ISSN No:-2456-2165

Harmonic Fusion: AI-Driven Music Personalization via Emotion-Enhanced Facial Expression Recognition Using Python, OpenCV, TensorFlow, and Flask

Moeez Rajjan [1] , Prajwal Deore [2] , Yashraj Mohite [3] , Yash Desai [4]
[1,2,3,4] Student , Department of Computer Engineering , Ramrao Adik Institute of Technology

Abstract:- The exciting rise of big data in recent years has drawn a lot of attention to the interesting realm of deep learning. Convolutional Neural Networks (CNNs), a key component of deep learning, have demonstrated their worth, particularly in the field of facial recognition [3]. This research presents a novel and creative technique that combines CNN-based microexpression detection technology with an autonomous music recommendation system [3] [1]. Our innovative algorithm excels at detecting minor facial microexpressions and then goes above and beyond by selecting music that perfectly matches the emotional states represented by these expressions.

Our micro-expression recognition model performs admirably on the FER2013 dataset, with a recognition rate of 62.1% [3]. We use a content-based music recommendation algorithm to extract some song feature vectors after we've deciphered the specific facial emotion. Then we turn to the tried-and-true cosine similarity algorithm to do its thing and recommend some music [3]. But it does not end there. This study isn't only about improving music recommendation systems; it's also about investigating how these systems may assist us manage our emotions [2] [1]. The findings of this study offer a great deal of promise, pointing to interesting prospects for incorporating emotion-aware music recommendation algorithms into numerous facets of our life."

Keywords:- *Deep Learning, Facial Micro-Expression Recognition, Convolutional Neural Network (CNN), FER2013 Dataset, Music Recommendation Algorithm, Emotion Recognition, Emotion Recognition In Conversation (ERC), Recommender Systems, Music Information Retrieval, Artificial Neural Networks, Multi-Layer Neural Network.*

I. INTRODUCTION

Deep learning has seen considerable use in today's information technology era, ranging from picture identification to image processing, with a special emphasis on face expression recognition [3]. Facial recognition, a rising field of study within the sphere of human-computer interaction, has made remarkable progress. However, when converting image processing advances into real-world contexts, its practical applicability frequently encounters limits. Image research usually focuses on improving recognition accuracy while ignoring secondary processes that have the ability to unlock the full value of image data [3]. This uncharted terrain in picture information highlights the necessity for a comprehensive approach that goes beyond mere accuracy enhancement and delves into real-world applicability.

This research offers a comprehensive strategy that combines deep learning techniques with a music recommendation system to bridge the gap between image analysis and practical effects. This strategy is based on employing convolutional neural networks (CNNs) for facial micro-expression identification to not only distinguish emotions but also to improve music recommendations [3]. The great potential of music to alter emotional states has increased its significance in various aspects of human existence [2]. Recognizing the strong relationship between emotions and music, our technique aims to offer musical selections that correspond to and influence the observed emotional state [2]. This effort to combine facial expression detection and music suggestions has the promise of providing a more immersive and enhanced user experience.

In tandem with this strategy, meticulous efforts are put into curating music databases, which include playlist crawls and manual annotations obtained from top music platforms [3]. By leveraging these datasets, we broaden the breadth of image processing outcomes, allowing for a more tailored and engaging user experience. This study aims to lay the groundwork for an innovative and integrated system that improves the user's emotional journey by smoothly combining the areas of facial expression detection and music suggestion.

The following sections of this paper delve into the methodology that underpins our approach, providing details on the design and training of the expression recognition model, the fusion of image processing results with the music recommendation algorithm, and the broader implications of our findings.

II. LITERATURE SURVEY

An extensive exploration into methods to discern users' behavioral and emotional states reveals a diverse landscape of research [2]. Various techniques, including facial expressions, gestures, body language, and speech analysis, have been employed to decode these emotional signals. Efforts to categorize emotional expressions on users' faces have spurred the development of different methodologies, involving feature extraction and classification algorithms [2].

