

Minia University

Faculty of Science

IT Department

**CHILD FACIAL EXPRESSION DETECTION**

* Supervisor:

**Dr. Ahmed Hamdy**

* Teacher Assistant:

**Dr. Sohier**

Prepared By:

Nesma AbdElBaki

**Abstract**

**This book delves into detecting children's faces between the ages of 6 and 12 years with mean age of 7.3 years, then knowing and understanding his feelings through his facial expressions.**

**To gain and absorb what is in this book, it would be good if you have knowledge of:**

* **linear algebra**
* **Programming Language (Python)**
* **Neural networks**
* **Deep learning**
* **SQL**

**As a quick introduction, this book will provide an introduction to the child, his needs, and methods of dealing with him to build a normal personality, and the benefit from him.**

**And an introduction to deep learning, its uses and how to use it in reality.**

**After this introduction, we will talk about the technology used in the project, planning, analysis and design and go deeper into the details of the project configuration.**

**This book targets those interested in deep learning and its uses, and targets institutions related to children and interested in developing it.**

**Contents**

**Abstract …………………………………………………………………………………………….………………………….…… 2**

**Contents ……………………………………………………………………………………………………………………….…...3**

**1. Feasibility Study**

**1.1 Overview ……………………………………………………………………..………………………..………...6**

**1.2 Introduction ………..………………………………………………………………………..………………….7**

**1.2.1 What is the Facial Expression Detection?**

**1.2.2 How does Facial Expression Detection work?**

# 1.2.3 Neural Networks

# 1.2.4 Convolutional Neural Network

**1.3 Description ………………………………………………………………………………………………………10**

**2. Planning and Analysis**

**2.1 Project Data ………………………………………………………………………………………..…………..12**

**2.2 Methods ……………………………………………………………………………………..………….……….15**

**2.3 A**[**scope of work (SOW)**](https://www.projectmanager.com/blog/sow-to-project-plan) **…………………………………………………………………………………….17**

