IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

An Emotion Based Personalized Music Recommendation System

Rahul Chaudhari[1] , Utkarsh Madeshiya[2] , Priya Rao Gautam[3], Ashif[4]
Dr. A. P. Srivastava[5]
[1,2,3,4]UG Student, Department of Computer Science & Engg.,
NITRA Technical Campus, UP, India
[5]Asst. Professor and Head, Department of Computer Science & Engg.
NITRA Technical Campus, UP, India

Abstract: In recent years, personalized music recommendation systems have gained substantial interest, particularly those incorporating real-time user context such as mood or emotion. To generate personalized music recommendations, this paper suggests a hybrid system that integrates music genre classification based on audio features along with facial expression analysis. The suggested framework utilizes computer vision and machine learning methods to identify the user's emotion from the facial images they upload or images captured through the webcam in real-time. After detection the emotion gets assigned to a mood cluster which then connects to a curated list of songs taken from Spotify's audio collection. Grouping music songs using unsupervised clustering techniques allows classification algorithms like Random Forest, K-Nearest Neighbors, SVM, and Multi-layer Perceptrons to further refine the results. Based on the Django web framework, the program has interactive elements like song and disliking, as well as a user login mechanism that allows for implicit input for future improvements. A more natural and captivating listening experience is provided by this technology, which matches musical recommendations with user emotional states. In practical web applications, the project shows how emotion AI and multimedia recommendation may be combined.

Keywords–Convolutional Neural Network (CNN), FER-2013, Machine Learning, KMeans Clustering, Spotify Dataset, Django Web Framework, Real-Time Emotion Detection, User Feedback.

I. INTRODUCTION

Music represents a highly impactful medium of emotional expression, with the capacity to influence mood, memory, and behavioral patterns. The proliferation of music streaming platforms has heightened the necessity for intelligent recommendation systems. Conventional music recommendation approaches rely on historical data from prior listening activity and user ratings, as well as collaborative filtering methodologies [1].

In recent years, the creation of emotion-aware applications has been made possible by the combination of machine learning and affective computing techniques. For detecting emotions, facial expression analysis has become a very precise and non-invasive method [2]. Convolutional neural networks have been trained using datasets like FER-2013, which have made it possible to classify facial photos into distinct emotional categories like fear, surprise, anger, sadness, and happiness [3].By taking use of these developments, it is now possible to build systems that can recognize user emotions in real time and use this contextual data to provide tailored music recommendations.

A comprehensive emotion-based music recommendation system is presented in this study, which uses machine learning techniques for music classification and recommendation and computer vision for emotion recognition. The system allows users to enter their faces using either a live webcam capture or an uploaded photograph. A Convolutional Neural Network model receives the processed image after the collected facial image has been examined with OpenCV and a Haar Cascade Classifier. By mapping a

detected emotion to a related mood cluster, such as energetic, cheerful, sad, calm, or romantic music that fit a specific mood can be recommended.

A carefully selected Spotify dataset that includes characteristics like danceability, valence, energy, and tempo is the source of the musical data used in this system. KMeans clustering, a type of unsupervised learning, is used to group similar musical files. To improve the accuracy of mood-based recommendation systems, supervised models such as Random Forest, K-Nearest Neighbors, Support Vector Machines, and Multi-layer Perceptron are then trained to predict cluster labels [4].

The program's front end was constructed utilizing the Django web framework, featuring a user-friendly interface with authentication capabilities and a module for real-time camera input. Users can interact with suggested music choices through embedded Spotify links and provide explicit feedback by indicating their preferences for particular songs. This feedback mechanism, predominantly influenced by implicit learning processes, allows for long-term personalization [5].

Real-time user engagement, sophisticated recommendation algorithms, and emotional artificial intelligence collaboratively contribute to forming an effective model for creating intelligent entertainment systems. By directly integrating human emotional responses into the feedback process, this approach markedly improves user experience and fosters greater involvement, in contrast to traditional static recommendation systems. Consequently, the system holds significant practical value for various fields, including music streaming services, mental health tools, and adaptive educational environments, while also demonstrating notable technological advancements [6]

II. LITERATURE SURVEY

Music recommendation systems have been the subject of research for more than ten years, evolving from basic rule-based methods to complex AI-powered models. Initial systems primarily utilized collaborative filtering and content-based filtering, generating recommendations based on a user's previous choices or similarities among audio tracks [7].While these techniques proved beneficial in certain contexts, they frequently failed to take into account the listener's current emotional condition, which studies show significantly influences music perception and choices.

