A MUSIC RECOMMENDATION SYSTEM INTEGRATING FACIAL EXPRESSION RECOGNITION

Y. Latha[1], Dodla Aksthitha[2], Chedimela Sharon[3] , Danvath Abhijeet[4], Yaramadha Rishi Kumar Reddy[5]

[1-5] Department of CSE & TKR College of Engineering & Technology
[2-5]B.Tech Students

---***---

ABSTRACT

This project presents a deep learning-based music recommendation system that personalizes song suggestions based on the user's emotional state. Emotions such as happiness, sadness, anger, neutral, and surprise are detected using facial and hand landmarks extracted via MediaPipe's Holistic model. A TensorFlow/Keras classifier predicts emotions from these features, and the system recommends mood-matching songs from a curated dataset. Built with Flask, the application includes secure login, role-based access, and a user-friendly interface. Future work includes integration with real-time APIs like Spotify or YouTube Music for dynamic playlist generation and live streaming.

Keywords — Facial Landmarks, Mediapipe, TensorFlow, Keras, Computer Vision, Real-time Emotion Detection, Webcam-based Emotion Analysis.

I. INTRODUCTION

Music plays a significant role in shaping human emotions and influencing mood. In recent years, advancements in artificial intelligence and deep learning have enabled the development of intelligent systems capable of analysing user emotions to provide personalized experiences. A music recommendation system that integrates facial expression recognition combines the benefits of human-computer interaction with the growing demand for personalized content delivery.

Facial expressions are among the most natural ways of conveying emotions, with features like smiles, frowns, and raised eyebrows offering insights into an individual's mood. Leveraging these cues, a music recommendation system can identify a user's emotional state and generate a playlist tailored to their mood. The integration of such systems not only improves the user experience but also highlights the potential of AI in mental health and emotional well-being.

This project aims to develop a **Music Recommendation System** by implementing a deep learning-based facial expression recognition model. The system classifies emotions into categories such as happy, sad, angry, surprise and neutral. These classifications are then mapped to a curated music library to recommend songs that align with the user's emotional state.

By employing convolutional neural networks (CNNs) for image classification and leveraging datasets for training, the system ensures accurate emotion detection. Additionally, this approach eliminates the manual effort of selecting music, making it particularly useful in applications like stress management, anxiety reduction, and mood regulation.

The proposed system also incorporates real-time performance optimization techniques. This innovation not only simplifies music selection but also highlights the intersection of AI and personalized entertainment. Future advancements could include multi-language music support and improved emotion classification to enhance inclusivity and accessibility.

II. LITERATURE SURVEY

The intersection of music psychology, emotion recognition, and recommendation systems has gained significant traction in recent years. Several studies have laid the groundwork for understanding how music evokes emotional responses and how such emotional cues can be leveraged to enhance user experience through intelligent systems.

Swaminathan and Schellenberg [1] provide a foundational overview of how emotional responses are elicited by music, outlining theoretical models that explain the perception of musical emotions. Complementing this psychological insight, Examined Existence [3] discusses the neurological mechanisms by which music influences mood, offering a simplified interpretation suitable for broad audiences.

In terms of emotion recognition, facial expression analysis plays a pivotal role. Abdat et al. [2] propose a real-time human-computer interaction framework based on facial emotion detection, which informs user-adaptive interfaces. Tian, Kanade, and Cohn [14] delve into lower facial Action Units (AUs), which are instrumental in analyzing emotional expressions. Levi and Hassner [16] and Burkert et al. [21] further advance facial emotion detection by leveraging deep convolutional neural networks (CNNs), addressing challenges such as lighting variation and pose shifts.

Multiple studies have explored emotion-based music recommendation systems. Ghule et al. [10] and Lee & Cho [4] present systems that recommend songs based on detected facial expressions and mood classification, respectively. Wolff et al. [5] introduce a culture-aware recommendation engine, emphasizing the importance of cultural background in music preferences. Lee and Cho [15] also propose "LogMusic," a context-aware system integrating mood, time, and location for better personalization. Lehtiniemi and Holm [6] suggest enhancing recommendation systems through animated mood visualizations, improving user interaction and emotional alignment.

