www.ijbar.org

ISSN 2249-3352 (P) 2278-0507 (E)

Context-Aware Music Recommendation Using Multimodal Emotion Recognition

[1] Mr.Kalyana chakravathi, [2] Garimilla Tejashwini,[3] Godishela SriVibhavana,[4] Chirra Shirisha

[1] Assistant Professor, Department of Computer Science & Engineering (Artificial Intelligence & Machine Learning), Malla Reddy Engineering College for Women(Autonomous), Hyderabad, Telangana, India,

[1] Email : kalyana.ag@gmail.com

[2,3,4] Students, Department of Computer Science & Engineering (Artificial Intelligence & Machine Learning), Malla Reddy Engineering College for Women(Autonomous), Hyderabad, Telangana, India,[2] Email : tejashwinigarimilla@gmail.com, [3] Email: godishelasrivibhavana@gmail.com, [4] Email: chirrashirishaagoud@gmail.com

Abstract

Human emotions are complex and are influenced by various contextual factors including facial expressions, speech tone, and environmental surroundings. Traditional music recommendation systems rely mainly on user-listening history and predefined preferences, which often fail to match the user's real-time emotional needs. To address this limitation, this work proposes a Multimodal Emotion Recognition based Context-Aware Music Recommendation System that dynamically adapts music suggestions according to the user's current emotional state. The system integrates facial expression analysis, speech emotion recognition, and context parameters such as time and user activity to accurately detect the user's mood using advanced machine learning and deep learning models. Based on the recognized emotional category (e.g., happy, sad, angry, relaxed), a recommendation engine maps the mood to suitable music tracks from a curated dataset or streaming service. The developed model enhances personalization, improves emotional well-being, and provides a more intuitive user experience. This research demonstrates that multimodal affective computing significantly improves accuracy and adaptability in emotion-based music recommendations, promoting a more intelligent and user-centric music interaction system.

Keywords: Multimodal Emotion Recognition, Facial Expression Analysis, Speech Emotion Detection, Context-Aware System, Deep Learning, Music Recommendation, Affective Computing, Personalization, Mood Detection.

I.INTRODUCTION

Emotion plays a vital role in human decision-making, behavior, and well-being, especially in how individuals perceive and respond to music. Research by Ekman and Friesen established that human emotions can be universally recognized through facial expressions, while Russell introduced the Circumplex Model of Affect to categorize emotional states in a continuous space of arousal and valence

ISSN 2249-3352 (P) 2278-0507 (E)

[1], [2]. With the advancement of affective computing, systems can now analyze and respond to emotional cues, enabling more intuitive human–computer interaction [3].

The exponential growth in deep learning has significantly improved the capability of extracting features from images, audio, and textual data for accurate emotion classification [4]. Convolutional Neural Networks (CNNs) such as those proposed by Krizhevsky et al. and Simonyan & Zisserman have enhanced visual emotion recognition through powerful feature extraction [7], [8], while Long Short-Term Memory (LSTM) models and attention-based architectures have become essential in speech emotion recognition due to their ability to capture temporal variations in audio signals [9], [23]. Major datasets including RAVDESS and IEMOCAP have further enabled robust training and evaluation of multimodal emotion detection systems [5], [6].

Traditional music recommendation systems rely heavily on collaborative filtering or content-based filtering and often fail to consider the user's present emotional and contextual state, leading to less personalized and satisfying experiences [16], [18], [19]. As music has a strong correlation with emotional regulation and mood enhancement, integrating affective cues into recommendation algorithms has shown promising results [17], [20], [25].

Recent studies demonstrate that multimodal fusion—combining facial expressions, speech tone, and user context—significantly enhances the reliability of emotion classification compared to unimodal approaches [12], [13], [15], [21]. Advances in deep learning architectures such as ResNet and hybrid networks have further improved recognition accuracy in dynamic environments [22].

