ISSN (Online): 2583 648X

I J I R I D

International Journal of Ingenious Research, Invention and Development

Volume 1 | Issue 3 | April 2023

Scientific Journal Impact Factor (SJIF 2023): 3.647

DOI: 10.5281/zenodo.8047671

www.ijirid.in

A Novel Approach for Facial Expression Recognition Using Deep Learning Techniques

Suraj V. Dhole[1], Bhavana Barbudhe[2]

[1]Assistant Professor, GH Raisoni University, Amravati (MH), India

Abstract: *Facial Expressions are an imperative characteristic of non-verbal communication, as we move towards digitization natural computer relations play a vital part. The emotional changes results in the difference in the expressions. This paper elaborates evolution of the Deep Convolutional Neural Network Model and Keras for building and training the Deep Learning Model. This paper aims to classify facial images into one of the seven face discovery classifiers using open CV and one of its classifiers for drawing the boundary box around the face to detect the correct expression.*

Keywords: Facial Expression Recognition, Deep Learning, Convolutional Neural Network, Regions of Interest.

I. INTRODUCTION

Facial Expression Recognition (FER) can be seen as a second step to face detection mechanism. Humans can express emotions through facial expressions, a part of nonverbal communication. When a machine communicates with people, FER can give more affinities and personalized service to people depending on their emotions, which eventually increases confidence and trust in people. We can express emotions in various ways, such as facial expressions, voices, physiological signals, and text. Machines capture the expressions through cameras and videos. Facial Expressions can be classified as surprised, happy, neutral, angry, sad, disgusted, and fearful. This paper elaborates on the development of a deep convolutional neural network model to classify facial expressions.

II. RELATED WORK

In [1], a novel Geometric features extraction method for facial expression recognition is proposed. ASM automatic fiducial point location algorithm is first applied to a facial expression image, and then calculating the Euclidean distances between the centre of gravity coordinate and the annotated fiducial points' coordinates of the face image. In order to extract the discriminate deformable geometric information, the system extracts the geometric deformation difference features between a person's neural expression and the other seven basic expressions. A multiclass Support Vector Machine (SVM) classifier is used to recognize facial expressions. Experiments indicate that the proposed method can obtain good classification accuracy, Linear Discriminant Analysis (LDA) is one of the principal techniques used in face recognition systems.

Linear Discriminant Analysis (LDA) is a well-known scheme for feature extraction and dimension reduction. It provides improved performance over the standard Principal Component Analysis (PCA) method of face recognition by introducing the concept of classes and the distance between classes. [3] provides an overview of PCA, the various variants of LDA and their basic drawbacks. The proposed method includes a development over classical LDA (i.e., LDA using wavelets transform approach) that enhances performance such as accuracy and time complexity. Experiments on the ORL face database clearly demonstrate this and the graphical comparison

ISSN (Online): 2583 648X

I J I R I D

International Journal of Ingenious Research, Invention and Development

Volume 1 | Issue 3 | April 2023

Scientific Journal Impact Factor (SJIF 2023): 3.647

DOI: 10.5281/zenodo.8047671

www.ijirid.in

of the algorithms clearly showcases the improved recognition rate in the case of the proposed algorithm. Many face recognition techniques have been developed during the past decades but the problem remains challenging, especially recognizing non-biological entities or avatars.

The local Binary Pattern (LBP) method is one of these techniques which has shown its superiority in recognizing faces. The original LBP operator mainly thresholds pixels in a specific predetermined window based on the grey value of the central pixel of that window. As a result, the LBP operator becomes more sensitive to noise, especially in near-uniform or flat area regions of an image. To deal with this problem a generalization of the LBP descriptor, Local Ternary Patterns (LTP), came to the present.

In [6] A. A. Mohamed and R. V. Yampolskiy introduce new locally adapted texture features for efficient avatar face recognition based on the original LTP operator. The proposed technique, Adaptive Extended Local Ternary Pattern (AELTP), shares with the original LTP descriptor being less sensitive to noise. However, AELTP is better as it determines the local pattern threshold automatically based on local statistics. Experiments conducted on two virtual world avatar face image datasets show that their technique performs better than original LBP, original LTP and Extended LTP (ELTP) in terms of accuracy.

[9] evaluates the performance both of some texture measures which have been successfully used in various applications and of some new promising approaches proposed recently. For classification, a method based on Kullback discrimination of sample and prototype distributions is used. The classification results for single features with one-dimensional feature value distributions and for pairs of complementary features with two-dimensional distributions are presented.

In [10], a new kernel-based manifold learning algorithm, called KDIsomap, is proposed for facial expression recognition. KDIsomap has two prominent characteristics. For one thing, as a kernel-based feature extraction method, KDIsomap can extract the nonlinear feature information embedded in a data set, as KPCA and KLDA do. For another, KDIsomap is designed to offer a high discriminating power for its low-dimensional embedded data representations in an effort to improve the performance of facial expression recognition. It's worth pointing out that in their work they focus on facial expression recognition by using static images from two well-known facial expression databases, but they do not consider the temporal behaviours of facial expressions, which can potentially lead to more robust and accurate classification results. Therefore, it is also an interesting task to explore the performance of temporal information on facial expression recognition in our future work.

Currently, facial expression recognition has become a dynamic research area. Several methods have been developed towards strong facial expression analysis, using several image acquisition, recognition and classification techniques. Facial expression analysis is an inherently multi-disciplinary domain and it is vital to look at it from all fields in order to have an insight on how to develop an efficient automatic facial expression recognition system. [11] has employed several advanced methods to enhance the recognition rate and execution time of facial expression recognition systems. Face detection has been carried out using the application of Viola–Jones descriptor. Originating an effective face representation from the initial face images is an important part of an effective facial expression analysis.

R. Shbib and S. Zhou. have tried to evaluate ASM features in order to label the appearance variation of expression images. Extensive results have shown that ASM features are strong and reliable for facial expression recognition. They have adopted AdaBoost to get the most discriminative facial features from a large facial feature. The best recognition rate is achieved by applying SVM classifier. However, this technique involves some limitations when it is applied to other datasets. In addition, facial features have been extracted by applying a pyramid ASM fitting technique in order to get the most discriminative facial features from large facial points. The geometrical shift among the estimated ASM feature points coordinates and mean shape of ASM is projected to the SVM classifier. Results have shown a satisfactory real-time and strong performance of the proposed approach.

ISSN (Online): 2583 648X

I J I R I D

International Journal of Ingenious Research, Invention and Development

Volume 1 | Issue 3 | April 2023

Scientific Journal Impact Factor (SJIF 2023): 3.647

DOI: 10.5281/zenodo.8047671

www.ijirid.in

III. EXPERIMENTAL SETUP

This research talks about seven face detection classifiers using open CV and one of its classifiers for drawing the boundary box around the face to detect the correct expression. For training the CNN models we have used 48x48 greyscale images from Kaggle's ICMP 2013-Fecial Expression Recognition (FER) dataset (https://www.kaggle.com/c/challenges-inrepresentation-learning-facial-expression-recognitionchallenge/data'). The FER dataset is divided into two folders called test and train, further divided into separate folders each containing one of the seven types of FER datasets. Each image has to be categorized into one of the seven classes that express different facial emotions. These facial emotions have been categorized as 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, and 6=Neutral.

Figure 1: Sample Images from the Dataset

The training dataset consists of the following images categorized into 7 folders

```
3995 angry images.
436 disgust images.
4097 fear images.
7215 happy images.
4965 neutral images.
4830 sad images.
3171 surprise images.
```

Figure 2: Training Dataset

ISSN (Online): 2583 648X

I J I R I D

International Journal of Ingenious Research, Invention and Development

Volume 1 | Issue 3 | April 2023

Scientific Journal Impact Factor (SJIF 2023): 3.647

DOI: 10.5281/zenodo.8047671

www.ijirid.in

The testing dataset consists of the following images categorized into 7 folders.

```
958 angry images.
111 disgust images.
1024 fear images.
1774 happy images.
1233 neutral images.
1247 sad images.
831 surprise images.
```

Figure 3: Training Dataset

Data Augmentation

To build the training model, training and validation batches are generated with the FER dataset image size 48x48 and batch size of 64 as per the memory size of CPU/GPU to speed up the training process. Image data augmentation is used to improve the performance and ability of the model to generalize. It's always a good practice to apply some data augmentation before passing it to the model, which can be done using Image Data Generator provided by Keras. ImageDataGenerator () class is used to accept values or images to treat the camera-captured image as a horizontal mirror image. These images are used to generate the training set. Test and training sets are generated by keeping the various parameters the same.

Training the CNN model

Here we designed a facial expression recognition system that used convolutional neural network architecture with fully connected layers in order to extract the facial features and then classify the extracted features based on the support vector machine classifier. We are creating blocks using Conv2D layer, Batch-Normalization, Max-Pooling2D, Dropout, and Flatten, and then stacking them together and at the end-use Dense Layer for output. CNN-SVM model architecture is as below:

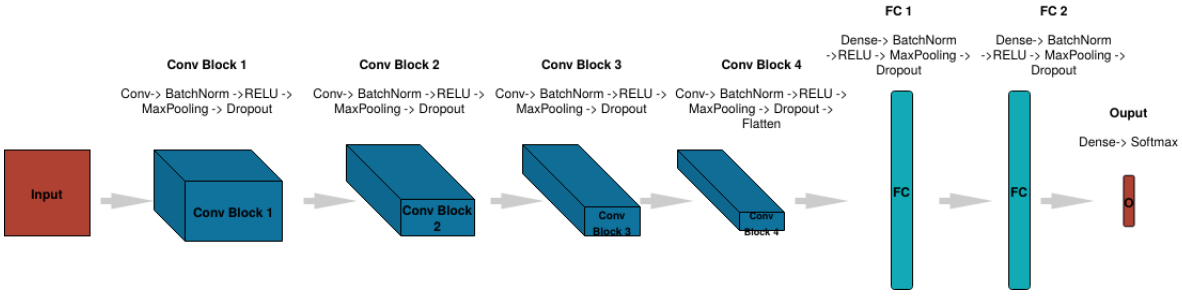


Figure 4: CNN Model

By following the above CNN architecture six activation layers are designed. Four convolution layer and 2 fully controlled layers. The ReLu function is used to increase the non-linearity in the images, maxpooling is used for dimensions' reduction of the image, dropout function to avoid over fitting of the training data. Flatten to convert image to 1- dimensional array. 1- dimensional array becomes the input to the fully controlled layers. Output layer has two techniques dense and softmax. In a normal CNN, at output layer we use sigmoid activation Also, for Multiclass classification problem we've to use squared hinge as loss function. Model summary as below:

ISSN (Online): 2583 648X

I J I R I D

International Journal of Ingenious Research, Invention and Development

Volume 1 | Issue 3 | April 2023

Scientific Journal Impact Factor (SJIF 2023): 3.647

DOI: 10.5281/zenodo.8047671

www.ijirid.in

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 48, 48, 32)        320

 batch_normalization (BatchN  (None, 48, 48, 32)       128
 ormalization)

 activation (Activation)     (None, 48, 48, 32)        0

 max_pooling2d (MaxPooling2D  (None, 24, 24, 32)       0
 )

 dropout (Dropout)           (None, 24, 24, 32)        0

 conv2d_1 (Conv2D)           (None, 24, 24, 64)        18496

 batch_normalization_1 (Batc  (None, 24, 24, 64)       256
 hNormalization)

 activation_1 (Activation)   (None, 24, 24, 64)        0

 max_pooling2d_1 (MaxPooling  (None, 12, 12, 64)       0
 2D)

 dropout_1 (Dropout)         (None, 12, 12, 64)        0

 conv2d_2 (Conv2D)           (None, 12, 12, 128)       73856

 batch_normalization_2 (Batc  (None, 12, 12, 128)      512
 hNormalization)

 activation_2 (Activation)   (None, 12, 12, 128)       0

 max_pooling2d_2 (MaxPooling  (None, 6, 6, 128)        0
 2D)

 dropout_2 (Dropout)         (None, 6, 6, 128)         0

 flatten (Flatten)           (None, 4608)              0

 dense (Dense)               (None, 128)               589952

 batch_normalization_3 (Batc  (None, 128)              512
 hNormalization)

 activation_3 (Activation)   (None, 128)               0

 dropout_3 (Dropout)         (None, 128)               0

 dense_1 (Dense)             (None, 256)               33024

 batch_normalization_4 (Batc  (None, 256)              1024
 hNormalization)

 activation_4 (Activation)   (None, 256)               0

 dropout_4 (Dropout)         (None, 256)               0

 dense_2 (Dense)             (None, 7)                 1799

=================================================================
Total params: 719,879
Trainable params: 718,663
Non-trainable params: 1,216
```

Figure 5: Model Summary

ISSN (Online): 2583 648X

I J I R I D

International Journal of Ingenious Research, Invention and Development

Volume 1 | Issue 3 | April 2023

Scientific Journal Impact Factor (SJIF 2023): 3.647

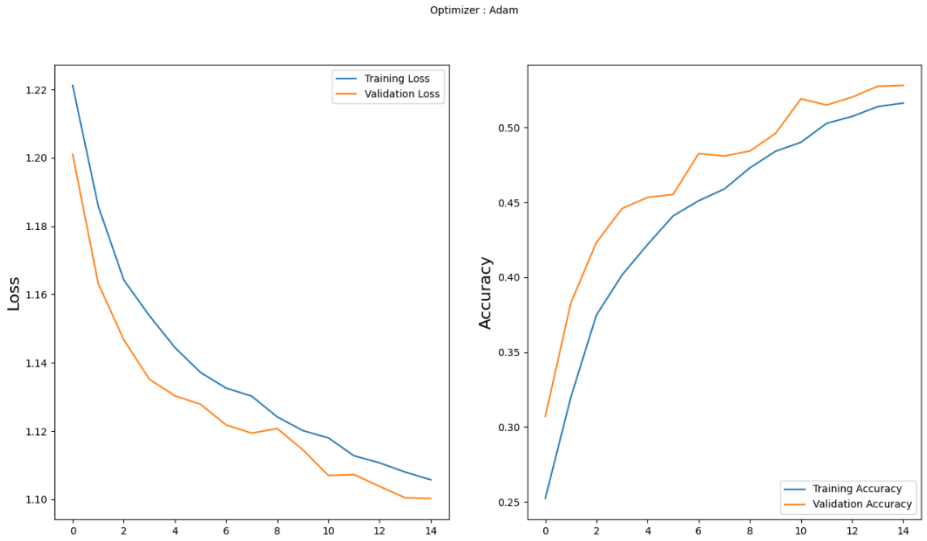
DOI: 10.5281/zenodo.8047671

www.ijirid.in

IV. RESULT AND DISCUSSION

The training loss is a metric used to assess how a deep learning model fits the training data. That is to say, it assesses the error of the model on the training set. Note that, the training set is a portion of a dataset used to initially train the model. Computationally, the training loss is calculated by taking the sum of errors for each example in the training set. On the contrary, validation loss is a metric used to assess the performance of a deep learning model on the validation set. The validation set is a portion of the dataset set aside to validate the performance of the model. The validation loss is similar to the training loss and is calculated from a sum of the errors for each example in the validation set. The graphs for training and validation loss and accuracy for the CNN model are shown below.

It has been observed that the designed CNN model can flawlessly detect facial expressions such as Happy, Sad, Neutral and Surprised.



Graph 1: CNN Training/ Validation Accuracy and Loss

Accuracy Score: Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observations to the total observations. Our model is a binary classifier which produces output with two classes, such as Positive or Negative, for given input data. A dataset used for performance evaluation is called a test dataset. It contains the actual labels for all images. These actual labels are used to compare with the predicted labels for performance evaluation after classification.

A binary classifier predicts all data instances of a test dataset as either positive or negative. This classification (or prediction) produces four outcomes – true positive, true negative, false positive and false negative.

- True positive (TP): correct positive prediction
- False positive (FP): incorrect positive prediction
- True negative (TN): correct negative prediction
- False negative (FN): incorrect negative prediction

Accuracy (ACC) is calculated as the number of all correct predictions divided by the total number of the dataset. The formula to calculate accuracy is as follows:

ISSN (Online): 2583 648X

I J I R I D

International Journal of Ingenious Research, Invention and Development

Volume 1 | Issue 3 | April 2023

Scientific Journal Impact Factor (SJIF 2023): 3.647

DOI: 10.5281/zenodo.8047671

www.ijirid.in

Accuracy = (TP + TN) / (TP + TN + FP + FN)

We perform two accuracy tests one on a large validation image set and another on a small validation image set. The accuracy is calculated for CNN based model. The results are shown using the confusion matrix as shown below.

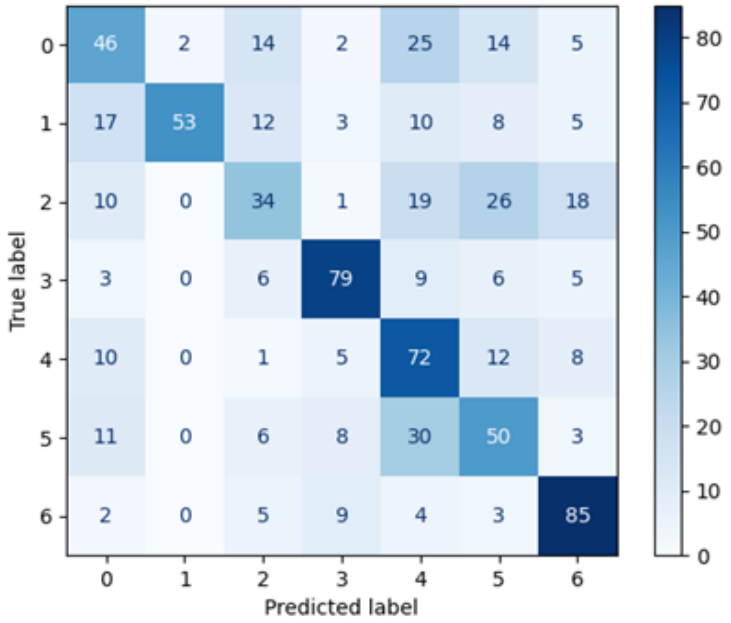


Table 1: Confusion Matrix for CNN-based Model for Test 1

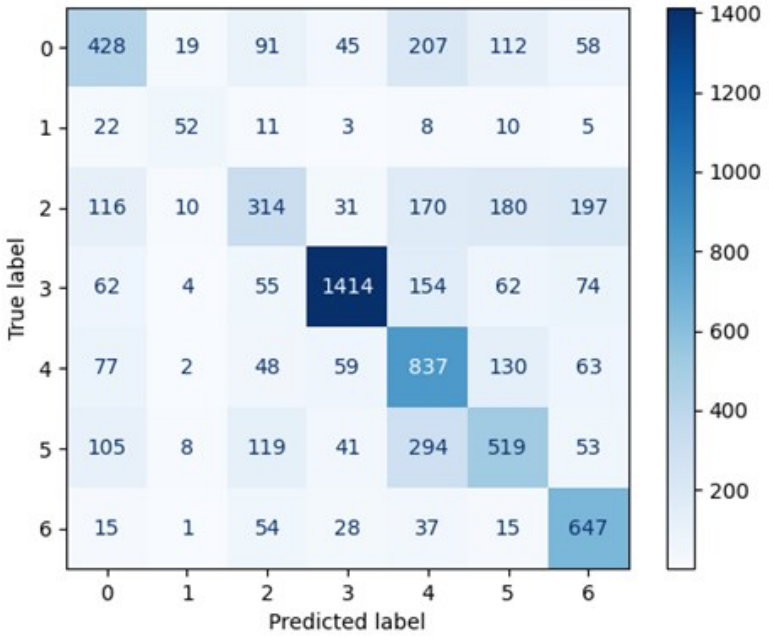


Table 2: Confusion Matrix for CNN-based Model for Test 2

ISSN (Online): 2583 648X

I J I R I D

International Journal of Ingenious Research, Invention and Development

Volume 1 | Issue 3 | April 2023

Scientific Journal Impact Factor (SJIF 2023): 3.647

DOI: 10.5281/zenodo.8047671

www.ijirid.in

IV. CONCLUSION

This paper presents a new facial decomposition for basic emotion state recognition. Based on facial landmarks regions of interest (ROI), corresponding to the main components of the face, are first extracted to represent the face image. proposed Convolutional Neural Network based architecture for facial expression recognition. There are 7 classes of facial expressions we tried to recognize. Using the FER database, we trained our model and calculated the accuracy of CNN based model. In future, some fine-tuning can be done along with using some other pre-trained model to improve accuracy.

REFERENCES

[1] G. Lei, X.-h. Li, J.-l. Zhou, and X.-g. Gong. Geometric feature-based facial expression recognition using multiclass support vector machines. In Granular Computing, 2009, GRC'09. IEEE International Conference on, pages 318–321. IEEE, 2009.

[2] K. Lekdioui, R. Messoussi, and Y. Chaabi. Etude et modelisation ´ des comportements sociaux d'apprenants a distance, ` a travers l'analyse ` des traits du visage. In 7eme Conf ` erence sur les Environnements ´ Informatiques pour l'Apprentissage Humain (EIAH 2015), pages 411– 413, 2015.

[3] P. Marasamy and S. Sumathi. Automatic recognition and analysis of human faces and facial expressions by LDA using wavelet transform. In Computer Communication and Informatics (ICCCI), 2012 International Conference on, pages 1–4. IEEE, 2012.

[4] S. S. Meher and P. Maben. Face recognition and facial expression identification using PCA. In Advance Computing Conference (IACC), 2014 IEEE International, pages 1093–1098. IEEE, 2014.

[5] A. Mehrabian et al. Silent Messages, volume 8. Wadsworth Belmont, CA, 1971.

[6] A. A. Mohamed and R. V. Yampolskiy. Adaptive extended local ternary pattern (aeltp) for recognizing avatar faces. In Machine Learning and Applications (ICMLA), 2012 11th International Conference on, volume 1, pages 57–62. IEEE, 2012.

[7] G. Molinari, C. Bozelle, D. Cereghetti, G. Chanel, M. Betrancourt, and ´ T. Pun. Feedback emotionnel et collaboration m ´ ediatis ´ ee par ordinateur: ´ Quand la perception des interactions est liee aux traits ´ emotionnels. In ´ Environnements Informatiques pour l'apprentissage humain, Actes de la conference EIAH´, pages 305–326, 2013.

[8] R. Nkambou and V. Heritier. Reconnaissance emotionnelle par l'analyse ´ des expressions faciales dans un tuteur intelligent affectif. In Technologies de l'Information et de la Connaissance dans l'Enseignement Superieur et l'Industrie´, pages 149–155. Universite de Technologie de ´ Compiegne, 2004. `

[9] T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of ¨ texture measures with classification based on featured distributions. Pattern recognition, 29(1):51–59, 1996.

[10] C. Shan, S. Gong, and P. W. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing, 27(6):803–816, 2009.

[11] R. Shbib and S. Zhou. Facial expression analysis using active shape model. Int. J. Signal Process. Image Process. Pattern Recognit, 8(1):9– 22, 2015.