Investigating Emotions, Social Roles, Well-being and Mobile Usage Behaviors During COVID-19 Home Confinement

NAN GAO*, Tsinghua University, China and University of New South Wales (UNSW), Australia
SAM NOLAN*, RMIT University, Australia
KAIXIN JI*, RMIT University, Australia
SHAKILA KHAN RUMI, RMIT University, Australia
JUDITH SIMONE HEINISCH, University of Kassel, Germany
CHRISTOPH ANDERSON, University of Kassel, Germany
KLAUS DAVID, University of Kassel, Germany
FLORA D. SALIM, University of New South Wales (UNSW), Australia

Pandemics and epidemics have been an essential part of human history. It has powerfully shaped human civilization, affecting politics, economics, medicine, and everyday human lives. Home confinement, one of the social isolation restrictions, has proven to be an effective way to prevent a pandemic and significantly reduce the spread of the virus. However, these measures have also led to negative psychological effects, which can be more severe than the physical discomfort caused by the pandemic itself. To address this issue, it is crucial to understand how being confined at home affects people's psychological states and their use of digital devices. To this end, we conducted an in-situ study with 32 participants living in states affected by COVID-19 lockdowns for three weeks. We examined human emotions, social roles, well-being, and mobile usage behaviors and explored the predictability of affective states using digital traces. Extensive experimental results show that COVID-19 home confinement is associated with decreased valence and increased work time, and that human well-being could be accurately inferred from mobile usage behaviors. We also present a set of interesting findings on how different factors of mobile usage affect human emotions. Overall, the findings of this research have great potential to inform the development of interventions and remote programs to support individuals in similar situations, such as those who are ill or undergoing post-operative home rehabilitation and home work.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing**; • **Applied computing**;

Additional Key Words and Phrases: Emotion, COVID-19 Lockdown, Home Confinement, Mobile Sensing, Social Roles, Well-being

*Equally contributing first authors

Authors' addresses: Nan Gao, nannn.gao@gmail.com, Tsinghua University, Beijing, China, 100084, University of New South Wales (UNSW), Sydney, Australia, 1466; Sam Nolan, samnolan555@gmail.com, RMIT University, Melbourne, Australia, 3000; Kaixin Ji, kaixin.ji@student.rmit.edu.au, RMIT University, Melbourne, Australia, 3000; Shakila Khan Rumi, shakilakhanrumi@gmail.com, RMIT University, Melbourne, Australia, 3000; Judith Simone Heinisch, judith.heinisch@comtec.eecs.uni-kassel.de, University of Kassel, Kassel, Germany, 34131; Christoph Anderson, c_anderson@posteo.de, University of Kassel, Kassel, Germany, 34131; Klaus David, david@uni-kassel.de, University of Kassel, Kassel, Germany, 34131; Flora D. Salim, flora.salim@unsw.edu.au, University of New South Wales (UNSW), Sydney, Australia, 1466.

2474-9567/2023/8-ART52 $15.00
https://doi.org/10.1145/1122445.1122456

ACM Reference Format:
Nan Gao, Sam Nolan, Kaixin Ji, Shakila Khan Rumi, Judith Simone Heinisch, Christoph Anderson, Klaus David, and Flora D. Salim. 2023. Investigating Emotions, Social Roles, Well-being and Mobile Usage Behaviors During COVID-19 Home Confinement. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 37, 4, Article 52 (August 2023), 28 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

The coronavirus disease 2019 (COVID-19) pandemic has resulted in a global lockdown, with many people experiencing home confinement for extended periods. Home confinement, defined as the restriction of individuals' movements and activities within their homes [58], is a common phenomenon that everyone may experience during their lives, whether due to illness or post-recovery [15, 42]. Previous research has shown that home confinement may have negative psychological effects, including stress, anxiety, and depression. [5, 72]. For instance, Bartoszek et al. [5] found that individuals who were quarantined were associated with high levels of depression, insomnia, loneliness, and everyday fatigue. Understanding the impact of home confinement on mental well-being is essential for developing interventions to support individuals during prolonged periods of confinement.

Existing approaches to measuring human mental well-being during home confinement typically rely on survey-based instruments [27, 72]. However, these approaches can be time-consuming and labor-intensive, and may also bring additional burdens to participants, leading to response bias (e.g., reflection bias and social desirability bias [28, 65]). Fortunately, recent advancements in mobile sensing technology offer a promising solution for predicting human mental well-being [19, 67]. For instance, Rashid et al. [67] modeled social anxiety by utilizing sparsely collected mobile sensor data from socially anxious participants, while Chowed et al. [19] measured depression, social anxiety, affect, and social isolation of undergraduate students using GPS signals collected through mobile devices. However, most sensing technologies are designed to monitor daily lives in general, rather than specifically for home confinement situations. As such, the patterns and daily behaviors captured by these technologies may differ significantly from those experienced during home confinement.

Studying the in-situ mental well-being and mobile usage behaviors of individuals confined to their homes can be challenging due to the variability of their degree of confinement, which may affect the accuracy and validity of collected data. Nevertheless, the COVID-19 lockdown presents a unique opportunity to study human well-being during home confinement. On one hand, the strict measures put in place have created similar confinement conditions for all participants due to policy mandates. On the other hand, the COVID-19 home confinement has led to an increasing reliance of digital technologies[63], with srategies such as stay-at-home and work-from-home being recommended worldwide [68]. As a result, people are increasingly turning to smartphones for social networks, online meetings, Internet surfing, or video games [68]. Recent research shows that COVID-19 lockdowns have caused around 70% of Internet users to increase their smartphone usage [76]. The reliance on technology during the COVID-19 lockdown has provided an excellent way to understand people's mental well-being during home confinement situations.

In this research, we aim to explore the relationship between people's mental well-being and their interactions with mobile devices during COVID-19 home confinement. Specifically, we are focusing on emotions and social roles because these aspects have a significant impact on overall well-being [21, 60]. To assess emotional states, we utilize the widely accepted *valence* and *arousal* dimensions within the *Circumplex Model of Affect* [73] to assess emotional states. Social roles, as defined by *Role Theory* [7], are associated with norms, expectations, rules and behaviors that provide valuable information about the user's current context, enabling us to understand their status, task boundaries and work-life balance. This, in turn, can have a significant impact on an individual's overall well-being and mental health. Furthermore, analyzing app usage records and extracting user activities from them allows for a more granular understanding of user behavior. As a result, our research questions are:

- **RQ1.** *What are people's emotions and social roles like during COVID-19 home confinement? To what extent does the COVID-19 lockdown influence people's emotions and work-life balance in their daily lives?*
- **RQ2.** *Is it possible to unsupervisedly extract user activity from app usage records? If so, how are these activities related to emotions, productivity and interruptions?*
- **RQ3.** *Can we model people's emotion from mobile usage during home confinement? If so, what are the most significant types of features for prediction?*

To answer the research questions, we conducted a three-week data collection during the COVID-19 lockdown in Melbourne, Australia. To answer RQ1, we analyzed patterns of perceived valence, arousal and social roles during COVID-19 home confinement using *Experience Sampling Method* (ESM) [6]. We also investigated the influence of the COVID-19 lockdown on emotions, social roles, and work-life balance by examining the end-of-day survey data. To address RQ2, we proposed the *App-based Information Gain Temporal Segmentation* (AIGTS) to extract user activities (e.g., video-watching, social, deep work) from app usage records by identifying segments and transition times between apps during the day. Finally, we demonstrated the predictability of perceived well-being from smartphone usage behaviors for RQ3. All scripts for this study are open source. [1] In summary, our contributions are as below:

- We collected a comprehensive dataset from 32 participants over a period of 3 weeks during COVID-19 home confinement in Melbourne, Australia. The participants completed ESM questionnaires and end-of-day surveys to report their emotions, social roles, well-being, etc. Their smartphone and desktop usage were recorded through an app running in the background. The resulting dataset includes 502,485 records of users' context and phone usage, along with 1,749 ESM responses and 265 end-of-day survey responses.
- We conducted an in-depth analysis of the participants' perceived valence, arousal, and social roles during the COVID-19 home confinement, and explore how lockdown affects people's well-being and work-life balance. We discovered that the lockdown tended to reduce valence and increase work hours due to interruptions or blurring of personal and work boundaries.
- We proposed an unsupervised activity recognition method called App-based Information Gain Temporal Segmentation (AIGTS). This method enables the extraction of user activities from app usage records. By identifying different segments of app sequences and applying clustering techniques, we were able to discover 16 distinct user activities in an unsupervised manner. We also discussed the correlations between these user activities and human well-being.
- We conducted extensive experiments to predict human well-being based on smartphone usage behaviors. Experimental results demonstrated the superior prediction performance of the proposed model on mental well-being. Additionally, we derived different factors that contribute to understanding well-being and discussed the top predictors of human emotion during home confinement.

The remainder of the paper is as below. Section 2 presents related work on the impact of COVID-19 home confinement on human well-being, as well as recent advances in mobile sensing technologies for predicting well-being. Section 3 describes the data collection procedures. In Section 4, we explore patterns of human well-being during COVID-19 home confinement and investigate how the lockdown affects emotion and work-life balance. Next, we extract user activities through app sequences and explore correlations between user activities and human well-being in Section 5. Then, in Section 6, we present experimental results for predicting mental well-being during home confinement and provide a detailed discussion of the findings. Section 7 discusses the implications and Section 8 lists the limitations of our work. Finally, we summarize this research in Section 9 and indicate the potential direction of future work.

[1] link to be shared after the double-blind review.

2 RELATED WORK

2.1 Impact of the Pandemic Restrictions on Human Well-being

Pandemics and epidemics have played an essential role throughout human history [81]. In the last century alone, there were Spanish flu (1918-1920), Asian flu (1957-1958), AIDS (1981-present day), H1N1 Swine flu (2009-2010), Ebola (2014-2016), COVID-19 (2019-present day), affecting people all over the world. Researchers explored the psychological symptoms arising from pandemics and epidemics, and found that restrictive measures (e.g., social distancing, quarantine and isolation) were associated with negative psychological effects [10, 18, 81], including post-traumatic stress symptoms [70], depression [31], stress [24], emotional disturbance [91], irritability[43], low mood [43], confusion [14], anger [51] and anxiety-induced insomnia [24]. Even after the quarantine ends, the psychological effects of quarantine can still be observed months or years later [37].

According to the World Health Organization [88], the COVID-19 pandemic has severally impacted human well-being. The potential for physically sick or discomfort from COVID-19 is considered less of concern than the psychological and social impact [34]. A growing body of literature has explored the impact of the COVID-19 on emotional and mental health [20, 41, 44, 89]. Daly et al. [20] explored the effects of COVID-19 on mental health in the UK. They found a significant increase in observed mental health problems, particularly for those aged 18-34 and women. Fountoulakis et al. [25] found that, during the COVID-19 lockdown in Greece, anxiety increased in more than 45% and suicidal thoughts increased in 10.40%. Khubchandani et al. [40] found the prevalence of depression (39%), anxiety (42%) and psychological distress (39%) after initial COVID-19 lockdown in the USA. The rate of serious mental health issues in the USA has more than doubled compared to pre-COVID-19 rates in the year 2019. Killgore et al. [41] found that loneliness has increased during the COVID-19 pandemic, especially for those under lockdown restrictions.

Despite the existing research on the impact of pandemic restrictions on human well-being, much of it is limited to specific regions and conducted during different phases of varying degrees of restrictive measures. This makes it challenging to fully comprehend the effects of strict home confinement on individuals' well-being. In light of this, the first research question (RQ1) focuses on studying the emotions, social roles and work-life balance change during COVID-19 home confinement, specifically in metropolitan Melbourne where the strictest restrictive measures (Stage 4) were implemented.

