Predicting ADHD using Smartphone Sensing Data

Ioana-Andreea Cristescu
Department of Computer Science
University of Richmond
Richmond, Virginia
ioanaandreea.cristescu@richmond.edu

Ying Zhu
Department of Computer Science
University of Richmond
Richmond, Virginia
ying.zhu@richmond.edu

***Abstract*—In this project, we explored a classification program for ADHD prediction. We used Android sensing collected through an Android smartphone application to predict ADHD in a small pilot sample. The approach is based on continuous behavioral data collected automatically from participants' phones. We then use supervised training to predict whether a patient has ADHD or not. As this is a novel research, we compare and contrast the performance (F1, accuracy, precision, and recall) of the various machine learning models such as SVM, Decision Tree, Random Forest, XGBoost, and Neural Networks.**

I. Introduction

ADHD (Attention Deficit Hyperactivity Disorder) is a neurodevelopmental disorder that affects both children and adults and is characterized by symptoms such as inattention, hyperactivity, and impulsivity. People with ADHD may struggle with paying attention to details, completing tasks, following instructions, organizing activities, and maintaining focus [1]. As the exact cause of this disorder is not yet fully understood, in our study, we focus on helping people prevent the negative impact of ADHD by exploring a classification problem for ADHD prediction. We collect sensing data recorded through an Android or iOS smartphone application from a small pilot sample of college students. Our approach is based on continuous behavioral data collected automatically from participants' phones. We then use machine learning to predict if a patient has ADHD or not.

II. Related Work

There are numerous studies that employ smartphone sensing data in assessing and detecting mental health diseases such as ADHD and depression [2], [3], [4], [5]. In the following, we take inspiration from related work and create something novel.

A previous study used smartphone sensing data collected through an Android app in order to predict ADHD symptoms [3]. Ware et. al. (2022a) used SMS data collected in time intervals of 7 days along with their ADHD symptoms scores. In our work, instead of using SMS data, we will utilize activity data recorded for each individual day or for a time interval that will be determined after performing data analysis. The ultimate goal of our project is to predict whether a patient has ADHD or not.

Ware et. al. (2022b) uses a sliding window approach in the attempt of capturing behavior patterns and changes. For example, for a day t, they would consider the data collected during the past n days, i.e., $[t-n+1, n]$. The value of n used in Ware et. al. (2022b) was $n = 7$ or $n = 14$. Similarly, we will attempt to capture variations in each user's behavior in between the completion of two surveys. The minimum time interval between two surveys that we take into consideration is 7 days. Surveys filled out at a faster rate will be discarded.

By working with smartphone sensing data that is susceptible to various unpredictable variables, we observe a significant amount of missing data. Yue et. al. (2021) presents an approach of dealing with GPS missing data collected through Android and iOS apps. Similar to their findings, our missing activity data can happen during the day or night due to scheduling of the operating system, failure of data capture by sensors, or mis-configuration by a participant. Using this information along with our own personal knowledge about the schedule of a college student during the weekdays and weekends, we will be able to develop an efficient way of substituting or discarding missing data.

III. Dataset Used

We have collected both sensing data and self-reported survey data for all phases of the research. Phase 1, with 7 participants, focused on collecting Android users' data from Spring 2022 to Summer 2022. Phase 2, with 8 participants, focused on collecting Android users' data from Winter 2022 to Spring 2023. And phase 3, with 15 participants, focused on collecting data iOS users' data from Spring 2022 to Spring 2023. We will focus on Phases 1 and 2 of the research in this paper.

A. Sensing Data

With consent, we have gathered sensing data from the app designed for both Android and iOS users, to cover the majority of the mobile phone choices of college students. The sensing data was collected through data sharing on the cell phones. The applications and phone setting has allowed the data to be anonymous. The sensing data include phone call data, message data, WiFi location data, activity, and app usage data. In this stage of the research, we will focus on the activity data. The data marks the sensing starting time, sensing type, activity, and confidence in the data. The "activity" element is described as "still", "tilting", "running", etc., corresponding to related human activities.