ISSN No:-2456-2165

The foundational work by Ekman and Friesen introduced Action Units (AUs), capturing both fleeting and enduring facial traits to elucidate the direct correlation between facial muscle movements and expressed emotions [2]. This led to the establishment of the Facial Action Coding System, which delineates 44 action units representing emotions with varying intensities [2]. The pursuit of algorithms based on distinctive features aligns with Ekman's recommendations [2].

In parallel, geometric-based approaches for emotional analysis have surfaced [2]. These approaches rely on facial markers such as eye corners, lip contours, and brow movements to extract defining characteristics [2]. The distances between these points generate a feature vector that dynamically changes with shifting emotional states [2]. This vector becomes instrumental in identifying emotions using Support Vector Machines (SVM) and Radial Basis Function Neural Networks (RBFNN) [2].

Efforts to classify music based on lyrical analysis have faced challenges due to language barriers and the nuanced emotional expressions conveyed through music [2].

The literature also features endeavors aiming to merge facial emotion recognition with music recommendation systems [1, 5]. These algorithms strive to personalize music choices by analyzing collected facial expressions from images, thereby reducing the time complexity for users managing extensive playlists [1]. This integration holds promise for enhancing the music experience by aligning it with the user's emotional journey [1]. Recent research has developed systems for emotional identification and music recommendation based on users' facial expressions, utilizing artificial neural networks for emotion classification and customized playlist suggestions [5].

Moreover, sophisticated methodologies like Convolutional Neural Networks (CNNs) have emerged as potent tools for recognizing emotions [2]. Leveraging deep learning, these networks demonstrate an aptitude for grasping subtle emotional cues and have gained traction in emotion recognition applications [2]. The interplay between facial expression analysis and music recommendation systems clearly presents significant potential for augmenting user engagement and delivering tailored experiences [1].

III. PROPOSED SYSTEM AND METHODOLOGY

➢ *Proposed System Overview*

This paper marks a groundbreaking stride in the digital era, where deep learning reshapes various domains like image recognition and facial expression analysis. It introduces an integrated system that unites emotion recognition seamlessly with music recommendation. Powered by Convolutional Neural Networks (CNNs), this system adeptly captures and interprets users' emotional expressions from facial photos in real-time. This capability serves as a bridge, merging emotions with auditory experiences, unlocking novel ways to connect with users and enhance their emotional well-being.

The central aim of this innovative system is to intuitively engage users by uplifting their emotional state through a customized music playlist. Emotions, whether happiness, sadness, neutrality, or surprise, profoundly influence human responses to music. Leveraging facial micro-expression recognition, our system aims to automatically tailor music selections to match users' current emotional state [4].

This visionary system operates harmoniously through a comprehensive set of integrated modules:

- Real-Time Capture: This module acts as the initial step, precisely capturing users' facial expressions through cameras, laying the groundwork for emotion recognition.
- Face Recognition: Using CNNs, this module extracts features from captured facial images, delving into the nuanced realm of facial expressions to capture emotional subtleties.
- Emotion Detection: Central to the system's cognitive abilities, this module conducts intricate feature analysis to discern emotional nuances from facial expressions, providing crucial emotional context for music recommendation.
- Music Recommendation: A pinnacle achievement, this module curates music recommendations based on recognized emotions, merging the emotional journey with auditory preferences for an enriched user experience [4].

➢ *Methodology*

- *Database Description*

The emotion detection process anchors on a Convolutional Neural Network (CNN) model trained on the FER2013 dataset. This dataset, comprising 30,219 grayscale facial images sized at 48x48 pixels, segregates emotions like happy, sad, angry, surprise, and neutral for comprehensive training. Utilizing this dataset, the system's goal is to leverage deep learning for accurate facial expression identification and interpretation, unraveling the emotional essence within [4].

The Emotion Extraction Module orchestrates image capture, feature extraction, and CNN-based analysis to reveal users' emotional states. Grayscale images, captured via cameras or webcams, undergo advanced feature extraction to untangle facial landmarks and expressions. A trained network interprets these features to determine and label the user's emotional state.

