JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Sung-Geet: Personalized Music Player Using Real-Time Facial Emotions

[1] Prof Dr.Prashant Nitnaware, [2]Simran Mallar, [3]Saniya Farash, [4]Shreya Gaikwad, [5]Harshinee Nadiminty

[1]Department of Computer Engineering,
Pillai College Of Engineering, New Panvel, India

ABSTRACT

Music recommendation systems have ample choice of songs and genres for a user to listen to, making it difficult for many users to make a selection. Keeping this in mind we propose a personalized emotion based music player that utilizes a person's facial expressions through a webcam to detect their emotion in real time. This approach not only personalizes the music selection process but also offers a more intuitive and interactive way for users to discover new songs and genres. When the webcam is activated, the system uses the MTCNN model for real-time emotion detection. When a face is detected, the system deploys the MobileNet model, a convolutional neural network pre-trained on the ImageNet dataset, to detect emotions, classify the emotions into different categories, and recommend music that aligns with the emotional state.

To train the emotion detection model, the FER 2013 dataset from Kaggle is used. These images capture a wide range of emotional states, providing a rich and diverse dataset for training the emotion detection model. The diversity of the images ensures that the model is robust and capable of handling various facial expressions, lighting conditions, and facial orientations, enhancing its overall performance and reliability. Lastly, when the user's emotion is determined, the system recommends a list of songs that best match their current mood. This personalized music player provides a mix of Hindi and English songs based on their emotional state. By considering the user's current emotions, the system significantly improves the overall music recommendation experience.

I. INTRODUCTION

Music recommendation systems have evolved into user tools, providing customized suggestions for songs and artists. Nevertheless, these systems don’t take into account the emotions of the user. Music recommendation systems will be improved if they incorporate emotions into them. In this study, we aim to develop and implement a real-time emotion-based music recommendation system. Its objective is to recommend music that would resonate with the user’s mood making it a personalized and enjoyable musical experience. For our system, after the emotions are detected, the user is provided with two choices of languages of music; Hindi and English so as to enhance customization for all their experiences.

A. Fundamentals:

In the digital age, the convergence of technology and human experience continues to drive innovation across various domains. One such domain is personalized music recommendation systems, where recent advancements have led to the emergence of emotion-based recommendation systems. At its core, an emotion-based music recommendation system seeks to redefine the music listening experience by offering personalized recommendations tailored to the user's current emotional state. This innovative system harnesses computer vision techniques, particularly facial recognition, to analyze and interpret users' emotional cues from their facial expressions. Once these emotions are discerned, the system leverages sophisticated algorithms to recommend music tracks that resonate harmoniously with the user's emotional disposition.

Creating an emotion-based music recommendation system involves combining key components that are crucial for how well the system works:

- **Data aggregation:** The music library consists of tracks labeled with emotions such as happy, sad, surprise, etc. For emotional data, the FER 2013 dataset is used, which includes thousands of images of facial expressions, each annotated with the corresponding emotion.
- **Emotion detection:** Advanced computer vision technology is used to capture and analyze facial expressions in real-time, while a machine learning model trained on the dataset is used to recognize different emotions from these facial expressions.
- **Music player:** The curated playlist corresponding to the detected emotions is displayed to the user in their chosen language for listening to the music.

B. Objectives:

The overarching objective of this endeavor encompasses a multifaceted approach aimed at redefining the music-listening experience:

- **Personalized recommendation:** The primary goal is to deliver personalized music playlist tailored to the user's current emotional state, thereby fostering a deeper emotional connection between the listener and the music they consume.
- **Utilizing Computer Vision:** By harnessing advanced computer vision algorithms, the system aims to analyze users' facial expressions and accurately infer their emotional states in real-time, enabling seamless and intuitive interactions.
- **Enhanced Listening Experience:** Ultimately, the objective is to enhance the overall music listening experience by providing recommendations that resonate harmoniously with the user's emotional state, potentially leading to mood enhancement and emotional well-being.

Furthermore, the versatility of the system extends beyond recreational contexts, with potential applications in healthcare settings to assist patients in managing their emotions and fostering emotional resilience.

II. LITERATURE SURVEY:

A. The Existing System:

In the emerging field of emotion-based music recommendation systems, dominant models mostly use Convolutional Neural Networks (CNN) for facial emotion detection. The algorithm is aimed at enhancing user involvement by providing a song playlist based on detected emotions. While most of these systems use static images in the detection of emotions, there is a noticeable lack of models that are efficient in processing real-time images, a characteristic that could greatly increase user participation.

Typically, the architectural design of these systems involves four main elements: face recognition; image normalization; emotion detection, and music suggestion. A vital step in this process is to gather and preprocess datasets as well as to create and train CNN. This model training can then be used to predict emotion and generate song suggestions.

Disadvantages of the existing system include:

- These models require numerous emotion-annotated datasets (which are of differing quality and diversity) for training purposes, a factor that may introduce bias into the system as it affects the ultimate performance of the model. Having biased datasets leads to incorrect predictions on emotions and this affects system reliability and inclusion.
- Time-consuming initialization and training of CNN models especially in large datasets may hinder real-time implementation of the system. Long training times can delay system deployment, thereby influencing overall user experience and responsiveness.
- Some models don't perform accurately on subtle or nuanced facial expressions thus reducing precision in recognizing emotions. Inaccurate recognition can lead to the wrong classification of emotions, thus affecting the accuracy of music recommendations made by EMAA (Emotion-based Music Auto Recommenders by EMAA) leading to customer dissatisfaction.

III. PROPOSED SYSTEM

The suggested music player applies a combination of deep learning, computer vision, and machine learning techniques to detect the facial expressions of users in order to recommend music that suits the facial emotion. A MobileNet model is employed by this system for facial emotion detection specifically using the fer2013 dataset which consists of grayscale images of faces with seven emotion categories.

Several approaches are used to improve the accuracy and avoid overfitting of the MobileNet model for facial emotion recognition such as Transfer Learning, Batch normalization, and Data augmentation. The web application has been developed with the Flask framework as the main interface. The emotion detection model captures the user's image in real-time and transmits it to the model. Once the user's emotion is detected, the webpage displays suggested songs that match those emotions. It also lets the user choose the language of songs, such as Hindi or English. This proposed system is capable of providing users with a more personalized and engaging music experience while benefiting both creators and providers.

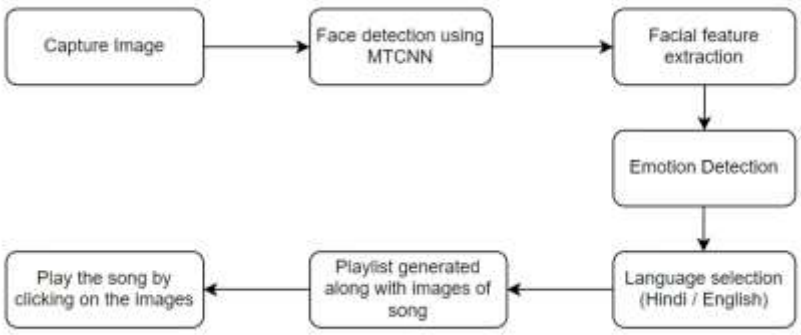


Fig.1:Proposed Fundamentals

1. Proposed system architecture

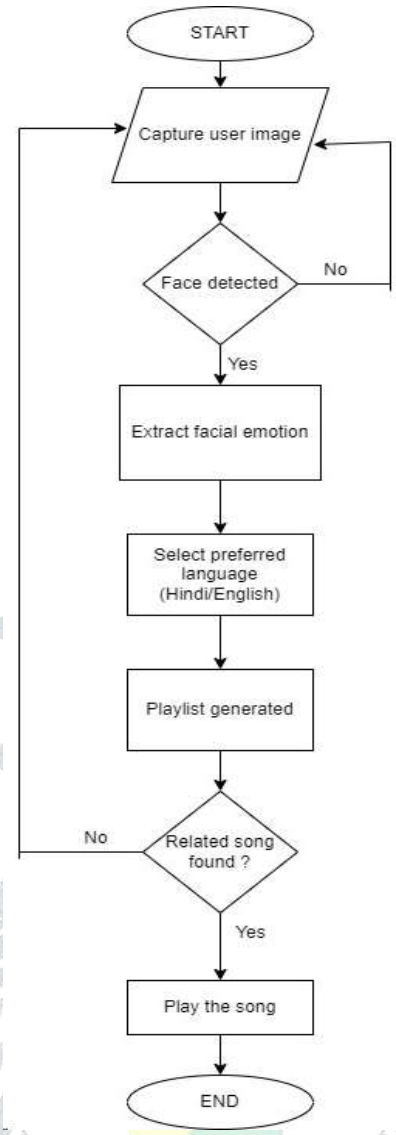


Fig. 2:Proposed Architecture

Advantages of the proposed system include:

- Music that empathizes: The system tailors its recommendations based on what a user feels at the moment thus making music selection more personal.
- Attitude regulator: Assuming cases where one switches off music due to a negative mood, this can be avoided if the machine suggests songs that are in tune with the particular emotion. For example: lyrics/songs for sadness or anger might include cheerful melodies or slow soothing tunes respectively.
- A larger emotional range: Facial recognition helps this system capture a wider spectrum of emotions than self-report measures found in traditional approaches. Thus, it is possible for more fine-grained and accurate recommendations about songs based on moods.

I. METHODOLOGY:

A. Data collection and preprocessing:

- **Data collection -** For training the emotion detection model, we used the FER 2013 dataset from Kaggle consisting of black and white pictures that show different facial expressions with each one being assigned to one of the seven emotions. These images are relatively small with sizes like 48 by 48 pixels.
- **Data augmentation -** Techniques like zooming, shearing, horizontal flipping, and rescaling were implemented.These techniques are employed to increase the effective size of the dataset by generating new variations of the existing images. Resizing pixel values to the range of [0, 1] during this process

helps standardize input data, improving training stability, and thereby enhancing the overall training process.

B. Emotion detection model:

- The dataset used for training the emotion detection model was formed by using the FER 2013 dataset. The dataset consists of approximately 35000 grayscale images, out of which 28821 images were used for training and another set of validation constituted 7066 images.
- Transfer learning was applied in constructing an emotion detection model using MobileNet as a pre-trained base model. This involved utilizing features learned from large datasets like ImageNet for face feature extraction. The fully connected layers of the model are then regularized using dropout to randomly disable neurons during training leading to a reduction in overfitting since it makes the network learn more solid features.
- Batch normalization is done on dense layer activations in order to stabilize and speed up the training process.
- Early stopping is also employed to stop training when validation loss stops decreasing hence, preventing the model from overfitting into the training data set.
- Data augmentation techniques applied helped the model to learn robust features and improve its generalization capabilities, reducing the risk of overfitting to the training data .
- The use of these methods collectively improves the accuracy of the MobileNet Model for Facial Emotion Recognition while at the same time avoiding overfitting risks.

C. Face detection model:

- When the web camera is on in real-time, it detects faces and tells us what emotions are conveyed by them.
- A real-time processing loop keeps capturing frames, detecting faces, classifying emotions in those faces, and showing them with the emotion labels overlaid on top of them.
- MTCNN (Multi-task Cascaded Convolutional Network), a lightweight image classification architecture, is used to accurately detect each frame's face and produce a bounding box of where it resides.
- A user can then exit the application gracefully after completing emotion detection using keyboard shortcuts (Ctrl + Q) which stops the webcam feed.
- The final emotion is predicted out of these seven emotions: happy, angry, neutral, sad, surprise, fear, and disgust; by taking the mode of the maximum frames of emotion detected.
- The final predicted emotion is then displayed on the screen, from which the user has the option to select either Hindi or English songs, according to their preference.

II. DATA SETS:

A. FER2013

The FER2013 data set is a widely used resource in the field of affective computing. It consists of a set of grayscale images of human faces labeled with seven emotional categories: angry, disgust, fear, happy, sad, surprise and neutral. Each image has been pre-processed so that the face in the image is aligned centrally with respect to size.

In this case, the images were transformed into grayscale so that all pixels have 'equal' intensity regardless of their colors or directions thus allowing easier computation on these pixels

The training set consists of 28,709 examples and the public test set consists of 3,589 examples.

