Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 5, April : 2024

PLAYING MUSIC AS PER FACIAL MOODS USING CNN ALGORITHM

Prasad Bhosle, Trinity College of Engineering and Research, Pune, India.
Aboli Mohite, Purva Mene, Prathamesh Abnave, Susmita Mali
Trinity College of Engineering and Research, Pune, India.

Abstract – Our project discusses a proposed solution for tracking a person's emotions using a network of cameras and adjusting lighting, music, and ambiance accordingly to improve their mood. It also mentions the use of computer vision and machine learning for facial expression recognition. Another aspect involves creating a mood-based music player that recommends songs based on the user's mood in real-time, enhancing customer satisfaction. The project highlights the growing importance of music in people's lives and the need for improved music emotion recognition methods using an artificial bee colony algorithm to enhance the structure of neural networks and achieve better recognition results and faster processing.

I. INTRODUCTION

In this stressful and struggling world of the 21st century mental health has become a major entity to be dealt with thus people are trying numerous tactics that will help them to analyse and calm their emotions. Like wise to deal with this major issue we have come up with an idea of music player based on emotions. Compared to its predecessors, the main advantage of CNN is that it automatically detects the important features without any human supervision. This is why CNN would be an ideal solution to computer vision and image classification problems.

Because their convolutional layers have fewer parameters compared with the fully connected layers of a traditional neural network, CNNs perform more efficiently on image processing tasks. CNNs use a technique known as parameter sharing that makes them much more efficient at handling image data.

The system's operation starts with a webcam connected to a microprocessor or device with sufficient
computational capabilities. This webcam captures images of users' faces, and these images are then processed through a trained model.[2]. Deep learning techniques are used to interpret these images, which can be challenging due to the high degree of similarity in facial features. This similarity results in significant covariance, making it difficult to distinguish different emotions accurately. Moreover, human accuracy in identifying the mood of another person typically ranges from 67% to 73%.[2]. However, despite these
challenges, the system is designed to effectively identify facial moods and provide accurate outputs.

II. RESEARCH SCOPE

Many music streaming platforms heavily depend on metadata such as song titles, genres, albums, and lyrics to facilitate music search functionalities. Nevertheless, some platforms are now integrating content-based methods, which analysis elements like melody, rhythm, and harmony to provide more advanced search options. Among these methods,
emotion-based approaches stand out as a
particularly intriguing avenue for the development of semantic music search engines, as they leverage
the deep emotional impact of music. Despite significant research efforts into emotion models and representations, only a few have managed to attain the requisite flexibility and robustness needed for effective emotion-oriented music applications.

LITERATURE SURVEY

This paragraph highlights the significance of music in human history and its integration into various aspects of daily life. [1] It emphasizes the impact of music on people's emotions and how it's used in different contexts, such as retail stores, exercise, relaxation, and medical care. [1] It discusses the evolution of music storage and retrieval, noting that traditional methods are no longer sufficient for personalized music retrieval,
leading to the need for emotion based music classification and retrieval. The paragraph also touches on the challenges of labelling music with emotions and the development of automatic music emotion recognition technology. [1] Various approaches, including machine learning algorithms, have been used to achieve music emotion recognition with recognition rates exceeding 60%. [1].

This passage discusses the significance of deep learning based Music Emotion Recognition and Classification (MER) as an active research field in Music Information Retrieval (MIR).[2] It highlights the challenges in manually retrieving and classifying music in a vast digital library and the importance of automatic retrieval systems, especially in a diverse and multicultural context. [2].

The passage emphasizes the role of music emotions in relation to nationality, age, mood, purpose, and cultural background, and how these factors influence the complexity of MER. Music's connection to human emotions and its therapeutic applications are also highlighted, along with the use of neurotransmitters and hormones to describe the emotional impact of music. [2].