2.2 Impact of the Pandemic Restrictions on Smartphone Usage

The outbreak of COVID-19 has interrupted normal activities and caused many organisations to close, forcing people to work from home employing technology [83]. Twitter even made an announcement to make working remotely a permanent option [11]. As people perform their duties by staying at home or working from home, they increasingly turned to digital devices to communicate, interact and conduct job responsibilities [68]. Compared to pre-lockdown, Internet service usage has increased from 40% to 100%, and usage of video conferencing apps such as Zoom increased tenfold [63, 76]. A recent global study [76] reported that around 70% of internet users used their smartphones or mobile phones more often as a direct result of the COVID-19 lockdown, suggesting that the COVID-19 lockdown policies may lead to the overuse of smartphones [68].

Smartphone overuse may affect the physical and psychological health of users, such as neck/lower back/shoulder pain, anxiety and depression [3, 12, 68]. Ratan et al. [68] pointed out that problematic smartphone usage affects the 'reward system' of the brain and subsequent harms will persist in the form of various emotional disorders even after the lockdown effects are lifted. Arora et al. [3] explored the problematic phone use and its impact on mental health during the COVID-19 lockdown, and found that 'Sadness' and 'anger' were the dominant emotions in addictive Twitter users, while 'valence' was the dominating emotion among non-addictive users. Busch et al. [12] revealed that excessive mobile phone use may severely impact people's mental health and human well-being.

Especially during the COVID-19 lockdown, people's mobile phone usage differs than normal, which increases the risk of overusing smartphones and affects their mental health [4].

2.3 Investigating User Activities from App Sequences

By analysing app sequences, researchers have been able to gain insights into various aspects of individuals' behaviours in daily lives, including social roles [1] and interruptiblity [61, 62]. Deploying the approach *'Application as a Sensor'*, Anderson et al. [1] found that app sequences on smartphones are correlated with individuals' private and work social roles. Similarly, Okoshi et al. [61, 62] explored app sequences data and applied machine learning techniques to predict when a break-point in activity occurs for intelligent notification delivery. These approaches focus on identifying breakpoints between apps as a means to understand interruptibility. However, none of the previous research studies have utilised breakpoints to identify activity segments and predict contextual variables (i.e., user activity) from these segments.

Moreover, previous studies often relied on simplistic methods of aggregating app usage, such as calculating the total amount of time an app is used over a specific period. However, those approaches fail to capture the nuances and transitions between different app activities, limiting understanding of user behaviours. Moreover, traditional self-report surveys may impose a burden on users, potentially resulting in biased or incomplete data. Conversely, analyzing app sequences in an unsupervised manner can yield a wealth of information pertaining to user behaviors and mental well-being. With this in mind, our research question RQ2 aims to explore the unsupervised extraction of user activity from app usage records.

2.4 Inferring Human Well-being Using Mobile Sensing Technologies

The advent of sensing technology has opened up new avenues for understanding and predicting human behaviours [50] and psychological stages, e.g., valence [36], mood [54, 55, 92], engagement [29, 35], social anxiety [19, 67], depression [82, 85], etc. Most studies integrate data from various sources for building comprehensive prediction models. For instance, Gao et al. [29] predicted emotional, cognitive and behavioral student learning engagement in an Australian high school using wearable sensors and environmental sensors. Wang et al. [85] measured depression of undergraduate students at Dartmouth College using passive mobile sensing data and wearable data. Jaques et al. [36] modeled the valence of MIT undergraduate students using physiological signals collected from wristbands and behavior data collected from mobile devices. Morshed et al. [55] presented to predict mood instabilities using passively sensed smartphone and wearable signals.

While utilizing multiple sensing sources can improve the accuracy of psychological state prediction, there is a growing interest in using mobile sensing data alone due to its convenience and ability to provide rich information about an individual's behavior, e.g., physical activity, communication patterns, and location. For example, Hintze et al. [33] analyzed mobile device usage characteristics such as session length, interaction frequency and daily usage in locked/unlocked state with respect to location context and diurnal pattern, to gain insights into how people use their smartphones and tablets. Madan [50] et al. designed a mobile sensing platform and found that location and communication logs could be used to model individual symptoms, long-term health outcomes, and diffusion of opinions in society.

Some studies demonstrated the potential of mobile sensing data alone for understanding psychological states, e.g., anxiety [19, 67, 82], mood [54, 92], emotion [80], social roles [1] and interruptibility [61, 62]. Zhang et al. [92] proposed an Android app to detect compound emotion utilizing smartphone usage data. Meegahapola et al. [54] explored the effect of geographical diversity on mood inference. They examined eight countries and compared the continent-specific, country-agnostic, and multi-country approaches trained for two mood inference tasks. Tlachac et al. [82] collected smartphone call logs and text logs during the COVID-19 pandemic for developing machine learning models to predict anxiety and depression. Tag et al. [80] developed an Andriod app to measure

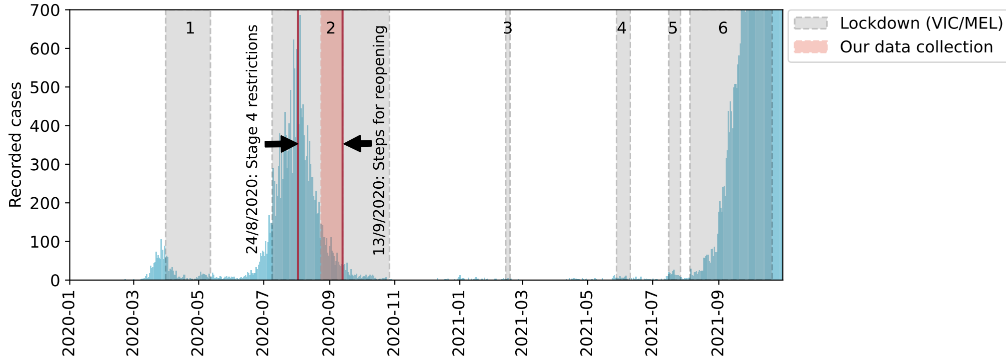


Fig. 1. The periods of six lockdowns and our data collection

emotion trajectories and examine how digital emotion regulation unfolds in naturalistic settings. Rashid et al. [67] utilized sparsely collected mobile sensor data to predict perceived social anxiety. Chowed et al. [19] explored the fine-grained GPS data collected through mobile devices to model depression, social anxiety, affect, and social isolation.

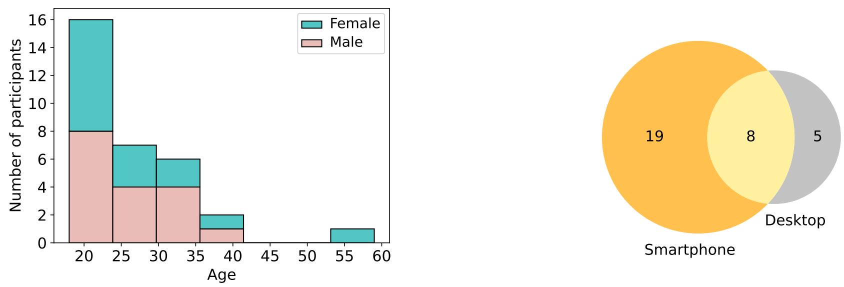
While numerous studies have utilized mobile sensing to predict human well-being, there is a significant research gap in the specific context of home confinement. Unlike daily life situations, individuals' behaviors may differ considerably during home confinement, such as increased reliance on digital devices, altered perception of time, and remote working. Although a few studies have examined psychological states using mobile sensing during the COVID-19 pandemic [59, 82], these works largely focus on broader environments rather than strictly home confinement. Investigating the unique characteristics of home confinement could provide valuable insights for understanding individuals in similar situations such as post-operative recovery, parental leave, aging or disability-related limited mobility. Therefore, we aim to address RQ3 by exploring the possibility of modeling emotions during home confinement and identifying significant features for prediction.

3 DATA COLLECTION

We conducted a field study in Melbourne, Australia, for three weeks, from 24 August 2020 to 13 September 2020. The study has been approved by the Human Research Ethics Committee at the authors' university. Prior to the data collection, we obtained informed consent from all participants, ensuring that they were aware of the study's purpose and that the dataset would be publicly released. The details about the background, participants, and collected data will be described in this section.

3.1 Background

To prevent the community spread of the COVID-19, Melbourne has been under six lockdowns since March 2020, with a total of 263 days of lockdowns, making it the city with the longest COVID-19 lockdown in the world [53, 79]. Our data collection was conducted during the second lockdown period with *Stage 4* restrictions (the strictest measures), which lasted from 9 July 2020 to 27 October 2020, with our study conducted from 24 August 2020 to 13 September 2020 (see Figure 1). Among the six lockdown periods, this was one of the longest and strictest lockdown periods, during which a state of disaster was declared on 2 August 2020, followed by *Stage 4* restrictions in metropolitan Melbourne [57].



(a) Age and gender

(b) Smartphone and desktop

Fig. 2. Demographic and device usage information of participants

Under the *Stage 4* restrictions in force at the time our data collection, a daily curfew was implemented from 8 PM to 5 AM, forbidding anyone from leaving their homes except for special reasons. Residents were only allowed to shop and exercise within 5 km of their homes, and all students returned to home-based learning. Other restrictions that were previously applied only to specific districts were extended to the entire Melbourne metropolitan area. On 13 September 2020, several restrictions were eased in Melbourne, including reduced curfews and loosened rules on outdoor exercise and social interactions. During home confinement due to the lockdown, people showed a strong reliance on technology, particularly for school and communication [68, 76]. This presents a unique opportunity for researchers to gain insights into the psychological states and smartphone usage behaviors of individuals during home confinement. The knowledge gained could prove invaluable in understanding individuals in similar home confinement situations, such as those who are ill or undergoing post-recovery home rehabilitation, and home-based work.

3.2 Participants

Participants were recruited through advertising and comprised of university students, professional workers, academic staff, and health workers. In total, we recruited 32 participants (17 males and 15 females), where most participants were between 18-40 years old (mean = 25.44, median = 23.00, std = 8.95). The age distribution of participants across genders can be found in Figure 2. It is worth noting that participants were primarily recruited through advertisements on social media platforms *Facebook* and *Discord*, which may have resulted in an over-representation of users of these platforms in our sample. While this could be considered a limitation of our study, it is important to note that our research did not specifically focus on social media usage behavior. Rather, we focus on individuals' psychological states and smartphone usage behaviors during COVID-19 home confinement. Therefore, we believe that the potential oversample threat to the validity of our results is minimized. However, it is still important to note that our findings are specific to this particular population subset and may not be necessarily generalizable to other populations. Future research should consider recruiting participants from a wider range of sources to improve the generalizability of the findings.

After the data collection, each participant was compensated a 25 AUD gift card as a token of appreciation for their time. Of the 32 participants, 27 recorded smartphone usage, 13 recorded desktop usage, and 8 recorded both smartphone and desktop app usage. Due to the relatively small number of participants using desktops, we primarily focused on analyzing the smartphone usage of the 27 participants. Analysis of desktop usage behaviors will be included at the end of the discussion and explored further in future work.