B. FER Plus

The Facial Expressions in the Wild (FER-Plus) dataset serves as a significant extension to the popular FER2013 dataset, providing a more comprehensive and challenging benchmark for facial expression recognition research. While FER-Plus inherits the core structure of FER2013, including grayscale images labeled with seven basic emotions, it significantly expands the dataset size (over 230,000 images compared to FER2013's 30,000) and incorporates images captured "in the wild" with variations in pose, lighting, occlusion, and image quality. This increased complexity and size address the limitations of FER2013, allowing researchers to develop more robust and generalizable models that can perform better on unseen real-world data. Furthermore, FER-Plus tackles the class imbalance issue present in FER2013 by employing data augmentation techniques, making it a valuable benchmark for comparing the performance of different facial expression recognition algorithms.

III. ALGORITHMS:

A. Face detection algorithms

1.Haar cascade

Haar cascade is an object detection algorithm meant to detect different objects in an image including faces. However, this often leads to the detection of not only faces but also edges and corners of objects in the background making it have about 48-50% accuracy on average. Its sensitivity to light further complicates its performance making it quite ineffective at capturing facial expressions, especially in environments with challenging lighting conditions. On the other hand, Haar cascade is mainly designed for object detection and hence may not give detailed landmark information required for deep emotion analysis.

2.MTCNN

MTCNN stands for Multi-task Cascaded Convolutional Networks, which is a deep learning model used widely for face detection, especially for images with varying scales and orientations. It utilizes a cascade structure that comprises three neural networks whose work is to detect faces. These networks progressively narrowed down the search for faces. Unlike the Haar Cascade, MTCNN was entirely focused on face detection alone. By effectively handling faces having different scales and orientations, this greatly improved overall model performance. MTCNN has significantly improved the accuracy of face detection while reducing false positives.

B. Model training algorithms

1.VGG16

In the task of facial emotion recognition, VGG16 had many challenges. Even though it is deep and complex, VGG16 struggled to achieve high accuracy; as it gave an accuracy of about 40%. It encountered overfitting as a key problem. Overfitting happened when the model managed to capture unnecessary patterns from the training data leading to its inability to generalize. Probably, this overfitting can be attributed to the complexity of the architecture of VGG16.

2.RESNET

The Resnet model was used for face emotion recognition and achieved an accuracy rate of around 42%. Although it boasted a design and included skip connections to address gradient vanishing issues in networks, ResNet faced significant challenges in accurately predicting emotions from facial expressions. The skip connections, crucial for helping ResNet effectively train networks, also raised the computational requirements leading to a greater need for memory and processing resources when compared to simpler models.

3.MobileNet

In MobileNet for emotion detection, transfer learning involves leveraging a pre-trained MobileNet model, originally trained on a large dataset like ImageNet, and adapting it to the specific task of recognizing emotions in images. The process typically involves removing the original classification layers of MobileNet and adding new layers for emotion classification. The pre-trained layers can be optionally frozen to retain their learned features, especially when the training data is limited. The model is then trained on a dataset of images labeled with emotions, where the weights of the new layers are adjusted to improve emotion prediction. Dropout and batch normalization techniques were also included so as to address overfitting and enhance the generalization ability of the network.

VII. IMPLEMENTATION

- We start by importing libraries such as cv2 for image processing, Keras, and TensorFlow for deep learning models as well as playing sound for the audio playback. The system has been deployed using Flask.
- The home function renders the main HTML template (index1.html) in the case of HTTP GET request.
- When the ‘check emotion’ button is clicked on, it activates the camera via the GET function, and this pop-up/ panel smoothly deals with Real-time emotion detection.
- The emotion detection panel will last on the screen for about 27 seconds or until it is manually closed.
- By then, only one of seven possible emotions: happy, sad, angry, fear, disgust and neutral should be identified as the user's final emotions.
- We have created a camera function that performs real-time emotion detection using a pre-trained Convolutional Neural Network (CNN) model with an MTCNN face detection model which takes frames from the webcam based on input data feeds to predict emotions remarked dynamically on the system.
- In doing so, we find out what final emotion is detected, and the user has to select between Hindi or English song recommendation buttons included accordingly.
- This script runs when executed directly as a Flask application. The ‘app.run(debug=True)’ statement launches the application in debug mode, aiding in development and troubleshooting.

VIII. UML Diagrams:

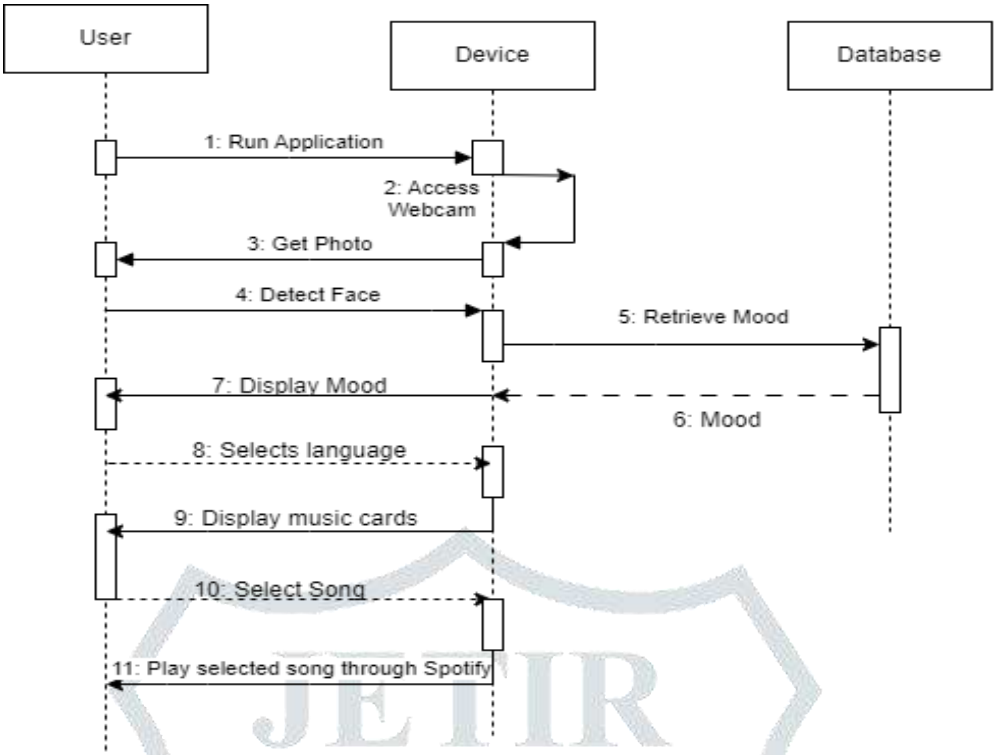


Fig. 3 : Activity Diagram

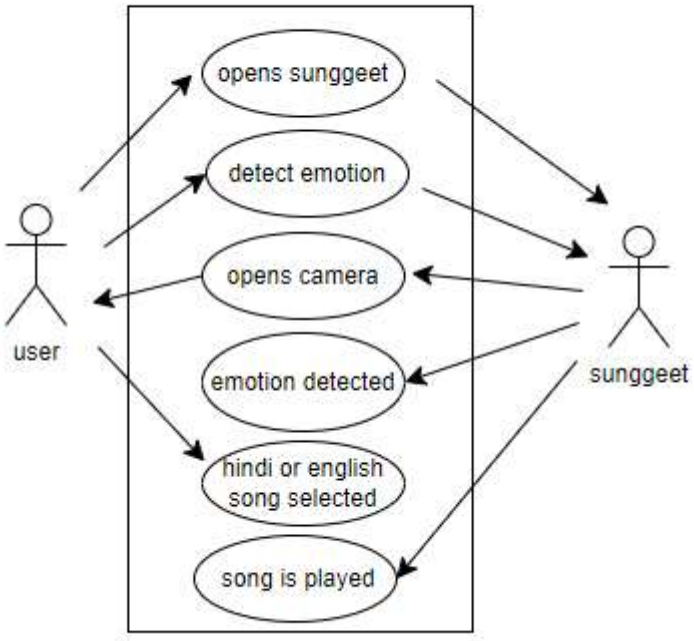
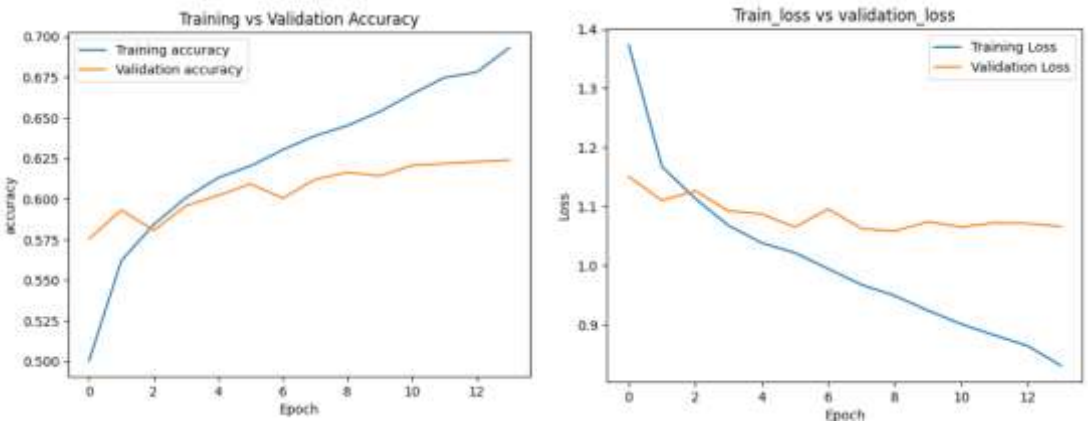


Fig. 4: Use Case Diagram

IX. RESULTS AND DISCUSSIONS:



(a) **(b)**

Fig.5(a). Accuracy graph , Fig5(b). Loss graph

In Fig.5(a), the graphs display the accuracy of our model, where the x-axis specifies the number of epochs and the y-axis specifies the accuracy.

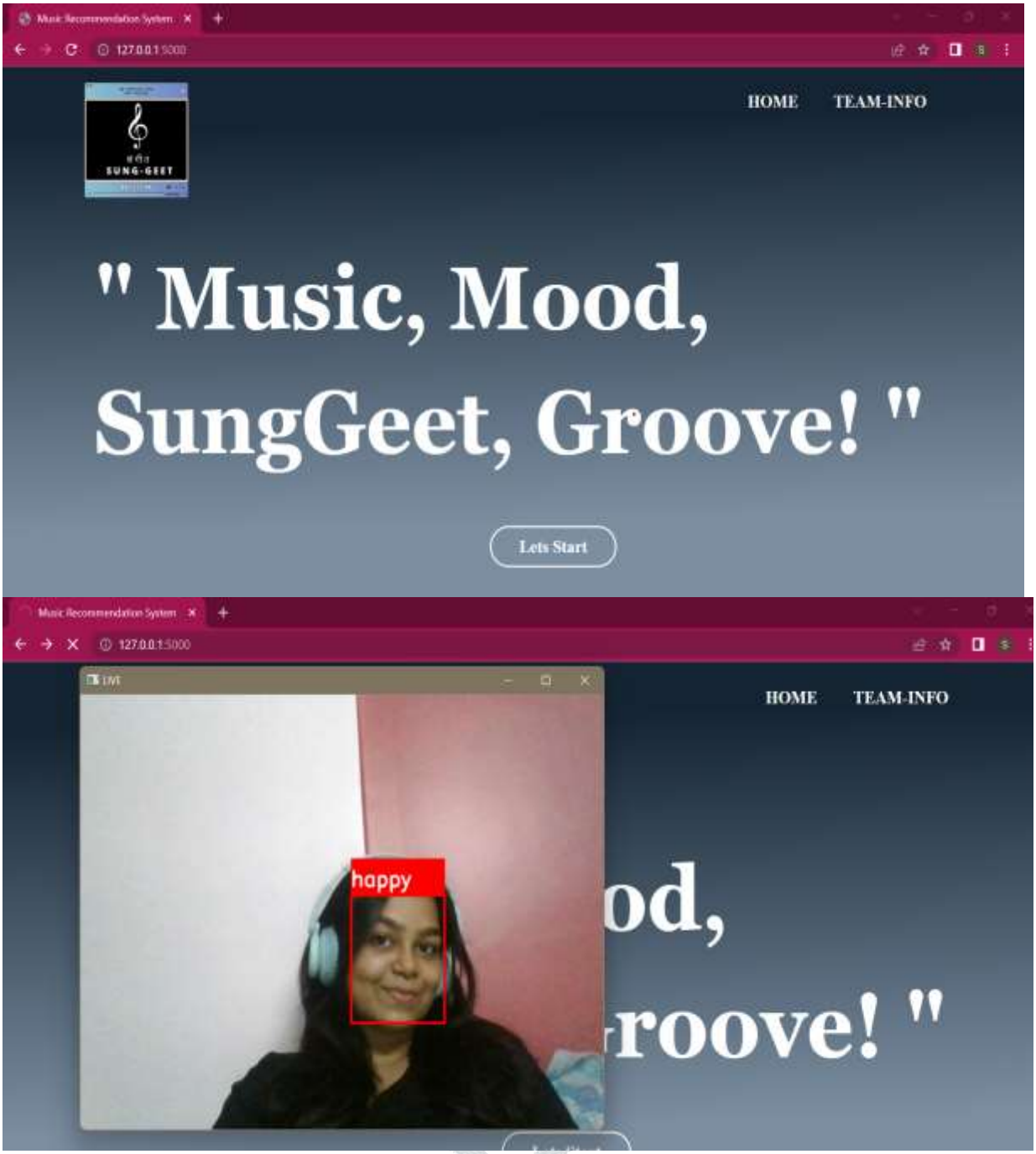
In Fig.5(b), the graph displays the training and validation loss of our model, where the x-axis specifies the number of epochs and the y-axis specifies the loss

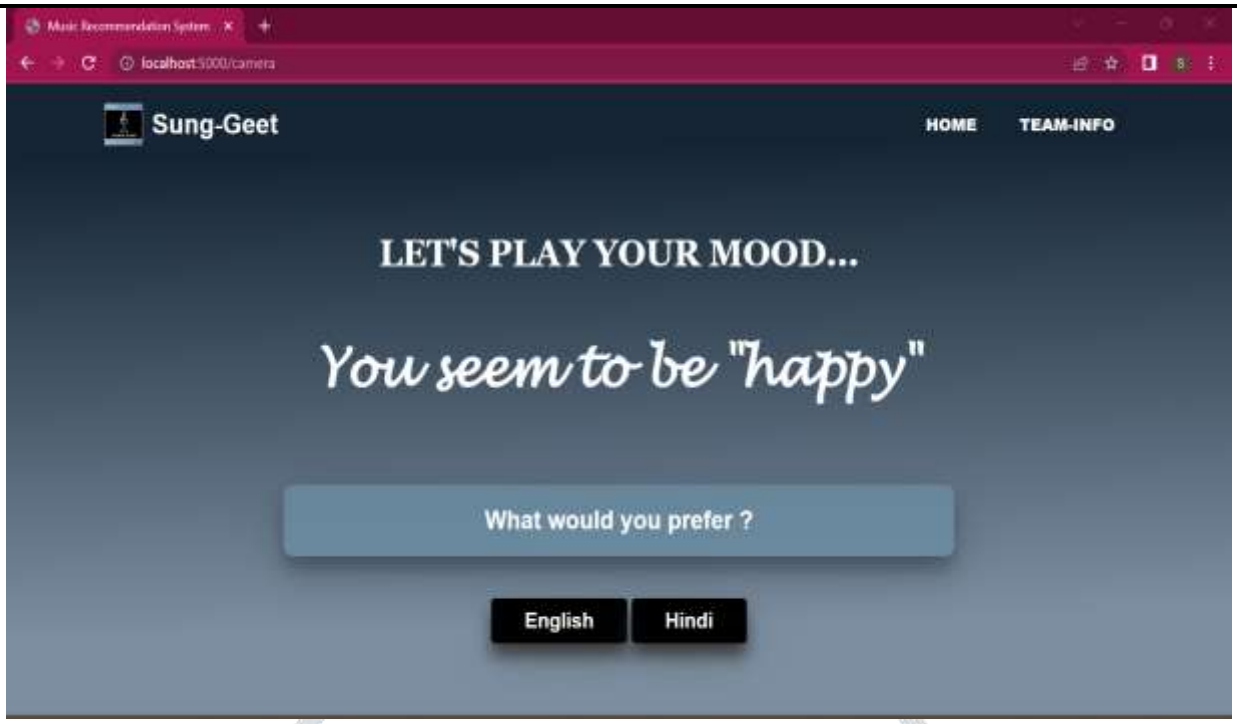
The use of transfer learning and other methods collectively improved the accuracy of the MobileNet Model for Facial Emotion Recognition, while at the same time avoiding overfitting risk. Our model has an accuracy of 70.13%.

Table 1 : Comparison of Accuracy

Sr No.	**Algorithms**	**Accuracy**
1.	VGG16	40%
2.	MobileNet	70.13%
3.	Resnet	42%

Sample of input and output screenshots:





X. CONCLUSION

It has long been known that music helps in mood modulation and acts as a stress relieving agent for many individuals. Recent technological advancements have provided opportunities for creating emotion-based music recommendation systems. A system may use facial recognition technology to detect the listener's facial expressions and infer their emotional state. It can then recommend music that matches or complements that emotional state, whether the listener is feeling upbeat and energetic, or calm and reflective. This personalized approach to music recommendation can enhance the listener's overall music listening experience, making it more enjoyable and meaningful.

REFERENCES

[1] Vijay Prakash Sharma, Azeem Saleem Gaded, Deevesh Chaudhary , Sunil kumar Shikha Sharma, "Emotion-Based Music Recommendation System" ,2021 9th International Conference on Reliability, Infocom Technologies and Optimization. Available at: https://ieeexplore.ieee.org/document/9596276

[2] Pranesh Ulleri, Shilpa Hari Prakash, Kiran B Zenith, Gouri S Nair, Jinesh Kannimoola "Music Recommendation System Based on Emotion", 2021 12th international conference of Computing Communication and Networking Technologies. Available at: https://ieeexplore.ieee.org/document/9579689

[3] Ankita Mahadik ,Vijaya Bharathi Jagan, Shambhavi Milgir, Prof.Vaishali Kavathekar, Janvi Patel ,"Mood based music recommendation system",June 2021. Available at: https://www.researchgate.net/publication/352780489_Mood_based_music_recommendation_system

[4] L. Shreya and N. Nagarathna, "Emotion Based Music Recommendation System for Specially-Abled," 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)., 2021 IEEE,Available at: https://ieeexplore.ieee.org/document/9574033

[5] Rahul ravi, S.V Yadhukrishna, Rajalakshmi, Prithviraj, "A Face Expression Recognition Using CNN & LBP",2020 IEEE. Available at: https://ieeexplore.ieee.org/document/9076422

[6] S Metilda Florence S and Uma M, 2020, "Emotional Detection and Music Recommendation System based on User Facial Expression", IOP Conf. Ser.: Mater. Sci. Eng. 912,06/2007. Available at: https://www.academia.edu/es/55562318/IRJET_Music_Recommendation_System_using_Emotion_Recognition

[7] Manas Sambare, FER2013 Dataset, Kaggle, July 19, 2020. Accessed on: September 9, 2020. [Online], Available at: https://www.kaggle.com/msambare/fer2013

[8] Mahmoudi MA, MMA Facial Expression Dataset, Kaggle, June 6, 2020. Accessed on: September 15, 2020. [Online]. Available at: https://www.kaggle.com/mahmoudima/mma-facial-expression

[9] S. Kiruthika and R. Meenakumari, "Music Recommendation System Based on User's Mood Using Machine Learning," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS). Available at: https://ieeexplore.ieee.org/document/9113086

[10] K. Sangeetha, R. Subashini, and R. Aruna, "Emotion-Based Music Recommendation System Using Machine Learning Algorithms," 2019 International Conference on Communication and Signal Processing (ICCSP). Available at: https://ieeexplore.ieee.org/document/8697882