Turkish Journal of Computer and Mathematics Education *Vol.13 No.03 (2022), 195-203*

Research Article

Convolutional Neural Network for the Recognition and Characterization of Emotions using Double Average Filtering and SELU activation – Valence Cognizance

F. Ludyma Fernando [a], Dr. John Peter [b]

[a] Research Scholar, Manonmaniam Sundaranar University, Palayamkottai
[b]Associate Professor and Head of the Computer Science Department, St. Xavier's College, Palayamkottai, Affltd to Manonmaniam Sundaranar University, Tirunelveli

Abstract: An Emotion being a complex psychological state that involves both experience and action becomes a challenge to be recognized accurately by programmable codes. This paper demonstrates the method of identifying each out of seven basic emotional states (happiness, surprise, fear, anger, fear, disgust, sadness and neutral) and characterizing them as either positive or negative (valence) from images in a given dataset. This has been achieved to a higher accuracy by a Convolutional Neural Network designed with Double Average filters and the SELU (Scaled Exponential Linear Unit) activation units. The images from the FER 2013 dataset is processed (converted to gray scale and the dimensions set to 48x48) and given as input to the CNN. The Double Average Filters remove the noises much more efficiently than Average Filters, since the process is repeated to give even lesser intensity variations between the pixels. The SELU activation used in the CNN gives an internal normalization on the filtered images, which results in a much better identifying of emotions than with other activation unit. The SELU in recent times, as mentioned by other researchers too, is a promising part of any networks that can be used in Machine Learning. The proposed novel CNN model has a training accuracy of more than 96.53%.

Keywords: Convolutional Neural Network, Double Average Filtering, Emotion Recognition

1.Introduction

Emotions govern our daily lives. They are a big part of human experience, and inevitably they affect our decision making. Every action or interaction done by us result in some emotion or the other. Sometimes there can be no emotion too. Humans tend to repeat the actions that make happy and avoid those that make them unhappy. These emotions which result from human actions or interactions can be of different intensities and magnitudes which paves way for distinguishing one emotion from the other. The distinctness of emotions lies in the fact that they can be expressed and can be understood or recognized. Suppose, if a person is dissatisfied by any product that was purchased, he would express his anger of his loss by writing hateful reviews for that product. Hence, when the review is read, the subjective information showcases the emotion (hatred/anger/disappointment) expressed by the author.

According to Psychological Theory, an Emotion is a complex psychological state that involves three distinct components: a subjective experience, a physiological response and behavioural or expressive response. This explains that an emotion can be felt by the person himself (Subjective Experience), can be expressed with actions (Physiological response) and through words or facial reactions (Behavioral/Expressive Response).

The above definition concludes two ideas with respect to Emotions:

1. *Subjectivity and Complexity*: An emotion is dependent on the person himself. The magnitude of the emotion may be deep to one and superficial to another (Subjectivity) Also, a person can experience a mixture of emotions at any given moment (Complexity).
2. *Response*: When an emotion is felt by a person, there is always a response associated with it. It may be physiological, gestural or facially expressive.

Many theories have been put forward by psychology researchers differentiating the emotions and their nature. The following are some of them:

- Paul Ekman's perception of emotions classified them to distinct ones (fear, disgust, anger, surprise, happiness and sadness. He also mentioned that emotions can be even more extensive.

- Robert Plutchik proposed a psycho-evolutionary classification approach for general emotional responses. He drew the famous 'Wheel of Emotions' to explain his proposal in a graphic way, which consisted of the 8 basic bipolar emotions: joy (vs) sadness, trust (vs) disgust, anger (vs) fear, surprise (vs) anticipation.
- Parrot identified about a 100 emotions based on physiological response and conceptualized them as a tree-structured list in 2001. Parrot defined these primary emotions: love, joy, surprise, anger, sadness and fear.

2. Emotion Recognition & Valence Cognizance

Emotion Recognition is the process of identifying human emotions from both facial and verbal expressions. There are six basic emotions (**Dilbag Singh, 2012**) – happiness, surprise, anger, fear, disgust and sadness. Valence, as used in psychology, is the measure of the emotion expressed. The measurement is considered positive or negative based on the intrinsic attractiveness or averseness (of an event, object or situation) respectively. This measurement is called *Valence*. Depending on the emotion felt, this term is used to characterize and categorize specific emotions. If the emotions are referred to as negative, it is said to have negative valence and if it is positive, it is said to have positive valence.For example, the emotion 'joy' has positive valence and the emotion 'fear' has negative valence. Positively/Negatively valenced emotions are evoked by positively/negatively valenced actions, objects and circumstances.

