2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR) | 978-1-6654-2521-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAITPR51569.2022.9844221

Facial Expression Recognition using Convolutional Neural Network

J.Avanija[1], K.Reddy Madhavi[2], G.Sunitha[3], Sreenivasa Chakravarthi Sangapu[4]
[1-4]*Sree Vidyanikethan Engineering College, AP, India*
[1]avans75@yahoo.co.in, [2]kreddymadhavi@gmail.com,
[3]gurramsunitha@gmail.com

Srujan Raju
CMR Technical Campus
Telangana,India.
ksrujanraju@gmail.com

Abstract **-** Facial Expressions pass on a lot of data outwardly instead of articulately. From the past few years, Facial Expression Recognition has been a challenging task in computer vision for Human-Machine Interaction as the way of expressing the emotions varies significantly. The main objective of Facial Expression Recognition (FER) systems is to detect an expressed emotion and recognize the same based on geometry and appearance features. Facial Expression Recognition is performed in four-stages namely pre-processing, face detection, feature extraction, and expression recognition to identify the seven key human emotions such as anger, disgust, fear, happiness, sadness, surprise and neutrality. The FER systems can be used in applications containing behavioural analysis on humans. This paper presents the comparison of different existing systems of Facial Expression Recognition.

Index Terms **-** *Deep Learning, Facial Expressions, Facial Expression Extraction, Expression Recognition*

I. INTRODUCTION

Human-Machine interaction is a field of study which focuses on designing computer technology to interact with humans. It is widely used in Human-Machine interaction, driver monitoring for autonomous driving, social communication, medical field, social robotics, education [2], real estate market analysis [16]. The main task of Facial expression Recognition is to detect human emotions by analysing facial expressions and to convey clear information about the emotions of people. Human brain can recognize emotions automatically without any delay but it is becomesl a challenging task for computers.

A machine can mimic the human brain and its functionalities by the advancements in technologies. Artificial intelligence, which refers to the simulation of human intelligence in machines, is accomplished by studying how the brain works, learns and the machines are programmed by the outcomes. By artificial intelligence, the communication between humans and computers gets simpler and simpler. Facial Expression Recognition is a technology which combines the machines with particular algorithms and uses them as a tool to detect human emotions. This can be achieved in different ways. Machine learning is one way to accomplish this by using different algorithms. Deep learning is another way which is called a subset of machine learning and it is inspired from the structure of the brain. Deep Neural Networks are more capable in extracting features from facial images.

In general, most of the researchers focus on the seven key human emotions. Sample images for the facial expressions are shown in Figure 2. Jianzhu Guo *et al.* [12] considered emotions by categorizing the 7 emotions into 50 categories such as happy, happily angry, happily surprised etc... in his work. Abdullah Talha Kabakus [1], Ninad Mehendale [8] and some of the researchers omitted the “neutral” expression as it is a challenging one which resembles almost all expressions.

The general architecture of Facial Expression Recognition systems contains 4 stages which are pre-processing, face detection, feature extraction and emotion classification. In Pre-processing, the input image is enhanced by reducing the noise and by some other techniques. In face detection, the input image is cropped by considering the face coordinates. In feature extraction, the image is transformed to represent the best features. This is an important stage as it results in influencing the performance of the system. The important features of an expression are eyes, nose, lips, ears and cheeks. Emotion classification is the last phase in which the extracted features are mapped with the labelled emotion classes to classify the facial images. Figure 1 presents the structure of the FER system.

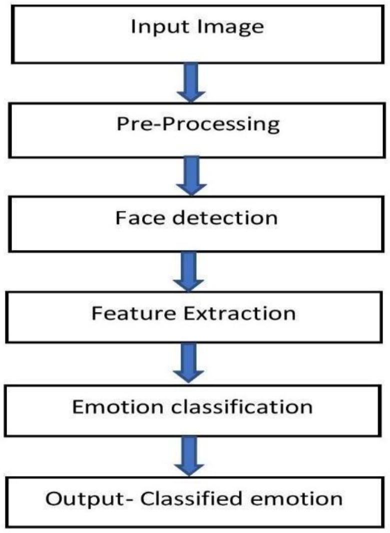


Fig. 1.Structure of Facial Expression Recognition

Fig. 2 Sample images of 7 Facial Expressions

II. LITERATURE SURVEY

This section focuses on some of the previous works that were done in this domain. In [14] Keyur Patel *et al.* discussed the comparison of various methods involved in facial expression recognition. The compared methods were Viola-Jones [4], [5], Principal Component Analysis for face detection, Local Binary Pattern (LBP) [11], [17], [18], Gabor filter, Scale Invariant Feature Transform (SIFT) [10] for feature extraction and Convolutional Neural Network (CNN), Support Vector Machine (SVM) [5], [10], Artificial Neural Network (ANN) for emotion classification.

Abdullah Talha Kabakus [1] proposed a model which detects the expressions effectively in various situations. The novedfhjl architecture was implemented by Python, OpenCV and some other open- source softwares containing image pre-processing, CNN modules (convolutional, pooling) and fully connected layer for classification. The system achieved an accuracy of 96% for CK+ dataset for 6 expressions but the accuracy was decreased to 85.71% when all seven expressions were considered and the analysis duration of the model is tolerable. In [2] Imane Lasri *et al.* proposed a model to recognize the student emotions from their faces and help teachers to modify their presentation accordingly. The proposed system used Haar cascades classifier which is trained by Adaboost algorithm to detect faces followed by CNN for emotion classification.

In [4] the author used Viola-Jones face detector along with scale normalization and gray level equalization for pre-processing the images. Kirsch edge operator was used to extract the edge information and the extracted features were trained by CNN to predict the expressions. Abir Fathallah [3] used CNN which is fine tuned with pre-trained model VGG. Sonia M. González-Lozoya [5] proposed a system that used Viola-Jones face detector that was integrated with VGG-face detector for face detection, pre- trained CNN.

In [6] the author in the proposed system used Multi cascaded convolutional neural networks (MTCNN) for real-time facial expression detection. The network model in the proposed system reduced the number of parameters in a convolutional neural network by replacing fully connected layers with Global Average Pooling and is suitable for the multi-classification of facial expressions. Even though it had achieved good results, it didn't perform well for images with light variations (dark and light) and blurred images. Si Miao [7] used Shallow CNN (SHCNN) to classify the static and micro expressions simultaneously without large datasets. The system overcomes the over fitting problem for small datasets. In [8] the author proposed a system that used two CNN's. First CNN is used for background removal from images by considering the shape, edges, textures along the faces and extracting facial features by generating the expressional vector. The second CNN for facial feature vectors. The background removal technique increased the accuracy compared to state-of-arts methods.

The author [9] proposed a deep model based on ensemble deep learning model. This method contained three sub- networks with variant depths. The recognition of the same emotion was performed by 3 sub-networks and probability vectors were obtained by these sub-networks and weighted average method was applied on those vectors to obtain the final vector. Argmax function was applied on the final vector to obtain the final class. The proposed system was simple and had less layers when compared to others but the computational efficiency was low and over-fitting may occur due to the use of the same training set for all sub-networks. In[10] the author proposed an approach that combined handcrafted features by Bag-of-visual-words (BOVW) and automatic features by CNN. Automatic features were obtained by VGG-face, VGG-f and VGG-13 CNN models were used and trained by Dense-Sparse-Dense training to prevent over-fitting. In the BOVW model, Dense SIFT and k-means clustering were used and finally SVM classifier was used for classification of emotions. This approach got a high accuracy of 75% for FER2013 and 59.58% for Affectnet. The limitation of the proposed system was that it can't be applied on video type of data.

Most of the systems were built by combining two or three models. In [11], the author proposed a Weight Mixture Deep Neural Network (WMDNN) for feature extraction in the proposed system which focused on two channels of facial images such as features of grayscale images were extracted by VGG16 model and features of Local Binary Pattern (LBP) by shallow CNN and finally, both were fused by weight and SoftMax classifier was used for emotion classification. The proposed system achieved an accuracy of 97% for CK+, 92.21% for JAFFE and 92.89% for oculu-casia. The ability of the proposed system to automatically extract features made the method implemented more easily than handcrafted features but there is need to simplify the network to speed up the algorithm. In [12] and [13], the authors focused on extending the facial expressions to 50 and 33 respectively. iCV-MEFED dataset was released by [12] in a challenge and some methods like Multi Modality network using visual and geometric information Though the methods achieved best accuracy, the processing time was slightly high. In [13], the author proposed a MSAU- NET, the model used dimensional model and the support vector machine (SVM) which made the image processing easier. Multi-Scale Action unit (MSAU) was also used for the fine grain facial recognition of the images. The Action Unit (AU) detection played a crucial role in discriminating the facial parts. Although the six basic facial expressions can be identified and recognized by the fine-grain model, the extra neutral facial expression was considered to be disadvantageous to this model.

Ai Sun *et al.* [15] proposed a system in which the facial expressions were recognized by three optimized active regions such as right eye, left eye and mouth instead of whole face region. For each active region a CNN was trained and classification was done by the frequency of each region. Face detection, landmark detection and rotation correction were done for searching optimized active regions. A decision level fusion was implemented for classification which increased the accuracy of CK+ (95%), JAFFE (96.57%) up to 2-3%. The recognition speed of the proposed system was fast but the searching time of active regions was too long. In [16] the author proposed a system for real estate market analysis using deep CNN. In trends and market preferences the response for change was highly flexible, but the accuracy for negative expressions was far more less when compared to positive expressions.

In [17] Jun Liu *et al,f*or feature extraction, a deep model was presented that incorporated Pyramid Histogram Orientation Gradient (PHOG), Edge Histogram Descriptor (EHD), and Local Binary Pattern (LBP).. The proposed model combined both VGG and the ResNet as the ResNet replaced the first part of VGG and SoftMax is used for classification. The limitation of the proposed system was it has to improve the rate of recognition for images with occlusion. Ji-Hae Kim *et al.* [18] proposed a system by combining the appearance feature based network that extracts the holistic features. These networks were built using CNN and achieved an accuracy of 96% for CK+ and 91.27% for JAFFE datasets. The main issue with the system was the cost of implementation was very high.

In [19], the author used a modular multi-channel deep CNN, consists of 4 steps such as image acquisition using OpenCV, face positioning and recognition. The model increased the performance of the network and has an accuracy of 68.4% for expression recognition, but in order to improve the performance, there is a need to reduce the size of the model. Shekar Singh [20] in the proposed system, the author used a CNN architecture without using any pre-processing techniques. In the proposed system Haar Cascade classifier was used for face detection and pre-trained dlib facial landmark detector was used for feature extraction. This architecture contains six convolutional layers along with RELU as an activation function. The accuracy of the proposed system on the FER2013 dataset is 61.7%. As there is no usage of any pre- processing techniques, the test accuracy is low.

Jia Xiang [21] proposed a system for facial expression recognition similar to the MTCNN framework. As we know MTCNN contains 3 networks (P-Net, R-Net, O-Net), the 3 tasks performed in this system were facial classification, boundary box regression and emotion classification. These networks were built using CNN and P-Net is trained with collected dataset and predicted results were given as input to R-Net. The predicted results of R-Net were given as training for O-Net which predicts the emotion.

In [22] data augmentation was involved in pre-processing and achieved an accuracy of 97% for CK+ and 86.7% for JAFFE. In [23] the system achieved an accuracy of 87% for both CK+ and JAFFE and proved that automated facial feature extraction methods were far more effective than hand-extraction feature methods. In [24], the pre-processing techniques used in the system were grayscale image conversion, face detection and applied emboss effect to images which made the images look like 3D. The system achieved an accuracy of 93% for CK+, 79.59% for JAFFE. In[24] the author used dropout technology to solve the problemof network overfitting.

The background analysis specified here reveals that the methods such as Viola-Jones, MTCNN, OpenCV for face detection, LBP, HOG for feature extraction and SVM, SoftMax for classification were used. Almost all the systems were built using CNN with different deep and light architectures. Moreover, most of the researchers focused on building the combination of two or three models by the utilization of pre-trained models for effective facial expression recognition. The issues for facial images such as images with different light variations [6], blurring [4], [6],occlusion [4], [17] and different face positions [3], [4] were not addressed in existing systems. Moreover, some of the existing models [1], [13], [21], [23], [24] did not provide accurate results for all expressions and generalization was not achieved in most of the cases.

III. METHODOLOGY

This section focuses on the various methods that were used in the previous works.

A. *Viola Jones Face Detection Algorithm*

It algorithm is broadly used to detect faces from images and it works well with human faces. Viola Jones algorithm initially detects the face on grayscale image and then identifies the face on a coloured image. It searches for the face within the outlined box by considering the features throughout the image by shifting the box.

Haar-Like features: Haar-Like features are digital image features used in Viola-Jones algorithm. It consists of light and dark regions. Edge, Line, four sided are the types of haar features shown in Figure 3 used in this algorithm. These features are used to extract useful information such as edges, straight and diagonal lines from an image that helps to identify the object. It produces the value of a feature by calculating the difference between the sum of pixels of light and dark regions.

- **Integral images:** As the dimension of generated haar feature is large and the computations of all features becomes heavy, integral images are used to reduce the processing time. It is utilized as a quick and efficient approach to calculate the sum of pixel values in an image or rectangular part of an image.
- ***AdaBoost training:*** Most of the calculated features are irrelevant so, AdaBoost algorithm is used to select the best among all features. The output of the algorithm is a strong classifier, which is the weighted sum of all weak classifiers which cannot classify on their own. These weak classifiers are found by iterating the algorithm 'x' number of times and in each iteration, it calculates and chooses the feature with least error rate.

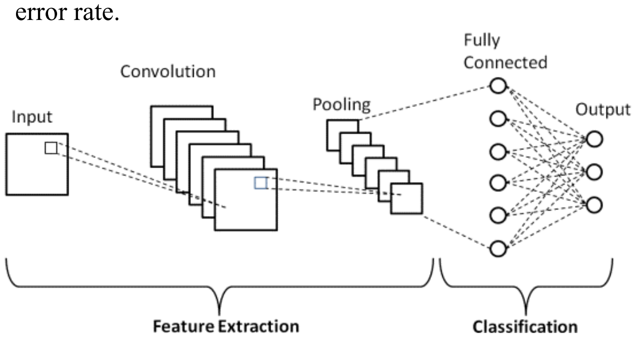


Fig. 3. Haar-Like features

- ***Cascading Classifier:*** Even the AdaBoost selects the best features but calculating the features for each region is still time consuming. A Cascade classifier is a multi-stage classifier in which each stage contains a classifier which is made up of some best features. The input forwarded from one stage to next only if the output of the classifier is positive, if it is negative then the input is discarded without forwarding. This procedure reduces the time consumption while calculating the features.

Viola Jones algorithm is fast and accurate but the training time is quite long. It is accurate for frontal face images but it is not effective in detecting the titled and turned faces.

B. *Convolutional Neural Network (CNN)*

Convolutional Neural Network is a deep learning neural network which is mainly used for image processing and classification. It is an algorithm which takes an image and is able to differentiate one from another with minimal pre-processing compared to other classification algorithms. Automatic detection of features without any human supervision is the main advantage of CNN compared to others. CNN architecture is built by using three types of layers: convolutional layer, pooling layer, fully connected layer. A convolutional layer can be followed by additional convolutional and pooling layers and the final layer is a fully connected layer. These layers are stacked together to form a deep model. The CNN architecture is shown in Figure 4.

- *Convolutional layer:* It acts as a feature extractor to extract the features from the input image. It contains learnable filters called kernels, which is a matrix of integers (trainable weights). The filter shifts by a stride throughout the image and performs a dot product with that portion of the image on which the filter is hovering in order to produce a feature map.
- *Pooling layer:* In order to reduce the dimensionality of the feature maps by selecting the best features, a pooling layer is used. There are two main types of pooling:
  - *i) Max Pooling:* It has a function which selects the pixel with maximum value from the selected input portion.
  - *ii) Average Pooling:* It has a function which calculates the average of all pixels from the select input portion and sends it to the output array.
- *Fully Connected layer:* It is a simple feed forward neural network which connects every neuron in one layer to the next layer. The output of the pooling layers is flattened and given as an input to this layer. This layer acts as a classifier to classify the images. It uses SoftMax, Sigmoid activation functions for image classifications. SoftMax activation function is widely used for image classification as it classifies the images by producing probability distributions.

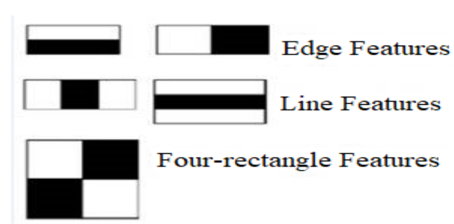


Fig.4. Convolutional Neural Network

In most of the papers they used CNN for feature extraction and classification purposes which gave effective results.

IV. Multi-Task Cascaded Convolutional Neural

Network(MTCNN)

Multi-task Cascaded Convolutional Neural Network (MTCNN) is used to detect faces and facial landmarks such as eyes, nose and mouth on images/videos. It is one of the most popular and most used face detectors because of its good generalization ability. It contains 3 stages which comprises deep neural networks such as Proposal Network (P-Net), Refine Network (R-Net) and Output Network (O-Net).

- *P-Net:* It is a shallow, fully connected network which does not contain any dense layers in its architecture. It creates multiple frames to scan the entire image, propose the candidate windows (which contains the face) and bounding box coordinates for that window. The highly overlapped windows are merged by using Non-Maximum Suppression (NMS).
- *R-Net*: The proposed candidate windows are given as an input to R-Net. It is a CNN, unlike P-Net it contains the dense layers as final layers in its architecture. It refines the candidate windows by reducing the number of candidates, adjusts the bounding boxes and again performs NMS. The output of R-Net is bounding box coordinates and the confidence of those boxes.
- *O-Net:* It is a complex CNN compared to the above two networks. The main aim of O-Net is to find the facial landmarks for the given candidate windows. It further reduces the number of candidate windows having low confidence. It outputs the adjusted bounding boxes and the coordinates of facial landmarks on images.

MTCNN is very accurate and it is effective in detecting faces with different sizes and rotations. But this is a bit slower when compared to Viola Jones algorithm.

D. *Local Binary Pattern*

Local Binary Pattern (LBP) is one of the recent texture descriptors. It creates a new image which represents the characteristics of the original image in a better way. It computes the local representation of textures on image, which is obtained by comparing each pixel with its neighbourhood pixels by taking a 3*3 windows. The procedure of this algorithm is

- Initially, the input image should be converted to gray-scale.
- For each pixel (gp) selects its p neighbours and considers the center pixel (gc) as threshold and performs a function ‘S’ which sets the adjusted pixel value as ‘1’ if that value is greater than or equal to center, otherwise ‘0’. A decimal equivalent is generated from these binary codes using the function shown in Figure 5 and this value is placed in gp.

$$LBP(gp_x, gp_y)\sum_{p=0}^{P-1} S(gp - gc) \times 2^p$$

Fig. 5 LBP operator

- As it uses 3*3 window, all pixel values range from 0-255 and a histogram is computed over the output LBP 2D array.

This LBP procedure was expanded to Circular LBP which is to use different numbers of radius and neighbours. This can be done by Binary Interpolation, which means if suppose a data point is present in between two pixels then it uses 4 nearest pixel (2*2) values to compute new data points. LBP texture descriptor is widely used because of its computationalsimplicity.

V. HISTOGRAM OF ORIENTED GRADIENTS

Histogram of Oriented Gradients (HOG) is a feature descriptor which focuses on the structure/shape of an object. It represents an image in compressed format by maintaining the shape and loses all insignificant features as shown in Figure 6. The steps involved to calculate the features are:

- Initially, it converts the image to grayscale and at the pre-processing stage, the image is resized to the width to height ratio to 1:2. Because, further the images are divided into 8 * 8 and 16 * 16 patches which are easy to calculate. The horizontal and vertical gradients are obtained by filtering images using “-1 0 1” kernels.
- Then its gradient magnitude and angle are calculated. From the obtained magnitude and angle bins are calculated for each using HOG.
- The 9 bins of histogram are compressed and converted to 9 feature vectors and this process continues for the entire image through sliding the window. Block normalization is applied to optimizethe features.
- In the final step all obtained vectors are combined to form features for the complete image.

This is a very powerful technique and achieved effective results in object detection because of its focus on shape of the object.

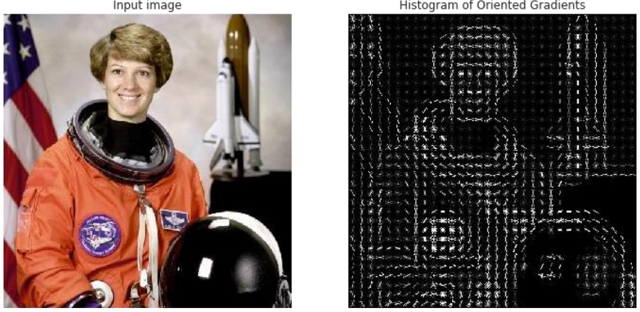
Fig. 5. Histogram of Oriented Gradients

VI. SUPPORT VECTOR MACHINE

Support vector Machine (SVM) is a supervised machine learning algorithm used for binary classification. It tries to find a hyperplane between the two classes for classification. In order to perform multi class classification, more than one SVM has to be implemented. This can be achieved by dividing multi classification problems into multiple binary classification problems. There are two approaches to achieve multiclass classification using SVM.

- *One-to-One:* In this SVM’s are implemented to find hyperplanes between every two classes. This means every hyperplane considers only two classes and neglects the rest. This approach requires “m(m-1)/2” SVM’s to classify m classes.
- *One-to-rest:* In this approach, it finds the hyperplanes that separate from the rest. This approach requires ‘m’ SVM’s to classify m classes as each SVM works to predict the membership among m classes.

SVM is used for classification in [5] and achieved best accuracy 98% for JAFFE and 94% for MUG datasets. In [10] the author implemented both global and local SVM’s by using One-to-Rest approach and concluded that the local SVM outperforms the global SVM.



VII. DISCUSSION

In [4] and [5], Viola Jones was used for face detection along with CNN using Fully connected layer and SVMrespectively, as classifiers. In [4], Viola Jones faced an issue with face positions along with some others but MTCNN [6] didn't face any issue regarding face position. The Classifiers SVM [5], SoftMax [18] produced 89.4%, 96.4% for CK+ and98.2%, 91.27% for JAFFE respectively. In [1], [22], CNN models produced 96.3%, 97.8% accuracy respectively. These CNN models differ in their architectures which influence the performance of the system. LBP images are used in [11], [18] in one of their models as they are combinations of models. These produced 97%, 96.46% for CK+ and 92.21%, 91.27% for JAFFE respectively. The accuracy for FER2013 dataset is comparatively less than other datasets due its size and real time facial images. The models [9], [11], [18] achieved best accuracy when compared to others which means the combination of models performs better than the individual models. So, the further focus should be on finding the best combination of models that increases the performance of the system. The summary of all FER system methods, performances and their limitations are shown in Table 1.

Table 1. Comparison of FER Methods

Methods	Accuracy	Limitation
CNN[1]	96.3%	Generalization problem
CNN[20]	61.7%	NoPre-processing techniques were included
Viola Jones + CNN[4]	88.56 %	It was not accurate for images with blurring, occlusion and different face positions
Viola Jones + CNN + SVM[5]	89.4%	Neutral expression was not considered in the proposed system.
CNN with Sub Networks[9]	96.4%	Training each network using the same data samples may cause a negative effect on recognition and may leads to overfitting.
MTCNN + CNN (with depth wise separable convolution layers) [6]	67%	It was not accurate for facial images with noise such as too strong/ dark lights and blurring
CNN[21]	97.8%	Generalization problem.
WMDNN (combines features from different channels such as grayscale and LBP)[11]	92.21%	The complexity of the network was high.
LBP + CNN + SoftMax (combines geometric and appearance features)[18]	91.27%	The cost of implementation was too high

VIII. CONCLUSION

This paper presents the comparison and detailed survey of various state-of-arts methods for Facial Expression Recognition systems. We have reviewed various previous works that are done in this domain. We have discussed various methods that were used in the previous works which are considered. A comparison has been done by considering the accuracy as a metric to evaluate the performance of the systems. We conclude that FER systems still face challenges with facial images that contain noise such as light variations, blurring, occlusion and the combination models performed better when compared to individuals.

REFERENCES

[1] Abdullah Talha Kabakus, "PyFER: A Facial Expression Recognizer Based on Convolutional Neural Networks", IEEE Access, vol. 8, pp. 142243- 142249, July 2020.

[2] Imane Lasri; Ano.uar Riad Solh; Mourad El Belkacemi, "Facial Emotion Recognition of Students using Convolutional Neural Network" IEEE conf… on ICDS, October 2019.

[3] Abir Fathallah; Lotfi Abdi; Ali Douik, "Face Expression recognition via Deep learning" IEEE conf… on computer system and applications, Nov 2017.

[4] Hongli Zhang; Alireza Jolfaei; Mamoun Alazab, “A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing”, IEEE Access, vol. 7, pp. 159081- 159089, October 2019.

[5] Sonia M. González-Lozoya; Jorge de la Calleja; Luis Pellegrin; Hugo Jair Escalante; Ma. Auxilio Medina & Antonio Benitez-Ruiz, “Recognition of facial expressions based on CNN features”, Multimedia tools and applications, 13987-14007, 2020.

[6] Ning Zhou; Renyu Liang; Wenqian Shi, “ A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection” , IEEE Access, vol. 9, pp. 5573 – 5584, December 2020.

[7] Si Miao; Haoyu Xu; Zhenqi Han; Yongxin Zhu; “Recognizing facial expressions using Shallow Convolutional Neural Network”, IEEE Access, vol. 7, pp. 78000-78011, June 2019.

[8] Ninad Mehendale, “Facial Emotion Recognition using Convolutional Neural Networks”, SN applied Sciences, February 2020.

[9] Wentao Hua; Fei Dai; Liya Huang; Jian Xiong; Guan Gu; “HERO: Human Emotions Recognition for Realizing Intelligent Internet of Things”, IEEE Access, vol. 7, pp. 24321-24332, February 2019.

[10] Mariana-Iuliana Georgescu; RaduTudor Ionescu; Marius Popescu, “Local Learning with Deep and Handcrafted Features for Facial Expression Recognition”, IEEE Access, vol. 7, pp. 64827-64836, May 2019.

[11] Biao Yang; Jinmeng Cao; Rongrong Ni; Yuyu Zhang, “Facial Expression Recognition Using Weighted Mixture Deep Neural Network Based on Double-Channel Facial Images", IEEE Access, vol. 6, pp. 4630-4640, December 2017.

[12] Jianzhu Guo; Zhen Lei; Jun Wan; Egils Avots; Noushin Hajarolasvadi; Boris Knyazev; Artem Kuharenko; Julio C. Silveira Jacques Junior; Xavier Baro; Hasan demirel; “Dominant and Complementary Emotion recognition from still Images of faces”, IEEE Access, vol. 6, pp 26391- 26403, April 2018.

[13] Liqian Liang; Congyan Lang; Yidong Li; Songhe Feng; Jian Zhao; “Fine Grined facial expression recognition in the wild”, IEEE Access, vol. 16, pp. 482-494, July 2020.Keyur Patel; Dev Mehta; Chinmay Mistry; Rajesh Gupta; Sudeep Tanwar; Neeraj Kumar; Mamoun Alazab; “ Facial sentiment Analysis Using AI techniques: State-of-art, Taxonomy and Challenges", IEEE Access, vol. 8, pp. 90495-90519, May 2020.

[14] Ai Sun, Yingjian Li, Yueh-Min Huang; Qiong Li & Guangming Lu , “Facial Expression Recognition using optimized Active regions”, Human centric computation and information sciences, November 2018

[15] MałgorzataRenigier-Biłozor; Artur Janowski; Marek Walacik & Aneta Chmielewska, “Human emotion recognition in the significance assessment of property attributes”, Journal of housing and built environment, March 2021.

[16] Jun Liu; Hongxia Wang; Yanjun Feng, “An End-to-End Deep Model with Discriminative Facial Features for Facial Expression Detection”, IEEE Access, vol. 9, pp. 12158-12166, January 2021.

[17] Ji-Hae Kim1, Byung-Gyu Kim 1, (senior member, IEEE), Partha Pratim Roy2, (Member, IEEE), and Da-Mi Jeong1; “Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure”, IEEE Access, vol. 7, pp. 41273-41285, March 2019.

[18] Zhou Yue; Feng Yanyan; Zeng Shangyou; Pan Bing; “Facial Expression Recognition Based on a Convolutional Neural Network”, IEEE conf… on Software Engineering and Service Science, 18-20 oct 2019.

[19] Shekhar Singh; Fatma Nasoz; “Facial Expression Recognition with Convolutional Neural Network”, IEEE conf… on computing and communication, 6-8 Jan 2020.

[20] Jia Xiang; Gengming Zhu; “Joint Face Detection and Facial expression Recognition with MTCNN”, IEEE conf. on International Science and Control Engineering”, 21-23 July 2017.

[21] Zhongzhao Xie; Yongbo Li; Xinmei Wang; Wendi Cai; Jing Rao; Zhenzhu Liu; “Convolutional Neural Networks for Facial Expression recognition with Few Training samples”, IEEE 37th Chinese Control Conference (CCC) ,25-27 July 2018.

[22] Minjun Wang; Zhihui Wang; Shaohui Zhang; Jiayu Luan; Zezhong Jiao; “Facial Expression Recognition Based on Deep Convolution Network”, IEEE conf… on Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 13-15 October 2018.

[23] Faten Khemakhem; Hela Ltifi; “Facial Expression Recognition using Convolutional Neural Network enhancing with pre-processing stages”, IEEE conf... on computer systems and applications, 3-7 November, 2019.

[24] Jiancheng Zou; Xiuling Cao; Sai Zhang; Bailin Ge; “A Facial Expression Recognition based on Improved Convolutional Network”, IEEE conf… on Intelligent Applied Systems of Engineering,26-29 April, 2019.