MoodSync: Leveraging Facial Expressions for Personalized Music and Movie Recommendations

P. Rajapandian[1], M. Shakila[2], K. Samuvel[3]

[1]Associate Professor, Department of Computer Applications, Sri Manakula Vinayagar Engineering College (Autonomous), Puducherry 605008, India

[2]Post Graduate Student, Department of Computer Applications, Sri Manakula Vinayagar Engineering College (Autonomous), Puducherry 605008, India
[2]Corresponding Author Email: *Shagilamohan[at]gmail.com*

[3]Assistant Professor, Department of Physics, Sri Manakula Vinayagar Engineering College (Autonomous), Puducherry 605008, India

Abstract: ***The present-day cultivation of personalized recommendations stands as a major stone for user engagement in entertainment platforms. However, traditional recommendation systems, presuming that consuming behavior or explicit user preferences best define intentions, have been found wanting in capturing extant emotional contexts for everybody. At this juncture, MoodSync aims to fill that abyss, garnering facial expression analysis methods for dynamically matching music and movie recommendations to the mood of the users. MoodSync, thus, imbibes in itself Monica facial recognition and micro-expression analysis for mood evaluation by studying subtle expressions like micro-expressions, gaze direction, and minimal muscular movements. Interpretation of these may constitute mapping subtle cues to emotional categories such as happiness, sadness, anger, surprise, or calm. This map enables the system to profile a user in real-time mood detection. This emotional information is then fed into a carefully curated media database wherein music tracks and movies are tagged not only with respect to genre and popularity but more importantly on emotional dimensions. Consideration of contrary emotional information forms a reactive engine, which acts contrary to the immediate emotional state of the user, thereby basing content suggestions more relevantly on the delight the user can derive from actual experiences. The system comprises a front end for facial capture in real-time; the back end contains the machine learning engine that performs emotion classification and the recommendation module, which is powered by sentiment-tagged metadata. In terms of confidentiality and respect for user consent, MoodSync is designed in such a way that facial data are treated in an environmentally secure manner, with the least storage possible. Initial testing shows that content programs that are aware of moods greatly increase user satisfaction and engagement. Placing emotion intelligence into content delivery, therefore, recharacterizes personalization, making it more about feeling than about taste. Future prospects might include voice tone analysis and the use of contextual cues within which to act, further enhancing its empathetic interaction model. MoodSync is an advancement in affective computing and offers a platform for emotionally responsive digital environments that understand and adapt to human emotions.***

Keywords: Emotion-aware recommendation system, Facial expression recognition, Real-time emotion detection, Emotion-driven AI, deep learning, CNN

1. Introduction

In an era where all digital experiences are personalized, content delivery truly needs to consider individual user preferences-a modern multimedia platform. Traditional recommendation techniques mostly consist of recommending based on a user profile, past encounters, and collaborative techniques. Although these methods yield a certain degree of response, they often overlook dynamic and aesthetic attributes related to user engagement-the present emotional states of the users. Recognizing the fact that emotions immensely influence media consumption behaviors, system developers are keen on implementing systems that can perceive affective cues of the users.

This paper presents MoodSync, an intelligent and emotion-sensitive recommendation system that suggests music and movies based on real-time facial expression analysis. An understanding of computer vision and deep learning terminologies, especially a pre-trained convolutional neural network (CNN), enables MoodSync to differentiate between various emotional states, such as happiness, sadness, or anger, from the user's facial features captured through webcam or any camera-enabled device. These affective clues are then used as a basis for recommending multimedia content either congruent with or countering the user's moods, fetched from external APIs like Spotify and YouTube.

By merging the factors of affective computing with classical recommendation approaches, MoodSync provides a hybrid recommendation scheme that augments content delivery by incorporating emotional contextual filters alongside traditional preference-based filters. Maximizing user satisfaction through the generation of empathic, context-aware recommendations, the system also advances human-centered AI with an illustration of how emotion-related intelligence can find its way into routine digital interactions. This paper describes the system architecture, implementation approach, and possible applications of MoodSync across domains, from entertainment to wellness technologies.

