iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IX **Month of publication:** September 2021

DOI: https://doi.org/10.22214/ijraset.2021.38064

Performance Comparison of Different Convolutional Neural Network Approaches for Facial Expression Recognition

Suhrid Shakhar Ghosh[1], Md. Yasir Arafat[2], Abdullah Al Noman Sarkar[3]
[1, 2, 3]Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology

Abstract: Facial expression is a non-verbal way of communication to express the human state of mind using facial muscles. Happiness, sadness, anger, surprise, disgust, fear, and neutral expressions are widely used in the field of medical rehabilitation, sentiment analysis, counseling, and so on inspiring researchers to develop effective models to classify the expressions effectively. LeNet5, AlexNet, Deep Model, Shallow Model, Deep CNN Model are some commonly used models that have been developed to recognize facial expressions using machine learning and deep learning. In this research, a new convolutional neural network model has been proposed and compared with the existing models. The FER-2013 dataset has been used to evaluate the performance using different metrics to find the efficiency of the models. The proposed model provides comparatively better accuracy than most of the existing models, which is 64.4%.

Keywords: Facial Expression, Image Classification, Convolutional Neural Network, Deep Learning, Non-verbal communication.

I. INTRODUCTION

Facial expression includes movement of face structure and face outlook[1]. For communication purposes, it is used both verbal and non-verbal way, using verbal approach is inherent and easy to understand whereas non-verbal communication is most opposite to another way. Understanding facial expression by machine is still a challenge [2]. Researchers are investing so many times in this era to recognize human facial expressions automatically with robust accuracy and efficient ways.

There are almost seven categories of expressions that are recognized so far including happiness, anger, sadness, neutral, disgust, fear, and surprise. Some of the expression identification rates are acceptable and comparatively high whereas some expressions identification rate is slightly poor and needs an improvement. Since the recognition of facial expression can be applied to directly Behavioral Science and Medical Rehabilitation, so accuracy rate and classification method convey an important role here. Recognition of facial expression is a challenging task. At the same time, developing an automated expression recognition system is getting a trend because of its tremendous application. Communicating with physically disable or autistic people is not so easy since they use non-verbal ways to exchange feelings. Developing a facial expression recognition system can make it easier of understanding autistic people's emotions. Besides, it’s has a direct impact on medical rehabilitation, behavioral science, and so on. Facial expression based security system is getting more popular and demandable because of their secured services for clients. Facial expression-based pattern unlocking devices are now in the marketplace and doing business billions of dollars.

To overcome the facial expression recognition challenge, a model of CNN will be designed so that outcome gives minimal error with comparatively high accuracy for seven categories of expressions. As a consideration, some factors like kernel size, hidden layers, input weight, dropout, etc. will be used to increase the accuracy rate higher than previously implemented methods.

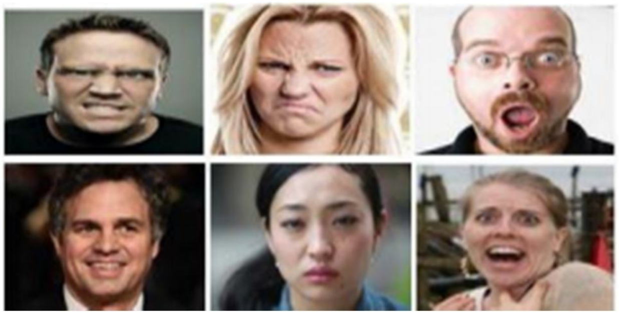

Fig 1: Different type facial expression [1]

II. LITERATURE REVIEW

In their research [1], Rajesh Kumar G A, Ravi Kant Kumar, Goutam Sanyal has mentioned two algorithms to recognize facial expression which are image pre-processing and then applying convolutional neural network. The steps for image preprocessing include- Haar-like feature extraction, remove irrelevant features using AdaBoost algorithm, and then cascade classifier to detect the face in the input image. The convolutional neural network with nine layers has been applied afterward which are – the input layer, Convolutional layer, ReLu Layer, Max-pooling, Convolution, Max-pooling, Convolution, Softmax, Output. The FER-2013 dataset was used to train and test the model. The Limitation of this work was the low accuracy rate which is only 48%.

