Multi-Modal Sleep Data and Next-Day Affect: A Machine Learning Comparison of Key Factors

Bradley J. Wheeler[1], Daley R. Fraser[1], Rachel Witt[2], Brant P. Hasler[2], Meredith J. Wallace[2]

[1]School of Computing & Information, University of Pittsburgh, Pittsburgh, PA; [2]Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pa

INTRODUCTION

- Sleep research harnesses advanced multi-modal technologies, producing millions of data points per participant.
- Machine learning offers powerful methods for modeling and analyzing these complex datasets.
- Selecting the right algorithms involves important trade-offs that greatly influence the outcomes.
- In this study, we apply machine learning to determine which data modalities are the strongest predictors of next-day affect.
- **GOAL: Framework to evaluate benefits and limitations of prominent machine learning algorithms abilities to model multi-model data for data-driven sleep research**

METHODS

Characteristic	Train Data	Validation Data	Test Data
n	646	193	173
Participants	99	31	30
CART, number (%)	36 (36.4)	12 (38.7)	11 (36.7)
SLATE, number (%)	35 (35.4)	10 (33.3)	11 (36.7)
CARRS, number (%)	28 (28.2)	9 (29.0)	8 (26.6)
Biological Female, number (%)	62 (62.6)	19 (61.3)	19 (63.3)
Age (years), mean (stdev)	16.3 (1.2)	16.4 (1.4)	16.3 (1.2)
Rest duration (mins), mean (stdev)	448.8 (102.9)	442.7 (97.9)	463.7 (91.1)
Sleep onset latency (mins), mean (stdev)	12.3 (16.0)	18.7 (26.2)	16.3 (17.5)
Snooze time (mins), mean (stdev)	14.6 (17.2)	14.0 (14.8)	15.4 (16.2)
Sleep efficiency (%), mean (stdev)	84.9 (6.5)	84.3 (7.5)	84.0 (6.7)
Wake after sleep onset (mins), mean (stdev)	44.2 (25.3)	40.6 (21.5)	45.1 (22.2)
Number of wake bouts (count), mean (stdev)	23.1 (9.4)	22.5 (9.6)	23.9 (9.2)
Scored total sleep time (mins), mean (stdev)	396.8 (91.4)	392.6 (90.5)	407.2 (84.9)
Number of sleep bouts (count), mean (stdev)	23.5 (9.5)	22.8 (9.6)	24.3 (9.2)
Percent immobile (%), mean (stdev)	87.4 (4.8)	87.7 (4.8)	87.0 (4.3)
Fragmentation index (%), mean (stdev)	23.5 (11.3)	22.7 (11.4)	24.9 (10.9)

Table 1. Characteristics of the training, validation, and testing datasets employed in this study. Datasets were stratified on outcome, biological sex, age, and study.

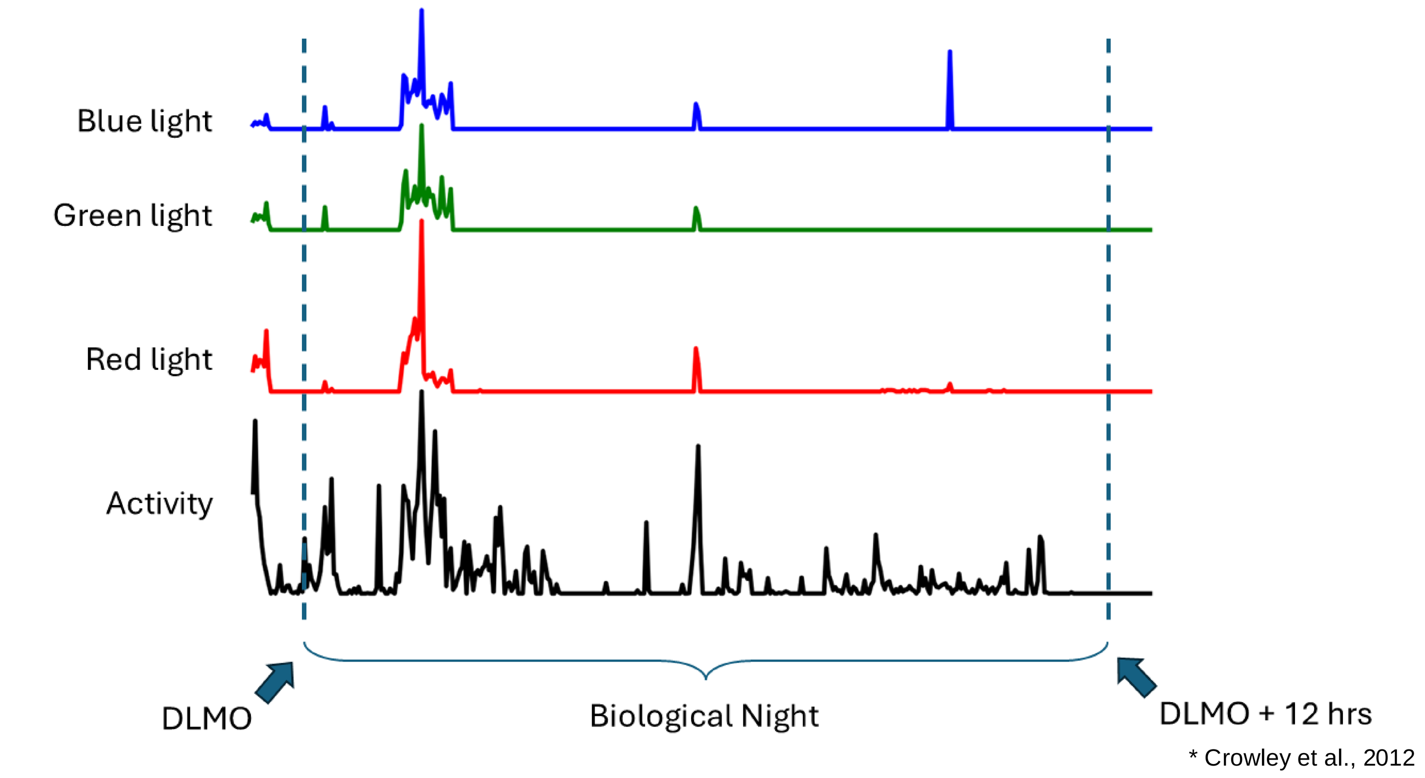

Figure 1. Source of the data for our experiment. Light and activity data between biological night was used engineered variable and raw data variable types, defined from the start of dim light melatonin onset.

$$outcome = \begin{cases} \textbf{If: } \text{mean pos affect for given day} >= \text{mean pos affect for study duration} \\ \quad \textbf{Then } 1 \\ \textbf{Else: } 0 \end{cases}$$

Figure 2. The outcome to predict is a binary indicator of whether a participant's overall positive affect for a given day is greater than their overall positive affect for the duration of the study, as recorded using the PANAS short form questionnaire.

Algorithm Type	Algorithms
Linear	Logistic Regression Lasso Ridge Elastic Net
Non-Linear	Support Vector Machine Random Forest Decision Tree Gradient Boosting Naïve Bayes K Nearest Neighbors
Deep Learning	Multilayer Perceptron Long Short-Term Memory

Table 3. The types of machine learning algorithms employed in modeling the data to predict the outcome.

Evaluation Pipeline:

For each algorithm in Table 3:

1. Use Table 1 train/val data to construct a model
2. Use Table 1 test data to compute model metrics
3. For each feature in Figure 1:
   a. Use Table 1 train/val data without select feature to construct a model
   b. Use Table 1 test data without select feature to compute model metrics
   c. Compute step 2 metrics – step 3b metrics to determine influence of select feature on outcome

RESULTS

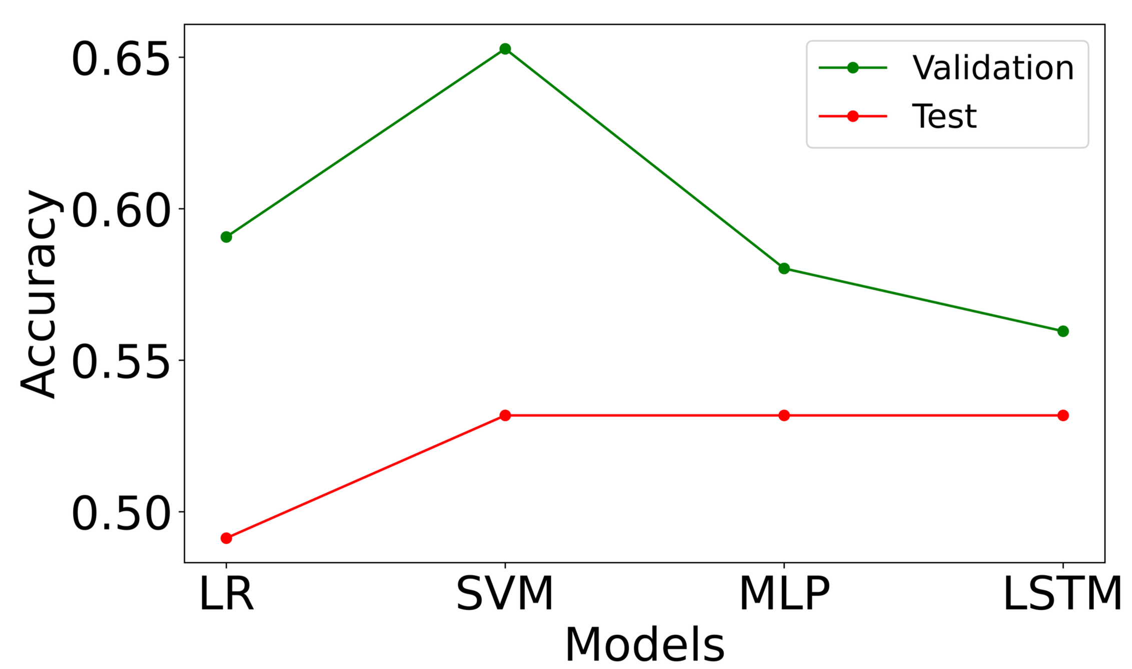

Figure 3. Accuracies on the validation and testing datasets for the machine learning algorithms selected for analysis

Algorithm	Accuracy	AUROC	Sensitivity	Specificity	Specificity – Sensitivity
Logistic Regression	0.4913	0.5362	0.2069	0.7791	0.5722
SVM	0.5318	0.5290	0.2414	0.8256	0.5842
MLP	0.5318	0.5277	0.3678	0.6977	0.3299
LSTM	0.5318	0.5309	0.6897	0.3721	-0.3176

Table 4. All metrics on the testing data dataset for the algorithms selected for the analysis

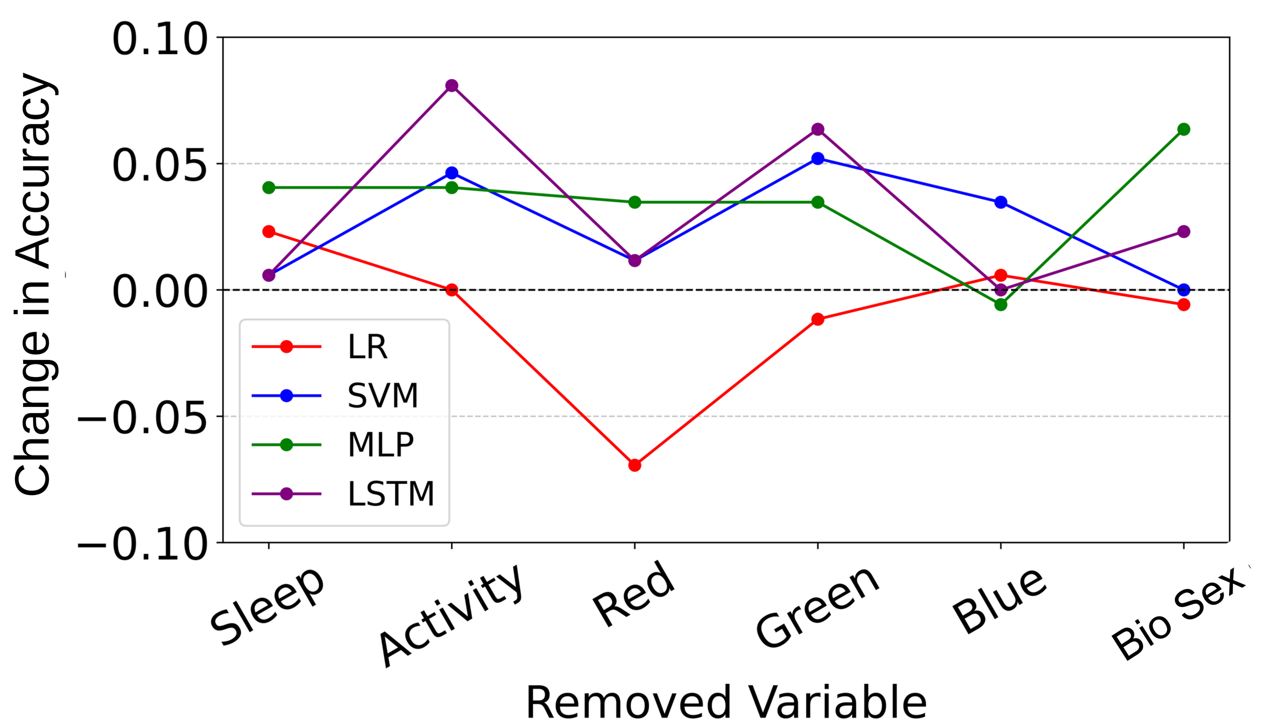

Figure 5. Difference in accuracy on the testing dataset when the algorithms use all the data and when one of the variable domains is removed. A higher delta indicates the removed variable is more informative in the model the machine learning algorithm produces.

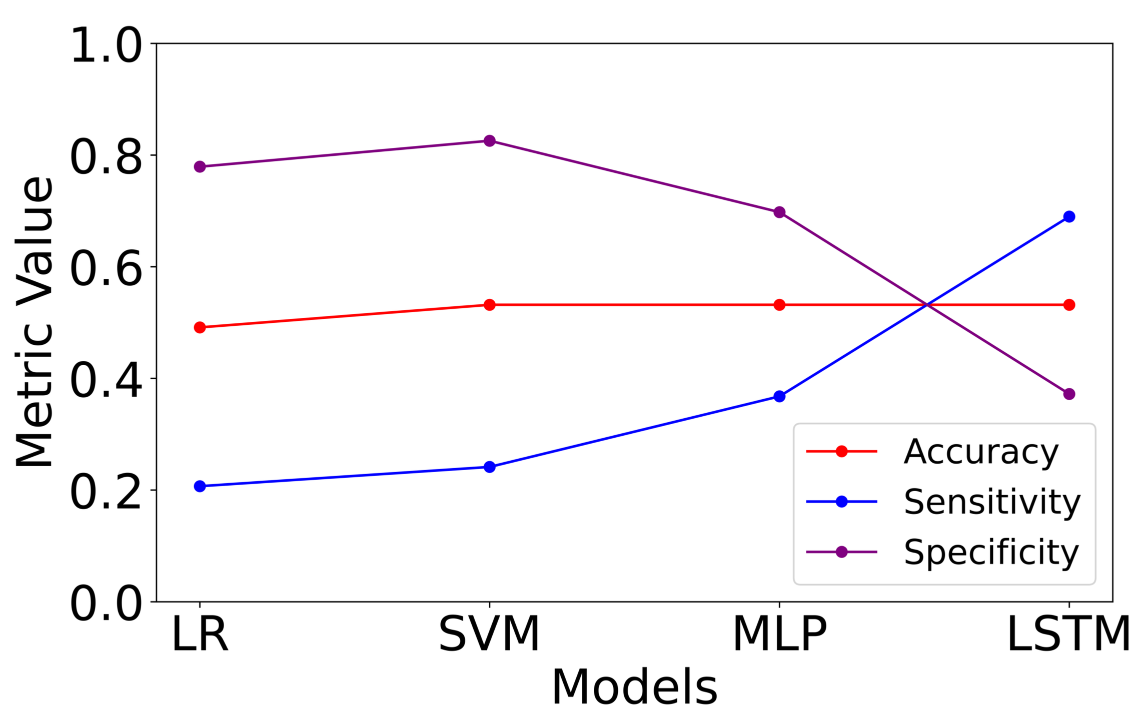

Figure 4. Accuracy, sensitivity, and specificity on the testing datasets for the machine learning algorithms selected for analysis

Variable	Mean All	STD All	Mean High	STD High	Top
Sleep	0.0188	**0.0166**	0.0174	0.0200	0.0058
Activity	**0.0419**	0.0332	**0.0559**	0.0218	**0.0809**
Red	-0.0029	0.0457	0.0193	**0.0133**	0.0116
Green	**0.0347**	0.0331	**0.0501**	**0.0145**	**0.0636**
Blue	0.0087	**0.0180**	0.0096	0.0219	0.0000
Bio Sex	0.0202	0.0315	0.0289	0.0322	0.0231

Table 6. The mean and standard deviation of the deltas in Figure 3 across algorithms for all four algorithms, the high performing algorithm (SVM, MLP, LSTM), and the top performing algorithm (LSTM). A higher mean indicates a more predictive variable domain. A lower standard deviation indicates better agreement between algorithms.

Summary

- Evaluating machine learning algorithms often requires multiple metrics to determine the best performer.
- Different research objectives can favor different algorithms, even when analyzing the same dataset.
- Our findings suggest that activity and green light are the strongest predictors of next-day affect; however, low and conflicting results across algorithms prevent definitive conclusions.
- Some data show weak or no associations with the outcome, regardless of the algorithm used. It is important to disclose this.
- There is an inverse relationship between algorithms that are more interpretable and computationally efficient, versus those that are more precise but require greater technical expertise to implement.

Funding Sources

- CARRS Pilot Study Grant (PI: Bradley J. Wheeler)
- R01 AA025626-05 (PI: Brant P. Hasler)
- R01 DA044143 (PI: Brant P. Hasler)
- RF1 AG056331-04 (PI: Meredith L. Wallce)

* Crowley SJ, Acebo C, Carskadon MA. Human puberty: salivary melatonin profiles in constant conditions. Developmental psychobiology. 2012 May;54(4):468-73.

Framework

Algo	Train Time	Hardware	Precision	Model Interpretability	Coding Skill
Linear	Secs	CPU	Low	High	Low
Non-Linear	Secs	CPU	Mod	Moderate	Low
MLP	Mins	CPU	High	Low	Mod
LSTM	Tens Mins – Hrs	GPU	Very High	Low	High