

# MoodCapture: Depression Detection using In-the-Wild Smartphone Images

SUBIGYA NEPAL\*, Dartmouth College, USA

ARVIND PILLAI\*, Dartmouth College, USA

WEICHEN WANG, Dartmouth College, USA

TESS GRIFFIN, Dartmouth College, USA

AMANDA COLLINS, Dartmouth College, USA

MICHAEL HEINZ, Dartmouth College, USA

DAMIEN LEKKAS, Dartmouth College, USA

SHAYAN MIRJAFARI, Dartmouth College, USA

MATTHEW NEMESURE, Dartmouth College, USA

GEORGE PRICE, Dartmouth College, USA

NICHOLAS JACOBSON, Dartmouth College, USA

ANDREW T. CAMPBELL, Dartmouth College, USA

MoodCapture presents a novel approach that assesses depression based on images automatically captured from the front-facing camera of smartphones as people go about their daily lives. We collect over 125,000 photos in the wild from N=177 participants diagnosed with major depressive disorder over a period of 90-days. Images are captured while participants respond to the PHQ-8 depression survey question: “*I have felt down, depressed, or hopeless*”. Our analysis explores important image attributes, such as, angle, dominant colors, location, objects and lighting. Our deep learning (DL) model using raw images yields an F1-score = 0.75 for detecting depression, while an ablation study demonstrates that hand-crafted rigidity parameters with machine learning (ML) provides greater discriminative power. Importantly, we evaluate user concerns of using MoodCapture technology to detect depression based on sharing photos, providing valuable insights into privacy concerns that inform the future design of in-the-wild imaged-based mental health assessment tools.

## ACM Reference Format:

Subigya Nepal\*, Arvind Pillai\*, Weichen Wang, Tess Griffin, Amanda Collins, Michael Heinz, Damien Lekkas, Shayan Mirjafari, Matthew Nemesure, George Price, Nicholas Jacobson, and Andrew T. Campbell. 2023. MoodCapture: Depression Detection using In-the-Wild Smartphone Images. 1, 1 (September 2023), 21 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

---

\*These authors contributed equally.

Authors' addresses: Subigya Nepal\*, Dartmouth College, Computer Science, USA; Arvind Pillai\*, Dartmouth College, Computer Science, USA; Weichen Wang, Dartmouth College, Computer Science, USA; Tess Griffin, Dartmouth College, Computer Science, USA; Amanda Collins, Dartmouth College, Computer Science, USA; Michael Heinz, Dartmouth College, Computer Science, USA; Damien Lekkas, Dartmouth College, Computer Science, USA; Shayan Mirjafari, Dartmouth College, Computer Science, USA; Matthew Nemesure, Dartmouth College, Computer Science, USA; George Price, Dartmouth College, Computer Science, USA; Nicholas Jacobson, Dartmouth College, Computer Science, USA; Andrew T. Campbell, Dartmouth College, Computer Science, USA.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2023 Association for Computing Machinery.

Manuscript submitted to ACM

Under Review at CHI 2024.

## 1 INTRODUCTION

Depression is a complex and pervasive mental health issue affecting millions of people worldwide. According to the World Health Organization (WHO), over 264 million people suffer from depression [33], making it a leading cause of disability and a major contributor to the overall global burden of disease. The consequences of depression extend beyond emotional distress [38], significantly impacting physical health [16, 32], social relationships [41], and occupational functioning [14]. In severe cases, depression can lead to suicide, accounting for nearly 800,000 deaths each year [2, 6]. The need for early detection and intervention in depression is critical, as timely identification of the condition allows individuals to access appropriate treatment and support, thereby improving clinical outcomes and reducing the risk of long-term complications [15, 39]. However, traditional methods of depression detection, such as self-report questionnaires and clinical interviews, often face limitations in terms of accuracy, timeliness, and accessibility. These approaches depend heavily on the subjective experiences and communication abilities of individuals, which can be influenced by various factors, including social desirability bias, stigma, and lack of self-awareness. In light of these challenges, the widespread use of smartphones in everyday life offers an unprecedented opportunity to explore alternative approaches to depression detection that are more objective, unobtrusive, and continuous. The vast amounts of data generated through daily smartphone usage, including images, text messages, and social media interactions, provide a rich and ecologically valid source of information that can be utilized to gain insights into individuals' mental states. Consequently, several studies have made use of smartphone sensing data to assess depression [10, 46].

In this paper, we report on MoodCapture, a novel and divergent approach from these prior studies that is solely based on periodically captured facial images and self-reported depression of smartphone users. This is made possible by the continuous improvement of smartphone cameras, which has enabled the capture of high-quality in-the-wild images, offering valuable information about users' emotions, expressions, and other visual cues related to mental health. Taking advantage of these advancements, recent progress in deep learning and image analysis, has the potential to transform the way we detect and monitor depression. By training deep learning models on extensive datasets of high-quality in-the-wild smartphone images, researchers can identify intricate patterns and features associated with depression, ultimately leading to the creation of more accurate, efficient, and personalized prediction tools. Moreover, the use of deep learning models can help address some of the challenges associated with traditional depression detection methods, such as, the reliance on subjective self-reports and the need for time-consuming clinical assessments. MoodCapture spontaneously captures face images 'in-the-wild', providing a direct window into people's expressions, contextual environments, and depressive states. Importantly, most of these spontaneous face images captured in our study show neutral poses, which is in direct contrast to images uploaded to social media sites. These social media images can be performative and influenced by biases, such as, social desirability and self-presentation, whereas the MoodCapture app is designed to capture authentic, unguarded facial expressions. This method minimizes the influence of self-awareness on emotions, thereby enhancing the credibility of our data and rendering our technique a more robust tool for detecting depression.

This paper contributes to the growing intersection of human-computer interaction (HCI) research and mental health assessment by investigating the potential of in-the-wild smartphone images and deep learning models for identifying depressive symptoms. We collected over 125,000 images from N=177 participants diagnosed with major depressive disorder over a period of 3 months, utilizing 87 distinct types of Android devices owned by users in the study. On average, each participant contributed 6 photos per day, resulting in a diverse and comprehensive dataset. We provide a comprehensive analysis of various image characteristics obtained from these images captured in-the-wild. We evaluate

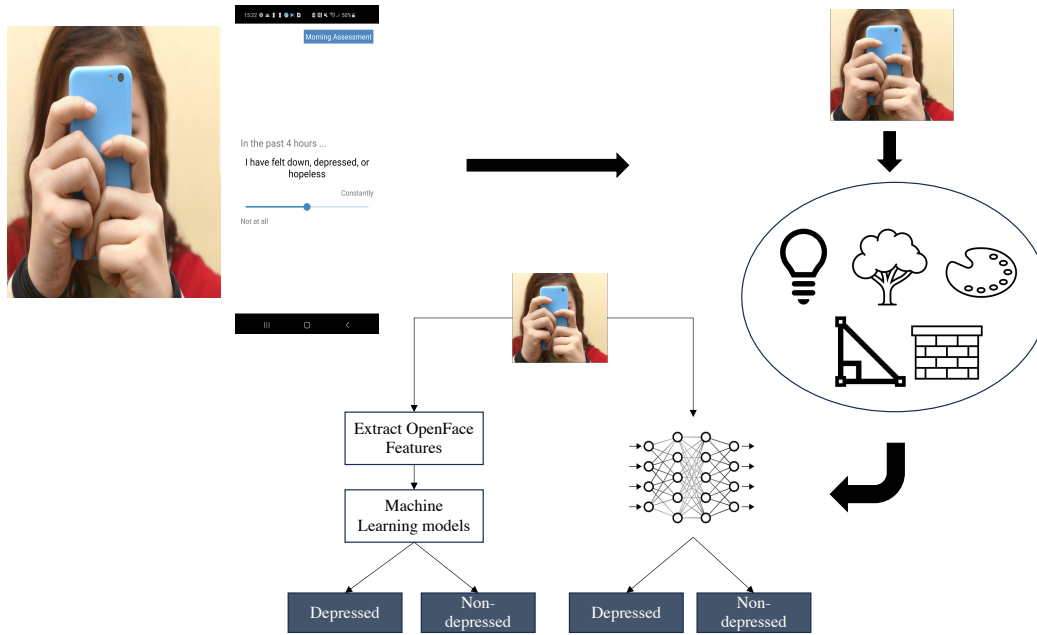


Fig. 1. MoodCapture Framework: Users answer the PHQ-8 depression survey questions using the MoodCapture Android App while the app takes bursts of photos using the front-facing camera (top-left) on the smartphone. Image characteristics are analysed using factors, such as, illumination, indoor vs. outdoors, phone angle, dominant image color, and background objects (top-right). For depression detection, OpenFace features are extracted to train machine learning models, while raw images are used to train deep learning models (bottom).

