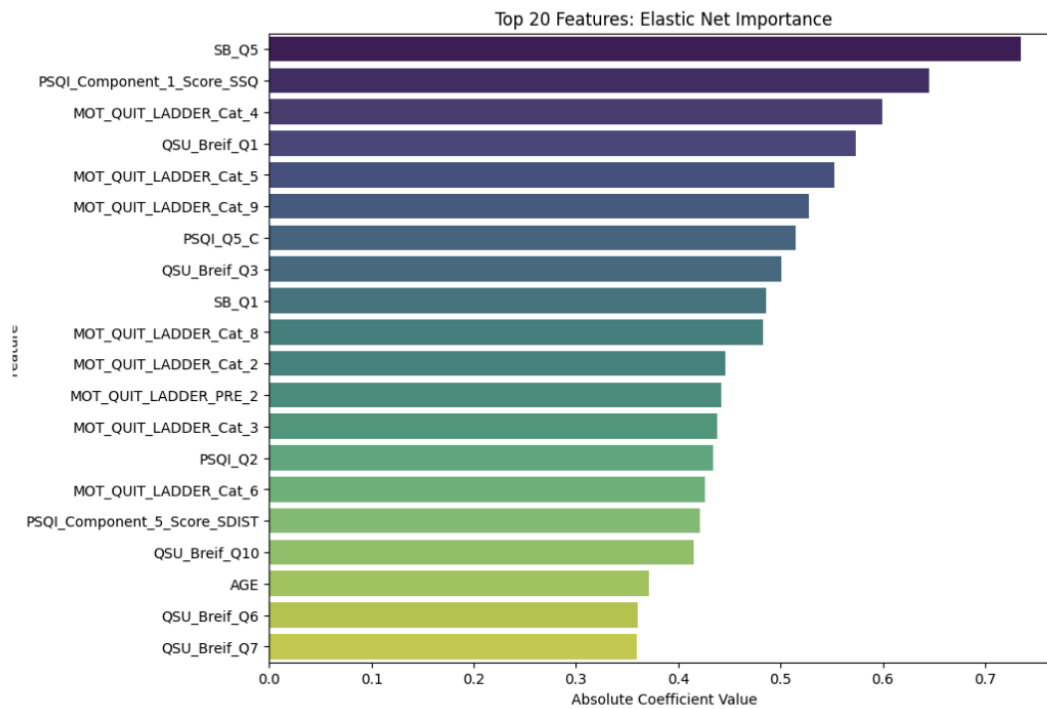


Fig. 4. Feature Importance for SB_Q4

3.5 Sleep Duration

Because sleep quality was a central focus of this extension, we also modeled Pittsburgh Sleep Quality Index (PSQI) component scores. For PSQI Component 3, which measured sleep duration, the SVM model produced a balanced accuracy of 67.47%. Logistic Regression with LOOCV performed especially well for this target, achieving 81.07% accuracy and a ROC-AUC of 0.8739. XGBoost reported 99.76% accuracy; however, the prediction distribution was uneven, suggesting that this result should be interpreted with caution.