Once emotions are decoded, the Audio Extraction Module recommends music or audio aligned with the user's expressed emotions. A personalized list of emotion-matched songs is presented to users, considering their preferences for an engaging musical experience.

The Emotion-Audio Integration Module seamlessly combines emotion-laden songs with the user's current emotional disposition. This module operates through a dynamic web interface leveraging technologies like PHP, MySQL, HTML, CSS, and JavaScript. It acts as the bridge for emotion-based audio selection, creating a harmonious blend between user sentiment and auditory pleasure [5].

ISSN No:-2456-2165

IV. MODULE DETECTION AND RECOMMENDATION

A. Emotion Detection Module: Deciphering the Language of Faces

In the domain of computer vision, face detection mirrors the complexity of deciphering the intricate language embedded in human expressions. It stands as a fundamental application, employing algorithms aimed at identifying faces or objects within images [4]. Consider these algorithms as the digital detectives, akin to Sherlock Holmes, finely tuned to spot faces amidst the visual cacophony. Face detection relies heavily on classifiers, the equivalent of the detective's magnifying glass. Their primary mission? Distinguishing whether an element within an image is a face (denoted as 1) or something else (denoted as 0). It's far from a trivial pursuit. These classifiers undergo rigorous training using extensive image datasets to achieve precision comparable to Sherlock's investigative skills.

Enter OpenCV, our trusted ally in this investigative journey, armed with two principal types of classifiers: Local Binary Pattern (LBP) and Haar Cascades [6]. Haar classifiers reign supreme in facial detection. Painstakingly trained on diverse facial data, they ensure pinpoint accuracy. Their ultimate goal is crystal clear: spotting faces within a frame while adeptly filtering out external distractions and noise. The secret lies in machine learning. Through cascade functions trained meticulously on input files, these classifiers use Haar Wavelet techniques, functioning as a metaphorical magnifying glass. These techniques break down image pixels into squares, employing machine learning methods to achieve accuracy akin to Sherlock's investigative prowess, a process aptly named "training data."

B. Feature Extraction: Unearthing the Hidden Gems

In the domain of deep learning, feature extraction resembles excavating precious gems within a chest of treasures. Envision a pre-trained network as an art connoisseur, discerning specific strokes over others. Input images traverse this network, pausing at set layers to acquire outputs as features—an artistic appreciation of a masterpiece, layer by layer. The initial layers in this convolutional network act as the connoisseur's focus on broader strokes, extracting high-level features through a limited filter set.

As we venture deeper, these layers emulate the connoisseur's magnifying glass, revealing intricate details. However, these filters are not ordinary; they specialize in capturing exquisite features, albeit with added computational intricacy [6].

C. Emotion Detection: Cracking the Emotion Code

Sure, let's delve into emotion detection—a fascinating process where Convolutional Neural Network (CNN) architecture takes the spotlight. Picture it as an ensemble of artists scrutinizing an image. These feature detectors, resembling artistic virtuosos, meticulously examine input images, unveiling distinct emotional strokes. Whether it's edges, lines, or curves, they dissect and analyze every aspect.

Now, the secret sauce—pooling. It's akin to stepping back to appreciate the broader canvas. Pooling ensures consistent outcomes even with minor input variations. It's about discerning patterns amidst chaos. Different pooling methods exist—like min, average, and max—but let's highlight max-pooling for its adeptness in capturing vital details.

Following this artistic analysis, the flattened inputs embark on a journey through a deep neural network. Imagine passing these artistic insights to a master storyteller proficient in deciphering the object's emotional state [6].

D. Music Recommendation Module: Crafting Musical Harmony

Ah, music - the universal language of emotions. In our quest to craft the perfect musical experience, we curate a database filled with Bollywood Hindi songs, each category brimming with 100 to 150 soul-stirring songs. You see, music isn't just a backdrop; it's a driving force behind emotions. So, when our emotion module detects a user's mood (say, they're feeling a bit blue), our system leaps into action. It recommends a curated playlist that resonates with their mood, effectively lifting their spirits.

