Thesis for Degree of
Master of Engineering

Facial Emotion Recognition using Deep Learning with Real-Time Convolutional Neural Networks

Graduate School of Industry
University of Ulsan

Department of Global Smart IT Convergence

AKHROR TUYAKOV ABDUSALOM UGLI

Facial Emotion Recognition using Deep Learning with Real-Time Convolutional Neural Networks

Supervisor: Chong, UiPil

A Dissertation

Submitted to
The Graduate School of Industry, University of Ulsan
In partial fulfilment of the Requirements
For the Degree of

Master of Engineering

by

AKHROR TUYAKOV ABDUSALOM UGLI

Department of Global Smart IT Convergence
Graduate School of Industry, University of Ulsan
Ulsan, South Korea

July 2022

Facial Emotion Recognition using Deep Learning with Real-Time Convolutional Neural Networks

This certifies that dissertation of Akhror Tuyakov Abdusalom Ugli is approved.

Committee Chair: Prof. Kim Dae Hwan

Committee Member: Prof. Park, Ju Chull

Committee Member: Prof. Chong, UiPil

Department of Global Smart IT Convergence
Graduate School of Industry, University of Ulsan
Ulsan, South Korea

July 2022

Abstract

Facial Expression Recognition while participating online video lessons, is a cutting-edge approach of gathering feedback from live facial expressions. It has been proposed that all educational institutions make the technology available, which would help to improve the quality of instruction by automatically monitoring the students' moods throughout the lecture and providing the lecturer with immediate feedback.

This study describes a program that concurrently gauges the learner's emotional state and captures on-screen activities in real time. This work consists of two main parts: facial expression recognition is the first section of this thesis. The FER2013 data set is used to train the convolutional neural network (CNN) for this section. The system accurately categorizes seven different emotions after being fed a series of face photos from the previous stage. In second part, analyze those facial expressions of students. We created an application utilizing the open-source JavaScript face-CNN-based API's trained Facial Expression Recognition Model, and we experimentally tested it to see if it could recognize facial expressions. The aim of this study is to help alleviate disruptions to learning and instruction caused by the pandemic by presenting an intuitive way to measure concentration, understanding, and engagement expected of a productive classroom.

Key words: Facial Expression Recognition, Convolutional Neural Networks, API, JavaScript, Real-time.

국문 초록

온라인 비디오 수업에 참여하는 사람들의 얼굴 표정 인식은 실시간 얼굴 표정으로부터 피드백을 받는 최첨단 접근 방식입니다. 많은 교육 기관에서 이 기술을 사용할 수 있도록 제안하며, 이는 강의 동안에 학생의 수강상태를 자동으로 모니터링함으로서 강사에게 즉각적인 피드백을 제공하여 강의 품질을 향상시키는 데 도움이 될 것입니다.

본 연구는 학습자의 감정 상태를 측정하고 화면상의 활동을 실시간으로 캡처하는 프로그램을 설명합니다. 이 작업은 두 가지 주요 부분으로 구성되는데, 첫 번째는 얼굴 표정 인식을 위하여 FER2013 데이터 세트를 이용하여 CNN(Convolutional Neural Network)을 훈련하는 부분입니다. 이 시스템은 시스템의 입력 사진으로부터 7 가지의 감정으로 정확하게 분류합니다.

두 번째 파트에서는 학생들의 표정을 분석합니다. 오픈 소스인 JavaScript face-CNN 기반 API 의 훈련된 얼굴 표정 인식 모델을 활용하여 애플리케이션을 만들고 얼굴 표정인식을 실험적으로 테스트했습니다. 이 연구의 목적은 수업시간에 학생들의 집중, 이해 및 참여도를 측정하는 데 직관적인 방법을 제시함으로써 전염병 등으로 인한 학습 및 수업 중단의 문제를 최소화하는 데 도움이 될 것입니다.

Acknowledgement

First and foremost, I would like to praise God for bestowing me with never ending patience, grace and will to pursue this work.

I would like to express my gratitude to Prof. Chong UiPil for all the support, help, guidance as well as for his valuable advice and encouragement during my thesis work. His observations and comments helped me to establish the overall direction to the research and to move forward with investigation in depth. I would like to thank him warmly and to express my sincere gratitude to him.

My sincere thanks to the University of Ulsan for giving me this chance to study at such a great university. Also, thanks to my committee members, prof. Kim Dae Hwan and prof. Park Ju Chull.

Last and not least, I must express my special thanks to my parents for their love, trust, understanding and every kind of support not only throughout my thesis but also throughout my life.

Content

Abbreviations

AI	Artificial Intelligence
DL	Deep Learning
ML	Machin Learning
FER	Facial Expression Recognition
ANN	Artificial Neural Networks
CNN	Convolutional Neural Networks
DLN	Deep Learning Networks
HCI	Human-Computer Interaction
BGR	Blue Green Red
RGB	Red Green Blue
ReLU	Rectified Liner Unit
GUI	Graphical User Interface
UX	User Experience
CK	Cohn - Kanade
API	Application Programming Interface

List of Tables

List of Figures

Chapter 1

1. Introduction

There are several ways to perceive information from one another, for example, verbal and non-verbal ways in human communication [1] . Verbal communication is the best way of language and transfers information through words; however, body language, gestures, and only non-verbal communication is used through facial expressions. Through their facial expressions, people may convey their emotional states in one of the most effective non-verbal ways possible. Human emotion detection is critical in the field of artificial intelligence (AI), human-computer interaction (HCI), intelligent robotics, and health care. The facial expression approach involves identifying facial expressions and categorizing them as happy, sad, angry, etc. In this study, I conclude different types of results from these evaluated emotions. In one case, confusion can be expressed when someone is having trouble or doesn't grasp something. The challenging point of human competence is the ability to identify the emotional state and feelings in this task.

According to the influence of globalization and the fast advancement of technology in the education field, considering the removal of barriers between students, new techniques and viewpoints have emerged for educational activities like online learning. With the use of this online learning platform, students may study in a more convenient and pleasant manner. Also, it helps teachers to undertake administrative tasks more efficiently, and they need to provide material that is highly engaging that is presented in a more accessible way.

