Child Mind Institute — Problematic Internet Use

Anvesh Kumar Chavidi, Eshita Gupta, Neha Thakur, Padhmavathy Cebolu Srinivasan, Saumya Varshney, and Sneha Karri

Department of Applied Data Science

San Jose State University

DATA 255 Deep Learning

Professor Simon Shim

December 09, 2024

Child Mind Institute — Problematic Internet Use

Introduction

In the present environment, which is increasingly digital, the interplay between the use of technology and children's and adolescents' health has emerged as a very crucial area of concern. Problematic internet use among youths is a serious public health problem that affects mental health and normal development processes, generally reducing the quality of life. Digital technologies bring many benefits, but the risks from excessive screen time are becoming increasingly well-documented. Traditional diagnostic procedures for such conditions are usually complex clinical assessments, which may be inaccessible, expensive, or difficult to use across socio-cultural boundaries, which further emphasize the need for more adaptable strategies.

This competition proposes a very novel strategy to use the measures of physical activity as an indirect measure for problematic internet use. The central hypothesis is that excessive technological engagement emerges through observable changes in physical behaviors and fitness indicators. The use of physical activity data brings multiple benefits, ease of fitness measurement, minimum clinical expertise in collecting the data, and the potential to leverage large datasets. These can be combined to develop effective personalized scalable predictive models contributing to a more holistic understanding of digital wellness. If successful, this approach may significantly enhance public health approaches in tackling problematic Internet use. Such predictive models could facilitate early intervention, supporting strategies that prevent poor mental health, create accessible mechanisms for screening, and facilitate healthier ways of engaging digitally.

The Healthy Brain Network (HBN) dataset was instrumental for this effort; it includes data on almost 5000 subjects aged 5-22, with rich physical activity measures based on accelerometer time-series measurements and extensive clinical and screening information. A variety of standardized assessments across multiple health domains are also included. The Severity Impairment Index (SII), derived from responses to the Parent-Child Internet Addiction Test, categorizes the levels of problematic internet use into four groups of severity: none, mild, moderate, and severe. Relating SII scores to physical activity metrics can enable the development of models that may predict problematic patterns of use before they escalate. The prevalence of missing data, imputation strategies, and feature engineering necessary for robust modeling will be further discussed in the report.

Litreture Survey/ Related work

The study by Cai et al., 2023 looks at the connection between student's mental health problems and problematic internet use (PIU). It summarizes research from several studies and shows that PIU is strongly linked to poor mental health outcomes, such as stress, anxiety, and depression. In order to address PIU and its effects on students' psychological health, the review highlights the necessity of early interventions and focused techniques. Potential processes that connect internet use behaviors to mental health issues are also covered.

The study by Saletti et al., 2021 assessed a number of intervention initiatives aimed at addressing youth problematic internet usage (PIU). It finds effective prevention techniques and emphasizes the value of early intervention to lessen the psychological and behavioral effects of PIU by methodically examining previous research. The authors stress the importance of age-appropriate, customized programs in lowering excessive internet use and encouraging better online practices.

The study by Pontes and Macur, 2021 examines the many forms of problematic internet usage (PIU) in teenagers and the psychosocial hazards that go along with it. It classifies teenagers according to how they use the internet and looks at connections to academic achievement, social isolation, and mental health issues. The study emphasizes the necessity of focused interventions made to fit certain PIU profiles in order to successfully address the underlying psychosocial problems.

The study by Restrepo et al., 2020 explores the connection between children's and adolescents' psychiatric illnesses and problematic internet use (PIU). It draws attention to the correlation between excessive internet behavior and mental health issues like melancholy, anxiety, and ADHD, which can impede social and academic functioning. The results highlight the value of PIU screening in mental health assessments and recommend using digital behavior management in therapeutic approaches.

The paper by Koczkodaj et al., 2022 suggests a unique strategy to prevent problematic internet use by combining differential evolution algorithms with peer assessment methodologies. It places a strong emphasis on identifying and reducing hazardous or excessive online activity through data-driven strategies. The study emphasizes how cooperative tactics, and computational tools might promote safe internet use.

The paper by Dresp-Langley, 2020 examines how digital technology affects kids' health, with an emphasis on mental, emotional, and physical health. It acknowledges the advantages of technology when used properly while also talking about the dangers of excessive screen time, such as mental health problems and developmental disabilities. The author supports rules that encourage kid-friendly online spaces and balanced digital use.

The study by Shi et al., 2021 explores the idea of "cyber-syndrome," combining Maslow's hierarchy of requirements with the cyber-physical-social-thinking space. It explores how human needs—from self-actualization to basic physiological needs—are met or disrupted by technology. A theoretical framework for comprehending digital dependency and its wider societal effects is put out by the authors.

The study by Cheng et al., 2019 examines the behavioral and psychological elements of problematic Facebook use, concentrating on those who feel powerless over their use and suffer detrimental effects on their lives. It recognizes trends of over-engagement and offers solutions to encourage more constructive online interactions. Insights on platform design considerations to lessen addictive behaviors are also provided by the research.

The paper by Tokunaga, 2015 applies communication theories to understand internet addiction and problematic internet use. It talks about how poor self-control, exacerbated by contemporary communication tools, leads to compulsive behaviors. The author highlights how personal psychology, and the media landscape interact, arguing that interdisciplinary approaches are necessary to combat PIU.

The paper by Cash et al., 2012 presents a comprehensive theoretical model for understanding internet addiction disorder, focusing on its classification, underlying mechanisms, and impact. The authors draw attention to research limitations and urge additional studies on preventive and intervention techniques. Potential parallels with other behavioral addictions are also covered in the research, which highlights the necessity of a standardized diagnostic approach.

Related Work

To capture the complex nature of Self-Identified Internet Usage (SII) or related problematic internet usage (PIU) behaviors, several research have concentrated on predicting and assessing these behaviors using a variety of models and assessment metrics. The following summarizes the approaches and insights from previous research.

Cai et al., 2023 conducted meta-analytic evaluations, in order to establish a connection between internet use behaviors and mental health outcomes. To find correlations between PIU and outcomes such as stress, anxiety, and depression, regression models were employed.

Tokunaga, 2015 emphasizes the role of self-regulation and media use patterns in PIU. Building on this, researchers have used regression models to link internet behaviors with psychosocial outcomes, deep learning models such as RNNs and LSTMs to analyze sequential usage data. Metrics like accuracy, recall, AUC-ROC, and F1 scores evaluate model performance, advancing the prediction of PIU and guiding targeted interventions. These methods align with Tokunaga's insights, highlighting PIU's multifaceted nature.

Montag et al., 2019 discusses the role of addictive design features in problematic internet use. Building on this, other researchers have predicted self-identified internet use (SII) using neural networks to analyze behavioral patterns, demographic data, and psychological scores. Deep learning models such as MLPs and RNNs have been employed to capture non-linear relationships and temporal patterns in user data. Evaluation metrics like accuracy, recall, and AUC-ROC have been widely used to measure model performance and effectiveness in identifying at-risk users.

Pontes and Macur, 2021 used deep neural networks (DNNs) to classify teenagers into discrete PIU profiles according to their behavioral and psychosocial traits. The researchers successfully identified latent risk categories by using the networks to extract significant elements from multi-dimensional input data, such as internet usage patterns and mental health markers.

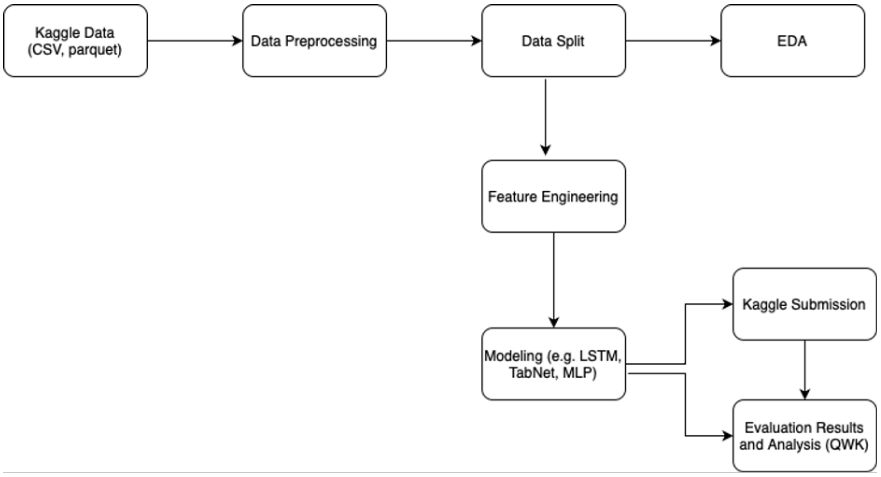
Koczkodaj et al., 2022 presented a hybrid deep learning architecture that dynamically predicts dangerous internet use patterns by fusing differential evolution algorithms with peer assessment feedback. Behavioral patterns were processed as temporal or spatial input using convolutional neural networks (CNNs), and the model's parameters were adjusted through evolutionary techniques to allow for greater adaptability.

Project Workflow

The project is structured to follow a step-by-step process, involving both Kaggle submission and evaluation metrics specific to this project. The first step of this process is to obtain the data from Kaggle. This is followed by a Data Preprocessing step, which performs data cleaning and schema corrections, details of which are covered in later sections. This is followed by an extensive Exploratory Data Analysis on the cleaned data. Before the data is ready for modelling, Feature Engineering is performed to combine various data sources. Then various Deep Learning modelling techniques are used to model the classification task, and the models are trained on the engineered features. Predictions made by each model are submitted to Kaggle for evaluation using the competition specific metrics. The results from each model and the Kaggle submission are then compared and analyzed. Figure 1 below shows the workflow of the project.

Figure 1

Project Workflow



Dataset Source and overview

Dataset Source

The competition uses a dataset from the Healthy Brain Network, which consists of approximately 5000 participants between age 5 to 22 years. This is an exciting possibility for interdisciplinary research where detailed information-from physical health to self-reported questionnaires and objective activity measurements-is available. Key features from the data include comprehensive physical activity data obtained through accelerometer (actigraphy) recordings, as well as in-depth clinical and research screening instruments. The primary objective of this competition is to accurately predict the Severity Impairment Index (SII), a standardized measure designed to quantify problematic internet use. The SII is derived from participant responses to the Parent-Child Internet Addiction Test (PCIAT) and classifies individuals into four categories of severity: None (0), Mild (1), Moderate (2), and Severe (3). By successfully modeling the SII, researchers can better understand the interplay between physical activity patterns, physiological measures, and internet use behaviors.