But how do we make the magic happen? Real-time emotion detection is the linchpin. It labels emotions like Happy, Sad, Angry, Surprise, and Neutral. These labels serve as our guiding stars, leading us to the perfect musical constellation. The songs are meticulously organized into folders using Python's trusty `os.listdir()` method. When you see the playlist, it's not just a list of songs; it's a symphony of emotions. Each song is ordered based on how often the user listens to it.

And the cherry on top? The GUI of our music player. It's like the stage where the emotions and music come together for a grand performance. Pygame, a multimedia library, takes center stage for audio playback. Functions like `playsong`, `pausesong`, `resumesong`, and `stopsong` manage the music, while Tkinter, the magician of GUI development, creates the visual magic [4].

E. Eigenfaces Approach: Recognizing the Unique You

When it comes to perceiving facial expressions, the Eigenfaces approach is our trusted ally. It zeroes in on the most significant parts of your face - the eyes, nose, cheeks, and forehead. Why? Because these areas change relative to one another when you express emotions. It's like recognizing a friend by their unique smile or the twinkle in their eye. The Eigenfaces algorithm works its magic by capturing these crucial features in faces and employing Eigenvalues and Eigenvectors to tell them apart. It's all about capturing the maximum variations in facial features among different faces, just like recognizing your friend in a crowd.

F. Face Detection and Recognition: The Art of Reading Faces

Recognizing facial expressions is all about reading the intricate stories etched on faces. To decode these tales, we process images with human faces, detecting emotions conveyed. Here, different algorithms step into the spotlight - Eigenfaces, Local Binary Patterns, Direct Cosine Transform, and Gabor Wavelets. They work like literary experts dissecting the nuances of facial expressions. OpenCV and the Eigenfaces algorithm are our trusty companions, detecting faces within images and unveiling the emotional cues hidden within them. It's like understanding the unspoken language of faces.

G. Music Feature and Recommendation: Crafting Musical Stories

Music recommendation isn't just about songs; it's about weaving musical stories that resonate with your soul. We delve into the depths of music, analyzing factors like artist, album, and mood. Artificial Neural Networks (ANNs), the maestros of our orchestra, take center stage. They classify songs into various categories, painting musical portraits based on diverse criteria. The Million Song Dataset by Kaggle serves as our training ground, offering metadata and triplet files brimming with song information and user interactions. It's a treasure trove of musical insights. With ANN-based methods, we deliver accurate classifications and recommendations, crafting musical journeys that dance with your emotions.

Emotion Detection Module
Equation for ReLU activation function:
f(x) = max(0, x)
Feature Extraction
Pooling operation equation:

```
Pooled Feature = max(Pooling Area)
```

Music Recommendation Module
Code for real-time emotion detection using CNN:

```python
# Sample code for real-time emotion detection
import cv2
import numpy as np

# Load pre-trained model
model = cv2.dnn.readNetFromCaffe(protoTxt, modelFile)

# Capture video from webcam
cap = cv2.VideoCapture(0)
while True:
  ret, frame = cap.read()
  if not ret:
    break

  blob = cv2.dnn.blobFromImage(cv2.resize(frame, (300, 300)), 1.0, (300, 300), (104.0, 177.0, 123.0))
  model.setInput(blob)
  detections = model.forward()

  # Extract emotion from detection results
  emotion = emotions[np.argmax(detections)]

  cv2.putText(frame, f'Emotion: {emotion}', (10, 30), cv2.FONT_HERSHEY_SIMPLEX, 0.7, (0, 0, 255), 2)
  cv2.imshow('Emotion Detection', frame)

  if cv2.waitKey(1) & 0xFF == ord('q'):
    break

cap.release()
cv2.destroyAllWindows()
```

V. RESULTS

The fusion of facial expression recognition with music recommendation systems is a captivating endeavor in recent research. Visionaries in this field employ Convolutional Neural Networks (CNNs) to create innovative solutions that bridge facial expression recognition and music recommendation.

Athavle et al. [4] designed a system that not only recognizes emotions like happiness, sadness, anger, surprise, and neutrality but also curates music playlists that match the user's mood seamlessly.

Yu et al. [3] explored the subtleties of facial micro-expressions, achieving a remarkable 62.1% recognition rate. They integrated this achievement into their music recommendation algorithm, creating a symphony of emotions in music.

Krupa et al. [2] elevated the CNN approach with their two-level CNN model, achieving recognition accuracies of up to 88% through meticulous optimization, emphasizing the importance of multi-level feature extraction.

Nareen Sai et al. [7] reported substantial recognition accuracies using a CNN architecture, highlighting the rising importance of deep learning in facial expression analysis and its integration with music recommendation frameworks.

Metilda Florence and Uma [5] conducted emotion recognition experiments based on user facial expressions. In one set, explicit instructions led to perfect accuracy when inner emotions matched facial expressions. In another set without guidance, accuracy varied widely, reflecting the diversity of human emotions.

This research underscores the growing significance of emotion recognition through facial expressions across disciplines. While achieving accurate classification of user emotions is well within reach with accuracy rates exceeding 80% in various scenarios, challenges persist, such as acquiring suitable image data and ensuring well-lit environments for precise predictions.

ISSN No:-2456-2165

Table 1
Average accuracy (%) of each detected emotion

Emotions Classified	Accuracy Rate (%)
Happy	90
Neutral	80
Surprised	77
Disgusted	62
Angry	50
Fearful	37
Sad	28

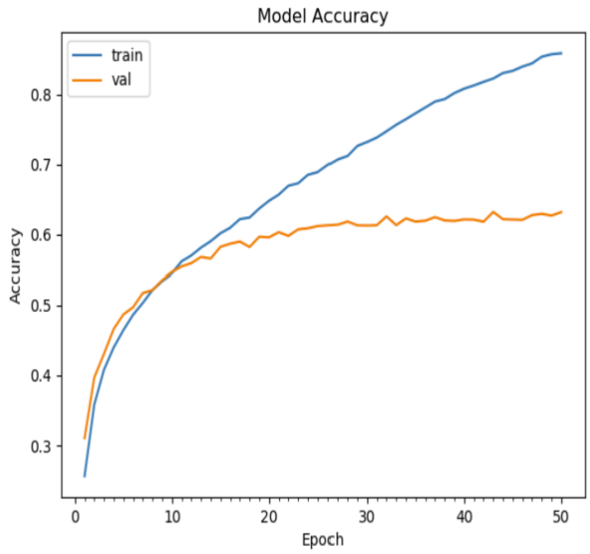


Fig 1 Model Accuracy

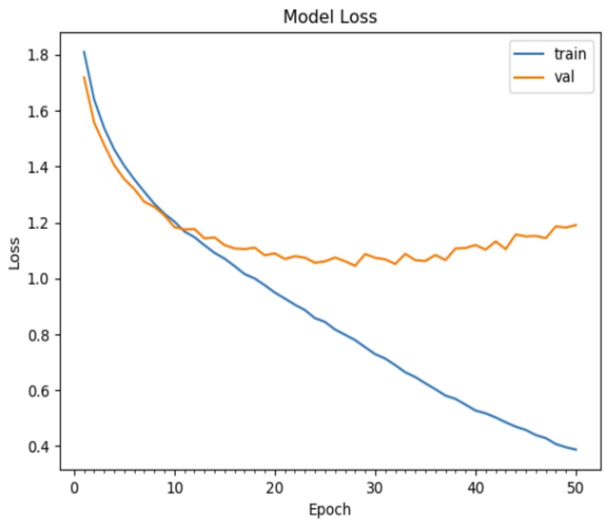
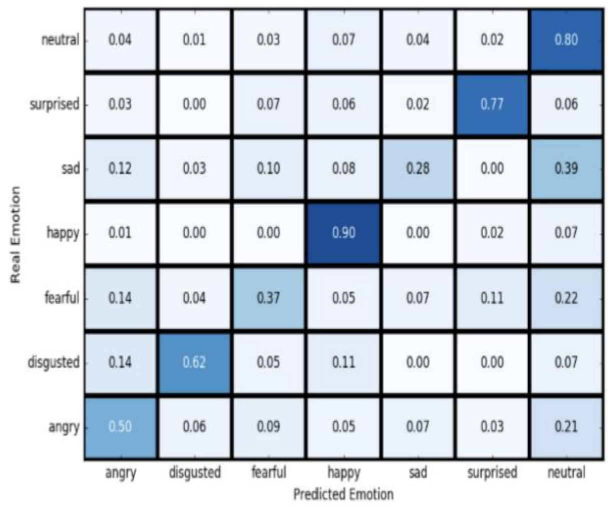
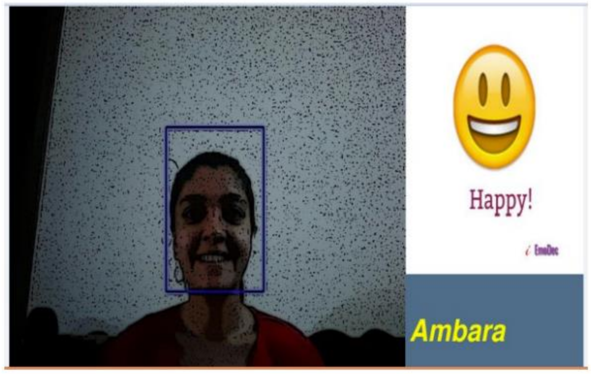


