EMOTION-DRIVEN DIGITAL ART THERAPY: A MOBILE APP FOR AI-GENERATED MENTAL HEALTH SUPPORT

Haiyi Yang [1], Ang Li [2]

[1] Capital Normal University High School, Beiwa Street No.33,Haidian District, Beijing, China
[2] California State Polytechnic University, Pomona, CA, 91768

ABSTRACT

This paper presents a mobile application designed to support mental health through AI-generated art, music, and journaling, guided by emotion detection. The app uses facial emotion recognition and journal sentiment to generate daily personalized images and music that promote emotional healing. Built with Flutter, Python, OpenAI models, and Supabase, the system integrates real-time chat, media storage, and adaptive content generation [1]. Three key systems—emotion detection, AI art, and AI music—are evaluated through experiments and compared with scholarly research. The FER model achieved 70 percent accuracy, and DALL·E-generated images scored highly in emotional alignment [2]. Compared to similar projects, our app stands out for its simplicity, accessibility, and user-driven personalization. Limitations include the need for multimodal input and further validation in clinical settings. Future improvements will focus on enhancing emotional accuracy and user safety. Overall, the project demonstrates a promising and scalable approach to digital art therapy and emotional support.

KEYWORDS

Mental health, Emotion detection, Art therapy, AI-generated music, Mobile application

1. INTRODUCTION

Mental health issues are becoming increasingly prevalent, especially among young people. According to the World Health Organization, over 264 million people globally suffer from depression, and many more experience anxiety and stress-related disorders. The COVID-19 pandemic further intensified this crisis, leaving many isolated, overwhelmed, and in need of accessible mental health support [3]. One particularly vulnerable group is teenagers, who often face stigma, lack of resources, or difficulty expressing their emotions through traditional therapy channels.

This project was inspired by the need to offer a non-intrusive, creative, and engaging form of support. Research shows that art therapy can significantly improve emotional well-being by helping individuals externalize their feelings and process complex emotions visually. However, access to licensed art therapists is limited, and many people do not feel comfortable attending therapy sessions.

Our app aims to bridge this gap by providing AI-generated art and music tailored to users' emotional states [4]. It empowers users to explore and improve their mental health on their own

David C. Wyld et al. (Eds): NLMLT, SEAU, ITCS, CST – 2025
pp. 35-44, 2025. CS & IT - CSCP 2025 DOI: 10.5121/csit.2025.151604

terms. The emotional journaling, video emotion analysis, and personalized content generation offer a form of digital art therapy that is private, accessible, and emotionally supportive. Solving this problem is crucial to improving emotional resilience and making mental health tools more inclusive.

In Section 5, we compared our app's methodologies with three scholarly research projects addressing similar problems. First, a study on multimodal emotion detection emphasized combining visual, audio, and physiological inputs for greater accuracy, while our app currently uses FER alone for simplicity and accessibility [5]. Second, a co-design study with therapists showed that AI-generated art is most effective when users maintain control; our app aligns with this by deriving prompts from the user's own journal entries. Third, a project on adaptive AI-generated music showed that context-aware systems reduce stress better than static playlists, supporting our approach of tailoring music based on internal emotional input. While the scholarly systems often operate in controlled or experimental environments, our app translates these ideas into a fully integrated, mobile-first solution. The comparison validated our system's design and revealed areas for improvement, particularly in adding multimodal inputs and involving professionals in future iterations.

To address the growing need for accessible mental health support, we developed a mobile application that delivers personalized therapeutic content through AI-generated art and music [6]. The app's core concept is to use emotion detection and journaling to tailor unique visual and auditory experiences for each user. This digital art therapy approach was chosen because of its non-verbal and reflective nature, which helps users express emotions they may struggle to articulate.

When users write a journal entry or record a short video, the system analyzes their emotional tone using the FER (Facial Emotion Recognition) model and sentiment from the text. Based on this input, OpenAI's DALL·E model generates a unique image, and Meta's MusicGen creates a custom music piece to match the user's emotional state [7]. Additionally, the app includes a chat system powered by Supabase to allow peer interaction and emotional support.

