IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Emotion Based Music Player

1Rushikesh Bawaskar, 2Soham Survase, 3Preetish Agarwal, 4Prateek Koul

5Professor S.P Mankar

1234Students, Department of Information Technology, MMCOOE, Maharashtra, India.

5 Professor, Department of Information Technology, MMCOE, India.

ABSTRACT

The activities of the human brain are impacted by listening to music. An emotion-based music player equipped with an automated playlist can assist users in maintaining a particular emotional state. This study puts forward an emotion-based music player that generates playlists based on photographs taken of the user. Sorting a playlist manually and annotating songs based on the current emotion can be time-consuming and tedious. Several algorithms have been developed to automate this process, but they are slow, increase the system's cost by requiring additional hardware, and have low accuracy. This paper proposes an algorithm that not only automates the audio playlist generation process but also classifies newly added songs. The primary task is to determine the current mood of the individual and play music accordingly making the system more efficient, fast, and automatic. The ultimate objective is to reduce the overall computational time and cost of the system while improving its accuracy. The primary aim is to improve an individual's mood if it is negative, such as feeling sad or depressed. This model is validated by evaluating the system against both user-dependent and user-independent datasets.

Keywords— Convolution neural network, Long Short term memory, Emotion detection, audio classification, hidden layers, Max-pooling.

INTRODUCTION:

Recognizing and expressing emotions in human communication systems is crucial. Humans possess the capacity to recognize and express emotions, while computers attempt to detect human emotions via image analysis or sensors. In our personal and professional lives, we interact with numerous people face-to-face or indirectly via phone calls, and it is sometimes necessary to be aware of the emotions of the person we are interacting with. Facial expressions and speech intonation play a significant role in conveying emotions. The physical characteristics and tone of the face can convey the energy in speech, which can be altered to express different emotions. Humans can effortlessly perceive these changes in signals, along with the information gathered from other sensory organs. This study examines the use of images, sensors, or speech to capture emotions.

Music is a crucial aspect of improving one's life, serving as a significant form of entertainment for music enthusiasts and listeners, and occasionally providing a therapeutic approach. "Where words fail, music speaks," and it has the ability to gradually transform a person's negative emotions into a positive mood.

An individual's emotions can be conveyed through various means, such as gestures, speech, facial expressions, and body language. To comprehend the user's mood, facial expressions are employed in our system. By utilizing the camera on a mobile device, we can capture the user's facial expressions. Numerous emotion recognition systems use the captured image as an input to identify the emotion. Neural networks are utilized in this application for emotion recognition.

IMPLEMENTATION:

The proposed algorithm centers around an automated music recommendation system that selects songs based on a person's mood or current emotions. Whenever the application is opened, the user's photo is captured, allowing for the detection of their current emotion. Based on the information obtained from the image, a song related to the emotion is played.

The music stored in the user's phone has already been sorted into seven distinct categories: happy, sad, angry, surprised, fearful, disgusted, and neutral. Newly added songs are dynamically classified into the appropriate mood. The algorithm comprises three modules: a Facial Expression Recognition module, a Song Emotion Recognition module, and a System Integration module. These modules are mutually exclusive, with the Facial Expression Recognition and Audio Emotion Recognition modules working independently of one another. The Facial Expression Recognition module identifies the user's emotion by capturing their facial expressions through the mobile device's camera, while the Audio Emotion Recognition module determines the emotional content of the music by analyzing its acoustic characteristics. Once the emotions have been recognized, the System Integration module plays a song that corresponds to the detected emotion. Hence, system integration module maps two modules to find the correct match of detected emotion.

Fisher Face Algorithm:

This image processing system employs PCA to reduce face space dimensions and then applies FDL (or LDA) to extract image features. We use this approach to maximize class separation during training, aiding in image recognition. For facial matching, we utilize the minimum Euclidean algorithm. It helps classify the user's expression, indicating their emotion.

Haar Cascade Algorithm:

This is a machine learning algorithm used for object detection in captured images, specifically designed as a cascade classifier. The classifier consists of multiple stages, each resembling a weak learner, which are simple classifiers known as boosting. When the label is positive, it progresses to the next stage, yielding the final result. The cascade classifier operates by identifying images based on labels, utilizing a collection of positive and negative images at various stages. Higher resolution images tend to produce better quality results due to their larger quantity. In our implementation, we utilize the haar cascade frontal face_default.xml to detect specific objects in the image, such as the nose, eyes, ears, and lips on a face. The Haar Cascade, developed by OpenCV, can also detect features from the source image. It works by training negative images onto positive images, superimposing them. Positive images only contain the object we want our classifier to categorize, while negative images encompass everything else, we do not wish to detect.