Dataset Overview

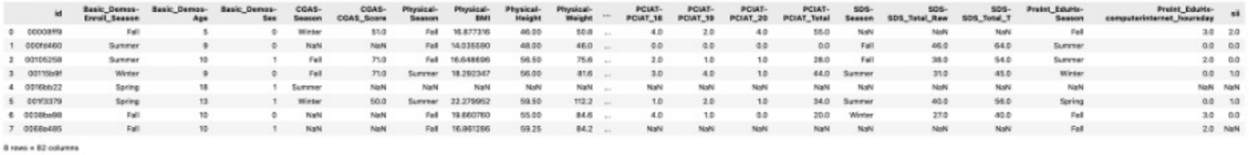
The data provided for this competition are structured into two formats, that is tabular data and time-series data. The tabular data, stored in CSV format, comprises numerous instruments and associated fields., including demographic profiles (e.g., age, sex, and enrollment season), physical measurements, fitness test results, and responses to questionnaires. It aslo has data related to internet use such as hours of daily computer or internet activity, offer insights into participants online behavior. Standardized assessments, including the Children's Global Assessment Scale (CGAS), furnish information on general functioning and mental health status. Physical measures (e.g., BMI, blood pressure), fitness test results from FitnessGram protocols, and body composition metrics obtained through Bio-Electric Impedance Analysis (BIA) contribute to a comprehensive view of each participant's physical health. Additional instruments cover subjective physical activity levels (from PAQ_A and PAQ_C questionnaires), sleep disturbances (SDS), and the PCIAT responses that inform the SII classification. Many of these instruments also record the season of data collection, potentially revealing seasonal trends or variations in health and behavior.

The time-series data is saved in parquet format, and it consists of continuous wrist-worn accelerometer recordings spanning multiple days. The actigraphy data provides a continuous stream of objective measures of participant activity and environment. Each record includes participant ID, a time index (5-second intervals), and acceleration data across three axes (X, Y, Z). Derived metrics, such as ENMO (Euclidean Norm Minus One) and anglez (arm angle), alongside contextual indicators like non-wear flags, light intensity, battery voltage, and temporal markers (time of day, day of week, quarter of the year, and days relative to the PCIAT administration), enable detailed characterization of physical activity patterns. These high-frequency measures can be integrated with the tabular data to examine potential correlations between daily activity profiles and problematic internet use severity.

The dataset is partitioned into training and test sets. The training set includes a train.csv file containing participant-level tabular data, some of which may be incomplete or have missing the SII value and corresponding series_train.parquet files holding the actigraphy data for each participant. The test set includes test.csv, where SII values are known but concealed for evaluation, along with series_test.parquet files containing the relevant participant-level actigraphy data. Additionally, the data_dictionary.csv file serves as a reference guide to all variables, their meanings, data types, and potential value ranges. Participant identifiers (id) are consistently used across the CSV and parquet files, ensuring a clear link between tabular and time-series data. Figure 2 3 4 5 shows the sample data for train.csv, test.csv, series_train.parquet and series_test.parquet files respectively.

Figure 2

Sample from train.csv

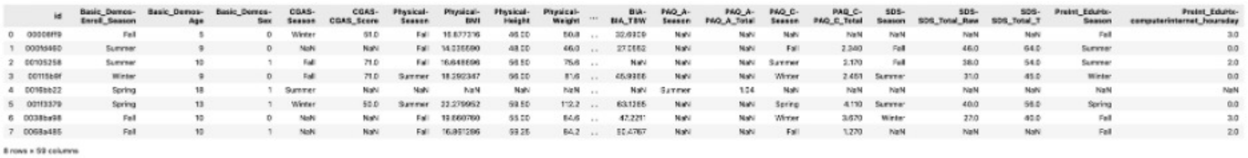


	id	Basic_Demos-Enroll_Season	Basic_Demos-Age	Basic_Demos-Sex	CGAS-Season	CGAS-CGAS_Score	Physical-Season	Physical-BMI	Physical-Height	Physical-Weight	...	PCIAT-PCIAT_18	PCIAT-PCIAT_19	PCIAT-PCIAT_20	PCIAT-PCIAT_Total	SDS-Season	SDS-SDS_Total_Raw	SDS-SDS_Total_T	PreInt_EduHx-Season	PreInt_EduHx-computerinternet_hoursday	sii
0	00008ff9	Fall	5	0	Winter	51.0	Fall	16.877316	46.00	50.8	...	4.0	2.0	4.0	55.0	NaN	NaN	NaN	Fall	3.0	2.0
1	000fd460	Summer	9	0	NaN	NaN	Fall	14.035590	48.00	46.0	...	0.0	0.0	0.0	0.0	Fall	46.0	64.0	Summer	0.0	0.0
2	00105258	Summer	10	1	Fall	71.0	Fall	16.648696	56.50	75.6	...	2.0	1.0	1.0	28.0	Fall	38.0	54.0	Summer	2.0	0.0
3	00115b9f	Winter	9	0	Fall	71.0	Summer	18.292347	56.00	81.6	...	3.0	4.0	1.0	44.0	Summer	31.0	45.0	Winter	0.0	1.0
4	0016bb22	Spring	18	1	Summer	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	001f3379	Spring	13	1	Winter	50.0	Summer	22.279952	59.50	112.2	...	1.0	2.0	1.0	34.0	Summer	40.0	56.0	Spring	0.0	1.0
6	0038ba98	Fall	10	0	NaN	NaN	Fall	19.660760	55.00	84.6	...	4.0	1.0	0.0	20.0	Winter	27.0	40.0	Fall	3.0	0.0
7	0068a485	Fall	10	1	NaN	NaN	Fall	16.861286	59.25	64.2	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Fall	2.0	NaN

8 rows × 82 columns

Figure 3

Sample from test.csv

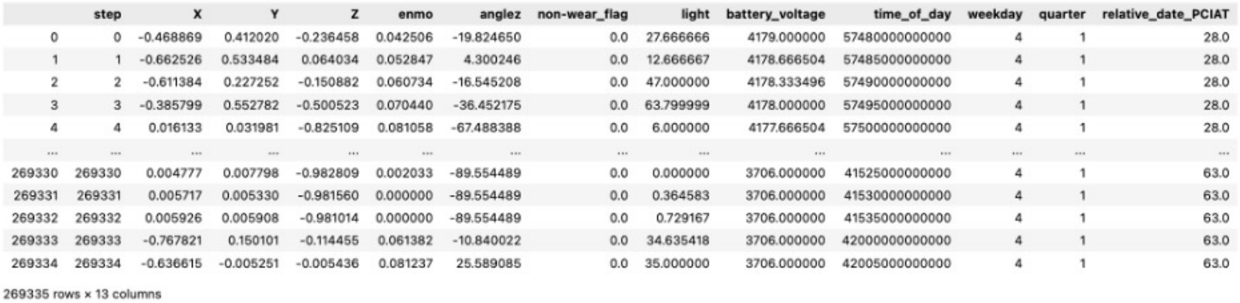


	id	Basic_Demos-Enroll_Season	Basic_Demos-Age	Basic_Demos-Sex	CGAS-Season	CGAS-CGAS_Score	Physical-Season	Physical-BMI	Physical-Height	Physical-Weight	...	BIA-BIA_TBW	PAQ_A-Season	PAQ_A-PAQ_A_Total	PAQ_C-Season	PAQ_C-PAQ_C_Total	SDS-Season	SDS-SDS_Total_Raw	SDS-SDS_Total_T	PreInt_EduHx-Season	PreInt_EduHx-computerinternet_hoursday
0	00008ff9	Fall	5	0	Winter	51.0	Fall	16.877316	46.00	50.8	...	32.6909	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Fall	3.0
1	000fd460	Summer	9	0	NaN	NaN	Fall	14.035590	48.00	46.0	...	27.0552	NaN	NaN	Fall	2.340	Fall	46.0	64.0	Summer	0.0
2	00105258	Summer	10	1	Fall	71.0	Fall	16.648696	56.50	75.6	...	NaN	NaN	NaN	Summer	2.170	Fall	38.0	54.0	Summer	2.0
3	00115b9f	Winter	9	0	Fall	71.0	Summer	18.292347	56.00	81.6	...	45.9966	NaN	NaN	Winter	2.451	Summer	31.0	45.0	Winter	0.0
4	0016bb22	Spring	18	1	Summer	NaN	NaN	NaN	NaN	NaN	...	NaN	Summer	1.04	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	001f3379	Spring	13	1	Winter	50.0	Summer	22.279952	59.50	112.2	...	63.1265	NaN	NaN	Spring	4.110	Summer	40.0	56.0	Spring	0.0
6	0038ba98	Fall	10	0	NaN	NaN	Fall	19.660760	55.00	84.6	...	47.2211	NaN	NaN	Winter	3.670	Winter	27.0	40.0	Fall	3.0
7	0068a485	Fall	10	1	NaN	NaN	Fall	16.861286	59.25	64.2	...	50.4767	NaN	NaN	Fall	1.270	NaN	NaN	NaN	Fall	2.0

8 rows × 59 columns

Figure 4

Sample from series_train.parquet

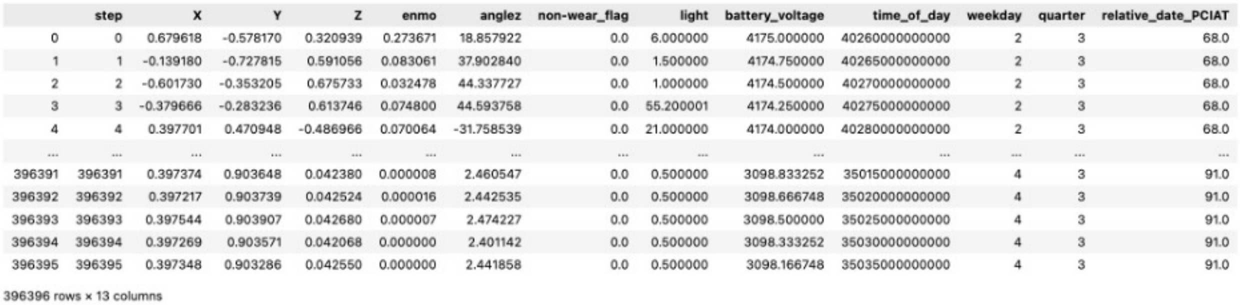


	step	X	Y	Z	enmo	anglez	non-wear_flag	light	battery_voltage	time_of_day	weekday	quarter	relative_date_PCIAT
0	0	-0.468869	0.412020	-0.236458	0.042506	-19.824650	0.0	27.666666	4179.000000	57480000000000	4	1	28.0
1	1	-0.662526	0.533484	0.064034	0.052847	4.300246	0.0	12.666667	4178.666504	57485000000000	4	1	28.0
2	2	-0.611384	0.227252	-0.150882	0.060734	-16.545208	0.0	47.000000	4178.333496	57490000000000	4	1	28.0
3	3	-0.385799	0.552782	-0.500523	0.070440	-36.452175	0.0	63.799999	4178.000000	57495000000000	4	1	28.0
4	4	0.016133	0.031981	-0.825109	0.081058	-67.488388	0.0	6.000000	4177.666504	57500000000000	4	1	28.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
269330	269330	0.004777	0.007798	-0.982809	0.002033	-89.554489	0.0	0.000000	3706.000000	41525000000000	4	1	63.0
269331	269331	0.005717	0.005330	-0.981560	0.000000	-89.554489	0.0	0.364583	3706.000000	41530000000000	4	1	63.0
269332	269332	0.005926	0.005908	-0.981014	0.000000	-89.554489	0.0	0.729167	3706.000000	41535000000000	4	1	63.0
269333	269333	-0.767821	0.150101	-0.114455	0.061382	-10.840022	0.0	34.635418	3706.000000	42000000000000	4	1	63.0
269334	269334	-0.636615	-0.005251	-0.005436	0.081237	25.589085	0.0	35.000000	3706.000000	42005000000000	4	1	63.0

