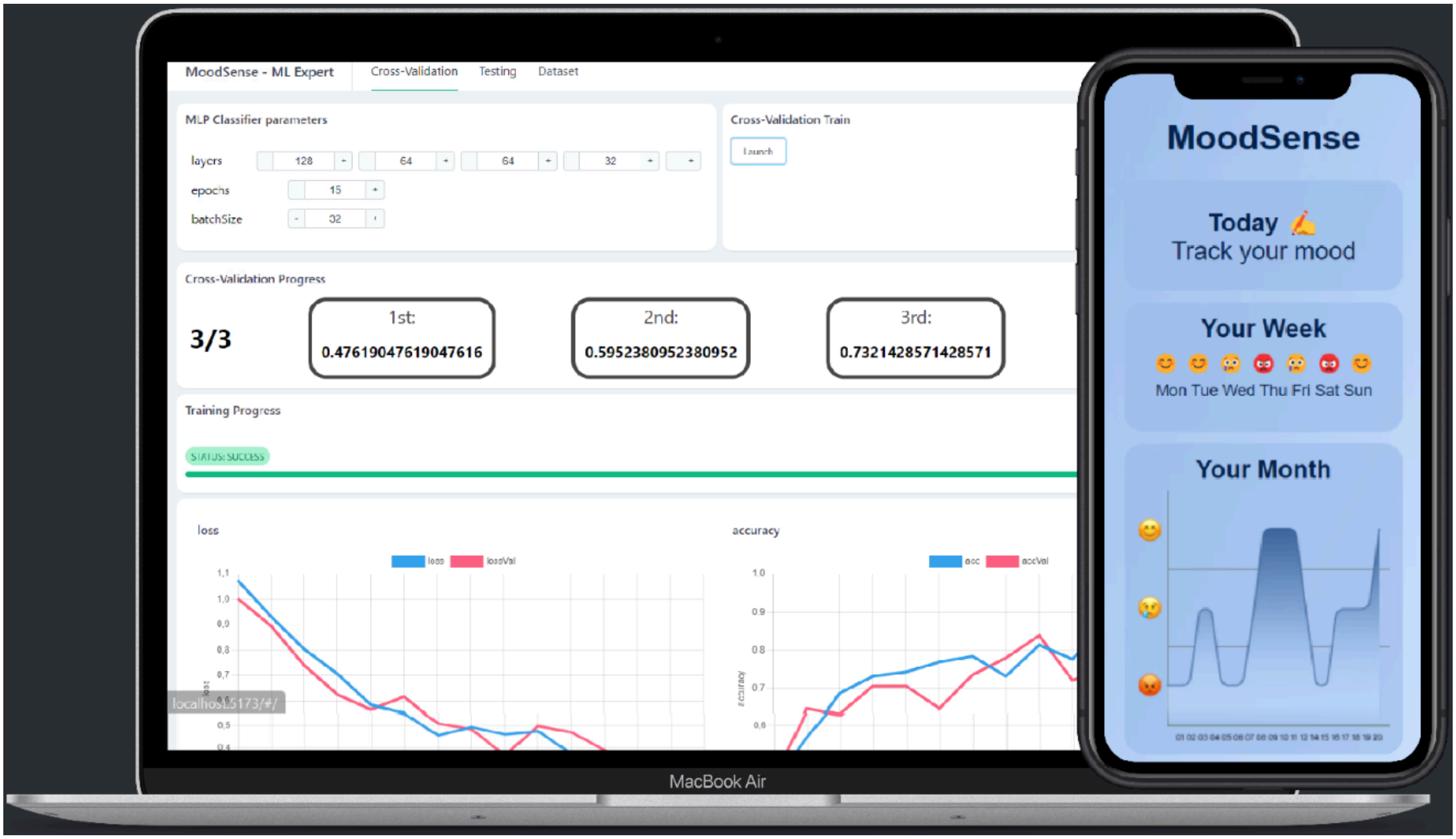
Interactive Machine Learning



Mood Tracking Using Machine Learning

Interactive Machine Learning

Fani-Marina Kalamara, Michelle Dutoit, Abel Henry-Lapassat

Introduction

Mood tracking has been widely studied as a pivotal factor in aiding psychotherapy and influencing the recognition of psychological phenomena, providing opportunities for more direct and personalized therapy. As the practice of mood tracking has undergone various advancements, particularly with the use of mobile platforms, there remains a gap regarding human-in-the-loop methodologies and the collaborative dynamics between experts and end-users. Recognizing the impact of mood on various psychological phenomena, and acknowledging the immense potential of Machine Learning in therapy-related scenarios, we introduce an interactive machine learning framework that incorporates an HCI perspective to involve Machine Learning experts, domain experts (therapists), and end-users (patients) into the mood tracking process for the assessment of psychological states. With this approach, we focus on the collaboration between the ML experts (who train and test the emotion recognition models), the therapists (who need to understand and enhance the ML predictions), and finally the end-users, who track their emotional states and are able to provide detailed qualitative descriptions about them. Our approach builds upon the existing AffectNet database, a large repository of facial expressions labeled with psychological terms related to emotions. Through a curation process, we refined the dataset to align with the emotional categories relevant to our scenario – happy, sad, and angry.

Literature Review

Mood tracking

Mood is an affective state that lasts anywhere from minutes to hours or even days and has a profound influence on various psychological phenomena such as cognition, memory, and stress. It is a fundamental aspect of human experience, impacting judgment, decision-making, thoughts, perception, health, and well-being. People may not be able to directly control their mood as it is both a physical and mental experience, and may be curious as to the causes of particular mood states. In recent years, mood tracking has emerged as a prominent area of interest within the quantified-self community, reflecting its significance in personal analytics and self-improvement endeavors [1].

