A PERSONALIZED MENTAL HEALTH SUPPORT SYSTEM USING AI-DRIVEN FACIAL EXPRESSION CLASSIFICATION AND REAL-TIME IMAGE GENERATION

Zeyu Zhang[1], Yu Sun[2]

[1]Santa Margarita Catholic High School, 22062 Antonio Pkwy, Rancho Santa Margarita, CA 92688
[2]Computer Science Department, California State Polytechnic University, Pomona, CA 91768

ABSTRACT

This research paper presents the development and evaluation of a personalized mental health support application that leverages AI-driven features for real-time user interaction [1]. The application includes components for facial expression classification and real-time image generation, both of which were subjected to rigorous testing through targeted experiments [2]. The first experiment evaluated the accuracy of the emotion recognition system, revealing strong performance with distinct emotions but highlighting challenges with subtle expressions. The second experiment tested the responsiveness of the image generation component, showing effective performance with simple inputs but identifying delays with more complex tasks. While the application demonstrates significant potential, especially in its ability to provide tailored emotional feedback and support, further refinement is needed to enhance accuracy, performance, and data security. The findings suggest that with continued development, this application could become a valuable tool in the field of mental health and emotional well-being.

KEYWORDS

Facial Expression Classification, AI-Driven Mental Health, Real-Time Image Generation, Emotion Recognition, Emotional Well-Being

1. INTRODUCTION

Mental health is a growing concern worldwide, affecting millions of individuals across various demographics [3]. With the rise of technology, the potential to use digital tools for mental health monitoring has become increasingly promising. However, despite the availability of numerous mental health applications, the current tools often fail to provide real-time, accurate assessments that can lead to timely interventions. The primary challenge lies in the lack of predictive models that can analyze and interpret behavioral data to forecast potential mental health issues before they become critical. According to the World Health Organization (WHO), depression and anxiety disorders alone cost the global economy $1 trillion each year in lost productivity [4]. Additionally, the pandemic has exacerbated mental health conditions, further stressing the need for an effective monitoring system. Given these alarming statistics, there is an urgent need to develop a more sophisticated system that can predict and monitor mental health status, offering timely interventions and support to those in need.

David C. Wyld et al. (Eds): NWCOM, ASOFT, ITCCMA, CSITY, BDIoT, MLNLP, SIGPRO, AIFZ– 2024
pp. 177-186, 2024. - CS & IT - CSCP 2024 DOI: 10.5121/csit.2024.141916

The methodologies explored offer different approaches to enhancing AI-driven mental health support systems [5]. Fei et al. (2020) employed deep convolutional networks for facial expression analysis, achieving high accuracy but encountering challenges in adapting to real-world, unstructured data. Muhammad et al. (2017) developed a facial-expression monitoring system with impressive accuracy in controlled conditions, but it struggled with dynamic, real-time environments. Roy et al. (2022) introduced a neuro-symbolic AI approach that improved explainability and reliability but required significant computational resources. Our project builds on these methodologies by integrating real-time data processing, enhancing adaptability, and streamlining the system for broader application, thereby addressing the limitations identified in these existing approaches.

To address the growing mental health crisis, this paper proposes the development of a predictive mental health monitoring system utilizing machine learning and artificial intelligence (AI) [6]. The system will analyze a range of data inputs, including behavioral patterns, social interactions, and physiological data, to provide a real-time assessment of an individual's mental health status. By leveraging AI's ability to process and learn from large datasets, the proposed system aims to predict potential mental health issues before they escalate. Unlike existing tools that focus primarily on self-reported symptoms, this system will integrate passive data collection and advanced analytics to deliver a more accurate and proactive approach. The proposed solution is expected to outperform traditional methods by offering personalized insights and recommendations, thus enabling timely and effective interventions. The integration of machine learning algorithms allows for continuous improvement of the system's predictive accuracy, making it a dynamic and robust tool for mental health professionals and users alike.

