A CNN-based Approach for Facial Emotion Detection

Sahana D*[1], Varsha K S[1], Snigdha Sen[2], Priyanka R[1]

1 Department of ECE, Global Academy of Technology, Bengaluru, Karnataka

2 Department of CSE, Global Academy of Technology, Bengaluru, Karnataka

{dayanandsahana@gmail.com,varshaacharya4129@gmail.com,
snigdha.sen@gat.ac.in, rameshpriyanka536@gmail.com}

Abstract: One of the most versatile ways in which individuals express their state of mind is through facial expressions. The advancement of deep learning-based technologies helped us to detect human emotion from images that can be used for understanding human feelings as well. The image can be static or can be captured through a web camera. The precise analysis of human facial expressions is necessary for a better understanding of human behaviour. With the recent progress in deep learning, CNN with its enhanced complex architecture is capable of emotion detection in a much better and more efficient way. In this paper, we experiment and demonstrate how to build a CNN predictor model using Tensor Flow that can predict the emotion from images of human facial expressions with satisfactory accuracy. Additionally, we also develop an application that asks for image input from the user and predicts the emotion from the given input image. Through this experiment, we are successful in demonstrating how CNN is an appropriate model for this task. Our work is beneficial in many applications such as lie detectors and student assessments to detect facial expressions very accurately.

Keywords: Convolution Neural Network, Deep Learning, Facial Emotion Recognition, Static image, Webcam.

1 Introduction

Emotions often arbitrate and ease interactions among human beings. A person’s state of mind is conceivable and determined by various modes such as speech patterns, gestures, and many other complex methods. However, the easier and more practical method is to examine facial expressions. Machine learning has shown tremendous performance in various areas like Data Analysis, Image Analysis, IoT, Health care, Astronomical Data Analysis [1][2][3][4], etc. Machine learning is a part of Artificial Intelligence that helps machines to understand patterns via a given dataset. Deep training is a subdivision of Machine Learning which is a very powerful technology capable of automatically extracting features from images or videos and handling complex data with high dimensions without human intervention. There are various Deep Learning algorithms like Recurrent Neural Networks RNNs, Generative Adversarial Networks GANs, Multilayer Perceptron MLPs [5], etc to serve various purposes. One such deep learning algorithm is the Convolution Neural Network CNN which is the type of artificial neural network, generally used to recognize and classify images or objects. Indeed, this algorithm is good for uprooting the characteristic of the image and is apt for image analysis subjects like image classification. CNN architecture consists of a heap of different surfaces that modifies input capacity into an output volume through a differential function. These multiple layers are made of artificial nodes in which each node gets weighted input data. The data is then passed into an activation function to output the result. CNN is proven to be an efficient recognition algorithm because of its striking features like simple structure, fewer training parameters, and adaptability. In this work, the TensorFlow framework is used to create a highly flexible CNN architecture to process pixel data and perform the task. When compared to the other deep learning algorithms, CNN has the advantage of requiring little pre-processing input. In this paper, we examine and explore how emotion recognition from facial expressions can be done using the CNN model which has numerous applications in real-life situations such as identification of driver's drowsiness, in medical research like autism therapy, and in mobile phones to automate clicking selfie.

The article is assembled into five sections:
The literature survey has been described in section 2. The dataset description goes on in section 3. Our presented framework is demonstrated in Section 4. Section 5 discusses the experimental setup and result of our trained model. Finally, we conclude with future work on Facial Emotion Recognition.

