Moodify: Detecting and Regulating Emotions using Deep Learning

M. Hemdan*, J. Boulis*, A. Hassan*, M. Ayman*
* CSCE Department, The American University in Cairo, Cairo, Egypt

***Index Terms*—Emotion Regulation, Deep Learning, Circomplex Model, Speech Recognition.**

I. Introduction

Emotions shape our lives. It gives us the sense and experience of the world. They affect us on many different levels, as they are significantly related to our behaviors, thoughts, and even mental health status. They are often influential and dynamic. Therefore, they sometimes go well and other times lead to other difficulties. Difficulties often result from the mishandling of negative emotions, leading to mental health issues like depression and anxiety. The development of such mental health issues can give rise to severe risks. Committing suicide is one.

According to the National Institute of Mental Health, the suicide rate has increased by 35.2% from 1999 to 2018 in the US [1]. Globally, around 703 thousand people die every year from suicide, according to the WHO [2]. The percentage of suicide among teenagers is even higher. It is considered the fourth leading cause of death in people aged 15 to 19 years. Causes include stress, relationship break-up, financial problems. Once a person cannot deal with this, his emotions are triggered, and his attention is only focused on that. In such a case, suicide is probable.

With this high level of danger, there has to be a global movement and action to prevent this and give attention to this problem. Several actions have been done in the Middle East area to promote happiness and care about the mental health status of the region's people. Examples of these movements include the establishment of the UAE's ministry of happiness by the crown prince Mohammed bin Rachid in 2016 to enhance his people's mental well-being. In KSA, similar inclinations have emerged. In the same year, crown prince Mohammed bin Salman established the General Entertainment Authority (GEA) led by the former Saudi prime minister Turki Al-El-Sheikh. The emergence of these two significant initiatives at around the same time is not just a mere coincidence, the signs and signals were getting worse quickly, and the emergence of such entities was the natural response from visionary world leaders.

With the rise of technology and the ubiquitousness of mobile devices, users are continuously generating data like written posts, recordings, images, videos, and even more. These chunks of data provide some information about the user's status and mood. With the level of attachment mobile devices have reached with the user, the data is even more indicative of the user's mental health status. Harnessing such data could be helpful to detect early symptoms of depression and early mood swings, which helps in taking early action. These data can be analyzed using artificial intelligence and deep learning to extract useful information indicating a user's mood status.

After detecting the user's emotion, some action has to be done in case of negative emotion to regulate or mitigate it. In psychology, there is a topic of study called emotion regulation which researches the nature of emotions and the strategies and techniques to regulate it along with their adaptiveness and effectiveness, which will be more elaborated in the following section. Emotions can be regulated through some set of actions. Based on the emotion and current day's activities, the system provides recommendations to the user to boost his mood.

A more detailed description of the system will be described below. Section II will provide background about the topic of emotion regulation. Section III will talk about the representation of emotion and which psychological model has been used to represent emotion. Section IV is denoted for emotion detection and the data used for this. In section V, more description will be given about the activity recommendation system. Section VI talks about the overall system, data flow, and the tools used for developing the application. Section VII is devoted to the future plans. Section VIII explains the business plans.

II. Emotion Regulation Background

Emotion Regulation is the process, whether it is extrinsic or intrinsic, of monitoring, evaluating, and modifying the emotions and the reactions associated with them. Self-regulation is the process of regulating one's emotions by himself. Emotion regulation includes both self-regulation and regulating other people's emotions. Every day, people are facing new challenges and potentially arousing stimuli. Emotion dysregulation is the difficulty of controlling the influence of emotions, which many people experience. Studies have shown an association between emotion dysregulation and depression and anxiety [5] [6].

In order to understand the techniques of emotion regulation, the emotion generation process should be made clear. Psychologists have described emotion generation in the following sequence: Situation, Attention, Appraisal, and Response. Situation is the beginning of the process when a person is involved in an emotionally relevant situation. The situation can be imaginary. When the person gives attention to the emotional situation, this is the Attention stage. Interpreting the emotional situation and evaluating it is the Appraisal phase. Last is the Response stage is the emotional response to that situation. The emotional response can even cause a change to the situation, which repeats the process [8].