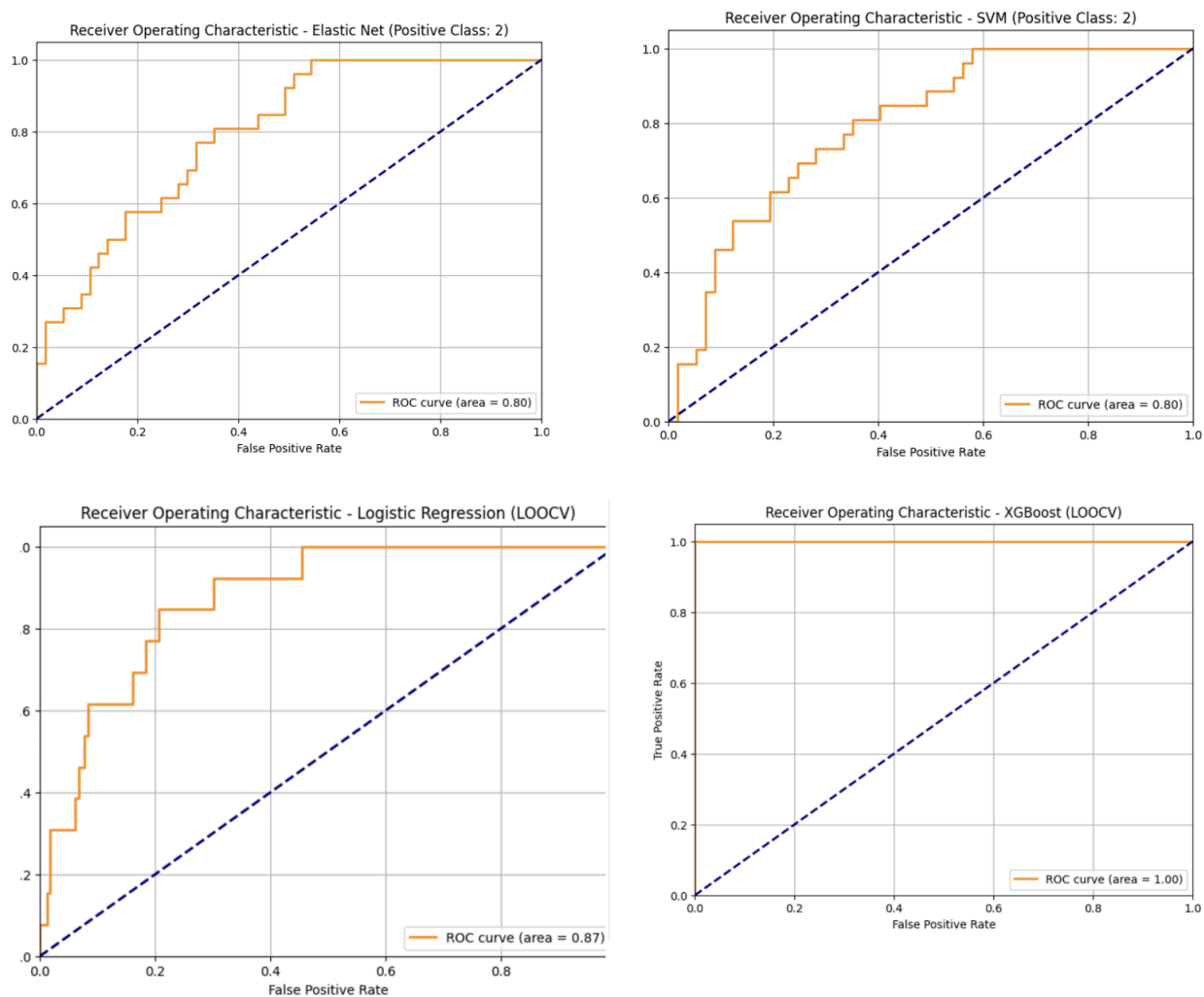


Fig. 5. ROC-AUC Curves for SVM, Elastic Net, Logistic Regression, and XGBoost utilizing PSQI Component 3

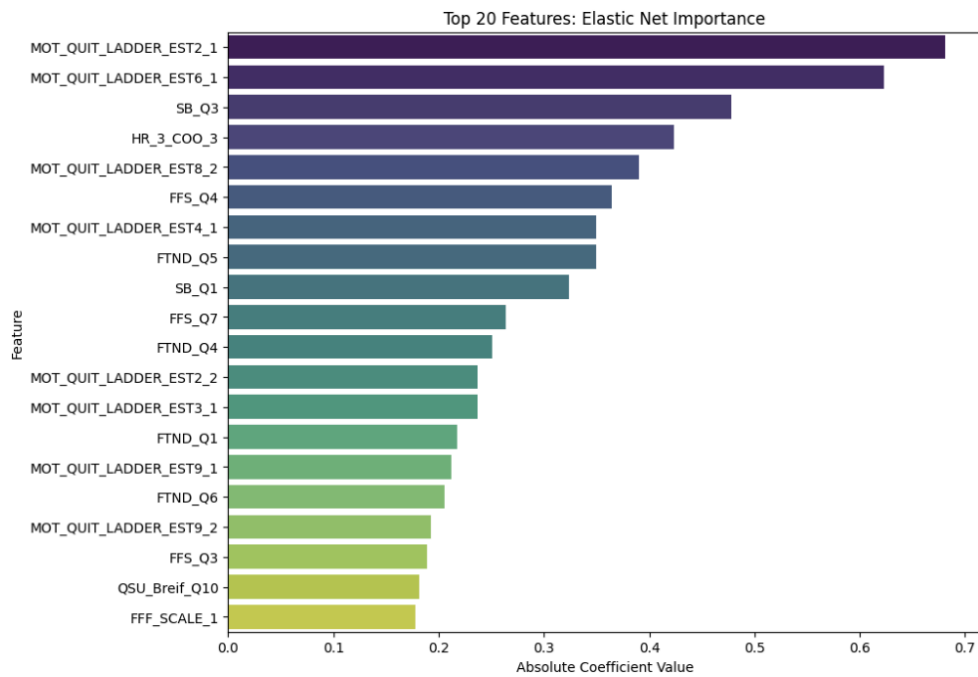


Fig. 6. Feature importance for PSQI Component 3

3.6 Sleep Quality

For PSQI Component 5, which measured sleep disturbance or sleep quality, the SVM model demonstrated an accuracy of 77.11%. For this target variable, the discretized Logistic Regression and XGBoost models underperformed relative to the SVM model, achieving 64.81% and 59.22% accuracy, respectively.

