Amrut Ranjan jena, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 12, Issue 4, (Series-I) April 2022, pp. 27-33

RESEARCH ARTICLE OPEN ACCESS

A Framework for Emotion Detection using Open Source Computer Vision and Convolutional Neural Network

Amrut Ranjan jena[1], Madhusmita Mishra[2], Ratnadeep Mazumder[3]
[1](Department of Computer Science, GNIT, Kolkata-700114, India)
[2] (Department of Computer Science, DSCSIT, Kolkata-700074, India)
[3](Department of Computer Science, GNIT, Kolkata-700114, India)

ABSTRACT
The Human emotions are the mental states of sentiments happen without conscious effort, and accompanied by physiological changes in facial muscles that result in facial expressions. The human-computer interaction uses nonverbal communication methods to know the emotion of a person through facial expressions, eye movement, and gestures. Besides, facial expression is a common procedure to find the mood because it transmits people's emotional states and feelings. Emotion recognition is a difficult task due to the facial patterns variety and complexity appears over face. The traditional computational methods fail to predict the emotions due to the variety and complexity of facial patterns. Therefore, machine learning methods are deployed to find better result for emotions detection. In this work, we have developed a model by using convolution neural networks, Open source computer vision, and tensorflow to detect the emotions of a person. Seven different emotions like anger, neutral, disgust, fear, happiness, sadness, and surprise are proposed by the model through the posture of mouth and eyes of a person.

Keywords - Emotion detection, Convolutional neural networks, Face detection, Facial emotion recognition, Open source computer vision

---

Date of Submission: 04-04-2022 Date of Acceptance: 19-04-2022

I. INTRODUCTION

Human Computer Interaction has become increasingly crucial as the usage of technology has increased. As a result, throughout the last decade, experts have focused their attention on facial emotion recognition system (FERS) by machines [1]. There is a demand for real-time applications that can identify and classify human expressions. This classification of emotions can be used to help the computers to understand user requirements [2]. So understanding and recognizing emotions is aided by facial expressions. Even for the term "interface" implies the centrality of the face in two-way communication. Reading facial expressions has been demonstrated in studies to significantly influence the interpretation of what has been said and influence the flow of communication [3]. Human emotion study stretches back to Darwin's pioneering work, and it has since drawn a slew of newcomers to the area [4]. Humans have seven fundamental emotions. These basic emotions include neutral, angry, disgusted, fearful, happy, sad, and surprised, and it is identified by the facial expression of a person. The book "The Expression of the Emotions in Man and Animals," by Charles Darwin (1872-1965), was a seminal work in the field of emotion studies [5]. This book was written to contradict Sir Charles Bell's (1844) assertion that particular muscles were developed to allow humans to express their emotions. The basic point of Darwin was that emotional manifestations are adaptable and developed. Emotional expressions, according to Darwin, not only developed as a component of the emotion process, but also served a vital communicating purpose. Emotion recognition systems are designed to apply emotion-related knowledge in a way that human-computer communication is improved and the person's experience is improved. Specialized systems can be built and used for further serious situations, such as aggression detection, stress detection, and frustration detection in medical applications. Consider having a psychologist use this strategy to discover what their patient is going through in order to make a more accurate diagnosis, because nonverbal communication such as body language and facial expressions disclose a lot more than words in this case. One of the most exciting and difficult issues in the field of artificial intelligence is computer vision. Computer Vision connects software to the pictures we perceive all around us. It allows software to understand and learn about the sights in its

surroundings [6].The term "open source computer vision" is abbreviated as "OpenCV." The architecture is composed of pre-programmed software, databases, and extensions that support the integration of computer vision applications. With a big developer community, it is one of the most widely used toolkits. It is well-known for the scale at which it creates real-world industrial use cases. OpenCV is based on the C/C++, Python, and Java programming languages and used to create computer vision software for Windows, Linux, macOS, Android, and iOS. Artificial neural networks (ANNs) are human brain inspired algorithms used in deep learning. Convolutional Neural Networks (CNNs) are the type of deep neural network that uses convolution as a mathematical operation. The system utilizes a two dimensional CNN for the recognition problem, because the dataset is made up of images [7]. In this work the deep convolutional neural network is used to classify seven unique facial expressions. A convolutional neural network known as ConvNet is a deep learning method which can receive an image as input, assign value (learnable weights and biases) to diverse components in the image, and distinguish between them [8]. In comparison to other image classification techniques, it is a deep artificial neural network which can recognize visual features out of an input image with minimum pre-processing. Every neuron is a crucial component of a CNN layer. They're linked together so that the outcome of one layer's neurons will become the input of the following neuron layers. As compared to all other classification methods, the amount of pre-processing required by CNN is significantly less. While basic approaches require hand-engineering of filters, ConvNets can learn these filters/characteristics with enough training. The back-propagation algorithm is used to calculate the partial derivatives of cost function [9]. Convolution is the process of creating a feature map by applying filters or kernels to an input image. A CNN model consists of many layers as presented in Figure 1.

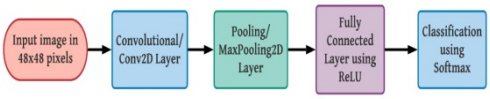

Fig. 1 CNN layers

Convolutional layer: In the context of a ConvNet, the main objective of convolution layer is to extract the features from the input image [10]. By learning image attributes with smaller sections of input data, convolution maintains the spatial connection between pixels. It computes the dot product of two matrices, one of which is the image and the other a kernel. A rectangle grid of neurons constitutes convolutional layers. Every neuron in the convolutional layer receives input from the rectangular section of the preceding layers; the weights for this rectangular area are the same as for each neuron. Figure 2 visualizes the convolutional layer.

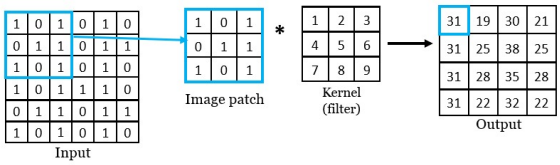

Fig. 2 CNN convolution layer

Pooling layer: Convolution spatial size feature is minimized by the pooling layer ([11]-[13]). Using dimensionality reduction, required data is reduced as per the computing capacity. It's useful for extracting rotational as well as positional stable dominant features that keeps the model's training process moving forward. There are 2 types of pooling: Maximum Pooling, and Average Pooling. The maximum value from the region of the image captured by the Kernel is obtained by max Pooling. Average Pooling, on either hand, produces the average of all values from the Kernel's section of the image. In this work, we consider the block's maximum as the final output to the pooling layer. As result, the convolution and pooling layers extract features from the input image, whereas the fully connected layer works as a classifier. In this work, we flatten the input image into a column vector after converting it to a suitable format. Every round of training uses back-propagation to send the flattened output to a feed-forward neural network. We pass our input image through the fully connected layer after flattening it. Figure 3 describes the max pooling layer.

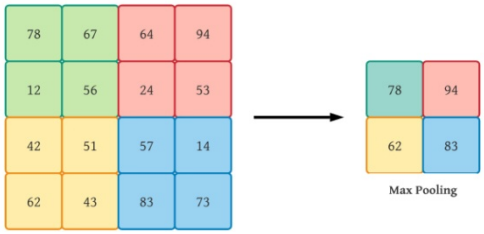

Fig. 3 Max pooling layer

The goal of this research work is to describe a way for creating the facial emotion recognition system (FERS) by applying deep CNN. Anger, happiness, sadness, surprise, fear, neutral, and disgust are the seven expressions of a person are classified by the model. The real time facial expressions images of a person are captured through

a webcam, and used to categories the emotion by the proposed model. This FERS can be used to analyze user expressions in order to improve the system's understanding of human needs.

II. METHODOLOGY

Deep learning models can attain state-of-the-art accuracy, even outperforming human performance in some instances. Deep learning technique is used in this model. "Keras" is a deep learning open library for facial expression recognition that was launched by "Google." The proposed model is using a robust CNN to recognise images. To recognize face expressions, a variety of datasets are employed. JAFFE Facial Expression Database, Cohn-Kanade AU-Coded Expression Database, CASIA NIR, Facial Expression Recognition Challenge Dataset, and CK+ Facial Expression Detection datasets are available. In this work, Facial Expression Recognition (FER-2013) dataset is used. It is taken from Kaggle repository. The dataset consists of 35,887 pictures with a 48x48 pixel restrictions. All of the photos in this dataset are grayscale portraits of humans. These faces have already been automatically added such that they are more or less identical in each picture and take up roughly the same amount of area. The goal is to categorize each face into one of the seven categories based on the emotion expressed in the facial expression (0=Angry, 1=Disgusted, 2=Fearful, 3=Happy, 4=Neutral, 5=Sad, 6=Surprise). There are 28,709 samples in the training set and 3,589 in the public and private testing set. The FER-2013 dataset has a large number of datasets; it has a greater prediction performance than other datasets like the CK+ dataset or the JAFFE dataset. Expanding the dataset size improves the training model's efficiency, leading to a high performance and confidence score. PyCharm IDE (Professional Edition) and Python 3.9 as the core interpreter is used to run the model.

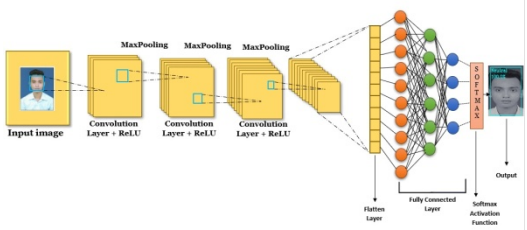

Fig. 4 CNN design with a sample input data

The model has three convolutional layers, three pooling layers for extracting features, one fully connected layer, and finally a softmax layer with seven emotion classifications. The input image is a grayscale facial image with a 48x48 pixel size. The FER-2013 dataset is used to train the model, and to evaluate loss accuracy. Photos from the FER-2013 dataset are considered that includes the seven expressions of facial emotions. The proposed model classifies emotion using "Keras" open library. Table 1 describes the CNN configuration fir the proposed model.

Table 1. CNN configuration

Layer Type	Input size/ Kernel size	Output Shape
Input Data	48x48	(None, 46, 46, 32)
Conv2D_1	3x3	(None, 44, 44, 64)
MaxPooling2D_1	2x2	(None, 22, 22, 64)
Dropout_1	0.25	(None, 22, 22, 64)
Conv2D_2	3x3	(None, 20, 20, 128)
MaxPooling2D_2	2x2	(None, 10, 10, 128)
Conv2D_3	3x3	(None, 8, 8, 128)
MaxPooling2D_3	2x2	(None, 4, 4, 128)
Dropout_2	0.25	(None, 4, 4, 128)
Flatten	-	(None, 2048)
Dense_1	-	(None, 1024)
Dropout_3	0.5	(None, 1024)
Dense_2	-	(None, 7)

Figure 5 shows the step of operations for the proposed model. Firstly, the model detects the face from the input image, which is then cropped and normalised to a size of 48x48 pixels. These facial images are then loaded into CNN as input. Then the facial expression is recognized and the result such as anger, disgust, fearful, happy, neutral, sad and surprised is achieved as shown in the figure 5.

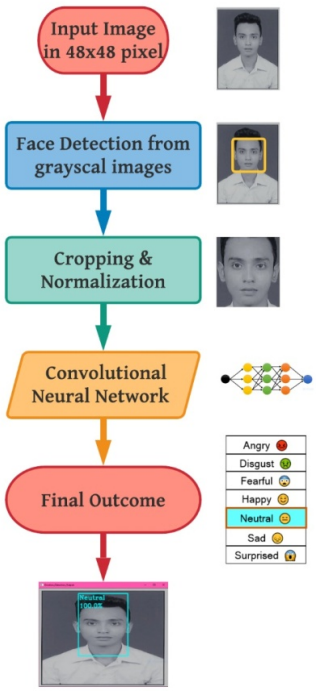

Fig. 5 Proposed model design

III. IMPLEMENTATION

The OpenCV package is used to take live frames from a web camera. The Haar cascades approach recognizes human facial emotions as shown in figure 6. The Haar cascade technique is a common edge and line detection tool. The framework of the final classifier, on the other hand, allows for an execution of the detector. In order to conduct an efficient result of classifiers, the Adaboost algorithm is used to selects a small number of significant features from a large set. TensorFlow and Keras API are used to create the CNN in the model. Inside Keras, image data generator class is used, to enhance the pictures quality. It offers a variety of augmentation options, including standardization, rotation, shifts, flips, and brightness changes. Then we added three convolutional layers, three pooling layers, and one fully connected layer to our CNN model. To add non-linearity in the CNN model, ReLU function is used at dropout layer. The main reason for adding this is to nullify those neurons; those have very less impact to the next layer while passing the data through it. Softmax function is used as the final activation function at output layer of the CNN ([14]-[16]). Because, softmax function creates a probability distribution from a set of values. Softmax is frequently employed as the activation for the output layer of a classification network due to its multiclass classification feature [17]. For better visualization, it is shown in figure 7.

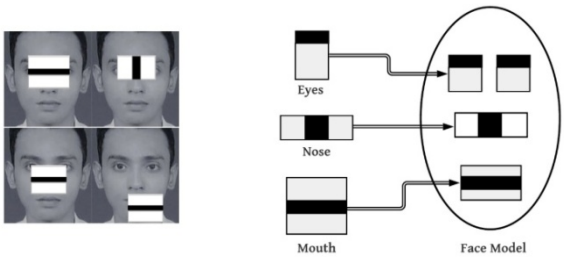

Fig. 6 Face detection using Haar cascade

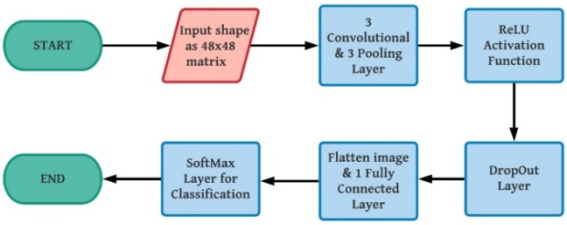

Fig. 7 CNN architecture for the proposed model

To build the CNN model, we partitioned the database into 80% of training dataset and 20% of testing dataset, and then used Adam optimizer to construct the model [18]. Keras examines each epoch to see whether the model outperformed the previous epoch's models. In addition to this, weights are saved into a file if the optimum case is obtained.

IV. RESULT ANALYSIS

The FER-2013 database is used to train the CNN model with the help of openCV and tensorflow. Table 2 shows the number of images in each emotion considered in FER-2013 dataset. The face picture captured through the webcam is rescaled to 48x48 pixels and converted to grayscale images by the model, which then used as inputs to the CNN model. At 100 epochs, the model reached 60% accuracy rate. The model training loss verses validation loss and training accuracy verses validation accuracy is represented in figure 8.

Table 2. Images under each class of FER-2013 dataset

Labels	Emotions	Number of images
0	Angry	4953
1	Disgust	547
2	Fearful	5121
3	Happy	8989
4	Neutral	6198
5	Sad	6077
6	Surprised	4002

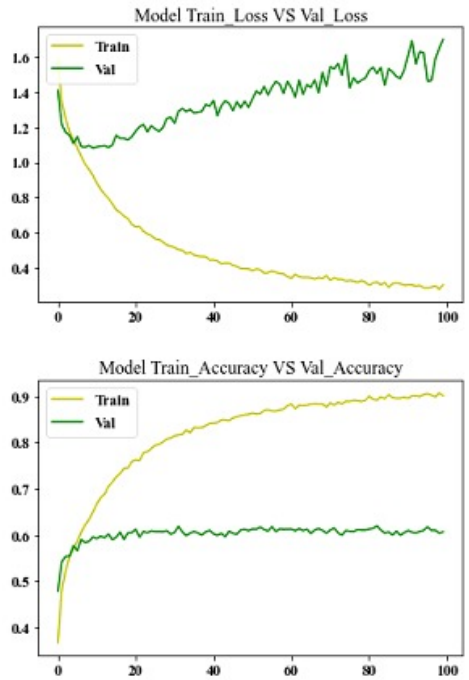

Fig. 8 Model loss and accuracy for training and validation

To evaluate the efficiency of the model, the

confusion matrix, precision, recall, F1-score, and support is calculated and presented in Table 3 [19]. Figure 9 shows the confusion matrix for the model.

Table 3. Model accuracy

	Precision	Recall	F1-score	Support
Anger	0.48	0.52	0.50	958
Disgust	0.76	0.50	0.61	111
Fear	0.50	0.39	0.44	1024
Happy	0.81	0.81	0.81	1774
Neutral	0.45	0.52	0.48	1247
Sad	0.78	0.74	0.76	831
Surprise	0.54	0.54	0.54	1233
Accuracy			0.60	7178
Macro avg	0.62	0.58	0.59	7178
Weighted avg	0.61	0.60	0.60	7178

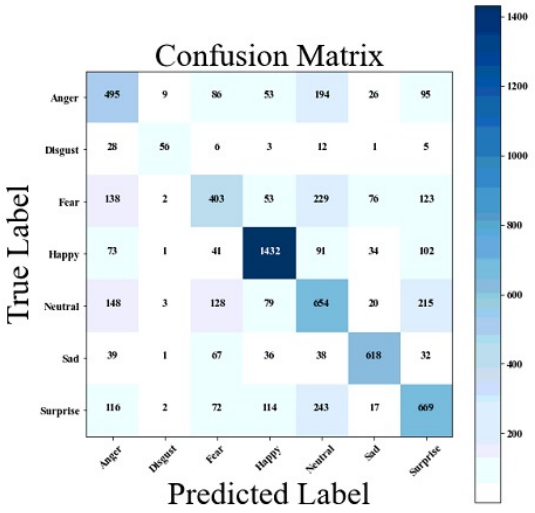

Fig. 9 Confusion matrix of the model

Figure 10 to 13 shows the emotion detection from static images by the model, and figure 14 to 16 shows the emotion detection from real time video by the model. From these results, it is found that the model classification accuracy vary from 91.7% to 100%.

Fig.10 Angry

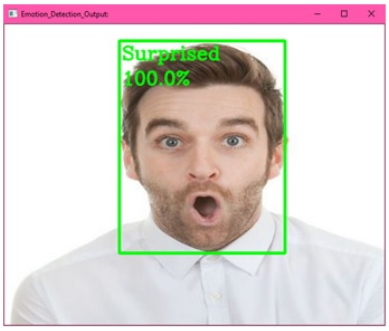

Fig. 11 Surprised

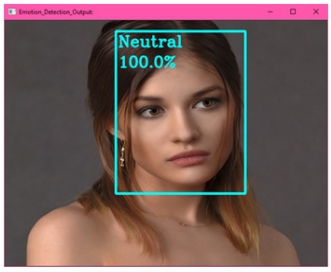

Fig. 12 Neutral

Fig. 13 Disgusting

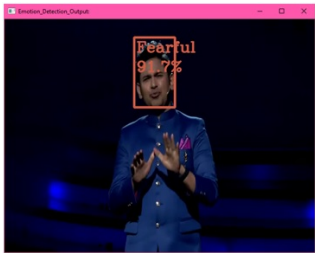

Fig. 14 Fearful

Fig. 15 Happy

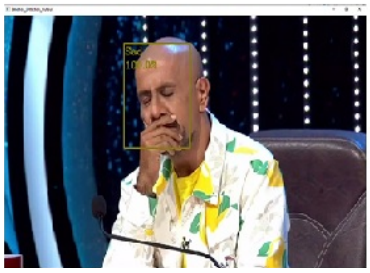

Fig. 16 Sad

V. CONCLUSION

An experienced human may often identify another human's feelings by assessing and glancing at that individual. Machines, on the other hand, are becoming increasingly sophisticated in today's world. Machines are currently aiming to act more human-like. If the machine has been trained to respond in the current sense of human sentiment.The machine can then act and react as if it were a human being. On the other hand, if a machine can recognise emotion, it can prevent a lot of problems.To find the emotion manifestation patterns, this framework is designed. This follows the framework step by step to reach the intended outcome.To follow the framework more effectively and recognise emotion expression patterns, as well as to use deep learning methods. The algorithms Keras, TensorFlow, and CNN are now being studied.Recognize faces in real time, facial detection with the Haar Cascade approach, and facial emotion classification with the CNN strategy are all possible. It has achieved in constructing a system with a general description of the study emotion using the CNN approach for the prediction of seven human facial expressions utilizing FER-2013 dataset. The FER-2013 dataset is used in the training phase, and feature extraction and facial prediction are done using the CNN technique. We used the FER-2013 dataset, which comprises 35.9k grayscale images, for this study. In the future, we will be able to utilize our own customized datasets to create it on an embedded platform like as an Arduino or Raspberry Pi, as well as IoT devices, IP cameras, and other similar devices. These devices, which have been trained using our custom datasets, can be deployed to defend the outside community from those who have evil intentions or goals. The result analysis section shows that the classification accuracy of the proposed model is 91.7% to 100%, and the model accuracy is 60%.

ACKNOWLEDGEMENTS

The authors are thankful to the publisher for publishing their article in the IJERA.

REFERENCES

[1]. D Li, G Wen. MRMR-based ensemble pruning for facial expression recognition. *Multimedia Tools and Applications.* 2018 Jun*;77(12)*:15251-72.

[2]. M. K Rusia, D. K. Singh. An efficient CNN approach for facial expression recognition with some measures of overfitting. *International Journal of Information Technology.* 2021 Dec*; 13(6):*2419-30.

[3]. Dricu M, Frühholz S. Perceiving emotional expressions in others: activation likelihood estimation meta-analyses of explicit evaluation, passive perception and incidental perception of emotions. *Neuroscience & Biobehavioral Reviews*. 2016 Dec 1; 71:810-28.

[4]. U Hess, P Thibault. Darwin and emotion expression. *American Psychologist*. 2009 Feb;*64(2)*:120.

[5]. Charles Darwin, *The expression of the emotions in man and animals* (University of Chicago press; 2015 Jul 31).

[6]. Mohammed MA, Abd Ghani MK, Hamed RI, Ibrahim DA. Review on Nasopharyngeal Carcinoma: Concepts, methods of analysis, segmentation, classification, prediction and impact: A review of the research literature. *Journal of Computational Science*. 2017 Jul 1; 21:283-98.

[7]. Y H Kwon, Yea-Hoon, Sae-Byuk , and Kim, Shin-Dug. Electroencephalography based fusion two-dimensional (2D)-convolution neural networks (CNN) model for emotion recognition system. *Sensors*. 2018 May; *18(5)*:1383.

[8]. P Murugan, Pushparaja. Hyperparameters optimization in deep convolutional neural network/bayesian approach with gaussian process prior. *arXiv preprint arXiv:1712.07233*. 2017 Dec 19.

[9]. Z Zhang, Zhifei. Derivation of backpropagation in convolutional neural network (cnn). *University of Tennessee, Knoxville, TN*. 2016 Oct 18.

[10]. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, & A. Rabinovich, (2015). Going deeper with convolutions. Proceedings of the *IEEE conference on computer vision and pattern recognition* (pp. 1-9).
[11]. H Gholamalinezhad, H Khosravi. Pooling methods in deep neural networks, a review. *arXiv preprint arXiv:2009.07485*. 2020 Sep 16.
[12]. A. B. Shetty, J. Rebeiro. Facial recognition using Haar cascade and LBP classifiers. *Global Transitions Proceedings*. 2021 Nov *1; 2(2)*:330-5.
[13]. A Seyeditabari, N. Tabari, W. Zadrozny. Emotion detection in text: a review. *arXiv preprint arXiv:1806.00674*. 2018 Jun 2.
[14]. B. Desmet, V. Hoste. Emotion detection in suicide notes. *Expert Systems with Applications*. 2013 Nov *15; 40(16)*:6351-8.
[15]. K. Sailunaz, M. Dhaliwal, J. Rokne,R. Alhajj. Emotion detection from text and speech: a survey. *Social Network Analysis and Mining*. 2018 Dec; *8(1)*:1-26.
[16]. A. R. Jena, R. Das, D. P. Acharjya. An integrated approach for prediction of radial overcut in electro discharge machining using fuzzy graph recurrent neural network. *International Journal of Embedded Systems*. 2021; *14(4)*:345-54.
[17]. Y Gao, H Wang, Z Liu. An end-to-end atrial fibrillation detection by a novel residual-based temporal attention convolutional neural network with exponential nonlinearity loss. *Knowledge-Based Systems*. 2021 Jan 5; 212:106589.
[18]. U M Khaire, R Dhanalakshmi. High-dimensional microarray dataset classification using an improved adam optimizer (iAdam). *Journal of Ambient Intelligence and Humanized Computing*. 2020 Nov; 11 (11):5187-204.
[19]. G. M. Foody. Status of land cover classification accuracy assessment. *Remote sensing of environment*. 2002 Apr *1; 80(1)*:185-201.