B. Self-reported Survey Data

Besides the sensing data, we also have the self-report questionnaire answered by each participant. Within the self-reported surveys, we list 18 questions about inattention, hyperactivity, and impulsivity. See the appendix for more details on all the questions.

Participants were asked to rate their responses to the questions above from a score of 1 to 4, representing "Never or rarely", "Sometimes", "Often", and "Very often" respectively. All 18 questions are listed in the Appendix. Each participant is required to answer the survey at the beginning stage of the research, to record as the baseline. Then, approximately every week, participants are required to upload their answers to the questionnaire.

Accompanied by the official diagnosis from physicians, we could further use the self-reported data to find correlations between sensing data and self-reported data later in the research.

IV. Approach

The approach we took in conducting this research consists of four stages as illustrated in Figure 1. After recruiting the participants with both Android and iOS devices, they were instructed to fill out a baseline questionnaire about their ADHD diagnoses and symptoms, followed by weekly surveys on inattention, hyperactivity, and impulsivity. At the same time, the participants were instructed on downloading the Android/iOS app for recording their activity. In the third step of our research, we pre-process the data, calculate an array of features, and use various machine learning models for ADHD prediction. In the fourth, and final step, we predict on unknown data by utilizing the models trained in the previous step.

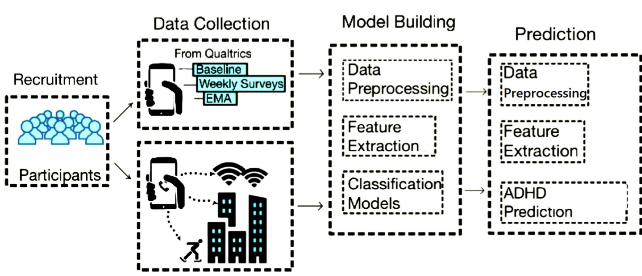

Fig. 1. Illustration of high-level approach.

V. Data Analysis

Since the data collection mechanism is different for Andriod and iOS platforms, in this project, we focused on the Android platform only, which includes phase 1 and phase 2 of the data collected.

A. Data Preprocessing

We first cleaned the anonymous data collected by removing the duplicates and filtering it based on the start and end date. In this research, we only focused on a certain time span of the participants' activities between Spring 2022 and Spring 2023. Then we visualize the data to determine the threshold of data to include in the research, as shown in V-B. We then extracted features, as explained in VI.

B. Data Visualization

We visualized the data collected by creating a histogram plot (Figure 2). The plot represents the frequency of the number of days between 2 self-reported surveys filled out by all participants. The x-axis will represent the number of days, such as 1, 2, ..., and 7 days. The y-axis will constitute the number of times the specific time interval occurred. With this plot, we can best select a threshold and discard the data that is below the threshold. From the 116 survey intervals, 78 show that participants filled out consecutive surveys with an interval of exactly 7 days. We, therefore, determine 7 to be the threshold and discard the data that is below 7.

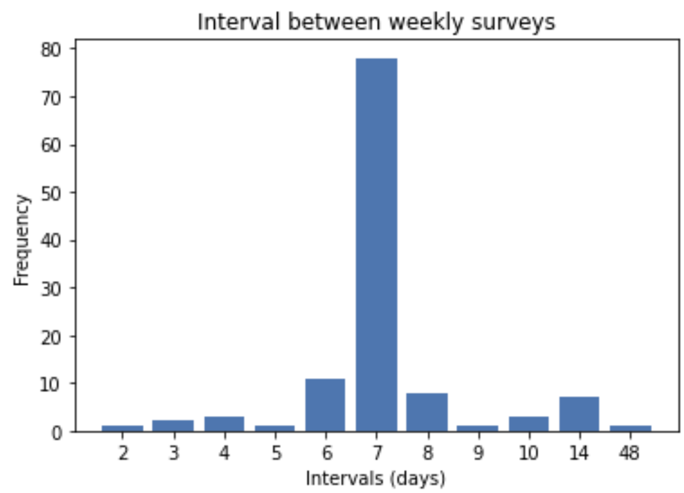

Fig. 2. Interval between weekly surveys.

