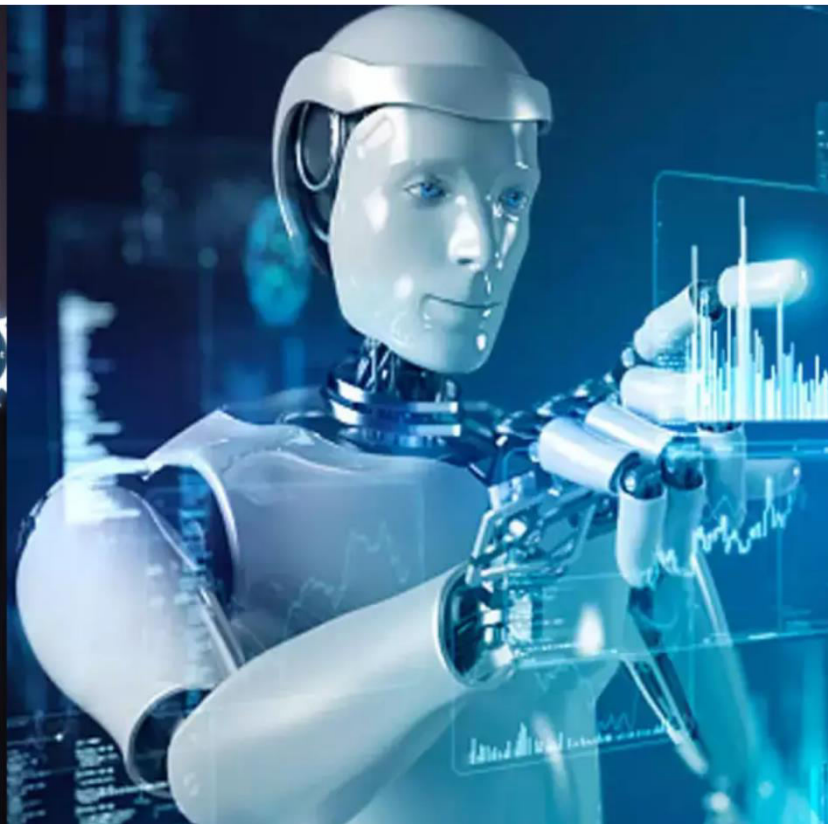




International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Sentiment Sounds

Arnav Jadhav, Aadya Parasnis, Shardul Vaidya, Bharat Patil

Polytechnic Student, Department of AI & ML Engineering, AISSMS Polytechnic College, Pune, India

Polytechnic Student, Department of AI & ML Engineering, AISSMS Polytechnic College, Pune, India

Polytechnic Student, Department of AI & ML Engineering, AISSMS Polytechnic College, Pune, India

Head of Department, Department of AI & ML Engineering, AISSMS Polytechnic College, Pune, India

ABSTRACT: Music has a strong effect on human emotions, and it has been extensively explored. Music evokes memories, heightens emotions, and, in certain situations, serves as treatment for patients. As artificial intelligence grows more widespread in all areas of the economy, the music industry is on the cusp of one of the most significant technological revolutions. In this paper, we describe the "Sentiment Sounds" system, a deep learning-based system that can detect human emotions from facial expressions and produce music based on the observed mood. Based on the ReLU as the foundation, this system intends to provide next-generation individualized music experiences. Beyond personal amusement, "Sentiment Sounds" has the potential to solve difficulties in mental health and productivity through therapeutic and adaptive soundscapes in real time.

The parts that follow discuss the inspiration for this innovation, a detailed examination of existing literature, the suggested approach, prospective applications, obstacles, and the transformative system's broad future possibilities. Through "Sentiment Sounds," we hope to demonstrate how technology may enhance the relationship between music and human emotions, opening up new avenues for participation and well-being.

KEYWORDS: Deep Learning, Emotion Detection, ReLU (Rectified Linear Unit), Facial Expression Analysis, Convolutional Neural Networks (CNNs).