Over the years, the researchers have found that human face (**Mukhopadhyay et. al, 2020**) is a good indicator of valence, especially with the corrugator and zygomatic activities, separating states such as pleasure/displeasure, joy/sadness, and like/dislike. While the space of the facial expressions is oceanic, the space formed by crossing all possible expressions, with all possible contexts and their demand to compute in real time is computationally intractable. No algorithms exist that describe how to precisely combine the many contributing channels into a full space of emotions. A different complex combination may need to be characterized for each emotion. Furthermore it is not sufficient to map the combination of signals emanating from the person whose emotion is being assessed and hence it is necessary to observe the context.

3. Proposed Model

The main objective is to develop a model that implements the data mining approach to detect emotions and their valence from the faces of subject's images from the FER 2013 Dataset by an Emotion Valence Cognizance Convolutional (**Keiron O'Shea and Ryan Nash ,2015**) Neural Network using the Double Average Filtering (**Kanchanadevi and P.R. Tamilselvi, 2020**) to set up the images for input to the CNN and the SELU activation (**Gunter et al, 2017**) to adjust the weights in the layers. The CNN acts as both the feature extractor as well as the classifier.

The FER 2013 dataset is used to detect emotion by recognition of facial expression. The .csv file pixels are converted into images and then the Double Average Filtering is used to preprocess the image. The images are resized and preprocessed to fit into the EVCNN model. Feature Extraction, MaxPooling and Padding are done to process the images. Since there are 7 classifications of emotions (as given in Table 3.1), there are 7 dense layers as the output layer. The final dense layer, is the SELU activation layer which is the novel idea used to increase the accuracy of the model. The model is best trained and saved as a .h5 file. Finally the SVM and the RBF Kernel is used to predict the Valence of the emotions (0 if the emotion is negative and 1 if the emotion is positive)

Value	Emotion	Valence
0	Anger	0
1	Disgust	0
2	Fear	0
3	Happiness	1
4	Sad	0
5	Surprise	1
6	Neutral	1

Table 3.1 : The Emotions , their Valence and their corresponding prediction values

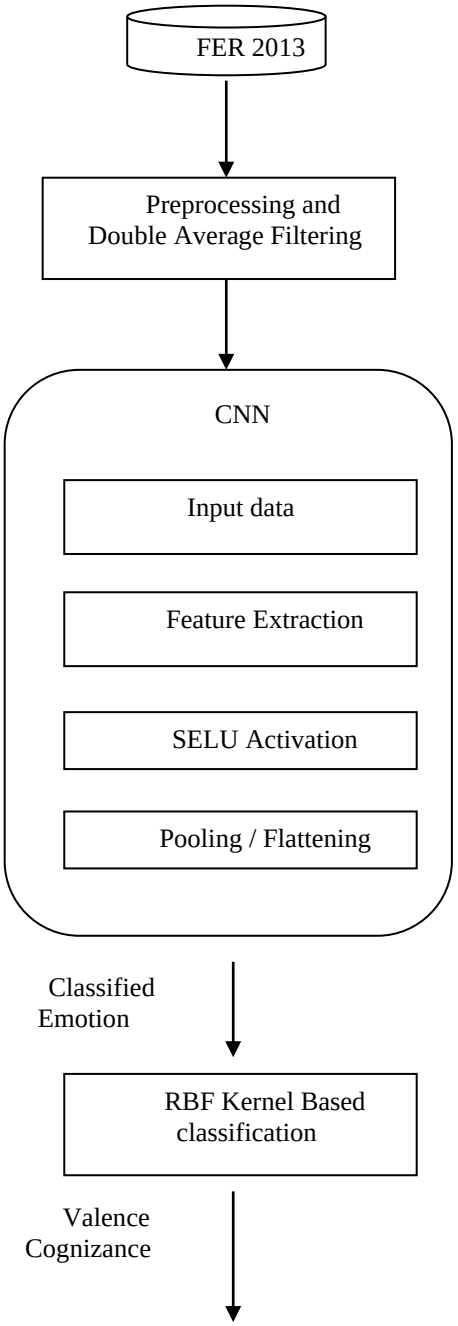


Fig.3.1: The Proposed Framework

4. **Double Average Filtering**

Filtering is a process where an image is modified or enhanced. Image processing operations implemented with filtering, include smoothing, sharpening, and edge enhancement. A special implementation of a low pass algorithm is the averaging filter. It calculates the output sample using the average from a finite number of input samples. It is helpful when there is a necessity to smooth data that carries high frequency distortion. In this project, the average filtering has been done twice and hence the name, Double Average Filtering. It was found to have an increased the quality of the input image by decreasing its noise.