Data Set:

To create a comprehensive dataset for emotion recognition, we obtained the Raw dataset by downloading images for seven distinct emotions from Google Images individually. In addition, we also gathered an Extra dataset from the Kaggle Datasets specifically for facial expression detection. The goal of obtaining these datasets is to train the algorithm to recognize different emotions accurately. Through extensive data collection and analysis, we can develop a more robust and reliable emotion recognition system that can be used in various applications, including music recommendation systems, mental health monitoring, and communication systems. By using multiple sources of data, we can increase the diversity and breadth of the dataset, making the algorithm more effective and accurate.

Trained Dataset:

To ensure the effectiveness and accuracy of the emotion recognition model, it undergoes a thorough training and testing phase. During the training phase, the model is taught using a set of pre-classified dataset that enables the model to learn and understand the characteristics of different emotions. The training dataset consists of images of faces with their corresponding expressions, and the model learns a set of weights that allow for the classification of facial expressions accurately. It is essential to ensure that the eyes are located in the center of the image, which aids in the recognition of facial expressions. This training phase is critical as it determines the model's ability to recognize emotions accurately and respond appropriately in real-world scenarios. The testing phase is then carried out to evaluate the model's performance and accuracy, ensuring that it can recognize and classify emotions accurately under various conditions. By undergoing these rigorous training and testing phases, we can develop a reliable and effective emotion recognition system that can be used in a wide range of applications.

For training, the sequence is:

1. Spatial normalization
2. Synthetic samples generation
3. Image cropping
4. Down-sampling
5. Intensity normalization.

Test Data:

At the time of testing, classifier takes images of face with respective eye center locations, and it gives output as predicted expression by using the weights learned during

For recognizing an unknown image (testing), the sequence is:

1. Spatial normalization

2. Image cropping

3.Down-sampling

4.Intensity normalization

Convolution Neural Network :

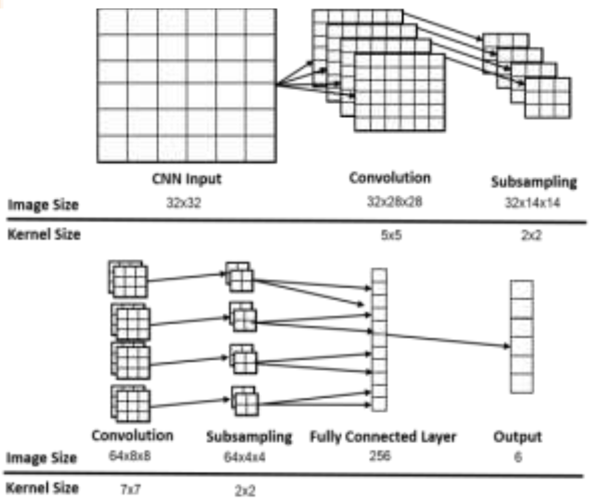

Figure 4: Architecture of proposed Convolution Neutral Network. It has five layers: first layer is convolution, second layer is sub-sampling, the third layer is convolution, fourth layer is sub-sampling, fifth layer is fully connected layer and final responsible for classifying facial image.

FINAL RESULTANT MODE:

The dataset containing photos is initially transformed into grayscale, to optimize and expedite the preprocessing and detection process. Each image is represented as a matrix of pixels, such as 48x48. The matrix of pixels is fed to the convolutional layers, which are the hidden layers of the model. During the process, maximum pooling is performed between each layer, which down-samples the input data or image, reducing its dimensions and enabling assumptions to be made about the features contained in subregions. This is essential to prevent over-fitting and reduce computational costs by reducing the number of parameters to learn. For instance, if the input image is represented as a 4x4 matrix, and the output required is 2x2, then pooling is carried out between all hidden layers. The processed data is then fed into a dense layer to avoid over-fitting by using the dropout technique. The output layer reveals the detected class. Assuming the detected expression is happy, the next phase is to choose one of the training datasets for the music model. The dataset is then trained according to the matching expression for playing music. For song classification, an LSTM neural network is utilized. One-hot encoding is performed to convert categorical variables into binary vectors, which accelerates and improves classification. Finally, the song is played based on the person's current mood.