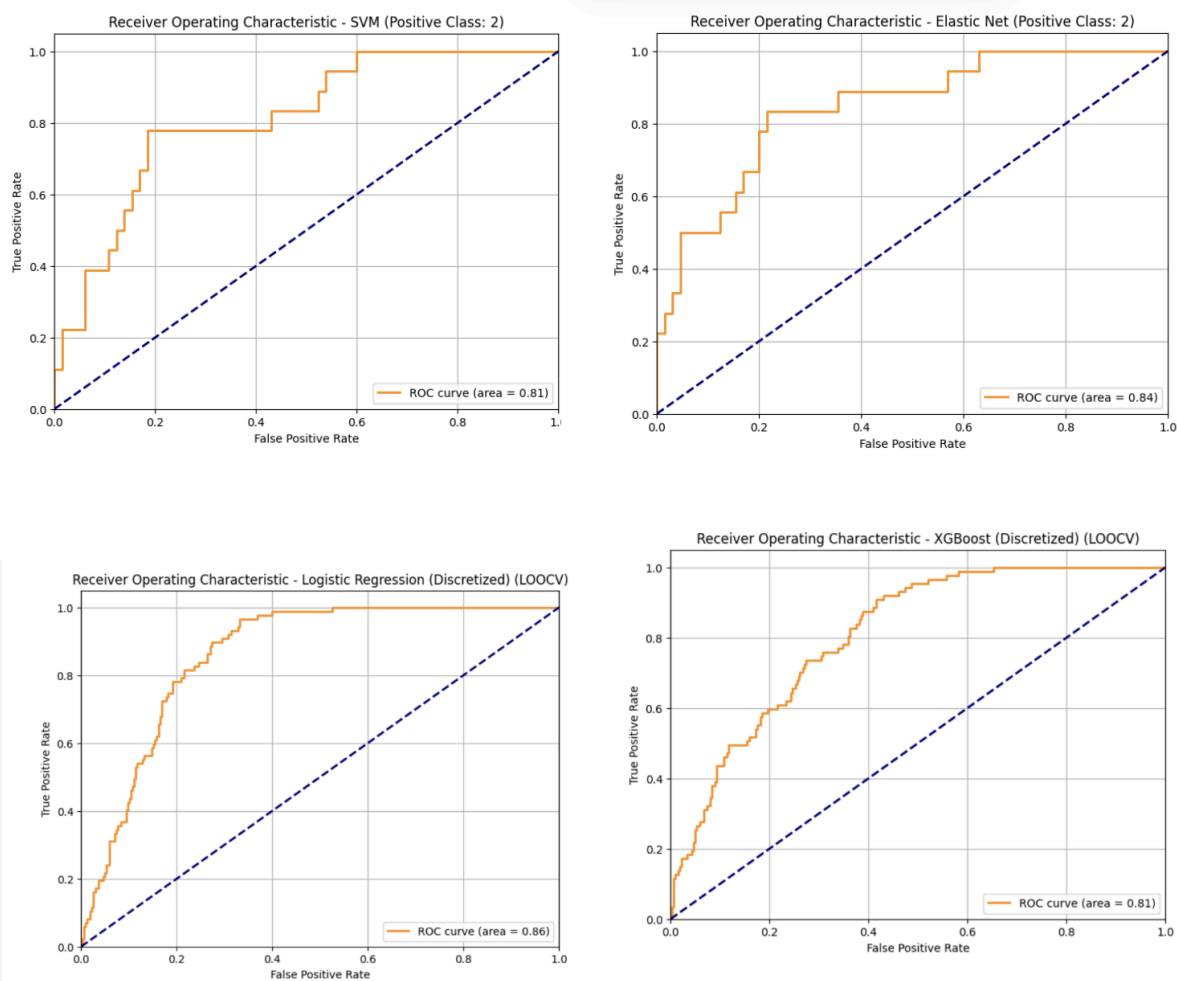


Fig. 7. ROC-AUC Curves for SVM, Elastic Net, Logistic Regression, and XGBoost utilizing PSQI Component 5