**2.3.1 CNN research…………………………..…………………………………………………………17**

**2.3.2 Theoretical research…………………………………………………………………………..18**

**2.4 Recourses ……………………………………………………………………………………………………….19**

**2.5 Constrains …………………………………………………………………………………………………….…19**

**2.6 Risks ………………………………………………………………………………………………………………..19**

**3. System Design**

**3.1 Actor Goal List…………………………………………………………………………………………………..21**

**3.2 Use Cases …………………………………………………………………………………………………………21**

**3.3 Use Case Diagram …………………………………………………………………………………………….22**

**4. Implementation**

**4.1 Data Capturing …………………………………………………………………….…………………………..24**

**4.2 Emotion Recognition ……………………………………………………………………..…………………28**

**4.2.1 Model …………………………………………………………………………………………………………28**

**4.3 Experimental framework and results ……………………………………………………….…….….32**

**4.4 Graphs ………………………………………………………………………………………………………….…..33**

**5. Report**

**5.1 Conclusion …………………………………………………………………………………………………..…. 23**

**5.2 Future Work …………………………………………………………………………………………………… 23**

**Proposal**

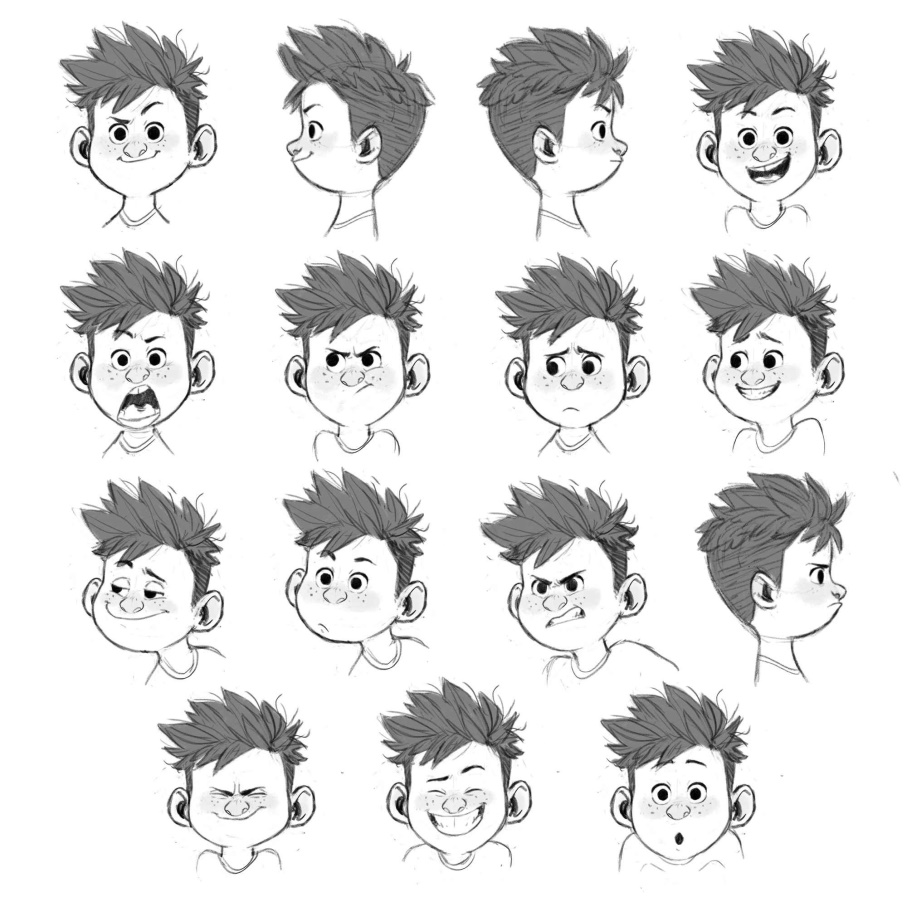
**Overview**

**In this section we will give the full picture of the project.**

**Because caring for children early is a future investment, our project serves them**

**The project is specially designed to recognize children's faces specifically and identify their different facial expressions to infer their simple feelings, and to closely identify their personalities so that we can direct them later, put them in their right places and focus on their interests.**

**So, the goal of the project is to create an algorithm capable of detecting children's feelings in a video clip, after it has been identified.**

****

**Introduction**

**Communication in any form i.e. verbal or non-verbal is vital to complete various daily routine tasks and plays a significant role in life.**

**Facial Expression Detection is the most effective form of non-verbal communication and it provides a clue about emotional state, mindset and intention.**

**1.3.1 What is the Facial Expression Detection?**

**Facial expression recognition or computer-based facial expression recognition system is important because of its ability to mimic human coding skills. Facial expressions and other gestures convey nonverbal communication cues that play an important role in interpersonal relations. These cues complement speech by helping the listener to interpret the intended meaning of spoken words. Therefore, facial expression recognition, because it extracts and analyzes information from an image or video feed, it is able to deliver unfiltered, unbiased emotional responses as data.**

**1.3.2 How does Facial Expression Detection work?**

**Facial expression recognition system is a computer-based technology and therefore, it uses algorithms to instantaneously detect faces, code facial expressions, and recognize emotional states. It does this by analyzing faces in images or video through computer powered cameras embedded in laptops, mobile phones, and digital signage systems, or cameras that are mounted onto computer screens. Facial analysis through computer powered cameras generally follows three steps:**

###### **1.Face detection**

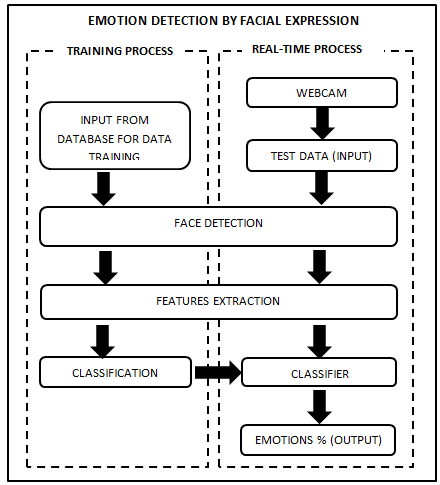
**Locating faces in the scene, in an image or video footage.**

###### **2.Facial landmark detection**

**Extracting information about facial features from detected faces. For example, detecting the shape of facial components or describing the texture of the skin in a facial area.**

###### **3.Facial expression and emotion classification**

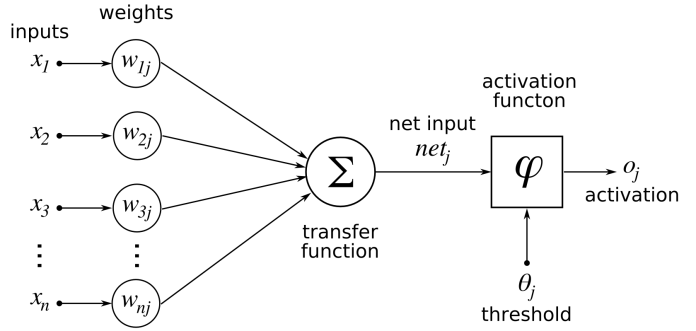
**Analyzing the movement of facial features and/or changes in the appearance of facial features and classifying this information into expression-interpretative categories such as facial muscle activations like smile or frown; emotion categories happiness or anger; attitude categories like (dis)liking or ambivalence.**

