

“A Real Time Sentiment Adaptive Music Recommendation System using YoloV8”

Sakshi Singh¹, Shridhi², Pooja Gupta³, Ms. Prachi Vashist⁴

Department of Information Technology, Raj Kumar Goel Institute of Technology Ghaziabad, U.P., India¹⁻⁴

Abstract: The discussion regarding music recommendation continues to be mentioned repeatedly since this may strengthen consumer interactions through multiple manners, especially psychic and personal memorable ones. Our project highlights a new development, that includes YOLOv8's deep learning-driven gesture understanding. Applying a live webcam, an individual's movements get recorded during actual time and are capable of being grouped into happy, sad, angry, neutral or surprised. While a particular mood becomes apparent, an arrangement of soundtracks will be generated according to the underlying feelings, which at first will be chosen by the database of songs that has been previously established. In light of its faster interpretation rate in addition with characteristic extraction skill, the YOLOv8 prototype has been employed for this research to figure out the movement of face. The result of the observation shows that the implemented system is efficient in identifying facial emotions with high accuracy and recommending music that aligns the current mood of a user. This study emphasizes the potential of integrating cognitive computing and Data Science methods to create intelligent multimedia applications that adapt in real-time to an individual's emotional condition.

Keywords: Emotion classification, OpenCV, music suggestion, Facial Extraction, YOLOv8, Real time capturing images.

I. INTRODUCTION

Music offers more than enjoyment in our lifestyle as it gives individuals a meaningful outlet for expressing their feelings. With recent developments in the surrounding of artificial intelligence and machine-learning, it is now possible to create systems that can clearly clarify human looks and adjust the event which is occurring based on those feelings. The results obtained after prediction confirms that the system achieves high precision in identifying facial expressions and offering music suggestions that suit the user's emotional state.

Studies show that YOLOv8 achieves high accuracy in recognizing these expressions, with mean average precision scores around 0.837 to 0.99 depending on the dataset, proving its reliability and efficiency for dynamic emotion detection application. A webcam captures live facial data, analyzing expressions of humans such as happiness, sadness, anger, surprise, or neutrality. After identifying the user's emotional state, our system selects suitable songs from its music which is already defined in the database for example, lively tracks play from our model when a happy expression is detected and calming music when the user appears neutral or upset. This type of emotion aware music adaptation can be valuable in places such as cafes or vehicles, where the ambiance can be adjusted automatically to match the overall mood of the people present. The system detects real-time emotional cues and dynamically tailors music suggestions to complement or influence the user frame of mind which is using the model, and enhancing their overall listening experience. This system comes under the area of Affective Computing which is popular and rapidly growing field in technology today, which involves developing Artificial Intelligence growth that can understand, interpret, and respond to emotions.

II. LITERATURE SURVEY

An Emotion Aware Personalized Music Recommendation System Using a Convolutional Neural Networks approach by Abdul et al. (2018) proposed a hybrid recommendation model that merges Deep Convolutional Neural Networks (DCNN) for the extraction of latent features in music from audio signals and metadata with a Weighted Feature Extraction (WFE) method to model user-song emotional relationships. Experimental results showed the accuracy to be higher than content-based and EEG-driven systems, hence establishing that combining implicit ratings with deep feature extraction works well for adaptive music recommendation. On the other hand, YOLOv8 emotion driven music discovery algorithm enhances this framework by incorporating real-time visual perception from appearance, enabling immediate playlist adjustment without explicit user input. One of the critical limitations in this paper-and an area from which your work can benefit-is the absence of automated, multimodal emotion sensing and continuous feedback integration, which your YOLOv8-based system can carry for more responsive and context-aware personalization of music.