the performance of a machine learning and deep learning models trained to predict depression based on these images, as shown in Fig. 1. At the end of the study period, we assess user acceptance by inquiring about participants' comfort levels and privacy concerns in sharing their photos for mental health assessment purposes. By integrating HCI principles and methodologies with mental health assessment, our research aims to foster the development of more user-centered, effective, and ethically sound tools for mental health assessment and intervention. The contributions of our work are as follows:

- We develop a first-of-its-kind passive-sensing image-based mobile app called MoodCapture that automatically collects in-the-wild smartphone images from participants' front-facing cameras, ensuring an unobtrusive data collection process and maintaining user privacy. Images are captured while participants respond to the PHQ-8 depression survey question: *"I have felt down, depressed, or hopeless"*. While users consent to have photos taken using the front-facing camera during the operation of the MoodCapture app they are not informed exactly when these photos are captured to promote in the moment naturalistic and authentic capture of users' expressions.
- We analyze various image characteristics providing insights into the visual properties of smartphone images that are captured in-the-wild, including, illumination (i.e., the amount and quality of light in the photos, such as, whether they are well-lit or poorly lit), indoor vs. outdoors, phone angle (i.e., the angle/orientation at which the smartphone is held when the photo is captured), dominant image color (i.e., the primary or most prevalent colors in the images) and objects (e.g., window, pillow, book) . We find the vast majority of images are taken

**Under Review at CHI 2024.**

in a well-lit environment, mostly indoors and from a low angle (i.e., the smartphone is being held by the user lower than their face and the phone front-camera is tilting upwards toward the face).

- We train and evaluate different machine learning and deep learning models on the collected images, demonstrating the feasibility and effectiveness of using in-the-wild smartphone images for predicting depression status (depressed vs non-depressed). Our best performing deep learning model obtains an F1-score of 0.75 based on raw images.
- We report on user acceptance with respect to the comfort levels of the participants in sharing their photos for mental health assessment, providing valuable insights into privacy concerns that inform the future design of in-the-wild image-based mental health assessment tools.

In addition to its relevance to the HCI community, our the MoodCapture study contributes to affective computing, which deals with recognizing, interpreting, and simulating human emotions. By leveraging computational methods and deep learning models to interpret emotional cues from images, our research contributes to the understanding and development of affective computing within the HCI field. Furthermore, our study has tangible, real-world implications, such as the potential benefits of early depression detection, timely interventions, improved clinical outcomes, and overall well-being for individuals.

This paper is structured as follows: Section 2, presents related works in depression detection and work that uses smartphone images. Section 3 details the MoodCapture study, participant demographics, and the analysis we perform to identify image characteristics and to detect depression. Section 4, discusses our results, while Section 5 describes the ethical considerations and user acceptance study. Section 6 discusses the study findings and its implications. Finally, Section 7 and Section 8, discuss the limitations of the study and provide some concluding remarks, respectively.

## 2 RELATED WORK

Depression is a pervasive mental health disorder that has traditionally been diagnosed through methods heavily reliant on individual self-reporting. Established tools such as the Beck Depression Inventory (BDI)[5] and the Hamilton Depression Rating Scale (HDRS)[25] serve as the benchmark for evaluating an individual’s mental health in both clinical and research domains. However, these instruments have some limitations. Their vulnerability lies in their dependence on individuals’ subjective recollections and articulations. The validity of these evaluations can be affected by social desirability bias, mental health stigmas, or an individual’s diminished self-awareness [13, 22, 42]. Given the inherent challenges of traditional tools, the research community has displayed strong interest in leveraging objective and continuous means of monitoring mental health. One promising avenue stems from the pervasive nature of smartphones. Their intrinsic ability to generate multifaceted data makes them an excellent candidate for unobtrusive depression detection. Many studies use passive data sources and derivatives of routine smartphone use to predict and ascertain depression indicators. By evaluating patterns in call logs, text messages, GPS coordinates, and overall smartphone activity, researchers have gained insights into behavioral shifts, social engagement frequencies, and alterations in daily routines, all of which can serve as indicators of deteriorating mental health [10, 46, 47]. Similarly, the growth of social media platforms provides ways to harness user-generated content for depression detection. In particular, analytical approaches using text and images have been applied to content from platforms like Facebook and Instagram. For instance, the linguistic attributes of posts can shed light on a user’s emotional state, sentiment, and overall mental well-being [8, 12]. Moreover, machine learning algorithms have been employed to decipher patterns and indicators of

**Under Review at CHI 2024.**

depression from visual content shared on these platforms. Such analyses often encompass aspects like colors, objects, scenes, and overall aesthetics [17, 19, 37].

Our study emphasizes the analysis of "in-the-wild" smartphone images, particularly those captured via front-facing cameras of smartphones. Given the spontaneous nature of such images, they offer a direct window into an individual's emotions, expressions, and environment. Thus, enhancing the accuracy of mental health assessments. In contrast to social media content, these images remain relatively free from biases like social desirability and self-presentation, which often affect traditional tools. Emotion recognition from facial features has received significant attention in computer vision, with applications spanning from education to healthcare [31]. Prior research has explored the efficacy of images for mental health assessments. These studies explored facial expressions, gaze patterns, and the overall composition of images to extract visual markers symptomatic of depression [26, 28, 30]. However, a significant portion of these studies are conducted in controlled environments or rely on participants deliberately capturing their images, which could inadvertently influence their emotional portrayal. For instance, in [26], photographs were captured using a tablet in a standardized clinical setting. Participants were asked to sit in front of a white background, remove any hats or glasses, and tie up long hair to expose their ears. They were then instructed to relax their facial expressions and look straight ahead. In [30], the authors employed a multi-modal deep Convolutional Neural Network (CNN), considering both facial expressions and body movements. They captured video using a 4K high-resolution camera in a controlled laboratory setting during psychotherapy sessions. Consequently, participants' expressions and body movements were analyzed in a highly regulated context. Numerous other studies have similarly relied on advanced devices for image capture, used video recordings, or incorporated additional signals (such as movement, audio) within controlled environments [20, 23, 34, 36, 48]. Our work aims to address these limitations by examining the feasibility of using spontaneously captured images from participants' smartphones, which could offer a more natural and less intrusive method for predicting depression. As smartphones have become an integral part of modern life, they are an ideal tool for unobtrusive and widespread data collection. By utilizing smartphone cameras to capture participants' images, our approach eliminates the need for controlled environments or deliberate image-taking, thereby reducing the potential for biased emotional portrayals. Furthermore, the widespread availability of smartphones enables our method to reach a larger and more diverse population, ultimately promoting greater accessibility and inclusivity in mental health assessments.

A limited number of past research have used truly "in-the-wild" smartphone images for mental health evaluation. For instance, Wang et al. [44] collected 5811 opportunistic photos in-the-wild from 37 students over 10 weeks using the front-facing camera of their phone. The study reported that depression scores significantly correlates with the students' facial expression and activity. While Wang et al. [44] were the first to use in-the-wild images from front-facing phone cameras to study mental health on a non-clinical population of college students, the authors state that there was insignificant signal in the images to predict self-reported depression. MoodCapture is inspired by this original work which was part of the StudentLife study [45] in 2013. Our progress is that a decade on from the StudentLife study phone cameras have seen significant advancements, leading to substantial differences in their capabilities compared to those from ten years ago, for example, new phone cameras typically offer much higher resolution and more megapixels than those from a decade ago resulting in sharper and more detailed face photos; advances in sensor technology and image processing have greatly improved low-light performance resulting in today's phone cameras capturing better quality face photos in low-light conditions; optical Image stabilization has become more common in smartphone cameras today, reducing the impact of shaky hands and resulting in smoother sharper photos, especially in low light; and finally front-facing cameras primarily designed for selfie shots have improved significantly in terms of resolution, image

**Under Review at CHI 2024.**

quality, auto-focus on the face. Other differences between Wang et al. [44] and our work is that we take advantage of massive advances presented by deep learning models and focus not on a non-clinical group but a clinical population.

Khamis et al. [24] studied the visibility of face and eye in 25,726 in-the-wild images of smartphone users and found that the full face is visible about 29% of the time. Authors stated that their state-of-the-art face detection algorithm performed poorly against photos taken from front-facing cameras. Similarly, other studies have used in-the-wild images to study visual attention and gaze of users [3]. Darvari et al. [11], on the other hand, used in-the-wild images from rear-facing cameras. The authors developed a smartphone application that allows users to periodically log their emotional state together with pictures from their everyday lives. They collected 3,305 mood reports with photos from 22 participants. Authors report finding context-dependent associations between objects surrounding individuals and their self-reported emotional state. However, the genuine spontaneity of these captures and their potential for unbiased mental health evaluation remain relatively unexplored. Our contribution to this growing field pivots on the innovative use of genuinely spontaneous, in-the-wild facial images for depression detection. By employing a passive-sensing mobile application that seamlessly captures images without the subject's acute awareness, we negate the potential influence of self-awareness on emotional representation. This strategy bolsters the ecological validity of our data source, making it a robust tool for depression detection. Further, we delve deep into the content and intrinsic characteristics of these images from a Human-Computer Interaction (HCI) standpoint. By comparing these characteristics with deep learning models trained on extensive datasets of spontaneously captured, high-quality images, we aim to further the connection between HCI research and mental health evaluation.