Table 1. An overview of the collected data

Category	Items	Questions / Descriptions	Options
Smartphone Usage	Application usage	Foreground application name, package identifier, category	N/A
	Notifications	Notification arrival time, content (hashed), sender (hashed), relationship	N/A
	Phone's states	Screen status, battery, ringer modes, GPS location, physical activities	N/A
ESM	Valence	How did you feel in the last hour? (Unhappy vs Happy)	5-Likert scale
	Arousal	How did you feel in the last hour? (Calm vs Excited)	5-Likert scale
	Social role	In the last 15 minutes, I was engaged in:	Work/private/both
	Interruptibility	I am currently available for other matters:	Work/private/both/none
	Task	The task I managed to progress the most in the last 90 minutes:	Multiple choices
End-of-day Survey	Productivity	I feel my day was productive.	5-Likert scale
	Interruptibility	I was continually interrupted today.	5-Likert scale
	Valence	Has the lockdown influenced your happiness today?	5-Likert scale
	Arousal	Has the lockdown influenced your energy today?	5-Likert scale
	Work-life balance	Has the lockdown impacted your work-life balance today?	5-Likert scale
	Valence	How has the lockdown influenced your happiness today?	Text (optional)
	Arousal	How has the lockdown influenced your energy today?	Text (optional)
	Work-life balance	How has the lockdown influenced your work life balance today?	Text (optional)
	Difference	How different do you feel today was from yesterday?	5-Likert scale
	Difference	How was today different from yesterday?	Text (optional)

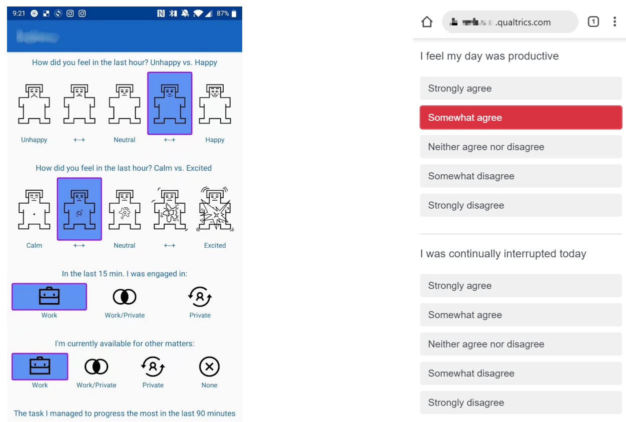
3.3 Procedures

Prior to data collection, participants were instructed to install the *Balance for Android* and *Balance for Desktop* applications [32] on their smartphones and desktops, respectively. These applications were specifically designed to record usage behaviors and enable background sensing and experience sampling, drawing inspiration from Van et al. [84]. Both applications were optimized to minimize their consumption of computational power, energy, and storage resources. The applications utilized ESM, popping up a short questionnaire to participants every 90 minutes when their devices were actively in use between 7 am and 10 pm. This helps to obtain unbiased samples of participants' activities, which served as the ground truth for our predictive models. Furthermore, an event-based ESM was triggered on the user's smartphone after a minimum interaction of 10 minutes, with subsequent event-based ESMs occurring every 30 minutes.

At the end of each day, participants received an end-of-day survey via email at 7:30 pm. This survey aimed to gather reflections on their well-being and the impact of the lockdown on their daily experiences. This provides valuable context and additional data beyond the ESM questionnaires. Overall, we collected 502,485 recordings for smartphone usage, 1,749 ESM responses, and 265 end-of-day responses.

3.3.1 Smartphone usage data. During the data collection, the background services of smartphones and desktops continuously tracked the following information: application usage (using *Android Accessibility Services* [2]), notifications (using *Android Notification Listeners* [3]), as well as the phone's states (i.e. screen/battery status, ringer modes, location updates and physical activities). Specifically, the mobile app usage behavior was recorded every time the user opens the app. For each notification, the application recorded the arrival time, hashed content (used for calculating the length of notification), the sender (private, work only, both or none), and participants' relationship to the sender (family, friend, work and none). Table 1 displays an overview of the collected data for smartphone usage, ESM and end-of-day survey. In total, we have collected 502,485 recordings for smartphone usage data from 27 participants.

[2]Accessibility Services: https://developer.android.com/reference/android/accessibilityservice/AccessibilityService
[3]Notification Listeners: https://developer.android.com/reference/android/service/notification/NotificationListenerService



(a) ESM (b) End-of-day survey

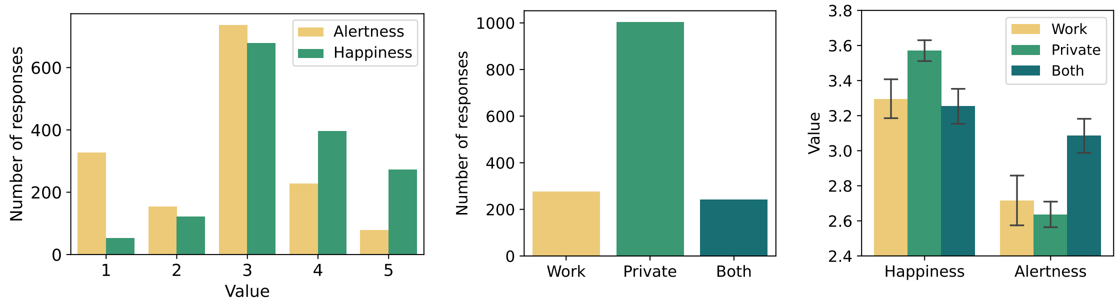
Fig. 3. The screenshots for the ESM and End-of-day survey

3.3.2 ESM responses. There are some popular questionnaires to measure well-being [8, 38], and the ESM questionnaire adapted from the *Self-Assessment Manikin* (SAM) [8] was employed in this research. In the ESM questionnaire, we collected information on valence, arousal, social role, interruptibility, and task from users. Especially, the item 'valence' was rated with a 5-point Likert scale from 'unhappy' to 'happy', and the item 'arousal' was also rated with a 5-point Likert scale from 'calm' to 'excited'. Both valence and arousal were used to represent the users' emotions based on the circumplex model [73]. Understanding the social role, interruptiblity and task help us understand how people manage their time and attention on the digital devices [2]. The detailed items of the ESM can be found in Table 1 and the screenshot of the ESM on users' smartphones was shown in Figure 3(a). Overall, we have collected 1,749 ESM responses from 29 participants. Of these, 1,522 responses were obtained from 25 participants using smartphones (2 participants did not contribute any ESM responses), while the remaining 227 responses were from 10 participants using desktops.

3.3.3 End-of-day survey. As shown in Figure 3(b), the end-of-day surveys were optional and included ten items to understand participants' well-being during the lockdown such as productivity, valence, energy, and work-life balance. In addition, we surveyed how participants feel today compared to yesterday. This information could describe the dynamics of people's experience during the COVID-19 lockdown. In the end-of-day survey, each item was rated with a 5-Likert scale or answered with free text input (see Table 1). Specifically, item 1 and item 2 were rated from 1 to 5 which indicates 'strongly agree', 'somewhat agree', 'neither agree nor disagree', 'somewhat disagree', 'strongly disagree', and items 3, 4, 5, 9 were rated from 1 to 5 which indicates 'a great deal', 'a lot', 'a moderate amount', 'a little', 'not at all'. In total, we collected 265 responses from 25 participants during the data collection period.

4 UNDERSTANDING EMOTIONS, SOCIAL ROLES, AND WORK-LIFE-BALANCE DURING COVID-19 HOME CONFINEMENT

In this section, we aim to answer RQ1: *"What are people's emotions and social roles like during COVID-19 home confinement? To what extent does the COVID-19 lockdown influence people's emotions and work-life balance in their daily lives?"*. To achieve this, we analyze 1522 ESM responses collected through smartphones from 25 participants,



(a) Valence and arousal

(b) Social roles

(c) Social roles and emotion

Fig. 4. The distribution of perceived valence, arousal, social roles and their relationship

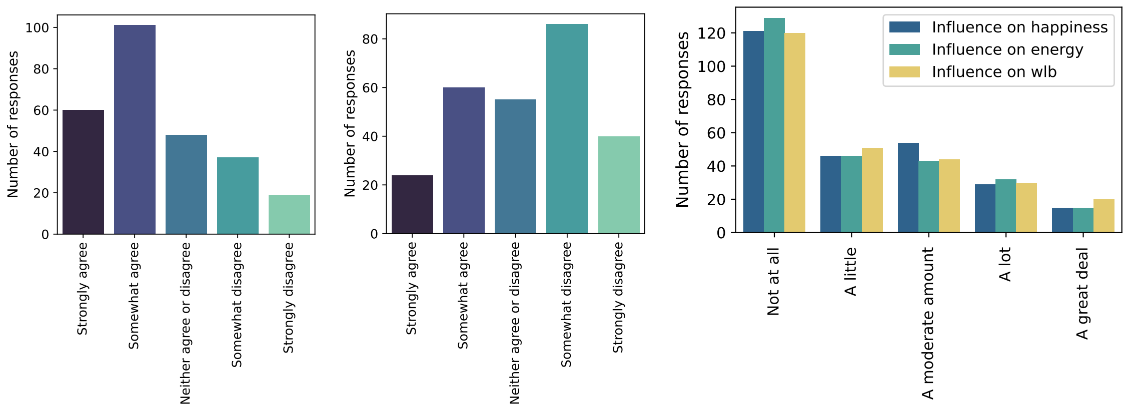
as described in Section 4.1. Furthermore, we explore the influence of COVID-19 lockdown by examining the end-of-day survey in Section 4.2.

4.1 The Perceived Valence, Arousal and Social roles

The distribution of perceived valence, arousal, and social roles from the ESM responses is depicted in Figure 4. On average, participants reported a valence level of 3.46 (SD = 1.01) and an arousal level of 2.72 (SD = 1.11). Usually, the relatively high valance and low arousal refer to a state of calmness, relaxation and peace. This could be attributed to various factors during the COVID-19 home confinement. One possible reason is that people may feel relieved to be safe at home, and cherish the opportunity to spend more quality time with their family and loved ones, away from the potential risks of infection outside. Furthermore, the absence of the usual stressors of daily life, such as commuting, work pressures, and social obligations, may be contributing to a general sense of calm and well-being. Additionally, some people may have found comfort in the sense of community and solidarity that emerged during the pandemic, as people came together to support one another through difficult times. All of the factors mentioned earlier may contribute to the high valence and low arousal state during COVID-19 home confinement. Nevertheless, it is crucial to acknowledge that our data collection was limited to a three-week period, as COVID-19 Stage 4 restrictions (strict home confinement) were lifted and steps for reopening commenced. Thereafter, in the event that strict home confinement lasts longer, it is possible that the emotional experiences of individuals may evolve as they adapt to the new normal and encounter new challenges. For instance, prolonged periods of confinement may lead to feelings of boredom, loneliness, and frustration.

Next, we investigate the distribution of social roles and their correlation with emotion during home confinement. A *social role* refers to a set of norms, expectations, rules, beliefs and behaviors that are associated with a particular social situation [69]. These roles can significantly influence users' behaviors and the type of information they prefer to receive [2]. As shown in Figure 4(b), smartphone users predominantly engaged in a *private-related role* during home confinement, which suggests that individuals tend to prioritize their personal lives and relationships while staying at home.

Since social roles play a crucial role in shaping individuals' well-being, we investigate the emotions of users in different social roles, as shown in Figure 4(c). We found that, when engaging in *private-related* roles, participants had the highest levels of valence and lowest levels of arousal compared to *work-related* roles and hybrid roles. This may be because private-related roles involve personal activities which can provide a sense of comfort and happiness, while work-related roles or hybrid roles may involve stress, deadlines, and pressure to perform. The



(a) Productivity (b) Interrupted (c) Influence on well-being

Fig. 5. The distribution of productivity, interruption, and influence during the lockdown

results of Welch's t-tests [93] further support these findings, indicating that there is a significant difference in valence levels between *private-related* roles and the other two roles (i.e., *private* vs. *both*, *private* vs. *work*), with p-value less than 0.05. The above findings provides insights into how individuals can structure their daily routines to promote positive emotions during periods of home confinement, and interventions aimed at improving well-being during home confinement may benefit from considering the types of social roles that individuals are engaging in [30].

4.2 The Influence of the COVID-19 Lockdown on Emotions, Social Roles and Work-life Balance

In total, 54.3%, 51.3% and 54.7% of end-of-day responses displayed that the lockdown has influenced their valence, energy, and work-life balance to some extent, as shown in Figure 5. For the influence of lockdown on valence, 13 participants reported that it had a negative effect, claiming they '*felt quite anxious and sad*' as '*staying at home has made me a little sad*'. They complained that they were unable to see friends (9 responses), see families (5 responses), go outside (10 responses), and feel bored (6 responses), e.g., '*I thought I might be able to see friends, but I did not*', '*I want to go outside, to the park*', '*I'm missing my family*', '*It's hard to say how it's specifically changed my happiness today because it's the build-up of it has gone on for so long. I'm bored and sick of my housemates and it feels like time has stopped* '. Nevertheless, two felt the lockdown has a positive effect, claiming that '*I am a lot happier*' and '*I was able to maximize my time to do some exercise and transition to work immediately*'.

