COVID Student Study: A Year in the Life of College Students during the COVID-19 Pandemic Through the Lens of Mobile Phone Sensing

Subigya Nepal
Dartmouth College
Hanover, NH, USA
sknepal@cs.dartmouth.edu

Weichen Wang
Dartmouth College
Hanover, NH, USA
weichen.wang.gr@dartmouth.edu

Vlado Vojdanovski
Dartmouth College
Hanover, NH, USA
vlado.vojdanovski.22@dartmouth.edu

Jeremy F. Huckins
Dartmouth College
Hanover, NH, USA
Biocogniv Inc.
Burlington, VT, USA
jeremy.f.huckins@dartmouth.edu

Alex daSilva
Dartmouth College
Hanover, NH, USA
alexander.w.dasilva.gr@dartmouth.edu

Meghan Meyer
Dartmouth College
Hanover, NH, USA
meghan.l.meyer@dartmouth.edu

Andrew Campbell
Dartmouth College
Hanover, NH, USA
andrew.t.campbell@dartmouth.edu

ABSTRACT

The COVID-19 pandemic continues to affect the daily life of college students, impacting their social life, education, stress levels and overall mental well-being. We study and assess behavioral changes of N=180 undergraduate college students one year prior to the pandemic as a baseline and then during the first year of the pandemic using mobile phone sensing and behavioral inference. We observe that certain groups of students experience the pandemic very differently. Furthermore, we explore the association of self-reported COVID-19 concern with students' behavior and mental health. We find that heightened COVID-19 concern is correlated with increased depression, anxiety and stress. We evaluate the performance of different deep learning models to classify student COVID-19 concerns with an AUROC and F1 score of 0.70 and 0.71, respectively. Our study spans a two-year period and provides a number of important insights into the life of college students during this period.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing**.

KEYWORDS

Mobile Sensing, COVID-19, Pandemic, Mental health, Digital Phenotyping

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04.
https://doi.org/10.1145/3491102.3502043

ACM Reference Format:
Subigya Nepal, Weichen Wang, Vlado Vojdanovski, Jeremy F. Huckins, Alex daSilva, Meghan Meyer, and Andrew Campbell. 2022. COVID Student Study: A Year in the Life of College Students during the COVID-19 Pandemic Through the Lens of Mobile Phone Sensing. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3491102.3502043

1 INTRODUCTION

The Coronavirus disease 2019 (COVID-19) has been disrupting the daily life of most people in the world by causing emotional, psychological and physical harm. In order to limit its spread, a strong series of measures were introduced, including curfews, lockdowns, stay-at-home advisories and directives asking people to self-quarantine and maintain social distance. Such measures have led to changes in several domains, from redefining workplace flexibility to the reshaping of college experiences of students. Specific to the college experiences – a topic we address in this paper – schools and colleges pivoted to online modes of teaching and learning. The vast majority of classes taught at colleges are being delivered online via Zoom, with students and teachers joining remotely from home or in some limited capacity from campuses. Such changes, however, have introduced or accentuated several stressors for students and faculty alike against the backdrop of the pandemic. Studies report that student's mental health is worsening since the start of the COVID-19 pandemic [21, 48] – the USA declared a national emergency on March 13, 2020. While the pandemic heightened existing stressors for students, it also introduced many new concerns, such as, social isolation [25], virtual fatigue [55], increased concern for family members' health (e.g., parents, grandparents, siblings), financial well-being and concerns for at-risk classmates [7, 48, 62]. A lack of social interaction alone has been linked in the general population to a wide variety of poor mental and physical health outcomes including depression, cardiac disease and early death [16]. It has been widely reported that many of these factors combine to impact the mental health and well-being of young adults during the pandemic.

In the study we report on in this paper, we seek to examine and quantify for the first time how the mental health and behavior of college students has changed during the COVID-19 pandemic in comparison to their life just prior to the pandemic. The study is driven by the following broad exploratory research questions:

- ***(Q1)*** *How does students' behavior change during the COVID-19 period compared with a year prior to the pandemic?* With this question, we aim to explore the adjustment students are making in their daily life during the pandemic. Behavior change here refers to the change in a number of mobile sensing features that we collect through students' phones across the two-year study; that is, in the pre-COVID-19 year and first COVID-19 year.
- ***(Q2)*** *Do all students react to the pandemic in a similar manner, i.e., do students experience similar behavioral changes during the pandemic?* It is possible that some students might be handling the pandemic differently than their counterparts in terms of behavioral change as well as their mental health. An insight into how students are reacting to the pandemic may help us identify differing groups and provide tailored support to them (e.g., digital interventions).
- ***(Q3)*** *Is concern for COVID associated with the mental health of students?* We ask students several EMAs (ecological momentary assessments) to capture their concern toward COVID. We believe it is important to identify whether such COVID concerns are related to student's mental wellbeing.
- ***(Q4)*** *How does the cohorts' COVID concern change as the pandemic progresses?* We would like to investigate whether the students have a consistently elevated concern about COVID across the first COVID-19 year.
- ***(Q5)*** *Based on the observation that COVID concern does indeed have an impact on student's mental health, viz.* ***(Q3)****, can we build a machine learning model to predict student's COVID concern?* Such a model would enable us to detect students who are more prone to mental health issues during the pandemic and provide a foundation for delivering proactive help to them. For instance, the output of the model could be used to opt people in for frequent check-ins with a mental health network or peer support network.

While the majority of the existing studies investigating students' behavioral changes during COVID-19 exclusively use self-reports, we utilize data from the real-time passive sensing StudentLife app [57] to identify changes in student behavior. The StudentLife app provides us access to behavioral and contextual information through various built-in sensors, which are less vulnerable to self-report bias. Note, the smartphone app used in our COVID student study has been validated and used in a number of prior clinical studies [2, 58, 59]. We use baseline sensing data from the mobile phones of N=180 undergraduate students for a year prior to the pandemic (called the *pre-COVID-19 baseline year* from March 2019 through February 2020) and compare it to the first 12 months of the pandemic (called the *first COVID-19 year* from March 2020 through February 2021). We directly compare and contrast the life of these 180 undergraduates during the pre-COVID baseline year and COVID year. Such a large amount of longitudinal data spanning two years in total allows us to draw important insights and robust conclusions associated with behavioral change of these students as a result of COVID-19. In addition, we use EMA to record ground truth input on mental state and COVID-19 concern from the students. We capture the self-reported level of depression, anxiety, stress, social life and self-esteem using EMA across the complete 24 month study period. During the COVID year, we also collect specific self-reports associated with students' concern about the ongoing COVID-19 crisis. Please note that we collect data from a cohort of undergraduate students with a demographic attending a selective university in the USA so we would urge the reader to be cautious in their interpretation of the results discussed in this paper, as we do not know exactly how well these results generalize to other populations. In this paper, we report on analysis of the data collected across the 24 months; the contributions of our work are as follows:

- To the best of our knowledge, we present the first mobile sensing study that investigates the objective behavioral changes of students as a result of COVID-19 pandemic, comparing the 12 months prior to the pandemic and the first 12 months of COVID. We consider N=180 undergraduate students and identify the changes in their behavior following the emergence of COVID-19 in the USA in March 2020. Specifically, as stated above, we consider the pre-COVID-19 baseline year and contrast it with the first COVID-19 year.
- While considering overall behavioral changes, we find that students' behaviors captured from passive sensing data shifted and time spent engaging in multiple activities substantially reduced during the pandemic. For example, we quantify that students are significantly less active (i.e., distance travelled decrease by 60%) during the first COVID-19 year but use their phone more (i.e., phone usage duration increases by 15%).
- We identify two distinct groups of students in the study population who have differing experiences during the pandemic in terms of behavioral changes and self-reported mental health and COVID concerns. Using cluster analysis, we observe that certain groups of students are much worse off and affected by COVID than their counterparts. These students have higher COVID concerns, depression, anxiety, stress and lower self-esteem. We also find that there are segments of students whose behavioral change during COVID is completely different from each other. For instance, while a cluster of students uses their phone more during the pandemic, students belonging to another cluster show a decrease in phone usage duration. We observe similar differences in other behaviors, such as, students' bedtime, wake-up time, sleep duration and still duration.
- We examine the relationship between COVID concerns and mental health, identifying a number of significant correlation (p-val < 0.05). Within our dataset, COVID-19 concern is moderately correlated with Patient Health Questionnaire

(PHQ)-4 (ρ=0.31), depression (ρ=0.35), anxiety (ρ=0.25) and self-reported stress (ρ=0.25).

- We observe that the majority of the students have a consistent COVID-19 concern, with 31% (N=56) of students always reporting a heightened COVID concern throughout the first COVID-19 year.
- Using students' self-reported COVID concern as the ground truth, we build a Fully Convolutional Neural Network that classifies the participants into either a higher or lower COVID concern class based on the mobile sensing data. We report a weighted AUROC and F1 score of 0.70 and 0.71, respectively. We also identify a number of important features from the model.

We are now in the second year of the COVID-19 pandemic and while many university students, faculty, staff and administrators are hopeful that our campuses will eventually return to normal it looks like we will be living with this pandemic and its variants for sometime to come. In fact, we have little visibility of the end of the pandemic with any certainty. As a result, we believe that the contributions discussed in this paper remains very relevant going forward in the next phase of the pandemic.

The structure of the paper is as follows. We start by discussing the related work on the use of passive sensing during COVID-19 to study behavior change in Section 2. Following this, we detail our study design, ground truth and dataset in Section 3 and report on the analysis on behavioral change of students in Section 4. Next, we investigate how the participants differ in their COVID-19 experiences and behavioral change by profiling them with the help of clustering in Section 5. We then examine the relationship between COVID-19 concern and self-reported depression, anxiety, stress, social level and self-esteem in Section 6. We explore the change in COVID-19 concern over time in Section 7. Following this, Section 8, explores different deep learning models for predictive analysis of COVID-19 concern, presenting our results and drawing insights associated with important modelling features. We discuss our overall findings and implications and the limitations of our work in Section 9 and 10, respectively. Finally, we present some concluding remarks in Section 11.

2 RELATED WORK

Mobile sensing is widely used for passive assessment of human behavior [19, 31, 32, 56]. Numerous studies find association between mobile sensing and different aspects of mental health [43, 57, 61], personality [12, 60], workplace behavior [11, 32, 35, 36], among other things. There are a growing number of papers reporting insights during the pandemic, however, the vast majority of these studies are based on purely self-reported inventories. In contrast, there are a smaller number of studies that use mobile sensing to study human behavior and contact tracing. Mobile sensing offers the advantage of passively collected contextual in-situ data, which can be used to make objective inferences in naturalistic settings. Capturing sensing data is particularly suited to assessing human behavior during the pandemic. In our case, this allows us to track students in our study and draw insights from their behavior in an unobtrusive and in-situ manner.

A number of mobile tracking apps have been used during the COVID period in order to facilitate contact tracing primarily through proximity sensing [29] to various degree of success and acceptance around the globe. Using Bluetooth signals, mobile apps can track and notify people if they have been in contact with individuals exposed to the COVID-19 virus [18] with the goal of helping to reduce the spread of COVID-19 in the general population. Some studies use such Bluetooth proximity data and Call Detail Records (CDR) to build contact networks for analyzing mobility of large populations [23, 51]. While this is specific to mobility, only a handful of other studies employ mobile sensing to study behaviors of participants. In a study of N=20 young adults over a seven day period pre- and during the COVID lockdown in Spain, Sañudo et al. [45] report finding a 68% decrease in the average number of steps per day, a 32% increase in smartphone use and a 7% increase in sleep duration during the lockdown period. This reduction in objectively assessed physical activity during COVID is also supported by other studies [38, 54]. One in particular from the United Kingdom, which tracked N=5395 participants from January 22 to June 17 2020, reports that the physical activity progressively decreased through the early phases of the pandemic, with a 47% decrease in the first full week of lockdown [30]. Huckins et al. [21] study N=217 undergraduate students using mobile sensing and compare their behavior from the past academic terms with their behavior during the first academic term impacted by COVID-19 (i.e., Winter 2020 term which started on January 6, 2020 and ended in the next 10 weeks). Compared with prior academic terms, the authors find that during the start of the pandemic, students were more sedentary, anxious and depressed. Furthermore, they observe a decrease in locations visited, decrease in physical activity and increase in phone usage. In a follow-up study, Mack et al. [28] report that among the same cohort of undergraduate students, increased anxiety and depression were significantly associated with rising interest in coronavirus and COVID fatigue search terms on Google.