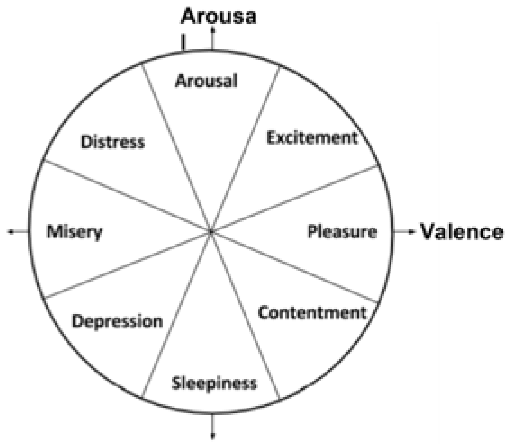

Fig. 1. The Circomplex Model

Several strategies have been developed to compact the influence of negative emotions. The strategies were developed for each stage of the emotion generation process separately. Psychologists often categorize them based on their timing with respect to the response time. Antecedent-focused techniques target the emotion before the response time, while the Response-focused techniques focus on the response after being generated.

Four main techniques correspond to each phase: Situation selection, Situation Modification, Attentional Deployment, and Response Modulation. Situation selection is the process of avoiding being involved in emotionally complicated situations from the beginning. Following this technique decreases the probability of experiencing the emotion (e.g., parents who removed their children from unpleasant situations). Situation Modification is the effort to modify the situation and its emotional impact by modifying the physical environment (e.g., injecting humor or extending a distance from the target person). Attentional Deployment is the direction of attention away from the situation or towards other content. It includes several techniques: Distraction, Rumination (passive attention towards thoughts, symptoms of distress), Worry (attention towards images in the mind of negative future), and Thought suppression. Response Modulation is the attempt to influence the physiological and behavioral response systems directly. It includes Drug use, exercise (used to mitigate the effects of the negative emotions), and sleep. Both Sleep and Exercise are adequate and well-studied Response Modulation techniques to mitigate negative emotions. Several psychotherapy techniques have been developed to implement these techniques, including CBT, EFT, MBCT, and DBT.

A. The process

Previous research has been done on Emotion Representation, and many measures have been proposed to measure emotion or effect like PANAS [12]. However, the problem with these measures is that they take much time as they are given in lengthy questionnaires, and they are often taken

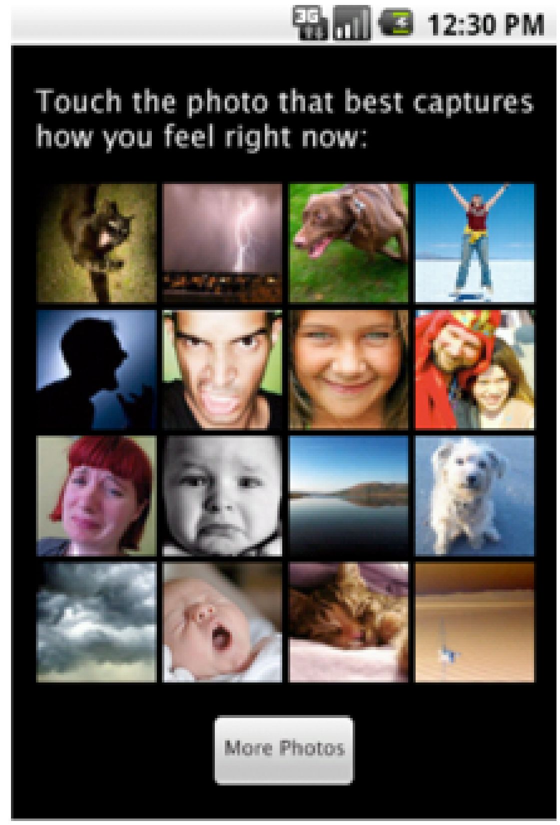

Fig. 2. The Pictures representing emotions in the PAM Model

High Arousal

Negative Valence	6	8	14	16	Positive Valence
	5	7	13	15	
	2	4	10	12	
	1	3	9	11	

Low Arousal

Fig. 3. Numbers corresponds to each picture from figure 2

at the beginning and end of the study. As the emotional states are continuously changing, the recall of the emotion is often distorted because of that. The literature proposed a solution to solve this issue: the Photographic Affect Meter or simply PAM [9], used in our Application. Emotion is often represented by two values: valence and Arousal. Valence measures the level of pleasure, and Arousal measures the level of activation. These values are best described by Russell's Circumplex Model of Affect [11] in which he showed the relationship between emotions and valence and Arousal and how valence and Arousal map to emotions, as shown in figure (1).

