Emotionrhythms: A Real-Time Music Player Based on Emotion Recognition

Yashashri Kamalakant Gosavi

PG Student, Dept. of Information Technology, Patkar-Varde College,

Mumbai, Maharashtra, India

Abstract - EmotionRhythms is a groundbreaking music player designed to automatically curate and play songs that align with the user's current emotional state. In today's fast-paced world, music holds a significant place in people's lives, often serving as a reflection of their emotions. However, traditional music players require users to manually sift through extensive playlists to find songs that match their mood, which can be time-consuming and frustrating. Recognizing that individuals frequently associate music with their emotional experiences, a system that intelligently organizes and recommends songs based on the user's emotional state would offer immense value. This innovative solution allows users to effortlessly enjoy music that resonates with their feelings, enhancing their listening experience. The proposed system focuses on four primary emotions—Happy, Sad, Angry, Fear, Disgust, Surprise and Neutral—which are detected using Convolutional Neural Networks (CNN), a machine learning technique specialized in facial expression analysis. Once the user's emotion is identified, the system seamlessly plays recommended songs from predefined directories, ensuring a personalized and emotionally attuned music experience.

***Keywords*— Convolutional Neural Networks (CNN), Emotion Recognition, Facial Expression Analysis, Music Recommendation System, Real-Time Emotion Detection, OpenCV**

I. INTRODUCTION

Music is an art form that involves the harmonious combination of sounds to create a pleasing auditory experience. It plays a vital role in the lives of many individuals, though people often have specific preferences when it comes to the type of music they enjoy. Music is not only a source of entertainment but also a powerful tool for fostering creativity and improving linguistic skills in children. Furthermore, music therapy has been shown to have therapeutic benefits for various groups, including cancer patients, children with ADHD, and individuals battling depression. As a result, many healthcare institutions have integrated music therapy into their treatment protocols to support emotional and mental well-being. The primary goal of EmotionRhythms is to streamline the process of selecting songs by automating recommendations based on the user's emotional state. To utilize the features of EmotionRhythms, users must first log in to the web application. Upon accessing the platform, users are presented with two choices: they can either manually select their current mood or allow the system to automatically detect their emotion through the webcam. Once the system identifies the user's mood, it seamlessly plays a song that aligns with the detected emotional state. The system leverages Convolutional Neural Networks (CNNs) for facial expression recognition, achieving an impressive accuracy rate of 80%, ensuring reliable and personalized music recommendations.

II. LITERATURE REVIEW

1) The "FACIAL EMOTION BASED MUSIC RECOMMENDATION SYSTEM USING COMPUTER VISION AND MACHINE LEARNING TECHNIQUES" is designed to enhance user experience by utilizing facial expressions for real-time music recommendations. This system employs computer vision and machine learning to precisely detect human emotions and tailor music suggestions accordingly. However, it encounters limitations, including difficulties in interpreting complex emotional states and concerns regarding privacy linked to facial recognition technology. To improve its effectiveness and address these issues, further development of the algorithms is required.

2) The research paper “A MACHINE LEARNING BASED MUSIC PLAYER BY DETECTING EMOTIONS” introduces a music player that leverages machine learning, particularly Convolutional Neural Networks (CNNs), for detecting emotions. This method provides improved computational efficiency and accuracy. By integrating CNNs with multilayer perceptron’s, the system reduces processing demands, which in turn boosts computational speed. However, it encounters challenges, including the requirement for large labelled datasets to optimize CNN performance and potential issues with feature visualization via back-propagation. While consolidating multiple

actions into a single step can streamline operations, it may also increase the complexity of the model and introduce new points of failure. Further refinement and validation of these techniques are crucial for optimizing real-time performance and ensuring effectiveness across diverse datasets.

3) EMOUSIC: A MUSIC PLAYER UTILIZING MACHINE LEARNING FOR EMOTION AND ACTIVITY presents Emousic, a music player designed to generate personalized playlists by leveraging machine learning techniques focused on users' emotions and activities. This innovative platform combines multiple data sources and employs cutting-edge data science strategies to determine mood and activity, delivering real-time, tailored music recommendations. However, it encounters several challenges, including privacy concerns, fluctuations in the accuracy of mood and activity recognition, and the interpretability of the underlying machine learning algorithms. Addressing these issues is vital for fostering user confidence and enhancing the system's performance.