2 Literature Survey

Many researchers have already published an enormous amount of information on the FER field. For example, in the late 20th century, the value of FER was identified by Charles Darwin in the book "The Expression of Emotions in Man and Animals". He mainly described emotions. Corneanu et al. fundamentally classified emotion recognition by multimodal approaches. He concentrated on methods and parameters used for emotion recognition. He focused on the classification of FER on the principles of parameterization and facial expression [6]. The first FER model was created by Matsugu et al. He operated it with the CNN model which generated robust identification and unconventional of the subject [7]. Concerning the Matsugu CNN model, the Fasel model consists of 2 CNN which were used to recognize facial expressions and face identity recognition. He performed on 5600 inert pictures of 10 topics and achieved an accuracy of 97.6% [8]. Anil and others developed the terse survey of the techniques; this describes emotion recognition and exact value on several databases. A succinct differentiation was made between 2D and 3D methods [9]. Mohammed et al. newly discovered an algorithm that was a combination of Extreme Learning Machine, Bilateral Two-Dimensional Principal Component Analysis, and curvelet-based algorithm. He succeeded in getting a very high rate by curvelet features which gave him vast replicas and therefore face emotion recognition was obtained [10]. The local Directional Number Pattern was discussed by Rivera and others. The techniques which had the competence of surpassing the rules were distinct from several systems [11]. Shan and others recommended the Local Binary Pattern because characteristics delivered at a fast rate with contrast to the Gabor wavelets.
Then focussing on SVM, he mainly focused on algorithms like linear discriminant analysis and template matching [12]. Yu and others suggested a method that had three states of the art face sensors continued with different deep CNN models. Merging of CNN was taken place by reducing the hinge loss and log-likelihood loss [13]. Using deep neural network emotion recognition was performed based on three architectures by Enrique Correa et al. His first architecture contained three complexity layers with two completely linked layers. Enrique Correa et al. operated second architecture with three connected layers than using two linked layers which made paced up the operation. He used three separate layers for the third architecture as a max-pooling layer, convolutional layer, and local contrast normalization, as time passed, they upgraded the third max-pooling layer to minimize the limit factor. They obtained precision for architectures of about 63%, 53%, and 63% considerately. The research

shows having a minimum mesh size decreases the process of an authentic network more than anticipated values. Therefore, it concluded that the second architecture was not determined as the other two architectures [14]. For FER, Kahou and others operated on a hybrid CNN-RNN architecture. The combination of CNN and RNN was used to create a hybrid model. As the study shows, they constructed three CNN types- 3x3 frame size, 5x5 filter size with three layers, and 9x9 filter size. They merged and operated feature level and decision level as a result they obtained remarkable upgrades. The cluster of CNN, RNN and mean of per structure deploy categorized methods were used to excel this architecture [15].

3 Description of Dataset used

Here, we have upskilled the model by using a dataset called FER2013 which is publicly available on Kaggle. The FER2013 is a foresighted dataset and it's frequently used in ICML and manifestos. This is one of the most difficult databases with the human point accuracy of 65% and maximal operating publish work performing efficiency of 75.2%. The dataset consists of 35,887 images that are assigned to 48×48 pixels in monochrome. Yet FER2013 is not an established index, as it includes pictures of the 7 facial interpretations with a circulation of Anger 4,953, Disgust 547, Fear 5,121, Happy 8,989, Sad 6,077, Surprise 4,002, and Neutral 6,198[16]. The priming activity is accomplished with a fixed framework which will generate a upskill miniature that is used as a forecast criterion and is stored in a folder with the appendix. The details tutoring procedure design entering training data and attestation facts. The instruction details are refined using the CNN method on create attribute descent that will be approximate to information affirmation. Interpretation conclusion will outcome skilled exemplary architectonics to reach the extremity era rate. The CNN designs residual modules and amalgamation in a complexity layer.

4 Our proposed framework

Our proposed workflow has been represented in fig1.In this work, we explore how Facial Emotion Recognition can be done in two ways such as,
I). Facial Emotion Recognition from a static image
II). Real-time Facial Emotion Recognition through webcam