****

**1.3.3 Neural Networks**

**It’s a technique for building a computer program that learns from data. It is based very loosely on how we think the human brain works. First, a collection of software “neurons” are created and connected together, allowing them to send messages to each other. Next, the network is asked to solve a problem, which it attempts to do over and over, each time strengthening the connections that lead to success and diminishing those that lead to failure.**

**1.3.4 Convolutional Neural Network**



**CNN is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.**

**The architecture of a Convent is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.**

**Description**

**Most of the existing works focus on analyzing adult faces, which ignore how to analyze facial expressions from baby facial images. And there are actually very few images of children. None of these datasets is specifically designed to explore the expression of babies. There are two main reasons for the lack of research on baby face analysis. The first reason is that the community has not realized the application values of analyzing baby’s facial expression. In fact, there are many applications of analyzing the facial expressions of children, such as advertising marketing for parents, intelligent family child care, and scientific parenting.**

**Another reason is that obtaining accurate expressions and analyzes of a child's face is a challenge and an adventure.**

**The main objective of managing this project is to care for the child, absorb his feelings and exploit them for his benefit and meet his needs in all parts of the academic, social, medical, artistic life ... etc.**

**Planning and Analysis**

**2.1 Project Data**

**Database was the selection of visual stimuli that can induce emotions in children. Considering ethical reasons and young age of children, the choice was made carefully and removed any stimuli that can have long term negative impact on the children. Due to these ethical reasons, not included emotion inducer clips for the negative expression of “anger” and selected very few clips to induce emotion of “fear” and “sadness”.**

**Due to this very reason, the proposed database contains more emotional clips of expressions of “happiness” and “surprise”.**

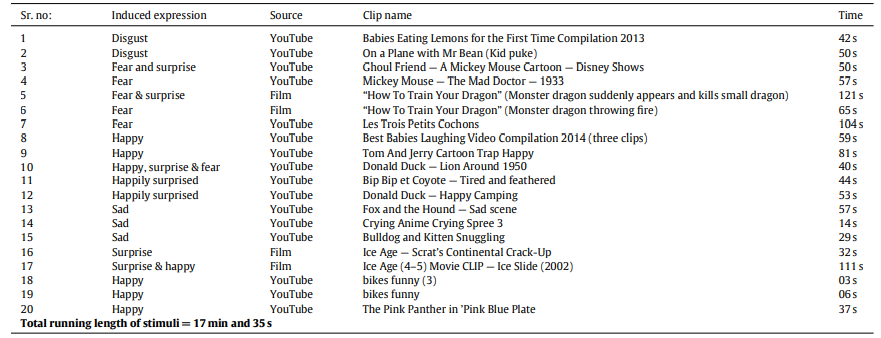
**Database used that contains movie clip of 12 ethnically diverse children. This unique database contains spontaneous / natural facial expression of children in diverse settings with diverse recording scenarios showing six universal or prototypic emotional expressions (happiness, sadness, anger, surprise, disgust and fear).**

**Children are recorded in constraint free environment (no restriction on head movement, no restriction on hands movement, free sitting setting, no restriction of any sort) while they watched specially built / selected stimuli. This constraint free environment allowed us to record spontaneous / natural expression of children as they occur.**

**The data is here:**

[**https://drive.google.com/drive/folders/1ERPIl\_Esd1b3v1xPUwIv425u1e8Qi-uy?usp=sharing**](https://drive.google.com/drive/folders/1ERPIl_Esd1b3v1xPUwIv425u1e8Qi-uy?usp=sharing)

* **List of stimuli selected as emotion inducers are presented in this table, Total running length of selected stimuli is 17 min and 35 s.**

****

**Generally, for evaluating and benchmarking different facial expression analysis algorithms, standardized databases are needed to enable a meaningful comparison. In the absence of comparative tests on such standardized databases it is difficult to find relative strengths and weaknesses of different facial expression recognition algorithms. Thus, it is utmost important to develop natural / spontaneous emotional database contains children movie clip / dynamic images.**

**This will allow research community to build robust system for children’s natural facial expression recognition.**

**Thus, this data is the best available because it is unconstrained, which is 208 dynamic film/photo clips of 12 ethnically diverse children showing spontaneous expressions in two environments ( laboratory/semester ) and ( home environment )**

