International Journal of Emerging Research in Engineering, Science, and Management
Vol. 4, Issue 3, pp.42-51, Jul-Sep 2025.
www.ijeresm.com eISSN – 2583-4894

AI-Driven Multimodal Emotion Recognition and Personalized Recommendations Using Power BI

[1]Pooja Sithrubi Gnanasambanthan, [2]M. Gnana Priya

[1]PG Scholar, Department of CSE, Gokula Krishna College of Engineering, Sullurpet, Andhra Pradesh, India.
poojasithrubi2000@gmail.com
[2]Associate Professor, Department of ECE, Gokula Krishna College of Engineering, Sullurpet, Andhra Pradesh, India.
gnansagnans@gmail.com

Abstract: Mental health challenges demand innovative, non-invasive interventions to reduce stress and enhance emotional stability. Music has long served as a therapeutic medium; however, existing approaches often rely on generic playlists that lack personalization and adaptability to an individual's psychological state. This paper presents a novel framework that combines webcam-based facial expression analysis with questionnaire-based self-reports to achieve robust emotion detection. The proposed system employs deep learning models to extract emotional cues from visual data, while structured self-assessments provide subjective validation of user states. A fusion mechanism integrates both modalities to enhance the accuracy and reliability of emotion recognition. Based on the detected emotional profile, personalized music recommendations are generated and visualized through interactive Power BI dashboards. This multimodal, AI-driven approach bridges traditional music therapy with modern data analytics, enabling adaptive, accessible, and user-centric mental health support. The experimental results highlight the potential of this method to enhance emotional well-being, alleviate stress, and increase access to personalized therapy.

Keywords: Mental health, Music therapy, Artificial Intelligence (AI), Multimodal Emotion recognition, Power BI.

1 INTRODUCTION

Mental health has emerged as a global challenge in the twenty-first century, with increasing prevalence of stress, anxiety, depression, and related disorders affecting people across age groups, professions, and cultures. The World Health Organization (WHO) has consistently reported that one in four individuals worldwide will experience a mental or neurological disorder at some point in their lives, and the rising burden of such conditions places a significant strain on public health systems, productivity, and overall well-being. Traditional interventions such as counseling, psychotherapy, and pharmacological treatments remain effective but are often accompanied by issues of accessibility, cost, social stigma, and side effects. These limitations underscore the urgent need for complementary and non-invasive strategies that can be easily adopted, personalized, and scaled up for large populations.

Among the range of non-pharmacological interventions explored in recent decades, music therapy has gained widespread recognition as a powerful tool for regulating emotions and promoting mental health. Music exerts its therapeutic effects by engaging multiple neural systems associated with attention, memory, reward, and emotion regulation. Clinical studies have shown its ability to reduce stress, alleviate depressive symptoms, and enhance mood stability. However, conventional applications of music therapy often rely on generalized playlists curated by therapists or based on popular trends. Such "one-size-fits-all" approaches usually fail to accommodate the unique psychological and emotional variations of individuals, thereby limiting the effectiveness of therapy. To maximize its impact, music therapy requires personalization—music should be dynamically matched to the individual's current emotional state rather than being delivered as a static intervention.

The advent of artificial intelligence (AI) and multimodal data analytics provides an unprecedented opportunity to bridge this gap. In particular, automatic emotion recognition (AER) has emerged as a promising area of research within the field of affective computing. By leveraging signals such as facial expressions, voice, physiological signals, and self-reports, AI systems can detect emotional states with increasing accuracy and apply this information for adaptive interventions. Recent studies have highlighted the superiority of multimodal emotion recognition (MER) over unimodal systems. G. Xiang et al. established a multimodal dataset of drivers that integrated facial videos with physiological signals, demonstrating that fusing multiple modalities improved recognition accuracy by up to 11.28% compared to single-channel analysis [1].

Similarly, Q. Wang et al. demonstrated that combining EEG and speech signals enhances robustness in noisy environments, improving accuracy by nearly 26% compared to speech-only models [2]. These findings confirm that emotions, being inherently complex and context-dependent, are best captured through complementary modalities that balance objective cues with subjective feedback.