Regarding the impact of lockdown on arousal, 11 participants claimed that the impact was negative, reporting feeling tired (11 responses), being lazy (5 responses) and losing motivation (3 responses), e.g., '*I don't feel as motivated to do productive things*', '*Emotionally tired, perhaps as a result of the lockdown*'. However, two participants claimed the lockdown gave them '*more energy*'. For the influence on work-life-balance, most participants reported that they faced more interruptions or distractions (14 responses), had difficulty separating work and life (11 responses), spent more time on work due to interruptions or blurring of personal and work boundaries (15 responses), e.g., '*I worked a bit more than I should've*', '*Everything blends together, separating work and life is really difficult*', '*Continuous interruptions, even my cat interrupts me*', '*Need to work longer hours due to interruptions*'. Overall, while a minority of responses showed that lockdown has a positive impact on their lives (e.g., more time

Table 2. App categories and top three frequently used apps

Upper Category	Category	Top App	Second App	Third App
Business	Business	Teams	LinkedIn	Zoom
Communication	Communication	Chrome	Discord	WeChat
Entertainment	Action	Among Us	Among Us	
	Casual	Scrap Collector	IdleCourier	
	Comics	TachiyomiJ2K		
	Travel & Local	Maps	PTV	
	Video Players & Editors	YouTube	Bilibili	YouTube Vanced
	Entertainment	Wow	Steam	RainbowSix
	Food & Drink	Uber Eats	Menulog	Mymacca's
	Health & Fitness	Samsung Health	Nike Training	
	Photography	Photos	Gallery	Photos
	Music & Audio	Google Play Music	Spotify	YouTube Music
	Role Playing	Fate/GO		
	Shopping	BIGAU	Gumtree	Depop
Productivity	Productivity	Excel	Outlook	PowerPoint
	Education	Canvas Student		
Reading	Books & Reference	Audible	AniDroid	MendeleyDesktop
	News & Magazines	Joey	Twitter	Sync Dev
Social	Social	Facebook	Instagram	Snapchat
Tools	Tools	Gboard	Explorer	Samsung Keyboard
Others	Auto & Vehicles	Android Auto		
	Finance	CommBank	CommSec	Up
	House & Home	Domain	Realestate	
	Lifestyle	Samsung Pay	Tinder	Hue
	Medical	MyTherapy	E4 realtime	
	Parenting	FamilyAlbum		
	Personalization	One UI Home	OnePlus Launcher	TouchWiz home

with family or gardening), most responses showed negative influences of lockdown on their valence, arousal and work-life balance.

Besides the qualitative feedback from participants, we adopted Spearman correlation [22] to study the relationship between the reported *Influence on Valence* (from end-of-day survey) and the average perceived valence for that day (from ESM responses). In this way, we can determine whether the lockdown had a positive or negative effect on valence. Likewise, we went through the same process in terms of users' arousal and social roles. For social roles, we correlate the reported *Influence on wlb* ('wlf' indicates work-life balance) to the percentage of the day spent on working (reported in 'work' or 'both' states). For example, if someone claimed that the lockdown significantly influenced their work-life balance (i.e., a score of 5, which becomes 3 after Likert transformation in the pre-processing), and they spent 50% of the day at work, then (50%, 3) would be a point in the correlation.

5 UNDERSTANDING APP USAGE BEHAVIORS DURING COVID-19 HOME CONFINEMENT

In this section, we address RQ2 *"Is it possible to unsupervisedly extract user activity from app usage records? If so, how are these activities related to emotions, productivity and interruptions?"*. First, we introduce the process of app categorization. Next, we conduct unsupervised activity recognition from app sequences and display the average activities for all clusters. In this end, we explore the correlation between user activities and perceived psychological metrics.

5.1 App Categorization

We categorized the apps using historical usage logs. For *Android* apps, we referred to the categories listed in the Google Play Store, excluding uncategorized apps. Windows apps were manually sorted into the same categories as Google Play Store apps. Less frequently used apps, with a cumulative usage time of less than one hour across all participants, were excluded. Overall, we recorded 28 app categories and consolidated them into 8 upper-level categories: *Business*, *Communication*, *Entertainment*, *Productivity*, *Reading*, *Social*, *Tools* and *Others*. A summary of the top three most frequently used apps in each category is provided in Table 2. Notably, Chrome is categorized as a *Communication* app due to the scarcity of browser apps in the Google Play Store.

Categorizing desktop apps within the same categories used for mobile apps proves challenging, as they are often used for work and software development (e.g., Pycharm, WinSCP etc). In light of this, we grouped several development and utility apps under the *Tools* Category. Overall, the top 10 *Tools* apps for both desktop and mobile include: *Gboard*, *Explorer*, *Samsung Keyboard*, *Always on Display*, *WinSCP*, *Pixel Launcher*, *Acrobat*, *pychar64*, *Google* and *Samsung keyboard*. It is worth mentioning that developer tools such as *Pycharm* and *WinSCP*, as well as the utility *Explorer*, fall under this category.

5.2 Unsupervised Activity Recognition From App Sequences

Segmentation. To study how the application usage is related to the aforementioned variables, we propose *App-based Information Gain Temporal Segmentation* (AIGTS), which focuses on differentiating activities within app usage. Figure 6 provides an illustration of how activities are extracted from app usage data. AIGTS draws inspiration from *Information Gain Temporal Segmentation* (IGTS) [74], an unsupervised method for segmenting multivariate time series data. IGTS works by identifying which splits in the multivariate time series minimize the weighted entropy across the series. The variables within the multivariate time series data corresponded to whether an app of a particular category was used.

Unlike IGTS focusing solely on the time series, AIGTS aims to identify different segments and transition times between apps throughout the day. It is capable of detecting when participants change their app usage patterns, such as transitioning from using Email and Browsers (both categorized under *Communication*) to browsing Reddit (*News and Media* category). This method allows for segmentation at any level of granularity desired, ranging from half a day to as little as half a minute for application activity. The granularity of segmentation is determined by the variable k, which represents the number of splits made in the time series. To ensure that segments are captured at a consistent level of granularity, we propose the following steps for calculating the optimal k in AIGTS:

(1) Calculate the knee point of the activity data for each day;
(2) Identify the average number of segments per minute for all days based on the knee point;
(3) Compute the median average segments per minute for all users;
(4) Multiply this median by the number of minutes per day (which varies from the first log to the last log) and choose this value as k.

These procedures guarantee that the average length of each segment will be consistent across all days.

Clustering. Next, AIGTS was conducted on all time series, resulting in a segmented list for each user. To extract features from these segments, *k-means* clustering [46] was applied. The distribution of different categories appearing in a segment was treated as an n dimensional vector, where n is the number of app categories. To identify the optimal number of segments, Silhouette analysis [47] was utilized with a maximum of 20 clusters. Especially, the label indicating which cluster a segment belongs to was used as a feature representing the activity within that segment. These segment labels were derived manually from the average distribution of apps in that cluster. In this paper, a segment that belongs to a particular cluster will be referred to as an ‘activity’.

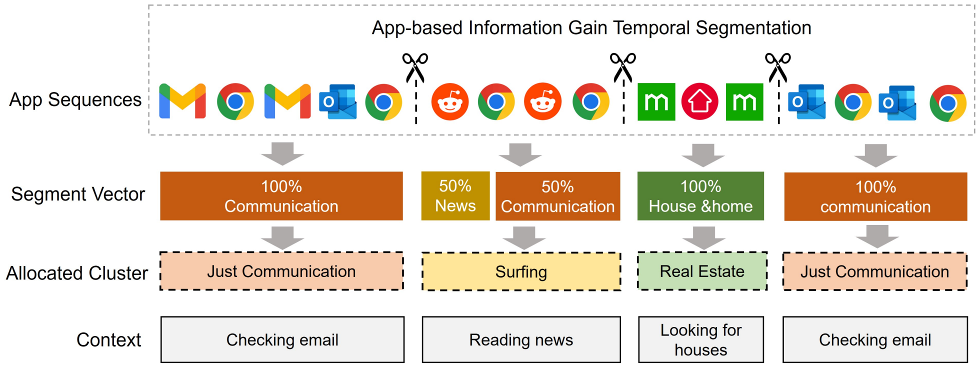


Fig. 6. An illustration of unsupervised activity recognition using app usage data

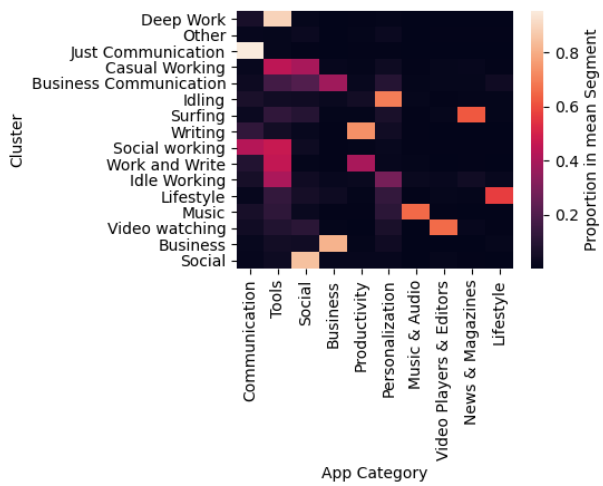


Fig. 7. Average activities for all clusters

Distribution of User Activities. Following the segmentation and clustering, we uncovered 16 clusters. The average duration of these user activities was 18 minutes. Figure 7 displays average activities for all clusters, where the average activity represents the average distribution of app categories appearing in each segment within the cluster. These clusters were manually named to describe the distribution of average activities. The x-axis displays the app categories that frequently appear in clusters, with infrequently appearing categories omitted. Notably, the cluster *Other* includes all uncategorizable segments and has roughly equal means across multiple categories. Additionally, the cluster *Idling* is used for *Personalization* apps because *Personalization* apps are typically launchers (see table 2), and indicate that the user is sitting on their device's home screen. Moreover, the most frequent user activities during the COVID-19 lockdown were *Just Communication* and *Idling*, accounting for 17% and 14.60% of activities, respectively. Conversely, the least frequent activities were *Business Communication* and *Music*, accounting for 1.22% and 1.65% of activities, respectively.

5.3 App Usage Correlation

Furthermore, we investigate the correlations between user activities with their *valence*, *arousal*, *productivity*, and *interruptions*. To understand whether any particular activity A was correlated with *valence*, we ran a Spearman correlation [22] between *valence* scores and the amount of time spent in user activity A in the previous hour. The same goes for the *arousal*, *productivity*, and *interruptions*, but the *productivity* and *interruptions* vary in duration from 1 hour to 24 hours, as these are calculated from the end-of-day survey and correspond to the value for a full day. Although the ESM responses also measure interruptibility, we used the interruptibility item from the end-of-day survey to assess the intensity of how the COVID lockdown affected well-being, as it directly captures the interruptions experienced throughout the day. The top correlation results for how different activities relate to usesr are displayed in Table 3. We also report the top predicting activities for each score and demographic group (i.e., male, female, ages 18-24, ages 25+).

As shown in Table 3, the correlations are weak with most activities, but few activities have moderate correlations. *Tools* related activities were positively correlated with overall *valence* and for both age groups. Particularly, *Idle Tools* have a relatively stronger correlation with valence in older adults, indicating an increase in the mood at work (*Idle Tools* include phone home screens and phone keyboards). *Surfing* by women and young age group are negatively correlated to *valence*, indicating reading news or surfing (e.g., reading COVID-19-related news) are negatively related to their valence. *Productivity* was negatively correlated with *Communication Tools* (e.g., keyboards and chat clients), and positively correlated with *Business Communication* (i.e., both Communication and Business apps such as Microsoft Teams). It can be seen that while both are Communication applications, whether or not they are used for work may imply productivity levels. *Interruptions* can be seen as the inverse of *productivity*, as it has positive correlation with *Communication Tools*, but the value is higher. The correlation reaches a moderate level for older adults while weakening for the young age group. This might cause by the difference in working loads between young and older age groups. *Lifestyle* negatively correlates with *interruptions*, especially for female and older adults.