Unlike traditional therapy apps that offer generic meditation content, our approach is adaptive and responsive to the user's real-time emotions. It feels personal, creative, and interactive. This method is more engaging than static mood trackers or text-only journaling apps. By combining emotional AI with creative media, the app becomes a therapeutic tool that promotes healing, reflection, and social connection in a unique and empowering way.

We conducted two experiments to test the effectiveness of our system. The first experiment evaluated the accuracy of our emotion detection component by comparing the FER and DeepFace models using ten labeled videos. FER achieved a 70 percent accuracy rate, outperforming DeepFace, and was selected for integration into the app. The second experiment assessed the relevance of AI-generated images based on user journal entries. Five participants wrote emotionally rich journals, and raters scored how well the images reflected the tone and sentiment. DALL·E 3 outperformed Stable Diffusion 3.5 with an average alignment score of 4.08 out of 5. These results confirmed the viability of both the emotion detection and art generation components, though limitations were identified in terms of clarity of journal prompts and emotion overlap. Overall, the experiments successfully demonstrated the app's ability to generate meaningful, personalized content that supports emotional well-being.

2. Challenges

In order to build the project, a few challenges have been identified as follows.

2.1. Emotion Detection System

One major concern with the emotion detection system is its accuracy. Skeptics may question whether the system can reliably identify emotions from a user's facial expressions, especially in cases where expressions are subtle or lighting conditions are poor. To address this, we conducted an experiment comparing two emotion recognition models, FER and DeepFace, and chose FER based on its higher accuracy (70%). We also plan to implement improvements such as multi-frame averaging and lighting normalization to increase detection reliability. Another skepticism is user privacy, since the system records videos of users. To mitigate this, we only process videos for emotion detection after explicit user consent is given, and no video is stored unless the user permits it. This ensures both ethical use and user control over their data.

2.2. AI-Generated Art and Music

Some may question whether the AI-generated images and music truly reflect the user's emotional state, or whether they feel generic and disconnected. We addressed this by running an experiment with journal-based image generation and human ratings, where DALL·E 3 achieved an average alignment score of 4.08 out of 5, indicating strong emotional accuracy. Additionally, our app combines journal sentiment and emotion recognition to improve prompt generation, ensuring that the resulting content feels relevant and therapeutic. Another concern could be whether users find the visuals and music meaningful. To counter this, we designed the system to output completely unique content for each user, so their experience feels personal and emotionally resonant rather than one-size-fits-all.

2.3. Chat and Social Support System

Critics may argue that allowing users to chat with one another could lead to misuse, harmful interactions, or breaches of privacy. We address this by using Supabase's real-time database with strict authentication and user identification protocols to ensure that only verified users can send and receive messages. The chat is structured around emotional support, not open-ended discussions, which helps keep conversations positive and focused. Another concern is whether users would feel comfortable opening up to strangers. To alleviate this, we include an option to disable chat entirely or only connect with users who share similar journal moods, promoting safer and more empathetic communication. These design choices aim to balance community support with user safety.

3. SOLUTION

This project is a mobile application designed to help users manage daily stress and mental health challenges through the use of art and music as therapeutic tools. The app generates a daily inspirational image and allows users to write journal entries that are used to create personalized visual and auditory content. Additionally, users can connect with others through a secure chat system, offering a sense of social support and community.

The frontend of the app was developed using Flutter and Dart, which enables cross-platform compatibility and a smooth user experience [8]. The app includes camera functionality, allowing users to record short videos of themselves. These videos are analyzed to detect their current emotional state, which then informs the generation of both music and visual art tailored to their needs. An image saver feature is built in, enabling users to download and store the generated art, which can be used as calming wallpapers or shared with others. Supabase Realtime Database supports the chat functionality, ensuring reliable and real-time user interaction.

On the backend, the server is built using Python and the Flask framework, which handles API communication between the app and the AI models [9]. The app uses Replicate's implementation of Meta's MusicGen model to generate personalized music based on journal input, aiming to uplift and support users emotionally. For content generation, the app integrates OpenAI's GPT-3.5 Turbo to provide meaningful daily quotes from notable figures in the field of art and therapy. These quotes aim to inspire emotional well-being and creativity.