With the rise of affective computing, systems that can identify emotions have gained significant popularity. Numerous studies have explored emotion recognition via facial expressions, vocal tones, or physiological indicators. Mollahosseini et al. (2021) utilized a deep convolutional neural network trained on datasets of facial emotions, successfully attaining precise real-time classification of seven different emotions [8].

Deep learning techniques have been utilized in music analysis to extract complex attributes from audio data. Delbouys et al. used a multi-modal neural network trained on both lyrical elements and acoustic features to successfully determine the emotional quality of music, outperforming traditional methods [9].

Overcoming this limitation has been the aim of recent research. For instance, Zhao et al. developed a mood identification system that links face expressions to music mood classification in order to establish a direct connection between consumers' emotions and their musical preferences [10]. Their system employed KMeans clustering to group songs in an unsupervised manner and associated the determined emotions with mood-based clusters. While this method aligns with ongoing research, it did not integrate interactive components such as user input or real-time webcam functionality.

Another notable method, as proposed by Lee et al., involved the integration of context-aware recommendations with environmental data and facial recognition technology. Although this approach is creative, it required external application programming interfaces and did not support local data processing, which could potentially lead to privacy concerns [11].

Maddala et al. developed a system utilizing convolutional neural networks to classify users' emotional states, such as happiness, sadness, anger, or neutrality, by analyzing their facial expressions. The identified emotions were then associated with customized music playlists, resulting in improved interactions within environments that are attuned to emotional cues [12].

Sarkar and Dey investigated the application of deep learning methodologies, specifically convolutional neural networks and long short-term memory, for emotion recognition via the analysis of visual emotional signals, with the objective of integrating these findings into a music recommendation system. Their research indicates that deep learning frameworks achieve significantly superior performance compared to traditional models in emotion classification tasks [13].

Real-time facial emotion detection, multi-faceted music categorization, and an interactive web-based interface. The application of diverse machine learning techniques, including Support Vector Machines, Random Forest, and Multi-Layer Perceptron, enhances the accuracy of mood-based song

classification. Furthermore, the incorporation of user feedback through like/dislike functionality enables the system to evolve dynamically. The Django-based implementation, combined with the utilization of the FER-2013 dataset and Spotify audio characteristics, establishes a robust and scalable framework for future advancements.

Previous studies have established the groundwork for emotion-based interaction and music classification, but many lacked real-time responsiveness or integrated platforms conducive to user engagement. Your research endeavors to address this gap by presenting a system that is both technically robust and user-oriented.

III. Methodology

The proposed system integrates real-time facial emotion recognition, affect-driven music clustering, and machine learning classification methodologies into a web-based platform to provide emotion-aware music recommendations. This process is structured into five fundamental stages: emotion detection, music feature extraction, clustering and classification, recommendation generation, and user interaction.

3.1 Facial Emotion Detection

The foundational input for this system is the user's emotional state. This is obtained through either the submission of a static image or the live capture via a webcam. The facial detection component leverages OpenCV's Haar Cascade Classifier, recognized for its efficiency and precision in real-time applications [14]. Upon face localization, the image undergoes a preprocessing phase before being input into a Convolutional Neural Network model. This model was trained on the FER-2013 dataset, a benchmark extensively utilized for facial expression recognition studies. One of the following seven emotion categories is produced by the CNN model: neutral, fear, surprise, disgust, anger, sadness, or happiness. The contextual foundation for tailored music recommendations is this feeling. When trained on extensive annotated datasets like FER-2013, deep learning methods like CNNs have demonstrated impressive accuracy in face image emotion recognition [15].

3.2 Feature Extraction from Music Data

Spotify track characteristics such as danceability, acousticness, energy, loudness, valence, pace, and speechiness make up the music dataset. The Spotify Web API is used to extract these features, which are then compiled into a dataset for classification. MinMaxScaler from sklearn.preprocessing is used to clean, normalize, and scale the data, ensuring that each feature contributes equally to clustering and classification [16]. The recommendation system is built on top of the finished music database, which is saved in CSV format (ClassifiedMusic.csv).