6 PREDICTING USER EMOTION AND SOCIAL ROLES FROM MOBILE USAGE DATA

In this section, we address RQ3 *"Can we model people's emotion from mobile usage during home confinement? If so, what are the most significant types of features for prediction?"*. First, we introduce the features extracted from mobile usage data and the prediction pipeline. We then list and discuss the prediction results of the general models and individual models. Finally, we examine the significant features for predicting human well-being.

6.1 Feature Extraction

Table 4 provides an overview of the features utilized in our research. Similar to previous mobile sensing studies [45, 75, 90], we extract contextual information, notifications-related information, physical activity, app, and activities inferred from apps. To process the app-usage data, we divide it into segments based on the ESM recording frequency (set at 90 minutes) and stored it in temporal sequences. Subsequently, we extract the features from these 90-minute segments. In order to handle cyclical features such as days (e.g., Monday) and times (e.g., Morning), we transforme them into sine and cosine values. Additionally, it is worth noting that app-usage patterns exhibit variations between weekdays and weekends [39, 45], prompting us to conduct separate analyses for these two categories. Regarding the time period per day, we divide it into six 4-hour segments, i.e., *Early Morning*, *Morning*, *Afternoon*, *Evening*, *Night*, and *Late Night*. Certain frequency-based features, such as notification-related or physical-activity-related features, are converted using quartiles, specifically the minimum, 25th percentile, 50th percentile, and 75th percentile. For instance, the number of interacted notification is categorized into groups as '0', '1-3', '4-9', and '10+'. Furthermore, the physical activities considered in our study only include active physical activities like bicycling, walking, and running.

Table 3. Top correlations for all variables

Groups	Variable	Lowest correlation ρ	Highest correlation ρ
All	Valence	Work and Write (-0.07)	Idle Tools (0.11)
	Arousal	Business (-0.05)	Social Tools (0.15)
	Productivity	Communication Tools (-0.24)	Business Communication (0.18)
	Interruptions	Lifestyle (-0.15)	Communication Tools (0.29)
Male	Valence	Just Communication (-0.08)	Business Communication (0.10)
	Arousal	Idling (-0.12)	Just Communication (0.11)
	Productivity	Communication Tools (-0.15)	Business Communication (0.34)
	Interruptions	Video Watching (-0.23)	Communication Tools (0.18)
Female	Valence	Surfing (-0.17)	Other (0.13)
	Arousal	Work and Write (-0.17)	Social Tools (0.21)
	Productivity	Communication Tools (-0.30)	Lifestyle (0.16)
	Interruptions	Lifestyle (-0.33)	Idling (0.41)
Age 18-24	Valence	Surfing (-0.13)	Just Tools (0.09)
	Arousal	Idle Tools (-0.17)	Social Tools (0.19)
	Productivity	Communication Tools (-0.21)	Work and Write (0.20)
	Interruptions	Social Tools (-0.10)	Communication Tools (0.16)
Age ≥25	Valence	Just Communication (-0.12)	Idle Tools (0.23)
	Arousal	Work and Write (-0.10)	Idle Tools (0.15)
	Productivity	Idle Tools (-0.25)	Social (0.20)
	Interruptions	Lifestyle (-0.24)	Communication Tools (0.42)

To capture user activities related to app-usages, we employed our proposed method, AIGTS, as detailed in Section 5.2. AIGTS allows us to extract both activities and their corresponding durations. Based on the extracted information, we compute the feature 'entropy_user_activity' by determining the proportion of each activity's time portion within the entire 90-minute time interval. In addition to extracting individual features such as 'num_noti', we also generate composite features by combing two existing features. For example, the feature 'num_communication' is calculated by selecting and combing the notification sources. This composite feature counts the number of notifications originating the *Business*, *Communication*, or *Social* app categories. Notifications that do not belong to these app categories are counted as 'app_noti'.

6.2 Prediction Pipeline

while emotion prediction can be addresed as a classification problem, with emotion levels categorized into two or three classes based on predefined thresholds [64, 92]. We believe that utilizing regression is a better approach becuase valence and arousal labels are inherently ordinal data, where the relative order holds significance. In this research, we establish a regression-based pipeline to forecast the emotion scores of participants, as detailed below.

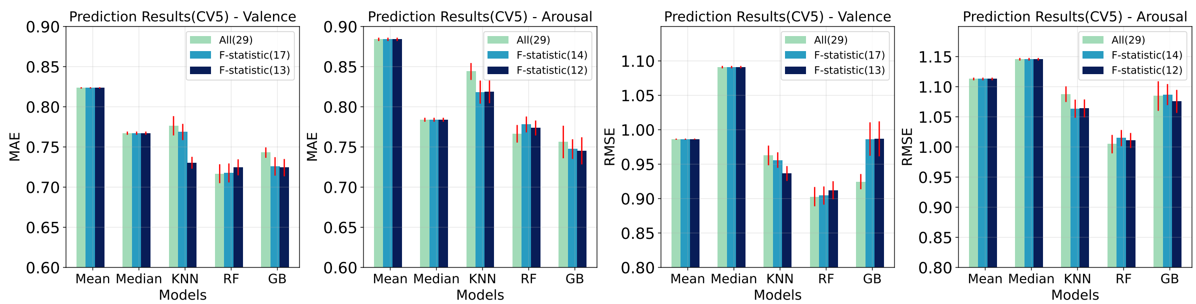
Emotion. The ESM questionnaires, completed by the participants every 90 minutes, serve as the ground truth in the regression models developed in this study. Specifically, the item 'valence' was rated on a 5-point Likert scale from 'unhappy' to 'happy', where 1 indicates 'unhappy' and 5 refers to 'happy'. Similarly, another item 'arousal' was also rated with a 5-point Likert scale from 1-5, with 1 indicates 'calm' and 5 means 'excited'. To predict the levels of valence and arousal, the app activities within 90 minutes prior to completing the ESM questionnaire were leveraged as features.

Table 4. The computed features from smartphone usage behaviors

Category	Feature	Description
Contextual Information	weekday	The day of the week (e.g., Monday, Tuesday)
	time_period	The time of the day (morning, afternoon, evening and midnight)
	sin_weekday	Sine Transformation of weekday
	cos_weekday	Cosine Transformation of weekday
	sin_time	Sine Transformation of time period
	cos_time	Cosine Transformation of time period
	is_weekend	Binary values describing whether is is weekend or not
Notification	num_noti	Number of notifications
	num_noti_interacted	Number of notifications interacted
	noti_cat	Categorical Transformation of the number of notifications
	noti_intr_cat	Categorical Transformation of the number of notifications interacted
	noti_interacted_pcg	The percentage of notification interacted to the total number of notifications
	num_communication	Number of notifications from the Communication Application
	main_contact	The main social relationship that contacted with
Physical Activity	nunique_activity	Number of unique physical activities
	main_activity	The main physical activity of the users
	num_activity	Number of physical activities
App	nunique_app	The number of unique apps
	num_app	The total number of apps
	nunique_app_cat	The number of unique app categories
	app_noti	The number of notification generated by non-communication apps
	nunique_app_noti	The number of unique apps sent notifications
	main_cat_app	The main category of apps
	second_cat_app	The second main category of apps
Activity Inferred from Apps	nunique_user_activity	The number of unique user activities
	num_user_activity	The number of user activities
	main_user_activity	The main user activity
	sec_user_activity	The second user activity
	entropy_user_activity	The entropy of different categories of user activity

Regressors. In this research, we aim to develop regression models that accurately predicts human emotions during home confinement. To accomplish this, we employ three commonly used regression models: *K-Nearest Neighbor* (KNN) [78] regressor , *Random Forest* (RF) [77] regressor, and *Gradient Boosting* (GB) [26] regressor. KNN is a widely utilized model known for its versatility across various domains. It works by finding the K nearest neighbors to a given data point and using their values to predict the target variable. RF and GB, on the other hand, are ensemble models based on decision trees. These models combine predictions from multiple trees. However, there is a distinction in their approach. GB constructs trees or aggregates results in a way thatminimizes the loss function, whereas RF constructs individual trees or aggregates results simultaneously.

Feature Selection. Various methods, such as SHAP [49], LIME [71], and the F-statistic (F) [66], can be used for feature selection. In this study, we employ the F-statistic to determine the appropriate number of features. While most of our features are numeric, the F-statistic measures the univariate linear dependence between each feature and the target variable, taking into account statistical significance. Generally, a higher F value indicates greater importance of the feature in predicting the target variable. In our experiments, we compare regression models using three feature sets: all features, features selected based on an F value greater than or equal to 5, and features selected based on an F value greater than or equal to 10.



(a) MAE for valence (b) MAE for arousal (c) RMSE for valence (d) MAE for arousal

Fig. 8. Prediction performance for valance and arousal with different feature sets, where the error bar presents a 95% confidence interval

Validation. To compensate for the limited richness of our data, we have implemented *nested cross-validation* (CV) [56] to train and evaluate our prediction model. This approach allows us to accurately estimate the unbiased generalization performance of the model. The nested CV consists of an inner loop and an outer loop. Initially, the dataset is divided into training, validation, and testing sets. This helps optimize the model's performance. In the outer loop, similar to previous human-centered studies [23, 29], we divide the data into multiple groups, each group representing one participant. To ensure unbiased evaluation, we apply *k-fold cross validation* [87] where k=5 on all groups. This means that the data from the same participants are not used in both the training and test sets simultaneously. In the inner loop, the remaining data groups are further split into three folds, with each fold serving as a validation set. We optimize the hyperparameters on the training set, evaluate them on the validation set, and select the best parameter settings based on the results from the three folds.

Baselines. Similar to previous human-centered studies [29, 86], we compare the proposed model with two baselines. The first baseline is the *Mean* baseline, which calculates the average score of user responses. The second baseline is the *Median* baseline, which calculates the median value based on the data distribution and treats it as a predicted value. To evaluate the performance of regression models, we employ two commonly used metrics: *Mean Absolute Error* (MAE) and *Root Mean Square Error* (RMSE) [17]. Generally, lower values of MAE or RMSE indicate better prediction performance.

6.3 Prediction Result

To address RQ3, we proposed several regression models in the experiment. First, a general prediction model was constructed by considering all participants in the study. Second, we built individual prediction models for each participant to account for the unique characteristics and patterns exhibited by different individuals. It is worth mentioning that only mobile phone data was utilized in this study. After excluding 19 records with invalid ESM data, a total of 1456 app-usage data records were obtained at 90-minute intervals for analysis.

6.3.1 General Model. The prediction performance for the general models is presented in Figure 8. The models used in the comparison include Mean baseline, Median baseline, KNN regressor, RF regressor, GB regressor. Among these models, the RF regressor demonstrated superior performance to both baselines in terms of MAE and RMSE for predicting valance and arousal. Interestingly, the RF regressors with different feature sets achieved comparable performance, with all feature sets yielding the best results. This can be attributed to the capability of RF to assign weights to features based on their importance, where incorporating more features leads to

Table 5. Prediction performance of the general models with different feature sets

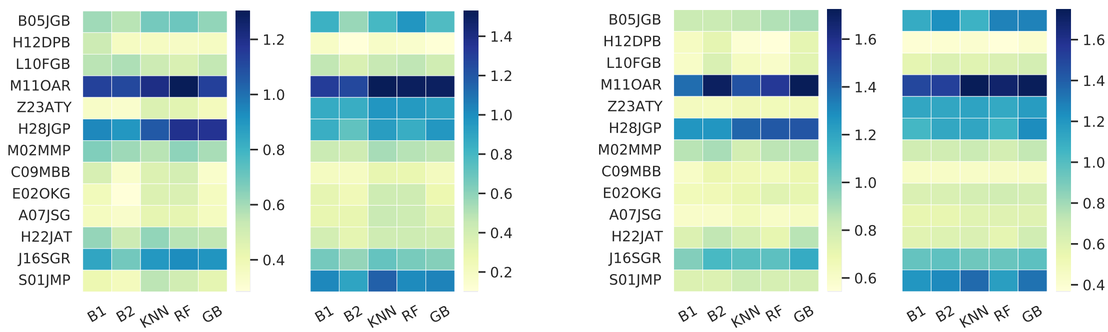
Prediction Result (RMSE/MAE)					
	Mean Baseline	**Median Baseline**	**KNN Regressor**	**RF Regressor**	**GB Regressor**
All features					
Valence (29)	0.986/0.824	1.091/0.767	0.963/0.776	**0.903**/**0.717**	0.925/0.743
Arousal (29)	1.113/0.884	1.146/0.784	1.088/0.844	**1.005**/0.766	1.084/**0.756**
Features with F-statistic >= 5					
Valence (17)	0.986/0.824	1.091/0.767	0.956/0.769	**0.904**/**0.717**	0.987/0.725
Arousal (14)	1.113/0.884	1.146/0.784	1.064/0.818	**1.015**/0.778	1.087/**0.747**
Features with F-statistic >= 10					
Valence (13)	0.986/0.824	1.091/0.767	0.936/0.730	**0.912**/0.725	0.987/**0.724**
Arousal (12)	1.113/0.884	1.146/0.784	1.064/0.819	**1.011**/0.774	1.076/**0.745**

improved prediction performance. Furthermore, it is observed that the proposed model generally exhibited better performance in predicting *valence* compared to arousal, the difference in performance may be attributed to the complexity of arousal, which is influenced by various factors such as physiological arousal and cognitive processing. These factors may not be fully captured by mobile usage behaviours alone.

