Digital Pulse

DETECTING PROBLEMATIC INTERNET USE

Ankit Kumar
Mangesh Singh
Siddharth Sahu

METHODS OF EVALUATING PIU

Internet Addiction Test

- Description: The IAT is a widely used tool that evaluates the level of internet addiction through 20 items rated on a five-point Likert scale, focusing on behaviors and symptoms associated with excessive internet use.

- Target Group: Adolescents and adults.
- Scoring: Total scores range from 0 to 100, with scores of 40 or above indicating problematic internet use.

Questionnaire-Children with Difficulties

- Description: This tool assesses daily difficulties experienced by children and adolescents, particularly in relation to PIU, examining how these difficulties correlate with internet use behaviors.

- Target Group: Children and adolescents.
- Scoring: The scoring method varies based on the specific items evaluated; typically involves summing scores across relevant dimensions.

Compulsive Internet Use Scale

- Description: The CIUS is a 14-item scale designed to assess compulsive internet use behaviors, featuring both self-report and parent-report versions for comprehensive data collection.

- Target Group: Children aged 8 to 10 years.
- Scoring: Items are rated on a Likert scale; total scores indicate levels of compulsive use.

PCIAT QUESTIONS

1. How often does your child disobey time limits you set for online use?
2. How often does your child neglect household chores to spend more time online?
3. How often does your child prefer to spend time online rather than with the rest of your family?
4. How often does your child form new relationships with fellow online users?
5. How often do you complain about the amount of time your child spends online?
6. How often do your child's grades suffer because of the amount of time he or she spends online?
7. How often does your child check his or her e-mail before doing something else?
8. How often does your child seem withdrawn from others since discovering the Internet?
9. How often does your child become defensive or secretive when asked what he or she does online?
10. How often have you caught your child sneaking online against your wishes?
11. How often does your child spend time along in his or her room playing on the computer?
12. How often does your child receive strange phone calls from new ""online"" friends?
13. How often does your child snap, yell, or act annoyed if bothered while online?
14. How often does your child seem more tired and fatigued than he or she did before the Internet came along?
15. How often does your child seem preoccupied with being back online when off-line?
16. How often does your child throw tantrums with your interference about how long he or she spends online?
17. How often does your child choose to spend time online rather than doing once enjoyed hobbies and/or outside interests?
18. How often does your child become angry or belligerent when your place time limits on how much time he or shes is allowed to spend online?
19. How often does your child choose to spend more time online than going out with friends?
20. How often does your child feel depressed, moody, or nervous when off-line which seems to go away once back online?

RELATION BETWEEN SII AND PIU

The cumulative PCIAT score is classified into four SII classes based on the score range it falls within.

Score Range	Level
0-30	Normal
31-49	Mild
50-79	Moderate
80-100	Severe

PROBLEM STATEMENT

Current assessments for problematic internet use are complex and require professional expertise.

Associations of Physical Activity, Screen Time with Depression, Anxiety and Sleep Quality among Chinese College Freshmen

Qi Feng, Qing-le Zhang, Yue Du, Yong-ling Ye, Qi-qiang He

Published: June 25, 2014 • https://doi.org/10.1371/journal.pone.0100914

Article | Authors | Metrics

Abstract
Introduction
Methods
Results
Discussion
Supporting Information
Author Contributions
References

Reader Comments
Figures

Abstract

Objectives

To investigate the independent and interactive associations of physical activity (PA) and screen time (ST) with depression, anxiety and sleep quality among Chinese college students.

Methods

A cross-sectional study was conducted in Wuhan University, China from November to December 2011. The students reported their PA, ST and socio-economic characteristics using self-administered questionnaires. Sleep quality was measured by the Pittsburgh Sleep Quality Index (PSQI). Depression and anxiety were assessed using the Self-rating Depression Scale (SDS) and Self-rating Anxiety Scale (SAS), respectively. Multivariate logistic regression models were used to estimate the odds ratios (ORs) and 95% confidence intervals (CIs) of the independent and interactive relationships of PA and ST with depression, anxiety and sleep quality.

Results

A total of 1106 freshmen (471 females and 635 males) aged 18.9±0.9 years were included in the study. After adjustment for potential confounders, high PA and low ST were independently associated with significantly lower risks for poor sleep quality (OR: 0.48, 95% CI: 0.30–0.78) and depression (OR: 0.67, 95%CI: 0.44–0.89), respectively. An interactive inverse association was observed for combined effects of PA and low ST on depression (OR: 0.62, 95%CI: 0.40–0.92) and sleep quality (OR: 0.51, 95%CI: 0.27–0.91). No statistically significant associations were found between PA, ST and anxiety among the participants.

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0100914

RESEARCH ARTICLE Open Access

Problematic internet use in children and adolescents: associations with psychiatric disorders and impairment

Anita Restrepo[1], Tohar Scheininger[1], Jon Clucas[2], Lindsay Alexander[1], Giovanni A. Salum[3], Kathy Georgiades[4], Diana Paksarian[5], Kathleen R. Merikangas[5] and Michael P. Milham[1,6,7,8*]

Abstract

Background: Problematic internet use (PIU) is an increasingly worrisome issue, as youth population studies are establishing links with internalizing and externalizing problems. There is a need for a better understanding of psychiatric diagnostic profiles associated with this issue, as well as its unique contributions to impairment. Here, we leveraged the ongoing, large-scale Child Mind Institute Healthy Brain Network, a transdiagnostic self-referred, community sample of children and adolescents (ages 5–21), to examine the associations between PIU and

Results: PIU was positively associated with depressive disorders (SR: aOR = 2.43, CI: 1.22–4.74, $p = .01$; PR: aOR = 2.56, CI: 1.31–5.05, p = .01), the combined presentation of ADHD (SR: aOR = 1.91, CI: 1.14–3.22, p = .01; PR: n.s.), Autism Spectrum Disorder (SR: n.s.; PR: aOR = 2.24, CI: 1.34–3.73, $p < .001$), greater levels of impairment (SR: Standardized Beta = 4.63, CI: 3.06–6.20, p < .001; PR: Standardized Beta = 5.05, CI: 3.67–6.42, p < .001) and increased sleep disturbances (SR: Standardized Beta = 3.15, CI: 0.71–5.59, p = .01; PR: Standardized Beta = 3.55, CI: 1.34–5.75, p < .001), even when accounting for demographic covariates and psychiatric comorbidity.

recommendations on internet use in U.S. youth.

Keywords: Internet addiction, Pediatric, Depression, ADHD, ASD, Impairment

https://link.springer.com/article/10.1186/s12888-020-02640-x

DATASET INFO

- **HBN Dataset**: Contains clinical screenings and research data, primarily physiological in nature.
- **Fitness Dataset:** Provides daily activity data from fitness bands (actigraphy) for each student, collected over 30 days to track physical behavior patterns.
- **Target**: The target variable (Sii) represents scores derived from the Parent-Child Internet Addiction Test (PCIAT), used to assess internet addiction levels.

Healthy Brain Networks Data

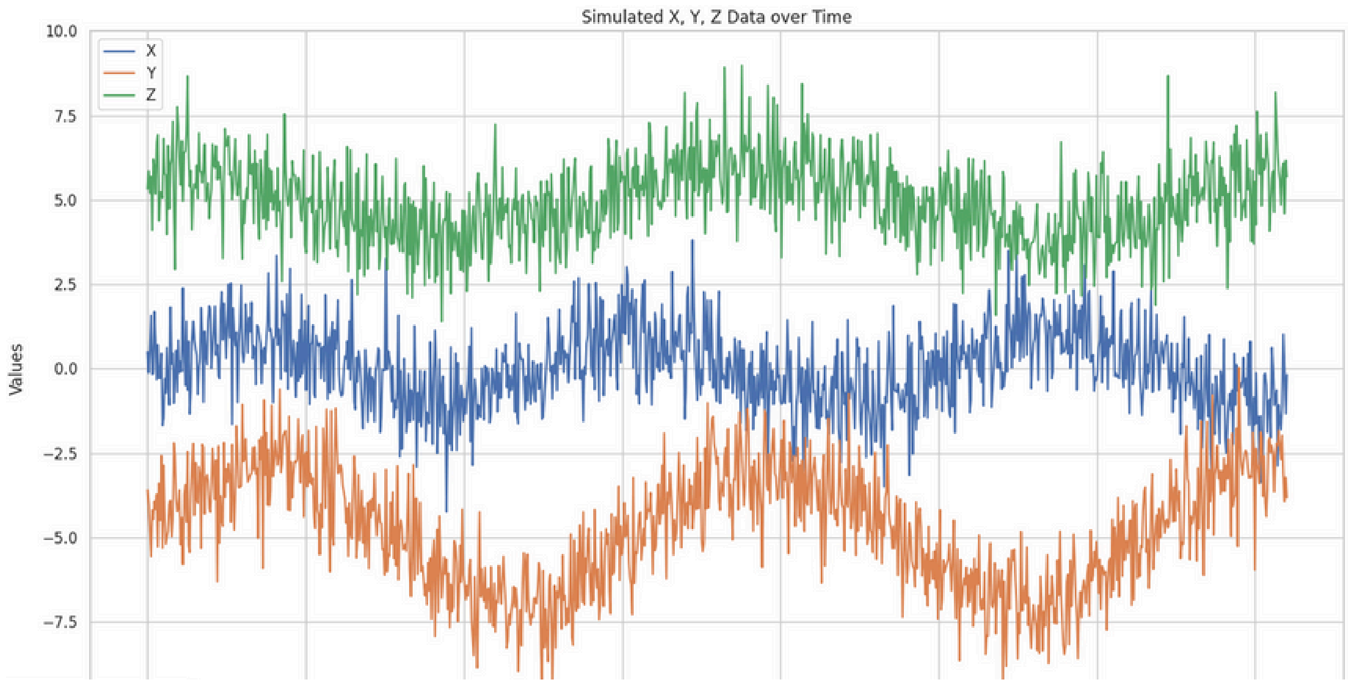
Actigraphy Files (Fitness Band Data)

https://link.springer.com/article/10.1186/s12888-020-02640-x

DATASET INFO

CGAS-Season	CGAS-CGAS_Score	Physical-Season	Physical-BMI	Physical-Height	Physical-Weight	...	PAQ_C-PAQ_C_Total	SDS-Season	SDS-SDS_Total_Raw	SDS-SDS_Total_T	PreInt_EduHx-Season	PreInt_EduHx-computerinternet_hoursday	sii
Winter	51.0	Fall	16.877316	46.00	50.80	...	2.1360	Spring	36.4	51.8	Fall	3.0	2.0
Spring	68.8	Fall	14.035590	48.00	46.00	...	2.3400	Fall	46.0	64.0	Summer	0.0	0.0
Fall	71.0	Fall	16.648696	56.50	75.60	...	2.1700	Fall	38.0	54.0	Summer	2.0	0.0
Fall	71.0	Summer	18.292347	56.00	81.60	...	2.4510	Summer	31.0	45.0	Winter	0.0	1.0
Summer	62.2	Spring	20.630858	63.36	164.56	...	1.9504	Spring	41.2	57.6	Spring	1.8	0.0
Winter	50.0	Summer	22.279952	59.50	112.20	...	4.1100	Summer	40.0	56.0	Spring	0.0	1.0
Spring	67.6	Fall	19.660760	55.00	84.60	...	3.6700	Winter	27.0	40.0	Fall	3.0	0.0
Spring	63.0	Fall	16.861286	59.25	84.20	...	1.2700	Spring	38.6	52.4	Fall	2.0	0.0
Spring	58.2	Spring	25.250308	69.40	168.56	...	2.6040	Spring	39.0	55.0	Summer	2.0	0.0
Summer	67.0	Spring	25.169036	63.65	158.96	...	1.9504	Spring	49.0	63.2	Spring	2.2	1.0