**However, the data volume is small and unreliable; there is no balance and a lot of noise**

**With the framework of automatic facial expression recognition based on the neural network, it achieved an average classification accuracy of 69% in our database.**

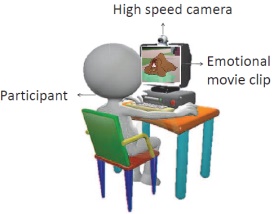
****

* **blended expressions**

**There are seventeen (17) video clips present in this database that have two labels, for example “happily surprised”, “ Fear surprise” etc. This is due to the fact that for young children different expressions co-occur / blended expressions or a visual stimulus was so immersive that transition from one expression to another expression was not pronounced.**

****

**2.2 Methods**

**In this section, the specific approach is illustrated. It can be divided into three parts: data collection and pre-processing feature extraction and classification and evaluation method.**

1. **Data Collection**

**Since the video is a more interactive experience than still photos**

**Visual stimuli that can stimulate feelings in a child, such as animation, have been carefully selected.**

**Then the child's expressions were recorded and approved after montage as data**

**The other part of the data is an Excel sheet that contains data for the children who were given the test (name and age).**

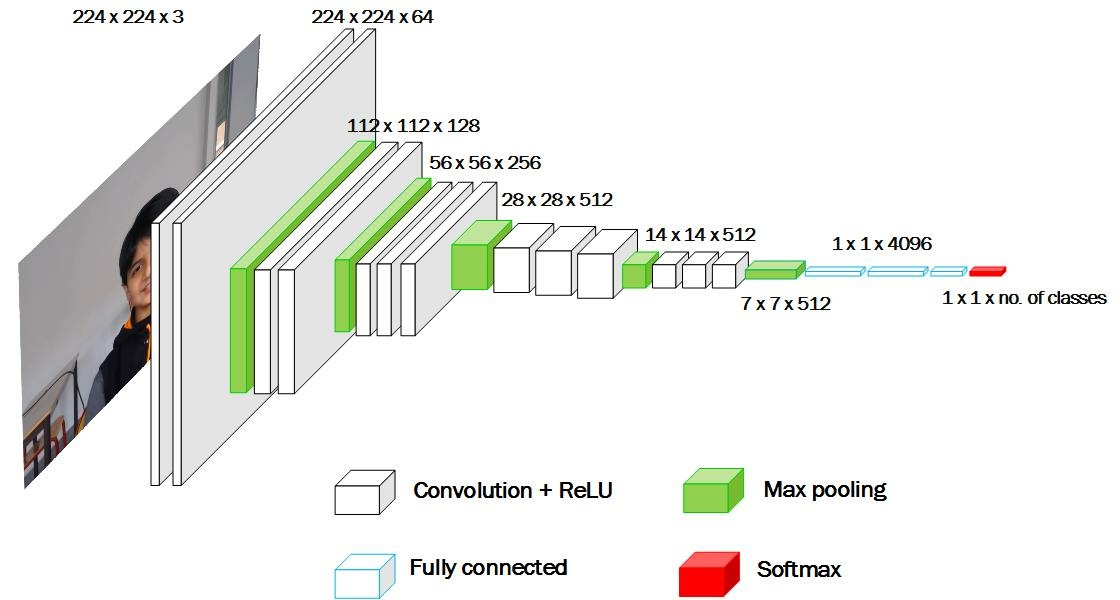
**And linking the two parts together, to move on to the second stage.**

1. **Pre-processing**

**Videos contain rich information. However, in order to train models more efficiently, only audios and facial crops from video frames are used in the training.**

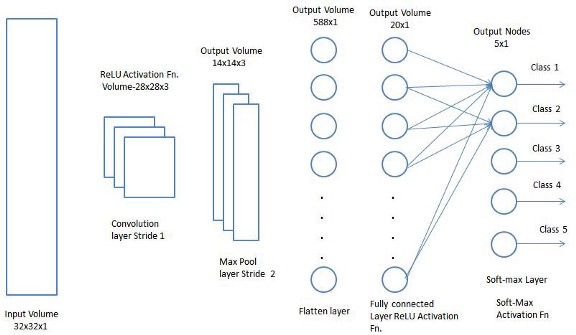
**In our case, we have 3 layers:**

* **Convolution layer: This layer makes use of filters, which are convolved with the image, producing activation or feature maps.**
* **Feature Pooling layer: This layer is inserted to reduce the size of the image representation, to make the computation efficient. The number of parameters is also reduced which in turn controls over-fitting.**
* **Classification layer: This is the fully connected layer. This layer computes the**[**probability**](https://deepai.org/machine-learning-glossary-and-terms/probability)**/ score learned classes from the extracted features from convolution layer in the preceding steps.**