In another line of work, Quer et al. [40] use passive sensing data from wearables in combination with COVID-19 symptom data to classify the participants into either a COVID-19-positive class (N=54) or COVID-19-negative (N=279) class. With the sole use of sensing data (i.e., heart rate, sleep and activity), the authors [40] obtain an AUC of 0.72. The combination of both sensing and symptom data leads to an AUC of 0.80 while classifying participants into these two groups. Perhaps the closest work to ours is the study done by Sun et al. [50]. The authors investigate passive sensing based behavioral changes of N=1062 participants recruited from Italy, Spain, Denmark, the United Kingdom and the Netherlands. They compare the behavioral change in the data collected from February 1, 2019 to July 5, 2020 across several time points: baseline, pre-lockdown and during lockdown. The authors [50] report that there is a decrease in mobility, an increase in the use of social media applications, a decrease in heart rate and an increase in sleep duration during the lockdown periods. While our study also reports on behavioral changes of participants, we consider the entire year's data right from the national emergency declaration in the USA on March 2020 to February 2021 – the first COVID-19 year in the United States. We compare the passively sensed behavior of that period with a full year prior to March 2020 (i.e., our baseline period is March 2019 to February 2020 – the pre-COVID-19 baseline year).

In addition, our study cohort is made up of young college students and therefore we offer insights into the behavioral change of a non-clinical population; the study by Sun et al. [50] contains participants having different medical conditions (e.g., major depressive disorder, multiple sclerosis). While the work by Sun et al. [50] uses parametric tests to compare the behaviors, we perform a variety of analysis in addition to the statistical tests and prediction.

3 COVID STUDENT STUDY METHODOLOGY

In what follows, we discuss the design of our COVID college student study, demographic information of the students that participated in the study, the ground-truth and the features used during the analysis.

3.1 Study Design

The two years of data collected, analyzed and discussed in this paper comes from a continuing, long-term mobile sensing study that is following N=220 college undergraduates at Dartmouth College across their 4 years of college life using smartphone sensing and self-report surveys.

Students were first recruited and consented just before they joined the university in their first year or during the first academic term of their first year attending the university (N=106 in 2017 and N=114 in 2018). We have been collecting and analyzing the data since the start date. All the participants in the study install the continuous mobile sensing app on their primary Android or Apple phone. They are asked to keep it installed and running on their phone for four years until they graduate college; this includes semester time and breaks from the university, including the summer period. The study is approved by Dartmouth College's Institutional Review Board (IRB). Participants are asked to answer a set of EMAs once every week and they are compensated 10 dollars per week for answering the EMAs, i.e., short surveys sent to their phone through the app. Students also complete a set of longer yearly and pre-post surveys.

The goal of this overarching four year long study is to better understand students' mental health and behavior through the lens of their complete college experience. As such, we ask students to self-report their anxiety, depression, self-esteem, social-levels and stress-levels. Most of these self-reports are delivered as an EMA through the mobile sensing application. The timeline of the pandemic on campus was as follows. During the Winter term of 2020, the COVID-19 virus started to spread globally. On March 10, 2020, our university suggested taking final exams remotely and asked all students to leave the campus as soon as possible. On March 13, COVID-19 was declared a national emergency in the United States. Winter term also ended on this exact date. In order to explore the changes in student's behavior during the pandemic, we started pushing new COVID follow-up surveys to the participants after IRB approval through the mobile app beginning March 18, 2020. This was in addition to the questions that were already being delivered as EMAs since the start of the study. Note that the COVID follow-up surveys were optional to answer for students. We discuss the surveys in more detail in Section 3.3 below.

3.2 Demographics

Out of the 220 participants, 180 completed at least one set of the COVID EMAs. Table 1 shows the demographics of the 180 students used in our analysis. The majority (68.9%, N=124) of our participants identify as females. In terms of race, 61.1% (N=110) are White, 23.4% (N=42) are Asians, 3.3% (N=6) are Black or African American, 2.8% (N=5) are American Indian/Alaska Native and 6.1% (N=11) belong to more than one race.

Table 1: Demographics of the participants. The table below lists the demographic composition of the students in our study.

Category	Count	Percentage
Sex		
Female	124	68.9%
Male	56	31.1%
Race		
White	110	61.1%
Asian	42	23.4%
Black or African American	6	3.3%
American Indian/Alaska Native	5	2.8%
More than one race	11	6.1%
Not reported	6	3.3%

3.3 Ground Truth

We collect self-reported data on the students' depression and anxiety scores, social levels, self-esteem and stress through EMAs. These questions were asked once a week at random times. Students also have the choice to open the app and answer the survey manually at any time they prefer. Depression and anxiety are tracked using the 4-item Patient Health Questionnaire (PHQ-4) for anxiety and depression [24]. State self-esteem is measured with three questions selected from the State Self-Esteem Scale, which includes a relevant question from each of the following categories: social, appearance and performance [27]. Stress is measured by asking "*Are you feeling stressed now?*" with a 5-point Likert scale with response labels ranging from "Not at All" to "Extremely." Self-reported time spent around others (social levels) is measured by asking "*Have you spent most of your time alone or with others today?*" with a 5-point Likert scale with response labels ranging from "Almost always alone" to "Almost always with others".

In this study, we focus on self-reported stress, given the complexities associated with measuring actual stress exposure. State self-esteem is considered a person's sense of their own worth or value at the current moment. The Patient Health Questionnaire (PHQ)-4 is frequently used as a brief measure with relatively good diagnostic performance for depression and anxiety, being comprised of the PHQ-2 and Generalized Anxiety Disorder-2 (GAD-2). As measured within this context, anxiety will be considered persistent and excessive worry. Similarly, depression is multifaceted, but we will focus on anhedonia and negative affect components, given their gross diagnostic abilities of anxiety and depression in college students [1]. Although focusing on these specific aspects of these mental health metrics is likely to be limiting, they allow for quick and more frequent assessment while maintaining rough diagnostic abilities. Note that we refer to PHQ-2 and GAD-2 as depression subscale and anxiety subscale throughout this paper.

At the start of the pandemic we added COVID-19 specific EMAs to understand the impact the current COVID-19 situation is having on the student's lives, social media usage, how supported they feel and mental health. Psychologists in our research team developed these short questions to take less than 1 minute of participants' time per week to maximize response rate. As mentioned, the COVID surveys were optional based on our amended IRB agreement. Students were asked to respond to the COVID related EMAs once a week at a random time, similar to other existing EMAs, discussed above. We list these COVID questions in Table 2. The first question, "*How concerned are you about COVID*"? is the most broadly applicable question and the main focus of our research moving forward. Notably, many of these questions were highly co-linear, such that if individuals were concerned about COVID, they also reported high concern for their family, others, etc. A simple PCA analysis showed that 8 of the 10 COVID questions (1-7, 10) loaded highly on the first component accounting for 40.4% of variance, while two questions (8 & 9) loaded highly on the second component, accounting for 14.1% of variance. Reliability measured in terms of Cronbach's α is 0.85 (lower confidence limit: 0.81, upper confidence limit: 0.87). While these questions were developed before COVID-19 related questions were publicly available from other groups, COVID-19 concern has been found to be related to a variety of factors in a cross-sectional group, including economic strain and stringency of self-quarantine behaviors [34]. By focusing on COVID-19 concern we are able to capture many aspects of students' lives, situations and opinions with one question, that we use here as ground truth.

3.4 Features

The smartphone sensing application runs in the background on both the iOS and Android phones. It passively collects data without any user interaction, although users can open the app if they wish to submit self-reports for the surveys without any prompt. The data is stored in the mobile locally and when the phone detects connection to the Internet, it uploads the data to HIPAA-compliant AWS servers before clearing the data from the device. Some of the features we collect are as follows:

Phone Usage. The mobile application records the number of phone locks and unlocks that the participants make. We derive (1) the number of phone locks and unlocks and (2) the average duration between phone locks and unlocks. These features act as a proxy of screen-time and phone usage. Researchers find that phone usage is correlated with depressive symptoms and anxiety [44, 59].

Mobility. The application samples GPS every 10 minutes, with the consideration of both energy conservation and data quality. We use this information to derive (1) the number of unique locations visited which are identified on the basis of DBSCAN [13], (2) distance travelled and (3) maximum distance from campus. Mobility features from mobile phones relate to anxiety and depression, based on several prior works [6, 44].

Physical Activity. We identify the activity that a participant is involved in with the help of activity recognition API supported by both iOS and Android phones.

Sleep. We derive sleep data from the phone. We compute (1) bedtime, (2) wake time and (3) sleep duration based on the method described in [8, 57]. Sleep data has a measurement error of +/- 32 minutes.

Semantic locations. We identify the home location of the participants based on where the students spend the majority of their time during the night. Also, we use a third party API [53] to attach semantic meanings to the raw Geo-location data. Several researchers report finding association between location types visited and mental health [4, 20].

Audio plays. The mobile app polls the system to start an audio session every 10 minutes. The system follows-up on the request by letting the app know whether there is any audio-based media (such as music, video) already playing on the phone. We use this information to compute (1) number of audio plays and (2) duration of audio plays. Prior research shows that people listen to music to regulate arousal and mood [46], so we include this in our analysis.

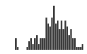
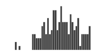
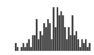
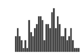
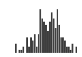
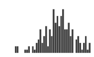
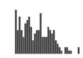
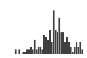
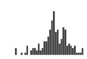
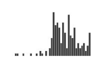
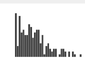
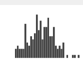
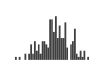
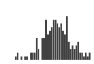
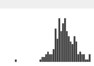
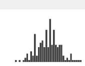
Regularity. We also compute the regularity index for each sensing feature using negation of approximate entropy [39]. Approximate entropy based regularity index considers the unpredictability of changes over time-series data. Meaning, time-series with higher uncertainty have higher irregularity. On the other hand, time series that are predictable have repetitive patterns, increasing their regularity. Regularity indexes are commonly explored in mental health sensing studies and many of them report finding a significant association between the two factors [6, 37].

4 Q1: HOW DOES STUDENTS' BEHAVIOR CHANGE DURING THE COVID-19 PERIOD COMPARED WITH A YEAR PRIOR TO THE PANDEMIC?

We begin our analysis by first exploring the change in behavior of the participants as a result of COVID. As we mentioned earlier in the Study Design section, the study has been ongoing since September 2017. Because of the availability of this longitudinally vast data, we are well-placed to perform a pre-post analysis investigating how students' behavior changes as they navigate through the pandemic.

In Figure 1, we visualize some passive sensing data collected since March 2019. The data is averaged over all participants throughout the year. The figures clearly show changes in students' behavior at the onset of the pandemic – we observe either a sudden uptick or a rapid step-down of the plots beginning around March 2020. COVID was declared a national emergency on March 13, 2020; therefore, we consider this specific date as the dividing point for our pre-post analysis. The vertical red line on the plot indicates this division. We can see that the sleep duration increases immediately following the national emergency declaration. Similarly, unlock duration also increases. However, the total distance travelled and walking duration decrease significantly. Perhaps what is interesting is that while other behaviors' trajectory seems to rise and fall, even reaching up to pre-March baseline levels in some cases, walking duration is down throughout the year and it barely reaches the baseline towards the end of March 2021. Although these figures are visually discriminating in terms of behavioral change, they are not best suited to compare each and every feature. In addition, it is also possible that the change in behavior beginning in March could

Table 2: Ground truth. The table below lists the ground truths that we collect as EMAs from the participants.

Ground Truth	Median	Std.	Range	Distribution
COVID				
1. How concerned are you about COVID-19?	4.00	1.48	1-7	
2. How much has the COVID-19 situation impacted your day to day activities in the last week?	4.00	1.80	1-7	
3. How much have you changed your behaviors in response to the COVID-19 situation in the last week?	4.00	1.82	1-7	
4. How concerned are you for yourself regarding COVID-19?	3.00	1.44	1-7	
5. How concerned are you for your classmates regarding COVID-19?	4.00	1.39	1-7	
6. How concerned are you for your family regarding COVID-19?	4.00	1.52	1-7	
7. How concerned are you about obtaining food, supplies, etc.?	2.00	1.53	1-7	
8. How supported do you feel?	4.00	1.49	1-7	
9. How much have you supported others?	4.00	1.36	1-7	
10. Is your social media usage: 1 (much less than normal) - 7 (much more than normal)	5.00	1.46	1-7	
PHQ-4				
1. Over the last 2 weeks, how often have you been bothered by the following problems? Feeling nervous, anxious or on edge; Not being able to stop or control worrying; Feeling down, depressed or hopeless; Little interest or pleasure in doing things;	2.00	2.80	0-12	
Self Esteem				
1. Right now, I worry about what other people think of me.	2.00	1.03	1-5	
2. Right now, I am pleased with my appearance.	3.00	1.02	1-5	
3. Right now, I feel as smart as others.	3.00	0.97	1-5	
Social Level				
1. Have you spent most of your time alone or with others today?	3.00	1.26	1-5	
Stress				
1. Are you feeling stressed now?	3.00	1.06	1-5	

