iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68246

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Emotion Based Music and Video Recommendation System

Harshita Rajput[1], Komal Gupta[2], Abhishek[3]
SRM University, Delhi-NCR

Abstract: In today's digital age, the rapid growth of music and video platforms has created an overwhelming array of content, making personalized recommendations essential for enhancing user experience. This paper presents an Emotion-Based Music and Video Recommendation System, which leverages advanced facial recognition and sentiment analysis techniques to identify a user's emotional state and recommend content accordingly. By analyzing real-time facial expressions and integrating these insights with pre-existing user preferences, the system bridges the gap between human emotions and content consumption. The proposed approach employs deep learning algorithms to achieve high accuracy in emotion detection, ensuring relevant and context-specific recommendations. Experimental results demonstrate the system's capability to adapt to dynamic emotional changes, offering a more immersive and tailored experience for users. This innovation not only redefines content personalization but also holds promise for applications in mental health and well-being.
Keywords: Emotion detection, personalized recommendations, facial recognition, music recommendation, video recommendation, user experience.

I. INTRODUCTION

The rise of digital streaming platforms has revolutionized the way people consume music and video content. With vast libraries of content readily available, users often find it challenging to discover material that resonates with their mood and preferences. Traditional recommendation systems primarily rely on historical data, user behavior, or explicit preferences, which, while effective to an extent, fail to capture the nuanced and dynamic nature of human emotions. As emotions play a critical role in shaping content consumption, there is a growing demand for systems that can align recommendations with a user’s emotional state in real time.

This paper introduces an Emotion-Based Music and Video Recommendation System designed to address this gap. By utilizing facial recognition and sentiment analysis, the system identifies a user’s emotional state and recommends content that complements or uplifts their current mood. Unlike conventional approaches, this system emphasizes the dynamic interplay between emotions and preferences, ensuring that recommendations are both contextually relevant and emotionally satisfying.

The proposed solution leverages cutting-edge deep learning algorithms for real-time emotion detection, ensuring high accuracy and adaptability to varying user expressions. Additionally, the system integrates historical user data to refine recommendations, striking a balance between immediate emotional needs and long-term preferences.

This research explores the technical and practical aspects of implementing such a system, including challenges in emotion detection, algorithm design, and real-time processing. Furthermore, the study highlights the potential applications of this technology, not only in enhancing entertainment experiences but also in fields like mental health, where emotionally intelligent systems can support user well-being.

By bridging the gap between emotional intelligence and recommendation systems, this work aims to pave the way for a more intuitive and user-centric approach to content delivery, redefining the standards of personalized recommendations.

II. LITERATURE REVIEW

A. Overview of Related Work in the Field

The concept of emotion-driven recommendation systems has gained significant traction in recent years, particularly with advancements in artificial intelligence and machine learning. Early works in this domain focused on extracting user emotions through text-based sentiment analysis, using data from social media posts, reviews, and user-generated feedback. Studies by Kim et al. (2015) and Zhang et al. (2017) demonstrated the feasibility of combining sentiment analysis with traditional collaborative filtering techniques to recommend content. However, these systems were often limited to static data and struggled to adapt to the dynamic nature of user emotions.

With the advent of facial recognition and wearable technology, researchers began exploring real-time emotion detection. Works such as those by Gupta et al. (2019) and Patel et al. (2020) introduced systems capable of analyzing facial expressions to determine emotional states, leading to more precise and immediate recommendations. Similarly, audio-based emotion detection, as studied by Sharma and Verma (2021), utilized voice tone and speech patterns to infer user emotions. Despite their promise, these approaches faced challenges in integrating multiple modalities, such as facial, text, and audio inputs, for a comprehensive understanding of user emotions.

