International Journal of IT-based Public Health Management
Vol.8, No.1 (2021), pp.9-18
http://dx.doi.org/10.21742/ijiphm.2021.8.1.02

Emotion Recognition using Facial Expression

Jay Naimesh Patel[1] and Jinan Fiaidhi[2*]

[1,2]Department of Computer Science, Lakehead University, Canada*
[2]jinan.fiaidhi@lakeheadu.ca*

Abstract

In recent years, research has shown an increased development of social networking applications. Social networking applications are recently getting wider interest among people of different ages. But due to irrelevant posts, the mood of a user can get affected. The project "Emotion Recognition Using Facial Expression" is based on a new concept where a person can filter friends' posts by emotions. The emotion would be detected from a facial expression. Using this application, a sad person's mood can be elevated as sad posts will be filtered out (as only happy posts will be seen). To develop the mobile application of this project, Android Studio was used. Python and TensorFlow were used to train the model. For data storage purposes, Firebase was used as it is compatible with any other platform.

Keywords: *Emotion recognition, Python, TensorFlow, Android studio*

1. Introduction

Numerous guardians are worried about what little children's openness to innovation can mean for their development. We realize that our preschoolers are rapidly learning new friendly and psychological capacities and we cannot let an addictive tendency of adhering to an iPad detriment this evolutionary advantage. However, pubescence is a crucial season of sped-up development, and all-around not many of us are focusing on what our youths' utilization of innovation means for them considerably clearer and more individual than a 3-year-old playing on father's iPhone. Specialists are worried about the pervasive influence of social media in juvenile life which is adding to tension and self-doubt in their lives [1].

Youngsters should accept there is caused to be concerned. The Royal Society for Public Health dispatched a survey of 14–24-year-olds in the United Kingdom to perceive what web-based media destinations meant for their wellbeing and prosperity. As indicated by the discoveries of the survey, Snapchat, Facebook, Twitter, and Instagram all added to more significant levels of despondency, nervousness, negative self-assurance, and detachment [2].

To avoid such problems of social media a new system or new technology should be used to make social media reliable to different age categories. It can be achieved by introducing neural networks in social media to extract features from the facial expression of humans to recognize emotions.

One or more movements or locations of the muscles under the surface of the human face represent a facial gesture. These muscle motions are used to express an individual's emotional state to different observers. Nonverbal contact requires facial gestures [2]. They are vital in day-to-day passionate correspondence just like how one speaks. They also portray inherent sentiments of an individual allowing him to reflect his inner state of mind. In a social setting,

Article history:
Received (May 5, 2021), Review Result (June 9, 2021), Accepted (August 30, 2021)

Print ISSN: 2205-8508, eISSN: 2207-3965 IJIPHM

when two individuals interact, their facial gestures speak for them. An exuberant individual's facial expression is strikingly different from that of a melancholic individual. As an outcome, the available data sets on exterior observable appearances are frequently employed in feeling acknowledgement programmable frameworks. Exploiting the most unveiled part of a human's anatomy 'The Face', PC vision frameworks (such as Cameras) can be employed to dissect the visage to reveal underlying emotions embodying an individual. Two key factors are influencing the nature of emotion recognition frameworks using cameras, namely Illumination environment and variances in the head position [3].

This idea of translation depends on the understanding that the presence of feelings is widespread across people groups just as human ethics, races, and societies. In an estimated type of outward appearance acknowledgment framework, an info detecting gadget, for example, a webcam or a fundamental camera acquired the information picture from a subject, and afterward, it communicates with the computer. After identification of the delegate highlights from the face district, the sincerely expressing facial picture is acquired, it is then pre-processed and a classification model is utilized to arrange them into one of the feelings such as Confusion, Disgust, Sadness, Fear, Surprise, Anger and Smiling (Happy). Many models were initially researched to achieve high accuracy but the best from that research for a classification problem Convolution Neural Network was found as the best suit for those kinds of problems.

The human emotion detection process involves the following steps (i) Pre-processing of facial expressions/gestures, (ii) Segmentation into various categories, (iii) Extraction of key features, and (iv) Classification using an algorithm. In this article, a Convolution Neural Network is chosen to train the system with various types of expressions and, as a result, to identify emotions using facial expressions. The videos are de-noised using a high boost filtering technique [4]. The motive of this research study is to extract and classify facial expressions of an individual into seven major states of emotions: Sadness, Anger, Disgust, Joy, Fear, Neutral, and Surprise. By recognizing the emotions of the human being, it will then show the relevant posts related to the emotions.