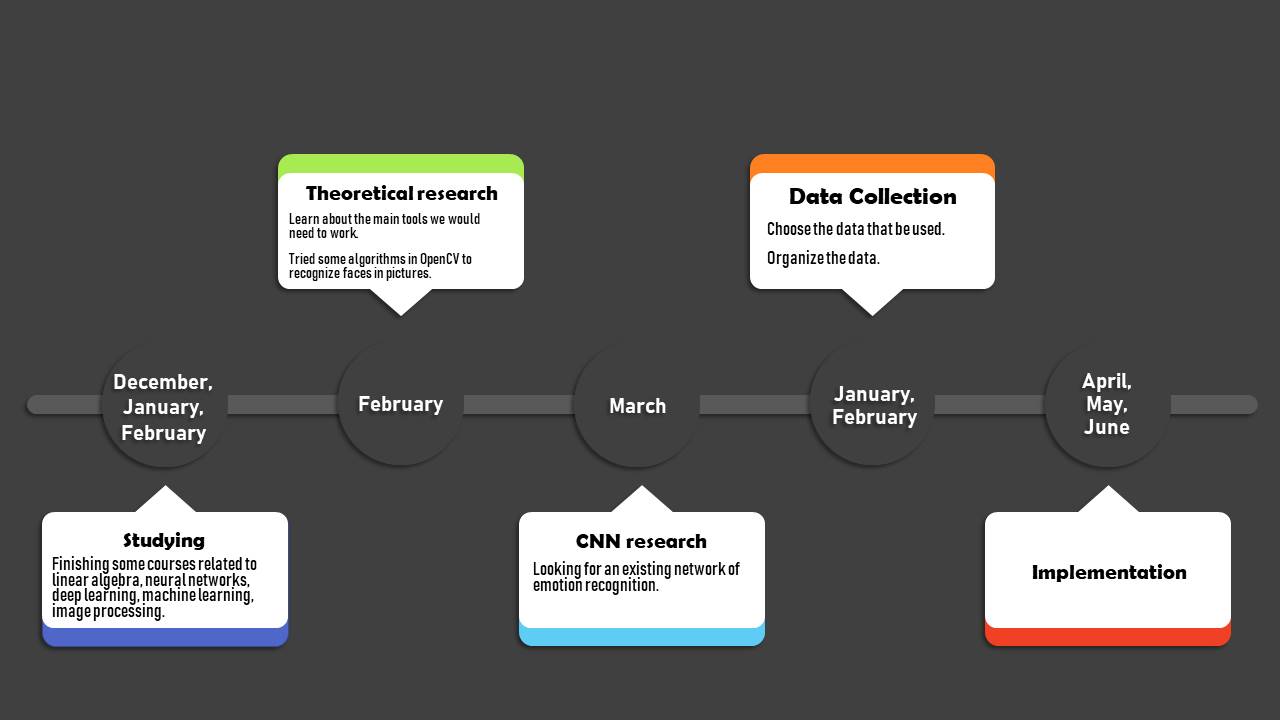
****

1. **feature extraction and classification and evaluation method**

**Before training classifiers, features must be extracted accordingly.**

****

**2.3 A**[**scope of work (SOW)**](https://www.projectmanager.com/blog/sow-to-project-plan)

****

**2.3.1 CNN research**

**We spend a long time on looking for an existing network of emotion recognition. Our plan was to find a network for testing results in a first time, without training. We found a lot of non trained networks. Occasionally, we found trained networks and tested them with labelled pictures of children that we found on internet. The experimental results were not enough satisfying, thus we continued our research. Finally, we found a performing network that detect faces and recognize emotions in videos: the face detection works with a model of haarcascades and the emotion detection works with a model trained on fer2013 database. We tried it on several examples of videos and it recognized pretty good the faces but not always the right emotion. Usually, happy emotion is recognized as neutral or sad as neutral, not surprising errors. We also tried it on a dataset of children pictures in good resolution and obtained pretty good results: 98% for happy, 55% for sad, 78% for neutral, 42% for angry, 35% for fear.**

**2.3.2 Theoretical research**

**When starting working on this project, I had to learn a lot of concepts about image processing and deep learning. I also had to learn about the main tools we would need to work with: opencv, haarcascades, tensorflow, keras.**