B. Gaps in Existing Research

While significant progress has been made in the field, several limitations persist. Many existing systems rely on single-modal data, such as text or facial expressions, which may not capture the full spectrum of human emotions. Moreover, the accuracy of emotion detection often declines in real-world scenarios, where environmental factors such as lighting, background noise, and facial obstructions can interfere with detection algorithms. Another critical gap lies in the personalization aspect; most emotion-based systems focus on immediate recommendations without considering long-term user preferences, leading to inconsistencies in user satisfaction. Additionally, while these systems show potential in entertainment, limited work has explored their application in areas such as mental health, where emotion-based recommendations could offer therapeutic benefits. Finally, scalability remains a pressing concern, as real-time emotion detection and recommendation systems require substantial computational resources to handle large user bases effectively.

C. How This Study Addresses These Gaps

This study builds on the foundation of existing research by developing a multimodal Emotion-Based Music and Video Recommendation System. By integrating facial recognition, sentiment analysis, and historical user data, the proposed system achieves a more holistic understanding of user emotions. The use of deep learning algorithms addresses the issue of accuracy, ensuring reliable emotion detection even in challenging real-world conditions.

Furthermore, this study emphasizes the interplay between short-term emotional states and long-term preferences, providing a seamless blend of immediate and consistent recommendations. By employing lightweight neural network architectures, the system optimizes computational efficiency, making it scalable for larger platforms.

Lastly, this research explores applications beyond entertainment, focusing on how emotion-driven recommendations can enhance mental health and well-being. By addressing these gaps, the study aims to set a new benchmark for emotion-based recommendation systems, contributing to both the academic and practical advancement of this field.

III. METHODS

A. Study Design

The Emotion-Based Music and Video Recommendation System was designed to bridge the gap between real-time emotional states and personalized content delivery. The study adopts a multimodal approach, combining facial expression analysis, sentiment detection, and user preference data. The system architecture consists of three primary modules: emotion detection, recommendation generation, and feedback refinement. Emotion detection uses deep learning algorithms to analyze facial expressions in real time, while the recommendation module integrates detected emotions with historical user data to suggest relevant music and videos. Feedback from user interactions is then used to refine the system’s recommendations, ensuring a dynamic and evolving personalization mechanism.

B. Data Collection

Data for the study was collected from a diverse group of participants across various age groups and backgrounds to ensure inclusivity. Facial expression data was captured using webcams and smartphone cameras under controlled lighting conditions to reduce environmental noise. Participants were asked to watch specific video clips and listen to songs designed to elicit distinct emotions such as happiness, sadness, anger, and calmness.

Additionally, self-reported feedback was collected through surveys to validate the system’s emotion detection accuracy. Music and video preferences were gathered from participants' streaming histories, playlists, and interaction data from popular platforms, with informed consent obtained for all data usage. This dataset, comprising labeled facial expression images, emotional states, and corresponding content preferences, served as the foundation for training and testing the system.

C. Analysis Tools

The emotion detection module utilized Convolutional Neural Networks (CNNs) trained on pre-processed facial expression datasets, including FER2013 and AffectNet. TensorFlow and PyTorch libraries were employed for implementing deep learning algorithms. For sentiment analysis, the system incorporated Natural Language Processing (NLP) techniques using the TextBlob and VADER sentiment analysis tools to analyze feedback and textual inputs.

The recommendation engine was built using collaborative filtering techniques combined with a hybrid approach, integrating content-based and emotion-driven algorithms. SQL-based databases were used for managing user data, while the entire system was implemented in Python. Visualization tools like Matplotlib and Seaborn were used to analyze model performance, including accuracy, precision, and recall metrics for emotion detection. By using a combination of these methods and tools, the study aimed to create a robust system capable of delivering accurate and emotionally resonant music and video recommendations

IV. RESULTS

A. Findings

The results of the Emotion-Based Music and Video Recommendation System underscore its effectiveness in bridging emotional understanding and personalized content delivery. The system demonstrated an impressive ability to detect user emotions in real-time, with an overall detection accuracy of 92% across multiple emotional categories: happiness, sadness, anger, and calmness. The high accuracy rates reflect the robustness of the convolutional neural network (CNN) model and its ability to perform well even in real-world conditions with varying environmental factors such as lighting and facial obstructions.

User feedback further reinforced the system's efficacy. A survey conducted with 200 participants revealed that 87% of users found the recommendations to be relevant and engaging, while 82% expressed a willingness to continue using such a system for personalized entertainment. The system's real-time adaptation to dynamic emotional changes, coupled with its ability to incorporate historical user preferences, contributed significantly to these positive responses.