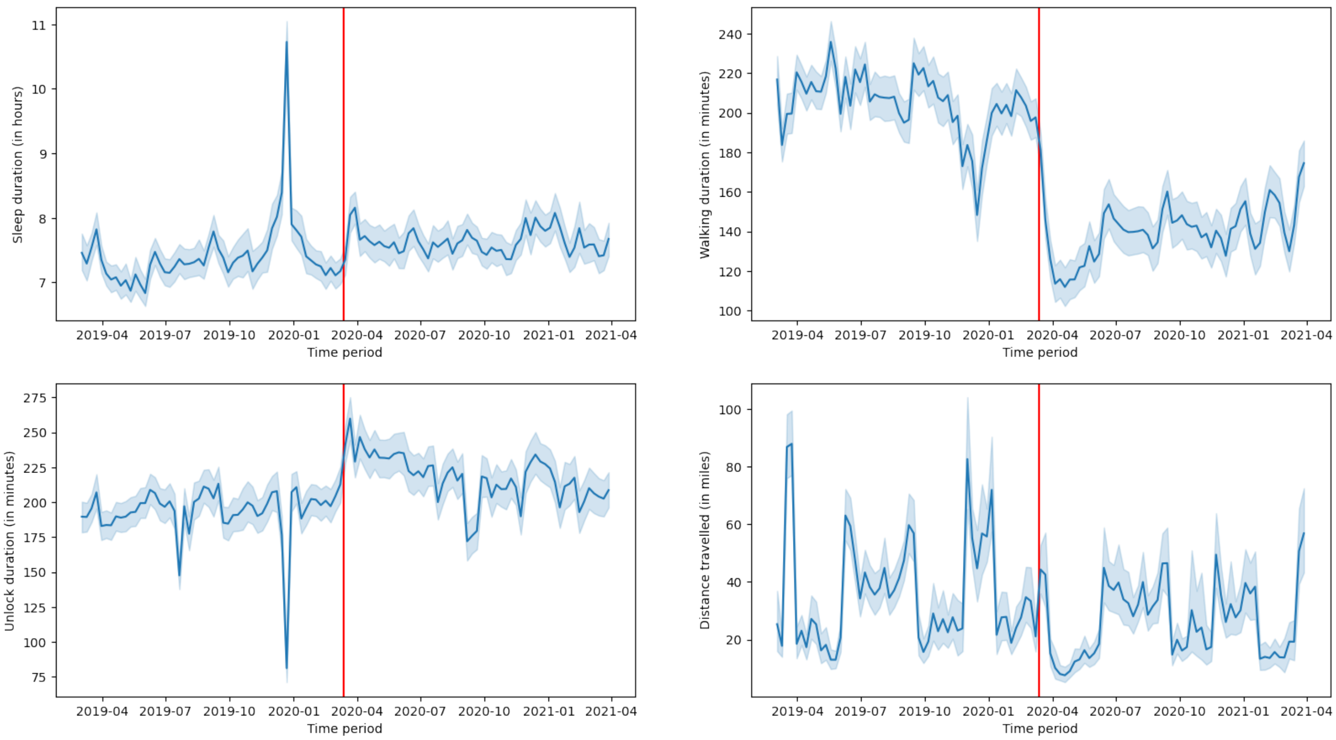


Figure 1: Changes in behavior pre- and post-March 13, 2020. The figure shows how several sensing features change over time as the COVID pandemic progresses in the United States (US). The vertical red line demarcates the period before and after March 13, 2020 when national emergency was declared in the US. The x-axis shows the time period and the y-axis, the feature. Note that the plot is drawn for weekly aggregated data, therefore, the horizontal bold blue line represents the overall average whereas the shaded regions represent the range of values. Plots generated over the entire data collection period are shown in the supplementary document.

simply be a result of other issues such as weather or the academic calendar. After all, the same day that the US declared national emergency, the university also ended its Winter semester. Therefore, for a more robust approach at identifying the behavioral changes that could be attributed to COVID, we take into consideration data belonging to an entire year. Note that we had a technical issue on the week of December 15, 2019 and only a handful of participants were active, as a result there is a sharp peak/valley around December 2019 that is most visible in the sleep duration and unlock duration plot. Please ignore the sudden peak and valley during that period, as it is heavily skewed.

More specifically, we compare the average value of the pre-COVID-19 baseline year with that of the first COVID-19 year. The use of data belonging to these two different periods allows us to perform an apples-to-apples comparison. We do this comparison with the help of Wilcoxon signed-rank test, as it is a non-parametric version of the paired t-test. We plot the result of the Wilcoxon signed-rank test in Figure 2; the y-axis contains the features that are significantly different between the two periods. Note, that the results being shown are statistically significant with a p-value of less than 0.05 after correcting for multiple comparisons using the Benjamini/Hochberg FDR correction procedure [3]. For the statistical test, we are only comparing daily averages. The divergent bars in the figure denote whether a particular feature changes positively or negatively during the first COVID-19 year (i.e., March 2020 - Feb 2021) compared to pre-COVID-19 baseline year (i.e., March 2019 - Feb 2020). The red dashed bars indicate a decrease, whereas the green solid bars indicate an increase. The x-axis shows percentage change, and the y-axis shows the name of the sensing behavior or feature. The x-axis shows the percentage change rather than absolute change in order to account for the differences in scales among different features.

While investigating the features that changed during the first COVID-19 year in comparison to the pre-COVID-19 baseline year, we find that the rate of decrease of features is more pronounced than the rate of increase. Some features that decrease with a higher rate include number of unique locations visited (-65%), distance travelled (-60%) and duration on bike (-50%). Similarly, number of phone unlocks (-24%), duration in vehicles (-21%) and walking duration (-33%) also decrease. On the other hand, there is an increase in time spent at home (+38%), number of audio played (+24%), maximum distance from campus (+21%), audio playing duration (+21%), phone unlock duration (+15%) and duration spent in a still position (+7%).

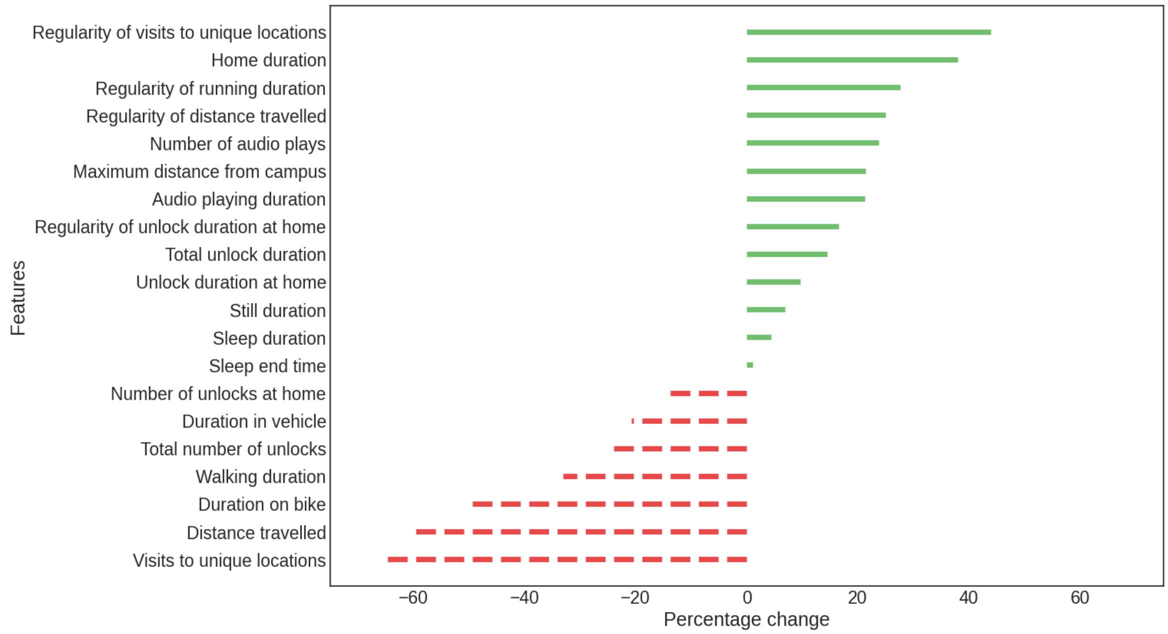


Figure 2: Features with significant differences between pre-COVID-19 baseline year vs first COVID-19 year. The figure shows the features that changed significantly and the associated percentage change. Red dashed bars denote that the feature decreased, whereas the green solid bars denote the feature increased in value during the COVID period. The comparison is based on the median value of the feature across the two time periods. The x-axis is the percentage change and the y-axis is the feature name. All results presented in this figure meet the following criteria: significant with a p-value of less than 0.05 after correcting for multiple comparisons using the Benjamini/Hochberg FDR correction procedure. Full statistical results are available in the supplementary document.

We also find that there is a small increase in sleep duration (+4%) and sleep end time (+2%). An increase in sleep end time means the participants wake up later than their baseline wake-up time. In terms of regularity, we find that the COVID-19 year is more regular than the pre-COVID-19 baseline year for a number of behaviors such as visits to unique locations (+44%), running duration (+27.6%), distance travelled (+25%) and unlock duration at home (+16.5%). It is interesting to note that while the number of unlocks decreases (-24%) during the pandemic period, the associated unlock duration (+15%) increases (aka screen time). This indicates that students are using their phones more in aggregate, with fewer but longer sessions in comparison to the baseline year.

5 Q2: DO ALL STUDENTS REACT TO THE PANDEMIC IN A SIMILAR MANNER?

In this section, we aim to better understand the behavior of the participants. We use cluster analysis to profile participants and generate descriptions of the clusters found. Participant profiling, sometimes also known as building personas, allows us to identify distinct groups of participants and draw inferences on the identified groups. Because the behavioral profile of students in the same group is similar, it helps us generate representative information for students belonging to each group. An unsupervised approach, such as, clustering, uses the inherent structure of the data to identify and group together similar data points. Therefore, it is suitable in our use-case where we want to identify and group together students with similar behavior without the consideration of any prior labels.

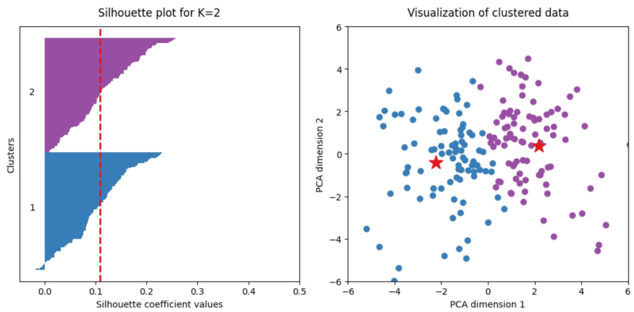
We cluster and profile students in two facets:

- *Based on self-reports.* We segment students based on their responses to the weekly EMAs we ask during the first COVID-19 year. We use the averaged response for the analyses.
- *Based on behavioral change.* We segment students based on the change in their behavior during the first COVID-19 year when compared with the pre-COVID-19 baseline year. In order to do this, we compute the difference between each student's first COVID-19 year features and pre-COVID-19 baseline year features. Then, we proceed with clustering and subsequent analyses on the resulting data. Note, features here refer to passive sensing data only.

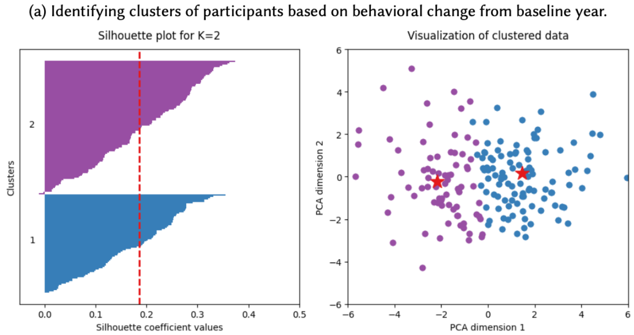
Given this breakdown, we explore whether certain groups of students in the study behave or deal with the pandemic differently than the other groups.

5.1 Clustering Process

Considering that we have many features and self-reports, we begin the procedure by first performing a Principal Component Analysis (PCA) to reduce the dimensionality of the data. We target 90% explanatory variance through principal components, retaining 19 features for behavioral mobile sensing data and 10 features for self-reports. We cluster the students using the principal components.



(a) Identifying clusters of participants based on behavioral change from baseline year.



(b) Identifying clusters of participants based on EMA response during COVID year.

Figure 3: Silhouette plot and clustered data. The figure on the left shows the silhouette plot when we cluster behavioral change and self-report data with number of clusters (K) = 2. The figure on the right is a representation of the clustered data points with two PCA loadings/dimensions. The vertical red dashed line represents the average silhouette score, red stars represent the centroid and the blue and magenta color indicate the data points belonging to two separate clusters.

Specifically, we use K-means clustering as our clustering algorithm, which aims to maximize the similarity within the data points in the same cluster and minimize the similarity between the data points in different clusters based on euclidean distance.

The key hyperparameter for K-means is K, the number of clusters to generate. We use the Silhouette method [42] to identify the optimal number of clusters K such that the data is segmented well among the clusters. In case of behavioral change facet, we obtain an average silhouette score of 0.11, 0.09 and 0.05 for K=2, 3 and 4, respectively. Similarly, for self-report facet, we obtain an average silhouette score of 0.18 for K=2, 0.14 for K=3 and 0.11 for K=4. In both cases, we find that K=2 leads to the highest average silhouette score. In Figure 3, we visualize the silhouette for K=2 for both the facets of behavioral change and self-reports. The x-axis shows the silhouette coefficient value, whereas the y-axis represents the clusters. Each colored point within the clusters is indicative of one data point. The average silhouette score is represented in the figure by a vertical red dashed line. Numerous points within each cluster score higher than the average silhouette score indicates that the grouping consists of similar points. Next to the silhouette score plot, we show a scatter plot of two principal components, color-coded based on the cluster they belong to after K-means clustering. The red stars represent the centroid of each cluster.