269335 rows × 13 columns

Figure 5

Sample from series_test.parquet



	step	X	Y	Z	enmo	anglez	non-wear_flag	light	battery_voltage	time_of_day	weekday	quarter	relative_date_PCIAT
0	0	0.679618	-0.578170	0.320939	0.273671	18.857922	0.0	6.000000	4175.000000	40260000000000	2	3	68.0
1	1	-0.139180	-0.727815	0.591056	0.083061	37.902840	0.0	1.500000	4174.750000	40265000000000	2	3	68.0
2	2	-0.601730	-0.353205	0.675733	0.032478	44.337727	0.0	1.000000	4174.500000	40270000000000	2	3	68.0
3	3	-0.379666	-0.283236	0.613746	0.074800	44.593758	0.0	55.200001	4174.250000	40275000000000	2	3	68.0
4	4	0.397701	0.470948	-0.486966	0.070064	-31.758539	0.0	21.000000	4174.000000	40280000000000	2	3	68.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
396391	396391	0.397374	0.903648	0.042380	0.000008	2.460547	0.0	0.500000	3098.833252	35015000000000	4	3	91.0
396392	396392	0.397217	0.903739	0.042524	0.000016	2.442535	0.0	0.500000	3098.666748	35020000000000	4	3	91.0
396393	396393	0.397544	0.903907	0.042680	0.000007	2.474227	0.0	0.500000	3098.500000	35025000000000	4	3	91.0
396394	396394	0.397269	0.903571	0.042068	0.000000	2.401142	0.0	0.500000	3098.333252	35030000000000	4	3	91.0
396395	396395	0.397348	0.903286	0.042550	0.000000	2.441858	0.0	0.500000	3098.166748	35035000000000	4	3	91.0

396396 rows × 13 columns

Exploratory Data Analysis and Pre-Processing

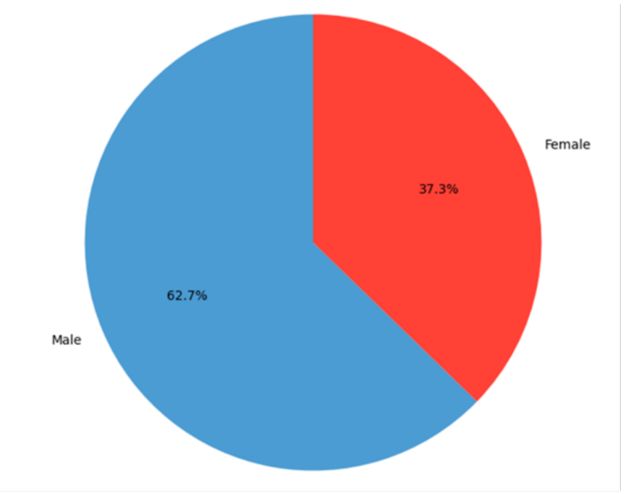
Exploratory Data Analysis

Below visualization are showing a detailed exploratory data analysis on the static survey and actigraphy dataset.

The pie chart in Figure 6 illustrates the gender distribution of participants as percentages, showing that 62.7% are male and 37.3% are female. This indicates that males constitute the majority of the participants.

Figure 6

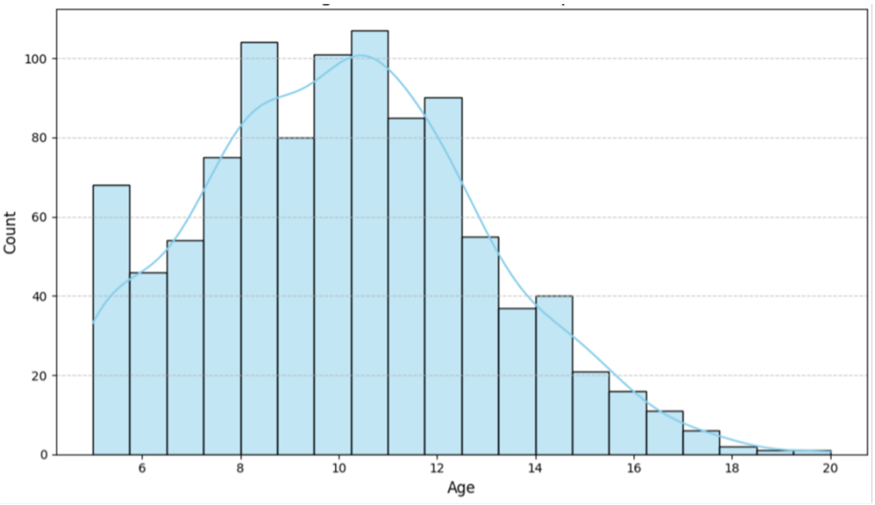
Gender distribution analysis



The chart in Figure 7illustrates the age distribution of participants and their respective counts. The participants range in age from 5 to 20 years, with the highest number of participants observed between the ages of 8 and 12. Participation gradually declines after the age of 10. The distribution is left-skewed, indicating greater participation among younger age groups.

Figure 7

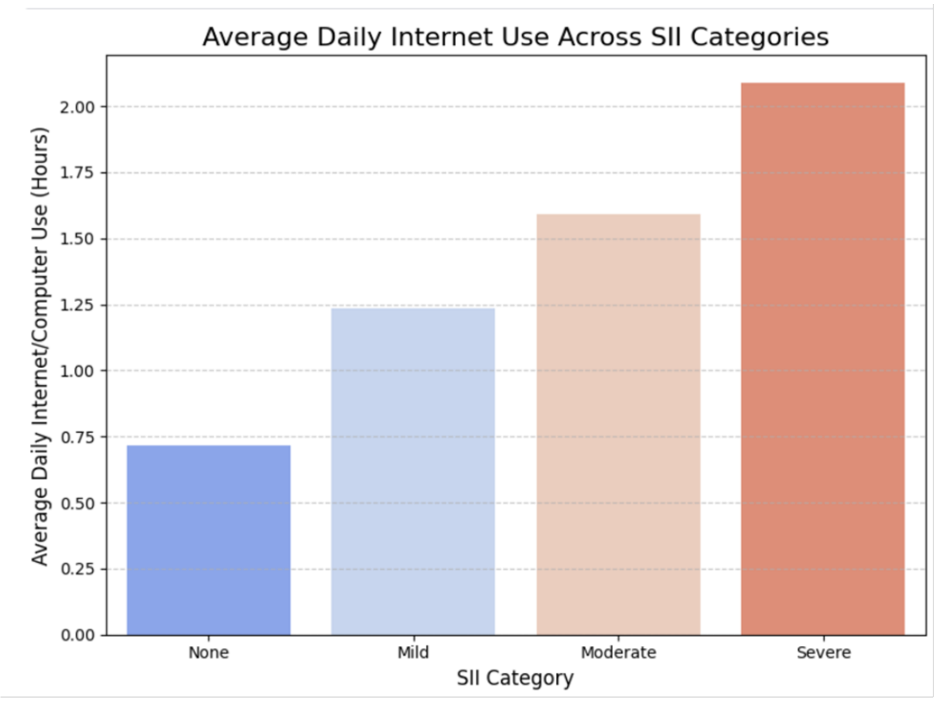
Age Distribution Among Participants



The chart in Figure 8represents the daily internet usage in hours among individuals across four SII categories: None, Mild, Moderate, and Severe. The analysis indicates that individuals with internet usage exceeding 2 hours per day tend to have a high Severity Impairment Index (SII), while those spending 1 to 2 hours are classified under moderate severity. Individuals with approximately 1.25 hours of usage are categorized as having mild SII, and those using the internet for less than 75 minutes per day fall into the "None" category, indicating no severity impairment. The data highlights a positive correlation between the severity of SII and daily internet usage, emphasizing that individuals who spend more than 2 hours online are more likely to experience greater severity impairment.

Figure 8

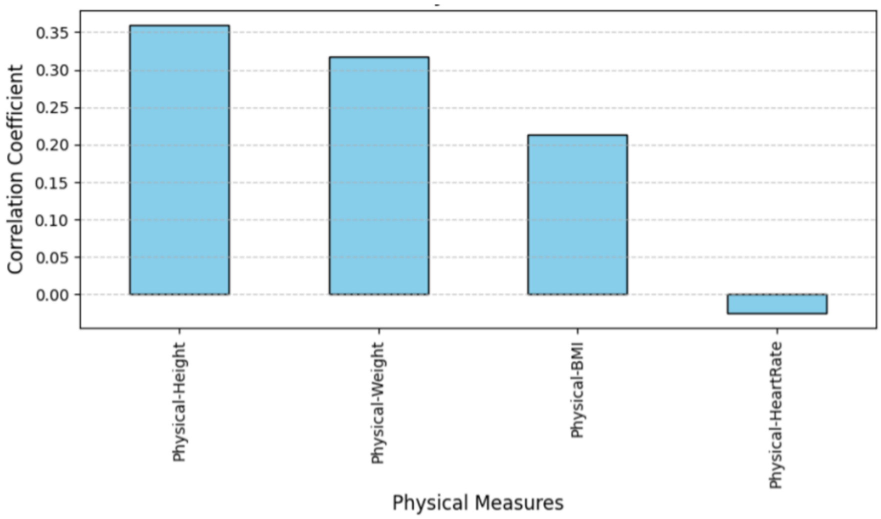
Average Daily Internet Usage Accross SII categories



The bar chart in Figure 9 shows the correlation coefficients of various physical measures compared to the Severity Impairment Index (SII). Physical height has the highest correlation at 0.35, indicating a strong positive relationship with SII. Physical weight follows closely, showing a moderate positive correlation. Physical BMI has a lower correlation of 0.20, reflecting a weaker relationship with SII. Lastly, physical heart rate has a correlation below 0.05, suggesting a negligible association. This analysis highlights that physical measures such as height and weight have a stronger connection to SII compared to other measures.