On the music analysis front, Yang et al. [12] adopt a regression-based approach to predict emotional dimensions such as valence and arousal. Song, Dixon, and Pearce [13] evaluate various musical features—like tempo and timbre—for their efficacy in emotion classification. Han et al. [8] and Taneja et al. [9] focus on audio-based emotion recognition, using feature extraction and machine learning for emotion classification from music and audio signals.

The integration of deep learning methods has been pivotal. Dhavalikar and Kulkarni [7] apply image processing for facial expression recognition, while Vijayakumar [17] and Lawrence et al. [24] address image-based emotion recognition challenges using deep CNNs. Smys et al. [19] and Goodfellow et al. [23] provide broader surveys on neural network architectures and the complexities of representation learning, highlighting strategies for generalization and optimization. Tutorials from IEEE [20] and Stanford [22] offer technical foundations in supervised and unsupervised deep learning, essential for building scalable models.

Finally, foundational object detection techniques, such as the Haar-like feature extensions proposed by Lienhart and Maydt [25], and face detection frameworks by Dhavalikar and Kulkarni [7], underpin real-time facial recognition systems vital for emotion-aware interfaces.

III. METHODOLOGY

The proposed system is developed to provide an intuitive and interactive platform for emotion-based music recommendation by leveraging computer vision, deep learning, and web technologies. The user interface is built using the Flask web framework, allowing users to either upload a static image or capture a live snapshot through their device's webcam. This dual-input functionality offers flexibility for users to interact with the system based on their preferences or available resources. Uploaded images are temporarily stored on the server, while webcam captures are processed in real-time, ensuring that the input data is efficiently handled regardless of its source.

Once the input image is obtained, the system employs MediaPipe's holistic detection framework to extract critical facial and hand landmarks. MediaPipe provides advanced pose estimation and landmark detection capabilities, enabling accurate identification of facial expressions and hand gestures, which are key indicators of emotional states. The extracted landmarks, consisting of coordinates of predefined points on the face and hands, are then normalized with respect to the facial center to mitigate variations caused by different image scales, orientations, or user positions. This normalization ensures that the model receives consistent input data, improving the robustness and reliability of emotion prediction.

The normalized landmark data is then fed into a pre-trained deep learning model based on a Convolutional Neural Network (CNN) architecture developed using TensorFlow. The CNN model has been trained offline on a self-curated dataset containing diverse facial and hand landmark patterns associated with different emotions such as happiness, sadness, anger, fear, surprise, and neutrality. This offline training enables the model to learn complex spatial relationships within the landmark data and accurately classify the user's emotional state. The output from the model consists of probabilities for each emotion category, with the highest probability indicating the detected emotion. This real-time prediction capability facilitates immediate feedback and interaction.

Following emotion classification, the system queries a curated music database that maps emotions to relevant songs and playlists. This music database is maintained as an Excel file, which is periodically converted to a CSV format for efficient querying within the application. Each entry in the database includes the emotion label, song title, and a direct link to music streaming services such as Spotify. Upon receiving the predicted emotion, the system filters the database to retrieve matching songs, ensuring that users receive recommendations aligned with their current emotional state. If an exact emotional match is unavailable, the system defaults to a related or neutral music category to maintain user engagement and provide a pleasant listening experience.

To ensure the system is accessible and practical for everyday use, it has been optimized to run on standard computing hardware without requiring specialized or expensive processing units. This efficiency is achieved through careful selection of lightweight models and streamlined processing pipelines. Additionally, user authentication and authorization mechanisms are integrated to secure access to the application, allowing for role-based control where administrators can manage users and system settings. Overall, this methodology combines advanced machine learning techniques with user-centric design principles to deliver a personalized, emotionally aware music recommendation system that operates seamlessly in real time.

