

How to Cite:

Kanagaraju, P., Ranjith, M. A., & Vijayasathy, K. (2022). Emotion detection from facial expression using image processing. *International Journal of Health Sciences*, 6(S6), 1368–1379. <https://doi.org/10.53730/ijhs.v6nS6.9748>

Emotion detection from facial expression using image processing

Dr. P. Kanagaraju

Assistant professor in Computer Science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode-637 215, Namakkal District, Tamil Nadu, India
Email: kanagaraju@ksrct.ac.in

M. A. Ranjith

Students of Computer Science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode-637 215, Namakkal District, Tamil Nadu, India
Email: ranjithalagesan2000@gmail.com

K. Vijayasathy

Students of Computer Science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode-637 215, Namakkal District, Tamil Nadu, India
Email: kvijayasathy123@gmail.com

Abstract---Facial expression recognition is a powerful tool for communicating our emotions, understanding, and intent with each other. It is an intelligent human-computer interaction technology. Various studies have been conducted to classify facial expressions. Six fundamental universal emotions can be expressed through facial expressions: happiness, sadness, anger, fearful, surprised, and neutral. In this project, emotion detection can be implemented in real time with the help of a webcam. Our work proposed a CNN-based VGG16 architecture for emotion detection systems. A model would be trained by using the FER-2013 dataset. Then the images from the dataset are first pre-processed, which includes operations such as image scaling, changing the colour mode, and so on. Following that, a CNN model with multiple layers was created. After that, the model would be trained with the specified dataset, resulting in the .h5 file, which is a pre-trained model file. Instead of repeatedly training the model, the results can be predicted using this file. Based on input, emotions can be categorised as whether they are happy, sad, angry, terrified, surprised, or neutral; and it is also focused on the real-time analysis of people's problems and providing solutions based on their emotions; i.e., it will automatically play a video-based solution when it detects if the person is being sad, angry, or afraid.

Keywords---convolutional neural network, image processing, facial expression recognition, real time testing, emotion based video.

Introduction

One example of nonverbal communication is facial expressions. To categorise these facial expressions, various studies have been carried out. Facial expressions can convey six basic emotions: happiness, sadness, anger, fearful, surprised, and neutral. Face expressions are the only way for humans to express their emotions. In any image sequence, one can instantly recognise a person's facial expression. But when looking for an automated recognition system, the systems already available are quite inadequate and incapable of accurately identifying emotions. As a result, identifying such emotions is essential because it has a variety of applications, including virtual reality gaming, online education, medical care, online security, etc.

FER is a thriving research topic in which many advances are being made in industries, including automatic translation systems and machine-to-human interaction. Individuals must communicate their emotions through facial expressions. Humans use facial expression recognition to transmit their emotions and intentions in the most powerful, natural, and fast possible way. In various cases, such as when they are hospitalised or have certain disabilities, humans may be unable to convey their feelings; thus, enhanced recognition of other emotional responses will result in more effective communication. Due to the emergence of IoT based smart environments in hospitals, smart homes, and smart cities, automatic human emotion identification has received a lot of attention recently. Using biometric indicators, a technology called facial expression recognition will identify the emotions shown on a person's face. More specifically, this technology can recognise the six main or universal expressions: neutral, anger, fear, joy, sadness, and surprise automatically. Due to its capacity to imitate human coding talents, FER, is becoming more and more important. In interpersonal connections, nonverbal clues like facial features and other gestures are crucial. These cues support speech by assisting the listener in analysing the meaning of the spoken words. Face expression recognition uses data from the image/video feed to gather and analyse emotional responses, allowing it to provide unfiltered, impartial response data. FER employs algorithms to recognise emotional states, decode facial expressions, and detect faces in real time. It does this by analysing faces through images and videos using cameras that are powered by computers that are embedded in laptops, cell phones, and computer screens. Facial recognition, facial landmark detection, and facial emotion categorization are the three processes for face recognition with computer-assisted cameras.

In this Deep Learning project, a model is built for classifying emotions from facial expressions in images using CNN and other python libraries. More than 35,000 different emotional images are included in the image dataset (Angry, Fear, Happy, Neutral, Sad, Surprise). In all, around 6 different classes are present in the dataset for image classification. The dataset is only 60.3 MB in size, so it won't take a long time to download. It contains two separate folders: training and

validation, where each folder consists of classes on various emotions. Inside the class, there have been a variety of images that portray human emotions.