Deep learning (DL) is a subfield of Machine learning (ML) method that instructs computers to learn by doing what comes naturally to people. In DL, a computer model learns to perform classification tasks directly from images, text, or sound based on artificial neural networks (ANN) that contain many layers, and it provides the appropriate context for grasping the lesson. DL models are trained by using a large set of labeled data and neural network architectures learn features directly from data without the need for manual feature extraction.   Learning facial expressions makes a significant contribution, and teachers utilize these expressions and technologies as a useful tool for further instruction and estimating the learning capacity of the student in a virtual class. Deep learning networks (DLN) and convolutional neural networks are two examples of deep learning techniques that we may use to recognize face emotions (CNN or ConvNet). One of the finest techniques for emotion recognition is CNN, and I have used CNN in this work.

We added CNN-based JavaScript Face API, this free-source library that is accessible on GitHub, for the practical portion of this thesis. Three of the models—Face Detection, Emotion Recognition, and Face Expression Recognition Model—were utilized in my study.

1.1 Purpose and Motivations

The motivating factor behind this project is the increased use online learning tools in education after pandemic. The COVID-19 pandemic [2] has triggered new ways of education. Within the entire world, educational institutions are ready to implement online learning platforms instated of delay the process of educating young generation and now, digital learning has emerged as a part of education that students and schools have had to adopt. As we saw above any lockdown cannot affect distance learning, however as with most modern technologies have advantages and disadvantages, e-learning also has its own set of pros and cons. This is difficult to control all students during the online lecture such as all students are following or not, which students have trouble to understand particular concept

On the other hand, the inspiration behind this study is to make an interactive platform where both sides: Students and instructors can communicate, and teachers can better understand their students' emotions to guide future decisions. Teachers can get better by being aware of students' emotions teachers may improve their weakness or help them to understand the particular topics. Also, teacher can evaluate the method that they used during online class.

1.2 Development of Online Learning systems

Providing the computers can communicate with one another on the internet network, it can be accessed from all over the world. Realizing web-based learning is not just uploading all materials on the web and then accessing them through other computers. However, it is not easy as imagined. The most important thing in online classes or meetings is interactivity between teachers and students. In an Online environment, monitoring is more difficult than in offline classes.

The majority of individuals believe that online learning is inferior than in-person instruction because educators have much more opportunities to observe every student in front of them and are better able to identify whether pupils are satisfied, perplexed, or unsatisfied with a certain subject. However, during online learning, the ability of the teacher to recognize students' emotions is really hard. Here a system is needed that must detect students' expressions and then provide real-time feedback to lecturers on their students' levels of engagement that which part of the online class is hard to understand and which part of that meeting is more satisfied by students in a virtual classroom.

Smart education includes online learning as a key component. Smart education gives more chances to teaching and learning styles. It has revolutionized education in both schools and businesses, allowing students and employees to learn at their own pace in a comfortable atmosphere. Furthermore, among researchers, educators, and businesses it has become an immense interest topic. Today's education is becoming more accessible, practical, and affordable thanks to advances in online learning. Students can choose from a variety of online courses and books to further their education online. By looking at numerous areas of online education, a rich and diversified educational system generates a significant body of study. Online learning has a significant influence on both the educational system and society at large; therefore it has the ability to change how learning is delivered in a Passarelli [3].

1.3 The limitations of online education

Additionally, it is evident that some facets of higher education, such as classroom contact with classmates and instructors, dorm life on campus, and university parties, cannot be "zoomed.". In Covid-19 Pandemic, the number of learners studied using online textbooks, thus in this circumstance, there is a tremendous need to develop a framework that recognizes students' emotions. After the pandemic, some universities have already started implementing a hybrid education model in that some classes are conducted online and others are offline. Online learning and facial expression analysis are the subjects of a study conducted by many researchers. They employ several techniques to determine facial emotions, such as CNN. As a result of facial emotion recognition research, there have been limitations of accuracy and recognizing emotions takes the most time while online classes.

1.4 Problem Statement

The large-scale availability of online education and materials such as books, articles, and contents have made learners' lifestyles easier because they may find many contents about a particular topic more conveniently. Depending on students, there are some dilemmas while learning online. While in actual courses, professors can immediately assess students' facial expressions, but when they are observing an online lecture, it may be more challenging to do so. A method that assesses a student's feelings when they follow the teacher is required since it gets challenging to discern the student's emotions.

In order to measure student interest in an onscreen task without the use of sophisticated gear, this study intends to establish a method. Some solutions to this issue include specialized eye tracking technology. However, I make use of a straightforward sensor, like the camera found in most computers. Upon completion of the exercise, a graph displaying the proportion of time spent staring at a screen and emotional state statistics is generated.

1.5 Research Objectives

This research aims to develop a system that can help online educators create an online education environment more efficiently. This work includes the following ideas:

- Making a screen recording of a student interacting with a learning item on a typical laptop while simultaneously capturing each individual student's reaction, head attitude, and mood.
- Identify the student's interests by analyzing their feelings on specific subjects

1.6 Related works

Since there has been a fair lot of research on the topic of emotion identification in static photographs, we were able to consult a number of articles for advice on best practices and methods to enhance our model. FER-2013 was developed as a component of a broader project that includes a Kaggle.com competition [4] . We were able to refer to that project in order to comprehend the dataset and the kinds of outcomes we would anticipate. In particular, the competition's winner successfully implemented a CNN with an output layer that flowed into a linear SVM to obtain a test accuracy of 71% [5] Zhang et al. achieved the highest accuracy 75.2%[6]

Chapter 2

Literature review

2.1 Deep learning for Computer vison

Computer vision is a study of building artificial systems that can process, perceive, and otherwise reason about visual data that could be images, videos, medical scans, or just any type of continuously valued signals [7] . The technique to be using computer vision across the field is deep learning. DL is the process of building artificial systems that can help someone worth bogging all to the goals of a computer. Due to its integration of all three artificial intelligence subtypes, deep learning is a distinctive approach to machine learning (Figure 1). Deep learning addresses challenges including scheduling, evaluating, and processing information. [8] .