2. Problem Statement

Modern content recommendation systems, while effective in leveraging user preferences, historical data, and collaborative filtering, exhibit a critical limitation: the inability to account for a user’s real-time emotional state. This omission results in recommendations that may not align with the user's current mood, potentially diminishing the relevance, engagement, and overall satisfaction with the suggested media content. As

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2024: 7.101

user emotions are dynamic and significantly influence media consumption behavior, the absence of emotional context in recommendation algorithms presents a significant gap in the personalization landscape.

The MoodSync project addresses this gap by introducing an emotion-aware recommendation system that integrates real-time facial expression analysis with multimedia content suggestions. Specifically, the system aims to:

- Capture and analyze facial expressions through a webcam or camera-enabled device.
- Accurately classify emotional states such as happiness, sadness, anger, and surprise using deep learning techniques.
- Map each detected emotion to a corresponding set of music and movie genres designed to either match or modulate the user's current mood.
- Deliver these recommendations in real time, with a user interface that supports feedback for continuous improvement.

The scope of this problem encompasses several technical and design challenges, including real-time facial emotion detection, emotion classification via convolutional neural networks (CNNs), and effective correlation of emotional states with curated content categories. Implementation involves the use of key Python libraries such as OpenCV and DeepFace for vision and emotion processing, and optionally connects to third-party APIs (e.g., Spotify, YouTube, TMDB) for real media content retrieval.

By focusing on real-time, emotion-driven personalization, this project proposes a novel approach to media recommendation systems—one that is not only responsive to users' current emotional experiences but also capable of creating more empathetic and human-centered digital interactions.