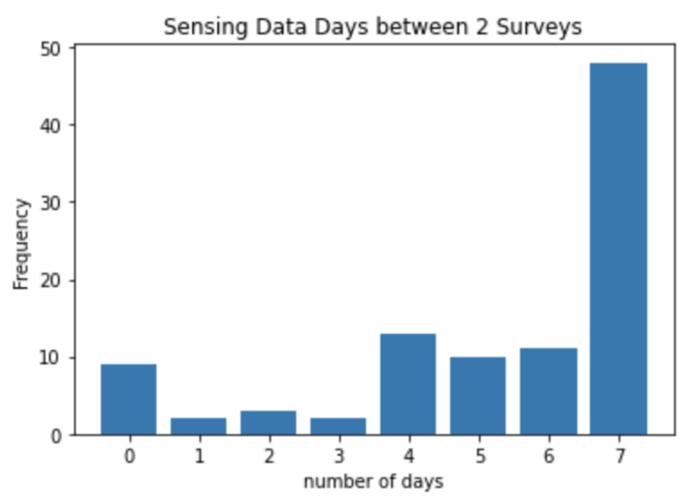

Fig. 3. Sensing Data Days between 2 Surveys.

With the remaining data, we created another histogram plot in order to analyze the number of days of recorded sensing data between 2 self-reported surveys (Figure 3). The aim is to establish a minimum threshold and discard the weeks that do not have sufficient data. We used the number of days as the x-axis, from 1 to 7 days, and the frequency as the y-axis. According to Figure 3, most participants have 7 days of sensing data between 2 self-reported surveys. However, there are over 30 weeks that have at least 4 days of sensing data. We establish 4 as the threshold, in order to include reasonable absences and not discard too much information.

VI. Methodology and Results

A. Data Scaling and Feature Extraction

We preprocess the Android data collected in weekly survey intervals where the distance between surveys is at least $n = 7$, with each week having at least 4 days of sensing data. Each user and each weekly interval is then mapped to their corresponding "Activity" features, ADHD label, and ADHD score. Once the features are calculated, we perform Min-Max scaling to normalize the range of independent variables.

1) "Activity" Features : Features are calculated as a weekly average depending on the activity type and the time of the day. The activity type can be further detailed into "in vehicle", "on bicycle", "running", "still", "tilting", "walking", and "unknown". In our project, we decided to concentrate on calculating features for the "still" and "running" activities, while discarding the data labeled as "unknown." To facilitate the calculation of features we sort the dataset in increasing order based on user ID and start time of the sensing data. The start time of the sensing data is recorded in Unix Timestamp. Considering that the levels of activity of a college student vary during the day, we divided the day into 4 intervals such that night ranges from 12:00 AM - 7:00 AM, morning ranges from 7:00 AM - 12:00 PM, afternoon ranges from 12:00 PM - 5:00 PM, and evening ranges from 5:00 PM - 12:00 AM. Therefore, that creates 8 features for each user during each weekly interval: "Still Night," "Still Morning," "Still Afternoon," "Still Evening," "Running Night," "Running Morning," "Running Afternoon," "Running Evening." Due to the lack of continuous sensing data and for an accurate prediction, we had to establish thresholds for each activity and each part of the day. That is, if we don't have any sensing data registered for a long period of time, we assign a limit to the amount of time the student might have done a certain activity.

- "Still"
  - Night: 7 hours
  - Morning: 1 hour
  - Afternoon: 1 hour
  - Evening: 3 hours
- "Running"
  - Night: 1 hours
  - Morning: 2 hour
  - Afternoon: 2 hour
  - Evening: 2 hours

All the calculations are done by subtracting two consecutive epoch timestamps. We add all the differences together as long as the next epoch timestamp is less than or equal to the established threshold and the user is performing the same type of activity. Once at least one of these two conditions are not met, we start recording the time for a "new" activity. Depending on the time interval between two epoch timestamps and time spent doing a certain activity before changing to a different activity or before having no recorded sensing data, we have the following scenarios.