4) The research paper "EMOTION RECOGNITION BASED ON FACIAL EXPRESSIONS USING CONVOLUTIONAL NEURAL NETWORK (CNN)" delves into the realm of facial emotion recognition through the application of Convolutional Neural Networks (CNNs), highlighting the rising relevance of human-computer interaction. It scrutinizes the difficulties encountered in Emotion Recognition Datasets and conducts experiments with CNN parameters to identify seven emotions expressed in human faces, leveraging the iCV MEFED dataset. While it presents fresh perspectives, the broader applicability of its results and the potential fluctuations in performance warrant further investigation and testing.

5) The research paper "FACIAL EMOTION RECOGNITION USING DEEP CONVOLUTIONAL NETWORKS" outlines a groundbreaking technique for identifying facial emotions via Deep Convolutional Networks (CNNs), which effectively learn features from facial expressions on their own. By utilizing facial action units (AUs) detected by CNNs, this method achieves a remarkable accuracy of 97.01% on the Cohn-Kanade database, exceeding the performance of standard CNN methods. However, challenges may arise regarding its use in diverse datasets and the necessary computational resources. Further testing and assessment of real-world applications are required.

6) The research paper "MUSIC RECOMMENDATION BASED ON FACE EMOTION RECOGNITION" outlines a unique strategy for recommending music through the recognition of facial emotions. It incorporates a Convolutional Neural Network for emotion identification and employs Pygame and Tkinter to facilitate music recommendations. This system automatically curate’s music playlists that reflect the user's emotional state, thereby promoting automation and efficiency in computation. However, there are concerns regarding the precision of emotion detection and the appropriateness of music suggestions, along with possible challenges in accommodating a wide range of user preferences and contexts. To enhance accuracy and user satisfaction, further validation and refinement are necessary.

III. SYSTEM DESIGN

Existing System

Existing systems for emotion-based music players often fall short in terms of advanced features when compared to the newly developed music player described in the provided content. These traditional systems typically offer basic functionalities, such as manual playlist selection and random song shuffling, but lack the capability to automatically detect and adapt music based on the user's emotional state. They often rely on simplistic algorithms or user input to generate playlists, without utilizing advanced machine learning techniques for emotion recognition. Furthermore, the user interfaces of these systems tend to be less interactive and customizable, resulting in a limited user experience. While they provide fundamental music playback features, they do not incorporate the sophisticated emotion-based auto-music playing functionality found in the newly developed music player. In summary, existing systems for emotion-based music players are relatively rudimentary and lack the comprehensive, user-centric features of the advanced system described. Below are the key drawbacks of current systems:

a) **Limited Features**: Existing systems often lack advanced capabilities, such as automatic mood detection and music adjustment. Instead, they rely on manual playlist selection and random song shuffling, which limits their effectiveness in providing a tailored listening experience.

b) **Lack of Personalization**: These systems frequently fail to deliver personalized music recommendations based on the user's emotional state. Without leveraging sophisticated machine learning techniques for emotion recognition, they cannot offer customized music experiences that resonate with individual users.

c) **Poor User Experience**: The user interfaces of existing systems are often less interactive and customizable, leading to a subpar user experience. They may lack intuitive design elements and features that enhance user engagement and satisfaction.
d) **Inaccurate Mood Detection**: Due to the absence of advanced algorithms and data-driven approaches, existing systems may struggle to accurately detect users' moods based on historical data. This can result in inconsistent music recommendations and a less immersive listening experience.

Proposed System

The proposed system, EmotionRhythms, integrates facial expression recognition with a music recommendation system. The system detects the user's emotions in real-time using OpenCV and CNN, and recommends songs based on the detected mood. The system is implemented as a web application, ensuring accessibility and ease of use. The key components of the system include:

- **Facial Expression Recognition**: Detects the user's facial expressions using OpenCV and CNN.
- **Music Recommendation**: Recommends songs based on the detected mood.
- **User Interface**: Provides a user-friendly interface for seamless interaction.
- **Database Management**: Uses PostgreSQL to store user data, playlists, and song metadata.

The system architecture is designed to be modular, allowing for easy integration of new features and improvements. The front-end is built using HTML, CSS, and JavaScript, while the back-end is powered by Django, a Python-based web framework. The system also incorporates RESTful APIs for seamless communication between the emotion detection module and the music recommendation engine.

IV. PROBLEM STATEMENT

Most existing music applications do not consider the user's facial expressions as an input parameter for recommending songs, which inspired me to develop a unique application that addresses this gap. The application I am creating utilizes artificial intelligence (AI) to detect users' facial expressions in real-time and recommend songs based on their detected emotions. While focusing on emotions makes the application stand out, I recognize that emotions are not the only factor in creating a successful software product. There are several edge cases that need to be addressed, such as users who may not be expressive or situations where poor lighting conditions hinder facial detection. To handle such scenarios, the application allows users to manually select their mood, similar to choosing a genre or artist in traditional music apps. Currently, the application supports seven primary emotions: happy, sad, angry, fear, surprise, disgust and neutral, which are the most commonly expressed feelings in real-life situations. I plan to expand this range in the future. The application prioritizes accuracy in emotion detection while also ensuring a seamless and visually appealing user interface to deliver an exceptional user experience.

V. PROBLEM SOLUTION

This research introduces the development of a music application that distinguishes itself by playing music based on the user's facial expressions, offering a novel alternative to conventional music applications. Unlike traditional methods that rely on manual song selection, wearable devices, or sound-based categorization, this system proposes an automated approach. The application is designed as a web-based platform using Django, a Python web framework, and employs PostgreSQL for database management to store essential user account details such as email, name, and password. Upon logging in, users are presented with two options: they can either manually select their mood or allow the webcam to automatically detect it. Once the mood is identified, the application seamlessly plays a song that matches the user's emotional state. For face detection, the system utilizes the Haar cascade algorithm, chosen for its simplicity, efficiency, and real-time processing capabilities. Since its introduction in 2001, Haar cascades have become one of the most widely used object detection algorithms in the OpenCV library, making it an ideal choice for this application. This innovative approach aims to enhance user experience by providing a more personalized and emotionally responsive music platform.

VI. METHODOLOGY

Machine learning (ML), a subset of artificial intelligence (AI), focuses on creating algorithms and models that enable computers to learn from data and make predictions or decisions without explicit programming. The core objective of machine learning is to allow systems to autonomously learn and improve from experience, identifying patterns and insights within data to support informed decision-making or predictions.

For this project, key libraries and frameworks such as OpenCV for facial recognition, scikit-learn, and TensorFlow for implementing machine learning algorithms will be utilized. These tools will be imported into the Python environment to facilitate the development of the system, ensuring efficient and accurate emotion detection and music recommendation capabilities.

6.1 Data Collection

The system uses the FER dataset for training the emotion detection model. The dataset contains facial expressions labelled with emotions such as happy, sad, angry, fear, surprise, disgust, and neutral. The dataset is pre-processed to ensure consistency and quality, including resizing images to 48x48 pixels and converting them to grayscale.

6.2 Splitting of Data

The dataset is split into training and testing sets (80:20 ratio) to ensure the model's accuracy and generalization. The training set is used to train the CNN model, while the testing set is used to evaluate its performance.

6.3 Modeling of Data

A Convolutional Neural Network (CNN) model is trained on the CK+ dataset to classify facial expressions into seven emotions. The model architecture consists of multiple convolutional layers, followed by max-pooling layers and fully connected layers. The model achieves an accuracy of 80% on the test set, demonstrating its effectiveness in emotion classification.

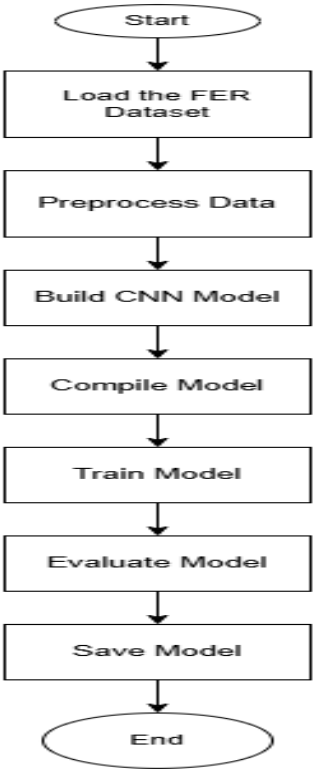

Fig.1. FER Dataset CNN Training Process

6.4 Facial Expression Recognition using OpenCV

The Facial Expression Recognition module in the EmotionRhythms system utilizes the Haar Cascade algorithm for real-time face detection, capturing the user's facial expressions through a webcam and processing the images using OpenCV. The detected face is pre-processed by converting it to grayscale and resizing it to a fixed dimension (e.g., 48x48 pixels) before being passed to the Convolutional Neural Network (CNN) model for emotion classification. The CNN model, trained on the FER dataset, classifies the facial expression into one of seven emotions: Happy, Sad, Angry, Fear, Surprise, Disgust, or Neutral. This real-time process ensures accurate emotion detection, which is then used to recommend songs that match the user's mood. While the system performs well under normal conditions, challenges such as poor lighting, facial obstructions, and subtle emotions may affect accuracy. Future improvements could include multi-modal emotion detection, advanced pre-processing techniques, and training on larger datasets to enhance robustness and performance.

Fig. 2. Facial expression detection.

6.5 Emotion Detection using CNN

The detected facial expressions are classified into emotions using a Convolutional Neural Network (CNN) model trained on labeled facial expression datasets such as FER-2013. The CNN processes input images through convolutional, pooling, and fully connected layers to predict emotions like happy, sad, angry, surprised, fearful, disgusted, or neutral. Based on the detected emotion, the system recommends songs from a custom-built database that I curated, where each song is categorized by mood, genre, tempo, and lyrics theme. For example, a "happy" emotion might trigger upbeat, energetic song recommendations, while a "sad" emotion could lead to slower, melancholic tracks. This personalized approach ensures that the recommendations are relevant and tailored to the user's emotional state, providing a unique and engaging music experience.

VII. IMPLEMENTATION & RESULTS

Following are the implementation and result screenshots of my project.

Home Page:

The homepage is the index page of the music recommendation system. Registration and login options are available on the homepage. By clicking on this option, users can register and login to their accounts, as shown in Figure 3.

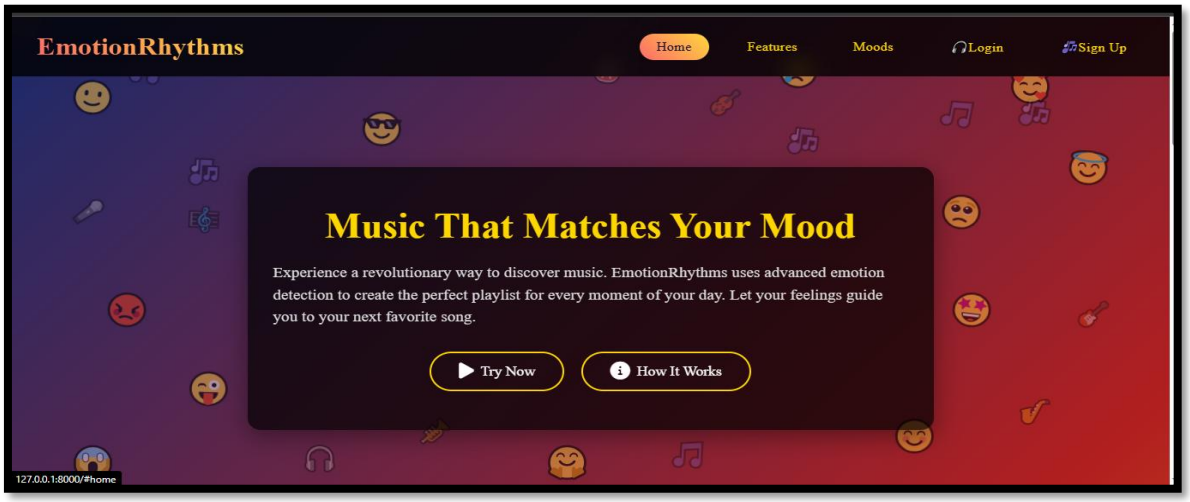

Fig. 3. Home Page

Registration Page:
This registration page is for users who wish to register themselves as users. They can successfully register after filling out all of the required information, as shown in Figure 4.

