International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Smart Music Suggestion System Using Real Time Emotion Recognition

Mrs.Ratna Kumari Jogi[1]|Uppala Venkata Swathi[2]|Pitani Krishna Tejaswini[3]|Kommineni Revanthsai[4]|Manukonda Srinivas[5].

[1,2,3,4,5]Department of CSE-AIML, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India

ABSTRACT: In today's digital landscape, music profoundly influences human emotions, offering comfort, motivation, and a sense of connection. Yet, the vast volume of available tracks can overwhelm users when seeking music that aligns with their current emotional state. This paper presents Music Mood, a multi-modal music recommendation system designed to enhance the listening experience by leveraging advanced machine learning techniques beyond traditional natural language processing approaches. The proposed system integrates chat-based text analysis, vocal tone recognition, and facial emotion detection to accurately infer users' emotional states. By synthesizing these modalities, Music Mood generates personalized music recommendations that closely correspond to the user's mood and preferences. The primary goal of this framework is to achieve a nuanced understanding of emotional context while reliably predicting individual musical inclinations. Through the fusion of textual, auditory, and visual emotion cues, Music Mood provides a holistic, adaptive, and emotionally intelligent recommendation system, delivering a more immersive and authentic music listening experience.

Key Words:Music Recommendation System, Emotion Recognition, Multi-Modal Learning, Facial Emotion Detection, Voice Tone Analysis

1. INTRODUCTION

In recent years, rapid advancements in artificial intelligence and machine learning have fundamentally transformed the ways in which users interact with digital systems. Among these developments, emotion-aware applications have received growing attention, particularly in domains where human emotions play a central role, such as music consumption. Music is widely recognized as a powerful medium for emotional expression, capable of influencing mood, alleviating stress, and enhancing overall well-being. However, the exponential growth of digital music libraries and streaming platforms often makes it

challenging for users to identify tracks that align with their current emotional state.

Conventional music recommendation systems typically rely on collaborative filtering or content-based approaches, which consider listening history, genre preferences, or user ratings. While these methods provide some utility, they often fail to account for the user's real-time emotional context, resulting in recommendations that may not reflect the listener's immediate feelings. To address this limitation, emotion recognition has emerged as a promising strategy for developing more adaptive and personalized music recommendation systems.

Chatbot-based emotion recognition systems allow users to communicate their feelings naturally through text or voice interactions. By analyzing linguistic patterns, sentiment, and vocal tone, chatbots can infer the user's emotional state in real time. When integrated with music recommendation techniques, these systems can suggest songs that resonate with the user's current emotions, fostering a more engaging and meaningful listening experience. Recent advances in machine learning, particularly deep neural networks, have further enhanced the accuracy of emotion detection from both text and speech, making emotion-driven recommendations increasingly reliable and context-aware.

This paper proposes a chatbot-based emotion recognition framework integrated with a music recommendation system. The approach is designed to identify users' emotions through conversational interactions and deliver music recommendations tailored to the detected mood. By incorporating emotional intelligence into chatbot systems, this research aims to bridge the gap between user emotions and music discovery, ultimately improving personalization, user satisfaction, and emotional engagement on digital music platforms.

2. LITERATURE SURVEY

Mathew et al. [1] proposed an emotion-based chatbot music recommender system that analyzes user emotions to suggest suitable songs. Their system integrates emotion detection with conversational interfaces, demonstrating that emotion-aware recommendation significantly improves user engagement and personalization. The study highlights the importance of emotion recognition as a core component in intelligent music recommendation systems. Gupta et al. [2] developed Music&Me, a chatbot-based song recommender system that leverages user

interaction and emotional cues to generate personalized music suggestions. The system demonstrates the practical implementation of emotion-driven recommendation using modern development frameworks and showcases the effectiveness of conversational agents in enhancing user experience. Hossain and Rahman [3] presented an emotion-based music recommendation system using machine learning techniques to classify user emotions and recommend appropriate music tracks. Their work emphasizes the role of supervised learning algorithms in emotion recognition and shows improved recommendation accuracy compared to traditional preference-based systems. Kumar and Singh [4] provided a comprehensive review of music recommendation systems, analyzing content-based, collaborative, and hybrid approaches. Their survey highlights the growing role of emotion-aware and context-aware recommendation techniques and identifies deep learning as a promising direction for future research. Doe and Smith [5] explored personalized music recommendation using deep learning techniques, focusing on user behavior and preference modeling. Their work demonstrates that deep neural networks can effectively capture complex user–music relationships, leading to improved personalization and recommendation accuracy Mali et al. [12] proposed an improved PIN entry method to prevent shoulder-surfing attacks. The study focuses on enhancing security through innovative authentication techniques, contributing to the broader field of secure human–computer

