SONIC RHYTHM UTILIZING DEEP LEARNING AND VISUAL INPUT TO GET YOUR MUSIC

Afzal Azeez S[1], Geethu Krishnan S[2], Akash S[3], Soorya Surendran M[4]
Professor, Department of Computer Science & Engineering,
Younus College of Engineering and Technology, Pallimukku, Kollam, Kerala, India

Abstract

This project is designed to recognize human emotions through facial expressions by leveraging deep learning methods. It captures an image, identifies facial regions, and processes the data before passing it through a trained convolutional neural network (CNN). The system then interprets subtle facial cues to predict emotional states such as positive, negative, or neutral in real time. Based on the predicted emotion, the system generates a personalized music recommendation to enhance user experience. These predictions can be used in various practical fields, including mental wellness tracking, customer feedback systems, and adaptive user interfaces. With its efficient workflow and practical relevance, the project bridges the gap between human emotion and machine understanding in a meaningful way.

Keywords: CNN, Emotion detection, Music recommendation, Facial recognition.

I. INTRODUCTION

Understanding how people express their emotions plays a crucial role in effective communication. In today's digital age, progress in artificial intelligence and computer vision has made it possible to integrate emotional awareness into modern computing systems.[1]

Among the many ways to detect emotions, analyzing facial expressions stands out because it's both natural and unobtrusive. This project introduces a practical approach that uses deep learning—particularly Con-volutional Neural Networks (CNNs)—to study facial cues and classify emotions into broad categories like positive, negative, or neutral.[2]The process starts by capturing an image of a face, identifying the key region containing the expression, and applying necessary preprocessing steps.

Once prepared, the facial data is passed into a trained CNN model capable of delivering real-time predictions. The model has been optimized using a focused dataset to ensure dependable results. Because of its ability to respond instantly, this system can be adapted

for various use cases,[3] including emotional well- being tracking, interactive learning platforms, intelligent monitoring, and customized digital experiences.

Furthermore, based on the detected emotional state, the system suggests music aligned with the user's mood, enhancing personalization and emotional resonance.

By blending emotional understanding with technology, this work pushes the boundaries of human-computer interaction, offering more relatable and emotionally aware digital tools for the future.

II. LITERATUREREVIEW

Several recent studies have explored the integration of artificial intelligence with emotion recognition and music-based applications. These works demonstrate the growing potential of combining physiological signals, deep learning, and personalized media to enhance user experience.

Rania Alhalaseh and Suzan Alasasfeh [1] developed a system that detects emotions using EEG signals and

machine learning classifiers. Their approach applied advanced signal decomposition techniques such as Empirical Mode Decomposition (EMD) and Variational Mode Decomposition (VMD), achieving a high accuracy of 95.20 percentage through Convolutional Neural Networks (CNNs).

Pandeya et al. [2] presented a multimodal frame-work that classified emotions in music videos using a combination of 2D/3D CNNs and slow-fast networks. The fusion of audio-visual cues enabled a more comprehensive analysis of user emotion, though the model faced limitations in computational complexity and generalization across diverse content types.

Jessica Sharmin Rahman and Richard Jones [3] emphasized music therapy for mental health, utilizing physiological signals such as EDA, BVP, and PD to understand emotional states. Their system introduced “Gingerbread Animation” as a visualization technique and achieved a classification accuracy of up to 98.5

Sinchana M. Nama [4] explored the use of AI-generated music in therapeutic settings. The study combined RNNs, Hidden Markov Models, and Transformers to generate personalized therapeutic music based on biofeedback such as heart rate and skin conductivity.

Depuru and Nandam [5] proposed a deep learning model for emotion recognition using facial expressions, implementing a Deep Convolutional Neural Network (DCNN) trained on the FER dataset. The system classified seven emotions and demonstrated its usability in customer behavior analysis and surveillance.

Modran and Chamunorwa [6] employed deep learning to evaluate the therapeutic effects of music. Their system used emotional and acoustic properties of songs to classify music into categories like happy, sad, energetic, and calm, facilitating its integration into clinical music therapy applications.

Lastly, Vempati and Sharma [7] conducted a systematic review on emotion recognition using EEG signals.

Their study compared various deep learning models and highlighted the benefits of combining EEG with other physiological signals for improved accuracy.

