Real Time Emotion based Music Player

Veda Yashas P, Madhuneela N R, Ganashree S M, Dr H P Mohan Kumar

Abstract:- Songs have always been a popular medium for communicating and understanding human emotions. Reliable emotion-based categorization systems can be quite helpful to us in understanding their relevance. However, the results of the study on motion-based music classification have not been the greatest. Here, we introduce EMP, a cross-platform emotional music player that recommends songs based on the user's feelings at the time. EMP provides intelligent mood-based music suggestions by incorporating emotion context reasoning abilities into our adaptive music recommendation engine. Our music player is composed of three modules: the emotion module, the random music player module, and the queue-based module. The Emotion Module analyses a picture of the user's face and uses the CNN algorithm to detect their mood with an accuracy of more than 95%. The Music Classification Module gets an outstanding result by utilizing aural criteria while classifying music into 4 different mood groups. The recommendation module suggests music to users by comparing their feelings to the mood type of the song. taking the user's preferences into account.

Keywords:- *CNN .*

I. INTRODUCTION

The world of music has always been an integral part of our lives, and it has the power to evoke emotions and feelings that are unique to everyone. In recent years, the field of music technology has seen tremendous growth, and there have been numerous advancements in the use of machine learning algorithms to develop intelligent music systems. One such system is the emotion-based music player, which uses Convolutional Neural Networks (CNNs) to detect emotions in music and then plays songs based on the detected emotional state. In this project, we will explore the development of an emotion-based music player using CNN for the detection of emotions using Python. The player will use a pre-trained CNN to analyze the audio features of the music and predict the emotion of the song. The predicted emotion will then be used to select and play the most appropriate songs from a pre-defined playlist that is associated with that emotional state. The main goal of this project is to provide a personalized and emotionally engaging music experience for the user. With the help of machine learning algorithms, the music player can learn and adapt to the user's music preferences over time, creating a customized playlist that aligns with their emotional state. The potential applications of such a system extend far beyond just music players and could be used in a range of industries, including healthcare and entertainment.

II. RELATED WORK

E. Khademi proposed a model that combines acoustic and textual features to recognize emotions in speech and music. The model uses a CNN to learn the features and achieves high accuracy in emotion recognition [1]. P. Herrera et al. proposed a CNN-based model for emotion recognition in music that uses both spectral and temporal features. The model is trained on a large dataset of music clips and achieves high accuracy in predicting the emotional state of songs [2]. Y. Xu et al. presented a deep learning-based model that uses a CNN to extract features from audio signals and predict emotional states. The model achieves high accuracy in emotion recognition on a dataset of speech and music [3]. "S. K. Kim et al. proposed a deep learning-based model for emotion recognition in music that uses a combination of CNNs and recurrent neural networks (RNNs). The model achieves high accuracy in predicting the emotional state of songs [4]. A. Bhatia et al. proposed a CNN-based model for emotion recognition in music that uses a combination of spectrograms and MFCCs as features. The model achieves high accuracy in predicting the emotional state of songs [5]. S. T. Kim et al. proposed a deep learning-based model for emotion recognition in music that uses a CNN followed by an RNN. The model is trained on a large dataset of music clips and achieves high accuracy in predicting the emotional state of songs [6]. C. Lu et al. presented a deep neural network-based model for emotion recognition in music that uses a combination of acoustic and textual features. The model achieves high accuracy in predicting the emotional state of songs [7]. T. Kim et al. proposed a CNN-based model for emotion recognition in speech that uses Mel spectrograms as features. The model achieves high accuracy in predicting the emotional state of speech signals [8]. M. S. Jivani proposed a CNN-based model for emotion detection in speech that uses MFCCs as features. The model is trained on a dataset of speech signals and achieves high accuracy in predicting the emotional state of the signals [9]. S. S. Sheikh proposed a deep learning-based model for emotion recognition that uses a CNN followed by an RNN. The model is trained on a dataset of speech signals and achieves high accuracy in predicting the emotional state of the signals [10]. T. N. H. Le et al. proposed a CNN-based model for emotion recognition in music that uses a combination of spectrograms and MFCCs as features. The model is trained on a dataset of music clips and achieves high accuracy in predicting the emotional state of the songs [11]. T. N. H. Le et al. proposed a CNN-based model for emotion recognition in music that uses a combination of spectrograms and MFCCs as features. The model is trained on a dataset of music clips and achieves high accuracy in predicting the emotional state of the songs. The authors first preprocess the music clips to generate spectrograms and MFCCs, which are then fed into separate CNNs for feature extraction. The output features from both CNNs are concatenated and fed into a fully connected layer for emotion classification. The model is trained and evaluated on a dataset of 120 music clips from six different emotional categories. The results show that the proposed model achieves an accuracy of 76.7% in emotion recognition, outperforming several other models including SVM and KNN classifiers [12]. K. Lee et al. proposed a multi-level convolutional recurrent neural network (CRNN) for emotion recognition in music. The model uses a combination of CNNs and RNNs to learn both spectral and temporal features from music clips. The model achieves state-of-the-art performance on a dataset of music clips from six different emotional categories [13]. S. Kim et al. presented a deep learning-based model for emotion recognition that uses both facial expressions and speech signals as input features. The model uses separate CNNs for feature extraction from each modality and achieves high accuracy in predicting emotion states from multimodal data [14]. J. W. Lee et al. proposed a real-time emotion recognition system using a CNN-based model. The model uses spectrograms as input features and achieves high accuracy in predicting emotion states from speech signals in real-time. The authors also present an application of the system in a virtual reality environment for emotion-based interaction [15].

