EmoGlass: an End-to-End AI-Enabled Wearable Platform for Enhancing Self-Awareness of Emotional Health

Zihan Yan
Zhejiang University
Hangzhou, China
zihanyan@zju.edu.cn

Yufei Wu
Zhejiang University
Hangzhou, China
yf.wu@zju.edu.cn

Yang Zhang
University of California, Los Angeles
Los Angeles, United States
yangzhang@ucla.edu

Xiang 'Anthony' Chen
University of California, Los Angeles
Los Angeles, United States
xac@ucla.edu

ABSTRACT

Often, emotional disorders are overlooked due to their lack of awareness, resulting in potential mental issues. Recent advances in sensing and inference technology provide a viable path to wearable facial-expression-based emotion recognition. However, most prior work has explored only laboratory settings and few platforms are geared towards end-users in everyday lives or provide personalized emotional suggestions to promote self-regulation. We present EmoGlass, an end-to-end wearable platform that consists of emotion detection glasses and an accompanying mobile application. Our single-camera-mounted glasses can detect seven facial expressions based on partial face images. We conducted a three-day out-of-lab study (N=15) to evaluate the performance of EmoGlass. We iterated on the design of the EmoGlass application for effective self-monitoring and awareness of users' daily emotional states. We report quantitative and qualitative findings, based on which we discuss design recommendations for future work on sensing and enhancing awareness of emotional health.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing**; **Ubiquitous and mobile computing systems and tools**.

KEYWORDS

Facial expression detection, Emotion sensing, Mental health, Mobile health, Wearable

ACM Reference Format:
Zihan Yan, Yufei Wu, Yang Zhang, and Xiang 'Anthony' Chen. 2022. EmoGlass: an End-to-End AI-Enabled Wearable Platform for Enhancing Self-Awareness of Emotional Health. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3491102.3501925

1 INTRODUCTION

Health psychology, especially psychosomatic/behavioral medicine, places a high priority on emotional health [14]. Negative emotions experienced over a long period of time may have severe implications [34, 59] and many emotional problems are left untreated until they become mental disorders, affecting nearly one billion people worldwide. [46]. Early detection and intervention of emotional disorders may have a significant impact on people's health and well-being [27]. However, people do not pay much attention to their emotions and the majority seldom seek help [67]. Moreover, many people cannot accurately perceive and report their emotions [10] and do not have the bandwidth to keep track of their own emotions around the clock.

Such a shortage of effective emotion detection and tracking mechanisms presents fertile research opportunities. One potential solution is to leverage wearable sensing that is becoming increasingly low-cost, ubiquitous, safe, and widely acceptable, thus providing a promising solution to detect, track and intervene in various health problems, ranging from physical health to mental health, including emotional health.

There are five channels through which emotions can be detected [29]: speech [33], text [49], facial expressions [1], body gestures/movements [6], and physiological states [68]. This paper focuses on and leverages facial expressions as direct indicators that have shown significant correlations with emotions and have been widely used in emotion-related diagnoses [23, 56, 96]. Prior work has explored various wearables supports for facial expression detection, including eyeglasses [57], earphones [18], and necklaces [17]. However, most existing systems are confined to laboratories, providing few insights on how well they would work in real-world settings [75]. Further, these systems did not explore interfaces to provide end-users with guidance and feedback in interpreting emotion detection results. Finally, prior work primarily focused on technical issues of facial expression recognition, with less attention on promoting emotional health based on wearable sensing of facial expression.

To fill these research gaps, we present EmoGlass (Figure 1), an end-to-end AI-enabled wearable platform consisting of custom-designed camera-equipped eyeglasses and a mobile application that can detect seven facial expressions, aiming to promote people's awareness of emotional health, enabling them to monitor their emotions, and providing personalized guidance for self-regulation.

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

ACM ISBN 978-1-4503-9157-3/22/04…$15.00
https://doi.org/10.1145/3491102.3501925

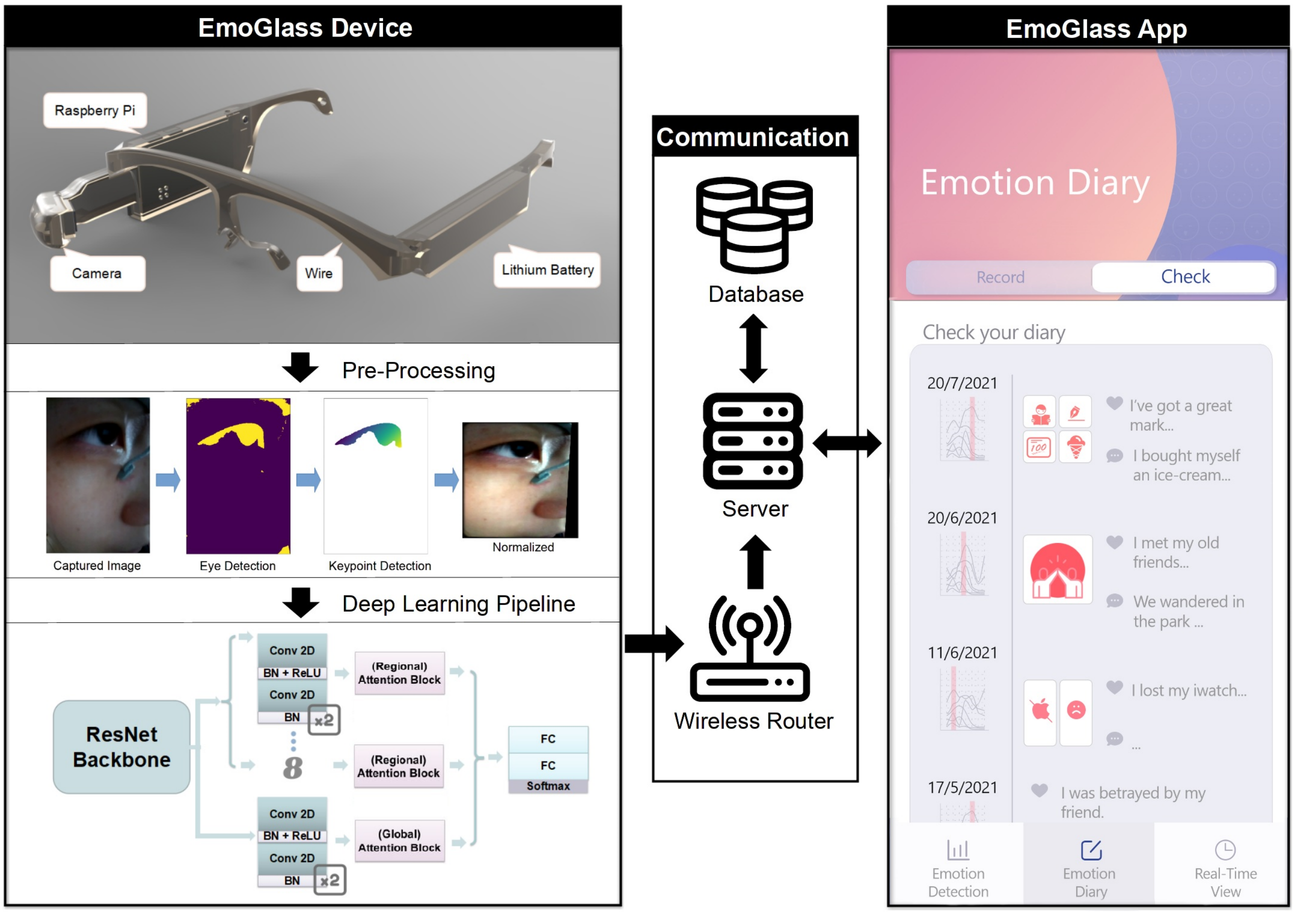

Figure 1: EmoGlass includes two major parts, EmoGlass device and EmoGlass mobile app. The camera-mounted and ACNN-embedded EmoGlass device is capable of recognizing facial expressions. The companion mobile app is used to enhance users' awareness of emotional health.

Specifically, we first built a wearable device in the form of a pair of glasses mounted with a camera and its supporting embedded system. To improve the robustness of our system in and out of the lab environments, we constructed three datasets from 15 participants, covering eyeglass remounting, various lighting conditions, as well as enhanced naturalness of facial expression. Based on these datasets, we developed per-user deep CNN models with attention mechanisms (ACNN) for facial expression recognition. We then developed a mobile application that allows users to self-monitor their emotion states tracked by the EmoGlass device, as aggregated and visualized on an hourly/daily/weekly basis. Users can also use the application to check and control when the camera is turned on/off and use an automatically-scheduled prompt to record their emotions and activities into an Emotion Diary, allowing them not only to review their emotional history but also to find suggestions in past activities to boost a positive emotion.

To validate the performance of the EmoGlass device and the first design iteration of EmoGlass application, we recruited the same 15 participants (because currently, EmoGlass' AI model is user-dependent) for a three-day out-of-lab study, where each participant wore EmoGlass at least three hours a day outside the lab. Participants were prompted every 15 minutes to report their emotions, which we use as ground truth labels. The overall accuracy of detecting seven facial expressions is 73.0%.

In the middle of the study, we reviewed feedback from the first seven participants and designed a second iteration of the EmoGlass application to address five key issues: 1) educating users about emotional health; 2) tracking events that trigger emotions and behaviors; 3) reminding/suggesting positive triggers to regulate emotions; 4) building users' trust of emotion detection; and 5) connecting emotional awareness and emotional health. We then deployed the EmoGlass application with the final design to the remaining eight participants and report qualitative findings, including user perceptions of wearable emotion-sensing, concerns of privacy, emotional awareness, and other key observations.

Overall, the system consists of:

- The EmoGlass system — an end-to-end AI-enabled wearable platform combining facial expression detection eyeglasses and a mobile application to enable self-monitoring of emotions to promote emotional awareness;
- Three datasets that iteratively extends in-the-lab controlled facial expression data with robustness for functioning outside controlled environments, including adding varied lighting conditions, device remounting, and naturally-occurred facial expressions;
- Iterative designs of the EmoGlass application that goes beyond visualizing detected emotions, with interactive features such as recording emotion-related activities and recommending past activities to help users regulate emotion;
- A three-day, 15-participant out-of-lab study to validate the EmoGlass system's technical performance and its feasibility in enabling end-users to self-monitor and regulate emotion.

The remaining paper is organized as follows (shown in Figure 2). Chapter 1 and 2 provide an overview of our study and related work. Chapter 3 introduces our hardware platform. The facial expression recognition (FER) model we used, describing the methods of data collection, model training and evaluation, and the performance of our model are described in Chapter 4. Chapter 5 describes the app in the first iteration, while Chapter 6 illustrates the out-of-lab study we conducted. Chapter 7 presents the initial feedback and corresponding design changes from the first round user study, along with overall findings and design recommendations. Chapter 8 includes discussions, limitations, and potential future work of our study, and Chapter 9 concludes our work.

2 RELATED WORK

EmoGlass is at the intersection of three related areas: smart eyewear for health sensing, emotion sensing & detection, and facial expression recognition on wearables.

2.1 Smart Eyewear for Health Sensing

The emergence of wearable devices and sensing technology has pushed traditional eHealth out of the clinical setting and has evolved into ubiquitous mHealth [36]. Eyewear comes in two main forms: eye masks and eyeglasses.

Health-related eye masks are often used for monitoring sleep [54] or critically ill patients. For example, A stroke prognostic tool called HealthSOS [37]. However, because eye masks will block users' view, eyeglasses present a greater potential for health sensing on a daily basis.

Eyeglasses trigger wide interests due to the unique wearing location, which can easily capture information of one's face and head [83], and its feasibility in the wild [15]. Smart eyeglasses have been employed in several application areas, including computer science, healthcare, education, industry, service, social science, and agriculture, among which, most attention has been focused on healthcare [45]. In addition to disease monitoring, it can also be used for monitoring neurological conditions due to its ability to capture images of pupils and neural signals. For example, Munusamy et al. demonstrated that telemedicine via smart eyeglasses is feasible and effective as an alternative to ward rounds for neurocritical ill patients [62]. Neuroglasses [80] have also been used for monitoring of neurodegenerative Parkinson's diseases [7]. Some researchers use eyeglasses to detect eye conditions, e.g., the MEMS-based (microelectromechanical systems) wearable eyeglasses, which have adequate sensitivity of retinal arteries and optic disc's observation for the detection of eye diseases [42, 43].

Outside of disease monitoring, a number of studies also focus on everyday health, most of which are related to diet and exercise. For example, Fitbyte — multimodal sensing eyewear — can track diet in unrestricted situations [13], which is also available in two other smart eyeglasses [35, 93]. On the other hand, exercise monitoring is a well-explored area with many mature commercial products [86]. For example, SOLOS [81] enables riders to obtain real-time data, including speed, cadence, heart rate and power zone. Similarly, Eyesight Raptor [26] can show information such as heart rate information. However, to the best of our knowledge, there has been little research on smart eyeglasses-based emotional health tracking that functions on a daily basis [57].

There are eyewear devices that support the measurement of mental health-related biomarkers, e.g., measuring the change in forehead and nose bridge temperatures in order to assess cognitive load [94]. EOG eyeglasses provide out-of-lab cognitive assessment function [19] and realize real-time emotion detection from single-eye images based on eyewear devices [88]. The EmotiGo eyewear system uses unobtrusive physiology-based emotion detection [74]. However, most of them target emotion sensing, with no further analysis of the identified emotions, such as temporal and statistical features, which can be important to emotional health applications.

2.2 Emotion Sensing & Detection

Accurately reporting emotion fluctuation is crucial for emotional health because psychologists found that identifying and labeling emotions is the starting point of emotion regulation [41]. Knowing what feelings we are going through can help us find suitable physical and behavioral responses to targeted emotions [12]. However, accurately understanding one's own emotion turns out to be challenging for many people [84], creating opportunities for wearable sensing technologies.