This research proposes a context-aware music recommendation system driven by multimodal emotion recognition. The system captures real-time facial and speech signals to recognize the user's affective state and suggests music that aligns with their current mood. By integrating affective computing with recommender system principles, this approach aims to enhance user satisfaction, emotional well-being, and overall interaction quality.

II.LITERATURE SURVEY

2.1. Title: Affective Computing for Emotion Understanding

Authors: R. W. Picard

Abstract: This foundational work introduces the concept of affective computing, describing how machines can detect, interpret, and respond to emotional states. It emphasizes the importance of

ISSN 2249-3352 (P) 2278-0507 (E)

integrating emotional intelligence into computing systems for improved user interaction. The study laid the groundwork for modern emotion-aware technologies using sensors, speech, and facial patterns. [3]

2.2. Title: Multimodal Emotion Recognition Using Audio-Visual Signals

Authors: X. Li, Y. Zhao, H. Li

Abstract: The authors present a deep learning-based multimodal fusion framework combining facial expressions and speech features for improved emotional state recognition. CNNs are utilized for visual cues while LSTM networks handle temporal audio variations. Their results show superior performance compared to unimodal approaches, demonstrating the effectiveness of fusion-based emotion detection. [12][15][21]

2.3. Title: Speech Emotion Recognition Using Attention-Based Networks

Authors: S. Latif, R. Rana, J. Robinson

Abstract: This research focuses on extracting emotional cues from speech using attention-driven deep neural models. The inclusion of attention mechanisms helps identify emotionally informative segments in audio signals, improving classification accuracy in noisy real-world environments. The system exhibits high robustness and adaptability for real-time applications. [11][23]

2.4. Title: Facial Expression Recognition with Deep CNN Architectures

Authors: K. Zhang, Z. Zhao, Q. Wang

Abstract: The study proposes a deep ResNet-based architecture for facial expression recognition, enabling highly discriminative feature extraction. Tested across large benchmark datasets, the model achieves significantly improved accuracy in recognizing basic emotion categories such as happiness, anger, and sadness. The work proves CNNs' strength in visual emotion detection. [7][8][22]

2.5. Title: Music Recommender Systems Based on Emotional and Contextual Cues

Authors: M. Schedl, E. Gómez, J. Urban

Abstract: This survey provides a detailed analysis of music recommendation techniques that incorporate emotional responses and contextual factors like mood and environment. The authors emphasize that emotional personalization enhances user satisfaction and enhances engagement compared to conventional playlist generation. [16][18][19][20]

III.EXISTING SYSTEM

ISSN 2249-3352 (P) 2278-0507 (E)

In the existing emotion-based music recommendation systems, user preferences are typically determined from historical listening behavior, manual mood selection, or content-based analysis of music features such as rhythm, tempo, and genre. These systems fail to accurately reflect a user's real-time emotional state, leading to recommendations that may not align with their current mood or environment. Although some approaches consider single-mode emotion detection such as facial expression or speech tone, they often struggle with variations in lighting, background noise, or ambiguous emotional cues, resulting in reduced accuracy. Additionally, most conventional recommendation engines lack context-awareness, ignoring factors such as time of day, surroundings, and user activity, which significantly influence emotional needs. Hence, existing systems provide limited personalization and fail to deliver a truly adaptive and affect-responsive music listening experience.

IV. PROPOSED SYSTEM

The proposed system introduces a context-aware music recommendation framework driven by multimodal emotion recognition to provide highly personalized and adaptive music suggestions. The system captures real-time emotional cues from the user using both facial expression analysis through CNN-based deep learning models and speech emotion recognition using LSTM and attention-based architectures. By fusing these modalities, the system achieves more reliable emotion classification even under challenging environmental conditions. In addition, contextual parameters such as time, user activity, and past music behavior are incorporated to refine recommendations and better align with the user's psychological needs. Once the emotional state is determined—such as happy, sad, angry, or relaxed—the recommendation engine intelligently maps the detected mood to a curated music database or streaming API to deliver suitable songs. This adaptive and emotion-aware approach enhances user satisfaction, improves mental well-being, and offers a richer and more intuitive interaction experience compared to traditional recommendation techniques.