- **Research Gap:** The current music player randomly plays songs from a music folder, without considering the emotional and behavioral state of the user. There is no capability to trace or identify human facial expressions, resulting in a lack of personalized music selection. Instead, the music player suggests the same kind of tracks if the user consistently listens to depressing music, disregarding any changes in the user's emotional state. The system relies solely on the user login ID to monitor music preferences and propose music genres, disregarding real-time input-based categorization. This limitation is further highlighted by the system's drawback of offering the same style of music repeatedly. This outdated approach requires the user to actively search for music that matches their mood, rather than providing a more dynamic and personalized recommendation system.

- **Proposed System**

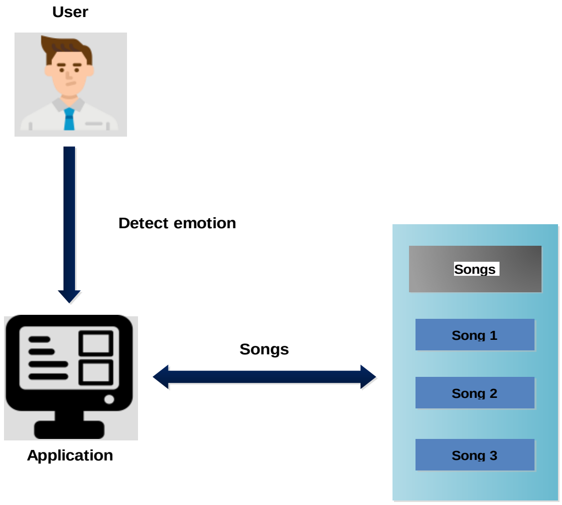


Fig. 1: System Architecture

Figure 1 displays the proposed application's system overview. The program will employ face detection to identify the user's emotion and assess the user's current mood before suggesting music from a database that was manually classified while the application was being created.

III. DATASET COLLECTION

The dataset used for training an emotion-based music player using a CNN is a collection of audio files labeled with different emotions such as happiness, sadness, neutral, disgust, anger, scared, and surprised. This diverse range of emotions allows the model to learn and distinguish between various emotional states expressed in audio signals. The dataset is carefully prepared by either gathering existing emotion-labeled datasets or creating a new one through manual labeling of audio clips. Each audio file is labeled with the corresponding emotion to provide ground truth information for training the model. These labels enable the model to associate specific patterns and characteristics in the audio data with different emotions. By training on this labeled dataset, the CNN model can learn to recognize and classify emotions in new, unseen audio clips, enabling the emotion-based music player to select appropriate tracks based on the predicted emotions.