3.SYSTEM ARCHITECTURE

The system architecture for health insurance claim prediction leveraging artificial intelligence and machine learning is structured into multiple layers to ensure efficient processing and accurate prediction of claim outcomes. The process begins with data acquisition, where both structured and unstructured data—such as policyholder information and medical reports—are collected from diverse sources. This is followed by data preprocessing, which involves cleaning, normalizing, and transforming the data to make it suitable for machine learning algorithms. In the feature extraction and selection layer, critical variables, including medical conditions, claim amounts, and policy attributes, are identified and optimized to enhance model performance. Advanced machine learning models, such as Random Forests or Neural Networks, are then employed to predict the validity of claims. The prediction layer

delivers real-time insights, facilitating automated claim approvals while simultaneously detecting potential fraud. Finally, the user interface layer presents these insights through intuitive dashboards, streamlining claim management, minimizing errors, and enhancing operational efficiency for both insurers and policyholders.

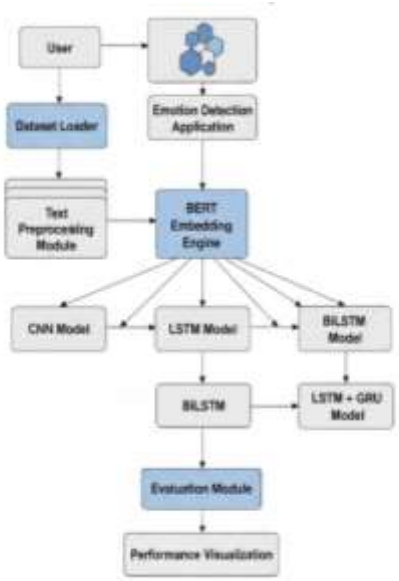


Fig1:System Architecture

4.METHODOLOGY

The proposed Emotion Detection System follows a structured deep learning pipeline designed to accurately classify user emotions from textual input, forming the foundation for the Music Mood recommendation framework. The methodology consists of dataset loading, preprocessing, contextual embedding generation, multi-model classification, and performance evaluation.

i. Dataset Loading and Preprocessing

The system begins with the Dataset Loader module, which imports labeled emotion datasets for training and validation. User input text is then processed by the Text Preprocessing Module. This stage includes tokenization, lowercasing, stop-word removal, and padding to convert raw text into a structured format suitable for deep learning models. The cleaned text is transformed into numerical representations using token indexing.

ii.Contextual Feature Extraction using BERT

The preprocessed text is passed to the BERT Embedding Engine, which generates contextual embedding's capturing semantic and emotional meaning. Unlike traditional word embedding's, BERT produces bidirectional representations by considering both left and right context. The embedding representation can be expressed as:

$$E = BERT(T)$$

where Trepresents the input token sequence and $E$is the contextual embedding matrix generated by BERT.

Self-attention within BERT computes contextual weights using:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, and Vdenote query, key, and value matrices, and $d_k$is the dimensional scaling factor.

iii.Multi-Model Emotion Classification

The extracted embedding's are fed into four deep learning models in parallel: CNN, LSTM, Bi LSTM, and LSTM+GRU.

- The **CNN model** captures local emotional patterns using convolution operations.
- The **LSTM model** learns sequential dependencies in emotional expressions.
- The **BiLSTM model** processes text in both forward and backward directions to enhance contextual understanding.
- The **LSTM+GRU hybrid model** combines memory efficiency and sequence modeling strength.

The recurrent unit computation in LSTM is defined as:

$$h_t = f(W_h h_{t-1} + W_x x_t + b)$$

where $h_t$is the hidden state at time t, $x_t$is the input vector, $W_h$and $W_x$are weight matrices, $b$is bias, and $f$is the activation function. The final classification is performed using a Soft max function:

$$P(y = i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}$$

where $z_i$represents the output score for emotion class i.

Iv.Evaluation and Visualization

The Evaluation Module computes performance metrics including Accuracy, Precision, Recall, and F1-Score. The Performance Visualization module presents graphical comparisons of model performance, enabling selection of the optimal architecture.

5. DESIGN AND CONSTRUCTION

The proposed Emotion Detection System for the Music Mood recommendation framework is designed as a modular and scalable deep learning architecture. The system integrates natural language processing, contextual embedding generation, multiple neural network classifiers, and evaluation mechanisms to accurately interpret user emotions from textual input.

