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.2023.52438

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

Deep Learning Approach for Facial Expression Recognition

Niharika .L[1], N.Charitha[2], N. Vivek[3], G. Sravan Kumar[4]
[1, 2, 3]*UG Scholar,* [4] *Assistant Professor, Department of Electronics and Communication Engineering, Maturi Venkata Subbarao Engineering College, Osmania University, Telangana, India*

Abstract: ***In recent decades, facial expression recognition has emerged as a hot topic with significant implications in the realm of human-computer interaction. The simplest way for humans to express their emotions is through facial expressions. Non-verbal communication relies heavily on facial expression. This study outlines deep learning-based Facial Expression Recognition (FER) algorithm. The performance of the FER approach is compared based on the number of expressions detected and the difficulty of CNN algorithms. The FER2013, CK+ databases were used in testing the design. CNNs (Convolutional Neural Networks) have as of late gained fame in the field of profound learning because of its superb plan and capacity to convey clever outcomes without the requirement for manual feature extraction from raw information. The suggested algorithm achieves a higher rate of recognition on four datasets.***
Keywords: Facial Expression Recognition (FER); CNN; Feature Extraction; Facial Expressions

I. INTRODUCTION

Facial expression is a significant non-etymological strategy for people to convey data. It has a high significance in the fields of human-PC collaboration, AI, and mental science as a division of expression recognition. The objective of expression recognition study (FER) is to help a machine to normally perceive looks. Ekman and Friesen [1] proposed six essential looks that are generally conveyed on the individual's face (angry, fear, disgust, surprise, happy, and sadness) during the last part of the 1990s. Various measure of exploration have been conveyed to date on the recognizable proof of these articulations. Human-PC cooperation innovation is a sort of innovation that involves PC hardware as the medium to make human-PC association. With the fast development of expression recognition and man-made consciousness lately, increasingly more examination in the field of human-PC communication innovation has been undertaken.[2][3] Facial expression recognition has a long history of purpose as a vital part of savvy human-PC cooperation. It's been utilized in businesses as assorted as right hand clinical, far off instruction, intuitive games, and public safety[4][5][6].

Facial expression recognition uses computer image processing technology to extract information representing facial expression features from the original input facial expression images and classifies the facial expression features based on human emotional expressions such as happiness, surprise, aversion, and neutrality [7], [8]. Facial expression identification is critical in the research of emotional measurement.

Artificial intelligence is making it easier for humans and computers to connect with one another. As a result, significantly supporting research into face expression recognition technology is vital for individual and societal evolution [9], [10]. Facial expression recognition is a technique for interpreting the inner emotion of a human face expression that employs a computer as a tool and integrates it with specific algorithms [11]. In the classroom, facial expression recognition can assist teachers capture and document students' emotional changes as they learn, as well as provide a better reference for teachers to teach students based on their ability.

In the field of traffic, facial expression recognition can be used to detect pilot or driver fatigue and to prevent traffic accidents by technological means.

Life management robots can better understand people's mental states and intentions by incorporating facial expression recognition into everyday life, and then respond appropriately, increasing the human-computer interaction experience. By incorporating gaze recognition into daily life, life executives robots can more likely comprehend individuals' psychological emotions and expectations, and then respond appropriately, thus strengthening the human-PC communication experience [12].

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

II. LITERATURE SURVEY

This section contains a survey of expression recognition algorithms proposed by various researchers, as well as an analysis of the datasets.

K.Lieu et al. [13] focused on face expression recognition and suggested a method that combines many progressive subnets. Each subnet is densely packed with CNN models that were individually created. The entire organization is brought together to build a design. The design was prepared and investigated utilizing FER2013 information tests, and a precision of 65.03 percent was accomplished.

Khorrami demonstrated in [14] that CNNs can get great precision in feeling characterization by utilizing a zero-inclination CNN on the prolonged CK+ and the Toronto Face Dataset. Aneja et al. [15] used deep learning to train a network to model the expression of human faces, one for animated faces, and one to transfer human photos into animated ones to produce a model of facial expressions for stylized animation characters.

In [16], Liu combined feature extraction and order in a single circling network, referring to the need for feedback from both areas. They applied their BDBN to CK+ and JAFFE with cutting-edge precision. Chung-Lin and Yu-Ming [17] proposed a Point Distribution Model (PDM) approach to dealing with look examination in the context of facial element extraction. The PDM evaluation looks into the measurable data of the arranged or named focuses' aspects across the preparation set. The proposed technique utilizes 180 pictures from 15 workers, with every individual exhibiting six motions, and afterward 12 pictures from each volunteer are picked.