- The time between the epoch timestamps is less than the threshold. In this situation, we keep adding the time difference to the total time. If the next activity is different from the current activity, then record the time and reset it to zero.
- The time between the epoch timestamps is more than the threshold. This situation creates two more scenarios. If the total time spent doing the activity is less than or equal to the threshold, then the total time becomes the threshold. If the total time spent doing the activity is more than the threshold, then add 10 more minutes to the total time. In both situations, record the time and reset it to zero.

2) ADHD Label: We mapped each user and each weekly interval to its corresponding ADHD label. The ADHD label represents whether a user has ADHD or not, and it was reported by the users themselves in the initial questionnaire. The label has value 1 if the patient has been diagnosed with ADHD, and value 0 if the patient has not been diagnosed with ADHD. By mapping each user to their ADHD label, we notice that we have 57 training examples with positive ADHD labels versus only 25 training examples with negative ADHD labels. In order to balance the class sample, we upsample the data representing the users that have not been diagnosed with ADHD. We perform this process by duplicating the 25 training examples with negative labels.

3) ADHD Scores: We mapped each user and each weekly interval to its corresponding ADHD score. The ADHD scores were calculated based on the 18 questions that the users were prompted to answer during the weekly surveys. Through the 18 questions the users self-reported their "inattention," "hyperactivity," and "impulsivity" symptoms. Each symptom could be ranked based on the frequency to which it was experienced during the past week: "never or rarely" (numeric score of 1), "sometimes" (numeric score of 2), "often" (numeric score of 3), and "very often" (numeric score of 4). In order to get the ADHD score, we summed over the scores of all 18 questions.

B. Feature Visualization

We further investigate the relations between the features computed and the ADHD diagnosis by plotting a series of scatter plots. We use a different color to represent each user. The survey dates are displayed on the x-axis, and activity duration in minutes on the y-axis.

Comparing Figure 4 and Figure 5, we can observe that both participants with ADHD and without ADHD tend to stay still for a longer period of time in the mornings. This might be attributed to the morning classes students were taking during the semester. Comparing Figure 6 and Figure 7, we observe that both participants with ADHD and without ADHD do not remain overly active during the night time. We cannot draw a clear correlation from the comparisons of the visualizations. We then further analyzed using correlation as shown in VII.

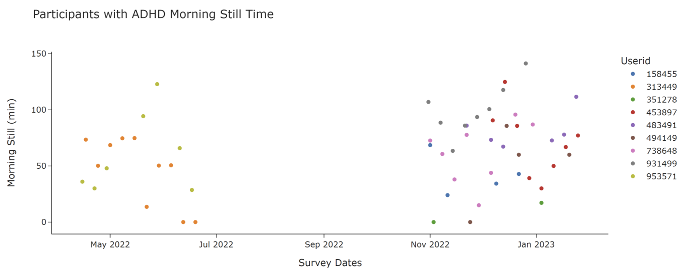

Fig. 4. Still duration during the morning (7 am - 12 pm) for participants with ADHD.

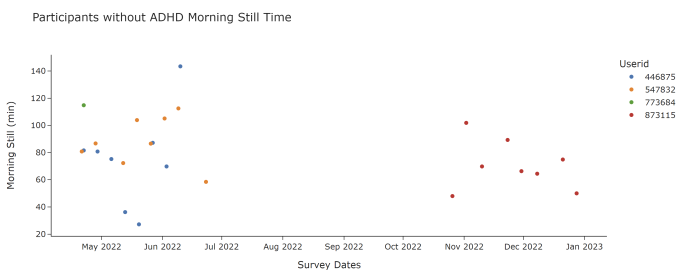

Fig. 5. Still duration during the morning (7 am - 12 pm) for participants without ADHD.