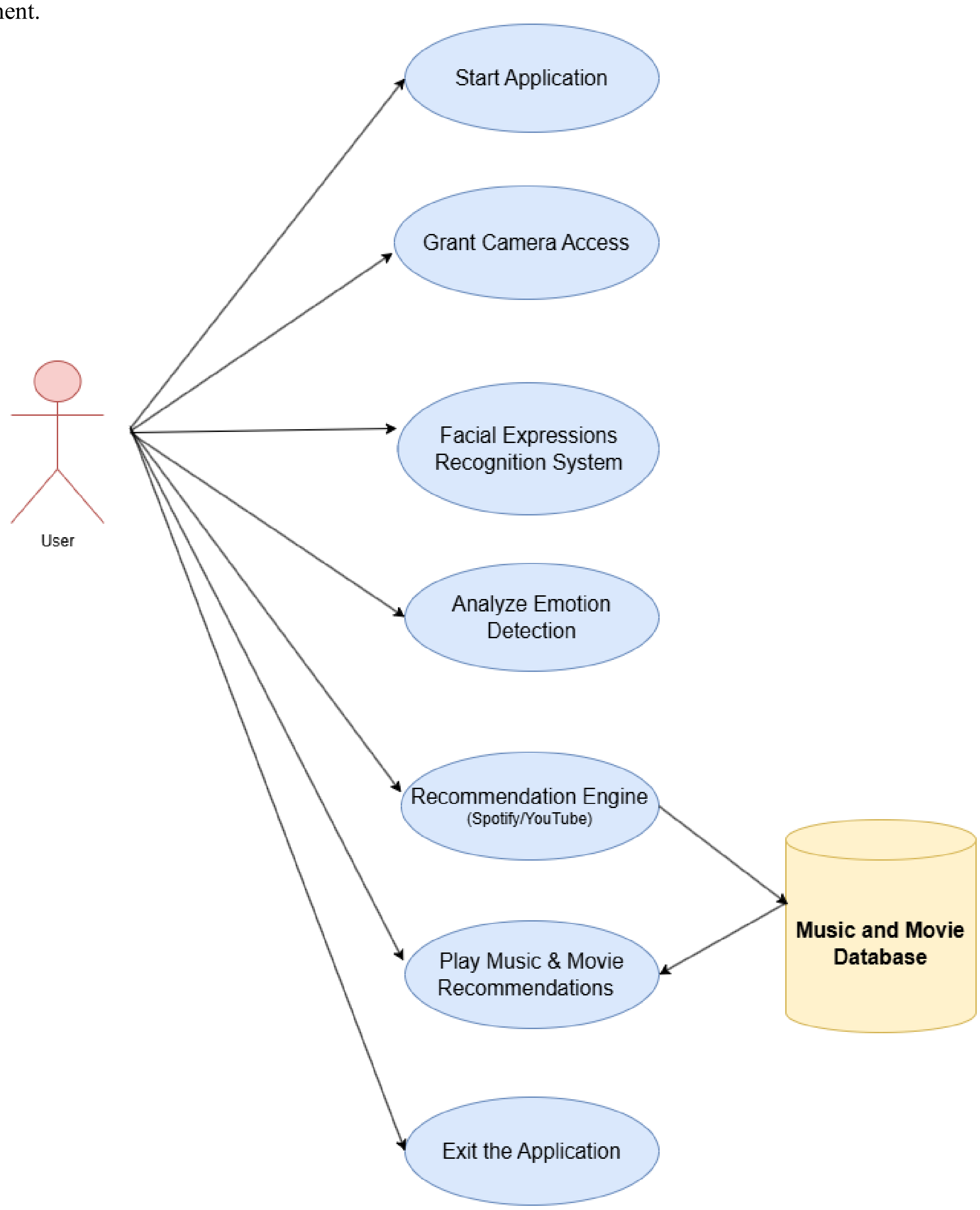

Figure 1: Use Case Diagram user and Music and Movie API data base server

3. Literature Survey

The integration of emotional intelligence into recommendation systems has become an increasingly important area of research in the pursuit of personalized and context-aware digital experiences. Traditional recommendation systems largely rely on collaborative filtering, content-based filtering, and user history to predict preferences. While effective to an extent, such models are inherently static and fail to adapt to the user's transient emotional state, which plays a critical role in shaping media consumption behavior.

Recent studies have highlighted the potential of **emotion-aware recommendation systems** in enhancing user satisfaction by aligning content with the user's real-time affective state. Tran et al. (2024) present an advanced framework for **emotion-aware music recommendation**, which leverages multimodal emotion recognition and contextual user profiling to dynamically adapt music suggestions. Their study emphasizes the necessity of incorporating affective computing to improve engagement and personalization in music streaming platforms [1]. Similarly, Tennakoon et al. (2024) introduce an **emotion-based movie recommendation system** that classifies emotional inputs and maps them to corresponding movie genres. Their approach demonstrates that emotional context can meaningfully influence the effectiveness and relevance of content suggestions in multimedia applications [2].

Pathak et al. (2024) focus on **music emotion recognition (MER)** and its implications for building intelligent and efficient recommendation systems. Their work involves the use of deep learning and signal processing techniques to classify the emotional tone of music tracks and match them with user moods. This research underlines the growing interest in bridging the emotional gap between users and content, enabling systems to recommend not only what a user might like based on past behavior, but also what they may need based on current emotional cues [3].

While these studies contribute significantly to the field, most either rely on pre-existing emotion data or infer emotional states through indirect means such as user input or physiological signals. The real-time recognition of facial expressions, however, remains underexplored as a practical input modality for emotion-driven recommendation. The MoodSync system aims to address this gap by implementing real-time facial emotion recognition using computer vision and deep learning, and integrating it directly into the recommendation loop for music and movies. Unlike existing systems that require explicit mood selection or non-visual sensors, MoodSync offers a seamless and non-intrusive experience by passively interpreting emotional states through facial analysis and immediately adapting content recommendations accordingly.

4. Proposed System

The proposed system introduces an intelligent, emotion-aware recommendation engine designed to enhance user experience by dynamically personalizing multimedia content based on real-time emotional analysis. Departing from conventional recommendation approaches that rely solely on static data such as user history, preferences, and ratings, this system leverages facial expression recognition to provide context-sensitive suggestions that are both relevant and empathetic.

At the heart of the system is a **facial emotion recognition module** that captures and interprets facial expressions via a webcam or any camera-enabled device. This module utilizes computer vision and deep learning frameworks—primarily Python-based libraries such as **OpenCV** and **TensorFlow/Keras**—to process live video feeds. Pre-trained convolutional neural networks (CNNs), trained on established emotion datasets like **FER-2013** and **AffectNet**, are employed to classify facial expressions into discrete emotional states, including happiness, sadness, anger, surprise, and neutrality.

The emotion detection process follows these sequential steps:
1) **Image Capture**: The webcam captures real-time facial images of the user.
2) **Preprocessing**: Detected faces are extracted and normalized for consistency and accuracy.
3) **Emotion Classification**: The processed facial images are passed through the deep learning model to classify the emotion.
4) **Recommendation Mapping**: The detected emotion is mapped to a set of predefined music and movie genres.