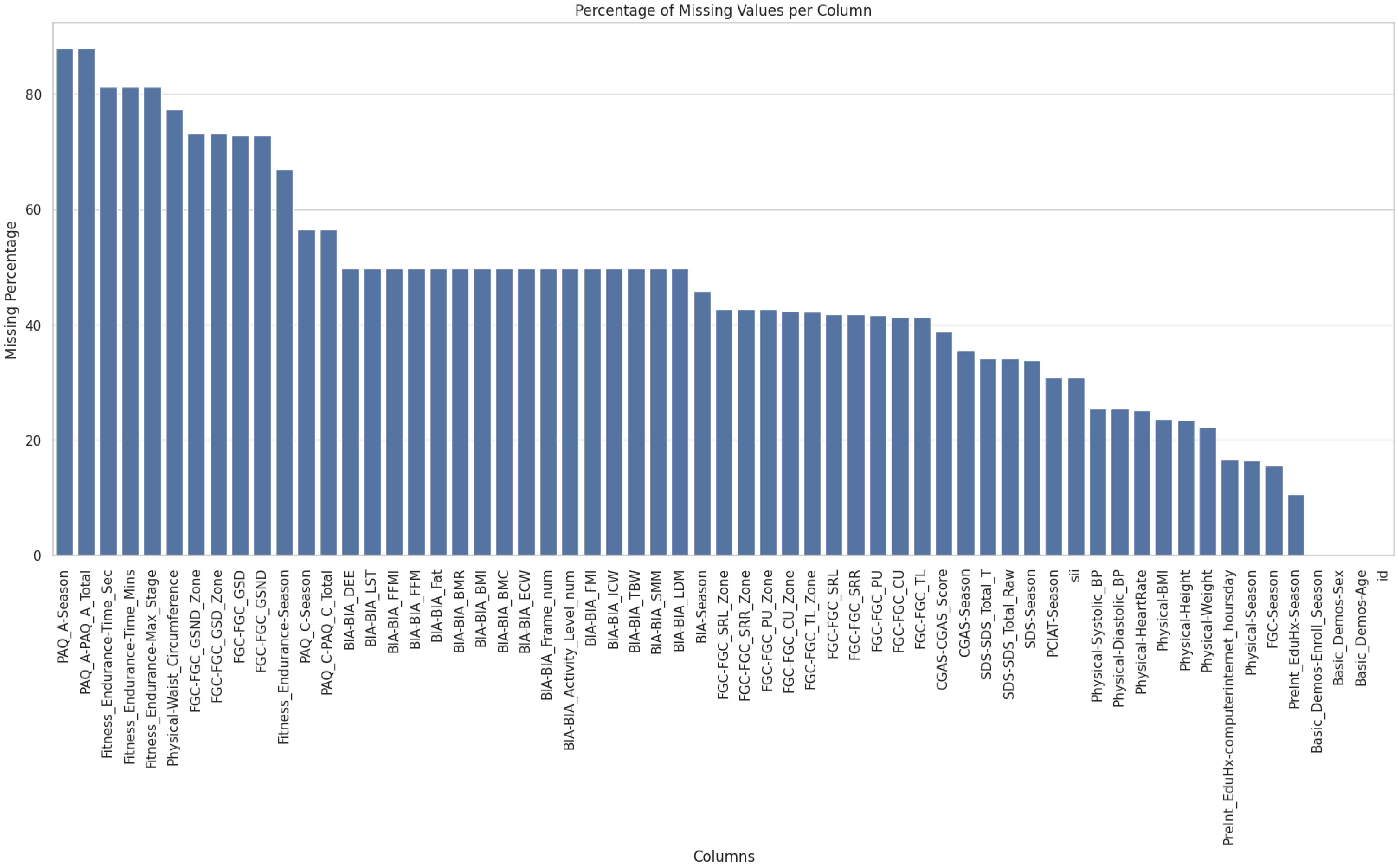
HBN Data



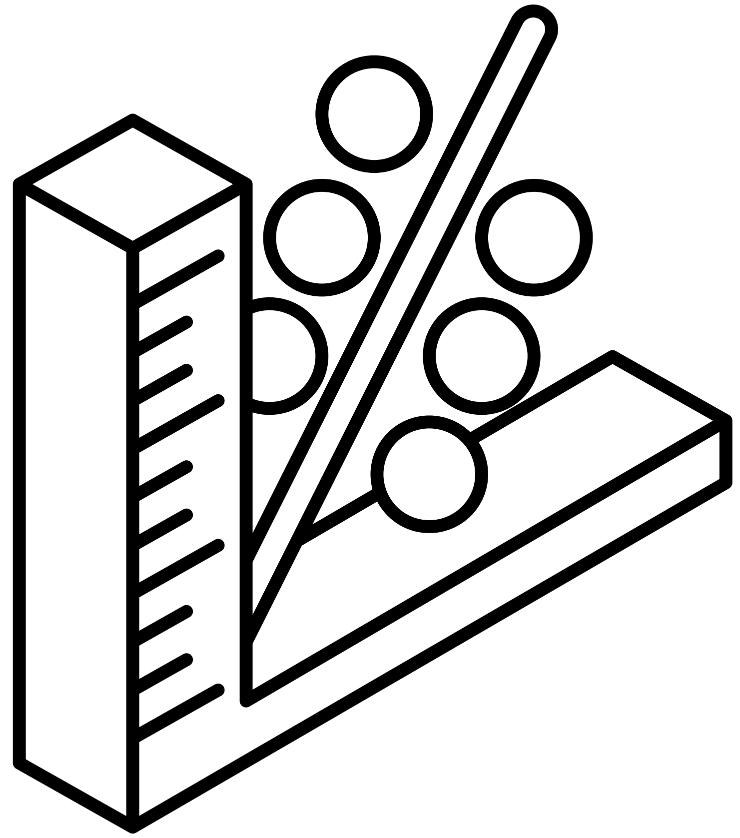
	step	X	Y	Z	enmo	anglez	non-wear_flag	light	battery_voltage	time_of_day	wee
0	0	0.679618	-0.578170	0.320939	0.273671	18.857922	0.0	6.000000	4175.00	40260000000000	
1	1	-0.139180	-0.727815	0.591056	0.083061	37.902840	0.0	1.500000	4174.75	40265000000000	
2	2	-0.601730	-0.353205	0.675733	0.032478	44.337727	0.0	1.000000	4174.50	40270000000000	
3	3	-0.379666	-0.283236	0.613746	0.074800	44.593758	0.0	55.200001	4174.25	40275000000000	
4	4	0.397701	0.470948	-0.486966	0.070064	-31.758539	0.0	21.000000	4174.00	40280000000000	
5	5	0.549772	-0.480966	-0.486315	0.090724	-31.631941	0.0	17.833334	4173.75	40285000000000	
6	6	0.262262	-0.575119	-0.508705	0.160436	-34.813435	0.0	14.666667	4173.50	40290000000000	
7	7	0.080009	0.131717	-0.968861	0.017968	-78.954483	0.0	12.000000	4173.25	40295000000000	
8	8	0.431136	0.197717	-0.603907	0.033459	-46.023037	0.0	1.000000	4173.00	40300000000000	
9	9	0.502958	0.176596	0.670719	0.0305[illegible]2	49.410473	0.0	8.000000	4172.75	40305000000000	

Actigraphy Data

COLUMNS WITH HIGH MISSING VALUES

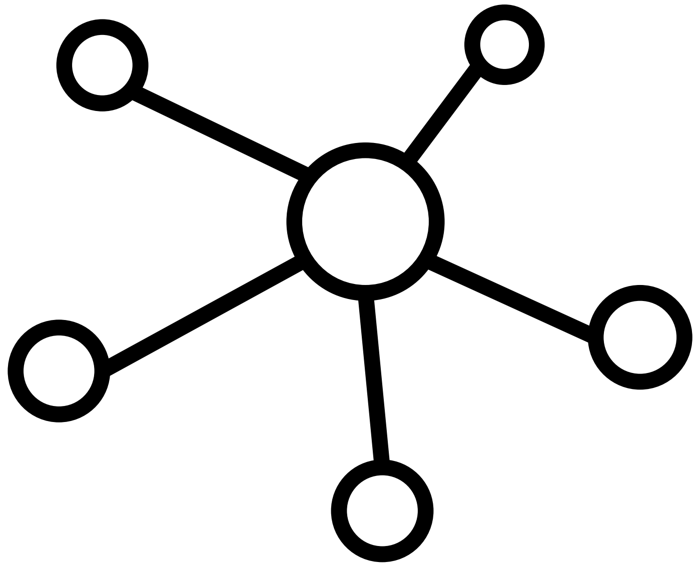


HANDLING MISSING DATA



Step 1: Removing Columns

- Dropped columns with more than 60% missing values to ensure data quality.