I. INTRODUCTION

Music serves as a universal language, capable of transcending cultural, geographical, and linguistic divides. Its capacity to evoke feelings and affect one's mindset has secured its role in human fate. Music, be it the soothing tunes of classical compositions or the energetic beats of modern pop songs, has historically served as a form of emotional expression and a medium of communication.

With the emergence of Artificial Intelligence (AI) and Machine Learning (ML), the classic connection between music and emotions has undergone a transformation. Music experiences can now be tailored in ways that were previously unachievable. AI systems have the capability to analyse large amounts of data, identify patterns, and adjust to personal preferences in ways that were once thought to be impossible.

II. MOTIVATION OF THE PROJECT

"Sentiment Sounds" was born out of the need for a more personalized and intuitive musical experience. Traditional music recommendation systems mostly rely on user behaviour, including playlists, likes, and preferences. Despite its effectiveness, this method lacks real-time adaptability.

Emotions are fleeting and largely incomprehensible. Better emotional well-being and a more engaging musical experience can result from a system that is able to recognize these feelings and adjust in real-time. This approach bridges the gap between static music recommenders and dynamic, emotion-featured soundscapes by combining deep learning and AI-powered automated information retrieval techniques. Given how much AI is influencing other businesses, the music sector ought to be the next to undergo a significant transformation. It has the ability to change how people choose, create, and listen to music by bridging the gap between human emotion and computer intelligence. Innovations like "Sentiment Sounds" serve as examples of what artificial intelligence

(AI) is capable of in the area of music personalisation. New opportunities will open for creativity and participation. performed to label each objectseparately. Finally, the set of selection criteria is applied to filter out non text regions. After text detection, text inpainting is accomplished by using exemplar based Inpainting algorithm.

III. LITERATURE SURVEY

Much progress has been made both in emotion detection and music recommendation systems. The current emotion detection methods are relying on a variety of methods, for example:

- Facial Emotion Recognition: Research studies by Ekman [1] have enabled the development of a framework for the recognition of innate facial expressions that express emotions such as experiencing happiness, sadness, anger, or surprise.
- Deep Learning Algorithms: Convolutional Neural Networks (CNNs) and activation functions such as ReLU have exhibited high accuracy in certain applications such as recognition tasks, including emotion classification [2].
- Music Recommendation Systems: There is for example, use of collaborative filtering and content-based filtering algorithms for music recommendations by companies such as Spotify and Apple Music. Nevertheless, such systems cannot be adapted in real time with respect to the user's emotional state.
- This work of convergence of these technologies is applied in "Sentiment Sounds" to develop a bridge system between emotion recognition and adaptive music recommendations.

IV. PROBLEM STATEMENT

Currently, the issue faced by the ongoing project is that most of the available music recommendation systems are not real-time adaptive. Most of them also fail to take into consideration the emotional state of users while making recommendations, failing to create a more realistic, personalized experience.

Emotion detection systems also exist however, they are mostly fragmented and compartmentalized from other applications such as music streaming. "Sentiment Sounds" brings these domains together in the creation of a new system for improving the user interface of mood-based music selections.

V. PROPOSED DEEP LEARNING ALGORITHMS

The heart of "Sentiment Sounds" is its emotional recognition algorithm using ReLU as activation function. The system architecture contains the following:

- Facial Data Acquisition: The camera feed is captured in real time to obtain the facial expression of the user.
- Preprocessing: Images are preprocessed by being converted to grayscale and their noise are removed to improve the detection accuracy.
- Feature Extraction: A CNN extracts facial features linked to emotions.
- Classification: The ReLU activation function is applied to the convolutional neural network (CNN) hidden layer to classify emotions effectively.
- Music Mapping: The recognized emotions are mapped to a pre-defined database music track, which are sorted into mood categories (i.e., happy, calm, and excited).
- Playback: The selected music track plays and gives instant feedback to the user.

The use of ReLU activation function is simple, and helps with the issue of vanishing gradients of the deep learning model.

VI. APPLICATIONS

"Sentiment Sounds" has numerous applications such as:

- **Mental Health:** Personalized music therapy for stress, anxiety, and depression management.
- **Entertainment:** Enhancing the experience of the user in the world of streaming services, gaming, and virtual reality.
- **Workplace Productivity:** Context-aware music to improve focus and productivity at work.
- **Retail and Hospitality:** Emotional background music to enhance customer experiences in stores, hotels, and restaurants.