In the course of this research, two critical experiments were conducted to evaluate key components of the application. The first experiment focused on assessing the accuracy of the facial expression classification system. Participants were recorded while expressing a range of predefined emotions, and the system's ability to accurately classify these emotions was tested. The results indicated that while the system performed well with distinct emotions like happiness and neutrality, it struggled with more subtle expressions such as fear and disgust, revealing areas for potential improvement in the underlying algorithms [7].

The second experiment aimed to evaluate the performance of the real-time image generation component. This experiment measured the time taken to generate images based on varying levels of input complexity, under both normal and high-load conditions. The findings showed that while the system was effective in generating simple and moderately complex images quickly, it experienced delays with more complex inputs. This highlighted the need for further optimization of the image generation algorithms to maintain responsiveness, particularly when handling multiple requests simultaneously.

2. CHALLENGES

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

2.1. Implementing Real-Time Data Processing

A major challenge anticipated in the development of this app is implementing real-time data processing, particularly for tasks such as emotion detection and image generation. The system needs to analyze large amounts of data, such as video frames, on the fly to produce accurate and timely results. This requires not only highly optimized algorithms but also the ability to efficiently utilize available hardware resources to prevent delays. To address this, we could focus on optimizing the code and using techniques that allow the app to handle tasks simultaneously, ensuring smooth and fast performance.

2.2. Ensuring Seamless Data Synchronization

Another critical challenge is ensuring seamless data synchronization across various components of the app, which may include emotion detection, image generation, and server communication. These components must interact smoothly to deliver a cohesive and integrated user experience. Any lag or data mismatch between these modules could result in inaccuracies or poor performance. To mitigate this risk, it may be necessary to design a robust communication protocol that facilitates reliable data exchange between components, along with implementing error handling and recovery mechanisms to address potential issues in real-time.

2.3. Creating a server

Another important challenge is creating a server that can grow with the app and keep user data safe. As the app might have many users and handle sensitive information, the server needs to be both powerful and secure. To manage this, we could use cloud services that automatically adjust to handle more users when needed. Additionally, strong security measures, like encrypting data and using secure logins, would be essential to protect user information. Regular checks and tests would help ensure the server stays reliable and secure as the app evolves.

3. SOLUTION

The application is designed to provide a smooth and interactive user experience, seamlessly integrating various components to enhance engagement and personalization. The user journey begins with a Splash Screen, which initializes the app and transitions into the Home Screen—serving as the main hub where users access features like the "Image of the Day" and navigate through the dashboard. The Log Screen plays a crucial role in collecting user input via surveys and journal entries, which feed into the app's analytical systems for personalized content and feedback.

The Art Screen is central to the app's creative interaction, allowing users to engage with AI-generated art or upload their own images. This screen connects to two AI engines: one for generating and enhancing visual content and another for analyzing facial expressions to provide real-time, emotion-based personalization [8]. The server architecture ensures seamless communication between these components, enabling real-time processing and secure data management, all while maintaining the scalability needed to handle varying user demands.

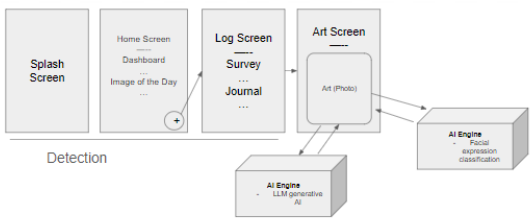

Figure 1. Overview of the solution

The AI-driven art generation engine is a key component of the application, designed to create and enhance visual content based on user interactions. This engine leverages advanced generative AI techniques to produce unique and personalized art pieces, contributing significantly to the user experience by offering content that resonates with the user's mental and emotional state.