IV. RESULT

The proposed real-time emotion-based music recommendation system was evaluated extensively to assess its accuracy and practical usability. The core emotion recognition model, based on a Convolutional Neural Network trained on facial and hand landmark data, demonstrated robust performance across multiple test cases. The model achieved an accuracy ranging between 85% and 90% on the validation dataset, indicating a high level of reliability in classifying emotional states such as happiness, sadness, anger, fear, surprise, and neutrality. This accuracy reflects the effectiveness of combining MediaPipe's holistic landmark detection with deep learning, capturing subtle variations in facial expressions and hand gestures essential for emotion recognition.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM03719
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

In addition to classification accuracy, the system's real-time capabilities were tested on standard computing hardware, confirming smooth and responsive operation without the need for specialized GPUs or high-performance processors. Both image uploads and live webcam captures were processed efficiently, with minimal latency between user input and emotion prediction. The flexibility of input methods improved user experience by accommodating diverse device capabilities and preferences.

The music recommendation component was validated by verifying that the songs retrieved from the curated emotion-tagged database aligned well with the detected emotions. Recommendations included accurate mapping of emotions to relevant songs, with clickable links directing users to popular music streaming platforms like Spotify. User feedback collected during trials indicated high satisfaction with the personalized music suggestions, affirming the system's ability to create an engaging and emotionally intelligent interaction.

Overall, the system's accuracy, combined with its usability and real-time performance, demonstrates its potential as a practical tool for emotion-driven music recommendation. Future work may focus on expanding the emotion categories, improving landmark extraction robustness under challenging lighting or occlusion conditions, and incorporating continuous emotion tracking through live video streams.