**Additionally, we had to get familiar with our working environments: we choose to use google colaboratory for our python project. After getting familiar with the concepts and the tools, we tried some algorithms in opencv to recognize faces in pictures and also in videos. This was our first result of face recognition in video:**

****

**Resources**

* **Hardware: Camera.**
* **Software: Datasets, Database server**

**Constraints:**

**Risks**

* **Errors in identification.**
* **Privacy.**
* **Misuse of data.**

**Design**

**3.1 Actor Goal list**

|  |  |
| --- | --- |
| Actor | Goal |
| * Administrator | * **Datasets Update** * **Validation Algorithms** |
| * User | * **Receives Analyzed Data** * **Studying Data** |
| * Technical | * **Maintain System** |

**3.2 Use case**

* **UC1: Datasets Update (Primary Actor [Administrator])**

**The administrator updates the data sets regularly whenever a child joins or leaves the institution.**

* **UC2: Validation Algorithms (Primary Actor [Administrator])**

**An administrator should check the algorithms to ensure they are working as planned.**

* **UC3: Receives Analyzed Data (Primary Actor [Users])**

**Receives the arranged data from the system.**

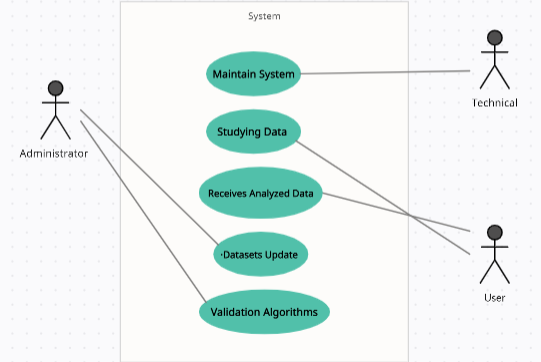
* **UC4: Study data to understand child (Primary Actor [Users])**

**He studies data and tries to absorb it to be used and exploited for the benefit of the child**

* **UC5: Maintain System (Primary Actor [Technical])**

**Ensure that there are no errors or lags in the application itself to reduce the risks from errors and delays, to make it more dependent on high performance.**

**3.3 Use case Diagram**



**Implementation**

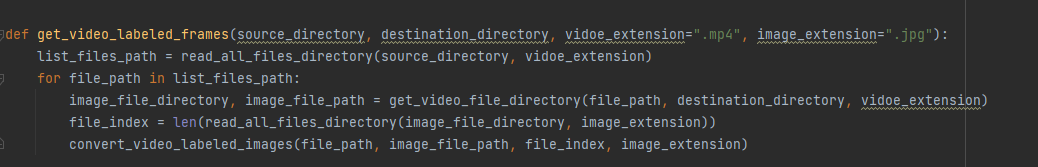
**4.1 Data Capturing**

**The first step that the program does is to capture the data from the source path in mp4 format and convert it to us in images jpg format**

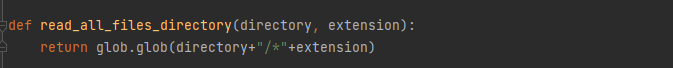
**And he will open it through the cascade in the opencv so that he can determine the face.**

****

**We read the videos from the source**

****

**Then build it in the list**

****

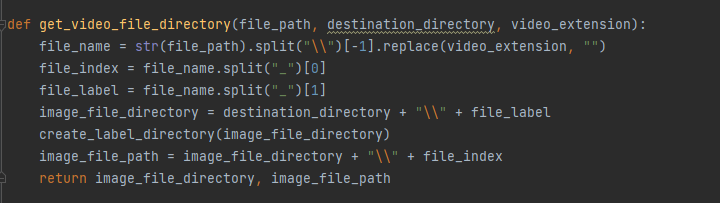
**children videos in a path called OriginalDataSet in this way:**