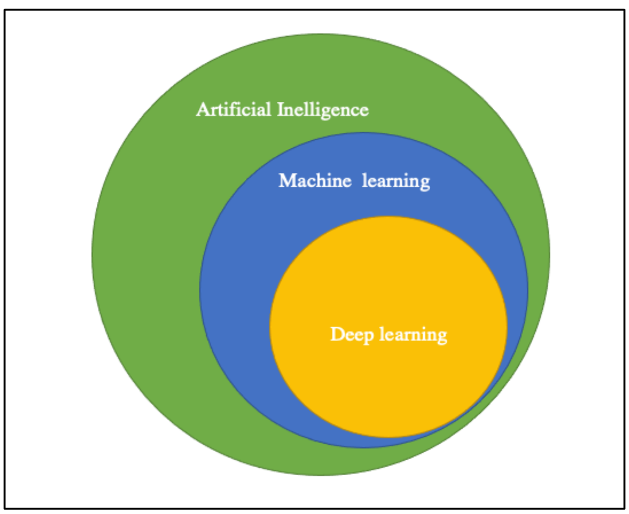

Figure 1. *Diagram of AI including DL and ML*

Feature learning, a set of techniques that enable computers to automatically detect the representations required for feature categorization from unstructured data, is the cornerstone of deep learning. Deep learning divides and distributes tasks among machine learning algorithms, which are organized in a network that can replicate the natural decision-making capabilities of the human brain [9] . As a result, deep learning is among the most effective methods for automatically recognizing characteristics.

2.2 Convolutional Neural Networks (CNN)

The parameter-sharing architecture and the concept of receptive field distinguish the convolutional neural network, which was first described by LeCun et al. [10]. in 1998. CNN has recently regained popularity and obtained exceptional state-of-the-art outcomes in the field of computer vision compared to traditional approaches. The power of CNNs to learn and extract features directly from raw input data (including distorted images) may be the reason for their growing popularity. Conventional machine learning and computer vision approaches require manually retrieved features. CNNs integrate the three FER phases (feature learning, feature selection, and classifier development) into one step and require minimum pre-processing. Furthermore, jobs that requiring calculation can reach promising results with minimal power consumption thanks to the graphical processing unit (GPU) technology.

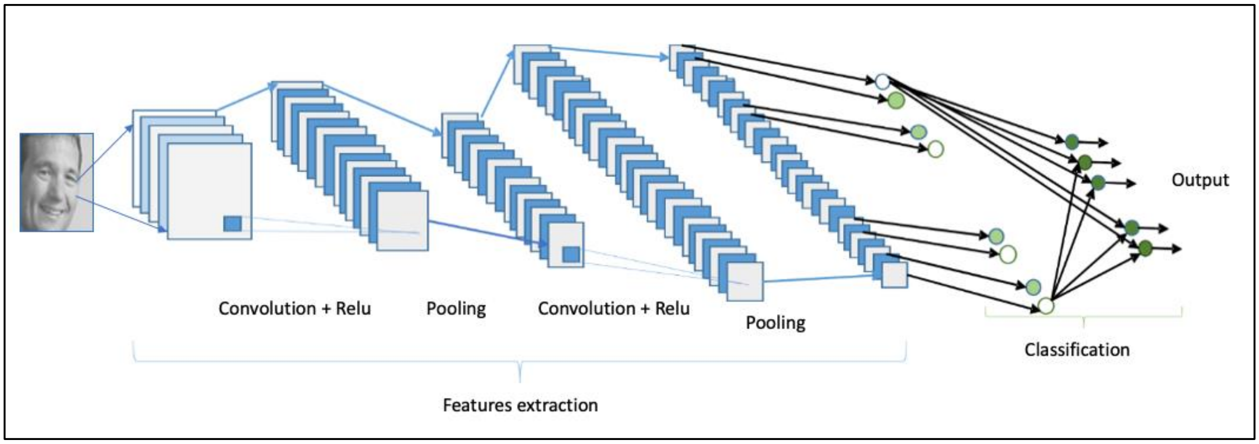

Figure 2. *Architecture of CNN*

2.2.1 Feature detection

The most popular use of CNN is for image analysis. Now, through the use of matrix multiplication and other concepts from linear algebra, CNN offers a more scalable method for performing picture classification and object identification tasks. The architecture of CNN is distinctive because it has main three different sorts of layers. A convolutional network's initial layer is the convolutional layer. Convolutional layers may be followed by pooling layers, but the fully-connected layer is always placed last. The CNN gets more complex with each layer as it recognizes more of the image [11] .

- *Convolution* activates various features of the input pictures by running them through a sequence of convolutional filters. This layer performs a dot product between two matrices, one of which is the limited portion of the receptive field and the other is a set of learnable parameters known as a kernel. While the kernel is smaller than an image, it is deeper. The kernel height and width, if the picture includes three channels, will be small spatially, but the depth will cover all three channels.
- *The pooling layer* substitutes the output of the network at particular locations using a summary statistic of nearby outputs. This minimizes the spatial dimension of the representation, which lowers the amount of computation and weights required. Throughout the pooling procedure, each slice of the representation is handled separately.
- *The fully-connected layer* helps in the mapping of representations between input and output.

These operations mentioned above are repeated over tens or hundreds of layers, with each layer learning to detect different features [12] .

2.2.2 Classification

The obtained feature vector will be fed to the face classification step, which needs a training phase to make a classifier or more capable of recognizing probe images [13]. There are three classifier categories: 1) Similarity-based ones known as the nearest neighbor rule, where the approach relies on grouping the similar patterns into the same class by establishing a distance metric; 2) Probabilistic approach using Bayes decision rule, which is a conditional probability model minimizing the misclassification probability. Text categorization has seen particular popularity with Naive Bayes classifiers; 3) Decision boundaries methods such as support vector machine, used for binary classification, underlay a given feature space into two zones representing two classes [14].

Random data can be categorized using CNN. Data are grouped according to aspects, traits, and qualities that they have in common with other sets of data in the same categories. A CNN is simpler to train than other classification algorithms since it uses fewer connections and parameters.

2.3 Overview of Facial expression

Facial expressions are one of the most fundamental factors of human communication. Not only thoughts and ideas are transmitted through the face, but also emotions. Facial expressions are configurations of various facial muscle movements that are used to determine a person's emotional state, such as anger, disgust, fear, joy, sadness, surprise, and, to a lesser extent, contempt, embarrassment, interest, pain, and shame. Explained face recognition in depth and provided various case studies conducted on the subject by other academics. According to psychologist Robert Plutchik [15] , there are eight fundamental emotions: happiness, confidence, surprise, fear, sorrow, anticipation, anger, and disgust. The wheel of emotions was designed by Plutchik to show the different connections between emotions. Plutchik only lists eight fundamental emotions, but the wheel demonstrates that there are many different degrees, giving rise to a wide spectrum of emotions. Plutchik argues that emotions are far more complicated than most people realize.

Any pair of adjacent fundamental emotions can be joined to form a new emotion. The 2D's opposite side sensory wheel, next to each other, are the two primary senses that make up the opposite feeling of the elicited emotion. Eight fundamental colors correspond to the eight basic expressions, and the color changes depending on how strong the feeling is [16] .