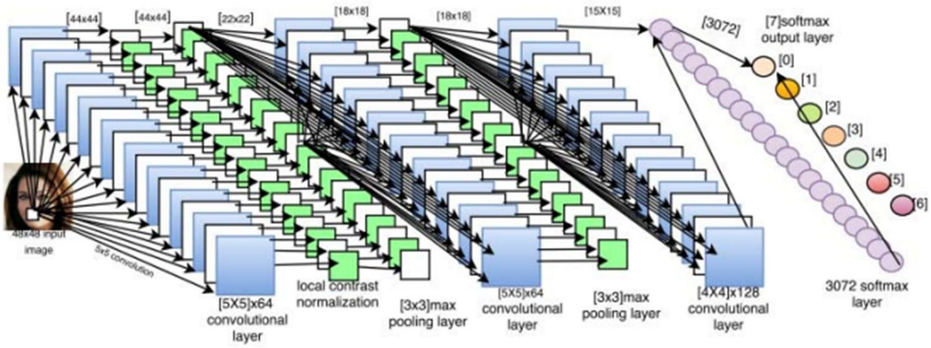

Fig 2: Complete architecture of Convolutional Neural Network [1]

In another research [2] Shima Alizadeh, Azar Fazel compared two facial expression recognition models namely shallow model and deep model which also used the FER-2013 dataset for training and testing of the models. In the shallow model, two convolutional layers along with a fully connected layer have been used. The convolutional layer-1 consists of thirty-two 3*3 filters whereas the convolutional layer-2 uses sixty-four 3*3 filters and both of them use a one-sized stride to format the layers. In the deep model, four convolutional layers have been used along with two fully connected layers. Each convolutional layer has one sized stride that formats the layer. The convolutional layers have thirty-two, twenty-eight, five hundred and twelve, and five hundred and twelve 3*3 filters respectively. The fully connected layers consist of 512 neurons and a hidden layer. The overall performance of the system was 55% and 64% respectively for the shallow model and deep model.

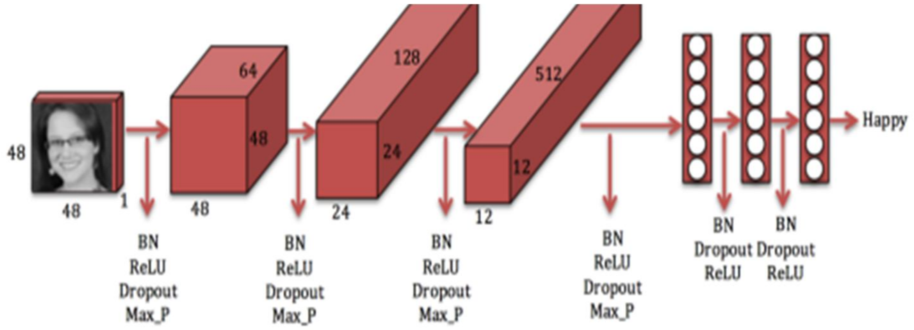

Fig 3: Complete architecture of Deep model [2]

Aneta Kartali, Miloš Roglić, Marko Barjaktarović, Milica Đurić-Jovičić, and Milica M. Janković used five discrete algorithms to recognize facial expression in their research[4]. The AlexNet CNN model first used image pre-processing which converts the images in cropped RGB format for desired size input. Each image is duplicated ten times so that the dataset can be enriched to use 70% as training data and the rest for testing. The convolutional neural network used here has five convolutional layers and three fully connected layers. Transfer learning has been used for removing overfitting and zero-mean Gaussian distribution with a deviation of 0.01 is used.

Guan Wang, Jun Gong proposed in their research[5] an improved Convolutional Neural Network based on the LeNet5 model. They extracted low-level features from the network and merged them with high-level features to build the classifier to avoid the constraints of low image information and noise interference under occlusion conditions.

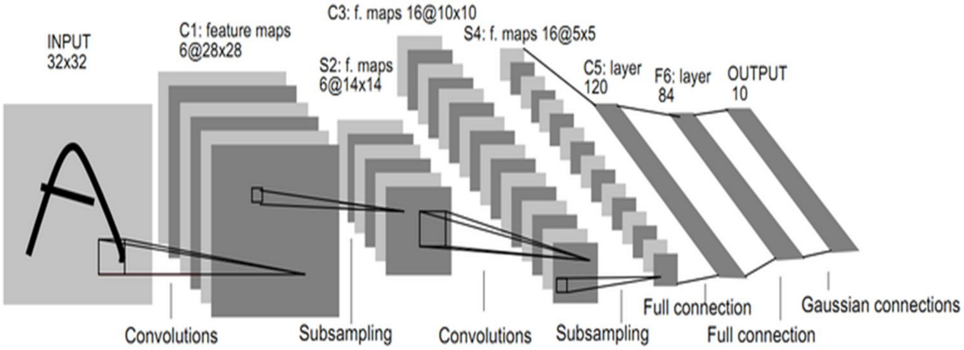

Fig 4: Complete architecture of LeNet-5 CNN Model [5]

III.PROPOSED MODEL

In this research, a new model has been proposed and compared with the other existing models to measure the performances in different metrics. The proposed facial expression recognition framework consists of data pre-processing and CNN Model.

