

Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation

Manjunath K G¹, Shreedevi Pramod² and Chiranth TP³

¹Siddaganga Institute of Technology, Tumkur/Department of Computer Science and Engineering, Tumkuru, Karnataka, India
Email: kgmanjunath@sit.ac.in

²⁻³Brindavan College of Engineering, Bagaluru main Road, Bangalore, India
Email: shreedevipramod@gmail.com, chiranthtp@gmail.com

Abstract— We came across many applications that are serve solutions and assistance for the music playlist creation recommendation. To provide more facility and the solutions to the music creation and recommendation of music and numerous technique approaches have followed in our proposed system. To determine the approached and developing the current state of human emotion and playing through the mood determined by the system. Our proposed system mainly focuses on playing and recommending the songs according to end user current mood or betterment of the mood.

Index Terms— Emotion, PSNR, facial expression, Open CV.

I. INTRODUCTION

According to the emotion, the music will be played from the predefined directories. Each sub-directory contains songs that corresponds to the emotion. Songs in the sub folders can be changed/replaced or deleted by the programmer depending on the requirements of user. At times it is possible that user might like different kinds of songs in certain mood. For example, when a user's emotion is detected to be Sad, then it is totally users choice what kind of mood does he/she want.

A. In[4], they have proposed a mood detection system. They have used modern CNN for building framework. Their architecture is fully-convolutional neural network contains 4 residual separable convolutions and each convolution is followed by a batch normalization operation and ReLU activation. Architecture has 60000 parameter, which is corresponds to reduction of 10x of Naive and 80X of original CNN. They have achieved 66% of accuracy in mood detection.

A literature survey on the developments in biometric image processing domain and techniques used for detecting face and recognizing facial emotions are reviewed and discussed below:

S. Amirhassan Monadjemi et al; proposed to enhance the performance of face detection and recognition systems. In face detection step, Skin color segmentation with YCbCr color space is used. Gaussian skin color model is used in color space image to segment the skin and non-skin pixels. Morphological operations are used to eliminate holes formed in eye, mouth, and nose in skin color segments. Gabor features extraction, dimension reduction using PCA, feature selection using LDA and SVM based classification are used to construct an efficient face recognition system with a high recognition rate [5]. regression model. The analyses show how the available information in a common ICU setting can be used to predict the progress of septic shock also in a very limited number of cases.

Anala M R et al., presented Viola Jones classifier method, Background subtraction method and skin color detection

on HSV color spaces. The Viola Jones classifier method gives good results for detecting faces but it takes more time for the detection and it does not give accurate results. The Background subtraction method and Skin color detection on HSV color space are more efficient when compared to the first method by giving accurate results for detecting faces in motion [6].

This section presents the existing methods for classifying emotion from the music. There are some relevant approaches which are concluded as follows.

Robert E. Thayer applies rhythm, tempo, intensity, pitch, and timbre to distinguish music emotions. He defines emotion types based on energy and stress. The energy of the music is a range between calm and energetic, while the stress is a range of happy and anxious/sad. This research divides the music emotions into 8 types; namely, exuberance, anxious/frantic, contentment, depression, calm, energetic, happy, and anxious/sad.

Y. Song et al apply an SVM-based approach for classifying the music emotions based on tags of the Last.FM website. There are 4 emotions provided in this research; namely, happy, angry, sad, and relax.

Mood Cloud, a real-time music mood visualization tool, classifies music emotions into 5 types; namely aggressive, happy, party, relax and sad. It applies the SVM library to analyze the emotion dataset. The result is then presented by using a Flash player.

M. T. Quazi applies the heart rate to classify human emotions. The research proposes 4 different types of human emotions; namely neutral, happy, sad and angry. The neutral emotion has the most stable heart rate (60 – 80 bpm), while the happy emotion has the highest variation rate (70– 140 bpm) which depends on the type of happiness. The sad emotion has the second highest variation rate (80 – 100 bpm). The heart rate of angry emotion is in the same range as the happy emotion, but it will not lower than 100bpm.

i.e. either positive or negative. For Example, if a user's current emotion is sad and his preference is set to positive, the application will suggest a happy song which contains a cheer up song element. On the other hands, if a user's preference is set to negative, a song in a sad mood will be suggested, and that song may be a heavy or pop song.

There are several applications that provides facilities and services for music playlist generation or play a particular song and in this process all manual work is involved. Now to provide there are various techniques and approaches have been proposed and developed to classify human emotional state of behavior. The proposed approaches have only focused on only some of the basic emotions using complex techniques like Viola and Jones. Several research papers giving a brief about the idea are: [1] in this paper, Authors states that, Music plays a really important role in human's lifestyle and within the modern advanced technologies. Usually, the user has got to face the task of manually browsing through the playlist of songs to pick. Here we are proposing an efficient and accurate model, that might generate a playlist supported current spirit and behavior of the user.

Existing methods for automating the playlist generation process are computationally slow, less accurate and sometimes even require use of additional hardware like EEG or sensors. Speech is that the most ancient and natural way of expressing feelings, emotions and mood and its and its processing requires high computational, time, and cost. This system supported real-time extraction of facial expressions also as extracting audio features from songs to classify into a selected emotion which will generate a playlist automatically such the computation cost is comparatively low.

[2] This paper proposes an intelligent agent that sorts a music collection supported the emotions conveyed by each song then suggests an appropriate playlist to the user supported his/her current mood. The user's local music collection is initially clustered supported the emotion the song conveys, i.e. the mood of the song. This is often calculated taking into consideration the lyrics of the song, also because the melody. Whenever the user wishes to get a mood-based playlist, the user takes an image of themselves at that instant. This image is subjected to facial detection and emotion recognition techniques, recognizing the emotion of the user. The music that best matches this emotion is then recommended to the user as a playlist. [3] In this paper, Authors states that, Nowadays, people tend to increasingly have more stress due to the bad economy, high living expenses, etc. taking note of music may be a key activity that assists to scale back stress. However, it's going to be unhelpful if the music doesn't suit the present emotion of the listener. Moreover, there's no music player which is in a position to pick songs supported the user emotion. To unravel this problem, this paper proposes an emotion-based music player, which is in a position to suggest songs supported the user's emotions; sad, happy, neutral and angry. The appliance receives either the user's pulse or facial image from a sensible band or mobile camera. It then uses the classification method to spot the user's emotion. This paper presents 2 sorts of the classification method; the guts rate-based and therefore the facial image-based methods. Then, the appliance returns songs which have an equivalent mood because the user's emotion. The experimental results show that the proposed approach is in a position to exactly classify the happy emotion because the guts rate range of this emotion is wide. [4] Authors says that, Digital audio is straightforward to record, play, process, and manage. Its ubiquity means devices for handling it are cheap, letting more people record and play music and speech. Additionally, the web has improved access to recorded audio. So,

the quantity of recorded music that folks own has rapidly increased. Most current audio players compress audio files and store them in internal memory. Because storage costs have consistently declined, the quantity of music which will be stored has rapidly increased. A player with 16 G bytes of memory can hold approximately 3,200 songs if each song is stored in compressed format and occupies 5 Mbytes. Effectively organizing such large volumes of music is difficult. People often listen repeatedly to a little number of favorite songs, while others remain unjustifiably neglected. We've developed Affection, an efficient system for managing music collections. Affection groups pieces of music that convey similar emotions and labels each group with a corresponding icon. These icons let listeners easily select music consistent with its emotional Content. Experiments have demonstrated Affection's effectiveness. [5] In this paper, a sensible music system is meant by recognizing the emotion using voice speech signal as an input.

The target of the speech emotion recognition (SER) system is to work out the state of emotion of a person's being's voice. This study recognizes five emotions-anger, anxiety, boredom, happiness and sadness. The important aspects in implementing this SER system includes the speech processing using the Berlin emotional database, then extracting suitable features and selecting appropriate pattern recognition or classifier methods to spot the emotional states. Once the emotion of the speech is recognized, the system platform automatically selects a bit of music as a cheer up strategy from the database of song playlist stored. The analysis results show that this SER system implemented over five emotions provides successful emotional classification performance of 76.31% using GMM model and an overall better accuracy of 81.57% with SVM model.

II. MOTIVATION

It was observed in a cross-database experiment [1] that raw features worked best with Logistic Regression for testing RaFD (Radboud Faces Database) database and Mobile images dataset. The accuracy achieved was 66% and 36% respectively for both using CK+ dataset as a training set. The additional features (distance and area) reduced the accuracy of the experiment for SVM (Support Vector Machine) from 89%. The algorithm that had been implemented generalized the results from the training set to the testing set better than SVM and several other algorithms. An average accuracy of 86% was seen for RaFD database and 87% for CK+ database for cross-validation=5. The main focus was feature extraction and analysis of the machine algorithm on the dataset. But accurate face-detection algorithms become very important if there are multiple people in the image. One of the works [10] was tested by deriving expression from the live feed via the system's camera or any pre-existing image available in the memory. It has been implemented using Python 2.7, OpenCV and NumPy. The objective was to develop a system that can analyse the image and predict the expression of the person. The study proved that this procedure is workable and produces valid results.