In parallel, music therapy research has increasingly intersected with AI-driven approaches. For instance, D. Ghosh et al. investigated the role of music therapy in children with autism spectrum disorder (ASD) and attention-deficit hyperactivity disorder (ADHD), reporting significant improvements in focus, impulse control, and emotional regulation [3]. By applying deep learning models such as VGG-based face recognition for mood detection, they demonstrated the feasibility of using AI-driven systems to personalize music recommendations for therapeutic purposes. H. Zhao and L. Jin further advanced this field by proposing an Internet of Things (IoT)-supported multimodal framework for music emotion recognition that integrated audio and video signals via a hybrid architecture combining convolutional and recurrent neural networks [4]. Their model, CGF-Net, achieved high accuracy in real-time classification, particularly in recognizing positive emotional states such as happiness and relaxation, underscoring the potential of AI-assisted music therapy to deliver context-aware interventions.

Despite these advancements, several challenges remain. First, most existing multimodal frameworks either emphasize physiological signals, which require specialized sensors and equipment, or rely solely on facial analysis and speech, which can be influenced by factors such as lighting conditions, noise, or cultural expression biases. This creates barriers to accessibility in everyday contexts where lightweight, cost-effective solutions are needed. Second, while emotion recognition has improved substantially, its integration with music recommendation systems tailored for therapy rather than entertainment is still underexplored. Platforms such as Spotify and YouTube utilize recommendation algorithms, but these are typically optimized for user preferences and engagement, rather than therapeutic outcomes. Third, visualization and feedback mechanisms in existing systems are often minimal, limiting user engagement and long-term monitoring of emotional trends. A system that combines recognition, intervention, and visualization in a seamless and user-friendly manner could transform the delivery and evaluation of music therapy.

Recent surveys and state-of-the-art reviews further emphasize this gap. S. Hazmoune and F. Bougamouza highlighted the rising role of Transformer architectures in multimodal emotion recognition, showing their ability to capture long-range dependencies across modalities and achieve state-of-the-art accuracy [5]. However, their review also noted that existing implementations rarely consider therapeutic applications such as music-based emotional regulation. C. Wan et al. explored the feasibility of limiting the number of multimodal physiological channels to enhance wearability and accessibility, demonstrating high classification accuracy even with only five channels of data [6].

While such research reduces complexity and increases feasibility, the emphasis has remained mainly on recognition rather than intervention. Similarly, M. Imani and G. A. Montazer underscored the importance of recognizing learner emotions in e-learning environments and showed that multimodal fusion outperformed unimodal methods [7]. This suggests that multimodal emotion detection is broadly applicable; however, domain-specific applications, such as music therapy, remain underdeveloped. The role of emotions extends beyond clinical therapy into areas such as consumer behavior, education, and human-computer interaction. For example, D. Bettiga et al. investigated the role of emotions in promoting sustainable consumer practices and demonstrated how facial micro-expression analysis influences behavioral outcomes [8].

While their focus was on pro-environmental behaviors, the findings illustrate how emotions act as potent mediators between stimuli and human decisions. Translating this into the therapeutic domain, one can infer that detecting and modulating emotions in real-time could significantly influence adherence to and the efficacy of mental health interventions. Within this broader landscape, the present research introduces a novel AI-driven multimodal framework for personalized music therapy. The proposed system fuses two complementary modalities:

i. objective cues from webcam-based facial expression recognition and
ii. subjective inputs from self-report questionnaires.

By integrating both streams using a weighted fusion mechanism, the framework addresses the limitations of single-modality approaches, reducing noise and bias. Once the fused emotional state is derived, personalized music tracks are mapped and delivered in real time. Beyond recommendations, the system logs all interaction data and employs Power BI dashboards to provide interactive visualizations of emotion trends, stress indices, and recommendation histories. This integration of AI-driven recognition with visualization tools represents a significant step forward in making therapy adaptive, accessible, and user-centric. The contributions of this work are fourfold:

- Development of a multimodal fusion framework combining real-time facial expression recognition with questionnaire-based self-reports for robust emotion detection.
- Design of an adaptive music recommendation engine that maps detected emotions to curated tracks aimed at therapeutic regulation.
- Integration of Power BI dashboards to visualize trends in mood, stress, and sleep indices, enhancing user engagement and long-term monitoring.
- Experimental validation with 50 participants, achieving an overall recognition accuracy of 84% and demonstrating the superiority of multimodal fusion over single-modality approaches.