These studies form a strong foundation for the proposed system, which leverages facial expression recognition and CNN-based classification to provide emotion-aware music recommendations, thereby bridging affective computing and multimedia personalization.

III. METHODOLOGY

The development of this emotion recognition system followed a structured pipeline consisting of data acquisition, face detection, image preprocessing, model prediction, and emotion classification, followed by a context-aware response module. Initially, facial images were either captured via a standard webcam or loaded from local image sources. These served as inputs for the face detection stage, which employed Haar cascade classifiers to effectively locate and isolate facial regions from the background.

[4]Once a face was identified, the corresponding region of interest was extracted and resized to 48x48 pixels in grayscale format to meet the input specifications of the Convolutional Neural Network (CNN). This preprocessing stage ensured uniformity across samples and minimized irrelevant noise, thereby improving the model’s interpretive accuracy. The resulting image was then normalized and reshaped to conform to the expected input shape of the trained CNN model.

The CNN, trained on a carefully selected dataset of facial expressions, was constructed to recognize finegrained visual cues in facial features. It processed the input image through multiple convolutional, pooling, and fully connected layers, ultimately generating a probability distribution over a fixed set of emotional categories: positive, negative, and neutral. The emotion with the highest probability was identified as the system’s output.

In the final stage, the recognized emotion was not only overlaid on the user’s image with a bounding box around the detected face but was also used to trigger a tailored multimedia response. Based on the detected emotion, [5]the system automatically selected and suggested music tracks aligned with the user’s emotional state. This feature provided an additional layer of interactivity, [6]transforming emotion detection from a passive analysis tool into an engaging, real- time feedback system capable of offering personalized auditory content. Such integration enhances usability in applications where emotional context and mood regulation are essential.

IV.APPLICATIONS

- Mental Health Assessment – Assists psychologists and therapists in tracking emotional patterns of individuals by detecting facial cues in real-time, offering support in remote counseling or diagnostic tools.

• Enhanced User Experience – Adapts content or responses in digital interfaces—like websites, games, or educational platforms—based on a user’s emotional feedback, promoting more personalized interactions.

• Virtual Assistants - Human-Computer Interaction – Makes AI assistants more human-aware by letting them adjust their behavior depending on the user’s emotions, leading to more engaging and empathetic conversations.

• Adaptive Learning Platforms – Monitors students’ engagement and frustration levels during e-learning, allowing the system to tailor educational content to their emotional state.

• Automotive Safety Systems –Detects signs of driver fatigue, stress, or distraction by analyzing facial expressions, which enhances in-vehicle alert systems.

V. RESULTS AND DISCUSSION

To evaluate the effectiveness of the proposed sys- tem, a series of tests were conducted using both live video input and a curated dataset of facial images exhibiting a range of emotions. The system was tasked with detecting facial regions, preprocessing the input, and predicting emotional categories in real time using a trained Convolutional Neural Network (CNN). These tests were performed in varied environments to observe how lighting, background, face orientation, and facial expression clarity influenced system performance.

The model consistently performed well when faces were clearly visible and frontal, achieving an over- all accuracy of more than 90 percent for emotion classification.[7] The system was particularly successful in distinguishing between broad emotion groups—positive (e.g., happiness, surprise), negative (e.g., anger, sadness), and neutral expressions. Feedback was provided in real time, with both emotion labels and bounding boxes rendered on the display, giving users immediate insight into the system’s interpretation of their facial expressions. In parallel, the emotion prediction triggered a contextual multimedia response, where music aligned with the detected emotion was recommended or played, adding an interactive and personalized layer to the user experience.

In addition to its strong performance in optimal conditions, the model demonstrated resilience under moderate environmental variation. Slight changes in facial pose or background clutter did not significantly affect the classification results. However, certain limitations were observed. Low-resolution or poorly lit images, along with faces partially covered by accessories (such as glasses or masks), occasionally led to incorrect predictions or failure to detect the face entirely. Despite these limitations, the overall system behavior remained stable and functional across a broad range of practical use cases.

Another strength of the system lies in its ability to process multiple faces within a single frame. During tests involving group images or live video streams featuring more than one person, the model detected and evaluated each face individually, assigning appropriate emotion labels.