VII. CHALLENGES

Although "Sentiment Sounds" have immense future impact, there are wide open challenges:

- **Accuracy of Emotion Detection:** The real-time emotion detection accuracy is high, since matched/mismatched recommendations cannot be given.
- **Privacy Issues:** The constant camera feed raises ethical and privacy concerns.
- **Cultural Differences:** Affective expressions and musical preference are different across cultures therefore, large data sets are needed for correct localization.
- **Hardware Constraints:** Real-time processing demands efficient hardware to avoid latency.
- **Dataset Bias:** All the training datasets should be heterogeneous in order to provide fairness and inclusiveness.

VIII. FUTURE SCOPE

The application range of "Sentiment Sounds" is not restricted to the present version. Future versions may comprise:

- **Integration with Streaming Platforms:** Integration into the Spotify, Apple Music, and YouTube Music for smooth integration.
- **Multi-Modal Emotion Detection:** Combining voice tone, body language and physiological signals to enhance precision.
- **Cross-Cultural Adaptability:** Expanding datasets and algorithms to accommodate different cultural contexts.
- **Wearable Devices:** Integration with smartwatches and AR/VR devices for a more enriching experience.
- **Therapeutic Applications:** Advanced research into the impact of music on mental health, allowing for targeted interventions for PTSD and ADHD, among others.

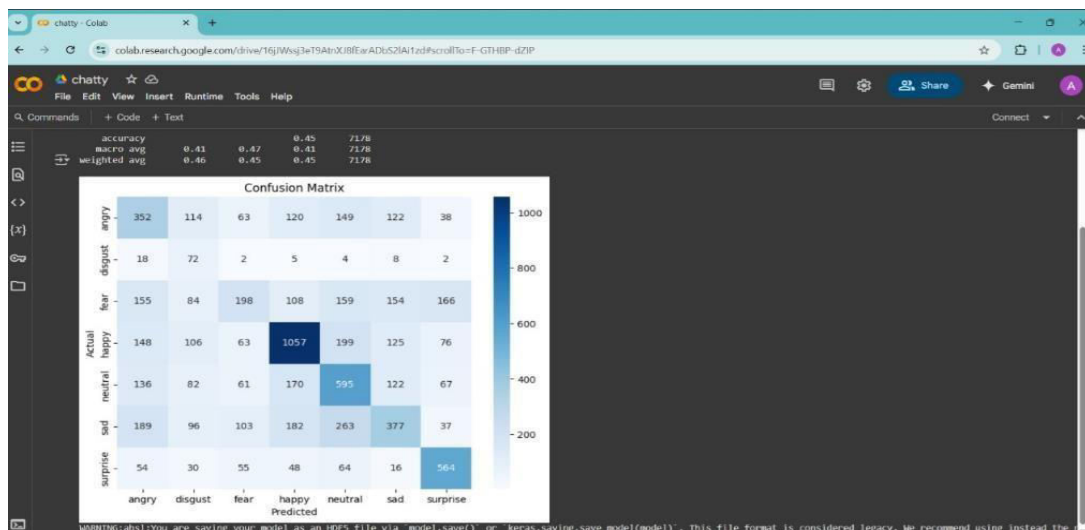
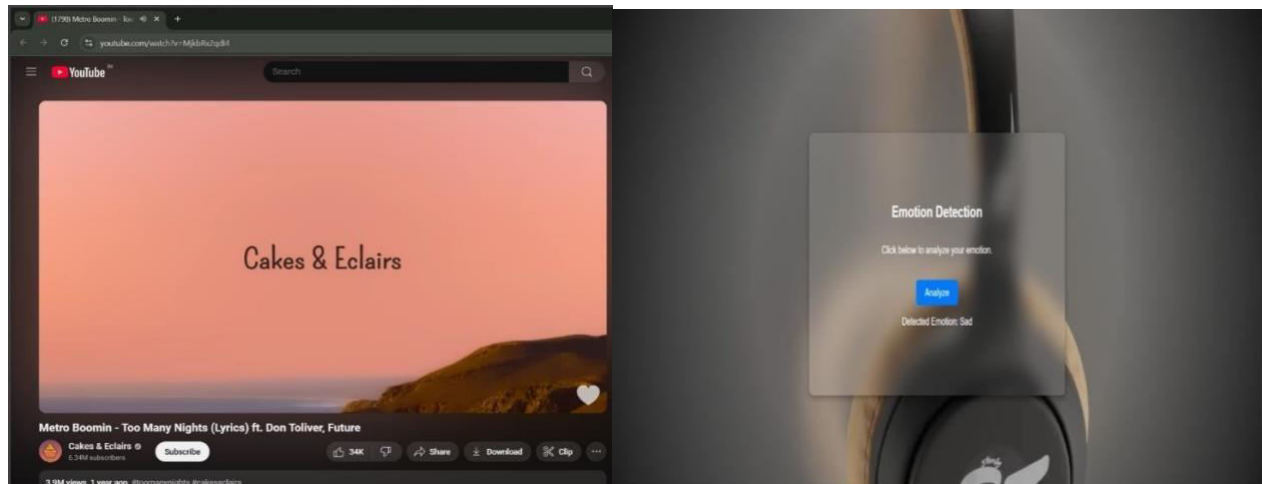
IX. CONCLUSION