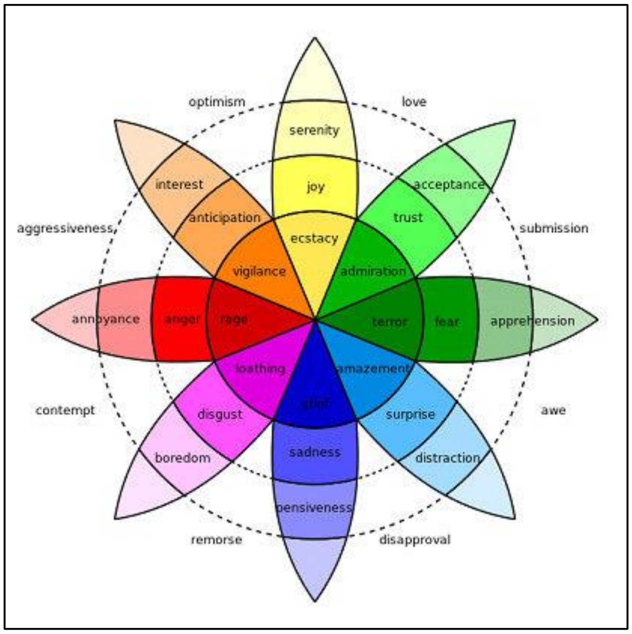

Figure 3. *Plutchik's wheel of expressions* [17]

The six fundamental emotions—surprise, fear, anger, disgust, sorrow, and happiness—were described by Paul Ekman. [18] . The three facial features that make up each expressed emotion are the brows-forehead, eyes-lids, and lower face.

Chapter 3

Methods and Datasets

Our study includes three main stages during one cycle of this whole architecture of our web application such as (i) face detection and extraction of representative frames with interval five-seconds during the whole video stream, (ii) expression recognition using a classification model on face photos with labels for seven distinct emotional states; (iii) storage of each student's individual emotional data and representation in a comprehensible manner.

A series of facial photos are used to capture the expressions in real time. Figure 4 depicts the study's overall flow.

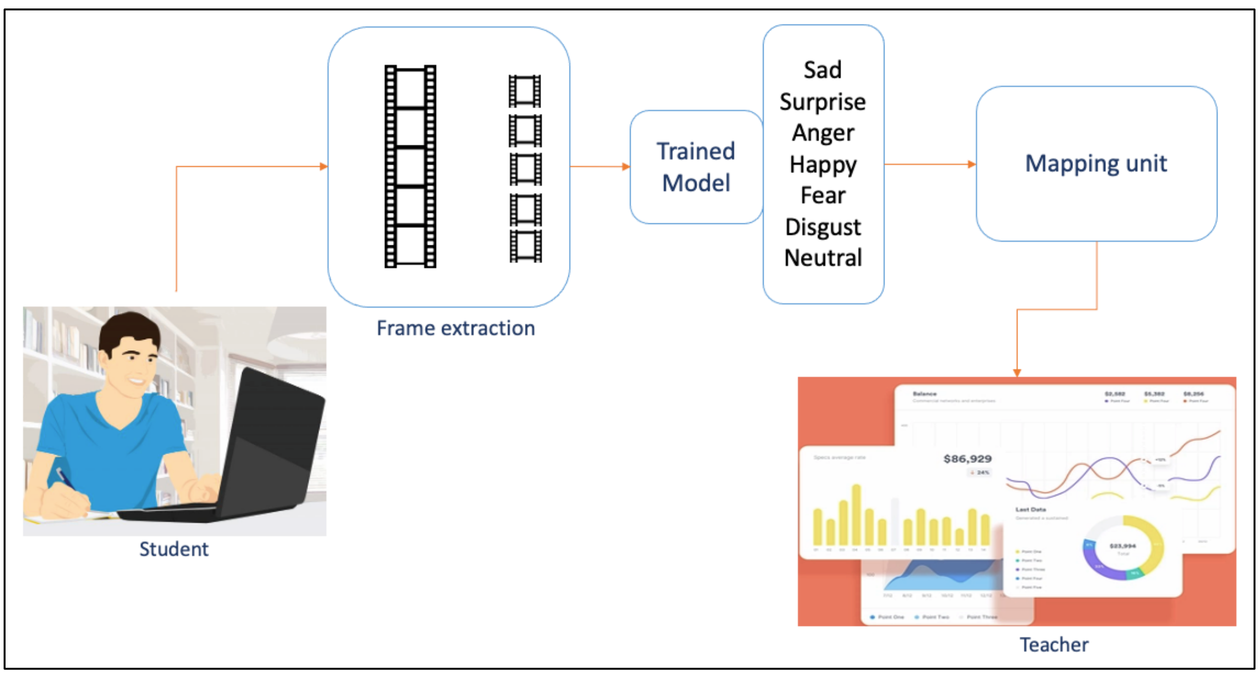

Figure 4. *Workflow visualization*

3.1 Customized Algorithm

In this study, the technology used with data augmentation is the Convolutional Neural Network. In Figure 3.2, there is a customized methodology that has been performed in my framework. The model initially takes an image from the dataset and uses a Cascade Classifier to find a face. If a face is discovered, it is forwarded to be preprocessed. Then the expanded dataset is loaded into CNN for class prediction.

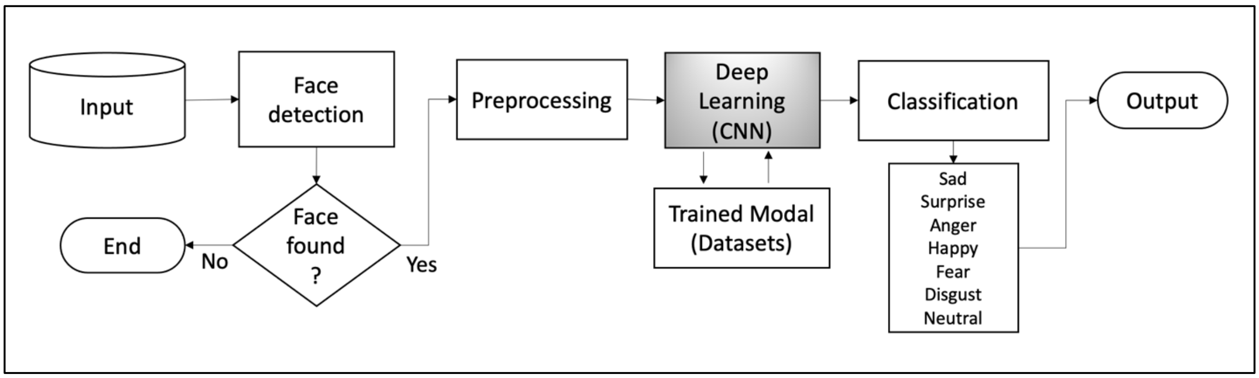

Figure 5 *Customized Algorithm*

3.1.1 Preprocessing

The data must be preprocessed in order for convolutional neural networks used for deep learning to make predictions that are more correct. In this framework, these preprocessing tasks typically involve:

- *Face detection and crop.* Face detection helps to find a face area in a given input image and then crop that area to use. Following face detection, the face area has been trimmed to remove background complexity and improve model training. In my case, I used an open-source method available on GitHub called Haar Feature-Based Cascade Classifier [19]

- *Grayscale conversion*. There also might be some scaling factors. After cropped face area that image size should have converted to the special size that the CNN expects. Pictures have been downsized to 9 pixels with 3 RGB colors. To reduce the complexity in pixel values, the images have been converted into BGR order having only one channel. So, it has become pretty much easy for the model to learn.

- *Image normalization.* A normalization process was performed on the dataset that method for matching a certain range of values by altering the pixel intensity levels. The process involves the contrast or histogram of an image is stretched in such a way that it allows the deep network to process the images more effectively.

- *Image augmentation*. It takes a lot of data to train a model with high accuracy and low error. But in reality, it's not that simple. If the dataset is large enough, it can extract additional features and compare them to the unlabeled data. Data augmentation is nothing, but given an input image, data augmentation can create multiple replicas of it by changing different properties of the image, such as performing image rotation, shift, flip, modify its brightness, change the angle etc.

Figure 6. *Data preprocessing* [20]

3.1.2 Proposed CNN Architecture

Convolutional neural networks differ from conventional neural networks in their architecture. Firstly, there are layers which consist of three dimensions such as height, width and depth. Finally, the output is summarized in a single possibility rating vector that is arranged along the depth dimension.

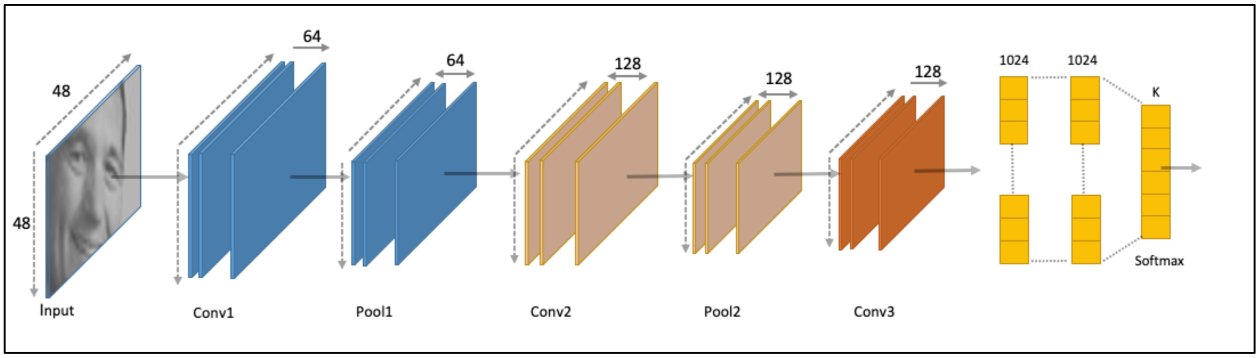

***Figure** 7. CNN layers*

In this thesis, I developed the CNN-based model to create the best accuracy while detecting facial expressions across a number of different parameters in a virtual lecture. The classification method used facial expressions has three convolution layers with 32, 64,128 and 256 filters, with a kernel size of 3x3. As model input, 48x48 sized photos were handed to the model. The model's input shape is (48,48,1), where 1 designates the input data's channel count. Since the photos have been changed to grayscale, there is just one channel. After the convolution layer, the model has a 2*2 pool size pooling layer, and max pooling has been selected. There are then four completely linked levels after that.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 48, 48, 32)	320
conv2d_1 (Conv2D)	(None, 48, 48, 64)	18496
batch_normalization (BatchNo	(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
conv2d_3 (Conv2D)	(None, 22, 22, 256)	295168
batch_normalization_1 (Batch	(None, 22, 22, 256)	1024
max_pooling2d_1 (MaxPooling2	(None, 11, 11, 256)	0
dropout_1 (Dropout)	(None, 11, 11, 256)	0
flatten (Flatten)	(None, 30976)	0
dense (Dense)	(None, 1024)	31720448
dropout_2 (Dropout)	(None, 1024)	0
dense_1 (Dense)	(None, 7)	7175

Figure 8 *CNN architecture*

In hidden layers, the ReLU activation function has been used. A dropout layer has been placed after each hidden layer, with the value of dropout set to 0.5. To avoid overfitting, it deactivates 50% of the hidden layer nodes at random. Finally, the model's output layer has seven nodes because it has seven classes. In the output layer, SoftMax was applied as an activation function.

A description of the convolutional neural network architecture developed for this study is shown in Figure 8.

3.1.3 Classification of facial expressions

After learning the deep characteristics, the final phase of FER is to assign the supplied face to one of the fundamental emotion categories. Unlike standard approaches, which separate the feature extraction and feature classification steps, deep networks may execute FER from start to finish. A loss layer is added to the network's end to handle the back-propagation error, allowing the network to output each sample's prediction probability without further processing. The most often used function in CNN is called SoftMax loss, which reduces the cross entropy between the calculated class probabilities and the actual distribution.

3.2 Datasets

Having enough labeled training data with as many different persons and environments as possible is essential for the development of a deep expression recognition system. Various standard facial datasets are available online:

Database	Samples	Expression distribution	Reference
CK+	593 images	6 basic expressions plus contempt and neutral	[21]
TED	112,234 images	6 basic expressions plus neutral	[22]
JAFFE	213 images	6 basic expressions plus neutral	[23]
FER2013	35,887 images	6 basic expressions plus neutral	[24]
ExpW	91,793 images	6 basic expressions plus neutral	[25]

Table 1. An overview of the facial expression datasets.

3.2.1 FER2013

In this work, we trained the publicly available dataset FER2013 that contains basic expressions and that are widely used for deep learning algorithms evaluation. The International Conference on Machine Learning released this collection in 2013. (ICML). The original authors of the open-source dataset FER2013 were Pierre-Luc Carrier, Aaron Courville, and others [24] . This dataset includes 48x48 pixel grayscale (BGR) images of face. These images are categorized each into seven different emotions: angry (4,593 images), disgust (547 images), fear (5,121 images), happy (8,989 images), sad (6,077 images), surprise (4,002 images), neutral (6,198 images) including test and train.

Figure 9. *Samples of FER2013 dataset*

This dataset is chosen because it includes photos displaying all seven emotional states and the data is balanced, it contributes favorably to the development of an effective CNN architecture.

Chapter 4

Results

4.1 Test the model

This section explains substantial the proposed classification method, the efficiency of the used model and the research findings in this chapter. Table 2 provides the environment of the system.

Category	Version
Operating System	macOS
CPU	Intel Core i5, 2.3 GHz
RAM	8 GB
Environment	Anaconda 4.7.5

Table 2. Experimental environment.

The framework has been written in python programing language, using the Jupiter notebook. There are some libraries which required in this work namely: Keras [26] – used for data argumentation; TensorFlow [27] – was used as system backend for built-in functions such as activation functions, optimizers; Matplotlib [28] – used to generate confusion matrix; NumPy [29] – was used every layer in CNN and OpenCV [30] – was mostly utilized for preparing images, such as grayscale conversion, image normalization.

4.1.1 Data preparation

For evaluation, the images in the FER2013 dataset are randomly shuffled that total training set and the public test set are used for training and validation respectively.

4.1.2 Training Parameters

To obtain a reliable test accuracy for the FER2013 dataset, a combination of experiments was conducted. It was first done using the optimizer with the ReLU activation function. The model was trained with multiple learning rates, including 0.001 and 0.0001, in a batch size of 64. However, the loss/accuracy curves showed some overfitting in the model because of all the pertinent learning rates. With a learning rate of 0.001, the best validation accuracy was attained at 66.56 percent. These models were all ran for 60 iterations.

4.2 Discussion and comparison

First, the first experiment is conducted to show the effectiveness of adopting the train. We compare the results of two scenarios that involve only from the pretrained weights using the FER2013 dataset.

As for the first-stage of training, this approach obtains high accuracy is achieved on training set but accuracy on validation set is stuck at 65.22%, which is approximately 2.5% lower than that of the highest-ranked approach [4] in the 2013 challenge on Kaggle (65 ± 5%).

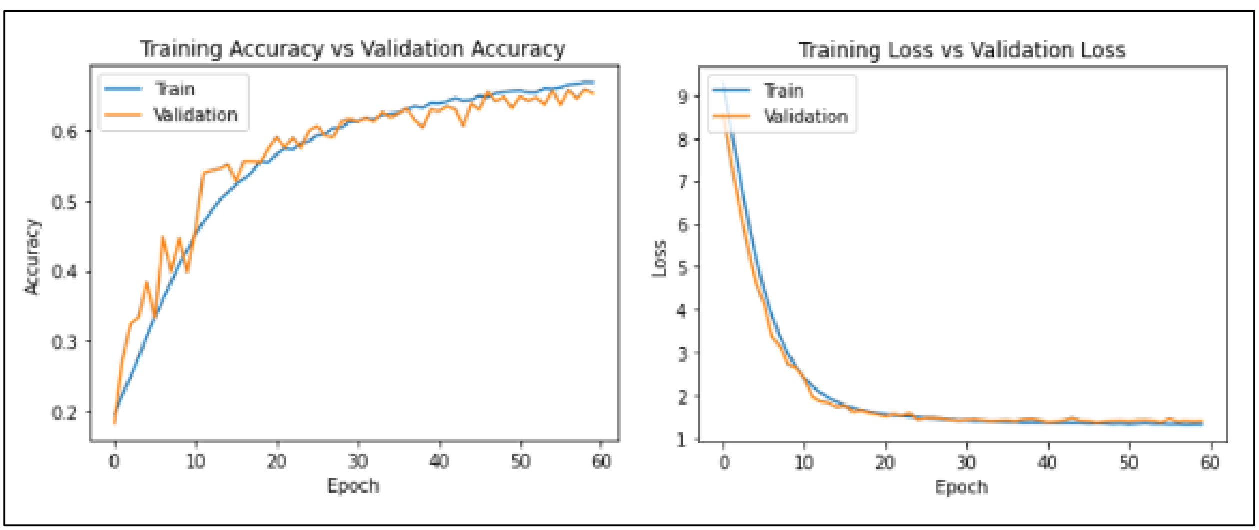

Figure 10. Loss/ Accuracy plot for the model with a learning rate of 0.001 for 60 epochs

Also no overfitting can see in the dataset hence is can be concluded that the inefficiency may be due to the unbalanced dataset. The total loss (which converges quickly during training) and the accuracy for validation are shown in Figure 10.

In the second experiment, we can see that the adoption of the model improved the train accuracy to 83.23% but the validation accuracy almost same as first experiment that has grown slightly (1.34%). The results presented above do not consider the center loss since its effectiveness is elaborated. This strategy improves the accuracy to the highest value at approximately 60 epochs. All changes shown with details below:

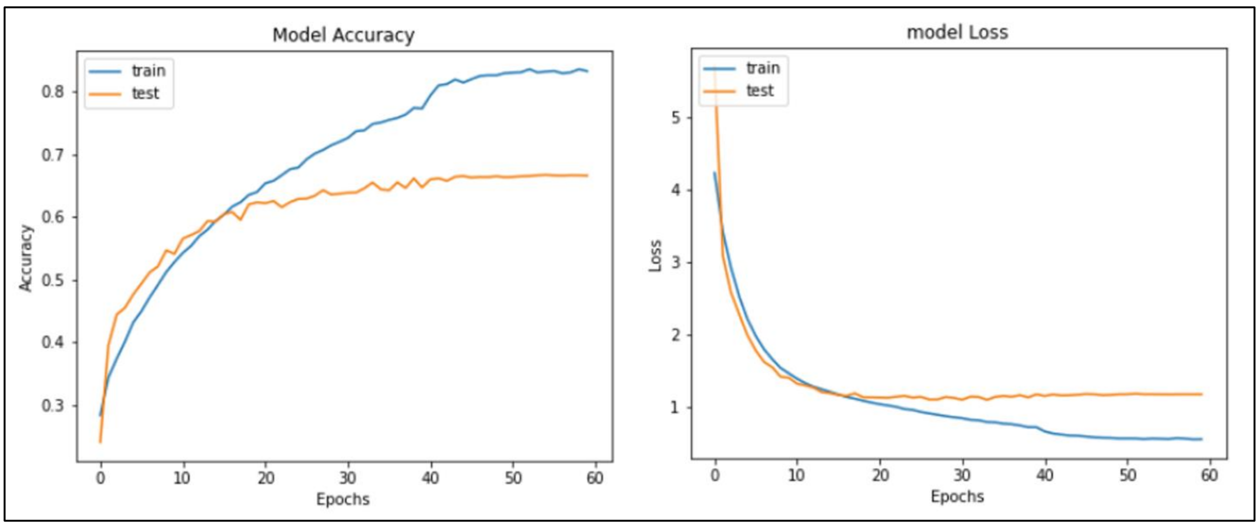

Figure 11. *Loss/ Accuracy plot for the model with a learning rate of 0.0001 for 60 epochs*