Figure 9

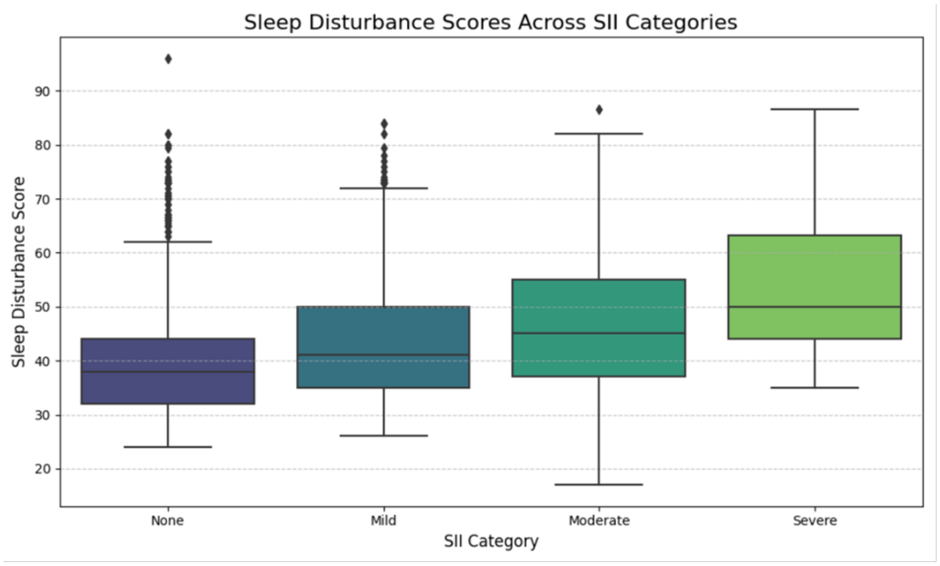
Correlation of Physical Measures with SII



The box plot in Figure 10illustrates the distribution of sleep disturbance scores across different SII categories. The median sleep disturbance score rises progressively from the "None" to the "Severe" category, with the IQR also widening, indicating greater variability in scores as severity increases. The plot highlights a positive correlation between SII severity and sleep disturbance, suggesting that individuals with higher severity impairment experience greater sleep disturbances.

Figure 10

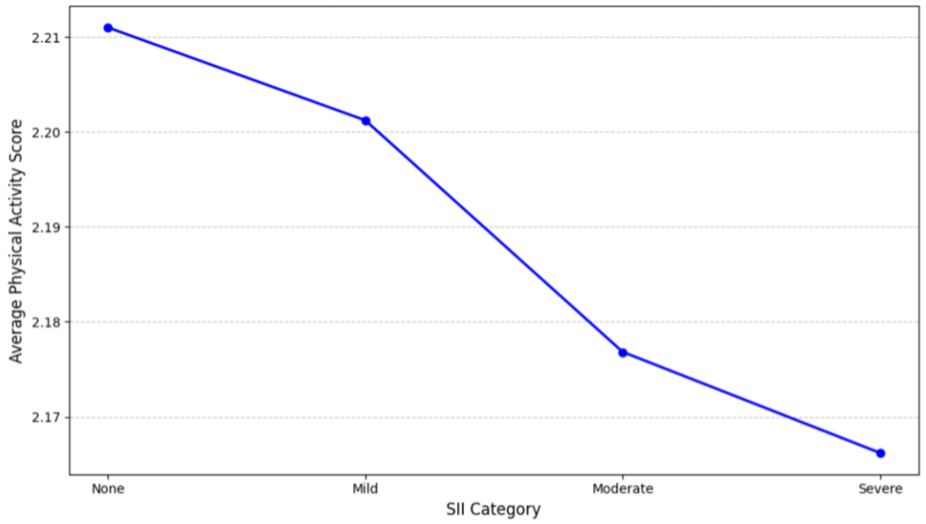
Sleep Disturbance score across SII categories



The line chart in Figure 11 depicts the average physical activity scores across different SII categories. Individuals in the "None" category have the highest score, approximately 2.21, with a gradual decline observed from Mild to Severe categories. This trend indicates that individuals without severity impairment tend to have better physical activity levels, reflecting good physical health, while those with higher severity impairment show lower activity levels, suggesting poorer physical health.

Figure 11

Physical Activity across SII categories



Data Preprocessing

Data Preparation. The dataset for this project is a combination of a parquet file containing physical activity data, including wrist-worn accelerometer readings, fitness assessments, and questionnaires, and a CSV file capturing internet usage and behavioral information. These datasets have been carefully merged on a common ‘id’ column to provide richer and more meaningful insights for prediction.

Handling Missing Values. To address missing values in both the training and testing datasets, features were separated into numerical and categorical groups to ensure appropriate imputation methods were applied. Missing numerical values were handled using the KNN imputer, which replaces them based on the values of the K nearest neighbors. For categorical features, missing values were substituted with the placeholder "unknown" and converted to the categorical data type for uniformity.

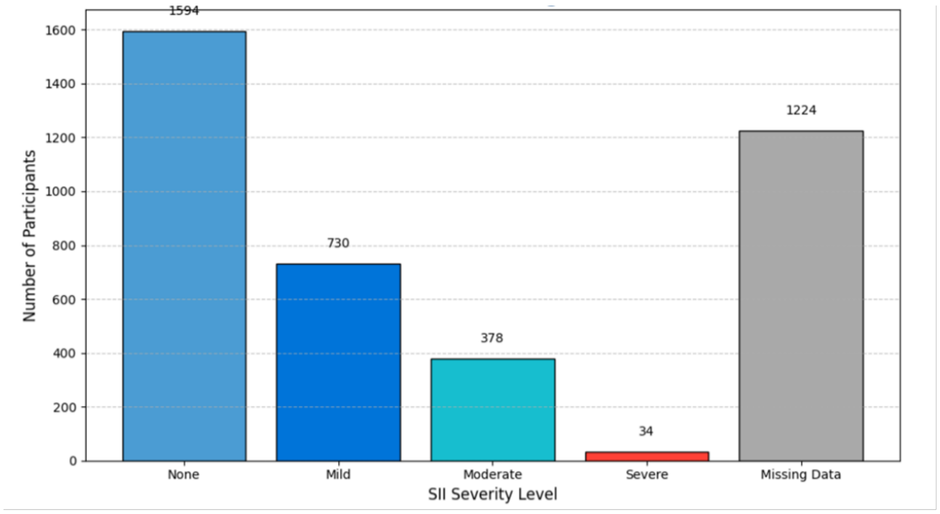
Feature Engineering. Descriptive statistics have been calculated for features from each parquet file and flattened into a single row to create an aggregated structure. Categorical data has been transformed into category type and encoded numerically using a custom mapping function, assigning a unique integer to each category. Separate mappings are applied to the train and test sets to prevent data leakage.

Feature Scaling. After addressing class imbalance, numerical features were scaled using the StandardScaler method, which normalizes features to have a mean of 0 and a standard deviation of 1. Models like TabNet are highly sensitive to the scale of feature values, so this scaling process improves training stability and accelerates convergence.

Handling class distribution. As illustrated in the Figure 12, the target variable in the dataset was imbalanced. To address this, the Borderline-SMOTE technique was employed. This advanced variant of the Synthetic Minority Oversampling Technique focuses on minority samples located near the decision boundary with the majority class, generating synthetic samples in these regions. This approach enhances the model's ability to classify the minority class more effectively.

Figure 12

Distribution of SII Target Variable



Problem Formulation

Technology has been growing at an exponential pace and has become an integral part of our daily lives. People of all ages are attuned to rapid advancements in technology and Internet has played a pivotal role. As digital technology with continuous use of internet becomes ingrained in our lives, it is extremely important to understand how problematic and compulsive internet usage has negatively impacted the physical health of people.

The Healthy Brain Network (HBN) has presented a comprehensive dataset with clinical samples for about 5000 volunteers aged 5-22 years who have undergone clinical and research screenings. The research aims to explore nuanced machine learning and deep learning methodologies to accurately predict the Severity Impairment Index(sii). SII is a standard measure of problematic internet usage, and it has four distinct level None, Mild, Moderate and Severe. Identifying biological markers is the end goal of this study then can help with advancements in the triage and treatment of various mental and social disorders.

Prediction of SII is an interesting task that requires us to understand the deep relationships between digital engagement and behavioral analysis. When we combine this with metrics on physical activity, we have a challenge to understand the complex relationships. Relationships between these intertwined metrics and how they contribute to the overall impairment index will help us understand how digital engagement is becoming problematic among adolescents and young adults. The end goal is to create a predictive model that can effectively transform multi-dimensional inputs into a multi-class classification.

Dataset comprises of demographic information, physical movement metrics, and metrics from clinical tests. Dataset also provides time series data from an accelerometer to track physical movements. The accelerometer data is from a wrist worn device and provides metrics for movements across, X, Y and Z axes. The primary metric SII measures the increasing levels of severity for the metal or social impairment.

Development of predictive models with the available dataset will require advanced data imputation and preprocessing techniques as there are a lot of missing elements in the dataset. The training set has missing target feature for almost 25% of the samples and the dataset columns are not consistent between test and train datasets. Clear information about each measure has been provided as part of the data dictionary and can be utilized to understand the overall structure of the data. The challenging part with this dataset would be the bridge the gap between clinical data and accelerometer data as the data is presented in different dimensions and time ranges.

Combining the data together from normal csv files and parquet files would require a deep understanding of the data and advanced techniques. Another important aspect to note is the feature “PCIAT-PCIAT_Total” which is available in training set but missing in test set. The target feature SII has been derived from this field as per the dataset information. However, as this is missing in test set, we need to exclude this and identify other features than can help us derive the target variable.

This necessitates utilizing advanced machine and deep learning models to understand the complex relationships between different metrics. The solution should be able to generalize well across a wide range of age groups while accounting for different development stages at each age.

The outcome of this study has a multi-layered significance in development of sophisticated and state of the art tools, that can help with early identification and intervention for mental and social disorders. This of great significance as digital engagement is here to stay and understanding the negative impact it is having on young and developing minds is the need of the hour. This will not only help prevent various development disorders in the young population, it will also help us to be better prepared for handling various emotional and physio-social challenges.

Model Selection

Implementation of Multi-Layer Perceptron (MLP)

Model Description. The main goal of Multi-Layer Perceptron (MLP) model is to predict the ordinal class labels which are applied to a multi-class classification issue. This feedforward neural network architecture uses several fully connected layers to process input information. The output layer generates class probabilities using a Softmax activation function, while each hidden layer introduces non-linearity using the ReLU activation function. This approach guarantees that the model can produce probabilistic outputs for multi-class classification and manage complex relationships in the data.

Model Architecture. Two hidden layers, an output layer, and an input layer that takes in numerical features from the dataset make up the MLP model. Each of the 64 neurons in the fully connected hidden layers has ReLU activation, which aids in identifying complex patterns in the data. The Softmax activation guarantees that the outputs reflect class probabilities, and the final output layer contains neurons equal to the number of target classes. Model complexity and computing efficiency are balanced in this simple yet powerful architecture.

Implementation. PyTorch is used to implement the MLP, and its forward pass and structure are defined by the MLPModel class. The dataset is preprocessed by utilizing Borderline-SMOTE for oversampling to correct class imbalance and StandardScaler to scale numerical characteristics. An 80-20 ratio is used to divide the data into training and validation sets. Managing class imbalance and guaranteeing uniform feature scaling for reliable model input are part of the training process. To increase generalization, the ensemble method trains five separate models, each of which starts with a unique set of random seeds.