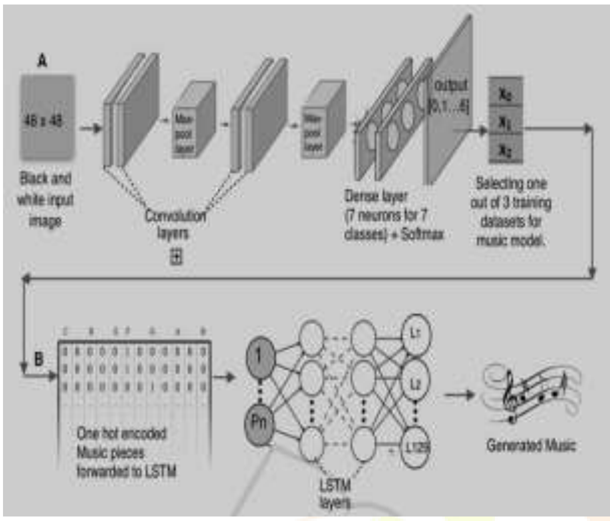

Figure: Final Model (Resultant) RESULT ANALYSIS

CONCLUSION:

The Emotion Based Music Player is an innovative solution that aims to enhance the music listening experience of users. The application provides a comprehensive set of features that meet the basic needs of music enthusiasts without causing any inconvenience. By leveraging advanced technologies, it enables greater interaction between the system and the user, making it a more personalized experience. One of its key features is its ability to capture the user's image using the phone's camera and detect their current emotion. Based on this analysis, the application suggests a customized playlist that matches the user's mood. Additionally, the application offers a unique feature that gradually transforms negative or bad thoughts into positive ones by adjusting the playlist to include songs with a more upbeat tone.

FUTURE SCOPE:

The Emotion Based Music Player has the potential for future enhancements, such as integration with Google Play Music to enable playing songs not available in local storage and speech-based access to the entire application. This system will be very beneficial for users searching for music that aligns with their mood and emotional state. It will reduce the time spent searching for music, which in turn will increase the overall accuracy and efficiency of the system. Moreover, the system will not only reduce physical stress but also be useful in music therapy systems, aiding music therapists in treating patients. In the future, it could be used to detect drowsiness in drivers and have other applications. With its advanced features, it will be a comprehensive system for music lovers and listeners.

REFERENCES:

[1] A. Savva, V. Stylianou, K. Kyriacou, and F. Domenach, “Recog nizing student facial expressions: A web application,” in 2018 IEEE Global Engineering Education Conference (EDUCON), Tenerife, 2018, p. 1459-1462.

[2] Bahdanau, D.; Cho, K.; and Bengio, Y. 2014. Neural machine translation by jointly learning to align and translate. In ICLR 2015.

using adjective noun pairs. In Proceedings of the 21st ACM international conference on Multimedia, 223–232. ACM.

[3] Borth, D.; Chen, T.; Ji, R.; and Chang, S.-F. Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content.

[4] Borth, D.; Ji, R.; Chen, T.; Breuel, T.; and Chang, S.-F. 2013. Large-scale visual sentiment ontology and detectors

[5] C. Tang, P. Xu, Z. Luo, G. Zhao, and T. Zou, “Automatic Facial Expression Analysis of Students in Teaching Environments,” in Biometric Recognition, vol. 9428, J. Yang, J. Yang, Z. Sun, S. Shan, W. Zheng, et J. Feng, Éd. Cham: Springer International Publishing, 2015, p. 439-447.

[6] Campos, V.; Jou, B.; and Giro-i Nieto, X. 2016. From pixels to sentiment: Fine- tuning cnns for visual sentiment prediction. arXiv preprint arXiv:1604.03489. 1. Cao, D.; Ji, R.; Lin, D.; and Li, S. 2014. A cross-media public sentiment analysis system for microblog. Multimedia Systems 1– 8.

[7] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, “The Faces of Engagement: Automatic Recognition of Student Engagementfrom Facial Expressions,” IEEE Transactions on

[8] Affective Computing, vol. 5, no 1, p. 86-98, janv. 2014. P. Ekman and W. V. Friesen, “Constants across cultures in the Face and emotionJournal of Personality and Psychology. Vol. 17, NO. 02