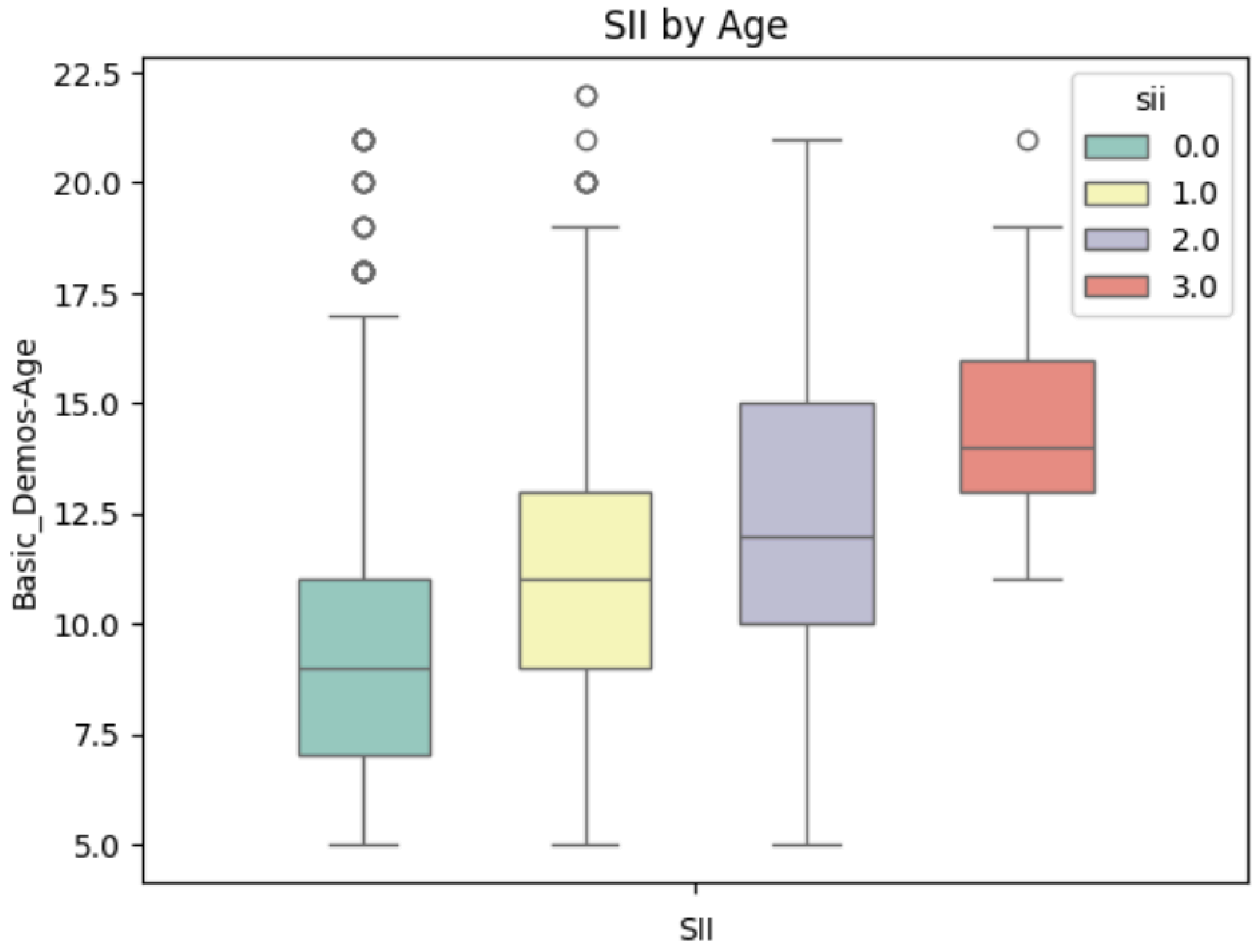


Step 2: Imputing Remaining Data

- For columns with less than 60% missing values, used KNN Imputer to fill in the gaps based on similar data points.

Insights....

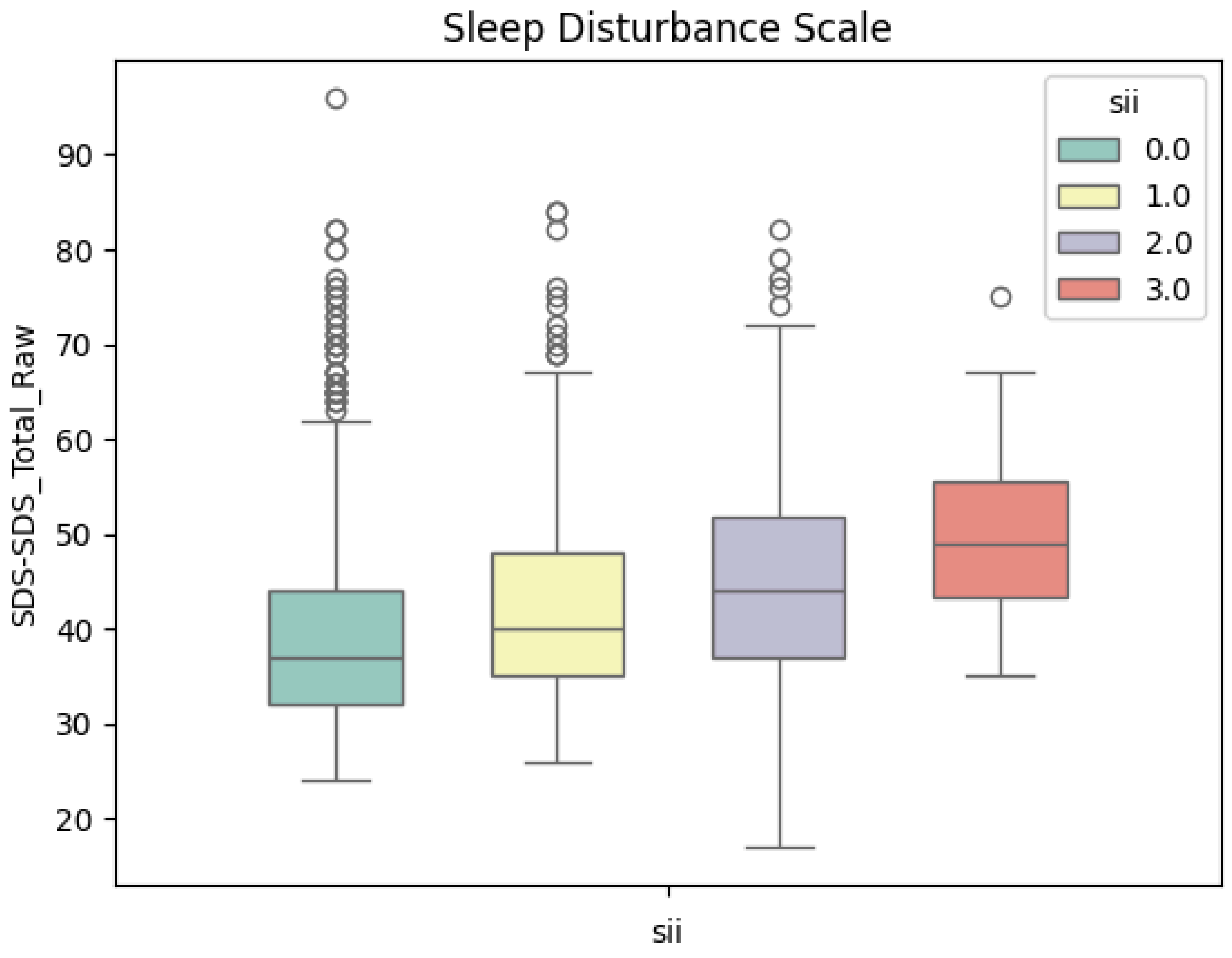
SII SCORES BY AGE



Individuals with higher SII scores (e.g., 3.0) tend to belong to older age groups, as reflected by the upward trend in median age.

Despite this trend, there is considerable overlap in ages across different SII categories, indicating that severity impairment can impact individuals across a broad age range.

SLEEP DISTURBANCE



The box plot demonstrates an increasing trend in the median SDS score as SII severity increases, highlighting the potential impact of sleep quality on problematic internet use.

Associations of Physical Activity, Screen Time with Depression, Anxiety and Sleep Quality among Chinese College Freshmen

Qi Feng, Qing-le Zhang, Yue Du, Yong-ling Ye, Qi-qiang He

Published: June 25, 2014 • https://doi.org/10.1371/journal.pone.0100914

Article **Authors** **Metrics**

Abstract
Introduction
Methods
Results
Discussion
Supporting Information
Author Contributions
References

Reader Comments
Figures

Abstract

Objectives

To investigate the independent and interactive associations of physical activity (PA) and screen time (ST) with depression, anxiety and sleep quality among Chinese college students.

Methods

A cross-sectional study was conducted in Wuhan University, China from November to December 2011. The students reported their PA, ST and socio-economic characteristics using self-administered questionnaires. Sleep quality was measured by the Pittsburgh Sleep Quality Index (PSQI). Depression and anxiety were assessed using the Self-rating Depression Scale (SDS) and Self-rating Anxiety Scale (SAS), respectively. Multivariate logistic regression models were used to estimate the odds ratios (ORs) and 95% confidence intervals (CIs) of the independent and interactive relationships of PA and ST with depression, anxiety and sleep quality.

Results

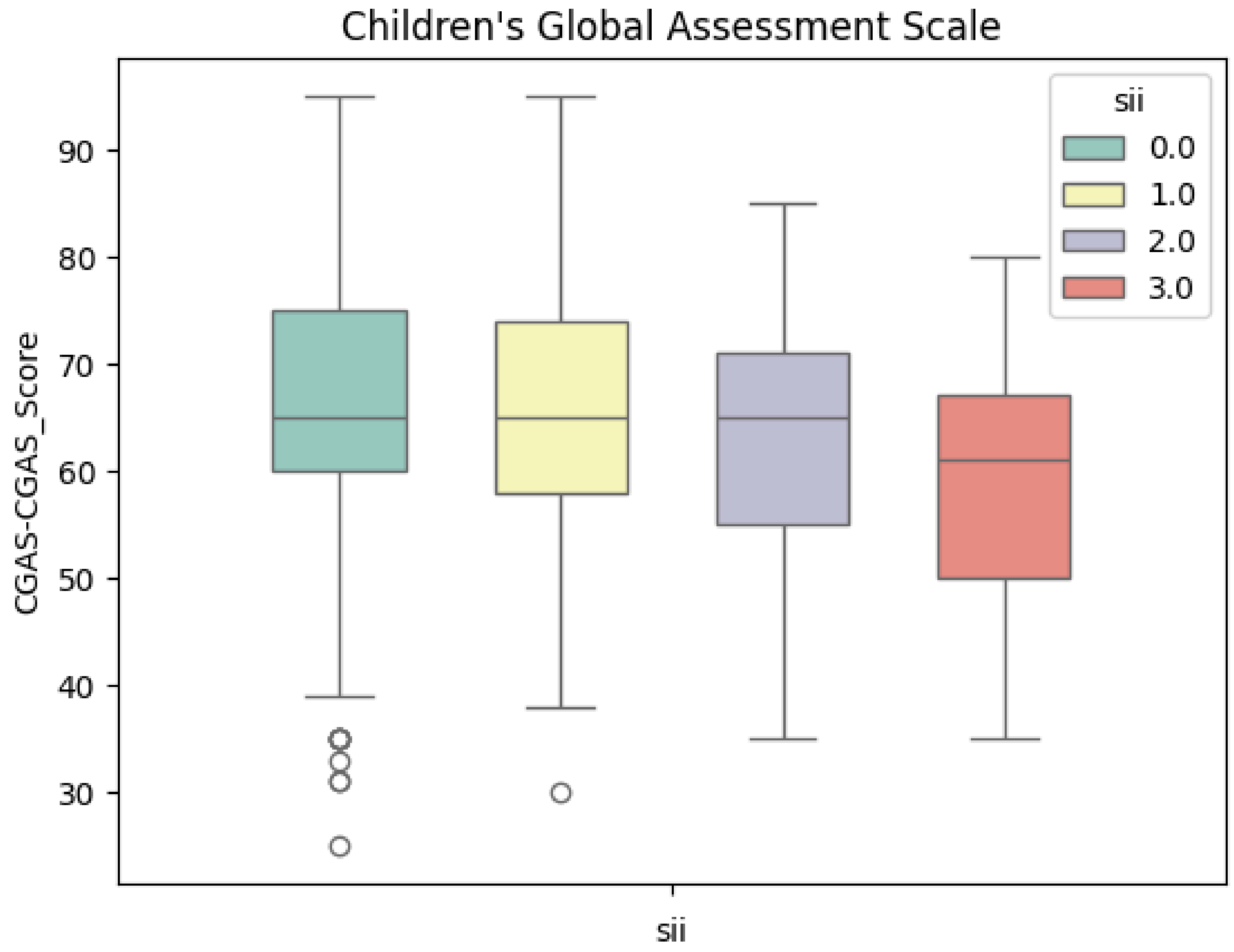
A total of 1106 freshmen (471 females and 635 males) aged 18.9±0.9 years were included in the study. After adjustment for potential confounders, high PA and low ST were independently associated with significantly lower risks for poor sleep quality (OR: 0.48, 95% CI: 0.30–0.78) and depression (OR: 0.67, 95%CI: 0.44–0.89), respectively. An interactive inverse association was observed for combined effects of PA and low ST on depression (OR: 0.62, 95%CI: 0.40–0.92) and sleep quality (OR: 0.51, 95%CI: 0.27–0.91). No statistically significant associations were found between PA, ST and anxiety among the participants.