The CNN Model is used to train and build an emotion recognition model that can be used in any application. With some improvements and advancements, a CNN has evolved from a machine learning architecture. A convolutional neural network model is used for automatic image classification and image labelling. CNN was chosen because of its superior architecture and performance. An input volume is converted into an output volume using a differentiable function through a series of multiple layers that form a CNN architecture (e.g., holding the class scores). The convolutional layer becomes CNN's most important factor. A set of convolutional filters with a constant receptive field make up the layer's parameters, and they are taken into account. The 2D activation map for all of these filters is produced by computing the matrix multiplication between both the filter input and the input after each filter is merged across the size and width of the image region during the forward pass. As an outcome, the network can learn filters which activate when a certain type of feature is detected at a particular spatial point in the input.

Literature Review

This paper proposes a light-weight convolutional neural network (CNN) for real-time and extensive facial emotion recognition to improve classification performance. By removing the fully connected layer, integrating the residual modules, and using depth-wise separable convolutions, their network model drastically decreases the number of parameters in the convolutional layer while also making the model more portable. Their model achieves good detection results by detecting photographs taken outside of the dataset, demonstrating that the model they created is effective for multi-classification of facial expressions. In place of the fully connected layer in the classic deep CNN model, their expression classification approach uses global average pooling. They combine residual modules with depth-wise separable convolutions in their model, which reduces the number of parameters and increases the model's portability. Finally, their model is tested with the FER-2013 dataset with an accuracy of 67%, and it has good detection and recognition effects on those figures out of the dataset [1].

This paper proposes an automatic video-based facial expression detection system to detect and classify human facial expressions from the provided image sequence. They created the integrated automatic system, which usually consists of the two elements: feature extraction and peak of the expression from frame detection. They recommend employing the Double Local Binary Pattern (DLBP) in their facial expression detection method in order to find peak expression characteristics from any frame in the video. The suggested DLBP approach can successfully decrease the detection time and has a considerably reduced dimensional size. Additionally, they suggested using the domain Logarithm-Laplace (LL) to obtain a more accurate facial for detection in order to address the fluctuations in illumination in LBP. For the first time, their method employs the Taylor expansion theorem.

The author presents a method for extracting key facial characteristics from the Taylor feature map. Researchers created the Taylor Feature Pattern (TFP), which is based on the LBP and Taylor expansions. According to experimental results on each of the Japanese Female Facial Expression (JAFFE) and Cohn-Kanade (CK) data sets, the proposed TFP technique for facial emotion feature extraction and LBP-based feature extraction algorithms are best suitable for real-time application [2]. This paper proposes an approach for deploying a CNN model that classifies the emotional expressions of physically disabled people (deafness, dumbness, and bedridden status) and autistic children based on facial characteristics and electroencephalograph (EEG) signals (CNN), and they also use long-short-term memory (LSTM) classifiers by developing an algorithm for real-time emotion recognition using virtual markers through an optical flow algorithm that works effectively in uneven lightning and subject head rotation (up to 25 degrees), in different backgrounds and various skin tones. Ten virtual markers are used to capture six facial emotions: happy, sadness, disgust, fear, anger, and surprise. Fifty-five students from UG (35 males and 25 females) with an average age of 22.9 years volunteered for the facial expression experiment. EEG signals were collected by 19 undergraduate students who volunteered. Haar-like characteristics are first employed for face and eye identification. Then, by using the Lucas-Kande optical flow algorithm, virtual markers are placed on certain areas of the subject's face and tracked. For facial expression categorization, each marker point's distance from the face's centre is used as a feature. EEG signal emotional categorization characteristics are based on the 14 signals acquired through EEG signal reader (EPOC) channels. Finally, before sending the features to the classifier LSTM and CNN, the features are fivefold. They used CNN to detect emotions using facial landmarks, and their greatest recognition rate was 99.81%. For recognising emotions in EEG signals, the LSTM has a maximum detection rate of 87.25 percent. [3].