### 3 METHODOLOGY

In what follows, we discuss the design of our MoodCapture study, demographic information of the individuals that participated in the study and the ground-truth used for analysis.

#### 3.1 Study Design

We recruited 181 participants from across the United States using targeted online advertisements on Google and Facebook. Each participant underwent a clinician-administered Structured Clinical Interview for DSM-5, and only those diagnosed with Major Depressive Disorder (MDD), without bipolar disorder, active suicidality, or psychosis, were eligible for the study. Upon qualification, participants installed our Android-based mobile sensing app on their devices, which gathered ecological momentary assessments during the 90-day study period. The app prompted participants to complete a brief Patient Health Questionnaire-8 (PHQ-8) survey about their depressive symptoms three times daily (morning, afternoon, and evening). As participants answered their daily surveys, the app was designed to discreetly capture a burst of up to 5 images using the front-facing camera. Specifically, images were taken when participants responded to the PHQ-8 item: *"I have felt down, depressed, or hopeless."* (see Fig. 2). We chose this question as we believed it would best capture participants' genuine emotions related to depression. The PHQ-8 is a validated inventory for measuring depression. For further information about the survey, please refer to the Ground Truth section.

During the onboarding process, we informed participants about the image capture procedure and emphasized that sharing their photos was optional. Upon launching the mobile app for the first time, participants were asked, *"To help us better understand your depressive symptoms, we would like to take a few photos in the background that capture your facial expressions while you fill out questionnaires. Do you give us permission to do this?"* Participants could respond with either "Yes" or "No." If they agreed to share their photos, the app captured images as they answered the ecological momentary assessment (EMA). If they opted not to share their photos, no images were captured. The image capture process was

**Under Review at CHI 2024.**

designed to be unobtrusive, with only a green dot at the top of the Android status bar/screen indicating camera usage – which users’ may or may not have observed. Participants did not see their face or receive any other indication that photos were being taken. This discreet image capture process ensured a seamless user experience without interrupting or obstructing the EMA flow. As stated earlier; while participants consented to have photos taken using the front-facing camera during the operation of the MoodCapture app in the study they were not informed exactly when these photos were captured promoting in the moment naturalistic and authentic capture of users’ faces and surroundings.

Participants were compensated \$1 for each completed EMA, with an additional \$50 bonus for achieving a completion rate of 90% or higher during the study period. Compensation was not dependent on sharing photos; participants were compensated regardless of their photo consent. The study was approved by the study institution’s Internal Review Board (IRB). Our analysis and predictive modeling focused on 177 out of the 181 participants who provided consent for their photos to be captured. We collected 125,335 images from these participants, excluding 15,063 photos that were either too blurry, contained no faces, featured children, or contained nudity.

### 3.2 Demographics

The majority of participants in our study identified as female (86.4%, N=153) followed by male (9.6%, N=17) and non-binary (2.8%, N=5). In terms of race, 83.6% (N=148) are White, 2.8% (N=5) are Asians, 4.5% (N=8) are Black or African American, 0.5% (N=1) are American Indian/Alaska Native and 6.7% (N=12) belong to more than one race. See Table 1 for the detailed breakdown.

Table 1. Demographics: Demographic composition of the participants in our study.

Category	Count	Percentage
<i>Sex</i>		
Female	153	86.4%
Male	17	9.6%
Non-binary	5	2.8%
Other (prefer to self-describe)	2	1.1%
<i>Race</i>		
White	148	83.6%
Asian	5	2.8%
Black or African American	8	4.5%
American Indian/Alaska Native	1	0.5%
More than one race	12	6.7%
Other (prefer to self-describe)	3	1.6%

### 3.3 Ground Truth

The PHQ-8, a well-validated instrument consisting of eight items, is widely used for measuring depression [27]. In our study, we adapted the PHQ-8 scale, which typically ranges from 0-4, to a continuous scale of 0-100 for more convenient smartphone responses (see Figure 2). A standard PHQ-8 score of 10 or higher (out of 27) signifies major depression [1]. In our continuous scale, this corresponds to a score of 334 or higher (i.e., 10/27 times 900). Thus, a PHQ-8 score exceeding 334 indicates depression.

To enhance the reliability and accuracy of the EMA responses, we employed a validation technique wherein the app randomly reversed one question in each PHQ-8 survey (thus adding an additional item), ensuring that participants are

**Under Review at CHI 2024.**

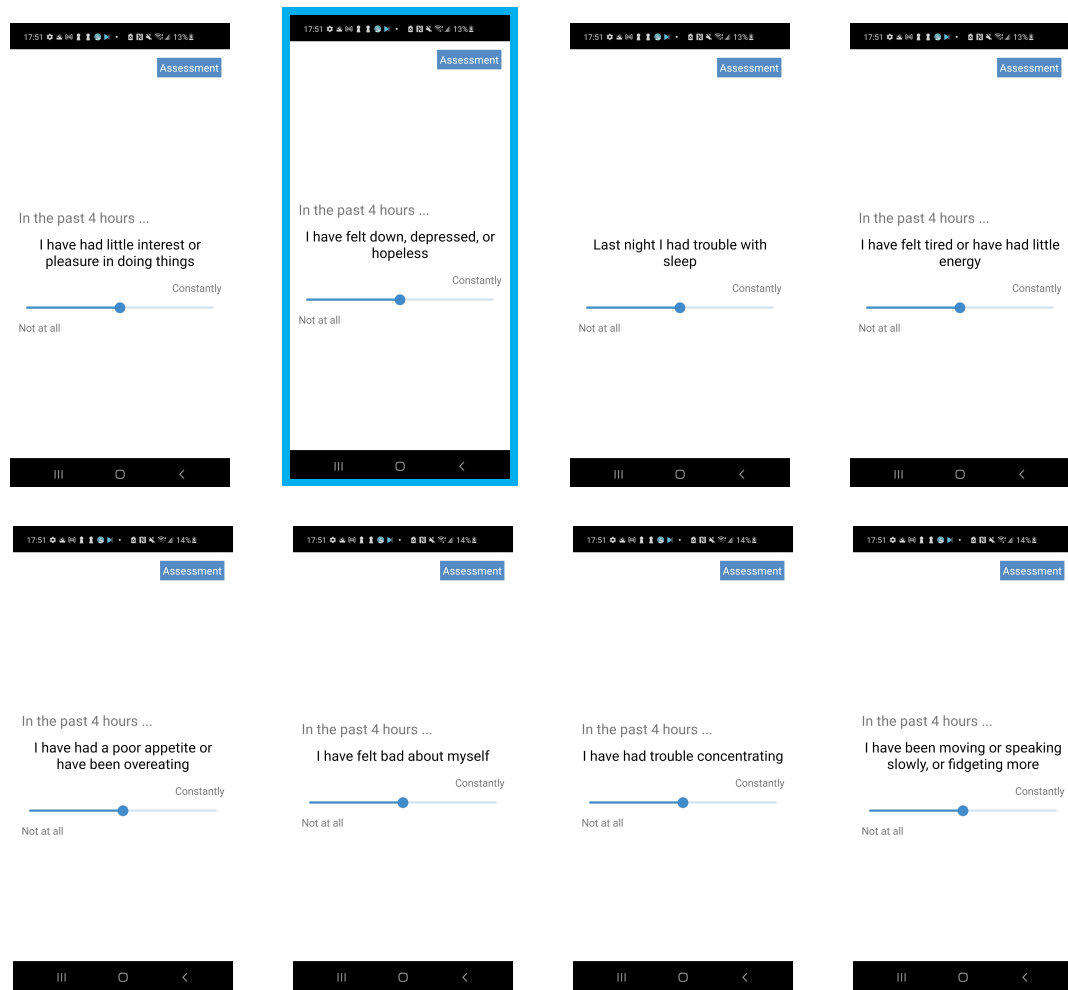


Fig. 2. PHQ-8 application screens for each item: Images are always captured while users respond to the PHQ-8 depression survey question (highlighted in cyan): “I have felt down, depressed, or hopeless”. While user’s consent to have photos taken using the front-facing camera during the operation of the MoodCapture app they are not informed exactly when these photos are captured to promote in the moment naturalistic and authentic capture of users.

attentive. We then compared the responses to the original and reversed questions; if there is a significant discrepancy, the response is excluded from our analysis. After applying this filtering process, we obtain a refined dataset comprising 31,215 ecological momentary assessments (EMAs). Since we captured a burst of images with each EMA response, we amassed 110,272 images in total. As depicted in Figure 3a, we divided our dataset into two groups: depressed (74,347 images,  $N=175$ ) and non-depressed (35,925 images,  $N=156$ ). On average, participants submitted 176 EMAs (stdev = 78) and 623 images (stdev = 278) per participant during the study period. It is crucial to note that all participants recruited for this study had major depressive disorder. Consequently, they reported being below the cut-off threshold on some days and above it on others. However, 19 participants consistently reported depression throughout the study.