4.3 Application

In this section, the web application development will be discussed. This application can help to educator to gather information about students during online classes. The strengths of our application are user friendly (easy to use), it does not show emotions only during class but it stores all information to database and portability that we can access to application using web browser on any device such as smartphone, tablet.

4.3.1 APP idea and documentation

The program was created to be a very simple program that runs in real-time and approximates the CNN model's 5 second latency. JavaScript Face-API [31], which is integrated by the CNN Framework, has been used to recognize facial expressions and facial landmarks. Face-api.js makes use of expression flow and speeds up the browser. A number of capabilities, including Face Recognition, Face Landmark Detection, Face Expression Recognition, and Age Estimation & Gender Recognition, are available in this open-source API.

As this framework is web-based, so the students should have a device with a web-camera and they need to allow camera access that to use this feature. Three steps made up the design concept.

- When the app is launched, it immediately launches the camera and looks for students' faces. The message "No face identified" will appear on the screen if no faces are found.
- The face detection method is the second phase. If a human face is found, the algorithm advances to the next stage and feeds the picture of the face to the CNN model that has already been trained.
- Once a face image has been put into the CNN model, it will categorize the facial expression and link it to the appropriate emoticon for that class. Finally, this emoticon will be a time-related graph.

4.3.2 Result of APP

We also create an implementation for real-time facial expression recognition from a regular webcam to evaluate the proposed system's capacity to run in real time. Faces are preprocessed using the same processes after the webcam is linked to the network (without data augmentation). Following preprocessing, the data is fed into the trained model of choice, which uses the best assessment result to accomplish the classification. The subject is requested to stand in front of the camera as shown on Figure 12 and make one of the basic facial expressions. The computation time for classifying a single frame is evaluated, and the results of the literature comparison are shown as graph in Figure 13.

In this application teacher can see students' emotion during virtual class which our application collects a state of student in every five second along with the time period. Our system captured the students' emotions during the time they spent in a certain lesson, which we refer to as the time length.

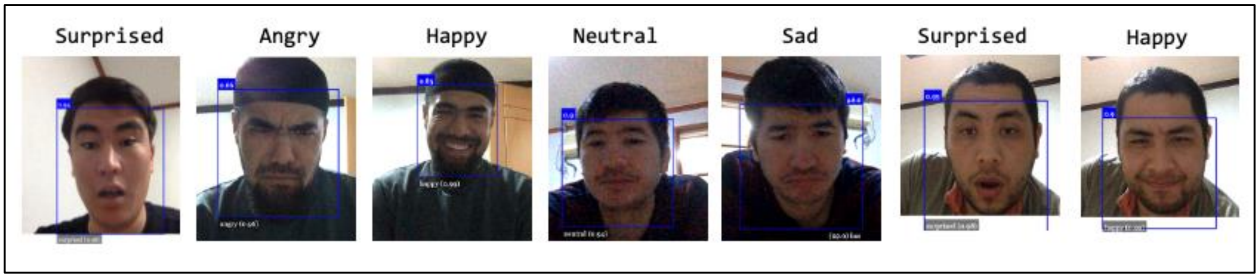

Figure 12. *Different emotion states*

In last part, for evaluators to see a graphical representation of statuses, we created a dashboard. A line graph has been provided on that dashboard. While taking any online lessons, the evaluator can see a visual representation of the student's facial expressions.