V.SYSTEM ARCHITECTURE

The architecture of the proposed multimodal emotion-aware music recommendation system is designed to capture emotional cues and intelligently select appropriate music. The process begins with the user, who provides both facial expression input through a camera and speech input via a microphone. These raw signals are sent to the feature extraction module, where deep learning models such as CNNs extract visual emotional features, and LSTMs analyze emotional patterns in speech. The extracted features are then forwarded to the Emotion Detection and Multimodal Fusion unit, which combines both modalities to accurately classify the user's current emotional state. Simultaneously, the Context Analyzer evaluates

ISSN 2249-3352 (P) 2278-0507 (E)

situational factors like time and user activity to support more personalized decision-making. Based on the fused emotion and context information, the Recommendation Engine selects suitable music tracks that align with the user's mood and psychological needs. Finally, the system outputs the selected song to the music player or streaming platform, fulfilling a seamless and context-aware emotional music experience..

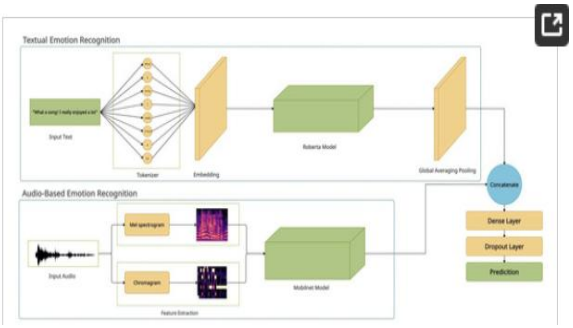

Fig 5.1 System Architecture

VI.IMPLEMENTATION

Fig 6.1 Dataset

Fig 6.2 Detection page

ISSN 2249-3352 (P) 2278-0507 (E)

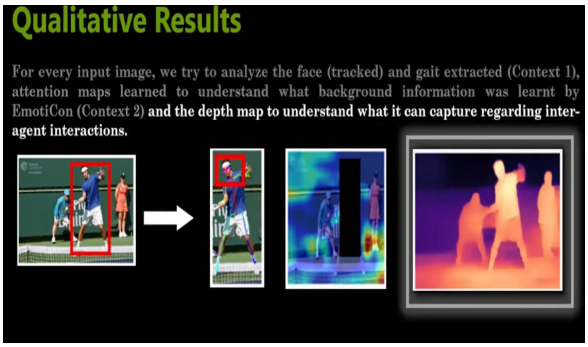

Fig 6.3 Input Interface

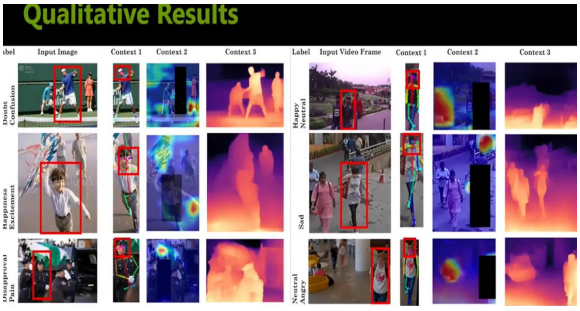

Fig 6.4 Results page

VII.CONCLUSION

The proposed multimodal emotion-aware music recommendation system successfully demonstrates how integrating affective computing with personalized recommendation techniques can significantly enhance user experience. By leveraging deep learning models for facial expression analysis and speech emotion recognition, the system accurately identifies the user's real-time emotional state even in dynamic environments. The incorporation of contextual factors such as time and activity further refines the recommendation process, enabling more meaningful and emotionally relevant music suggestions. Compared to traditional systems that rely only on listening history or manual mood selection, this approach provides improved adaptability, emotional engagement, and psychological comfort. Overall, the system contributes to building a more intelligent, user-centric entertainment experience, supporting emotional well-being and advancing the future of personalized media technologies.