Figure 1. The figure depicts different facial expression and their emotions on the same person

2. Literature review

Emotions play a major part in a person's lifespan. An emotion's significance is mirrored in each facial expression. When opposed to the facial expression of older people, the facial expression of babies has a distinct sense. Every person's emotion should be treated with caution since they can contribute to a variety of situations. Human facial expressions can be closely observed during company sales to anticipate a customer's emotions. Criminals' facial expressions are critical in investigations and deciding whether the person in detention is providing truthful or false information [5].

The arrangement of highlights that are considered for ex- traction and the classifier that is utilized for the assignment of grouping are both similarly essential to decide the exhibition of an outward appearance acknowledgment framework. For an inadequately chosen set of highlights, now and again, even a decent grouping calculation can't give an ideal outcome. Along these lines, choosing better highlights has consistently stayed a pre-imperative for high characterization precision and great outcome [2].

The feature extraction model can be employed in a legal setting for an examination of lawbreakers and to ascertain the veracity of their data based on their associated facial expressions. For example, the melancholic demeanor of lawbreakers and criminals can effectively be utilized to ascertain their self-destructive inclination, thus pre- venting any untoward act/incident. It has gotten fundamental to robotize the way toward recognizing the feelings of people by utilizing outward appearances. The highlights extraction strategy assumes a significant part during the time spent mechanization of distinguishing feelings utilizing outward appearances [4]. To separate the highlights appropriately, the picture must be fragmented well. It is mandatory to include neural organizations for preparing and testing [3].

Individuals now and then watch out for express an inclination that is in their psyche yet can't generally communicate through actual articulations similarly as others do which in outcome leads to blended and complex passionate states which makes it hard for an articulation acknowledgment framework to perceive.

The features extraction technique is critical in automating the process of sensing emotions through facial expressions. The picture must be correctly segmented to remove the features. To train and test the classifier correctly convolution neural network is used. The emotion detection convolution classifier architecture can be seen from [Figure 2].

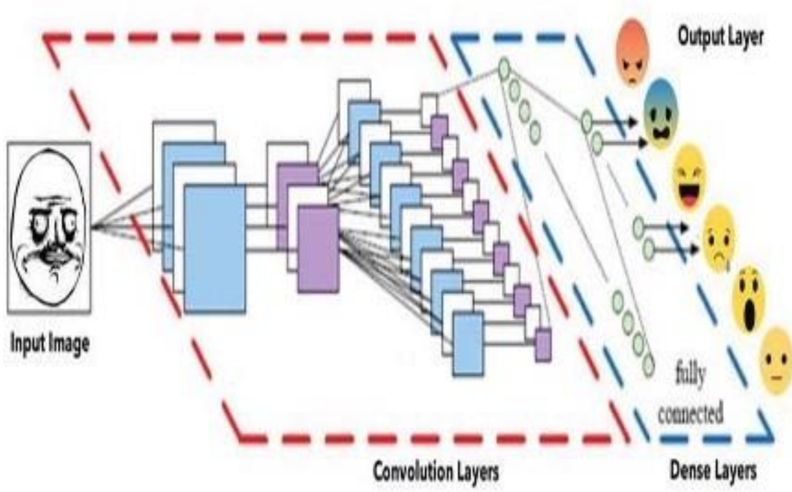

Figure 2. Convolution neural network architecture

2.1. Emotion taxonomy

Emotion researchers and psychologists have developed a variety of theories for categorizing feelings, ranging from commonly expressed simple emotions to culturally diverse nuanced emotions. Two models have dominated facial expression analysis among the different models of emotion research: Ekman's simple collection of emotions [5] and Russell's circumplex model of affect [6]. In 1971 [5], Ekman and Freisen suggested six prototypical fundamental emotions: anger, disgust, fear, happiness, sadness, and surprise, all of which are uniformly expressed and remembered.

The universality of these essential feelings was further confirmed by cross-cultural research in [7], which had its origins in Charles Darwin’s universality thesis. This categorical definition has gained prominence, and it benefits from the fact that humans can easily recognize and identify facial expressions synonymous with basic emotions.