In the original study, the authors used certain pictures as emotion indicators so that users could choose the picture similar to their feelings without having to enter valence and arousal, as shown in figure (2). Then, the chosen pictures are converted into their respective emotions and then valence and arousal. The mapping between images and the circumplex model graph is done using figure (3), where 1 indicates the top left image, 2 the one on its right, and so on. For example, the first picture maps to the valence of -2 and arousal of -2.

The main problem with their approach is that valence and arousal are often discretized. The valence and arousal are measured discretely, either 2, 1, -1, -2. This might not give an accurate representation of the emotion.

III. Emotion Detection Module

Emotion Detection requires a representative measure of a user's emotional status throughout the day. Several methods have been used to detect emotions. Some use images, and others use voice to predict emotion. The main problem with image input is that it captures an instantaneous state. It does not measure emotion in the long term, throughout the day. To do that, the user has to take several pictures throughout the data to infer his emotion, which is very tedious. However, the problem is that people change their behavior/ emotion if they know they are being observed. This is called the Hawthorne effect [3]. Thus, the emotions extracted are not representative of the actual emotions of the person.

As a result, we tend to use a more representative value that can give the emotional state in the long run over the day. Voice is the best candidate for that. Voice can be extracted and analyzed using speech recognition and emotion detection models to infer emotions. The voice input can be used all over the day without the user's active participation. This makes it easy for the user. Although this might raise some privacy concerns, as the user's voice is being recorded, the speech recognition and emotion detection is done in real-time without the need to store any files, so no data is being transmitted outside the mobile phone. The detection is done locally. The model is trained in the cloud. For precise detection, it is downloaded to the app to do the detection locally.

The main problem that may face the voice input is the silence state. A user can interact with social media, post on Facebook and Twitter, and does not speak. In such a case, the model will not be able to detect the emotion of the person. A good solution is to rely on written information. The user writes posts on social media, texts some friends, or even takes notes. Such information may indicate something about his status. For example, a user can text his friend that he is not feeling well. In such a case, the app analyzes this text using sentiment analysis techniques to infer the user's emotion. Regarding privacy, these data are extracted and analyzed without saving and without transferring the data to the cloud; the model runs locally to get the emotion. A more clear data flow of the emotion detection module is illustrated in figure (4)

Regarding the Emotion detection module, we have used the ParallelDots API, a pre-trained emotion detection deep learning model implemented on the cloud through an API. It analyzes the text sent to it and sends back the emotions, represented as a set of probabilities. The resulting emotions are different from the ones used in the Circumplex model.

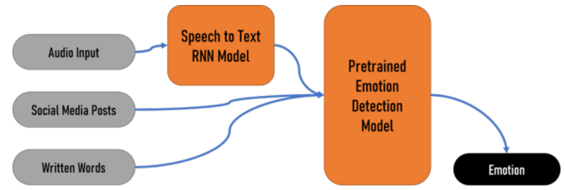

Fig. 4. Emotion Detection Module

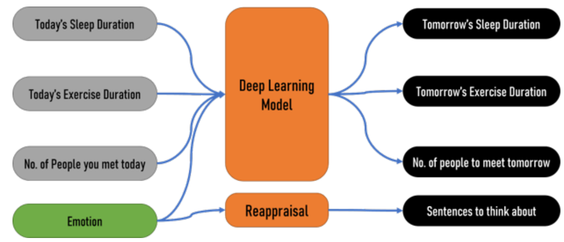

Fig. 5. Activity Recommendation Module

Thus, we search for these emotions' valence and arousal in the ANEW dictionary by Bradly [7]. For example, the emotion fear corresponds to 2.26 valence and 3.68 arousal. Both these values will be multiplied by the probability of fear from the first API, say 0.68. Two float values will arise, so we would have to map them into emotions to act as an input. Here comes the function of the second API, which takes continuous (float) valence and arousal values and converts them into emotions according to the diagram of the circumplex models. These APIs help us build our system to take continuous valence and arousal, which is more realistic, unlike the discrete PAM values used in the dataset.