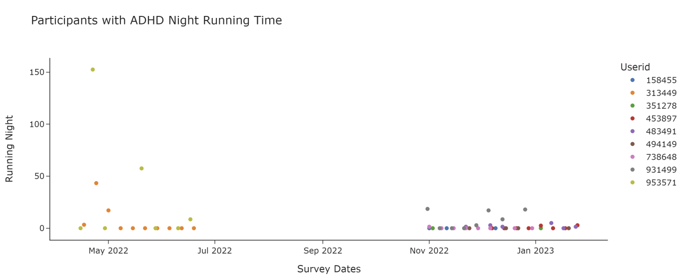

Fig. 6. Running duration during the night (12 am - 7 am) for participants with ADHD.

C. Classification Methodology and Results

We compare and contrast the performance and limitations of several machine learning models for ADHD prediction. We use the Scikit-learn module for our analysis. First, we split the dataset with 107 training examples into a ratio of $7:3$ for the training and test datasets. Then, we train the SVM, Decision Tree, Random Forest, XGBoost, and Neural Networks models and calculate their accuracy, precision, recall, and F1 score

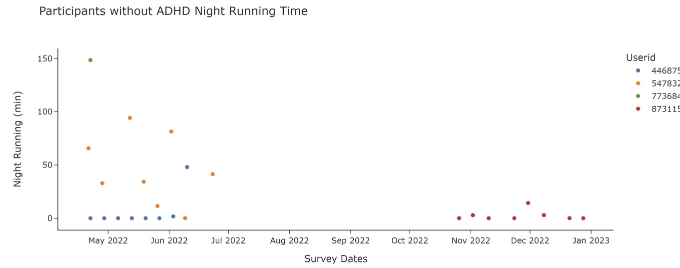

Fig. 7. Running duration during the night (12 am - 7 am) for participants without ADHD.

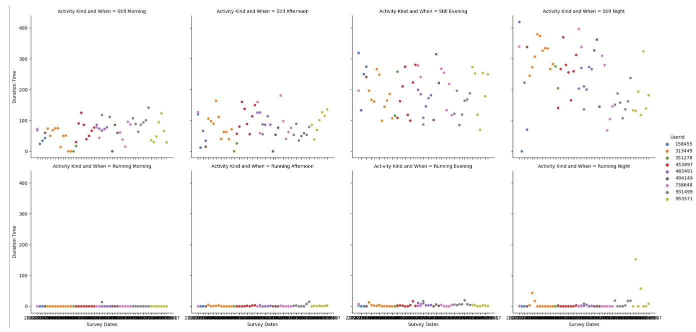

Fig. 8. Still and running durations during different parts of the day for participants with ADHD.

(Figure 10). The Decision Tree, Random Forest, and XGBoost algorithms are trained using the default parameters. For the SVM model, we perform classification using linear kernel, polynomial kernel of degree 2, and RBF kernel for values of the penalty parameter $C = [1, 10, 100, 1000]$. Lastly, we train two Neural Networks algorithms with dense layers by using the binary cross entropy loss function (one network contains two hidden layers with 5 and 3 neurons, and the second network contains only one hidden layer with 4 neurons). The chosen architectures can be explained by the fairly small dataset. We utilize the "ReLU" activation function for the hidden layers and the "sigmoid" function for the output layer.

Figure 10 presents the classification results. We observe that the ADHD labels are predicted accurately with an F1 score as high as 0.86 for SVM with RBF kernel, Random Forest, and XGBoost. Moreover, we can notice that the algorithm that performed the worst was the Neural Network with one hidden layer with an F1 score of 0.73.

VII. Evaluation and Analysis

We explored Pearson Correlation between the ADHD self-reported scores and the eight features we used in the model. The correlation measures the strength of the linear relationship between the two variables described above.

