iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IX **Month of publication:** September 2024

DOI: https://doi.org/10.22214/ijraset.2024.64140

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 IX Sep 2024- Available at www.ijraset.com

Emotion Based Music Player through Face Recognition System

Miss. Harsha Varyani[1], Prof. R. B. Late[2], Prof. N. G. Dharashive[3]
M. S. Bidve Engineering College, Latur, Maharashtra, India

Abstract: ***The goal of this work is to build an emotion detect system which can analyze basic facial expression of human. In this project a method is presented for mood detection based on humans face emotions. The proposed method used the humans face to identify the mood of that human and finally using this result play the audio file which related to human's emotion. Firstly system takes the human face as input then the further process will going on. Face detection and eye detection is carried out. After that using feature extraction technique to recognize the human face. This method helps to recognize the human's emotion using feature of face image. Through the feature extraction of lip, mouth, and eyes, eyebrow, those feature points are found. If the input face wills matches exactly to the emotions base datasets face then we can identify the humans exact emotion to play the emotion related audio file also we will fetch the news data based on user preferences using API. Recognition under different environmental conditions can be achieved by training on limited number of characteristics faces. The proposed approach is simple, efficient, and accurate. System play's very important role in recognition and detection related field.***
Keywords: Face Detection, Feature Extraction, Face Emotion.

I. INTRODUCTION

Face detection and identification are crucial aspects of human-computer interaction, involving the challenging task of distinguishing and recognizing facial features. This process is essential for any recognition algorithm, as the detection rate directly impacts the recognition stage. Detecting and localizing faces in images, especially amidst noise, is a complex task that requires precise algorithms to ensure accuracy. Mood detection based on facial emotions is a current topic of interest, addressing various challenges such as varying poses, lighting, and expressions in facial images. Despite advancements, face emotion detection remains challenging due to factors like pose variation and illumination changes. The goal is to use facial images to determine human mood and subsequently play corresponding audio files, utilizing face recognition techniques to match input images with trained data. The proposed approach is simple, efficient, and accurate. This system gives accurate result as compare to existing approach. System play's very important role in recognition and detection related field.

A. Background

Emotional aspects have more impact on Social intelligence like communication understanding, decision making and also helps in understanding behavioral attitude of human. Emotion play important role during communication. Emotion recognition is implemented out in diverse way; it may be verbal or non-verbal. Voice (Audible) is verbal way of communication & Facial expression, action, body postures and gesture is non-verbal form of communication. Human can recognize emotions without any meaningful delay and effort but recognition of facial expression by machine is a big challenge.

B. Aim and Objectives

This paper works to find the mood of human using facial emotion, comparative study of popular face expression recognition techniques & phases of facial expression. So this work gives the brief introduction towards techniques, application and challenges of emotion recognition system.

Objective of proposed system:

1) To find the emotion of human using facial features
2) To identify the face using training feature data
3) To study the human face extraction methods
4) To fetch the online news data based on user preference

©IJRASET: All Rights are Reserved

C. Motivation

With the help of this research we can identify the human mood using human emotions. So this project is as important to individual as much to public too.

II. LITERATURE SURVEY

The work presents the study of various famous and unique techniques used for facial feature extraction and emotion classification. Various algorithms of facial expressions research are compared over the performance parameters like recognition accuracy, number of emotions found, Database used for experimentation, classifier used etc [1].

This work presents a system for automatic facial expression recognition and emotion classification using the `Viola Jones Face Detection' technique for face localization. Feature vectors are optimized with a subset feature selection technique, and the combined features are classified using SVM, Random Forest, and KNN classifiers. [2].

The proposed technique use three steps face detection using Haar cascade, features extraction using Active shape Model (ASM) and Ada-boost classifier technique for classification of five emotions anger, disgust, happiness, neutral and surprise [3].

In this work implement an efficient technique to create face and emotion feature database and then this will be used for face and emotion recognition of the person. For detecting face from the input image we are using Viola-Jones face detection technique and to evaluate the face and emotion detection KNN classifier technique is used [4].

This paper objective is to display needs and applications of facial expression recognition. Between Verbal & Non-Verbal form of communication facial expression is form of non-verbal connection but it plays pivotal role. It expresses human related or filling & his or her mental situation [5].

In this proposed system it is attention on the human face for recognizing expression. Many techniques are available to recognize the face image. This technique can be adapted to real time system very easily. The system briefly displays the schemes of capturing the image from web cam, detecting the face, processing the image to recognize few results [6].

In this work, adopt the recently introduced SIFT flow technique to register every frame with respect to an Avatar reference face model. Then, an iterative technique is used not only to super-resolve the EAI representation for each video and the Avatar reference, but also to improve the recognition performance. Also extract the features from EAIs using both Local Binary Pattern (LBP) technique and Local Phase Quantization (LPQ) technique [7].

In this study, a frame of emotion recognition system is developed, including face detection, feature extraction and facial expression classification. In part of face detection, a skin detection process is support first to pick up the facial region from a complicated background. Through the feature detection of lip, mouth, and eyes, eyebrow, those feature points are launch [8].

This work introduces a new facial emotion recognition technique using Haar transform and adaptive Ada-boost for face identification, and PCA with a minimum distance classifier for face recognition. It explores two methods for facial expression recognition: one using PCA and KNN, and the other using NMF and KNN.[9].

The paper presents a Lightweight Emotion Recognition (LER) model that addresses the limitations of traditional models by using a densely connected convolution layer and advanced compression techniques to reduce parameters without losing accuracy. It also incorporates multichannel input preprocessing to improve feature extraction from facial images.[10]

The thesis reviews various methods for extracting facial expressions and emotional grading, comparing algorithms based on precision, emotional range, databases, and classification techniques. One highlighted method uses the Viola-Jones algorithm for face detection and geometric and appearance-based methods for feature extraction, with Support Vector Machines (SVM) for classifying emotions like happiness, sadness, anger, and surprise. This approach is praised for its accuracy and efficiency in different conditions.[11].

III. PROPOSED SYSTEM

The proposed method detects human emotions from facial images and uses these results to play corresponding audio files. The system first captures a face image, then performs face and eye detection. Feature extraction techniques identify facial features like lips, mouth, eyes, and eyebrows to determine emotions. If the input face matches the dataset's emotions, the system plays the related audio file. Training on a limited number of characteristic faces allows detection under various environmental conditions.

A. Advantages Proposed System

1) Detection the mood of human using human face image.
2) Detect eye of human face image
3) Face image emotions identification using image features will be achieved

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

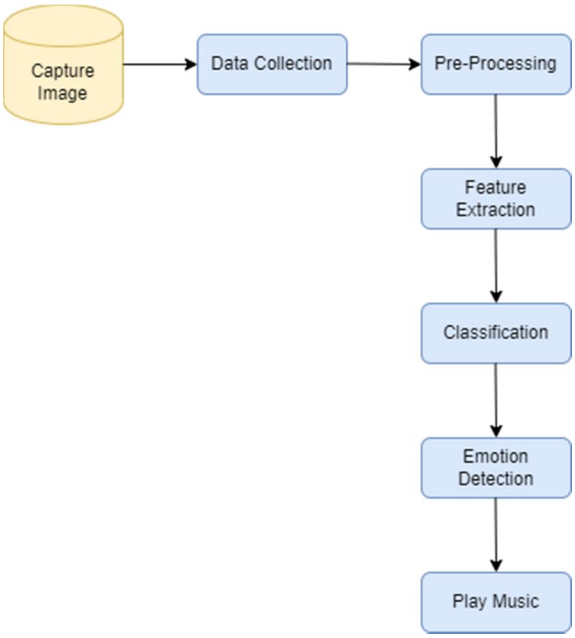


Fig 1. Proposed system architecture

IV. ALGORITHM

A. Haar Cascade Algorithm

Haar Cascade Algorithm is a machine learning based algorithm proposed by Paul Viola and Michael Jones in which the cascade image is trained by providing a lot of positive and negative images,that is used to detect the object in images. This algorithm needs a lot of positive images (images of faces) and negative images i.e. images without faces to train the classifier. Haar features are used to extract the features from images. First set of two rectangle features is responsible for finding out the edge, Second set of 3 rectangle features is responsible for finding out if there is a lighter region surrounded by the region and the same if implemented conversely.
Third set of 4 rectangle features is responsible for finding out the change of pixel intensity across diagonal. Every feature has a single value which is obtained by subtracting the sum of the pixel under the white rectangle from the sum of the pixel under the black rectangle. All possible locations of every kernel are used to calculate the feature. To calculate each feature we need to find sum of the pixel under the white and black rectangle for solving these integral images are introduced. It makes the calculation of the sum of the pixels simple.

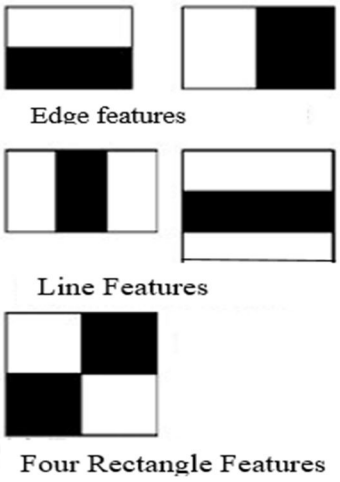


Fig 2.Haar Features

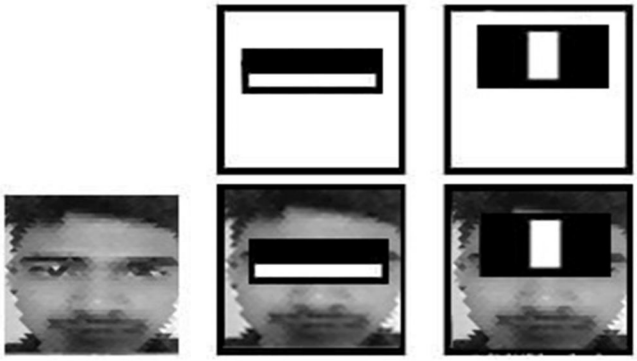


Fig -3: Haar Features

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

Among all the features calculated, many features are irrelevant. If we take the following example first feature focuses on property that describes a region of the eye that is often darker than the region of nose and cheeks. The second feature focuses on the eye region is darker than the bridge of the nose ,but if the same window when applied on cheeks or any other place is irrelevant, so to minimize such irrelevance AdaBoost is used. To minimize the irrelevance in a feature ,we apply every feature on every training image. For each feature, we get the best threshold value which discriminates faces into positive and negative. We select the feature which has a minimum error rate. The final classifier is the weighted sum of these different weak classifiers .It is called weak because it alone is not able to classify an image but together with others, they form a strong classifier. The final setup has 6000 approximate features. But implementation of 6000 features is a time-consuming process, so the further process is carried out. In an image, most of the part is non-face part.So we check if a window has a face region. If it is not then it gets discarded and no further process will be carried out on that region.So the possibility of finding face increases. Cascade classifier is used for this, instead of implementing 6000 features on a window. A group of features are applied step by step. If window fails at first stage window gets discarded. If it is passed then second stage of features is applied and it continues the process, and the face gets detected.

B. Convolutional Neural Networks (CNNs) and DenseNet Architecture

Convolutional Neural Networks (CNNs) are deep learning algorithms used mainly for image recognition and classification. They automatically learn spatial hierarchies of features from images using convolutional layers, pooling layers, and fully connected layers.

a) *Convolutional Layers:* These layers apply a set of filters (kernels) to the input image, performing convolution operations to extract features such as edges, textures, and patterns. Each filter produces a feature map, highlighting different aspects of the input image.

b) *Pooling Layers:* Pooling layers reduce the spatial dimensions of the feature maps, typically using operations like max pooling or average pooling. This down-sampling helps in reducing the computational complexity and controlling overfitting.

c) *Fully Connected Layers:* After several convolutional and pooling layers, the high-level reasoning in the network is done via fully connected layers. These layers take the flattened feature maps and output the final classification scores.

1) DenseNet Architecture

DenseNet, short for Densely Connected Convolutional Network, introduces a novel connectivity pattern within CNNs. Unlike traditional CNNs, where each layer has a single connection to the next layer, DenseNet connects each layer to every other layer in a feed-forward manner. This dense connectivity pattern enhances information flow and gradient propagation throughout the network.

a) *Dense Blocks:* A Dense Block consists of multiple convolutional layers, where each layer receives the concatenated feature maps of all preceding layers as input. This dense connectivity ensures maximum information flow between layers, improving learning efficiency and mitigating the vanishing gradient problem.

b) *Transition Blocks:* Between dense blocks, transition blocks are used to reduce the dimensions of the feature maps. These blocks typically include Batch Normalization, ReLU activation, a 1x1 convolutional layer for dimensional reduction, and an AveragePooling2D layer for down-sampling.

c) *Growth Rate:* The growth rate in DenseNet determines the number of filters added per layer within a dense block. A typical growth rate is set at 32, meaning each layer adds 32 filters to the feature maps.

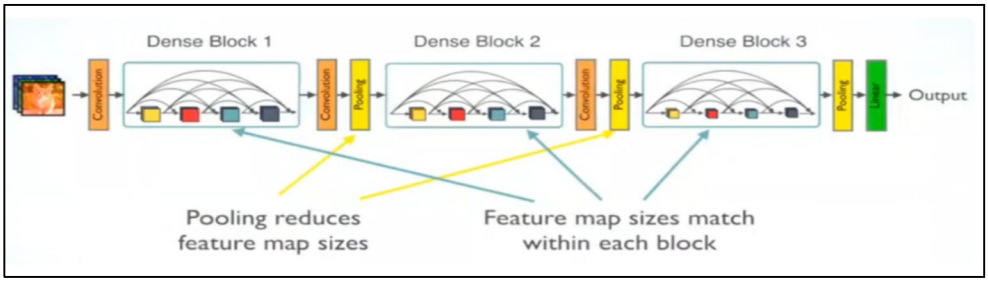


Fig.3.8.DenseNet Architecture

d) *Global Average Pooling and Classification:* After the final dense block, a Global Average Pooling layer is used to reduce each feature map to a single value. This is followed by a fully connected layer with softmax activation for classification. In the proposed architecture, the final output layer is designed to classify inputs into 8 different emotion categories.

2) *Detailed Architecture of the Proposed DenseNet*

a) *Dense Block 1:* Comprises 6 convolutional layers, each followed by Batch Normalization and ReLU activation.
b) *Transition Block 1:* Includes Batch Normalization, ReLU activation, a 1x1 convolutional layer, and an AveragePooling2D layer to reduce dimensions.
c) *Dense Block 2:* Contains 12 convolutional layers.
d) *Transition Block 2:* Similar to Transition Block 1, it reduces dimensions.
e) *Dense Block 3:* Consists of 24 convolutional layers.
f) *Transition Block 3:* Reduces dimensions before the final dense block.
g) *Dense Block 4:* Comprises 16 convolutional layers.
h) *Global Average Pooling:* Reduces each feature map to a single value.
i) *Fully Connected Layer:* With softmax activation for classification into 7 emotion categories.

C. HOG Algorithm

Face recognition involves two main tasks: verification, where a person's face is compared to a database for access, and identification, where a face is matched against multiple faces in the database. The Histogram of Oriented Gradients (HOG) method enhances this process by dividing an image into small cells (typically 8x8 pixels) to capture local details, which are essential for identifying unique features and analyzing the overall shape and structure of a person. Once the image is divided into cells, the next step is to compute the gradient orientation for each pixel within a cell. This involves calculating the direction of the greatest rate of change in intensity for each pixel. An M-bins histogram is then used to accumulate these orientations across all pixels in the cell.

This histogram represents the distribution of gradient orientations within the cell, capturing the local edge directions. These edge directions are essential for understanding the contours and outlines of the person in the image. Each cell's histogram is then created, representing the distribution of gradient orientations within that cell. These histograms are crucial as they capture the local edge directions, which are indicative of the shape and structure of the person. To ensure that the process is robust to changes in illumination and shadowing, the local contrast in overlapping blocks of cells is normalized. This normalization step adjusts the histograms to account for variations in lighting, making the HOG process more reliable under different lighting conditions.

After normalization, all the cell histograms are concatenated into a single vector, forming the final HOG feature vector. This vector is a comprehensive representation of the overall edge distribution and shape information of the person in the image. It combines the local details captured in each cell into a unified descriptor that can be used for further analysis or object detection tasks.When we get the intermediate image next step is extraction and creation of histogram. We use grid x and grid y to divide the obtained image into multiple grids. After this process ,we can extract the histogram of each region .Our image is in greyscale every histogram i.e. the histogram in every grid will contain around 256 positions which represents the intensity of the pixel.

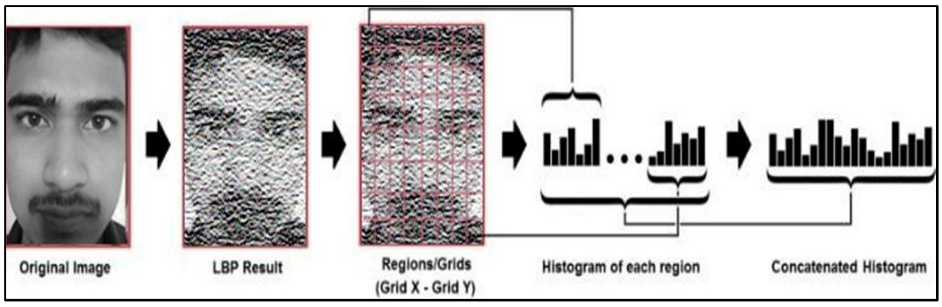


Fig. 3.9 HOG

After getting the histogram of all the grids, we need to concatenate them and a final histogram is get created .this final histogram gives characteristics of the original image .now the algorithm is trained and histograms are created. When we give an input image, we perform the same steps and create the histogram of an input image, and if we want to find if the face is present in the dataset we just compare the input image's histogram with a histogram of images in the dataset. We can use chi-square, Euclidian distance, the absolute value approaches to compare the histogram. We use threshold and the 'confidence' to estimate if the algorithm has correctly recognized the image. We can say that the algorithm has successfully recognized if the confidence is lower than the threshold defined.

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

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

V. MATHEMATICAL MODEL

The color distribution information can be captured by the low-order moments, using only the first three moments: mean, variance and skewness, it is found that these moments give a good approximation and have been proven to be efficient and effective in representing the color distribution of. These first three moments are defined as:

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} P_{ij}$$

$$\sigma_i = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_{ij} - \mu_i)^2}$$

$$S_i = \left[\frac{1}{N}\sum_{j=1}^{N}(P_{ij} - \mu_i)^3\right]^{\frac{1}{3}}$$

Where, Pij is the value of the i^{th} color channel of the j^{th} image pixel. Only 3 x 3 (three moments for each color component) matrices to represent the color content of each image are needed which is a compact representation compared to other color features.

Step1: Smooth the image with a Gaussian filter to reduce noise and unwanted details and textures.

$$g(m,n) = G_\sigma(m, n) * f(m, n)$$

Where

$$G_\sigma = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{m^2+n^2}{2\sigma^2}\right)$$

Step2: Compute gradient of g (m, n) using any of the gradient operations (Roberts, Sobel, Prewitt, etc) to get:

$$M(n, n) = \sqrt{g_m^2(m,n) + g_n^2(m,n)}$$

And

$$\boldsymbol{\theta}(m, n) = tan^{-1}\left[g_n(m, n) / g_m(m, n) /\right]$$

Step3: Threshold M:

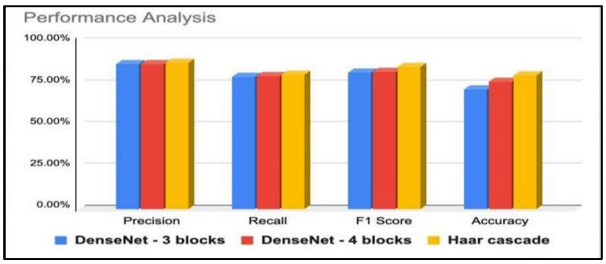
$$M_T(m, n) = \begin{cases} M(m,n) \; if \; M(m,n) > T \\ 0 \qquad\qquad Otherwise \end{cases}$$

According to co-occurrence matrix, there are several textural features measured from the probability matrix to extract the characteristics of texture statistics of remote sensing images. Correlation measures the linear dependency of grey levels of neighboring pixels.

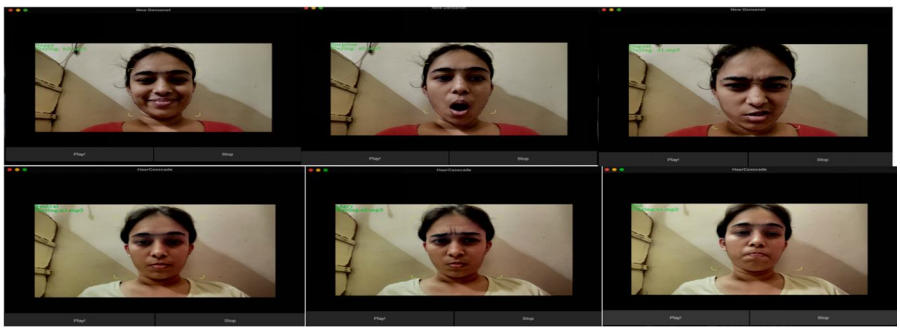
$$\text{Correlation} = \frac{\sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1}(i,j)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

VI. COMPARISON

Technique	DenseNet-3	DenseNet-4	Haar cascade
Precision	86.83%	86.86%	87.85%
Recall	79.45%	79.86%	80.63%
F1 Score	81.90%	82.14%	85.21%
Accuracy	71.73%	76.14%	79.98%



VII. RESULT



©IJRASET: All Rights are Reserved

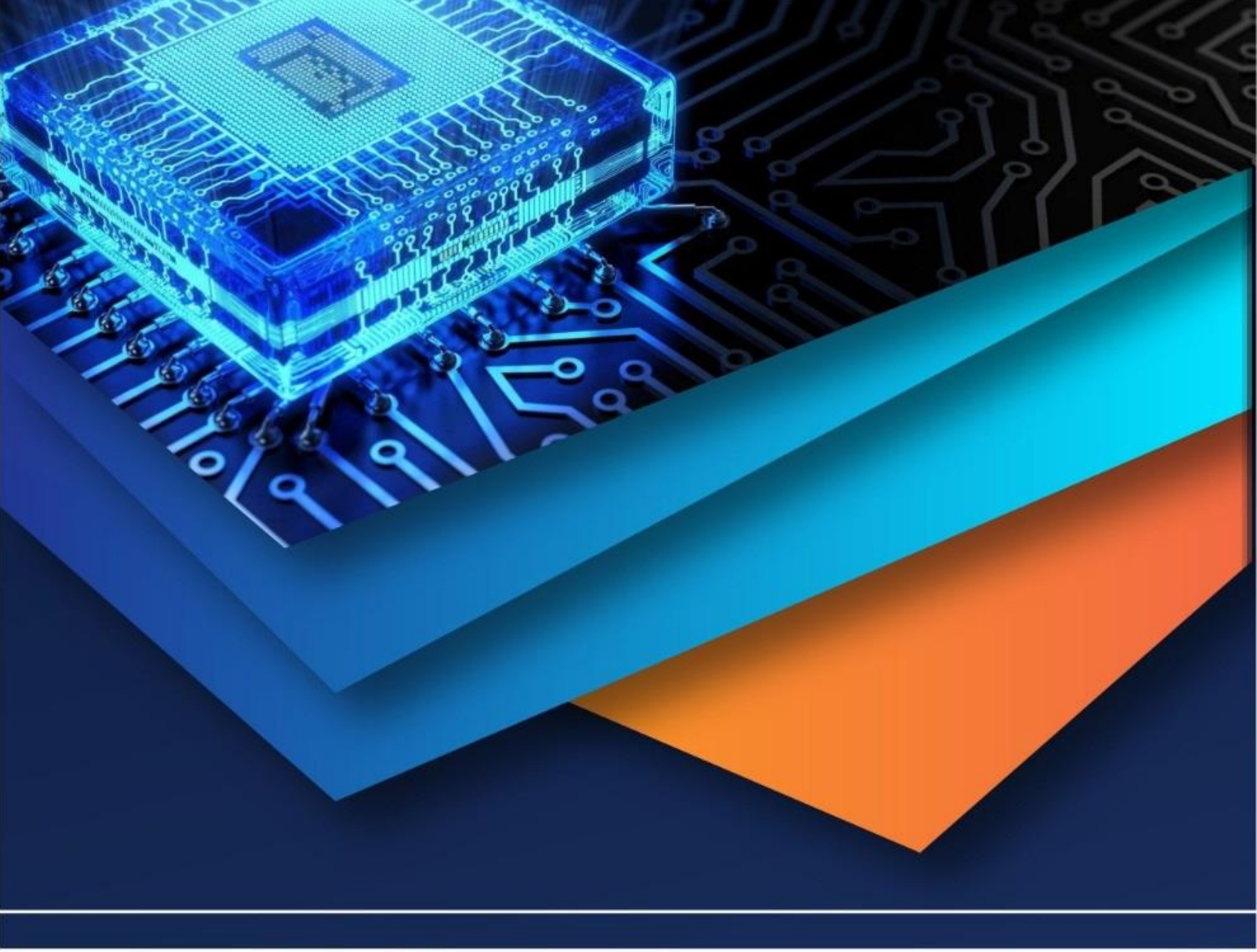
VIII. CONCLUSION

This work proposes a method for detecting human emotions from facial images by modifying an algorithm to extract and match features with a training dataset. Additionally, it uses these results to play audio files and fetch news data based on user preferences via an API. The system aims to improve recognition accuracy and processing time for large face databases, highlighting its potential role in emotion detection applications.

REFERENCES

[1] Bharati Dixit, Arun Gaikwad, "Facial Features Based Emotion Recognition". ISSN (e): 2250-3021, ISSN (p): 2278-8719 Vol. 08, Issue 8 (August. 2018)
[2] J Jayalekshmi, Tessy Mathew,"Facial expression recognition and emotion classification system for sentiment analysis". 2017 International Conference.
[3] Suchitra, Suja P.Shikha Tripathi, "Real-time emotion recognition from facial images using Raspberry Pi II". 2016 3rd International Conference
[4] Dolly Reney, Neeta Tripathi, "An Efficient Method to Face and Emotion Detection". 2015 Fifth International Conference.
[5] Monika Dubey, Prof. Lokesh Singh**, "**Automatic Emotion Recognition Using Facial Expression: A Review**".** International Research Journal of Engineering and Technology (IRJET) Feb-2016.
[6] Anuradha Savadi Chandrakala V Patil**, "**Face Based Automatic Human Emotion Recognition**".** International Journal of Computer Science and Network Security, VOL.14 No.7, July 2014.
[7] Songfan Yang, Bir Bhanu,"Facial expression recognition using emotion avatar image"**.** 2011 IEEE International Conference.
[8] Leh Luoh**,** Chih-Chang Huang, Hsueh-Yen Liu, "Image processing based emotion recognition"**.** 2010 International Conference.
[9] Jiequan Li**,** M. Oussalah**,** "Automatic face emotion recognition system"**.** 2010 IEEE 9th International Conference.
[10] GUANGZHE ZHAO,HANTING YANG, MIN YU, "Expression Recognition Method Based on a Lightweight Convolutional Neural Network".10.1109/ACCESS.2020.2964752, IEEE Access
[11] Bharati Dixit, Arun Gaikwad, "Facial Features Based Emotion Recognition". ISSN (e): 2250-3021, ISSN (p): 2278-8719 Vol. 08, Issue 8 (August. 2018)

©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)