**Under Review at CHI 2024.**



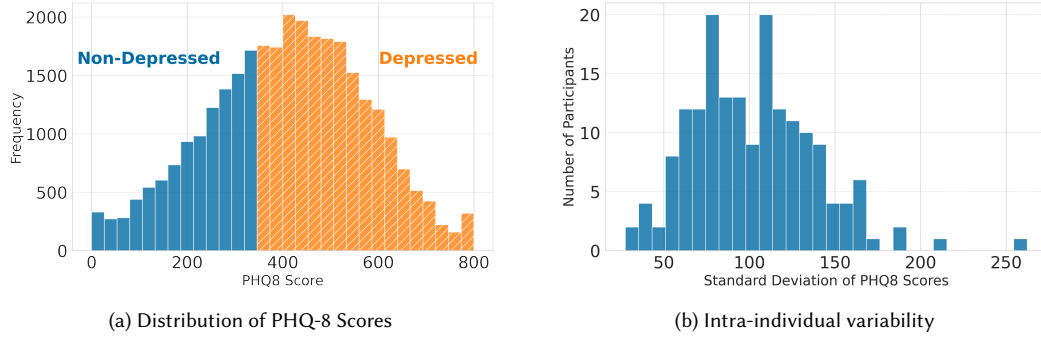


Fig. 3. PHQ-8 score statistics: Figure (a) depicts the distribution of the PHQ-8 score reported by the participant and which group each response falls on (i.e., Depression or No Depression). Figure (b) showcases the variability of PHQ-8 scores among participants over the duration of the study.

Figure 3b shows the variability of PHQ-8 scores among participants i.e., Intra-individual variability. It provides insight into the fluctuations in a participant’s scores over time. On average, participants’ scores varied around their own mean by approximately 101.92 points, with the variability ranging widely from a standard deviation of 27.56 points to as high as 262.24 points. This suggests that some participants had relatively stable scores over time, while others exhibited more pronounced fluctuations.

### 3.4 Image Characteristics

We gather in-the-wild images captured by participants using a diverse range of smartphones with varied configurations and camera placements. Specifically, participants used 87 different types of smartphones, manufactured by 9 different brands, including 46 Samsung models, 10 Google models, 11 Motorola models, and several other brands such as LG Electronics and OnePlus. A total of 107 Samsung, 36 Google, and 19 Motorola devices were used in the study, among others. The images were stored in JPEG format with varying resolutions, depending on the device and the camera used. The most common resolutions were 3648x2736 (used by 57 participants, 32%), 3264x2448 (used by 52 participants, 29%), and 2640x1980 (used by 16 participants, 9%). The resolution ranged from 1920x1080 to 4656x3488. The file size of the images is associated with their resolution – the average file size was 1.12 MB, with a range of 66 KB to 6.02 MB. Our naturalistic approach at capturing image ensures ecological validity and represents users’ natural behavior while engaging with their devices in different environments.

To examine the characteristics of these images, we analyze factors such as phone angle, dominant color, lighting condition, photo location, and background elements present in the photos. The in-the-wild smartphone images offer a unique glimpse into the multitude of ways users interact with their devices and surroundings. However, extracting meaningful insights from these images demands a refined approach that acknowledges the diverse contexts in which they are captured. To achieve this, we utilize the BLIP [29] visual question answering (VQA) model, an advanced AI tool specifically designed for image analysis and answering questions about image content and context. BLIP is recognized as a state-of-the-art method for visual question answering tasks. With the help of the VQA model, we explore the following characteristics:

**Image Angle:** By inquiring about the image angle, we gain an understanding of user interaction dynamics with their devices. Varying angles, such as high or low, offer insights into users’ physical engagement with their smartphones.

**Under Review at CHI 2024.**

High, low, or level angle refers to the perspective from which an image is captured or taken with respect to the subject in the frame. A high angle shot is taken from above the subject, looking down on it. A low angle shot is taken from below the subject, looking up at it. A level angle shot is taken from the same height as the subject, capturing it at eye level. We asked the VQA: *“Is the image taken from a high, low, or level angle?”*.

**Dominant Colors:** Colors are crucial for establishing the context of an image. To identify dominant colors in the images and understand the users’ environments, we asked the VQA: *“What is the dominant color of the image?”*.

**Lighting Condition:** Lighting conditions in an image reveal important information about the user’s ambient environment. Using the VQA model, we classified images based on their lighting as well-lit, dimly lit, or poorly lit. We asked the VQA: *“Is the image well-lit, dimly lit, or poorly lit?”*.

**Photo Location:** The location context (indoors or outdoors) can significantly influence user-device interactions. We determined the location context of images with the help of the VQA model by asking: *“Is the photo taken indoors or outdoors?”*.

**Background Objects:** Identifying specific objects in the background can provide valuable information about the user’s context and activities. We queried the VQA model about the background objects to recognize and categorize various elements within the images. We asked the VQA: *“What are the background objects in the photo?”*.

**Number of People in the Image:** In order to evaluate the social context of the images, we employed the VQA model to determine the number of people present in each image. This information provides insight into users’ social interactions and their surroundings during device usage. We asked: *“How many people are in the image?”*.

By leveraging the BLIP VQA model, we are able to extract structured insights about the content and context of in-the-wild images, enhancing our understanding of user behavior and interaction with their devices in diverse settings.

### 3.5 Depression Detection

In this study, we aim to accurately identify depression from facial images by utilizing both machine learning and deep learning techniques. Prior to training, we divide our dataset into three distinct subsets: a training set with 127 participants, a validation set with 15 participants, and a hold-out test set comprising 35 participants. It is important to note that the hold-out test set is not employed for training purposes or hyperparameter tuning. Given that most of our participants get categorized into both the depressed and non-depressed groups based on their EMA responses over a 90-day period, we adopt the Leave Subjects Out approach. This method ensures that all images associated with a single participant are exclusively used for either training, validation, or testing the model, but not mixed among the subsets. As a result, our model’s performance is highly robust.

**3.5.1 Machine Learning.** To facilitate machine learning approaches, we extract 711 facial features using OpenFace [4], a well-validated feature set for depression detection that has been employed in a variety of studies [18, 35, 40]. The extracted features consists of 2D and 3D facial landmarks, head pose, eye gaze, facial expressions represented by action units, and rigid and non-rigid shape parameters. Consequently, we train three machine learning models to predict whether a facial image belongs to the depressed or non-depressed class using OpenFace features. The methods we utilize include Logistic Regression (LR) [21], Random Forest (RF) [7], and Extreme Gradient Boosting (XGB) [9]. To gain

Under Review at CHI 2024.

valuable insights into the effectiveness of different hand-crafted feature sets in real-world settings, we also conduct an ablation study (Section 4).

**3.5.2 Deep Learning.** Deep learning models are capable of learning useful features directly from raw images. Pre-trained computer vision models trained on large-scale datasets can capture image features that are transferable to other domains. As a result, we examine the performance of various Efficient Net [43] and InceptionResNetv3 variants, which were previously trained on the ImageNet and VGGFace2 datasets, respectively. Upon observing that the EfficientNet B0 (EffNet) model provided the best performance while other models were underfitting our dataset, we decided to further fine-tune EffNet for depression prediction. We implement EffNet using the PyTorch framework, freezing all layers during the training process except for blocks 6 and 7. The model is fine-tuned using binary cross-entropy loss, the Adam optimizer (with a learning rate of 0.0001), a batch size of 256, and for a total of 50 epochs. This fine-tuning process allows the model to learn and adapt to the specific characteristics of our depression detection dataset, potentially improving its performance and generalizability.