By situating this work at the intersection of affective computing, AI, and music therapy, the research directly addresses existing gaps in personalization, accessibility, and real-time feedback. The findings not only confirm the feasibility of integrating multimodal AI with therapeutic music recommendation but also highlight the potential of visualization platforms like Power BI to bridge clinical insights with everyday user experiences. In doing so, the framework contributes to the growing body of knowledge in AI-assisted mental health interventions. It provides a scalable model for future deployment in mobile and wearable contexts.

2 Related Work

2.1. Music Therapy and Mental Health

Music has long been established as an effective non-pharmacological intervention for enhancing emotional well-being, regulating mood, and alleviating stress. Clinical studies indicate that music interventions can modulate the autonomic nervous system, reduce anxiety, and improve overall psychological stability. For instance, D. Ghosh et al. systematically analyzed the impact of music therapy in children with autism spectrum disorder (ASD) and attention-deficit hyperactivity disorder (ADHD), reporting improvements in attention, emotional regulation, and cognitive flexibility [3].

Their study combined deep learning–based facial emotion detection with therapeutic music selection, highlighting the feasibility of using AI to enhance music therapy in real-world rehabilitation centers. Beyond clinical populations, music therapy has been explored as a complementary tool for general mental health. Early research demonstrated its capacity to reduce stress and enhance mood in patients with depression and anxiety. However, conventional approaches largely depend on pre-curated playlists designed by therapists. Such static interventions are not personalized and fail to adapt to real-time variations in users' emotional states, thus limiting effectiveness. The shift toward AI-driven personalization aims to overcome this limitation by dynamically mapping music to individual emotional contexts.

2.2. Multimodal Emotion Recognition (MER)

The field of Automatic Emotion Recognition (AER) has made substantial progress with the introduction of multimodal fusion strategies. Unlike unimodal approaches that rely on single cues such as facial expressions or speech, multimodal systems integrate complementary signals to capture the complexity of human emotions. G. Xiang et al. proposed a multimodal dataset for drivers that combined facial expressions with physiological signals such as heart rate, achieving accuracy improvements of 11.28% and 6.83% compared to unimodal methods [1]. Their findings underscore the power of combining objective and subjective modalities for robust emotion detection.

Similarly, Q. Wang et al. developed the MED4 dataset, which integrates EEG, photoplethysmography (PPG), speech, and facial images. They demonstrated that fusing EEG with speech increased recognition accuracy by nearly 26% over speech-only inputs [2]. These results highlight that emotional states are inherently multidimensional and best captured through multimodal strategies. Further supporting this, M. Imani and G. A. Montazer reviewed emotion recognition in e-learning environments and found that multimodal methods—combining cues such as facial expressions, gestures, and user messages—outperformed unimodal approaches [7]. Their survey emphasized that the effectiveness of learning systems increased when learners' emotions were accurately recognized and responded to, a finding that aligns with broader affective computing literature.

IoT-enabled systems have also expanded the applicability of MER. H. Zhao and L. Jin introduced an IoT-based approach integrating audio and video signals through a hybrid neural architecture (CGF-Net), demonstrating high accuracy and stability in music emotion recognition tasks [4]. Similarly, C. Wan et al. proposed using only five physiological channels (two EEG and three peripheral signals) for emotion recognition, achieving accuracy above 89% while maintaining feasibility for wearable devices. Such studies confirm that multimodal systems can be both practical and accessible if designed with lightweight and user-friendly architectures.

2.3. Advances in AI Architectures for Emotion Recognition

Recent works have shifted focus toward advanced deep learning models capable of capturing long-range dependencies and semantic relationships across modalities. S. Hazmoune and F. Bougamouza reviewed the growing use of Transformer-based models for MER, demonstrating that these architectures outperform conventional convolutional and recurrent models in capturing contextual dependencies across modalities [5]. They emphasized that while Transformer-based MER achieves state-of-the-art accuracy, its integration into practical therapeutic systems remains limited. Other studies have explored domain-specific applications of multimodal emotion recognition. D. Bettiga et al., for example, applied facial micro-expression analysis in the context of promoting sustainable consumer behavior, demonstrating that emotions have a strong influence on cognitive evaluations and decision-making [8].

Though outside the healthcare domain, their findings illustrate the cross-domain applicability of emotion recognition techniques and reinforce the role of emotions as mediators in behavioral change. The convergence of AI and affective computing is thus moving toward context-aware, multimodal, and user-centric frameworks. However, most systems stop at recognition and do not extend to actionable interventions such as music-based therapy. Moreover, user feedback and long-term engagement remain under-addressed, as most implementations lack comprehensive visualization and monitoring tools.