The outward appearances related to these particular feelings have directed the investigations to identify the outward appearance acknowledging through the recent forty years, and this kind of feeling subspace has been the most famous model for figuring emotion. Russel [8] recommended another order worldview of human feeling in which enthusiastic states are portrayed by circles in two-dimensional bipolar space (agreeableness obnoxiousness, excitement rest) as opposed to discrete gatherings. Rage, for example, maybe viewed as a symbol of intense displeasure and relatively high arousal. The Circumflex Model of Russell [6] is depicted in [Figure 3.]

Figure 3. The circumflex model of Russell

3. Methodology

(1) Eco-system

The objective of this system is to develop a social networking mobile application similar to Instagram but with the addition of an emotion recognition feature. The new feature aims at recognizing the 7 types of emotions of the person using the mobile camera. The system will predict the type of emotion depicted from the facial expression of a person by employing the machine learning technique. As per the emotion recognized by the system, the person will be shown mood-elevating posts. For an instance, if a person is in a sad mood, then only happy posts can be seen as the main aim of the project is to show posts relevant to the emotions of any person. Firstly, the model would be trained using Python and TensorFlow, and then the trained model would be deployed in the android application by converting it model into TensorFlow lite (for emotion recognition). The Firebase would be used as cloud storage as it was easy to connect with any platform. Moreover, the prediction would take place in the firebase custom model as it is easy to start prediction in the cloud rather than doing it on a local device i.e., without GPU. [Figure 4] shows the flow diagram of the system.

(2) Dataset

The dataset consists of the gray-scale image of faces with a resolution of 48x48 pixels. The faces have been consequently registered so that they are appropriately focused in each image and take up about the same amount of space. The aim is to categorize each face into one of seven groups based on the emotion shown in the facial expression (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

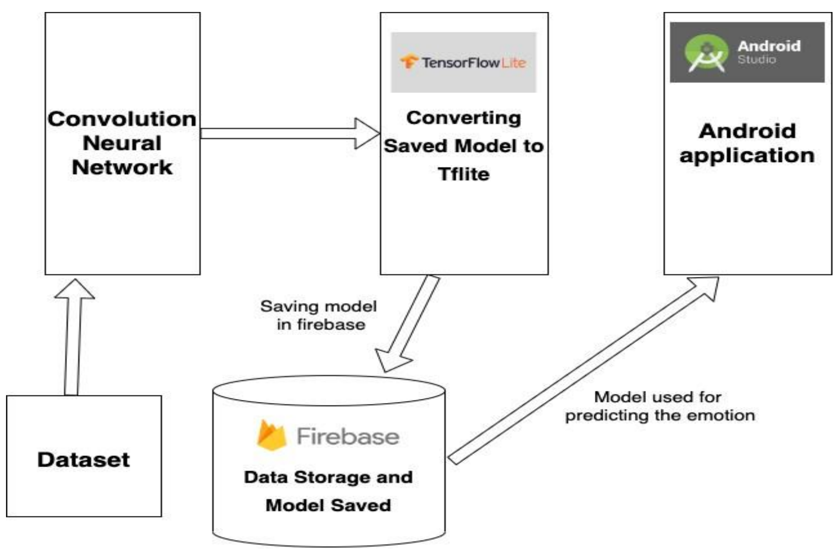

Figure 4. Flow diagram of emotion recognition using facial expression

“Emotion” and “pixels” are two columns in the train.csv file. The 1st column i.e., emotion column has a numeric code for the emotion present in the picture, in a range of 0 to 6, inclusive. For each image, the “pixels" column contains a series of characters enclosed in quotes. The contents of this series of characters are pixel values in row main sequence, separated by spaces. The only column in test.csv is “pixels,” and the job is to estimate the emotion column.

	emotion	pixels	usage
0	emotion	pixels	Usage
1	0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121...	Training
2	0	151 150 147 155 148 133 111 140 170 174 182 15...	Training
3	2	231 212 156 164 174 138 161 173 182 200 106 38...	Training
4	4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1...	Training
5	6	4 0 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84...	Training
6	2	55 55 55 55 55 54 60 68 54 85 151 163 170 179 ...	Training
7	4	20 17 19 21 25 38 42 42 46 54 56 62 63 66 82 1...	Training
8	3	77 78 79 79 78 75 60 55 47 48 58 73 77 79 57 5...	Training
9	3	85 84 90 121 101 102 133 153 153 169 177 189 1...	Training

Figure 5. Details of “train.csv" file from a dataset

(3) Libraries used

a. Keras: It is an efficient and powerful open-source package to develop and evaluate deep learning models.

b. Sklearn: It is a useful library for machine learning. It consists of many efficient tools for statistical modeling and machine learning such as classification, regression, clustering, and dimensionality reduction.

c. Pandas: To manipulate the numerical data from the dataset. The method of translating data into something which a machine may comprehend is called pre-processing. Preprocessing steps needs to be carried out to overcome the fitting issue. Mainly, the preprocessing steps involve tokenization, stemming, vectorization. But before proceeding further, the data from the dataset needs to be filtered. The dataset comprises sentences or maybe just a single word as a part of a movie review. So, only the full sentence needs to be taken for our training and testing purposes. The dimensions of the dataset are also reduced.

(4) Preprocessing steps

Preprocessing steps involve dropping the unnecessary columns using Pandas library so that false training and prediction of the model are avoided. To train a model the pixels need to be split with their respective label. X and Y NumPy array would return to train the model. Moreover, X and Y are split into train and test to ready the data for training the model.

The grayscale facial images are achieved by converting the NumPy array that consists of pixels to an image and plotting that. The images can be seen in [Figure 6]

(5) Proposed model

Firstly, the dataset is imported. After importing, the data needs to be pre-processed before passing it into a convolution neural network where it is classified. Once the data is preprocessed, features are extracted from the data and the model is trained. Now, the trained model is used to classify the image that is imported into the system. As it is a multi-class classification problem the output i.e., the classification of the test image would be from 0-6 classes. The flow diagram of the proposed model can be seen in [Figure 7].

Figure 6. NumPy array consisting of pixels converted to an image with the appropriate label

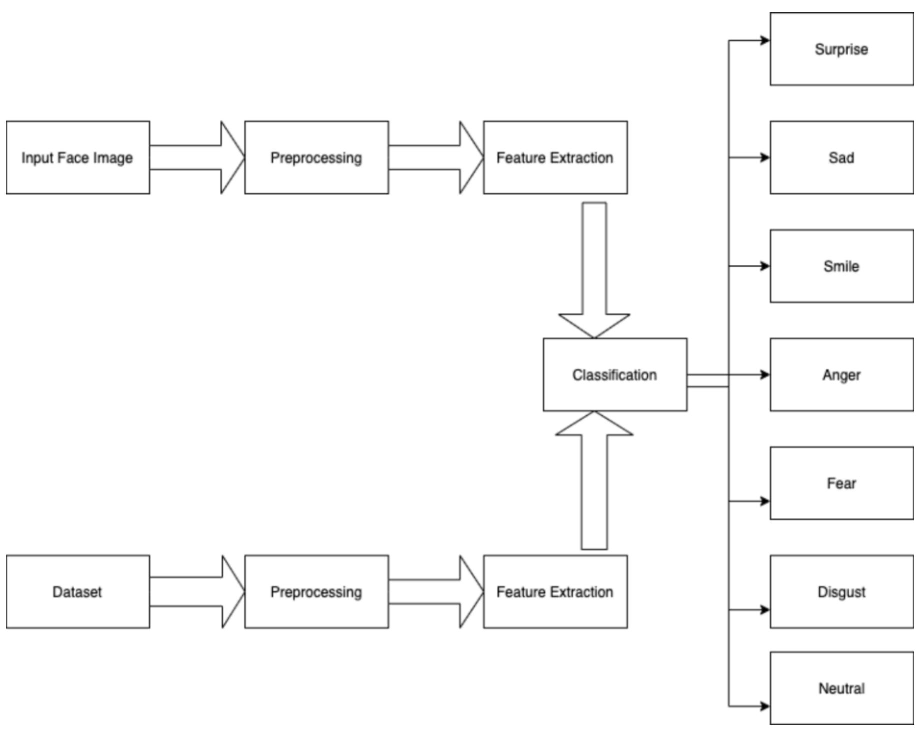

Figure 7. Flow diagram

A convolution neural network model is defined to train and test the movie review dataset. Keras is used to create a 2D convolution neural network. Before making the model, the data needs to be converted from raw data to NumPy array. The input shape is changed so that it can fit in our proposed model. The convolution layer, Maxpooling layer, flatten layer, batch-normalization are used to process the input. A batch size of 64 is used to train the model. The best optimizer found from experimental analysis such as Adam is used, and the suitable learning rate is chosen to have the appropriate accuracy for our model. The layers can be seen in [Figure 8].

4. Results experimental analysis and results

To get the accurate result for the proposed model, I used the best-suited hyper-parameters and optimizers and derived the graph for training validation accuracy and training validation loss. The hyper-parameter is shown in [Table 1].

Table 1. Hyper-parameter

Hyper-Parameter	Values
Epoch	20
Optimizer	Adam
Batch Size	64
Number of Layers	19
Kernel Size	5
Learning Rate	le-3

The training validation accuracy and training-validation loss graph are shown in [Figure 9] and [Figure 10] respectively. The model performed well as discussed in the proposed system.

```
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 48, 48, 64)        1664
conv2d_1 (Conv2D)            (None, 48, 48, 64)        102464
batch_normalization (BatchNo (None, 48, 48, 64)        256
max_pooling2d (MaxPooling2D) (None, 24, 24, 64)        0
conv2d_2 (Conv2D)            (None, 24, 24, 128)       204928
conv2d_3 (Conv2D)            (None, 24, 24, 128)       409728
batch_normalization_1 (Batch (None, 24, 24, 128)       512
max_pooling2d_1 (MaxPooling2 (None, 12, 12, 128)       0
conv2d_4 (Conv2D)            (None, 12, 12, 256)       295168
conv2d_5 (Conv2D)            (None, 12, 12, 256)       590080
batch_normalization_2 (Batch (None, 12, 12, 256)       1024
max_pooling2d_2 (MaxPooling2 (None, 6, 6, 256)         0
flatten (Flatten)            (None, 9216)              0
dense (Dense)                (None, 128)               1179776
batch_normalization_3 (Batch (None, 128)               512
activation (Activation)      (None, 128)               0
dropout (Dropout)            (None, 128)               0
dense_1 (Dense)              (None, 7)                 903
activation_1 (Activation)    (None, 7)                 0
=================================================================
Total params: 2,787,015
Trainable params: 2,785,863
Non-trainable params: 1,152
```

Figure 8. CNN model layers

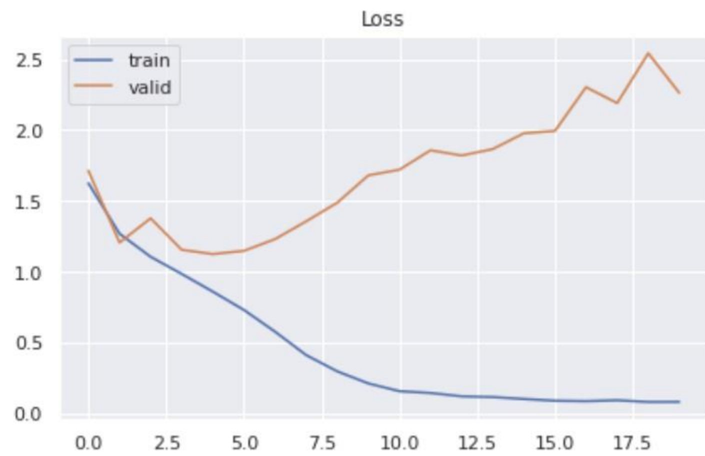

Figure 9. Training-validation accuracy

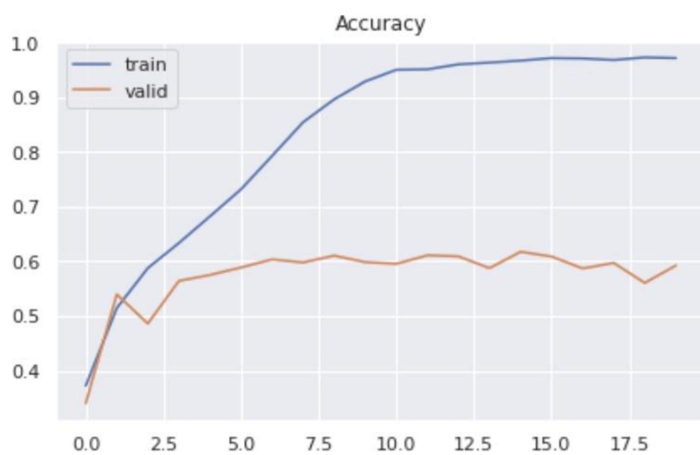

Figure 10. Training-validation loss

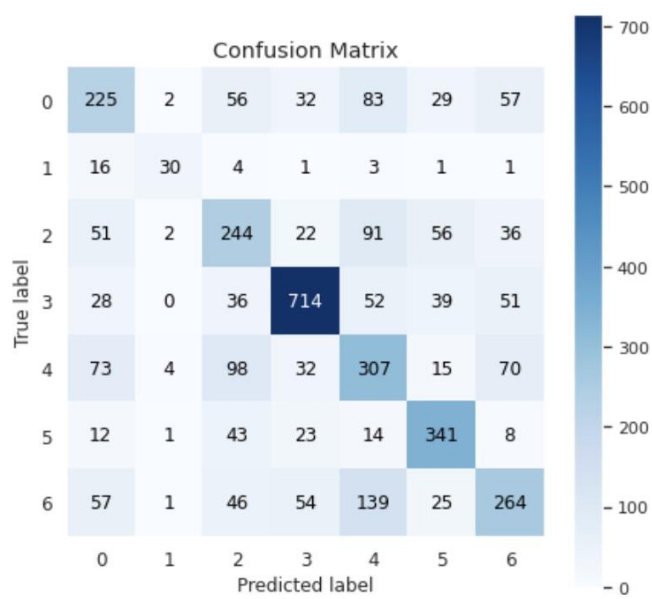

Figure 11. Confusion matrix

Moreover, the confusion matrix was formed to determine the performance of the trained classification model on the testing data for which the exact values or true values are known. The confusion matrix of the performed classification model can be seen in [Figure 11].

The model performed well and meet the expectation and achieved the training accuracy of 97%.5.

5. Conclusion

Facial expression analysis for emotion recognition is a difficult task in processing an image and human-computer interaction. During the tenure of the past two decades, extensive research has been undertaken in this area, and in recent years, it has attracted a lot of popularity by its variety of applications and implementations in many of the domains. This paper shows one of the best applications where we can use emotion recognition using facial recognition to avoid mood change in any age category while using the social media platform. The proposed system can be useful for further research to improvise the current social media platforms for the betterment of users.

Acknowledgment

We would like to acknowledge that the efforts made by Professor Dr. Jinan Fiaidhi were remarkable. Not only she did introduce us to the right path but also encouraged us in the overall completion of the project with ease. We are thankful for her efforts and piece of advice for the betterment of the project.

REFERENCES

[1] https://childmind.org/article/how-using-social-media-affects-teenagers/, Rachel Ehmke

[2] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, "Emotion recognition using facial expressions," Procedia Computer Science, vol.108, pp.1175-1184, **(2017)**

[3] B. Y. L. Li, A. S. Mian, W. Liu, and A. Krishna, "Using Kinect for face recognition under varying poses, expressions, illumination, and disguise," In 2013 IEEE workshop on applications of computer vision (WACV), IEEE, pp.186-192, **(2013)**

[4] K. Suresh Babu and S. Vemuru, "A low-cost software-defined radio-based cognitive radio test-bed for LTE networks," International Journal of Engineering and Technology (UAE), vol.7, (3.1 Special Issue 1), pp.51-55, **(2018)**

[5] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," Journal of Personality and Social Psychology, vol.17, no.2, pp.124, **(1971)**

[6] J. A. Russell, "A circumplex model of affect," Journal of Personality and Social Psychology, vol.39, no.6, pp.1161, **(1980)**

[7] P. Ekman, "Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique," pp.268, **(1994)**

[8] T. Matlovic, P. Gaspar, R. Moro, J. Simko, and M. Bielikova, "Emotion's detection using facial expressions recognition and EE," In 2016 11th international workshop on semantic and social media adaptation and personalization (SMAP), IEEE, pp.18-23, **(2016)**

[9] D. Matsumoto and H. S. Hwang, "Facial expressions," **(2013)**

[10] J. Jaworek-korjakowska, and P. Kleczek, "Eskin: Study on the smartphone application for early detection of malignant melanoma," Wireless Communications and Mobile Computing, **(2018)**, DOI: 10.1155/2018/5767360