Table 5 displays the prediction results of the general models for *valence* and *arousal*. Regarding the *valence* dimension, our results indicate that the RF regressor achieved the best predictive performance across three different feature sets. For the regression model using all features, the RMSE obtained by the RF algorithm was 0.903, surpassing the Median baseline by 17.23%. Furthermore, the MAE of the RF regressor using all feature sets was 0.717, outperforming the mean baseline by 12.99% and the median baseline by 6.52%. Similar results were observed for the models utilizing the top 17 features. Regarding the *arousal* dimension, our results indicate that the RF regressor utilizing the all features achieved the best performance in terms of RMSE, with a value of 1.005. This outperformed the Median baseline by 12.30%. These outcomes suggest that the selected features serve as effective predictors of *arousal* for participants during home confinement.

6.3.2 Individual Models. The individual models for both *valence* and *arousal* were trained using all 29 features. However, due to some participants having insufficient data recorded, the models were only trained on those with more than 50 data points. This resulted in a total of 13 participants and 1220 data points being used. The prediction results for these models, with the ID of participants indicated on the y-axis, can be seen in Figure 9. Overall, we can observe variations in the prediction performance of *valence* and *arousal* across different participants. For instance, participant 'M11OAR' consistently exhibits higher MAE and RMSE on both *valence* and *arousal* than the other participants, e.g., 'H12DPB' and 'E02OKG'. This variability suggests that individual differences play a significant role in determining the accuracy of the predictions, highlighting the need to consider personalized factors when developing prediction models for emotion in home confinement situations.

Additionally, it is worth noting that while the proposed *valence* models exhibited better prediction performance than baseline models for certain participants (e.g., 'M02MMP', 'L10FGB'), there were participants who did not achieve satisfactory results. The possible reasons may be two-fold: firstly, individual differences play a crucial role in determining *valence*, and it may not be straightforward to accurately predict *valence* solely based on mobile usage behaviors for some participants. This highlights the complexity and subjectivity of *valence* and the need for a more nuanced understanding of individual emotional experiences; secondly, insufficient data availability may have limited the performance of the predictive models for some participants. Having a larger and more diverse dataset would enable better training and improve the performance of the predictive models. These findings contribute to our understanding of personalized prediction models for well-being, which could



(a) MAE for valence (b) MAE for arousal (c) RMSE for valence (d) RMSE for arousal

Fig. 9. Prediction performance of individual models using all features

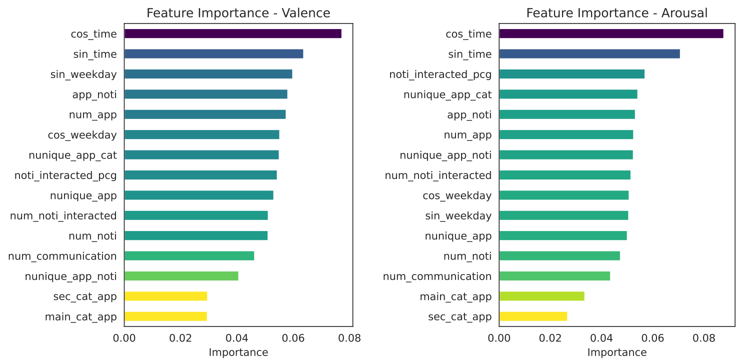
ultimately lead to more tailored interventions and support systems to enhance individuals' well-being during periods of home confinement.

6.4 Feature Importance

To interpret the key factors influencing the prediction of *valence* and *arousal* during home confinement, we examine the feature importance of the best-performing models, which in this case are the RF regressors. For brevity, we focus on the top 15 features for each model, as shown in Figure 10. The RF regressors utilize impurity-based feature importance, which measures how much including a particular feature reduces impurity across the decision tree nodes. A higher score indicates a more important feature, as it leads to a greater reduction in impurity. A higher score indicates a more important feature, as it leads to a greater reduction in impurity. As depicted in Figure 10, we observe that the most important features for both *valence* and *arousal* are similar, but with different weights. These features primarily include time-related features, app-category features, and notification features. These findings suggest that these variables play a significant role in accurately predicting participants' *valence* and *arousal* levels during home confinement.

Both models consistently identify 'sin_time' and 'cos_time' as the top two important features, indicating that the time period of the day significantly influences participants' *valence* and *arousal*. Additionally, the day of the week ('weekday') impacts both models, although it has less significance in the *arousal* model. Surprisingly, the feature 'is_weekend' does not appear among the top 15 features, suggesting that while there is still variation in terms of time and day, weekdays may not have a significant impact on participants' *valence* and *arousal* levels. These findings differ slightly from the observations reported in [39], which found that app-usage patterns varied greatly vary by time and day. The discrepancy may be attributed to the unique circumstances of home confinement, where individuals may experience a loss of time perception and a blurring of the distinction between weekdays and weekends. This aligns with the findings presented in [16], which suggest that the perception of time and daily routines can be disrupted during periods of confinement.

Above findings have important implications for understanding the factors that contribute to human well-being, particularly in the context of home confinement. They provide valuable insights that can inform the development of interventions aimed at promoting emotional wellness during such periods. By considering the impact of app usage, notifications, and time-related factors, we can also design targeted strategies to enhance individuals' emotional well-being and overall quality of life.



(a) Valence (b) Arousal

Fig. 10. Top 15 features based on impurity score with RF regressors

7 IMPLICATIONS

Our study examines the impact of home confinement during the COVID-19 lockdown on individuals' well-being, social roles, and work-life balance.Through the use of unobtrusive mobile phone sensing, we aim to predict participants' well-being and social roles. The findings from our study are significant in raising awareness about the importance of mental health not only during periods of home confinement but also in similar situations such as post-operative recovery, parental leave, aging, or limited mobility related to disability. These findings can also play a crucial role in the early detection and treatment of mental illnesses that may arise during these circumstances. Additionally, policymakers can utilize our research to develop appropriate strategies that minimize the negative effects of pandemic restrictive policies on mental health. To achieve this, quarantine or isolation time should be kept as short as possible, and voluntary quarantine could be promoted to reduce long-term mental health complications, as recommended by Brooks et al. [10].

In the post-pandemic era, the traditional office-only working arrangement has shifted to more flexible arrangements, with an increasing number of people opting to work from home [48]. Studies have shown that individuals who continued to work from home after the pandemic exhibited similar behaviors to those during the pandemic [52], underscoring the importance of studying people's behaviors in high technology-reliance contexts. Our study was conducted in a city with the longest home confinement period, offering optimal experimental conditions to investigate the influence of app usage on well-being while minimizing confounding effects from other variables, such as nature and physical social interactions. Furthermore, our study contributes to understanding the mental health benefits and consequences of working from home or relying heavily on technology. For instance, tracking mental well-being status provides a means of measuring work-life balance and enables individuals to manage their working schedules or arrangements (e.g., work-from-home or half-office half-home) more effectively.

In addition, this study provides a valuable opportunity for researchers to study the psychological states and smartphone usage behaviors of individuals under home confinement conditions, which has important implications for various fields. For example, it may inform the development of interventions to support individuals in similar situations, such as those willing or undergoing post-operative home rehabilitation. Understanding how individuals cope with home confinement and the role that technology plays in their lives can help healthcare professionals or researchers develop effective interventions and design effective remote support programs to improve the overall quality of care for patients in similar situations. Furthermore, the findings from this study may have broader implications for our understanding of human behavior and mental health during times of crisis, informing the development of new policies and interventions to support individuals in future crises.

8 LIMITATIONS

We have identified three main limitations in this research. One significant limitation is the selection of time and place. Regarding time, we were unable to compare user well-being and digital device usage data before COVID-19 lockdown with the data during the lockdown due to the unpredictable nature of lockdown policies. However, we have mitigated this limitation by allowing users to report the impact of the lockdown on their well-being and work-life balance through self-reported questionnaires. These data clearly demonstrate the effects of lockdowns on people's lives, as discussed in Section 4. In terms of geography, our data collection was limited to Melbourne, where the COVID-19 lockdown with Stage 4 restrictions was applied. While the specific measures taken by different countries and regions vary, and the overall mental well-being of populations may differ, studies conducted in other countries have also demonstrated the impact of COVID-19 on human well-being and mental health [9, 10, 13]. Therefore, our research still holds value in informing government officials or managers to make appropriate decisions and highlighting the importance of mental health care during home confinement situations.

The second limitation is the relatively short duration of the data collection period. We chose to collect data for only three weeks in order to minimize interruptions from ESM questionnaires and end-of-day surveys, with the goal of reducing the burden on participants during the challenging realities of COVID-19 home confinement. Despite the short duration, we were still able to collect a substantial amount of data, including 502,485 records of users' context and phone usage, 1,749 ESM responses, and 265 end-of-day survey responses. It is worth noting that the delivery of end-of-day survey link through emails allowed users to complete the survey voluntarily at their convenience. Although the number of end-of-day survey responses is small, the optional nature of the surveys, especially text questions, ensures the credibility and reliability of the collected data. However, due to the limited three-week period of our data collection, it is important to acknowledge that if strict home confinement lasts longer, the emotional experiences of individuals may evolve as they adapt to new challenges.

Finally, the third limitation of this study is the number and representativeness of the participants. A total of 32 participants were recruited, with 27 using smartphones, 13 using desktops, and only 8 using both. Since the majority of participants use mobile phones for data collection, the majority of our background application data analysis focuses on mobile usage data. Additionally, participants were primarily recruited through Facebook and Discord advertisements, which may have led to a population that is skewed towards tech-savvy information workers. Additionally, there is a possibility of oversampling individuals who were 'relieved to be safe at home and cherish the opportunity to spend more quality time with their family,' as these individuals may have had more availability to participate in the study. It is important to acknowledge that the COVID-19 lockdown presented challenges in recruiting participants. However, despite these challenges, we were able to collect sufficient data from each participant to conduct our analysis.

9 CONCLUSION

In this study, we delved into how human emotions, social roles, and mobile usage behaviors were affected during the COVID-19 home confinement period. Our proposed models yielded promising results in predicting human emotions during home confinement. Moreover, we uncovered intriguing insights regarding the correlation between different factors and human well-being. To address our research objectives on home confinement, we formulated three research questions and key findings are summarized below.