Figure 11 presents the correlation results. Among all eight features, we observe that only "Running Night (12 am - 7 am)" has a positive correlation with ADHD scores. Among the remaining features, "Still Night" has the largest negative correlation, followed by "Running Afternoon" (12 pm - 5

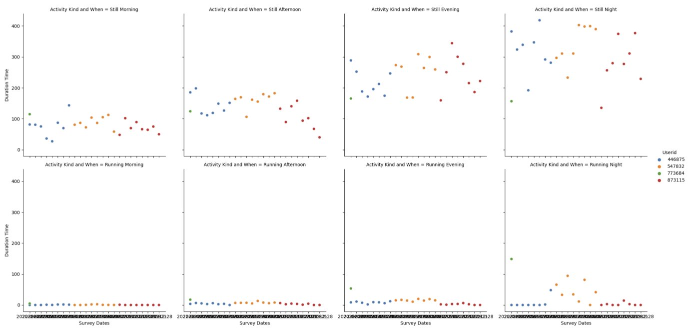

Fig. 9. Still and running durations during different parts of the day for participants without ADHD.

Models		F1	Precision	Recall	Accuracy
SVM	Linear Kernel (C=1)	0.85	0.81	0.89	0.82
	Polynomial Kernel of degree 2 (C=100)	0.84	0.84	0.84	0.82
	RBF Kernel (C=100)	0.86	0.89	0.84	0.85
Decision Tree		0.76	0.78	0.74	0.73
Random Forest		0.86	0.89	0.84	0.85
XGBoost		0.86	0.89	0.84	0.85
Neural Network	One hidden layer	0.73	0.58	1	0.58
	Two hidden layers	0.84	0.84	0.84	0.82

Fig. 10. ADHD prediction results using SVM with linear kernel, polynomial kernel of degree 2, RBF kernel, Decision Tree, Random Forest, XGBoost, and Neural Networks.

pm). These results are intuitively sound, as hyperactivity and impulsivity are key symptoms of ADHD. According to CDC [1], people with ADHD "often unable to play or take part in leisure activities quietly" and "are often 'on the go' acting as if 'driven by a motor'". These descriptions correspond to being excessively active during nighttime when rest is supposed to take place. These descriptions also explain the still afternoon correlation of "-0.2606", when participants with ADHD are less likely to be still during afternoons.

VIII. Limitations

A. Small data sample size

In this stage of the research, we have only conducted research on 15 participants with Android phones. We realize the limitations of small data sizes. With only 15 participants, it is challenging to set a proper threshold, which will eliminate the noisy data, while still preserving the remaining data and permitting the model to generalize well (avoiding overfitting).

B. Outliers

With a small data size, it is hard to distinguish outliers in the dataset. For example, among participants with ADHD, only 1 user (with user id = 953571) stays excessively active during night time (12 am - 7 am), with an active time of 152.5 minutes and 57.5 minutes on 2 days, while most participants with ADHD do not run at night. User 953571, being a potential outlier, can drive the Pearson correlation to be positive in the analysis for "Running Night" and ADHD diagnosis (VII).

Features		Pearson Correlation
Still	Morning	-0.1773
	Afternoon	-0.2606
	Evening	-0.1193
	Night	-0.3486
Running	Morning	-0.1262
	Afternoon	-0.2718
	Evening	-0.1234
	Night	0.0481

Fig. 11. Pearson correlation between ADHD (0 or 1) and corresponding feature for all eight features.

IX. Future Work

A. Linear regression

In our project, we performed classification to decide whether a user has ADHD or not. One potential future work is to go beyond that and determine the level of ADHD a person has. For that, we would perform linear regression and predict the ADHD scores.

B. More features

As mentioned in III-A, "Activity" sensing data includes various actions. While in our research we concentrated on time spent "still" or "running," we could calculate more features that would represent the amount of time users spent "in vehicle," "biking," "walking," or "tilting." At the same time, we would like to include other types of sensing data like location and app usage. These new features could offer more insight about the behavior of people with ADHD, while also improving the performance of our machine learning models.

C. iOS data

We will continue this work with a larger dataset (collected across both android and iOS platforms) in our upcoming research study.