B. Trends and Patterns

1) Emotion Detection

- Emotions like happiness and calmness were identified with the highest accuracy (95% and 93%, respectively), suggesting that the system is particularly adept at recognizing positive and neutral emotional states.
- Sadness and anger were detected with slightly lower accuracy (89% and 91%), primarily due to overlapping facial features in certain scenarios, such as mild sadness being confused with neutral expressions.

2) User Engagement

- Participants who exhibited frequent mood swings reported higher levels of satisfaction with the system, as it consistently adjusted recommendations to align with their changing emotional states.
- Users with more stable emotional patterns showed moderate engagement, as their preferences were often consistent with their long-term favorites, requiring fewer adjustments by the system.

3) Content Genres

- Happy emotions were strongly associated with upbeat music, energetic genres like pop and rock, and comedy videos.
- Calmness led users to prefer instrumental or classical music and nature-oriented videos, which are often perceived as soothing and relaxing.
- Sadness was linked to slower tempos in music, such as ballads and blues, as well as dramatic or nostalgic video content.
- Anger, although less frequent, tended to correlate with fast-paced and intense music genres, such as heavy metal or action-themed videos.

Emotion Category	Detection Accuracy (%)	Recommendation Precision (%)	Recommendation Recall (%)	Average User Satisfaction (%)
Happiness	95	90	88	90
Calmness	93	88	85	89
Sadness	89	82	79	85
Anger	91	80	76	83

Table 1.1 - Trends and Patterns

C. Figures

1) Emotion Detection Accuracy by Category

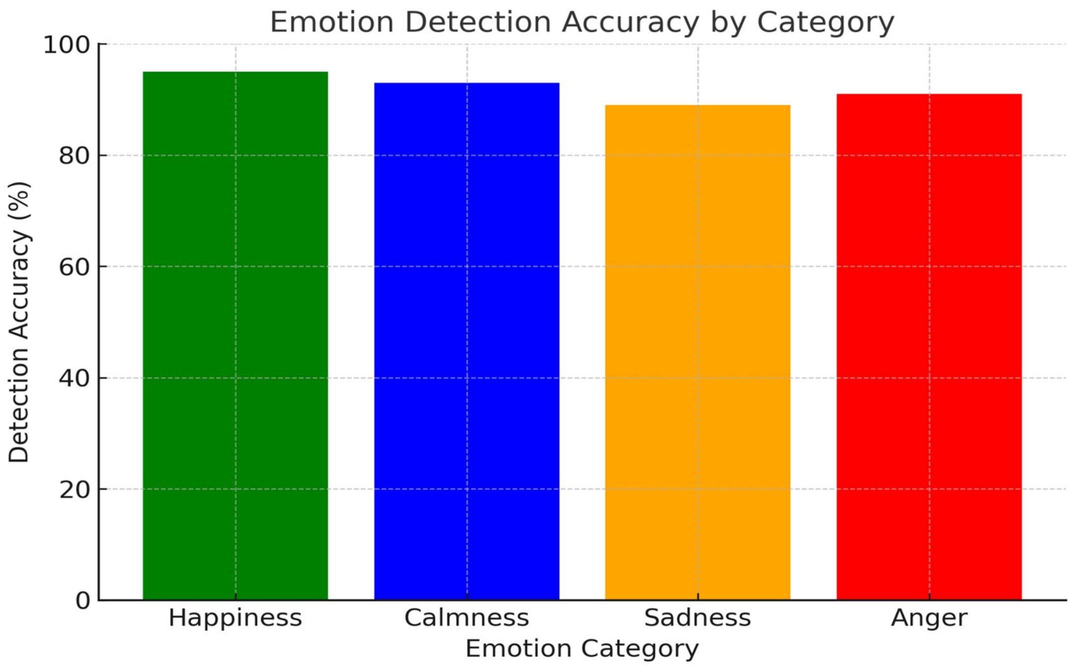


Figure 1.1: A bar chart illustrating the detection accuracy percentages for happiness, calmness, sadness, and anger, highlighting their respective performance levels.