To sort and fit the highlights separated from facial pictures, the Action Parameters (AP) Classifier is utilized. The proposed strategy had a precision pace of 84.41 percent. R. Kumar et al.[18] analysed a method for detecting emotions (Frame by Frame) using a DCNN, and it represents the various varieties in degrees of power of human sentiments passed on through faces, going from extremely low to raised measures of feelings. This strategy was prepared utilizing the FER-2013 dataset. In the appraisal, the proposed strategy delivers great results. Yang et al. [19] proposed a Facial Recognition Standard and examined the future dataset for face discovery. Commented on Facial Landmarks in the Wild, Face Detection Dataset Benchmark, and PASCAL FACE are a couple datasets who need more information to prepare. Basic examples are produced by face discovery techniques in various ways. Both positive and negative examples are helpful in the preparation test. The FreNet is a deep learning-based structure for look acknowledgment portrayed in this paper.

As opposed to convolutional neural networks in the spatial space, FreNet acquires picture improvement benefits in the recurrence area, like powerful investigation and primary imitation expulsion [20]. [21] presents a progressive Bayesian subject model in light of posture to resolve the troublesome issue of multi-client facial expression.

Prior to seeing articulation, the technique integrates nearby appearance highlights with worldwide mathematical data and becomes more acquainted with moderate portrayal. It offers an incorporated answer for multi-utilitarian facial expression by imparting a bunch of capacities to various stances, bypassing the singular preparation and boundary change of each stance, permitting it to be extended to a critical number of postures.

III. AIM AND OBJECTIVES

A. Aim

To design a neural network for Facial expression recognition and predict the emotion for an input image.

B. Objectives

1) To design a neural network.
2) To predict the emotion of an image.
3) To display the predicted expression.

C. Proposed System

Facial emotion recognition is the process of detecting human emotions from facial expressions. The human brain recognizes emotions automatically, and software has now been developed that can recognize emotions as well. This technology is becoming more accurate all the time, and will eventually be able to read emotions as well as our brains do. AI can detect emotions by learning what each facial expression means and applying that knowledge to the new information presented to it. Emotional artificial intelligence, or emotion AI, is a technology that is capable of reading, imitating, interpreting, and responding to human facial expressions and emotions.

IV. METHODOLOGY

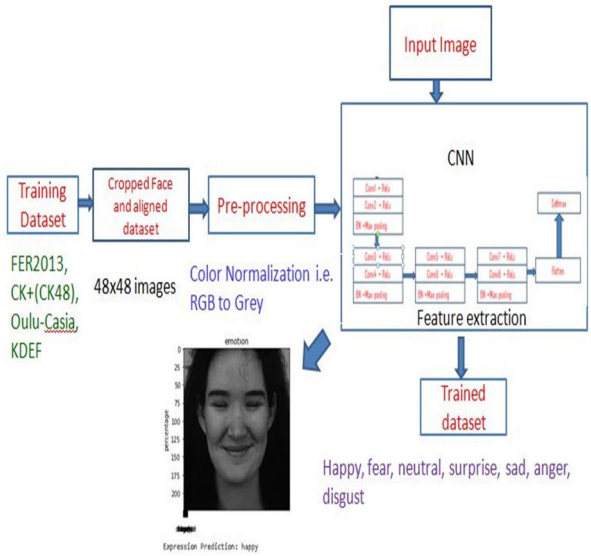


Fig.1 shows the block diagram of the model and how it can be executed is shown in the execution flow below.

