iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** I **Month of publication:** January 2025

DOI: https://doi.org/10.22214/ijraset.2025.66597

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

Predicting Human Behavior by Voice and Face Modulation

Soundarya S [1], Lalitha S[2]
[1, 2]*Dept of Electronics and Communication Engineering, B M S College of Engineering, Bangalore, Karnataka, India*

Abstract: ***The secret to clearly expressing one's views and ideas is communication. Speech is the most prevalent and efficient means for humans to communicate out of all the many types of communication. The Internet of Things (IoT) age is evolving quickly, making increasingly sophisticated technology accessible for daily use. Basic wearable gadgets to sophisticated self-driving cars automated systems used in a variety of industries are examples of these applications. Intelligent apps mostly employ voice- based input, are interactive, and require little user effort to operate. Because of this, these computer programs must be able to fully understand human speech. The speaker's age, race, gender, language, and mood may all be inferred from a speech percept. An emotion detection system is integrated with a number of current voice recognition systems utilised in Internet of Things applications to assess the speaker's emotional state. One of these techniques for identifying human emotions (such as fear, neutral, anger, pleasure, disgust, sorrow, and surprise) is based on the technique for analysing facial expressions employing the popular deep learning.***

I. INTRODUCTION

Emotion recognition touted as a fundamental part of human communication, is found to play a pivotal in a range of domains including human-computer interaction, affective computing, and sentiment analysis. Detecting emotions accurately from different modalities such as facial expressions, voice tonality, and textual content is a complex yet essential task. This project focuses on the development of a comprehensive "3- way Emotion Detection" system, integrating three distinct modalities: facial expression analysis using Convolutional Neural Networks (CNNs), voice-based emotion detection utilizing the Librosa library, and text-based emotion recognition employing advanced Natural Language Processing (NLP) techniques. There holds great promise for enhancing the efficacy of human-computer interaction systems and advancing our understanding of emotional cues in digital communication. One of these techniques for identifying human emotions is deep learning-based facial expression identification. The goal would be in accurately accurately determining the state by understanding the facial expressions in a jiffy. Through the following approach, CNN is trained using labelled face photos taken from the expressions dataset. The suggested CNN model could then help decide in which facial expression is made. Artificial intelligence (AI) has emerged as a prominent and essential field that has gained significant attention from researchers and programs. It has rapidly seen as an major aspect in our lives through various applications such as chatbots, digital assistants like Siri, and other technological systems. Among the powerful applications of AI, face recognition techniques stand out, as evidenced by Google Photos' ability to group photos of individuals.

This paper proposes the integration of these two concepts, creating a system that recommends music, podcasts, and movies based on facial emotion, text, and voice recognition. Emotion recognition has a broader scope in fields like robotics, enabling efficient sentiment analysis without the need for human involvement. The motivation emanate from the emotion recognition necessity in technological applications. As human-computer interaction becomes more pervasive, recognising and reacting to human emotions are critical for creating user-centric and emotionally intelligent systems. Existing systems often rely on single- modal approaches, which may lack the depth and accuracy needed for a nuanced understanding of emotions. This project is motivated by the need to overcome these limitations and create a more robust emotion detection system by combining information from facial expressions, voice tonality, and textual content.

A. Phases in Facial Expression Recognition

Images of different facial expressions are used to train the facial expression recognition system using supervised learning. picture capture, face identification, picture preliminary processing, feature extraction, and categorisation are among the steps that make up the system's training and testing phases. The following is a summary of the procedure:

1) Image Acquisition

Either a dataset or real-time camera capture are used to gather images with face emotions.

©IJRASET: All Rights are Reserved

2) Face Detection

- Image Pre-processing:

It involves normalisation against changes in pixel location or brightness as well as noise reduction.

a) Color Normalization
b) Histogram Normalization

- Feature Extraction

The most crucial aspect of a pattern classifying task is choosing the feature vector. The key characteristics are then extracted from the facial picture following pre-processing. Scale, position translation, and light level fluctuations are among the intrinsic issues with picture categorisation.

3) Recommendation System

The recommendation system in this project combines advanced algorithms and manual curation to ensure a personalized and enriching user experience. Manually registered links, such as those for healthcare professionals, media souces, fast-consuming content sites like YouTube, play an important part in enhancing relevance along with reliability of recommendations.

