EAI Endorsed Transactions
on Context-aware Systems and Applications
Research Article EAI.EU

Facial Sentiment Recognition using artificial intelligence techniques

Vuong Xuan Chi[1*], Phan Cong Vinh[1*]

[1]Faculty of Information Technology, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam.

Abstract

Facial emotion recognition technology is used to analyze and recognize human emotions based on facial expressions. This technology uses deep learning models to classify facial expressions, eyes, eyebrows, mouth, and other facial expressions to determine a person's emotions. The application of facial emotion recognition in the field of education is a potential way to evaluate the level of student absorption after each class period. Using cameras and emotion recognition technology, the system can record and analyze students' facial expressions during class. In this paper, we use the Convolutional Neural Network (CNN) algorithm combined with the linear regression analysis method to build a model to predict students' facial emotions over a period of time camera recorded.

Keywords: Facial Sentiment Recognition, Convolutional artificial neural network, Linear regression, Satisfied prediction

Received on 02 September 2023, accepted on 21 September 2023, published on 22 September 2023

doi: 10.4108/eetcasa.v9i1.3930

* Corresponding author. Email: vxchi@ntt.edu.vn, pcvinh@ntt.edu.vn

1. Introduction

In recent years, many studies related to facial recognition and analysis have been widely applied. Through Deep Learning models, many works have produced prediction results with the state of the high performance. In addition, hardware systems have also integrated microchips that support human face recognition through the camera on that device. The application of Deep Learning for emotional analysis is also being built and developed to control and predict human behavior through faces [2].

Facial emotion recognition is one of the developments in facial image recognition, however, many definitions are not really clear. With the division of seven types of emotions as Happiness, Satisfaction, Surprise, Fear, Sadness, Anger, and Indignation, Matsumoto [1] proposed, however, Mase and Pentland [3] said, there are 4 types of emotions 'Happy', 'Sad', 'Angry' and 'Surprise' expressed more clearly; Other types of emotions are often ambiguous and highly dependent on the observer's experience (i.e., cannot be precisely quantified). The Radboud Faces Database divides facial emotions into 8 categories: 'Happy', 'Sad' 'Fear', 'Angry', 'Surprise', 'Disgust', 'Neutral' and 'Scorn'. The Kaggle FER-F2013 dataset [4] only has 7 types of emotions: ‘Happy’, ‘Sad’ ‘Fear’, ‘Angry’, ‘Surprise’, ‘Disgust’, and ‘Neutral’. Facial recognition applications are mainly applied to detect strangers entering illegally through smart cameras integrated into existing systems. Some banks have also applied facial recognition when making transactions at banks. withdrawal point. Many universities have also tested and built facial recognition systems, but they are only at the testing level and have not yet been deployed. In addition, most current timekeeping and attendance systems are based on fingerprint timekeeping machines.

Using facial emotion recognition technology in education can bring a number of benefits: instead of having to rely on student feedback through surveys or interviews, this technology allows assessment Automatically assess student satisfaction and absorption levels, it helps schools save time and effort; Can provide immediate feedback after each class, helping lecturers and schools immediately identify problems and adjust teaching methods to increase student absorption.

2. Deep learning modeling approach

2.1 Facial Sentiment Recognition based on Deep Learning model

Deep learning (DL) is a subset of machine learning methods. DL focuses on building computer models that use "deep" architectures (visualized as having many layers) to learn complex representations of data and perform complex tasks based on data with a lower level of abstraction, by data layering and nonlinear transformations [5]. Some popular deep learning algorithms include ANN, which consists of many layers of neurons (computational units) connected to each other. These neural networks are capable of learning hidden features and mapping inputs to desired outputs. In addition to CNN architecture, deep learning network models have many other architectural forms such as fully connected feedforward layers, RNN, LSTM, GRU, DBN... [6][8][10] a simple DL network architecture with 3 layers:

Input layer: This layer receives input data and transmits it over the network. In the face recognition problem, the input layer can be the pixels of an image containing a face.