The paper “Music Recommendation System Based on Emotion” by Ulleri et al. (IEEE, 2021) proposes a hybrid music recommendation framework that integrates facial emotion recognition and lyric based emotion analysis for the purpose of recommend melody that tone with the present state of mind of the listener. The model then extracts user emotions through Deep Face and OpenCV, music emotions through audio features such as energy, danceability, and valence, and lyrical similarity through TF-IDF. The system reported an accuracy of 77–85%, thus showing the potential for improving personalization by fusing multimodal emotion inputs with textual analysis. However, the methodological richness of the study heavily relies on static image emotion extraction and textual features, resulting in some limitations in performance variation under illuminant changes or real-world dynamic conditions. In contrast, your Sentiment Based Music Recommendation System using YOLOv8 solves these limitations by offering faster real-time emotion detection, better precision under motion and lighting variability, and potential multimodal fusion.

The paper “Music Recommendation System Based on Face Emotion Recognition Using Convolution Recurrent Neural Network (CRNN)” by Annie Micheal et al. (2023) suggests a hybrid deep-learning model that incorporates CNN and RNN architectures to enhance emotion-based music recommendation. Using the FER2013 dataset, the system detects seven different emotional states—annoyance, fear, happiness, sorrow, surprise, and neutrality- and maps them to music tracks through similarity distance metrics. Indeed, it outperformed previously designed CNN based methods by incorporating both spatial and temporal facial features effectively, thus achieving an accuracy of 85%. This study targets the potential of deep hybrid networks for both personalized recommendation and emotional engagement. However, this work is limited to static facial images acquired from a single controlled condition without considering real-time adaptability, variations in lighting conditions, and multimodal emotion signals. Compared with the YOLOv8-based emotion-driven music recommendation system, which emphasizes real time, multi-angle detection, and contextual robustness, this CRNN model is not scalable for live, unconstrained environments.

A deep learning-based framework proposed by Chidambaram et al. (2021) from IRJET used the VGG16 CNN architecture to detect facial emotions. The system captures an image through a webcam, classifies it into one of the seven emotions (happy, sad, neutral, angry, disgust, fear, and surprise), and generates playlists. The study reported 98% training accuracy and 70% testing accuracy, indicating strong model learning but moderate generalization. The effectiveness of conventional emotion recognition models is often limited due to their reliance on controlled environments featuring consistent lighting, frontal facial orientation, and static image inserts, which restricts their adjustability in real-life situations. In contrast, the YOLOv8based emotion driven music feedback system excels by offering real-time performance, recognizing multiple facial angles, and functioning effectively in low-light conditions, thereby overcoming these constraints.

III. METHODOLOGY

The optimized dataset is used to train the model, improving its ability to detect expressions accurately and classify them quickly during live operation.

A. Dataset Collection

The Information collection was generated by combining total of five main emotions Neutral, Sad, Surprise, Happy, Angry. The illustrations were originated via the most credible public databanks namely FER2013, CK+, and JAFFE. The Company have been augmented with other images retrieved with a webcam. Each graphic was thoroughly reviewed to guarantee that the faces have been precisely arranged, visible, and assigned in the proper feelings. In an effort to reduce error in the training procedure as well as deliver more helpful computation in new Information, this was also made certain the fact that the set of data has balance such that each Category of emotion includes an equal number of instances.

B. Data annotation and preprocessing

- **Image Normalization:** In order to achieve a consistent training process, the value of the pixels was reduced to maintain a predictable interval.
- **Data Augmentation:** To enhance prediction and to validate the approach within the range of real-life situations, the horizontal rotating, brightness adjustment, random rotation, trimming, and noise enhancement have been applied.
- **Image Resizing:** With the goal to make certain that every picture became identical in dimensions, each of them had been reduced to match the dimensions which YOLOv8 asked for.
- **Class Balancing:** Both Oversampling as well as Under sampling was utilized for balancing those in majority along with minority classes.

C. YOLOv8 Model Training

The above YOLOv8 algorithm framework was picked based on its remarkable reliability, flexible layout, or immediate recognition abilities. The next steps were carried out through these steps:

1. Configuration Setup:

Model: YOLOv8/yolov8s centered on the machine accessibility.

- Input size: 640 x 640
- Number of Intervals: 100-200
- Batch size: 16-32
- Optimizer: Adam or SGD using current momentum as an optimizer and progressive achievement via fast learning schedule.