To detect human emotions, various sensing methods within human-computer interaction and affective computing fields are widely adopted [77]. Categorized by the source of information, there are majorly five basic methods that are based on facial expression [4], voice [87], physiological signal [30], text [2] and body movement [90].

One particular challenge of emotion sensing is to obtain ground truth [75]. While human labelling is adopted on large scale databases like AFEW [22] and AffdexNet [60], researchers often ask participants to self-assess emotion and use their reports as ground truth [31]. To solve a lack of details [32, 65], researches also use open-ended questions [55], close-ended questions [64, 79] and scales [31].

Alternatively, some others have adopted real-time ground truth for emotion detection based on facial expressions [18], e.g., full-face based detection API [58]. However, this method limits people's activity and is unsuitable for out-of-lab study, because they need to make sure a user's face can be captured by the camera. For in-lab-elicited emotions, some researchers use multiple methods simultaneously to generate ground truths and find consensus across methods to

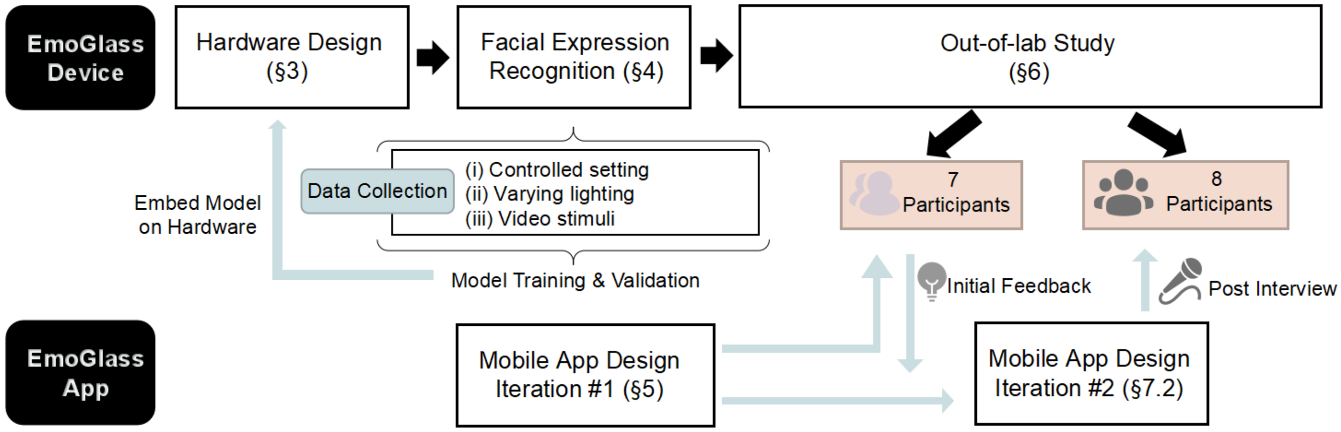

Figure 2: Overview of the development and study process of EmoGlass, corresponding to specific sections in the paper. To be specific, we describe Hardware Design, Facial Expression Recognition, and Mobile APP Design Iteration #1 before Out-of-lab Study. In the middle of the Out-of-lab Study, we provided Mobile Application Design Iteration #2 based on interview results.

increase accuracy. For example, Rattanyu et al. report that they adopted the rating of stimuli for ground truth when matched with alternative sourced labels [71].

Researchers hope to detect people's emotions in their daily lives, instead of in a controlled environment (laboratory setup), as emotions have a special role in people's daily lives, such as driving most decisions [75], affecting social situations [70], and physical health [47]. In this case, wearables equipped with pervasive sensors, which can be used almost everywhere without affecting normal activities, gain wide interests. We intend to fill this gap as we enable emotion monitoring via facial expression recognition eyeglasses while alleviating the challenges of sensing people's emotions in the field [78].

2.3 Facial Expression Recognition on Wearables

Due to the importance of facial expressions, automatic facial expression recognition (FER) is now one of the fundamental tasks and research focus in computer vision, with an accuracy as high as 99.8% [4] on a classic dataset under laboratory-controlled conditions. Nevertheless, it is difficult for wearables to capture the frontal face of wearers without blocking their vision. Alternatively, researchers have placed cameras on eyeglasses [3, 25], headphones [18], and neck pieces [17] to track facial expressions by capturing part of or even just the contour of a user's face. Besides ordinary RGB cameras, researchers also use near-infrared cameras [17] to detect facial expression. Camera-based sensors need to overcome several challenges to achieve FER, including variations of shadows, illumination, head pose, and individuals' facial expressions [76]. Other non-RGB cameras, including thermal cameras [48], depth cameras [85], and RGB-D cameras [51], are also used for FER, yet no research has instrumented these cameras on wearables for facial expression recognition.

To classify facial expressions, or to measure the intensity of each facial expression from face images or videos, researchers have used a variety of computational models. While some choose to use traditional methods like dictionary learning [53], many recent methods are based on convolutional neural networks (CNN). Researchers also use temporal convolution (TCN) [95] or recurrent layers like long short-term memory (LSTM) [89] and gated recurrent units (GRU) [40] to learn temporal information in the video clips. Notably, Li et al. [52] use CNN with an attention mechanism to highlight regional features. To understand the way the CNN model detects facial expressions, Mousavi et al. [61] used de-convolution to visualize CNN and discussed the invariance, redundancy, and filtering for deep networks and compared the representation learned with facial action coding system (FACS).

Prior research has also leveraged low-dimensional multichannel sensing techniques to detect facial deformation. For example, Iravantchi et al. [38] used the acoustic interferometry-based mask to sense nine-class face gestures with 89.0% accuracy. Photo reflective sensors [57] and pressure sensors [50, 57] have also been used in FER by tracking skin deformation. The accuracy of these methods, however, drops off a cliff when the device is removed and put back on. Masai [57] reported their accuracy based on photo reflective sensors mounted smart eyewear decreased by 14.7% from one-time use to uses on different days. Other wearables leverage bioelectricity, such as electromyography (EMG) [30], to sense facial expressions, although this method can be easily disturbed by other factors, such as blinking and the sensor requires close contact to the skin. For the consideration of robustness and comfort (as a wearable component), we chose the camera as the sensor. Further, as detailed later, we go beyond prior work by proposing a systematic method (based on feature analysis of CNN models) to find the optimal camera angle.

3 HARDWARE DESIGN

Specifically, the convolution and pooling layers in a CNN play a role in extracting features, thus representing feature effectiveness. For CNN, reconstructed images indicate features learned, which have been commonly used in selecting configurations that yield optimal learning performance and minimal overfitting

3.1 Form Factor

3.1.1 Iterative Design of Form Factor. We went through a design iteration that yielded three prototypes. The first iteration (Figure 3a) looked like regular eyeglasses with full frames, but we found that the frames would easily obscure the images captured by the camera. Thus, we designed the second version (Figure 3b), which allowed the user to remove the lens. However, we discovered that the weight is mostly on the right side of the eyeglass, making it less comfortable and unbalanced to wear. To resolve this issue, we placed the battery and the control board on separate sides of the eyeglasses, resulting in our final prototype (Figure 3c).

3.1.2 The Choice of Camera. There are a few requirements for the camera: 1) wide-angle, encompassing as many facial features as possible in one image; 2) CSI interface for compatibility with the camera interface of Raspberry Pi Zero; 3) autofocus, maintaining image quality even when the camera's position is shifted; and 4) small to avoid being obtrusive. Most off-the-shelf IR, RGB-IR, RGB-D cameras, and endoscopes are difficult to meet all the requirements, so we choose an RGB camera, specifically, a 120-degree wide-angle RGB camera module based on OV5647 [66].

3.2 Fabrication and Hardware Assembly

Once the design was finalized, we 3D printed the eyeglasses' holder for our camera-mounted system using an Ultimaker S5 with PLA. To capture visible deformations caused by muscle movement on the user's faces, we positioned the camera (Figure 3c) at the front-right corner of the eyeglasses pointing towards a wearer's face. The camera sits on an adjustable seat that can be rotated up-down and left-right. This connection structure with two degrees of freedom makes it easy to adjust camera angles and the resultant captured images. The camera is controlled by a Raspberry Pi Zero W with 512 MB RAM (Figure 3c), capable of transmitting the detection result via Wi-Fi or Bluetooth, and loaded with a 32 GB Micro SD card to save the recording. The Raspberry Pi Zero W has a BCM2835 SoC loaded with one ARM11 core and Broadcom VideoCore IV GPU, providing enough computational power to run the deep learning model and thus achieve real-time inference.

The battery is installed on the left side of our eyeglasses (Figure 3c), using two wires to connect to the devices on the right side, as Figure 3c shows. The whole device is powered by a 500 mAh 3.7 V replaceable Lithium-polymer cell, regulated by an independent tiny DC-DC module to provide 5 V power for Raspberry Pi. Moreover, to make it safer and more comfortable to wear, all electronic devices except the part below the lens of the camera are encapsulated inside the frame of eyeglasses, which also protected the delicate components from being exposed.

To achieve real-time data display on our mobile application, our eyeglasses opportunistically upload data to the server whenever connections are established – specifically when the network is reachable and the new data is available. The post-processing steps of the model's prediction, including softmax, denoise, and so on, are performed on the phone.

As the camera is mounted with an adjustable angle, we need to determine which angle is optimal for detecting facial expressions. One common approach is to compare the accuracy of models trained by images from each angle. However, such a method may be model-dependent and takes a relatively long time to collect data for each possible camera angle.

To achieve a time- and cost-effective angle selection, we conducted a feature analysis based on the visualization of convolutional networks. Specifically, the convolution and pooling layers in a CNN play a role in extracting features, thus representing feature effectiveness. For CNN, reconstructed images indicate features learned, which have been used in selecting configurations that yield optimal learning performance and minimal overfitting [61]. We applied this approach in this experiment. Since the layers in the evaluation network are also the first few layers in our final model, the effectiveness for feature extraction is the same, allowing us to train a smaller model on a smaller dataset and predicting the effectiveness of the larger model on the larger dataset. We built a simple CNN consisting of five convolutions with rectified linear unit activation and pooling layers, just like a ResNet with only two blocks and without skip connections, followed by FC and softmax to make it a classification model, and trained on the frames captured, then visualized using aforementioned deconvolutions method [61, 82, 91]. This allowed us to see the quality of the feature extraction (Figure 4) that indicate the optimal camera angle. We conducted a pilot study with three participants to test five angles, providing us a diverse set of facial features (i.e., "eyes", "nose" and "mouth"). It took roughly three minutes to finish the data collection for each angle, resulting in a total study length of 15 minutes. In the interest of time, participants only performed three facial expressions: natural, happy, and sad. The reconstructed images highlight the facial feature important for CNN, which indicated that the area around the eye and mouth is rich in features. Therefore, we chose the angle corresponding to the third image that the contours of features are most similar to those of different regions of the face (Figure 4c).

4 FACIAL EXPRESSION RECOGNITION

4.1 Data Collection

To bridge low-cost in-lab data and variable in-the-wild conditions, our data collection consists of three iterative experiments: E1) in a controlled environment with ample ambient and uniformed light; E2) in a controlled environment under lighting from six directions, two distances, and two colors; and E3) in a semi-controlled emotion-triggering environment using videos as stimuli. Building a dataset successively in such a three-fold process allows us to quantitatively understand how EmoGlass' facial emotion recognition model can adapt to changes caused by lighting (E1 → E2) and the naturalness of expression (E1 & E2 → E3).

We recruited 15 (7M/8F) participants via an intranet notices service from a local university, $M_{age} = 21.1$, $SD_{age} = 1.6$. The participants come from different backgrounds spanning Computer Science, Design, Administrative Management, English, Remote Sensing, Pharmacy, Medicine, and Agronomy). Because our EmoGlass and normal eyeglasses cannot be worn at the same time, we required our participants to have normal vision or could wear contact lenses. All the participants signed the consent form before the experiment and were paid $50 each in the form of Amazon gift cards. All participants went through all three experiments, and a user-specific model was trained and validated for each participant.

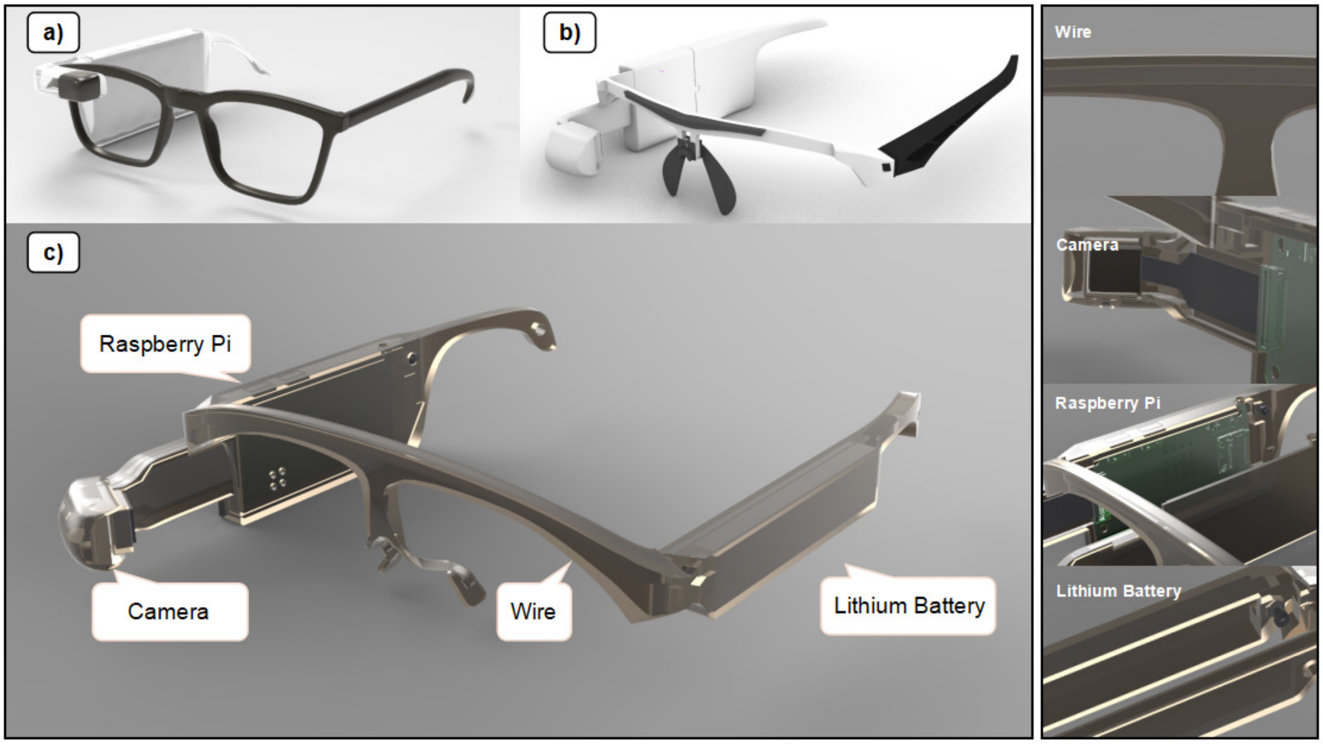

Figure 3: The three eyeglasses prototypes in our iterative design include (a) the first version, in which the frames will block the view; (b) the second version, in which the weight is unbalanced; (c) the final version. The final version contains four main parts: the camera which is used to capture the partial face, the Raspberry Pi which is used to run the ACNN model, and the Lithium Battery which is used to power the system.