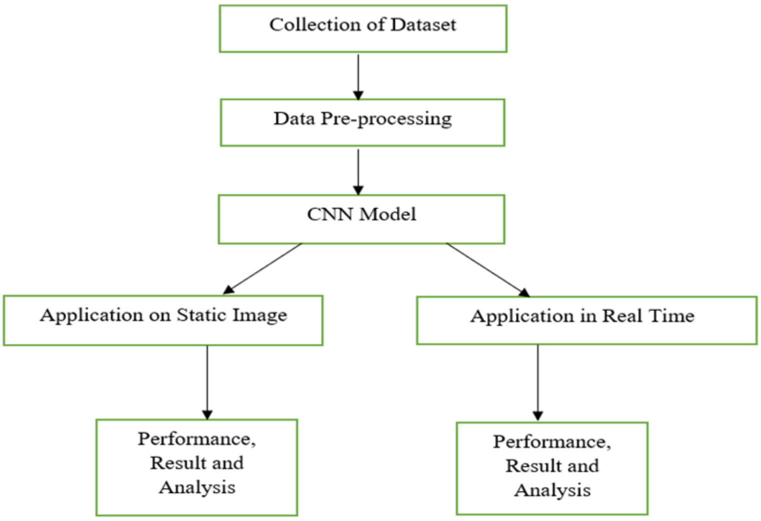

Collection of Dataset
Data Pre-processing
CNN Model
Application on Static Image
Application in Real Time
Performance, Result and Analysis
Performance, Result and Analysis

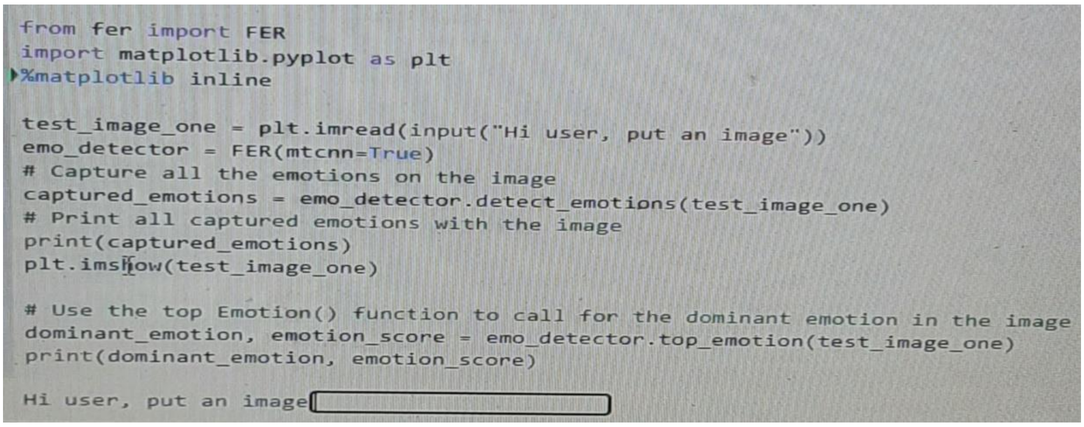
Fig. 1. Proposed workflow

5 Experimental setup and result

The whole project has been carried out using Python in Jupyter Notebook. It is a web application for making and sharing computational archives. It offers a basic, smoothed-out, report-driven insight for executing machine learning algorithms.

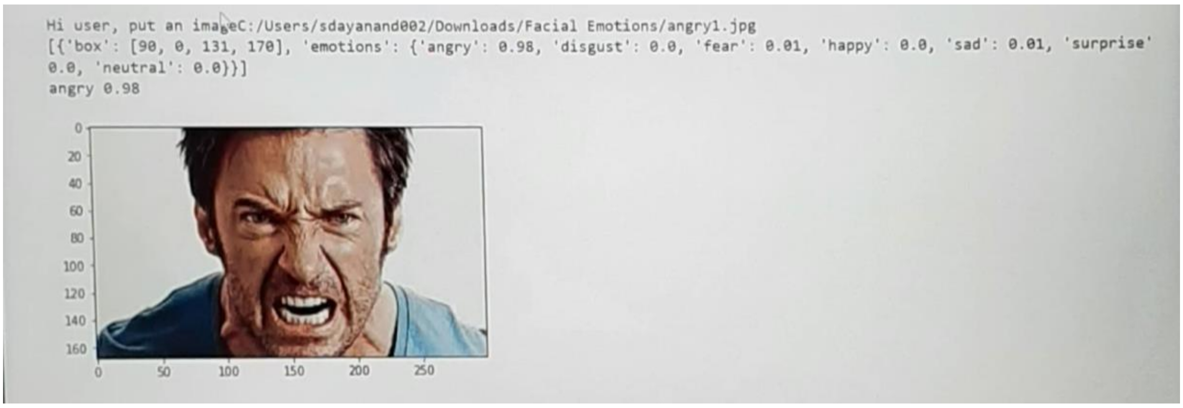
I Facial Emotion Recognition from a static image

A program is developed by using an open-source python library called Facial Emotion Recognizer (FER) for sentiment analysis. The primary function of any sentiment analysis is to isolate the polarity of the input (textual content, speech, facial features, and many more) and understand whether the primary sentiment presented is positive, negative, or neutral. The FER() constructor is set by giving it a Multi-Task Cascaded Convolution Network which is a type of neural network to identify faces and facial expressions. When the Multi-Task Cascade Convolution Network is initialized to ‘True’ the model detects a face, and when it is initialized to ‘False’ the function makes use of the OpenCV Haarcascade classifier. On program execution, it first asks for user-defined image input and once the user gives the input of his choice it specifies different emotions along with intensity levels in the output. The emotions are categorized into 7 categories namely ‘Fear’, ’Neutral’, ’Disgust’, ’Sad’, ’Happy’, ’Anger’, and ‘Surprise'. Each emotion is calculated, the result is placed on a scale of 0 to 1 and finally, the dominant emotion with the highest score is displayed. In Fig.2 we have shown the sample screenshot of user-defined image input. In Fig.3 we have shown the representation of scores of various emotions on a scale of 0 to 1 and dominant emotion with the highest score displayed at the bottom.

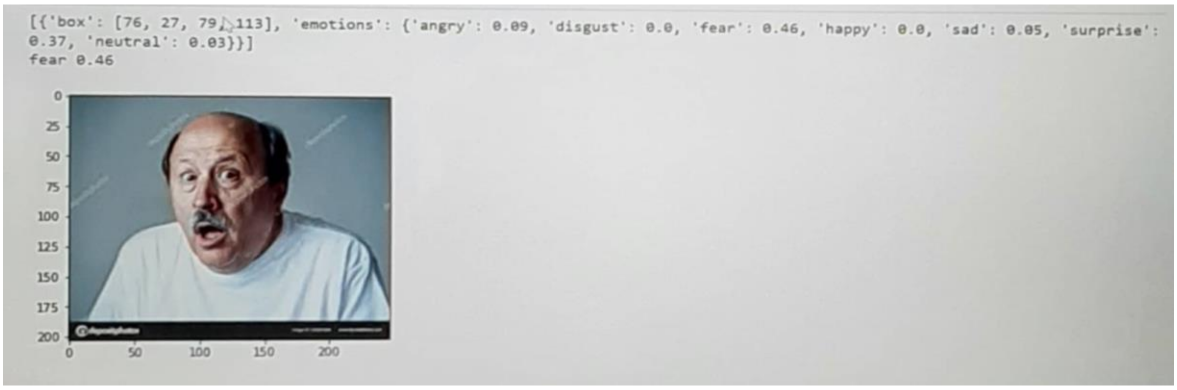


```
from fer import FER
import matplotlib.pyplot as plt
%matplotlib inline

test_image_one = plt.imread(input("Hi user, put an image"))
emo_detector = FER(mtcnn=True)
# Capture all the emotions on the image
captured_emotions = emo_detector.detect_emotions(test_image_one)
# Print all captured emotions with the image
print(captured_emotions)
plt.imshow(test_image_one)

# Use the top Emotion() function to call for the dominant emotion in the image
dominant_emotion, emotion_score = emo_detector.top_emotion(test_image_one)
print(dominant_emotion, emotion_score)

Hi user, put an image
```

Fig. 2. Sample screenshot of user-defined image input



```
Hi user, put an imageC:/Users/sdayanand002/Downloads/Facial Emotions/angry1.jpg
[{'box': [90, 0, 131, 170], 'emotions': {'angry': 0.98, 'disgust': 0.0, 'fear': 0.01, 'happy': 0.0, 'sad': 0.01, 'surprise'
0.0, 'neutral': 0.0}}]
angry 0.98
```

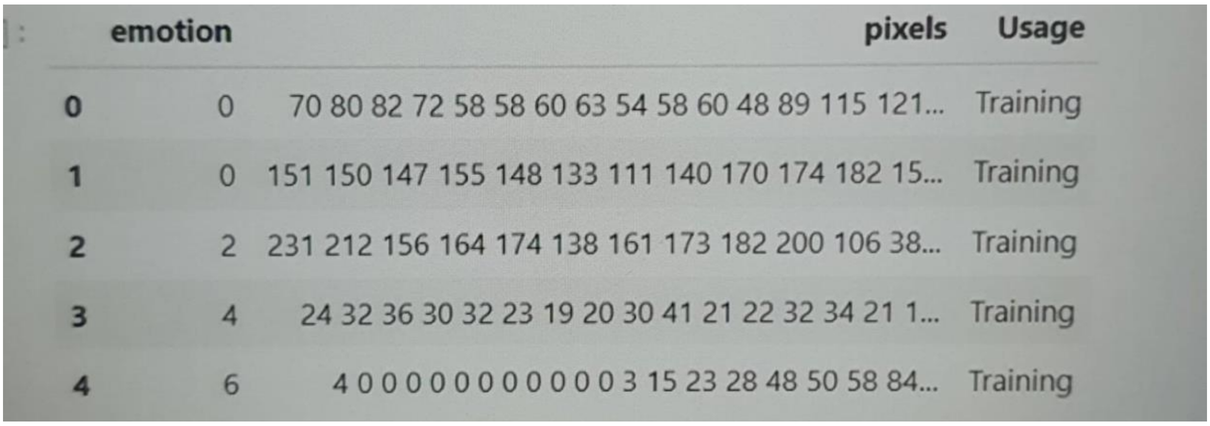


```
[{'box': [76, 27, 79, 113], 'emotions': {'angry': 0.09, 'disgust': 0.0, 'fear': 0.46, 'happy': 0.0, 'sad': 0.05, 'surprise':
0.37, 'neutral': 0.03}}]
fear 0.46
```

Fig. 3. Representation of scores of various emotions on a scale of 0 to 1 and the predominant emotion with the highest score is shown at the bottom.

II Real-time Facial Emotion Recognition

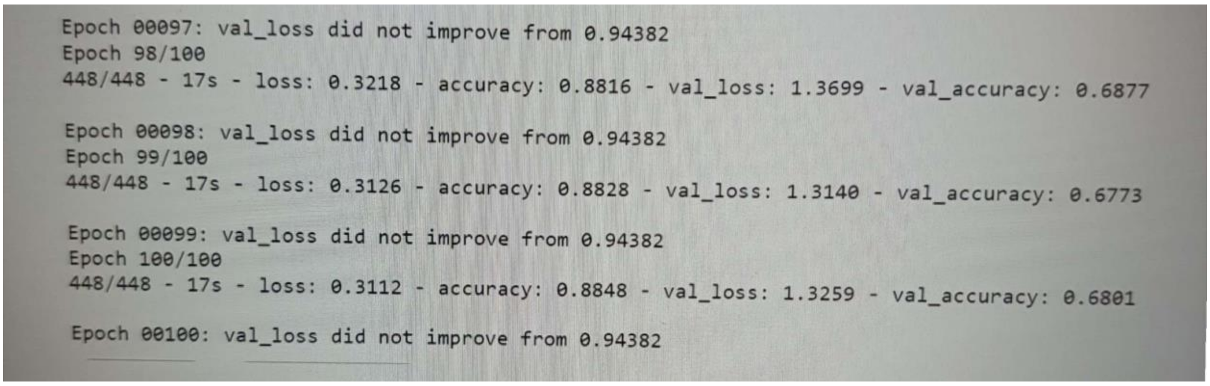
A CNN model with various convolutional filters working and examining the entire feature matrix is built using functional API to carry out the dimensionality reduction. Dimensionality reduction is done to convert the high dimensional dataset into the lesser dimensional dataset. A set of deep neural networks is created to analyse visual imagery. Each concurrent layer of the neural network is connected to the input neurons. Artificial neurons or nodes in CNN's accept the image pixels as input in the form of arrays. Since CNN's are feedforward networks, the progression of data takes place only in one direction, from their inputs to their outputs.
Fig.4 contains the snapshot of the loading dataset and Fig.5 depicts the Training model with the batch size of 64 and 100 epochs.



	emotion	pixels	Usage
0	0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121...	Training
1	0	151 150 147 155 148 133 111 140 170 174 182 15...	Training
2	2	231 212 156 164 174 138 161 173 182 200 106 38...	Training
3	4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1...	Training
4	6	4 0 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84...	Training

Fig. 4. Loading dataset

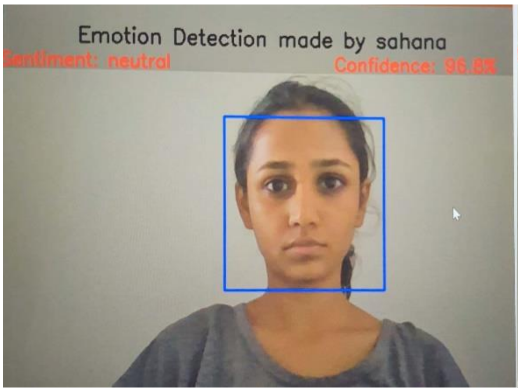
The model is trained by using the FER2013 dataset in which emotions are classified into
7 categories: 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise and 6=Neutral[17].
We have used 100 epochs with a batch size of 64 to train our model.

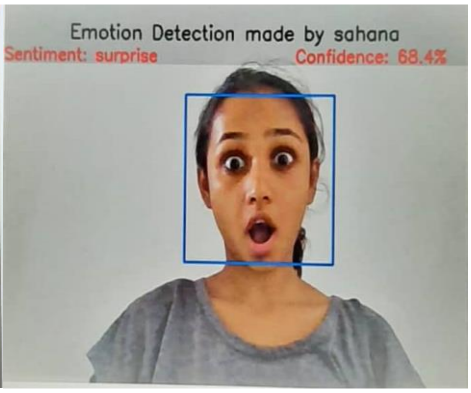


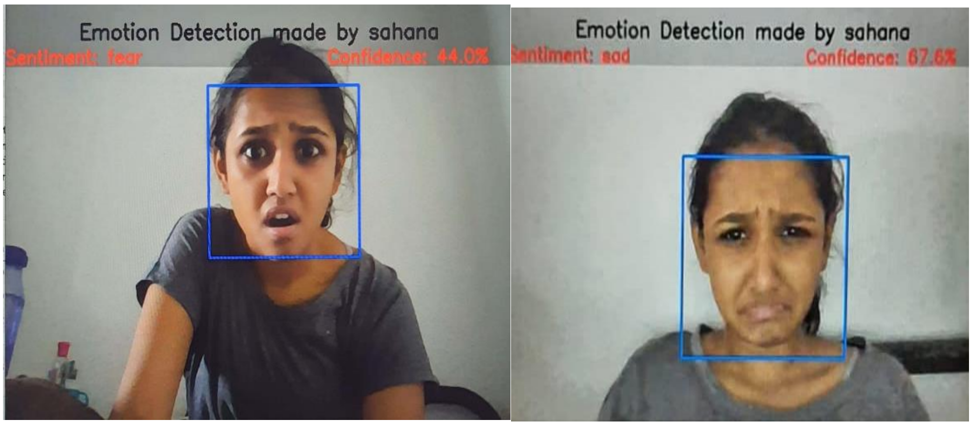
```
Epoch 00097: val_loss did not improve from 0.94382
Epoch 98/100
448/448 - 17s - loss: 0.3218 - accuracy: 0.8816 - val_loss: 1.3699 - val_accuracy: 0.6877

Epoch 00098: val_loss did not improve from 0.94382
Epoch 99/100
448/448 - 17s - loss: 0.3126 - accuracy: 0.8828 - val_loss: 1.3140 - val_accuracy: 0.6773

Epoch 00099: val_loss did not improve from 0.94382
Epoch 100/100
448/448 - 17s - loss: 0.3112 - accuracy: 0.8848 - val_loss: 1.3259 - val_accuracy: 0.6801

Epoch 00100: val_loss did not improve from 0.94382
```

Fig. 5. Training the model using 100 epochs with the batch size of 64

Inside the model, blocks are created using the Conv2D layer, Batch-Normalization, Max-Pooling2D, Dropout, and Flatten which are stacked together. Batch normalization is used to enhance the strength and performance of neural networks by allowing inputs with unit variance and zero means. Pooling decreases the dimensionality of characteristics while holding on to the most important data like the image hidden-layer output matrix. Dropout minimizes overfitting and voids the contribution of a few neurons towards the succeeding layer by arbitrarily not renovating the weights of some nodes. Finally, Flatten is used to transform multi-dimensional input arrays into a single-dimensional long continuous linear vector to classify the image. In the end, we use the Dense layer to get the output from the preceding layer and provide the output to the next layer. We have used a Python library called Facial Emotion Recognizer to identify faces and predict emotions. With the usage of Haar-Cascade, the position of faces is detected and cropped. We have used OpenCV to read frames and process the image. Image augmentation is done to enhance the overall performance and capability of the model to generalize. Finally, with the execution of code, the webcam automatically turns on and is capable of predicting emotion by looking into the facial expression of a person in front of the camera or from the human face images located on the webcam which is popularly known as Real-Time Facial Emotion Recognition. From Fig 6 and 7 it is very evident that our proposed CNN model is capable of predicting human emotions from various facial expressions.



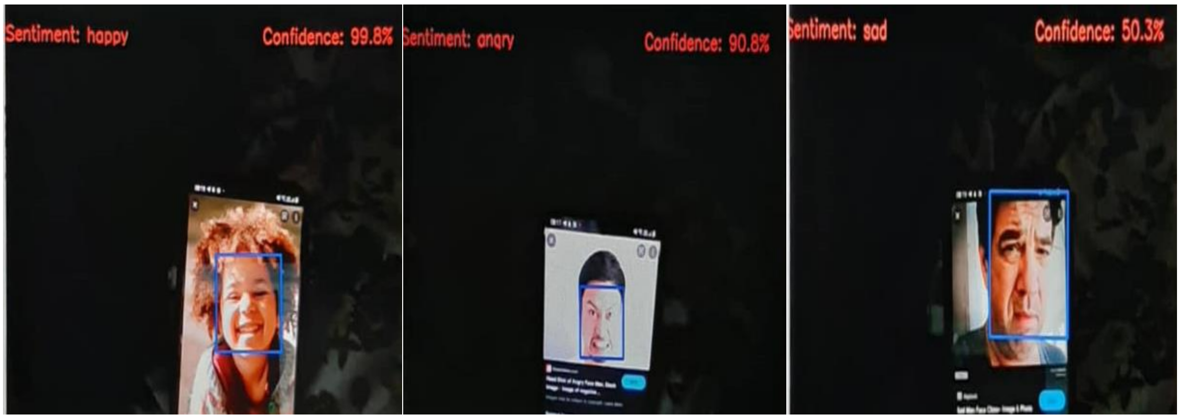




Emotion Detection made by sahana
Sentiment: fear
Confidence: 44.0%

Emotion Detection made by sahana
Sentiment: sad
Confidence: 67.6%

Fig. 6. Output window containing different human facial expressions along with emotion prediction and confidence



Sentiment: happy
Confidence: 99.8%

Sentiment: angry
Confidence: 90.8%

Sentiment: sad
Confidence: 50.3%

Fig. 7. Depiction of emotion and confidence predicted when images are shown to the webcam from a mobile phone

Conclusion

Facial expressions depict a non-verbal communication method that is vital in interpersonal relations. The results revealed above show how CNN is capable of understanding facial characteristics and detecting facial emotion. This technology can be implemented in many real-life situations such as lie detectors, mood-based learning for students, and detection of masked individuals[18]. Facial Emotion Recognition can be widely used in areas such as diagnosing mental illness and detecting social or physiological interactions between people. Furthermore, the study shows that Facial Emotion Recognition could provide society with better regard and contribute to Human-Robot Interface (HRI) interactions in the upcoming days. As a part of our future work, we plan to explore this technology in the medical field, specifically in the Psychology domain, to find out emotional states based on the observation of optical and audial nonverbal signs or gestural signs. Nonverbal signs or gestural signs include voice, postural, facial, and signs displayed by a person.

References

[1] Sen, Snigdha, et al. "Astronomical big data processing using machine learning: A comprehensive review." Experimental Astronomy (2022): 1-43.

[2] Sandeep, V. Y., Snigdha Sen, and K. Santosh. "Analysing and Processing of Astronomical Images using Deep Learning Techniques." 2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONNECT). IEEE, 2021.

[3] Sen, Snigdha, et al. "Implementation of neural network regression model for faster redshift analysis on cloud-based spark platform." International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Springer, Cham, 2021.

[4] Monisha, R., Snigdha Sen, Rajat U. Davangeri, K. S. Sri Lakshmi, and Sourav Dey. "An Approach Toward Design and Implementation of Distributed Framework for Astronomical Big Data Processing." In Intelligent Systems, pp. 267-275. Springer, Singapore, 2022.

[5] https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm

[6] Corneanu C.A., Simón M.O., Cohn J.F., Guerrero S.E. Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. IEEE Trans. Pattern Anal. Mach. Intell. 2016;**38**:1548–1568.doi: 10.1109/TPAMI.2016.2515606.

[7] Matsugu M., Mori K., Mitari Y., Kaneda Y. Subject independent facial expression recognition with robust face detection using a convolutional neural network. Neural Netw. 2003;**16**:555–559. doi: 10.1016/S0893-6080(03)00115-1.

[8] Fasel B. Robust face analysis using convolutional neural networks; Proceedings of the 16th International Conference on Pattern Recognition; Quebec City, QC, Canada. 11–15 August 2002; pp. 40–43.

[9] Anil J., Suresh L.P. Literature survey on face and face expression recognition; Proceedings of the 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT); Nagercoil, India. 18–19 March 2016; pp. 1–6.

[10] Mohammed A.A., Minhas R., Wu Q.J., Sid-Ahmed M.A. Human face recognition based on multidimensional PCA and extreme learning machine. *Pattern* Recognit. 2011;**44**:2588–2597. doi: 10.1016/j.patcog.2011.03.013.

[11] Rivera A.R., Castillo J.R., Chae O.O. Local directional number pattern for face analysis: Face and expression recognition. IEEE Trans. Image Process. 2013;**22**:1740–1752. doi: 10.1109/TIP.2012.2235848.

[12] Shan C., Gong S., McOwan P.W. Facial expression recognition based on local binary patterns: A comprehensive study. Image Vis. Comput. 2009;**27**:803–816. doi: 10.1016/j.imavis.2008.08.005.

[13]Yu Z., Zhang C. Image-based static facial expression recognition with multiple deep network learning; Proceedings of the 2015 ACM on International Conference on Multimodal

Interaction; Seattle, WA, USA. 9–13 November 2015; New York, NY, USA: ACM; 2015. pp. 435–442.

[14] Kahou S.E., Pal C., Bouthillier X., Froumenty P., Gülçehre Ç., Memisevic R., Vincent P., Courville A., Bengio Y., Ferrari R.C., et al. Combining modality specific deep neural networks for emotion recognition in the video; Proceedings of the 15th ACM on International Conference on Multimodal Interaction; Sydney, Australia. 9–13 December 2013; New York, NY, USA: ACM; 2013. pp. 543–550.

[15]Ebrahimi Kahou S., Michalski V., Konda K., Memisevic R., Pal C. ICMI '15, Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM; New York, NY, USA: 2015. Recurrent Neural Networks for Emotion Recognition in Video; pp. 467–474.

[16]I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee et al., "Challenges in representation learning: A report on three machine learning contests," in International Conference on Neural Information Processing. Springer, 2013, pp. 117–124.

[17] https://www.analyticsvidhya.com/blog/2021/11/facial-emotion-detection-using-cnn/

[18] Kumar, S., Yadav, D., Gupta, H. et al. Towards smart surveillance as an aftereffect of COVID-19 outbreak for recognition of face masked individuals using YOLOv3 algorithm. Multimed Tools Appl (2022). https://doi.org/10.1007/s11042-021-11560-1