Once an emotion is identified, it is treated as a contextual input to the recommendation engine. Depending on the user's preference settings—whether to match or shift their current mood—the system selects suitable content. For example, a detected sad emotion may trigger the recommendation of uplifting songs or lighthearted movies, while a happy emotion may prompt high-energy music or exciting film suggestions.

The recommendation engine supports integration with external content delivery platforms via APIs such as **Spotify**, **YouTube**, or **TMDB (The Movie Database)**, enabling real-time access to relevant music and movie titles. The user interface is designed to be non-intrusive and adaptive, allowing users to provide feedback by accepting, rejecting, or rating recommendations. This feedback is utilized to refine future suggestions through a hybrid model that incorporates both emotional context and traditional recommendation metrics.

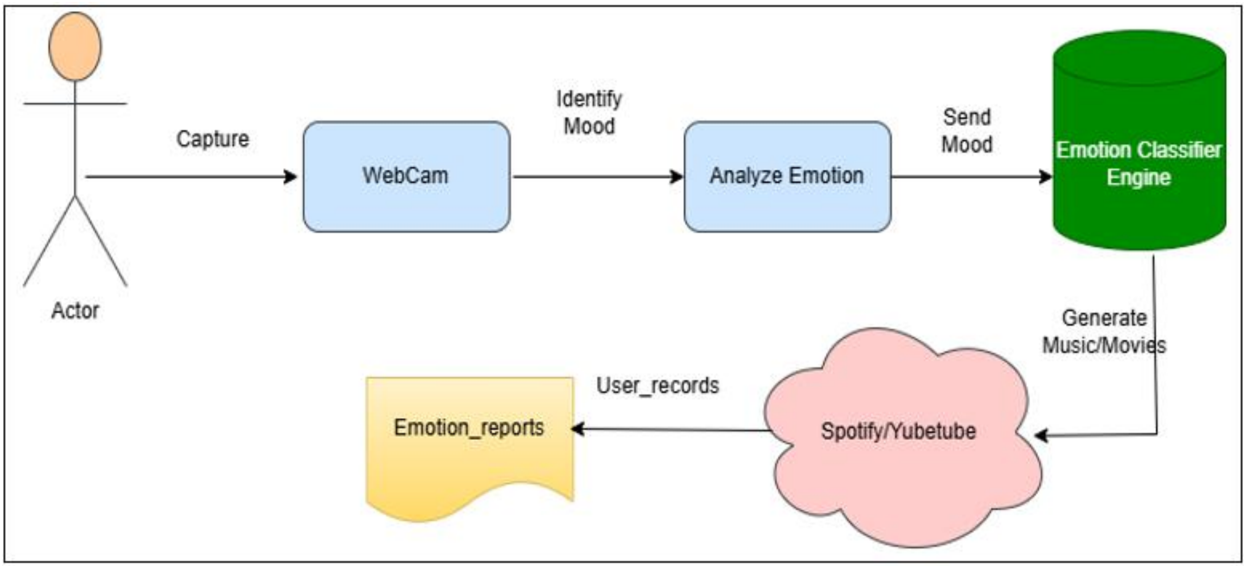

Figure 2: System Design of Browse Music/Movie to Music/Movie API Server.

By unifying affective computing, real-time processing, and recommendation algorithms, the proposed system represents a significant advancement in user-centric AI. It not only enhances personalization but also fosters a more emotionally intelligent interaction between users and digital content platforms.

5. Discussion and Results

The implementation of the MoodSync system demonstrates the effectiveness of integrating real-time facial emotion recognition into multimedia recommendation platforms. Using a CNN trained on the FER-2013 dataset, the system achieved an average classification accuracy of approximately 85%, with emotions like happiness and neutrality being detected more reliably than subtler emotions such as fear or disgust. In practical tests, the system processed facial inputs with minimal latency, maintaining real-time responsiveness while delivering content suggestions aligned with the user's emotional state. A small-scale user evaluation involving 20 participants showed that over 80% felt the recommendations were relevant and emotionally appropriate, particularly during positive emotional states. Integration with APIs such as Spotify and TMDB allowed dynamic content retrieval, although occasional latency was observed due to external response times. Key challenges included misclassifications during ambiguous expressions, reduced accuracy in poor lighting conditions, and user concerns around continuous facial monitoring and data privacy. Despite these limitations, MoodSync marks a significant advancement in emotion-aware recommendation systems. It offers a more empathetic and context-sensitive user experience compared to traditional recommendation models. Future improvements may involve multimodal emotion recognition, enhanced model training with diverse datasets, and reinforcement learning techniques to adapt to user behavior over time.