4.1.1 *Experiment 1: Controlled environment with uniformed light.*

Based on an approach similar to [69], we used a series of pre-recorded videos (hereafter referred to as emotion cards) portraying facial expressions for the participant to imitate. We used six emotion cards, each lasting five minutes, altogether taking 30 minutes for this experiment.

Before data collection began, we first introduced and described the emotion cards and provided a practice session to help participants get familiar with each facial expression and the pace of the experiment.

Specifically, we considered six types of basic emotions (happy, sad, disgust, angry, surprised, and fearful) and a neutral state. Psychologist Paul Eckman identified these six basic emotions were universally experienced across all cultures [24] and, most of the time, people are in neutral, non-expressionless states. In the experiment, the order of emotion cards is randomized. Participants were asked to follow the facial expression shown on the screen. Two progress bars were shown on the screen: one was the current emotion-imitating progress (5 seconds) and the other tracked the progress of the current session.

In the real world, people might take eyeglasses on and off as frequently as they need, which constantly changes the relative position of the camera and the face. Therefore, to capture this behavior, participants were asked to remount the eyeglasses after each session. The participants were free to take a break between sessions. We deleted data points where facial expressions were interfered with by activities such as coughing, swallowing, or licking lips.

4.1.2 *Experiment 2: Controlled environment with varying directional light.*

Lighting effects, which will affect shadow position, image color and brightness, signal-to-noise ratio, and so on, have been shown to have an impact on image classification in facial expression recognition [39]. To solve this problem, we built extended the previous dataset with various lighting effects [44], which is relatively reliable and low-cost, because CNN has been proved to have the ability to deal with light varying by using more training samples [44].

This experiment followed the same procedure as the previous experiment, except the experimenter had to switch lights during the process. In each session, we adjusted the light angle and light intensity by repositioning the portable light sources across 12 predefined locations (Figure 5) and changed the light color by toggling light settings. Note that, across the 12 locations, there were six unique angles (or direction of light) and along each angle the light was placed either close or far from the participant (thus varying the light intensity based on distance). In total, we had 12 sessions, with each session featuring one unique light location and two colors.

4.1.3 *Experiment 3: Semi-controlled emotion-triggering environment.*

Initially, we used emotions as keywords to select 60 videos as stimuli

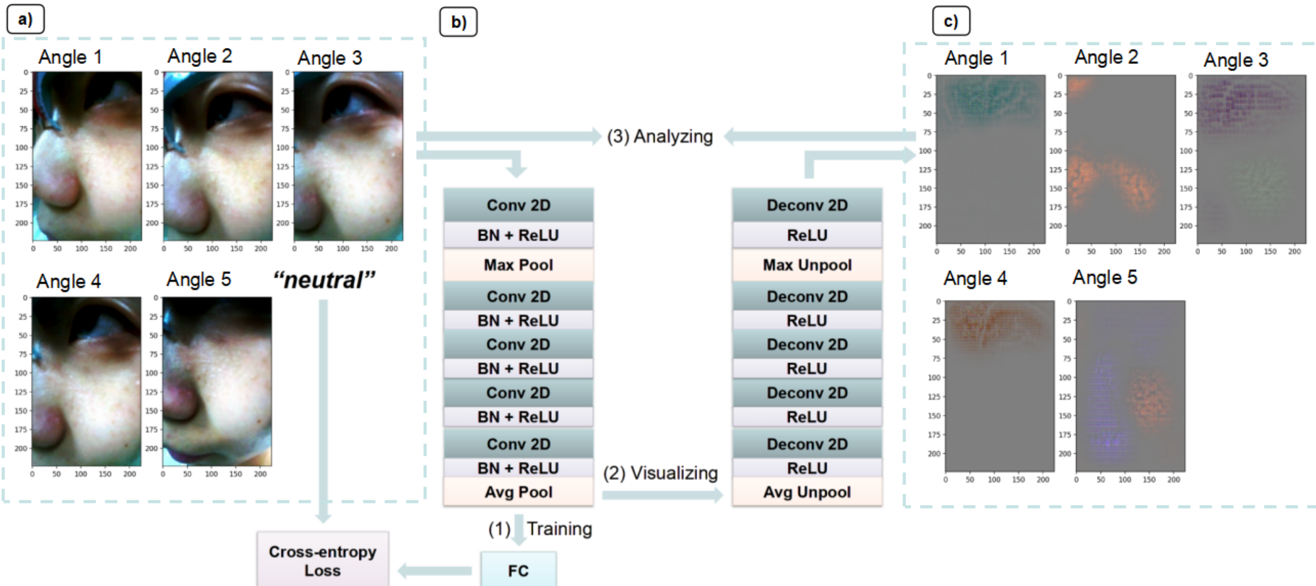

Figure 4: The process to assess camera angle. 1) Small datasets with five different angles are collected (a), then fed into a convolutional network (b) for training. 2) We use parameters from the trained model to reconstruct and visualize feature (c) using a de-convolution network. 3) We interpret the result by analyzing the similarity of input images and reconstructed features and determine the optimal camera angle with the most usable feature.

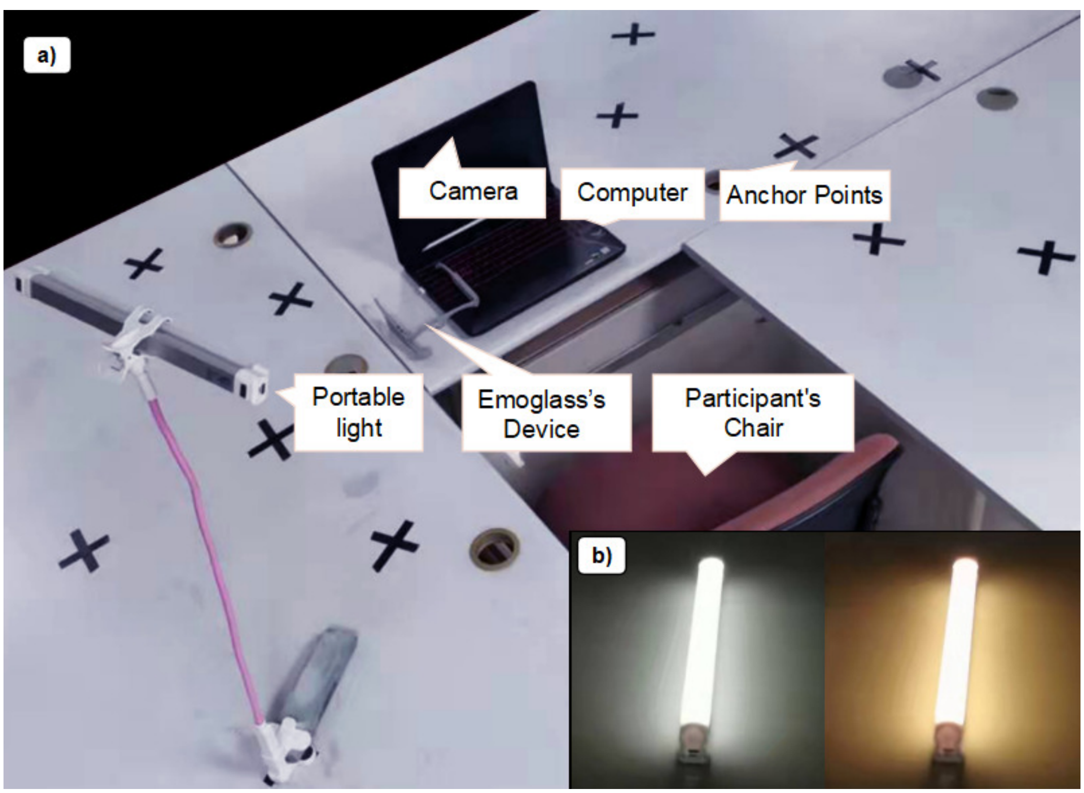

Figure 5: The setup of three data collection: a) the setup of the experimental scene, and b) two light colors used in the experiment.

from YouTube, including 10 videos for each non-neutral expression and none for neutral expression because it does not need to be triggered.

Next, four experimenters watched the videos and scored (1-10) how much the video can trigger certain facial expressions. Based on the scores, we chose the top five videos with the highest ratings for each emotion category and concatenate them to generate a 30 min stimuli video. We asked the participants to watch the video and recorded their emotions using Affectiva API.

4.2 Ground Truth Acquisition

To acquire the ground truth of participants' facial expressions, we recorded full-face videos using a separate camera. We did not use labels of videos directly because the video pieces might not arouse corresponding emotions from the videos' beginning to end. We then performed a temporal segmentation [21] using 68 feature points of the face. If the feature points of the two consecutive frames have decreased to a pre-defined threshold, we considered that the faces corresponding to the two frames exhibited the same expression. In this way, we segmented the camera streams by distinct expressions and reduced the amount of data that needed to be labeled. There are three common methods to obtain ground truth for facial expressions: 1) using the tag of emotion card; 2) using computer-vision-based facial expression detection API [18], and 3) manual labeling by the experimenter.

We tried all three methods in an early pilot study. First, we found several discrepancies in results between the emotion card and the API method. Additionally, our participants reported that they have difficulty imitating the standard facial expressions indicated in the emotion cards. Moreover, the API itself also has a limited performance (sometimes even failing to detect faces). As a result, we eventually used emotion cards and API methods as a filtering process. These two ground truths' labeled ground truths were used only when they agreed with each other; the remaining instances were manually labeled (i.e., the third method).

4.3 Image Pre-processing and Data Augmentation

Since the eyeglasses may shift during uses and users might often remount eyeglasses in their daily lives, we prepossessed each image to incorporate such variances. Specifically, we first applied a threshold and morphological calculations on the gray-scaled image (Figure 6b) to extract the dim part in the images, including pupils, hair, and shadows (Figure 6c). We calculated connected components in the image and used a position-and-area-related weighing function to choose the most appropriate component (Figure 6d). We used another weighing function to calculate the position of two key points (at the corner of eyes, Figure 6e, Figure 6f). The image is then positioned according to the position of two key points(Figure 6g). We applied a rectangle-shaped mask to remove the effect of edge areas and histogram-scaled to lessen the effect of the luminaries' environment (Figure 6h). We also resized the image into 224x224 (Figure 6i).

Our data augmentation increased the robustness against camera angle changes due to motion. Specifically, we applied affine transformation on our dataset to enlarge the dataset itself with enough noise while not increasing experiment time. The affine transform we applied is constructed from a scaling (scaling factor range from 0.8 to 1.2, distributed uniformly) with 60% probability, a rotating (six degrees at most in both directions, distributed uniformly) with 60% probability, and a translating (30 pixels at most in both axes, distributed uniformly) with a 60% probability.

Through the pilot study, we found that the color of ambient light will create a halo of color around the image for the light tint. Due to the complexity of the light color, we chose to augment the data set by using some virtual light effect overlay, including random (linear) color correction and random gamma correction.

4.4 Deep Learning pipeline

4.4.1 Model selection. Because the main part of the image is the cheek that cannot provide much helpful information, assigning different weights to different areas is important to highlight regional features. We add regional CNNs with an attention net to calculate the attention (weight) based on the work of [52] to make it an ACNN and enable the ability to highlight regional features.

4.4.2 Network Architecture. The model we used is based on the work of [52], which added regional CNN to generate regional feature embedding by using a pre-defined deterministic method (regions around specific key points in face mesh in their case) besides global feature embedding. We replaced the VGG-16 backbone in the model [52] with ResNet-18 to reduce parameter count, and adopt fixed-position windows to capture regional features since the camera we used is mounted on participants' heads and stayed at a constant relative position. We reused the first few blocks from ResNet-18, including the initial convolution and pooling layers and four basic residual blocks, which encode a three-channel 224x224 image into a feature vector sized 128x28x28. To better focus on regional features such as mouth and eyes features, we created eight different windows sized 128x12x12 separating in the feature space (see Figure 7b) and use two two-layer residual blocks (the parameter of blocks is independent across different windows) to extract regional features within the region. An attention net consisting of pooling and convolution is used to learn the attention within the region. We use sigmoid to normalize the calculated attention value and use the value to weigh (to multiply) the features. We also used another unit including residual blocks and another attention net to use global features. The result of the global unit and eight regional units are concatenated, and a linear layer is used to encode the feature into a 1024-value vector.

A final linear layer and $Softmax$ operator are used to make the classification. We used batch normalization after each convolution in the network to help the training. The initial parameters of our model are given randomly.

4.5 Training & Validation

We trained and evaluated models on multiple datasets and tested the effectiveness of transfer learning techniques to increase the robustness of models based on a controlled and semi-controlled setting. For validation sets, we chose the data from the last session in experiment 1 and experiment 2, and the 17% data from the tail in experiment 3 (referred to as V1, V2, and V3, respectively). The

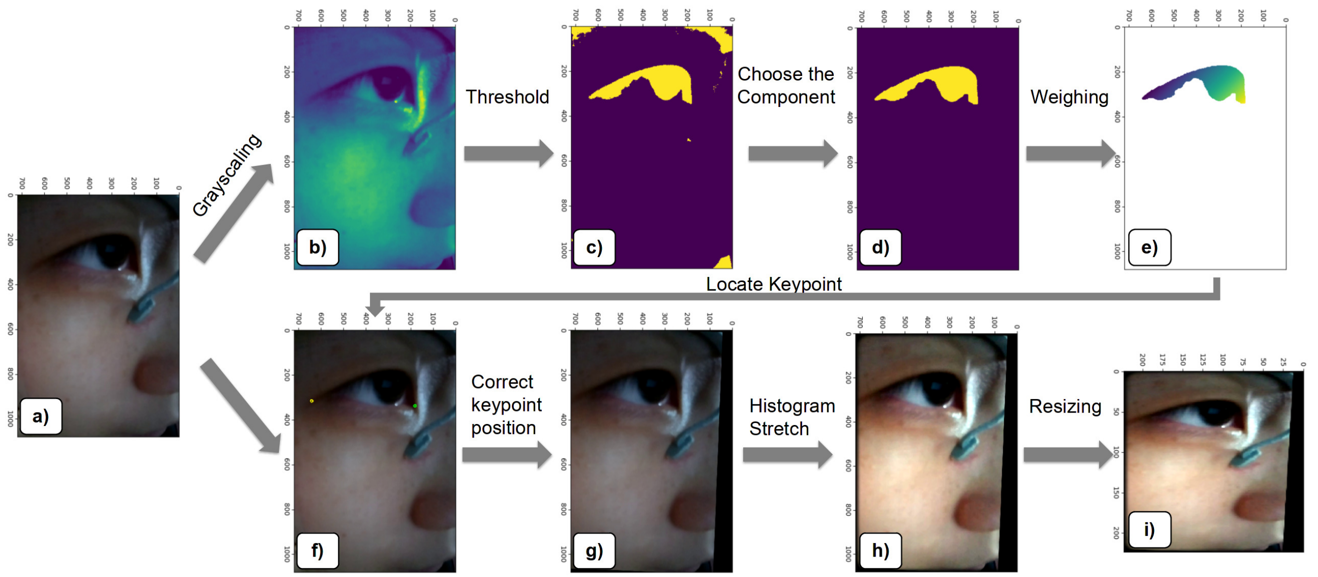

Figure 6: Image pre-processing workflow. To locate the key points (f) from input (a), we apply threshold (c) on the grayscaled image (b) then choose the eyes component (d) and locate the key points using weighing function (e). We correct the displacement and rotation of the image (g), equalize the light (h), and adjust the size (i).

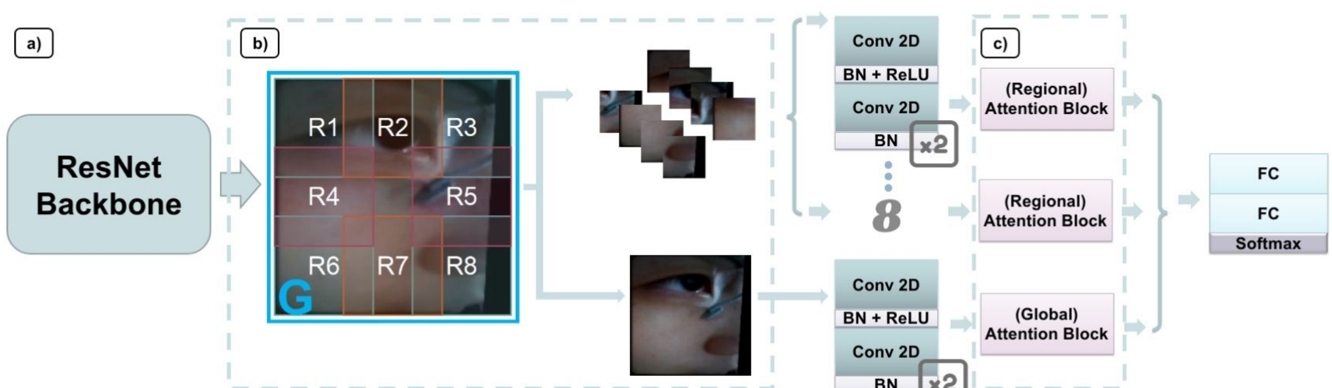

Figure 7: Network architecture. a) the ResNet backbone, b) the decomposition of pictures, c) the attention blocks.

optimizer we used for this network is Adam, with the cosine annealing learning rate scheduler. We used cross-entropy to represent classification loss.

We first trained our per-user models (referred to as M1) for all 15 participants independently with random initial parameters on the dataset collected in experiment 1 (T1) and validate the result on all three validation sets (V1, V2, V3). The result is presented in and Table 1 (detailed accuracy data for each participant is provided in the appendix). Across all 15 participants, M1 achieves 85.0% accuracy (SD=7.9%) on the validation set from experiment 1 (V1). The highest accuracy is 97.1% (from P11). There are four accuracy lower than 80% (68.1% for P9, 73.9% for P10, 76.0% for P5, 79.4% for P8). The median accuracy is 88.1%.

We then transferred M1 onto experiment 2's dataset (T2), training a new model M2. We validated M2 on the second validation set (V2). V2 reached an accuracy of 80.6% across 15 participants, both significantly below M1 on V2 (note that T2 and V2 come from experiment 2, while M1 is trained on T1 from experiment 1). This suggests that it becomes more difficult to distinguish between emotions when the lighting is not the same. This result suggests that by transfer learning, we can transfer the effectiveness in a controlled laboratory environment into the effectiveness in a less-controlled environment just by adding a certain amount of extra data (even though this extra data alone cannot train a new model from scratch).

In the third step, we transferred M2 onto the third experiment's dataset (T3), and validated it on the third validation set (V3). The

	Accuracy on V1	Accuracy on V2	Accuracy on V3
trained on T1	85.0% (SD=7.9%)	71.6% (SD=11.1%)	62.2% (SD=8.4%)
transferred onto T2	/	80.6% (SD=8.1%)	64.5% (SD=8.7%)
transferred onto T3	/	/	80.9% (SD=9.1%)

Table 1: Average accuracy on the dataset. We trained the model on training set 1 (T1) and transferred on training set 2 (T2) and later training set 3 (T3). The accuracy is calculated on validation sets (V1, V2 and V3). The accuracy trends in the figure illustrate the need to collect data sets for different lighting conditions and natural expressions.

mean accuracy reached 80.9%, which is close to the mean accuracy on the second experiment. However, the worst accuracy is only 57.0% (P15), while the highest accuracy hits 98.0% (P4). The standard deviation of accuracy is 9.1%, higher than that in both experiment 1 (7.9%) and experiment 2 (8.1%). One explanation is that the accuracy in this section is affected by 1) the differences between controlled (following emotion cards) and naturally-occurring (triggered by YouTube videos) facial expressions and 2) the unbalanced number of samples in different categories of expressions, since each participant's amount of facial expression could be quite different even as they watched the same video clip.

Figure 8: Average confusion matrix for the first model validated on the first validation set. We averaged each item in the confusion matrix for each participant and then normalize the array.

We also validated M1 on V2 + V3 and validated the M2 on V3 as a baseline for transfer learning. Theoretically, transfer learning should help the model adapt to a new dataset without too many samples and training costs and result in a rise in the accuracy on the same validation dataset. However, we were surprised to find that the accuracy of the second validation set dropped after transfer learning on T2 for P4 (86.8% to 84.6%) and P11 (73.3% to 64.3%), though the average accuracy of all participants increased from 71.6% to 80.6%. One possible explanation could be that some participants might have followed emotion cards differently in experiment 2 due to fatigue or other factors. On V3, M1 and M2 reach a mean accuracy of 62.2% and 64.5%, respectively. In comparison, after transfer learning, M3 on V3 reaches the mean accuracy of 80.8%. Such an increase suggests how T3 made a major difference in transferring models originally trained on T1 and T2.

Notably, the distribution of accuracy of the third dataset T3, which includes more natural facial expressions, is significantly correlated with the accuracy of only trained on emotion cards (Pearson correlation, R=0.6288, p=0.012<0.05), suggesting the dataset we collected plays an important role in improving the accuracy of sensing human's facial expression in their daily life.

5 MOBILE APP DESIGN: ITERATION #1

We describe our initial design and implementation of the EmoGlass mobile application (Figure 9), which we used for the subsequent out-of-lab study and iterated later with the first seven participants. As shown in Figure 9, this first version of the EmoGlass application consists of the following functionalities:

- **Start/stop control of the device** is accessible on the home page and the recording page. The home page (Figure 9a) allows users to either start the recording by pressing the *start* button, or view the recorded data by selecting a date in the calendar. The recording page is shown when users are actively recording their expressions. Users can press the *end* button to stop the recording.
- **Emotion stats** is presented with different presentation and scale. Weekly report page (Figure 9b) provides line chart visualizations of weekly data, while also presenting the average proportion of occurrence for each emotion on each day as the vertical axis.
- **Daily report** page (Figure 9c and Figure 9d) illustrates emotion data within a day in various visualizations including a polygon graph, a line chart, and a qualitative evaluation report that describes the emotion comprehensively. The *see more* button at the bottom allows users to check detailed information within the specific hour at the date. Specific time page (Figure 9e) presents the intensity of each emotion within the specific hour as a line graph with a granularity of one minute.
- **Prompt and diary** Probe page (Figure 9f) asks users to choose an emotion that best represents their emotion at the time, which we use to assess the accuracy in the out-of-lab study and also to compose the Diary page (Figure 9g) that renders the recorded emotion diary for viewing later. The

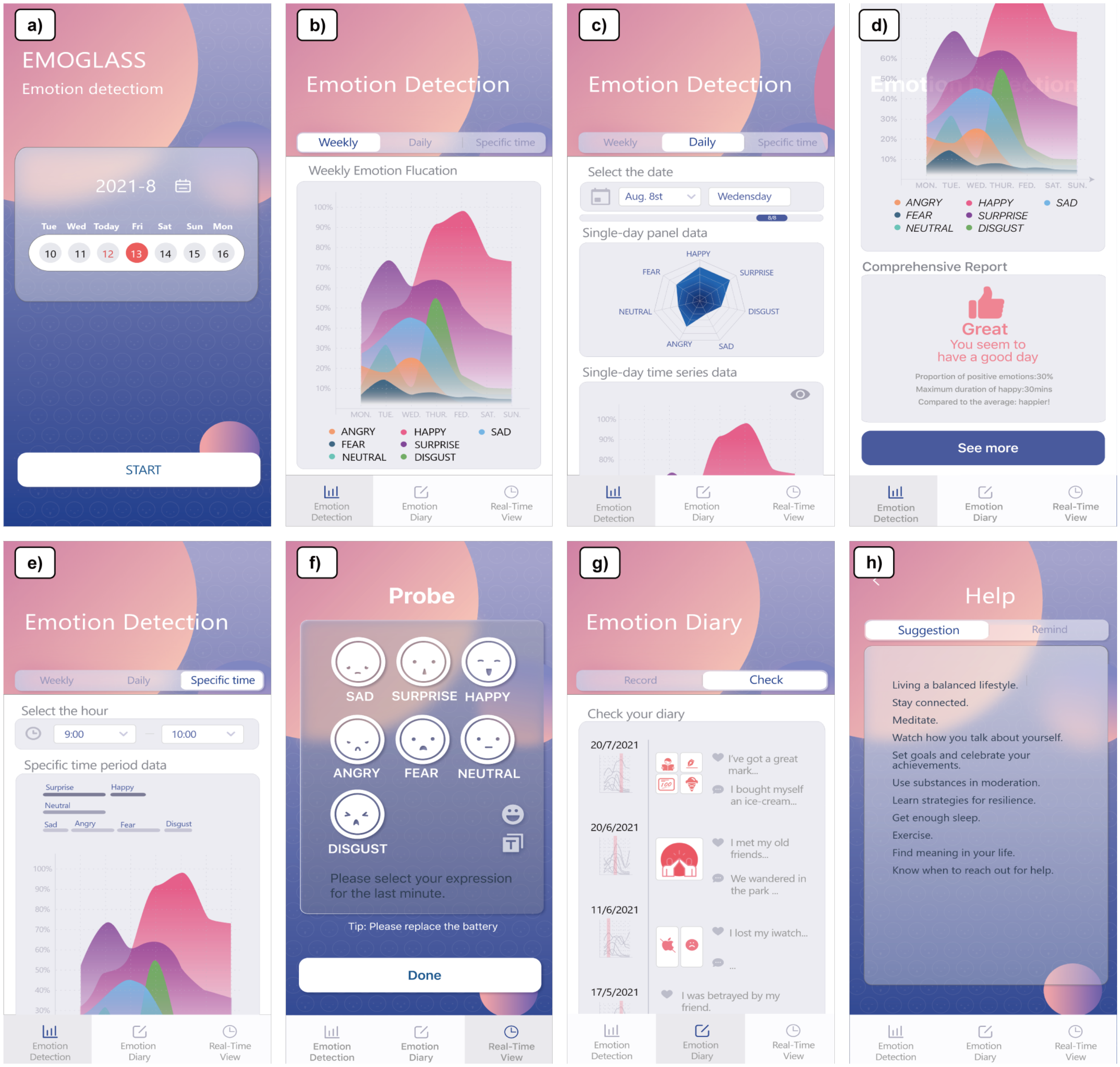

Figure 9: EmoGlass workflow with initial UI design. a) The home page, b) Weekly report page, c) & d) Daily report page, e) Specific time page, f) Probe page, g) Diary page, and h) Help and suggestion page.

probe page pops up every 15 minutes during the out-of-lab study.

- **Help and suggestions** page (Figure 9h) provides general guidance and suggestions to users to further learn about and adjust their emotional health.

When each user opened the application for the first time, we used the unique ID of the user to register on the application. The application was built using React Native and Expo could be run on both iOS and Android platforms. To improve the offline performance of our app, we cached the latest statistics and provided users information of their connection history and status, for example, "last seen 1.5 hours ago".

6 OUT-OF-LAB STUDY

To investigate the feasibility of EmoGlass, we conducted an out-of-lab study that first evaluated the facial expression recognition performance of EmoGlass given the out-of-lab scenarios. Next, we leveraged feedback from seven participants to iterate our mobile application design. Finally, we deployed the redesigned application and collected more feedbacks, reported as overall findings and design recommendations.

We recruited the same 15 participants from the data collection experiments. We provided each participant with one EmoGlass device and two rechargeable lithium batteries. We helped them install EmoGlass mobile application on their phone. It is worth mentioning that we divided 15 participants into two groups — seven participants using mobile application iteration #1 and the other eight using mobile application iteration #2.

The out-of-lab study consists of several key steps:

- Introduction. We kicked off the study by introducing the background and motivation of EmoGlass to each participant. Participants then filled out a survey for collecting their demographic and general emotional health-related information. Then we introduced them to how to use the EmoGlass device and mobile application. Moreover, we told them to keep the camera location as consistent as possible and make sure there is a Wi-Fi connection during usage. By observing the user records on the app's backend server, we monitored the application use situation of participants.
- Out-of-lab sessions. Following the introduction, we asked each participant to use EmoGlass for at least three hours per day for the next three days. They could choose wherever they would want to wear EmoGlass. They were also free to remount the EmoGlass during that period. We reminded them to replace the battery before it ran out (after around one hour of use). Following an experience sampling approach, a pop-up menu showed up in our app every 15 minutes, prompting users to tag their current emotions.
- Post-study interview. At the end of the three days, we met with participants again and conducted a semi-structured interview asking about their experience and feedback. The interview was structured around five high-level topics: 1) Notion of emotional health, 2) Self-perception of emotional Fluctuation, 3) Wearable sensors, 4) Mobile application, and 5) The overall end-to-end platform.

7 RESULTS & FINDINGS

7.1 Facial Expression Recognition: Technical Performance

During the three-day-long out-of-lab study, 36 prompts were given to each participant when they were wearing the device. Note that we paused the experience sampling when eyeglasses were off the participants. We used the labels provided by participants as ground truth data points, which were used in calculating the facial expression recognition accuracy (Figure 11a). The average accuracy across 15 participants is 73.0% (SD=18.0%, Median=80.6%).

Notably, three participants yielded accuracies below 50%, namely P4 (47.2%), P11 (44.4%), and P15 (30.6%). According to their self-report, P4 walked on darker roads at night for almost 20 minutes, P11 felt uncomfortable with the glasses and adjusted the position of the glasses several times, and P15 reported that he took some conference calls while wearing glasses. Because our device cannot provide them with accurate emotional detection feedback, to avoid introducing bias caused by sensing inaccuracy, we excluded these three participants later in the qualitative analysis.

We collected 224 samples in total, including 115 neutral-marked, 69 happy-marked, and 19 sad-marked samples. Figure 11b shows the percentages of expressions in our collected samples. F1 scores for each class (for all participants) are given in Figure 10.

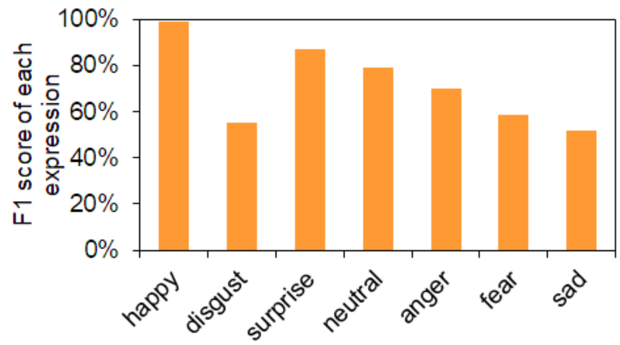

Figure 10: F1 score of facial expression recognition in the out-of-lab study.

There are several possible reasons for such a low accuracy. For P4 and P15, the accuracy of the validation set in the third dataset is either significantly too high (98.0% for P4, Z-score=1.87, p=0.031<0.05) or significantly too low (57.0% for P15, Z score=-2.62, p=0.004<0.05), suggesting the model may be overfitting on that dataset (for P4) or failed to learn enough features (for P15). For P11, the accuracy on validation sets is 84.3% (Z score=0.37, p=0.710935, not significantly out of the population).

For P11, in T3-V3, the accuracy for neutral is 67.1%, quite low compared with happy (98.0%) and surprise (95.2%). In the out-of-lab study, P11 reported 17 happy, 16 neutral, 2 anger and 1 surprised emotion in total. The trained model successfully identified 15 happy prompts out of 17, but misclassified all neutral prompts as sad, which caused the overall accuracy to drop to 44.4%. One possible explanation is that the Affectiva API we used for ground truth labels in experiment 1, 2 and 3 may have misinterpreted the expression for sad for P11.

7.2 Mobile App Design: Initial Feedback and Iteration #2

In the middle of the out-of-lab study, we gathered and analyzed seven participants' feedback and designed the second iteration of the mobile application. We describe below what we learned during the interview and how we adjusted our mobile application.

7.2.1 The need for educating users about emotional health. We found that most participants had an incomprehensive understanding of emotional health. First, participants tended to focus only on the manifestations of emotions. For example, P2 said, "Being able to express all kinds of emotions is emotional health". Second, participants sometimes overly emphasized emotional stability. P6 said, "Emotional health is mainly emotional stability, without too much volatility". Additionally, we found that participants tended to

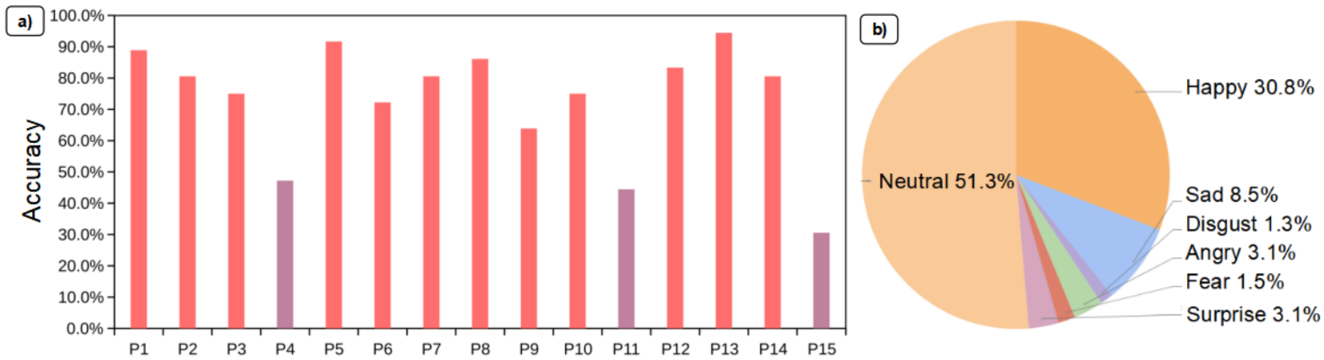

Figure 11: (a) Accuracy of facial expression recognition in the out-of-lab study, accuracy lower than 50% (P4, P11, P15) is marked in purple color (b) the average distribution of participants' prompted emotions.

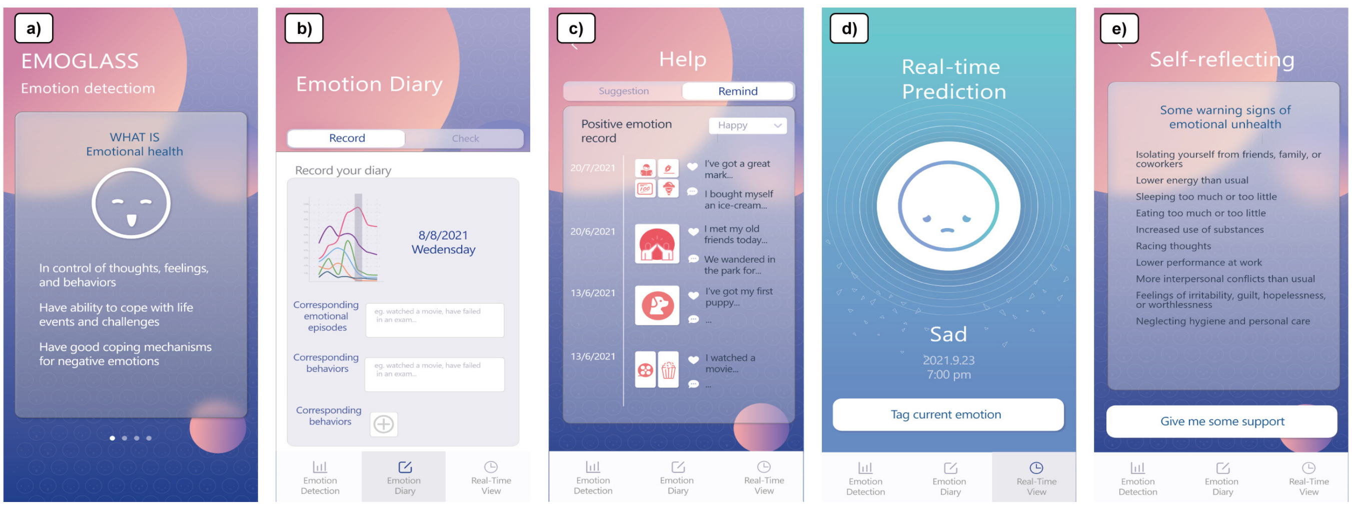

Figure 12: Added feature in EmoGlass Mobile App in Iteration #2, including a) a new landing page with the description of emotional health, b) a new page for tagging triggers on the emotion graph, c) a help page for emotion regulation help by using recorded triggers, d) a new page providing real-time prediction function, and e) a new page for self-diagnostic advice.

neglect the ability to regulate emotions as part of emotional health. Only P5, P7 and P3 mentioned that emotional health includes allowing ourselves to adjust by not dwelling on negative emotions for too long. Moreover, participants neglected the control of behavior. Only P7 reported that "Emotional health is about not being immersed in negative emotions for too long and not developing into negative behaviors". Furthermore, we found that participants lacked the knowledge of how awareness of one's emotion might play an important role in assessing and improving one's emotional health, and the go-to solution was always "getting an evaluation with a psychologist" or "fill out a survey".

- **Design Change #1: an introduction of what emotional health is and is not** (Figure 12a), which presents the definition of emotional health and clarifications of common misconceptions found in participants' responses.

7.2.2 Visualizing emotion is not enough. Users need to be contextualized with triggers and bodily responses. Some participants tended to associate their emotions with reasons of emotion triggering. For example, P5 reported, "My emotion is easily affected by life pressure and academic pressure. If I have many deadlines today, I will feel like I am not well today." Most participants mentioned that, to be aware of past emotions, they need to recall what happened during that period. For instance, P2 said, "I think my worst month of the year was exam month." Some other participants mentioned that they only noticed their emotions when their negative emotions built up enough to affect their behaviors. For instance, P7 said, "I tend to analyze emotions on the basis of my bodily response. If I do not move all day and feel sluggish, I know I feel blue." P8 and P9 reported that they do not want to chat with anyone when they are down. However, no participants could fully recall and connect events to form a clear and complete mapping of emotion triggers—emotion—bodily response.

- **Design change #2: 'Emotion Diary' for recording emotion triggers** (Figure 12b) when checking daily report.

Specifically, a user can select a certain period on the emotion fluctuation graph to tag particular emotions. A prompt will guide the user to write a description for this event and upload corresponding pictures if needed. Compared with prompt-based activity recording, this design is much more convenient and non-disruptive, as the user does not need to respond to the prompt throughout the day. Besides, this function can also help users to keep track of little things that go unnoticed when happening. For instance, P12 said, "When I saw the daily report, especially some emotions that match with what I expected. I would rack my brain to recall what happened during that time. In the process, I found some happiness that I had overlooked."

7.2.3 Reminding/suggesting positive activities to regulate emotions. Participants noted there were recurring emotion triggers. P7 said, "I always feel [the same] when encountered with same incidents over and over again. However, I cannot realize it in time. If EmoGlass can record similar incidents and advise me to change my response, my overall emotional condition could be improved." However, participants rarely made use of such recurring events, e.g., taking actions that had been shown before to result in positive emotions. Unlike physical health that can be improved under some general suggestions, the approaches that make people happy are somewhat personally different, yet some participants would look for general guidance rather than their personal history for help. P1 said, "I have searched for them on the Internet, but it seems just work for others except for me." P6 said, "I do not remember what kinds of things can make me happy again, nothing, nothing at all." It could be difficult to re-trigger happiness when one is in a low emotion, thus recording and reusing positive-emotion-related episodes might be helpful.

- **Design change #3: Help-Remind (Figure 12c) supports users by reminding them of activities that happened during their previous positive emotion**. To get suggestions about what to do for promoting positive emotions, a user can search by keywords (e.g., happiness) what actions were recorded in the past.

7.2.4 The trust of emotion detection. Some participants with a technical background (e.g., P9) mentioned that they understood how EmoGlass works. However, other participants expressed doubts, because they do not think the facial expression can always represent the underlying emotions, especially in the public scenes, and six emotion is not able to cover all human emotions. For example, P3 said, "Emotions are a complicated thing. I feel that the facial expression reflects the emotion that people want to express, not the true emotion". We also asked our participants who had doubts whether they trusted other health sensing wearables. P5 said, "At the beginning, I also did not trust sports bracelets. I just used them for fun. However, one night at 12 o 'clock, my teacher suddenly advanced the deadline of the assignment, and I found that my exercise bracelet showed that my blood pressure rose instantly. This case, consistent with my expectation, made me believe in sports bracelets". A few other participants also mentioned that their trust in the results depended on whether the device's feedback matched their expectations during the period when they knew exactly what their status was.

Besides, P1 said some feedback and interaction can improve users' trust, "I trust Apple Watch because when I click the button to measure heart rate, the watch will vibrate after one minute, which gives me a feeling that the watch indeed measures my heart rate carefully."

- **Design change #4: Real-time detection (Figure 12d) allows users to check what EmoGlass thinks their emotion is around this very moment**.

7.2.5 Connecting emotional awareness and emotional health. EmoGlass aims at enhancing people's awareness of their emotions, which is critical for maintaining emotional health. However, some participants were unsure about the difference between these two concepts, i.e., how to judge emotional health according to EmoGlass' emotion detection and quantitative reports. For example, P14 asked if "The mobile can tell me the ups and downs of my emotion, and it can tell me how my emotion has changed in a long time, but I think it's more like emotion management." P10 asked for a more detailed description, "If I'm having a bad day and I get angry ... it could be a sign of emotional health because I'm working through my negative emotions." P2, P7, P8 also asked similar questions about how to identify emotional health problems.

- **Design change #5: the Self-reflecting feature (Figure 12e) presents guidance for users to further perform self-diagnosis of their emotional health**, which goes beyond showing emotional detection results by providing more information about signs of emotional unhealthiness.

7.3 Feedback to Mobile App Redesign, Overall Findings, and Design Recommendations

As we report findings after we deployed the redesigned application to the remaining eight participants, we highlight insights specific to design changes added in the second iteration as bold texts in parentheses. Overall, participants responded positively to the idea of EmoGlass, including a systematic exposure to the concept of emotional health (**Intro to Emotional Health & Self-Reflection**), the availability of emotional awareness integrated into wearables, rich feedback, visualizations, and interactive features (**Help-Remind & Emotion Diary-Record**) in the mobile application that goes beyond traditional measuring methods (e.g., survey).

7.3.1 How did participants respond to using wearables for emotion detection? All participants recognized the value of using wearable devices to detect emotions. They felt the detection mechanism is more transparent, as P2 said, "I can understand the working mechanism of wearables and the way it produces results, so I am more willing to believe it than traditional scales." However, before we add the real-time function, it was a little hard for non-technical participants to understand the system. P9, P14, and P13 mentioned that the real-time prediction function gives them a clear sense of how each data point is generated (**Real-Time Prediction**). Over half of the participants mentioned the convenience of wearables. For example, P3 said, "It is troublesome to fill out a form and go to the doctor. However, if I have wearables, the only thing I need to do is wearing them." P8 said, "I used an emotion management app before, but I had to manually input all the emotions. It was so troublesome that I didn't want to use it after a few days. EmoGlass is very convenient" (**Emotion Diary-Record**). P6 said, "The report of wearables has exact numbers, which is more nuanced than scale or other methods". P7 mentioned the value of identifying problems early on — "Many people do not take the initiative to assess their emotional problem until it is serious enough to realize it. But wearables are a timely reminder."

Some participants also showed their concern towards wearables: 1) The presence of wearables might change people's behavior. P1 said," If a user understands the logic of the product and prefers to be healthy, they will subconsciously hide their genuine emotions by expression suppression." 2) The competitive relationship between wearables. P14 said, "Different wearable devices, similar in form, each has only a few functions. But there is usually only one type of wearable that users are willing to use." 3) Social appropriateness. More than half of the participants mentioned they are not willing to wear such eyeglasses on social occasions, for example, meeting with people at work.

- **Design recommendations**. A wearable emotional sensor can provide access to rich behavior data always available, but the form factor needs to be as lightweight as possible to minimize how many users notice the device. One idea is making it modular, e.g., the EmoGlass device can be iterated as an accessory to regular eyeglasses or other smart eyewear. Finally, due to the sensitive nature of emotion sensing, the wearable should be easy and quick to turn off on certain social occasions.

7.3.2 How was participants' concern of privacy? When designing the hardware, we consider the privacy issue: 1) We positioned the camera so that it mostly captures the face, not the environment. Before the experiment began, the participants were shown images from the camera to ensure that no sensitive information would be captured. 2) The server, which is used to ease connectivity problems, does not parse the data, and the results calculated by the Raspberry Pi model are directly sent to and analyzed on the user's phone.

However, three participants still had concerns, even though they knew the cameras would capture only partial faces. P1 said, "I'm afraid to go to the bathroom with these camera eyeglasses. What if they slip off and the camera changes direction?" P7 and P13 also expressed the same concerns about the change of camera angle during long-term use and suggested that "maybe you can add a function on the mobile application, allowing users to check the picture captured by the camera whenever they want."

Thirteen participants mentioned their concerns about data leakage — it was not the face images they were worried about but the emotional data. For example, P5 asked, "Can data be transferred without going through the server?" P12 suggested that "Maybe you can make the eyeglasses work offline. In this case, I will be more willing to use it".

- **Design recommendations**. Besides adopting best practices for protecting users' privacy (which EmoGlass already did), it is also important to show and explain to users how such privacy-protecting mechanisms work, e.g., allowing users to see what is or has been recorded and showing how data is directly transmitted between the glasses and the phone.

7.3.3 How was participants' emotional awareness? First, EmoGlass enabled participants to periodically reflect on their emotions. P2, P8, P10 mentioned that when they see the daily report, they would analyze why the emotion of the day was the way it was and what transpired. Some participants said that, later in the experiment, they would actively compare their emotions with those of the previous days to see any abnormalities. For instance, P3 said, "I really like the weekly report, I can compare the data before and after. Maybe you can provide a monthly report, it will be interesting."

Second, most of the participants found EmoGlass's detection and visualization helpful and informative. For example, P1 and P14 who self-reported being overly sensitive mentioned that after seeing their daily emotion report, they recalled what happened at each emotion turning point to reflect on emotionally-related events. In this way, EmoGlass enabled them to be more mindful about their emotional sensitivity. Three participants reported that the application recorded some of their unintentional emotion fluctuations. P3 said, "I thought I enjoyed being alone, but I found that when I was working overtime alone, a colleague suddenly came to me. Even though we don't know each other very well, I actually smiled for a long time. Maybe the company made me happy") (**Emotion Diary-Record**).

- **Design recommendations**. Finally, participants also identified needs for new emotion sensing functionalities that can be added to EmoGlass. P13 found that EmoGlass can provide very direct feedback about emotion, however, currently there is no support for sensing a mixture of emotions such as boring and agitated, and hidden emotions, which are not shown on the face. Some users also had higher expectations for EmoGlass. P12 said, "I feel that EmoGlass can further strengthen the guiding function ... Perhaps the ultimate goal of EmoGlass is not necessarily to keep people dependent on the EmoGlass but to guide behavior and enable users to have self-awareness." Future design can involve other "off-device" activities, e.g., expressive writing, to help users gain independence in maintaining awareness of their own emotions.

7.3.4 Would participants use EmoGlass in their daily lives? The six participants reported that they would habitually use EmoGlass in their daily life as a reference for tracking their emotions over time and as a preventative tool, since they did not know when they might have emotional issues in the future. The other participants mentioned that they would prefer using EmoGlass only in specific scenarios. For example, P7 reported that he would only use EmoGlass in private places because, in these situations, he will be more relaxed to release his emotions instead of suppressing them due to social considerations. P1 would use our system in social situations to check whether his facial expressions are socially appropriate. P4 proposed that we set up service sites of EmoGlass in schools or shopping malls and let people try them out for a while instead of long-term use to democratize the notion of emotional health for more people. Three participants reported a surprising usage scenario — informing others how they feel by using this application. For example, P2 said, "My boyfriend has no idea about my emotion. Sometimes I tell him I'm angry, but he doesn't believe me. If I had this application, I'd send him a screenshot of my emotion report when I got mad." The other two participants also expressed their desire for promoting empathy via sharing emotion using our system.

- **Design recommendations**. Given the various preferred ways of using EmoGlass, future design can incorporate personalized modes, such as a "private" mode that only turns on the device in private locations, a "social" mode that indicates, in real-time, the user's facial expression, and a "sharing" mode that notifies closed ones when certain emotions occur.

8 DISCUSSIONS, LIMITATIONS AND FUTURE WORK

In the out-of-lab study, we chose an experience sampling frequency of one prompt every 15 minutes for not wanting to interrupt participants' daily activities. However, this sampling is relatively sparse, which yielded much fewer data points than ones collected in controlled environments. The amount of data needed to achieve statistical significance needs to be increased, which we hope to explore in the future with long-term deployment studies of EmoGlass with a larger user base.

We also do not set detailed scenes for the out-of-lab study. We let participants act in accordance with their normal behavior and daily settings. However, such an approach is prone to error. For example, being in the dark or chatting for long periods of time can significantly decrease the detection accuracy of our platform. Moreover, it made us miss an opportunity to gain more insights into other factors, such as social norms. In the future, we can give more detailed tasks, including working in public places or staying alone at home, etc.

Additionally, our facial expression detection is based on sensing local regions of users' faces. In comparison, another mainstream approach, which is sensing face contours, has several innate shortcomings. First, the camera needs to be strategically angled to have a field of view that includes both the user's face and the background. Having both faces and backgrounds in captured video feeds might result in technical difficulties in foreground-background separations and privacy concerns for users. However, sensing face regions has their own set of drawbacks, the accuracy is not so high due to the relatively small sensing area, which we hope to mitigate by exploring future work that can be innately more robust at sensing faces, such as using depth cameras.

Although the current accuracy does not seem high enough for real-world applications and only a small portion of participants would continue to use the system, most users liked our concept. We believe that the EmoGlass platform can appear more user-friendly by refining the hardware and algorithm. Although this is only a preliminary attempt, our result can support that the method we proposed is better than visualizing data only, and can truly trigger people's emotional awareness rather than just inform results. Our design findings also lay a foundation to further sensing system design in the emotion-related fields, such as affective computing. Besides, by using end-to-end concept [92], our work goes beyond monitoring and visualization, and tried to guide the change of perception and thus lead to the change of behaviors [28, 63]. In this case, our platform might be a solution to overcome people's overrelying on AI [16], which might exert long-term use of monitoring wearables. Furthermore, our system can contribute to other topics, such as emoji typing, silent speech, eye-tracking-based attention measurement, and mHealth monitoring.

It is well-known that Deep Learning requires a large amount of training data, which we achieved via our three-part iterative data collection studies. We found user-dependent models to be most effective, as different individuals' facial expressions might differ significantly. However, this process requires calibrations that demand user efforts. To mitigate this issue, we hope to explore leveraging synthesized data from virtual cameras and textured 3D models of users' faces, which can be easily built with depth-sensing techniques (e.g. Apple Face ID and LiDAR). We envision that EmoGlass can easily fine-tune its model parameters using data synthesized from end users' face models during a brief initial calibration. Additionally, we plan to build user-independent models that can work across users without calibrations for use scenarios that require readily available emotion detection and logging.

Figure 13: We envision future extension of our work retrofitting to a wide array of garments and jewelries, such as a cap.

We acknowledge that EmoGlass' hardware platform has much room to improve. First, the power conception can be optimized by implementing energy-aware scheduling (e.g., entering sleep mode when no environmental stimuli are detected). It is also possible to leverage solar cells to harvest energy from ambient lights. Additionally, more sensors other than cameras can be leveraged in the future, resulting in a multisensory system that could yield more accurate results. Moreover, EmoGlass could be transformed into a snap-on device that could retrofit existing jewelry or clothing (e.g., caps as shown in Figure 13). Finally, we can implement all computation on our embedded system and a smartphone with Bluetooth, without transmitting data to a server, to reduce users' concern of privacy.

Currently, we only rely on facial expression as the single indicator of emotion. In the future, we can make better use of the images that are captured, e.g., leveraging the images of the pupil to infer emotions that are otherwise hard to read from facial expressions [8, 9]. Moreover, synthesizing multiple indicators to more comprehensively sense emotion is also a problem worth discussing in further work.

While we make the assumption, we admit that facial expressions and emotions are not necessarily congruent. Some works in social psychology argue that the mapping from facial expressions to emotions is not necessarily universal, because how people communicate anger, disgust, etc, varies substantially across cultures, situations, and even across people within a single situation [11]. However, although the association may vary with culture and is loose enough to consist of many alternative accounts, the facial expressions and emotion labels are associated [73]. Some works [20] also claim that social and interpersonal scenes do affect the suppression and expression of emotion. However, people will not be in these situations all the time. In the future, we can change the dataset and algorithm to solve this problem because continuous facial expression tracking can distinguish the "true smile" and "fake smile" [5]. All in all, there are tons of shreds of evidence supporting the coherence between emotion and facial expression [72]. While the facial expressions are not equivalent to emotions, using them is still a suitable choice to stand for emotion. Additionally, other methods of emotion detection, which are introduced in our related work, entail complex equipment, such as physiological sensing, which can be easily disrupted by noise in the environment and user motion. In comparison, sensing facial expressions, which can be easily done with computer vision, is more practical in the real world.

Finally, we would like to investigate sensing and logging methods that can handle complex emotions – ones that contain multiple basic emotions. In fact, participants in our study reported complex emotions. We asked participants to describe their overall feelings for today before starting the study, and we received two main types of responses. One was well-defined basic emotions such as "happy" and "unhappy" while the other was more ambiguous, for example, "overwhelmed", "overloaded", "frustrated" and so on. These results indicated that complex emotions are an essential part of people's daily lives and thus important to monitor, which we plan to explore in future work.

9 CONCLUSION

In this paper, we present EmoGlass, the first end-to-end AI-enabled wearable platform that consists of a pair of facial expression detection eyeglasses and a mobile application to enable self-monitoring of emotion to promote emotional awareness. To improve EmoGlass' robustness in out-of-lab environments, we collected three datasets featuring slight changes of camera angle due to remounting, various lights conditions, and natural expressions rather than posed faces in a controlled lab space. We conducted a three-day out-of-lab study (N=15) to evaluate the performance of EmoGlass. We also iterated our application design based on user feedback in this study. We included some unique features in the application to help people better understand and regulate their emotions based on personal positive triggers. We present a quantitative analysis of our platform performance and qualitative findings based on participants' feedback. Finally, we discuss design recommendations for future work on sensing and enhancing awareness of emotional health. Overall, we believe that EmoGlass provides a powerful tool through which we can generate rich insights for the research community to leverage in using wearable sensing to address emotional health challenges.

ACKNOWLEDGMENTS

This work was funded in part by the National Science Foundation under grant IIS-1850183 and IIS-2047297.

REFERENCES

[1] Ralph Adolphs. 2002. Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and cognitive neuroscience reviews* 1, 1 (2002), 21–62.

[2] Md Shad Akhtar, Asif Ekbal, and Erik Cambria. 2020. How Intense Are You? Predicting Intensities of Emotions and Sentiments using Stacked Ensemble [Application Notes]. *IEEE Computational Intelligence Magazine* 15, 1 (2020), 64–75. https://doi.org/10.1109/MCI.2019.2954667

[3] Hiroaki Aoki, Ayumi Ohnishi, Naoya Isoyama, Tsutomu Terada, and Masahiko Tsukamoto. 2021. FaceRecGlasses: A Wearable System for Recognizing Self Facial Expressions Using Compact Wearable Cameras. In *Augmented Humans Conference 2021* (Rovaniemi, Finland) *(AHs'21)*. Association for Computing Machinery, New York, NY, USA, 55–65. https://doi.org/10.1145/3458709.3458983

[4] Mouath Aouayeb, Wassim Hamidouche, Catherine Soladié, Kidiyo Kpalma, and Renaud Séguier. 2021. Learning Vision Transformer with Squeeze and Excitation for Facial Expression Recognition. *CoRR* abs/2107.03107 (2021), 1–13. arXiv:2107.03107 https://arxiv.org/abs/2107.03107

[5] Akshay Asthana, Stefanos Zafeiriou, Shiyang Cheng, and Maja Pantic. 2014. Incremental face alignment in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, IEEE, 1859–1866.

[6] Anthony P Atkinson, Mary L Tunstall, and Winand H Dittrich. 2007. Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. *Cognition* 104, 1 (2007), 59–72.

[7] M Avalle, V Belotti, M Frascio, and R Razzoli. 2021. Development of a wearable device for the early diagnosis of neurodegenerative diseases. *IOP Conference Series: Materials Science and Engineering* 1038, 1 (feb 2021), 012033. https://doi.org/10.1088/1757-899x/1038/1/012033

[8] Areej Babiker, Ibrahima Faye, Kristin Prehn, and Aamir Malik. 2015. Machine learning to differentiate between positive and negative emotions using pupil diameter. *Frontiers in psychology* 6 (2015), 1921.

[9] Serdar Baltaci and Didem Gokcay. 2016. Stress detection in human–computer interaction: Fusion of pupil dilation and facial temperature features. *International Journal of Human–Computer Interaction* 32, 12 (2016), 956–966.

[10] Lisa Feldman Barrett. 2006. Valence is a basic building block of emotional life. *Journal of Research in Personality* 40, 1 (2006), 35–55.

[11] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M Martinez, and Seth D Pollak. 2019. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest* 20, 1 (2019), 1–68.

[12] Lisa Feldman Barrett, James Gross, Tamlin Conner Christensen, and Michael Benvenuto. 2001. Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition and Emotion* 15, 6 (2001), 713–724. https://doi.org/10.1080/02699930143000239 arXiv:https://doi.org/10.1080/02699930143000239

[13] Abdelkareem Bedri, Diana Li, Rushil Khurana, Kunal Bhuwalka, and Mayank Goel. 2020. FitByte: Automatic Diet Monitoring in Unconstrained Situations Using Multimodal Sensing on Eyeglasses. 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, 1–12. https://doi.org/10.1145/3313831.3376869

[14] M. Berking and P. Wupperman. 2012. Emotion regulation and mental health: recent findings, current challenges, and future directions. *Current Opinion in Psychiatry* 25, 2 (2012), 128.

[15] Taryn Bipat, Maarten Willem Bos, Rajan Vaish, and Andrés Monroy-Hernández. 2019. *Analyzing the Use of Camera Glasses in the Wild*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290605.3300651

[16] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–21.

[17] Tuochao Chen, Yaxuan Li, Songyun Tao, Hyunchul Lim, Mose Sakashita, Ruidong Zhang, Francois Guimbretiere, and Cheng Zhang. 2021. NeckFace: Continuously Tracking Full Facial Expressions on Neck-mounted Wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–31.

[18] Tuochao Chen, Benjamin Steeper, Kinan Alsheikh, Songyun Tao, François Guimbretière, and Cheng Zhang. 2020. C-Face: Continuously Reconstructing Facial Expressions by Deep Learning Contours of the Face with Ear-Mounted Miniature Cameras. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 112–125. https://doi.org/10.1145/3379337.3415879

[19] George Chernyshov, Kirill Ragozin, Benjamin Tag, and Kai Kunze. 2019. EOG Glasses: An Eyewear Platform for Cognitive and Social Interaction Assessments in the Wild. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (Taipei, Taiwan) *(MobileHCI '19)*. Association for Computing Machinery, New York, NY, USA, Article 67, 5 pages. https://doi.org/10.1145/3338286.3344418

[20] Elizabeth Chervonsky and Caroline Hunt. 2017. Suppression and expression of emotion in social and interpersonal outcomes: A meta-analysis. *Emotion* 17, 4 (2017), 669.

[21] Fernando De la Torre, Joan Campoy, Zara Ambadar, and Jeffrey F Cohn. 2007. Temporal segmentation of facial behavior. In *2007 IEEE 11th International Conference on Computer Vision*. IEEE, IEE, IEEE, 1–8.

[22] A. Dhall, R. Goecke, Simon Lucey, and T. Gedeon. 2011. Acted Facial Expressions In The Wild Database. *ANU Computer Science Technical Report Series* TR-CS, 11 (2011), 02.

[23] Fifth Edition et al. 2013. Diagnostic and statistical manual of mental disorders. *Am Psychiatric Assoc* 21 (2013), 992 pages.

[24] Paul Ekman. 1992. An argument for basic emotions. *Cognition and Emotion* 6, 3-4 (1992), 169–200. https://doi.org/10.1080/02699939208411068 arXiv:https://doi.org/10.1080/02699939208411068

[25] Mohamed Elgharib, Mallikarjun B. R., Ayush Tewari, Hyeongwoo Kim, Wentao Liu, Hans-Peter Seidel, and Christian Theobalt. 2019. EgoFace: Egocentric Face Performance Capture and Videorealistic Reenactment. *CoRR* abs/1905.10822 (2019), 10 pages. arXiv:1905.10822 http://arxiv.org/abs/1905.10822

[26] Everysight LTD. 2021. *Rapter - Everysight*. Everysight LTD. Accessed: 2021-09-07.

[27] Steven R Forness, Loretta A Serna, Elizabeth Nielsen, Katina Lambros, Mary Johnell Hale, and Kenneth A Kavale. 2000. A model for early detection and primary prevention of emotional or behavioral disorders. *Education and treatment of children* 23, 3 (2000), 325–345.

[28] Jon Froehlich, Tawanna Dillahunt, Predrag Klasnja, Jennifer Mankoff, Sunny Consolvo, Beverly Harrison, and James A. Landay. 2009. UbiGreen: Investigating a Mobile Tool for Tracking and Supporting Green Transportation Habits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1043–1052. https://doi.org/10.1145/1518701.1518861

[29] Jose Maria Garcia-Garcia, Victor M. R. Penichet, and Maria D. Lozano. 2017. Emotion Detection: A Technology Review. In *Proceedings of the XVIII International Conference on Human Computer Interaction* (Cancun, Mexico) *(Interacción '17)*. Association for Computing Machinery, New York, NY, USA, Article 8, 8 pages. https://doi.org/10.1145/3123818.3123852

[30] Anna Gruebler and Kenji Suzuki. 2014. Design of a wearable device for reading positive expressions from facial emg signals. *IEEE Transactions on affective computing* 5, 3 (2014), 227–237.

[31] Han Wen Guo, Yu Shun Huang, Jen Chien Chien, and Jiann Shing Shieh. 2015. Short-term analysis of heart rate variability for emotion recognition via a wearable ECG device. In *2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*. IEEE, IEEE, 262–265. https://doi.org/10.1109/ICIIBMS.2015.7439542

[32] Cheng He, Yun-jin Yao, and Xue-song Ye. 2017. An Emotion Recognition System Based on Physiological Signals Obtained by Wearable Sensors. In *Wearable Sensors and Robots*, Canjun Yang, G. S. Virk, and Huayong Yang (Eds.). Springer Singapore, Singapore, 15–25.

[33] Nazia Hossain and Mahmuda Nazin. 2018. EmoVoice: Finding My Mood from My Voice Signal. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (Singapore, Singapore) *(UbiComp '18)*. Association for Computing Machinery, New York, NY, USA, 1826–1828. https://doi.org/10.1145/3267305.3277832

[34] Jie Hu and Kenneth J Gruber. 2008. Positive and negative affect and health functioning indicators among older adults with chronic illnesses. *Issues in mental health nursing* 29, 8 (2008), 895–911.

[35] Qianyi Huang, Wei Wang, and Qian Zhang. 2017. Your glasses know your diet: Dietary monitoring using electromyography sensors. *IEEE Internet of Things Journal* 4, 3 (2017), 705–712.

[36] Kevin Hung, C. C. Lee, and Sheung-On Choy. 2015. *Ubiquitous Health Monitoring: Integration of Wearable Sensors, Novel Sensing Techniques, and Body Sensor Networks*. Springer International Publishing, Cham, 319–342. https://doi.org/10.1007/978-3-319-12817-7_15

[37] Iqram Hussain and Se Jin Park. 2020. HealthSOS: Real-Time Health Monitoring System for Stroke Prognostics. *IEEE Access* 8 (2020), 213574–213586.

[38] Yasha Iravantchi, Yang Zhang, Evi Bernitsas, Mayank Goel, and Chris Harrison. 2019. *Interferi: Gesture Sensing Using On-Body Acoustic Interferometry*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300506

[39] P Kalaiselvi and S Nithya. 2013. Face recognition system under varying lighting conditions. *IOSR Journal of Computer Engineering (IOSRJCE)* 14, 3 (2013), 79–88.

[40] Kai Kang and Xin Ma. 2019. Convolutional Gate Recurrent Unit for Video Facial Expression Recognition in the Wild. In *2019 Chinese Control Conference (CCC)*. IEEE, IEEE, 7623–7628. https://doi.org/10.23919/ChiCC.2019.8866235

[41] Maryam Karimi. 2021. A Review of Emotions Makes up The Bulk of Life. *International Journal of Advanced Studies in Humanities and Social Science* 10, 3 (2021), 156–160.

[42] Neelam Kaushik, Takashi Sasaki, Yoshiki Takahashi, Toru Nakazawa, and Kazuhiro Hane. 2017. MEMS-based retinal-imaging system for visual health monitoring. In *24th International Display Workshops, IDW 2017*. International Display Workshops, International Display Workshops, Sendai, Japan, 1319–1321.

[43] Neelam Kaushik, Takashi Sasaki, Yoshiki Takahashi, Toru Nakazawa, and Kazuhiro Hane. 2019. MEMS-based wearable eyeglasses for eye health monitoring. *Biomedical Physics & Engineering Express* 6, 1 (2019), 015006.

[44] Suleman Khan, M Hammad Javed, Ehtasham Ahmed, Syed AA Shah, and Syed Umaid Ali. 2019. Facial recognition using convolutional neural networks and implementation on smart glasses. In *2019 International Conference on Information Science and Communication Technology (ICISCT)*. IEEE, IEEE, IEEEE, 1–6.

[45] Dawon Kim and Yosoon Choi. 2021. Applications of Smart Glasses in Applied Sciences: A Systematic Review. *Applied Sciences* 11, 11 (2021), 4956.

[46] Jina Kim, Daeun Lee, and Eunil Park. 2021. Machine Learning for Mental Health in Social Media: Bibliometric Study. *J Med Internet Res* 23, 3 (2021), e24870.

[47] Kavitha Kolappa, David C Henderson, and Sandeep P Kishore. 2013. No physical health without mental health: lessons unlearned?

[48] Marcin Kopaczka, Raphael Kolk, and Dorit Merhof. 2018. A fully annotated thermal face database and its application for thermal facial expression recognition. In *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, IEEE, 1–6. https://doi.org/10.1109/I2MTC.2018.8409768

[49] Uros Krcadinac, Philippe Pasquier, Jelena Jovanovic, and Vladan Devedzic. 2013. Synesketch: An open source library for sentence-based emotion recognition. *IEEE Transactions on Affective Computing* 4, 3 (2013), 312–325.

[50] Hao Li, Laura Trutoiu, Kyle Olszewski, Lingyu Wei, Tristan Trutna, Pei-Lun Hsieh, Aaron Nicholls, and Chongyang Ma. 2015. Facial performance sensing head-mounted display. *ACM Transactions on Graphics (ToG)* 34, 4 (2015), 1–9.

[51] Jing Li, Yang Mi, Gongfa Li, and Zhaojie Ju. 2019. CNN-Based Facial Expression Recognition from Annotated RGB-D Images for Human–Robot Interaction. *International Journal of Humanoid Robotics* 16, 04 (2019), 1941002. https://doi.org/10.1142/S0219843619410020 arXiv:https://doi.org/10.1142/S0219843619410020

[52] Yong Li, Jiabei Zeng, Shiguang Shan, and Xilin Chen. 2018. Occlusion aware facial expression recognition using cnn with attention mechanism. *IEEE Transactions on Image Processing* 28, 5 (2018), 2439–2450.

[53] Wang Yan Li Ming, Peng Xiujiao. 2018. Facial Expression Recognition Based on Improved Dictionary Learning and Sparse Representation. *Journal of System Simulation* 30, 1, Article 28 (2018), 8 pages. https://doi.org/10.16182/j.issn1004731x.joss.201801004

[54] Sheng-Fu Liang, Chin-En Kuo, Yi-Chieh Lee, Wen-Chieh Lin, Yen-Chen Liu, Peng-Yu Chen, Fu-Yin Cherng, and Fu-Zen Shaw. 2015. Development of an EOG-based automatic sleep-monitoring eye mask. *IEEE Transactions on Instrumentation and Measurement* 64, 11 (2015), 2977–2985.

[55] Xi Lu, Xiaohang Liu, and Erik Stolterman Bergqvist. 2019. "It Sounds like She is Sad": Introducing a Biosensing Prototype That Transforms Emotions into Real-Time Music and Facilitates Social Interaction. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312784

[56] Guorong Ma, Chu Wang, Yanli Jia, Jiawei Wang, Bingren Zhang, Chanchan Shen, Hongying Fan, Bing Pan, and Wei Wang. 2018. Electrocardiographic and electrooculographic responses to external emotions and their transitions in bipolar i and ii disorders. *International journal of environmental research and public health* 15, 5 (2018), 884.

[57] Katsutoshi Masai, Yuta Sugiura, Masa Ogata, Kai Kunze, Masahiko Inami, and Maki Sugimoto. 2016. Facial Expression Recognition in Daily Life by Embedded Photo Reflective Sensors on Smart Eyewear. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (Sonoma, California, USA) *(IUI '16)*. Association for Computing Machinery, New York, NY, USA, 317–326. https://doi.org/10.1145/2856767.2856770

[58] Daniel McDuff, Abdelrahman Mahmoud, Mohammad Mavadati, May Amr, Jay Turcot, and Rana el Kaliouby. 2016. AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 3723–3726. https://doi.org/10.1145/2851581.2890247

[59] Susan Mineka, David Watson, and Lee Anna Clark. 1998. Comorbidity of anxiety and unipolar mood disorders. *Annual review of psychology* 49, 1 (1998), 377–412.

[60] Ali Mollahosseini, Behzad Hasani, and Mohammad H. Mahoor. 2019. AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. *IEEE Transactions on Affective Computing* 10, 1 (2019), 18–31. https://doi.org/10.1109/TAFFC.2017.2740923

[61] Nima Mousavi, Henrique Siqueira, Pablo Barros, Bruno Fernandes, and Stefan Wermter. 2016. Understanding how deep neural networks learn face expressions. In *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, IEEE, IEEE, 227–234.

[62] Thangaraj Munusamy, Ravindran Karuppiah, Nor Faizal A Bahuri, Sutharshan Sockalingam, Chun Yoong Cham, and Vicknes Waran. 2021. Telemedicine via smart glasses in critical care of the neurosurgical patient—COVID-19 pandemic preparedness and response in neurosurgery. *World Neurosurgery* 145 (2021),

e53–e60.

[63] Elizabeth L. Murnane, Xin Jiang, Anna Kong, Michelle Park, Weili Shi, Connor Soohoo, Luke Vink, Iris Xia, Xin Yu, John Yang-Sammataro, Grace Young, Jenny Zhi, Paula Moya, and James A. Landay. 2020. Designing Ambient Narrative-Based Interfaces to Reflect and Motivate Physical Activity. 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, 1–14. https://doi.org/10.1145/3313831.3376478

[64] Bahareh Nakisa, Mohammad Naim Rastgoo, Andry Rakotonirainy, Frederic Maire, and Vinod Chandran. 2018. Long Short Term Memory Hyperparameter Optimization for a Neural Network Based Emotion Recognition Framework. *IEEE Access* 6 (2018), 49325–49338. https://doi.org/10.1109/ACCESS.2018.2868361

[65] Grzegorz J. Nalepa, Krzysztof Kutt, Barbara Giżycka, Paweł Jemioło, and Szymon Bobek. 2019. Analysis and Use of the Emotional Context with Wearable Devices for Games and Intelligent Assistants. *Sensors* 19, 11 (2019), 24 pages. https://doi.org/10.3390/s19112509

[66] OmniVision Technologies Inc. 2009. 5 Megapixel image sensor, OV5647 datasheet, V1.0.

[67] World Health Organization et al. 2001. *Mental disorders affect one in four people.* World Health Report. Accessed: 2021-08-27.

[68] Christian Peter, Eric Ebert, and Helmut Beikirch. 2005. A Wearable Multi-sensor System for Mobile Acquisition of Emotion-Related Physiological Data. In *Affective Computing and Intelligent Interaction*, Jianhua Tao, Tieniu Tan, and Rosalind W. Picard (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 691–698.

[69] Christian Peter, Eric Ebert, and Helmut Beikirch. 2005. A Wearable Multi-sensor System for Mobile Acquisition of Emotion-Related Physiological Data. In *Affective Computing and Intelligent Interaction*, Jianhua Tao, Tieniu Tan, and Rosalind W. Picard (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 691–698.

[70] Carolyn R Pool. 1997. Up with emotional health. *Educational leadership* 54, 8 (1997), 12–14.

[71] Kanlaya Rattanyu, Michiko Ohkura, and Makoto Mizukawa. 2010. Emotion monitoring from physiological signals for service robots in the living space. In *ICCAS 2010*. IEEE, IEEE, 580–583. https://doi.org/10.1109/ICCAS.2010.5669914

[72] Rainer Reisenzein, Markus Studtmann, and Gernot Horstmann. 2013. Coherence between emotion and facial expression: Evidence from laboratory experiments. *Emotion Review* 5, 1 (2013), 16–23.

[73] James A Russell. 1994. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological bulletin* 115, 1 (1994), 102.

[74] Mohammad Nasser Saadatzi, Faezeh Tafazzoli, Karla Conn Welch, and James Henry Graham. 2016. EmotiGO: Bluetooth-enabled eyewear for unobtrusive physiology-based emotion recognition. In *2016 IEEE International Conference on Automation Science and Engineering (CASE)*. IEEE, IEEE, 903–909. https://doi.org/10.1109/COASE.2016.7743498

[75] Stanisław Saganowski, Anna Dutkowiak, Adam Dziadek, Maciej Dzieżyc, Joanna Komoszyńska, Weronika Michalska, Adam Polak, Michał Ujma, and Przemysław Kazienko. 2020. Emotion Recognition Using Wearables: A Systematic Literature Review - Work-in-progress. In *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, IEEE, 1–6. https://doi.org/10.1109/PerComWorkshops48775.2020.9156096

[76] Najmeh Samadiani, Guangyan Huang, Borui Cai, Wei Luo, Chi-Hung Chi, Yong Xiang, and Jing He. 2019. A Review on Automatic Facial Expression Recognition Systems Assisted by Multimodal Sensor Data. *Sensors* 19, 8 (2019), 27 pages. https://doi.org/10.3390/s19081863

[77] Anvita Saxena, Ashish Khanna, and Deepak Gupta. 2020. Emotion Recognition and Detection Methods: A Comprehensive Survey. *Journal of Artificial Intelligence and Systems* 2 (01 2020), 53–79. https://doi.org/10.33969/AIS.2020.21005

[78] Philip Schmidt, Attila Reiss, Robert Dürichen, and Kristof Van Laerhoven. 2019. Wearable-based affect recognition—A review. *Sensors* 19, 19 (2019), 4079.

[79] Philip Schmidt, Attila Reiss, Robert Dürichen, and Kristof Van Laerhoven. 2018. Labelling Affective States "in the Wild": Practical Guidelines and Lessons Learned. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (Singapore, Singapore) *(UbiComp '18)*. Association for Computing Machinery, New York, NY, USA, 654–659. https://doi.org/10.1145/3267305.3267551

[80] Andrea Sciarrone, Igor Bisio, Chiara Garibotto, Fabio Lavagetto, Mehrnaz Hamedani, Valeria Prada, Silvia Stara, Angelo Schenone, Federico Boero, and Gianluca Gambari. 2020. NeuroGlasses: A Wearable Prototype for Early Detection of Neurological Symptoms. In *2020 5th International Conference on Smart and Sustainable Technologies (SpliTech)*. IEEE, IEEE, 1–6. https://doi.org/10.23919/SpliTech49282.2020.9243713

[81] Solos Technology Limited. 2021. *Solos Smart Glasses: One for All Glasses for Smart People.* Solos Technology Limited. Accessed: 2021-09-07.

[82] Weinan Song, Yuan Liang, Jiawei Yang, Kun Wang, and Lei He. 2021. T-Net: Learning Feature Representation With Task-Specific Supervision For Biomedical Image Analysis. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, IEEE, 1924–1928. https://doi.org/10.1109/ISBI48211.2021.9433945

[83] Benjamin Tag, Jamie A Ward, Yuji Uema, and Kai Kunze. 2019. Eyewear 2019: Third Workshop on Eyewear Computing - Focus: Social Interactions. 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, 616–618. https://doi.org/10.1145/3341162.3347761

[84] MM Tugade, LF Barrett, and JJ Gross. 2006. Matters of feeling precisely: Emotional granularity and emotion regulation. *Unpublished manuscript* 0, 0 (2006), 0 pages.

[85] Md. Zia Uddin, Mohammed Mehedi Hassan, Ahmad Almogren, Mansour Zuair, Giancarlo Fortino, and Jim Torresen. 2017. A facial expression recognition system using robust face features from depth videos and deep learning. *Computers & Electrical Engineering* 63 (2017), 114–125. https://doi.org/10.1016/j.compeleceng.2017.04.019

[86] K Vijayalakshmi, S Uma, R Bhuvanya, and Annamalai Suresh. 2018. A demand for wearable devices in health care. *International Journal of Engineering and Technology* 7, 1.7 (2018), 01–04.

[87] Kun-Ching Wang. 2014. The Feature Extraction Based on Texture Image Information for Emotion Sensing in Speech. *Sensors* 14, 9 (2014), 16692–16714. https://doi.org/10.3390/s140916692

[88] Hao Wu, Jinghao Feng, Xuejin Tian, Edward Sun, Yunxin Liu, Bo Dong, Fengyuan Xu, and Sheng Zhong. 2020. EMO: Real-Time Emotion Recognition from Single-Eye Images for Resource-Constrained Eyewear Devices. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services* (Toronto, Ontario, Canada) *(MobiSys '20)*. Association for Computing Machinery, New York, NY, USA, 448–461. https://doi.org/10.1145/3386901.3388917

[89] Zhenbo Yu, Guangcan Liu, Qingshan Liu, and Jiankang Deng. 2018. Spatio-temporal convolutional features with nested LSTM for facial expression recognition. *Neurocomputing* 317 (2018), 50–57. https://doi.org/10.1016/j.neucom.2018.07.028

[90] Haris Zacharatos, Christos Gatzoulis, and Yiorgos L. Chrysanthou. 2014. Automatic Emotion Recognition Based on Body Movement Analysis: A Survey. *IEEE Computer Graphics and Applications* 34, 6 (2014), 35–45. https://doi.org/10.1109/MCG.2014.106

[91] Matthew D. Zeiler and Rob Fergus. 2014. Visualizing and Understanding Convolutional Networks. In *Computer Vision – ECCV 2014*, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, 818–833.

[92] Haoqi Zhang, Andrés Monroy-Hernández, Aaron Shaw, Sean Munson, Elizabeth Gerber, Benjamin Hill, Peter Kinnaird, Shelly Farnham, and Patrick Minder. 2014. WeDo: End-To-End Computer Supported Collective Action. *Proceedings of the International AAAI Conference on Web and Social Media* 8, 1 (May 2014), 639–642. https://ojs.aaai.org/index.php/ICWSM/article/view/14567

[93] Rui Zhang and Oliver Amft. 2017. Monitoring chewing and eating in free-living using smart eyeglasses. *IEEE journal of biomedical and health informatics* 22, 1 (2017), 23–32.

[94] Fangqing Zhengren, George Chernyshov, Dingding Zheng, and Kai Kunze. 2019. Cognitive Load Assessment from Facial Temperature Using Smart Eyewear. 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, 657–660. https://doi.org/10.1145/3341162.3348391

[95] Lei Zhong, Changmin Bai, Jianfeng Li, Tong Chen, and Shigang Li. 2020. Facial Expression Recognition Method Based on a Part-Based Temporal Convolutional Network with a Graph-Structured Representation. In *Artificial Neural Networks and Machine Learning – ICANN 2020*, Igor Farkaš, Paolo Masulli, and Stefan Wermter (Eds.). Springer International Publishing, Cham, 609–620.

[96] Tie Hua Zhou, Gong Liang Hu, and Ling Wang. 2019. Psychological disorder identifying method based on emotion perception over social networks. *International journal of environmental research and public health* 16, 6 (2019), 953.