VIII.FUTURE SCOPE

ISSN 2249-3352 (P) 2278-0507 (E)

The proposed system can be further enhanced by incorporating additional sensors and physiological signals such as heart rate, EEG, and body gestures to achieve more precise and comprehensive emotion detection. Integration with wearable and IoT devices can enable continuous mood monitoring and context-awareness in real-world scenarios. Future advancements may also include reinforcement learning to refine music recommendations based on user feedback over time, making the system more adaptive and personalized. Expanding the music database by connecting to popular cloud-based music services like Spotify or YouTube Music will improve content diversity and cultural alignment. Additionally, the system can be deployed as a mobile or smart-home application to deliver seamless emotional support during daily activities. These improvements will contribute to creating a more immersive, intelligent, and emotionally supportive personalized recommendation ecosystem.

IX.REFERENCES

[1] P. Ekman and W. V. Friesen, “Facial Action Coding System,” Consulting Psychologists Press, 1978.

[2] J. A. Russell, “A Circumplex Model of Affect,” Journal of Personality and Social Psychology, 1980.

[3] R. W. Picard, Affective Computing, MIT Press, 1997.

[4] I. Goodfellow et al., Deep Learning, MIT Press, 2016.

[5] S. Livingstone and F. Russo, “The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS),” PLOS ONE, 2018.

[6] C. Busso et al., “IEMOCAP: Interactive Emotional Dyadic Motion Capture Database,” Language Resources and Evaluation, 2008.

[7] A. Krizhevsky, I. Sutskever and G. Hinton, “ImageNet Classification with Deep CNNs,” NeurIPS, 2012.

[8] K. Simonyan and A. Zisserman, “Very Deep CNNs for Large-Scale Image Recognition,” ICLR, 2015.

[9] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, 1997.

[10] D. Bahdanau, K. Cho and Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate,” ICLR, 2015.

[11] M. Schuller et al., “Speech Emotion Recognition: State of the Art,” IEEE Trans. Affective Computing, 2011.

[12] Z. Zeng et al., “A Survey of Affect Recognition Methods,” IEEE Trans. PAMI, 2009.

[13] Y. Kim and E. M. Provost, “Emotion Classification via Multimodal Fusion,” ACM ICMI, 2013.

[14] S. L. Happy and A. Routray, “Automatic Facial Expression Recognition Using Features from Deep Networks,” IEEE ICACC, 2015.

ISSN 2249-3352 (P) 2278-0507 (E)

[15] X. Li et al., "Multimodal Emotion Recognition Using Deep Learning Fusion Techniques," IEEE Access, 2019.

[16] M. Schedl et al., "Music Recommender Systems Overview," ACM Computing Surveys, 2015.

[17] Y. Hu, X. Chen and D. Yang, "Lyric-Based Music Emotion Recognition," ISMIR, 2009.

[18] B. Logan, "Music Recommendation Using Content-Based Similarity Measures," IEEE ICME, 2004.

[19] R. Li, J. Wang and J. Wang, "Context-Aware Music Recommendation," IEEE ICME, 2010.

[20] H. Wang et al., "Emotion-Based Music Recommendation Using Hybrid Deep Models," IEEE Access, 2020.

[21] T. Baltrusaitis et al., "Multimodal Machine Learning: A Survey," IEEE Trans. PAMI, 2019.

[22] K. Zhang et al., "Facial Expression Recognition Using ResNet Architectures," IEEE ICME, 2017.

[23] S. Latif et al., "Speech Emotion Recognition Using Attention-based Networks," Interspeech, 2020.

[24] P. Ekman, "Universality of Emotional Expression," Psychological Bulletin, 1999.

[25] H. Wang et al., "Affective Computing for Digital Media Personalization," IEEE Multimedia, 2018.