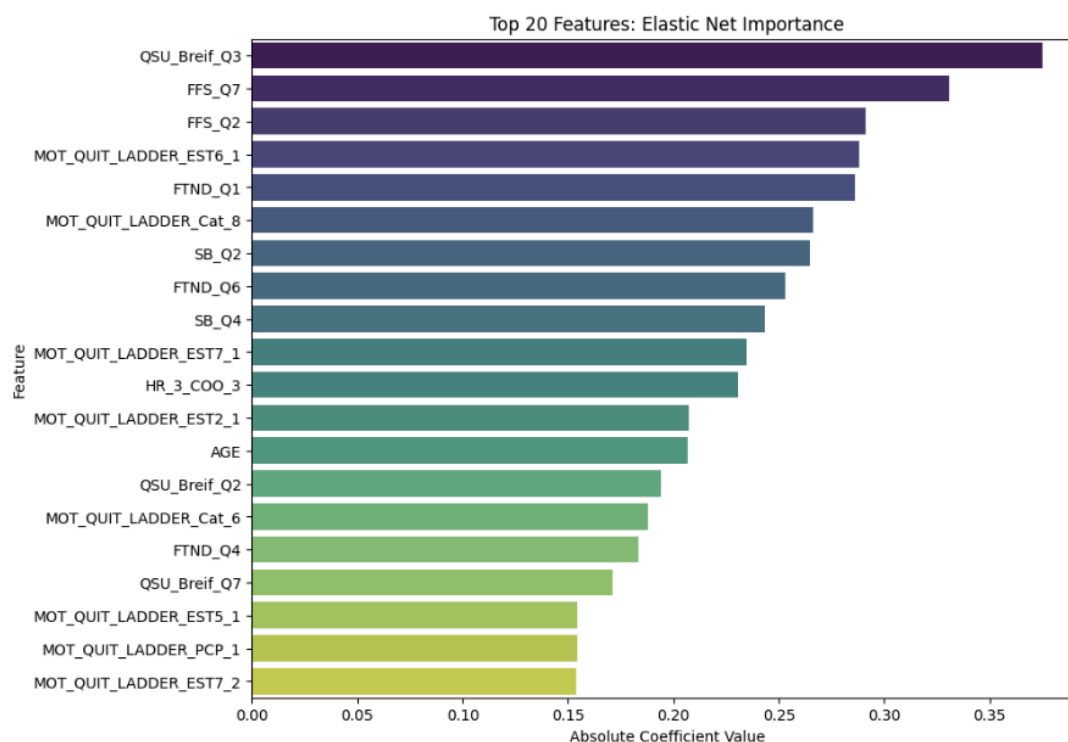


Fig. 8. Feature importance for PSQI Component 5

3.7 Perceived Relapse Risk

When modeling the PSRQ_RISK_SCORE, the SVM and Elastic Net models achieved high accuracies of 92.77% and 95.18%, respectively. However, the feature importance charts indicated that these models may have been affected by uneven data distribution and heavily weighted input columns. Because of this imbalance, PSRQ_RISK_SCORE was treated as a less reliable predictive target compared to SB_Q4 or the PSQI components.

Across the balanced Elastic Net models, variables related to overall sleep quality, such as PSQI_Component_1_Score_SSQ, and specific cessation survey items, such as SB_Q5, consistently appeared among the highest absolute coefficient rankings. These rankings indicated

that self-reported sleep disruption and smoking-related survey responses were among the more influential features in the extension models.

3.8 Motivation to Quit Score

Finally, we attempted to predict a stratified motivation to quit score using the MOT_QUIT_LADDER variable. Scores were separated into lower motivation scores, ranging from 1 to 5, and higher motivation scores, ranging from 6 to 10. The XGBoost model achieved 60% predictive accuracy with LOOCV when using only data from the FSS and PSQI measures. This result showed moderate predictive performance for estimating self-reported motivation to quit from the available sleep-related and survey-based variables.

Discussion

Overall, the models suggested that sleep-related variables tended to be relatively useful predictors of smoking-related behavior. While other variables, such as previous smoking habits and quitting motivation, were also explored, they were less consistent as predictors of smoking behavior. As a result, we chose to focus our main analysis on the variables and models that produced more balanced and interpretable results. Notably, while models based on perceived relapse risk score (PSRQ) variables produced higher accuracy than the sleep-related variables, these results appeared to be skewed by uneven data distribution and heavily weighted input columns. Because of this, the PSQI variables were treated as more reliable and consistent predictors of cessation-related outcomes than PSRQ alone.