III. PROPOSED SYSTEM

A novel face emotion recognition system from video frames is proposed. Detection and recognition of emotion depends on the input video frames taken. The quality of the image is found using image quality metrics. Initially Face region is detected from the human face by an algorithm and skin color segmentation with RGB color space is used to extract the face skin and non-skin regions. After the segmentation, morphological operations are applied to extract the boundary of the non-skin regions namely eyes and mouth. Then the emotions are recognized by calculating the area of the mouth region. The proposed methodology involves; image acquisition, pre-processing, face detection, segmentation, and morphological processing and area calculation. Algorithm Convert video into frames, Read the input video frame image, Convert the image into grayscale image. Enhance the input image with median, wiener and Gaussian filters find the best filter based on PSNR, RMSE values. Apply viola-jones algorithm to detect the face region. Use bounding box method and crop the face region. Use threshold value to extract non skin regions. Apply morphological operations to extract continuous boundaries of non-skin region. Mask the boundary from the original image. Extract the mouth region. Area is calculated from extracted mouth region. Recognize facial emotions based on the value of area.

The system first checks for the presence of a face in the input using the face detection process, then classifies the input and generates an output which is an Emotion (mood) based on the expression extracted from the real time graphical input.

The main aim of this work is to play the music according to the emotion of the user, which is detected from the facial expression of the User. This is made possible by using 2 technologies: One is the machine learning algorithms which are used to detect the emotion of the user from their facial expression.

The other one is the web development tools for developing the UI or frontend of the music player.

The music player opens with an interactive UI and offers 3 modes for the user to select from Random mode, Emotion mode, Queue mode.

In Random mode the music is played randomly irrespective of the emotion. In queue mode, the music is played according to the playlist defined by the user. The main feature of the project is the emotion mode in which the music is played according to the emotion of the user. As soon as the user selects the emotion mode the algorithm accesses the device's camera and using the OpenCV library of python the image of the user is captured and sent to the algorithm for emotion analysis. an application that runs based on the idea that we can detect a person's mood based on the expression on his face.

Python a versatile programming language doesn't come pre-installed on your computer devices. Python was first released in the year 1991 and until today it is a very popular high-level programming language. Its style philosophy emphasizes code readability with its notable use of great whitespace. The object-oriented approach and language construct provided by Python enables programmers to write both clear and logical code for projects. This software does not come pre-packaged with Windows.

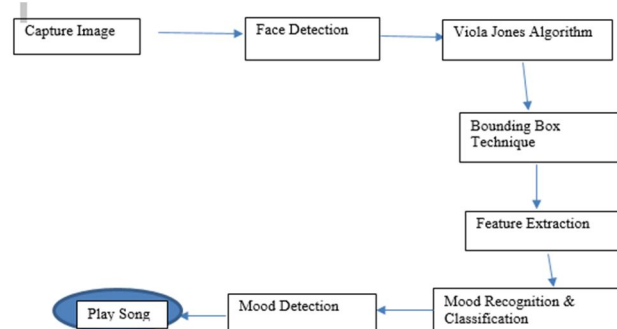


Figure:1 System Flow diagram

IV. CONCLUSION

In this paper, we have discussed that how our proposed system recommend music based on facial expression using machine learning algorithms. The proposed system is also scalable for recommending music based on facial expressions by using techniques after collecting data. The system is not having complex process to recommend music that the data like the existing system. Proposed system gives genuine and fast result than existing system. Here in this system we use machine learning algorithms to recommend music based on real time facial expression. This proposed system is developed with the help of a Machine learning module in Python Script. We have used OpenCV, TensorFlow, Keras, Web Browser and of course Python Scripting to develop this automated system. This will eventually help the person and calm the user mood appreciate to their wise choice that will help him to improve their mood. The songs will be played with a particular emotion playlist according to the emotion detected. The research done in the area of human psychology has proven that music has an intense effect on the listeners.