## 4 RESULTS

## 4.1 Image Characteristics

Our analysis using the VQA model reveal many insights into diverse features of real-world smartphone images. These images serve as glimpses into user interactions and surroundings. In terms of capture angle, the images predominantly favored a low angle, with approximately 96.08% falling into this category. Conversely, a mere 3.92% were captured from a high angle, suggesting a specific user posture or device interaction habit in the majority of instances. Dissecting the dominant colors present, we found that ‘white’ emerged as the prevailing color, characterizing roughly 67.51% of the dataset. Other noticeable colors included ‘black’ at 8.70%, while a combined representation of ‘brown’, ‘blue’, ‘gray’, and ‘yellow’ accounted for approximately 18%. A diverse array of other hues constituted the remaining 5.75%, emphasizing the richness of user environments. The lighting conditions under which these images were taken were also revealing. A

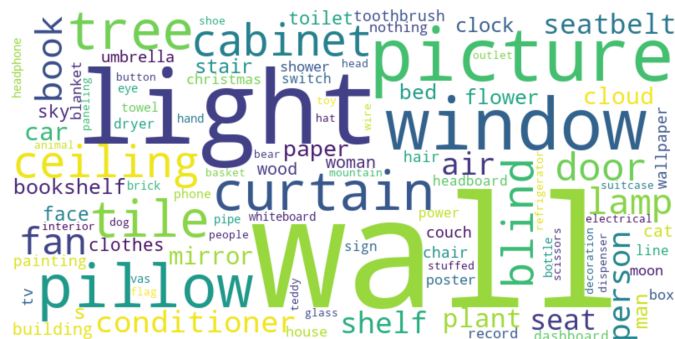











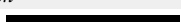







Fig. 4. Background objects: Word Cloud showing the range of objects detected in the background of the images captured.

vast majority (80.63%) were captured under well-lit conditions, indicating optimal settings for smartphone interaction. The dimly lit and poorly lit categories followed with 10.36% and 9.01%, respectively, showcasing the varied ambient conditions in which users interact with their devices. Furthermore, in terms of photo location, an impressive 95.08% of the images were taken indoors, signifying the primary environment for user-device interaction. The outdoor segment,

constituting 4.92%, provided insight into the more dynamic and mobile interactions users might experience. Notably, 95.81% of the captured images featured only one person. Regarding background objects, we discovered that walls, lights, pictures, and windows were the most common elements. The presence of terms such as "pillow" could imply individuals reclining, while words like "plant," "moon," "flower," and "cloud" might suggest outdoor settings. Overall, it appears that a significant number of images were captured indoors against plain backdrops, possibly within homes or offices. To visually represent these background objects, we have created a word cloud, which can be seen in Figure 4.

Table 2. Image Characteristics: Different characteristics of the image captured, such as image angle, dominant colors, lighting conditions, photo location and number of people present.

Characteristics	Count
<i>Image Angle</i>	
High angle 	105,949 (96.08%)
Low angle 	4,323 (3.92%)
<i>Dominant Colors</i>	
White 	744,14 (67.51%)
Black 	9,586 (8.70%)
Brown 	6,053 (5.49%)
Blue 	5,809 (5.27%)
Gray 	5,197 (4.72%)
Other 	9,213 (8.31%)
<i>Lighting Conditions</i>	
Well lit 	88,843 (80.57%)
Dimly lit 	11,418 (10.35%)
Poorly lit 	10,011 (9.08%)
<i>Photo Location</i>	
Indoors 	104,800 (95%)
Outdoors 	5,472 (5%)
<i>Number of People in the Image</i>	
One 	105,657 (95.81%)
Two 	523 (0.47%)
Three + 	8 (0.01%)
None 	4084 (3.71%)

## 4.2 Predictive Analysis

Investigating depression detection using machine learning and deep learning methods enables us to assess the capabilities of MoodCapture in naturalistic conditions comprehensively. From Table 3, we observe that EffNet, a deep learning model, outperforms traditional machine learning models, obtaining a balanced accuracy of 0.62 and an F1-score of 0.75. These results suggest that the ability of deep learning methods to automatically learn useful features from raw data is crucial for enhancing detection performance in this context. Furthermore, EffNet demonstrates superior performance in identifying positive depression cases while maintaining precision-recall balance (precision=0.73 and recall=0.77). Achieving such a balance is essential in real-world applications where both false positives and false negatives have significant implications. Note that balanced accuracy thoroughly evaluates a model's performance on both ends: it captures the true positive rate (how well the model detects actual cases of depression) and the true negative rate (how well the model identifies non-depressed cases). In contrast, the F1-score predominantly centers its assessment on the harmonic mean of precision (how many identified as depressed truly are) and recall (how many actual depression

**Under Review at CHI 2024.**

cases the model catches). Attaining a balanced accuracy level of 0.62 along with a F1-score of 0.75 at the same time is a clear indicator of the model’s capability. It signifies that our model can discern between depressed and non-depressed individuals with a good degree of accuracy, a characteristic that becomes paramount in practical, real-world applications.

Table 3. Performance: Depression detection using machine learning and deep learning methods

Method	Balanced Accuracy	F1-score	Recall	Precision
LR	0.56	0.65	0.61	0.71
RF	0.57	0.74	0.73	<b>0.75</b>
XGB	0.50	0.70	0.75	0.65
EffNet	<b>0.62</b>	<b>0.75</b>	<b>0.77</b>	0.73

Among traditional machine learning methods, we observe that the Random Forest (RF) model performs the best, achieving a balanced accuracy of 0.57 and an F1-score of 0.74. While RF’s ability to detect positive classes is on par with EffNet, the ability to distinguish classes is lower. The performance of RF highlights the importance of considering ensemble methods, which combine multiple decision trees to improve generalization and reduce overfitting. Logistic Regression (LR) also performs reasonably well, with a balanced accuracy of 0.56. This finding suggests that linear classifiers can be useful in larger datasets and should not be overlooked when dealing with complex problems. The performance of LR emphasizes the relevance of simpler models, which can provide interpretable results and require less computational resources.

Notably, from Table 3, it is evident that RF performs considerably better than Extreme Gradient Boosting (XGB), indicating that tree bagging is superior to boosting in our case. Understanding the strengths and weaknesses of various methods can help researchers and practitioners make informed decisions when designing depression detection systems. In conclusion, our investigation into depression detection using machine learning and deep learning methods provides valuable insights into the performance and potential of different techniques when applied to MoodCapture data in naturalistic conditions. The results emphasize the importance of considering a range of methods, from deep learning models capable of learning complex features to traditional machine learning techniques that offer interpretability and simplicity. By carefully selecting and fine-tuning these models, we can improve the overall performance and applicability of depression detection systems in real-world scenarios.

### 4.3 Ablation Study

In this analysis, we aimed to determine if specific OpenFace feature sets are more useful for depression detection by evaluating the performance across the seven groups (Facial action units, Gaze, Eye landmarks, Pose, Rigidity Parameters, 2D and 3D landmarks). From Table 4, we make several interesting observations that provide insights into the utility of individual feature sets.

First, we notice that many feature sets perform better than the whole, indicating that only some specific features in the image are useful for depression detection. This finding suggests that a more focused approach to feature extraction and selection may improve overall performance. Second, we observe that facial action units are less discriminative than other features. This result may be attributed to the presence of partial face images, which are common in front-facing cameras, thus hindering the effectiveness of action units in detecting depression. Third, we find that eye landmarks outperform gaze features, suggesting that minor eye expressions are more beneficial for depression detection compared

**Under Review at CHI 2024.**

to the direction in which the user is looking. This observation highlights the importance of capturing subtle facial changes when developing depression detection systems.

Table 4. Ablation Study: Investigation depression detection of OpenFace feature sets using a random forest

Feature Set	Balanced Accuracy	F1-score	Recall	Precision
Facial Action Units	0.57	0.68	0.65	0.71
Gaze	0.56	0.68	0.66	0.70
Eye Landmarks	0.61	0.71	0.68	0.74
Pose	0.58	0.71	0.71	0.71
Rigidity Parameters	<b>0.66</b>	<b>0.74</b>	0.72	<b>0.76</b>
2D Landmarks	0.58	0.65	0.56	0.73
3D Landmarks	0.58	0.72	<b>0.74</b>	0.71

Rigidity shape parameters emerged as the most useful feature set, obtaining a balanced accuracy of 0.66 and an F1-score of 0.74. Although rigidity parameters outperform our deep learning method in discriminative power (balanced accuracy 0.66 vs. 0.62), their predictive power on the positive class is slightly lower (F1-score 0.74 vs. 0.75). Furthermore, they achieve a good precision-recall balance with values of 0.72 and 0.76, respectively. These parameters capture the placement of the face in the image (rigid: e.g., scaling, translation, rotation) and describe facial expressions and deformations (non-rigid: e.g., expression, wider or taller faces). By comparing 2D and 3D landmarks from Table 4, we observe that 3D features have better predictive power on the positive class, suggesting that depth features are important for depression detection. This finding emphasizes the potential benefits of incorporating 3D facial information when developing and refining depression detection systems.

In conclusion, the ablation study provides valuable insights into the utility of specific feature sets for depression detection. By understanding the strengths and limitations of individual features, researchers and practitioners can make informed decisions when designing and implementing depression detection systems, ultimately improving overall performance and applicability in real-world scenarios.