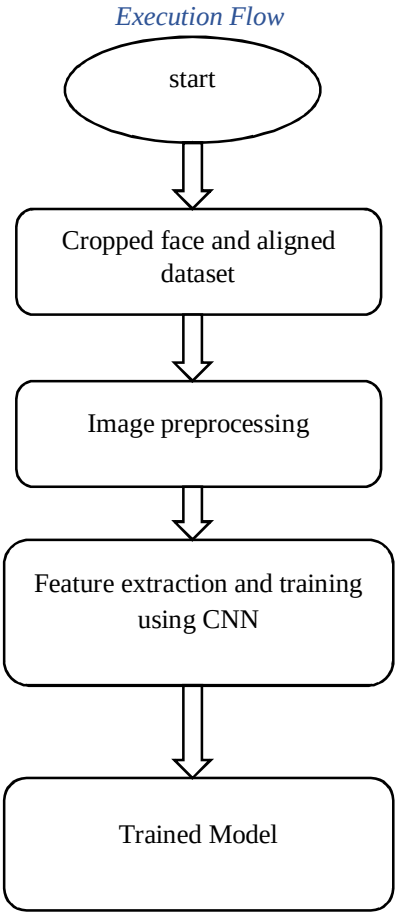


Fig. 2: shows the execution flow of the model and the description of each step is as follows:

A. Cropped Face and aligned Dataset

1) Face Detection is valuable while detecting a facial picture. Face Detection is done to train a dataset by using a classifier. In this phase, images in the dataset are being cropped so that the face can be analysed properly and these images are aligned such that their size will be 48x48 pixels.

2) In this phase, images in the dataset are being cropped so that the face can be analyzed properly and these images are aligned such that their size will be 48x48 pixels.

B. Image Pre-processing

1) Picture pre-handling incorporates the expulsion of commotion and standardization against the variety of pixel position or brilliance.

2) Here we have used color normalization so as to turn all the RGB images into Grey scale 48x48 images.

C. Feature extraction and training using CNN

1) After preprocessing, the picture of the face is utilized to remove the significant highlights. Scale, posture, interpretation, and varieties in brightening level are on the whole innate issues in picture order.

2) The CNN algorithm is used to extract the important features.

D. Network Architecture

CNN is a neural network that extracts input image features and then categorizes the image characteristics using another neural network. The feature extraction network makes use of the input image. The neural network uses the extracted feature signals for classification. The neural network classification then operates on the image information to generate the output. Fig.3 shows the CNN architecture used in this paper.

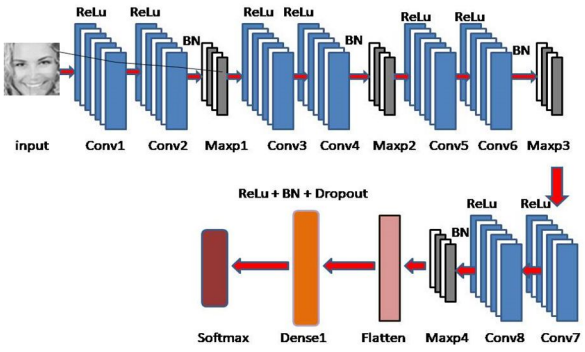


Fig.3: CNN's overall architecture for Facial expression recognition.

CNN, which is made up of neurons with learnable biases, is used to demonstrate image recognition and classification. Every filter that accepts input performs convolution, which is then followed by non-linearity. As shown in Fig.3, the proposed CNN structure includes Convolutional, batch normalization, ReLU, max pooling, Flatten, and Dense layers. We have used keras framework to train the network in this project.

The convolutional layer is a basic component of a convolutional network. The Convolution layer's primary goal is to extract features from the input image. A collection of learnable neurons represents the image used as input. The following convolutional layer receives input data as feature maps. The convolution layer, which has the qualities of local connection and value sharing, lies at the heart of CNN. The specific strategy is to quantify on the top input layer utilizing the convolution core one by one [27][28].

The units that utilise the rectifier are included in the ReLU. It is an operation that sets all non-positive values in the map to zero. For the sake of clarity, we will assume that y is the neuron input and that the rectifier is given as f(y) = Maximum (0,y) for neural networks.

The layer of pooling decreases the significance of every activation map, however it can have more subtleties. The images provided as information are divided into a collection of non-covering square forms, and each area is sub-examined by a non-direct activity such as medium or most notable. It's wedged between the convolution layers in the middle. Instead of being distributed throughout the system, batch normalization performs input on a 0 to 1 scale.

The flatten module converts the contribution from N Dimension to 1 Dimension fully without affecting the example size.

Dropout is a training strategy in which neurons are ignored at random. They are "disappeared" at random. This implies that their interest in downstream neuron enactment is dealt with sequentially on the forward path, and any weight refreshes are not relevant on the neuron on the opposite way.

The dense layer is the layer in which the layers are plotted and the highlights are thoroughly examined. The dense layer is the final sign for the image to be treated during the experimental stage.