X. Conclusions

In conclusion, this project aimed to explore a novel approach for ADHD prediction using Android sensing data collected through a smartphone application. The use of continuous behavioral data allowed for a more comprehensive understanding of participants' daily routines and behaviors, leading to more accurate predictions. Our study employed various machine learning models, including SVM, Decision Tree, Random Forest, XGBoost, and Neural Networks. We analyzed the collected data and extracted relevant features for prediction. The performance of the models was evaluated using several metrics, including F1 score, accuracy, precision, and recall. Among the models, SVM with RBF kernel, Random Forest, and XGBoost performed the best with an F1 score of 0.86, precision of 0.89, recall of 0.84, and accuracy of 0.85. The Pearson correlation offers some insight into the relations between features extracted and ADHD diagnosis (0 or 1) and provides an intuitive explanation. Overall, the results of this pilot study demonstrate the potential of using mobile sensing data and machine learning models for ADHD prediction. Limitations and further research can indicate future exploration and implications to improve the current method for broader applications.

References

[1] CDC. Symptoms and Diagnosis of ADHD — CDC — cdc.gov. https://www.cdc.gov/ncbddd/adhd/diagnosis.html. [Accessed 17-Apr-2023].

[2] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, pages 3–14, 2014.

[3] S. WARE, L. E. KNOUSE, I. DRAZ, and A. ENIKEEVA. Predicting adhd symptoms using smartphone sensing data. 2022a.

[4] S. Ware, C. Yue, R. Morillo, C. Shang, J. Bi, J. Kamath, A. Russell, D. Song, A. Bamis, and B. Wang. Automatic depression screening using social interaction data on smartphones. *Smart Health*, 26:100356, 2022b.

[5] C. Yue, S. Ware, R. Morillo, J. Lu, C. Shang, J. Bi, J. Kamath, A. Russell, A. Bamis, and B. Wang. Fusing location data for depression prediction. *IEEE transactions on big data*, 7(2):355–370, 2018.

XI. Appendices

We added some important details of the research in the appendices such as the survey questions, codes, etc.

Appendix

We include the 18 questions from III-B (Inattention), section 2 (hyperactivity), and section 3 (impulsivity).

Section 1 (Inattention)	Never or rarely	Some-times	Often	Very often
1. Fail to give close attention to details or make careless mistakes in my work or other activities	1	2	3	4
2. Difficulty sustaining my attention in tasks or fun activities	1	2	3	4
3. Don't listen when spoken to directly	1	2	3	4
4. Don't follow through on instructions and fail to finish work or chores	1	2	3	4
5. Have difficulty organizing tasks and activities	1	2	3	4
6. Avoid, dislike, or am reluctant to engage in tasks that require sustained mental effort	1	2	3	4
7. Lose things necessary for tasks or activities	1	2	3	4
8. Easily distracted by extraneous stimuli or irrelevant thoughts	1	2	3	4
9. Forgetful in daily activities	1	2	3	4
Office Use Only (Section 1) Total Score__________ Symptom Count__________				
Section 2 (Hyperactivity)	Never or rarely	Some-times	Often	Very often
10. Fidget with hands or feet or squirm in seat	1	2	3	4
11. Leave my seat in classrooms or in other situations in which remaining seated is expected	1	2	3	4
12. Shift around excessively or feel restless or hemmed in	1	2	3	4
13. Have difficulty engaging in leisure activities quietly (feel uncomfortable, or am loud or noisy)	1	2	3	4
14. I am "on the go" or act as if "driven by a motor" (or I feel like I have to be busy or always doing something)	1	2	3	4
Office Use Only (Section 2) Total Score__________ Symptom Count__________				

Section 3 (Impulsivity)	Never or rarely	Some-times	Often	Very often
15. Talk excessively (in social situations)	1	2	3	4
16. Blurt out answers before questions have been completed, complete others' sentences, or jump the gun	1	2	3	4
17. Have difficulty awaiting my turn	1	2	3	4
18. Interrupt or intrude on others (butt into conversations or activities without permission or take over what others are doing)	1	2	3	4
Office Use Only (Section 3) Total Score__________ Symptom Count__________				