## 5 ETHICAL CONSIDERATIONS AND USER ACCEPTANCE

In studies involving sensitive mental health data, it is paramount to address the ethical implications to safeguard participants' privacy, confidentiality, and well-being. Our primary goal was to prioritize the security and confidentiality of the data. We securely stored all collected data and granted access only to specific team members. We took great care in removing all personally identifiable information by implementing a thorough anonymization process. To respect privacy, any image that unintentionally captured subjects or nudity was identified during a review by two team members and subsequently deleted. We understand the sensitive nature of mental health and made sure to maintain transparency with our participants. They were informed about the study's purpose, methodology, and expected outcomes. This approach not only sought their permission but also ensured they felt comfortable and safe throughout the process. We further clarified that their compensation was unrelated to their photos.

At the end of the study, we asked participants about their comfort levels with automated front-facing photo capture during surveys. This was optional, so we have responses from only 172 out of the 181 participants that were recruited. Approximately 45% of participants were comfortable, while 38% felt it was intrusive or uneasy, and the remaining 17% were neutral. Participants in the study expressed various concerns regarding the photo bursts, which can be summarized

**Under Review at CHI 2024.**

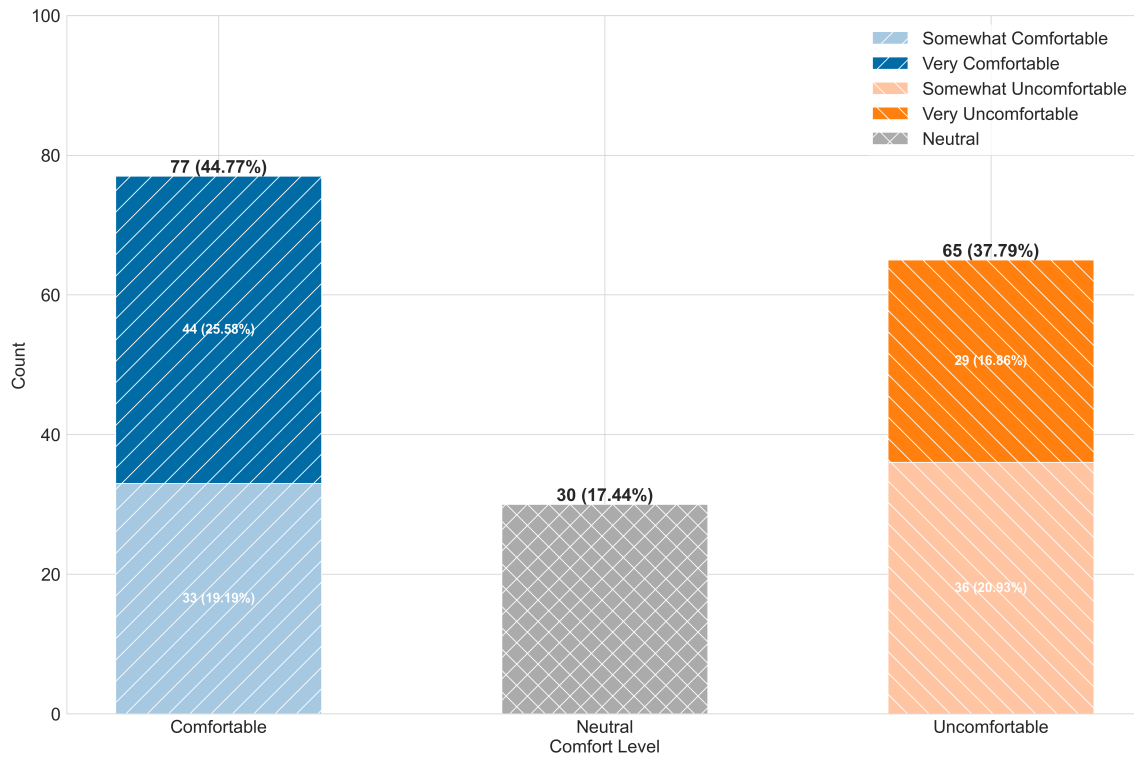


Fig. 5. Comfort Level: Participant's comfort with the automated capture of their photos.

into a few key themes. While we acknowledge these concerns, it is important to note that the study followed strict privacy and data protection guidelines.

- (1) **Privacy and Surveillance:** Participants felt uncomfortable with the idea of being watched or monitored, as it evoked a sense of intrusion into their personal space. One participant mentioned, *"I don't like being watched. I'm already paranoid when it comes to cameras."*
- (2) **Appearance and Self-Esteem:** Several participants mentioned their discomfort with having their photos taken due to concerns about their appearance. One participant stated, *"I don't want people to see photos of me"*, while another said, *"I am very uncomfortable with my appearance when I'm depressed."*
- (3) **Inappropriate Situations:** Participants worried about the possibility of photo bursts being taken during inconvenient or inappropriate moments. One participant shared, *"If I was comfortable and at home, during some of them I may not have been completely covered."*
- (4) **Data Security:** Although participants were aware of the study's data protection measures, some still expressed concerns about the safety and storage of their images. One participant expressed, *"The idea of my picture being out there...although I know it was to be analyzed with AI."*
- (5) **Lack of Control:** Participants felt uneasy about not being able to review, approve, or delete the photos taken during the bursts, as well as not knowing when the camera was active. A participant shared, *"Having pictures*

**Under Review at CHI 2024.**

*taken and not knowing what they looked like or if they were embarrassing is an uncomfortable thing to think about.”*

In summary, participants’ concerns mainly revolved around privacy, self-esteem, potential inappropriate situations, data security, and control over the images. It is essential to consider these concerns when designing and implementing studies involving photo bursts or similar data collection methods to ensure participants’ comfort and trust in the research process. Acknowledging the sensitive nature of our research, we offered participants the option to delete their photos at the end of the study if they felt uncomfortable. Interestingly, no participants chose this option, highlighting the trust they placed in our research process and commitment to ethical conduct. We remain keenly aware of the potential for technology misuse, especially in unauthorized surveillance or data mining scenarios. We have taken measures to minimize such risks, emphasizing that our technological developments are primarily intended as health aids, not tools for unwarranted monitoring. Further, to address participants’ concerns regarding privacy and data security, one possible solution could be leveraging the capabilities of AI chips on smartphones. By conducting all image classification and processing on the device itself, no images would need to be transmitted or stored externally. This approach could significantly alleviate users’ concerns about their images being stored or accessed by unauthorized parties. As AI technology continues to advance, incorporating on-device processing capabilities into our research methodology may not only increase user trust and comfort but also pave the way for a new generation of privacy-focused health aids. In line with our commitment to ethical conduct, we will continue to explore and implement such technological advancements to ensure the protection of participants’ data and privacy in our research.

## **6 DISCUSSION**

### **6.1 Summary of results**

Our study investigated the potential of using in-the-wild smartphone images and deep learning models for detecting depression, aiming to contribute to the development of user-centered and unobtrusive mental health assessment tools. The results of our analysis provided valuable insights into the characteristics of in-the-wild images, the performance of machine learning and deep learning models in depression detection, and user acceptance of such approaches.

The image characteristics analysis revealed that most images were captured from a low angle, indoors, and under well-lit conditions. These findings highlighted the participants’ natural behavior with their smartphones, emphasizing the importance of considering real-world HCI dynamics in designing mental health assessment tools. The analysis also showed a diverse range of dominant colors and background objects, further reflecting the richness of user environments. Our result demonstrates that the EffNet deep learning model significantly outperforms traditional machine learning models in detecting depression from facial images, achieving a balanced accuracy of 0.62 and an F1-score of 0.75. This outcome highlights the effectiveness of deep learning models, which have the ability to automatically learn valuable features from raw data, particularly when analyzing facial images captured in uncontrolled, real-world settings. The promising performance scores obtained in this study are even more noteworthy considering that the facial images were captured using a diverse range of smartphone devices – 87 different models from 9 distinct brands. As the camera quality of these devices varies significantly, it is important to note that the results may be influenced by factors such as image clarity and auto-focus capabilities. Despite these potential limitations, our findings support the ecological validity of the study and emphasize the potential of deep learning models like EffNet in accurately detecting depression from facial images, even when captured in less-than-ideal conditions. Moreover, our ablation study demonstrated the importance of specific OpenFace feature sets, such as rigidity shape parameters and 3D landmarks, in depression

**Under Review at CHI 2024.**



detection. By focusing on these features, researchers can potentially improve the overall performance of mental health assessment tools. In terms of user acceptance, we found diverse responses regarding participants' comfort levels with automated front-facing photo capture. While some participants were comfortable with the process, others felt uneasy due to concerns related to privacy, self-esteem, inappropriate situations, data security, and control over the images. These concerns highlight the need for careful consideration of ethical implications in designing and implementing studies involving photo bursts or similar data collection methods.