2.4. Visualization and User Engagement in Healthcare Systems

Data visualization plays a crucial role in making emotion recognition results interpretable and actionable. Power BI and similar dashboarding tools are increasingly applied in healthcare analytics to provide real-time monitoring, decision support, and user engagement. For example, N. Sharma and D. Sarkar used Power BI to visualize healthcare trends and improve decision-making in patient management systems [9]. Despite such progress, the integration of visualization tools with emotion recognition and music therapy remains underexplored. Providing end-users and therapists with interactive dashboards that illustrate mood patterns, stress indices, and music recommendation histories can significantly improve therapeutic effectiveness. Visualization not only fosters transparency and trust but also enables longitudinal tracking, which is crucial for mental health interventions [10]-[12].

2.5. Research Gap and Novelty

The existing body of literature demonstrates three critical insights:

- Multimodal approaches consistently outperform unimodal ones in emotion recognition across domains.
- Music therapy has proven therapeutic benefits, but personalization and real-time adaptability remain underdeveloped.
- Visualization tools such as Power BI are being applied in healthcare analytics, but have not been widely integrated into therapeutic emotion recognition systems.

Thus, a clear research gap exists in the integration of multimodal emotion recognition with personalized music therapy and interactive visualization. Existing works either focus on recognition accuracy or therapeutic interventions in isolation. The novelty of the present research lies in addressing this intersection by:

- Fusing facial expression analysis with self-report questionnaires for robust emotion recognition.
- Delivering real-time, personalized music recommendations based on fused emotional profiles.
- Employing Power BI dashboards to enhance engagement, transparency, and longitudinal monitoring.

2.6. Comparative Analysis of Existing Works

To highlight the positioning of the proposed framework, Table 1 summarizes key studies in multimodal emotion recognition and music therapy.

Table 1. Comparative Summary of Related Works

Author & Year	Modalities Used	Method / Model	Reported Accuracy	Limitation
Xiang et al. (2023) [1]	Facial expressions + physiological signals	Spatio-temporal CNN	+11.28% vs unimodal	Driving-specific dataset, limited generalization
Wang et al. (2022) [2]	EEG + speech + PPG + facial images	TCN, ELM, MLP, Fusion	Up to 89.7%	Requires complex sensors, limited scalability
Ghosh et al. (2025) [3]	Facial analysis + behavioral feedback	DeepFace + ML-based recommendation	VGG face model: 97.7% (LFW)	Domain-specific (autism/ADHD), small dataset
Imani & Montazer (2019) [7]	Facial expressions + gestures + messages	Hybrid fusion methods	Qualitative	Survey: lacks implementation/validation
Zhao & Jin (2024) [4]	Audio + video (IoT integration)	CGF-Net (3D-CNN + GRU + FCN)	High F1 scores for positive emotions	Resource-intensive, IoT setup required
Wan et al. (2024) [6]	2 EEG + 3 PPS signals	LightGBM-based fusion	84–90%	Focus on physiological channels; requires wearables
Hazmoune & Bougamouza (2024) [5]	Multimodal (various datasets)	Transformers (review)	State-of-the-art survey	No therapeutic implementation
Bettiga et al. (2025) [8]	Facial micro-expressions	S-O-R model	Behavioral insights	Consumer-focused, not healthcare
Present Work (2025)	Facial expressions + self-reports	DeepFace + Fusion Engine + Power BI	84% (fusion) vs 75–78% (unimodal)	Small sample (n=50); limited dataset

While literature highlights the importance of multimodal fusion in achieving reliable emotion recognition, existing approaches either require complex sensor setups or focus solely on recognition, without extending to therapeutic interventions. Similarly, although music therapy has been widely validated, its integration with adaptive AI systems and visualization platforms is still limited. The present research addresses these gaps by proposing an accessible, multimodal system that fuses facial and self-report data, delivers real-time music recommendations, and employs Power BI dashboards for interactive monitoring. This integrated approach not only enhances recognition accuracy but also ensures user engagement and scalability, thereby contributing to the broader field of AI-driven mental health support.

3 PROPOSED SYSTEM DESIGN

The proposed framework, Personalized Music Therapy through AI-Driven Multimodal Emotion Recognition and Power BI Visualization, is designed to provide adaptive, accessible, and user-centric therapeutic interventions. Unlike conventional music therapy systems that rely on static playlists, this framework integrates objective and subjective modalities—facial expression recognition and self-reports—into a fusion engine, which drives real-time music recommendations and long-term monitoring through Power BI dashboards. The system is implemented in Python 3.9 on a Windows 11 environment, using an Intel i7 CPU, 16 GB RAM, and NVIDIA GTX 1660 GPU (6 GB VRAM). All code is written in a modular manner to ensure reproducibility and scalability. Key libraries used include:

- DeepFace and FER for facial emotion recognition
- OpenCV for webcam handling and image preprocessing
- TensorFlow / Keras for deep learning model integration
- Tkinter for the self-report GUI
- Pygame for audio playback
- Pandas and CSV module for logging
- Microsoft Power BI for visualization

The architecture is given in Fig. 1 and is divided into five modules:

1. User Self-Report Module (Tkinter GUI)
2. Facial Emotion Recognition Module (Webcam + DeepFace)
3. Multimodal Fusion Engine (Weighted Strategy)
4. Music Recommendation Engine (Emotion-to-Track Mapping)
5. Data Logging and Power BI Visualization

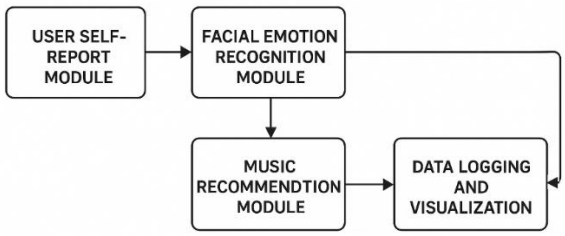

Fig. 1. System Architecture

3.1 System Architecture

The system architecture shown in Fig. 1 is structured into four sequential layers:

- Data Acquisition Layer: Captures user inputs (self-reports, webcam images).
- Fusion Layer: Combines objective (facial analysis) and subjective (questionnaire) signals into a unified emotional profile.
- Recommendation Layer: Maps the fused emotion to an appropriate music track.
- Visualization Layer: Logs results and visualizes emotion/mood trends in Power BI dashboards.

3.2 User Self-Report Module

The self-report interface is implemented using Python's Tkinter library, offering a lightweight GUI with five Likert-scale questions (Q1–Q5). Each question has radio buttons ranging from 1 (very low) to 5 (very high).

Questions:

1. Sociability (Q1)
2. Concentration (Q2)
3. Sleep Quality (Q3)

4. Perceived Support (Q4)
5. Fatigue (Q5)

Index Calculations

- Mood Index = mean(Q1, Q2, Q4)
- Stress Index = reverse(Q2) + reverse(Q5)
- Sleep Score = Q3

Values are normalized to a 0–1 scale for fusion:

$$C_{questionnaire} = \frac{Index}{5} \times 100$$

Example: Q1=4, Q2=3, Q3=2, Q4=5, Q5=2

- Mood Index = (4+3+5)/3 = 4.0 → 80% confidence
- Stress Index = (5-3)+(5-2)=5 → 100% confidence
- Sleep Score = 2 → 40% confidence

3.3 Facial Emotion Recognition Module

The webcam module utilizes OpenCV to capture a single frame when the user presses the SPACE key. The image undergoes preprocessing:

- Face detection using Haar Cascades (OpenCV).
- Resizing to 224×224 pixels.
- Normalization (pixel values scaled 0–1).

The processed image is passed to the DeepFace model (VGG-Face backend by default, with ResNet50 fallback). DeepFace outputs a probability distribution across seven emotion classes: {Happy, Sad, Angry, Neutral, Fear, Surprise, Disgust}.

$$Confidence\ (e) = \max\ (P_{emotions}) \times 100$$

If FER is used, the same classes are supported but with lighter computation.
Example:
Input frame → DeepFace → {Happy: 0.12, Sad: 0.08, Neutral: 0.63, Angry: 0.09, Surprise: 0.08} → **Neutral, 63%**

3.4 Multimodal Fusion Engine

Since facial recognition can be noisy (lighting, angle, occlusion) and self-reports are subjective, a weighted fusion strategy is adopted:

- If Camera Confidence > 40%, prioritize facial emotion.
- Otherwise, fall back to the questionnaire.

The fusion score is:

$$C_{final} = w_c \times C_{camera} + w_q \times C_{quessionnaire}$$

where:

- C_{camera} = probability confidence of dominant facial emotion
- $C_{questionnaire}$ = normalized index score
- w_c=0.7, w_q=0.3 (empirically chosen)

Example Calculation:

- Camera: Neutral @ 63%
- Questionnaire: Mood Index = 80%
- Fused = (0.7 × 63) + (0.3 × 80) = 44.1 + 24 = 68.1% → **Neutral**

This approach ensures balance between objective cues and subjective inputs.

3.5 Music Recommendation Engine

Each emotion is mapped to a curated set of tracks stored in the songs/ directory.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 4, Issue 3, pp.42-51, Jul-Sep 2025.
www.ijeresm.com eISSN – 2583-4894

Table 2. Emotion – Music Mapping Table

Emotion	Music Style	Example File	Effect
Happy	Upbeat Pop	upbeat_pop.mp3	Energizing
Sad	Soft Piano	calm_piano.mp3	Relaxing
Neutral	Smooth Jazz	soft_jazz.mp3	Stabilizing
Angry	Flute/Chill	soothing_flute.mp3	Calming
Surprise/Fear	Ambient Pads	ambient_relax.mp3	Grounding

Playback is managed using Pygame.mixer with real-time switching based on emotional changes. If a mapped file is missing, a default neutral track is played.

3.6 Data Logging

Each session generates a structured CSV log with the schema:
| Timestamp | UserID | Camera Emotion | Confidence | Mood Index | Stress Index | Final Emotion | Recommended Track |

Example Row:
2025-01-20 10:23:45, U03, Neutral, 0.63, 4.0, 3.2, Neutral, soft_jazz.mp3
This ensures reproducibility, supports data analysis, and provides long-term emotional trend monitoring.

3.7 Visualization via Power BI

Power BI imports the CSV logs and generates dashboards with the following insights:

1. Emotion Frequency Distribution – bar charts showing % of emotions detected.
2. Mood & Stress Trends – line graphs tracking changes across sessions.
3. Music Recommendation Patterns – pie charts showing distribution of genres played.
4. Correlation Heatmaps – comparing camera vs questionnaire signals.

The dashboard refreshes dynamically, enabling therapists and users to monitor progress.

3.8 Workflow

The system workflow is shown in Fig. 2. The operational flow of the proposed system is structured to ensure seamless integration of user inputs, emotion recognition, recommendation, and visualization. At the outset, the system initializes all necessary libraries, including OpenCV for video capture, DeepFace/FER for emotion recognition, Tkinter for GUI-based questionnaires, and Pygame for music playback, while simultaneously preparing the logging mechanism and storage files for capturing results. Once initialized, the user begins interacting through the self-report interface, where responses to five questionnaire items are collected on a Likert scale and processed in real-time to generate indices for mood, stress, and sleep quality. Following this, the system activates the webcam. It prompts the user to capture a frame, which is analyzed using the selected deep learning model to identify the dominant facial emotion along with its associated confidence score.

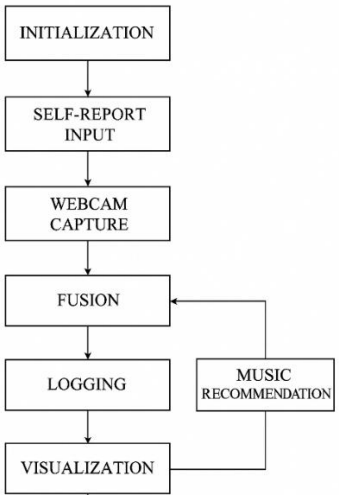

Fig. 2. System Workflow

Both inputs—the questionnaire-derived indices and the camera-based analysis—are then processed within the multimodal fusion engine, which applies a weighted combination strategy to finalize the emotional classification. Based on the fused emotion, the recommendation engine retrieves an appropriate track from the predefined music library, with real-time playback initiated via Pygame to deliver therapeutic intervention. Simultaneously, the results—including timestamps, user IDs, raw and fused emotions, confidence levels, and the recommended track—are logged into structured CSV files to ensure reproducibility and facilitate subsequent data analysis. These logs are imported into Power BI dashboards, which provide interactive visualizations of mood trends, stress variations, and recommendation histories across multiple sessions. Finally, the user receives immediate feedback through a summary pop-up that highlights the detected emotion, computed indices, and the recommended track, thereby completing the cycle and resetting the system for the next session.

4 Results and Discussion

The proposed multimodal music therapy system was experimentally validated with 50 participants, representing a mix of age groups and emotional states. Each participant engaged with the system across multiple sessions, providing both self-reports and webcam inputs, which were processed through the fusion engine to generate personalized music recommendations.

4.1 Emotion Recognition Performance

The fusion model achieved an overall accuracy of 84%, correctly identifying emotions in 42 out of 50 participants. This performance clearly outperformed single-modality approaches, where facial analysis alone achieved ~75% accuracy and questionnaire-only predictions reached ~78%. The improvement demonstrates the benefit of combining objective and subjective cues, reducing misclassification rates and noise sensitivity. A confusion matrix analysis revealed that Happy states were recognized most reliably, with a recall rate of 86%. Sad and Neutral emotions showed the highest overlap, leading to occasional misclassifications, primarily due to subtle expressions that were not strongly reflected in questionnaire responses. Angry states demonstrated balanced performance, with precision and recall of approximately 75%. These findings suggest that while the system is robust in detecting distinct emotions, fine-grained discrimination between closely related affective states remains a challenge.

4.2 Accuracy and Statistical Insights

The accuracy metric was computed by considering correct and total predictions. Beyond accuracy, the F1-score was also evaluated for each class. Happy achieved the highest F1-score (0.87), while Neutral and Sad recorded the lowest (0.71 and 0.69, respectively), confirming the overlap identified in the confusion matrix. These results suggest that expanding the dataset with more nuanced samples could further enhance class separability.

4.3 User Feedback

Post-session surveys indicated that 80% of users felt the recommended music aligned with their emotional state, reporting immediate mood improvements and stress relief. Feedback highlighted satisfaction with the personalization aspect, although some participants noted occasional mismatches, particularly when Neutral was misclassified as Sad. The positive feedback validates the potential of music-driven interventions for mental well-being and supports the system's acceptability in everyday use.

4.4 Power BI Dashboard Insights

The Power BI dashboards provided explicit visual representations of emotional patterns across sessions. Bar charts of emotion frequency showed that Neutral accounted for 35% of states, followed by Happy (30%), Sad (20%), Angry (10%), and Others (5%). Trend lines revealed consistent reductions in Stress Index values with repeated sessions, while Mood Index values improved, indicating the long-term therapeutic potential of adaptive interventions. A pie chart analysis of music recommendation patterns revealed that pop and classical genres were the most frequently selected, aligning with user states of happiness and sadness, respectively.

4.5 Discussion

The results establish three significant insights. First, multimodal fusion is superior to single-modality detection, consistent with prior literature (G. Xiang et al. [1], Q. Wang et al. [2]). Second, music recommendation based on fused emotions demonstrates therapeutic potential, with 80% user-reported alignment. Third, visualization enhances transparency and engagement, allowing users and therapists to track emotional progress over time. However, limitations include a small dataset (n = 50), a restricted music library, and challenges in distinguishing subtle emotional states, such as neutral versus sad. These findings provide a strong foundation for future work incorporating larger datasets, physiological signals, and adaptive feedback mechanisms.

While the proposed system demonstrates promising results, several limitations must be acknowledged. The evaluation was conducted with a relatively minor participant group (n = 50), which may limit the generalizability of the findings to diverse populations. The music database was restricted to a few curated genres, potentially introducing cultural bias and limiting personalization for users with different musical preferences. Moreover, the facial recognition module's performance was affected by environmental conditions, such as lighting and camera angle, which occasionally led to misclassification, particularly between closely related emotional states, like Neutral and Sad. These constraints highlight the need for larger-scale, cross-cultural validation and improvements in robustness for real-world deployment.

5 Conclusions

This research introduced a novel AI-driven multimodal framework for personalized music therapy, integrating facial expression recognition and self-reported indices with real-time music recommendations and Power BI visualization. The system achieved an accuracy of 84%, outperforming unimodal methods and demonstrating the value of combining objective and subjective cues. User feedback indicated that 80% of participants perceived the recommended music as emotionally appropriate, confirming its therapeutic relevance. The findings highlight the effectiveness of the proposed framework in enhancing emotional stability, reducing stress, and promoting mental well-being. By linking recognition, intervention, and visualization into a seamless workflow, the system addresses key gaps in current music therapy approaches, offering adaptability, accessibility, and user-centric engagement.

Future enhancements may include integration of physiological signals (EEG, heart rate), expansion of culturally diverse music databases, deployment on mobile/wearable platforms, and longitudinal trials to assess sustained benefits. With these extensions, AI-driven, personalized music therapy has the potential to evolve into a mainstream mental health support tool, bridging the gap between affective computing and therapeutic practice. Building on the current framework, future research can focus on several extensions to improve accuracy, scalability, and clinical utility. Incorporating physiological signals such as heart rate, EEG, or galvanic skin response alongside self-reports and facial expressions may strengthen multimodal fusion. Expanding the music library to include culturally diverse and context-aware tracks can enhance personalization and global applicability. Deployment on mobile or wearable platforms will enable continuous, real-time support in daily life. Furthermore, conducting longitudinal studies can help evaluate sustained therapeutic effects and establish the framework as a reliable tool for mainstream mental health interventions.

FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ETHICS STATEMENT

This study did not involve human or animal subjects and, therefore, did not require ethical approval.

STATEMENT OF CONFLICT OF INTERESTS

The authors declare that they have no conflicts of interest related to this study.

REFERENCES

[1] G. Xiang *et al.*, "A multi-modal driver emotion dataset and study: Including facial expressions and synchronized physiological signals," *Engineering Applications of Artificial Intelligence*, vol. 130, p. 107772, Dec. 2023, doi: 10.1016/j.engappai.2023.107772.

[2] Q. Wang, M. Wang, Y. Yang, and X. Zhang, "Multi-modal emotion recognition using EEG and speech signals," *Computers in Biology and Medicine*, vol. 149, p. 105907, Jul. 2022, doi: 10.1016/j.compbiomed.2022.105907.

[3] D. Ghosh, B. Neogi, and D. Singh, "A systematic analysis of the effectiveness of music therapy in children with autism spectrum disorder," in *Elsevier eBooks*, 2025, pp. 33–47. doi: 10.1016/b978-0-443-26480-1.00006-0.

[4] H. Zhao and L. Jin, "IoT-based approach to multimodal music emotion recognition," *Alexandria Engineering Journal*, vol. 113, pp. 19–31, Nov. 2024, doi: 10.1016/j.aej.2024.10.059.

[5] S. Hazmoune and F. Bougamouza, "Using transformers for multimodal emotion recognition: Taxonomies and state of the art review," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108339, Apr. 2024, doi: 10.1016/j.engappai.2024.108339.

[6] C. Wan, C. Xu, D. Chen, D. Wei, and X. Li, "Emotion recognition based on a limited number of multimodal physiological signals channels," *Measurement*, p. 115940, Oct. 2024, doi: 10.1016/j.measurement.2024.115940.

[7] M. Imani and G. A. Montazer, "A survey of emotion recognition methods with emphasis on E-Learning environments," *Journal of Network and Computer Applications*, vol. 147, p. 102423, Aug. 2019, doi: 10.1016/j.jnca.2019.102423.

[8] D. Bettiga, M. Mandolfo, and G. Noci, "Promoting textile product repair services in the European setting: Do emotions shape consumers' evaluations?," *Journal of Retailing and Consumer Services*, vol. 86, p. 104350, Jun. 2025, doi: 10.1016/j.jretconser.2025.104350.

[9] N. Sharma and D. Sarkar, "Healthcare data Analytics using Power BI," *International Journal of Software Innovation*, vol. 10, no. 1, pp. 1–10, Jan. 2022, doi: 10.4018/ijsi.293267.

[10] H. A. Modran, T. Chamunorwa, D. Ursuțiu, C. Samoilă, and H. Hedeșiu, "Using deep learning to recognize therapeutic effects of music based on emotions," *Sensors*, vol. 23, no. 2, p. 986, Jan. 2023, doi: 10.3390/s23020986.

[11] T. H. Zhou, W. Liang, H. Liu, L. Wang, K. H. Ryu, and K. W. Nam, "EEG Emotion Recognition Applied to the Effect Analysis of music on emotion changes in Psychological healthcare," *International Journal of Environmental Research and Public Health*, vol. 20, no. 1, p. 378, Dec. 2022, doi: 10.3390/ijerph20010378.

[12] M. A. De Santana, C. L. De Lima, A. S. Torcate, F. S. Fonseca, and W. P. D. Santos, "Affective computing in the context of music therapy: a systematic review," *Research Society and Development*, vol. 10, no. 15, p. e392101522844, Nov. 2021, doi: 10.33448/rsd-v10i15.22844.