iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52372

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 11 Issue V May 2023- Available at www.ijraset.com

Hybrid Features Extraction for Emotion Detection using Deep Convolutional Neural Network

Manju Rathod[1], Prof. Jitendra Musale[2], Yadnesh Haribhakta[3], Mayuri Koli[4], Sharwari Surve[5]
[1, 2, 3, 4, 5]*Department of Computer Engineering, Anantrao Pawar College of Engineering and Research, Pune, India*

Abstract: ***These Human facial expressions communicate much information more visually than verbally. Recognition of facial expressions is essential for human-machine interaction. Applications for automatic facial expression recognition include but are not limited to, comprehending human behaviour, spotting mental illnesses, and creating artificial human emotions. Human-computer interaction, autonomous cars, and a wide range of multimedia applications all rely heavily on facial expression detection. In this study, we present a reusable architecture for recognizing human emotions based on facial expressions. Two machine learning algorithms that may be learned in advance for use in real-time applications make up the framework. At first, we use AdaBoost cascade classifiers to find instances of faces in the photos. The extracted neighborhood difference characteristics then serve as a representation of a face's features according to its localized appearance data. Instead of only looking at intensity data, the NDF predicts multiple patterns depending on the connections between surrounding areas. Despite a high identification rate, facial expression recognition by computers remains difficult. Based on geometry and appearance, two widely used approaches for automated FER systems are often used in the literature. Pre-processing, face detection, feature extraction, and expression classification are the four steps that typically make up facial expression recognition. Various deep learning techniques (convolutional neural networks) were used in this effort to recognize the primary seven human emotions: anger, fear, happiness, surprise, and neutrality.***

Keywords: Face recognition, Feature extraction and selection, CNN, classification, emotion detection, sentiment classification

I. INTRODUCTION

They detect facial moods based on time series picture inputs. Calculate your mood depending on your score or weight using the class label. Successfully apply the supervised learning technique to the test model based on the training data. Use the suggested system's highest level of accuracy. They could turn that student's negative mindset into a good one. Our aspirations increased dramatically with the development of contemporary technology, which has no limitations. A lot of research is being done nowadays in the area of digital images and image processing. The rate of development has been exponential and is constantly rising. In the modern world, image processing research spans various disciplines and has several applications. Signal processing in the area of image processing uses pictures as both input and output signals. Facial expression identification is among the most critical uses of image processing. Our facial expressions communicate our emotions. In interpersonal communication, facial expressions are crucial. Our faces communicate our emotions via nonverbal scientific gestures called facial expressions. The generation needs automatic facial expression detection, essential for robotics and artificial intelligence. This project aims to create an automated facial expression recognition system that can recognise and categorise human face photographs with expressions into seven distinct classes, such as neutral, angry, fearful, happy, and surprised.

II. LITERATURE SURVEY

According to [1] the technique for gauging the learners' mental state during an online learning session. Confusion, discontent, contentment, and frustration are four complicated emotions we have recognised as combinations of fundamental emotions. We have also shown that more is needed to judge emotion based on a single picture capture. To evaluate the learner's mental state, we considered a window of six picture frames taken by the camera. Considering six picture frames is also influenced by the idea that human emotions are continuous and not particularly discrete, i.e., they do not change suddenly but require some time (albeit not much) to change.

According to [2] The machine learning technique uses specific significant extracted characteristics to model the face. Therefore, it will only be able to recognize emotions with a high degree of accuracy if the features are hand-engineered and rely on past knowledge. In this study, convolutional neural networks (CNN) were created to identify facial expressions of emotion.

©IJRASET: All Rights are Reserved

The importance of facial expressions in nonverbal communication may be seen in how they represent a person's inner emotions. Numerous studies have been conducted on computer modelling of human emotion by various types of researchers. Nevertheless, the human visual system still needs to catch up. In this study, a face's eye and lip regions were first identified using the Viola-Jones technique and afterwards with the aid of a neural network. Additionally, neural network methods, deep learning models, and machine learning approaches are used for emotion identification.

According to [3] an overview of the state of human emotional stress biomarkers and the primary prospective biomarkers for wearable affective system sensors in the future are presented. Emotional stress has been identified as a significant factor in societal issues, including quality of life, crime, health, and the economy. While electroencephalography, physiological parameter techniques, and blood cortisol testing are the gold standards for assessing stress, they are often intrusive or uncomfortable. They are not appropriate for wearable real-time stress monitoring. Instead, cortisol in biofluids and VOCs released from the skin are sensible and helpful signals for sensors to identify emotional stress episodes.

According to [4] graph-based technique for recognising emotions that uses facial landmarks. Several pre-processing stages were implemented based on the suggested strategy. Face key points must undergo pre-processing before face expression characteristics can be retrieved. Face detection using the Haar-Cascade method, landmark implementation using a media-pipe face mesh model, and model training on seven emotional classes are the three primary processes in recognising emotions on masked faces. For model training, the FER-2013 dataset was used. For unmasked looks, an emotion detection model was created. The top half of the face was then marked with markers. After retrieving landmark locations and faces, we recorded emotional class landmark coordinates and exported them to a comma-separated values (CSV) file. After that, the dynamic classes received model weights. Last, a web camera application evaluated a landmark-based emotion identification model for the upper face regions on photos and in real time.

According to [5] the FER2013 dataset's single network with the best classification accuracy. Adopting the VGGNet architecture, we carefully adjust its hyperparameters and test several optimisation techniques. Recognising facial expressions that communicate fundamental emotions like fear, pleasure, contempt, etc., is known as facial emotion recognition. It is helpful in human-computer interactions and may be used in customer satisfaction surveys, online gaming, digital advertising, and healthcare.

According to [6] A face expression identification approach based on a convolutional neural network (CNN) and image edge detection is suggested to circumvent the laborious procedure of explicit feature extraction in conventional facial expression recognition. The edge of each picture layer is retrieved during the convolution process after the facial expression image has been normalised. The maintain the edge structure information of the texture picture, the extracted edge information is placed on each feature image. The maximum pooling approach is then used to reduce the dimensionality of the retrieved implicit features. Finally, a Softmax classifier is used to classify and identify the test sample image's expression by carefully combining the Fer-2013 facial expression database with the LFW data set; a simulation experiment is created to test the resilience of this approach for facial expression identification against a complicated context.

According to [7] A language learner's conversational ability may be assessed using multimodal activities, including speech content, prosody, and visual cues. Although linguistic and auditory components have been well studied, less attention has been paid to visual factors, including eye contact and facial expressions. We first built a dataset of 210 Japanese English learners' online video interviews with comments on their speaking skills to develop an automated speaking proficiency evaluation system using multimodal traits.

According to [8] Multi-block deep learning was used in a self-management interview programme to discern user emotions. The multi-block deep learning technology helps the user learn by sampling the major facial areas (eyes, nose, mouth, and so on), which are essential for emotion analysis from face recognition. It is in contrast to the basic framework for learning about whole-face photographs. The multi-block technique is used to sample using multiple AdaBoost learning. This process also includes similarity evaluation for efficient block image screening and verification. The suggested model's performance is compared to AlexNet, previously mainly used for facial recognition. The specified area's identification rate and extraction time are compared as comparison items.

According to [9] the core concept of the automated attendance system is to employ facial recognition more conveniently than other biometric systems. The following three approaches—Viola-Jones (Haar Cascades), Histogram Oriented Gradients (HOG), and Convolution Neural Network (CNN)—are evaluated based on face detection accuracy. In terms of accuracy, the Viola-Jones method did better than the others. Face detection and identification are performed in real-time attendance systems using the Viola-Jones and CNN algorithms, respectively. The benefit of the offered method is that it may overcome obstructions, including alignments, very apparent faces, and poor lighting.

According to [10] a video-based attendance system is made using real-time face recognition. The system offers simultaneous multi-user attendance and facial liveness identification. Facial data may be automatically collected by the technology and stored in the database with attendance information. The FaceNet algorithm is the foundation for the system's face recognition component. In contrast, the MTCNN (Multitask Convolutional Neural Network) technique is the foundation for the system's face detection component. The TensorFlow framework is used to develop the method's face liveness detection component based on the ERT (Ensemble of Regression Trees) algorithm and can detect user blinking. The attendance system uses Python, while the user interface is made with the Qt library.

III. PROPOSED SYSTEM

The system design that uses the face expression recognition detection administration optimization algorithms also deals more explicitly with the public approach. Systems use cameras to identify students' faces, which are stored on hard drives. Utilizing feature extraction and selection techniques, the CNN framework conducts training modules and stores each feature in the tainting database. To identify the face in the input picture that has been provided, then to take a frame for object recognition. The test feature uses the entire training dataset to produce similarity weights for each item and then recommends the real student id based on the consequences obtained. The system automatically updates the attendance for each student based on the provided ID. Convolutional Neural Network (CNN) is used for both training and testing.

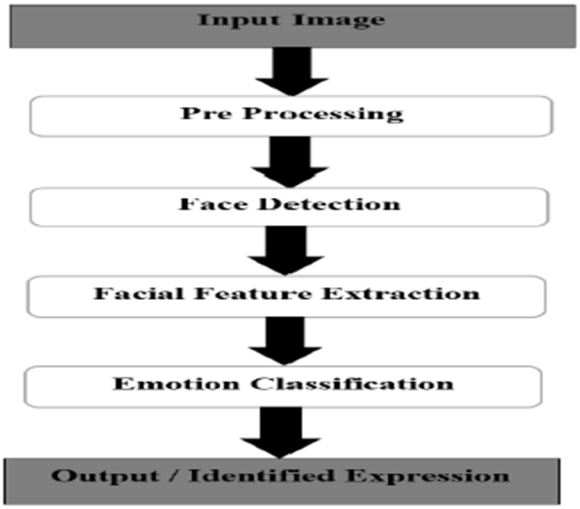


Figure 1: Proposed Architecture

IV. METHODOLOGY

1) *Module 1:* Image Acquisition: Images used for facial expression recognition are static images or image sequences. Images of face can be captured using camera.
2) *Module 2:* Face detection: Face Detection is useful in detection of facial image. Face Detection is carried out in training dataset using Haar classifier called Voila-Jones face detector and implemented through Opencv. Haar like features encodes the difference in average intensity in different parts of the image and consists of black and white connected rectangles in which the value of the feature is the difference of sum of pixel values in black and white regions.
3) *Module 3:* Feature Extraction: Selection of the feature vector is the most important part in a pattern classification problem. The image of face after pre-processing is then used for extracting the important features. The inherent problems related to image classification include the scale, pose, 17 translation and variations in illumination level
4) *Module 4:* Testing: Using any free train model using CNN classifier we achieved weight system recommend the max function presents itself as a probability for each emotion class.

In order to implement the 80-20 strategy, both the training dataset and the testing dataset have been split into 80% and 20% of the total data. This is done in order to determine whether or not the model is accurate when applied to data that it has never been exposed to previously. Because of this, we were able to evaluate whether or not we were over-fitting the training dataset and, if so, whether or not we needed to train for longer epochs and at a slower learning rate. If the accuracy rate was higher than the training accuracy, then we knew that we needed to train for longer epochs.

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

V. MATHEMATICAL MODEL

A mathematical model is a description of a system using mathematical concepts and language. A model may help to explain a system and to study the effects of different components of a system to predict the behavior of a system. The mathematical modeling for our system is as follows:

A. Training Process

Input: Training dataset Train-Data [], Many activation functions [], Threshold Th
Output: Extracted Features Feature set[] for a trained module that has been finished.
Step 1: Set the data input block d[], the activation function, and the epoch size.
Step 2: Features-pkl ← Feature-Extraction (d[])
Step 3: Feature-set [] ← optimized (Features-pkl)
Step 4: Return Feature-set []

B. Testing Process

Input: Extracted features of testing instances set Data [i.....n], Train data policies PSet[41].....T[n]
Output: Normal or attack.
Steps:

1) For each (Data [i] into Data) choose n attributes from Data [i] using below formula,

$$\text{Treeset}(k) = \sum_{k=1}^{n} \text{attribute}\,[D[i]k \ldots\ldots D[n]n]$$

2) For each (PSet [i] from PSet),

$$\text{Train}[m] = \sum_{m=1}^{n} \text{attribute}\,[T[i]k \ldots\ldots T[n]n]$$

3) Evaluate train and test instances using below formula,

$$\text{Treeset}[k].\text{weight} = \text{similarity}(\text{Treeset}[k] \sum_{m=1}^{n} \text{Trainset}\,[m]$$

4) If (Treeset[k]: weight > Th),
Treeset[k].class ← Train[m]: class
Break;

5) Return Treeset[k].class

VI. EXPERIMENT AND RESULTS

Intel i7 CPU 2.7 GHz has used with 16 GB Random Access Memory for execution. The RESENT (32,50, 101 and 152) version has used for experimental investigation of proposed systems including 5G network. The major factors has considered execution time (including data processing, data uploading and downloading etc.), memory consumption, network overhead and energy for valuation the efficiency of proposed systems.

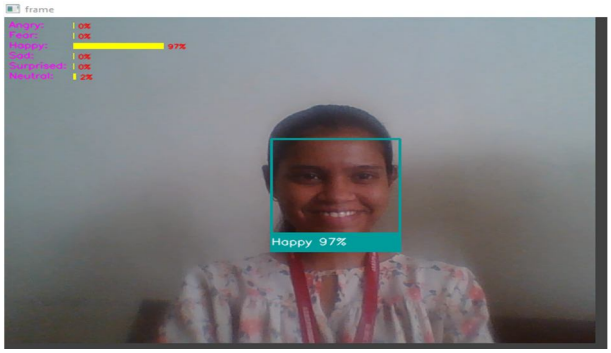


Figure 2: Image Testing experiment 1

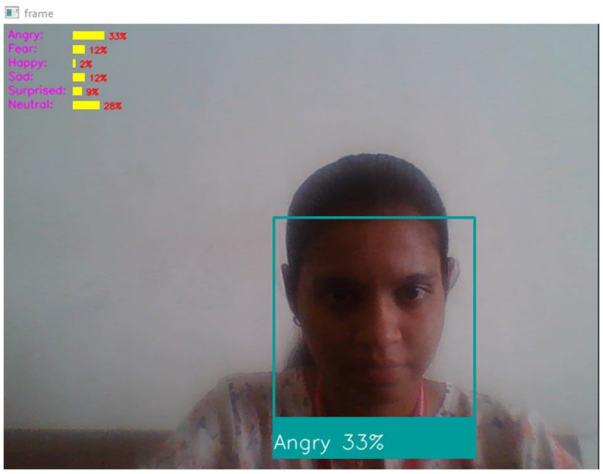


Figure 3: Image Testing experiment 2



Figure 4: Image Testing experiment 3

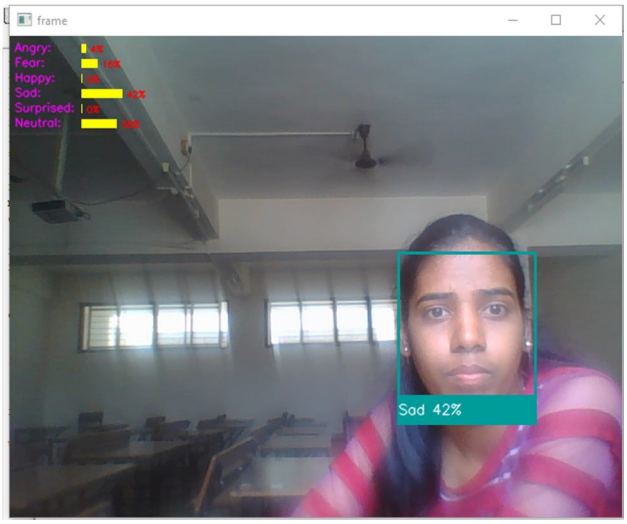


Figure 5: Image Testing experiment 4

The entire process of developing an automatic attendance system in detail is explained in this section. The system consists of Windows over which OpenCV 2.0 runs. Program is developed in Python 3.6 or higher version for the algorithm and implementation in real time. Is used in our system to predict and detect faces in an image. Face recognition libraries comes with standard in Python Library.

Table 1: Various experimental analysis for emotion detection with different data samples

Data Samples	Accuracy	Precision	Recall	F-score
50	94.20	95.30	96.10	96.30
100	94.30	95.70	95.30	95.10
150	94.15	95.80	96.40	94.95
200	92.40	93.55	94.90	96.30
250	91.50	92.60	93.90	94.85
300	91.80	93.10	94.20	94.60
500	92.10	93.05	94.70	95.80

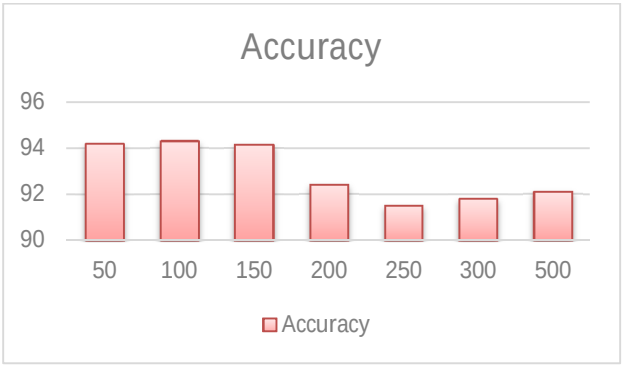


Figure 6: Various experimental analysis for emotion detection with different data samples

The above Table 1 demonstrates how accurate the recommended model, which is a combination of VGG16 and ResNet, is. Based on the validation dataset, the suggested model achieves an accuracy of 0.68, whereas the VGG16 classifier achieves 0.41 and the ResNet50 classification model achieves 0.44. The suggested model achieves an accuracy of 0.81 on the testing dataset, whereas the VGG16 classification model achieves just 0.36 and the ResNet50 classifier achieves 0.49. According to the findings of this research, the recommended framework, which was trained on a specialized dataset, provides a higher level of accuracy compared to the other two pre-trained models. Figure 5 depicts the confusion matrix that was discovered after analyzing the photographs in the test set depicting a variety of feelings such as surprise, joy, anger, contempt, fear, anxiety, and grief. These feelings include.

VII. DATASET DESCRIPTION

A proposed model for the CNN with seven layers:

After the information has been imported and partitioned, it is subjected to a preprocessing step in which the values are normalized so that they are all contained within the range [0, 1], and the values are converted to the needed size as well as what the network wants. This step ensures that the data is compatible with the network. Because the data had previously existed in a variety of sizes and forms, we were required to turn it into a standardized shape. It has been discovered that carrying out preprocessing on the data results in increased accuracy. The convolutional neural network with seven layers of convolutional processing generated an accuracy of 0.81 for the outcomes it produced.

A. VGG16 Model

This is a classifier that has already been pre-trained, and it was constructed using the ImageNet dataset. In this particular setup, we began by installing the topmost layer, and then, in order to get a higher level of fine tuning, we followed up with the installation of the softmax layer. Because it is pretrained on such a wide range of classes, the model exhibited a test accuracy of 0.35, which is slightly lower than the proposed convolutional neural network structure.

B. Model ResNet-50

Because the ResNet model can also function as a pre-trained classifier, it is a very helpful tool to have at your disposal when dealing with photographs.

We also tested out this pre-trained classifier so that we could do a comparison of the different degrees of accuracy. This model contains fifty layers, and it is unusually deep since it was trained on both data that is scene-centric and data that is object-centric (MS COCO). This model achieved a testing accuracy of 0.48 when applied to our one-of-a-kind dataset.

C. Algorithm

During the training process, we used a specialized convolutional framework that was trained on our own produced dataset in conjunction with pre-trained versions of the ResNet and VGG16 models. Figure 4 shows the pre-trained version of the VGG16 classifier that was created with the help of the ImageNet dataset. On the other hand, we constructed our own proposed seven-layer customized convolutional neural network model in order for it to be trained on our very own dataset. Despite the fact that it was trained on a unique dataset, this model performs in a manner that is similar to that of the VGG16 and ResNet models. The dependability of these three models is being evaluated, and moving forward, the one that fares the best in terms of performance will be used to anticipate the emotional state of the visual imagery using the GUI. Another essential component is how much data a deep learning model uses throughout its training phase; the outcomes will be of higher quality the more data it processes. In order to train the specialized convolutional neural network architecture, we used a dataset that was created specifically for this purpose. The framework of the convolutional neural network is expanded with additional layers by our unique model, which does this so that the prediction performance may be improved. As a direct result of this, the dataset generates superior results, and the prediction may be considered to be much more accurate.

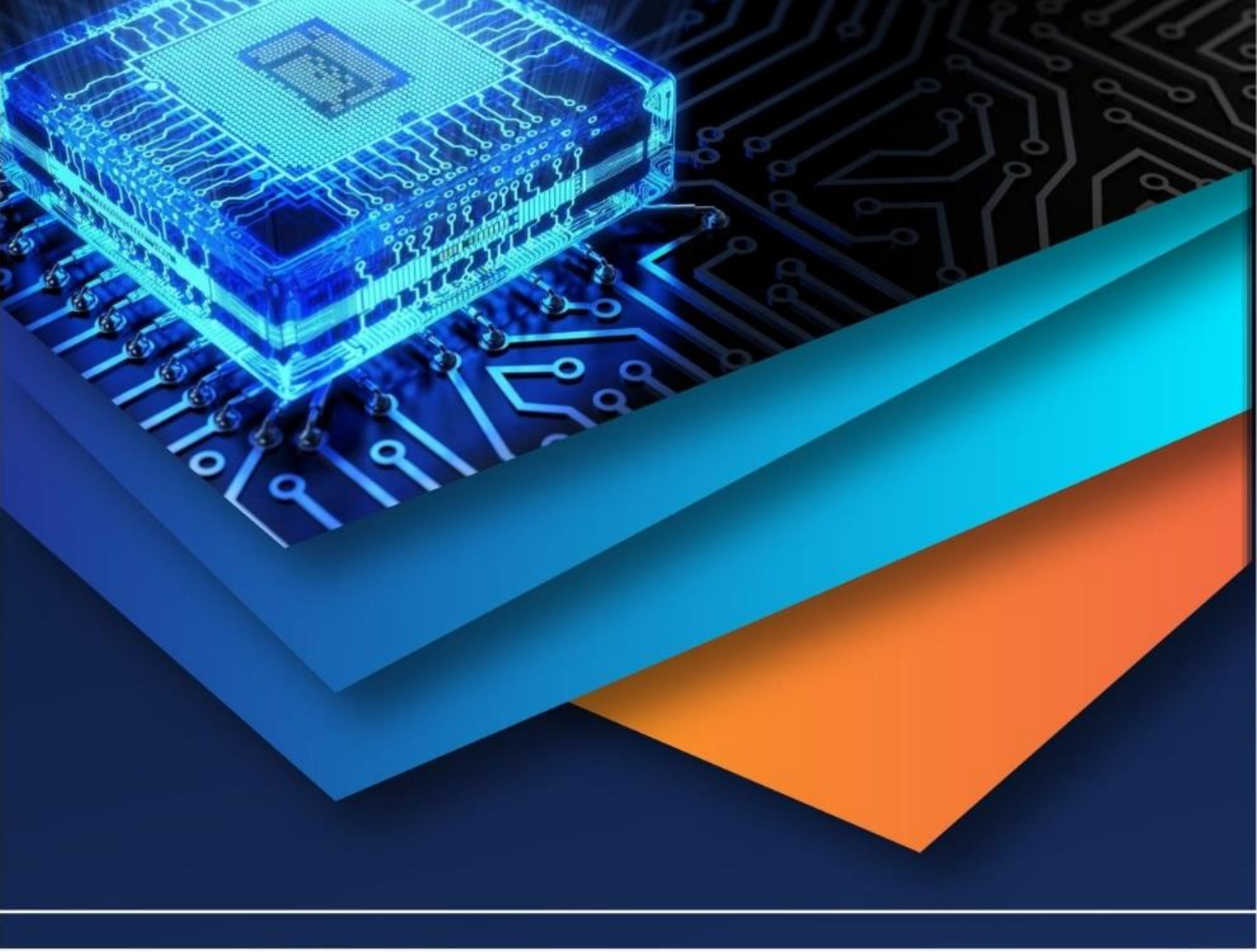
VIII. CONCLUSION

This study presents a robust face recognition model based on mapping behavioural variables with physiological biometric parameters in the facial expression recognition system. Geometrical components reconstructed as the basis matching template for the identification system are related to the physiological aspects of the human face with relevance to diverse emotions such as pleasure, fear, rage, and surprise.

The system's behavioral component links the mentality that underlies various expressions as its property foundation. In algorithmic genetic genes, the property bases are detached as a disclosed and concealed category. The gene training set assesses each face's ability to express uniquely and offers a robust expressional recognition model in biometric security.

REFERENCES

[1] Mukhopadhyay, Moutan, et al. "Facial emotion detection to assess Learner's State of mind in an online learning system." Proceedings of the 2020 5th international conference on intelligent information technology. 2020.
[2] Ali, Md Forhad, Mehenag Khatun, and Nakib Aman Turzo. "Facial emotion detection using neural network." the international journal of scientific and engineering research (2020).
[3] Zamkah, Abdulaziz, et al. "Identification of suitable biomarkers for stress and emotion detection for future personal affective wearable sensors." Biosensors 10.4 (2020): 40.
[4] Farkhod, Akhmedov, et al. "Development of Real-Time Landmark-Based Emotion Recognition CNN for Masked Faces." Sensors 22.22 (2022): 8704.
[5] Khaireddin, Yousif, and Zhuofa Chen. "Facial emotion recognition: State of the art performance on FER2013." arXiv preprint arXiv:2105.03588 (2021).
[6] Zhang, Hongli, Alireza Jolfaei, and Mamoun Alazab. "A face emotion recognition method using convolutional neural network and image edge computing." IEEE Access 7 (2019): 159081-159089.
[7] Saeki, Mao, et al. ”Analysis of Multimodal Features for Speaking Proficiency Scoring in an Interview Dialogue.” 2021 IEEE Spoken Language Technology Workshop (SLT). IEEE, 2021.
[8] Shin, Dong Hoon, Kyungyong Chung, and Roy C. Park. ”Detection of emotion using multi-block deep learning in a self-management interview app.” Applied Sciences 9.22 (2019): 4830.
[9] Patil, Payal, and S. Shinde. "Comparative analysis of facial recognition models using video for real time attendance monitoring system." 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, 2020.
[10] Huang, Shizhen, and Haonan Luo. "Attendance System Based on Dynamic Face Recognition." 2020 International Conference on Communications, Information System and Computer Engineering (CISCE). IEEE,



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)