**C:\Users\Nesma\OneDrive\Pictures\Screenshots\2021-07-06 (9).png**

**Label class**

**We need to extract these two pieces of information So we cut the file name into two parts**

**label = file index class = file label**

****

**And I will start making a new directory for each video, in which the video cutter (frames) will be placed.**

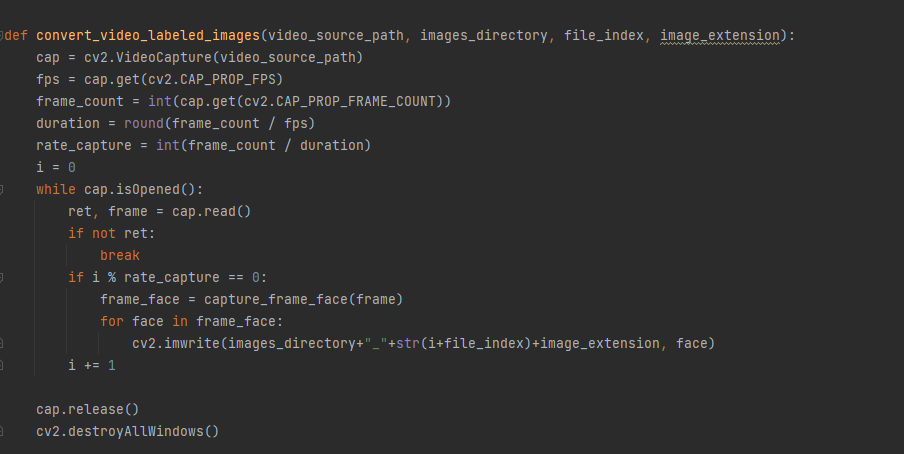
**Here a problem appears**

**In each video, I cannot take every frame from the video, for one second, for example, has 10 frames, and the face is fixed and does not give any reaction, so the data will be filled with great noise and the results will not be great**

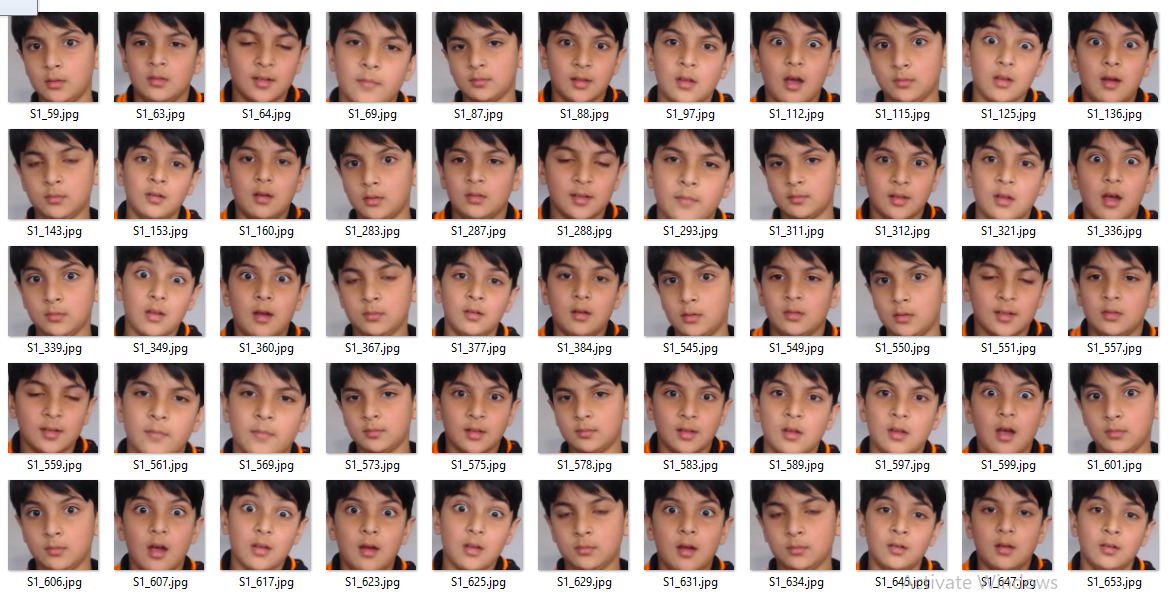
**To solve this we will use the time split, so that if the video is 3 seconds, I only need 6 frames of it..**

**Here another problem arises**

**The videos are short, and some of them come with the emoji after the first second, and some of them come after the second, and so on, we will also have about 4 noise frames.**

****

**And this is the noise**

****

**These frames are supposed to be from a video of the expression (surprise).**

**Here we address another problem with our model and a very important point:**

* **Balance Data**

**I am a strong believer in using extreme examples to prove points. In this case, I’d like to provide the following example to demonstrate why data balance matters in classification algorithms:**