Hidden layer: This layer is the core of the network and performs calculations to learn complex features from input data. Deep learning architectures often have many consecutive hidden layers to learn high level data's representations.

Output layer: This layer creates the output of the model after going through hidden layers. In the face recognition problem, the output layer can predict the location and identity of faces in the image.

Depending on the type of model and specific task, there may be additional components in the deep learning architecture, i.e., recurrent neural networks, skip connections, dimensionality reduction layers, and many other types of layers to enhance model performance in each specific problem.

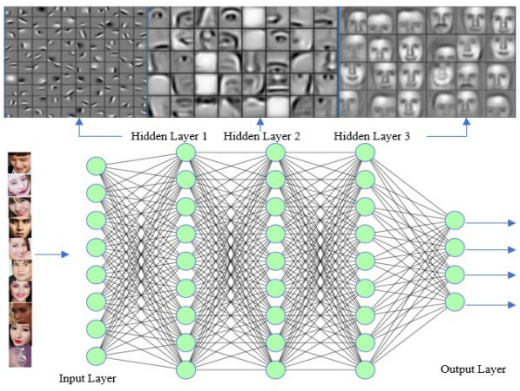

Figure 1. Deep learning model in face recognition

A typical deep learning model [7] is used in human face recognition, in which the input data of the network can be data in the raw form of RGB pixels (even without preprocessing). The features are combined and formed into small details in the first hidden layer, then continue to be reconstructed and combined with large details in the second hidden layer, and finally, the feature images of the whole face are in the third hidden layer. The output layer gives an assessment of the probability of which class (person) the face belongs to, as shown in Figure 1.

2.2. Deep learning application for facial emotion recognition

There are many types of deep learning algorithms applied to the problem of facial emotion detection [9]. Some popular types of deep learning models are as follows:

- CNN is one of the most popular types of deep learning models in image processing and computer vision. CNN networks are capable of learning complex features from image data and can automatically extract important features related to emotions.
- RNN is a type of deep learning model often used in processing time series or linked data. RNN can be used to process image sequences containing facial expressions in videos.
- Long Short-Term Memory (LSTM): a variant of RNN, designed to handle the problem of disappearing information in long data strings.
- Multimodal Embedding Neural Network: This is a type of model that combines data from many different information sources, such as facial images and voice sounds, to detect facial emotions comprehensively. faceted and multi-dimensional.
- Transfer Learning Neural Network: a technique that reuses a model previously trained on a similar task such as face detection or emotion classification, then fine-tunes the model for the specific problem.

The first dataset for this problem is CK+ with only 593 image series, the MMI dataset also only has 740 images and 2900 videos. Some recently appeared datasets have a larger number of samples such as EmotionNet [11] with 1 million samples or AffectNet [12] with 450 thousand samples. The datasets also vary in the number and way of classifying emotions, as well as in calculating the performance of the classification methods.

Widely applied, but the problem of detecting facial emotions is still a big challenge, and the accuracy of current systems is still quite low. According to the CNN model of Liu et al, for the new MMI dataset, it is about 78.5% [13]; According to Vielzeuf et al, the VGG16-LSTM model for the new AffectNet dataset achieved 48.6% [14].

3. Proposed evaluation method

3.1 Proposed evaluation of emotional state recognition

Linear regression is widely used in many different fields to determine the linear relationship between a dependent variable and one or more independent variables. In this article, the author uses a linear regression equation to predict the emotion rate after identifying the facial emotions of a number of students, and from there, evaluate the emotions through each state. draw conclusions about whether students are satisfied or not when participating in a class. Once you have finished implementing CNN, the next steps will be:

Step 1. Split the video into small segments (T), for example, 60 seconds.
Step 2. Randomly take 1 image/1 second in the divided video T, and calculate the percentage of each emotion.
Step 3. Implement a linear regression model to calculate the accuracy and loss coefficient of each emotion at a specific time.
Step 4. Calculate the total percentages (accuracy) of each emotion of nT
Step 5. Compare the accuracy of the Sum of Proportions of each emotion.