Figure 2. Screenshot of the daily image

```python
def get_image(info, show=False):
    print('generating image...')
    prompt = f"Create an image based on the following information from a journal entry or
             f"The image should visually convey and enhance mental health awareness by inc
             f" symbolize healing, support, and resilience. Use calming colors, soothing s
             f"symbols to create a therapeutic tool that helps in relaxation and self-refl
             f"aim to evoke positive emotions and provide comfort, making it a valuable re
             f" awareness and support."
    response = client.images.generate(
        model="dall-e-3",
        prompt=prompt,
        size="1024x1024",
        quality="standard",
        n=1,
    )

    image_url = response.data[0].url
    response = requests.get(image_url)
    img = Image.open(BytesIO(response.content))

    if show:
        img.show()

    # Convert image to base64 string
    buffered = BytesIO()
    img.save(buffered, format="JPEG")
    img_str = base64.b64encode(buffered.getvalue()).decode('utf-8')

    return img_str
```

Figure 3. Screenshot of code 1

This code sample is responsible for generating images based on user input, such as a journal entry or survey response. It begins by creating a prompt that guides the AI to generate an image that conveys mental health awareness through symbols of healing, support, and resilience. The AI model, dall-e-3, processes this prompt and generates an image, which is then retrieved and displayed to the user. Additionally, the image is converted to a base64 string for easy transmission or storage. This process is triggered whenever a user interacts with the Art Screen, ensuring that the content generated is both unique and relevant to the user’s current state.

The facial expression classification engine is a critical component of the application, designed to analyze users' facial expressions in real-time and determine their emotional states. This analysis helps the application provide personalized feedback and content that aligns with the user’s current emotional state, enhancing the overall user experience. The component operates by processing video input from the user, detecting facial features, and classifying emotions based on established models of facial expression recognition.

```python
def get_emotion(location_videofile):
    # Build the Face detection detector
    face_detector = FER(mtcnn=True)
    # Input the video for processing
    input_video = Video(location_videofile)

    # Analyze every frame of the video for emotions
    processing_data = input_video.analyze(face_detector, display=False,

    # Convert analyzed data into a dataframe
    vid_df = input_video.to_pandas(processing_data)
    vid_df = input_video.get_first_face(vid_df)
    vid_df = input_video.get_emotions(vid_df)

    # Summarize the emotions detected throughout the video
    emotions = {
        'Angry': sum(vid_df.angry),
        'Disgust': sum(vid_df.disgust),
        'Fear': sum(vid_df.fear),
        'Happy': sum(vid_df.happy),
        'Sad': sum(vid_df.sad),
        'Surprise': sum(vid_df.surprise),
        'Neutral': sum(vid_df.neutral),
    }

    # Determine the most prominent emotion
    dominant_emotion = max(emotions, key=emotions.get)
    return dominant_emotion
```

Figure 4. Screenshot of code 2

The code sample provided handles the core task of facial expression classification. It starts by setting up a facial detection model (FER) that can identify and analyze faces in video frames [9]. The video is processed frame by frame to detect emotions such as anger, disgust, fear, happiness, sadness, surprise, and neutrality. These emotions are quantified and stored in a DataFrame for further analysis.

The code then aggregates the emotions detected throughout the video to identify which emotion was most prominent during the user's interaction [10]. The output of this process is the dominant emotion, which the application uses to tailor its responses and content, providing a personalized experience that aligns with the user's emotional state.

This classification runs in real-time as the user interacts with the camera, ensuring that the feedback is immediate and relevant. The accuracy and efficiency of this component are crucial to the app's success in delivering meaningful and responsive user experiences.

The user data logging and analysis component is essential for capturing and processing user inputs such as survey responses and journal entries. This component enables the application to gather valuable insights into the user's mental and emotional state over time, which can be used to personalize content, provide feedback, and track progress. By logging this data consistently, the app builds a comprehensive profile for each user, allowing for more informed and tailored interactions.

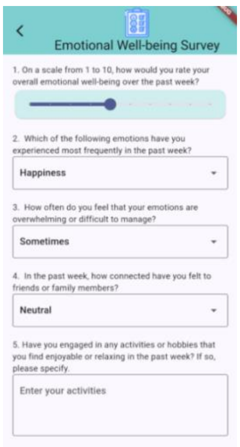

Figure 5. Screenshot of the survey

```python
def log_user_data(user_id, entry_type, content):
    # Create a timestamp for the log entry
    timestamp = datetime.now().strftime('%Y-%m-%d %H:%M:%S')

    # Structure the data to be logged
    log_entry = {
        'user_id': user_id,
        'entry_type': entry_type,  # e.g., 'journal' or 'survey'
        'content': content,
        'timestamp': timestamp
    }

    # Save the log entry to the database
    database.insert('user_logs', log_entry)

    # Optionally, analyze the data immediately after logging
    analyze_user_data(user_id)

def analyze_user_data(user_id):
    # Fetch the user's logs
    user_logs = database.fetch_all('user_logs', {'user_id': user_id})

    # Perform analysis on the data, e.g., detecting trends or emotional pa
    analysis_result = perform_sentiment_analysis(user_logs)

    # Store or use the analysis result for further actions
    database.update('user_profiles', {'user_id': user_id}, {'analysis': ar
```

Figure 6. Screenshot of code 3

The code sample provided illustrates how the app logs and analyzes user data. When a user submits a journal entry or completes a survey, the app records this information in a structured format, including the type of entry, the content, and a timestamp. This data is then stored in a database, forming part of the user's profile.

Immediately after logging the data, the app can trigger an analysis function that reviews the user's entries over time. For example, sentiment analysis might be performed on journal entries to detect emotional trends. The results of this analysis are then stored and can be used to tailor future interactions or provide feedback to the user. This process ensures that the app not only records data but also actively uses it to enhance the user's experience, offering insights and support based on their logged activities.

4. EXPERIMENT

4.1. Experiment 1

This experiment aims to evaluate the performance and responsiveness of the real-time image generation component within the app. Given the importance of this feature in delivering personalized and timely content, it is essential to ensure that the system can generate images efficiently, even when processing inputs of varying complexity.

To test the accuracy of the facial expression classification system, we will set up an experiment where a diverse group of participants will be recorded while expressing a range of predefined emotions (e.g., happiness, sadness, anger, surprise). These video recordings will serve as the input data for the classification engine.

The experiment will involve comparing the system's predictions against the known emotions that the participants were instructed to express. This will allow us to measure the system's accuracy by calculating the percentage of correct classifications. The control data, consisting of the participants' predefined emotions, will serve as the benchmark for evaluating the system's performance.

The experiment is designed to cover a wide variety of facial expressions, including subtle variations, to assess the robustness of the classification engine across different emotional states.

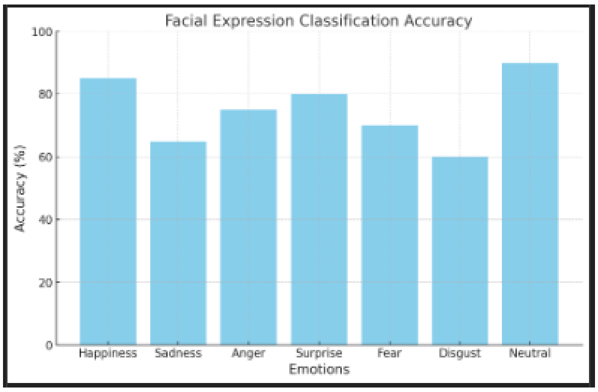

Figure 7. Figure of experiment 1

The analysis of the experiment's data will focus on key statistical measures such as the mean and median accuracy rates across all tested emotions. For instance, if the system correctly identifies 85% of happy expressions but only 65% of sad expressions, this discrepancy would be explored to understand the underlying causes.

The lowest accuracy might be observed in emotions that are typically more challenging to distinguish, such as fear versus surprise, while the highest accuracy might be seen in more distinct expressions like happiness. Unexpected results could point to issues like model bias or limitations in the training data. Understanding these variations will help in refining the classification engine, ensuring more consistent and reliable performance across all emotions.

4.2. Experiment 2

This experiment aims to evaluate the performance and responsiveness of the real-time image generation component within the app. Given the importance of this feature in delivering personalized and timely content, it is essential to ensure that the system can generate images efficiently, even when processing inputs of varying complexity.

This experiment aims to evaluate the performance and responsiveness of the real-time image generation component within the app. Given the importance of this feature in delivering personalized and timely content, it is essential to ensure that the system can generate images efficiently, even when processing inputs of varying complexity.

To evaluate performance, the time taken to generate each image will be measured, starting from the initial request to the final rendering. Additionally, the system’s ability to handle concurrent requests will be tested by generating multiple images simultaneously. This will help determine the scalability and efficiency of the image generation process under different load conditions.

Predefined prompts of varying complexity will be used as control data to ensure consistent testing. The results of these tests will provide insights into how well the system manages real-time demands and whether further optimization is needed to enhance responsiveness.

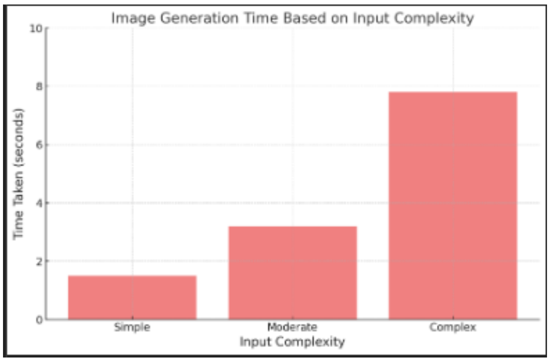

Figure 8. Figure of experiment 2

The analysis of the experiment's data indicates a clear correlation between input complexity and image generation time. Simple images were generated in approximately 1.5 seconds, moderate images in 3.2 seconds, and complex images in 7.8 seconds. This trend suggests that while the system is effective at generating images in real-time, there is a need for optimization, particularly when dealing with complex inputs that require more extensive processing.

The experiment also tested the system's performance under load conditions, where multiple images were generated simultaneously. The results showed that the system could handle concurrent requests without significant degradation in performance, although additional resource management strategies may be needed to ensure consistent responsiveness.

Overall, this experiment confirms that the real-time image generation component is capable of delivering personalized content efficiently, but it also highlights areas where further improvements could enhance the system's overall performance.

5. RELATED WORK

One notable approach to AI-driven mental health support involves the use of deep convolutional networks for emotion analysis, as demonstrated by Fei et al. (2020)[11]. Their method employs a deep convolution network to analyze facial expressions and detect mental health conditions. The system utilizes features extracted from the Fully Connected Layer 6 of AlexNet, combined with a Linear Discriminant Analysis Classifier to classify emotions. This method has shown higher accuracy and efficiency compared to other state-of-the-art deep learning models, including VGG16, GoogleNet, ResNet, and AlexNet itself. However, its reliance on pre-defined databases limits its adaptability to real-world scenarios where data can be more varied and unstructured. Moreover, the model's complexity may require significant computational resources, making it less suitable for low-power devices or real-time applications. Our project seeks to improve on this by integrating real-time processing and scalability, allowing the system to operate efficiently in diverse environments.

Another approach is presented by Muhammad et al. (2017), who developed a facial-expression monitoring system aimed at improving healthcare services in smart cities[12]. Their system uses a bandlet transform for feature extraction, followed by a center-symmetric local binary pattern (CS-LBP) technique and classification using a Gaussian mixture model and support vector machine. This method achieved an impressive 99.95% accuracy in recognizing facial expressions. However, the system's focus on static image processing may limit its effectiveness in dynamic, real-time environments where facial expressions can change rapidly. Additionally, the system's high accuracy is dependent on controlled conditions, which may not always be replicable in real-world settings. Our project aims to address these limitations by incorporating dynamic data inputs and improving the system's adaptability to varied and less controlled environments.

The third methodology is explored by Roy et al. (2022) through a neuro-symbolic AI approach for mental healthcare[13]. This method integrates symbolic clinical knowledge into neural AI systems to improve the reliability and explainability of AI-driven interventions in mental health. The hybrid system leverages both neural networks for pattern recognition and symbolic reasoning for decision-making, addressing some of the challenges faced by purely data-driven AI models, such as a lack of interpretability and explainability. However, the complexity of combining symbolic reasoning with deep learning requires substantial computational power and extensive training data, which could limit the system's accessibility and scalability. Our project builds on this by streamlining the integration of AI models with real-time data processing and enhancing the system's scalability for broader application in diverse environments.

6. CONCLUSIONS

The application, while promising, faces several limitations that need to be addressed for it to reach its full potential. One significant challenge is the variability in the accuracy of facial expression recognition, particularly with subtle or complex emotions such as fear and disgust [14]. The current system struggles to consistently identify these emotions, which could hinder its ability to provide accurate and meaningful feedback. Improving this aspect would require refining the underlying algorithms and expanding the training dataset to include a broader range of facial expressions and environmental conditions. Additionally, the real-time image generation component, although effective, can experience delays when processing complex inputs or handling multiple requests simultaneously. Optimizing the algorithms and enhancing resource management could mitigate these performance issues, ensuring the system remains responsive and reliable under all conditions.

Another critical area of concern is data privacy and security. Given the sensitive nature of the information being processed, robust encryption methods and secure data handling practices are essential to maintaining user trust and ensuring compliance with relevant regulations. The system must be designed to safeguard user data at every stage, from collection to storage and analysis. This is particularly important in the context of mental health, where users are likely to be sharing highly personal information. Moving forward, additional improvements could include integrating more advanced AI models to deepen emotion recognition capabilities, as well as expanding the app's features to offer predictive analytics for mental health monitoring [15]. Incorporating user feedback mechanisms could also help tailor the system more closely to individual needs, further enhancing the personalization and effectiveness of the user experience.

In conclusion, the development of this application represents a significant step forward in leveraging AI to enhance mental health support and user interaction. While there are areas that require further refinement, the current system demonstrates a strong foundation for providing personalized, real-time responses that are both meaningful and impactful. With continued development and optimization, this application has the potential to become a valuable tool in the field of mental health and emotional well-being.

REFERENCES

[1] 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.

[2] De la Torre, Fernando, and Jeffrey F. Cohn. "Facial expression analysis." Visual analysis of humans: Looking at people (2011): 377-409.

[3] Escobar, Javier I., and William A. Vega. "Mental health and immigration's AAAs: where are we and where do we go from here?." The Journal of Nervous and Mental Disease 188.11 (2000): 736-740.

[4] Kessler, Ronald C., et al. "Epidemiology of anxiety disorders." Behavioral neurobiology of anxiety and its treatment (2010): 21-35.

[5] Carr, Sarah. "'AI gone mental': engagement and ethics in data-driven technology for mental health." Journal of Mental Health 29.2 (2020): 125-130.

[6] Ghahramani, Zoubin. "Probabilistic machine learning and artificial intelligence." Nature 521.7553 (2015): 452-459.

[7] Matsumoto, David, and Hyisung C. Hwang. "Judgments of subtle facial expressions of emotion." Emotion 14.2 (2014): 349.

[8] Yang, Liping, et al. "Google Earth Engine and artificial intelligence (AI): a comprehensive review." Remote Sensing 14.14 (2022): 3253.

[9] Hao, Long, and F. Jiang. "A new facial detection model based on the Faster R-CNN." IOP conference series: materials science and engineering. Vol. 439. No. 3. IOP Publishing, 2018.

[10] Mahlke, Sascha. "Understanding users' experience of interaction." Proceedings of the 2005 annual conference on European association of cognitive ergonomics. 2005.

[11] Fei, Zixiang, et al. "Deep convolution network based emotion analysis towards mental health care." Neurocomputing 388 (2020): 212-227.

[12] Muhammad, Ghulam, et al. "A facial-expression monitoring system for improved healthcare in smart cities." IEEE Access 5 (2017): 10871-10881.

[13] Roy, Kaushik, et al. "Tutorial: Neuro-symbolic ai for mental healthcare." Proceedings of the Second International Conference on AI-ML Systems. 2022.

[14] Li, Shan, and Weihong Deng. "Deep facial expression recognition: A survey." IEEE transactions on affective computing 13.3 (2020): 1195-1215.

[15] Samek, Wojciech, Thomas Wiegand, and Klaus-Robert Müller. "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models." arXiv preprint arXiv:1708.08296 (2017).