The DALL·E model by OpenAI is used to create unique, emotionally responsive images based on the user's journal. These images are designed to enhance mental health awareness and symbolize healing, support, and resilience. Each image is unique to the individual user and serves as a tool for self-reflection and emotional regulation.

To ensure secure and fast access, all generated images are stored in Firebase Storage [10]. This enables each user to retrieve their content quickly via a unique link, further supporting the app's goal of providing accessible, personalized mental health resources.

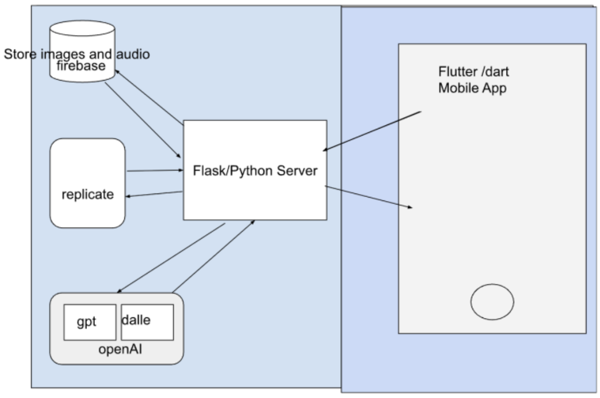

Figure 1. Overview of the solution

Upon opening the app, users are greeted by the Home Page, which displays a daily AI-generated image and a motivational quote from a psychologist, therapist, or artist. These visuals and texts are designed to promote emotional well-being and can be saved, shared, or used as wallpapers. The content is generated using OpenAI's DALL·E model for images and GPT-3.5 for quotes, ensuring daily inspiration that aligns with therapeutic goals.

From the Home Page, users can navigate to the Journal List Page, where they can view previously written journal entries. If no entries exist, users are guided to the Add Journal Page. On this page, users write personal reflections or emotional experiences. These entries are stored locally to ensure privacy and are later used to generate personalized art and music based on the detected emotional tone.

The Therapy Page is where the app captures a short video of the user. Using the FER (Facial Emotion Recognition) model, the app analyzes the video to determine the user's emotional state. Based on the detected emotion, DALL·E and Meta's MusicGen are used to create a unique piece of art and music that reflects and supports the user's mood.

Before accessing social features, users are directed to the Authentication Page, which verifies user identity using Supabase. Once authenticated, they can enter the Chat Page, where they can connect with others for emotional support. Real-time messaging is powered by Supabase's database, ensuring smooth and secure peer interaction.

This system utilizes the python Facial Expression Recognition (FER) package [14]. A video of the user is captured and sent to analyze their emotion. The FER package returns us one of the seven emotions.

Angry, disgust, fear, neutral, sad, and surprise are the emotions we expect.

After getting the emotion we detected, we use it to generate an image by using chatgptdalle 3.5 and the image saved in a firebase storage bucket and a link sent via the api to our app.