In conclusion, our research highlights the potential of using in-the-wild smartphone images and deep learning models for depression detection, offering a more objective, unobtrusive, and continuous approach to mental health assessment. By carefully considering the insights gained from our analysis and addressing the ethical implications, researchers and practitioners can work towards developing user-centered, effective, and ethically sound tools for mental health assessment and intervention.

## 6.2 Implications

The findings from our study hold significant implications for various stakeholders, including researchers, practitioners, and policymakers in the fields of mental health, digital health, human-computer interaction (HCI), and public health.

Our research highlights the potential of utilizing smartphone images and deep learning models as a supplementary method for mental health assessment. This innovative approach encourages the exploration of alternative ways to assess mental health that can complement traditional tools such as self-report questionnaires and clinical interviews. While our data were collected from participants who had major depressive disorder, the results pave the way for future research to investigate the broader applicability of these methods, potentially leading to a better understanding of depression and improved mental health support over time. Consequently, promoting timely access to appropriate interventions and support systems. From an HCI perspective, the study underscores the importance of considering user acceptance and user-centered design when developing mental health assessment tools that utilize smartphone images and deep learning models. Understanding users' concerns and preferences is crucial for creating tools that are more likely to be adopted and used by those in need of support. This focus on user acceptance can inspire the HCI community to design mental health assessment tools that balance effectiveness, privacy, and user engagement, leading to the development of more accessible and inclusive digital mental health solutions.

In the broader context of public health, the study's findings emphasize the importance of leveraging technology and innovative methods to address mental health challenges. As mental health disorders continue to impact individuals and communities worldwide, adopting novel approaches like the one presented in our study can contribute to more effective prevention strategies, early intervention, and resource allocation. This could ultimately lead to better mental health outcomes and overall well-being for individuals across various demographic and cultural contexts. In summary, the implications of our study extend well beyond the immediate findings, offering valuable insights for a range of stakeholders working at the intersection of mental health, digital health, and human-computer interaction. By considering user acceptance, exploring the potential of smartphone images for mental health assessment, and recognizing the broader public health context, our study contributes to the development of more effective, user-friendly, and contextually appropriate mental health assessment tools with the potential to improve the lives of individuals affected by depression.

## 7 LIMITATIONS

Our study while providing valuable insights into the use of in-the-wild smartphone images and deep learning models for depression detection, has some limitations that should be acknowledged. First, our study's dataset may be limited in size

**Under Review at CHI 2024.**

and diversity, as it consists of a relatively small number of participants. A larger and more diverse sample would provide a more robust representation of the general population and enhance the generalizability of the findings. Furthermore, the study relies on self-reported data, such as depression scores, which may be subject to biases, including social desirability and recall bias. Future research could be significantly enhanced by including more objective measures of mental health, such as clinical evaluations or physiological indicators. It is important to highlight that all participants in our study had received clinical diagnoses for MDD. However, we relied on self-reported data for tracking daily depression levels, which facilitated more consistent monitoring. Our study also focused exclusively on a clinically depressed cohort. Including healthy individuals in the dataset would have been beneficial for developing a more comprehensive and accurate prediction model. A randomized controlled trial (RCT) with healthy controls or incorporating a diverse cohort of individuals not experiencing depression could provide valuable insights into the differences between depressed and non-depressed individuals and improve the model's ability to distinguish between them. Future research should consider expanding the dataset to include both depressed and healthy individuals, which can contribute to the development of more effective and precise mental health assessment tools.

Another limitation is that the study primarily focuses on the analysis of in-the-wild smartphone images and their relationship with depression. However, there may be other factors, such as social interactions, physical activity, and environmental context, that could provide additional insights into depression detection. Integrating these factors into future research may help to develop more holistic and accurate prediction models. Deep learning models, while powerful and effective, can often be considered as "black-box" models with limited interpretability. This may make it difficult to understand the specific features or patterns that the model has identified as being related to depression. Future research could explore the use of more interpretable models or techniques to provide insights into the underlying mechanisms linking visual cues and depression. Lastly, the use of in-the-wild smartphone images for mental health assessment raises ethical and privacy concerns, which need to be carefully considered when designing and implementing such tools. Ensuring user consent, data security, and transparency in the use of personal data is crucial for maintaining trust and fostering the adoption of these tools. Addressing these limitations in future research can help to further advance our understanding of the relationship between smartphone images, deep learning models, and depression detection, contributing to the development of more effective, user-centered, and ethically sound mental health assessment tools.

## 8 CONCLUSION AND FUTURE WORK

Through this study, we have demonstrated the potential of using in-the-wild smartphone images and deep learning models to detect depression, offering valuable insights for mental health assessment, HCI and digital health. With this, we aim to pave the way for more effective and user-centered mental health assessment tools. Addressing the limitations of our study and building upon its findings, future research can contribute to the development of more robust, accurate, and ethically sound mental health assessment tools that have the potential to improve the lives of individuals affected by depression.

When we embarked on designing our MoodCapture study to investigate whether high-resolution face capture from phones could assess mood, we were acutely aware of the ethical issues surrounding our research and the potential privacy concerns of a population that included individuals diagnosed with depression. As discussed in the section on Ethical Considerations and User Acceptance, our study was meticulously designed to safeguard user privacy throughout, and we sought their evaluations of the MoodCapture app post-study. This invaluable feedback forms the foundation for future work in image-based mood detection which we believe is a promising technology. One direction we plan to pursue as our next step involves utilizing on-phone AI chips that are now available on top-end smartphones to run

**Under Review at CHI 2024.**

deep learning models directly on the device, ensuring that images never leave the phone. Additionally, we intend to explore the combination of this on-device prediction approach with federated deep learning, where models are trained without sharing raw data across a network in a central entity such as a server or cloud. This approach could effectively address security concerns associated with centralized data collection and the privacy issues our participants raised during the acceptance study.

## REFERENCES

- [1] 2019. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. *BMJ* (April 2019), l1781. <https://doi.org/10.1136/bmj.l1781>
- [2] Awuni Prosper Mandela Amaltinga and James Fenibe Mbinta. 2020. Factors associated with depression among young people globally: a narrative review. *International Journal Of Community Medicine And Public Health* 7, 9 (Aug. 2020), 3711. <https://doi.org/10.18203/2394-6040.ijcmph20203949>
- [3] Mihai Băce, Sander Staal, and Andreas Bulling. 2020. Quantification of users' visual attention during everyday mobile device interactions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [4] Tadas Baltrušaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. IEEE, 59–66.
- [5] Aaron T Beck, Robert A Steer, Gregory K Brown, et al. 1987. *Beck depression inventory*. Harcourt Brace Jovanovich New York.
- [6] José Manoel Bertolote, Alexandra Fleischmann, Diego De Leo, and Danuta Wasserman. 2003. Suicide and mental disorders: do we know enough? *British Journal of Psychiatry* 183, 5 (Nov. 2003), 382–383. <https://doi.org/10.1192/bjp.183.5.382>
- [7] Leo Breiman. 2001. Random forests. *Machine learning* 45 (2001), 5–32.
- [8] Stevie Chancellor and Munmun De Choudhury. 2020. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine* 3, 1 (2020), 43.
- [9] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785–794.
- [10] Prerna Chikersal, Afsaneh Doryab, Michael Tumminia, Daniella K Villalba, Janine M Dutcher, Xinwen Liu, Sheldon Cohen, Kasey G Creswell, Jennifer Mankoff, J David Creswell, et al. 2021. Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing: a machine learning approach with robust feature selection. *ACM Transactions on Computer-Human Interaction (TOCHI)* 28, 1 (2021), 1–41.
- [11] Victor-Alexandru Darvari, Laura Convertino, Abhinav Mehrotra, and Mirco Musolesi. 2020. Quantifying the relationships between everyday objects and emotional states through deep learning based image analysis using smartphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–21.
- [12] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Proceedings of the international AAAI conference on web and social media*, Vol. 7. 128–137.
- [13] Jorge Arias de la Torre, Gemma Vilagut, Antoni Serrano-Blanco, Vicente Martín, Antonio José Molina, Jose M Valderas, and Jordi Alonso. 2020. Accuracy of Self-Reported Items for the Screening of Depression in the General Population. *International Journal of Environmental Research and Public Health* 17, 21 (Oct. 2020), 7955. <https://doi.org/10.3390/ijerph17217955>
- [14] M Deady, D A J Collins, D A Johnston, N Glozier, R A Calvo, H Christensen, and S B Harvey. 2021. The impact of depression, anxiety and comorbidity on occupational outcomes. *Occupational Medicine* 72, 1 (Oct. 2021), 17–24. <https://doi.org/10.1093/occmed/kqab142>
- [15] David M Fergusson and Lianne J Woodward. 2002. Mental health, educational, and social role outcomes of adolescents with depression. *Arch. Gen. Psychiatry* 59, 3 (March 2002), 225–231.
- [16] Ralph R. Frerichs, Carol S. Aneshensel, Patricia A. Yokopenic, and Virginia A. Clark. 1982. Physical health and depression: An epidemiologic survey. *Preventive Medicine* 11, 6 (Nov. 1982), 639–646. [https://doi.org/10.1016/0091-7435\(82\)90026-3](https://doi.org/10.1016/0091-7435(82)90026-3)
- [17] Venkata Rama Kiran Garimella, Abdulrahman Alfayad, and Ingmar Weber. 2016. Social media image analysis for public health. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 5543–5547.
- [18] Yuan Gong and Christian Poellabauer. 2017. Topic modeling based multi-modal depression detection. In *Proceedings of the 7th annual workshop on Audio/Visual emotion challenge*. 69–76.
- [19] Sharath Chandra Guntuku, Daniel Preotiuc-Pietro, Johannes C Eichstaedt, and Lyle H Ungar. 2019. What twitter profile and posted images reveal about depression and anxiety. In *Proceedings of the international AAAI conference on web and social media*, Vol. 13. 236–246.
- [20] Weitong Guo, Hongwu Yang, Zhenyu Liu, Yaping Xu, and Bin Hu. 2021. Deep Neural Networks for Depression Recognition Based on 2D and 3D Facial Expressions Under Emotional Stimulus Tasks. *Frontiers in Neuroscience* 15 (April 2021). <https://doi.org/10.3389/fnins.2021.609760>
- [21] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. 2013. *Applied logistic regression*. Vol. 398. John Wiley & Sons.
- [22] Melissa Hunt, Joseph Auriemma, and Ashara C. A. Cashaw. 2003. Self-Report Bias and Underreporting of Depression on the BDI-II. *Journal of Personality Assessment* 80, 1 (Feb. 2003), 26–30. [https://doi.org/10.1207/s15327752jpa8001\\_10](https://doi.org/10.1207/s15327752jpa8001_10)
- [23] Manju Lata Joshi and Nehal Kanoongo. 2022. Depression detection using emotional artificial intelligence and machine learning: A closer review. *Materials Today: Proceedings* 58 (2022), 217–226.