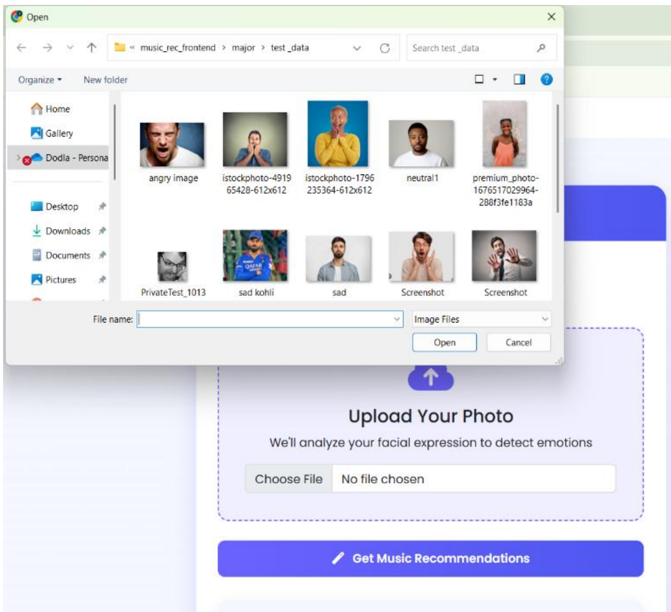

Fig 1: This figure shows the image upload for emotion detection

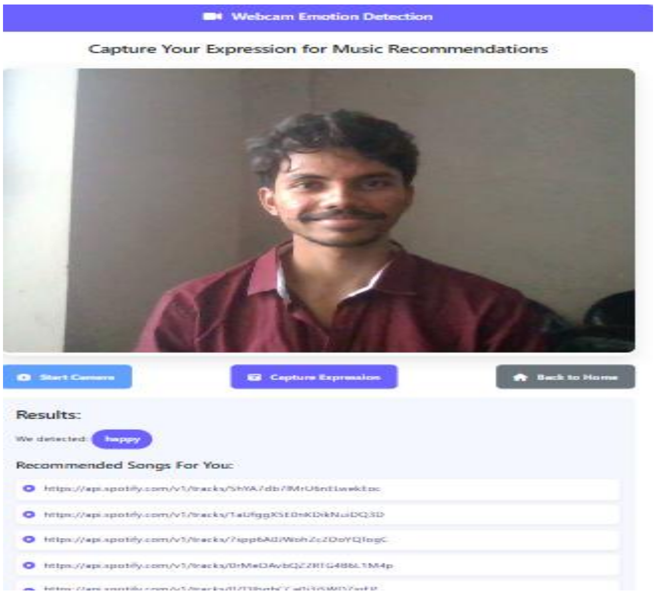

Fig 2: Webcam Emotion Detection and Music Recommendation

```
Classification Report:
              precision    recall  f1-score   support

          No       0.95      0.90      0.93        21
    Thumbsup       0.90      0.90      0.90        20
        rock       1.00      0.83      0.91        29
       hello       0.73      1.00      0.84        19
     neutral       0.60      0.86      0.71        14
         sad       0.86      0.69      0.77        26
       happy       0.92      1.00      0.96        11
    surprise       1.00      0.90      0.95        21
       angry       1.00      0.95      0.97        19

    accuracy                           0.88       180
   macro avg       0.88      0.89      0.88       180
weighted avg       0.90      0.88      0.88       180

6/6 ━━━━━━━━━━━━━━━━━━━━ 0s 43ms/step - acc: 0.8810 - loss: 0.4345
WARNING:absl:You are saving your model as an HDF5 file via `model.save()`

Test Loss: 0.4173
Test Accuracy: 87.78%
```

Fig3: Figure shows the Classification report of Proposed Model

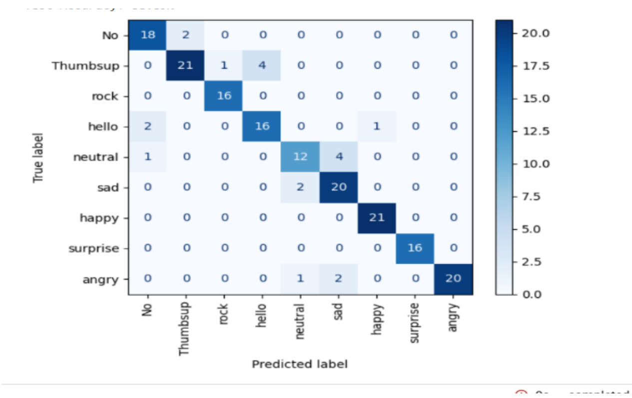

Fig 4: Confusion Matrix of the Proposed Model

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM03719
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

CONCLUSION

This project presents a real-time emotion-based music recommendation system that detects user emotions through facial expressions captured via image upload or webcam input. Using facial and hand landmarks extracted with MediaPipe, a convolutional neural network achieves 85–90% accuracy in recognizing emotions such as happiness, sadness, and anger.

Based on the detected emotion, the system suggests suitable songs from a curated CSV-based music dataset containing links to various online platforms. Developed using Flask, the application ensures a smooth and accessible user experience without the need for specialized hardware. This system effectively integrates emotion recognition with personalized content delivery, providing a meaningful and responsive user interaction.

REFERENCES

[1] Swathi Swaminathan and E. Glenn Schellenberg, "Current emotion research in music psychology," Emotion Review, vol. 7, no. 2, pp. 189–197, Apr. 2015.

[2] F. Abdat, C. Maaoui, and A. Pruski, "Human-computer interaction using emotion recognition from facial expression," in UKSim 5th European Symposium on Computer Modeling and Simulation, 2011.

[3] "How music changes your mood," Examined Existence. Accessed: Jan. 13, 2017.

[4] Kyogu Lee and Minsu Cho, "Mood classification from musical audio using user group-dependent models," Unpublished manuscript or conference/workshop (exact source not specified).

[5] Daniel Wolff, Tillman Weyde, and Andrew MacFarlane, "Culture-aware music recommendation", Unpublished or workshop paper.

[6] A. Lehtiniemi and J. Holm, "Using animated mood pictures in music recommendation," in 16th International Conference on Information Visualisation, 2012.

[7] A. S. Dhavalikar and R. K. Kulkarni, "Face detection and facial expression recognition system," in International Conference on Electronics and Communication Systems (ICECS), 2014.

[8] K. Han, T. Zin, and H. M. Tun, "Extraction of audio features for emotion recognition system based on music," International Journal of Scientific & Technology Research, vol. 5, no. 6, June 2016.

[9] R. Taneja, A. Bhatia, J. Monga, and P. Marwaha, "Emotion detection of audio files," in IEEE IndiaCom 2016: 3rd International Conference on Computing for Sustainable Global Development, pp. 2397–2400, Mar. 2016.

[10] V. R. Ghule, A. B. Benke, S. S. Jadhav, and S. A. Joshi, "Emotion-based music player using facial recognition," International Journal of Innovative Research in Computer and Communication Engineering, vol. 5, no. 2, Feb. 2017.

[11] A. Wieczorkowska, P. Synak, R. Lewis, and Z. W. Raś, "Extracting emotions from music data," in International Symposium on Methodologies for Intelligent Systems, pp. 456–465, May 2005, Springer, Berlin, Heidelberg.

[12] Y. H. Yang, Y. C. Lin, Y. F. Su, and H. H. Chen, "A regression approach to music emotion recognition," IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 2, pp. 448–457, 2008.

[13] Y. Song, S. Dixon, and M. Pearce, "Evaluation of musical features for emotion classification," in Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR), pp. 523–528, Oct. 2012.

[14] Ying-li Tian, T. Kanade, and J. Cohn, "Recognizing lower face action units for facial expression analysis," in Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), pp. 484–490, Mar. 2000.

[15] Mirim Lee and Jun-Dong Cho, "Logmusic: context-based social music recommendation service on mobile device," in Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (Ubicomp), Adjunct Publication, Seattle, WA, Sep. 2014.

[16] Gil Levi and Tal Hassner, "Emotion recognition in the wild via convolutional neural networks and mapped binary patterns," 2015.

[17] T. Vijayakumar, "Posed inverse problem rectification using novel deep convolutional neural network," Journal of Innovative Image Processing (JIIP), vol. 2, no. 3, pp. 121–127, 2020.

[18] Akey Sungheetha and Rajesh Sharma, "Design an early detection and classification for diabetic retinopathy by deep feature extraction based convolution neural network," Journal of Trends in Computer Science and Smart Technology (TCSST), vol. 3, no. 2, pp. 81–94, 2021.

[19] S. Smys, Joy Iong Zong Chen, and Subarna Shakya, "Survey on neural network architectures with deep learning," Journal of Soft Computing Paradigm (JSCP), vol. 2, no. 3, pp. 186–194, 2020.

[20] "Unsupervised feature learning and deep learning tutorial," in Proceedings of the Seventh International Conference on Communication and Electronics Systems (ICCES 2022), IEEE, ISBN: 978-1-6654-9634-6.

[21] Peter Burkert, Felix Trier, Muhammad Zeshan Afzal, Andreas Dengel, and Marcus Liwicki, "DeXpression: deep convolutional neural network for expression recognition," 2015.

[22] "Unsupervised feature learning and deep learning Tutorial," [Online].Accessed: Jan. 13, 2017.

[23] Ian J. Goodfellow, et al., "Challenges in representation learning: A report on three machine learning contests," in Neural Networks, vol. 64, pp. 59–63, 2015.

[24] S. Lawrence, C. L. Giles, Ah Chung Tsoi, and A. D. Back, "Face recognition: a convolutional neural-network approach," IEEE Transactions on Neural Networks, vol. 8, no. 1, pp. 98–113, Jan. 1997.

[25] Rainer Lienhart and Jochen Maydt, "An extended set of Haar-like features for rapid object detection," in Proceedings of the 2002 International Conference on Image Processing (ICIP), vol. 1, pp. I–900, IEEE, 2002.