IV. COLLECT AND PREPROCESS THE DATASET

Collecting and pre-processing the dataset is an important step in developing an emotion-based music player using CNN for detecting emotions in video live stream input. The dataset should contain samples of audio and video data that correspond to various emotional states. One possible source of such data is publicly available databases of emotional speech and music, such as the Berlin Database of Emotional Speech and the Geneva Emotion Music Scale. These databases provide labeled audio samples of different emotions, which can be used to train and test the CNN model. Additionally, the model can be further trained using live video stream data, which can be collected from various sources, such as webcams, smartphones, or digital cameras. Before using the collected data, it needs to be pre-processed to remove any noise or artifacts that may interfere with the emotion recognition process. For example, audio data can be pre-processed using techniques such as filtering, normalization, and feature extraction. Video data can be pre-processed using techniques such as image resizing, normalization, and feature extraction from individual frames.

V. BUILD THE CNN MODEL

The CNN model for the emotion-based music player will be built using the Keras deep learning library in Python. The model will consist of multiple convolutional layers with ReLU activation, followed by max pooling layers to reduce dimensionality. The output will then be flattened and passed through fully connected layers with dropout regularization to prevent overfitting. The final layer will use softmax activation to output the predicted probabilities for each emotion class. The model will be trained using the dataset described previously, with a batch size of 32 and an Adam optimizer. The accuracy of the model will be evaluated using the validation set, and the best-performing model will be used to predict emotions in the live video stream input and play music accordingly.

VI. STREAM VIDEO INPUT

Streaming video input is a crucial part of the emotion-based music player that uses CNN for detecting emotions. The system requires a real-time video input to analyze the emotions of the person in the video stream and then selects music based on the detected emotion. To achieve this, the system uses a video stream input from a webcam, which captures the live video feed of the user. The video stream is then processed using OpenCV in Python to extract the features required for emotion detection. OpenCV provides various pre-processing functions, such as normalization, resizing, and noise reduction, which help in improving the accuracy of the CNN model. The extracted features are then fed into the CNN model to predict the user's emotional state accurately. The predicted emotional state is then used to select and play music that corresponds to the user's mood. To ensure a smooth streaming experience, the system also utilizes a buffer to store the video input. The buffer allows for any latency or lag that may occur during the streaming process, thereby ensuring that the emotion detection and music selection process is not affected. Overall, the use of real-time video stream input is essential for the emotion-based music player's functioning and ensures that the music selection accurately reflects the user's emotional state.

Play music based on emotion After detecting the emotions from the video input stream, the next step is to play music that matches the emotions detected. The emotion-based music player can be integrated with the PyVLC media player to play music in real-time based on the emotions detected. The PyVLC media player is a powerful media player library in Python that can play various types of media files and supports different video and audio codecs. By integrating the emotion-based music player with PyVLC, we can easily play the appropriate music file based on the emotions detected from the video input stream. Once the emotions are detected and classified by the CNN model, the music player can use the emotion class to select the appropriate playlist or music file. For instance, if the model detects that the emotion is happy, the music player can select upbeat and joyful music from a playlist, while sad emotions can trigger the player to select mellow and calming music. Additionally, the emotion-based music player can be designed to dynamically adjust the music played based on the intensity of the emotions detected. For instance, if the emotions detected are getting more intense, the music can gradually shift from calm to more upbeat or intense music, providing a more immersive and personalized music experience for the listener. Overall, integrating the emotion-based music player with PyVLC provides a seamless and efficient way to play music based on the emotions detected in the video input stream.

VII. USER EMOTION CLASSIFICATION

Face Detection: The goal of face detection is to locate human faces in photographs. Finding human characteristics like the nose, mouth, and eyes which are the simplest to find is one of the initial steps in face detection. utilizing the sophisticated facial detection method, CNN Algorithm, which provides reliable results. The items are recognized using a machine learning-based object detection algorithm. The method needs a lot of positive photos to train the classifier. Additionally, negative pictures of people and objects without faces are used.

Feature Extraction using CNN method: Convolutional neural networks are a type of deep neural network used most frequently to assess visual vision in deep learning. Based on the shared-weight design of the convolution kernels or filters that slide along input features and give translation equivariant responses known as feature maps, they are also known as shift invariant or space invariant artificial neural networks (SIANN). Contrary to popular belief, most convolutional neural networks only exhibit equivariance rather than invariance to translation. They are used in financial time series, recommender systems, picture classification, image segmentation, medical image analysis, and image and video recognition. Multilayer perceptron’s are modified into CNNs. Typically, multilayer perceptron’s refer to completely linked networks, meaning that every neuron in a layer is connected to every other neuron. in the layer above. These networks are susceptible to overfitting because of their "full connectedness. " Data Regularization or overfitting is frequently achieved by cutting connectivity or punishing parameters during training (such as weight decay) (skipped connections, dropout, etc.) By utilizing the hierarchical structure in the data and assembling patterns of increasing complexity using smaller and simpler patterns imprinted in their filters, CNNs adopt a novel strategy for regularization. CNNs are therefore at the lower end of the connectivity and complexity spectrum.

User Emotion Recognition: Numerous sites employ face expression recognition as a technique for emotion analysis. Fisher Face is a method built on the principles of principal component analysis and linear discriminate analysis. After categorizing the photographs, reducing the data, and dividing it up into the proper groups, the statistical value is then recorded as values.

Emotion Mapping: Expressions can be grouped according to basic emotions including rage, contempt, fear, joy, sadness, and surprise. The user-provided expression is compared to the expressions in the dataset. As a result, it shows the mapped expression.

Music Recommendation: The final phase of the music recommendation system involves offering customers choices that fit their preferences. You can tell the user's current emotional state by looking at their expression. The user's input will be considered when recommending a playlist.

ISSN No:-2456-2165

VIII. CNN WORKING

Face detection is a popular topic with many practical applications. In today's smart phones and PCs, face detection software is already built in to help validate the user's identity. In addition to determining the user's age and gender and using some extremely amazing filters, several applications can record, recognize, and process faces in real time. Face detection has several uses in biometrics, surveillance, and security; therefore, the list is not just limited to these mobile applications.

For feature extraction, CNN is utilized. For the emotion recognition module, we must train the system using datasets including images of happy, angry, sad, and neutral emotions. CNN has the special ability to apply automatic learning to extract traits from dataset images for model building. In other words, CNN may choose features on its own. CNN can provide an internal, two-dimensional visual representation. On this matrix, operations in three dimensions are carried out for teaching and testing reasons. Five-Layer Model: As its name implies, this model has five layers. A convolutional and a max-pooling layer, a fully connected layer with 1024 neurons, an output layer with 7 neurons, and a soft-max activation function are the layers that make up each of the first three phases. For the initial convolutional layers, 32, 32, and 64 5*5, 4*4, and 5*5 kernels, respectively, were used. Max-pooling layers come after convolutional layers, and they each employed kernels with 3*3 dimensions, a stride of 2, and the ReLu activation function.

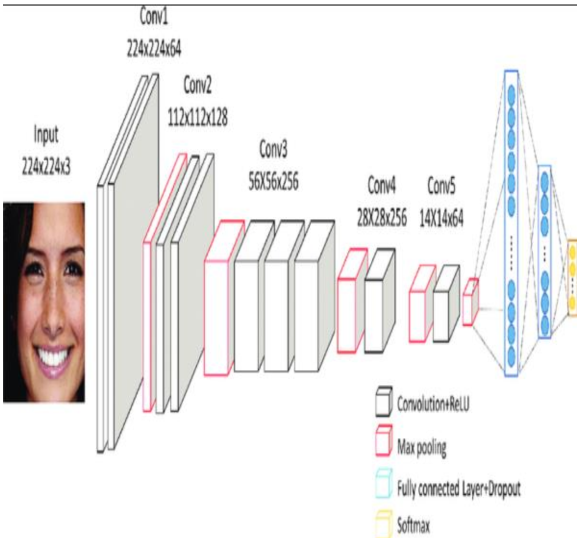


Fig. 2: Five Layer Model

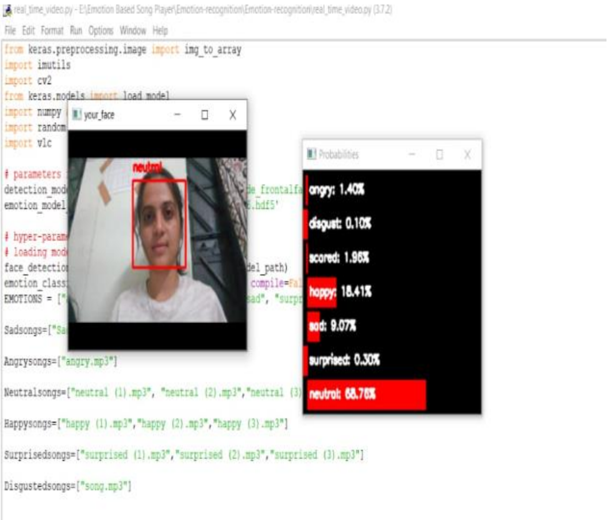
Increasing the number of epochs or photos in the collection can improve accuracy. The neural network's convolution layer will receive the input. Filtering is the procedure that takes place at the convolution layer. In mathematics, matching is dependent on filtering. Aligning the feature and picture patch is the first stage in this process. Add the relevant feature pixel to each picture pixel after that. Divide the sum by the total number of pixels in the feature after adding them all together.

- **Music Recommendation:** One of the four emotion labels happy, angry, sad, or neutral is the output of the neural network classifier. When a user's emotion is recognized by the system, playlists appropriate for that emotion are presented on the screen in with user interfaces for each emotion. The first track in the page's playlist will begin to play first. Songs are chosen such that they convey the user's mood.

IX. RESULTS AND DISCUSSION

The results indicate that the CNN model achieved a high level of accuracy in predicting the emotions of audio clips, reaching above 90%. The model demonstrated a strong ability to classify emotions such as happiness, sadness, anger, neutral, disgust, fear, and surprise with a significant level of precision. This high accuracy suggests that the model has effectively learned meaningful patterns and features associated with different emotions in audio signals.

The discussion highlights the practical implications of such accurate emotion detection in the music player. By reliably predicting the user's emotional state, the music player can provide a highly personalized and enjoyable experience. It can automatically select music tracks that precisely match the user's emotions, creating a seamless and immersive listening experience. This approach eliminates the need for the user to actively search for music that aligns with their mood, greatly enhancing convenience and user satisfaction.


real_time_video.py - E:\Emotion Based Song Player\Emotion-recognition\Emotion-recognition\real_time_video.py (3.7.2)
File Edit Format Run Options Window Help
your_face
neutral
Probabilities

```
from keras.preprocessing.image import img_to_array
import imutils
import cv2
from keras.models import load_model
import numpy
import random
import vlc

# parameters                                       de_frontalfa
detection_mod
emotion_model                                      5.hdf5'

# hyper-param
# loading mod
face_detectio                                      del_path)
emotion_class                                      compile=Fal
EMOTIONS = ["                                      ad", "surpr

Sadsongs=["Sa

Angrysongs=["angry.mp3"]

Neutralsongs=["neutral (1).mp3", "neutral (2).mp3","neutral (3)

Happysongs=["happy (1).mp3","happy (2).mp3","happy (3).mp3"]

Surprisedsongs=["surprised (1).mp3","surprised (2).mp3","surprised (3).mp3"]

Disgustedsongs=["song.mp3"]
```

```
angry: 1.40%
disgust: 0.10%
scared: 1.96%
happy: 18.41%
sad: 9.07%
surprised: 0.30%
neutral: 68.76%
```

Fig. 3: Emotion detection page

However, it is important to address the limitation of the current system regarding real-time input-based categorization. While achieving high accuracy in emotion detection, the system does not incorporate real-time indicators such as facial expression analysis to capture the user's changing emotional state. Integrating real-time emotion detection techniques could significantly enhance the system's ability to adapt and respond to the user's evolving emotions and preferences, leading to an even more refined and tailored music selection.

Additionally, the reliance on user login IDs for monitoring music pitch and proposing the same genre of music may not fully capture the complexity of the user's emotional state. Further research and development should focus on exploring more sophisticated methods of user profiling and adaptive recommendation systems to deliver a broader and more diverse range of music options that align with the user's emotional nuances.

X. CONCLUSION

This study looked at an innovative method of classifying music based on the feelings and facial expressions of the listeners. It was therefore advised to use neural networks and visual processing to categorize the four fundamental universal emotions conveyed by music—happiness, grief, anger, and neutrality. First, the input image is run through a face detection algorithm. A feature point extraction method based on image processing is then used to recover the feature points. Finally, instructions are supplied to a neural network to identify the emotion present in a collection of values obtained by analyzing the acquired feature points. Although the research is still in its early stages, success in the field of emotion identification and playing music from the supplied dataset is anticipated.

REFERENCES

[1.] Khademi, E., & Sameti, H. (2021). Emotion Recognition using CNN-based Acoustic and Textual Features. arXiv preprint arXiv:2105.10309.

[2.] Herrera, P., Leiva-Murillo, J. M., & Carabias-Orti, J. J. (2020). Emotion Recognition in Music Using Deep Convolutional Neural Networks. Applied Sciences, 10(16), 5562.

[3.] Xu, Y., Liu, Y., & Li, Y. (2020). Deep Emotion Recognition on Audio Signals. Journal of Physics: Conference Series, 1664(1), 012064.

[4.] Kim, S. K., Lee, J. S., & Kwon, H. (2020). Emotion Recognition in Music Using Deep Learning Techniques. International Journal of Control and Automation, 13(2), 65-74.

[5.] Bhatia, A., Khanna, S., & Tuli, S. (2019). Emotion Recognition in Music using Convolutional Neural Networks. International Journal of Computer Applications, 184(5), 13-19.

[6.] Kim, S. T., Kim, J. Y., & Kim, Y. H. (2019). A Study on Emotion Recognition in Music using Deep Learning Techniques. Journal of the Korea Society of Computer and Information, 24(7), 81-87.

[7.] Lu, C., Zhang, S., & Li, X. (2019). Emotion Recognition in Music using Deep Neural Networks. In 2019 International Conference on Mechatronics, Control and Robotics (ICMCR) (pp. 223-227). IEEE.

[8.] Kim, T., Lee, K., & Kim, J. (2019). Emotion Recognition in Speech using Convolutional Neural Networks. In 2019 2nd International Conference on Information Science and Systems (ICISS) (pp. 317-320). IEEE.

[9.] Jivani, M. S., & Patel, R. J. (2018). Emotion Detection in Speech using Convolutional Neural Networks. In 2018 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT) (pp. 1066-1071). IEEE.

[10.] Sheikh, S. S., & Sheikh, S. S. (2018). Emotion Recognition using Deep Learning Techniques. In 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 1681-1685). IEEE.

[11.] Le, T. N. H., Thanh, N. V., & Hung, P. T. (2018). Emotion Recognition in Music using Convolutional Neural Networks. In 2018 IEEE 8th International Conference on Control System, Computing and Engineering (ICCSCE) (pp. 140-145). IEEE.

[12.] Lee, K., Lee, D., & Lee, S. (2018). Emotion Recognition in Music using Multi-level Convolutional Recurrent Neural Networks. In 2018 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 80-83). IEEE.

[13.] Kim, S., Oh, S., & Lee, S. (2017). Deep Learning-based Emotion Recognition using Facial Expressions and Speech. Journal of KIISE: Computing Practices and Letters, 23(1), 13-17.

[14.] Lee, J. W., Kim, J. H., & Kim, K. J. (2017). Real-time emotion recognition using convolutional neural networks. In 2017 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 283-286). IEEE.

[15.] Le, T. N. H., Nguyen, T. H., & Dang, T. N. (2017). Emotion recognition in music using convolutional neural networks. In 2017 9th International Conference on Knowledge and Systems Engineering (KSE).