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0100914

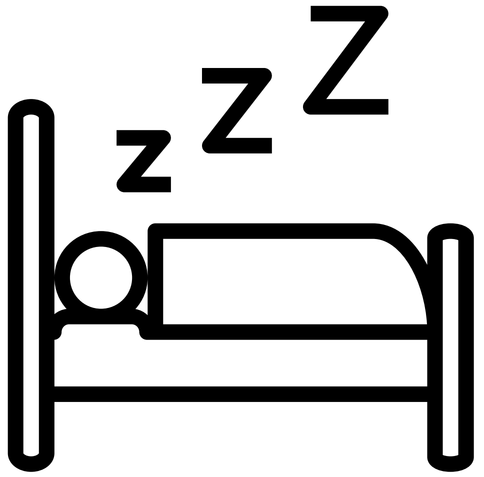
CHILDREN'S GLOBAL ASSESSMENT SCALE AND SII CORRELATION

The Children's Global Assessment Scale (CGAS) is a measure developed by Schaffer and colleagues at the Department of Psychiatry, Columbia University to provide a global measure of level of functioning in children and adolescents. The measure provides a single global rating only, on scale of 0-100.

Higher the SII (proxy for Problematic Internet usage) lower is the CGAS scores, indicating reduced functioning.



FEATURES FROM ACCELEROMETER DATA



Percentage of Sleep Time

$$P_{sleep} = \frac{Instances of Sleep}{Total Instances}$$

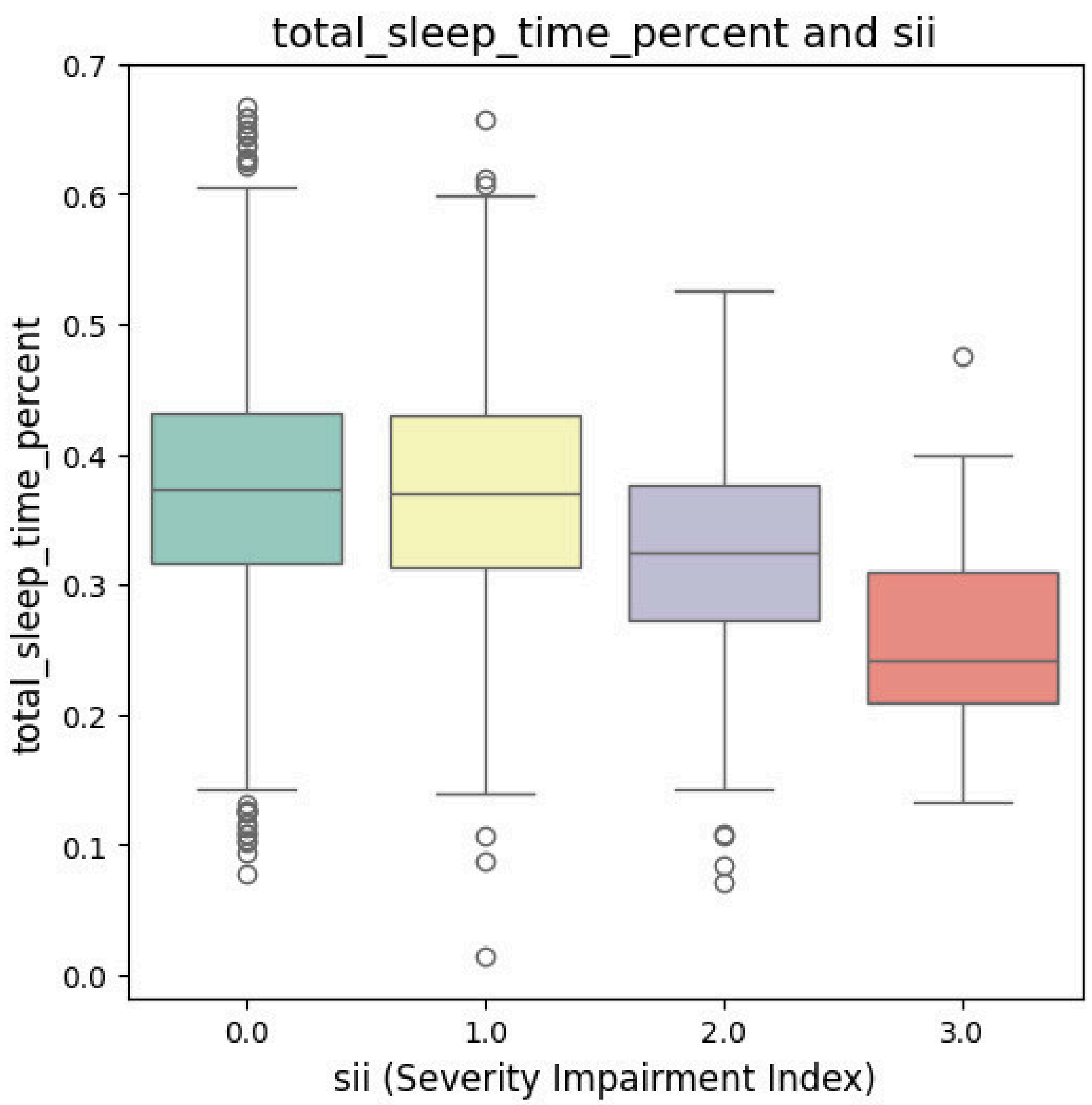
Percentage of Sitting Time

$$P_{sitting} = \frac{Instances of Sitting}{Total Instances}$$

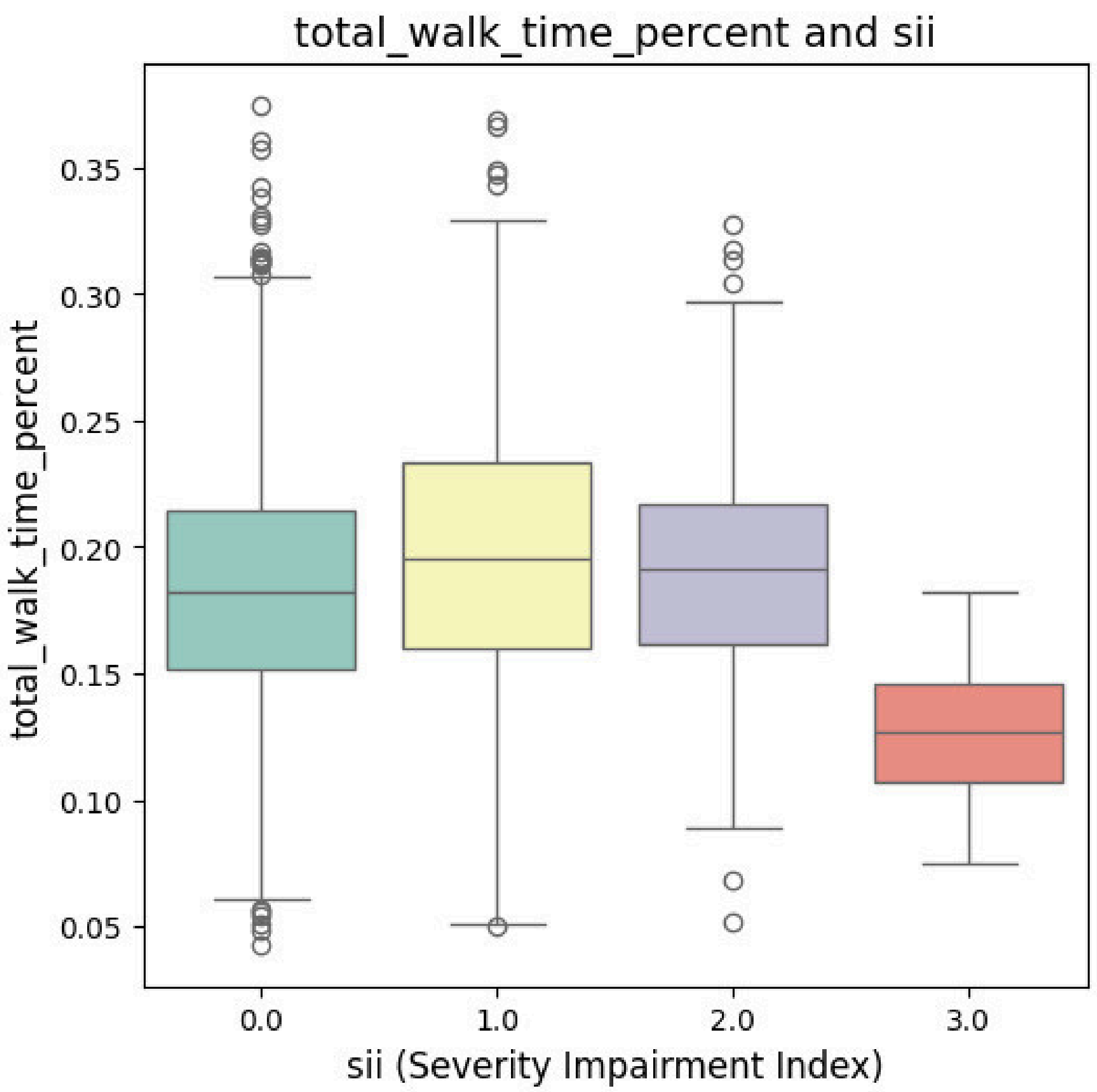
Percentage of Walking Time

$$P_{walking} = \frac{Instances of Walking}{Total Instances}$$

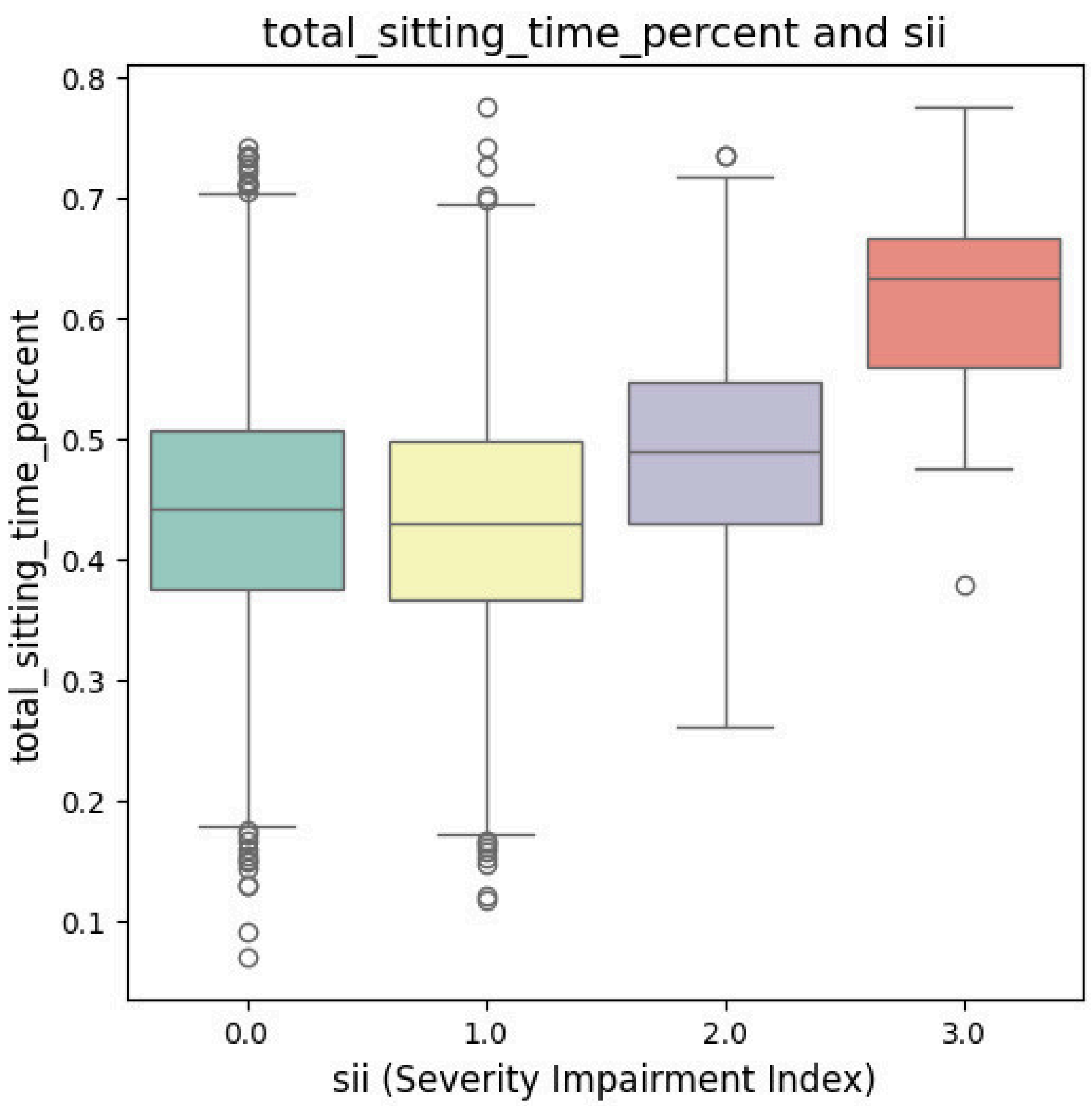
SII AND ACTIVITY PATTERNS



Higher SII scores show a decrease in total sleep time percentage and greater sleep disturbances.



As SII increases, there is a visible decline in total walk time percentage, indicating lower engagement in physical activity.



Higher SII scores are associated with an increase in total sitting time percentage, reflecting a more sedentary lifestyle.

MERGING DATA

- **Feature Integration:** Actigraphy data provided temporal activity patterns, while Healthy Brain Networks data offered cognitive and neurological features.

- **Final Dataset:** The combined dataset was prepared for comprehensive modeling, ensuring a richer representation of predictors for SII.

Healthy Brain Networks Data

+

Actigraphy Files (Fitness Band Data)

MODEL SELECTION FOR PREDICTING SII

LightGBM: A Highly Efficient Gradient Boosting Decision Tree

XGBoost: A Scalable Tree Boosting System

Tianqi Chen
University of Washington
tqchen@cs.washington.edu

Carlos Guestrin
University of Washington
guestrin@cs.washington.edu

We call the new GBDT algorithm with GOSS and EFB *LightGBM*[2]. Our experiments on multiple public datasets show that LightGBM can accelerate the training process by up to over 20 times while achieving almost the same accuracy.