After grouping students into clusters based on multiple facets, we investigate how the data differ across each cluster and aim to provide a descriptive summary of the students as represented by the clusters. We do the clustering based on principal components and then describe the clusters based on the original data (i.e., the self-report, derived features) that we find to be significantly different between the two clusters. We compare the data among the two clusters based on the Kruskal-Wallis test. As we cannot make any assumption about the normality of the values we consider, we use the non-parametric Kruskal-Wallis test. The independent variable for the test is the cluster the data belongs to and the dependent variable is the behavioral feature/self-report. In Table 3 and 4, we list the top significant behavioral features and self-reports. All duration related features are in minutes and distances in miles. Note, that the listed data are all statistically significant with a p-value of less than 0.05 after correcting for multiple comparisons using Benjamini/Hochberg FDR correction procedure.

Table 3: Profiling based on EMA. The table below shows the differences between the clusters when participants are clustered on the basis of their self-reports, as per the result of Kruskal Wallis test. C_1 and C_2 represents the mean scores of two clusters, and *mean* refers to the mean of all the students. The *C.I.* column indicates the 95% confidence interval for the overall mean. Full statistical results are available in the supplementary document. Statistical significance reported after Benjamini/Hochberg FDR correction (*** $p < .001$, ** $.001 \leq p < .01$, * $.01 \leq p < .05$).

Self-reports	C_1	C_2	Mean	C.I.	H statistic
PHQ-4	1.78	4.21	2.91	[2.58, 3.24]	58.04***
Depression	0.94	2.18	1.52	[1.34, 1.68]	60.76***
Anxiety	0.84	2.03	1.40	[1.22, 1.57]	48.17***
Stress	2.42	3.06	2.72	[2.61, 2.81]	38.95***
Self-esteem	12.00	11.16	11.60	[11.32, 11.88]	8.02***
COVID-1	3.48	4.82	4.10	[3.94, 4.26]	75.30***
COVID-2	4.04	5.32	4.64	[4.45, 4.82]	46.01***
COVID-3	3.64	4.86	4.21	[4.02, 4.39]	43.69***
COVID-4	2.67	3.95	3.27	[3.09, 3.43]	57.78***
COVID-5	3.14	4.34	3.70	[3.55, 3.85]	66.20***
COVID-6	3.77	5.21	4.44	[4.26, 4.61]	68.58***
COVID-7	2.14	3.41	2.73	[2.55, 2.91]	49.64***
COVID-9	3.98	4.50	4.22	[4.06, 4.38]	9.88***
COVID-10	4.64	5.30	4.95	[4.79, 5.10]	21.92***

5.2 Examining Clusters based on Self-reported Differences

After performing the Kruskal-Wallis test, we find that social level and COVID-8 question (i.e., *How supported do you feel?*) are no longer significantly different between the two clusters. Therefore, we list only the remaining significant self-reports in Table 3. The C_1 and C_2 columns are the average value of the associated self-report within the clusters C_1 and C_2, respectively. The column identified as *mean* is the average value of the corresponding self-report for all the students. We observe that students in cluster C_2 (N=85, 47%) report higher scores in all COVID questions than participants in C_1 (N=95, 53%). In fact, cluster C_2 students score higher than the average score for each of the COVID questions. Cluster C_2 students also score higher than cluster C_1 students in PHQ-4, depression subscale, anxiety subscale and stress. The only self-report where students in cluster C_1 score higher (mean=12.00) than students in cluster C_2 (mean=11.16) is self-esteem.

It appears that the algorithm finds similarity in grouping together participants with higher scores in PHQ-4 and COVID questions. Interestingly, students with lower PHQ-4 and COVID scores have higher self-esteem than their counterparts. All in all, we find that C_2 students are much worse off and affected by COVID than C_1 students. Note that there is also a statistically significant difference between the number of self-reports submitted by the students belonging to the different clusters (Kruskal Wallis; p-val=0.025, H statistic=5.03). On average, students in cluster C_1 provided 65 self-reports per person, whereas the ones in cluster C_2 submitted 53 self-reports each. Meaning, students who are much worse off and affected by COVID provide fewer self-reports than the remaining students.

Table 4: Profiling based on behavior change. The table below shows the differences between the clusters when students are clustered on the basis of the change in their sensing data, as per the result of Kruskal-Wallis test. C_1 and C_2 represents the mean scores of two clusters, and *mean* refers to the mean of all the students. The *C.I.* column indicates the 95% confidence interval for the overall mean. Full statistical results are available in the supplementary document. Statistical significance reported after Benjamini/Hochberg FDR correction (*** $p < .001$, ** $.001 \le p < .01$, * $.01 \le p < .05$).

Features	C_1	C_2	Mean	C.I.	H statistic
Phone unlock duration per day	-7.93	57.14	26.15	[17.53, 34.77]	74.14***
Sleep start time	-18.60	12.75	-2.18	[-10.78, 6.37]	13.17***
Number of phone unlocks per day	-36.20	-15.74	-25.48	[-30.23, -20.74]	29.44***
Sleep end time	36.38	-4.00	15.23	[3.09, 27.38]	7.67**
Sleep duration	55.20	-16.74	17.44	[0.44, 34.2]	30.93***
Duration of audio plays per day	10.72	83.73	48.78	[35.85, 61.70]	66.61***
Still duration at home	95.09	-22.30	31.32	[9.24, 53.34]	33.91***
Distance travelled per day	-69.10	-9.74	-37.13	[-60.53, -13.74]	5.42*
Walking duration per day	-76.89	-34.36	-54.61	[-62.15, -47.05]	38.17***
Still duration per day	113.13	42.65	76.21	[66.21, 86.21]	63.92***
Regularity in duration on vehicle	-0.01	-0.16	-0.09	[-0.13, -0.05]	11.99*

5.3 Describing Clusters on Behavioral Change

We observe that cluster C_2 has the majority of the participants, i.e., N=94 (52%) whereas cluster C_1 has N=86 (48%) of the participants. Table 4 reports some top significant features and how their values differ among the clusters. Note, we report on aggregated features for the entire 24-hour day. As we mentioned earlier, for behavioral change facet, we subtract the first COVID-19 year features from the pre-COVID-19 baseline year and then proceed with clustering. This allows us to investigate the propensity with which the features change for a certain group of students in comparison to other groups. Negative numbers in Table 4 denote that the associated behavior decreases during the first COVID-19 year, whereas positive numbers indicate an increase in the behavior.

We notice several differences between the clusters. Students in cluster C_1 have an overall decrease in their phone usage duration (mean=-7.93 minutes) compared with their baseline, whereas students in cluster C_2 show an increase in their phone usage duration (mean=57.14 minutes). While the phone unlock duration differs among the two clusters, the number of unlocks is decreasing for the both of them, more so for students in cluster C_1 (mean=-36.20) than for students in cluster C_2 (mean=-15.74). While it may appear surprising that there is an increase in phone unlock duration but a decrease in number of phone unlocks for C_2, it is entirely possible because students could be using their phone for a longer period of time in a single go (i.e., unlock/lock); that is, there are fewer unlocks of the phone but when the phone is unlocked usage lasts longer (i.e., more screen time). Contrast this behavior with when the students are on campus. It makes sense that students might perform more locks and unlocks while moving around campus for example, visiting social places, cafeterias, gyms, in and around classes, talking with friends they might meet around campus, etc. In terms of sleep, we find that students in cluster C_1 go to bed earlier than baseline (mean=-18.60 minutes) and wake up later (mean=36.38 minutes) whereas students in cluster C_2 stay up longer than they did in the past-year (mean=12.75 minutes) and wake up earlier (mean=-4.0 minutes). As a result, the sleep duration for students in cluster C_1 increases (mean=55.20 minutes) whereas it decreases for students in cluster C_2 (mean=-16.74 minutes).

Students in both the clusters have an increase in audio play duration (discussed in Section 3.4), although cluster C_2 has a larger increase in audio play duration (mean=83.73 minutes) compared to cluster C_1 (mean=10.72 minutes). Interestingly, still duration at home increases for students in cluster C_1 (mean=95.09 minutes) but decreases in cluster C_2 (mean=-22.30 minutes). The distance travelled decreases by a higher rate for students in cluster C_1 (mean=-69.10 miles) than it does for participants in cluster C_2 (mean=-9.74 miles). Similarly, walking duration also decreases by a higher rate for cluster C_1 students (mean=-76.89 minutes) compared to C_2 students (mean=-34.36 minutes). However, the overall still duration increases for students in both clusters, more so for C_1 students (mean=113.13 minutes) than for C_2 students (mean=42.65 minutes). Students in cluster C_1 also have higher still duration than the average of all the students. Higher rate of decrease in distance travelled and walking duration whereas higher rate of increase in still duration at the same time might indicate that cluster C_1 students used to be more active in the past-year than C_2 students. We investigate this using only the data of pre-COVID-19 baseline and find that although C_1 students are slightly more active in past-year than C_2 students, the difference is not statistically significant (p-value > 0.05). However, when we focus only on COVID-19 year data, we find a significant difference between the two clusters in their activity level (p-value = 0.02). We observe that the students in cluster C_2 are more active even during COVID period compared to C_1 students (p-value=0.02). We also compare all the pre-COVID-19 features of the two clusters and find that there is not any statistically significant difference between the behaviors of students belonging to the two clusters prior to the pandemic (p-value > 0.05).

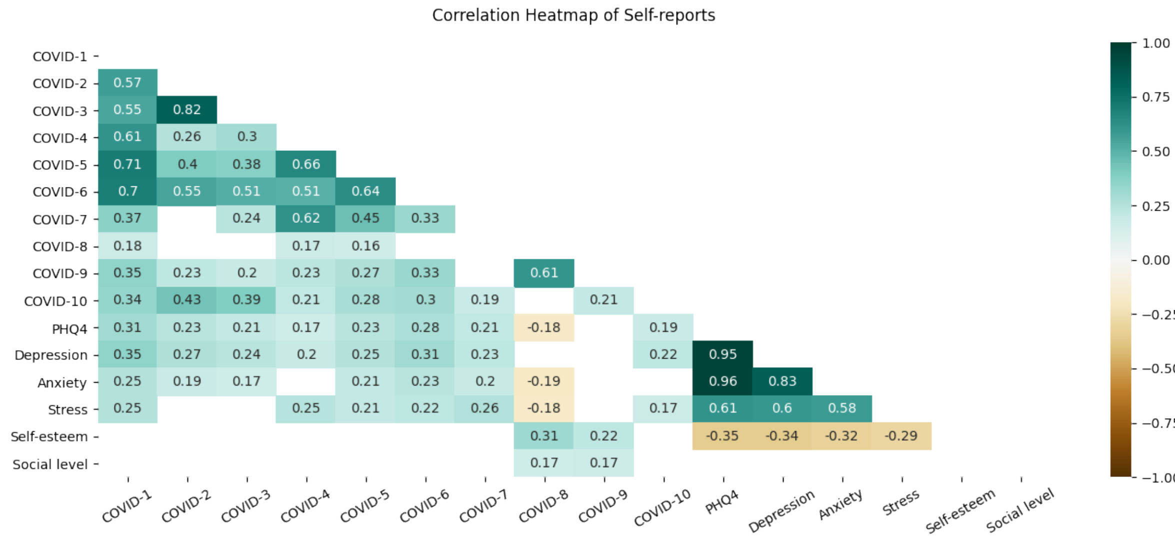


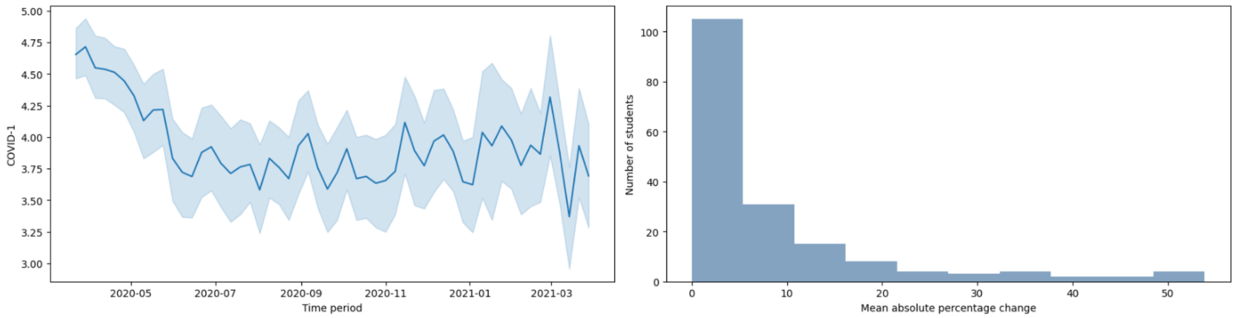
Figure 4: Correlation among different EMAs. The figure shows the correlation coefficient of all the EMAs that we collect. The darker green the tint, the stronger the relation between the variables. White, blanked out grids indicate that the results are not statistically significant (i.e., p-value is greater than 0.05). Full statistical results are available in the supplementary document.

