Recognition of Facial Emotions using Convolutional Neural Networks

K Shivaraj, Pawan Kumar R, Malatesh K R, Lavanya A, S Fouziya, Mohammad Ali P H, Tejesh

ABSTRACT

Every day, people interact with one another through expressing their emotions. Body language, conversational tone, and facial expressions may all be used to communicate this message. Face expression is the most natural, clearest, and consistent way to convey one's feelings, hence it has a significant impact on the transmission of emotional content. Because of the similarity of human facial expressions, it may be difficult to identify one from the other even with trained eyes. There are numerous similarities between being surprised and afraid. As a consequence, it will be impossible to predict a person's face expression. This project aims to develop real-time facial expression-based emotion recognition applications. In this Deep Learning-based research, Convolutional Neural Networks (CNN) are used. MobileNet is used to train the model to recognise. Happiness, grief, surprise, and disgust are the four unique facial emotions that may be identified from one another. Because of this, the identification accuracy of this research was an astounding 85%. In the near future, other face expressions may be added to the previously developed software.

1. INTRODUCTION

It's crucial to remember that face expressions may convey human feelings. Individuals use expressions of emotion to convey their ideas and feelings in a clear and concise way because they are born with the ability to do so. This kind of communication is known as nonverbal communication. The expressions on a person's face may provide others hints about how to approach them. Human-computer interaction relies on the ability to recognise and comprehend the emotions of others.

The physical and mental health of a person have a significant impact on their emotional state. As a result, it is referred to as interpersonal communication. Despite the fact that people can express their emotions directly through their facial expressions, the similarity of many diverse expressions makes it challenging to accurately identify an emotion. For instance, the expressions of astonishment and terror are similar. In addition, expressions of fear and astonishment are similarly comparable. Therefore, these similarities might cause mistaken recognition to occur while using the unaided eye. An application for automatic emotion recognition is desperately needed because it can be difficult to recognize emotions traditionally.

Known as K-Nearest Neighbors, this method of emotion identification has an accuracy rate of over 85% and has been extensively employed in research. However, even though the KNN method consumes a large amount of memory, it is not very fast. High-accuracy and quick detection are two of the benefits of CNN, a Deep Learning-based approach. More than 90% accuracy has been achieved using CNN in automated emotion identification systems like this.

2. LITERATURE SURVEY

Yu and Zhang were able to achieve an accuracy of 0.612 using a five-layer ensemble CNN. After they trained their models using FER-2013, they employed Static Facial Expressions 2.0 (SFEW) to improve them. SFEW annotated video frames were scanned using a set of three face detectors, which the researchers used to identify and extract faces. Voting and data disruption were two of the methods used to enhance CNN's object recognition. Using stochastic pools, the researchers were able to get even better results with the little quantity of data they had.

Kahou et al. built a model that can recognise objects in both moving and still images by using a CNN-RNN architecture. The videos were created using AFEW 5.0 and the pictures were taken from the FER-2013 and the Toronto Face Database. The IRNNs, which are made up of linear units that have been rectified, were utilised in place of the LSTM units (ReLUs). These IRNNs were a simple solution to the issue of vanishing and bursting gradients. It's hard to believe, but their accuracy rate was just 0.528%.

Convolutional layers, max pooling and four Inception layers make up Mollahosseini et alnetwork. .'s Analysis of data from seven different datasets, including the FER-2013 dataset, was carried out utilising this network of computers. AlexNet networks were also compared to see how well they performed with the same datasets. The MMI and FER-2013 datasets were found to be the best matches for their respective designs, but the other five datasets also performed well. However, the MMI and FER-2013 datasets yielded superior outcomes. The FER-2013 dataset has an accuracy score of 0.664.

Ming Li and colleagues propose using a neural network model to address two issues associated with FERs based on still photos: the inter-variability of emotions between participants and the misclassification of emotions. Convolutional neural networks are used to build the model, with the first one learning from facial expression datasets, and the second one using DeepID to learn about a person's identify. Following the concatenation of these two networks, the Tandem Facial Expression (TFE) Feature is conveyed to the fully connected layers. Evaluation of the proposed model was carried out using both the Extended CohnKanade (CK+) and the FER+ databases. The CASIA-WebFace database was used to detect the characteristics. After 200 training cycles, the model achieved an accuracy of 71.1% on the FER2013 dataset and an accuracy of 99.31% on the CK+ database. Results from these studies show that the model is superior to a wide variety of cutting-edge techniques when applied to the CK+ and FER+ datasets.

It is suggested by Ming Li and colleagues that a neural network model be used to address the inter-variability and misclassification of emotions in FERs based on still photographs. With the use of two convolutional neural networks, the model can learn about an individual's identity by using facial expression datasets for one network and DeepID data for the other. The Tandem

Facial Expression or TFE Feature is conveyed to the fully connected layers following the concatenation of these two networks in order to generate a new model. Extensive CohnKanade (CK+) and Extended FER+ databases were used to test the proposed concept. The CASIA-WebFace database was used to determine the identifying characteristics of the individual. It took 200 training cycles to achieve an accuracy of 71.1% on the FER2013 dataset and 99.31% on the CK+ database for the model. For CK+ and FER+ datasets, the model outperforms a wide variety of cutting-edge techniques, according to these studies.

Facial expression recognition is an issue that has been tackled by a number of academics in the past. Tests were run on several models to determine the degree of accuracy that each model could achieve on one dataset, FER-2013, rather than a combination of numerous datasets.

3. METHODOLOGY

A web camera is used to record video of a person. The movie is broken down into individual frames, which are then used as input into a classifier so that the desired feeling may be extracted.

Three modules make up this system. Those are:

i. The camera records 30 frames of live action video at a frame rate of one per second during the preprocessing stage (fps). BGR is the abbreviation for the format in which the frames are stored. A grayscale version is created, which makes it much simpler to work with on a computer.
ii. Face detection: It is possible to utilise OpenCV's pre-trained classifier, known as Haar-cascade, to accurately recognise faces. Face coordinates will also be returned as a result. The picture will be cropped to show just the subject's face using these coordinates.
iii. Classifier: A licence has been issued making use of a Convolution Neural Network (CNN) model that has previously been constructed. The attached document contains the code that was used for the training of the model. Data that has been trained specifically for this purpose is used. Because of this, a list is generated that details the percentage of chance that each of the seven feelings will occur. The utmost output that is anticipated to be produced is the sensation that corresponds to the required output, which is the value that is highest among these.

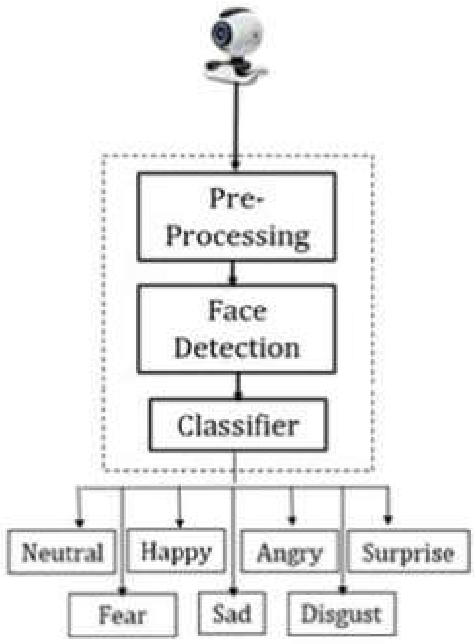


Fig. 1: High level Design of classifying emotions using video Input

The Convolutional Neural Network, or CNN, is a very accurate method for detecting objects since it is based on deep learning (Liam Schonevel 2021). CNN operates on numerous levels, each of which is in charge of a certain kind of change. Convolutional processing is used for feature extraction's initial layer. A convolutional technique is then used to find out how the pixels in a final picture are related to each other. Filters are applied to the convolution of an image in order to perform operations such as edge detection, blurring, and sharpening. ReLu's nonlinearity will be advantageous to ConvNet. The data that our ConvNet is trying to learn from is made up entirely of nonnegative linear values.

If the picture is too huge, the pooling layer will also attempt to reduce the total number of parameters. Each map's size may be reduced by subsampling or down sampling, a technique referred to as "spatial pooling." Using this strategy, you can be certain that none of the important details will be overlooked. The most common methods for spatial pooling are maximum, average, and sum. Before being sent to the layer, the matrix is first converted to a vector and then sent to the layer as a whole. This process is well-represented by analogies to neural networks. As seen in Figure 2, CNN's fundamental structure is depicted.

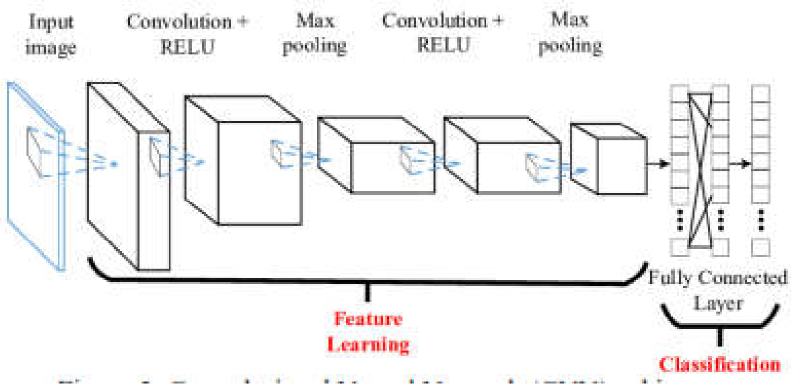


Figure 2. Convolutional Neural Network (CNN) architecture.

4. RESULT AND ANALYSIS

The testing technique is carried out to ensure that the application's intended purpose is realised. The reliability of an emotion reading is determined by analysing a picture of a person's face captured with the camera of a mobile device and sent to an application. A total of eighty photographs are utilised in the accuracy test of the programme.

There were a total of 80 photos utilised in the testing of the correctness of the system. False results are shown in Table 1, which suggests that the algorithm is incorrectly recognising the emotion. The results are misleading, since the algorithm is unable to reliably identify the correct emotion from an image of a facial expression. For example, the software will recognise a good result even if it was initially expected to be bad. Facial similarities between the candidates lead to inaccurate results in an online application. Based on the computation of average accuracy, a conclusion may be reached about the overall accuracy performance. Using Equation 1 as a guide, you may calculate the accuracy.

$$Accuracy = \frac{Number\ of\ correct\ prediction}{Total\ number\ of\ all\ cases} * 100\% \quad (1)$$

For every picture handed in for testing, a forecast is made for how many times that photo will be used in future predictions. The accuracy rate is 85 percent when the quotient is multiplied by 100. The application's total accuracy rate is 86%.

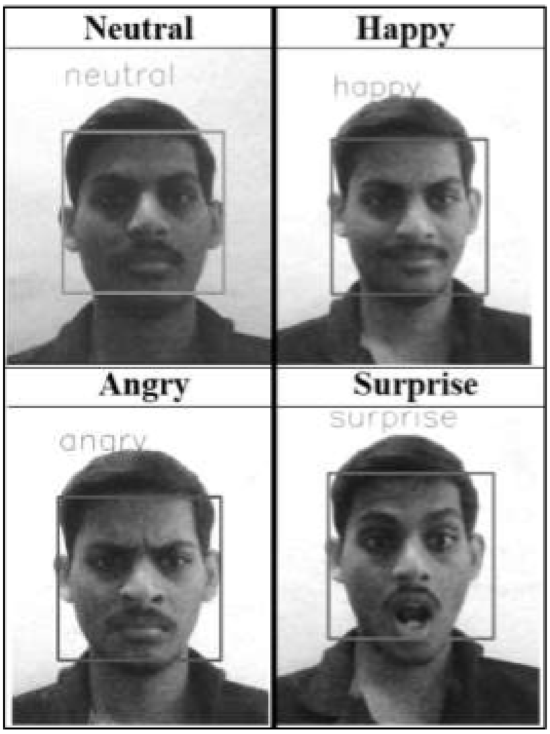
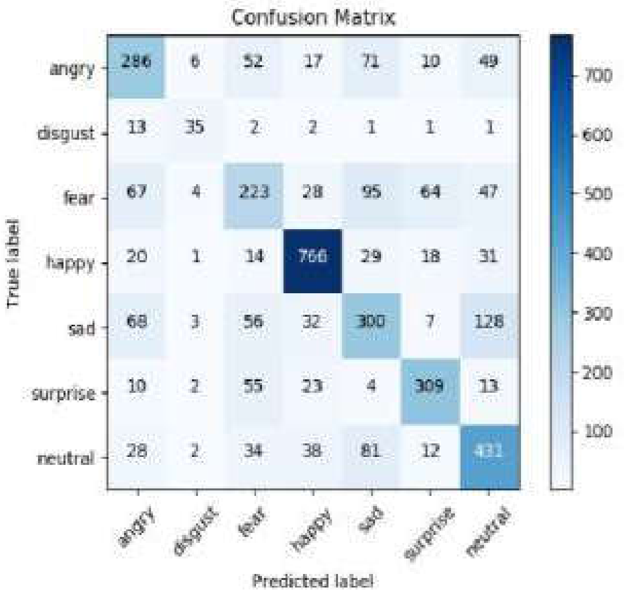


Fig. 3: Sample results of emotion classification of unseen data

Table 1 presents the confusion matrix for the many feelings that people experience. The matrix provides a comparison of the actual values of each emotion with the predicted values for that emotion. According to the matrix, there is a 0.14 percent probability of getting angry and sad confused with one another. Both surprise and fear have a 0.13 percent risk of being mistaken with one another. Because of the diagonal values, we have a more accurate ability to anticipate a variety of feelings. The word "happy" has the highest level of accuracy, coming in at 87 percent. With an accuracy rating of just 42 percent, fear is the emotion that can be recognised with the least amount of reliability.

Target Class	Predicted Class							Accuracy
	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral	
Angry	0.58	0.01	0.11	0.03	0.14	0.02	0.11	58%
Disgust	0.24	0.64	0.04	0.04	0.01	0.02	0.01	64%
Fear	0.13	0.01	0.42	0.05	0.18	0.12	0.09	42%
Happy	0.02	0.00	0.02	0.87	0.03	0.02	0.04	87%
Sad	0.11	0.01	0.09	0.05	0.51	0.01	0.22	51%
Surprise	0.02	0.00	0.13	0.06	0.02	0.74	0.03	74%
Neutral	0.04	0.01	0.05	0.06	0.13	0.02	0.69	69%

1(a)



1(b)

Table. 1 (a): Confusion matrix of the model showing ratio

1 (b): Confusion matrix of the model showing numbers

CONCLUSION

As a direct consequence of this, a software application that uses Convolutional Neural Networks (CNNs) to identify emotions was developed. The computer programme is able to differentiate between four distinct feelings, including surprise, disgust, melancholy, and surprise. The Convolutional Neural Network (CNN) is a beneficial expression that describes humans. It used the MobileNet algorithm with a one-of-a-kind dataset and was assessed using a confusion expression. An average accuracy rate of 92.50 percent was achieved by the application that was created. It was able to achieve a sensitivity of 85.00 percent while simultaneously achieving a specificity of 95.00 percent. Emotional recognition was therefore made possible for CNN, which resulted in positive outcomes that may help CNN in its work on succession planning. In the near future, it is expected that CNN's performance will be enhanced by merging it with any other AI method.

REFERENCES

[1] Liam Schoneveld, Alice Othmani, Hazem Abdelkawy. (2021). Laveraging recent advances in deep learning for audio-Visual emotion recognition. *Pattern Recognition Letters 146*, 1-7.

[2] Yishu Liu, Guifang Fu. (2021). Emotion recognition by deeply learned multi-channel

textual and EEG features. Futures Generation Computer Systems 119, 1-6.

[3] Neha Jain, Shishir Kumar, Amit Kumar, Pourya Shamsolmoali, Masoumeh Zareapoor. (2018). Hybrid deep neural networks for face emotion recognition. Pattern Recognition Letters 115, 101-106.

[4] Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, Remigiusz J. Rak. Emotion recognition using facial expressions. International Conference on Computational Science 2017;

[5] Shervin Minaee, Amirali Abdolrashidi. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. 2019;

[6] Nithya Roopa. S. Emotion Recognition from Facial Expression using Deep Learning. International Journal of Engineering and Advanced Technology (IJEAT) 2019; 8(6S):.

[7] Arkaprabha Bhandari, Nikhil R Pal. (2021). Can edges help convolutional neural networks in emotion recognition?. Neurocomputing 433, 162-168.

[8] Poorna S S, Anjana S, Parvathy Varma, Anjana Sajeev, Arya K Ca , Renjith Sb , G J Nair a. Facial Emotion Recognition using DWT based Similarity and Difference features. Proceedings of the Second International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) 2018; ():

.

[9] Firoz Mahmud, Bayezid Islam, Arfat Hossain, Pushpen Bikash Goala. Facial Region Segmentation Based Emotion Recognition Using K-Nearest Neighbors . Proceeding of the International Conference on Innovation in Engineering and Technology (ICIET) 2018; ():

[10] Marina Chatterjee. A Quick Introduction to KNN Algorithm. https://www.mygreatlearning.com/blog/knn-algorithm-introduction/ (accessed 10 May 2020).

[11] Wafa Mellouk, Wahida Handouzi. Facial emotion recognition using deep learning: review and insights. Procedia Computer Science 2020; 175(): .

[12] Ratnalata Gupta, Vishwamitra LK. (2020). Facial expression recognition from videos using CNN and feature aggregation. Materials Today: Proceedings.

[13] Jagli, P. M., & Shetty, M. P. (2019). Human Emotion Recognition using Machine Learning.

[14] International Journal of Trend in Scientific Research and Development, 101-103.

[15] Rassadin, & Savchenko. (2017). Compressing deep convolutional neural networks in visual emotion recognition. National Research University Higher School of Economics, Laboratory of Algorithms and Technologies for Network Analysis, 207-213.

[16] Stephanie Glen. Sensitivity vs Specificity and Predictive Value. https://www.statisticshowto.com/probability-and-statistics/statistics definitions/sensitivity- vs-specificity-statistics/ (accessed 15 June 2020).