**Under Review at CHI 2024.**

- [24] Mohamed Khamis, Anita Baier, Niels Henze, Florian Alt, and Andreas Bulling. 2018. Understanding Face and Eye Visibility in Front-Facing Cameras of Smartphones used in the Wild. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. <https://doi.org/10.1145/3173574.3173854>
- [25] Kenneth A. Kobak. 2010. Hamilton Depression Rating Scale. , 1 pages. <https://doi.org/10.1002/9780470479216.corpsy0402>
- [26] Xinru Kong, Yan Yao, Cuiying Wang, Yuangeng Wang, Jing Teng, and Xianghua Qi. 2022. Automatic Identification of Depression Using Facial Images with Deep Convolutional Neural Network. *Medical Science Monitor* 28 (June 2022). <https://doi.org/10.12659/msm.936409>
- [27] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. 2001. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine* 16, 9 (2001), 606–613.
- [28] Young-Shin Lee and Won-Hyung Park. 2022. Diagnosis of Depressive Disorder Model on Facial Expression Based on Fast R-CNN. *Diagnostics* 12, 2 (Jan. 2022), 317. <https://doi.org/10.3390/diagnostics12020317>
- [29] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. 2022. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International Conference on Machine Learning*. PMLR, 12888–12900.
- [30] Dongdong Liu, Bowen Liu, Tao Lin, Guangya Liu, Guoyu Yang, Dezhen Qi, Ye Qiu, Yuer Lu, Qinmei Yuan, Stella C. Shuai, Xiang Li, Ou Liu, Xiangdong Tang, Jianwei Shuai, Yuping Cao, and Hai Lin. 2022. Measuring depression severity based on facial expression and body movement using deep convolutional neural network. *Frontiers in Psychiatry* 13 (Dec. 2022). <https://doi.org/10.3389/fpsyt.2022.1017064>
- [31] Wafa Mellouk and Wahida Handouzi. 2020. Facial emotion recognition using deep learning: review and insights. *Procedia Computer Science* 175 (2020), 689–694. <https://doi.org/10.1016/j.procs.2020.07.101>
- [32] James S Olver and Malcolm J Hopwood. 2013. Depression and physical illness. *Medical Journal of Australia* 199, S6 (Oct. 2013). <https://doi.org/10.5694/mja12.10597>
- [33] World Health Organization. 2023. Mental disorders. <https://www.who.int/news-room/fact-sheets/detail/mental-disorders> Accessed: [2023].
- [34] A. Pampouchidou, K. Marias, M. Tsiknakis, P. Simos, F. Yang, and F. Meriaudeau. 2015. Designing a framework for assisting depression severity assessment from facial image analysis. In *2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. 578–583. <https://doi.org/10.1109/ICSIPA.2015.7412257>
- [35] Anastasia Pampouchidou, Olympia Simantiraki, C-M Vazakopoulou, Charikleia Chatzaki, Matthew Pedititis, Anna Maridaki, Kostas Marias, Panagiotis Simos, Fan Yang, Fabrice Meriaudeau, et al. 2017. Facial geometry and speech analysis for depression detection. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 1433–1436.
- [36] Alexander Ramos-Cuadros, Luis Palomino Santillan, and Willy Ugarte. 2021. Evaluating the Depression Level Based on Facial Image Analyzing and Patient Voice. In *International Conference on Information and Communication Technologies for Ageing Well and e-Health*. Springer, 35–55.
- [37] Andrew G Reece and Christopher M Danforth. 2017. Instagram photos reveal predictive markers of depression. *EPJ Data Science* 6, 1 (Aug. 2017). <https://doi.org/10.1140/epjds/s13688-017-0110-z>
- [38] Jonathan Rottenberg, James J. Gross, and Ian H. Gotlib. 2005. Emotion Context Insensitivity in Major Depressive Disorder. *Journal of Abnormal Psychology* 114, 4 (Nov. 2005), 627–639. <https://doi.org/10.1037/0021-843x.114.4.627>
- [39] Samuli I Saarni, Jaana Suvisaari, Harri Sintonen, Sami Pirkola, Seppo Koskinen, Arpo Aromaa, and Jouko Lönnqvist. 2007. Impact of psychiatric disorders on health-related quality of life: general population survey. *Br. J. Psychiatry* 190 (April 2007), 326–332.
- [40] Guramritpal Singh Saggu, Keshav Gupta, KV Arya, and Ciro Rodriguez Rodriguez. 2022. DepressNet: A Multimodal Hierarchical Attention Mechanism approach for Depression Detection. *Int. J. Eng. Sci.* 15, 1 (2022), 24–32.
- [41] Ziggi Ivan Santini, Ai Koyanagi, Stefanos Tyrovolas, Catherine Mason, and Josep Maria Haro. 2015. The association between social relationships and depression: A systematic review. *Journal of Affective Disorders* 175 (April 2015), 53–65. <https://doi.org/10.1016/j.jad.2014.12.049>
- [42] Georg Schomerus, Charlotte Auer, Dieter Rhode, Melanie Luppa, Harald J. Freyberger, and Silke Schmidt. 2012. Personal stigma, problem appraisal and perceived need for professional help in currently untreated depressed persons. *Journal of Affective Disorders* 139, 1 (June 2012), 94–97. <https://doi.org/10.1016/j.jad.2012.02.022>
- [43] Mingxing Tan and Quoc Le. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*. PMLR, 6105–6114.
- [44] Rui Wang, Andrew T. Campbell, and Xia Zhou. 2015. Using Opportunistic Face Logging from Smartphone to Infer Mental Health: Challenges and Future Directions. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers* (Osaka, Japan) (*UbiComp/ISWC'15 Adjunct*). Association for Computing Machinery, New York, NY, USA, 683–692. <https://doi.org/10.1145/2800835.2804391>
- [45] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. 3–14.
- [46] Rui Wang, Weichen Wang, Alex DaSilva, Jeremy F Huckins, William M Kelley, Todd F Heatherton, and Andrew T Campbell. 2018. Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–26.
- [47] Xuhai Xu, Prerna Chikersal, Afsaneh Doryab, Daniela K Villalba, Janine M Dutcher, Michael J Tumminia, Tim Althoff, Sheldon Cohen, Kasey G Creswell, J David Creswell, et al. 2019. Leveraging routine behavior and contextually-filtered features for depression detection among college students. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–33.

**Under Review at CHI 2024.**

- [48] Xiuzhuang Zhou, Kai Jin, Yuanyuan Shang, and Guodong Guo. 2018. Visually interpretable representation learning for depression recognition from facial images. *IEEE transactions on affective computing* 11, 3 (2018), 542–552.