5. The Emotion Valence Cognizance CNN

The Conv2D class constructor from Keras [8] has been used for the model and it includes the following parameters:

Filters: The first required Conv2D parameter is the number of filters the Convolutional layer will learn. These filters are primarily used for *Feature Extraction*. In Emotion Valence Cognizance CNN model, 1024 filters in the hidden layers for feature extraction and 4096 filters with 7 dense layers for classification, which obviously gives a good learning phase for the model. Max pooling is to be done at each step to reduce the spatial dimensions of the output volume. As the spatial volume is decreasing the number of filters used is increasing. The number of filters used with each layer can be accounted to the power of 2 as the values.
Kernel size: The second required parameter for conv2D is the Kernel size, specifying the width and height of the 2D convolution window. It is an odd integer – (3,3).(**Li et al, 2020**)
Padding: This parameter can be used in case either when the spatial volume of the output needs to be reduced (valid) or if the spatial volume of the output needs to be the same as the input volume (same)
In our case, *Max Pooling* is used to reduce the output volume. It is a pooling operation that calculates the maximum value for patches of a feature map, and uses it to create a down sampled (pooled) feature map. It is usually used after a Convolutional layer.

6. Activation Functions

Two activation functions have been used with this model: The ReLU and the SELU.
Rectified Linear Unit(ReLU): The ReLU is the most commonly used activation function.
The ReLU equation can be given by:

$$ReLU(x) = max(0,x)$$

The above equation tells that, if the input is less than 0, the input sets to 0. If the input is greater than 0 the input sets to the input itself, which means there can be no negative values while processing. Hence this activation is used in the hidden layers of the CNN model.

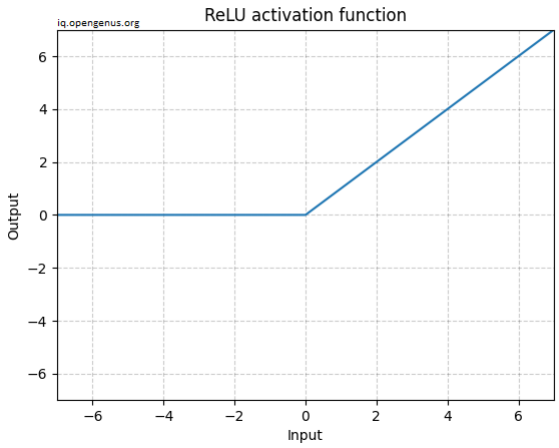


Fig.6.1.: Rectified Linear Unit (ReLU)

ReLU is used in many CNN models and helps facilitate the working of the same, because of its ability to rule out negative values. Though ReLU has its role in the model, there is a novel idea of an activation unit introduced in this model which is the SELU(**Flora Sakketou and Nicholas Ampazis ,2010**).
Scaled Exponential Linear Unit (SELU): This activation function is one of the newer functions. The best part about the SELU activation is that of internal normalization by the use of *LeCun_Normal* and AlphaDropout for weight initialiaiztion and application of dropouts respectively. The authors of the SELU activation have calculated two values : alpha (α) nad lambda (λ) value for the equation. The values are:

$\alpha \approx 1.673263242354377284817042991671$

$\lambda \approx 1.050700987355480493419334985294$

The equation for the SELU activation:

$$SELU(x) = \lambda \begin{cases} x & if\ x > 0 \\ ae^x - \alpha & if\ x \leq 0 \end{cases}$$

(i.e) if the input value is greater than zero, the output value becomes *x* multiplied by λ. If the input value is less than or equal to 0, there is a function that tends to 0, which is the output y, when x is zero. Essentially when x less than 0, the alpha is multiplied with the exponential of the x-value minus the alpha value, and then we multiply by the lambda value.
The most important aspect of SELU function is that it is *self-normalizing*. To be direct, first the mean is subtracted, and then divided by the standard deviation. So the components of the network (weights, biases and

activations) will have a mean of zero and a standard deviation of one after the normalization. This is the output of the SELU activation function.

There is an advantage of using the mean of zero and the standard deviation of one. Let us assume that the initialization function LeCun_Normal initializes the parameters of the network as a normal distribution. In the case of SELU, the network will be normalized entirely. Essentially, when multiplying or adding components of such a network, the network will be still considered as a normal distribution. In turn, the whole network and its output in the last layer is also normalized. The graph of a normal distribution with a mean of 0 and a standard deviation of 1.

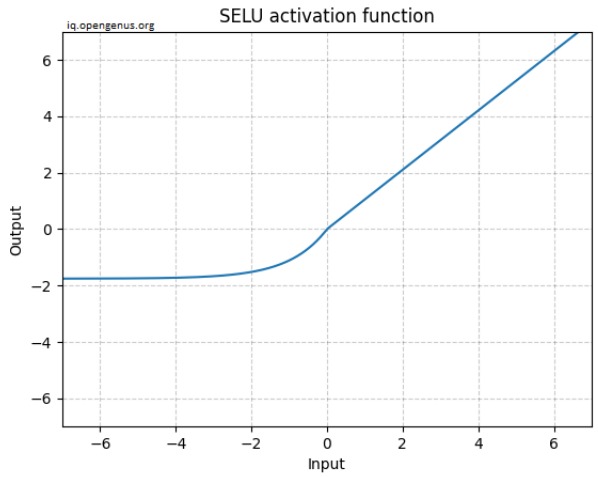


Fig. 6.2.:Scaled Exponential Linear Unit (SELU)

The output of the SELU is normalized, which could be called the internal normalization, hence the fact that all the outputs are with a mean of zero and a standard deviation of one. This is different from external normalization, where the batch normalization and other methods are used. The internal normalization actually happens with SELU because the variance decreases when the input is less than 0 and increases when the input is greater than 0. The standard deviation is the square root of variance, and hence the result is 1.

SELUs allow to construct a mapping g with properties that lead to Self Normalizing Neural Networks (SNN)(**Gunter et al, 2017**). SNN cannot be derived by another activation functions other than SELU, which has

1. negative and positive values for controlling the mean
2. saturation regions(derivatives approaching zero) to dampen the variance of its too large in the lower layer
3. a slope larger than 1 to increase the variance, if it is too small, in the lower layer
4. a continuous curve.

The differentiated function for the SELU activation is given as:

$$SELU'(x) = \lambda \begin{cases} 1 & if\ x > 0 \\ ae^x & if\ x \leq 0 \end{cases}$$

(i.e), if x>0, then the output will be y.But if x less than 0 then the alpha value is simply multiplied by the exponential operation on x.The SELU function is used as the activation unit in the output layer of the proposed model

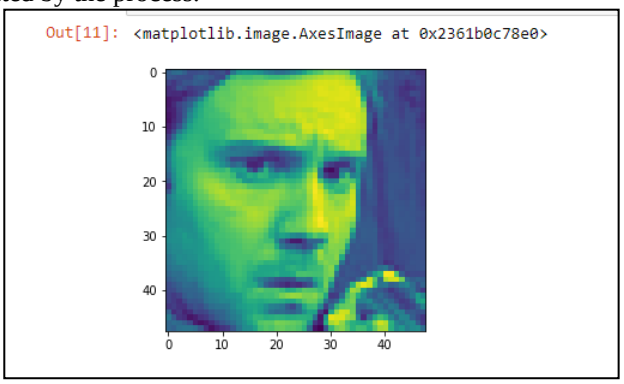
7. EVCNN - Valence Cognizance

The valence of the emotions are characterized as positive and negative. For this purpose the emotions are first labeled according to the nature of the emotions. The positive emotions are labeled 0 (inclusive of the neutral emotion) and the negative emotions are labeled as 1. The training data for the Valence Cognizance is given via a RBF Kernel in a SVM(**Zhang et al, 2020**), which classifies whether the emotion recognized is positive or negative. The haarcascade algorithm is implemented to detect the face. Since the output of the CV2 function is in BGR mode, it is necessary to convert it back to the RGB before using the haarcascade algorithm.

Once the face is detected, the image has to be preprocessed and fed into the newly designed EVCNN model for prediction. This gives as output an array which consists of the values of the seven emotions (Table 3.1). The array value turns 1 for the recognized emotion. The recognized emotion is compared with the labeled emotions and the valence is calculated.

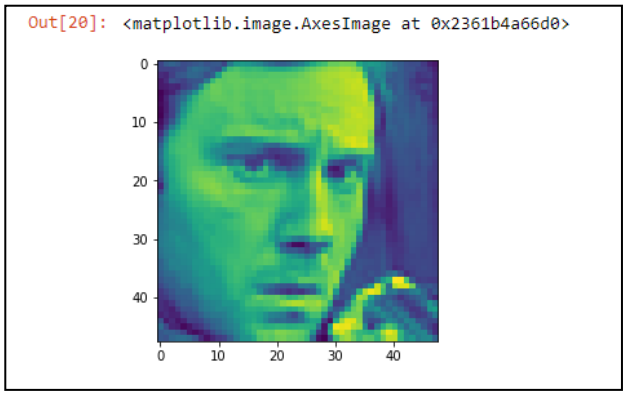
8. Results

1. The use of Double Average Filtering for noise reduction gave the following result, as opposed to not using the same. The following images are from the training set of the FER2013, for which Fig 8.1 has the Double Average Filtering applied while Fig 8.2 has not. Though the images seem to be similar, the values have been impacted by the process.



[70., 80., 82., ..., 52., 43., 41.]

Fig. 8.1. Before Double Average Filtering



[-132.43881, -117.93227, -111.67004...,-140.58823, -153.72404, -159.8687]

Fig. 8.2. After Double Average Filtering

2. The Prediction of the emotions was tested with the Tesing set of images of the FER 2013 dataset. Then a sample prediction was made for a random image, Fig 8.3 for which the predictions result. of the corresponding emotion is given by Fig 8.4:

Fig. 8.3 Sample Image

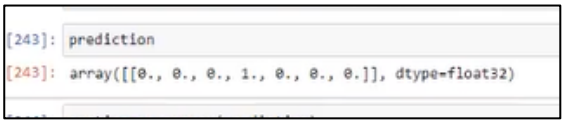


Fig. 8.4 Prediction Result

The array's fourth element has turned 1, for which the labeled emotion is *'happiness'*.

3. The valence of the recognized emotion (happiness) for the Fig 8.3, is calculated by first locating the face in the image, because the face is the primary source for emotion recognition as given in Fig 8.5.

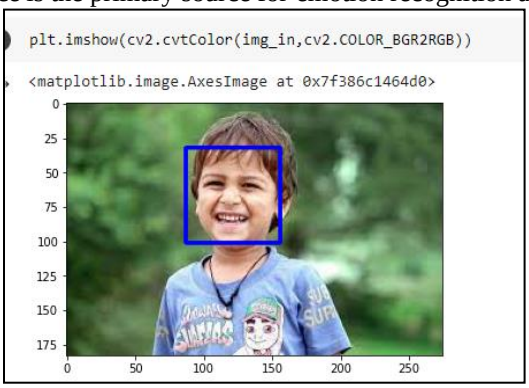


Fig 8.5 : Identifying the face

Then, the area of the bounding box is taken out, enlarged and converted into RGB to get the correct facial points to detect the emotion as in Fig 8.6:

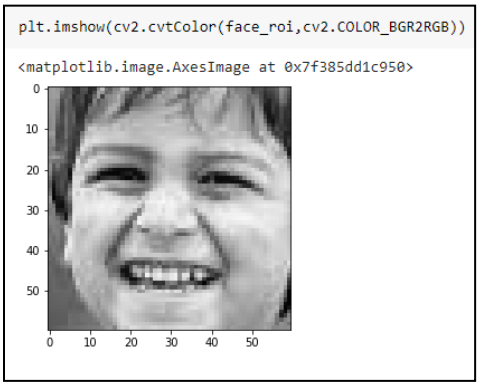


Fig 8.6 : The cropped face from the image

Using the code, and implementing the model, the emotion has been recognized and the valence calculated. Since happiness is a positive emotion, its valence has been calculated to be 1, as in Fig 8.7

Fig 8.7 : Valence Cognizance

4. This training phase gave an accuracy of about **96.53%** , though the number of epochs of training the whole dataset of about 35k images was only 50.
5. The testing phase gave an accuracy of **76.88 %,** which is lower than the training phase (As it usually is) but better than the previous findings for the same dataset.

Convolution+Residual+Attention [10]	64.4%
CNN [11]	65%
ENSEMBLE[12]	75.8%

VGG [13]	73.28%
VGG+GAP[14]	69.46%
CNN[15]	65.97%
CNN[16]	70.14%
CNN[17]	65%
Proposed model (EVC-CNN)	**76.88%**

Table 8.1: Testing phase accuracy of EVCNN

9.Conclusion

Though ReLU had to be used in the hidden layers, the SELU sets the standard for better accuracy due to its ability to self normalize. With SELU, internal normalization is faster than external normalization, which means the network converges faster. Vanishing and exploding gradient problem is impossible. It is a relatively new activation function and it needs to be explored more, which may be a disadvantage.

References

Li, J., Jin, K., Zhou, D., Kubota, N., & Ju, Z. (2020). Attention mechanism-based CNN for facial expression recognition. Neurocomputing, 411, 340-350.

Flora Sakketou(B) and Nicholas Ampazis(B) Department of Financial and Management Engineering, University of the Aegean, Chios, Greece , On the Invariance of the SELU Activation Function on Algorithm and Hyperparameter Selection in Neural Network Recommenders
DOI: 10.1007/978-3-030-19823-7_56

Zhang, J., Yin, Z., Chen, P., & Nichele, S. (2020). Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. Information Fusion, 59, 103-126.

Mukhopadhyay, M., Pal, S., Nayyar, A., Pramanik, P. K. D., Dasgupta, N., & Choudhury, P. (2020). Facial Emotion Detection to Assess Learner's State of Mind in an Online Learning System. Proceedings of the 2020 5th International Conference on Intelligent Information Technology. doi:10.1145/3385209.3385231

Keiron O'Shea 1 and Ryan Nash 2 1 Department of Computer Science, Aberystwyth University, Ceredigion, SY23 3DB keo7@aber.ac.uk 2 School of Computing and Communications, Lancaster University, Lancashire, An Introduction to Convolutional Neural Networks, arXiv:1511.08458v2 [cs.NE] 2 Dec 2015

Dilbag Singh, Gwangju Institute of Science and Technology · Electrical Engineering and Computer Science Concentration, Human Emotion Recognition System, DOI:10.5815/ijigsp.2012.08.07

Günter Klambauer Thomas Unterthiner Andreas Mayr Sepp Hochreiter(2017).Self-Normalizing Neural Networks LIT AI Lab & Institute of Bioinformatics, Johannes Kepler University Linz A-4040 Linz, Austria {klambauer,unterthiner,mayr,hochreit} @bioinf.jku.at

Khopkar, Apeksha & Adholiya, Ashish. (2021). Facial Expression Recognition Using CNN with Keras. Bioscience Biotechnology Research Communications. 14. 47-50. 10.21786/bbrc/14.5/10.

Kanchanadevi1 and P.R. Tamilselvi(2020). Preprocessing using image filtering method and techniques for medical image compression techniquesb,Department of Computer Science, Periyar University, India,Department of Computer Science, Government Arts and Science College, Komarapalayam, India

Gupta, A., Arunachalam, S., &Balakrishnan, R. (2020). Deep self-attention network for facial emotion recognition. Procedia Computer Science, 171, 1527-1534.

Mellouk, W., &Handouzi, W. (2020). Facial emotion recognition using deep learningfor review and insights. Procedia Computer Science, 175, 689-694.

Khanzada, A., Bai, C., &Celepcikay, F. T. (2020). Facial expression recognition with deep learning. arXiv preprint arXivfor2004.11823.

Khaireddin, Y., & Chen, Z. (2021). Facial emotion recognitionfor State of the art performance on FER2013. arXiv preprint arXivfor2105.03588.

Kusuma, G. P., Jonathan, A. P. L., & Lim, A. P. (2020). Emotion recognition on fer-2013 face images using fine-tuned vgg-16. Advances in Science, Technology and Engineering Systems Journal, 5(6), 315-322.

Zahara, L., Musa, P., Wibowo, E. P., Karim, I., & Musa, S. B. (2020, November). The facial emotion recognition (fer-2013) dataset for prediction system of micro-expressions face using the convolutional neural network (cnn) algorithm based raspberry pi. In 2020 Fifth International Conference on Informatics and Computing (ICIC) (pp. 1-9). IEEE.

Jaiswal, A., Raju, A. K., & Deb, S. (2020, June). Facial emotion detection using deep learning. In 2020 International Conference for Emerging Technology (INCET) (pp. 1-5). IEEE.

Agrawal, A., & Mittal, N. (2020). Using CNN for facial expression recognitionfor a study of the effects of kernel size and number of filters on accuracy. The Visual Computer, *36*(2), 405-412.