These findings partially support the University of Bristol's approach of using sleep behavior as a predictor of smoking and cessation behavior. The original study emphasized the possibility that sleep metrics, collected through actigraphy and ecological momentary assessment (EMA), could help identify moments when a participant may be at higher risk of relapse. Our results could not fully replicate this approach because the publicly available data did not include the temporal participant-level data needed to reconstruct the original Chronic Sleep Approach (CSA). However, the fact that sleep-related summary variables still performed relatively well in several models suggests that sleep may remain an important factor in understanding smoking cessation outcomes.

The extension models also helped clarify which types of outcomes were more reliable to predict. SB_Q4, which measured whether participants had made a quit attempt in the last 12 months, produced relatively balanced results across multiple models. Similarly, PSQI Component 3 and PSQI Component 5 provided useful information about sleep duration and sleep quality. In contrast, PSRQ_RISK_SCORE produced very high model accuracies, but these results were less reliable because they appeared to depend on uneven distributions or dominant weighted features. This distinction is important because a model with higher accuracy is not necessarily stronger if its predictions are imbalanced or overly dependent on a small number of variables.

While we were able to partially validate the University of Bristol's analytical approach, several constraints limited our analysis and extension. Most importantly, the lack of time intervals in the available dataset limited our modeling options. Because the original JITAI system depended on identifying risk based on changing sleep and behavioral patterns over time, the

absence of chronological data prevented us from testing whether a participant's previous night of sleep could predict near-term smoking risk. This limitation is especially significant because the original study's intervention was designed to respond to momentary or daily changes, not only general participant-level characteristics.

Additionally, the relatively small dataset increased the risk of overfitting. Although leave-one-out cross-validation allowed us to test each participant individually, small datasets can still produce unstable model performance, especially when models are trained on many survey-based and weighted variables. Some of our highest reported accuracies, including the Random Forest replication model and certain XGBoost or PSRQ-based models, should therefore be interpreted cautiously. These results may reflect patterns in the available dataset rather than a model that would generalize well to a larger population.

Another limitation is that many of the variables were based on self-reported measures or total scores. While these variables are useful for understanding participant behavior and attitudes, they do not provide the same level of detail as repeated EMA responses or actigraphy measurements collected across time. Because of this, our study was better suited to identifying broad associations between sleep-related variables and smoking-related outcomes than predicting imminent relapse risk.

Future research should use larger datasets with repeated daily measurements of sleep, craving, mood, and smoking behavior. Access to temporal EMA and actigraphy data would allow researchers to more directly evaluate whether changes in sleep quality or sleep duration predict short-term relapse risk. Future studies should also compare model performance across balanced metrics, not accuracy alone, to avoid overstating results from imbalanced data. Overall, our

findings suggest that sleep-related variables may be valuable in smoking cessation prediction, but stronger conclusions require larger, more detailed, and time-based datasets.

References

- Callaghan, L., Yong, H.-H., Borland, R., Cummings, K. M., Hitchman, S. C., & Fong, G. T. (2021). What kind of smoking identity following quitting would elevate smokers' relapse risk? *Addictive Behaviors*, 112, Article 106654. <https://doi.org/10.1016/j.addbeh.2020.106654>
- Intille, S., Haynes, C., Maniar, D., Ponnada, A., & Manjourides, J. (2016). μ EMA: Microinteraction-based ecological momentary assessment (EMA) using a smartwatch. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 1124–1128). Association for Computing Machinery. <https://doi.org/10.1145/2971648.2971717>
- Matthews, J., McConville, R., Stone, C., Isotalus, H., Willis, K., Durant, C., Wilson, S., & Attwood, A. (2022, August 11). *Using sleep and physical activity metrics to predict smoking relapse risk factors to inform a just-in-time-adaptive intervention: A machine learning approach* (Study protocol, Version 1.0). University of Bristol, School of Psychological Science. <https://osf.io/9yr8n/>
- Patterson, F., Grandner, M. A., Malone, S. K., Rizzo, A., Davey, A., & Edwards, D. G. (2017). Sleep as a target for optimized response to smoking cessation treatment. *Nicotine & Tobacco Research*, 21(2), 139–148. <https://doi.org/10.1093/ntr/ntx236>
- Thai, J. M., Charles, J. J., Smit, T., Redmond, B. Y., Giordano, T. P., Ditre, J. W., Garey, L., & Zvolensky, M. J. (2025). The role of sleep difficulties in the relation between pain and smoking cessation processes. *Journal of Behavioral Medicine*, 48(4), 671–682. <https://doi.org/10.1007/s10865-025-00580-z>