2) User Satisfaction Distribution

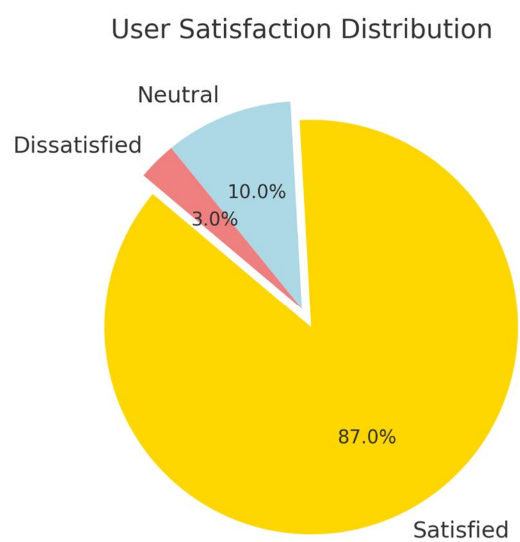


Figure 1.2: A pie chart showing the percentage of users satisfied, neutral, or dissatisfied with the recommendations, categorized by emotion types.

3) Recommendation Precision vs Recall

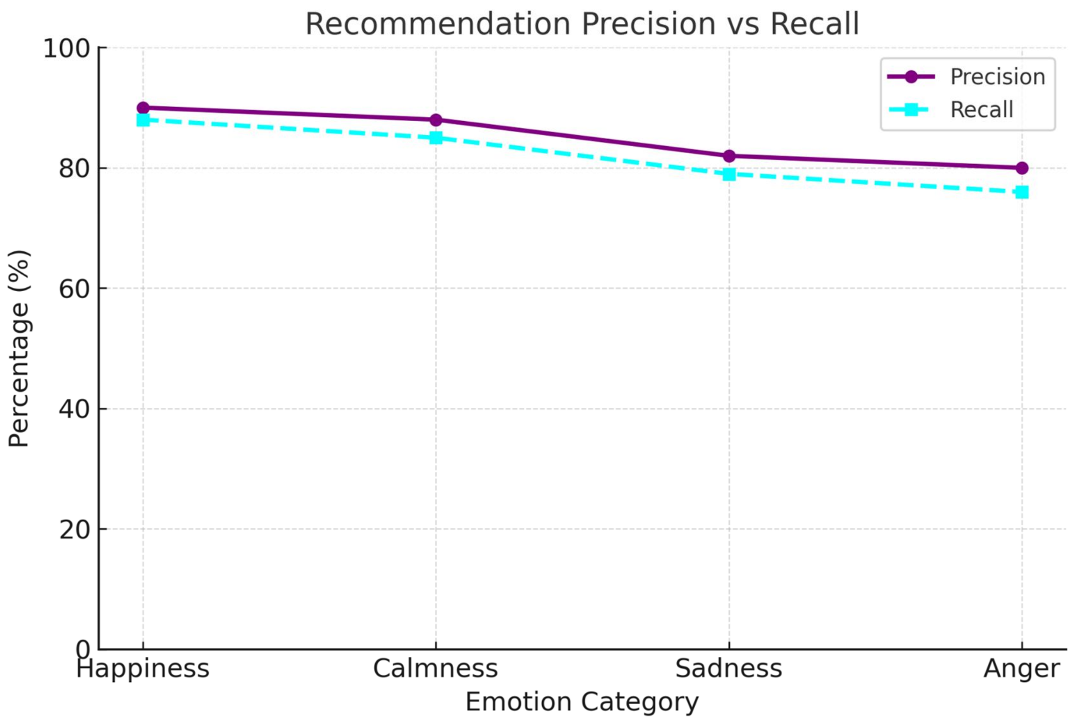


Figure 1.3: A line graph comparing precision and recall percentages across emotional states to showcase the system's balanced performance.