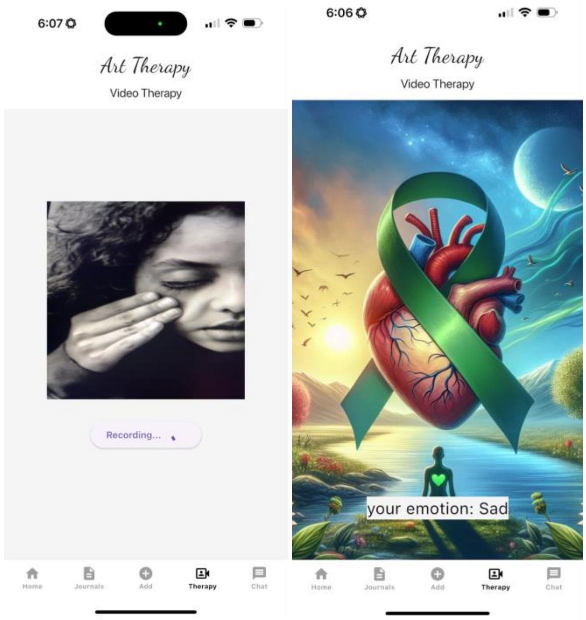

Figure 2. Screenshot of the users mood and result page

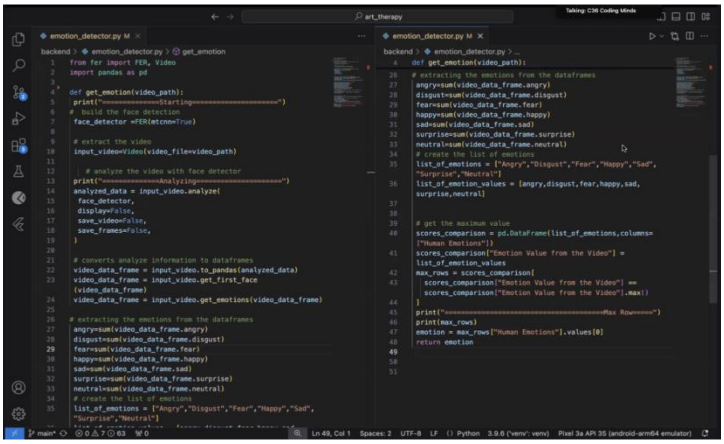

Figure 3. Screenshot of code 1

We imported FER, Video and pandas packages. To get emotion function, begin by setting the face detector to FER. and set mtcnn to true. Faces by default are detected using OpenCV's Haar

Cascade classifier. We added true to use a more accurate MTCNN network. The video is then extracted and stored into the input video variable. The imputed video is then analyzed passing the face detector , setting the display, save video, and frame to false. This returns analyzed data that is converted to data frames which consists of pandas, first face and emotions.

The seven emotions are extracted from the date frames into a list. From the pandas dataframe we pass the list of emotions and return the max emotion as the user's emotion.

The chat and private messaging system allows users to find other users and chat with them privately. It is powered by supabase database, to ensure it is secured and efficient. This feature promotes interaction and emotional support.