IV. Activity Recommendation

A. Dataset Description

The dataset used for this project is the StudentLife dataset. This dataset has been collected at Dartmouth College, where 60 students participated in a ten week-study. The dataset contains the exercise and sleep durations, social interactions, and deadlines for each student.

B. Previous research

Previous research had been done on this dataset to provide activity recommendations for students. Paper [10] implemented a Reinforcement learning algorithm using Q-Learning to give the best recommendations to the user. The recommendations are the amount of time needed for exercise, sleep, and the number of people to socialize with. The reward is the emotion of the person. The reward is high when the person is happy, emotion towards high valence and arousal, and low when these values are approaching the negative axis.

C. Our System

Our system has used Deep Neural Networks for multi-label classification, as shown in figure (5). The DNN model then classifies among the exercises to pick (low, medium, high) and sleep duration periods. The model has been trained using the data of a converged RL model; hence, the recommendations should be somehow beneficial to the model since it increased the reward in the RL model. Regarding the emotion input, the main problem with Paper [10] is that the user enters the mood. There is no automatic detection of emotions. An emotion detection module is implemented in our system, and the output is provided to the activity recommendation module.

V. System Details

A. Implementation

First, a User Interface of the application is made on Figma with a prototype version for user interaction simulation. This User Interface is then converted into an actual application with a front end and a back end. The application's front end is made using React Native, while the APIs have been implemented using NodeJS. A database is implemented using MySQL to store users' information like psychiatric bookings and psychiatric available appointment times. A NoSQL database like MongoDB is implemented on the server to store chat. An API is implemented on the server to query from the database and send it back to the user's app or modify the database for new users or newly booked appointments. A login Authentication service is used for login and sign-up purposes from Google Firebase.

B. System Modules

The app system is divided into three main modules: the emotion detection module, activity recommendation module, and psychiatric help module.

1) Emotion Detection module: The emotion detection module converts the device's sound into text using the speech-to-text RNN model. This text is then concatenated with other text like keyboard text input and the user's Twitter posts, depending on the user's preferences. In the case of Twitter posts, the user has to provide the username and password to the Twitter API to allow the app to fetch posts from the user's account. In the case of the keyboard input, the user has to install an additional keyboard application. This keyboard sends the written text to the main app locally for analysis. The text from the three inputs is then concatenated and provided to the sentiment analysis deep learning model to get the emotion. The model is implemented inside the application, so no data is getting out of the application. However, for the time being, a pre-trained model from an online service has been used. This service provides an API to send text and receive an accurate measure of emotion. The API only provides a limited range of emotions with their probabilities. Conversion is then done to convert this into a representation that can work with the circumplex model, more explained in the emotion detection module section.

2) Activity Recommendation Module: The second module is the activity recommendation module which recommends specific activities to the user (e.g., recommended sleep duration, recommended exercise duration, and the number of people to socialize with) for the current day. The Activity Recommendation Deep Learning model is implemented using TensorFlow. It was trained on the StudentLife dataset. After that, the trained model's weights are then saved to be loaded again in the application source code using the TensorFlowJS library. The application can then call the model at any time to output the recommendation.

For the reappraisal part, some hard-coded statements are used to encourage the user to think about the problem depending on the emotion detected. For example, if the user is feeling sad, he properly remembers something in the past. The app then encourages the user to reframe this event in the user's mind to accept that or even think of even worse alternatives.

Other recommendations like getting outside, recommending locations to go outside, socializing events, or times of meditation are left as future plans to add them.

3) Psychiatric Help Module: The last module is the psychiatric module, where users interact with psychiatrists for help. The user can book an appointment with a psychiatrist for a fee. The user can chat with the psychiatrist via the application or call him in video or voice. Several services are needed for this: a chatting service, video streaming service.

VI. Future Plans

Several features are yet to be explored and implemented.

A. Federated Learning for better model accuracy

The main problem with the current implementation of the activity recommendation module is that deep learning often requires a large amount of data. The dataset does not provide a sufficient amount of data for training a good model. A good technique is to use the data generated from the users in the production phase to train a global deep learning model. This model is then distributed to the edge devices for training the model using their local data. After training, the weights of the local models are then concatenated together to update the global model using some cryptographic techniques, like the ones used in federated learning (e.g., Secure Aggregation).