6 Q3: IS CONCERN FOR COVID ASSOCIATED WITH THE MENTAL HEALTH OF STUDENTS?

After clustering the students based on their self-reported responses to the EMAs, we observe that students who score higher in PHQ-4 and COVID questions are grouped together by the algorithm. This begs whether these EMAs are related. The COVID questions primarily ask students about their COVID concerns, its impact on their behavior and whether the students feel supported during the pandemic. We want to explore if these are somehow related to the depression, anxiety and stress that students report during the pandemic. To investigate the relationship among the EMAs, we perform a correlation analysis. We use Spearman's rank-order correlation since the responses are measured in Likert scale, which is ordinal in nature. We show the result of the correlation analysis in Figure 4. Note, that the figure contains only statistically significant results, i.e., all the presented correlations are significant with a p-value of less than 0.05 after adjusting using Benjamini/Hochberg FDR correction procedure. Non-significant and identity correlations are blanked out and appear as a white space in the heatmap.

The x-axis and the y-axis represent the self-reports and the values contained therein indicate the correlation coefficient. We observe that the COVID-1 question (on the x-axis) is strongly correlated with most of the other COVID questions: COVID-2 (ρ=0.57), COVID-3 (ρ=0.55), COVID-4 (ρ=0.61), COVID-5 (ρ=0.71), COVID-6 (ρ=0.7) and moderately correlated with most of the remaining COVID questions: COVID-7 (ρ=0.37) and COVID-9 (ρ=0.35). We also find that COVID concern (COVID-1) is moderately correlated with change in social media usage, i.e., COVID-10 (ρ=0.34). Note that the median score of COVID-10 responses is 5 (on a scale of 1-7, where 4 is normal), indicating that the majority of the participants had an increase in their social media usage during the pandemic. 104 participants responded with a score higher than 4, while 76 participants responded with 4 or less. COVID-1 question also has the strongest correlation with PHQ-4 (ρ=0.31) among all the COVID questionnaires. It has moderate correlations with both the depression sub-scale (ρ=0.35) and anxiety sub-scale (ρ=0.25) and also with the stress question (ρ=0.25). Self-esteem is only positively correlated with COVID-8 (ρ=0.31) and COVID-9 (ρ=0.22). Interestingly, it has a negative correlation with PHQ-4 (ρ=-0.35) and stress (ρ=-0.29). Social level is weakly correlated with COVID-8 (ρ=0.17) and COVID-9 (ρ=0.17) only.

The correlation analysis shows that the COVID-1 question, "*How concerned are you about COVID-19?*", appears to have moderate-to-strong correlation with all the other remaining COVID questions. In some manner, such correlation is expected, since the question asks students about their COVID concern in a much broader sense. While other questions get specific about how concerned students are about their family members, or the availability of food, or the support they receive, a direct question asking about their overall concern can be assumed to at least capture some aspect of information from all these more specific questions. Because of the broader sense of the question, the COVID-1 question is also moderately correlated with PHQ-4, its subscales and stress. Among all the other specific questions about COVID, COVID-1 has the highest correlation with PHQ-4 and both of its subscales on depression and anxiety. *Motivated by this important insight, we examine in section 8 whether we can use passive sensing data solely from students' phones to predict a student's response to the COVID-1 question; that is, whether we can predict a student's COVID concern from their phone data.* Because COVID concern is correlated with PHQ-4, depression, anxiety and stress, we believe that it is an important problem to explore. If we can predict COVID concern of students, we may be able to detect



(a) COVID concern over time.

(b) Percentage change in COVID concern.

Figure 5: Overall COVID-1 response and percentage change. Figure (a) represents the change in COVID concern of participants over time. The solid blue line represents the average value, whereas the shaded regions represent the range of values. Figure (b) is the histogram of percentage change in students' COVID-1 responses. The x-axis represents the absolute mean percentage change and the y-axis is the count of students.

vulnerable student population and provide proactive help to them during the pandemic (e.g., digital interventions).

7 Q4: HOW DOES THE COHORTS' COVID CONCERN CHANGE AS THE PANDEMIC PROGRESSES?

From the earlier section, we learned that among all the COVID related questions that we ask the students, COVID concern (i.e., COVID-1) has the strongest association with their mental health. Knowing that it is an important factor in play for students' mental health during the pandemic, we want to explore how it changes over time as the pandemic progresses. In Figure 5a, we visualize student's responses to COVID-1 question beginning on March 2020 all the way up to March 2021. As we can see on the plot, students start with a highly elevated concern for COVID in March 2020. The overall concern goes down after June 2020 but it still stays consistent around 4 (on a range of 1-7). We observe a few peaks during early September, mid-November, mid-December in the year 2020 and also at the start of the year 2021. COVID concern goes to 3 or below for a few participants around March 2021. In terms of the response rate of the participants, we have 4163 responses to COVID-1 in total from 180 participants with each participant submitting 23 self-reports on average. We observe that 56 out of 180 participants always report a heightened COVID concern of 4 or higher throughout the first year of COVID-19 pandemic.

Next, we explore the volatility in the COVID-1 responses of the students. To do this, we generate the percentage change in each student's responses for the COVID-1 question in the order they are received. Then we take an average of all the percentage changes to arrive at the mean percentage change for each student's COVID-1 response. The final number thus calculated represents how consistent or variable students are in their self-reported COVID concern. In Figure 5b, we show a histogram of mean absolute percentage change of the students. The overall average percentage change is 8%. The majority of the students' responses are consistent (N=128 have $\leq$ 8% change) with only 29 students' mean percentage change being higher than 15%. In addition, when we consider the change in response based only on the difference between the first COVID-1 self-report at the beginning of March 2020 and the final COVID-1 response at the end of March 2021, we find that majority of the students (N=104) start with a higher COVID concern but by their final response, they report a much lower COVID concern. However, of the remaining 76 students, 33 have an increase in their COVID concern in comparison to their initial self-report, whereas 43 students' report having the same concern they had when they started out in March 2020.

8 Q5: CAN WE BUILD A MACHINE LEARNING MODEL TO PREDICT STUDENT'S COVID CONCERN?

In this section, we examine whether we can predict COVID concern of the students based solely on their mobile sensing data. Understanding this is important for two reasons: 1) it could shed some light into the relationship between behavioral features and COVID concern and 2) if we can achieve satisfactory performance, such a model could be useful as part of an early detection system to identify vulnerable students and provide proactive help to them. During the COVID-19 pandemic, behavior of individuals is quite diverse, with some individuals rarely leaving their home and others continuing on about their daily life as if little had changed. These distinct behaviors are likely due to a variety of factors ranging from COVID-19 case count in the local area, local laws and personality traits. To optimally capture these differences with one single questions, we focus on the first question in the COVID-19 survey for this predictive analysis, i.e., "*How concerned are you about COVID-19?*". As mentioned in section 6, this broader COVID question is correlated with depression, anxiety and stress; and, motivated by this insight, we examine whether we can use passive sensing data solely from students' phones to predict a student's response to the COVID-1 question; that is, whether we can predict a student's COVID concern from their phone data. First, we approach this

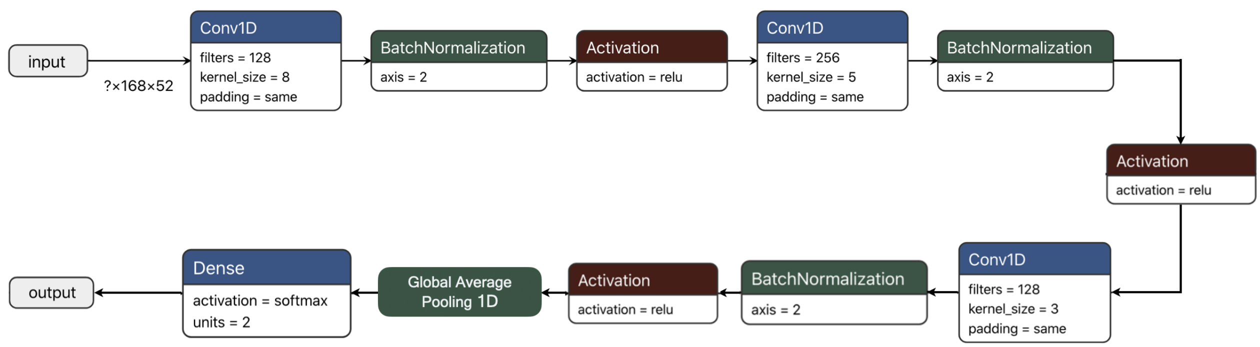


Figure 6: Architecture of Fully Convolutional Neural Network (FCNN). The figure shows the architectural overview of the deep learning model that outperforms all the other models we try. The blue boxes represent the layers in the network; Conv1D is referring to 1D convolutional layers and Dense, to fully connected layer.

problem as a simple binary classification problem, dividing the self-reported COVID-19 concern into either higher or lower class using a median split. As discussed earlier in the paper, students responded to the survey question using a Likert scale ranging from 1 to 7. The median score is 4 and after dividing into two classes based on the median, we have 2886 self-reports in the higher COVID concern class, and 1277 self-reports in the lower COVID concern class.

As part of the pre-processing step, we impute the features that we collect based on forward fill (firstly) and backward fill (if there are still missing values) approach, which are a standard set of techniques used to handle missing values in time series modeling. This imputation approach carries the most-recent non-missing observation forward (or backward) and replaces the missing value with it. After imputation, we standardize all the features within- and between-person. In terms of modeling, we treat the predictive task as a time series classification problem. As deep learning approaches have done exceptionally well in sequential data, such as text, audio, financial data and time series, we explore the utility of several state-of-the-art deep learning time series classification models within our dataset. Firstly, we convert our dataset into a time series suitable format. For each self-report, we take into account the last 7 days' data with the granularity of an hour; that is, each sample/data point contain 168 values for a single feature (i.e., 7 times 24 values). Because there are 52 features in total, each data point is therefore an array of size 168x52. We divide the dataset into train, validation and test splits of 60:40. Meaning, we use 60% of the data to train the model, and the remaining 40% of the data to validate (20%) and test (20%) the model.

We ran several deep learning based models that have been previously shown to work well for a multivariate time series classification task. These include, Time-Convolutional Neural Network (CNN) [64], InceptionTime [22], Multi-Channel Deep Convolutional Neural Network (MCDCNN) [65], Residual Neural Network (ResNet) [63], Multi-Layer Perceptron (MLP), Time Warping Invariant Echo State Networks (TWIESN) [52], Long short-term memory (LSTM) [49] and Fully Convolutional Neural Networks (FCNN) [63]. Models are evaluated using Area under the curve (AUROC), precision score, recall score and F1 score metrics. The performance of these models is listed in Table 5. We observe that FCNNs outperform other deep learning based approaches, obtaining a well-balanced overall score across all the evaluation metrics; specifically, an AUROC of 0.70, F1 score of 0.71 and precision and recall score of 0.75 and 0.70, respectively.

Table 5: Prediction result. The table below lists results of different deep learning models performing predictions on the COVID concern question. Note, all the reported metrics are weighted to account for label imbalance, resulting in an F1-score that is not between precision and recall.