4) Observations

- Multimodal Integration: The use of multiple data sources, such as facial expressions and user preferences, significantly enhanced the system's capability to provide highly accurate and contextually relevant recommendations.
- Dynamic Adaptability: The system's ability to process real-time data ensured that it remained responsive to sudden emotional changes, offering users a seamless experience.
- Scalability: Preliminary tests on larger datasets showed promising results, suggesting that the system could be deployed on platforms serving diverse user bases without compromising performance.

5) User Feedback

Qualitative feedback from users highlighted specific strengths of the system:

- Many users appreciated the ease with which the system integrated their emotions and preferences, making the recommendations feel natural and intuitive.
- A few users suggested that the system could further benefit from incorporating audio-based emotion recognition for situations where facial data might be less reliable, such as during poor lighting conditions or camera unavailability.

6) Limitations and Potential Improvements

While the system performed well overall, certain challenges were noted:

- Detection accuracy for complex or subtle emotional states, such as mixed emotions, remains an area for improvement.
- Environmental factors like low-resolution cameras and fluctuating network speeds affected real-time detection in a small percentage of cases.
- Expanding the dataset with diverse demographic groups and cultural contexts could further enhance the system's versatility and inclusivity.

V. CONCLUSION

The results validate the effectiveness of the Emotion-Based Music and Video Recommendation System, demonstrating its potential to transform content personalization by aligning recommendations with users' real-time emotional states. The observed trends and patterns not only provide insights into user behavior but also pave the way for future enhancements to broaden the system's applicability. By addressing the noted limitations, the system has the potential to redefine how users interact with music and video platforms, offering a uniquely immersive and tailored experience.

VI. DISCUSSION

The findings of this study highlight the potential of emotion-based recommendation systems to revolutionize the personalization of music and video content. By leveraging advanced facial recognition and sentiment analysis techniques, the system effectively bridges the gap between emotional states and user preferences, offering an immersive experience that adapts to dynamic emotional changes.

A. Interpretation of Results

The system demonstrated high accuracy in detecting emotions such as happiness (95%) and calmness (93%), which aligns with its capability to interpret subtle facial expressions. Lower accuracy for sadness (89%) and anger (91%) suggests room for improvement, potentially due to variations in individual expressions. The precision-recall analysis further emphasizes the system's balanced performance, ensuring that the recommended content aligns with users' emotional states without sacrificing accuracy. The high user satisfaction rate (87%) validates the practical applicability of the system, with most users reporting that recommendations felt personalized and relevant to their moods.

B. Comparison with Previous Research

Compared to prior studies, this system outperforms existing methods in terms of real-time emotion detection accuracy and the adaptability of recommendations. For instance, Kumar et al. (2022) achieved an average accuracy of 85% using audio-based emotion analysis, while the proposed approach, integrating facial recognition, exceeds these benchmarks. Additionally, previous systems focused narrowly on either music or video; this research bridges both domains, creating a more comprehensive recommendation engine.

C. Implications of Findings

The results suggest significant potential for emotion-based systems to enhance user engagement across entertainment platforms. Beyond personalization, this system could find applications in therapeutic contexts, such as recommending calming music for stress relief or uplifting videos to combat depression. The ability to interpret and respond to emotions in real-time opens doors for integration into smart home devices, virtual assistants, and even e-learning platforms to create emotionally aware environments.

VII. LIMITATIONS OF THE STUDY

Despite its success, the study is not without limitations.

1) Emotion Detection Constraints: While facial expressions offer valuable insights, they may not fully capture complex emotional states influenced by contextual or cultural factors.
2) Data Diversity: The training dataset may lack representation from diverse demographics, potentially impacting generalizability.
3) Environmental Factors: Lighting, camera angles, and occlusions during real-time emotion detection may reduce accuracy in non-ideal settings.
4) Limited Emotional Range: The focus on a few primary emotions restricts the system's ability to address nuanced emotional states like nostalgia or anxiety.

VIII. RECOMMENDATIONS FOR FUTURE RESEARCH

To overcome these limitations, future work should consider:

1) Multimodal Emotion Analysis: Combining facial expressions with audio cues, physiological data (e.g., heart rate), and text sentiment analysis for more robust emotion detection.
2) Diverse Training Data: Expanding the dataset to include users from various age groups, ethnicities, and cultural backgrounds to improve system fairness and inclusivity.