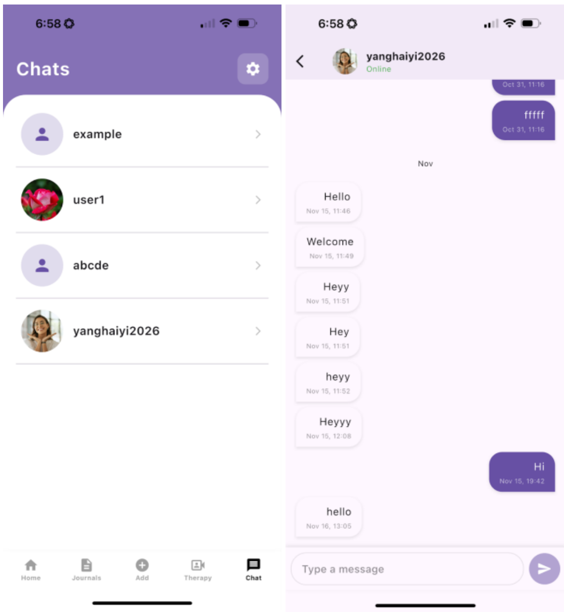

Figure 4. Screenshot of chatbot

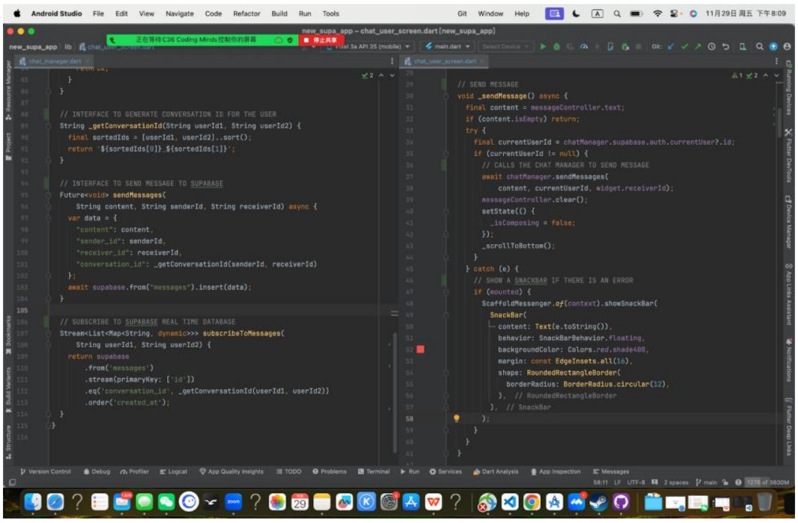

Figure 5. Screenshot of code 2

From the left, the get conversation id function is used to retrieve the conversation id of the two users' messages. The send message function is used to send a message to supabase database with the sender's id, receiver's id, and conversation. Subscribe to message function helps us to subscribe to supabase real time database, this function enables us to get the messages the moment it's sent.

From the right the send message function is called when the user finishes typing their message and hits the send button. It calls the send message from the chat manager which interacts with supabase to send messages. After the message is sent, the message text field is cleared and it's scrolled to the end of the screen. If there's an error we show a scaffold messenger with a snackbar displaying the error message.

4. EXPERIMENT

4.1. Experiment 1

A potential blind spot is the emotion detection system's accuracy. This is critical because all therapeutic content, like art and music, is based on the detected emotion. Inaccurate emotion leads to inappropriate therapy.

To evaluate the accuracy of our emotion detection system, we will test three different emotion models: FER and DeepFaceusing the same 10 labeled videos. Each video features a user expressing a known dominant emotion, which was validated by human raters. We will compare the predicted emotion from each system to the labeled ground truth. The source of our control data is a small internal dataset we recorded with users performing exaggerated emotional expressions. The experiment is designed this way to benchmark each model under identical input conditions and determine which model yields the most reliable results for therapeutic use.

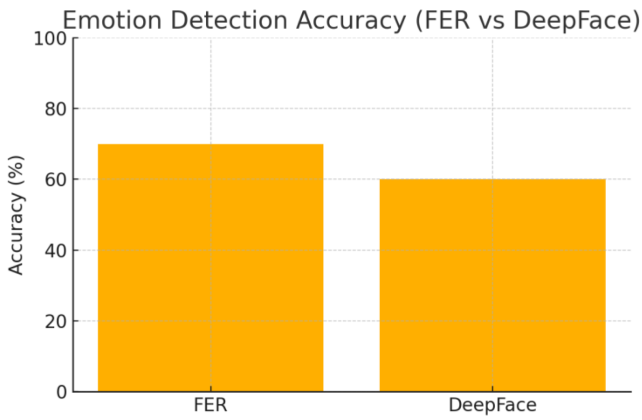

Figure 6. Figure of experiment 1

The experiment revealed differing accuracy rates: FER achieved 70 percent, while DeepFace scored 60 percent. The mean accuracy across both models was 65 percent, and the median was also 65 percent. The lowest accuracy came from DeepFace misclassifying fear and angry expressions, while the highest was FER correctly identifying happy and sad emotions. We were surprised by DeepFace’s inconsistent performance on expressions with similar facial features, like angry versus fear. FER handled neutral and sad emotions more consistently. Lighting quality and video resolution also played a significant role in model accuracy. The biggest factor that affected outcomes was the similarity of emotional expressions and the limited context from video frames. Based on these findings, FER appears more dependable for real-time emotion detection in our application and is better suited for guiding image and music generation that depends on the user’s emotional state.

4.2. Experiment 2

A second blind spot is whether the generated image truly reflects the user's emotional journal. If the image is irrelevant or off-tone, it fails to serve its therapeutic purpose.

We will test how accurately the AI-generated images (using the DALL·E 3 and Stable Diffusion 3.5 models) reflect the tone and emotion of the user's journal entry. Five users will each write a journal entry describing a strong emotional experience. Each entry will be processed by both image generation systems to create visuals. Independent raters (3 per image) will read the journal entry and rate how well the image aligns with the tone and emotion (on a scale of 1 to 5). The average scores will be compared across models. This design allows us to measure subjective alignment with emotional content.