B. Customized recommendations

We plan to further customize our recommendations in the future to achieve better personalization. For example, we have three output recommendations: the duration of sleep, the number of people to socialize with, and the number of hours to walk or exercise. We plan on building a database for each country so that the landmarks and attractions spots near the user are saved and recommended along with the hours to walk so that a user can walk there and enhance his/her mood. For example, one client might need to walk in a calm place with lots of green areas to improve his mood, while another needs to be in a crowded place. Such places cannot be recommended unless we have the database mentioned above and the client's location to recommend places based on his mood and needs. Also, we plan on building a database per user so that his/her favorite contacts are saved, and based on his/her mood and state; we might recommend him/her not only to chat to some people but to offer him/her names of people he/she likes but never connected with for a while now. On the contrary, the client's mood might be that he/she needs to connect with someone he has recently chatted with so. Instead, that person is also recommended, and for that, a database per client is required.

VII. Business Plan

According to a research report from UnivDatos Market Insights, the Mental Health Apps Market is expected to grow at a CAGR of 20.5% from 2021-2027 to reach US$ 3.3 billion by 2027 [4]. In the market, Mental Health Apps can be viewed as the distribution of mental healthcare facilities in computer applications. These apps have gained tremendous attention recently because of several reasons. First, with the breakout of the covid-19 pandemic, the digital markets have been suddenly boosted, where all activities need to be digitalized to reach the masses. Second, the pandemic's extreme conditions led to a decrease in human interactions and isolation, which is a problem that technology can play a significant role in solving. Hence, there is a real market need for such apps, and we are trying to satisfy such a need.

The global mental health app market is segmented on the type of platform, type of subscription, end-user, and geography. Regarding the subscription of the Moodify app, it follows a freemium business model, which means that we offer basic/limited features to the users at no cost and then charge a premium for supplemental or advanced features. To elaborate, users will access a set of limited recommendations such as sleeping, exercising, walking, and listening to songs. These recommendations, tailored to the user's mood, will act as an incentive to pay for the premium features. These features include having more personalized recommendations and more resources, such as chatting with a therapist at any time if the app indicates long periods of low mood.

Regarding the end-user of the Moodify app, it targets the emotional and mental health domain and is designed for people of all age groups and genders who are mentally healthy and want to:

- Find internal balance and become happier.
- Enhance their self-awareness by controlling and monitoring their mood swings, cultivate positive thinking, and break bad habits.
- Self-monitoring and self-improvement of their psychological state (including thoughts and feelings) while tracking progress.
- Explore their mood patterns and handle stress
- Control their feelings and thoughts, as well as anxiety, depression, stress, and sleep disorders.

Finally, regarding our targeted user's geographical area, we mainly focus on users in the MENA region since we tailor some recommendations like "go to an outdoor recreational place" to real places in the MENA region such as UAE, Egypt, and KSA.

References

1 https://www.nimh.nih.gov/health/statistics/suicide.

2 https://www.who.int/news-room/fact-sheets/detail/suicide.

3 https://catalogofbias.org/biases/hawthorne-effect/, Jan 2018.

4 Mental health apps market report, share amp; size, analysis 2021-2027, Aug 2021.

5 A. Aldao and S. Nolen-Hoeksema. Specificity of cognitive emotion regulation strategies: A transdiagnostic examination. *Behaviour Research and Therapy*, 48(10):974–983, 2010.

6 A. Aldao, S. Nolen-Hoeksema, and S. Schweizer. Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, 30(2):217–237, 2010.

7 M. M. Bradley, P. J. Lang, M. M. Bradley, and P. J. Lang. Affective norms for english words (anew): Instruction manual and affective ratings, 1999.

8 J. J. Gross. *Handbook of Emotion Regulation*. The Guilford Press, 2014.

9 J. P. Pollak, P. Adams, and G. Gay. Pam. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011.

10 J. A. Ramos Rojas, J. Rosas, Y. Shen, H. Jin, and A. K. Dey. Activity recommendation: Optimizing life in the long term. In *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–10, 2020.

11 J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

12 D. Watson, L. A. Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: The panas scales. *Journal of Personality and Social Psychology*, 54(6):1063–1070, 1988.