This paper proposes a method that extracts the facial features automatically and that is useful for FER tasks by using a weighted mixture deep neural network (WMDNN). Numerous preprocessing methods, such as face identification, rotation correction, and data augmentation, are employed to restrict the FER areas. WMDNN is used to process two channels of facial images, including LBP images and facial gray-scale images. The VGG16 architecture is implemented to extract facial expressions automatically and the elements from facial grayscale photos. The network is trained with fine-tuning by initialising the parameters, which are collected from the ImageNet database. A shallow CNN is built to extract the facial expression characteristics automatically based on the LBP facial images. The final recognition result based on fused features is obtained using SoftMax classification. The proposed algorithm can correctly identify the six fundamental facial emotions: joy, sadness, anger, disgust, fear, and surprise, according to their testing results. The average recognition accuracy scores for the benchmarking data sets "CK+," "JAFFE," and "Oulu-CASIA" are 97%, 92.20%, and 92.30%, respectively [4].

This paper proposes a method for classifying emotions based on face features and NSLBP properties. Gabor wavelet transform with the chosen sizes and orientations are used to generate the local features for each frame. An ensemble classifier uses these traits to determine where the face region is located. The

regions of the mouth and eyes are recognised using the ensemble approach and the distinctive qualities of each and every pixel on the face. The properties of the mouth and eyes are extracted using normalised semi-local binary patterns. To recognise the mood of the person from facial expression, the multiclass Adaboost algorithm is utilised to identify and classify the discriminative characteristics. When compared to the existing approaches, the new methods proposed by them perform much better on the RML, CK, and CMU-MIT databases, owing to their innovative characteristics [5].

This paper proposes research that offers a revolutionary image-based face expression recognition system. There are two main steps involved in their proposed system they are feature recognition and FER. In order to reduce the unpredictability of appearance changes, Haar like features are used in the face identification process, The FER technique collects HOG characteristics from each face area and then uses a SVM to categorise the final facial expression. A facial region is detected by Haar-like features. Then, the facial region of interest (ROI) is reset. The new facial ROI is used to generate HOG features. Based on SVM, FER is carried out on the retrieved HOG features. The ROI rearrangement technique reduced environmental change and the person's hierarchical structure, while facial expression categorization enhanced the classification rate. In the studies, the system correctly identified a person's facial expression, and the proposed method achieved an F1 score of 0.8759 [6].

Methodology

There are four phases in this methodology. They are: Emotional Database, Image Preprocessing, CNN Architecture, and Testing at Real Time.

Emotion Database

There are many open-access facial expression datasets available online. The emotion dataset has been downloaded through the Kaggle repository. In our project, the dataset used for training the model is the FER-2013 dataset. FER-2013 is a large dataset which is publicly accessible on Kaggle's FER Challenge. The FER-2013 dataset contains 35667 facial expressions. Among these, the train set is 28273, the public validation set is 3533, and the private validation set is also 3533. Each figure is composed of a grayscale image with a fixed size of 48 x 48. There are 6 expressions, which correspond to digital labels 0-5 respectively: 0, anger; 1, fear; 2, happiness; 3, neutral; 4, sad; 5, neutral. In the train set, there are 3995, 4097, 7215, 4830, 3171, and 4965 figures of the six kinds of expressions respectively.

Image Preprocessing

Image preprocessing refers to the steps taken to format images before they are used by model training and inference. The images from the dataset are subjected to image preprocessing. Preprocessing includes face detection and illumination correction, and performing some operations like image resizing, changing the colour mode, etc. The purpose of preprocessing is to increase the image quality so that we can better analyse it.

CNN Architecture

CNNs have been employed in a wide range of computer vision applications, including FER. The CNN-based VGG16 model was developed using the proposed CNN architecture. The network mimics the VGG16 architecture used in the classification of 2D facial expression data and includes the 13 convolutional layers with elu as an activation function, five max-pooling layers, and three fully connected layers. The convolutional layers have a kernel size of 3x3 and are batch normalised and stacked together, followed by a max-pooling layer with a kernel size of 2x2 and a stride of 2. After all the operations of convolutional layers and pooling layers, each frame was fed to the fully connected layers and the prediction of each frame was processed with the softmax classifier as six different facial emotion states.