Traditionally, mood assessment relied on pen-and-paper questionnaires or technological tools such as Ecological Momentary Assessment (EMA) for tracking and modeling human behavior [2]. However, with the widespread adoption of smartphones, researchers have increasingly turned to mobile platforms for measuring mood and emotion [3,4,5,6]. While self-reported measures remain standard practice, the utilization of smartphones enables researchers to explore implicit mood detection and automatic prediction, thereby overcoming limitations associated with traditional methods, such as user fatigue in EMA.

The effectiveness of smartphone-based mood measures is contingent upon their length, quality, and user experience. Longer measures may exacerbate user fatigue and compliance issues, impacting response quality and overall study outcomes. Consequently, researchers have sought to enhance the user interface (UI) and interaction design of mood-tracking apps, often favoring shorter measures to improve usability and user engagement.

Despite the benefits of shortened or modified measures, the validity and reliability of such measures must be validated to ensure the accuracy of mood data and subsequent predictive models. Failure to validate measures can yield arbitrary data, incorrect modeling, and unreliable predictions, undermining the utility of mood-tracking technologies [3].

Users of mood-tracking apps are primarily driven by negative life events or mental health shifts, aiming to understand and improve their emotional well-being. These apps facilitate

self-awareness and retrospection, allowing users to review past emotions and understand patterns. Users tend to document positive moods more than negative ones and prefer simple, personalized tracking interfaces with visualized data trends. However, existing apps lack interpretive guidance and intervention support, highlighting areas for improvement [7].

Facial expression recognition for mood tracking

According to previous studies, nonverbal cues account for two thirds of interpersonal communication. Among the many nonverbal components of human communication, facial expressions are an important source of information. This lends itself to using FER for mood tracking as FER can recognise the emotion and classify it for mood tracking [8]. According to Sariyanidi et al.'s study (2014), a number of elements, including accurate emotion recognition algorithms, efficient representation techniques, and precise face registration, enhance the efficacy of FER approaches. Previous work has already used FER for mood detection, including for well-being recognition systems to monitor stress and mood status in office environments. They utilized diverse data sources, including surveys and behavioral patterns, to classify well-being accurately. Employing machine learning algorithms such as SVM, they demonstrate promising accuracy levels, exceeding 83% in both generic and personalized models [10].