ng used as a stand-alone predictor, it
o real-world production pipelines for
prediction [15]. Finally, it is the de-
le method and is used in challenges
ze [3].

In this paper, we describe XGBoost, a scalable machine learning system for tree boosting. The system is available as an open source package[2]. The impact of the system has

on many machine learning challenges. We propose a novel sparsity-aware algorithm for sparse data and weighted quantile sketch for approximate tree learning. More importantly,
we provi
sion and
By combi
of examp

method, deep neural nets, was used in 11 solutions. The success of the system was also witnessed in KDDCup 2015, where XGBoost was used by every winning team in the top-10. Moreover, the winning teams reported that ensemble methods outperform a well-configured XGBoost by only a small amount [1].

Abstract

Gradient Boosting Decision Tree (GBDT) is a popular machine learning algorithm, and has quite a few effective implementations such as XGBoost and pGBRT. Although many engineering optimizations have been adopted in these implementations, the efficiency and scalability are still unsatisfactory when the feature dimension is high and data size is large. A major reason is that for each feature, they need to scan all the data instances to estimate the information gain of all

https://www.kdd.org/kdd2016/papers/files/rfp0697-chenAemb.pdf

https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

MODEL SELECTION FOR PREDICTING SII

Objective: Identify the most accurate model for predicting **SII** before addressing class imbalance.

Models Compared: Random Forest, XGBoost, and LightGBM.

Performance Results: LightGBM outperforms others with the highest accuracy of **87.11%.**

Model	Accuracy (%)
Random Forest	66.04
XGBoost	84.27
LightGBM	**87.11**

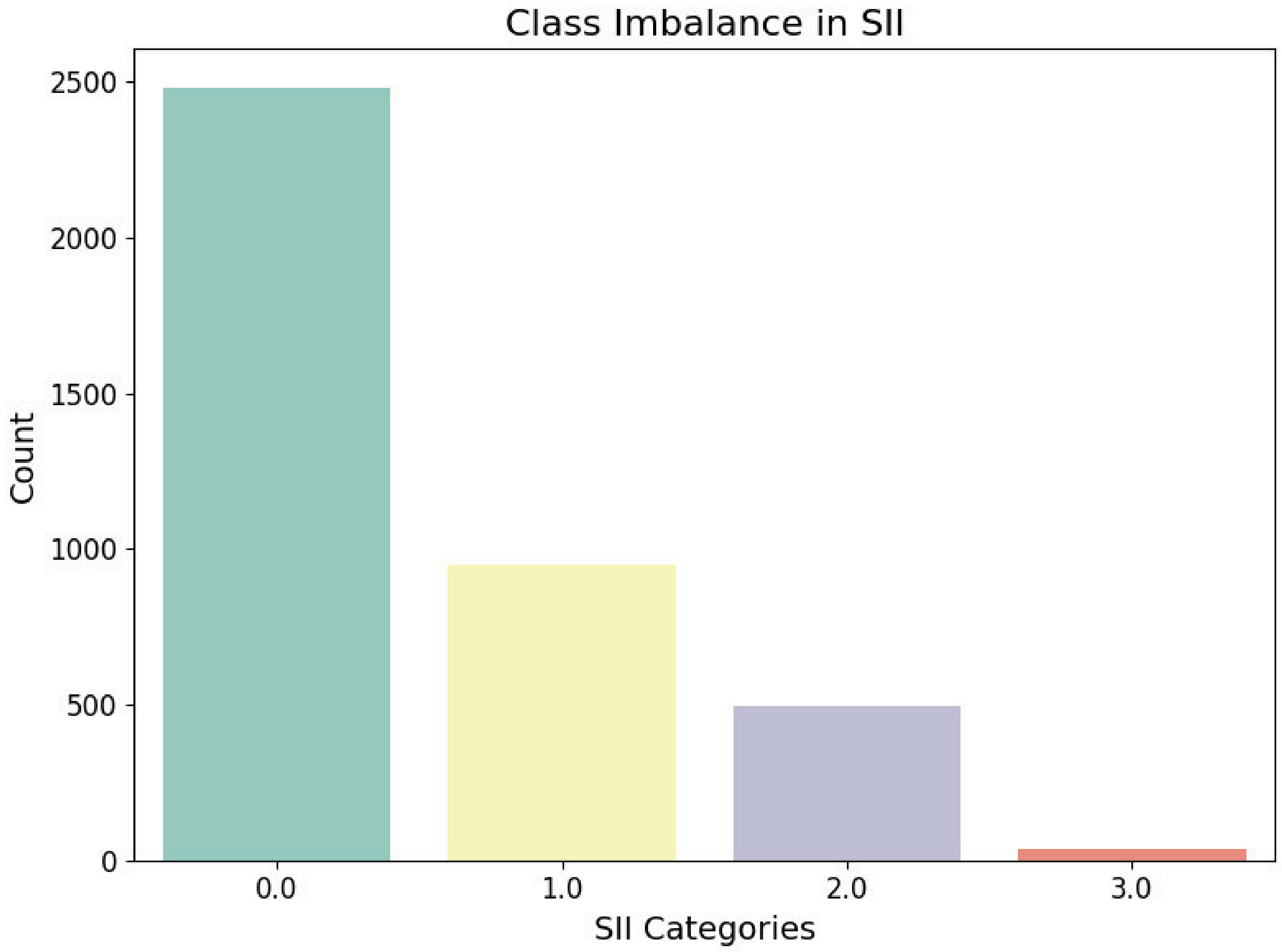
Conclusion: LightGBM is the preferred choice for its superior performance in the preliminary evaluation.

HANDELLING CLASS IMBALANCE

The dataset reveals a significant class imbalance in the SII categories, with the majority of data points concentrated in lower levels of internet addiction.