In response to RQ1, we found that participants reported relatively high levels of valence but relatively low levels of arousal, suggesting that participants experienced a state of calmness and relaxation during COVID-19 home confinement. Additionally, our analysis suggested that the COVID-19 lockdown has a detrimental effect on valence and work-life balance, with participants experiencing longer work hours due to interruptions and a blurring of personal and work boundaries. Regarding RQ2, we segmented app usage data and identified the most frequent user activities 'Just Communication' and 'Idling', potentially attributed to increased virtual communication and available free time. We also found that tools-related activities positively correlated with valence, suggesting people used productivity apps effectively. Moreover, interruptions displayed a positive association with Communication Tools, possibly due to the social nature of these tools leading to more interruptions during work hours. Moving on to RQ3, we developed regression models to predict people's valence and arousal based on their mobile usage during home confinement. These models exhibited superior performance compared to baseline models. Notably, our findings revealed that time-related features emerged as the primary predictors for both valence and arousal.

Despite the challenges we faced in collecting data during COVID-19 lockdowns, we believe that gaining insights into human well-being and mobile usage behaviors during home confinement is crucial. This understanding is not only valuable for policymakers in implementing effective measures during future outbreaks, but also for individuals to become more self-aware of their well-being and mental health. Moving forward, our research will expand to investigate users' desktop usage behavior. By comparing the differences in usage behavior across different platforms, we aim to gain a deeper understanding of people's daily lives and work-life balance.

10 ACKNOWLEDGMENTS

We sincerely thank the anonymous reviewers.

REFERENCES

[1] Christoph Anderson, Judith S. Heinisch, Sandra Ohly, Klaus David, and Veljko Pejovic. 2019. The Impact of Private and Work-Related Smartphone Usage on Interruptibility. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (London, United Kingdom) *(UbiComp/ISWC '19 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 1058–1063. https://doi.org/10.1145/3341162.3344845

[2] Christoph Anderson, Clara Heissler, Sandra Ohly, and Klaus David. 2016. Assessment of social roles for interruption management: a new concept in the field of interruptibility. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct*. 1530–1535.

[3] Anshika Arora, Pinaki Chakraborty, and MPS Bhatia. 2021. Problematic use of digital technologies and its impact on mental health during COVID-19 pandemic: assessment using machine learning. In *Emerging Technologies During the Era of COVID-19 Pandemic*. Springer, 197–221.

[4] Ibrahim Arpaci and Tugba Kocadag Unver. 2020. Moderating role of gender in the relationship between big five personality traits and smartphone addiction. *Psychiatric Quarterly* 91, 2 (2020), 577–585.

[5] Adrian Bartoszek, Dariusz Walkowiak, Agnieszka Bartoszek, and Grzegorz Kardas. 2020. Mental well-being (depression, loneliness, insomnia, daily life fatigue) during COVID-19 related home-confinement—A study from Poland. *International journal of environmental research and public health* 17, 20 (2020), 7417.

[6] Daniel J. Beal. 2015. ESM 2.0: State of the Art and Future Potential of Experience Sampling Methods in Organizational Research. *Annual Review of Organizational Psychology and Organizational Behavior* 2, 1 (April 2015), 383–407. https://doi.org/10.1146/annurev-orgpsych-032414-111335

[7] Bruce J Biddle. 1986. Recent developments in role theory. *Annual review of sociology* 12, 1 (1986), 67–92.

[8] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (March 1994), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

[9] Annette Braunack-Mayer, Rebecca Tooher, Joanne E Collins, Jackie M Street, and Helen Marshall. 2013. Understanding the school community's response to school closures during the H1N1 2009 influenza pandemic. *BMC public health* 13, 1 (2013), 1–15.

[10] Samantha K Brooks, Rebecca K Webster, Louise E Smith, Lisa Woodland, Simon Wessely, Neil Greenberg, and Gideon James Rubin. 2020. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *The lancet* 395, 10227 (2020), 912–920.

[11] Dana Brownlee. 2020. Twitter, Square Announce Work From Home Forever Option: What Are The Risks? https://www.forbes.com/sites/danabrownlee/2020/05/18/twitter-square-announce-work-from-home-forever-optionwhat-are-the-risks/

[12] Peter André Busch and Stephen McCarthy. 2021. Antecedents and consequences of problematic smartphone use: A systematic literature review of an emerging research area. *Computers in human behavior* 114 (2021), 106414.

[13] Grazia Caleo, Jennifer Duncombe, Freya Jephcott, Kamalini Lokuge, Clair Mills, Evita Looijen, Fivi Theoharaki, Ronald Kremer, Karline Kleijer, James Squire, et al. 2018. The factors affecting household transmission dynamics and community compliance with Ebola control measures: a mixed-methods study in a rural village in Sierra Leone. *BMC public health* 18, 1 (2018), 1–13.

[14] Maureen A Cava, Krissa E Fay, Heather J Beanlands, Elizabeth A McCay, and Rouleen Wignall. 2005. The experience of quarantine for individuals affected by SARS in Toronto. *Public Health Nursing* 22, 5 (2005), 398–406.

[15] Elizabeth L Cay, N Vetter, AE Philip, and Pat Dugard. 1972. Psychological status during recovery from an acute heart attack. *Journal of psychosomatic research* 16, 6 (1972), 425–435.

[16] Nicola Cellini, Natale Canale, Giovanna Mioni, and Sebastiano Costa. 2020. Changes in sleep pattern, sense of time and digital media use during COVID-19 lockdown in Italy. *Journal of sleep research* 29, 4 (2020), e13074.

[17] Tianfeng Chai and Roland R Draxler. 2014. Root mean square error (RMSE) or mean absolute error (MAE). *Geoscientific model development discussions* 7, 1 (2014), 1525–1534.

[18] Jounghwa Choi. 2020. Taylor, S.(2019). The psychology of pandemics: Preparing for the next global outbreak of infectious disease. Newcastle upon Tyne, UK: Cambridge Scholars Publishing. *Asian Communication Research* 17, 2 (2020), 98–103.

[19] Philip I Chow, Karl Fua, Yu Huang, Wesley Bonelli, Haoyi Xiong, Laura E Barnes, and Bethany A Teachman. 2017. Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. *Journal of medical Internet research* 19, 3 (2017), e62.

[20] Michael Daly, Angelina Sutin, and Eric Robinson. 2020. Longitudinal changes in mental health and the COVID-19 pandemic: Evidence from the UK Household Longitudinal Study. (2020).

[21] Krista De Castella, Philippe Goldin, Hooria Jazaieri, Michal Ziv, Carol S Dweck, and James J Gross. 2013. Beliefs about emotion: Links to emotion regulation, well-being, and psychological distress. *Basic and applied social psychology* 35, 6 (2013), 497–505.

[22] Joost CF De Winter, Samuel D Gosling, and Jeff Potter. 2016. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological methods* 21, 3 (2016), 273.

[23] Elena Di Lascio, Shkurta Gashi, and Silvia Santini. 2018. Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018),

1–21.
[24] Clete DiGiovanni, Jerome Conley, Daniel Chiu, and Jason Zaborski. 2004. Factors influencing compliance with quarantine in Toronto during the 2003 SARS outbreak. *Biosecurity and bioterrorism: biodefense strategy, practice, and science* 2, 4 (2004), 265–272.
[25] Konstantinos N Fountoulakis, Maria K Apostolidou, Marina B Atsiova, Anna K Filippidou, Angeliki K Florou, Dimitra S Gousiou, Aikaterini R Katsara, Sofia N Mantzari, Marina Padouva-Markoulaki, Evangelia I Papatriantafyllou, et al. 2021. Self-reported changes in anxiety, depression and suicidality during the COVID-19 lockdown in Greece. *Journal of affective disorders* 279 (2021), 624–629.
[26] Jerome H Friedman. 2002. Stochastic gradient boosting. *Computational statistics & data analysis* 38, 4 (2002), 367–378.
[27] Martha Leticia Gaeta, Laura Gaeta, and Maria del Socorro Rodriguez. 2021. The impact of COVID-19 home confinement on Mexican university students: Emotions, coping strategies, and self-regulated learning. *Frontiers in Psychology* 12 (2021), 642823.
[28] Nan Gao, Mohammad Saiedur Rahaman, Wei Shao, and Flora D Salim. 2021. Investigating the reliability of self-report data in the wild: The quest for ground truth. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*. 237–242.
[29] Nan Gao, Wei Shao, Mohammad Saiedur Rahaman, and Flora D. Salim. 2020. n-Gage: Predicting in-class Emotional, Behavioural and Cognitive Engagement in the Wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–26.
[30] Michelle Hadjiconstantinou, Jo Byrne, Danielle H Bodicoat, Noelle Robertson, Helen Eborall, Kamlesh Khunti, and Melanie J Davies. 2016. Do web-based interventions improve well-being in type 2 diabetes? A systematic review and meta-analysis. *Journal of medical Internet research* 18, 10 (2016), e270.
[31] Laura Hawryluck, Wayne L Gold, Susan Robinson, Stephen Pogorski, Sandro Galea, and Rima Styra. 2004. SARS control and psychological effects of quarantine, Toronto, Canada. *Emerging infectious diseases* 10, 7 (2004), 1206.
[32] Judith S Heinisch, Nan Gao, Christoph Anderson, Shohreh Deldari, Klaus David, and Flora Salim. 2022. Investigating the Effects of Mood & Usage Behaviour on Notification Response Time. *arXiv preprint arXiv:2207.03405* (2022).
[33] Daniel Hintze, Philipp Hintze, Rainhard D Findling, and René Mayrhofer. 2017. A large-scale, long-term analysis of mobile device usage characteristics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 1–21.
[34] Emily A Holmes, Rory C O'Connor, V Hugh Perry, Irene Tracey, Simon Wessely, Louise Arseneault, Clive Ballard, Helen Christensen, Roxane Cohen Silver, Ian Everall, Tamsin Ford, Ann John, Thomas Kabir, Kate King, Ira Madan, Susan Michie, Andrew K Przybylski, Roz Shafran, Angela Sweeney, Carol M Worthman, Lucy Yardley, Katherine Cowan, Claire Cope, Matthew Hotopf, and Ed Bullmore. 2020. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *The Lancet Psychiatry* 7, 6 (June 2020), 547–560. https://doi.org/10.1016/S2215-0366(20)30168-1 Publisher: Elsevier.
[35] Sinh Huynh, Seungmin Kim, JeongGil Ko, Rajesh Krishna Balan, and Youngki Lee. 2018. Engagemon: Multi-modal engagement sensing for mobile games. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 2, 1 (2018), 1–27.
[36] N. Jaques, S. Taylor, A. Azaria, A. Ghandeharioun, A. Sano, and R. Picard. 2015. Predicting students' happiness from physiology, phone, mobility, and behavioral data. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. 222–228. https://doi.org/10.1109/ACII.2015.7344575
[37] H Jeong, HW Yim, YJ Song, M Ki, JA Min, J Cho, JH Chae, et al. 2016. Mental health status of people isolated due to Middle East Respiratory Syndrome. Epidemiol Health. 2016; 38: e2016048.
[38] Stephen Joseph, P Alex Linley, Jake Harwood, Christopher Alan Lewis, and Patrick McCollam. 2004. Rapid assessment of well-being: The short depression-happiness scale (SDHS). *Psychology and psychotherapy: Theory, research and practice* 77, 4 (2004), 463–478.
[39] Kleomenis Katevas, Ioannis Arapakis, and Martin Pielot. 2018. Typical phone use habits: Intense use does not predict negative well-being. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–13.
[40] Jagdish Khubchandani, Sushil Sharma, Fern J Webb, Michael J Wiblishauser, and Sharon L Bowman. 2021. Post-lockdown depression and anxiety in the USA during the COVID-19 pandemic. *Journal of Public Health* 43, 2 (2021), 246–253.
[41] William DS Killgore, Sara A Cloonan, Emily C Taylor, Daniel A Lucas, and Natalie S Dailey. 2020. Loneliness during the first half-year of COVID-19 Lockdowns. *Psychiatry research* 294 (2020), 113551.
[42] Tom Kitwood and Kathleen Bredin. 1992. Towards a theory of dementia care: personhood and well-being. *Ageing & Society* 12, 3 (1992), 269–287.
[43] Sing Lee, Lydia YY Chan, Annie MY Chau, Kathleen PS Kwok, and Arthur Kleinman. 2005. The experience of SARS-related stigma at Amoy Gardens. *Social science & medicine* 61, 9 (2005), 2038–2046.
[44] Jiayu Li, Danni Zhan, Yuhong Zhou, and Xuemei Gao. 2021. Loneliness and problematic mobile phone use among adolescents during the COVID-19 pandemic: The roles of escape motivation and self-control. *Addictive behaviors* 118 (2021), 106857.
[45] Tong Li, Tong Xia, Huandong Wang, Zhen Tu, Sasu Tarkoma, Zhu Han, and Pan Hui. 2022. Smartphone app usage analysis: Datasets, methods, and applications. *IEEE Communications Surveys & Tutorials* 24, 2 (2022), 937–966.
[46] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. 2003. The global k-means clustering algorithm. *Pattern recognition* 36, 2 (2003), 451–461.
[47] Rosa Letı, M Cruz Ortiz, Luis A Sarabia, and M Sagrario Sánchez. 2004. Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. *Analytica Chimica Acta* 515, 1 (2004), 87–100.