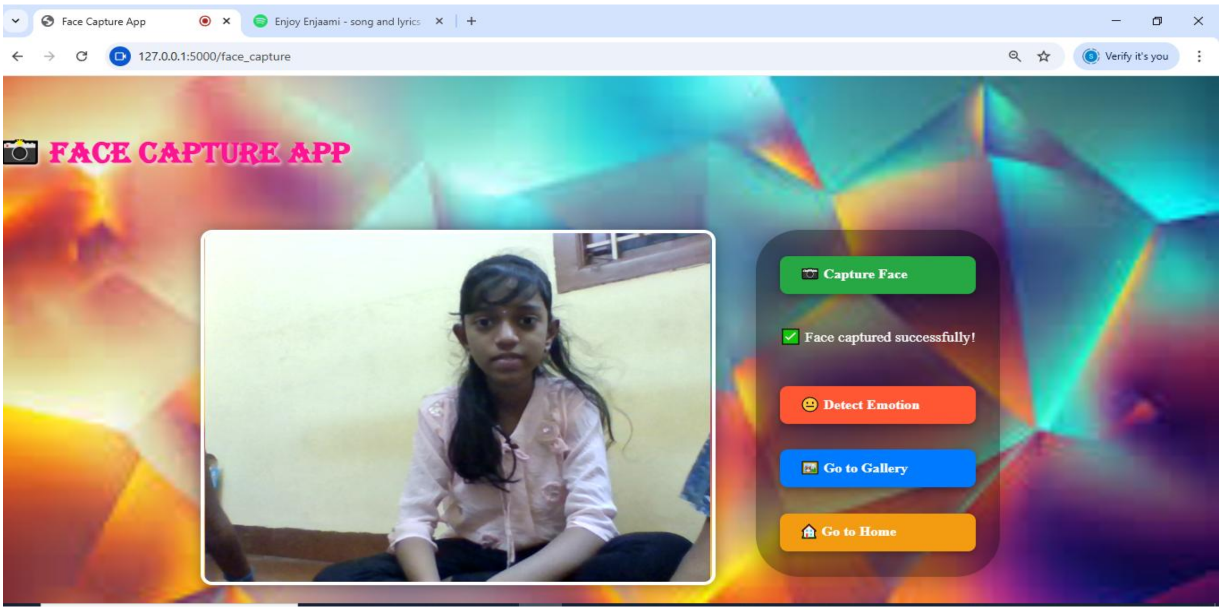

Figure 3: User interface for Face Captured Page

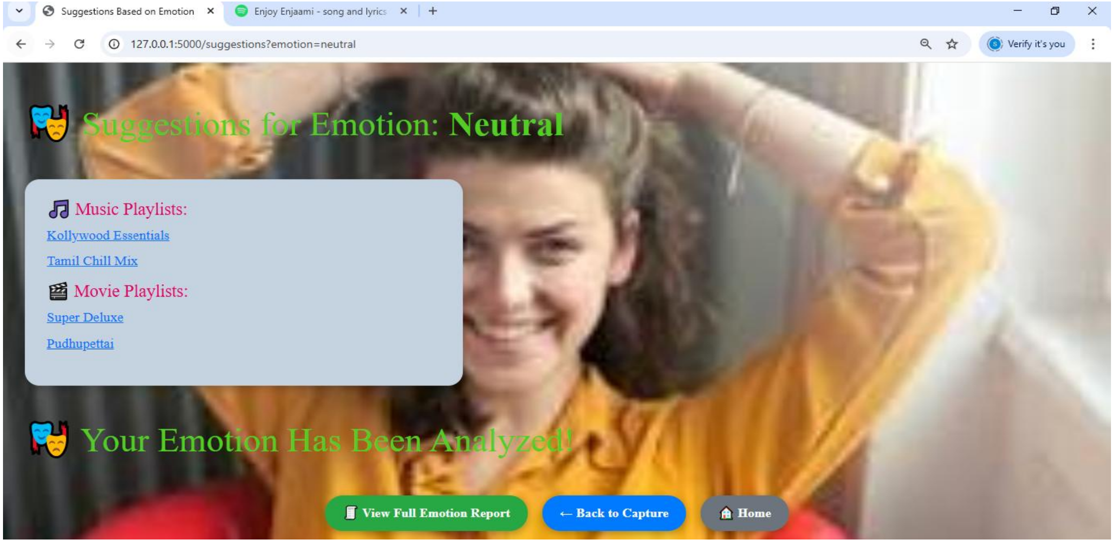

Figure 4: User interface for Song Suggestion Page

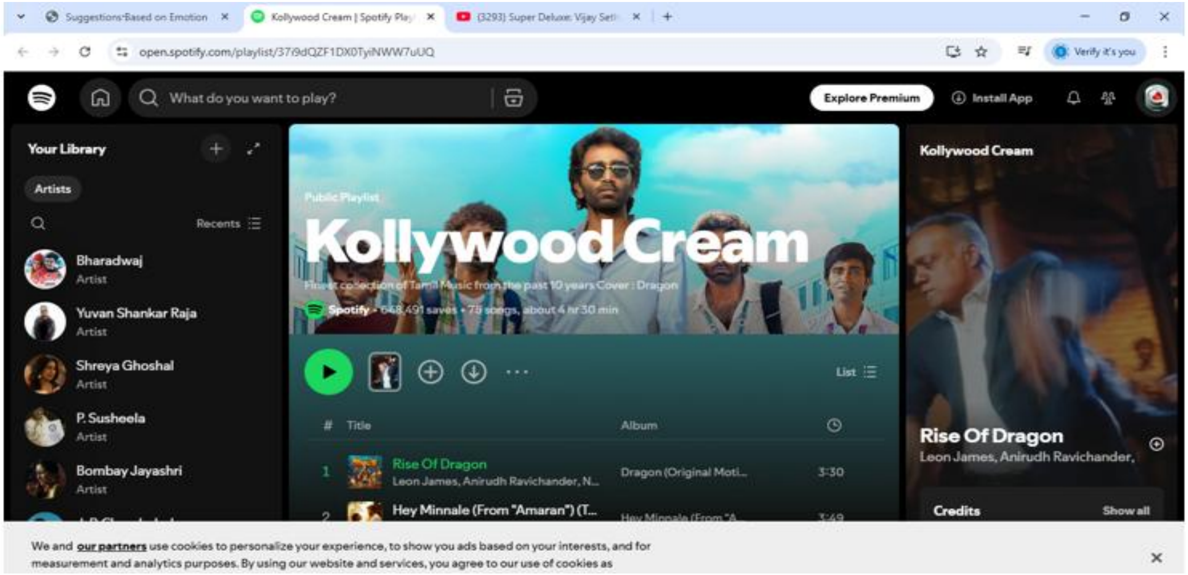

Figure 5: User interface for Song Playlists Page

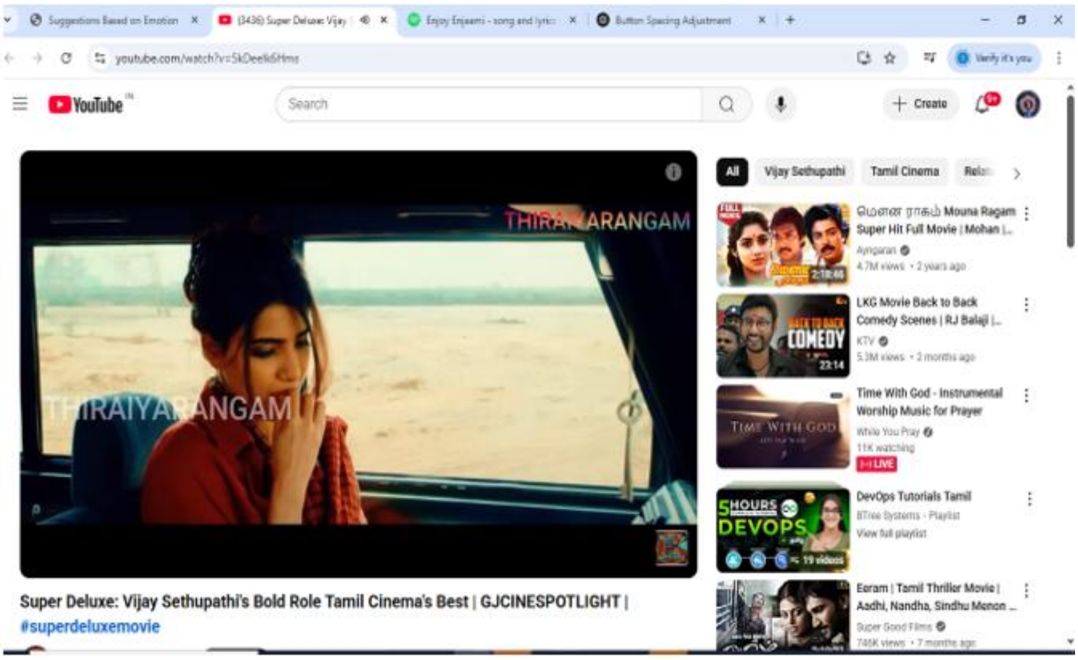

Figure 6: User Interface for Movie Playlists Page

6. Conclusions

This paper presented MoodSync, an intelligent, real-time emotion-aware recommendation system that enhances multimedia personalization through facial expression analysis. By leveraging computer vision and deep learning techniques, the system is capable of accurately detecting users' emotional states and mapping them to suitable music and movie content. This approach addresses the limitations of conventional recommendation engines, which often overlook the user's current emotional context and provide static, history-based suggestions.

The results demonstrate that integrating real-time emotion recognition significantly improves user satisfaction, relevance of recommendations, and overall user engagement. The system's ability to adapt dynamically to emotional feedback allows for more empathetic and human-centered content delivery, particularly in scenarios where user preferences are fluid or unexpressed.