"Sentiment Sounds" is the largest step toward integrating emotion detection and music recommendation systems. The system based on a deep learning algorithm in the integration of ReLU activation function achieves a novel, real-time user experience advantage for users. However, also in this case, due to difficulties related to privacy and cultural variability of trends, the scope of implementation is enormous and disruptive. This marks the beginning of tomorrow when music would not only entertain but also heal, motivate, and connect.

ACKNOWLEDGMENT

We would like to express their gratitude to all of the researchers, developers, and academics whose contributions have laid the groundwork for advances in emotion identification and music recommendation systems. Special thanks go to the teams behind open-source deep learning frameworks like TensorFlow and PyTorch, which played a significant role in the development of this project. Second, a heartfelt thank you to everyone who worked with and reviewed "Sentiment Sounds," providing vital comments and assistance during the creation process.

OUTPUT



X. METHODOLOGY

The methodology adopted for the development of the "Sentiment Sounds" project follows a systematic and modular approach combining facial emotion recognition with music recommendation. It consists of several stages—data collection, model training, real-time emotion detection, emotion-to-music mapping, and final playback execution. This approach ensures the system is adaptable, efficient, and suitable for real-world applications.

- 3.1 Data Acquisition and Preprocessing
- The first phase involves gathering facial expression data. For this purpose, the FER-2013 dataset was used, which includes thousands of labeled grayscale facial images categorized into basic emotion classes (e.g., happy, sad, angry, surprise, fear, disgust, and neutral). Images are resized to a uniform dimension (48x48 pixels), normalized, and converted to grayscale to reduce processing complexity.
- 3.2 Model Development and Training
- A Convolutional Neural Network (CNN) was designed for the classification of facial emotions. The architecture includes several convolutional layers with ReLU activation, followed by max-pooling layers to reduce dimensionality. Dropout layers are added to prevent overfitting. The output layer uses softmax activation to

classify the image into one of the predefined emotion categories. The model is trained using the Adam optimizer and categorical cross-entropy as the loss function.

- 3.3 Real-Time Emotion Detection
- After the model is trained, it is integrated with OpenCV to work with real-time webcam input. Frames are captured, faces are detected using Haar Cascade Classifier, and the region of interest is cropped and passed to the CNN model for emotion classification. The detected emotion is then displayed on the screen in real-time.
- 3.4 Emotion-to-Music Mapping
- Each classified emotion is mapped to a specific playlist or category of music stored locally or accessible via a streaming API. For instance, a happy emotion might trigger upbeat and energetic tracks, whereas a sad emotion would lead to calm or soothing music. A random selection function ensures variety while preserving emotional consistency.
- 3.5 User Interface and Integration
- A simple user interface (UI) was built using Python's Tkinter module, allowing the user to start or stop the detection, view current emotions, and manage playback. The system is lightweight and requires only basic hardware: a webcam and speakers/headphones.
- 3.6 Testing and Evaluation
- The system was tested under various lighting conditions, emotional expressions, and facial angles to evaluate the accuracy and responsiveness. Performance metrics such as accuracy, confusion matrix, and F1-score were analyzed, and user feedback was collected to ensure usability and satisfaction.
- This modular methodology ensures that the system remains scalable, reliable, and easily upgradable for future developments such as multi-modal emotion input and cloud-based music recommendation

XI. HARDWARE/SOFTWARE USED

Note on Hardware Usage:

This project is completely software-based and does not involve any custom hardware components. The only hardware utilized is a standard webcam integrated into or connected to a computer system. The application is designed to run on general-purpose computing devices, making it accessible and cost-effective for a wide range of users without the need for specialized hardware.

- 3.1 Data Acquisition and Preprocessing
- The first phase involves gathering facial expression data. For this purpose, the FER-2013 dataset was used, which includes approximately 28,700 labeled grayscale facial images categorized into seven basic emotion classes: anger, sadness, happiness, neutral, fear, disgust, and surprise. Images are resized to a uniform dimension (48x48 pixels), normalized, and converted to grayscale to reduce processing complexity.
- 3.2 Model Development and Training
- A Convolutional Neural Network (CNN) was designed for the classification of facial emotions. The architecture includes several convolutional layers with ReLU activation, followed by max-pooling layers to reduce dimensionality. Dropout layers are added to prevent overfitting. The output layer uses softmax activation to classify the image into one of the predefined emotion categories. The model is trained using the Adam optimizer and categorical cross-entropy as the loss function.
- 3.3 Real-Time Emotion Detection
- After the model is trained, it is integrated with OpenCV to work with real-time webcam input. Frames are captured, faces are detected using Haar Cascade Classifier, and the region of interest is cropped and passed to the CNN model for emotion classification. The detected emotion is then displayed on the screen in real-time.
- 3.4 Emotion-to-Music Mapping
- Each classified emotion is mapped to a specific playlist or category of music stored locally or accessible via a streaming API. For instance, a happy emotion might trigger upbeat and energetic tracks, whereas a sad emotion would lead to calm or soothing music. A random selection function ensures variety while preserving emotional consistency.
- 3.5 User Interface and Integration
- A simple user interface (UI) was built using Python's Tkinter module, allowing the user to start or stop the detection, view current emotions, and manage playback. The system is lightweight and requires only basic

hardware: a webcam and speakers/headphones.

- 3.6 Libraries and Packages Used
- OpenCV: For camera access and face detection
- NumPy: For efficient array manipulation
- TensorFlow & Keras: For model building, training, and evaluation
- Tkinter: For building a basic graphical interface
- Playsound / VLC Module: For local music playback based on emotional output
- 3.7 Testing and Evaluation
- The system was tested under various lighting conditions, emotional expressions, and facial angles to evaluate the accuracy and responsiveness. Performance metrics such as accuracy, confusion matrix, and F1-score were analyzed, and user feedback was collected to ensure usability and satisfaction.
- This modular methodology ensures that the system remains scalable, reliable, and easily upgradable for future developments such as multi-modal emotion input and cloud-based music recommendations.

REFERENCES

- [1] Ekman, P., "Basic Emotions," in Handbook of Cognition and Emotion Wiley, 1999).
- [2] Winning ImageNet classification at human-level performance, "Surpassing human-level performance on ImageNet classification", in Proceedings of the IEEE International Conference on Computer Vision, 2015.
- [3] Kotte, K. R., & Panyaram, S. (2025). Supply Chain 4.0: Advancing Sustainable Business. Driving Business Success Through Eco-Friendly Strategies, 303.
- [4] Spotify Research, "Emotion-Based Music Recommendation," Accessed 2024.
- [5] From Emotion to Sentiment", IEEE Transactions on Affective Computing, 2017.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details