The CNN's final layer employs a classifier known as Softmax. This classifier represents a challenging multi-yield characterization method. When a particular example is handled in the framework, each neuron generates a value between 0 and 1, indicating the chance that the example belongs to that class. As a result, the class associated with the neuron having the highest yield esteem is chosen as the characterisation execution.

Parameters Used

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 48, 48, 64)	1664
conv2d_2 (Conv2D)	(None, 48, 48, 64)	102464
batch_normalization_1 (Batch	(None, 48, 48, 64)	256
max_pooling2d_1 (MaxPooling2	(None, 24, 24, 64)	0
conv2d_3 (Conv2D)	(None, 24, 24, 128)	73856
conv2d_4 (Conv2D)	(None, 24, 24, 128)	409728
batch_normalization_2 (Batch	(None, 24, 24, 128)	512
max_pooling2d_2 (MaxPooling2	(None, 12, 12, 128)	0
conv2d_5 (Conv2D)	(None, 12, 12, 256)	819456
conv2d_6 (Conv2D)	(None, 12, 12, 256)	1638656
batch_normalization_3 (Batch	(None, 12, 12, 256)	1024
max_pooling2d_3 (MaxPooling2	(None, 6, 6, 256)	0
conv2d_7 (Conv2D)	(None, 6, 6, 512)	3277312
conv2d_8 (Conv2D)	(None, 6, 6, 512)	6554112
batch_normalization_4 (Batch	(None, 6, 6, 512)	2048
max_pooling2d_4 (MaxPooling2	(None, 3, 3, 512)	0
flatten_1 (Flatten)	(None, 4608)	0
dense_1 (Dense)	(None, 128)	589952
batch_normalization_5	(Batch (None, 128)	512
activation_1 (Activation)	(None, 128)	0
dropout_1 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 7)	903
activation_2 (Activation)	(None, 7)	0

Total params: 13,472,455
Trainable params: 13,470,279
Non-trainable params: 2,176

Fig 4: Parameters

Fig 4 shows the parameters in the CNN architecture at each of the layer.

E. Implementation Details

The kernel sizes in operations are set to 3x3 and to 5x5 in advance. For all convolutions, there are 72 feature maps. To train this network, we utilize TensorFlow, Adam optimizer and keras framework. For data augmentation, the training photos are cropped into 48 x 48 patch pairs and horizontally flipped. To train this model we used datasets called FER2013,CK+,Oulu-Casia,KDEF. We used python code to implement this method in Kaggle platform.

V. RESULTS

A. Datasets

1) *FER2013:* It is a wild facial expression dataset including 28689 training shots and 3589 testing photos. There are seven expressions in FER2013: anger, contempt, fear, happy, neutral, surprise, and sadness. Because of the varying lighting, occlusions, stances, and low resolution, the FER2013 dataset is more challenging to analyze than the prior three. Furthermore, this dataset has a large number of unclear labels. We compare our network's performance on FER2013 to that of commonly used CNNs without model pre-training. On the training set, we train our model, and on the test set, we calculate recognition accuracy. [20]
2) *CK+:* It consists of 327 labeled sequences of 108 participants, each of which begins with neutral and ends with peak state. Happiness, disgust, fury, fear, surprise, sadness, and contempt are the seven different face expressions in CK+. [20]

Expression Prediction Of images

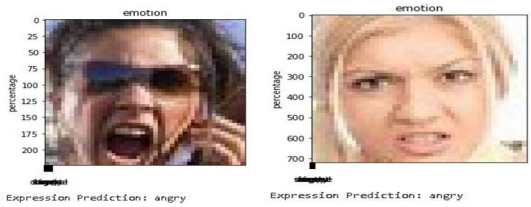


Fig 5: Angry

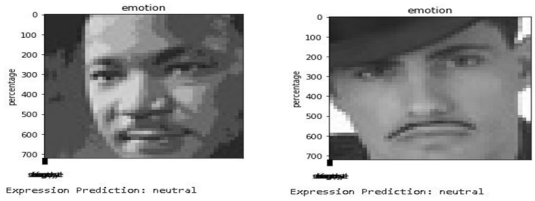


Fig 6: Neutral

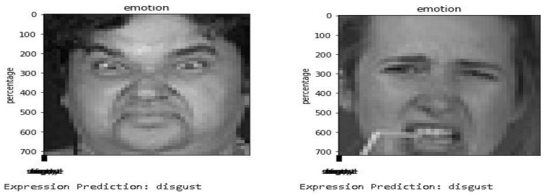


Fig 7: Disgust

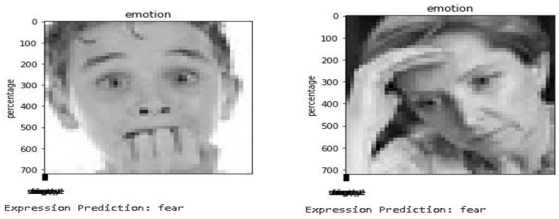


Fig 8: Fear

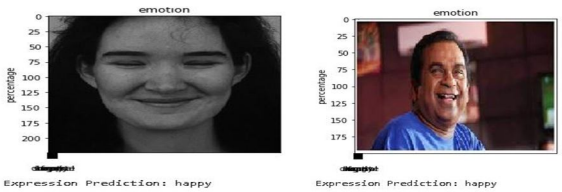


Fig 9: Happy

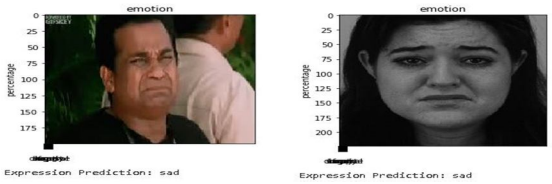


Fig 10: Sad

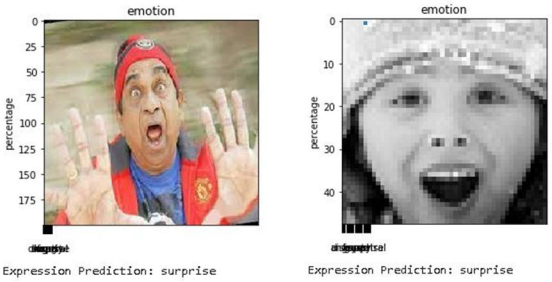


Fig 11: Surprise

S.No	Dataset	Training Accuracy	Test Accuracy
1.	CK+(CK48)	98.78	90.43
2	FER 2013	97.67	66.98

Table 1: Accuracies of two datasets on the CNN network

FER2013 Accuracy(%)	
Unsupervised Domain Adaptation	65.5%
VGG+SVM	66.13%
GoogleNet	65.42%
FER on SoC	66.8%
This Method	66.98%

Table2: Comparisons of Accuracies on FER2013 dataset on the CNN network

Methods[CK+]	Accuracy[%]
CSPL [35]	89.6
3DCNN [33]	85.8
This method	90.43

Table 3: Comparisons of Accuracies on CK+ dataset on the CNN network

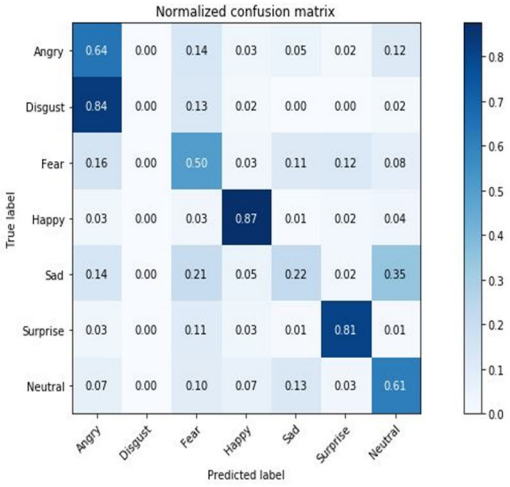


Fig 12: Confusion Matrix of FER2013 on network

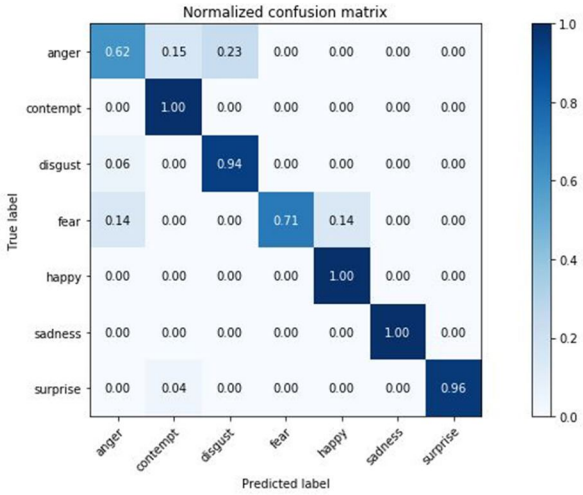


Fig 13: Confusion Matrix of CK+ on network

VI. CONCLUSION

The most important parameters for nonverbal communication are facial expressions. In the fields of training, workplaces, and, in the mental states of the patients can be effortlessly broke down, and medicinal activities can be done at a quicker rate. Here we have used multiple layers by varying the kernel sizes to train the CNN network. Hence from the observations, we have achieved 90.43%, 66.98% as testing accuracies for CK+, FER2013 datasets respectively. The trial results show that the proposed strategy mix beats a portion of the cutting edge techniques with regards to exactness.

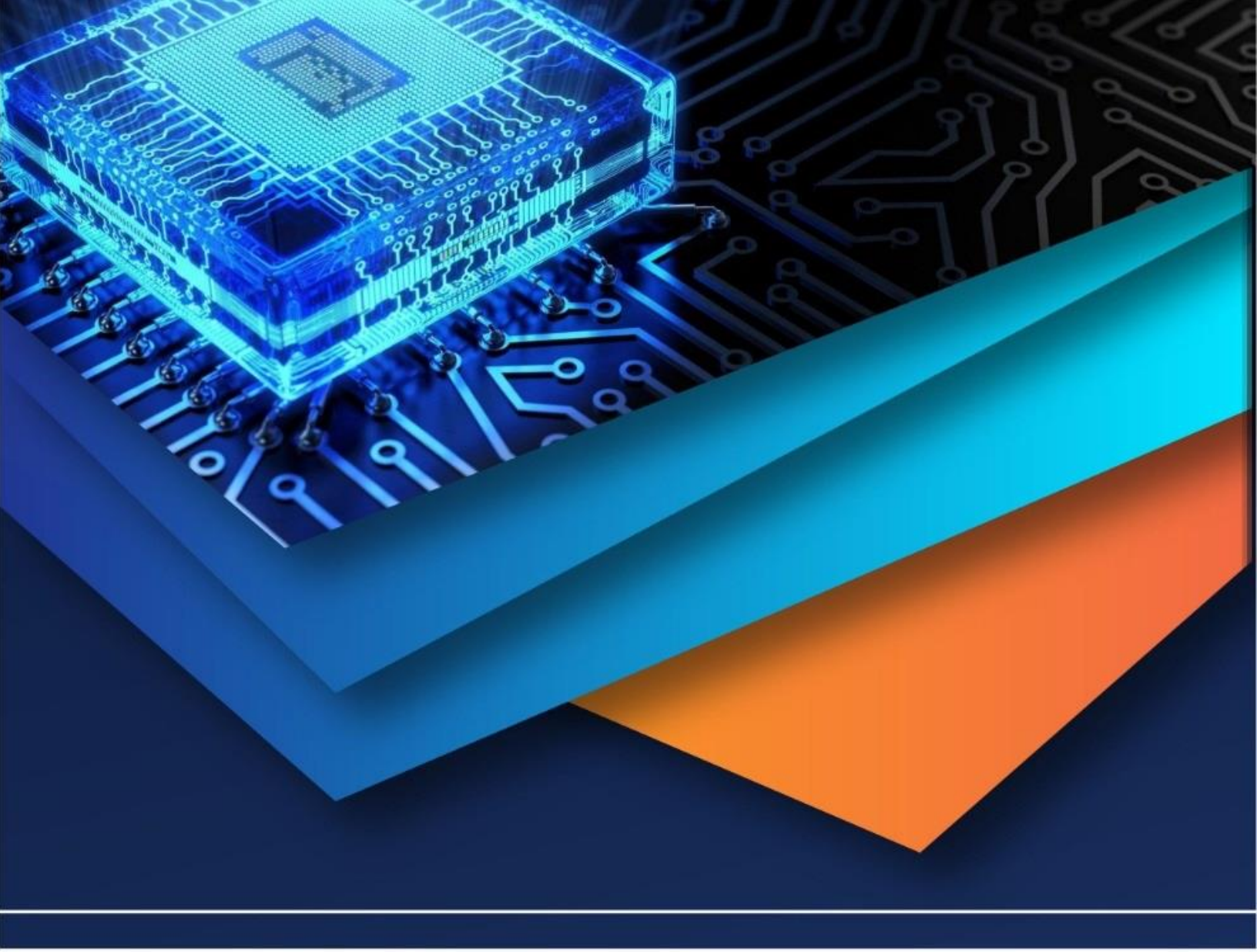
VII. FUTURE SCOPE

Face,Age, Gender and Emotion discovery can be applied in a wide scope of uses, including human-PC communication, bio-metric security, etc. As a result, it offers insight into AI technology, which simulates the human mind using various supervised and unsupervised machine-learning methodologies.

REFERENCES

[1] Paul Ekman and Wallace V Friesen. “Constants across cultures in the face and emotion.” In: Journal of personality and social psychology 17.2 (1971), p. 124.
[2] Raja Majid Mehmood, Ruoyu Du, and Hyo Jong Lee. “Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors”.
[3] Tengfei Song, Wenming Zheng, Cheng Lu, Yuan Zong, Xilei Zhang, and Zhen Cui. “MPED: A multi-modal physiological emotion database for discrete emotion recognition”. In: IEEE Access 7 (2019), pp. 12177–12191.

[4] Erdenebileg Batbaatar, Meijing Li, and Keun Ho Ryu. “Semantic-emotion neural network for emotion recognition from text”. In: IEEE Access 7 (2019), pp. 111866–111878.
[5] Yuanhui Zhang, Yan Li, Bo Xie, Xiaolu Li, and Junjiang Zhu. “Pupil localization algorithm combining convex area voting and model constraint”. In: Pattern Recognition and Image Analysis 27.4 (2017), pp. 846–854 .
[6] Hongying Meng, Nadia Bianchi-Berthouze, Yangdong Deng, Jinkuang Cheng, and John P Cosmas. “Time-delay neural network for continuous emotional dimension prediction from facial expression sequences”. In: IEEE transactions on cybernetics 46.4 (2015), pp. 916–929.
[7] Madhumita Takalkar, Min Xu, Qiang Wu, and Zenon Chaczko. “A survey: facial micro-expression recognition”. In: Multimedia Tools and Applications 77.15 (2018), pp. 19301–19325.
[8] Mehmet Sira¸c Ozerdem and Hasan Polat. “Emotion recognition based on ¨ EEG features in movie clips with channel selection”. In: Brain informatics 4.4 (2017), pp. 241–252.
[9] Sergio Escalera, Xavier Bar´o, Isabelle Guyon, Hugo Jair Escalante, Georgios Tzimiropoulos, Michel Valstar, Maja Pantic, Jeffrey Cohn, and Takeo Kanade. “Guest editorial: The computational face”. In: IEEE Transactions on Pattern Analysis and Machine Intelligence 40.11 (2018), pp. 2541–2545.
[10] Xiang Yu, Shaoting Zhang, Zhennan Yan, Fei Yang, Junzhou Huang, Norah E Dunbar, Matthew L Jensen, Judee K Burgoon, and Dimitris N Metaxas. “Is interactional dissynchrony a clue to deception? Insights from automated analysis of nonverbal visual cues”. In: IEEE transactions on cybernetics 45.3 (2014), pp. 492–506.
[11] Filippo Vella, Ignazio Infantino, and Giuseppe Scardino. “Person identification through entropy oriented mean shift clustering of human gaze patterns”. In: Multimedia Tools and Applications 76.2 (2017), pp. 2289– 2313.
[12] Hongli Zhang, Alireza Jolfaei, and Mamoun Alazab. “A face emotion recognition method using convolutional neural network and image edge computing”. In: IEEE Access 7 (2019), pp. 159081–159089.
[13] Kuang Liu, Mingmin Zhang, and Zhigeng Pan. “Facial expression recognition with CNN ensemble”. In: 2016 international conference on cyberworlds (CW). IEEE. 2016, pp. 163–166.
[14] Pooya Khorrami, Thomas Paine, and Thomas Huang. “Do deep neural networks learn facial action units when doing expression recognition?” In: Proceedings of the IEEE international conference on computer vision workshops. 2015, pp. 19–27.
[15] Deepali Aneja, Alex Colburn, Gary Faigin, Linda Shapiro, and Barbara Mones. “Modeling stylized character expressions via deep learning”. In: Asian conference on computer vision. Springer. 2016, pp. 136–153.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY