IJCSNT, Vol.8, No.1, 2019
DOI-10.18486/ijcsnt.2019.8.1.01

Facial Expression Recognition using Deep Convolutional Neural Network

SHRADHA DUBEY
Department of Computer Science & Engineering and Information Technology,
Madhav Institute of Technology and Science, Gwalior India
e-mail: dubeyshradha29@gmail.com,

MANISH DIXIT
Department of Computer Science & Engineering and Information Technology,
Madhav Institute of Technology and Science, Gwalior India
e-mail: dixitmits@gmail.com

Abstract *- Human Facial Expression Recognition (FER) is always been a challenging task for researchers. Face expressions play an important role in non-verbal communication. Nowadays, deep learning is achieving great attention in this area. But it is a tedious task to build a simple and effective architecture in deep learning. The need is to develop an architecture which is fast to train and achieves good accuracy. Therefore, for detecting human facial expressions deep learning model using a convolutional neural network(CNN) is used in image classification as these models are effective and achieve high accuracy and effictiveness for the facial expression recognition problems. This work is proposed to acquire results improvement on FER-201, JAFFE, and CK+ dataset. Python platform is used for implementation and execution of the proposed work.*

Keywords*-FER, CNN, Shallow CNN, Deep CNN.*

I. Introduction

The emotions which bring faltering in the face muscles are known as facial expressions. Emotion is a word used to represent a feeling of a person at that instant of time[1]. These are very much useful in the area of computer vision and the interaction between humans and computers. The human-computer interaction(HCI) is highly inspired by human-human interaction. It has its application in many areas such as can be used for security purposes, medical management, driver safety and many more. From the facial expressions on can be able to predict the mood or behavior of the person thus, it plays a major role in non-verbal communication[8].

During a group discussion, whoever puts more emphasis on his/her words or speak in a sentimental or energetic tone are much more noticed as compared to others. The same procedure takes place with the nonverbal visual communication.

The facial images here are categorized into the different facial emotion categories, namely, happy, fear, angry disgust, sad, surprise and neutral are considered in this study. Presently, a deep neural network is extensively used in all the image processing and computer vision techniques as it has the potential to simply handle spatial images[23]. The deep neural network is also able to give better results with a large and variety of datasets. The objective of this work is to propose a deep neural network model that includes considerable layers of convolution and profound residual blocks for recognition of human facial emotion. Convolutional Neural Networks(CNN) was designed to simplify the feature selection procedure and produce many accurate results as compared to other existing methods. The work presented here may be considered as the groundwork for overall automated classification of human emotion system for use in

various applications including E-learning, emotion surveillance, lie detection, pain assessment and so on. The rest of the paper is divided as in the second section various related studies have been seen. Then in the rest of the paper, the proposed work with the experimental results has been discussed. Lastly, the paper is concluded along with its future scope.

II. Related Study

Pengyuan Liu et.al.[2019] They have presented Cause Emotion Action Corpus(CEAC) and proposes two novel experiments named as emotion causality and emotion inference[1]. The CEAC not only used to determine emotions but it is also useful for analyzing cause events and action events. They have analyzed the baseline performance which shows that there is much more improvement is needed in both the tasks.

Tarik A. Rashid[2018] In this paper, the author has used different classifiers namely Multilayer Perceptron (MLP), Decision Tree and CNN to determine emotion recognition accuracy and came to the result that CNN produces the best recognition accuracy. They have first used preprocessing of data followed by balancing the unbalanced dataset, next the significant features were extracted for emotion recognition and finally, these features are applied in a classifier model as input[2].

Hakan Boz et.al.[2018] The authors have implemented the Artificial Intelligence(AI) technique for recognizing human emotions. They have trained the system using the Vortex Optimization Algorithm by using different types of Human Recognition datasets and uses the Cascade Feedforward Artificial Neural Network.

Saeed Turabzadeh et.al.[2018] In this literature, the emotions have been recognized in a real-time environment and the dataset used is implemented from videos. To detect the emotions at a faster frame rate in real time environment the field-programmable gate array (FPGA) is developed. Digilent VmodCAM camera sensor is used in this study and ASTM Apartan-6 FPGA is used to build the model[3]. For displaying video in real time as well as for predicting labels of the emotions the graphical user interface is used.

Mostafa Mohammadpour et.al. [2017] — In this work, CNN is developed for facial expressions recognition. They have worked on seven basic human emotions and uses the Cohn-Kanade database and achieved better results. The novelty of their work is that they have used facial Action Units(AUs) to achieve more accurate results.

MD. Zia Uddin et.al.[2017] In this paper, for recognizing the facial expressions a depth camera-based novel technique is implemented. For efficient emotion detection, the rank of each pixel in the depth picture is calculated here by eight local directional pixels. For each pixel in a depth image, the eight surrounding directions and eight histograms are implemented by using calculated ranks. They have then concatenated these histograms for depth image feature representation[10]. The various techniques are used in this literature for the better features and at last, they have trained the features using deep learning approach for obtaining much more accurate results.

Viraj Mavani et.al.[2017] A CNN is designed on the different datasets for determining human emotion detection. The model is trained and tested independently on the different facial expression datasets[9]. Further, for predicting saliency the method used by them is the Deep Multi-Layer Network and they have also observed the general confusion as exhibited by humans between various expressions.

III. Datasets

Deep Neural Network generally have the neccessity of huge amount of data for training. Moreover, the performance of the model depends on the selection of images used for training data, therefore, the selection of high qualitative and quantitative data must be done to achieve good performance. There are various types of facial recognition dataset that are available for human emotion detection.

FER-2013 dataset, which consists of about 36,000 images of 48x 48-pixel resolution which indicates that the width and height of the image is 48. The six basic emotions plus neutral emotion is used for prediction and are labled from 0 to 7 as angry, disgust, fear, happy, sad, surprise[24] and neutral respectively and batch size of 64 is considered for the input processing task. Following figure shows some samples from every class of facial expression . Along with the labels defined for images from 0 to 6, the images in the dataset are split up into three distinct sets which are training, validation, and test sets. These sets contains 28709 images for training, 3589 images for testing, and the remaining 3589 is used for validation. This set of images is choosen as it is a challenging dataset. The images are not aligned properly and a part of them are not correctly labeled and these challenges make the classification task tough as the model should be robust and well formed.

Besides FER 2013 dataset, the Extended CohnKanade (CK+)[15], and the Japnese Female Facial Expression (JAFFE)[16,19], are used. JAFFE database contains the 7 facial expressions which has been discussed earlier , as the name suggests JAFFE, contains 10 Japanese female models. This database consists of 60 Japanese subjects of seven basic emotions. 60 Japanese subjects rated each picture on six emotional adjectives. On the other hand, CK+ dataset consists of images of both male and female. CK+ provides rules and baseline outcomes for the tracking of facial feature, the unit of action and recognition of emotion.

The datasets is distinguished on the basis of quantity, quality, and ‘cleanness’ of the images. The emotions ‘in the wild’ are shown FERC-2013 set which makes the images harder to recognize but the robustness of the model can be achieved due to its large size. On the other hand, the set of pictures in JAFFE and CK+ dataset are posed. Following figures shows the sample images from JAFFE, FER-2013, CK+ dataset respectively.

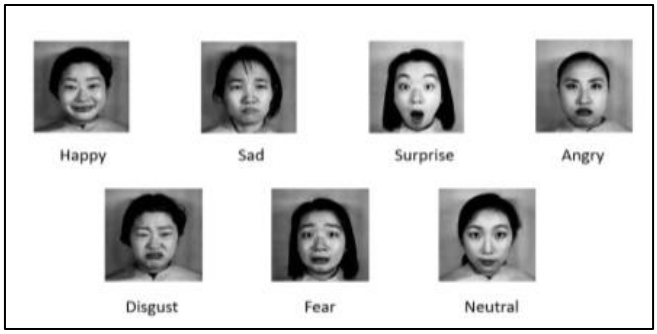

Fig.1. A JAFFE database sample images along with their corresponding emotions

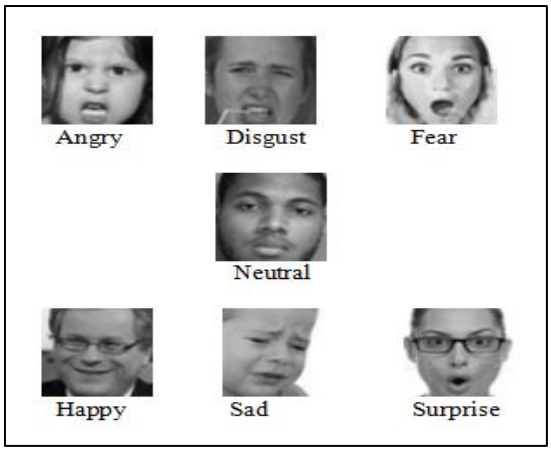

Fig.2. A FER2013 database sample images along with their corresponding emotions.

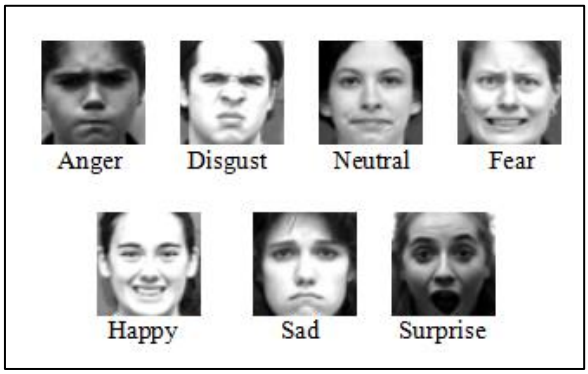

Fig.3. A CK+ database sample images along with their corresponding emotions

IV. Proposed Work

The CNN learning techniques namely the shallow and deep CNN learning technique is compared. In this, "Shallow" neural networks is a term used to describe NN that usually has only one hidden layer[25] as compared to deep NN which have several hidden layers, often of various types. In

the recent study, it has been analyzed that deep NN with the correct architecture performs better than shallow ones with the same computing power[8](eg. No. Of neurons or links). Comparision of machine learning approach and deep learning to differentiate facial expression is shown in fig.4.

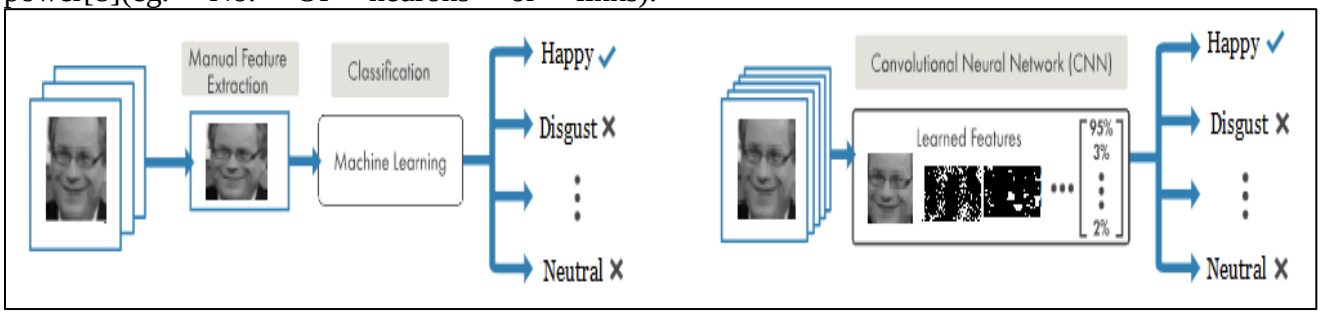

Fig.4. Comparision of a machine learning approach (left) to categorizing facial expression with deep learning (right).

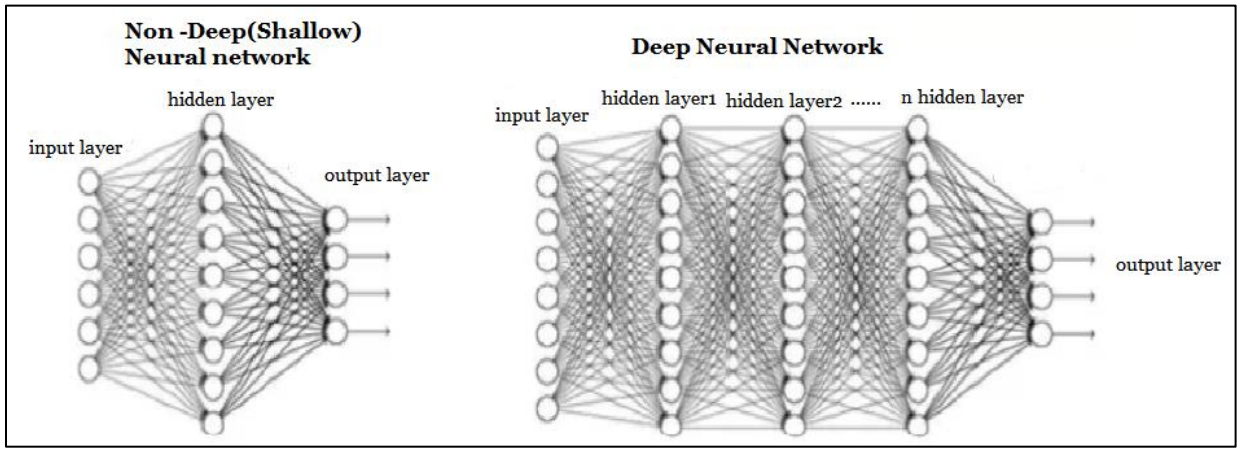

Fig.5. A simple Shallow and Deep Neural Network Architecture

Here, the convolutional neural network(CNN) is used as a deep neural network architecture for FER[10]. Convolutional Neural Networks are very comparable to normal Neural Networks, they consist of neurons that have weights and biases. Every neuron in this network acquires some inputs, then performs its dot product and deliberately succeeds it with a non-linearity. A single differentiable score function is still expressed throughout the network from the raw image pixels on one end to class results at the other. Convolutional Neural Networks take advantage of the reality that the input consists of images and they restrict the architecture sensibly[3]. Specifically, unlike a standard Neural Network, a ConvNet's layer have neurons organized in three dimensions which represent width, height, depth(activation volume).

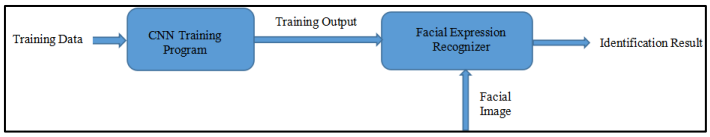

Fig.6. System Design

Firstly, before giving the data for the input, the is processed and the human face is detected, after the detection the face is being cropped and the is used for the further processing. In the proposed model system, out of the two architectures discussed, the shallow CNN architecture consists of convolutional layer and fully connected(FC) layer[25], the number of both the layers is two and

one respectively. The convolutional layer used priorly is consist of 64-3x3 filters. In addition to these filters, batch normalization to protect the input f rom being scattered, 2-dimension maxpooling of 2x2 filter and dropout function to reduce overfitting is used. Similarly, the second convolutional layer architecture, with 128-3×3 filters, batch normalization, dropout and 2d-maxpooling laeyer as used earlier. Lastly there is a FC layer, with 512 neurons in its hidden layer and for the working of loss function softmax is used.

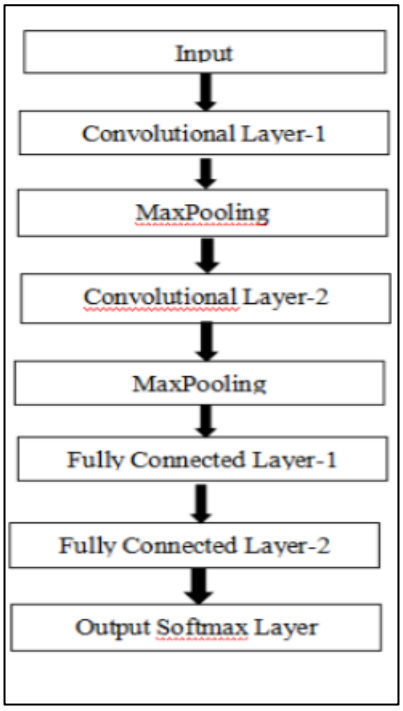

Fig.12. CNN architecture for Shallow neural network

In the proposed, in deep CNN[2], the network contains 2D convolutional layers and fully connected layer which are four and three respectively in number. The convolutional layer, designed in the architecture contains 3×3 kernel size, along with batch normalization, dropout and ReLU as activation function. The convolutional layer is accompanied by max-pooling with a pooling window of 2X2 and stride as 2×2 which discards 75 per cent of the activation by downsampling each depth slice in the input by 2 in width and height. After this a flatten layer is used which flattens the input from ND to 1D without affecting the batch size. The deep network uses the cross entropy loss along with Adam optimizer. Generally the system operates in 80-10-10 ratio for training-validation-test sets respectively.

Apart from the FER-2013 dataset the deep neural network is also trained and tested on the JAFEE and CK+ dataset and achieved a better recognition accuracy on all the datasets as compared to earlier designed models (SVM, Shallow CNN) for FER.

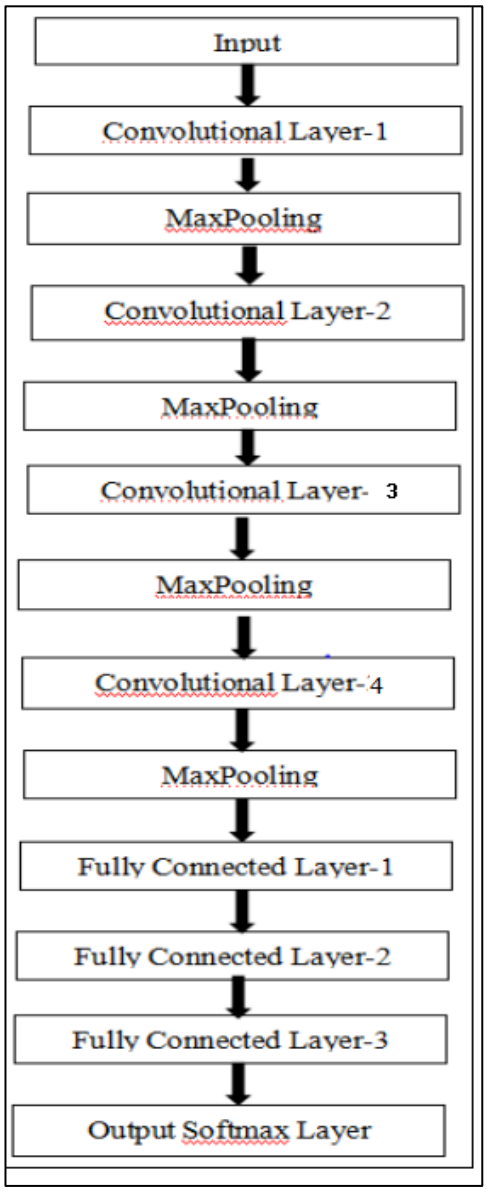

Fig.8. CNN architecture for Deep CNN

Convolutional Neural Network(CNN) or ConvNet are made up of neurons that have learnable weights and biases. CNN architectures creates a specific hypothesis the images are the inputs which allows us to encode certain characteristics into the architecture. These then make the forward feature more effective to execute and significantly decrease the quantity of parameters in the network. The various layers of the architecture is described below-

a) **Convolutional Layer** Convolution is a mathematical operation that combines two features to create a third function. For extracting features from the input image the initially used layer in the CNN architecture is the convolutional layer [7]. This layer works on an input layer and to generate a feature map filter is used. It takes the input as an image matrix and a filter or kernel.

The matrix dimension of an image is in the form of (h x w xd)

where, h, w, and d represent the height, weight, depth respectively. The basic design of convolution layers is shown in fig.9.

The product of this image matrix dimension is done with the filter of the form (f_h x f_w x d). Here, the depth of the image remains unchanged. Finally, the output is obtained which is represented as (h- f_h+1) x (w- f_w + 1) x 1 and this is known as volume dimension output.

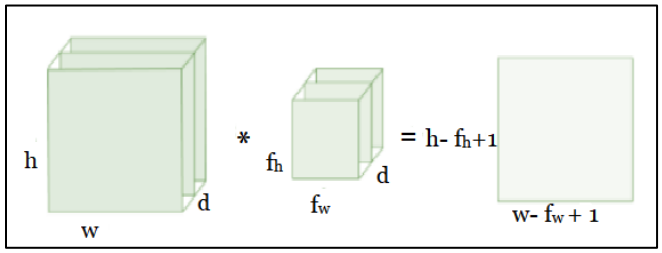

Fig.9. A basic design of Convolutional Layer

b) **Pooling Layer** In the ConvNet architecture, the spatial size is reduced to scale down the parameter amount as well as the computation power of the network and hence the overfitting in the network is also be controlled. On every input depth slice, the pooling layer works independently and resizes it spatially, using the different pooling operations. Here we use the MAX pooling which is commonly used with a stride size of 2.To calculate the largest patch of every feature map the max pooling operation is performed. This gives better performance results in computer vision work as compared to other pooling techniques like average pooling. The function of pooling layer is shown in fig.10. The pooling layer is computed on the basis of following parameters[11].