The paragraph underscores the integral role of music in people's lives, both as a source of enjoyment and for medical purposes, such as stress relief. [3] It mentions the advancement of high-end music players with enhanced features, including volume control,
modulation, and genre selection. The goal is to create a platform that plays music based on a person's current mood using facial expression recognition in the context of artificial intelligence (AI) and machine learning (ML).[3] The process involves dimensionality
reduction for facial expression recognition, starting with face detection and then extracting facial features like the eyes, nose, and mouth to identify emotions. [3] The paragraph also highlights the impact of music on human emotions and cognitive functions, emphasizing its role in enhancing mindfulness and mood. [3] The connection between musical preferences and individual characteristics and the growing use of digital emotion detection in multimedia content like music and movies is noted. The proposed recommender system is described, which can identify user emotions and suggest appropriate songs based on their mood. It uses the Kaggle Look Acknowledgment dataset for emotion detection and Bollywood Hindi songs for the music library. [3] Facial emotion detection
is implemented using a Convolutional Neural Network with high accuracy, approximately 95.14%. This system aims to improve users' emotional states through music recommendations based on their detected feelings. [3].

This passage discusses the growing
trend of sharing life experiences and opinions through images and videos on social networks,
which has led to an explosion of online social data. [4] Analysing these multimedia data, particularly at the emotional level, has gained significant importance, as emotions can influence decision making and have wide ranging applications from marketing to political forecasting. [4] While there has been progress in textual sentiment analysis, emotion analysis of social images poses several challenges. One major challenge is the "affective gap" where finding features to express emotions in images remains a key issue. Previous efforts have typically focused on predicting dominant emotions that are common among viewers, without considering individual and subjective evaluations.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 5, April : 2024

PROPOSED ARCHITECTURE

The face expressions play important role in detecting face emotions, they depend upon different physical and mental situations. The emotion recognition system is trained using supervised learning approach which undergoes training and testing of dataset. The general process of facial emotion recognition system includes: Face Detection, Image Preprocessing, Feature Extraction, and Classification.

Fig.1. Proposed Architecture

Proposed Methodology: The system you've described is focused on using facial expression recognition to create music playlists that match the user's emotional state. The main steps and components of the system are as follows:

1) Real-Time Capture: The system captures the users face accurately in real time. This is the initial step to get input for emotion recognition. In this we are using the camera to capture the users face.
2) Facial Recognition: The captured user's face is then used as input for the facial recognition module. This step involves using a convolutional neural network (CNN) that is trained to analyse the user's facial features. Convolutional Neural Networks (CNNs) are designed to automatically and adaptively learn spatial hierarchies of features from input images through convolutional layers. This enables them to capture intricate patterns and representations, making them highly effective for tasks like image classification, object detection, and segmentation.
3) Emotion Detection: In this phase, the features extracted from the user's facial image are used to identify emotions. The system then provides music recommendations based on the detected emotions. Feature extraction most important part of classification and emotion recognition. After the preprocessing of an image the facial features with high expression intensity are extracted like eyebrows, forehead wrinkles, nose, jawline, mouth corner. The facial feature extraction method id carried out by Local Binary pattern algorithm. The local binary pattern technique operates by pointing the pixels of an image and comparing it with the neighbourhood pixels by using binary number.
4) Music Recommendation: This module recommends songs based on the user's emotions and the desired music ambiance. The module plays the songs on the basis of emotions like happy, sad, Angry, etc.

CNN Algorithm

A Convolutional Neural Network (CNN) is a specialized deep learning model tailored for tasks involving image recognition and processing. Its architecture comprises several layers, including convolutional layers, pooling layers, and fully connected layers.[3]. At the heart of a CNN lie the

convolutional layers, where filters are applied to input images to extract essential features like edges, textures, and shapes. These convolutional layers help the network discern intricate patterns within the data. Following the convolutional layers, the feature maps undergo down sampling through pooling layers. These layers reduce the spatial dimensions while preserving crucial information, thus enhancing computational efficiency and preventing overfitting. [2]. Lastly, the output from the pooling layers is fed into one or more fully connected layers, which play a pivotal role in making predictions or classifying the image based on the extracted features. Inspired by findings in neuroscience, CNNs consist of layers of artificial neurons, or nodes. These nodes compute the weighted sum of inputs and produce activation maps,
highlighting significant features within the dataset. During the processing of data within a CNN, each layer returns activation maps, pinpointing important features. For instance, when presented with an image, the CNN identifies features based on pixel values, such as colours, and generates function.