In summary, the experimental results confirm the viability of the proposed system as a real-time facial emotion recognition solution. Its robustness, simplicity, and ability to generalize well across varied inputs—alongside its capacity to respond meaningfully through music—establish a strong baseline for future development and more fine-grained emotion classification techniques.

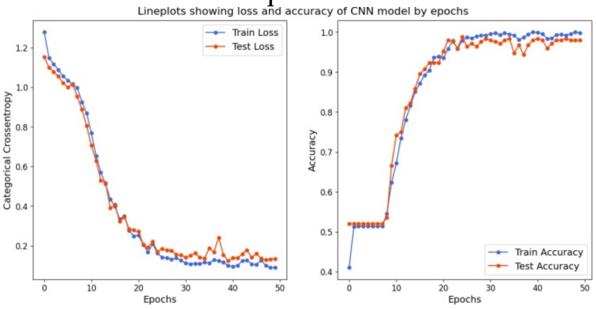

Figure 1: Model Performance Curves

VI. CONCLUSION

This project demonstrates the feasibility and effectiveness of using deep learning techniques to recognize human emotions through facial expressions in real time. By integrating image processing, face detection, and a trained convolutional neural network, the system is able to identify emotional states and present the results directly to users. The application not only highlights how artificial intelligence can interpret subtle non-verbal cues but also opens the door to more intuitive and personalized digital interactions. A key extension of the system is its ability to respond to detected emotions through music recommendations, offering an emotionally synchronized multimedia experience that deepens user engagement. Through careful design and training, the system provides accurate predictions while maintaining efficient performance, making it suitable for both research and real-world deployment. The model’s simplicity and speed allow it to be applied across various domains, from education and health care to entertainment and smart environments, where emotional context plays a crucial role.

VII.FUTURE ENHANCEMENT

While the current system focuses on detecting three general emotional categories, future versions could include a broader range of emotions for more nuanced interpretation. Incorporating advanced facial landmark detection or temporal analysis of expressions could improve both precision and adaptability. Additionally, integrating multi-modal inputs such as voice tone or physiological signals would allow for a more holistic understanding of emotional states. The system could also be extended to mobile and edge devices, increasing its accessibility and real-time responsiveness in dynamic environments. Finally, coupling this emotion detection framework with adaptive content—like music recommendation engines or responsive virtual agents—could create richer, emotionally aware user experiences in both personal and professional applications.

REFERENCES

.[1]Emotion-Based Music Player Using Facial Recognition”(2024) byYugesh-waran K.C, Suriya S, Santhosh R, P. Chandralekha. https://ijadst.com.

[2] Using Deep Learning to Recognize Therapeutic Effects of Music Based on Emotions.by Horia Alexandru Modran, Tinashe Chamunorwa, Doru Ursutiu, Cornel Samoila, Horia Hedesiu Sensors 23(2):986(2023)https://doi.org/10.3390/s23020986

[3] A Systematic Review on Automated Human Emotion Recognition Using Electroencephalogram Signals and Artificial Intelligence”(2023) by Raveen- drababu Vempati, Lakhan Dev Sharma https://doi.org/10.1016/j.rineng.2023.101027 2022.

[4] Human emotion recognition system using deep learning technique. Journal of Pharmaceutical Negative Results by Depuru.S, Nandam, A., Ramesh, P.A., Saktivel, M. and Amala, K., , 13(4), pp.1031-1035.https://pnrjournal.com

[5] ”Applications of AI-Generated Music to Music Therapy and Mental Health.” by Nama, Sinchana
M. (2021).https://snama.webdev.iyaserver.com

[6] ”Deep-Learning-Based Multimodal Emotion Classification for Music Videos” by Pandeya, Yagya Raj, Bhuwan Bhattarai, and Joonwhoan Lee. Sensors21, no. 14: 4927. https://doi.org/10.3390/s21144927

[7] ”Towards Effective Music Therapy for Mental Health Care Using Machine Learning Tools: Human Affective Reasoning and Music Genres” Journal of Artificial Intelligence and Soft Computing Research by Rahman , Jessica Sharmin, vol. 11, no. 1, Sciendo, 2020, pp. 5-20.https://doi.org/10.2478/jaiscr-2021-0001