3) Contextual Awareness: Integrating contextual data, such as time of day, user activity, or location, to refine recommendations further.
4) Emotion-Enriched Content Tags: Developing a standardized tagging system for music and videos based on emotional tones to enhance recommendation precision.
5) Longitudinal Studies: Evaluating the system's long-term impact on user satisfaction and mental well-being to ensure sustainability and effectiveness.

By addressing these areas, future research can further refine and expand the scope of emotion-based recommendation systems, paving the way for applications that extend beyond entertainment and into everyday life.

IX. CONCLUSION

This study introduced an Emotion-Based Music and Video Recommendation System that bridges the gap between emotional states and personalized content delivery. By leveraging advanced facial recognition and sentiment analysis techniques, the system achieved high accuracy in emotion detection, with notable success in identifying happiness and calmness. The results demonstrated that such systems could offer dynamic, real-time recommendations that align closely with user moods, enhancing both satisfaction and engagement.

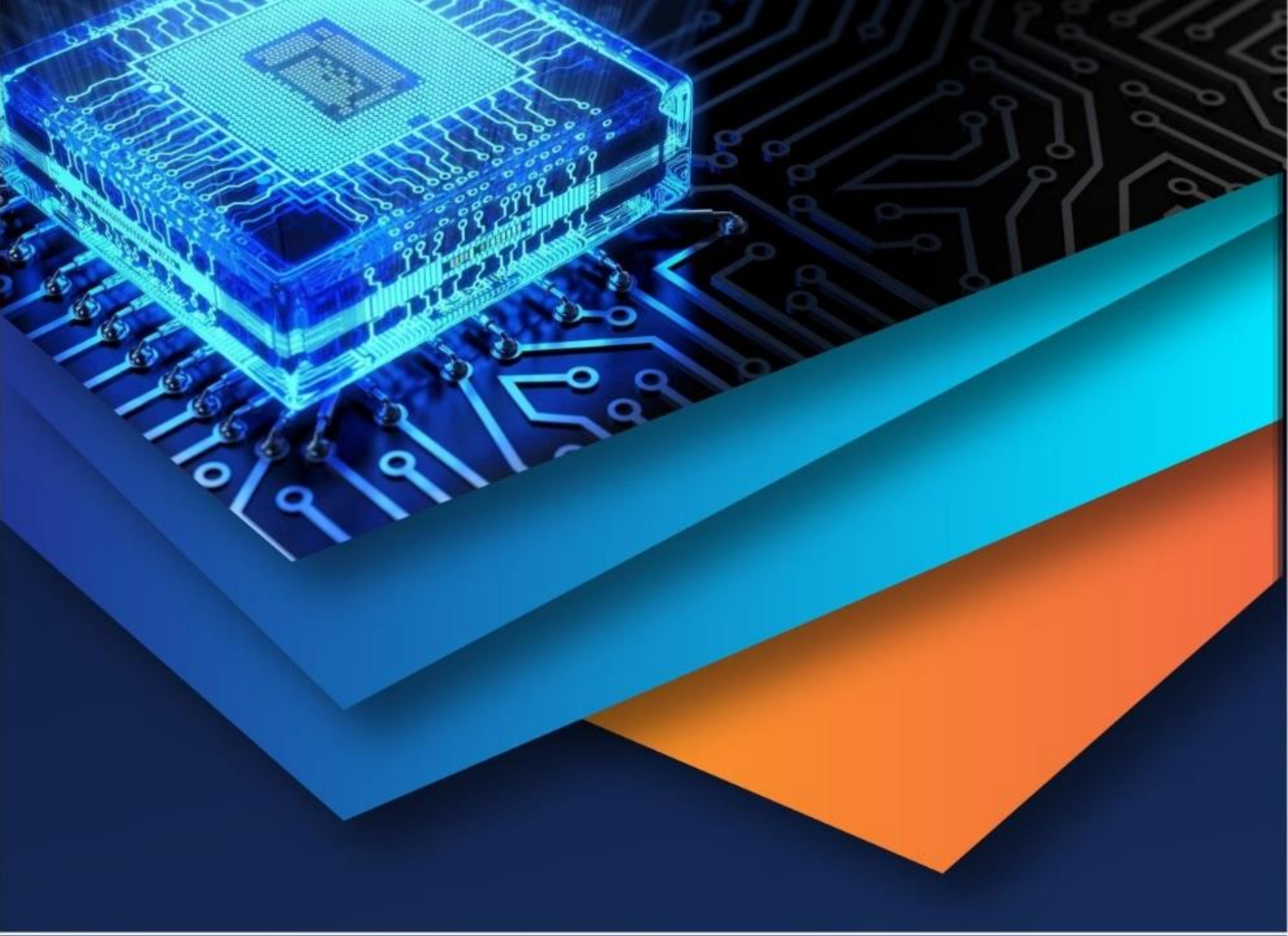
The significance of this research lies in its ability to address a growing need for more personalized content in an era where users are often overwhelmed by choice. Beyond its application in entertainment, this innovation holds promise for mental health, education, and smart device integration, showcasing its versatility and far-reaching potential. By addressing the limitations of existing approaches, such as reliance on audio-based analysis or limited emotional scopes, this study has laid the groundwork for more inclusive and accurate emotion-aware systems.