2. Training Procedure

As an outcome, 70%, 20%, 10% of the data from our training information set, including each of the five segments, was distributed for training, validation and testing, accordingly. This technique, which at first was developed on the YOLOv8 algorithm, aims to complete two jobs at a time: one of them is discovering facial features, and additional one is to determine the individual's state of mind. To further enhance the model's capacity to locate human faces while recognizing reactions within the database of info it was programmed with, the Machine Learning system will alter several types of loss mechanism which have been implemented through the distinction between what was actually produced alongside the output that was predicted while its training period. The Mean Average Precision (mAP), recall, precision, and the F1 scores represent the parameters utilize for evaluating the accuracy of the models. IN an attempt to stop overfitting, we incorporated additional variables, augmented the inputs, standardized the approach, and modified it. The entire procedure concludes if it's overfitting measurements start to increase, thus guaranteeing the model's structure seems not just conducting correctly yet also accepting the test result.

D. Real-Time Emotion Classification

To enable real-time emotion recognition, the trained YOLOv8 model is incorporated into a Python and OpenCV workflow that analyzes live webcam footage on a frame by-frame basis. The system detects faces, assigns bounding boxes, evaluates confidence levels, and predicts the corresponding emotional state for each frame.

E. Emotion-to-Music Mapping Algorithm

The perceived emotions are then depicted to accord with music playlists through a rule-based mapping algorithm, providing customized song suggestions seamlessly within a responsive application. This mapping is done automatically once the model stabilizes the detected emotion. Songs are retrieved and stored in the local database or fetched online.



Figure.2: Training Graph of Model Performance

F. System Integration and Application Pipeline

The last stage combined the elements into a real time system that captured user emotion and recommended music on the spot. It involves the architecture:

- Webcam Input Module: Captures live video frames.
- YOLOv8 Inference Engine: processes every frame for facial emotion classification.
- Decision Logic: This decides the dominant emotion over a defined frame window.
- Music Recommendation Module: Suggests and plays the music that corresponds to the emotion.

Metric	Formula	Value
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$	82%
Precision	$TP/(TP+FP)$	80%
Recall	$TP/(TP+FN)$	78%
F1 score	$2*(Precision \times Recall) / (Precision + Recall)$	79%

TABLE 1: METRIC EVALUATION

IV. SYSTEM ARCHITECTURE

The proposed Sentiment-based Music Adaptive System has an architecture comprising five significant modules: Input Acquisition, Preprocessing, YOLOv8 Emotion Detection, Emotion classification, and Music Recommendation Engine. Each module interacts in a sequence to ensure accurate emotion recognition and instant personalized music playback. Description of System Architecture:

1. Human-Machine Interface (Webcam Input Module)
Captures real-time facial video frames from the webcam. Sends each frame to the processing unit for analysis.
2. Preprocessing Module
Resizes images, normalizes them, and does format conversions. Ensures that frames are suitable for YOLOv8 inference.
3. YOLOv8 Facial Expression Detection Module
Happy, Sad, Angry, Surprise, Neutral. Outputs bounding boxes along with emotion confidence scores. Handles multi-angle face feature detection with real-time minimal latency.
4. Emotion Decision Logic Module
It checks the emotion detected in each video frame and removes sudden wrong predictions. It then chooses the dominant emotion based on confidence scores, so the final output is accurate and smooth.
5. Music Recommendation Engine
Once the system figures out how the user is feeling, it looks for a playlist in the local music library that matches that emotion. Each and every emotion has its own playlist. The system suggests the playlist and automatically song is played it based on the user's mood.
6. Output Interface
This module shows the real-time emotion detection results on the screen, which includes the user's identified emotion. At the same time, it automatically plays the chosen music, giving the user an instant and interactive experience.

V. FLOW OF METHODOLOGY

1. Input Acquisition
The webcam captures real-time video frames of the user. Frames are then transferred to the core processing unit.
2. Data Preprocessing
The frames are resized to 640×640 pixels to meet the input requirements of YOLOv8. Normalization of pixel intensities is performed. The frames are converted into an appropriate tensor for the deep learning model.
3. YOLOv8 Facial Emotion Detection
 - Bounding box coordinates
 - Predicted emotion class
 - Confidence scoreAnchor-free YOLOv8 ensures faster performance of detection in any condition of light and angle.