The model should be compiled by using Adam as the optimizer, loss as a categorical cross entropy, and metrics as accuracy after it has been created. The model can be fit for training and validation after it has been compiled. Here the batch size is set as 128 with 55 epochs. A model overview of the mentioned CNN architecture is shown in Figure 1. Once the training has been completed, there is a need to evaluate the model and compute its loss and accuracy. Finally, the model can be saved as an .h5 file. Instead of repeatedly training the model, this pre-trained model file might be utilised to generate predictions. It classifies the emotions according to the input provided and the system should display whether they are happy, sad, angry, fearful, surprised, neutral or not.

Model: "sequential"		
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 48, 48, 64)	640
batch_normalization (Batch Normalization)	(None, 48, 48, 64)	256
conv2d_1 (Conv2D)	(None, 48, 48, 64)	36928
batch_normalization_1 (Batch Normalization)	(None, 48, 48, 64)	256
max_pooling2d (MaxPooling2D)	(None, 24, 24, 64)	0
dropout (Dropout)	(None, 24, 24, 64)	0
conv2d_2 (Conv2D)	(None, 24, 24, 128)	73856
batch_normalization_2 (Batch Normalization)	(None, 24, 24, 128)	512
conv2d_3 (Conv2D)	(None, 24, 24, 128)	147384
batch_normalization_3 (Batch Normalization)	(None, 24, 24, 128)	512
max_pooling2d_1 (MaxPooling2D)	(None, 12, 12, 128)	0
dropout_1 (Dropout)	(None, 12, 12, 128)	0
conv2d_4 (Conv2D)	(None, 12, 12, 256)	295168
batch_normalization_4 (Batch Normalization)	(None, 12, 12, 256)	1024
conv2d_5 (Conv2D)	(None, 12, 12, 256)	590080
batch_normalization_5 (Batch Normalization)	(None, 12, 12, 256)	1024
conv2d_6 (Conv2D)	(None, 12, 12, 256)	590080
batch_normalization_6 (Batch Normalization)	(None, 12, 12, 256)	1024
max_pooling2d_2 (MaxPooling2D)	(None, 6, 6, 256)	0
dropout_2 (Dropout)	(None, 6, 6, 256)	0
conv2d_7 (Conv2D)	(None, 6, 6, 512)	1180160
batch_normalization_7 (Batch Normalization)	(None, 6, 6, 512)	2048
conv2d_8 (Conv2D)	(None, 6, 6, 512)	2359808
batch_normalization_8 (Batch Normalization)	(None, 6, 6, 512)	2048
conv2d_9 (Conv2D)	(None, 6, 6, 512)	2359808
batch_normalization_9 (Batch Normalization)	(None, 6, 6, 512)	2048
max_pooling2d_3 (MaxPooling2D)	(None, 3, 3, 512)	0
dropout_3 (Dropout)	(None, 3, 3, 512)	0
conv2d_10 (Conv2D)	(None, 3, 3, 512)	2359808
batch_normalization_10 (Batch Normalization)	(None, 3, 3, 512)	2048
conv2d_11 (Conv2D)	(None, 3, 3, 512)	2359808
batch_normalization_11 (Batch Normalization)	(None, 3, 3, 512)	2048
conv2d_12 (Conv2D)	(None, 3, 3, 512)	2359808
batch_normalization_12 (Batch Normalization)	(None, 3, 3, 512)	2048
max_pooling2d_4 (MaxPooling2D)	(None, 1, 1, 512)	0
dropout_4 (Dropout)	(None, 1, 1, 512)	0
flatten (Flatten)	(None, 512)	0
dense (Dense)	(None, 64)	32832
batch_normalization_13 (Batch Normalization)	(None, 64)	256
dropout_5 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 64)	4160
batch_normalization_14 (Batch Normalization)	(None, 64)	256
dropout_6 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 6)	390
Total params: 14,768,826		
Trainable params: 14,739,622		
Non-trainable params: 8,704		

Figure 1. Model Summary for the proposed CNN architecture

Testing at Real Time

After training on the specified CNN model, the trained model has been tested in real time. First, human faces were found using a webcam and the Haar Cascade library. After that, the model is investigated to determine which classes the detected images fall into. The probability of the facial expression belonging to which class was shown on a separate screen, and the emotion of which class has a higher chance was overwritten on the Haar Cascade frame as a result of the predictions. The system should display the emotions of the people, and also, if the detected emotion is sad, angry, or afraid, the system should play a solution video to help the user overcome such an emotion.

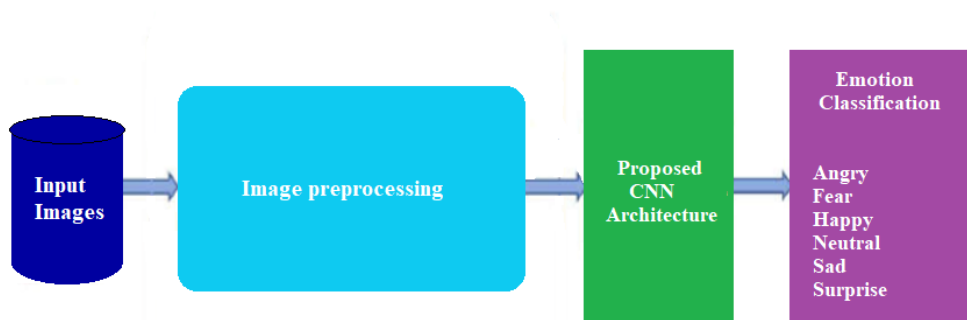


Figure 2. Block diagram for emotion detection from facial expression using image processing

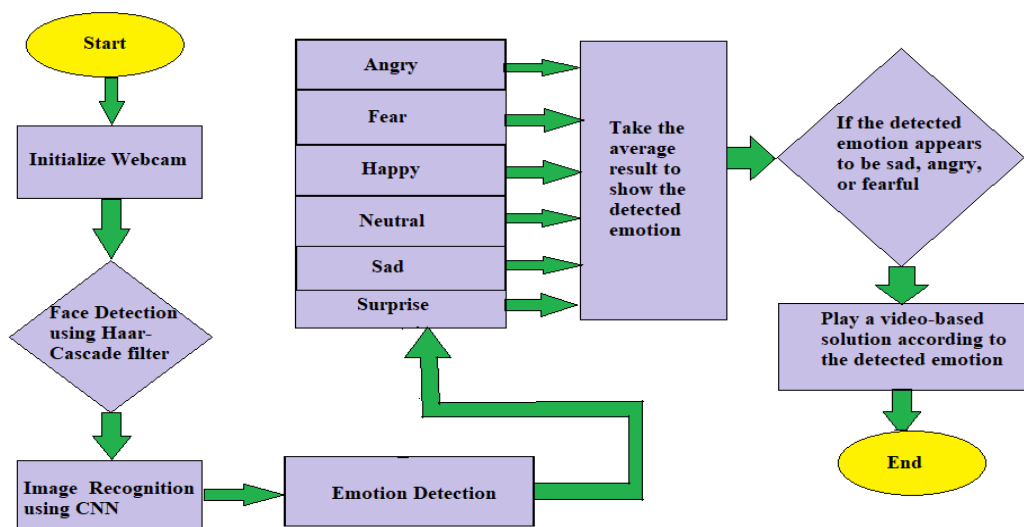


Figure 3. Flow chart for emotion detection from facial expression using image processing

Result and Discussion

This work has successfully created and achieved a result of being able to detect emotions from facial expressions. The FER-2013 dataset is used to train a model and generate a pre-trained model. It is now time to classify the relevant emotions using OpenCV and a webcam to test the model that was built for real-time emotion recognition. Also, it focused on people's real-time problems and provided solutions by analysing their emotions and knowing that: 1) If they are stressed or in depression when their emotions seem like sadness, then the system will automatically play a video to overcome that problem. 2) If they are scared of something when their emotions seem like fear, then the system will automatically play a video to get rid of that fear. 3) If they are angry at someone and their emotions seem like anger, then the system will automatically play a video to calm the person.

In this study, the VGG16 architecture was trained using the Keras and TensorFlow libraries, and a proposed deep learning model was employed to predict the emotional states. The AMD Ryzen 5 5500U processor was used for experiments and training the dataset. The proposed VGG16 model was set with the mentioned parameters. Figure 4 and Figure 5 show performance metrics (Model Accuracy: training accuracy and validation accuracy; Model Loss: training loss and validation loss) of the proposed architecture during training and testing. According to experimental results, training loss was found to be 1.0740; training accuracy was found to be 59.72%; validation loss was found to be 0.9299; and validation accuracy was found to be 64.52%.

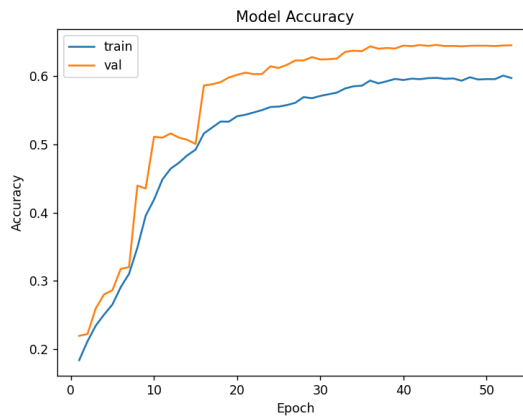


Figure 4. The accuracy graph of a model

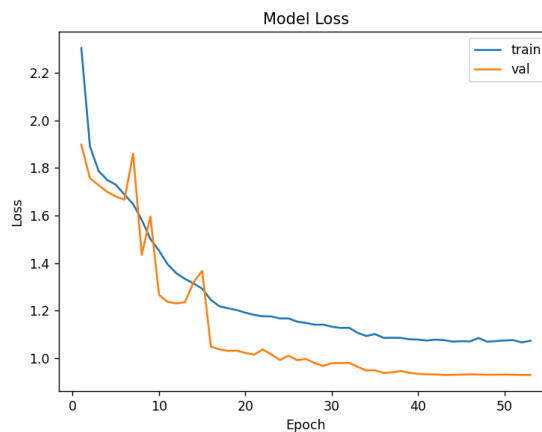


Figure 5. Loss graph of a model

In order to determine which emotions are easiest to distinguish and which emotions are more difficult to distinguish, it is necessary to calculate the confusion matrix. Figure 6 depicts the confusion matrix of our proposed architecture.

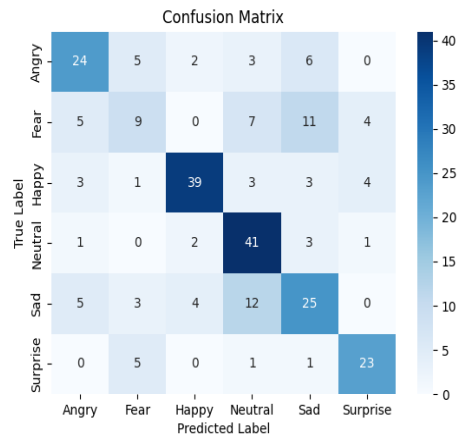


Figure 6. Confusion matrix for our proposed work

According to Figure 6. In the confusion matrix, the proposed VGG16 architecture model is more accurate at prediction of neutral, happy, sad, angry, and surprised, less accurate at prediction of fear emotion states.

Conclusion

In the area of computer vision, facial recognition is still a hard problem to solve. There are various investigations and research on emotion recognition. Emotions are an integral method of expressing our judgements and decisions in daily life, and this work aims to recognise and detect exactly these emotions. This work is capable of recognising 6 integral emotions: happy, sad, anger, fear, neutral, and surprise, and it will automatically play a video when it detects the following emotions: sad, angry, fear. In this work, emotion detection from facial expressions is proposed on the basis of image processing with the help of convolutional neural networks. FER-2013, which served as a dataset for the investigation, has been used to evaluate several information bases. The model is set up for training using the Keras and TensorFlow libraries, and it is tested on the FER dataset to see how accurate it is during training and validation. Then, the real-time test model has the functionality to query each image that occurs every second. Our model was able to achieve a 64.52% accuracy rate.

Future Work

Furthermore, the better results and the better accuracy for the CNN classifier to get the full, highly accurate result than the previous result can be detected using other deep learning algorithms and some machine learning algorithms. Also, we plan to play a video directly from YouTube based on the people's emotions. In order to make a user-friendly application, we plan to design it as a website and help people overcome their problems.

References

1. Ning Zhou, Renyu Liang, and Wenqian Shi (2020), "A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection", *IEEE Access*, vol. 9, pp. 5573-5584.
2. Aya Hassouneh, Mutawa A.M., and Murugappan M. (2020), "Development of a Real-Time Emotion Recognition System using Facial Expression and EEG based on machine learning and deep neural network methods", *Science Direct on Informatics in Medicine Unlocked*, vol. 20.
3. Yuanyuan Ding, Qin Zhao, Baoqing Li, and Xiaobing Yuan (2017), "Facial Expression Recognition from Image Sequence Based on LBP and Taylor Expansion", *IEEE Access*, vol. 5, pp. 19409-19419.
4. Biao Yang, Jinmeng Cao, Rongrong Ni, and Yuyu Zhang (2017), "Facial Expression Recognition using Weighted Mixture Deep Neural Network Based on Double-Channel Facial Images", *IEEE Access*, vol. 6, pp. 4630-4650.
5. [5]Ithaya Rani P., and Muneeswaran K. (2018), "Emotion Recognition based on facial components", *Indian Academy of Science*.
6. Guihua Wen, Tianyuan Chang, Huihui Li, and Lijun Jiang (2020), "Dynamic Objectives Learning for Facial Expression Recognition", *IEEE Transactions on Multimedia*, vol. 22, pp. 2914-2925.
7. Pawel Tarnowski, Marcin Kolodziej, Andrzej Majkowski, and Remigiusz J. Rak (2017), "Emotion recognition using facial expression", *Science Direct on Procedia Computer Science*, vol. 108, pp. 1175-1184.
8. Charvi Jain, Kshitij Sawant, Mohammed Rehman, and Rajesh Kumar (2018), "Emotion Detection and characterization using Facial Features", 3rd International Conference and workshops on Recent Advance and Innovations in Engineering (ICRAIE).
9. Shivam Gupta (2018), "Facial emotion recognition in real-time and static images", 2nd International Conference on Inventive System and Control (ICISC).
10. Leonid Ivanovsky, Vladimir Khryashchev, Anton Lebedev, and Igor Kosterin (2017), "Facial Expression Recognition Algorithm Based on Deep Convolution Neural Network", 21st Conference of Open Innovations Association (FRUCT).
11. [11] Sanghyuk Kim, Gwon Hwan An, and Suk-Ju Kang (2017), "Facial Expression Recognition System using Machine Learning", International SoC Design Conference (ISOCC).
12. Alaramzana Nujun Navaz, Serhani Mohamed Adel, and Sujith Samuel Mthew (2019), "Facial Image Pre-Processing and Emotion Classification: A Deep learning Approach", *IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*.
13. Mehemet Akif Ozdemir, Berkay Elaggoz, Aysegul Alaybeyoglu, Reza Sadighzadeh, and Aydin Akan (2019), "Real Time Emotion Recognition from Facial Expressions using CNN Architecture", *Medical Technologies Congress (TIPTKNO)*.
14. Parvathi D.S.L., Leelavathi N., Ravikumar J.M.S.V., and Sujatha B. (2020), "Emotion Analysis using Deep Learning", *International Conference on Electronics and Sustainable Communication Systems (ICESC)*.
15. Yassine Rabhi, Makrem Mrabet, Farhat Fnaiech, and Mounir Sayadi (2018), "A Real-time Emotion Recognition System for disabled persons", 4th

- International Conference on Advanced Technologies for Signal and Image Processing (ATSIP).
16. Shekhar Singh, and Fatma Nasoz (2020), "Facial Expression Recognition with Convolutional Neural Networks", 10th Annual Computing and Communication workshop and Conference (CCWC).
 17. Kustina, K.T., Dewi, G.A.A.O., Prena, G.D., Suryasa, W. (2019). Branchless banking, third-party funds, and profitability evidence reference to banking sector in indonesia. *Journal of Advanced Research in Dynamical and Control Systems*, 11(2), 290-299.
 18. Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2021). Get vaccinated when it is your turn and follow the local guidelines. *International Journal of Health Sciences*, 5(3), x-xv. <https://doi.org/10.53730/ijhs.v5n3.2938>
 19. Kusumawati, A. H., Wulan, I. R., & Ridwanuloh, D. (2020). Formulation and physical evaluation sheet mask from red rice (*Oryza Nivara*) and virgin coconut oil (*Cocos Nucifera* L). *International Journal of Health & Medical Sciences*, 3(1), 60-64. <https://doi.org/10.31295/ijhms.v3n1.148>