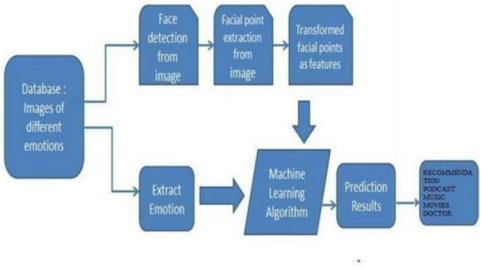


Fig 1 Facial Emotion Recognition Using Machine Learning

This fusion of algorithmic intelligence and human touch sets our system apart. It's like having a wise kind of friend alongside you, one who reads our emotions with discerning eyes and offers thoughtful process suggestions that not only entertain but also uplift and guide. It's a system that transcends the limitations of the ordinary, recognizing that true personalization lies not only in what you're recommended, but also in how it aligns with the intricate tapestry of our emotions.

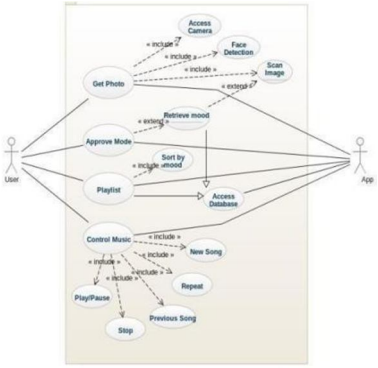


Fig 2 Use Case Diagram

The image is a data flow diagram (DFD) that shows how the system works to recognize a user's emotion and play music based on it. The DFD is divided into four levels, with each level showing more detail about the system.

It shows the possible ways that some person can interact with this build. The system in this case is a music player. The user can access the camera, scan an image, get a photo, retrieve the mood, and approve the mode. The usercan also control the music player by playing, pausing, repeating, stopping, and playing the previous song. The user can also access the database and create a new song.

Use case diagrams are an excellent tool for illustrating the various ways a user may engage with a system. They can be utilised to distinguish between the various use cases and user categories of the system. A higher-level perspective of the system can also be provided with the use of use case diagrams.

Here is a more detailed explanation of the different use cases in the diagram:

- Access Camera: This use case allows the user to access the camera on their device. This could be used to takea picture of a song to add to the playlist.
- Scan Image: This use case allows the user to scan an image. This could be used to scan a barcode on a CD orto scan a QR code that links to a song.
- Get Photo: This use case allows the user to get a photo from their device. This could be used to add a photo tothe playlist or to use as the album art for a song.
- Retrieve Mood: This use case allows the user to retrieve their mood. This could be used to create a song list that relates to the user's mood. Approve Mode: This use case affords the user to approve the mode. This could be used to approve the mood that was retrieved or to approve a new mode that the user created.
- Previous Song: This use case helps the user to play the previous song in the playlist.
- Control Music: This use case lends the user to control the music player. This could include playing, pausing, repeating, stopping, and playing the previous song.
- Access Database: This use case allows the user to access the database. This could be used to add songs to the playlist or to get information about songs.
- Create a New Song: This use case empowers the user to create a new song. This could be done by adding a title, or artist name and album to the song.
- Access Camera: The user grants the app permission to access the device's camera.
- Scan Image: The app captures an image of the user's face using the camera.
- Face Detection: The app analyzes the captured image to detect facial expressions.
- Retrieve Mood: Based on the detected facial expressions, the app attempts to identify the user's current mood. Sort by Mood: The app uses identified mood attribute to sort the music library and recommend songs that match the user's mood.
- Get Photo: (Extension of Retrieve Mood) The user can choose to save the captured image for later use. Approve Mode: (Extension of Retrieve Mood) The user can confirm or reject the app's mood assessment. Control Music: The user can play, pause, stop, skip to the next/previous song, and repeat songs using playback controls.
- Playlist: The user can access and manage playlists, including creating new playlists, adding songs to existing playlists, and removing songs from playlists.
- Access Database: The app accesses the music library database to retrieve music files based on user selections or recommendations.

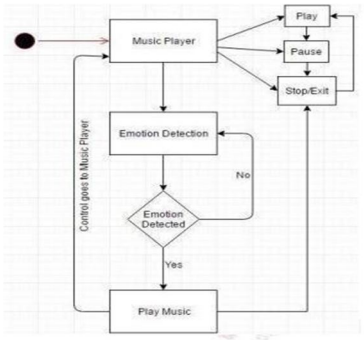


Fig 3 Activity Diagram

The activity diagram illustrates the flow of actions and interactions within the web application for 3-way emotion detection. It starts with the "User Interaction" and branches based on the user's actions, including uploading an image, recording audio, entering text.

- ✓ Facial Expression Detection: If the user uploads an image, the system processes the image using facial expression detection to analyze and detect emotions from the facial expressions. The detected emotions are then displayed to the user.
- ✓ Voice-Based Emotion Detection: If the user records audio, the system processes the audio using voice-based emotion detection to analyze and detect emotions from the voice. The detected emotions are then displayed to the user.
- ✓ Text-Based Emotion Detection: If the user enters text, the system processes the text using text-based emotion detection to analyze and detect emotions from the text. The detected emotions are then displayed to the user.

II. RELATED WORK

A. *Convolution Neural Network*

Convolution neural networks are a special type of feed-forward NN where the visual brain serves as inspiration for the connections between the layers. One type of deep neural network used for visual imaging analysis is the Convolution Neural Network (CNN). They may be used in natural language processing, picture categorisation, video and image recognition, and more. The initial extraction layer information with a picture input is called convolution. Convolution uses little squares of input data to learn visual attributes while maintaining the relation among pixel entities. This mathematical process requires two inputs, which can be an image array and the filter or kernel. Each input image will pass through multiple layers of convolution with filters (kernels) to produce output feature maps.

Basically, the convolution neural networks have 4 layers :- convolution layers, ReLU layer, pooling layer and the fully connected layer.

B. *Convolutional Layer*

A tinge of the photographs taken with the use of convolution layers computer later which has read the picture in set. These are referred to as features or filters. The convolutional layer becomes far more adept at identifying similarities than complete image matching scenarios when these partial feature matches are sent in around similar location in both the pictures. The newly input photographs are compared with filters; provided they match, the image can be appropriately labelled. In this case, align the features and the picture, then split the entire amount the feature has pixels by the combined sum of pixels, then pixel multiplication of the picture by the corresponding feature pixel. We make a map and arrange the filter's values where they belong. In the same way, movement of the feature to each other location in the image and see how the feature corresponds with each of those locations. Ultimately, the result will be a matrix.

C. *ReLU Layer*

We remove any negative values from the filtered images and replace them with zero in the rectified linear unit, or ReLU layer. This is done to keep the numbers from accumulating up to zeroes. This transform function does not activate a node if its supplied value exceeds a preset threshold and eliminates any negative numbers from the matrix. The output will be zero if the value entered is less than zero.

D. *Pooling Layer*

We decrease or diminish the image's size in this layer. Here, we choose the window size initially, then we stroll your window over your filtered photographs after mentioning the necessary stride. Next, extract the highest values from every window. By doing this, the layers will be pooled and the picture and matrix will both get smaller. The reduced size matrix is sent into the fully connected layer.

E. *Fully Connected Layer*

After it has gone through the convolutional, ReLU, and pooling layers, we must stack all of the layers. The input picture is classified using the fully connected layer. Unless a 2x2 matrix is obtained, these layers must be repeated if necessary. The real categorisation then takes place in the fully linked layer at the end.

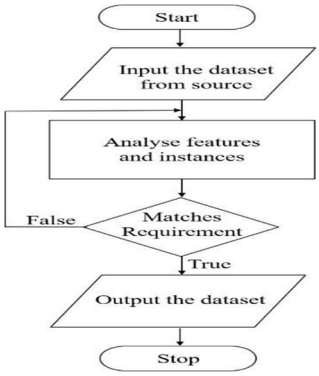


Fig 4 Flowchart For Data Acquisition

Figure 4 shows how the data collection flowchart looks. A thorough analysis is performed after the data set is gathered from a source. Only when the image satisfies our standards and is unique is it chosen for training or testing.

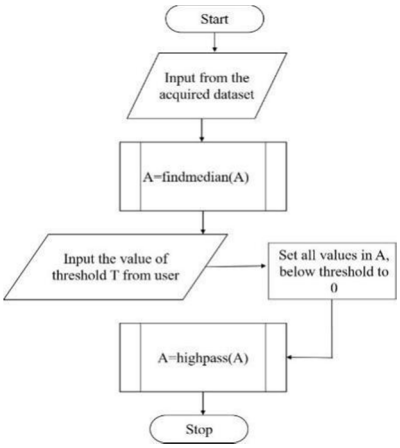


Fig 5 Flowchart for the pre-processing module

Figure 5 shows the flowchart for preparatory processing the images derived from the previous phase's final output. The image is transformed from Colour to greyscale to facilitate processing. After that, a high-pass filter is used to enhance the finer features, global basic thresholding is used to eliminate the background, and an averaging filter is used to eliminate noise.

The first step in the process is to recognise the face in the input image, followed by identifying official features involve eyes and mouth. These features would be subjected to specific filters and transformations [1].

The image classification system typically involves two stages: Requires labeled data: This means we need a dataset of images which has each image is already correctly assigned with category process: learning from the labeled data that can identify patterns and relationships between the image pixels and the corresponding category is the model's objective [2].

This study focuses on addressing investigates various Convolutional Neural Network (CNN) parameters and designs for the identification of the following seven feelings in human faces: Contempt, rage, fear, disgust, joy, grief, and surprise. It also discusses the difficulties related to Emotion Recognition Datasets. The iCV MEFED (Multi-Emotion Facial Expression Data set) is selected as the key data set for this research due to its novelty, interest, and high level of difficulty. The initial CNN network architecture consists of the following: Convolutional Layers, four Max Pooling Layers, one Dropout Layer, and two Fully Connected Layers[3].

This paper explains about an autonomous face emotion categorisation system that extracts features using Speeded Robust Features (SURF) and Convolutional Neural Networks (CNN). The suggested model's 91% accuracy makes it possible to identify human emotions using facial expressions in an efficient manner[4].

The study of Facial Emotion Recognition entails three main phases which have Pre-processing, feature extraction and classification stages. For the application, they have contrasted the various deep learning architectures available in Keras. By using Transfer Learning from well-known pre-trained models like VGG16, ResNet152V2, Xception, and InceptionV3, it makes use of Deep Facial Features in photos. A data set that includes the Cohn-Kanade Dataset (CK+) and the Japanese female facial emotion (JAFFE) data set may be used to assess the performance of the bottleneck features that are generated for the input photos [5].

Facial recognition is an automated identification technology that requires facial features, including statistical or geometric features. It combines digital image processing, video processing, and other related technologies Research has revealed that the human brain's ability to recognize faces involves the coordinated activity of thousands of neurons. Machine learning along with deep learning algorithms have harnessed this understanding to develop intricate networks that mimic neural functions[6].

Facial emotion recognition has been revolutionized by deep learning techniques, and this study aimed to develop a Deep Convolution Neural Network (DCNN) which can accurately predict five different facial emotions. This uses two convolution layers and dropout layers to prevent over fitting, and the input image is reformed to 32 x 32 and processed through these layers, later going with ReLU activation and pooling. A second convolution layer produces a 2-dimensional array that houses feature values, which gets flattened and fed to the neural network's dense layers[7].

The automatic facial expression detection system that focuses on seven parameters: anger, fear, disgust, sadness, contempt, surprise and happiness. The system utilizes recorded data of images and employs an experimental algorithm tested on both the AdvancedVisual Database and a natural building database.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The method involves dividing the image into an matrix, adjusting grayscale and scale, and rotating the lower area of the image[8].
The rise of social media has led to an overwhelming amount of textual data, necessitating an approach to extract information that is detrimental from it. Aimens system proposes a solution to find emotions such as happy, sad, and angry from contextual conversations. The system implements the Long short-term memory (LSTM) model with word2vec and doc2vec embeddings, achieving an F-score of 0.7185 [9].
TheOpen SMILE software is used for feature extraction, and the emotional features are found by calculating statistics using 12 functions from 16 real acoustic features. The classifier considered here achieves an accuracy of 89%. The system is established by signals via speech through an interface, performing end- point detection, extracting acoustic features, applying statistical functions, and inputting them in classification model. The program called WeChat is used to gather speech signals in MP3 format, whichis converted to WAV format before feature extraction. The system works on accumulating speech signals within 60 seconds, these would be sent to the background, which returns a JSON string including possible errors and the category of emotion of the speech. The study shows the identification results of the speaking emotion recognition model using the WeChat program, that is more suitable for the creation of conversational emotion recognition systems due to its user-friendly design [10].

III. RESULTS

The web application for emotion detection was successfully implemented, incorporating facial expression detection, voice-based emotion detection, The system was tested using various inputs to evaluate its performance and accuracy.

A. Voice-Based Emotion Detection

The Librosa library used for voice-based emotion detection performed well in analyzing audio recordings and detecting emotions based on tone, pitch, and intensity. The model showed good accuracy in detecting emotions such as joy, fear, surprise, etc., from the audio inputs.
The base code emotion detection from real time audio signal is written in python and run in Python 3.7.
This run.py file contains the code for importing emotion packages and web application that enables by user to give the input voice.

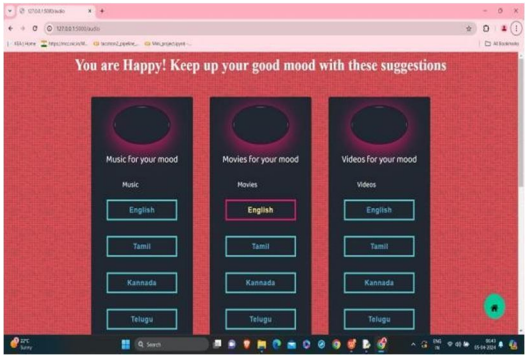


Fig. 6 Describing Happy emotion

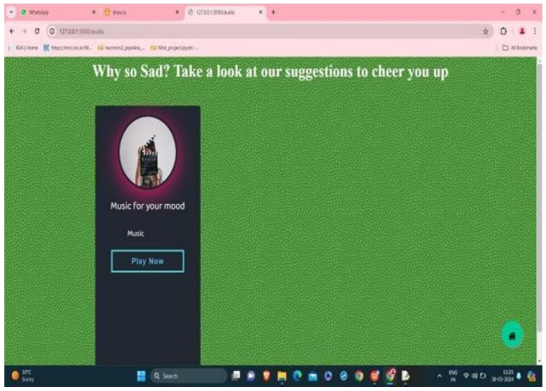


Fig. 7 Describing Sad emotion

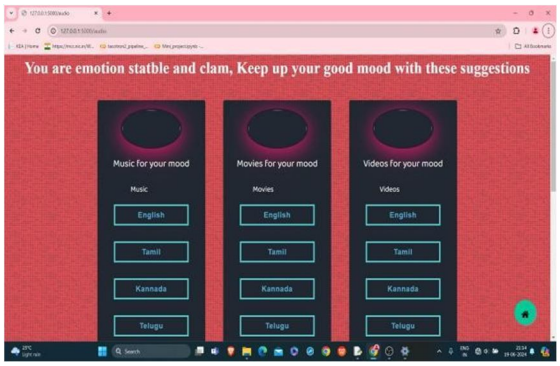
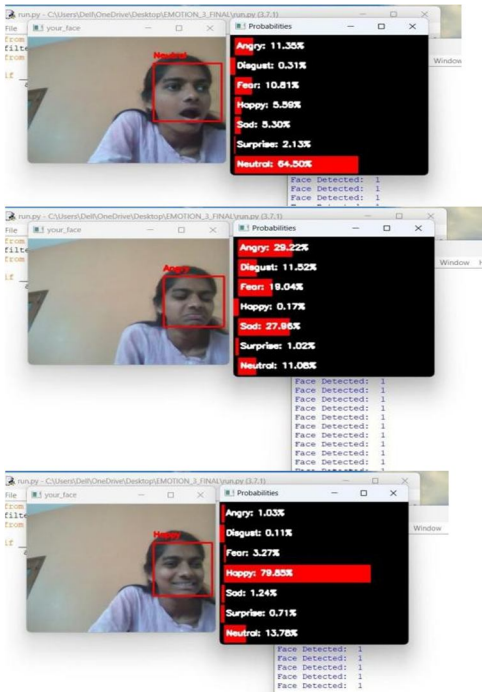


Fig. 8 Describing Stable emotion

B. Facial Expression Detection

The VGG16 model used for facial expression detection achieved high accuracy in classifying facial expressions into different emotional states, including happiness, sadness, anger, etc. The model was able to detect subtle changes in facial expressions and provide accurate results.

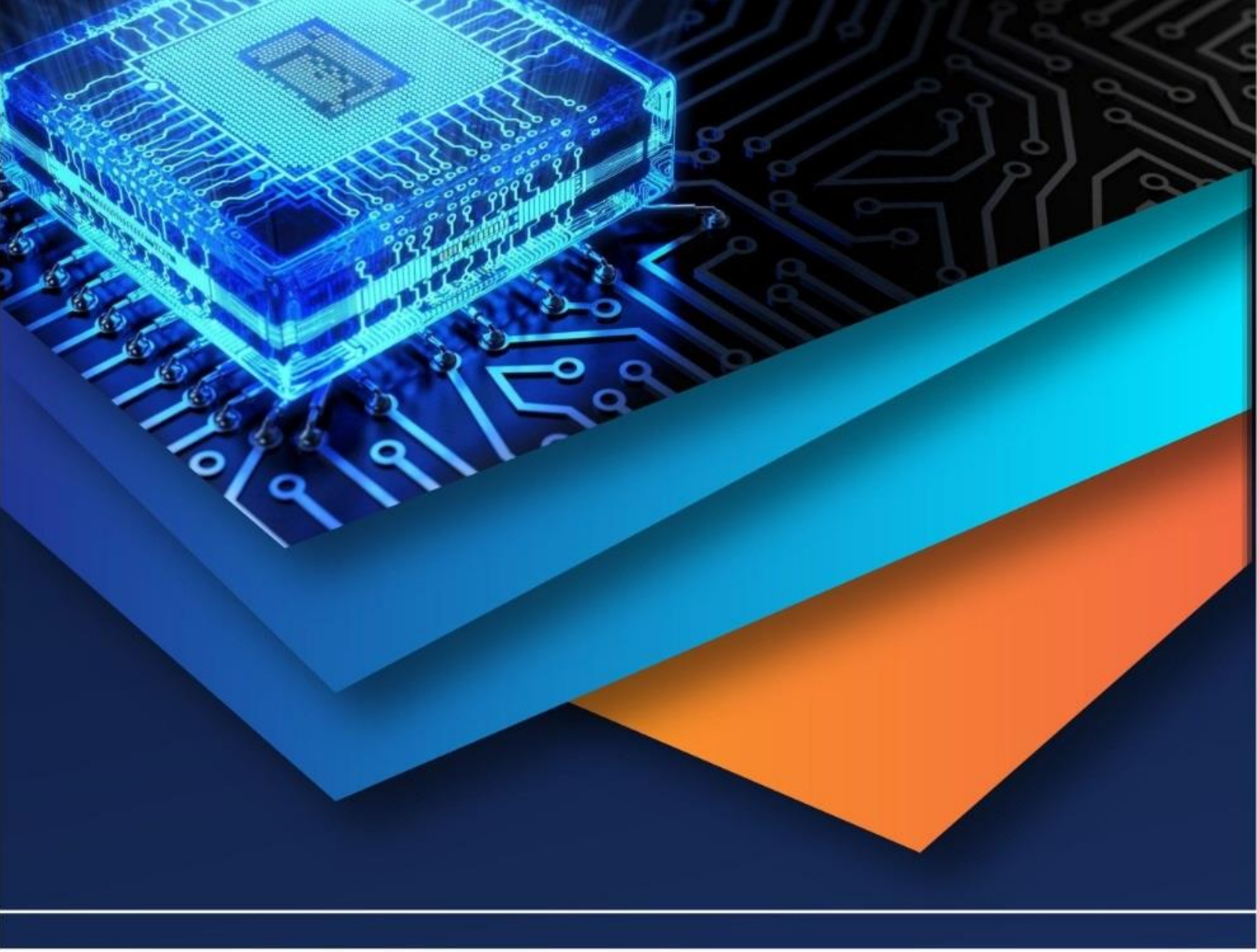


REFERENCES

[1] Raut, Nitisha, "Facial Emotion Recognition Using Machine Learning" (2018). Master's Projects. 632. https://doi.org/10.31979/etd.w5fs- s8wd
[2] Hemanth P,Adarsh ,Aswani C.B, Ajith P, Veena A Kumar, "EMO PLAYER: Emotion Based Music Player", International Research Journal of Engineering and Technology (IRJET), vol. 5, no. 4, April 2018, pp. 4822-87.
[3] Music Recommendation System: "Sound Tree", Dcengo Unchained: Sıla KAYA, BSc.; Duygu KABAKCI, BSc.; Işınsu KATIRCIOĞLU, BSc. and Koray KOCAKAYA
[4] BSc. Assistant : Dilek Önal Supervisors: Prof. Dr. İsmail Hakkı Toroslu, Prof. Dr. Veysi İşler Sponsor Company: ARGEDOR
[5] Tim Spittle, lucyd, GitHub, , April 16, 2020. Accessed on: [Online], Available at: https://github.com/timspit/lucyd
[6] A. Abdul, J. Chen, H.-Y. Liao, and S.-H. Chang, "An Emotion-Aware Personalized Music Recommendation System Using a Convolutional Neural Networks Approach," Applied Sciences, vol. 8, no. 7, p. 1103, Jul. 2018.
[7] Manas Sambare, FER2013 Dataset, Kaggle, July 19, 2020. Accessed on: September 9, 2020. [Online], Avalilable at: https://www.kaggle.com/msambare/fer2013
[8] MahmoudiMA, MMA Facial Expression Dataset, Kaggle, June 6, 2020. Accessed on: September 15, 2020. [Online], Available at: https://www.kaggle.com/mahmoudima/mma- facial-expression
[9] Dr. Shaik Asif Hussain and Ahlam Salim Abdallah Al Balushi, "A real time face emotion classification and recognition using deep learning model", 2020 Journal. of Phys.: Conf. Ser. 1432 012087

[10] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, San Francisco, CA, USA, 2010, pp. 94-101, doi: 10.1109/CVPRW.2010.5543262.
[11] Puri, Raghav & Gupta, Archit & Sikri, Manas & Tiwari, Mohit & Pathak, Nitish & Goel, Shivendra. (2020). Emotion Detection using Image Processing in Python.
[12] Patra, Braja & Das, Dipankar & Bandyopadhyay, Sivaji. (2013). Automatic Music Mood Classification of Hindi Songs.
[13] Lee, J., Yoon, K., Jang, D., Jang, S., Shin, S., & Kim, J. (2018). MUSIC RECOMMENDATION SYSTEM BASED ON GENRE DISTANCE AND USER PREFERENCE CLASSIFICATION.
[14] Kaufman Jaime C., University of North Florida, "A Hybrid Approach to Music Recommendation: Exploiting Collaborative Music Tags and Acoustic Features", UNF Digital Commons, 2014.
[15] D Priya, Face Detection, Recognition and Emotion Detection in 8 lines of code!, towards data science, April 3, 2019. Accessed on: July 12, 2020 [Online], Available at: https://towardsdatascience.com/face-detection- recognition-and-emotion-detection-in-8-lines-of- code-b2ce32d4d5de
[16] bluepi, "\Classifying Different Types of Recommender Systems, November 14, 2015. Accessed on: July 7, 2020. [Online], Available on:https://www.bluepiit.com/blog/classifying- recommender- systems/#:~:text=There%20are%20majorly%20s ix%20types,system %20and%20Hybrid%20recommender%20system
[17] T. L. Nwe, S. W. Foo, and L. C. De Silva, "Speech emotion recognition using hidden Markov models," Speech Commun., vol. 41, no. 4, pp. 603–623, Nov.2021.
[18] S.S.Narayanan, "Toward detecting emotions in spoken dialogs,"in IEEETrans. in Speech Audio Process., vol. 13, no. 2, pp. 293–303, Mar. 2021.
[19] B. Yang and M. Lugger, "Emotion recognition from speech signals using new harmony features," Signal Processing, vol. 90, no. 5, pp. 1415–1423, May2020.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)