Models	Performance scores			
	Precision	Recall	F1	AUROC
Time-Convolutional Neural Network (CNN)	0.54	0.74	0.63	0.50
InceptionTime	0.68	0.59	0.62	0.60
Multi-Channel Deep Convolutional Neural Network (MCDCNN)	0.63	0.67	0.65	0.52
Residual Neural Network (ResNet)	0.73	0.57	0.60	0.64
Multi-Layer Perceptron (MLP)	0.54	0.74	0.63	0.50
Time Warping Invariant Echo State Network (TWIESN)	0.80	0.74	0.63	0.50
Long short-term memory (LSTM)	0.70	0.61	0.62	0.54
Fully Convolutional Neural Network (FCNN)	**0.75**	**0.70**	**0.71**	**0.70**

The architecture of our FCNN model is similar to the one proposed by Wang et al. [63] – it consists of 3 1D convolution layers and one dense output layer. The architecture is shown in Figure 6. As shown in the figure, each convolutional block (or layer) is followed by a batch normalization operation, with the result fed into a rectified linear unit (ReLU) activation function. After the third and final convolutional block, we perform global average pooling, which averages the output of the final convolutional block over the time dimension; thereby, flattening the output to one long vector that is passed through a fully connected layer to the output to make the prediction. The fully connected (or dense) output layer at the end uses a softmax function to return probabilities for the two

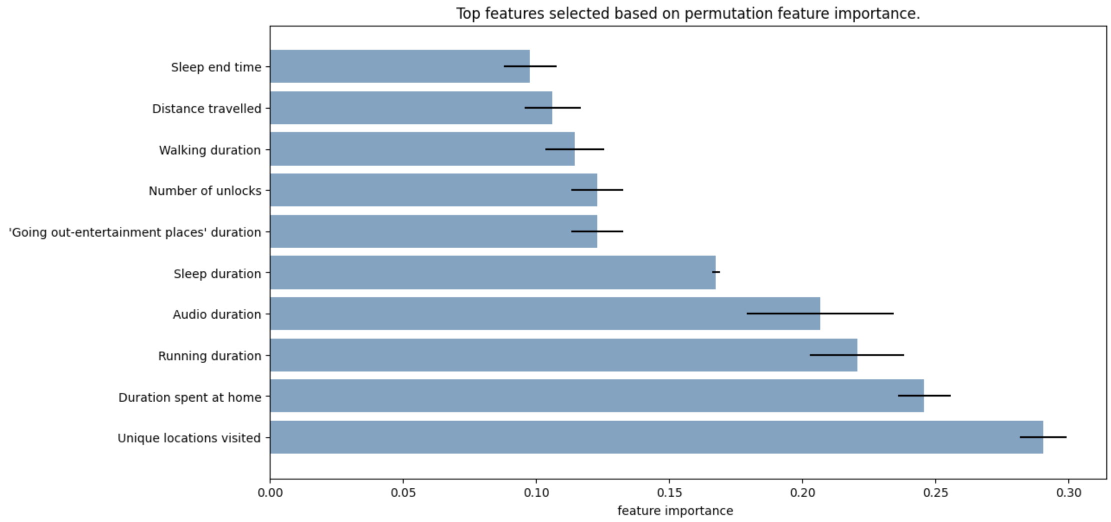


Figure 7: Important features. The figure shows some of the most important features obtained from permutation feature importance approach. The x-axis is the feature importance score. Note that the error bars represent the standard deviation of the importance scores over 5 different runs of permutation.

classes. Figure 6 also shows the attributes of each layer. The input to the model is a 3D tensor where each data point is of size 168x52. If there were 100 data points, for instance, we would have an input tensor of size 100x168x52. All convolutional layers use a stride of 1 and same padding (aka zero padding) so that the size of the output after the convolution operation is the same as the input. The first convolutional layer contains 128 filters with a kernel size of 8. The second convolutional layer contains 256 filters with a kernel size of 5, followed by the final convolutional layer, which contains 128 filters and kernel of size 3. There is no pooling or regularization operation, other than the global average pooling prior to the dense layer. We use a batch size of 8 and 200 iterations during the training phase. Models are optimized using Adam optimization. For hyperparameters, we start with some baseline parameters and play around with manual hyperparameter tuning, focussing on specific search candidates. We do not perform any intense hyperparameter tuning, such as, random search or Bayesian optimization approaches.

A natural next step after modeling is to understand which features play an important role in prediction of COVID concerns for students. This is particularly challenging in the case of deep models because they represent "black-box" techniques. Making deep models more interpretable is an important and active research problem in the deep learning community. For our analysis, we use a frequently implemented model agnostic and easy-to-understand approach to investigate feature importance and interpretability. Model agnostic approaches of interpretability do not depend on the specifics or the technicalities of the model, but solely at their predictions/outputs. One of such approaches that we employ is termed "permutation feature importance" [15]. This approach focuses on a single feature at a time and randomly shuffles all its values. We then compare the score of the model when predicting on the original dataset (i.e., with no shuffling) versus when predicting on the modified dataset (where a single feature has been shuffled).

Intuitively, this means that we are trying to break the relationship between the feature and the target such that the change in prediction score after the shuffle, should be indicative of how important that feature is for that particular model. We perform the same process for each feature separately, where only one feature is shuffled every single time. In addition, the permutation is repeated for each feature 5 times. We report some of the most important features based on this process in Figure 7. The x-axis represents the feature importance score, whereas the y-axis shows the feature name. The error bars represent the standard deviation of the scores over the 5 permutations we perform for each feature. Based on permutation feature importance technique discussed, we find that the number of unique locations visited, duration spent at home, running duration, audio play duration and sleep duration matter the most when predicting student COVID concern.

9 DISCUSSION

9.1 Summary of results

Throughout this paper, we examine students' behavior and how it is influenced by the COVID-19 pandemic period, a period of unprecedented societal change. While a number of studies have investigated the behavioral changes of several population groups during the COVID-19 pandemic, we consider the entire year's data right from the national emergency declaration in the USA in March 2020 to February 2021 – the first COVID-19 year in the United States. We compare the pre-COVID-19 baseline year with an entire first year of COVID-19, shedding light on several important behavioral changes captured by the mobile sensing features; this approach differentiates our study to other similar studies, because we collect objective, longitudinal data spanning a two-year period. We find that students' behaviors shifted and time spent engaging in multiple physical and social activities substantially reduced during the pandemic. Sensing features such as number of unlocks at home, duration in vehicle and walking duration have a negative change during the pandemic compared with that of the baseline year. However, features such as time spent at home, total audio playing duration and total unlock duration increase during the pandemic. Several of our findings that relate to the behavioral trends of the students are supported by prior studies. For instance, consistent with our finding that there is a drop in walking duration, duration on bike and distance travelled, existing work that use objective data report finding a decrease in physical activity levels during the pandemic [21, 30, 38]. Similarly, the increase in phone usage and sleep duration that we observe replicates work by other researchers within populations they study [45, 50]. It is logical that people would be socializing less (i.e., not going out as often as they used to) if they are following social distancing directives and thus would experience increased time spent at home and a longer duration of being stationary. These students are less likely to be traveling, leading to an expected decrease in walking duration, duration in vehicle and distance travelled, as well as visits to unique locations. As a result of not getting out of the house as much and be social in-person, students may resort to the use of social media apps and phones to keep themselves engaged with their friends and families, increasing the phone usage duration as we observe in our analysis. Also, the self-reported responses for social media usage (i.e., COVID-10 question) indicate that majority of the students in our study had an increase in their social media usage during the pandemic.

In addition to the broad analysis, we perform several analyses that are unique to our work and has not been pursued by other studies focussed on COVID-19. For instance, we cluster participants based on their survey responses and mobile sensing data in order to determine if there are certain groups of students who react differently to the pandemic than their counterparts. When clustering based on self-reports, we find that students who have a higher-than-average PHQ-4, COVID responses and stress level, fall in one cluster while students who score lower on all these self-reports fall in the remaining cluster. We also find that students who fall in the "low" cluster have higher self-esteem compared to students in the "high" cluster. The result from this analysis suggests that within our undergraduate student cohort, high COVID concern (as obtained from the COVID questionnaire) is related to higher depression, anxiety and stress. This is expected because COVID-19 is traumatic for many, with actual physical, social and economic consequences. Studies show that COVID-19 is associated with mental illness [14, 26], increasing depression and anxiety among several populations including college students [7]. This is also observed within our dataset, when we investigate the relationship between the self-reports. Our correlation analysis shows that COVID concern and PHQ-4 have a positive relationship. Specifically, COVID concern is positively correlated with PHQ-4, its depression subscale (PHQ2), its anxiety subscale (GAD2) and stress. It is worth mentioning that we observe a negative correlation of COVID-8 question (i.e., *How supported do you feel?*) with stress, anxiety and depression. This finding is interesting, particularly, in context of real-world stressors, such as, the COVID-19 pandemic as it may relate with the idea of the 'stress buffering hypothesis'. The idea suggests that social support relates to better mental health specifically in times of stress because people can feel like they have others to support them during those challenging times [9]. Following this, we cluster students based on their behavior change and observe several differences between the clusters. While phone usage duration increases for students in cluster 2, we find that students in cluster 1 spend less time using their phone. Cluster 1 students go to bed earlier, wake up later and sleep for longer durations compared to their baseline behavior. However, cluster 2 students go to bed later, wake up earlier and sleep for shorter duration than they did in the pre-COVID baseline year. In addition, we explore the change in COVID-1 responses of the students. We find that the majority of the students are consistent in their COVID-1 responses and most importantly, 31% of them report a heightened COVID concern throughout the first COVID-19 year.

Motivated by our understanding that the COVID-1 question reflects students' overall COVID-19 concern related to mental health, we perform a predictive analysis of COVID concern. This further highlights the uniqueness of our work. While prior analyses compare changes to the baseline behavior, in the predictive analysis work, we only consider behavior during the pandemic period. We obtain a satisfactory result, achieving an AUC of up to 0.70. Using the permutation feature importance technique, we identify features that are important to our model. It is, however, important to understand that feature importance values do not reflect the intrinsic value of a feature towards prediction, it simply shows how important the feature is for our particular model. We observe that activity related features turn out to be important, along with phone usage, visits to entertainment places, sleep and distance travelled. In other words, how active or sedentary one's life is, the phone engagement and sleep plays a role in identifying COVID concern. This also makes sense through the lens of mental health, since prior research shows that mental health is related with physical activity, phone usage and sleep [47, 59]. There is also a relationship between mobility and mental health [33], and relatively recently researchers report finding that the number of unique places frequented relates to positive mental health [17]. Overall, our findings reveal a number of interesting insights.

9.2 Implications

Our work offers several important implications. First, we find association between different COVID experiences (based on self-reports) with the student's depression, anxiety, stress, self-esteem and social-level. This sheds some light on how the pandemic impacts the mental health of college students. Being able to understand the concern and experiences of students can help mitigate behaviors and factors that induce it and lower the associated negative effects. In addition, to the best of our knowledge, the behavior change analysis is the first of its kind comparing young adult population's pre-COVID-19 baseline year with their first COVID-19 year. Observations based on the clustering approach show how different groups of young college students react to the pandemic. While existing studies may have looked into how people, in general, are affected by the pandemic, by profiling participants we are able to address gaps in our understanding of how subgroups of specific populations may respond to traumatic events differently. The lack of published studies identifying mental health or behavior clusters related to COVID-19 behaviors highlights the need to identify such segments within our populations and study them separately instead of generalizing findings towards all communities and populations. The insights from such studies would be more far-reaching and useful for each population segment, as public policies (such as improving wellbeing, mental health) could then be targeted in a personalized manner towards the segment of population under consideration and their specific experiences. With the help of passive sensing and other ubiquitous technologies, we may be able to identify the population groups while being sensitive to the individual differences and provide more tailored help and suggestion as forms of digital interventions. This adds another implication of our current work – the findings may be applied to help make recommendations to individuals during time of crisis, such as a pandemic. For instance, intervening Human Computer Interaction (HCI) designs could make use of our findings to provide interventions, such as, cognitive behavioral therapy (CBT) and suggestions for lifestyle modification and self-care focusing on aspects, such as, moderating phone usage, maintaining physical activity levels and sleep. This is especially important because such variables have been found to be related to mental health and well-being.

On a broader level, passive sensing devices, such as phones, provide the ability to understand change in behavior as a result of the pandemic as well as lockdowns and quarantines that come along with it. As a result, they can not only be used to monitor the consequences of future pandemics and the social distancing measures as they are implemented, but can also be used to determine if behaviors return to baseline as the pandemic wanes and life returns to normal. Therefore, future studies could make use of the devices for public health promotion before as well as during and after the pandemic. In addition, these ubiquitous devices provide mobility and other behavioral information which can be used to understand pandemic trends (e.g., COVID fatigue) and act proactively. Sensing technologies can help us understand novel events such as COVID-19 and can be implemented before researchers have time to develop questions specific to the novel event. Perhaps more importantly, we show that the data available from passive devices can provide an estimate of how concerned someone is about the COVID-19 pandemic. This is useful because, for one, it can be used to detect students who might be more prone to mental health issues during the pandemic (COVID concern has correlation with depression/anxiety) allowing us to provide proactive help to them. This could be achieved by integrating the application with college health centers with participant consent. On a larger scale, since pandemics lead to an increase in suicidal ideation, substance use and depression [10, 41], such applications could also be useful in identifying the most vulnerable people across an entire population. We envision this as an early detection system that could provide individualized suggestions to help people navigate difficult times, or notify a provider with their consent so they can reach out to the individual to provide assistance. Such an intervention system is part of our future work. Finally, viewing large-scale changes in behaviors that vary over time and across individuals with respect to COVID-19 represents a critical perspective, such that behaviors are not stationary and that the models to predict behaviors in the future should be closely monitored for drift.

A number of critically important issues surface when considering the possible future integration of mobile sensing into predictive and intervention systems (e.g., personalized intervention, looping in student heath centers, etc). These include many design and systems implementation issues, such as, transparency (i.e., the user is fully informed about the app function and data usage), validity (i.e., issues such as predictive performance, replication of results and importantly generalizability of predictive engines) and critically important, privacy (i.e., designing systems that protect the privacy of users). A large amount of future research and larger scale studies are required to address these open challenges.