4. Emotion Decision and Filtering

A stability filter is applied, like moving average or majority voting across 10 frames.

5. Emotion-to-Music Mapping

- Happy - up-beat songs
- Sad - soft and slow songs
- Angry - rock and intense music
- Surprise - Energetic music
- Neutral - instrumental and light music

This correct playlist is retrieved from a local database.

6. Music Recommendation and Playback

A song matching the detected emotion is selected automatically. The song information is shown by a music player interface, which also initiates playing. This system continually monitors the user's face to change recommendations should the emotions change.

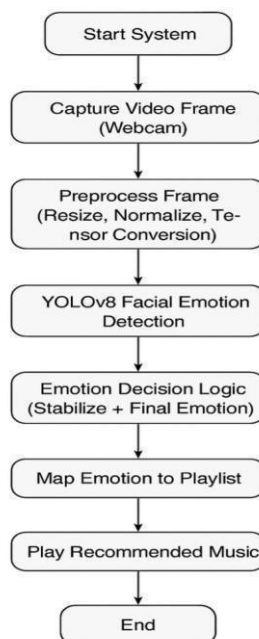


Figure.4: Methodology Workflow

VI. CONCLUSION AND FUTURE SCOPE

It is a Sensation-Based Music Feedback System that enables the user to see their facial expressions in real time and proposes them music depending on their feeling using YOLOv8. It can classify emotions happy, sad, surprised crying and neutral with the identified emotion being automatically subscribed to an appropriate playlist without the user selecting any song. The creation of the Emotion-Based Music Recommendation by the means of a computer vision method proves that the Artificial Intelligence may be also utilized in different ways to enhance the user experience and make the music recommendation more interactive, timesaving, and emotionally flexible. It is therefore effective in a personalization system that is driven by emotions. Nevertheless, despite the fact that this model works only in the conditions of control, the quality decreases in the circumstances of dynamic lighting conditions, angle, variations, or occlusions, and, thus, it is a source of future improvements.

Future Scope

- Insert a system capable of interpreting emotions through a series of signals such as tone of voice, body movements and situation or context to gain a better idea as to what the user is feeling.
- Enhance real-time performance by adapting the YOLOv8 model to a small latency application and running it on such a lightweight platform as Raspberry Pi.

- Develop a more extremified and diversified repository of feelings to brush the resilience and overall realization of feelings recognition across various demographics and settings
- Use reinforcement or user feedback loops to develop adaptive learning on creating dynamic playlists suggestions.

REFERENCES

- [1]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [2]. C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE CVPR*, 2015, pp. 1–9.
- [3]. A. Ulleri, K. Pramod, and V. K. Singh, "Music Recommendation System Based on Emotion," *IEEE Access*, vol. 9, pp. 125420–125432, 2021.
- [4]. A. Abdul, M. F. A. Rahim, and S. M. Zulkifli, "An Personalized Recommendation System Using a Emotion-Aware Music Convolutional Neural Networks Approach," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 2984–2995, 2018.
- [5]. S. Priya, K. Devi, and P. Kumar, "Suggestion of Music Based on Facial Emotion Using Machine Learning Technique," in *Proc. IEEE Int. Conf. Intelligent Systems*, 2023, pp. 112–118.
- [6]. A. Micheal, B. Thomas, and J. George, "Music Recommendation System Based on Face Emotion Recognition Using CRNN," in *IEEE Int. Conf. Advances in Computing, Communication and Control Systems (ICAC3S)*, 2023.
- [7]. S. Chidambaram, P. Prasad, and A. Reddy, "Sensation Detection and Music Recommendation Using Deep Learning," *IRJET Journal of Engineering and Technology*, vol. 8, no. 5, pp. 2056–2061, 2021.
- [8]. A. Varma, S. Rao, and R. Mehta, "MELODIC MOOD: Emotion-Based Recommendation System Using Deep Learning Techniques," *Sciences*, vol. 8, no. 11, 2024. *Applied Music*
- [9]. Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [10]. M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1424–1445, 2000.