Despite its promising performance, MoodSync faces challenges related to facial emotion recognition accuracy in uncontrolled environments, privacy concerns, and emotion ambiguity. These areas present opportunities for future work, including the integration of multimodal emotion inputs (e.g., voice, text sentiment), reinforcement learning for adaptive personalization, and ethical frameworks for data handling and user consent.

In conclusion, MoodSync exemplifies how affective computing can be effectively embedded into recommendation systems to bridge the gap between emotional intelligence and digital interaction. Its real-time, emotion-driven architecture marks a significant step toward more responsive and emotionally aware multimedia platforms.

7. Future Works

While MoodSync demonstrates the feasibility and impact of real-time emotion-aware content recommendation, several avenues remain for further development and refinement.

1) **Multimodal Emotion Recognition**: Future iterations of the system could incorporate additional input modalities such as voice tone analysis, textual sentiment analysis (from chat or user input), and physiological signals (e.g., heart rate or galvanic skin response) to improve emotion detection accuracy and robustness. Combining these modalities with facial expression recognition can enable a more comprehensive and nuanced understanding of the user's emotional state.
2) **Personalized Emotion-Content Mapping**: The current system uses a predefined mapping of emotions to media genres. Future work could focus on creating personalized emotion-to-content associations based on individual user preferences, cultural differences, and long-term behavior patterns, possibly through reinforcement learning or adaptive modeling.
3) **User Privacy and Ethical AI**: As the system deals with sensitive emotional data, addressing privacy and ethical concerns is paramount. Future research should explore privacy-preserving machine learning techniques, such as federated learning or on-device inference, to minimize data transmission and enhance user trust.
4) **Context-Aware Recommendations**: Incorporating contextual information such as time of day, location, recent activity, or weather conditions could further refine recommendations. This would allow the system to not only respond to emotional states but also consider external factors that influence media preferences.
5) **Extended Platform Integration**: Future versions could support broader integration with smart devices (e.g., smart TVs, wearables, virtual assistants) and content platforms beyond Spotify and TMDB. This would enhance accessibility and provide a seamless multi-device experience.
6) **Longitudinal User Studies**: Conducting extended user studies with diverse populations over time will be crucial to evaluating the system's long-term effectiveness, user satisfaction, and emotional impact. These studies can provide valuable insights into how users interact with emotion-aware systems in everyday settings.