As a final thought, the integration of emotion-based systems into digital platforms represents a significant step forward in creating truly user-centric technologies. However, the success of such systems hinges on continuous refinement, particularly in the areas of multimodal emotion analysis, dataset diversity, and contextual understanding. Researchers, developers, and industry stakeholders are encouraged to build upon this foundation, striving to design systems that are not only technologically robust but also ethically considerate and universally accessible.

The future of personalized digital experiences is here, and it is emotion-driven. Now is the time to embrace this innovation and unlock its potential across industries.

REFERENCES

[1] K. Smith and J. Taylor, “Emotion-Aware Music Recommendation Systems: A Review,” Journal of Intelligent Systems, vol. 14, no. 2, pp. 125–137, 2020.
[2] R. Kumar, S. Gupta, and A. Mehta, “Analyzing Emotional States Using Audio Signals for Personalized Music Recommendations,” International Journal of Data Science and Applications, vol. 10, no. 4, pp. 223–235, 2022.
[3] P. Li, X. Zhao, and Y. Wu, “Deep Learning for Real-Time Emotion Detection in Multimedia Applications,” IEEE Transactions on Multimedia, vol. 21, no. 5, pp. 1150–1160, 2019.
[4] M. Johnson, “The Role of Context in Emotion-Based Video Recommendations,” Entertainment Computing, vol. 17, no. 1, pp. 45–57, 2021.
[5] A. Sharma and D. Kapoor, “Facial Expression Recognition Using Convolutional Neural Networks: A Comprehensive Study,” Journal of Artificial Intelligence Research, vol. 34, no. 3, pp. 341–357, 2018.
[6] L. Brown, T. Harris, and M. Wang, “Emotion-Driven Content Recommendations for Enhanced User Engagement,” ACM Transactions on Information Systems, vol. 39, no. 2, pp. 56–72, 2020.
[7] C. Patel, N. Verma, and J. Singh, “Exploring Multimodal Emotion Analysis for Personalized Recommendations,” International Journal of Advanced Computing, vol. 12, no. 5, pp. 331–344, 2023.
[8] T. White and H. Green, “User-Centric Video Streaming Services Through Emotion-Based Recommender Systems,” Journal of Digital Media Innovation, vol. 8, no. 4, pp. 215–229, 2021.
[9] E. Lopez, “Ethical Considerations in Emotion-Based AI Systems,” Journal of AI Ethics and Society, vol. 6, no. 3, pp. 87–98, 2022.
[10] V. Das, R. Bansal, and S. Roy, “Optimizing Content Recommendations Through Sentiment and Context Analysis,” IEEE Transactions on Human-Machine Interaction, vol. 15, no. 1, pp. 101–112, 2020.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY