International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT2390356

Emotional Analysis using Deep Learning

D. Hari Krishna, Kottada Rakesh, Abhishek Kaveli, Kamsani Karthikeya, Maladoddi Pavan Kalyan

Department of ECE, B.V.R.I. T, Narsapur, Andhra Pradesh, India

ARTICLEINFO

Article History:

Accepted: 20 June 2023

Published: 05 July 2023

Publication Issue

Volume 9, Issue 4

July-August-2023

Page Number

06-14

ABSTRACT

Emotions are mental states that accompany physiological changes in the face, resulting in facial expressions. Sympathy, anger, worry, joy, fright, and other significant emotions are a few examples. Facial expressions play a significant role in non-verbal communication because they encapsulate a person's emotions. There has been a great deal of research done on computer modelling of human emotions. Computer modelling of human emotions has been made possible by computer technology. However, it is still in its infancy. The authors attempted to overcome limitations and create new opportunities as well as gain a better understanding and implement this simple form of human interaction in proposed computer-aided world. It has been made possible to evaluate and interpret genuine facial expressions in real time thanks to new techniques for collecting facial expressions and quick, highresolution pictures. The FER (Facial Expression Recognition) method currently relies on motionless frames, which makes it very hard to recognize foreground from background in the absence of motion information. This study describes a real-time facial expression identification system that detects faces using HAAR cascading classification and classifies facial expressions using convolutional neural networks. The system utilizes a webcam to dynamically display emotion text and accurately categorizes seven major emotions, including anger, disgust, fear, happiness, sadness, surprise, and neutrality. Real-time facial expression recognition may be utilised in a number of real-world applications, including as airport security, trade, and medical monitoring.

Keywords- Convolutional Neural Network, Emotion Recognition, HAAR Cascade, Facial Expressions

I. INTRODUCTION

People connect with one another and carry out their relationships through their emotions. It is a non-verbal form of expression that, like body language, may spice up a conversation. The seven universal emotions—happiness, surprise, neutrality, anger, sadness, disgust, and fear—are acknowledged by many cultures. Surprisingly, cross-cultural debates may still be found when dealing with more complex statements that may have more than one manner of expressing themselves. As a result, the availability of a mechanism for recognising mood based on facial expressions would eliminate these disparities.

Innovation like this would be more advantageous for its privacy, refreshment, and feedback applications. Recognizing facial emotions is a complex task because of the key drawbacks of i) No dataset for the system to be trained. ii) The input image can be a static image or an intermediate frame into a facial expression. The Convolutional Neural Network model received video stream input from a camera. It is critical to optimize the network for the model to perform properly. Because of the ambient circumstances that might impact accuracy, it becomes more difficult. When compared to other types of neural networks, Convolutional Neural Networks have been demonstrated to be more effective. CNN is a layer-wise algorithm. It has convolutional layers, each of which is responsible for bringing out features from the training set's pictures. To accomplish shift variance, extracted features are max pooled. The complexity increases as you move from one layer to the next. The convolutional filters in the topmost convolutional layer reflect the object's representative properties. Fully connected layers classify the outputs of many filters from convolutional layers.

II. RELATED WORK

Bhadana L et al [1], proposed that solely on categorizing emotions into three groups: neutral, negative, and positive. They put forward a CNN model which achieved an accuracy rate of 81%. The FER2013 and cohn-kanade (CK+) extended datasets were utilized in the study. Due to FER's numerous applications in the areas of computer vision, robotics, and human-computer interaction, it has recently become a popular research topic. In recent years, Human-Computer Interaction, or FER, has grown in importance as a study area. CNN was utilised in real-time to categorise human emotions from dynamic facial expressions, with a test accuracy of 57%, demonstrating the utility of utilising neural networks in this manner. Minaee S et al [2], Emotion identification is based on speech and EEG signals. The electrical activity of neurons in the brain is used in EEG. Speech recognition in multiple languages has gotten a lot of attention. The use of speech, emotion, and EEG together has been used very little in research.