Scenario

In our scenario, we have a machine learning expert Lea, a psychologist Amelie and a patient named James. James got the recommendation of Amelie to track his mood to make it possible to find patterns in mood, and review the mood of James during the therapy session. To make it easier to track the moods an interactive machine learning application is used.

In the beginning the model is trained based on data from James. This data is collected in collaboration between Lea and James, where they take multiple pictures of him over a week. The model is then trained on this data and was previously trained with the AffectNet dataset [11,12] which can recognize emotions. WIth this a more accurate model is trained that can track James emotions.

During the application, the app reminds James to take a picture of him multiple times a day, a number that was set in the beginning and can be changed. Then after the picture was taken, or anytime that works for James he can review which mood was registered and can also refine it, if it was not correct. [This helps the model to become even more accurate].

In therapy sessions, Amelie and James can look at the mood tracker and review it together. In the UI of the therapist the system also shows the certainty of the result and if it is uncertain between two results, so the therapist can take a look at it.

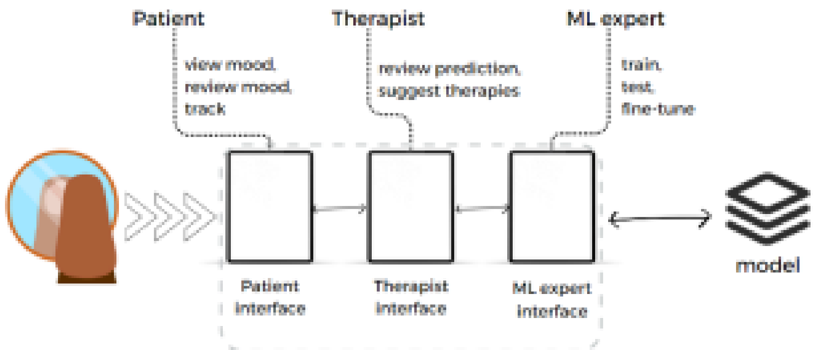
Dataset

The dataset employed in our study is AffectNet [11,12], a large repository of facial expressions comprising approximately 0.4 million images, each with a resolution of 96x96 pixels. AffectNet is labeled by “affects”, a psychological term denoting facial expressions, encompassing eight distinct emotion categories: neutral, happy, sad, angry, fear, surprise, disgust, contempt. The creation of the database involved querying 1250 emotion-related keywords in 6 different languages across 3 distinct search engines. The retrieved images were classified through 2 deep neural networks in the categorical model, and were subsequently annotated by expert human labelers, contributing to the dataset’s robustness. For our study, we focused on refining the dataset to enhance its specificity of emotion categories relevant to our scenario. Therefore, we selected 3 emotion categories: happy, sad, angry, aligning with our study’s scope. To curate

the dataset, each team member participated in the selection of the images that would be included in each set - specifically, each team member selected 50 images (per category) for the training set, and 25 images (per category) for the test set. This curation process was taken to optimize our dataset’s composition and accuracy in the training of our model, and to minimize biases during the selection.

Proposed Design

Our proposed design is a unified framework for interactive mood tracking (to facilitate the therapy of psychological phenomena), using machine learning. The IML platform will enhance end-user participation as well as collaboration between patients, therapists, and machine learning experts, capturing a multi-sided approach:



First, the overall framework features a smart mirror that will be present in the patient's personal space and will send real-time mood data to the patient's digital interface, facilitating a continuous feedback with human-in-the-loop. The smart mirror will assist in refining the model by adapting the captured images to the patient's unique characteristics.
The patient digital interface will daily and dynamically adjust, based on the captured images of the smart mirror. There will be an overview of the weekly moods of the patient, accompanied with visualizations. Users will engage with the platform to review the captured mood, being able to confirm the prediction, or correct it by personally tracking their emotional state with their camera and logging additional qualitative information (detailed descriptions) about their emotional states. In this way, the patients will actively participate in the ongoing model improvement, ensuring a holistic representation of their mental well-being. Additionally, each prediction will be presented to the patients with the associated percentages, allowing them to validate or refine the predictions.
The machine learning expert interface will further facilitate collaboration, by enabling the upload of instances, allowing for emotional label selection and saving the respective images to the corresponding dataset (train/test). The design will also feature a comprehensive dataset table view, accompanied by graphical representations and scatter plots for detailed analysis. When new images are added through the patient's interface, the machine learning expert will be able to audit the classification of the model and to make precise adjustments. Simultaneously, the machine learning expert will constantly train and fine-tune the model for personalized accuracy. During the crucial first week of personalization, where the model needs to adjust to the patient's unique characteristics, a cross-validation approach will be employed, allocating 2/3 of the data for training and 1/3 for testing, ensuring a solid foundation for the subsequent model refinement.
Finally, the therapist interface will provide a window to the model's decision making process, explaining how the model reached a specific decision. This feature will enable the therapists to

make corrections and collaborate with the ML expert to ensure transparency in the model's functioning. Simultaneously, there will be visualizations of the daily/weekly/monthly patient moods and psychological descriptions - from the interface's functionalities, the therapist will be able to suggest tailored solutions based on each patient's psychological state, as well as talk to them through dedicated channels (chat, video).

Implementation

ML Expert Interface

Initially, adaptation to Marcelle's components required meticulous attention to ensure a seamless interaction. We opted for the dashboard layout as it is the best suited for desktop usage. Subsequently, we crafted 3 pages to align with our proposed design, each describing a precise workflow and purpose:

1. Dataset Page

The objective here was to facilitate extensive data visualization and manipulation while ensuring user comprehension and swift interactions. The workflow begins with the user accessing the primary dataset (training set), where via a dataset table, he can remove or duplicate instances. Additionally, they have the option to add instances from an image upload. They can then choose to save these instances into the training set, the test set, or both. Further down the page, a scatter plot illustrates the data, allowing users to identify anomalies and relocate them from the table to the upload section to move them into the testing set.

2. Cross-Validation Page

This segment aims to enable users to train their models interactively, incorporating Cross-Validation training to align with our design. The workflow starts with the user reviewing the model's parameters (pre-optimized defaults from our testing). Subsequently, they initiate Cross-Validation, monitoring the evolving learning curves for each fold and evaluating the performance of the trained classifier via a confusion Matrix. Given our initial struggle with implementing Cross-Validation -especially how to merge the 3 trainings-, we sought assistance from Jules, who provided an existing example from which we derived the solution, tailored to our program's requirements.

3. Testing Page

This section proved to be the most time-consuming and challenging. Its purpose is to retrieve models from previous training and conduct tests to facilitate comparison. This can be achieved using either the aforementioned test dataset or a single uploaded image from any source. The workflow entails the ML Expert selecting a model from the list, loading it, and then choosing to activate the toggle widget for real-time predictions on individual instances, visualizing the results on a confidence plot, or leaving it inactive to view the confusion matrix of the loaded model on the test set. The primary hurdle encountered here was loading the saved model for testing and prediction, which was eventually overcome after considerable efforts. Pros of that it we had to deeper understand marcelle's functioning, which might serve for future improvements of our project.

Patient Interface

The patient interface consists of a mobile app and a Marcelle Wizard that allows interaction with the machine learning dataset. Due to project time constraints, the app interface was built using HTML to mimic a mobile app. It features a functional button opening the Marcelle Wizard for mood tracking, along with hard-coded elements providing weekly and monthly mood overviews, aiding users in identifying mood patterns.

The Wizard component streamlines the interface to avoid user confusion and seamless integration into the mobile app. User workflow within the wizard involves: 1. Viewing the latest ML model entry and its prediction.

2. Reviewing and potentially correcting entries, with the ability to modify labels and add mood descriptions. The mood description is planned to be stored within the app and accessible to both users and therapists.
3. The final page offers motivation for continued mood tracking and provides recommendations.
Some challenges that were encountered during implementation were to access the prediction accuracy of the last entry of the ML dataset, which was then left out. Another challenge was to add logic to the wizard component, so that the patient can choose to change the label of the data. This was bypassed by including it always, and allowing the user to add extra information so it is practical even if they don't want to change the label. Additionally, the performance of loading the last dataset entry is quite slow, as the algorithm loads the whole data set first and then takes the last instance. Another solution couldn't be found in the documentation, unfortunately.