Typically, in image analysis, a CNN initially detects fundamental features like edges. Subsequently, these features are passed to subsequent layers, which progressively identify more complex patterns such as corners and colour groups.[3]. This iterative process continues until a final prediction is made, leveraging the hierarchical structure of the network to discern increasingly intricate details within the image.

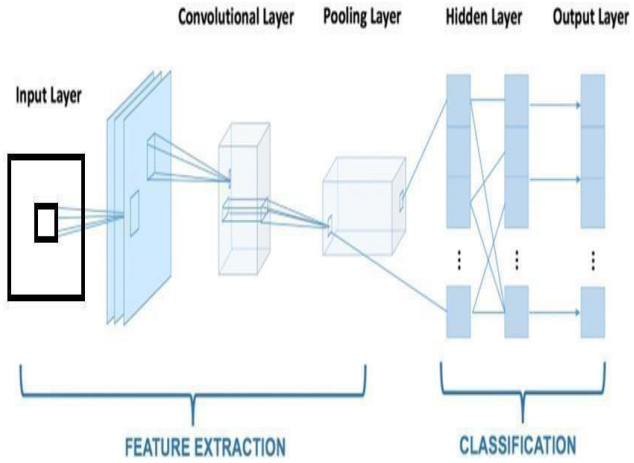


Fig.2. CNN Architecture

III. RESULT ANALYSIS

Input Layer:

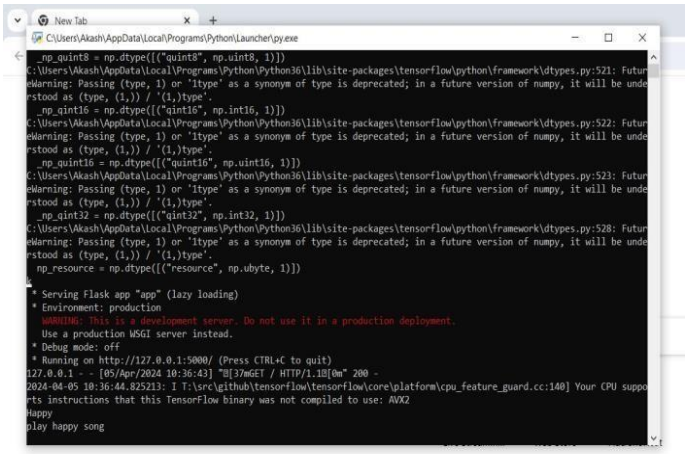
The input layer receives the raw image data.

Images are typically represented as grids of pixels with three colour channels (red, green, and blue – RGB). The dimensions of the input layer match the dimensions of the input images (e.g., 28x28x1 for a 28×28-pixel image with RGB channels).

Convolutional Layers (Convolutional and Activation):

Convolutional layers consist of multiple filters
(also called kernels). Each filter scans over the Fig.1. Output input image using a sliding window. Convolution operation calculates the dot product between the filter and the region of the input. Activation functions (e.g., ReLU – Rectified Linear Unit) introduce non-linearity to the network. Multiple convolutional layers are used to learn hierarchical features. Optional: Max Pooling layers reduce the spatial dimensions (width and height) to reduce computational complexity.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 5, April : 2024



Pooling Layers:
Pooling layers (e.g., Max Pooling or Average Pooling) reduce the spatial dimensions of feature maps while retaining important information.
Pooling helps to make the network more robust to variations in the position or size of objects in the input.
Flatten Layer:
A flatten layer reshapes the output of the previous layers into a 1D vector, allowing it to be input to a dense layer.
Fully Connected Layers:
After several convolutional and pooling layers, CNNs typically have one or more fully connected layers (also called dense layers). Fully connected
layers combine high-level features learned from previous layers and make final predictions. In classification tasks, these layers output class probabilities .

Fig.3. Neutral Face

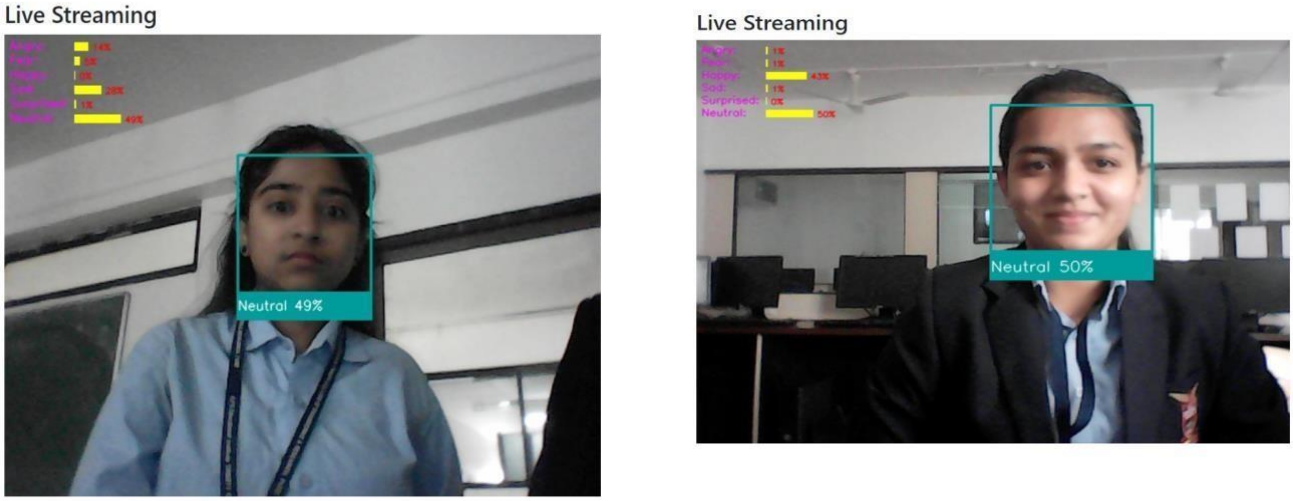


Fig.2. Happy Mood

IV. CONCLUSION

Emophony is a groundbreaking system that seamlessly integrates with numerous music platforms, making it effortless for users to discover music tailored to their current emotional state. Users receive a curated list of songs based on their emotions, eliminating the need for manual track

searches. Emophony efficiently directs users to gaana.com, where they can instantly play their chosen music with a simple click. The system utilizes affective computing methods to accurately predict and cater to users' changing emotional states. It offers a diverse selection of music, enriching the listening experience by aligning songs with users' emotions. Moreover, the paper introduce Music Information Retrieval (MIR) network designed to classify music based on emotional characteristics.

V. REFERENCES

1] Huang, C.; Zhang, Q. Research on Music Emotion Recognition Model of Deep Learning Based on Musical Stage Effect. Sci. Program. 2021.
2] Shikhar C. Maheshwari, Amit H. Choksi and Kaiwalya J. Patil. Emotion Based Ambiance And Music Regulation using Deep Learning, 2020.
3] Mohak Kedia, Vaibhav Bhagat, Yash Khannade, Nayan Laturiya. Emotion Based Ambiance And Music Regulation Dept. Of Computer Science & Engineering, 2023.
4] Sicheng Zhao, Hongxun Yao, Member, IEEE, Yue Gao, Senior Member, IEEE, Guiguang Ding, and Tat-Seng Chua. Predicting Personalized Image Emotion Perception In Social Network. 2018.
5] Han, X.; Chen, F.; Ban, J. Music Emotion Recognition Based on a Neural Network. 2023.