Training Configuration. By reducing the log-loss between the predicted and true class labels, CrossEntropyLoss optimizes the model for multi-class classification during the training phase. With a learning rate of 0.001, the Adam optimizer is used for effective gradient updates. To reduce the loss, the model iteratively updates weights throughout 50 epochs of training in batches of size 32. To further improve prediction robustness and reliability, a group of five models is trained. To ensure diversity in learnt representations, the ensemble uses a weighted voting method to aggregate predictions. The CrossEntropyLoss, which calculates the discrepancy between true labels and predicted class probabilities, is essential to training the MLP model. The model modifies its parameters by using this loss in order to increase the probability of accurate predictions. This loss function guarantees that the model produces calibrated probabilities for each class, making it especially appropriate for multi-class classification problems.

Evaluation. Several metrics, such as Mean Squared Error (MSE), Precision, Recall, and F1 Score (weighted averages), are used to assess the model's performance on the validation set. In order to account for the ordinal character of the classes, the agreement between predictions and true labels is also measured using Quadratic Weighted Kappa (QWK).

TabNet Ensemble Classifier

Model Description. The TabNet Classifier, built using PyTorch, leverages sequential attention to dynamically select the most relevant features at each decision step. This mechanism enhances interpretability and enables efficient learning by focusing on the most important features. Designed specifically for tabular data, TabNet has been shown to outperform traditional methods on such datasets Arik and Pfister, 2019.

Model Architecture.

- n_d: Dimension of the prediction layer
- n_a: Dimension of the attention layer
- n_steps: Number of successive steps in the network
- gamma: A scaling factor (greater than 1) for attention updates
- lambda_sparse: Regularization parameter to enforce feature sparsity
- mask_type: Masking function used ("entmax")
- optimizer_fn: Optimizer used for training (AdamW) Falbel et al., 2024

Training Configuration. Five TabNet models were trained using different random seeds. Each model was trained for a maximum of 200 epochs with early stopping (patience of 20 epochs) to prevent overfitting. A batch size of 1024 and a virtual batch size of 128 were used during training.

Prediction and Evaluation. Ensemble predictions were generated for the test set by averaging the outputs of all trained models to produce the final predictions. The model's performance was evaluated using metrics such as Mean Squared Error (MSE), Accuracy, Precision, Recall, F1 Score, and Quadratic Weighted Kappa.

Custom Implementation of a TabNet Regressor

Model Description. The other model used is TabNet with custom implementation of TabNet wrapper, a deep learning network specifically designed for tabular data. TabNet incorporates neural network expressiveness with decision tree interpretability, focusing on salient feature interactions using an attention mechanism. In such a way, it handles structured data effectively while retaining the ability for interpretability. A custom TabNet wrapper provides further extensions to this model by including preprocessing, cross-validation, and optimized training, thus making it appropriate for robust and scalable modeling tasks.

Model Architecture.

- Shared Feature Layers: Process the input features through shared layers to generate general embeddings for downstream tasks.
- Sequential Decision Steps: Multiple stages of decision-making are performed with an attention mechanism that dynamically selects the most relevant features.
- Sparse Feature Selection: Sparsity loss is applied to encourage the use of only essential features, hence enhancing interpretability.
- Encoder-Decoder Framework: Representation learning for tasks such as dimensionality reduction by reconstruction is supported.

Implementation. Decision steps (n_steps) Configured to test multiple values like 3, 5, and 7. Feature embedding dimensions, n_d and n_a tried were 32, 64, 128, among others. Gamma (gamma) for attention weighting was selected as 1.5 and 2.0.

Training Configuration.

- Pre-processing: Imputation: Missing values were handled using KNN and median imputation.
- Feature Engineering: Inclusion of interaction-based features, namely BMI-related ratios and hydration status.

- Dimensionality Reduction: Autoencoders compress high-dimensional time-series features into compact representations.
- Normalization: StandardScaler was applied to normalize numerical features.
- Cross-Validation: 5-fold stratified cross-validation was used to ensure that the evaluation is well-balanced among the target classes.

Optimization.

- Loss Function: MSE was used for this model.
- Optimizer: Adam optimizer with a learning rate of 0.02 and weight decay of 1×10^{-5}.
- Scheduler: ReduceLROnPlateau dynamically adjusted the learning rate based on validation performance
- Early Stopping: Training stopped after 20 epochs of no improvement in validation loss.
- Hardware: Training performed on GPU and CPU; GPU significantly accelerated performance and produced better results.
- Prediction: Predictions across cross-validation folds were averaged to improve robustness.
- Threshold optimization mapped continuous outputs to discrete classes, improving class separation.

Prediction and Evaluation. Predictions were generated by averaging the results across all cross-validation folds. For multi-class targets (0, 1, 2, 3), threshold optimization was employed using the Nelder-Mead method to find the best thresholds for mapping predictions to discrete classes. Quadratic Weighted Kappa (QWK):

- QWK was used as the primary evaluation metric, ensuring the model's predictions align closely with the ground truth on a weighted scale.

- Threshold optimization improved the QWK score by finding the best class mapping for continuous predictions.

Cross-validation results:

- Mean Train QWK: Evaluated model fit on training data.
- Mean Validation QWK: Assessed generalization on unseen data.

Threshold Analysis:

Thresholds were iteratively optimized for each hyperparameter configuration to test the model's ability to predict binary classes effectively.

Implementation of Long Short-Term Memory (LSTM)

Model Description. LSTM is a type of RNN (Recurrent Neural Network) that can handle sequential data with long range dependencies. It is an advanced neural network that uses "gates" to capture short- and long-term memory. The gates in the LSTM architecture are named as "Forget", "Input" and "Output" and it helps to prevent issues like exploding / vanishing gradients that is common with standard RNNs. It is able to retain and process information over multiple time steps. LSTMs can be effectively employed with use cases involving NLP and time series analysis. LSTM was introduced in 1997 by Sepp Hochreiter and Jürgen Schmidhuber with an aim to overcome issues with traditional neural networks.

Model Architecture.

- LSTM layer: Processes sequential data as per provided input size, hidden units and stacked layers
- Input: batch size, seq_length, input_size
- Output: batch_size, seq_length, hidden_size
- Dropout Layer: Prevents overfitting by performing regularization by dropping a fraction of activations

- Fully Connected Layer: This is the layer used for prediction. It maps the LSTM output to the provided output size, ie. Number of output classes.
- Input: batch_size, hidden_size
- Output: batch_size, output_size

Implementation. Pytorch was used to implement as baseline LSTM model with 151 as the input size, 2 LSTM layers, hidden size of 128 and output size of 4. Default LSTM configuration that uses tanh as activation function for cell states and sigmoid for gates. Prior to training, the data was preprocessed using different techniques like normalization, imputation and dimensionality reduction using T-SNE. One-hot encoding was performed for categorical columns and parquet data was loaded and reduced using T-SNE.

Training Configuration. The batch sizes were 32, 64, 128 with epochs as 10, 20, 50, 100 with early stopping with a threshold of 5,10, 15 and finally with no early stopping. CrossEntropyLoss was used as the loss function and Adam optimizer with learning rates of 0.02, 0.01 and 0.0001. The final pre-processed data was split into training and validation sets with 0.2 test size and random seed as 42 and stratified on the target variable. SMOTE was used to prevent class imbalance by generating synthetic samples to prevent overfitting on the majority class. The preprocessing and training were performed primarily on CPU as the data volume was low and it did not justify usage of GPU for this use case.

Prediction and Evaluation. Metrics like Accuracy and Quadratic Weighted Kappa (QWK) were used to review the results and measure the model's performance on validation dataset. The model showed decent performance on training dataset with a mean QWK of 0.3150, however, it struggled with validation set primarily in the absence of metrics that were missing between the training and test datasets. Due to absence of data, the metrics specific to "Parent-Child Internet Addiction Test" were removed from the training dataset during the final feature reduction, this caused a significant drop in the overall accuracy for the validation set. The model was able to generate a QWK score of 0.3283 on the validation set, while it generated a score of QWK of 0.31 on the unseen test dataset on the Kaggle competition. The training was early stopped at 37 epochs as the validation loss was rising.

Implementation of Siamese Long Short-Term Memory (LSTM)

Model Description. Since the dataset consists of both static features and time series features for a subset of participants, an approach is required to process these 2 features together using a single Neural Network architecture. A Multi-layer Perceptron or MLP is a well-proven approach to model tabular data, while Long Short-Term Memory or LSTM is known to work well in modeling time-series data. The tabular data is in the shape (num_samples, num_features), where num_samples is the total number of participants in the study who have a valid SII score, and num_features are the cleaned, imputed features obtained from the data cleaning process. Similarly, the time series data is in shape (num_samples, time_steps, num_features). The sequence dimension, time_step, is present in the dataset with a data point every 5 seconds for around 30 days for each participant who decided to wear the device for recording actigraphy data. Since the time_steps in this case is very big, the data points are aggregated to a day using mean, standard deviation, minimum and maximum functions. For participants who didn't have corresponding time-series data point, a dummy datapoint filled with zeros of shape (time_steps, num_features) was used. This ensured that there is a time series component for every participant in the study, and both the data formats could be given as input to the model. The LSTM model then processes the time series data, before collapsing the time_step dimension using Global Max-Pooling operation. The pooled time series features are then merged with the static features inside the model.

Model Architecture. The Siamese LSTM network takes as input 2 batches of data –

- Static features of the shape (batch_size, num_static_features)
- Time series features of the shape (batch_size, time_steps, num_ts_features)

The architecture consists of the following layers.

- embed_layer and embed_relu – A PyTorch Linear layer with input features of size 38 (number of static features after cleaning) and output features set to 128, which is followed by ReLU activation.

- lstm_layer – A LSTM block with 2 LSTM layers, which takes as input features of shape 38 (number of time series features), and outputs features of size 64. The dropout between the LSTM layers is set to 0.2, and due to its bidirectional nature, the total number of features output by the LSTM block is 64 * 2 = 128.

- pool_layer – This layer applies a Global Max Pooling operation on the output from LSTM layer to collapse the time_steps (dim=1) dimension, resulting in the data shape being (batch_size, 128).

- concat_layer – This layer combines the output from embed_relu and pool_layer along the feature dimension (dim=1), resulting in the data shape being (batch_size, 256).

- hidden_layer, hidden_relu and dropout_layer – This layer transforms the features in data from 256 to 64, applies a ReLU activation, followed by a Dropout layer with rate set to 0.2.

- class_layer – This layer takes as input data of shape (batch_size, 64) and outputs the logits for each class. Hence the data shape at the output of this layer is (batch_size, 4).

- The time series features are padded with zeros along the time_steps dimension to ensure the shape of entire batch remains consistent for efficient tensor processing.