Handling this imbalance is crucial to:

- Ensure Model Accuracy
- Improve Generalization
- Fair Representation



CLASS IMBALANCE

Handling class imbalance problem using oversampling techniques: A review

Publisher: IEEE

Cite This | PDF

Anjana Gosain ; Saanchi Sardana All Authors

Abstract
Document Sections
I. Introduction
II. Oversampling Techniques Used
III. Evaluation Metrices
IV. Dataset Analysis
V. Experimental Results
Show Full Outline

Abstract:
The objective of classifier is to classify objects of a data set into one or more classes based on its characteristics. In real life applications, classifiers are applied on data sets which are unbalanced i.e. some classes having very less number of instances known as minority classes as compared to other classes known as majority classes. Classification algorithms are highly accurate for the majority classes but significantly less accurate for the minority classes. Unbalanced data sets have a negative effect on classification performance of traditional classification algorithms. Analyzing such problem is called class imbalance problem. To solve Class Imbalance Problem different techniques have been proposed at the Data level, Algorithm level and at the Hybrid level. Most commonly used data balancing techniques are over and under sampling for handling the class imbalance problem. In our paper we compare various oversampling techniques which are SMOTE (Synthetic minority oversampling approach), ADASYN, Borderline-SMOTE, Safe-Level SMOTE by applying different classifiers to the problem and observing various performance metrics.

proposed at the Data level, Algorithm level and at the Hybrid level. Most commonly used data balancing techniques are over and under sampling for handling the class imbalance problem. In our paper we compare various oversampling techniques which are SMOTE (Synthetic minority oversampling approach), ADASYN, Borderline-SMOTE, Safe-Level SMOTE by applying different classifiers to the problem and observing various performance metrics.

Here is a list of techniques we used for handling class imbalance:

1. Undersampling
2. SMOTE (Synthetic Minority Oversampling Technique)
3. Oversampling
4. Class Weights
5. ADASYN (Adaptive Synthetic Sampling)

HANDELLING CLASS IMBALANCE

Challenge: Improving model performance by addressing data class imbalance.

Model Used: LightGBM (Gradient Boosting Framework)

Method	Accuracy (%)
Undersampling	58.35
SMOTE	64.69
Oversampling	75.02
Class Weights	**77.29**
ADASYN	76.32

DATA COLLECTION

1
Health Data

Collected easily accessible information like Physical-BMI, Physical-Height, and Physical-Weight.
Retrieved additional data from previous medical records, including Physical-Diastolic_BP and Physical-Systolic_BP.

.

2
Screen Time Data

Used as a proxy for time spent sitting or lying down (essentially no physical activity).
Primarily intended for verifying the data collected from the actigraphy dataset.

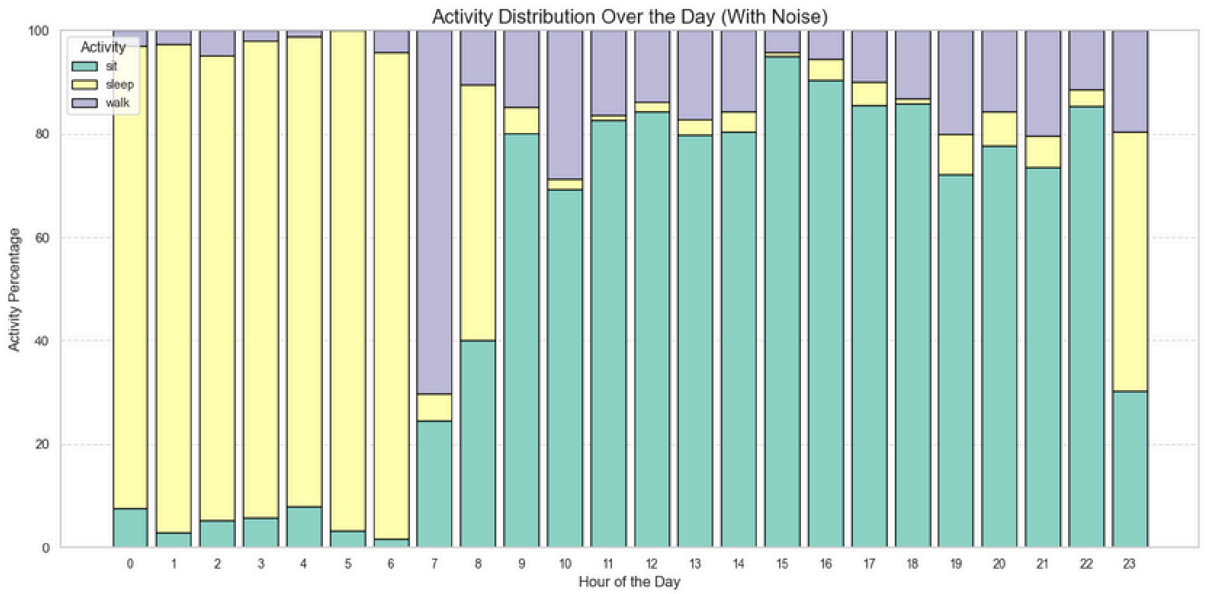
3
Actigraphy Data

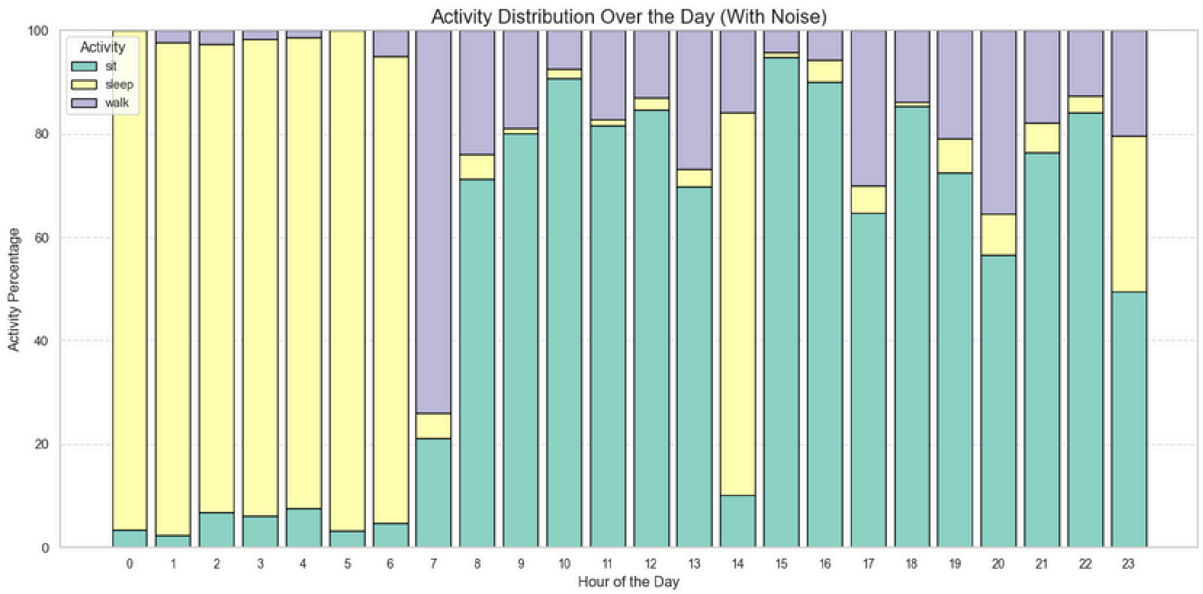
Collected through fitness bands, which classify activity levels and patterns.

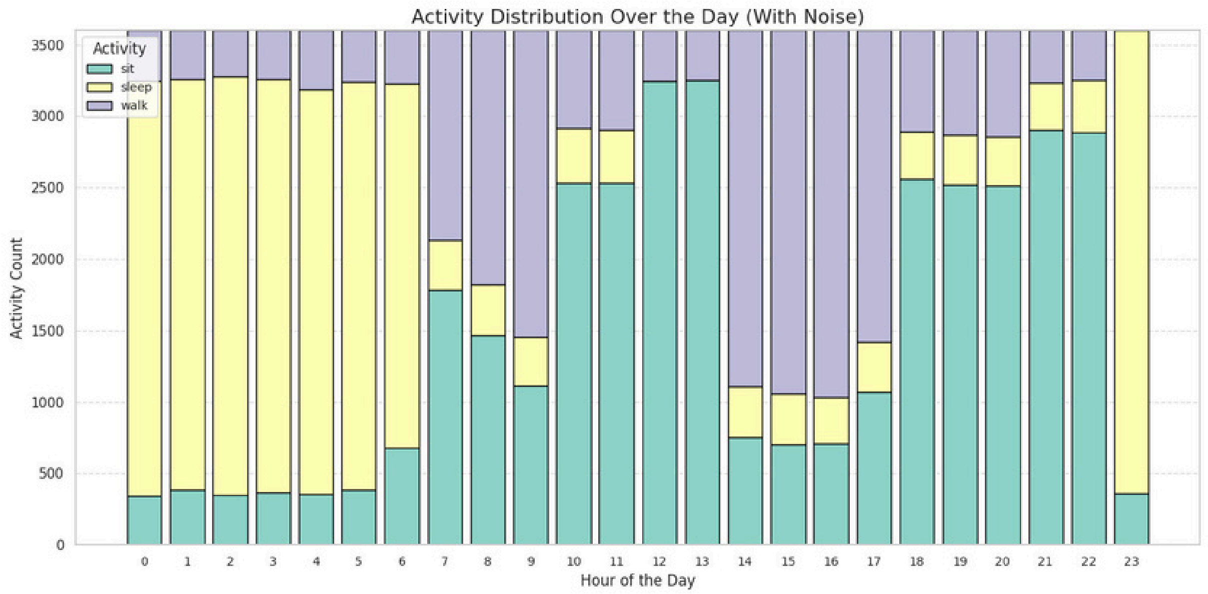
Table 2. Raw Data Table Column Definitions

Column	Description
timestamp	Milliseconds from 01.01.1970 (UNIX epoch in milliseconds)
running_counter	A running 16-bit number to check no data is missing
ppg1	Green PPG signal
ppg2	Ambient for green PPG signal
ppg3	Not in use
ppg4	Not in use
acc1	Acceleration in one dimensions
acc2	Acceleration in one dimensions
acc3	Acceleration in one dimensions
rssi	Signal strength during the transfer of the file containing this line
savetime	Time according to the phone when the file was transferred (for SQL table only, not csv)

INFERENCE ON THE COLLECTED DATA

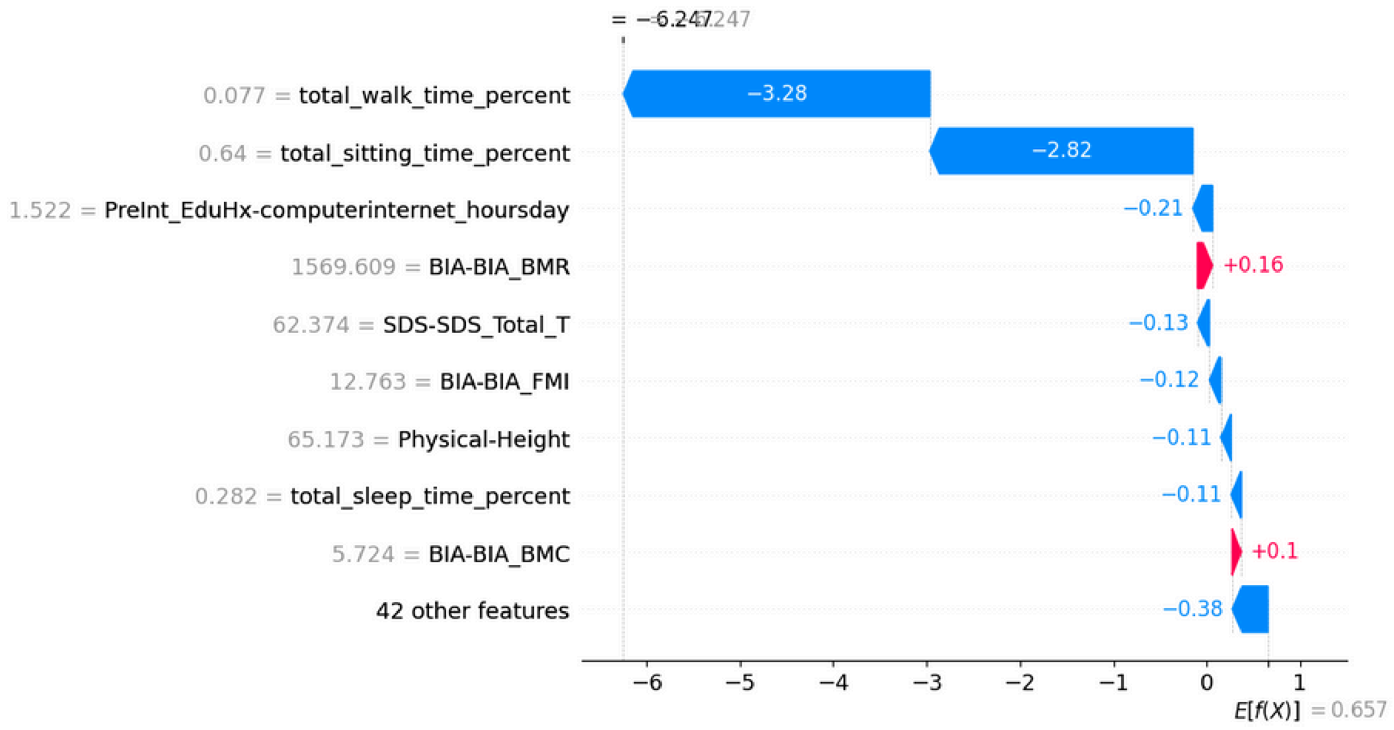


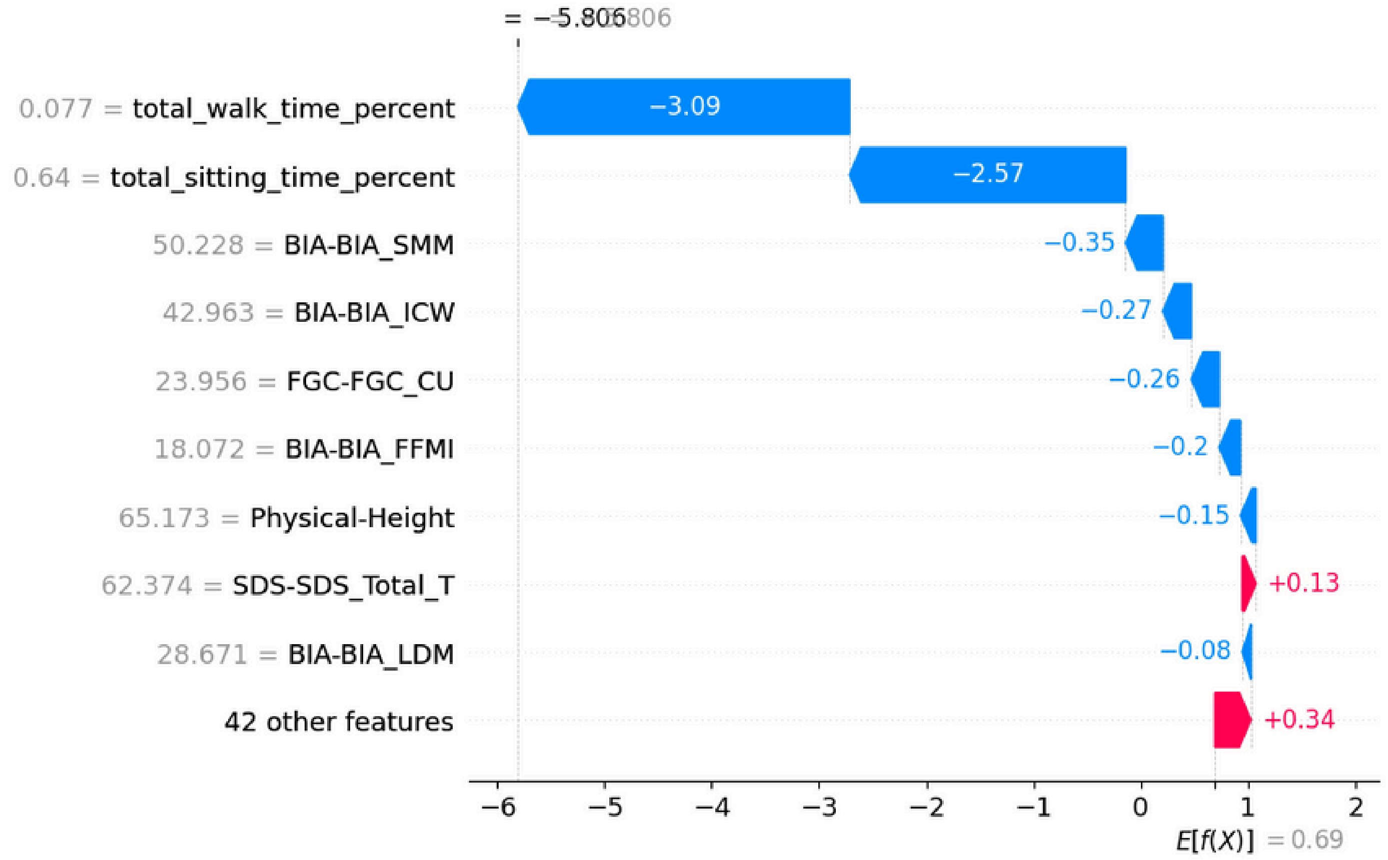


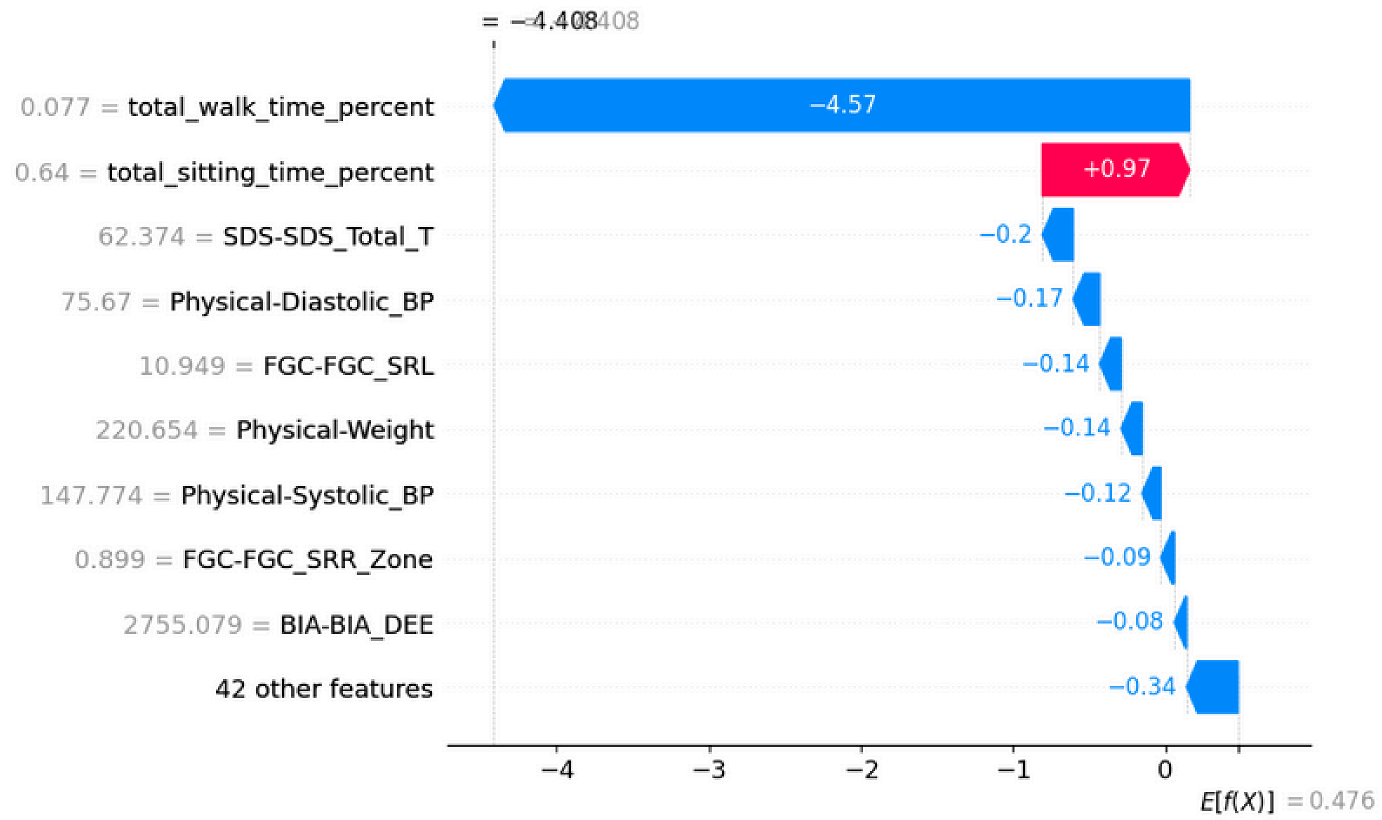


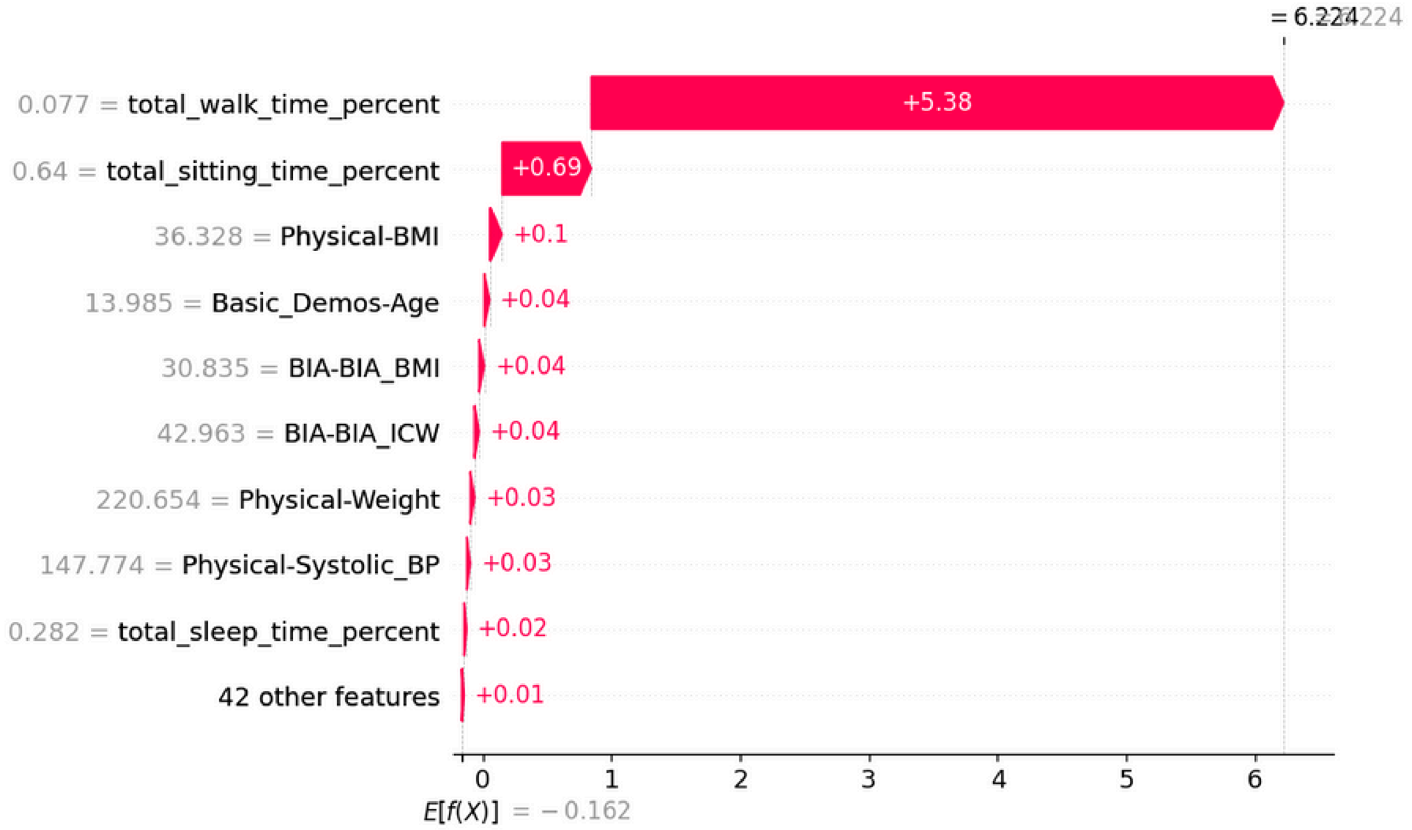
The trained model classified all participants with an SII score of 3, indicating problematic internet usage. This aligns with the observed average screen time of **8 hours 27 minutes**, coupled with minimal physical activity.

INFERENCE ON THE COLLECTED DATA





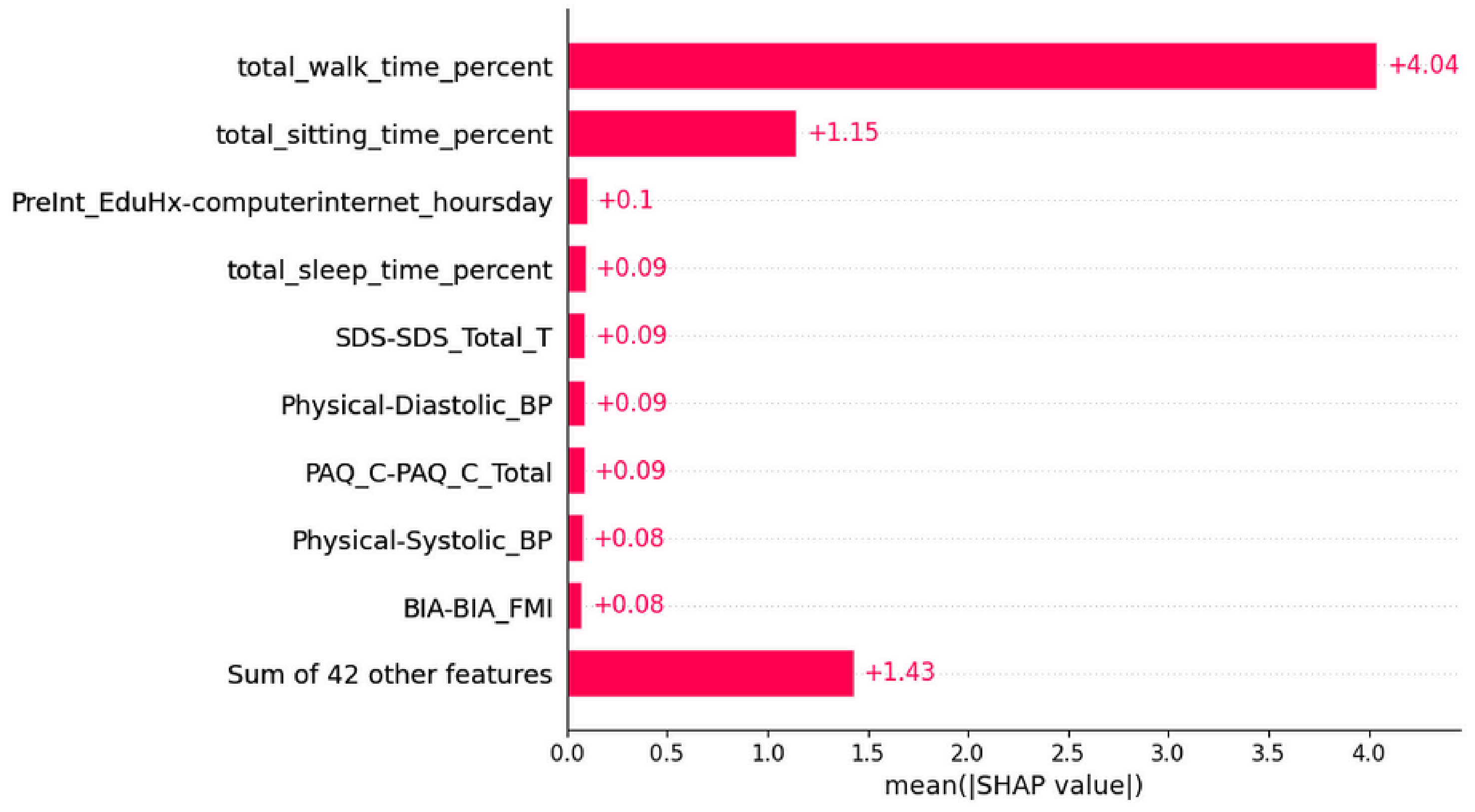




AGGREGATING SHAP VALUES

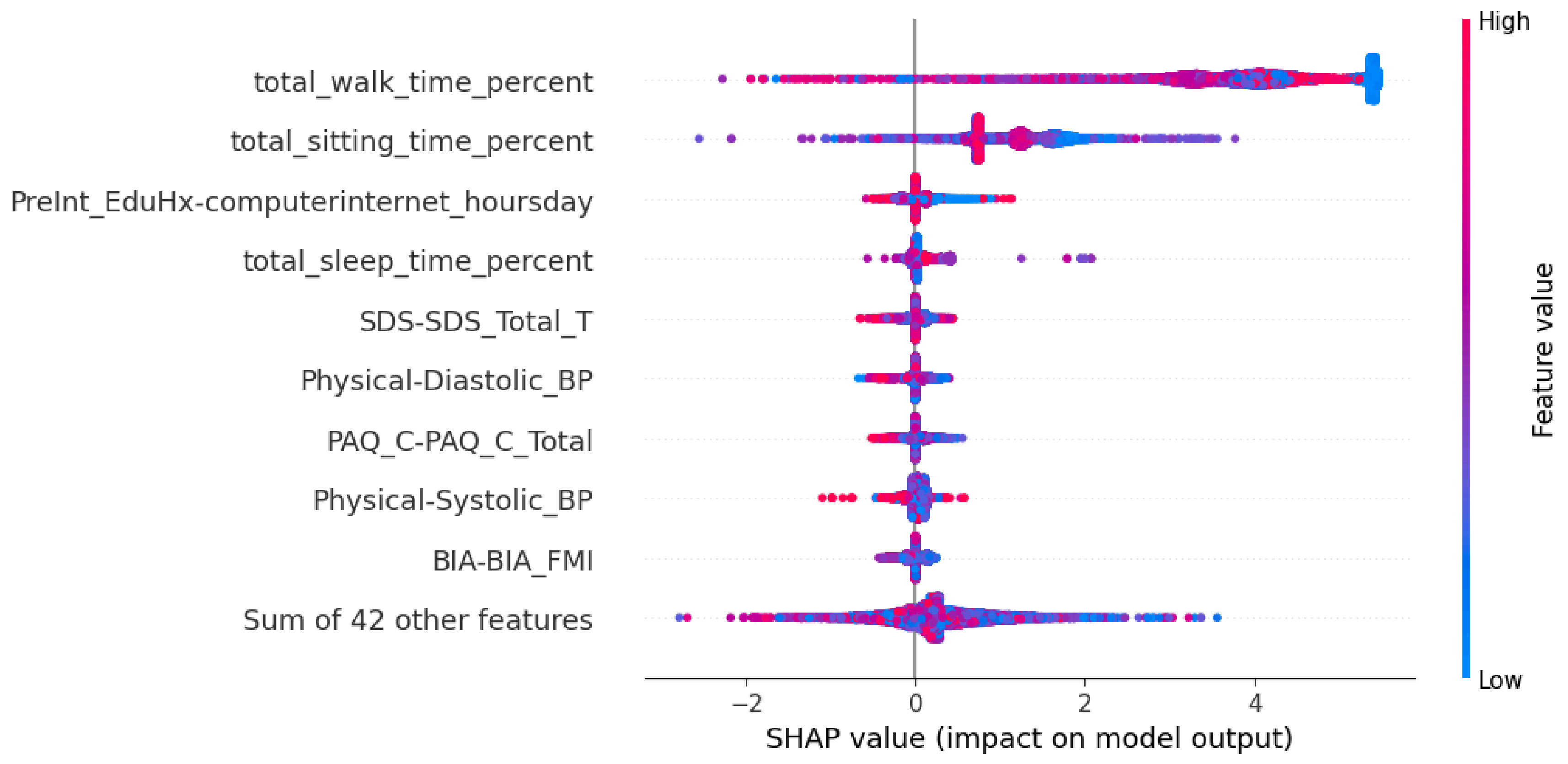
Focusing on Predicted Class

We have different plots for each class, analyzing this can be bit tedius so we will focus on the values of only the predicted class. This graph shows the contribution of features to the predicted class.



Chaos in Bee-swarm Plots

In the bee-swarm plot we do not see a clear relationship between shap values and feature values because features will have different relationships depending on the predicted class.



IMPACT

- Our model, trained on both physiological and actigraphy data, streamlines the tedious and complex conventional process of assessing problematic internet usage.
- Following the assessment, we can provide timely interventions based on relevant features to enhance the well-being of individuals.

Thank you!