A. Data pre-processing

In this research, the FER-2013 dataset has been used for training and testing purposes. This dataset consists of 48*48-pixel grayscale face images that contain 28,709 training samples and 3,589 testing samples. These samples are further subdivided into seven categories (Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral). The distribution of the dataset is shown in Table I.

TABLE I
Distribution of dataset

Categories	No. of Training Example	No. of Testing Example
Angry	3995	467
Disgust	436	56
Fear	4097	496
Happy	7215	895
Sad	4965	653
Surprise	4830	415
Neutral	3171	607

As the class samples in the dataset are imbalanced, the class data samples were balanced using data balancing techniques such as replication, reduction, and sample transformation, etc. Later, the data samples were preprocessed to extract facial landmark features [13] that will be fed into the CNN. Finally, a 10 fold cross-validation approach is used during model training and performance testing. In the next section, a comprehensive discussion on the details of the proposed model has been provided.

B. Convolutional Neural Network Model

In this study, careful consideration has been taken during the development of the CNN model framework as the CNN model needs to accurately recognize seven separate categories of expression which makes it a multi-class classification problem. Furthermore, several regularization techniques such as weight decay and dropout have been employed to prevent model overfitting toward the training samples.

The overall proposed model consists of three convolutional layers. In each of the layers, batch normalization has been incorporated to stabilize the model as well as to make the convergence faster. The rectified linear activation function (ReLU) has been used as an activation function in each of the layers. The model parameters are regularized with a decaying weight and a dropout rate of 50% is also to be used in the hidden layers to further prevent model overfitting. Later, the output from the three consecutive CNN layers is subjected to a flattening process to extract the features vector. Finally, this feature vector is feed into the fully connected layer with softmax activation that maps the features to the seven facial expression categories. The overall proposed facial expression recognition model can be represented as followings:

1) Convolutional Layer-1: 64, 3*3 Convolution layer with a pooling layer of 2*2. Batch normalization and ReLU activation function were also used.
2) Convolutional Layer-2: 128, 3*3 Convolution layer with a pooling layer of 2*2. Batch normalization and ReLU activation function were also used. Dropout was also used.
3) Convolutional Layer-3,4: 256, 3*3 Convolution layer with a pooling layer of 2*2. Batch normalization and ReLU activation function were also used. It was used twice as the third and fourth convolutional layers.
4) Flattening: The data from the convolutional layers were flattened into a vector to be fed to the fully connected layer.
5) Fully Connected Layer: Three fully connected layers with 128, 64, and 7 neurons were used to process the flattened data received from the convolutional layers.
6) Activation function: The softmax activation function was used to classify the result from fully connected layers to generate output.

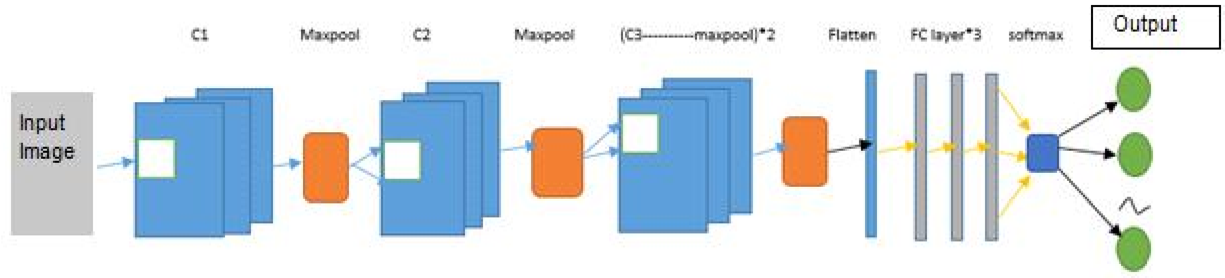

Fig. 5: Proposed model structure

The proposed CNN model architecture is given below:

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 46, 46, 64)        640
_________________________________________________________________
activation (Activation)      (None, 46, 46, 64)        0
_________________________________________________________________
batch_normalization (BatchNo (None, 46, 46, 64)        256
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 23, 23, 64)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 21, 21, 128)       73856
_________________________________________________________________
dropout (Dropout)            (None, 21, 21, 128)       0
_________________________________________________________________
activation_1 (Activation)    (None, 21, 21, 128)       0
_________________________________________________________________
batch_normalization_1 (Batch (None, 21, 21, 128)       512
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 10, 10, 128)       0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 8, 8, 256)         295168
_________________________________________________________________
activation_2 (Activation)    (None, 8, 8, 256)         0
_________________________________________________________________
```