Training Configuration. To train the model, 5-fold cross validation was used to train 5 models with the same initial weights and get an unbiased estimate of the model performance. These 5 models were hence each trained with a different subset of the dataset and evaluated on a different subset of the dataset. The following hyperparameters were used to train these models

- Loss Function – CrossEntropyLoss with the following class weights – [0.4292, 0.9360, 1.8123, 20.0956] to ensure the class imbalance in the dataset is taken into consideration.

- Seed – 42, used everywhere to ensure the models are all initialized in the same way.
- Epochs – Each model was trained for max of 2000 epochs.
- Patience – Each model was Early Stopped with a patience of 100 epochs while monitoring the validation loss.
- Batch Size – Each model was trained with a batch size of 64.
- Optimizer – AdamW with default learning rate of 0.001 and weight_decay of 0.01.

Prediction and Evaluation. To make predictions on the test dataset, the following approach was utilized.

- Feed the test dataset to each of the 5 Siamese LSTM models trained above.
- Each model makes class predictions for each sample in the test dataset
- Take a majority vote across the 5 class predictions for each sample in the test dataset to get the final class predictions.
- Save the predicted SII to submission.csv along with the participant IDs. The submission.csv is then used by the Kaggle competition to calculate the quadratic weighted kappa (QWK score), which provides an estimate on the agreement between possible outcomes.

Ensemble of Voting Regressor with TabNet

Model Description. The model is designed using the ensemble of different models under the Voting Regressor to improve predictions on the target variable. It makes use of the strengths provided by each of LightGBM, XGBoost, CatBoost, and TabNet regressors fine-tuned with specific hyperparameters. A wide range of data pre-processing-form feature engineering to encoding for categorical and numerical variables-kickstarts the architecture. Also, dimensionality reduction is carried out on time series data through an autoencoder, compacting the meaningful features in a smaller dimensional space. A voting regressor then sums these different models with optimized weights. Further assessments include quadratic weighted Kappa among others and optimized threshold performance to achieve exact classifications. Ensemble takes advantage of the varied predictions across these models to allow significant enhancement in predictive capabilities of one model over another in improving its predictive powers.

Model Architecture. Below explaination shows the architecture used to design the voting regressor model with ensemble and TabNet.

- Time-series features extracted using an autoencoder for dimensionality reduction.
- Base Regressor – XGBoost, LightGBM, Catboost and TabNet
- Ensemble with Voting Regressor - Combines predictions from these models and creates a weighted voting mechanism to aggregate predictions for final output.
- Optimization - Threshold optimization and cross validation to enhance classification accuracy.
- Feature Importance and Interpretability – Contributing features analyzed using tree-based regressor outputs.
- Evaluation - Metrics like Quadratic Weighted Kappa (QWK) and Residual analysis and confidence interval estimation are calculated for predictions.

Training Configuration.

- For Tree based regressor-
  - learning_rate – set to 0.05 for stable learning rate.
  - balancing complexity and overfitting.
  - n_estimators – set to 200 for preventing overfitting and maintaining model complexity.
  - tree_method – gpu_hist is used for faster computation.
- For TabNet Model-

- n_d (decision layer width) and n_a (attention layer width – set to 64 to balance performance
- n_steps - Decision steps are set to 5 in the network.
- gamma: Set to 1.5 for sparsity regularization in feature selection
- lambda_sparse: It is set to 1e-4 to use sparse features
- optimizer_params - Learning rate of 0.02 with weight decay of 1e-5

Prediction and Evaluation. The best QWK score, as calculated in this notebook, is 0.494. This has been optimized by threshold optimization on the out-of-fold prediction by trying to match the predicted and actual values of sii in the dataset.

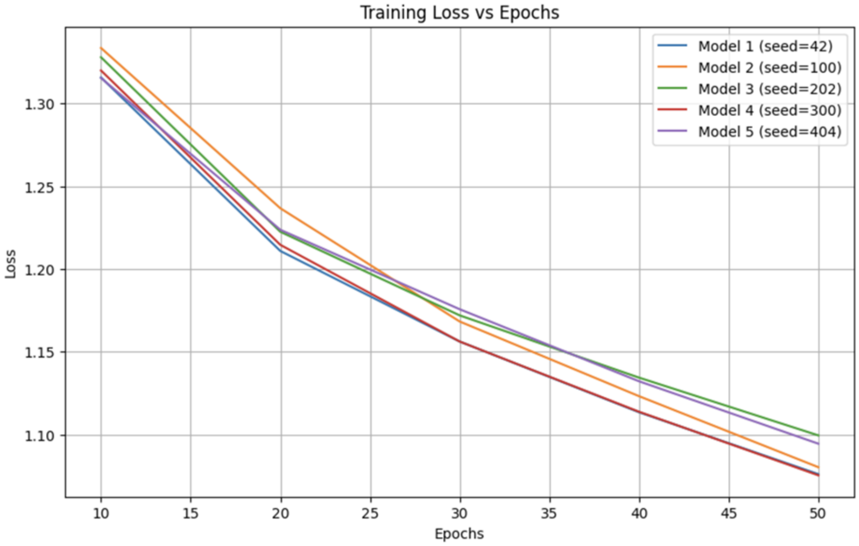
Results, Analysis and Visualization

Multi-Layer Perceptron (MLP). Metrics like MSE, Precision, Recall, and F1 Score, which assess the model's predictive performance on the validation set, show how accurate and consistent the model is. The model efficiently reduces prediction errors, as evidenced by its validation MSE of 0.9398. Weighted Precision, Recall, and F1 Scores of 0.5688, 0.5255, and 0.5393 correspondingly demonstrate its balanced performance. Furthermore, the model's sensitivity to ordinal correlations between class labels is highlighted by its Quadratic Weighted Kappa (QWK) score of 0.3694, underscoring its applicability to the issue. These outcomes show that the model can manage multi-class classification tasks while taking the target labels' ordinal character into consideration.

The training loss for five models of Multi-Layer Perceptron (MLP) was started with various random seeds plotted across 50 epochs shown in figure 13. The training loss is shown on the y-axis, and the epochs are represented on the x-axis. Effective learning is demonstrated by the steady decrease in loss seen in all models. Due to variances in parameter initialization, there are some variations in loss values among models; nonetheless, by epoch 50, all models converge to comparable loss levels. While Model 2 (seed=100) begins with the largest loss and converges a little more slowly at first, Model 4 (seed=300) attains the lowest loss. All models show consistent learning and convergence, indicating a solid training process overall.

Figure 13

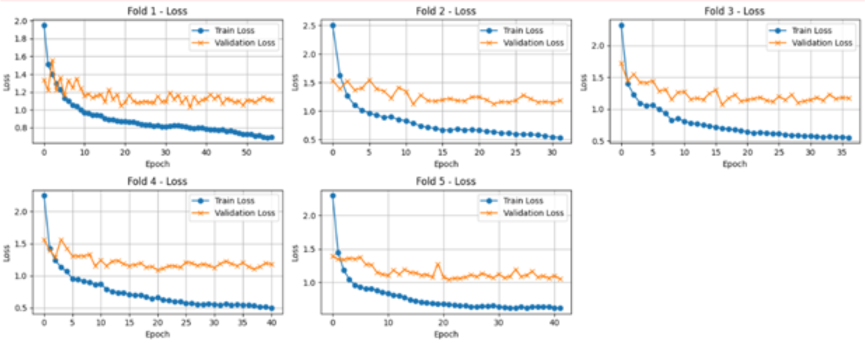
Training Loss Across Epochs for Different Model Seeds



TabNet Ensemble Classifier. The performance of the TabNet model was evaluated using multiple metrics, averaged across five TabNet models. The model achieved a QWK score of 0.42, reflecting a moderate level of agreement between the predicted and actual classes. A validation MSE of 0.60 indicates a lower error rate in predictions. The recall and precision scores, both at 58%, demonstrate the model's capability to accurately identify positive instances. Additionally, the F1 score of 57% highlights the model's ability to maintain a good balance between precision and recall. Figure 14 depicts the training and validation process for each TabNet Classifier. The training loss exhibits a steady and rapid decline across all folds, maintaining a smooth trajectory. Similarly, the validation loss decreases alongside the training loss, with minor fluctuations before stabilizing after a few epochs, aided by the implementation of early stopping to prevent overfitting. The uniformity observed across all folds highlights the model's reliability in generating consistent predictions.

Figure 14

Training Loss Across Epochs for Different Model Seeds



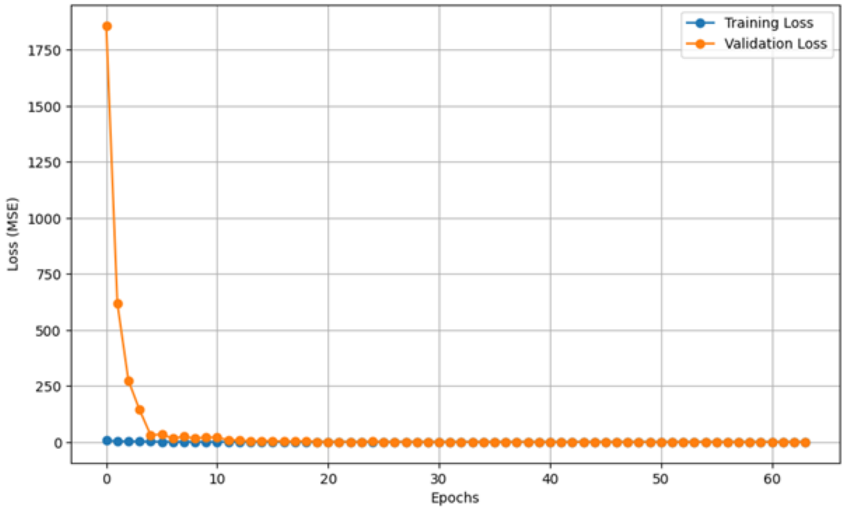
Custom Implementation of a TabNet Regressor. The model achieved a Mean Train QWK of 0.5491, indicating moderate alignment between the predictions and ground truth on the training data. However, the Mean Validation QWK of 0.4469 highlights a drop in generalization performance, suggesting room for improvement in the model's ability to predict unseen data accurately. After threshold optimization, the Optimized QWK Score improved to 0.497, demonstrating the effectiveness of fine-tuning decision thresholds to enhance prediction quality. While the results are promising, further efforts in hyperparameter tuning, feature engineering, and addressing class imbalances could improve performance.

The training and validation loss curves demonstrate the performance of the model over multiple epochs as shown in figure 15. At the initial phase both the training and validation losses decrease rapidly during the first few epochs, highlighting effective initial learning by the model. After a few epochs, the loss curves converge and stabilize, indicating that the model has reached optimal learning without overfitting. The validation loss closely follows the training loss, showcasing good generalization to unseen data.

Performance metrics like the training loss values approach close to zero, showing that the model can fit the training data well and low and stable validation loss indicates that the model is not overfitting and will be strong enough to handle unseen data. Quadratic Weighted Kappa (QWK) metric further confirmed the model's performance, with a high QWK score indicating effective prediction capabilities.

Figure 15

Visualization of training and validation loss



Long Short-Term Memory (LSTM).. The model achieved a mean QWK score of 0.3150 on the training dataset, while on validation dataset it achieved a score of 0.3283. As is visible from figure 16 below, the model's performance started to plateau around 20th epoch, and it started to overfit from 25th epoch onward as is visible from the plot. The learning rate was reduced to prevent early overfitting and also weight decay was added, however, it did not help. This caused early stopping to trigger around epoch 30 as we were training with a patience of 10. There is continuous improvement in the QWK score for the validation dataset as the epochs progressed as seen in figure 18. This suggests that additional preprocessing or feature reduction might be needed to get better performance. The accuracy for the training dataset was rising continuously till the 30th epoch suggesting that the model had started to learn the patterns of the training dataset instead of being able to generalize on unseen data.

Another important point to note is that critical data had to be removed due to mismatch between test and training datasets which negatively impacted performance. To be able to predict unseen data, the size of the training and test sets should match, however, if we retained the datapoints it would have caused an issue during prediction with unseen data. This suggests additional information about how to generate the missing metrics for the test dataset would help here as they can be a crucial differentiator.

Figure 16

Visualization of training and validation loss over Epochs

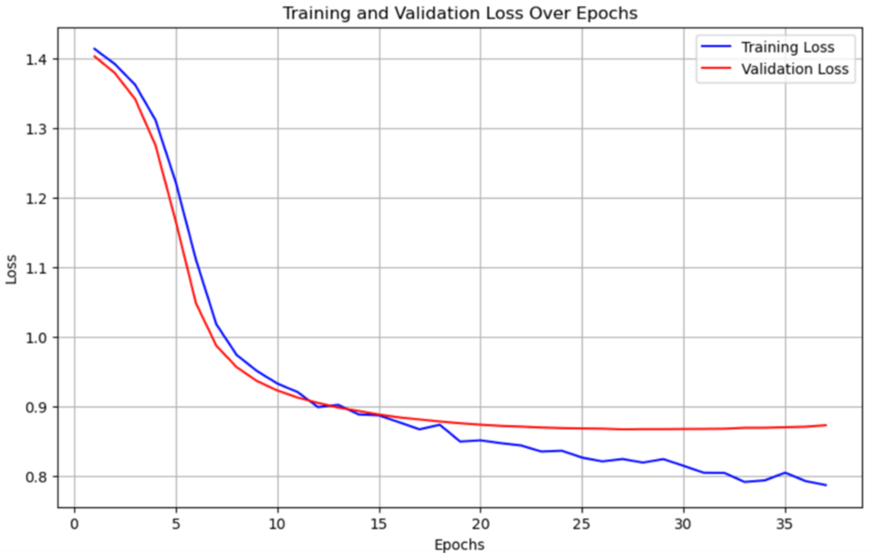


Figure 17

Visualization of training and validation Accuracy over Epochs

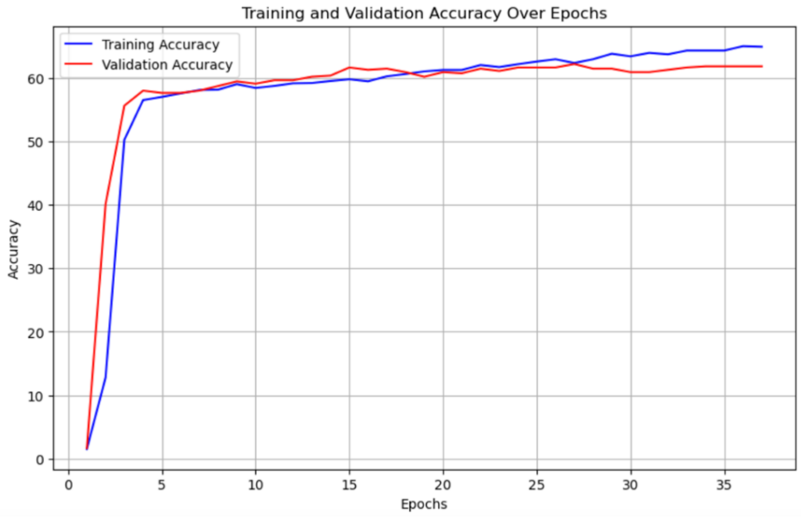
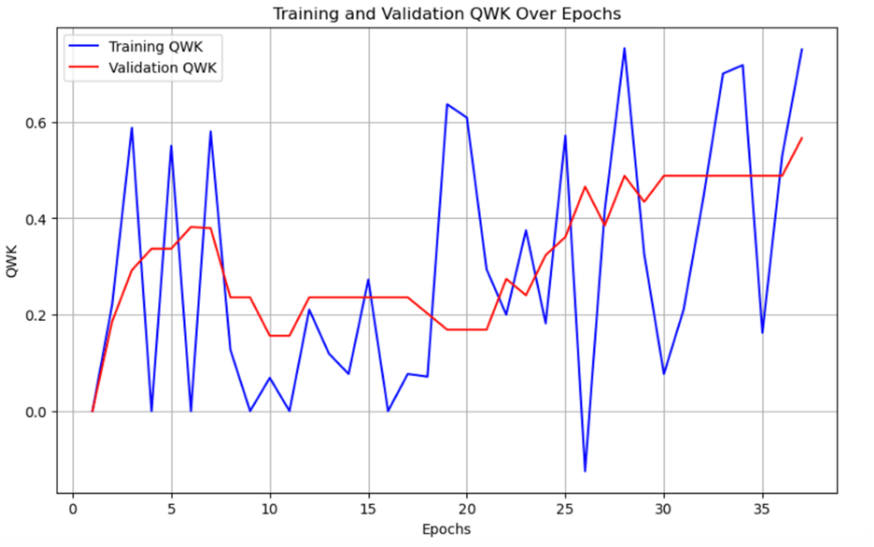


Figure 18

Visualization of training and validation QWK over Epochs



Siamese Long Short-Term Memory (LSTM).. The predictions on the train dataset made by Siamese LSTM Network models achieved a QWK score of 0.297, while test dataset achieved a QWK score of 0.327. Shown below figure 19 are the progress of loss and accuracy for training and validation splits across the epochs for each fold of the dataset. As we can see, the model started overfitting quite quickly, and the validation loss kept increasing after a few epochs, while the validation accuracy remained almost constant through the epochs. Overall, the 5 models combined achieved an unbiased estimated accuracy of 0.48 on the training partition of the dataset.

Figure 19

Train and Validation loss at each fold

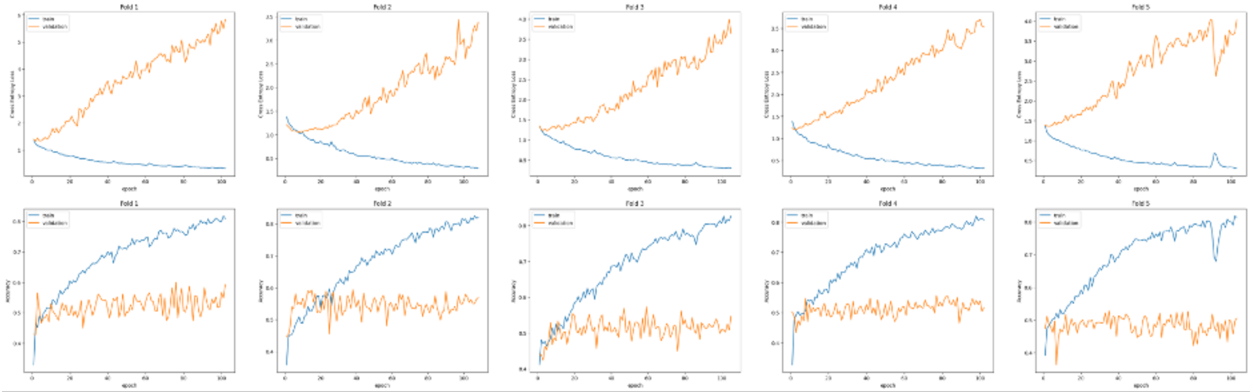
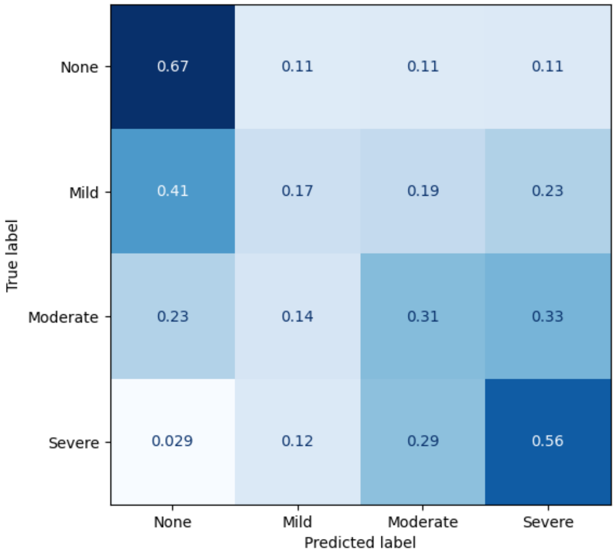


Figure 20

Confusion Matrix with respect to the True Labels in the Train dataset



	None	Mild	Moderate	Severe
None	0.67	0.11	0.11	0.11
Mild	0.41	0.17	0.19	0.23
Moderate	0.23	0.14	0.31	0.33
Severe	0.029	0.12	0.29	0.56

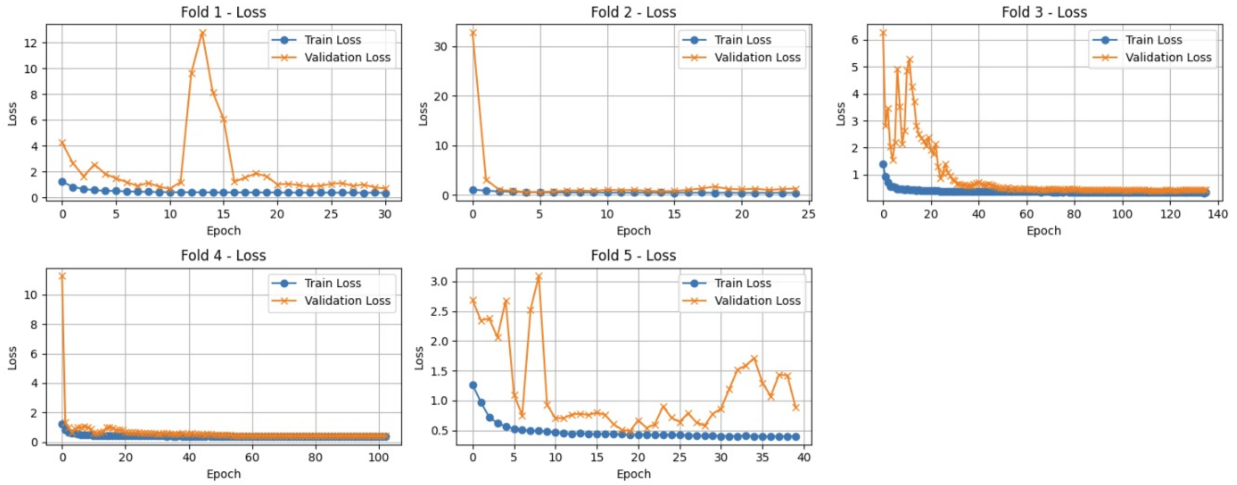
Since the dataset has a large class imbalance, it is important to ensure that the model has a good recall for the minority classes. Hence, the unbiased predictions of the train dataset labels made by the Siamese LSTM network model were checked for performance using a confusion matrix weighted with respect to the true labels. From the confusion matrix in figure xx, the model achieved a good recall on the None class, correctly identifying 67% of the samples belonging to the None class. The models then performed the best on Severe class, successfully recalling 56%

of the samples belonging to this class. From the confusion matrix, the models together were having a hard time predicting the Mild and Moderate classes. The models predicted 33% of the Moderate samples as Severe, while 41% of the Mild samples were predicted as None. Hence it can be expected that the models together are making the most mistakes when the samples belong to Mild and Moderate classes.