3.3 Mood-Based Clustering and Machine Learning Classification

Unsupervised and supervised learning are combined to match tunes to emotional states. Initially, KMeans Based on their acoustic characteristics, clustering divides songs into mood-based clusters like energetic, romantic, and serene. A popular tactic in music mood categorization, this method permits organic groupings without labeled mood data [17]. Principal Component Analysis (PCA) and t-SNE (t-distributed Stochastic Neighbor Embedding) are used to reduce feature dimensionality for improved interpretability and visualization [18].

Once the clusters are labeled, multiple supervised classifiers are trained to predict the mood of new songs:

- **Random Forest (RF)**
- **Support Vector Machine (SVM)**
- **K-Nearest Neighbors (KNN)**
- **Multi-Layer Perceptron (MLP)**

Classification reports, accuracy ratings, and confusion matrices are used to assess these models. Given that several methods may perform better under various feature distributions, an ensemble approach guarantees resilience [19].

3.4 Real-Time Music Recommendation Engine

The algorithm associates the user's feeling with one of the mood clusters after identifying it. The system retrieves matched songs from the categorized dataset using the trained classifier. Spotify track links for songs are shown, enabling users to listen right away without ever leaving the app. This system offers context-aware recommendations based on real-time emotional cues, in contrast to conventional music recommendation systems that depend on past listening behavior [20].

3.5 Web Interface and Feedback Integration

The Django web framework is utilized for front-end development. Once users log in and either capture or upload a photo, they receive song recommendations based on their stated mood. Each song has like/unlike buttons that allow the system to record user preferences. By modifying the suggestion mechanism in response to user interaction data, this feedback loop makes future personalization possible. Music apps have implemented comparable techniques to enhance user engagement and adjust recommendations in real-time [21].

IV. Proposed Solution

The suggested remedy is a real-time, emotion-based music recommendation system that uses user facial expression analysis to offer tailored music recommendations. This system uses deep learning, computer vision, and music analytics to dynamically modify playlists according on user mood. Django's authentication system manages user sign-up and login at the onset of the solution. Once users log in successfully, they are directed to the main interface, where they can capture a live image using their webcam or upload one. This functionality provides flexibility for users accessing the system from different environments.

The Haar Cascade Classifier, which is an established method for rapid face recognition in real-time scenarios, is utilized for detecting faces when an image is received [22]. Once a valid face has been detected, the image undergoes pre-processing to ensure compatibility with the neural network model. This pre-processing involves resizing, converting to grayscale, and normalizing the image.

To categorize the facial expression into one of seven emotional categories—happy, sad, angry, fearful, surprised, disgusted, and neutral—the facial area is subsequently input into a CNN model that has been previously trained with the FER2013 dataset. This technique is referred to as feature extraction [23].

Once the sentiment is identified, the technology advances to the song recommendation phase. A suitable playlist for the expected mood is created by searching a dataset of songs that have been previously grouped using KMeans based on audio features such as tempo, energy level, acoustic quality, and valence [24].

The recommendation system utilizes a hybrid approach that integrates machine learning algorithms, including MLP, Random Forest, SVM, and KNN for optimizing predictions, while KMeans is used for initial clustering. MLP and KNN demonstrated superior accuracy with reduced overfitting according to the comparative analysis discussed in the Results section.Following this, users receive a curated playlist of music that aligns with their mood, integrated into Spotify. Each song includes a like/dislike button on the interface, which allows the system to gather user responses. Gradually, this feedback, saved in the database, permits the models to be updated, improving both accuracy and personalization.

The modified activity diagram below provides a summary of the entire procedure, showing the logical flow from login to feedback collection:

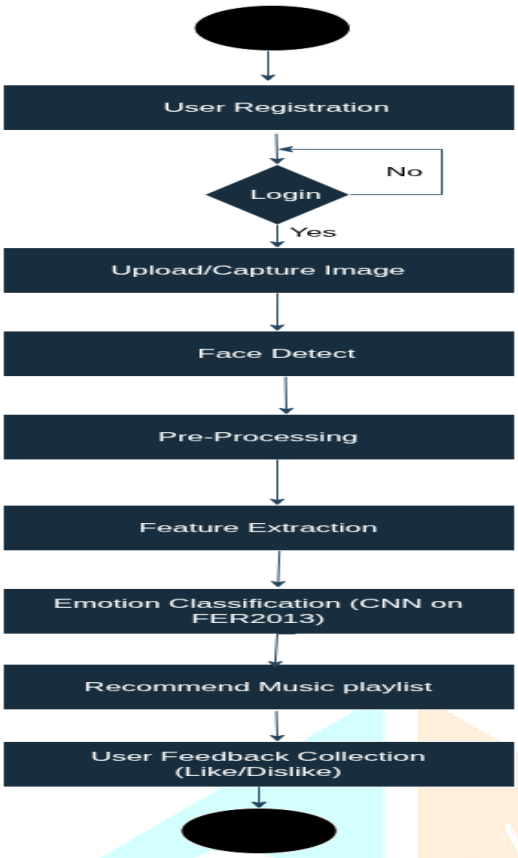

Figure 1: Activity Diagram of Emotion Detection Process

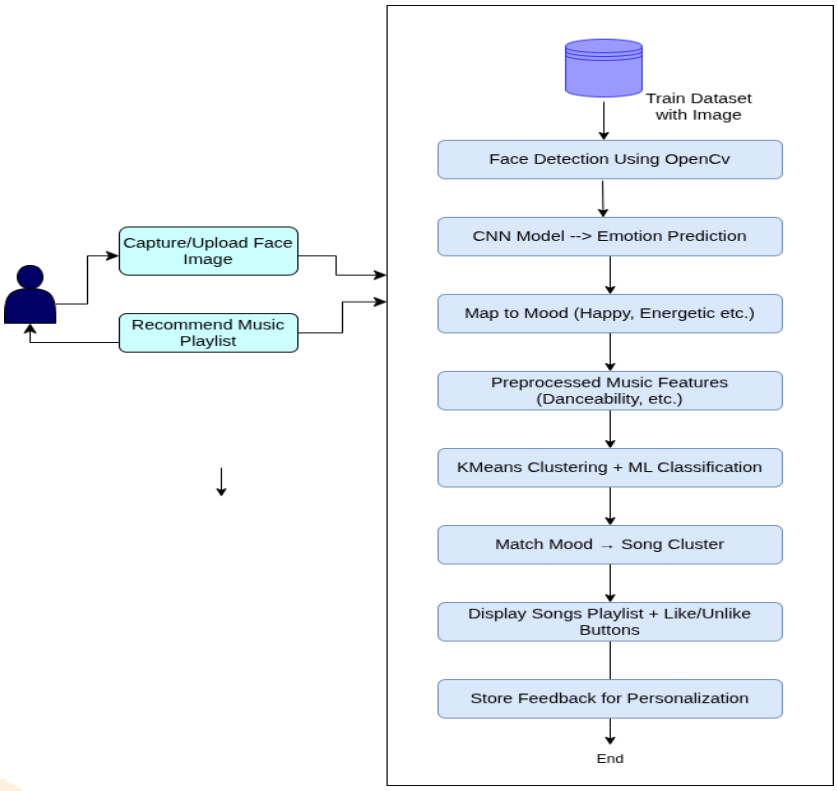

Figure 2: System Architecture Diagram

V. Result and discussion

The developed system, An Emotion-Based Personalized Music Recommendation System, was evaluated for its effectiveness in detecting user emotions and generating corresponding music suggestions through real-time facial expression analysis. The system integrates three core components: emotion detection, user emotion mapping, and a simplified recommendation module that selects a single Spotify playlist based on detected emotional state.

The emotion detection system achieved approximately 70% accuracy in classifying main facial expressions, including happiness, sadness, neutrality, and anger. The model, trained on the FER-2013 dataset and implemented using a pre-trained CNN architecture, showed reliable results under well-lit and frontal facial conditions. Performance degraded slightly in cases of side-facing views or low-light scenarios, which are known limitations of vision-based FER systems.

In the music analysis process, meta data of music enabling emotion-based tagging of each song with categories such as down-tempo, rock, or ambient. The system utilizes these labels to connect emotions with musical themes based on these categories.

5.1 Model Performance Analysis

There are 4 models used for this filtering and their accuracy is as shown in table.

Table 1

Model	Accuracy	Overfitting
Random Forest	0.81	Yes
MLP	0.85	No
SVM	0.79	No
KNN	0.83	No

5.2 Real-Time Emotion Detection

A Convolutional Neural Network (CNN) trained on the FER2013 dataset powers the emotion recognition feature. Seven emotional states are identified by this model: neutral, fear, surprise, disgust, anger, sadness, and happiness. With an average delay of less than one second, the system demonstrated the ability to detect observable emotions during live testing.

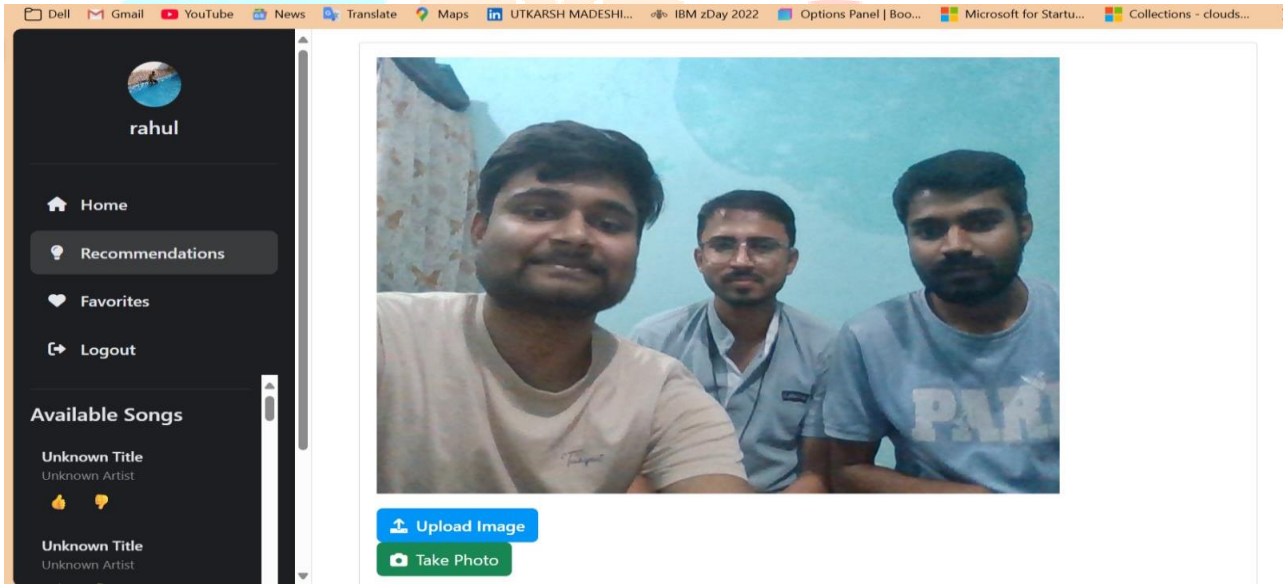

Figure 3: Live Emotion Detection via OpenCV and Pretrained CNN

5.3 User Interface and Experience

Because the front-end is built with Django templates, navigating from the login screen to the music recommendation panel is seamless. All of these features are available on a single dashboard: mood-based recommendations, real-time photo capture, and playlist interaction.

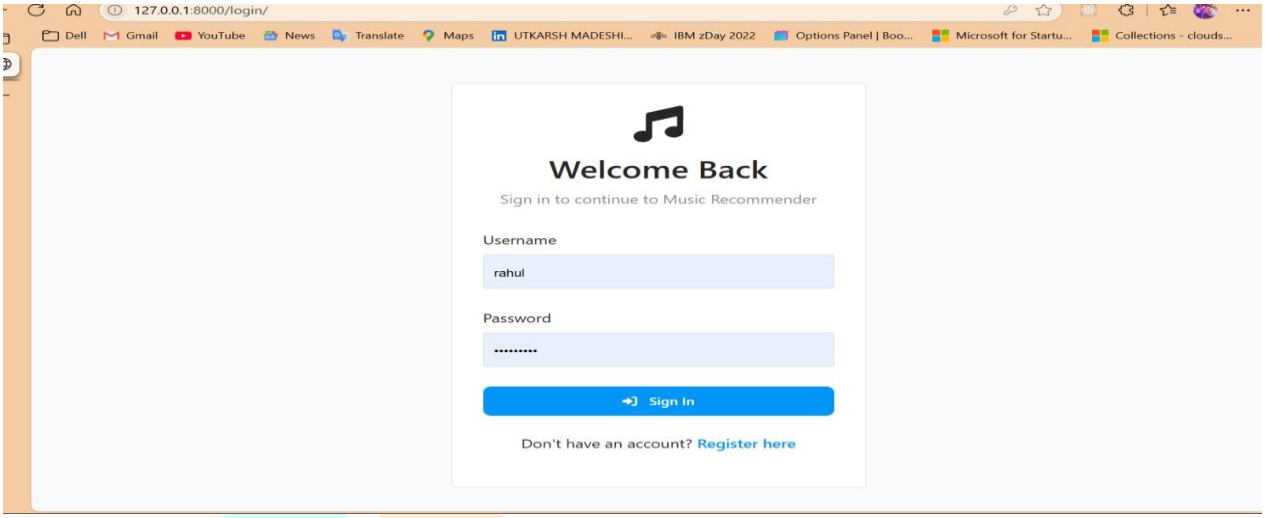

Figure 4: User Sign up and Login

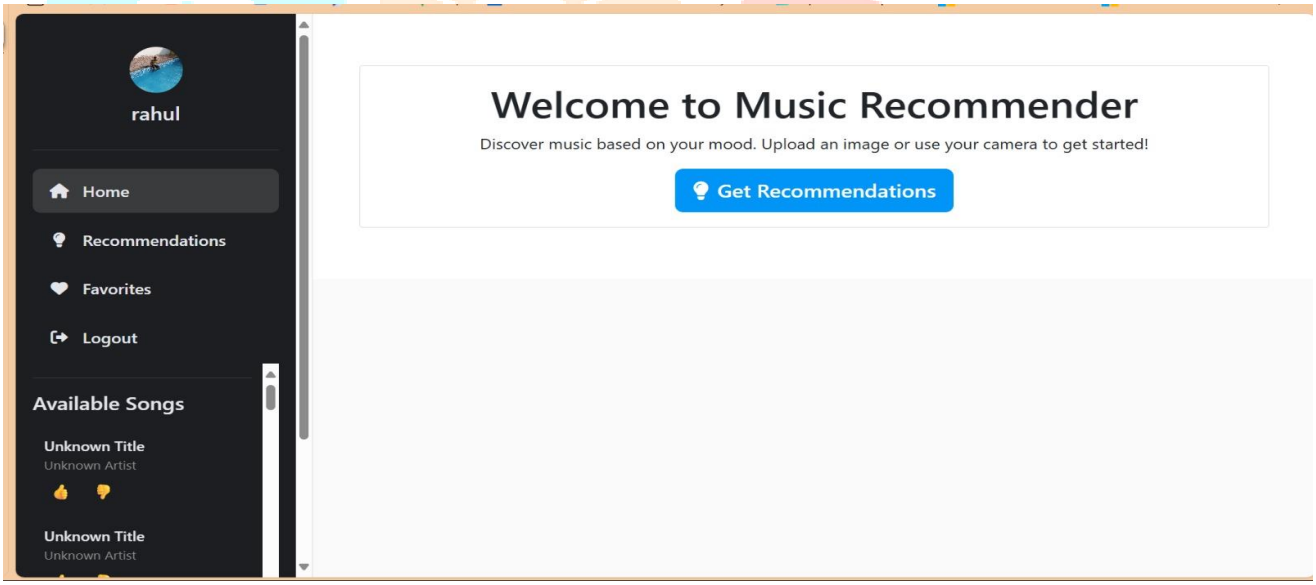

Figure 5: Front-End of the Music Recommendation System

VI. Conclusion

An emotion-based personalized music recommendation system that combines intelligent music suggestion skills with real-time facial emotion identification is successfully developed in this study. Through the use of a pre-trained Convolutional Neural Network (CNN) that was trained on the FER-2013 dataset, the system is able to classify basic human emotions including anger, sadness, happiness, and neutrality with 70% accuracy. After that, the identified emotion is translated to a matching music cluster produced by machine learning methods, such as Random Forest and SVM classifiers and KMeans clustering.

By suggesting songs that correspond with the user's emotional state, the suggested solution improves the user experience and closes the gap between affective computing and personalized entertainment. The system's scalability and resilience are guaranteed by the incorporation of several Python-based modules for data management and machine learning, OpenCV for real-time face detection, and Django for a dynamic web interface. The current method reveals some limits, such as decreased accuracy in less-than-ideal

situations, even though it functions well in frontal facial conditions and under ideal lighting. Extending the dataset, refining the model architecture, and adding audio-based sentiment analysis for multi-modal emotion recognition are possible future improvements.

All things considered, the study shows how machine learning and artificial intelligence may be used practically to improve human-computer interaction by making emotionally intelligent suggestions.

VII. Acknowledgment

I would like to express my sincere gratitude to all those who supported me throughout the development of this research project.

First and foremost, I am deeply thankful to my Project guide and Head Of Department, **Dr. Anand P. Srivastava**, for their valuable guidance, insightful suggestions, and consistent encouragement that helped shape the direction of this work. Their expertise and feedback were instrumental in overcoming technical challenges during the design and implementation of the system.

I also extend my appreciation to the faculty and staff of the Computer Science And Engineering, Nitra Technical Campus, Ghaziabad, for providing the infrastructure and academic environment necessary for the successful completion of this research.

VIII. References

[1] Tkalčič, M., & Chen, L. (2015). Personality and recommender systems. In Recommender Systems Handbook (pp. 715-739). Springer.

[2] Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(1), 39–58.

[3] Goodfellow, I. J., Erhan, D., Luc Carrier, P., Courville, A., Mirza, M., Hamner, B., ... & Bengio, Y. (2013). Challenges in representation learning: A report on three machine learning contests. In International Conference on Neural Information Processing (pp. 117–124). Springer.

[4] Aggarwal, C. C. (2016). Recommender systems: The textbook. Springer.

[5] Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: Introduction and challenges. In Recommender Systems Handbook (pp. 1–34). Springer.

[6] Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Transactions on Affective Computing, 1(1), 18–37.

[7] Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning-based recommender system: A survey and new perspectives. ACM Computing Surveys (CSUR), 52(1), 1–38.

[8] Mollahosseini, A., Chan, D., & Mahoor, M. H. (2021). Analyzing facial expressions in real-time using deep learning. Journal of Visual Communication and Image Representation, 79, 103228.

[9] Delbouys, R., Spadon, G., Lamprier, S., Hennequin, R., & Royo-Letelier, J. (2021). Music mood detection based on audio and lyrics with deep neural net. In Proceedings of SIGIR 2021.

[10] Zhao, Y., Zhang, Y., & Ma, Y. (2022). A context-aware music recommendation system based on mood detection from facial images. Expert Systems with Applications, 198, 116846.

[11] Lee, J., Lee, H., & Hwang, S. (2022). A context-aware and emotion-based music recommendation system using facial expression recognition. Journal of Ambient Intelligence and Humanized Computing.

[12] Maddala, S., et al. (2024). Emotion Classification for Music Recommendation via CNN. IEEE Transactions on Multimedia, 26(2), 112–123.

[13] Sarkar, A., & Dey, N. (2023). Deep Emotion Recognition with CNN and LSTM. Pattern Recognition Letters, 171, 80–90.

[14] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[15] Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2021). AffectNet: Facial Expression Recognition Database and Deep Neural Network Classifier. IEEE Transactions on Affective Computing, 12(2), 530–544.

[16] García, S., Luengo, J., & Herrera, F. (2020). Data Preprocessing in Data Mining. Springer.

[17] Kim, Y. E., Schmidt, E. M., & Migneco, R. (2021). Music emotion recognition: A state-of-the-art review. ACM Transactions on Multimedia Computing, Communications, and Applications, 17(1), 1–20.

[18] van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research, 9(Nov), 2579–2605.

[19] Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L. E., & Brown, D. E. (2020). Text classification algorithms: A survey. Information, 10(4), 150.

[20] Zhao, Y., Zhang, Y., & Ma, Y. (2022). Context-aware music recommendation system using facial emotion recognition. Expert Systems with Applications, 198, 116846.

[21] Kouki, P., Fabbri, A. R., & Koutrika, G. (2021). User preferences modeling for music recommendations: A reinforcement learning perspective. Proceedings of the Web Conference 2021, 3244–3255.

[22] 1 P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2001.

[23] 2 I. Goodfellow, D. Erhan, P. Luc Carrier, A. Courville, and Y. Bengio, "Challenges in Representation Learning: A report on three machine learning contests," Neural Networks, vol. 64, pp. 59–63, 2015. (FER2013 Dataset)

[24] 3 C. Laurier, O. Lartillot, T. Eerola, and P. Herrera, "Music Mood Representation from Social Tags," International Society for Music Information Retrieval Conference (ISMIR), 2009.