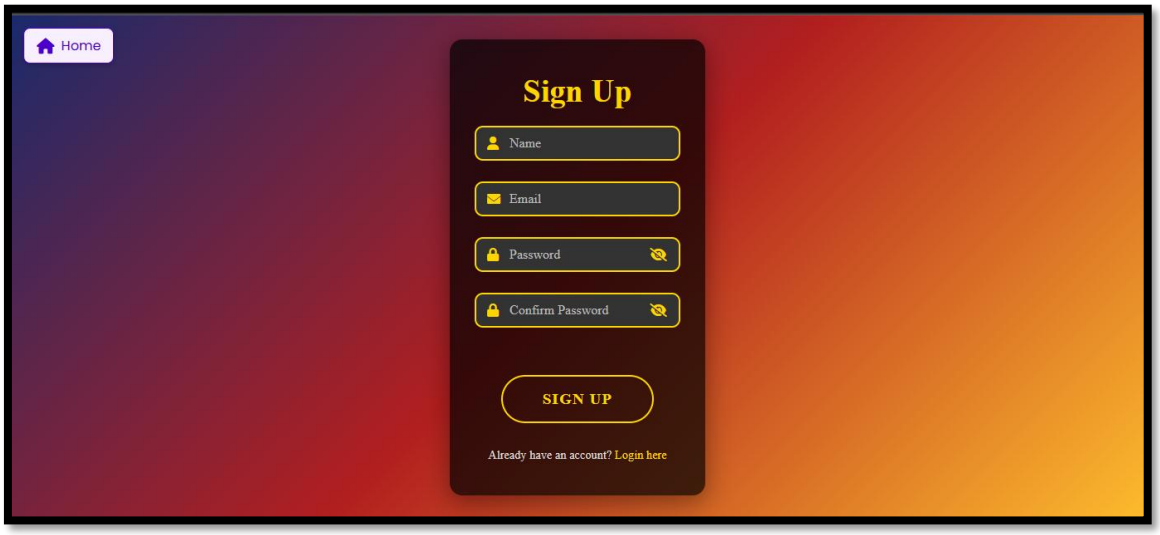

Fig. 4. Registration Page

Login Page: As shown in Figure 5, a user can log in to their profile account after filling in the right information, such as their email and password. The information is sent to the database to check for a match.

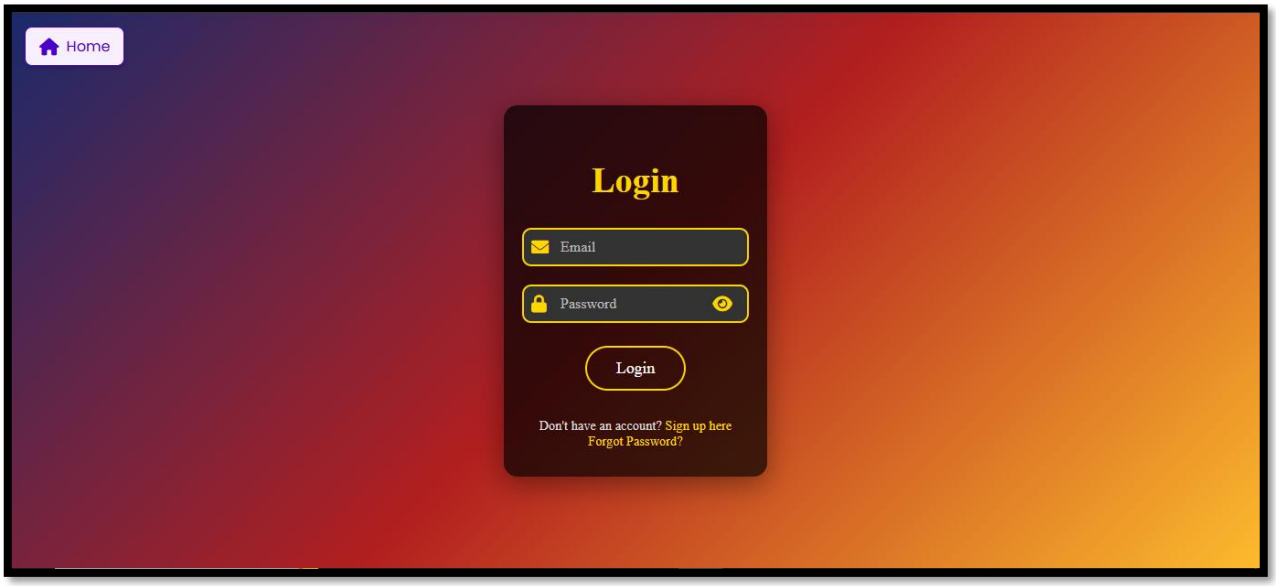

Fig. 5. Login Page

Music Genre Page: As shown in Figure 6, a user can select your preferred music genre such as Hindi, Marathi, English, or Punjabi.

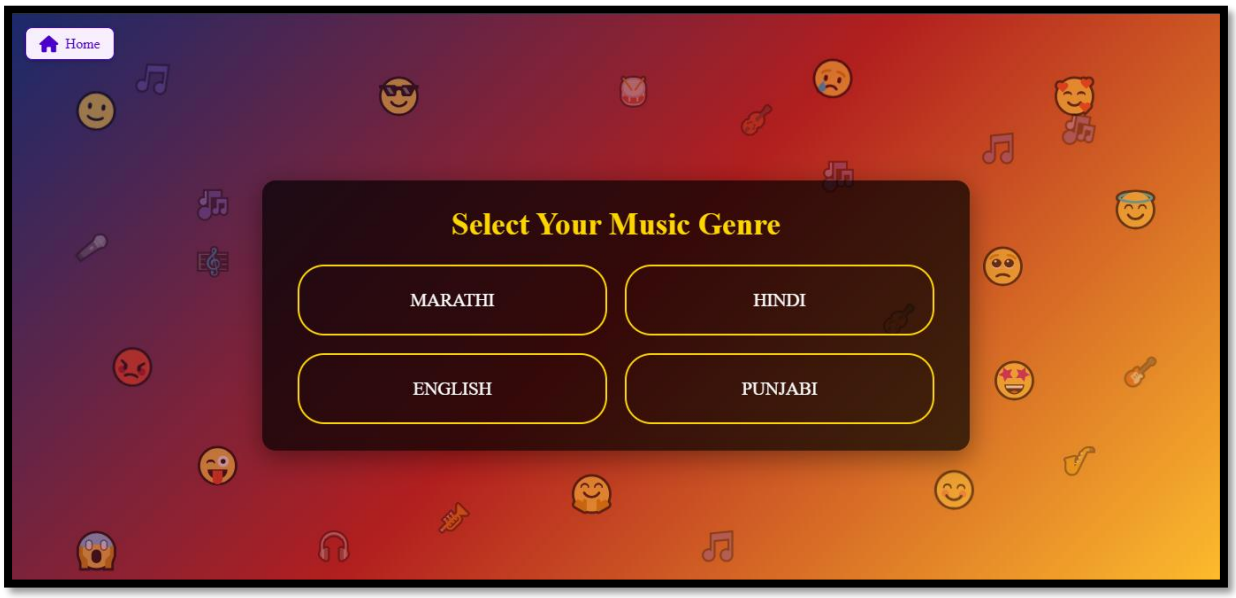

Fig. 6. Music Genre Page

Select the Mood Page: As shown in Figure 7, a user can select a mood manually or click on the Start Detection Button for Real-Time Detection.

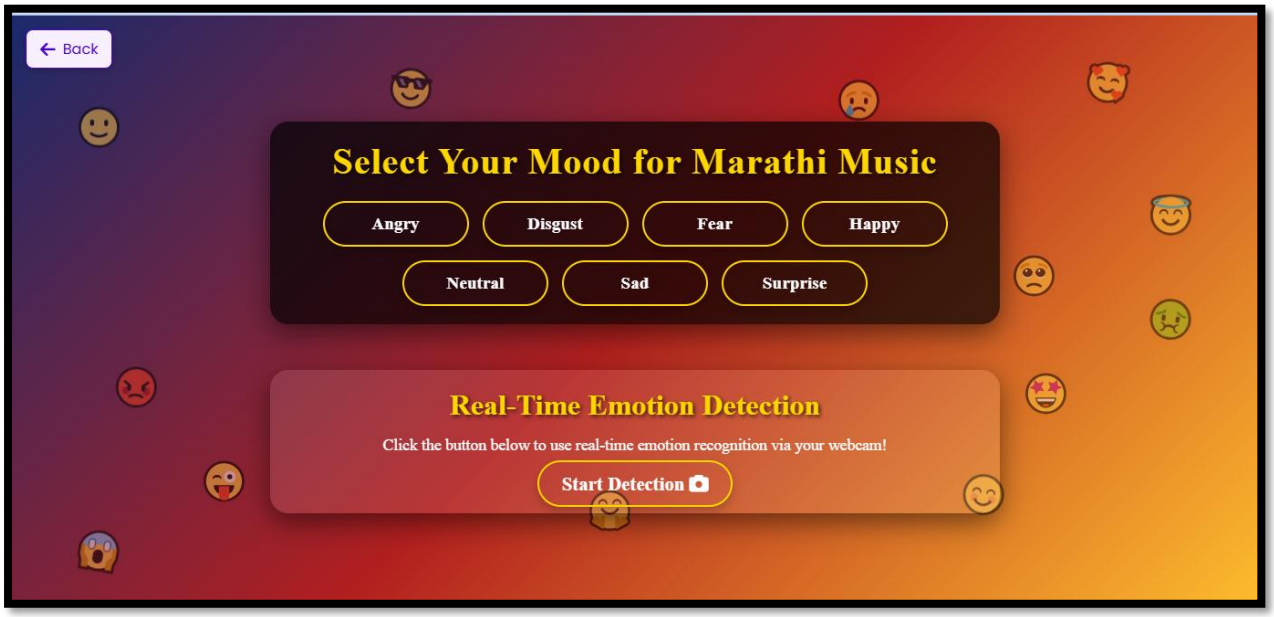

Fig. 7. Select the Mood Page

Real-Time Detection Page: As shown in Figure 8, a user can click on the start detection the detection is start and stop the detection then stop it and automatically play the song.

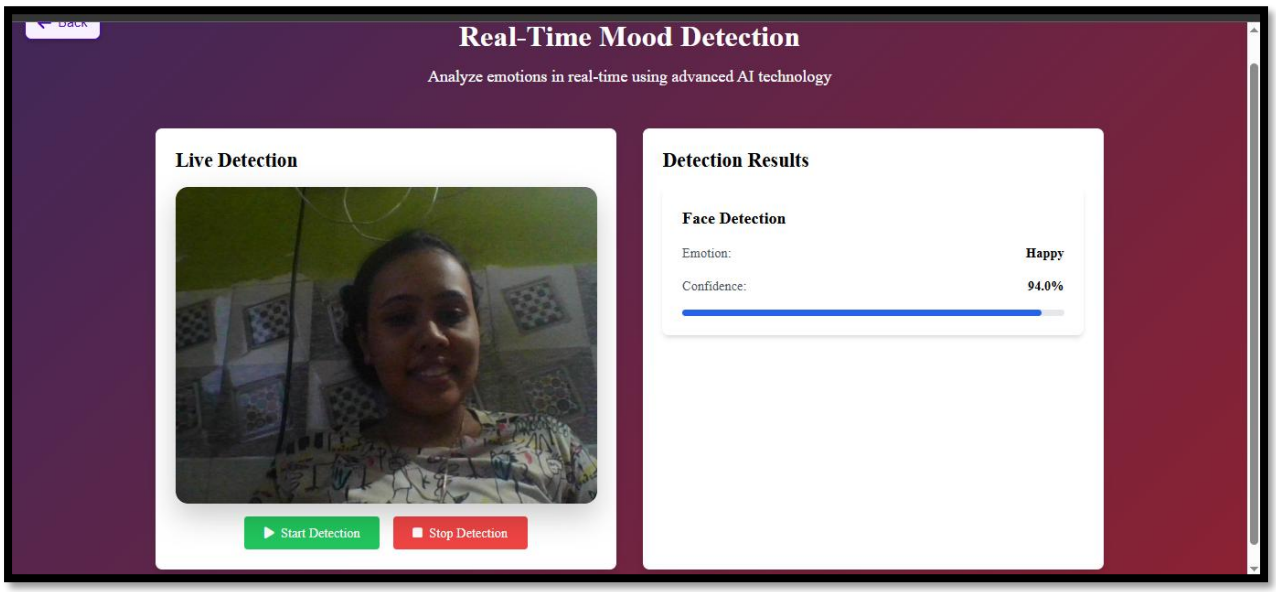

Fig. 8. Real-Time Detection Page

Play song Page: As shown in Figure 9, Song play automatically according to music genre selection and mood detection.

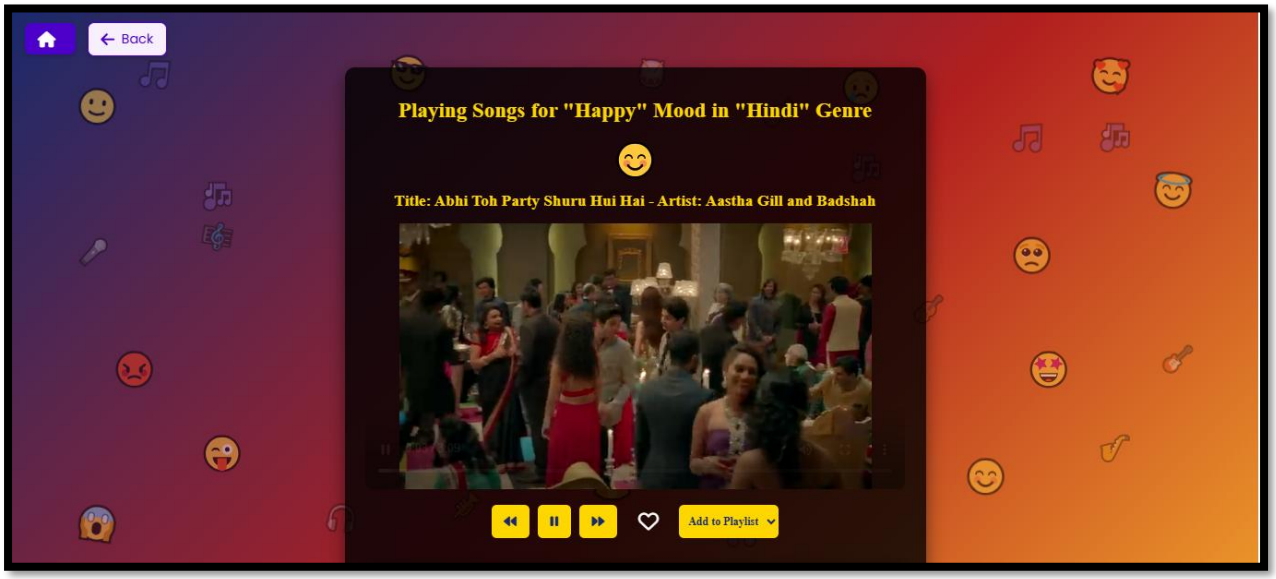

Fig. 9. Real-Time Detection Page

VIII. ANALYSIS

Our analysis of the EmotionRhythms, which utilizes Convolutional Neural Networks (CNN) and OpenCV, demonstrated an average accuracy rate of 80% in identifying basic emotions. The system exhibited consistent performance across various datasets, unaffected by demographic factors such as age, gender, or ethnicity. It also maintained reliable emotion recognition accuracy despite variations in lighting conditions and background noise. These results underscore the potential of CNN-based emotion detection systems enhanced by

OpenCV to significantly improve user experiences. However, further research is necessary to refine model parameters and tackle challenges posed by complex real-world environments, ensuring even greater accuracy and robustness in emotion recognition.

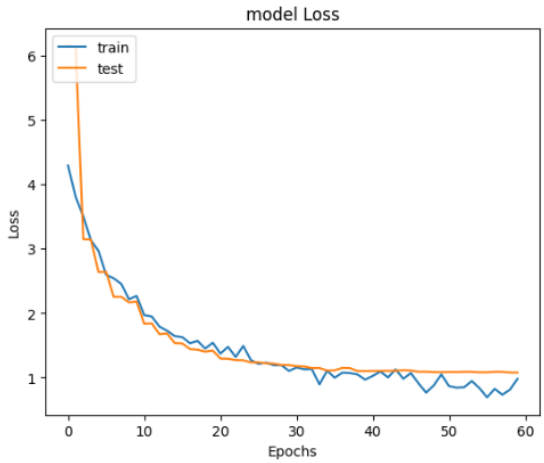

Fig. 10. Training vs Testing Accuracy.

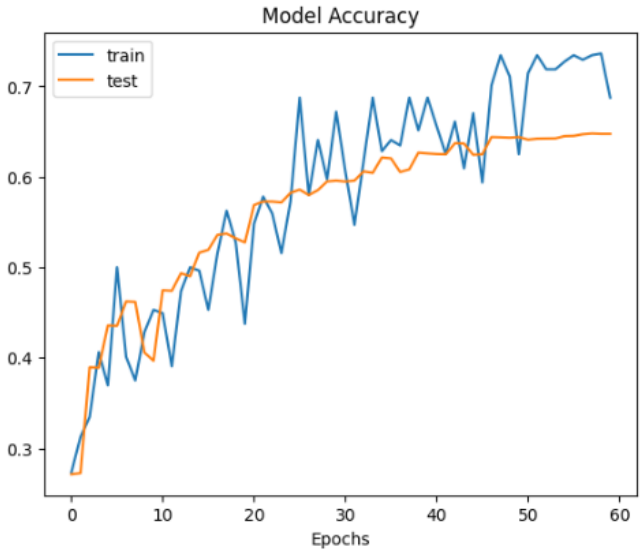

Fig. 11. Training vs Testing Accuracy.

IX. CONCLUSION

EmotionRhythms is a revolutionary music player that bridges the gap between technology and human emotions. By leveraging facial expression recognition and machine learning, the system provides personalized music recommendations based on the user's emotional state. The system enhances emotional well-being and offers a unique and immersive music experience. Future work will focus on expanding the song database, integrating with popular music streaming platforms, and addressing privacy concerns related to facial recognition. The system has the potential to revolutionize the way we experience music, offering a more personalized and emotionally intelligent listening experience.

REFRENCES

[1] Dr. Shaik Asif Hussain and Ahlam Salim Abdallah Al Balushi, "A real-time face emotion classification and recognition using deep learning model", 2020 Journal of Physics: Conference Series.

[2] P. Lucey et al., "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[3] Raut, Nitisha, "Facial Emotion Recognition Using Machine Learning," 2018 Master's Projects.

[4] Patra, Braja & Das, Dipankar & Bandyopadhyay, Sivaji, "Automatic Music Mood Classification of Hindi Songs," 2013.

[5] Londhe RR and Pawar DV, "Analysis of facial expression and recognition based on a statistical approach," International Journal of Soft Computing and Engineering, 2012.

[6] Zeng Z, Pantic M, Roisman GI, and Huang TS, "A survey of affect recognition methods," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008.

[7] Gokul Krishnan K, Parthasarathy M, Sasidhar D, and Venitha E, "Emotion detection and music recommendation system using machine learning," International Journal of Pure and Applied Mathematics, 2018.

[8] Shlok Gilda, Husain Zafar, Chintan Soni, and Kshitija Waghurdekar, "Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation," presented at the IEEE WiSPNET 2017 conference.