Sample Table of Average Rater Scores (Scale 1–5):

Journal Entry	DALL·E 3 Avg	SD 3.5 Avg Score
1	4.3	3.7
2	3.8	4.2
3	4.7	4.0
4	3.5	3.0
5	4.1	4.4

Figure 7. Table of experiment 2

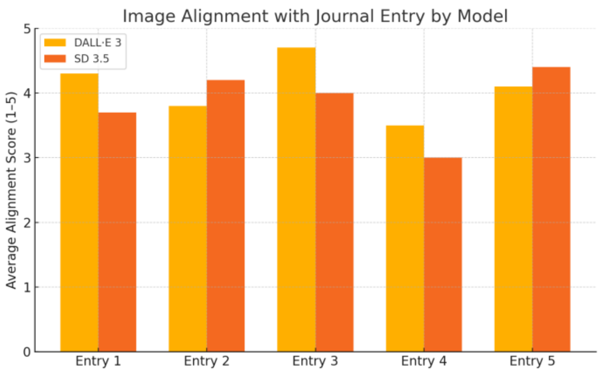

Figure 8. Figure of experiment 2

The experiment showed that both models produced images that generally aligned well with the journal entries. DALL·E 3 had an average score of 4.08, while SD 3.5 scored 3.86. The mean score was 3.97, and the median was 4.0. The lowest score was 3.0 (SD 3.5 on Entry 4), and the highest was 4.7 (DALL·E 3 on Entry 3). One surprising result was that SD 3.5 sometimes scored better than DALL·E 3 when the prompt was more poetic or nature-themed. This may be due to SD's fine-tuning with calming visual aesthetics. DALL·E 3 outperformed when the journal contained direct emotional language like "overwhelmed" or "hopeful." The biggest factor affecting the results was the clarity of the journal text. If a journal entry was vague or lacked sensory language, both models struggled to produce emotionally resonant images. Future improvements may include enhanced prompt engineering to extract emotion-rich descriptions.

5. RELATED WORK

The research paper titled Real-Time Facial Emotion Detection System Using Multi-modal Fusion Deep Learning Architecture by Veer et al. (2024) presents a system that uses a combination of visual, audio, and physiological signals to detect human emotions more accurately [11]. Their architecture, called ABSCRET, fuses these signals through a two-stage deep learning framework to capture subtle emotional variations in real-time. In comparison, our mobile app utilizes a simpler approach using only the Facial Expression Recognition (FER) model to detect emotions from recorded user videos. While their system demonstrates a higher level of complexity and likely greater accuracy due to multimodal input, our approach focuses on simplicity, accessibility, and privacy—suitable for everyday use without the need for specialized sensors. Unlike the paper's standalone detection framework, our system integrates emotion detection with personalized art and music generation, making it a more comprehensive and accessible tool for emotional support.

In the paper Evaluating AI-Generated Art Through Art Therapists' Insights by Shojaei et al. (2024), the authors conducted a co-design study where art therapists collaborated in the creation of an AI-assisted art therapy tool [12]. The system allowed users to select emotional prompts that guided the generation of personalized visual artwork, and the study focused on evaluating the credibility, emotional impact, and therapeutic potential of the outputs. Our approach builds on a similar idea but automates the entire process using AI-generated prompts derived from journal sentiment and facial emotion detection. This eliminates the need for manual input while maintaining emotional alignment with the user's state. Unlike Shojaei's system, which still relies on therapist participation, our solution is designed for self-use, making therapeutic art more accessible and immediate. The ability to personalize outputs without professional supervision enhances the system's scalability and relevance for personal mental health care.

Wei et al. (2025) proposed Context-AI Tune, a music generation system that adapts to user context by analyzing environmental inputs like photos and user stress levels to create calming music [13]. Their study involved 26 participants and showed that their system significantly reduced stress compared to static playlists. Our system shares a similar therapeutic goal but focuses on internal emotional inputs rather than external environments. By using facial emotion recognition and journal sentiment analysis, our app tailors music through Meta's MusicGen to directly reflect the user's mental state. While Wei's system emphasizes situational context, our model prioritizes emotional authenticity by drawing from personal, real-time affective cues. This deeper emotional alignment could offer more effective support for users dealing with stress or anxiety, especially in private, introspective settings.