Ensemble of Voting Regressor with TabNet. This plot in Figure 21 shows the training versus validation loss curves obtained in the course of training the model using the methodology of 5-fold cross-validation. Each subplot corresponds to one fold and, therefore, plots the evolution of the loss values across epochs. Within each subplot, the blue line represents the training loss, and the orange line represents the validation loss. Within most folds, the training losses go down, which signifies effective learning. The validation loss has mostly gone down, with incidents of overfitting or instability during an epoch. This underlines fold-to-fold variation in the distribution of fold data-for example, folds like Fold 2 managed to stabilize earlier, while others did not. Overall, the model has converged, with the validation loss for the most part tracking the training loss toward the end of most folds.

Figure 21

Train and Validation loss at each fold



Result Analysis

The Table 1 below shows the key metrics calculated for the various models: Quadratic Weighted Kappa (QWK) Score, which calculates the agreement between the predicted vs. actual rankings, and the Kaggle Submission Score indicative of the leaderboard performance on unseen test data. Among the presented models, TabNet Voting Regressor Ensemble has the highest QWK of 0.450 and the highest Kaggle score of 0.494, reflecting its best predictive alignment and robustness. The performance of the TabNet Wrapper is also good in terms of QWK at 0.477 but with a relatively low Kaggle score of 0.303. The sequential models-LSTM and Siamese LSTM-underperform in QWK but generally bring along a middle performance in Kaggle scores, placing Siamese LSTM at 0.327. Overall, the ensemble TabNet Voting Regressor shows that this effective combination of several models within one ensemble outperforms every other approach.

Table 1

Model Evaluation Comparison

Model Name	**TabNet Ensemble Classifier**	**TabNet wrapper**	**LSTM**	**MLP**	**Siamese LSTM**	**TabNet Voting Regressor Ensemble**
QWK (Quadratic Weighted Kappa) Score	0.4271	0.477	0.3283	0.3694	0.297	0.450
Kaggle Submission Score	0.391	0.303	0.310	0.316	0.327	0.494

Conclusion

This project demonstrated the effectiveness of leveraging advanced machine learning models, such as TabNet and its custom wrapper, alongside sequential models like LSTM and MLP, for tackling multi-class classification problems with tabular and time-series data. The TabNet Voting Regressor Ensemble emerged as the most robust model, achieving the best Quadratic Weighted Kappa (QWK) and Kaggle scores, highlighting the power of ensemble methods in enhancing predictive performance. While the TabNet Wrapper and MLP showed strong capabilities in learning feature interactions, sequential models like LSTM struggled to generalize effectively on tabular data. The integration of feature engineering, dimensionality reduction through autoencoders, and threshold optimization played a pivotal role in improving model interpretability and accuracy. This project underscores the importance of tailored model selection and optimization strategies for achieving reliable predictions in challenging multi-class problems. Overall, to conclude TabNet Voting Regressor Ensemble achieved highest QWK and scored top on leaderboard with 0.494 score.

Future Work

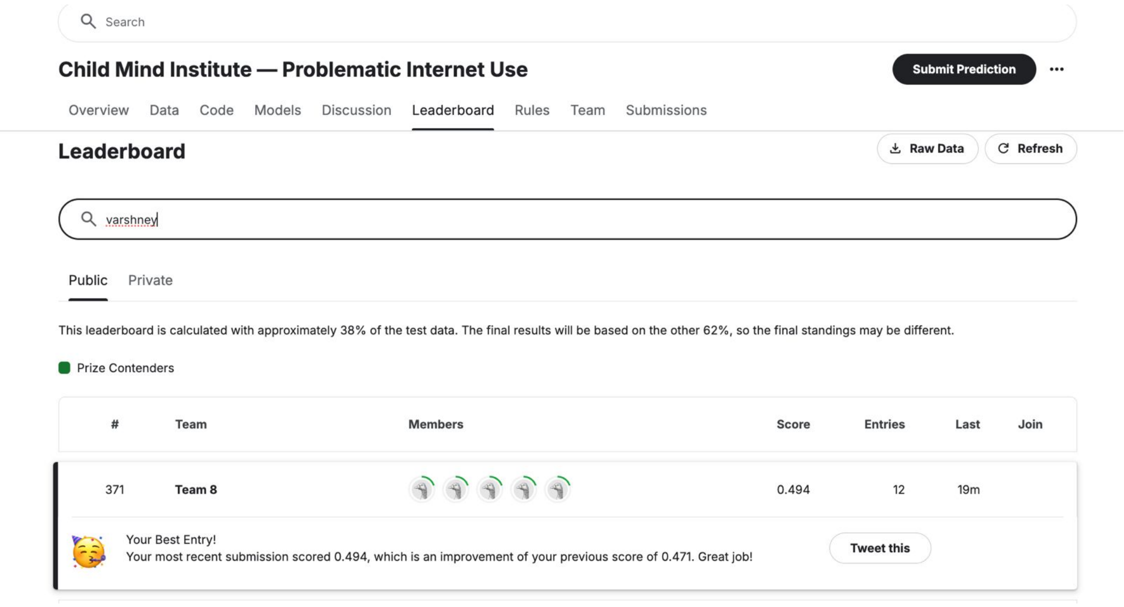
Future work can focus on further enhancing model performance by experimenting with advanced feature engineering techniques, such as generating interaction features or leveraging domain-specific knowledge for feature creation. Incorporating additional time-series preprocessing methods, such as frequency domain analysis, could improve the representation of actigraphy data. Hyperparameter tuning across all models, particularly for ensemble approaches, may yield better predictions. Exploring transfer learning with pre-trained models for time-series tasks and integrating explainable AI methods to understand model predictions better could also add value. Finally, addressing hardware-specific inconsistencies and scaling the solution for larger datasets would ensure robustness and broader applicability.

Kaggle Competition Ranking

Below Figure 22 is showing the kaggle comeptition ranking of 371 out of total participants of 15,404. There are around 300 participants with the same submission score.

Figure 22

Kaggle Competition Ranking



Child Mind Institute — Problematic Internet Use

Leaderboard

This leaderboard is calculated with approximately 38% of the test data. The final results will be based on the other 62%, so the final standings may be different.

🟩 Prize Contenders

#	Team	Members	Score	Entries	Last	Join
371	Team 8		0.494	12	19m	

Your Best Entry!
Your most recent submission scored 0.494, which is an improvement of your previous score of 0.471. Great job!

References

Arik, S. Ö., & Pfister, T. (2019). Tabnet: Attentive interpretable tabular learning. *arXiv preprint arXiv:1908.07442*. https://doi.org/10.1609/aaai.v35i8.16826

Cai, Z., Mao, P., Wang, Z., Wang, D., He, J., & Fan, X. (2023). Associations between problematic internet use and mental health outcomes of students: A meta-analytic review. *Adolescent Research Review*, *8*(1), 45–62. https://doi.org/10.1007/s40894-022-00201-9

Cash, H., Rae, D. C., Steel, H., & Winkler, A. (2012). Internet addiction: A brief summary of research and practice. *Current psychiatry reviews*, *8*(4), 292–298. https://doi.org/10.2174/157340012803520513

Cheng, J., Burke, M., & Davis, E. G. (2019). Understanding perceptions of problematic facebook use: When people experience negative life impact and a lack of control. *Proceedings of the 2019 CHI conference on human factors in computing systems*, 1–13. https://doi.org/10.1145/3290605.3300429

Dresp-Langley, B. (2020). Children's health in the digital age. *International journal of environmental research and public health*, *17*(9), 3240. https://doi.org/10.3390/ijerph17093240

Falbel, D., RStudio, Regouby, C., Fridgeirsson, E., Haarmeyer, P., & Verweij, S. (2024). Tabnet: Fit 'tabnet' models for classification and regression (version 0.6.0) [r package].

Koczkodaj, W. W., et al. (2022). Combating harmful internet use with peer assessment and differential evolution. *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, 1–6. https://doi.org/10.1109/ICECET55527.2022.9873437

Montag, C., Lachmann, B., Herrlich, M., & Zweig, K. (2019). Addictive features of social media/messenger platforms and freemium games against the background of psychological and economic theories. *International journal of environmental research and public health*, *16*(14), 2612. https://doi.org/10.3390/ijerph16142612

Pontes, H. M., & Macur, M. (2021). Problematic internet use profiles and psychosocial risk among adolescents. *PLoS One*, *16*(9), e0257329. https://doi.org/10.1371/journal.pone.0257329

Restrepo, A., Scheininger, T., Clucas, J., Alexander, L., Salum, G. A., Georgiades, K., & Milham, M. P. (2020). Problematic internet use in children and adolescents: Associations with psychiatric disorders and impairment. *BMC psychiatry*, *20*, 1–11. https://doi.org/10.1186/s12888-020-02640-x

Saletti, S. M. R., Van den Broucke, S., & Chau, C. (2021). The effectiveness of prevention programs for problematic internet use in adolescents and youths: A systematic review and meta-analysis. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *15*(2). https://doi.org/10.5817/CP2021-2-10

Shi, F., Ning, H., & Dhelim, S. (2021). A tutorial of cyber-syndrome viewed from cyber-physical-social-thinking space and maslow's hierarchy of needs. *arXiv preprint*, *arXiv:2111.02775*. https://doi.org/10.48550/arXiv.2111.02775

Tokunaga, R. S. (2015). Perspectives on internet addiction, problematic internet use, and deficient self-regulation: Contributions of communication research. *Annals of the International Communication Association*, *39*(1), 131–161. https://doi.org/10.1080/23808985.2015.11679174