V. RESULT AND DISCUSSION

Below screen showing dataset details used to train above algorithms

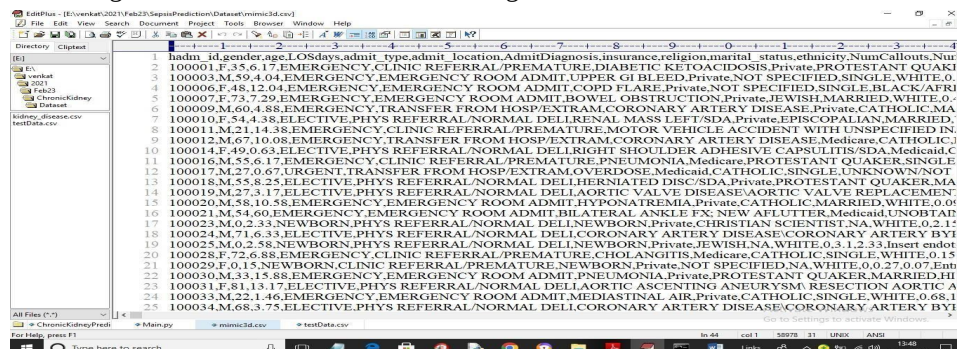


Figure:2 MIMIC-III Dataset

In above dataset screen first row contains column names and remaining rows contains dataset values and from above dataset we are using Hospital Expired column as Target Value. We have coded this project using JUPYTER notebook and below are the code and outputscreens with blue colour comments.

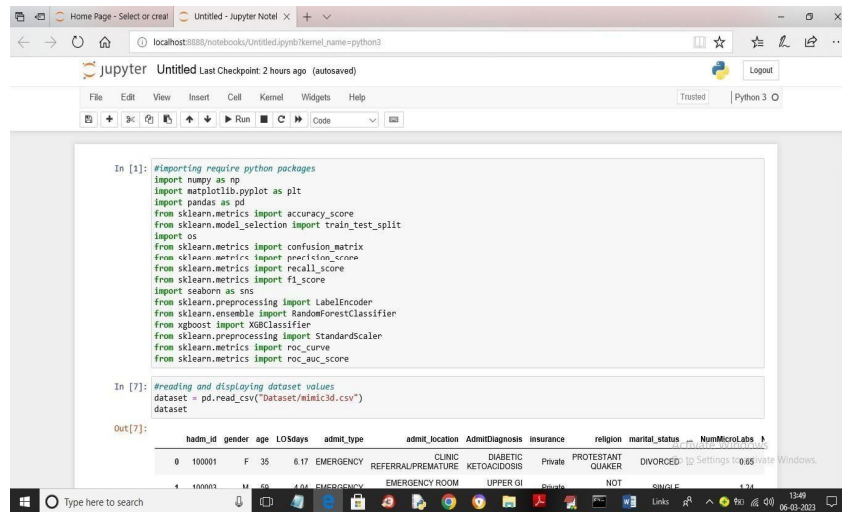


Figure 3: Importing Python Packages

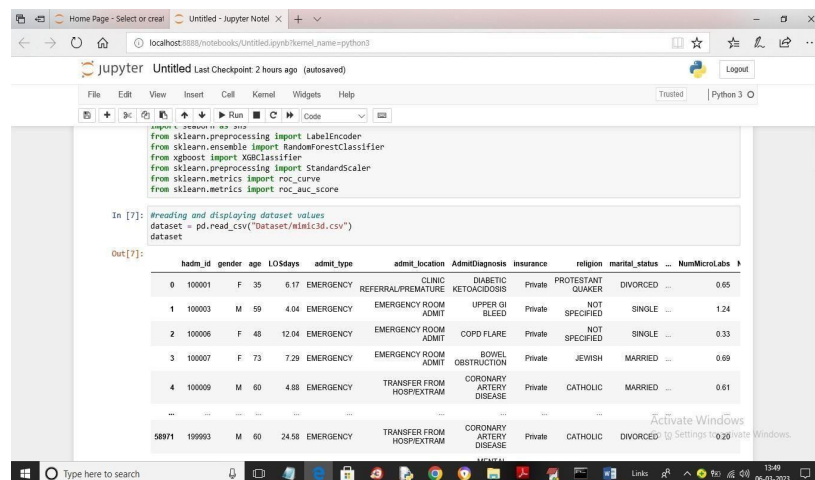


Figure 4 : Reading and displaying dataset values

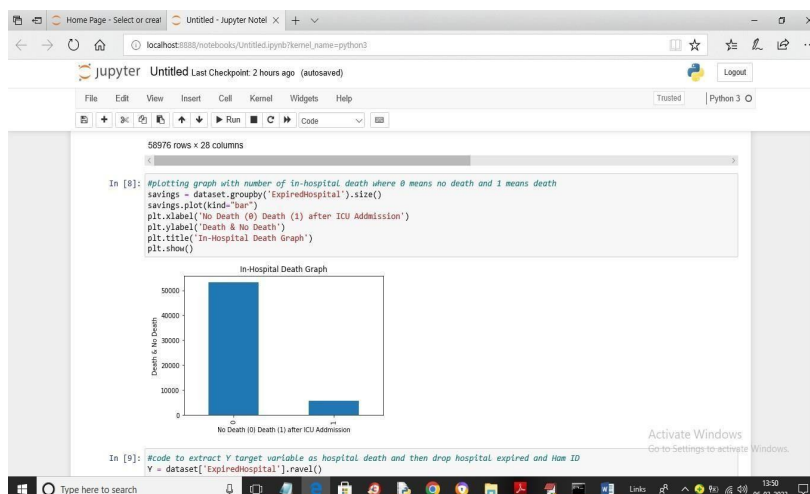


Figure 5: In-Hospital Death Graph

In above screen finding and plotting graph of ‘No-Death & in-hospital death’ where 0 meansno death and 1 means death which is in X-axis and death count in y-axis.

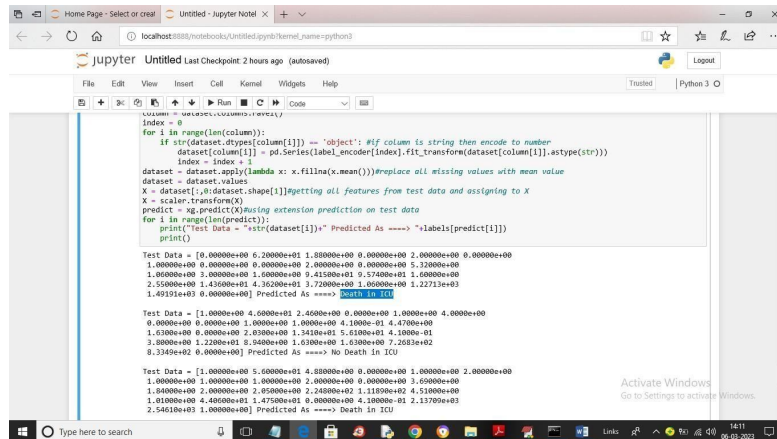


Fig 6: Predicting death

In above screen we are reading test data and then predicting death or no death and in above screen in square bracket we can see test values and after arrow symbol = $\Rightarrow$ we can see predicted values as ‘Death in ICU’ or ‘No Death in ICU’.

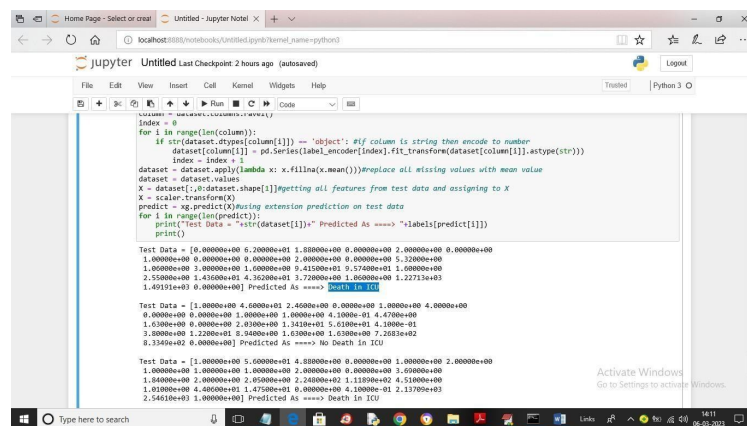


Fig 6: Predicting death

In above screen we are reading test data and then predicting death or no death and in above screen in square bracket we can see test values and after arrow symbol $\Rightarrow$ we can see predicted values as ‘Death in ICU’ or ‘No Death in ICU’.

REFERENCES

- [1] Healthcare Department, “Healthcare Department news,” 2 February 2018. [Online].
- [2] M. T. Quazi, “Human Emotion Recognition Using,” MASSEY UNIVERSITY, PALMERSTON NORTH, New Zealand, 2012.
- [3] “How music changes your mood”, Examined Existence. [Online]. Available: <http://examinedexistence.com/how-music-changes-yourmood/>. Accessed: Jan. 13, 2017.
- [4] Ujjwalkarn, “An intuitive explanation of Convolutional neural networks,” the data science blog, 2016. [Online]. Available: <https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>. Accessed: Jan. 13, 2017.
- [5] Punith Kumar M B And Dr. P.S. Puttaswamy, "Video to Frame Conversion of TV News Video By Using Mat Lab", International Journal of Advance Research in Science And Engineering, IJARSE, vol. No.3, issue no.3, march 2014 issn-2319-8354(e) 95.
- [6] Kanchan Lata Kashyap, Sanjivani Shantaiya "Noise Removal of Facial Expression Images Using Wiener Filter", National Conference on Emerging Trends in Computer Science and Information Technology (ETCSIT) 2011 proceedings published in International Journal of Computer Applications (IJCIA).