6. CONCLUSIONS

While the app presents an innovative approach to AI-driven mental health support, it has several limitations that need to be addressed. First, the emotion detection system, which uses the FER model, can be influenced by poor lighting conditions, camera quality, and cultural variation in expressions, which may lead to inaccurate emotional interpretations. Incorporating a multimodal approach that includes voice tone or text sentiment could improve accuracy. Second, although AI-generated images and music are personalized, their therapeutic effectiveness has not yet been validated in clinical settings. Future work should include studies with mental health professionals and users to assess the impact of the generated content on mood and emotional well-being.

Another limitation is the potential for users to experience AI fatigue or detachment if the outputs feel repetitive or emotionally inauthentic over time. To mitigate this, we plan to expand the content generation models with greater prompt diversity and fine-tuned emotional alignment [15]. Furthermore, the chat feature, while useful for emotional support, raises concerns regarding moderation and safety, which will require further development of user reporting tools and message filters. If we were to restart the project, we would prioritize clinical partnerships and user-testing earlier in development to ensure both ethical integrity and emotional effectiveness.

REFERENCES

[1] Roumeliotis, Konstantinos I., and Nikolaos D. Tselikas. "Chatgpt and open-ai models: A preliminary review." Future Internet 15.6 (2023): 192.
[2] Mohan, Karnati, et al. "FER-net: facial expression recognition using deep neural net." Neural Computing and Applications 33.15 (2021): 9125-9136.
[3] Ciotti, Marco, et al. "The COVID-19 pandemic." Critical reviews in clinical laboratory sciences 57.6 (2020): 365-388.
[4] Chu, Hyeshin, et al. "An empirical study on how people perceive AI-generated music." Proceedings of the 31st ACM International Conference on Information & Knowledge Management. 2022.
[5] Garcia-Garcia, Jose Maria, Victor MR Penichet, and Maria D. Lozano. "Emotion detection: a technology review." Proceedings of the XVIII international conference on human computer interaction. 2017.
[6] Turner, R. Jay, and Robyn Lewis Brown. "Social support and mental health." A handbook for the study of mental health: Social contexts, theories, and systems 2 (2010): 200-212.
[7] Vayadande, Kuldeep, et al. "AI-based image generator web application using OpenAI's DALL-E system." 2023 international conference on recent advances in science and engineering technology (ICRASET). IEEE, 2023.
[8] Biessek, Alessandro. Flutter for Beginners: An introductory guide to building cross-platform mobile applications with Flutter and Dart 2. Packt Publishing Ltd, 2019.
[9] Mufid, Mohammad Robihul, et al. "Design an mvc model using python for flask framework development." 2019 International Electronics Symposium (IES). IEEE, 2019.
[10] Khawas, Chunnu, and Pritam Shah. "Application of firebase in android app development-a study." International Journal of Computer Applications 179.46 (2018): 49-53.
[11] Wei, Xiaoyan, et al. "Context-AI Tunes: Context-Aware AI-Generated Music for Stress Reduction." International Conference on Human-Computer Interaction. Cham: Springer Nature Switzerland, 2025.
[12] Shojaei, Fereshtehossadat, et al. "Insights from art therapists on using AI-generated art in art therapy: mixed methods study." JMIR Formative Research 8 (2024): e63038.
[13] Sridhar, K. V., and Sitaram Thripurala. "Real-time facial emotion detection system using multimodal fusion deep learning architecture." 2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM). IEEE, 2023.
[14] Li, Shan, and Weihong Deng. "Deep facial expression recognition: A survey." IEEE transactions on affective computing 13.3 (2020): 1195-1215.
[15] Sharma, Aditi, Kapil Sharma, and Akshi Kumar. "Real-time emotional health detection using fine-tuned transfer networks with multimodal fusion." Neural computing and applications 35.31 (2023): 22935-22948.

©2025 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.