Fig 2 Model Loss



Performance matrix of the final model

Fig 3



Output of one of the emotions from the model

Fig 4

VI. CONCLUSION

In the realm of music recommendation systems, the integration with facial emotion recognition techniques has sparked substantial advancements. This technological marriage not only refines the precision in recognizing users' feelings through facial cues but also refines the emotional depth of music recommendations.

Take, for instance, Yu et al. [3], who introduced a model based on convolutional neural networks (CNNs) to decode facial microexpressions. This achievement hit a recognition rate of 62.1%, laying the groundwork for a groundbreaking recommendation algorithm. Its real-time emotion recognition crafts music suggestions tailored to users' current emotional states, not just based on their past history.

Then there's Krupa et al. [2] and their emotion-sensitive smart music recommender. It extends beyond music, delving into a deeper understanding of users. By integrating chatbot interactions and facial expression-based emotion detection, they've fashioned a system reaching beyond songs, embracing areas like driver assistance, lie detection, surveillance, and mood-based learning.

Gilda et al. [1] offer a smart music player seamlessly blending facial emotion recognition into mood-based music recommendations. It's not just smart; its striking 97.69% accuracy minimizes user effort in playlist creation, emphasizing technology adapting to emotions.

Athavle et al. [4] pioneered a music recommendation system that doesn't merely shuffle tracks but orchestrates mood transformations. From happiness to surprise, it gauges emotions and crafts playlists in sync. It's an application demonstrating music's magical influence on moods, enhancing user experiences.

Collectively, these studies highlight the potential of merging facial emotion recognition with music recommendation. While paving remarkable paths, challenges persist. The precision in recognizing microexpressions demands enhancement, and issues like adverse lighting conditions cast shadows. Nonetheless, these hurdles are mere stepping stones toward more personalized and emotionally synchronized music experiences.

In essence, the intersection of emotions and music yields magic. These advancements extend beyond melodies, influencing domains like mental health therapy and gaming. As researchers refine these systems, their promise to enhance user well-being and satisfaction grows, ensuring technology resonates more profoundly with the human heart.

FUTURE SCOPE

The evolution of uniting emotions with music recommendation systems continues, paving the way for unexplored territories ripe with innovation:

Exploring Nuanced Emotions: What if the system could discern even the subtlest shades of disgust and fear? Future research could expand the spectrum of recognized emotions, incorporating these intricate sentiments. It's about technology comprehending not only smiles but also the intricate tapestry of human emotions.

Illuminating Dark Spaces: Adverse lighting and low-quality camera resolutions present hurdles. Future systems should excel in any setting, ensuring emotions are never concealed in the dark. Imagine a system that perceives your emotions even in dimly lit environments.