- It receives a volume size of $W_1 \times H_1 \times D_1$, where W_1, H_1, D_1 represents the width, height, and depth of the accepted input respectively.
- The two parameters are needed: their spatial extent which is represented as G, the stride S,
- Finally, the output is produced of volume $W_2 \times H_2 \times D_2$, where $W_2 \times H_2 \times D_2$ is calculated as-

$W_2=(W_1-G)/S+1$ eq.(1)

$H_2=(H_1-G)/S+1$ eq.(2)

$D_2=D_1$ eq.(3)

- Introduces zero parameters since it computes a fixed function of the input
- For Pooling layers, it is not usual to pad the input using zero-padding

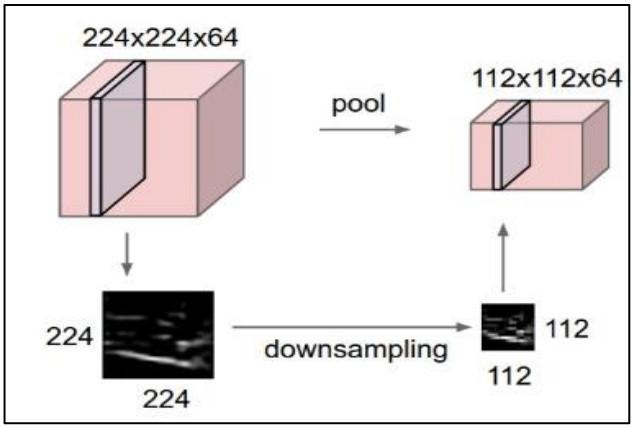

Fig.10. The function of Pooling Layer

c) **Normalization Layer** We normalize the input layer by changing and scaling the activations, the input is normalized. For instance, if the features are in the range 0 to 1 and some are from 1 to 1000 then in order to speed up the learning they should be normalized Batch normalization decreases the quantity by the value of the hidden unit (covariance shift)[13].

d) **Fully Connected(FC) Layer** The input of the FC layer is the result which we obtained from the final pooling layer.The output of the final pooling layer acts as an input to the fully connected layer. The feature vector in the input is represented as a fully connected layer and it holds the information which is essential for the input. During training, this feature vector is being used to determine the loss, and help the network to get a train[11]. Each conv layer holds several filters that represent one of the local features. The FC layer holds composite and aggregated information from all the conv layers that matters the most.FC layer uses SoftMax as the final classification layer to predict the given input category[13]. The result generated by the FC layer with neuron size 'n' and with input x will be as follows:

$$F(x) = A(\sum_{i=1}^{n} W * x)$$

where 'A' is an Activation function,'W' is a Weight matrix.

e) **Rectified Linear Unit(ReLU)** The ReLU, is not a distinct element of the method of CNN. The aim of implementing the feature of the rectifier function is to enhance the non-linearity of our images[17]. The need to do this is because images obtained are generally non-linear in terms of their pixels transition, the color or edge difference, etc. Rectified linear unit is shown in fig.(11)

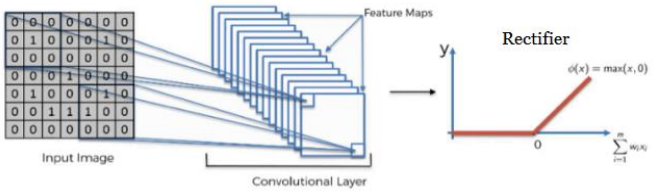

Fig.11.Rectified Linear Unit Operation.

V. Experimental results

The architectures compared here for facial expression recognition are SVM, shallow neural networks and deep CNN using all the three datasets. We are fitting the model using appropriate epochs with 0.001 learning rate and it is compiled using categorical_crossentropy as the loss function to get the better classification results. We have concluded that better identification accuracy is achieved using the deep Convolutional Neural Network. The following figures show the results obtained from different datasets along with their test results which are acquired by testing the result on different images.

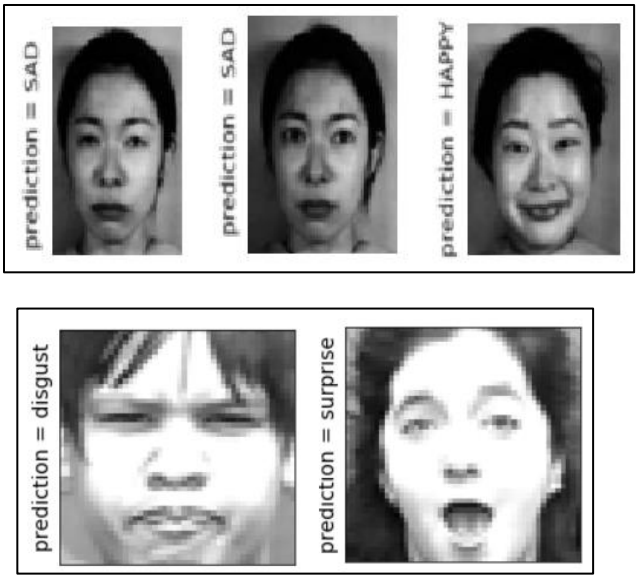

Fig.12. Recognition results of different datasets after training the data

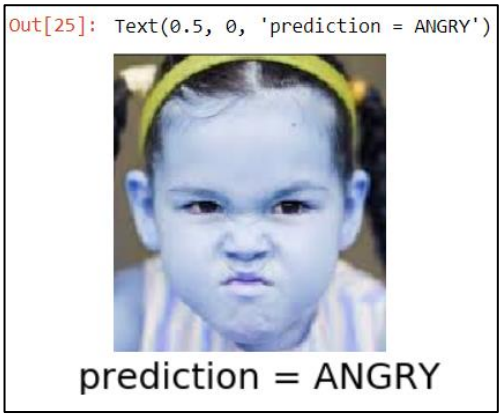

Fig.13. Prediction Results after testing the model

Table1. Accuracy Recognition of different techniques on various datasets

Recognition Accuracy			
Dataset	**SVM**	**Shallow CNN**	**Deep CNN**
FER-2013	50.5%	56.56%	66.75%
JAFFE	95.60%	94.40%	97.90%
CK+	95.10%	95.00%	97.29%

Recognition accuracy is observed on various datasets as shown in Table1 which shows Deep CNN performance accuracy is morw in case of all selected datasets

VI Conclusion

In this study, the human facial expressions have been recognized and tested on the various image datasets which make an interactive environment between humans and computers. Various types of datasets are analyzed and results are compared with different algorithms that are used for facial expression recognition to recognize different emotions in a better way. The work then has been tested on various pictures. Among the different FER algorithms, we have concluded that the deep Convolutional Neural Network which uses the full face as an input in this study is determined to produce the best recognition accuracy.

References

[1] Pengyuan Liu,, Chengyu Du , Shuofeng Zhao, "Emotion Action Detection and Emotion Inference: the Task and Dataset",arXiv:1903.06901v1 [cs.CL] 16 Mar 2019.

[2] Tarik A. Rashid, "Convolutional Neural Networks based Method for Improving Facial Expression ,publication/328737102

[3] Saeed Turabzadeh, Hongying Meng, "Facial Expression Emotion Detection for Real-Time Embedded Systems"MDPI,26 January 2018

[4] Hakan Boz,Utku Kose, "Emotion Extraction from Facial Expressions by Using Artificial Intelligence Techniques",BRAIN – Broad Research in Artificial Intelligence and Neuroscience, Volume 9, Issue1 (February, 2018), ISSN 2067-8957.

[5] Shan Li and Weihong Deng, "Deep Facial Expression Recognition: ASurvey",arXiv:1804.08348v2 [cs.CV] 22 Oct 2018

[6] Rizwan Ahmed Khan,Alexandre Meyer,Hubert Konik, "Saliency-based framework for facial expression recognition",Higher Education Press and

Springer-Verlag GmbH Germany, part of Springer Nature 2018

[7] Shima Alizadeh,Azar Fazel, “Convolutional Neural Networks for Facial Expression Recognition”,arXiv:1704.06756v1 [cs.CV] 22 Apr 2017.

[8] Mehdi Ghayoumi, “A Quick Review of Deep Learning in Facial Expression”,Journal of Communication and Computer 14 (2017).

[9] Viraj Mavani,Shanmuganathan Raman,Krishna P Miyapuram, “Facial Expression Recognition using Visual Saliency and Deep Learning”,2017 IEEE International Conference on Computer Vision Workshops

[10] Md. Zia Uddin,Weria Khaksar,Jim Torresen,“Facial Expression Recognition Using Salient Features and Convolutional Neural Network” 10.1109/ACCESS.2017.2777003,December 22, 2017.

[11] Mostafa Mohammadpour,Hossein Khaliliardali, “Facial Emotion Recognition using Deep Convolutional Networks” IEEE 4th I International Conference on Knowledge-Based Engineering and Innovation (KBEI) Dec. 22QG, 2017.

[12] Dinh Viet Sang,Nguyen Van Dat,Do Phan Thuan, “Facial Expression Recognition Using Deep Convolutional Neural Networks”,2017 9th International Conference on Knowledge and Systems Engineering(KSE).

[13] Christopher Pramerdorfer, Martin Kampel, “Facial Expression Recognition using Convolutional Neural Networks: State of the Art”,arXiv:1612.02903v1 [cs.CV] 9 Dec 2016.

[14] Mehang B. Patel,Dipak L .Agrawal, “A Survey Paper on Facial Expression Recognition System”February 2016, Volume 3, Issue 2 JETIR (ISSN-2349-5162)

[15] Daniel Llatas Spiers, “Facial emotion detection using deep learning”,Uppsala University,June 2016.

[16] Haval Abdulkarim Ahmed,Tarik A. Rashid,Ahmed T. Sadiq, “Face Behavior Recognition Through Support Vector Machines”,(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 1, 2016.

[17] Ali Mollahosseini, David Chan, and Mohammad H. Mahoor, “Going Deeper in Facial Expression Recognition using Deep Neural Networks”arXiv:1511.04110v1 [cs.NE] 12 Nov 2015.

[18] Evangelos Sariyanidi, Hatice Gunes, and Andrea Cavallaro, “Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition”,IEEE Transactions on pattern analysis and machine intelligence, Vol. 37, No. 6, June 2015.

[19] Anchal Garg , Dr. Rohit Bajaj, “Facial Emotion Recognition and Classification Using Hybridization Method ”,International Journal of Engineering Research and General Science Volume 3, Issue 3, May-June, 2015

[20] S.Ramya, G.Sophia Reena, “Facial Emotion Recognition Using Optimization Technique ”,International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 11, November 2013.

[21] Maedeh Rasoulzadeh, “Facial Expression recognition using Fuzzy Inference System”,International Journal of Engineering and Innovative Technology (IJEIT) Volume 1, Issue 4, April 2012.

[22] Jagdish Raheja, “Human Facial Expression Detection From Detected in Captured Image using Back Propagation Neural Network”,IJCSIT,Vol.2,No.1,February2010.

[23] Heechul Jung1, Sihaeng Lee, “Development of Deep Learning-based Facial Expression Recognition System”.

[24] Hong-Wei Ng, Viet Dung Nguyen, “Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning”, proceedings of the ACM International conference on Multmodal Interaction, pp. 443-449, 2018.

[25] Deepika Dubey, GS Tomar, “Echelon Based Pose Generalization of Facial Image Approaches”, Asia Pacific Journal of Convergent Research Interchange, Vol.3, No.1, pp.63-75, March 2017.

[26] Junnan Li ; Edmund Y. Lam, “Facial expression recognition using deep neural networks”. IEEE International Conference on Imaging Systems and Techniques (IST), pp.1-4, 2015

[27] Ivan Gogic, Martina Manhart, “Fast facial expression recognition using local binary features and shallow neural networks”. The Visual Computer, pp.1-16, 2018.

[28] Deepika Dubey, GS Tomar, H. Kim "Scrutinized Study on Face Recognition by Pose invariant Methodology", International Journal of future Generation Communication and Networking, vol. 9, No. 9, pp.333-342, 2016.

Authors Profile

Manish Dixit received his B.E. in Computer Technology from Barkatullah University, Bhopal, M.E in Communication Control and Networking from MITS, Gwalior and PhD from Rajiv Gandhi Technical University, Bhopal. He is currently working as Professor in the Department of Computer Science and Engineering and Information Technology, MITS, Gwalior, India. He has organized more than 10 International Conferences of IEEE in different capacities. He has presented more than 85 research papers in National and International Conferences and Journals. He is a Fellow Member of IETE, Senior Member of IEEE and Secretary IEEE MP Subsection. He is member of IET, IAENG and CSI.

Shradha Dubey received her B.E. in Computer Science and Engineering from Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal in the year of 2016. She is currently a PG research scholar in the Department of CSE/IT, Madhav Institute of Technology and Science, Gwalior, India.She is a member of IEEE.