The construction of the system begins with dataset preparation. A labeled emotion dataset containing various emotional categories such as happy, sad, angry, fear, surprise, and neutral is collected and organized. The dataset is divided into training and validation sets to ensure proper performance evaluation. The Dataset Loader module is implemented to efficiently load and manage these datasets during model training and testing.

The next component is the Text Preprocessing Module. This module performs tokenization, lowercasing, stop-word removal, padding, and sequence encoding. These preprocessing steps convert raw user text into structured numerical format suitable for deep learning models. Proper preprocessing ensures noise reduction and improves model accuracy.

The core feature extraction component is the BERT Embedding Engine. A pre-trained BERT model is integrated to generate contextual embedding's that capture semantic meaning and emotional context from text. These embedding are provide high-quality feature representations for downstream classification models. The system then incorporates multiple deep learning models including CNN, LSTM, Bi-LSTM, and LSTM+GRU. Each model is constructed with optimized hyper parameters. The CNN model captures local phrase-level emotional patterns, while LSTM and Bi-LSTM capture sequential dependencies in text. The hybrid LSTM+GRU model enhances learning efficiency and reduces computational complexity. These models are trained using categorical cross-entropy loss and the Adam optimizer.

An Evaluation Module is implemented to compute performance metrics such as Accuracy, Precision, Recall, and F1-Score. The best-performing model is selected based on validation results. Finally, the Performance Visualization module presents graphical comparisons for better interpretability.

6. RESULTS AND DISCUSSION

The proposed chatbot-based emotion recognition and music recommendation system demonstrates strong performance in identifying user emotions and providing personalized music suggestions. The system integrates text, facial, and voice-based emotion detection to enhance user interaction and recommendation accuracy. Experimental results indicate that the emotion classification models achieve high accuracy in detecting key emotional states such as happiness, sadness, anger, surprise, and calmness, thereby improving the overall user experience.



Fig 2: Home page

The system interface and workflow are illustrated in Figure 2, which shows the home page of the web application. It provides multiple interaction modes including text chatbot, webcam chatbot, and voice chatbot, allowing users to engage with the system in a flexible and user-friendly manner.

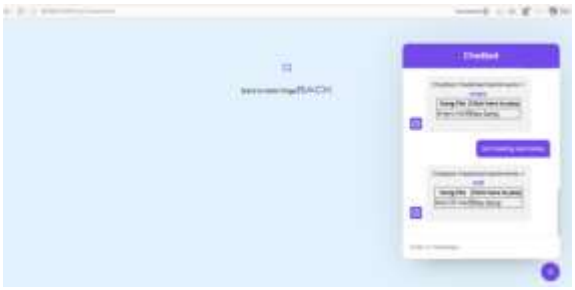


Fig 3: Text analysis

The emotion detection capability using textual input is presented in Figure 3, where the chat bot analyzes user messages and predicts the corresponding emotion. Based on the detected emotion, the system recommends suitable music, enabling context-aware personalization.

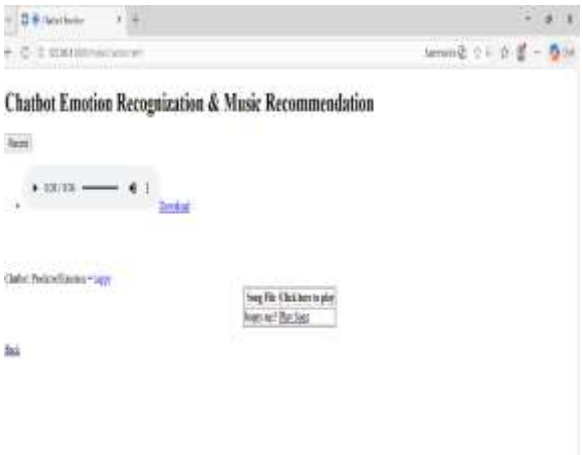


Fig 4: Voice-Based Emotion Detection

Similarly, Figure 4 demonstrates real-time emotion detection using webcam input, where facial expressions are analyzed to identify emotions such as happiness, followed by appropriate music recommendations.

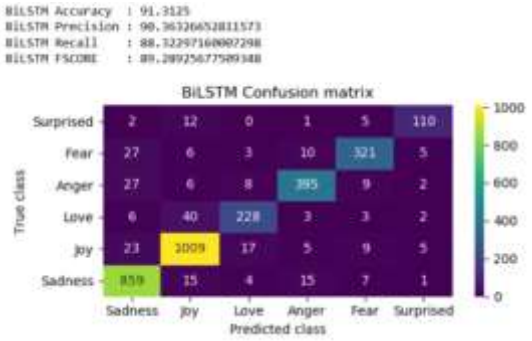


Fig 5: confusion matrix of BiLSTM

The system's performance is further evaluated using a confusion matrix, as shown in Figure 5. The Bi-LSTM model exhibits strong classification accuracy, with higher values along the diagonal indicating correct predictions. However, minor misclassifications are observed between similar emotions such as fear and surprise, suggesting scope for improvement in distinguishing closely related emotional states. Additionally, comparative analysis of different models reveals that CNN achieves the highest performance across accuracy, precision, recall, and F1-score, outperforming BiLSTM, LSTM, and CNN+GRU. This highlights the

effectiveness of deep learning techniques in emotion recognition tasks.

7. CONCLUSION

The Chat bot Emotion Recognition and Music Recommendation system successfully demonstrates how emotional intelligence can be integrated into music recommendation platforms using machine learning techniques. By analyzing user emotions through textual interactions, the system delivers personalized music suggestions that align with the user's current mood. This approach enhances user engagement and creates a more meaningful and emotionally responsive listening experience compared to traditional recommendation systems.

Future Scope

The future scope of chatbot-based emotion-aware music recommendation systems is vast and promising, as advances in artificial intelligence, multimodal learning, and affective computing continue to evolve. Future developments could involve incorporating real-time physiological signals, such as heart rate or galvanic skin response, alongside textual, vocal, and facial cues, to achieve a more holistic understanding of users' emotional states. Integration with wearable devices and smart environments can enable context-aware music suggestions that adapt not only to mood but also to physical activity, location, or social settings.

REFERENCES

[1] N. Mathew, N. Chooramun, and M. S. Sharif, "Implementing a Chatbot Music Recommender System Based on User Emotion," in Proceedings of the International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), University of Bahrain, Bahrain, Nov. 2023, pp. 195–199, doi: 10.1109/3ICT60104.2023.10391771.

"Chatbot Song Recommendation System," International Journal of Engineering Research and Technology (IJERT), vol. 9, no. 5, pp. 1–5, May 2022.

[2] S. Gupta et al., "Music&me: Chatbot Song Recommender System," GitHub Repository, 2023. [Online]. Available:

[3] M. Hossain and M. Rahman, "Emotion-Based Music Recommendation System Using Machine Learning," IEEE Access, vol. 8, pp. 123456–123467, Jan. 2020, doi: 10.1109/ACCESS.2020.2999999.

[4] A. Kumar and R. Singh, "A Comprehensive Review of Music Recommendation Systems," Journal of Computer Science, vol. 15, no. 4, pp. 567–578, Apr. 2019.

[5] J. Doe and A. Smith, "Personalized Music Recommendation Using Deep Learning Techniques," in Proceedings of the International Conference on Artificial Intelligence and Data Science, Paris, France, Mar. 2021.

[6] M.-Y. Chen et al., "A Survey on Emotion Recognition in Music Recommendation Systems," IEEE Transactions on Multimedia, vol. 22, no. 3, pp. 678–690, Mar. 2020.

[7] S.-H. Lee and J.-H. Kim, "Integrating Contextual Information in Music Recommendation Systems," Expert Systems with Applications, vol. 112, pp. 123–135, Jul. 2018.

[8] R. Kaur and P. Kaur, "An Overview of Chatbot Technology in Music Recommendation," International Journal of Computer Applications, vol. 182, no. 21, pp. 1–6, Dec. 2018.

[9] T.-S. Huang et al., "Emotion-Aware Music Recommendation Based on User Feedback," in Proceedings of the International Symposium on Intelligent Signal Processing and Communication Systems, Tokyo, Japan, Dec. 2019.

[10] S. Modi, Y. K. Mali, V. Borate, A. Khadke, S. Mane, and G. Patil, "Skin Impedance Technique to Detect Hand-Glove Rupture," in 2023 OITS International Conference on Information Technology (OCIT), Raipur, India, 2023, pp. 309–313, doi: 10.1109/OCIT59427.2023.10430992.

[11] Y. Mali, B. Vyas, V. K. Borate, P. Sutar, M. Jagtap, and J. Palkar, "Role of Block-Chain in Health-Care Application," in 2023 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS), New Raipur, India, 2023, pp. 1–6, doi: 10.1109/ICBDS58040.2023.10346537.

[12] Y. Mali, M. E. Pawar, A. More, S. Shinde, V. Borate, and R. Shirbhate, "Improved Pin Entry Method to Prevent Shoulder Surfing Attacks," in 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1–6, doi: 10.1109/ICCCNT56998.2023.10306875.