```
batch_normalization_2 (Batch (None, 8, 8, 256)        1024
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 4, 4, 256)        0
_________________________________________________________________
conv2d_3 (Conv2D)           (None, 2, 2, 256)        590080
_________________________________________________________________
activation_3 (Activation)    (None, 2, 2, 256)        0
_______________________________________________________________
batch_normalization_3 (Batch (None, 2, 2, 256)        1024
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 1, 1, 256)        0
_________________________________________________________________
flatten (Flatten)           (None, 256)              0
_________________________________________________________________
dense (Dense)               (None, 128)              32896
_________________________________________________________________
activation_4 (Activation)    (None, 128)              0
_________________________________________________________________
batch_normalization_4 (Batch (None, 128)              512
_________________________________________________________________
dropout_1 (Dropout)         (None, 128)              0
_________________________________________________________________
dense_1 (Dense)             (None, 64)               8256
_________________________________________________________________
activation_5 (Activation)    (None, 64)               0
_________________________________________________________________
batch_normalization_5 (Batch (None, 64)               256
_________________________________________________________________
dense_2 (Dense)             (None, 7)                455
=================================================================
Total params: 1,004,935
Trainable params: 1,003,143
Non-trainable params: 1,792
```

IV.RESULT ANALYSIS

The dataset used in this experiment is the FER-2013 Dataset which consists of 48*48-pixel grayscale face images. The training set contains 28,709 examples and the testing set contains 3,589 examples which are divided into seven categories (Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral). The distribution of the dataset is shown in Table I.

TABLE III
Distribution of Balanced dataset after preprocessing

Categories	No. of Training Example	No. of Testing Example
Angry	3000	467
Disgust	3000	56
Fear	3000	496
Happy	3000	895
Sad	3000	653
Surprise	3000	415
Neutral	3000	607

The performance of the models (Shallow model, Deep Model, Deep CNN Model, LeNet5, AlexNet, and proposed model) are shown in the following confusion matrices.

TABLE IIIII
Confusion Matrix for shallow model

	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	216	2	68	82	59	6	34
Disgust	12	20	10	8	3	0	3
Fear	89	0	214	57	79	13	44
Happy	56	0	70	699	36	8	26
Sad	140	0	123	98	211	6	75
Surprise	36	0	98	39	18	207	17
Neutral	88	0	91	129	86	4	209

TABLE IVV
Confusion Matrix for Deep model

	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	254	2	46	30	95	6	30
Disgust	12	30	2	1	9	0	1
Fear	46	0	216	15	146	13	31
Happy	40	0	29	705	53	8	48
Sad	98	2	70	33	381	6	57
Surprise	20	0	25	25	18	320	7
Neutral	76	1	48	46	86	12	263

TABLE V
Confusion Matrix for Deep CNN model

	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	166	2	52	88	116	5	38
Disgust	9	24	6	5	6	0	6
Fear	46	0	153	82	139	14	62
Happy	19	0	12	739	80	4	41
Sad	60	0	44	103	345	2	99
Surprise	21	0	35	45	51	236	27
Neutral	49	0	33	100	154	1	270

TABLE VI
Confusion Matrix for LeNet5

	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	173	3	54	69	82	18	68
Disgust	14	21	7	6	4	2	2
Fear	70	1	167	52	77	59	70
Happy	52	0	26	623	78	22	94
Sad	98	3	74	69	271	14	124
Surprise	22	3	37	26	20	290	17
Neutral	73	5	49	90	84	25	281

TABLE VII
Confusion Matrix for AlexNet

	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	152	10	109	35	94	22	45
Disgust	12	21	7	6	6	2	4
Fear	41	3	205	24	110	55	58
Happy	44	6	132	483	95	39	96
Sad	59	3	153	52	248	27	111
Surprise	17	1	73	14	20	257	33
Neutral	48	4	88	62	157	26	222

TABLE VIII
Confusion Matrix for Proposed Model

	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	271	0	34	14	95	14	39
Disgust	11	28	3	2	8	0	4
Fear	44	0	217	12	121	32	70
Happy	33	0	16	694	54	16	82
Sad	62	1	51	19	422	10	88
Surprise	12	0	17	20	18	327	21
Neutral	54	1	35	34	139	4	340

The model performance based on the F1 measure is given in Table IX.

TABLE IX
Model Performance based on F1 measure

Model Name	Precision	Recall	F1 Score
Shallow Model[2]	0.615	0.457	0.524
Deep Model [2]	0.636	0.584	0.609
Deep CNN Model [2]	0.608	0.495	0.495
LeNet5 [5]	0.505	0.479	0.492
AlexNet [4]	0.447	0.425	0.436
Proposed Model	0.610	0.682	0.644

From table IX, it can be seen that the proposed CNN model greatly outperforms the previous state-of-the-arts facial expression classification methods such as shallow model[2], deep model[2], LeNet5[5], AlexNet[4]. All the result presented in Table IX is taken by averaging the outcomes from the 10 fold cross-validation. The proposed method consistently provided better precision, recall, and F1 score than the other methods. The reason for the superior performance of the proposed model can be attributed to several factors such as dataset balancing, use of 10 fold cross-validation, and enhanced regularization to reduce overfitting.

V. CONCLUSIONS

Facial expression recognition is a very popular yet challenging task. Numerous researches over the years have developed various facial recognition model majority of which lacks desired accuracy and suffers from overfitting. With an aim to address the issues of the prior studies, this research has proposed a convolutional neural network model architecture and later validates the performance of the proposed model by testing it against 5 state-of-the-arts facial expression recognition techniques. During the experimental analysis, the F1 Score is used as the performance metric to represent the comparative False positives and False negatives rates of the models. It is evident from the experimental result that the proposed model is able to provide better accuracy in facial expression recognition as it greatly outperforms all the other techniques in terms of the F1 Score, attaining an overall accuracy improvement of 4%.

REFERENCES

[1] Rajesh Kumar G A, Ravi Kant Kumar, Goutam Sanyal, “Facial Emotion Analysis Using Deep Convolutional Neural Network”, International Conference on SignalProcessing and Communication(ICSPC17’)-July 2017.

[2] Shima Alizadeh, Azar Fazel, “Convolutional Neural Networks for Facial Expression Recognition”, International Conference on Signal Processing and Communication(ICSPC17’)-22 Apr 2017.

[3] Chunling Cheng, Xianwei Wei, “Emotion Recognition Algorithm Based on Convolution Neural Network ”, International Conference on Intelligent Systems and Knowledge Engineering (ISKE’)-2017

[4] Aneta Kartali, Miloš Roglić, Marko Barjaktarović, Milica Đurić-Jovičić, and Milica M. Janković, “ Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches”,2018 14th Symposium On Neural Networks and Applications(Neural), Belgrade, Serbia, November 20-21 2018

[5] Guan Wang, Jun Gong, “ Facial Expression Recognition Based on Improved LeNet-5 CNN”, The 31st Chinese Control and Decision Conference (2019 CCDC)

[6] Medium. 2021. Deep Learning: Feedforward Neural Network. [online] Available at: <https://towardsdatascience.com/deep-learning-feedforward-neural-network-26a6705dbdc7> [Accessed 12 September 2021].

[7] X. Zhang, W. Pan, and P. Xiao, "In-Vivo Skin Capacitive Image Classification Using AlexNet Convolution Neural Network," 2018 IEEE 3rd International Conference on Image, Vision, and Computing (ICIVC), Chongqing, 2018, pp. 439-443, DOI: 10.1109/ICIVC.2018.8492860.

[8] Docs.gimp.org. 2020. 8.2. Convolution Matrix. [online] Available at: <https://docs.gimp.org/2.8/en/plug-in-convmatrix.html> [Accessed 16 December 2020].

[9] Medium. 2020. Basic Overview Of Convolutional Neural Network (CNN). [online] Available at: <https://medium.com/dataseries/basic-overview-of-convolutional-neural-network-cnn-4fcc7dbb4f17> [Accessed 16 December 2020].

[10] Medium. 2020. The Most Intuitive And Easiest Guide For CNN. [online] Available at: <https://towardsdatascience.com/the-most-intuitive-and-easiest-guide-for-convolutional-neural-network-3607be47480> [Accessed 16 December 2020].

[11] "Challenges in Representation Learning: Facial Expression Recognition Challenge | Kaggle", Kaggle.com, 2020. [Online]. Available: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data. [Accessed: 13- Dec- 2020].

[12] N. Yu, P. Jiao and Y. Zheng, "Handwritten digits recognition base on improved LeNet5," The 27th Chinese Control and Decision Conference (2015 CCDC), Qingdao, 2015, pp. 4871-4875, DOI: 10.1109/CCDC.2015.7162796.

[13] Q. T. Ngoc, S. Lee, and B. C. Song, “Facial Landmark-Based Emotion Recognition via Directed Graph Neural Network,” *Electronics*, vol. 9, no. 5, p. 764, May 2020. https://doi.org/10.3390/electronics9050764