iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.62870

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 12 Issue V May 2024- Available at www.ijraset.com

Tune into Emotions: A Study ofMusical Therapy's Influence onFacial Expression Recognition

Sampada Huria[1], Shashwat Mishra[2], Dravyanshu Singh[3], Mr. Ashwini Kumar[4]
Graphic Era Deemed to be University, Dehradun

***Abstract:** This research paper introduces an innovative approach to music recommendation systems, utilizing facial expression analysis for the delivery of personalized music suggestions. Through the utilization of machine learning algorithms, the system assesses the user's emotional state by analyzing facial expressions, subsequently offering music recommendations aligned with their mood. The proposed methodology incorporates a deep learning-based model for the detection and classification of facial expressions, complemented by a collaborative filtering algorithm to generatepersonalized music recommendations*
Keywords: ***personalized music, deep learning, music recommendation***

I. INTRODUCTION

Music, a universal language that transcends borders and cultures, has long been recognized for its profound impact on human emotions and well-being. It is capable of bringing about feelings of pleasure, relaxation and even relief. By using this emotional influence, music therapy has emerged as a treatment modality enabling it to leverage the inherent link between musicand people's emotions in order to improve their mental and personal well-being.

Music therapy is a branch of evocative art therapy, in which music can be used to deal with emotions, thinking and social needs. In a variety of contexts this innovative healing approach has proved promising, from alleviation of stress and anxiety to enhancing brain function and emotional expression. Musical therapy, which is an excellent choice for those interested in a comprehensive well-being, often offers individuals the safe place where they can examine and express their emotions.

This project proposes a novel system for detecting a person's mood through facial emotion recognition using Convolutional Neural Networks (CNNs) and subsequently generating a music playlist based on the identified mood. Overcoming the challenges faced by machines in interpreting facial expressions, the project leverages advancements in CNN technology. This dataset ensures robustness across diverse facial expressions. The system aims to seamlessly integrate facial emotion analysis with personalized music recommendations, enhancing user engagement and offering a unique, immersive experience. A person normally signs his or her mood by means of gesture such as hand movement, face expression and voice tone. We're focusing primarily on the face for this project because of how important it is. There are frequent links between mood and music. Depending on the song, we may feel joyful, depressed, energized, or relaxed. Music therapy has been examined because of the important effect that music may have on outlook and quality of life, to manage several health conditions, improve mood or reduce stress. The program uses a camera to take real-time pictures of your face, detect facial features such as emotion detection, mood interpretation and predict emotions before it connects with the Playlist Player module that will inform you what music is playing to improve your mood. We're using 4 here, Basic emotions are Anger, Surprise, Happy and Sad. This system shall be applicable to users who wear spectacles, in contrast with the provision of this feature which is not provided for by other such systems.

II. LITERATURE REVIEW

The integration of emotion detection with music recommendation systems has emerged as a fascinating area of research. Several studies have proposed diverse methodologies, each addressing the intricate relationship between facial expressions and personalized music suggestions. In [1], S. Metilda Florence and M. Uma (2020) explored emotional detection for personalized music recommendation based on user facial expressions. Their system comprises three modules: Emotion-Audio extraction, Audio extraction, and Emotion-Emotion extraction. Despite its innovative approach, the system faced challenges with accuracy due to a limited picture dataset and the requirement for a well-lit environment with at least 320p image quality.

Similarly, H. Immanuel James et al. (2019) in [2] proposed an emotion-based music recommendation system through facial expression analysis. Their methodology includes using a linear classifier for face detection, gradient boosting for facial landmark map creation, and a multiclass SVM classifier for emotion classification. However, this approach struggled with accurately recording all emotions due to a limited dataset and potential generalizability issues with handcrafted features.

©IJRASET: All Rights are Reserved

In another approach, Yusuf Yaslan et al. (20XX) [3] focused on emotion learning through signals from wearable devices, such as galvanic skin response and photoplethysmography. They aimed to predict arousal and valence from multi-channel physiological signals. This study highlights the potential of wearable computing devices in capturing emotional states for music recommendations. Ramya Ramanathan et al. (20XX) [4] developed an intelligent music player using emotion recognition. They utilized convolutional neural networks for face detection and basic emotion recognition, identifying emotions such as happy, sad, anger, excitement, surprise, disgust, fear, and neutral. This work underscores the application of deep learning techniques in emotion-based music recommendation systems. These studies collectively underscore the evolving landscape of emotion-based music recommendation systems, each contributing valuable insights and methodologies. Common challenges include dataset limitations, generalizability concerns, and the quest for an accurate emotion-to-music mapping. To address these challenges, our proposed work aims to leverage a larger, more diverse dataset and employ advanced deep learning techniques for improved emotion detection and music recommendation accuracy. By integrating multiple data sources and enhancing the robustness of emotion classification, we seek to develop a more reliable and user-friendly emotion-based music recommendation system.

III. PROBLEM STATEMENT

In the field of music therapy, the field of facial expressionanalysis is one of the most interesting avenues of inquiry. The emotional response of individuals to music is often reflected in their facial expressions during therapy sessions. There are valuable insights about the effectiveness of music therapy interventions that can be gained from these expressions. As a result, our primary research concern is the following: How can the identification of facial expressions by means of artificial intelligence techniques help us to understand how musictherapy affects our emotions?

IV. PROPOSED SYSTEM

The planning process helps us express the interaction betweenthe user and musician. The purpose of this system is to capturethe face with the camera. The captured images are fed into a neural network that predicts emotions. Then, the system uses the detected emotions to generate a list of songs. The primary objective of our application is to provide a music playlist that aligns with the user's mood, whether it be happy, sad, neutral, orsurprising.

The system is designed to detect different emotional states, andif the user displays a negative emotion, it will offer a selection list containing the most suitable music to uplift their mood.

There are four modules in the facial recognition-based musicrecommendation system:

1) Real-time capture: In this module, the system capturesthe user's face in real-time.

2) Face recognition: Here, the user's face serves as input,and convolutional neural networks are programmed to measure features of user images.

3) Emotion detection: This section identifies different emotional behaviours by extracting features from the user's visual expressions.

4) Aesthetic perception: The module for accepting comments assesses the user's mood based on the typeof song feedback.

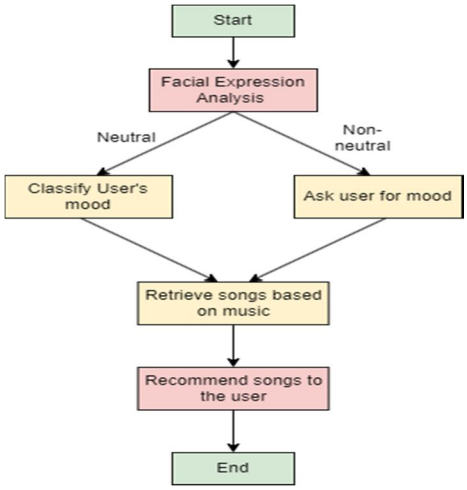


Fig.1 Block diagram of the proposed system

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

V. METHODOLOGY

A. Database Description

We employed the Kaggle dataset, FER2013, to construct a convolutional neural network model. Thedataset is divided into training and testing data. The training file contains 24,176 images, and the test file contains 6,043 images. These 48x48 pixel grayscale images of human faces are labeled with five emotions:happy, sad, angry, surprised, and neutral. Faces are consistently positioned in each image. FER-2013 incorporates both exposed and unexposed avatars, sourced from Google image searches and emotion-related expressions. Notably, the dataset presents challenges due to conflicting information, affecting emotions like happiness, sadness, anger, neutrality,surprise, hatred, and fear.

In addressing these challenges, we utilized the SoftMaxweight loss method for its relative balance in training.

However, to overcome limitations related to group features, we employed the categorical cross-entropyloss function, evaluating error rates during each iteration.

B. Emotion Detection Module

1) Face Detection

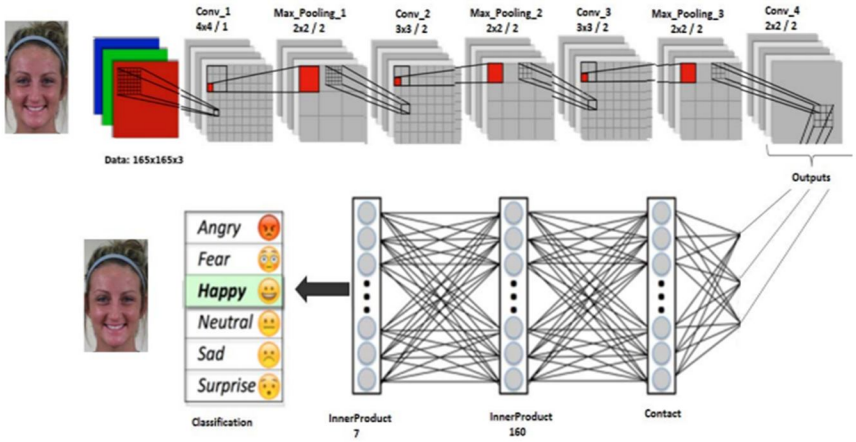
This detection operates seamlessly on both videos and photos, utilizing segmentation to distinguish the presence (1) or absence (0) of a face in an image. The classifier undergoes training with multiple images to enhance face identification accuracy. OpenCV employstwo classification methods: LBP (Local Binary Pattern)and Haar Cascades. Specifically, the Haar classifier is employed for face detection, involving training with transformed face data to enable recognition of various faces. Face detection primarily aims to identify external noise and reduce it to detect faces within the frame. This machine learning-based approach involves training a stepwise function with input methods, leveraging Haarwavelet technology. This technology learns image pixels as frames through a function, employing machine learning techniques to refine training data witha high degree of accuracy.

2) Feature Extraction

In the process of feature extraction, a pre-trained convolutional neural network (CNN) is employed as asequential model to capture significant details from images. This involves passing an image through thenetwork, halting at the initial layer, and utilizing the resulting output based on specified criteria.

The convolutional network's initialization phase extractshigh-level features using various filters, gradually increasing their number as the process deepens, albeitat a higher computational cost. The output of this model manifests as a custom map, representing the average features of each layer beyond the first.

Visualizing feature maps aids in comprehending the distinctive features each convolution layer captures. The generation of these feature maps involves applyingfilters or feature detectors to either the input image or the feature map outputs from preceding layers. The integration of effort and discrimination learned fromneural networks contributes to refining this feature extraction process. This methodology proves instrumental in discerning the importance of specificfeatures for image classification.



©IJRASET: All Rights are Reserved

3) CNN units are Described Below: Convolutional Layer

a) A Convolution Layer

Ci (i network layer) is parameterized by its number N of convolution cards.
Mi j (j ∈ {1, ..., N}), the size of the convolution kernels.
Kx × Ky (often square), and the connection diagram in the previous layer
Li−1: Each convolution card
Mi j is the result of a convolution sum of cards previous layer.
Mi−1 j by its respective convolution kernel. In the case of a fully connected card to the cards of the previous layer, the result is calculated by the equation 1.

$$\sum_{n=1}^{N} M_j^i = \phi\left(b_j^i + \sum_{n=1}^{N} M_n^{i-1} * k_n^i\right)$$

where * is the convolution operator.

b) Pooling layer

In the classical architectures of convolutional neural networks, convolution layers are followed by subsampling layers. A sub-sampling layer reduces the size of cards and introduces invariance to (low) rotations and translations can appear as input. The output of max-pooling layer is given by the maximum activation value, in the input layer for different regions of size Kx × Ky non-overlapping. Similarly to a convolution layer, a bias is added, and the result is passed to the transfer function φ defined above.

c) Fully Connected Layer

After several max pooling and convolutional layers, the high-level reasoning in the neural network is done via fully connected layers. Neurons in a fully connected layer have full connections to all activations in the previous layer, as seen in regular Neural Networks. Their activations can hence be computed with a matrix multiplication followed by a bias offset. Table I present network configuration, the patch size of data input is 165×165. Then, the convolutional and Max Pooling layers are chosen with different kernel sizes (4×4, 3×3, 2×2) and different strides (1, 2)

4) Emotion Detection

In the realm of emotion detection, a Convolutional Neural Network (CNN) architecture, plays a pivotal role. This architecture utilizes filters or capture devices on input images to generate maps through the application of the Rectified Linear Unit (ReLU) activation function. Feature detectors or filters within the CNN detect crucial image elements like edges, vertical and horizontal lines, folds, and more, therebydefining various features.

Pooling, a process that maintains output consistency despite minor input variations, is then applied using methods such as minimum, average, or maximum pooling. It illustrates the feature extraction from each layer in the CNN, showcasing the neural network's ability to discern and interpret image features. In the context of emotion recognition, the CNN model, treated somewhat as a black box, is trained to learn features from input images without explicitly defining them. When faced with a user's image, the trained CNNmodel predicts the corresponding behavior and assigns labels to the image, facilitating emotion recognition.

C. Music Recommendation Module

In the music recommendation module, a Bollywood Hindi song database has been curated, comprising 100 to 150 songs for each emotion category. Recognizing the impact of music on mood, the system tailors music playlists based on the user's real-time emotions, as determined by the mood module-ranging from happy and sad to angry, surprised, and neutral. The process involves connecting emotion tags tospecific folders within the music library, and the os.listdir() method in Python facilitates this association. Table 1 showcases song titles categorized under emotions, such as "Happy" and "Sad," within thedatabase. For instance, when the mood is identified as 'Happy,' the system dynamically changes the directory to the corresponding folder, providing a message to the user and generating a playlist from the available happy songs using os.listdir(). The playlist is then displayed in the graphical user interface (GUI). To manage the music playback, the script utilizes the Pygame library, offering functions like play, pause, resume, and stop. Variables such as playlist, song status, and root store the names of all songs, the statusof the current song, and the main GUI window, respectively. Tkinter is employed to create the GUI foremotional awareness. This integration ensures a personalized and responsive music recommendation system based on the user's emotional state.

VI. RESULT AND ANALYSIS

In the evaluation of various algorithms, including Support Vector Machines (SVMs), Extreme Learning Machines (ELMs), and Convolutional Neural Networks (CNNs) on the Fer2013 dataset, a comprehensive comparison is presented in Table 2. Each algorithm is assessed based on its validation and test accuracy.

Notably, the use of Convolutional Neural Networks significantly enhances visual perception performance,as indicated by the higher accuracy values.

Table 2: Analysis and Evaluation of Algorithm Accuracyon Fer2013 Dataset

S.no.	Title	Author	Method	Accuracy	Gap
1	Smart music player integrating facial emotion recognition and music mood recommendation	Shlok Gilda, Husain Zafar, Chintan Soni, Kshitija Waghurdekar	convolutional neural networks (CNNs)	90.23 %	No user preferences
2	Facial Expression Recognition via Deep Learning	Abir Fathallah, Lotfi Abdi, Ali Douik	Visual Geometry Group model (VGG) + CNNs	93.33 %	complexity of human expression variations.
3	Music Genre Classification and Recommendation by Using Machine Learning Techniques	Ahmet Elbir, Hilmi Bilal Çam, Mehmet Emre Iyican, Berkay Öztürk, Nizamettin Aydin	Support Vector Machine (SVM)	72.60 %	No collaborative filtering
4	Face Expression Recognition Using CNN & LBP	Rahul Ravi; S.V Yadhukrishna; Rajalakshmi prithviraj	LBP and CNN	80.30 %	challenge of accurately recognizing facial expressions
5	**Tune into Emotions: A Study of Musical Therapy's Influence on Facial Expression Recognition**	Ashwini Kumar, Shashwat Mishra, Dravyanshu Singh,Sampada Huria	FER Dataset and CNN	85 %	Challenge of more accuracy

Table 3 outlines the hyperparameters employed in training the CNN network. These parameters includebatch size, class quantity, optimizer, learning rate, number of epochs, number of layers, activation functions (Relu, SoftMax), and the loss function (Categorical-cross-entropy).

Class	Test	Train	Total
Angry	958	3995	4953
Disgust	111	436	547
Fear	1024	4097	5121
Happy	1774	7215	8989
Neutral	1233	4965	6198
Sad	1247	4830	6077
Surprised	831	3171	4002

Table 3: Hyperparameters of Training the CNN Network

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

These hyperparameters provide crucial insights into theconfiguration and tuning of the CNN model, showcasing the specifications that contribute to its improved performance. The details encompass batch processing, optimization methods, activation functions, and more, which collectively play a pivotal role in the model's successful training and accurate prediction

VII. CONCLUSION

Through an extensive review of existing literature, it is evident that numerous methodologies for leveraging visual cues in music have been explored by researchers and developers, providing a foundation for our work. The goals of our system have been delineated based on these insights, aligning with the evolving capabilities ofAI applications. The system we present focuses on elucidating the intricate connection between music anduser mood, aiming to guide users in selecting music that positively influences their emotional state. By employing a practical system capable of capturing user emotions such as happiness, sadness, anger, neutrality,or confusion, we streamline the process of tailoring music recommendations.

Acknowledging the resource-intensive nature of processing large files, our approach emphasizes cost- effectiveness and adherence to standard equipment, avoiding unnecessary complexity. The motivation behind our facial recognition-based music recommendation system is to simplify the user's task of creating and managing playlists. By integrating facial recognition technology, our system offers a user-friendly and efficient solution to enhance the overall music listening experience, aligning with the advancinglandscape of AI applications.

VIII. FUTURE SCOPE

While the system currently works well, there is still room for future improvements to improve the overall customization and user experience. One way to improve is to better customize the app to personal preferences and allow users to set everything more personally and ideal. In addition, exploring other ways of knowing emotions, including excluding emotions such as hate and fear , can help provide greater understanding again from the user's perspective, increasing the accuracy of the system. In addition, future developments will go beyond the aesthetic view and contribute to music therapy. By improving the body's ability, it can become an important tool for music therapists to treat patients with stress, anxiety, pressure and injury challenges. The widespread use of this system has the potential to benefit human health thanks to the healing power of music. Problems caused by camera poor quality provide opportunities for improvement. The aim of these decisions for future development is to make the system versatile, relevant and adaptable to various situations and user needs.

REFERENCES

[1] Ramya Ramanathan, Radha Kumaran, Ram RohanR, Rajat Gupta, and Vishalakshi Prabhu, an intelligent music player based on emo-tion recognition, 2nd IEEEInternational Conference on Computational Systems and Information Technology for Sustainable Solutions2017. https://doi.org/10.1109/CSITSS.2017.8447743

[2] Shlok Gilda, Husain Zafar, Chintan Soni, Kshitija Waghurdekar, Smart music player integrating facial emotion recognition and music mood recommendation, Department of Computer Engineering, Pune Institute ofComputer Technology, Pune, India, (IEEE),2017. https://doi.org/10.1109/WiSPNET.2017.8299738

[3] Deger Ayata, Yusuf Yaslan, and Mustafa E. Kamasak, Emotion-based music recommendation system using wearable physiological M. Athavle et al. ISSN (Online) : 2582-7006 International Conference on Artificial Intelligence (ICAI-2021) 11 Journal of Informatics Electrical and ElectronicsEngineering (JIEEE) A2Z Journals sensors, IEEE transactions on consumer electronics,vol. 14, no. 8, May 2018. https://doi.org/10.1109/TCE.2018.2844736

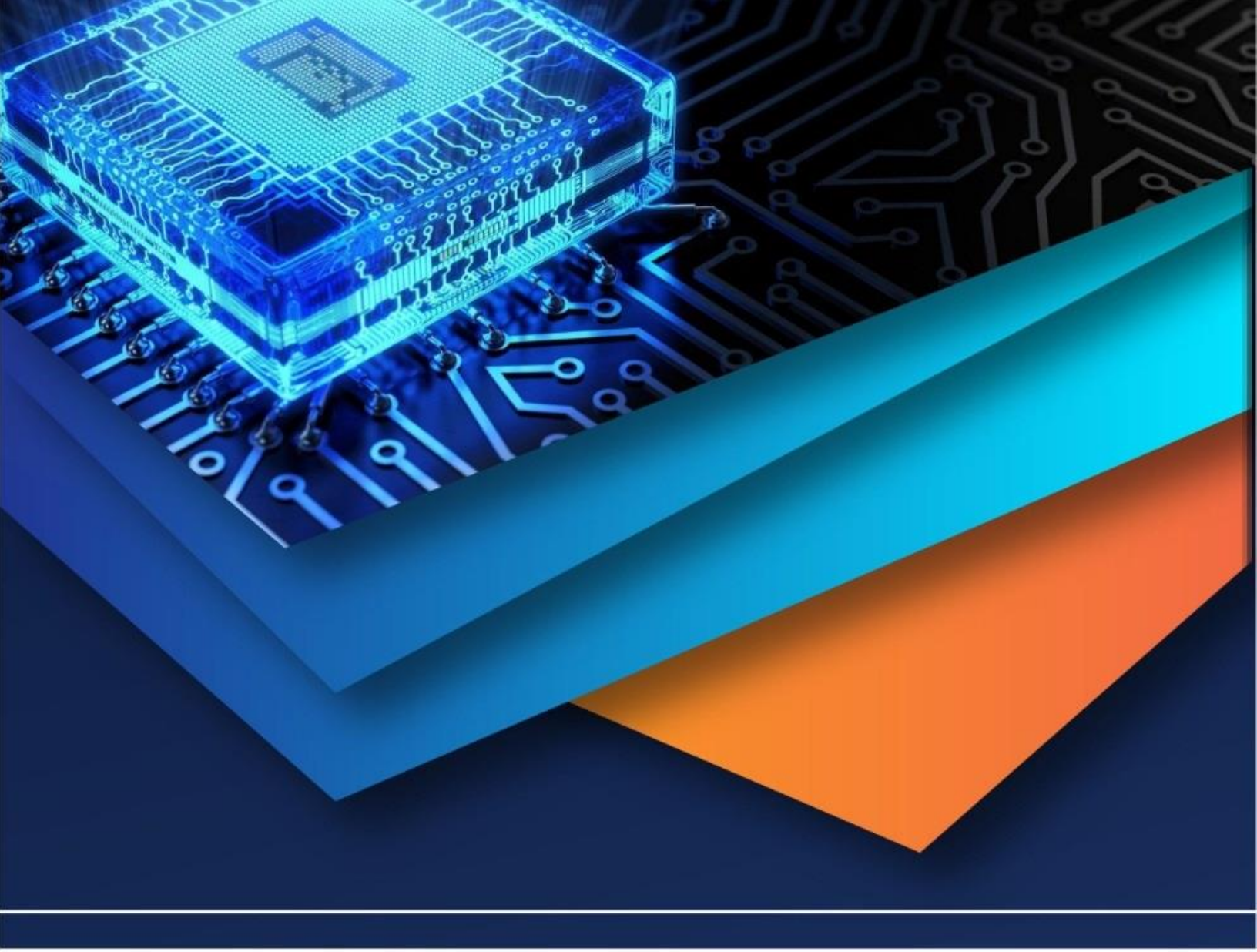
[4] Ahlam Alrihail, Alaa Alsaedi, Kholood Albalawi,Liyakathunisa Syed, Music recommender system forusers based on emotion detection through facial features, Department of ComputerScience Taibah University, (DeSE), 2019. https://doi.org/10.1109/DeSE.2019.00188

[5] Research Prediction Competition, Challenges in representation learning: facial expression recognition challenges, Learn facial expression from an image, (KAGGLE).

[6] Preema J.S, Rajashree, Sahana M, Savitri H, Review on facial expression-based music player,International Journal of Engineering Re-search & Technology (IJERT), ISSN-2270181,Volume 6, Issue 15, 2018.

[7] AYUSH Guidel, Birat Sapkota, Krishna Sapkota, Music recommendation by facial analysis, February 17,2020.

[8] CH. sadhvika, Gutta.Abigna, P. Srinivas reddy, Emotion-based music recommendation system, Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad; InternationalJournal of Emerging Technologies and Innovative Research (JETIR) Volume 7, Is-sue 4, April 2020.



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)