10 LIMITATIONS

Our work has several limitations, the most glaring of which is the fact that the study is investigating a small sample of students with specific demographics in one particular college in the USA. Therefore, we would urge caution against making any sweeping generalizations based on our results. We need to perform further studies to validate whether we can obtain similar findings across different populations (e.g., other universities). However, even though the sample size is relatively small, we find significant relationships from the data, which support our results. Given the limited demographic representation in the current population, we likely underestimate the number of clusters existing in across the entire population. We believe that this reinforces the importance of the current cluster analysis and suggest that other researchers should include similar analyses in their work. Further, as we mention throughout the paper, several of the results we observe are also obtained by earlier researchers among different populations across different countries. It is possible that we may have introduced self-selection bias by studying only that population of students who choose to take part in it. We also use self-reported responses to different inventories, some of which are created in-house and are not validated effectively, likely introducing several biases in the ground truth. Another limitation relates to the reliability of APIs (Application Programming Interface) and algorithms used by the mobile phones in this study: we deduce the level of physical activity (e.g., steps, activities, mobility, etc.) based off of data received from mobile phones without detail knowledge of any inherent error in these measurement and inference systems. For example, most behavioral studies like ours use the output from iOS/Android activity recognition without the companies reporting the actual performance metrics of these algorithms. It would be helpful if Apple and Google published this device/algorithm performance data on a regular basis. What we do know is that these algorithms are trained using a massive amount of data and likely offer population scale performance suitable for in the wild studies like ours [5].

Some limitations of our work also open up opportunities for future research. Although we use a standard empirical approach for identifying optimal cluster arrangement, it may be preferable to evaluate clusters or goodness of fit based on domain expertise. It could be especially more crucial when studying a sensitive and important topic such as mental health. While our modeling approach for prediction is sound, we have no way of identifying whether the architecture has been overfitted for our specific dataset. We handle overfitting as much as possible within our dataset, but we cannot know for certain if it is achieving high performance only in our dataset. Similar studies need to be conducted in the field to investigate whether the modeling approach can be generalized to other similar problems. Finally, we would also argue that the measures we use to collect the data are not sufficient to conduct a thorough analysis of the results. Use of multimodal sensing (such as wearables, phones, Bluetooth trackers, social media) along with the combination of several other ground truths would lead to a more holistic view of the participants' behavior enabling us to make strong conclusions.

11 CONCLUSION

In this paper, we studied the impact of COVID on behavioral changes of undergraduate students with objective mobile sensing data alongside several self-reported measures. We compared their behavior in the year prior to COVID with their behavior during the first year of COVID. With the initial analysis grouping everyone together, we found that there is mostly a negative change in student's behavior, leading to a more sedentary lifestyle during the first COVID-19 year. To better understand the changes, we investigated whether there are certain subgroups of students that reacted differently to the pandemic. Profiling students based on a clustering approach revealed several distinct characteristics among subgroups of students. We found that students with high COVID concern have a heightened PHQ-4, anxiety, depression and stress. We identified a particular group of students who are more concerning because of their heightened COVID concern and high score in mental health metrics. Correlation analysis corroborated our finding showing that there is a moderate correlation between COVID concern and PHQ-4, depression, anxiety and stress level of students. We also trained a deep learning model to predict the COVID concern of the students, obtaining an AUROC of 0.70.

We believe our findings pave the way for further research to investigate how pandemics impact the mental health of young adults at colleges. Importantly, we are now in the second year of the COVID-19 pandemic and while many students are hopeful that their lives will eventually return to normal in the near future, it looks like we may be living with the pandemic and its variants for sometime to come. In fact, we have little visibility of the end of the pandemic with any certainty. As a result, we believe that the insights from our work on finding connections between phones and human behaviors associated with the pandemic remains very relevant going forward, in particular for students, faculty, researchers and university administrators who are now dealing with the challenges associated with the next phase of the pandemic.

ACKNOWLEDGMENTS

This work was supported by National Institute of Mental Health, grant number 5R01MH059282.

REFERENCES

[1] B. Arroll, F. Goodyear-Smith, S. Crengle, J. Gunn, N. Kerse, T. Fishman, K. Falloon, and S. Hatcher. 2010. Validation of PHQ-2 and PHQ-9 to Screen for Major Depression in the Primary Care Population. *The Annals of Family Medicine* 8, 4 (July 2010), 348–353. https://doi.org/10.1370/afm.1139

[2] Dror Ben-Zeev, Rachel Brian, Rui Wang, Weichen Wang, Andrew T Campbell, Min SH Aung, Michael Merrill, Vincent WS Tseng, Tanzeem Choudhury, Marta Hauser, et al. 2017. CrossCheck: Integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse. *Psychiatric rehabilitation journal* 40, 3 (2017), 266.

[3] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57, 1 (1995), 289–300. http://www.jstor.org/stable/2346101

[4] Mehdi Boukhechba, Philip Chow, Karl Fua, Bethany A Teachman, and Laura E Barnes. 2018. Predicting Social Anxiety From Global Positioning System Traces of College Students: Feasibility Study. *JMIR Mental Health* 5, 3 (July 2018), e10101. https://doi.org/10.2196/10101

[5] Andrew Campbell. 2021. Personal communication with Andrew Campbell who worked in the Android sensing group, Google, 2016-2017. Private Communication.

[6] Luca Canzian and Mirco Musolesi. 2015. Trajectories of Depression: Unobtrusive Monitoring of Depressive States by Means of Smartphone Mobility Traces Analysis. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 1293–1304. https://doi.org/10.1145/2750858.2805845

[7] Wenjun Cao, Ziwei Fang, Guoqiang Hou, Mei Han, Xinrong Xu, Jiaxin Dong, and Jianzhong Zheng. 2020. The psychological impact of the COVID-19 epidemic on college students in China. *Psychiatry Research* 287 (2020), 112934. https://doi.org/10.1016/j.psychres.2020.112934

[8] Zhenyu Chen, Mu Lin, Fanglin Chen, Nicholas D. Lane, Giuseppe Cardone, Rui Wang, Tianxing Li, Yiqiang Chen, Tanzeem Choudhury, and Andrew T. Campbell. 2013. Unobtrusive sleep monitoring using smartphones. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*. 145–152.

[9] Sheldon Cohen and Thomas A. Wills. 1985. Stress, social support, and the buffering hypothesis. *Psychological Bulletin* 98, 2 (Sept. 1985), 310–357. https://doi.org/10.1037/0033-2909.98.2.310

[10] Mark É. Czeisler, Rashon I. Lane, Emiko Petrosky, Joshua F. Wiley, Aleta Christensen, Rashid Njai, Matthew D. Weaver, Rebecca Robbins, Elise R. Facer-Childs, Laura K. Barger, Charles A. Czeisler, Mark E. Howard, and Shantha M.W. Rajaratnam. 2020. Mental Health, Substance Use, and Suicidal Ideation During the COVID-19 Pandemic — United States, June 24–30, 2020. *MMWR. Morbidity and Mortality Weekly Report* 69, 32 (Aug. 2020), 1049–1057. https://doi.org/10.15585/mmwr.mm6932a1

[11] Vedant Das Swain, Koustuv Saha, Hemang Rajvanshy, Anusha Sirigiri, Julie M Gregg, Suwen Lin, Gonzalo J Martinez, Stephen M Mattingly, Shayan Mirjafari, Raghu Mulukutla, et al. 2019. A multisensor person-centered approach to understand the role of daily activities in job performance with organizational personas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–27.

[12] Yves-Alexandre de Montjoye, Jordi Quoidbach, Florent Robic, and Alex Pentland. 2013. Predicting Personality Using Novel Mobile Phone-Based Metrics. In *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer Berlin Heidelberg, 48–55. https://doi.org/10.1007/978-3-642-37210-0_6

[13] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise.. In *Kdd*, Vol. 96. 226–231.

[14] Catherine K. Ettman, Salma M. Abdalla, Gregory H. Cohen, Laura Sampson, Patrick M. Vivier, and Sandro Galea. 2020. Prevalence of Depression Symptoms in US Adults Before and During the COVID-19 Pandemic. *JAMA Network Open* 3, 9 (Sept. 2020), e2019686. https://doi.org/10.1001/jamanetworkopen.2020.19686

[15] Aaron Fisher, Cynthia Rudin, and Francesca Dominici. 2019. All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. arXiv:1801.01489 [stat.ME]

[16] Erika Friedmann, Sue A Thomas, Fang Liu, Patricia G Morton, Deborah Chapa, Stephen S Gottlieb, Sudden Cardiac Death in Heart Failure Trial (SCD-HeFT) Investigators, et al. 2006. Relationship of depression, anxiety, and social isolation to chronic heart failure outpatient mortality. *American heart journal* 152, 5 (2006), 940–e1.

[17] Aaron S. Heller, Tracey C. Shi, C. E. Chiemeka Ezie, Travis R. Reneau, Lara M. Baez, Conor J. Gibbons, and Catherine A. Hartley. 2020. Association between real-world experiential diversity and positive affect relates to hippocampal–striatal functional connectivity. *Nature Neuroscience* 23, 7 (May 2020), 800–804. https://doi.org/10.1038/s41593-020-0636-4

[18] Enrique Hernandez-Orallo, Pietro Manzoni, Carlos Tavares Calafate, and Juan-Carlos Cano. 2020. Evaluating How Smartphone Contact Tracing Technology Can Reduce the Spread of Infectious Diseases: The Case of COVID-19. *IEEE Access* 8 (2020), 99083–99097. https://doi.org/10.1109/access.2020.2998042

[19] Jin-Hyuk Hong, Ben Margines, and Anind K. Dey. 2014. A Smartphone-Based Sensing Platform to Model Aggressive Driving Behaviors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 4047–4056. https://doi.org/10.1145/2556288.2557321

[20] Yu Huang, Haoyi Xiong, Kevin Leach, Yuyan Zhang, Philip Chow, Karl Fua, Bethany A. Teachman, and Laura E. Barnes. 2016. Assessing social anxiety using gps trajectories and point-of-interest data. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM. https://doi.org/10.1145/2971648.2971761

[21] Jeremy F Huckins, Alex W DaSilva, Weichen Wang, Elin Hedlund, Courtney Rogers, Subigya K Nepal, Jialing Wu, Mikio Obuchi, Eilis I Murphy, Meghan L Meyer, Dylan D Wagner, Paul E Holtzheimer, and Andrew T Campbell. 2020. Mental Health and Behavior During the Early Phases of the COVID-19 Pandemic: A Longitudinal Mobile Smartphone and Ecological Momentary Assessment Study in College Students. *Journal of Medical Internet Research* (May 2020). https://doi.org/10.2196/20185

[22] Hassan Ismail Fawaz, Benjamin Lucas, Germain Forestier, Charlotte Pelletier, Daniel F. Schmidt, Jonathan Weber, Geoffrey I. Webb, Lhassane Idoumghar, Pierre-Alain Muller, and François Petitjean. 2020. InceptionTime: Finding AlexNet for time series classification. *Data Mining and Knowledge Discovery* 34, 6 (Sep 2020), 1936–1962. https://doi.org/10.1007/s10618-020-00710-y

[23] Katherine Klise, Walt Beyeler, Patrick Finley, and Monear Makvandi. 2021. Analysis of mobility data to build contact networks for COVID-19. *PLOS ONE* 16, 4 (April 2021), e0249726. https://doi.org/10.1371/journal.pone.0249726

[24] K. Kroenke, R. L. Spitzer, J. B.W. Williams, and B. Lowe. 2009. An Ultra-Brief Screening Scale for Anxiety and Depression: The PHQ-4. *Psychosomatics* 50, 6 (Nov. 2009), 613–621. https://doi.org/10.1176/appi.psy.50.6.613

[25] Leodoro J. Labrague, Janet Alexis A. De los Santos, and Charlie C. Falguera. 2021. Social and emotional loneliness among college students during the COVID-19 pandemic: The predictive role of coping behaviors, social support, and personal resilience. *Perspectives in Psychiatric Care* (Jan. 2021). https://doi.org/10.1111/ppc.12721

[26] Jianbo Lai, Simeng Ma, Ying Wang, Zhongxiang Cai, Jianbo Hu, Ning Wei, Jiang Wu, Hui Du, Tingting Chen, Ruiting Li, Huawei Tan, Lijun Kang, Lihua Yao, Manli Huang, Huafen Wang, Gaohua Wang, Zhongchun Liu, and Shaohua Hu. 2020. Factors Associated With Mental Health Outcomes Among Health Care Workers Exposed to Coronavirus Disease 2019. *JAMA Network Open* 3, 3 (March 2020), e203976. https://doi.org/10.1001/jamanetworkopen.2020.3976

[27] Mark R. Leary and Roy F. Baumeister. 2000. The nature and function of self-esteem: Sociometer theory. In *Advances in Experimental Social Psychology Volume 32*. Elsevier, 1–62. https://doi.org/10.1016/s0065-2601(00)80003-9

[28] Dante L Mack, Alex W DaSilva, Courtney Rogers, Elin Hedlund, Eilis I Murphy, Vlado Vojdanovski, Jane Plomp, Weichen Wang, Subigya K Nepal, Paul E Holtzheimer, Dylan D Wagner, Nicholas C Jacobson, Meghan L Meyer, Andrew T Campbell, and Jeremy F Huckins. 2021. Mental Health and Behavior of College Students During the COVID-19 Pandemic: Longitudinal Mobile Smartphone and Ecological Momentary Assessment Study, Part II. *J Med Internet Res* 23, 6 (4 Jun 2021), e28892.

[29] Tania Martin, Georgios Karopoulos, José L. Hernández-Ramos, Georgios Kambourakis, and Igor Nai Fovino. 2020. Demystifying COVID-19 Digital Contact Tracing: A Survey on Frameworks and Mobile Apps. *Wireless Communications and Mobile Computing* 2020 (Oct. 2020), 1–29. https://doi.org/10.1155/2020/8851429

[30] Hannah McCarthy, Henry W W Potts, and Abigail Fisher. 2021. Physical Activity Behavior Before, During, and After COVID-19 Restrictions: Longitudinal Smartphone-Tracking Study of Adults in the United Kingdom. *Journal of Medical Internet Research* 23, 2 (Feb. 2021), e23701. https://doi.org/10.2196/23701

[31] Shayan Mirjafari, Hessam Bagherinezhad, Subigya Nepal, Gonzalo J. Martinez, Koustuv Saha, Mikio Obuchi, Pino G. Audia, Nitesh V. Chawla, Anind K. Dey, Aaron Striegel, and Andrew T. Campbell. 2021. Predicting Job Performance Using Mobile Sensing. *IEEE Pervasive Computing* 20, 4 (2021), 43–51. https://doi.org/10.1109/MPRV.2021.3118570

[32] Shayan Mirjafari, Kizito Masaba, Ted Grover, Weichen Wang, Pino Audia, Andrew T. Campbell, Nitesh V. Chawla, Vedant Das Swain, Munmun De Choudhury, Anind K. Dey, Sidney K. D'Mello, Ge Gao, Julie M. Gregg, Krithika Jagannath, Kaifeng Jiang, Suwen Lin, Qiang Liu, Gloria Mark, Gonzalo J. Martinez, Stephen M. Mattingly, Edward Moskal, Raghu Mulukutla, Subigya Nepal, Kari Nies, Manikanta D. Reddy, Pablo Robles-Granda, Koustuv Saha, Anusha Sirigiri, and Aaron Striegel. 2019. Differentiating Higher and Lower Job Performers in the Workplace Using Mobile Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 2, Article 37 (June 2019), 24 pages. https://doi.org/10.1145/3328908

[33] Sandrine R. Müller, Heinrich Peters, Sandra C. Matz, Weichen Wang, and Gabriella M. Harari. 2020. Investigating the Relationships between Mobility Behaviours and Indicators of Subjective Well–Being Using Smartphone–Based Experience Sampling and GPS Tracking. *European Journal of Personality* 34, 5 (Sept. 2020), 714–732. https://doi.org/10.1002/per.2262

[34] Lorene M. Nelson, Julia F. Simard, Abiodun Oluyomi, Vanessa Nava, Lisa G. Rosas, Melissa Bondy, and Eleni Linos. 2020. US Public Concerns About the COVID-19 Pandemic From Results of a Survey Given via Social Media. *JAMA Internal Medicine* 180, 7 (July 2020), 1020. https://doi.org/10.1001/jamainternmed.2020.1369

[35] Subigya Nepal, Gonzalo J. Martinez, Shayan Mirjafari, Stephen Mattingly, Vedant Das Swain, Aaron Striegel, Pino G. Audia, and Andrew T. Campbell. 2021. Assessing the Impact of Commuting on Workplace Performance Using Mobile Sensing. *IEEE Pervasive Computing* 20, 4 (2021), 52–60. https://doi.org/10.1109/MPRV.2021.3112399

[36] Subigya Nepal, Shayan Mirjafari, Gonzalo J Martinez, Pino Audia, Aaron Striegel, and Andrew T Campbell. 2020. Detecting Job Promotion in Information Workers Using Mobile Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–28.

[37] Mikio Obuchi, Jeremy F Huckins, Weichen Wang, Alex daSilva, Courtney Rogers, Eilis Murphy, Elin Hedlund, Paul Holtzheimer, Shayan Mirjafari, and Andrew Campbell. 2020. Predicting Brain Functional Connectivity Using Mobile Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–22.

[38] Jean Louis Pépin, Rosa Maria Bruno, Rui-Yi Yang, Vincent Vercamer, Paul Jouhaud, Pierre Escourrou, and Pierre Boutouyrie. 2020. Wearable Activity Trackers for Monitoring Adherence to Home Confinement During the COVID-19 Pandemic Worldwide: Data Aggregation and Analysis. *Journal of Medical Internet Research* 22, 6 (June 2020), e19787. https://doi.org/10.2196/19787

[39] S. M. Pincus. 1991. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences* 88, 6 (March 1991), 2297–2301. https://doi.org/10.1073/pnas.88.6.2297

[40] Giorgio Quer, Jennifer M. Radin, Matteo Gadaleta, Katie Baca-Motes, Lauren Ariniello, Edward Ramos, Vik Kheterpal, Eric J. Topol, and Steven R. Steinhubl. 2020. Wearable sensor data and self-reported symptoms for COVID-19 detection. *Nature Medicine* 27, 1 (Oct. 2020), 73–77. https://doi.org/10.1038/s41591-020-1123-x

[41] Mark A. Reger, Ian H. Stanley, and Thomas E. Joiner. 2020. Suicide Mortality and Coronavirus Disease 2019—A Perfect Storm? *JAMA Psychiatry* 77, 11 (Nov. 2020), 1093. https://doi.org/10.1001/jamapsychiatry.2020.1060

[42] Peter J. Rousseeuw. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20 (Nov. 1987), 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

[43] Sohrab Saeb, Emily G Lattie, Konrad P Kording, and David C Mohr. 2017. Mobile phone detection of semantic location and its relationship to depression and anxiety. *JMIR mHealth and uHealth* 5, 8 (2017), e112.

[44] Sohrab Saeb, Mi Zhang, Christopher J Karr, Stephen M Schueller, Marya E Corden, Konrad P Kording, and David C Mohr. 2015. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *Journal of Medical Internet Research* 17, 7 (July 2015), e175. https://doi.org/10.2196/jmir.4273

[45] Borja Sañudo, Curtis Fennell, and Antonio J. Sánchez-Oliver. 2020. Objectively-Assessed Physical Activity, Sedentary Behavior, Smartphone Use, and Sleep Patterns Pre- and during-COVID-19 Quarantine in Young Adults from Spain. *Sustainability* 12, 15 (July 2020), 5890. https://doi.org/10.3390/su12155890

[46] Thomas Schäfer, Peter Sedlmeier, Christine Städtler, and David Huron. 2013. The psychological functions of music listening. *Frontiers in Psychology* 4 (2013). https://doi.org/10.3389/fpsyg.2013.00511

[47] Jussi Seppälä, Ilaria De Vita, Timo Jämsä, Jouko Miettunen, Matti Isohanni, Katya Rubinstein, Yoram Feldman, Eva Grasa, Iluminada Corripio, Jesus Berdun, Enrico D'Amico, and Maria Bulgheroni and. 2019. Mobile Phone and Wearable Sensor-Based mHealth Approaches for Psychiatric Disorders and Symptoms: Systematic Review. *JMIR Mental Health* 6, 2 (Feb. 2019), e9819. https://doi.org/10.2196/mental.9819

[48] Changwon Son, Sudeep Hegde, Alec Smith, Xiaomei Wang, and Farzan Sasangohar. 2020. Effects of COVID-19 on College Students' Mental Health in the United States: Interview Survey Study. *Journal of Medical Internet Research* 22, 9 (Sept. 2020), e21279. https://doi.org/10.2196/21279

[49] Ralf C. Staudemeyer and Eric Rothstein Morris. 2019. Understanding LSTM - a tutorial into Long Short-Term Memory Recurrent Neural Networks. *CoRR* abs/1909.09586 (2019). arXiv:1909.09586 http://arxiv.org/abs/1909.09586

[50] Shaoxiong Sun, Amos A Folarin, Yatharth Ranjan, Zulqarnain Rashid, Pauline Conde, Callum Stewart, Nicholas Cummins, Faith Matcham, Gloria Dalla Costa, Sara Simblett, Letizia Leocani, Femke Lamers, Per Soelberg Sørensen, Mathias Buron, Ana Zabalza, Ana Isabel Guerrero Pérez, Brenda WJH Penninx, Sara Siddi, Josep Maria Haro, Inez Myin-Germeys, Aki Rintala, Til Wykes, Vaibhav A Narayan, Giancarlo Comi, Matthew Hotopf, and Richard JB Dobson and. 2020. Using Smartphones and Wearable Devices to Monitor Behavioral Changes During COVID-19. *Journal of Medical Internet Research* 22, 9 (Sept. 2020), e19992. https://doi.org/10.2196/19992

[51] Miklos Szocska, Peter Pollner, Istvan Schiszler, Tamas Joo, Tamas Palicz, Martin McKee, Aron Asztalos, Laszlo Bencze, Mor Kapronczay, Peter Petrecz, Benedek Toth, Adam Szabo, Attila Weninger, Krisztian Ader, Peter Bacskai, Peter Karaszi, Gyozo Terplan, Gabor Tuboly, Adam Sohonyai, Jozsef Szoke, Adam Toth, and Peter Gaal. 2021. Countrywide population movement monitoring using mobile devices generated (big) data during the COVID-19 crisis. *Scientific Reports* 11, 1 (March 2021). https://doi.org/10.1038/s41598-021-81873-6

[52] Pattreeya Tanisaro and Gunther Heidemann. 2016. Time Series Classification Using Time Warping Invariant Echo State Networks. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. 831–836. https://doi.org/10.1109/ICMLA.2016.0149

[53] Here Technologies. [n.d.]. HERE. https://here.com/. Accessed: 2019-07-27.

[54] Geoffrey H. Tison, Robert Avram, Peter Kuhar, Sean Abreau, Greg M. Marcus, Mark J. Pletcher, and Jeffrey E. Olgin. 2020. Worldwide Effect of COVID-19 on Physical Activity: A Descriptive Study. *Annals of Internal Medicine* 173, 9 (Nov. 2020), 767–770. https://doi.org/10.7326/m20-2665

[55] Scott Toney, Jenn Light, and Andrew Urbaczewski. 2021. Fighting Zoom Fatigue: Keeping the Zoombies at Bay. *Communications of the Association for Information Systems* 48, 1 (2021), 10.

[56] Rafael Wampfler, Severin Klingler, Barbara Solenthaler, Victor R. Schinazi, and Markus Gross. 2020. *Affective State Prediction Based on Semi-Supervised Learning from Smartphone Touch Data*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376504

[57] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. 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*. 3–14.

[58] Rui Wang, Weichen Wang, Min SH Aung, Dror Ben-Zeev, Rachel Brian, Andrew T Campbell, Tanzeem Choudhury, Marta Hauser, John Kane, Emily A Scherer, et al. 2017. Predicting symptom trajectories of schizophrenia using mobile sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–24.

[59] Rui Wang, Weichen Wang, Alex daSilva, Jeremy F Huckins, William M Kelley, Todd F Heatherton, and Andrew T Campbell. 2018. Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–26.

[60] Weichen Wang, Gabriella M. Harari, Rui Wang, Sandrine R. Müller, Shayan Mirjafari, Kizito Masaba, and Andrew T. Campbell. 2018. Sensing Behavioral Change over Time: Using Within-Person Variability Features from Mobile Sensing to Predict Personality Traits. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 141 (sep 2018), 21 pages. https://doi.org/10.1145/3264951

[61] Weichen Wang, Shayan Mirjafari, Gabriella Harari, Dror Ben-Zeev, Rachel Brian, Tanzeem Choudhury, Marta Hauser, John Kane, Kizito Masaba, Subigya Nepal, Akane Sano, Emily Scherer, Vincent Tseng, Rui Wang, Hongyi Wen, Jialing Wu, and Andrew Campbell. 2020. Social Sensing: Assessing Social Functioning of Patients Living with Schizophrenia Using Mobile Phone Sensing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376855

[62] Xiaomei Wang, Sudeep Hegde, Changwon Son, Bruce Keller, Alec Smith, and Farzan Sasangohar. 2020. Investigating Mental Health of US College Students During the COVID-19 Pandemic: Cross-Sectional Survey Study. *J Med Internet Res* 22, 9 (17 Sep 2020), e22817. https://doi.org/10.2196/22817

[63] Zhiguang Wang, Weizhong Yan, and Tim Oates. 2016. Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline. arXiv:1611.06455 [cs.LG]

[64] Bendong Zhao, Huanzhang Lu, Shangfeng Chen, Junliang Liu, and Dongya Wu. 2017. Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics* 28, 1 (Feb. 2017), 162–169. https://doi.org/10.21629/jsee.2017.01.18

[65] Yi Zheng, Qi Liu, Enhong Chen, Yong Ge, and J. Leon Zhao. 2014. Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks. In *Web-Age Information Management*. Springer International Publishing, 298–310. https://doi.org/10.1007/978-3-319-08010-9_33