H.Kumar et al [3], Frameworks such as keras and tensor flow were suggested by them. They utilize parameters of ANN (Artificial Neural Networks) and CNN (Convolution Neural
Networks) that are readily available to achieve the desired results. ANN and CNN typically rely on backpropagation, a widely used algorithm, which calculates the function's gradient to determine how the parameters can be utilized to reduce errors that may impact model efficiency. Overfitting is a common issue that arises. According to A. Santra et al. [4], The suggested technique begins by training a convolutional neural network (CNN) is designed to identify different facial expressions and classify them into five basic emotional states, namely Joyful, Astonished, melancholic, Furious, Indifferent, revulsion and temper. The Haar filter is then used to identify and retrieve facial characteristics. Then, an audio synthesiser receives the anticipated group emotions and outputs audio. This model's accuracy is 65%. According to J. C. Kim et al. [5], the proposed technique is looking at video clips and recognising facial emotions. The content can be divided into three sections. The first part call for gathering data on the motion flow and face landmark flow. The second part is the classification of facial motion flow using CNN. The third part is the classification of facial geometry flow using SVM. The expression sequence is additionally categorised by combining.

III. METHODOLOGY

The paper will be structured into 3 distinct functional modules namely.

i. Data processing

ii. Face registration
iii. Emotion classification

The FER system was developed using The FER2013 dataset was used in a Kaggle competition on FER. This dataset has a complete count of all the data of 35,887 tagged a number of pictures, of which some are included 3589 are test images and 28709 are train images. The dataset was also used for the final exam of the competition and includes 3589 private test images. The black and white images in the FER2013 dataset are 48x48 pixels in size. Viewpoint, lighting, and scale are all different between the images in the FER2013 collection. Indicators of emotion include 0 for rage, 1 for disgust, 2 for fear, 3 for joyfulness, 4 for neutral, 5 for sorrowness, and 6 for surprise. Table 1 displays the number of photos with varied emotional states.

TABLE 1. NUMBER OF IMAGES WITH VARIOUS EMOTIONS

S.No	Emotions	Number of Images	Label
1	Angry	4593	0
2	Disgust	547	1
3	Fear	5121	2
4	Happy	8989	3
5	Neutral	6198	4
6	Sad	6077	5
7	Surprise	4002	6

Fig. 1. Images of various emotions from a dataset

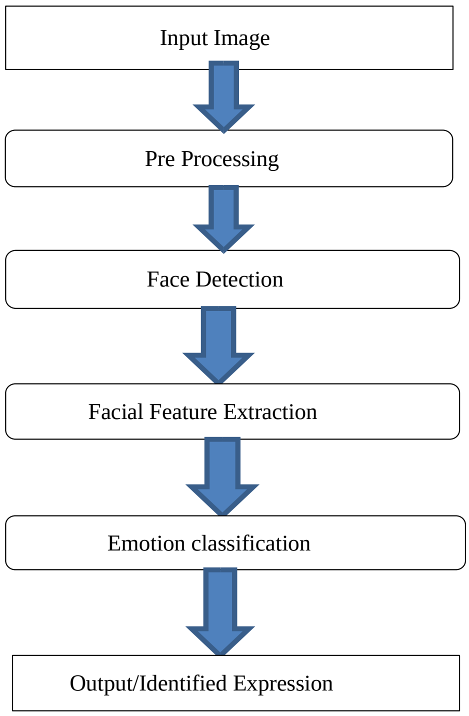

Fig 1. Architecture of proposed model.

The dataset has three different sets: Training, Public Test, and Private Test. The Training and Public Tests are used to create the model, while the Private Test is used to assess it. The images of various emotions from a dataset are shown in figure 2.

DATA PROCESSING

The input image to the FER system could differ in size, colour, and illumination. To help the algorithm provide more accurate and timely findings, some image pre-processing was done. The pre-processing techniques used are scaling, normalizing, and converting the image to grayscale. First, normalization is a technique used to even out an image's lighting and enhance the face's appearance. Whereas when a coloured image is scaled to grey, the amount of light falling on the image determines the value of each pixel. Grey scaling is utilized because coloured images are more difficult for computers to process. Finally, cropping entails removing extraneous portions of an image. Additionally, it reduces the amount of memory needed.

FACE REGISTRATION

The initial step in any FER system is face detection. Using the Haar technique, the target was found. The Viola-Jones detectors, sometimes referred to as the HAAR cascades classifier method, are classifiers that can identify a particular object in an image or video after training. They are instructed to differentiate between assertive and unfavourable pictures of the face. Haar cascade algorithm has proven to be a quick and accurate way to find things in pictures. Three areas of the face that are dark are identified by the Haar characteristics, including the brows. Then, using the Haar cascades and quick pixel computation, the computer is trained to recognize two dark patches on the face. These patches successfully remove extraneous background information from the image.

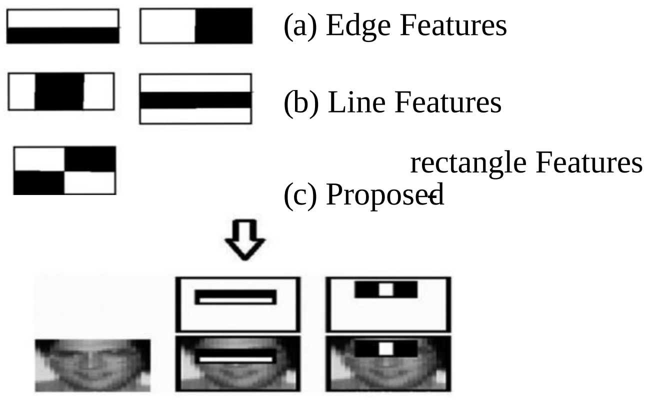

Fig. 3. Different types of features

The darker area's pixel values are reduced by the ones present in the lighter section. It is more likely that a HAAR feature has been found the nearer the value is to 1. Figure 2 displays pictures of various emotions taken from a dataset.

EMOTION CLASSIFICATION

The technique's main objective was to test multiple alternative architectures on the CNN to maximise overfitting while one way to enhance precision is by utilizing the validation set. The steps in the categorising of emotions are as follows:

Data Separation:

The FER2013 dataset is categorized into three usages: training, public testing, and private testing as indicated by the "Usage" label. The Training and Public Test sets were utilized to create the model, whereas Private Test set was utilized to appraise it.

Model creation and training:

The convolution layer generates a learnable filter that randomly convolves the input. The process creates a dot product between each input local area and the filter. A feature map—a 3D volume comprising various filters—is the product. The pooling layer in Convolutional Neural Network (CNN) is useful while decreasing the spatial dimensions of the input layer, thereby reducing computing expenses. The fully connected layer links every neuron from the preceding layer to the neurons in output layer, and the final layer (output layer) size corresponds to the count of classes. To prevent overfitting, an AF (activation function) is utilized. In convolution neural network, ReLU activation function is preferred due to its constant gradient of 1, meaning that most errors are communicated back during back-propagation.

$$f(x) = \max(0, x) \quad \text{------(1)}$$

Without any calculations, ReLU provides the gradient directly. This greatly aids in back-propagation, saving a significant amount of time and processing resources. Since most real-time data is not linear, neural networks must include nonlinearity in their layers. There is no saturation with ReLU because it has a linear function for x>0.

• SoftMax - The SoftMax function normalises an N-dimensional vector of real values into a range of values between 0 and 1.

$$\sigma(z)_i = \frac{e^{zi}}{\sum_{j=1}^{k} e^{zj}} \qquad \text{----------(2)}$$

$\boldsymbol{\sigma}$ = SoftMax
z = input vector, made up of (z0, ... zK)
$\boldsymbol{e^{zi}}$ = standard exponential function for an input vector
$\boldsymbol{K}$ = number of classes in the multi-class classifier
$\mathbf{e^{zj}}$ = standard exponential function for output vector

Each layer of inputs is normalized through batch normalization, which further increases the network's stability and training efficiency. Additionally, the batch normalization keeps the activation standard deviation and mean activation close to one and zero, respectively. This entail figuring out pixel values between 0 and 1 based on the image's brightness. The values of the brighter part's pixels are deducted from the values of the darker area. It is more likely that a HAAR feature has been found the nearer the value is to 1.

IV. EVALUATION MODEL

The validation set, which contains 3589 images, was used to test the model created during the training phase.

Modify to classify real-time images.

Real-time photos can be analysed for emotion recognition using transfer learning. Weights and values from the model built during the training phase can be used to construct new facial emotion detection challenges. FER becomes quicker for real-time photos because the model created already has weights. The model's convolutional architecture is displayed in figure 4.

Transfer learning is a method used in the field of machine learning that uses the information learned from training one model to inform the learning of a related model. A model can be trained on a sizable dataset of facial expressions to identify various emotions when it comes to emotion recognition.

The model's weights and values can be used as a starting point for developing new models those can be detect emotions in photographs once it was trained. Amount of time and resources needed to train a new model from scratch can be greatly decreased with this method.

Additionally, using weights that have already been trained can help the model be more accurate. This is due to the pre-trained model having already learned to recognise facial expression patterns that are characteristic of many persons.

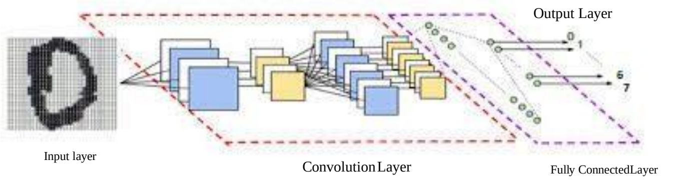

Fig. 4. Convolutional Architecture of the Model

V. RESULTS AND DISCUSSION

A test set of photographs from the original dataset and recently acquired webcam images were used to evaluate the rendering of the proposed image pre-processing ability. The results from both cases indicated accurate predictions on the proposed validation dataset. The performance metrics graph depicted the variations in different parameters across 50 epochs, demonstrating the effectiveness of the system. proposed model predicted outcomes are shown in figure 6.

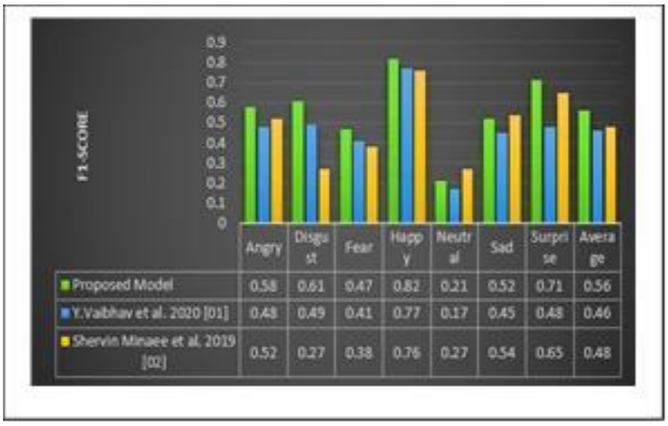

Fig. 5. Comparison

The proposed system of Emotional Analysis using Deep Learning is a significant improvement over the existing systems. It offers higher accuracy compared to the existing system.

TABLE 2. CLASSIFICATION TABLE

Emotions	Precision	Recall	F1 Score
Rage	0.51	0.67	0.58
Disgust	0.70	0.55	0.61
Fear	0.48	0.46	0.47
Joy	0.79	0.86	0.82
Neutral	0.26	0.18	0.21
Sorrow	0.52	0.52	0.52
Surprise	0.71	0.71	0.71
Average	56%	58%	56%

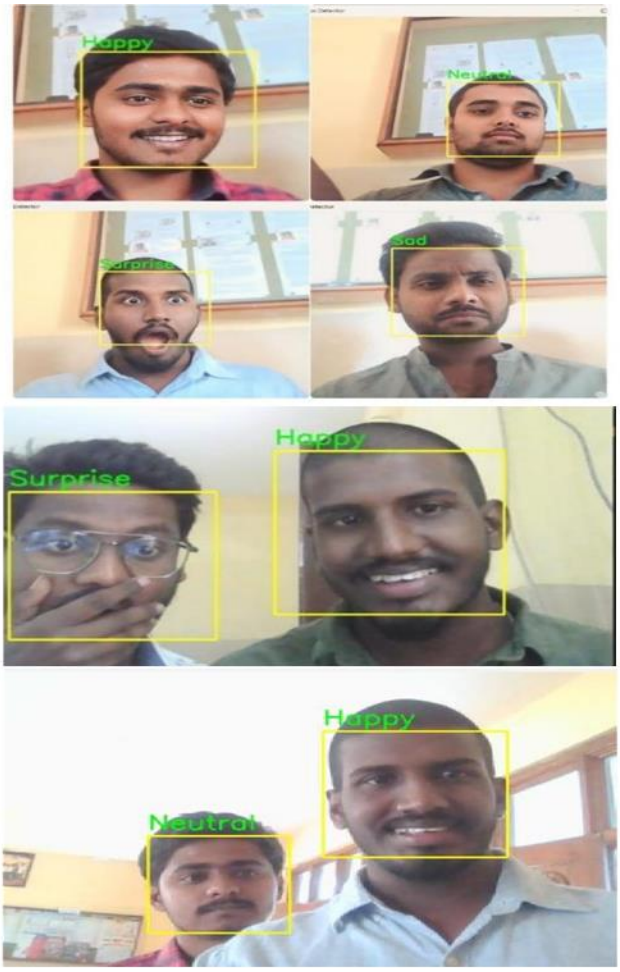

Fig. 7. The confusion matrix of the proposed model

The confusion matrix of the recommended model predicts various emotions such as rage, disgust, fear, joy, sorrow, surprise, and neutrality. You can find confusion matrix of the proposed model shown in Figure 7, and the classification report in Table 2.

True Negative (TN): The test result was unfavourable and indicated a negative result.

True Positive (TP): The test result was favourable and indicated a positive result.

False Negative (FN): The test result was affirmative, but it indicated a negative outcome in the future.

False Positive (FP): The test results showed a negative outcome; however, it was anticipated to be positive.

The performance of the proposed model is evaluated using the following metrics.

RECALL: A classifier's effectiveness lies in its capacity to identify all positive instances. To quantify this, we measure the true positives against the total number of correct predictions and incorrect predictions for each category.

$$\text{Recall} = \frac{TP}{TP+FN} \text{ -----------------(3)}$$

PRECISION: A classifier's capacity to identify an instance as negative instead of positive is known as negative predictive value. It is calculated for each group by dividing true positives by the sum of true positives and false positives. Precision is a measurement of how accurate positive forecasts are.

$$\text{Precision} = \frac{TP}{TP+FP} \text{ ----------(4)}$$

F1 SCORE: The F1 score is a mathematical calculation that combines precision and recall into a single score. The score value ranges from 0.0 to 1.0, with a high score value specify better performance. Unlike accuracy measures, F1 scores consider both precision and recall, making them a more reliable metric for comparing classifier models. As a rule, it's best to use the weighted average of F1 scores when comparing models, rather than relying solely on global accuracy.

$$\text{F1 Score} = \frac{2*(\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \text{ -------(5)}$$

The suggested methodology uses a webcam to anticipate a person's sentiment in real-time. By building the model with the CNN algorithm, this is accomplished. The camera detected several faces in the shot, this was successfully accomplished. The emotion that was generated was presented alongside the face in written form, and the text included any of the seven emotions that are universally acknowledged: Joyful, Astonished, Melancholic, Furious, Indifferent, Revulsion, and Terror. Since everything was dynamic instead of using still images like in earlier systems, the implementation was challenging. The output accuracy frequently deviated from the accuracy and varied greatly based on the lighting and camera level. Furthermore, finding an appropriate dataset was challenging. For the model to be trained to produce effective accuracy, a sizable dataset was needed.

In this situation, the right label is frequently the second most likely emotion when the model predicts erroneously. This study describes a face expression recognition system that uses a strong face recognition model. The model maps behavioural features with physiological biometric parameters to recognize diverse facial expressions such as joy, sorrow, fear, rage, surprise, and disgust. The geometrical structures of the human face that correspond to these expressions are used as a basis for the recognition system's template.

VI. REFERENCES

[1]. Lakshmi Bhadana ,P.V.Lakshmi , D.Rama Krishna ,G.Surya Bharti , Y.Vaibhav,'Real Time Facial Emotion Recognition With Deep Convolutional Neural Network'Journal Of Critical Reviews,Vol.7, pp.2394-5125, 2020.

[2]. Minaee S, Minaei M, Abdolrashidi A.” Deep-emotion: Facial expression recognition using the attentional convolutional network”,Vol.9,pp.3046, 2021.

[3]. Harshit Kumar, AyushElhance , Vansh Nagpal,'Facial Expression Recognition System',https://www.researchgate.net/publication/353414620_Facial_Expression_Recognition_System.

[4]. Arpita Santra , Vivek Rai , Debasree Das , Sunistha Kundu, 'Facial Expression Recognition Using Convolutional Neural Network', International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN,Vol.10, pp. 2321-9653, 2022.

[5]. Assoc Prof. Mrs. Jyothi S Nayak, Preeti G, ManishaVatsa, Manisha Reddy Kadiri, Samiksha S,' Facial Expression Recognition: A Literature Survey', International Journal of Computer Trends and Technology (IJCTT) – Vol.48 pp.2231-2803, 2017.

[6]. Prince Awuah Baffour, Henry Nunoo-Mensah, Eliel Keelson, Benjamin Kommey,“A Survey on Deep Learning Algorithms in Facial Emotion Detection and Recognition'.JurnalIlmiahBidangTeknologiInformasi dan Komunikasi, Vol.7 No.1, pp 2502-3470,2022.

[7]. Jin-Chul Kim, Min-Hyun Kim, Han-Enul Suh, 'Hybrid Approach for Facial Expression Recognition Using Convolutional Neural Networks and SVM', Appl. Sci,Vol.12, pp 5493, 2022.

[8]. ImaneLasri , Mourad El Belkacemi, Anouar Riad Solh, 'Facial Emotion Recognition of Students using Convolutional Neural Network', IEEE, Vol.03, pp.978-1-7281, 2019.

[9]. Md. Forhad Ali, Mehenag Khatun, Nakib Aman Turzo,' Facial Emotion Detection Using Neural Network ', International Journal of Scientific and Engineering Research,Vol.11, pp.2229-5518,2020.

[10]. Lakshmi Bhadana ,P.V.Lakshmi , D.Rama Krishna ,G.Surya Bharti , Y.Vaibhav,'REAL TIME FACIAL EMOTION RECOGNITION WITH DEEP CONVOLUTIONAL NEURAL NETWORK'JOURNAL OF CRITICAL REVIEWS,Vol.7, pp.2394-5125, 2020.

[11]. GuoluCao',Yuliang Ma,'XiaofeiMeng,'Emotion Recognition Based On CNN', Proceedings of the 38th Chinese Control Conference,Vol.38, pp. 8627-8630, 2019.

[12]. Rajesh Kumar G, Ravi Kant Kumar , Goutam Sanyal,' Facial Emotion Analysis using Deep Convolution Neural Network',International Conference on Signal Processing and Communication,Vol.17, pp.369-374, 2017.

[13]. Hai-Duong Nguyen, Sun-Hee Kim, Guee-Sang Lee, Hyung-Jeong Yang, In-Seop Na, and Soo-Hyung Kim, “Facial Expression Recognition Using a Temporal Ensemble of Multi-level Convolutional Neural Networks”,IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, pp.1949-3045,2019.

[14]. Song Luqin, “A Survey of Facial Expression Recognition Based on Convolutional Neural Network”,IEEE,pp.7281-1651, 2019.

[15]. Mingjie Wang, Pengcheng Tan,Xin Zhang, Yu Kang , CanguoJin, Jianying Cao1, “Facial expression recognition based on CNN”,Facial expression recognition based on CNN - IOPscience

[16]. Xiaoguang Chen, Xuan Yang, MaosenWang,Jiancheng Zou,“Convolution

Neural Network for Automatic Facial Expression Recognition", IEEE International Conference on Applied System Innovation, pp.978-1-5090-4897-7, 2017.

[17]. NING ZHOU , RENYU LIANG , AND WENQIAN SHI, " A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection", School of Electronic and Information Engineering, Lanzhou Jiaotong University,Vol.9, pp.5573-5584, 2021.

[18]. Tanoy Debnath1 , Md. Mahfuz Reza , Anichur Rahman1, Amin Beheshti , Shahab S. Band & HamidAlinejad Rokny, "Four layer ConvNet to facial emotion recognition with minimal epochs and the signifcance of data diversity". https://www.nature.com/articles/s41598-022-11173-0

[19]. Fuguang Yao and Liudong Qiu, "Facial Expression Recognition Based on Convolutional Neural Network Fusion SIFT Features of Mobile Virtual Reality", https://doi.org/10.1155/2021/5763626

[20]. Ninad Mehendale, "Facial emotion recognition using convolutional neural networks", https://doi.org/10.1007/s42452-020-2234-1

[21]. Hongli Zhang, Alireza Jolfaei, and Mamoun Alazab, "A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing", IEEE Access, Vol.7, pp.159081-159089, 2019.

[22]. Apeksha Khopkar and Ashish Adholiya Saxena, "Facial Expression Recognition Using CNN with Keras", Thomson Reuters ISI Web of Science Clarivate Analytics USA and Crossref Indexed Journal, Vol.14, pp.47-50, 2021.

[23]. Mehmet Akif OZDEMIR, Berkay ELAGOZ, AysegulALAYBEYOGLU, Reza SADIGHZADEH and Aydin AKAN, "Real Time Emotion Recognition from Facial Expressions Using CNN Architecture",

[24]. Xiaoguang Chen, Xuan Yang, Maosen Wang, Jiancheng Zou, "Convolution Neural Network for Automatic Facial Expression Recognition", 7 IEEE International Conference on Applied System Innovation, pp.814-817, 2017.

[25]. Saad Saeed, Asghar Ali Shah, Muhammad Khurram Ehsan, Muhammad Rizwan Amirzada, Asad Mahmood , and Teweldebrhan Mezgebo, "Automated Facial Expression Recognition Framework Using Deep Learning" https://doi.org/10.1155/2022/5707930

[26]. V. S. Amal, Sanjay Suresh , G. Deepa, " Real-Time Emotion Recognition from Facial Expressions Using Convolutional Neural Network with Fer2013 Dataset"002C, https://link.springer.com/chapter/10.1007/978-981-16-3675-2_41

Cite this article as :

D. Hari Krishna, Kottada Rakesh, Abhishek Kaveli, Kamsani Karthikeya, Maladoddi Pavan Kalyan, "Multiple Restaurant Table Booking System", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 4, pp.06-14, July-August-2023. Available at doi : https://doi.org/10.32628/CSEIT2390356

Journal URL : https://ijsrcseit.com/CSEIT2390356