Proposed Evaluation

Our proposed evaluation plan involves an initial expert study with machine learning experts and psychologists, in order to evaluate the User Interface and identify potential gaps in our solution. The key variables that will be assessed during the expert study will include the ease of use/usability, completeness and, most crucially, the overall understandability. The study will also encompass the use of a think aloud protocol, where participants will be encouraged to verbally express any thoughts, difficulties, surprises and problems with the design. Apart from that, there will be a post-task interview to further investigate the main challenges, areas difficult to comprehend, and missing elements of our UI.
Following the expert study, a pilot study with non-experts will be conducted, in order to assess the UI from the user side. For this study, there will be a set of tasks that users will have to complete. In the meantime, we will collect the number of mistakes, successes and a set of observations utilizing variables such as the ease of use/usability, completeness and understandability.
The final step will include a more extensive user study of 2 months, which will involve the users and the therapists. In this study, the evaluation will be conducted through interviews and diary studies, focusing on variables such as the ease of use with both the smart mirror and the digital user interface, the overall effectiveness and efficiency, the identification of mistakes, and some performance metrics. Overall, understandability remains a central variable in our proposed evaluation for the success of our interactive machine learning framework.

Link to repositories and videos:

Machine Learning Expert UI: https://github.com/Jvaljer/IML-MoodTracker Patient UI: https://github.com/michelledtt/iml_user_ui
Demo videos:
https://drive.google.com/drive/folders/1EZ9PjZgyT0Pm9b887OrjLyvSIYYdpPwO?usp=sharing 5

References

[1] Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. 2014. Understanding quantified-selfers' practices in collecting and exploring personal data. In Proceedings of

the SIGCHI Conference on Human Factors in Computing Systems (CHI ’14). Association for Computing Machinery, Toronto, Ontario, Canada, 1143–1152. https://doi.org/10.1145/2556288.2557372

[2] Saul Shiffman, Arthur A. Stone, and Michael R. Hufford. 2008. Ecological Momentary Assessment. Annual Review of Clinical Psychology
4,1(2008),1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

[3] Torkamaan, H., & Ziegler, J. (2020). Mobile mood tracking: An investigation of concise and adaptive measurement instruments. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(4), 1-30.

[4] Mehrab Bin Morshed, Koustuv Saha, Richard Li, Sidney K. D’Mello, Munmun De Choudhury, Gregory D. Abowd, and Thomas Plötz. 2019. Prediction of Mood Instability with Passive Sensing. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3,3(Sept.2019),75:1–75:21. https://doi.org/10.1145/3351233

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

[6] Xiao Zhang, Fuzhen Zhuang, Wenzhong Li, Haochao Ying, Hui Xiong, and Sanglu Lu. 2019. Inferring Mood Instability via Smartphone Sensing: A Multi-View Learning Approach. In Proceedings of the 27th ACM International Conference on Multimedia (MM ’19). Association for Computing Machinery, Nice, France, 1401–1409. https://doi.org/10.1145/3343031.3350957

[7] Schueller, S. M., Neary, M., Lai, J., & Epstein, D. A. (2021). Understanding people’s use of and perspectives on mood-tracking apps: interview study. JMIR mental health, 8(8), e29368.

[8] Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. *sensors*, *18*(2), 401.

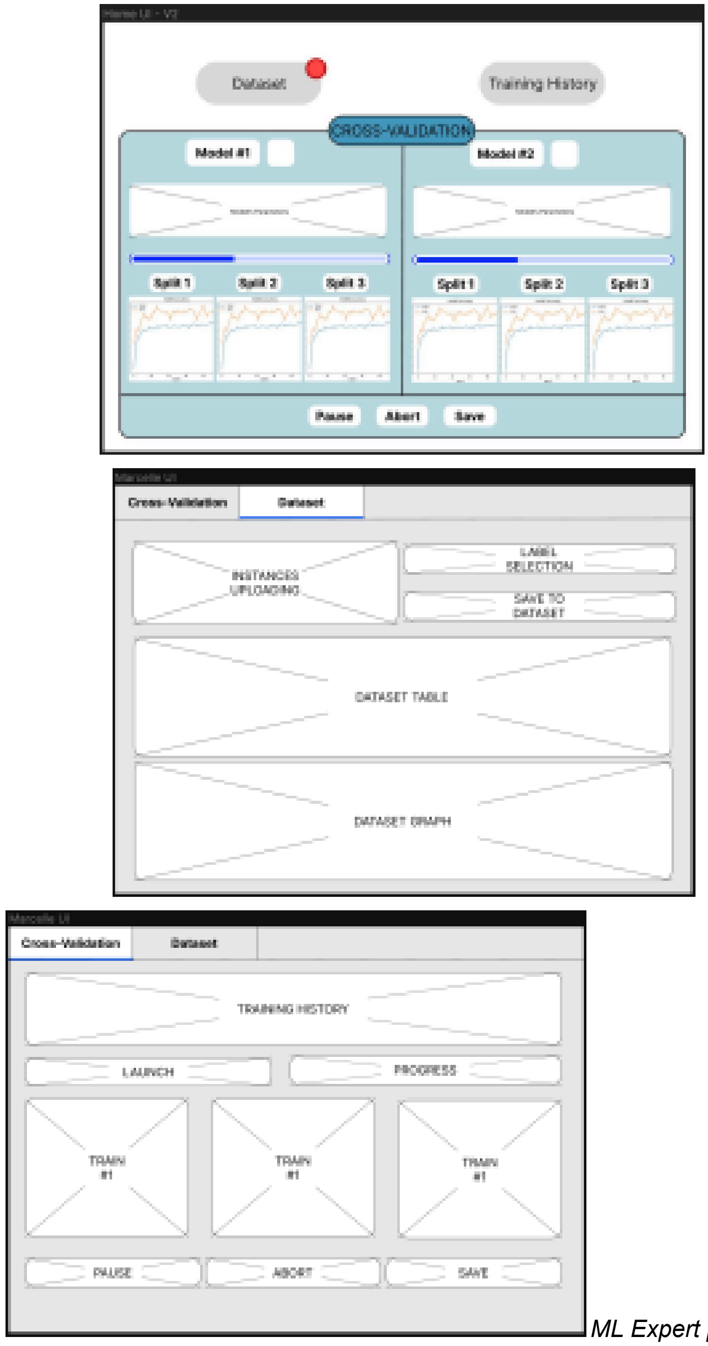
[9] Sariyanidi, E., Gunes, H., & Cavallaro, A. (2014). Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE transactions on pattern analysis and machine intelligence*, *37*(6), 1113-1133.

[10] Lawanot, W., Inoue, M., Yokemura, T., Mongkolnam, P., & Nukoolkit, C. (2019, January). Daily stress and mood recognition system using deep learning and fuzzy clustering for promoting better well-being. In *2019 IEEE International Conference on Consumer Electronics (ICCE)* (pp. 1-6). IEEE.

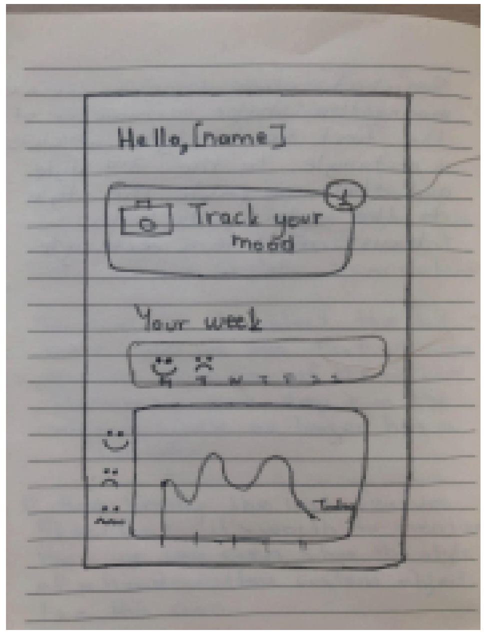
[11] AffectNet Training Data. Accessed February 2024. Retrieved from Kaggle: https://www.kaggle.com/datasets/noamsegal/affectnet-training-data

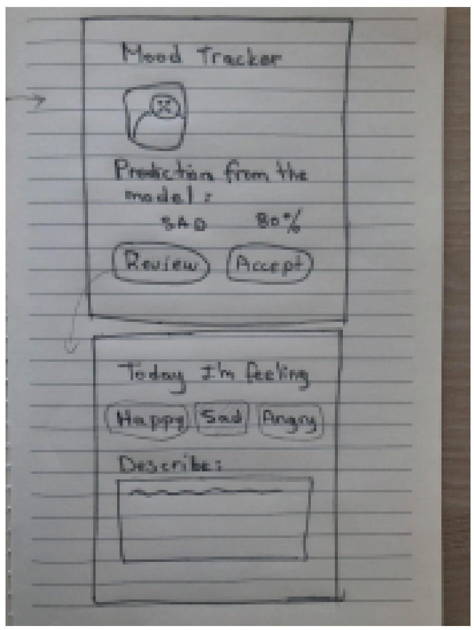
[12] Mollahosseini, A., Hasani, B., & Mahoor, M.H. (2017). AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. *IEEE Transactions on Affective Computing*, 10, 18-31.

Annex

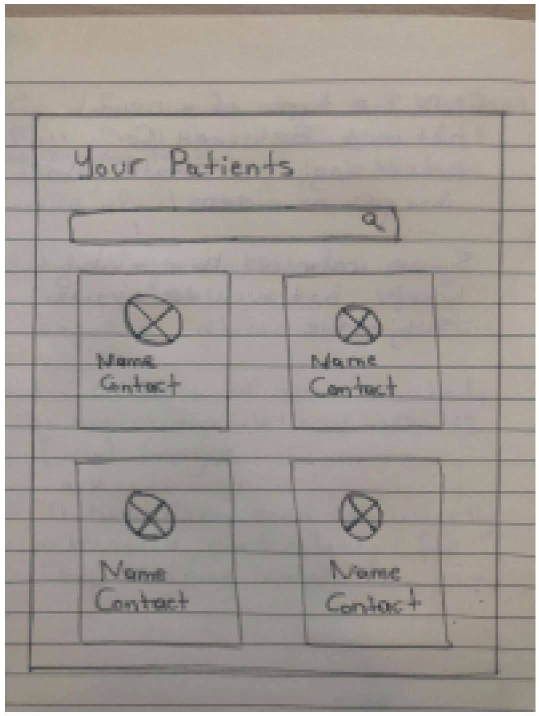


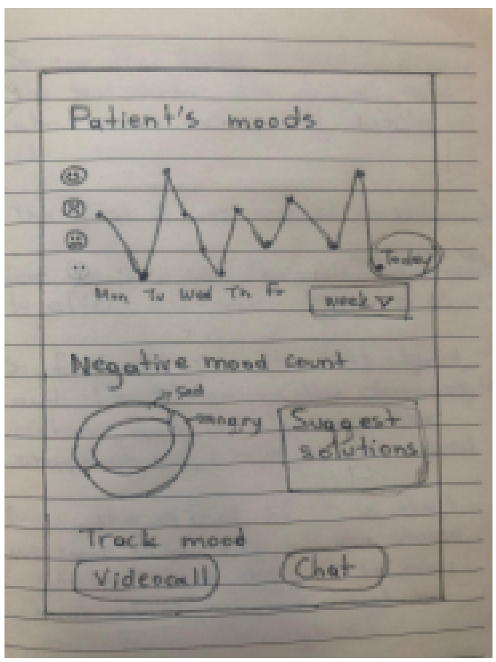
ML Expert prototype





Patient prototype





Therapist prototype