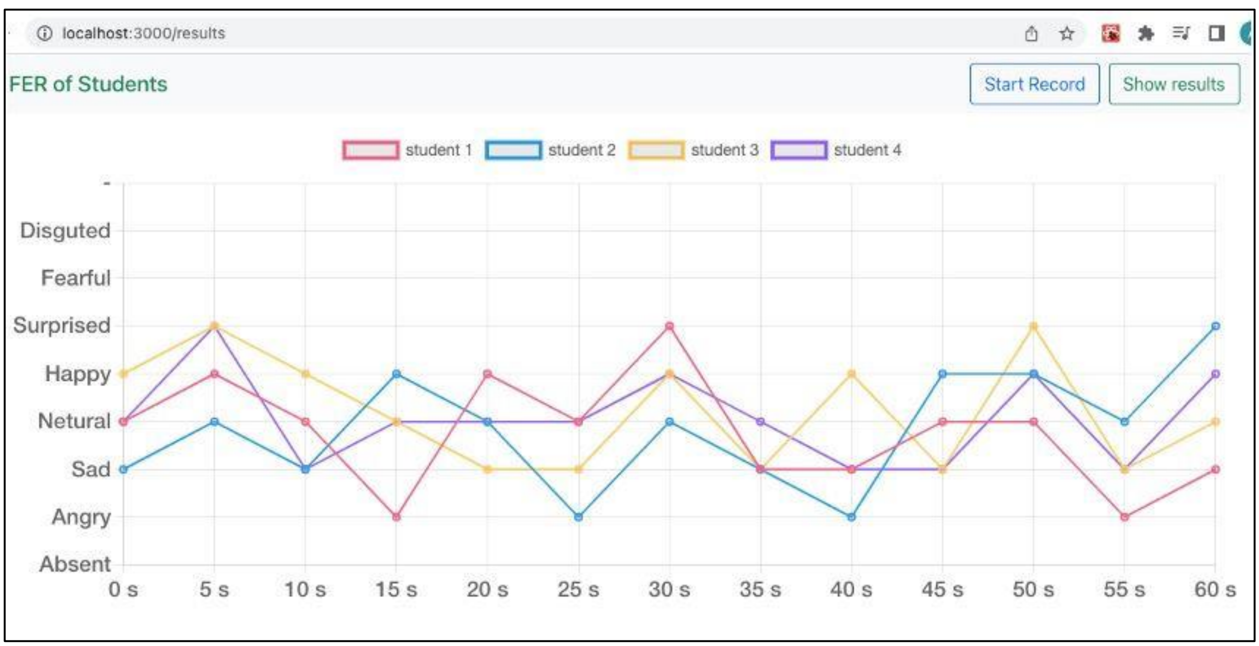

Figure 13. *Graphical view of emotion and time.*

Chapter 5

Conclusion and Future works

5.1 Conclusion

In current age of research and technology, using emotion detecting software during online classes is a successful strategy. Although this method was not available a few years ago. There are many online study programs available nowadays as the majority of educational systems transition to an online learning model. Through the many online learning platforms, there are numerous chances for learners to begin studying anything at anytime, anywhere. We need to comprehend the feelings of students when they are studying because our educational system is shifting toward an online learning platform. In this work, we attempted to address this issue by employing a cutting-edge technology to detect learners' facial expressions during studying via online applications. This expression detection approach can also be used to replace outdated feedback techniques.

By utilizing the idea of machine learning based on a CNN model, this study primarily takes into account the educational component. The experiments revealed respectable training and validation accuracy. We cannot anticipate a big increase in overall accuracy in this study just by increasing the number of repetitions (the accuracy increased slightly and not significantly).

5.2 Future works

We will expand this study in the future by involving additional students in order to undertake a more thorough examination. Our research now focuses on students, but we plan to expand it to include teachers in the future and create a platform that connects students and teachers. We can combine our software with other online learning environments so that they may acknowledge the emotions of their users and do better analyses using that data. In order for the teachers to accurately assess the participants, we will try to expand the number of classes of facial expressions and data for the assessment phase.

In the future, the model will need to be improved to predict more accurately. The appearance of a student's face should not have a significant impact on their ability to operate efficiently. The audience members' level of enjoyment may be gauged by changing the emotional state to a new value. Emotional intensities may also be included in emotion labels. This would enable the network model to distinguish among a happy and a very happy facial feeling. A dataset that is unaffected by size, angle of view and lighting was used to train the system model.

References:

[1] Di Pilato, A.; Taggio, N.; Pompili, A.; Iacobellis, M.; Di Florio, A.; Passarelli, D.; Samarelli, S. "*Deep Learning Approaches to Earth Observation Change Detection",* pp. 13-20, 2021.

[2] World Health Organization, https://covid19.who.int/data

[3] L. Deng, D. Yu et al., "*Deep learning: methods and applications,"* Foundations and Trends in Signal Processing, vol. 7, no. 3–4, pp. 197–387, 2014.

[4] Research Prediction Competition- https://www.kaggle.com/c/challenges-in-representation- learning-facial-expression-recognition-challenge

[5] Y. Tang "*Deep learning using Support Vector Machines*" in International Conference on Machine Learning Workshops, 2013

[6] Z. Zang , P. Luo, C. Loy and X.Tang "*Learning Social Relation Traits from Face Images*" in Proc, Conference on Computer Vison 2015.

[7] Justin Johnson, "*Introduction to Deep Learning for Computer Vision*", EECS 498.008 , 2022

[8] S. Skansi, *Introduction to Deep Learning: From Logical Calculus to Artifical Intelligence*, Springer International Publishing, 2018.

[9] P. Kim, MATLAB Deep Learning, Apress, 2017.

[10]. Yann LeCun et al. "*Gradient-based learning applied to document recognition*". In: Proceedings of the IEEE 86.11, pp. 2278–2324, 1998.

[11] Ajitesh Kumar, *"CNN Basic Architecture for Classification & Segmentation*", Retrieved from https://vitalflux.com/cnn-basic-architecture-for-classification-segmentation, September, 2021.

[12] Stanford University's Course — CS231n: *Convolutional Neural Network for Visual Recognition* by Prof. Fei-Fei Li, Justin Johnson, Serena Yeung

[13] Mohamed Kas. "*Development of handcrafted and deep based methods for face and facial expression recognition*" UTBM Montbéliard, pp. 14-29, 2021.

[14] N. Piao, R.-H. Park, *"Face recognition using dual difference regression classification*", IEEE Signal Processing Letters 22 (12) 2455–2458, 2015

[15] Plutchik, Robert, "*Emotions and Life: Perspectives from Psychology, Biology, and Evolution"*, Washington, DC: American Psychological Association, 2002.

[16] Robert Plutchik's *"Wheel of Emotions*" Retrieved from https://study.com/academy/lesson/robert-plutchiks-wheel-of-emotions-lesson-quiz.html, 2015.

[17] Pico, I. *The Wheel of Emotions, by Robert Plutchik* | PsicoPico. Retrieved from http://psicopico.com/en/la-rueda-las-emociones-robert-plutchik, 2016.

[18] P. Ekman, "*Universals and Cultural Differences in Facial Expressions of Emotion"* Nebraska Symposium on Motivation, vol. 19, pp. 207-282, 1972.

[19] Open source data, https://github.com/opencv/opencv/tree/master/data/haarcascades

[20] Tawsin Uddin Ahmed, "*Facial Expression Recognition using Convolutional Neural Network with Data Augmentation", 2014-4251*

[21] Open source dataset, https://www.kaggle.com/code/shawon10/ck-facial-expression-detection/data

[22] M. Susskind, A. K. Anderson, and G. E. Hinton, "*The Toronto face database*," Department of Computer Science, University of Toronto, Toronto, ON, Canada, Tech. Rep, vol. 3, 2010.

[23] Open source dataset, https://www.kasrl.org/jaffe_download.html

[24] Open source dataset, https://www.kaggle.com/datasets/msambare/fer2013

[25] Z. Zhang, P. Luo, C. L. Chen, "*From facial expression recognition to interpersonal relation prediction,*" International Journal of Computer Vision, vol. 126, no. 5, pp. 1–20, 2015

[26] Keras – deep learning API, https://keras.io/about.

[27] TensorFlow – open source platform. https://www.tensorflow.org/overview.

[28] Matplotib – visualization with Python, https://matplotlib.org.

[29] Numpy – fundamental package for scientific computing with Python, https://numpy.org.

[30] Open Source Computer Vision, https://docs.opencv.org/4.x/index.html

[31] Open source data, MIT licensed, https://github.com/justadudewhohacks/face-api.js/