Personalizing Melodies: Collaborative filtering techniques hold the promise of a more personalized musical journey. These systems could tune in not just to your mood but also your musical preferences, creating a symphony that deeply resonates with your inner self.

Healing Through Harmony: Beyond entertainment, the future might witness these systems stepping into the domain of music therapy. Picture therapists utilizing emotion detection to craft sessions alleviating stress, anxiety, depression, or trauma. It's the fusion of technology and mental well-being.

As these advancements unfold, it's crucial to strike a balance between refining algorithms and practical application. The goal is clear: to create technology that not only comprehends us but also touches the strings of our emotions, enhancing our experiences and addressing our emotional well-being.

The future promises a symphony of possibilities, where innovation and emotions coalesce, ensuring a more enriching and harmonious tomorrow.

LIMITATIONS

The system has its share of limitations, contributing to a nuanced understanding of its capabilities. Primarily, the system's emotional understanding remains confined within the limits of its dataset. This constraint restricts its capacity to discern the entire breadth of human emotions, emphasizing the critical role of comprehensive and diverse data in training such systems.

Another significant factor influencing the system's performance is lighting. Similar to the nuances of photography, the system thrives in well-lit environments where it can accurately detect and interpret facial expressions. However, in dimly lit surroundings, its precision might be compromised, indicating the importance of optimized lighting conditions for optimal functionality.

Moreover, the quality of the images greatly impacts the system's accuracy in emotional interpretation. It prefers images of higher resolution, preferably at least 320p, to vividly capture and decode emotional nuances. Clearer and sharper images contribute to a more accurate portrayal of emotions, underscoring the importance of image quality in the system's operations.

Recognizing these limitations is pivotal as it propels ongoing efforts towards system enhancement. The system continually strives to evolve, seeking improvements that surmount these challenges, fostering a deeper and more precise alignment with human emotional expressions.

REFERENCES

[1]. S. Gilda, H. Zafar, C. Soni and K. Waghurdekar, "Smart music player integrating facial emotion recognition and music mood recommendation," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 2017, pp. 154-158, doi: 10.1109/WiSPNET.2017.8299738.

[2]. K. S. Krupa, G. Ambara, K. Rai and S. Choudhury, "Emotion aware Smart Music Recommender System using Two Level CNN," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2020, pp. 1322-1327, doi: 10.1109/ICSSIT48917.2020.9214164.
[3]. Z. Yu, M. Zhao, Y. Wu, P. Liu and H. Chen, "Research on Automatic Music Recommendation Algorithm Based on Facial Micro-expression Recognition," 2020 39th Chinese Control Conference (CCC), Shenyang, China, 2020, pp. 7257-7263, doi: 10.23919/CCC50068.2020.9189600.
[4]. M. Athavle, D. Mudale, U. Shrivastav, and M. Gupta (2021). "Music Recommendation Based on Face Emotion Recognition." Journal of Informatics Electrical and Electronics Engineering, Vol. 02, Iss. 02, S. No. 018, pp. 1-11. 2021. https://doi.org/10.54060/JIEEE/002.02.018
[5]. S. Metilda Florence and M. Uma, "Emotional Detection and Music Recommendation System based on User Facial Expression," presented at the 3rd International Conference on Advances in Mechanical Engineering (ICAME 2020), IOP Conf. Series: Materials Science and Engineering 912 (2020) 062007, IOP Publishing, doi: 10.1088/1757-899X/912/6/062007.
[6]. Samuvel, D. J., Perumal, B., & Elangovan, M. (2020). Music recommendation system based on facial emotion recognition. 3C Tecnología. Glosas de innovación aplicadas a la pyme. Edición Especial, Marzo 2020, 261-271.
[7]. Nareen Sai, B., Sai. Vamshi, D., Pogakwar, P., Seetharama Rao, V., Srinivasulu, Y. "Music Recommendation System Using Facial Expression Recognition Using Machine Learning." International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 10, Issue VI, June 2022. ISSN: 2321-9653. DOI: https://doi.org/10.22214/ijraset.2022.44396.