[48] Susan Lund, Anu Madgavkar, James Manyika, Sven Smit, Kweilin Ellingrud, Mary Meaney, and Olivia Robinson. 2021. The future of work after COVID-19. *McKinsey Global Institute* 18 (2021).

[49] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017).

[50] Anmol Madan, Manuel Cebrian, Sai Moturu, Katayoun Farrahi, et al. 2011. Sensing the" health state" of a community. *IEEE Pervasive Computing* 11, 4 (2011), 36–45.

[51] Zdravko Marjanovic, Esther R Greenglass, and Sue Coffey. 2007. The relevance of psychosocial variables and working conditions in predicting nurses' coping strategies during the SARS crisis: an online questionnaire survey. *International journal of nursing studies* 44, 6 (2007), 991–998.

[52] Stijn A A Massar, Alyssa S C Ng, Chun Siong Soon, Ju Lynn Ong, Xin Yu Chua, Nicholas I Y N Chee, Tih Shih Lee, and Michael W L Chee. 2021. Reopening after lockdown: the influence of working-from-home and digital device use on sleep, physical activity, and wellbeing following COVID-19 lockdown and reopening. *Sleep* 45, 1 (10 2021). https://doi.org/10.1093/sleep/zsab250 arXiv:https://academic.oup.com/sleep/article-pdf/45/1/zsab250/42462400/zsab250_suppl_supplementary_materials.pdf zsab250.

[53] Alexia Matheson, Claire J McGannon, Atul Malhotra, Kirsten R Palmer, Alice E Stewart, Euan M Wallace, Ben W Mol, Ryan J Hodges, and Daniel L Rolnik. 2021. Prematurity rates during the coronavirus disease 2019 (COVID-19) pandemic lockdown in Melbourne, Australia. *Obstetrics and gynecology* 137, 3 (2021), 405.

[54] Lakmal Meegahapola, William Droz, Peter Kun, Amalia De Götzen, Chaitanya Nutakki, Shyam Diwakar, Salvador Ruiz Correa, Donglei Song, Hao Xu, Miriam Bidoglia, et al. 2023. Generalization and Personalization of Mobile Sensing-Based Mood Inference Models: An Analysis of College Students in Eight Countries. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–32.

[55] Mehrab Bin Morshed, Koustuv Saha, Richard Li, Sidney K. D'Mello, Munmun De Choudhury, Gregory D. Abowd, and Thomas Plötz. 2019. Prediction of Mood Instability with Passive Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3 (Sept. 2019). https://doi.org/10.1145/3351233 Place: New York, NY, USA Publisher: Association for Computing Machinery.

[56] Andreas C Müller and Sarah Guido. 2016. *Introduction to machine learning with Python: a guide for data scientists.* " O'Reilly Media, Inc.".

[57] Yara Murray-Atfield and Joseph Dunstan. 2020. Curfew and exercise limits for Melbourne residents, regional Victoria moved to Stage Three lockdown. https://www.abc.net.au/news/2020-08-02/victoria-coronavirus-restrictions-imposed-death-toll-cases-rise/12515914

[58] Marco Narici, Giuseppe De Vito, Martino Franchi, Antonio Paoli, Tatiana Moro, Giuseppe Marcolin, Bruno Grassi, Giovanni Baldassarre, Lucrezia Zuccarelli, Gianni Biolo, et al. 2021. Impact of sedentarism due to the COVID-19 home confinement on neuromuscular, cardiovascular and metabolic health: Physiological and pathophysiological implications and recommendations for physical and nutritional countermeasures. *European journal of sport science* 21, 4 (2021), 614–635.

[59] 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 *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 42, 19 pages. https://doi.org/10.1145/3491102.3502043

[60] Mikael Nordenmark. 2004. Multiple social roles and well-being: a longitudinal test of the role stress theory and the role expansion theory. *Acta Sociologica* 47, 2 (2004), 115–126.

[61] Tadashi Okoshi, Jin Nakazawa, and Hideyuki Tokuda. 2014. Attelia: Sensing User's Attention Status on Smart Phones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 139–142. https://doi.org/10.1145/2638728.2638802 event-place: Seattle, Washington.

[62] T. Okoshi, J. Ramos, H. Nozaki, J. Nakazawa, A. K. Dey, and H. Tokuda. 2015. Attelia: Reducing user's cognitive load due to interruptive notifications on smart phones. In *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 96–104. https://doi.org/10.1109/PERCOM.2015.7146515

[63] Neena Pandey, Abhipsa Pal, et al. 2020. Impact of digital surge during Covid-19 pandemic: A viewpoint on research and practice. *International journal of information management* 55 (2020), 102171.

[64] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. *arXiv preprint cs/0205070* (2002).

[65] Delroy L Paulhus. 1991. Measurement and control of response bias. (1991).

[66] PT Pope and JT Webster. 1972. The use of an F-statistic in stepwise regression procedures. *Technometrics* 14, 2 (1972), 327–340.

[67] Haroon Rashid, Sanjana Mendu, Katharine E Daniel, Miranda L Beltzer, Bethany A Teachman, Mehdi Boukhechba, and Laura E Barnes. 2020. Predicting subjective measures of social anxiety from sparsely collected mobile sensor data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–24.

[68] Zubair Ahmed Ratan, Sojib Bin Zaman, Sheikh Mohammed Shariful Islam, and Hassan Hosseinzadeh. 2021. Smartphone overuse: A hidden crisis in COVID-19. *Health Policy and Technology* 10, 1 (2021), 21.

[69] Steven P Reise and Niels G Waller. 2009. Item response theory and clinical measurement. *Annual review of clinical psychology* 5, 1 (2009), 27–48.

[70] Diane L Reynolds, JR Garay, SL Deamond, Maura K Moran, W Gold, and R Styra. 2008. Understanding, compliance and psychological impact of the SARS quarantine experience. *Epidemiology & Infection* 136, 7 (2008), 997–1007.

[71] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.

[72] Paula Rodríguez-Fernández, Josefa González-Santos, Mirian Santamaría-Peláez, Raúl Soto-Cámara, Esteban Sánchez-González, and Jerónimo J González-Bernal. 2021. Psychological effects of home confinement and social distancing derived from COVID-19 in the general population—a systematic review. *International Journal of Environmental Research and Public Health* 18, 12 (2021), 6528.

[73] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[74] Amin Sadri, Yongli Ren, and Flora Salim. 2017. Information gain-based metric for recognizing transitions in human activities. *Pervasive and Mobile Computing* (2017). https://doi.org/10.1016/j.pmcj.2017.01.003

[75] Zhanna Sarsenbayeva, Gabriele Marini, Niels van Berkel, Chu Luo, Weiwei Jiang, Kangning Yang, Greg Wadley, Tilman Dingler, Vassilis Kostakos, and Jorge Goncalves. 2020. Does Smartphone Use Drive Our Emotions or Vice Versa? A Causal Analysis. 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/10.1145/3313831.3376163

[76] K Sebire. 2020. The coronavirus lockdown is forcing us to view'screen time'differently. That's a good thing. *The Conversation* (2020).

[77] Mark R Segal. 2004. Machine learning benchmarks and random forest regression. (2004).

[78] Yunsheng Song, Jiye Liang, Jing Lu, and Xingwang Zhao. 2017. An efficient instance selection algorithm for k nearest neighbor regression. *Neurocomputing* 251 (2017), 26–34.

[79] Scott Stansfield, Arsheeya Rattan, Ben W Mol, Daniel L Rolnik, and Atul Malhotra. 2022. Impact of the COVID-19 pandemic and multiple community lockdowns on total live birth rates and preterm births in Melbourne, Australia. *Australian and New Zealand Journal of Obstetrics and Gynaecology* (2022).

[80] Benjamin Tag, Zhanna Sarsenbayeva, Anna L Cox, Greg Wadley, Jorge Goncalves, and Vassilis Kostakos. 2022. Emotion trajectories in smartphone use: Towards recognizing emotion regulation in-the-wild. *International Journal of Human-Computer Studies* 166 (2022), 102872.

[81] Dalila Talevi, Valentina Socci, Margherita Carai, Giulia Carnaghi, Serena Faleri, Edoardo Trebbi, Arianna di Bernardo, Francesco Capelli, and Francesca Pacitti. 2020. Mental health outcomes of the CoViD-19 pandemic. *Rivista di psichiatria* 55, 3 (2020), 137–144.

[82] ML Tlachac, Ricardo Flores, Miranda Reisch, Katie Houskeeper, and Elke A Rundensteiner. 2022. DepreST-CAT: Retrospective Smartphone Call and Text Logs Collected during the COVID-19 Pandemic to Screen for Mental Illnesses. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–32.

[83] Aruna Tyagi, Anoop Kumar Prasad, and Dinesh Bhatia. 2021. Effects of excessive use of mobile phone technology in India on human health during COVID-19 lockdown. *Technology in society* 67 (2021), 101762.

[84] Niels van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The Experience Sampling Method on Mobile Devices. *Comput. Surveys* 50, 6 (Dec. 2017), 93:1–93:40. https://doi.org/10.1145/3123988

[85] 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. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1 (March 2018). https://doi.org/10.1145/3191775 Place: New York, NY, USA Publisher: Association for Computing Machinery.

[86] 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. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–21.

[87] Tzu-Tsung Wong and Po-Yang Yeh. 2019. Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering* 32, 8 (2019), 1586–1594.

[88] World Health Organization and others. 2020. *Mental health and psychosocial considerations during the COVID-19 outbreak.* Technical Report. World Health Organization.

[89] Jiaqi Xiong, Orly Lipsitz, Flora Nasri, Leanna M. W. Lui, Hartej Gill, Lee Phan, David Chen-Li, Michelle Iacobucci, Roger Ho, Amna Majeed, and Roger S. McIntyre. 2020. Impact of COVID-19 pandemic on mental health in the general population: A systematic review. *Journal of Affective Disorders* 277 (2020), 55 – 64. https://doi.org/10.1016/j.jad.2020.08.001

[90] Ye Xu, Mu Lin, Hong Lu, Giuseppe Cardone, Nicholas Lane, Zhenyu Chen, Andrew Campbell, and Tanzeem Choudhury. 2013. Preference, context and communities: a multi-faceted approach to predicting smartphone app usage patterns. In *Proceedings of the 2013 International Symposium on Wearable Computers*. 69–76.

[91] Mi-Kyung Yoon, Soon-Young Kim, Hye-Sun Ko, and Myung-Soo Lee. 2016. System effectiveness of detection, brief intervention and refer to treatment for the people with post-traumatic emotional distress by MERS: a case report of community-based proactive intervention in South Korea. *International journal of mental health systems* 10, 1 (2016), 1–5.

[92] Xiao Zhang, Wenzhong Li, Xu Chen, and Sanglu Lu. 2018. MoodExplorer: Towards Compound Emotion Detection via Smartphone Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 176 (jan 2018), 30 pages. https://doi.org/10.1145/3161414

[93] Donald W Zimmerman and Bruno D Zumbo. 1993. Rank transformations and the power of the Student t test and Welch t'test for non-normal populations with unequal variances. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 47, 3 (1993), 523.