3.2 Linear regression equation

The goal of all supervised learning models in machine learning is to find a prediction function that minimizes the error value compared to the ground truth. The ground truth here is the value of the target variable y. This error is measured through loss functions. Training a machine learning model essentially boils down to finding the extreme value of the loss function [15] [16].

In the forecasting problem, use the MSE (Mean Square Error) function as the loss function. This function has a value equal to the average of the sum of squared errors between the predicted value and the ground truth. Suppose we consider a univariate regression equation including observations whose dependent variable is $\mathbf{y} = \{\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^n\}$ and input variable $\mathbf{x} = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^n\}$. Vector, $\mathrm{w} = (\mathrm{w}^0, \mathrm{w}^1)$ with w^0, w^1 are the slope coefficient and the estimated coefficient, respectively.
The univariate linear regression equation:

$$\hat{y}_i = f(x_i) = w_0 + w_1 * x_i$$

where (xi, yi) is the ith data point.
The goal is to find a vector that minimizes the error between the predicted and actual values. That is, minimizing the loss function is the MSE function:

$$\mathcal{L}(\mathbf{w}) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$= \frac{1}{2n}\sum_{i=1}^{n}(y_i - w_0 - w_1 * x_i)^2$$

where $\mathcal{L}(\mathrm{w})$ represents a loss function of w under the condition that we know the input is vector x and the dependent variable vector y. We can find the extreme value of the equation based on the derivative with respect to w0 and w1 as follows:

- Derivative with respect to w0: (1)

$$\frac{\delta\mathcal{L}(\mathbf{w})}{\delta w_0} = \frac{-1}{n}\sum_{i=1}^{n}(y_i - w_0 - w_1 * x_i)$$

$$= \frac{-1}{n}\sum_{i=1}^{n} y_i + w_0 + w_1\frac{1}{n}\sum_{i=1}^{n} x_i$$

$$= -\bar{\mathbf{y}} + w_0 + w_1\bar{\mathbf{x}} = 0$$

- Derivative with respect to w_1 : (2)

$$\frac{\delta\mathcal{L}(\mathbf{w})}{\delta w_1} = \frac{-1}{n}\sum_{i=1}^{n} x_i(y_i - w_0 - w_1 * x_i)$$

$$= \frac{-1}{n}\sum_{i=1}^{n} x_i y_i + w_0\frac{1}{n}\sum_{i=1}^{n} x_i + w_1\frac{1}{n}\sum_{i=1}^{n} x_i^2$$

$$= -\mathbf{xy} + w_0\bar{\mathbf{x}} + w_1\overline{\mathbf{x}^2} = 0$$

- Equation (1), we present:

$$w_0 = \bar{\mathbf{y}} - w_1\bar{\mathbf{x}}.$$

Substituting into equation (2) we can calculate:

$$\begin{aligned} -\overline{\mathbf{xy}} + w_0\bar{\mathbf{x}} + w_1\overline{\mathbf{x}^2} &= -\overline{\mathbf{xy}} + (\bar{\mathbf{y}} - w_1\bar{\mathbf{x}})\bar{\mathbf{x}} + w_1\overline{\mathbf{x}^2} \\ &= -\overline{\mathbf{xy}} + \bar{\mathbf{y}}\bar{\mathbf{x}} - w_1\bar{\mathbf{x}}^2 + w_1\overline{\mathbf{x}^2} \\ &= 0 \end{aligned}$$

Thence inferred:
Substitute in to calculate:

$$w_0 = \bar{\mathbf{y}} - w_1\bar{\mathbf{x}}.$$

4. Results and discussions

4.1 Dataset

The FER2013 dataset [4][18] is a widely used dataset in the field of facial emotion recognition (Facial Expression Recognition). This is a dataset of human face images with corresponding emotion labels. FER2013 includes images of faces collected, under different conditions with different facial expressions. The data totals about 35,887 facial images. All images in this data set are 48x48 pixels in size and are divided into 3 parts: a training set of 28,709 images, a public test set of 3,589 images, and a public test set of 3,589 images. The private test set

includes 3,589 images. Figure 2 shows the Model classes in the data set.

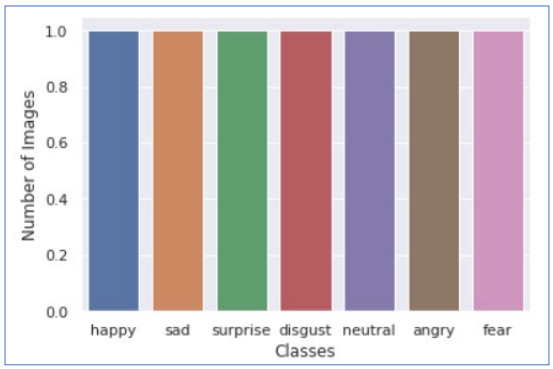

Figure 2. Labels in the training dataset

Each image in the dataset is labeled with one of the following emotions: 'Happy', 'Sad' 'Fear', 'Angry', 'Surprise', 'Disgust', and 'Neutral' as shown in Figure 3 and the number of images. Images of the emotions of the training set and test set are shown in Figure 4.

Figure 3. Visualize Images before Assigning Label

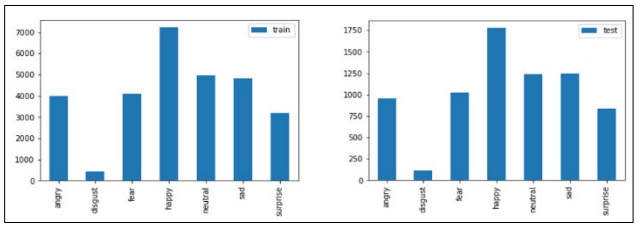

Figure 4. Visualize the number of images from train and test sets

4.2 Facial Emotion Recognition based to CNN

First, normalize the pixel value to [0,1] and convert the image to standard size, 48x48 pixels. Continue to split the dataset into a training dataset and a test dataset. Then, build the CNN model with main layers such as the convolution layer, ReLU activation layer, pooling layer, fully connected layer, and Softmax output layer. Choose the number of convolution layers, number of neurons, and hyperparameters appropriate to the specific problem. Then, train the CNN model on the training set. During training, the model will learn features from image data and predict the corresponding emotions, Max-pooling layer is used after every two convolutional layers [17], [19], [20]. Figure 5 is a diagram depicting facial emotion recognition from the dataset.

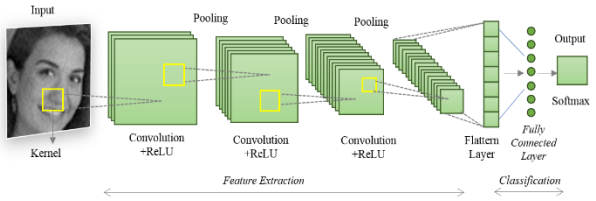

Figure 5. CNN model for Facial Emotion Recognition

4.3 Accuracy and loss function

Test and evaluate the model, using the test set to evaluate the model's performance. Evaluation can be done using metrics such as accuracy, Loss function, confusion matrix, and other metrics related to the accuracy of emotion classification. In experiments, the accuracy and loss function are shown to be used together to evaluate and monitor the performance of CNN. Figure 6 shows two graphs evaluating accuracy and loss when training and testing the model.

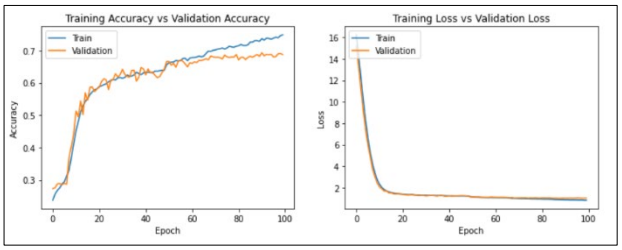

Figure 6. Evaluating model performance via Accuracy and Loss

Input the image extracted from the camera into the model and get the corresponding emotion prediction. After running the test model from the images of the video obtained, the results based on the training set predict the student's emotional state. Figure 7 below is the three emotions Neutral, Happy, and Angry predicted from a student in the classroom:

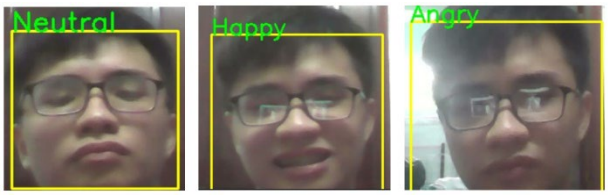

Figure 7. Prediction Result

4.4 Results of facial status analysis on a period of time

Analyzing and synthesizing facial states over a period of time, the author uses linear regression to create a prediction model based on the percentage of occurrence of the 7 'Happy' facial emotional states, 'Sad' 'Fear', 'Angry', 'Surprise', 'Disgust', and 'Neutral' were obtained from the CNN method.

In reality, if camera analysis extracts an image, we cannot evaluate the emotional state over a long period of time, for example the entire course of a student's class. In this experiment, the author took a short period of time to analyze the ratio of facial emotional states.

Suppose in 1 second, we randomly take 1 image, testing with a time of 60 seconds (T_1), so we have 60 images on T_1. During this time period, emotions are calculated as a percentage of occurrence for each second over 60 seconds. These emotional states are calculated from the extracted images. The first is Happy, which has the highest occurrence rate of ~90%, followed by Neutral ~9%, the third is Sad ~1%, the remaining states have a rate of 0. Figure 8 visualizes the percentage level average of HAPPY status.

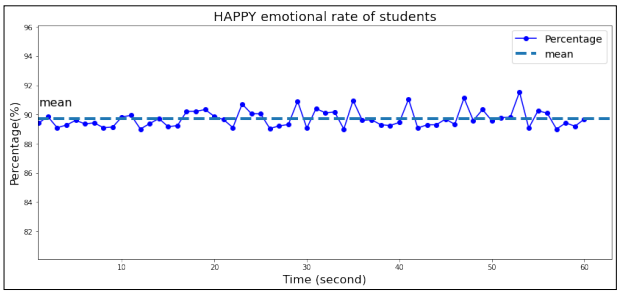

Figure 8. Evaluating the status by emotion HAPPY in 60 seconds

Evaluating the performance of the model by Accuracy and mean square error in T_1, HAPPY state:

```
w1:   -0.00244487718894251
w0:   90.60833239809607
```

Similar to the Neutral emotional state, Figure 9 visualizes the percentage level average of NEUTRAL status:

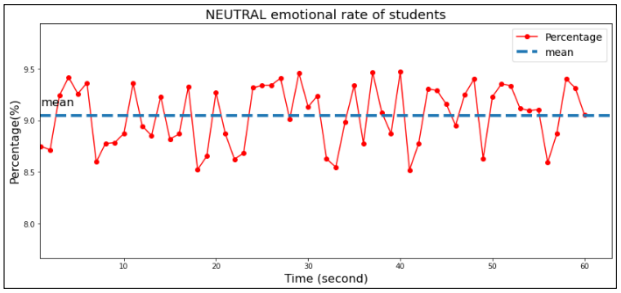

Figure 9. Evaluating the status by emotion NEUTRAL in 60 seconds

Neutral emotional state obtained:

```
w1:   0.0027309736940815737
w0:   8.966680923163844
```

Similarly, with the emotional Sad state, Figure 10 visualizes the percentage level average of SAD status.

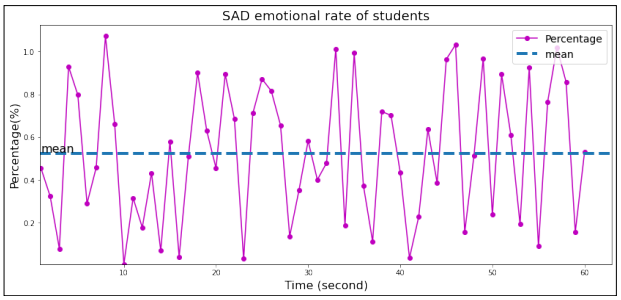

Figure 10. Evaluating the status by emotion SAD in 60 seconds

Variance and precision:

```
w1:   0.002333053667685484
w0:   0.454398166468926
```

For the remaining emotional states, the percentage of occurrence is not present. Continue to choose 60 images over the next 60 seconds (T_2), and experiment similarly as above, Neutral has the highest appearance rate ~80%, followed by Neutral ~20%, the remaining states account for the highest proportion ~ 0. Figure 11 visualizes the two states Happy and Neutral of the next 60 seconds (T_2).

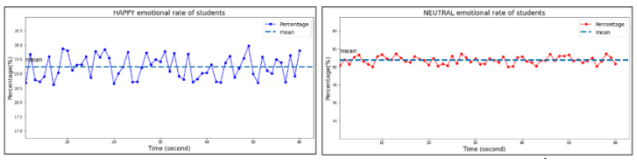

Figure 11. Evaluating the status by emotion HAPPY and NEUTRAL in 60 seconds

In T_2, HAPPY emotional state:
w1: 0.00019752267554824925
w0: 19.228659145512445
NEUTRAL emotional state:
w1: 0.000854584043345456
w0: 80.72030254234464

Finally, to get the accuracy of the appearance rate of emotional states from students' faces, compare the average of all the rates of emotions of the same type from the analyzed T times (60 seconds):

$$\boldsymbol{avAccS} = \frac{1}{n}\sum_{T=1}^{n}(T_1 w_0 + T_2 w_0 + \cdots T_n w_0)$$

Compare the average state accuracy (**avAccS**), HAPPY and NEUTRAL emotional rates of the total of the first 60 seconds (T_1) and 60 seconds (T_2) continued:

$$\boldsymbol{avAccHappy} = \frac{\mathbf{90.60833239809607 + 19.228659145512445}}{\mathbf{2}}$$

$$\boldsymbol{avAccNeutral} = \frac{\mathbf{8.966680923163844 + 80.72030254234464}}{\mathbf{2}}$$

Thence inferred: ***avAccHappy > avAccNeutral***

The HAPPY facial emotional state appeared more than the NEUTRAL facial emotional state extracted from the camera for 2 minutes.
Thereby, we can predict the facial emotions of students in the classroom over a long period of time.

5. Conclusion

Research Deep Learning algorithms to analyze features to enrich original training data. Thereby, choose a suitable CNN algorithm to build a simulation program to analyze facial emotions and make accurate predictions of facial emotions. Furthermore, combined with CNN, the author uses a linear regression model to evaluate facial emotions in a video. The article has made some new contributions in applying Deep Learning to recognize facial emotions, applied to assessing student satisfaction in the classroom. The research results will contribute to improving the position and quality of scientific research at Nguyen Tat Thanh University.

Acknowledgment

We would like to thank Nguyen Tat Thanh University for the support of time and facilities for this study.

References

[1] Matsumoto, David, and Hyi Sung Hwang. *Reading facial expressions of emotion*, Psychological Science, 2011.
[2] D. Yanga, Abeer Alsadoona, P.W.C. Prasad, A. K. Singhb, A. Elchouemic, *An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment*, Elsevier B.V, 2018
[3] K. Mase, A. Pentland. *Recognition of facial expression from optical flow*, IEEE TRANSACTIONS on Information and Systems, Vol E74-D, No10, pp. 3474-3483, 1991
[4] I Goodfellow, D Erhan, PL Carrier, A Courville, M Mirza, B Hamner, W Cukierski, Y Tang, DH Lee, Y Zhou, C Ramaiah, F Feng, R Li, X Wang, D Athanasakis, J Shawe-Taylor, M Milakov, J Park, R Ionescu, M Popescu, C Grozea, J Bergstra, J Xie, L Romaszko, B Xu, Z Chuang, and Y. Bengio. *Challenges in Representation Learning: A report on three machine learning contests*. arXiv 2013, 2013.
[5] Bengio, Yoshua. *Learning Deep Architectures for AI, Foundations and Trends in Machine Learning*: Vol. 2: No.1, pp 1-127, 2009.
[6] Paul Viola and Michael Jones. *Rapid Object Detection using a Boosted Cascade of Simple Features* IEEE, 2001.
[7] Honglak Lee, Roger Grosse, Rajesh Ranganath and Andrew Y. Ng. *Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations*, ICML, 2009.
[8] Soad Almabdy, Lamiaa Elrefaei, *Deep convolutional neural network-based approaches for face recognition*, Appl. Sci, Doi:10.3390/app9204397, 2019.
[9] Keyur Patel, Dev Mehta, Chinmay Mistry, Rajesh Gupta, Sudeep Tanwar, Neeraj Kumar, Mamoun Alazab. *Facial Sentiment Analysis using AI Techniques: State-of-the-Art, Taxonomies, and Challenges*, IEEE Access, Doi 10.1109/ACCESS.2020.2993803, 2020.
[10] Wafa Mellouk, Wahida Handouzi. *Facial emotion recognition using deep learning: review and insights*, Elsevier B.V, 2020.
[11] C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez. *Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild*, in Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016.
[12] Imane Lasri, Anouar Riad Solh, Mourad El Belkacemi. *Facial Emotion Recognition of Students using Convolutional Neural Network*, IEEE, 2019.
[13] A. Mollahosseini, B. Hasani, and M. H. Mahoor. *Affectnet: A database for facial expression, valence, and arousal computing in the wild*, IEEE Transactions on Affective Computing, vol. PP, no. 99, pp. 1-1, 2017.
[14] X. Liu, B. Kumar, J. You, and P. Jia, *Adaptive deep metric learning for identity-aware facial expression recognition*, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, pp. 522-531, 2017.
[15] V. Vielzeuf, S. Pateux, and F. Jurie, *Temporal multimodal fusion for video emotion classification in the wild*, Proc.

ACM International Conference on Multimodal Interaction, pp. 569-576, 2017.

[16] Vladimir Cherkassky, Yunqian Ma. *Selecting the Loss Function for Robust Linear Regression*, NC 2569 Under Review in Neural Computation, Revised June 10, 2002

[17] Xiaogang Su, Xin Yan, Chih-Ling Tsai. *Linear regression*, Volume 4, May/June 2012 Wiley Periodical s, Inc, https://doi.org/10.1002/wics.1198, 2012.

[18] Shuang Liu, Dahua Li, Qiang Gao, Yu Gong, *Facial Emotion Recognition Based on CNN*, IEEE, 2020.

[19] S. Turabzadeh, H. Meng, R. Swash, M. Pleva, and J. Juhar, *Facial Expression Emotion Detection for Real-Time Embedded Systems*, Technologies, vol. 6, no. 1, p. 17, Jan, 2018.

[20] J. Flores. *Training a TensorFlow model to recognize emotions, Available*: https://medium.com/@jsflo.dev/training-a-tensorflow-model-to-recognize-emotions-a20c3bcd6468, 2018