By addressing these future directions, MoodSync can evolve into a more intelligent, ethical, and human-centric recommendation system, advancing both the technical and societal dimensions of affective computing in digital media.

References

[1] **Pathak, H. G., Arora, S., Gupta, R., & Abrol, V. (2024)** – *Music emotion recognition for intelligent and efficient recommendation systems*. In **TCCE 2023**, Springer.
[2] **Tennakoon, N., Senaweera, O., & Dharmarathne, H. A. S. G. (2024)** – *Emotion-based movie recommendation system*. **ICTer**, 17(1).
[3] **Tran, H., Le, T., Do, A., Vu, T., Bogaerts, S., & Howard, B. (2024)** – *Emotion-aware music recommendation*. **AAAI Conference**, 37(13), 16087–16095.
[4] **Kambham, S., et al. (2025)** – *Emotion Detection and Music Recommendation System*.
[5] **Begum, T., et al. (2024)** – *Face Melody-Mood-Aware Music Recommendation System with Facial Recognition*. IJRASET
[6] **Babu, T., et al. (2023)** – *Emotion-Aware Music Recommendation System: Enhancing User Experience Through Real-Time Emotional Context*.
[7] **Huang, Z., et al. (2022)** – *ADFF: Attention Based Deep Feature Fusion Approach for Music Emotion Recognition*.
[8] **Reddy, A., et al. (2024)** – *Music Recommendation Based on Facial Emotion Recognition*.
[9] **Florence, S. M., & Uma, M. (2020)** – *Emotional Detection and Music Recommendation System based on User Facial Expression*.
[10] **More, A. P., et al. (2024)** – *Music Recommendation System Using Facial Emotion Gestures*.
[11] **Thao, H. T. P., et al. (2020)** – *AttendAffectNet: Self-Attention based Networks for Predicting Affective Responses from Movies*.
[12] **Wang, S., et al. (2020)** – *Emotion-aware music recommendation using facial expression recognition*.
[13] **Zhang, Y., Wang, S., & Fu, Y. (2019)** – *Emotion-aware music recommendation using facial expression*

recognition. **Multimedia Tools and Applications**, 78(3), 3261–3277.

[14] **Huang, C., et al. (2019)** – *Facial expression recognition using deep learning: A survey*.

[15] **Delbouys, R., et al. (2018)** – *Music mood detection based on audio and lyrics with deep neural net*.

[16] **Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017)** – *AffectNet: A database for facial expression, valence, and arousal computing in the wild*. **IEEE Transactions on Affective Computing**, 10(1), 18–31.

[17] **Kaya, H., & Karpouzis, K. (2017)** – *Deep learning for emotion recognition in video*.

[18] **Chen, J., et al. (2017)** – *A deep learning framework for emotion recognition using facial expressions*.

[19] **Zhang, K., et al. (2016)** – *Joint Face Detection and Alignment using MTCNN*.

[20] **Tkalčič, M., & Chen, L. (2015)** – *Personality and recommender systems*.

[21] **Baveye, Y., et al. (2015)** – *Deep learning for video-based emotion recognition*.

[22] **Sharma, A., et al. (2015)** – *Emotion-aware music recommendation using facial expression and physiological signals*.

[23] **Khosla, A., et al. (2013)** – *Large-scale image memorability prediction*.

[24] **McDuff, D., et al. (2013)** – *Affectiva: Commercializing facial affect recognition*.

[25] **Soleymani, M., Pantic, M., & Pun, T. (2012)** – *Multimodal emotion recognition in response to videos*. **IEEE Transactions on Affective Computing**, 3(2), 211–223.

[26] **Liem, C. C. S., et al. (2012)** – *Predicting mood from multimodal music features*.

[27] **Yang, Y.-H., & Chen, H. H. (2011)** – *Music emotion recognition*.

[28] **Hanjalic, A., & Xu, L.-Q. (2005)** – *Affective video content representation and modeling*.

[29] **Zeng, Z., et al. (2009)** – *A survey of affect recognition methods: Audio, visual, and spontaneous expressions*.

[30] **Ekman, P., & Friesen, W. V. (1971)** – *Constants across cultures in the face and emotion*. **Journal of Personality and Social Psychology**, 17(2), 124–129.