**Here we are trying to predict the baby's facial expression, let's focus on the results**

**The value of (y) we are trying to predict from another reference which is the child's facial expression**

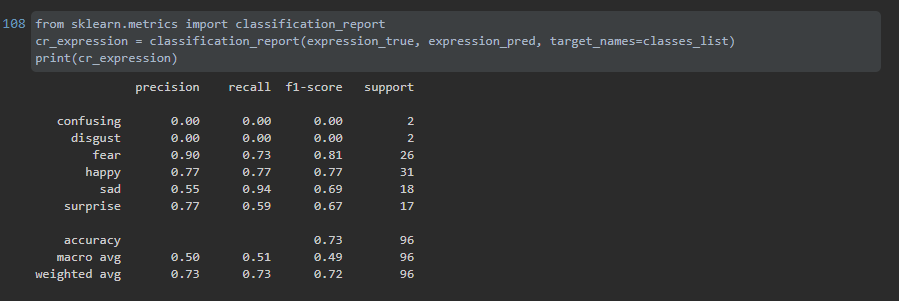
**And if we had 500 faces as data and 499 showed angry facial expressions while only one showed happiness**

**Will our model have 100% accuracy?**

**Actually no, it proves to be 99.99% accurate as 0.01% is due to a happy face**

**Well, the 499-to-1 distribution doesn't do the trick because we don't have the required level of data to understand what might lead to this error.**

**Therefore, balanced data gives us the same amount of information to help predict each category and thus gives a better idea of ​​how to respond to the test data, And this is what we have..**

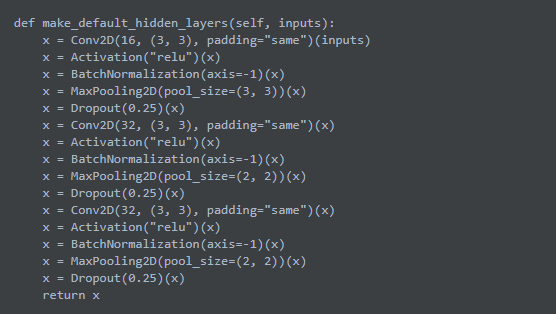
****

**4.2 Emotion Recognition**

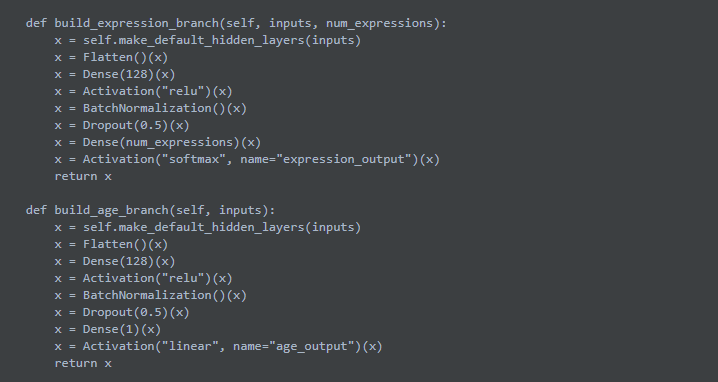
* **Model**

**For our emotion recognition algorithm, we used a CNN model named “expression\_net”**

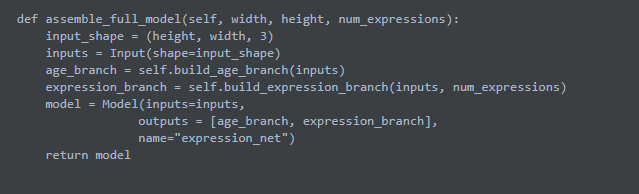
**1)**

****

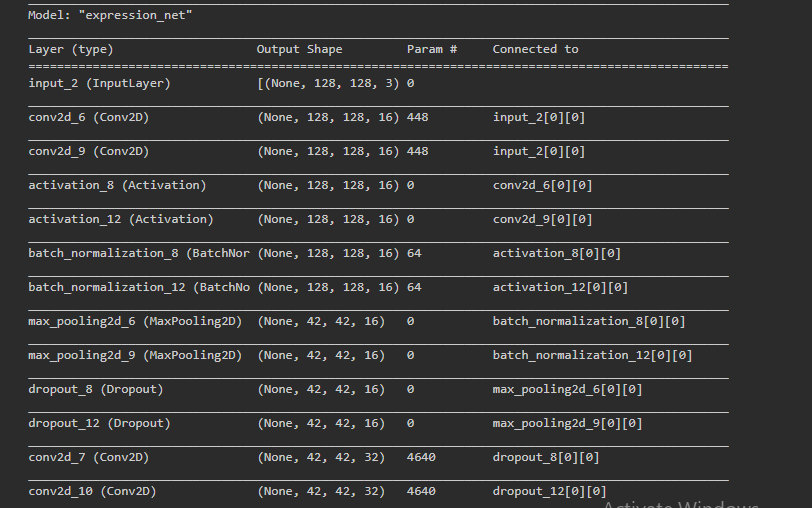
**2)**

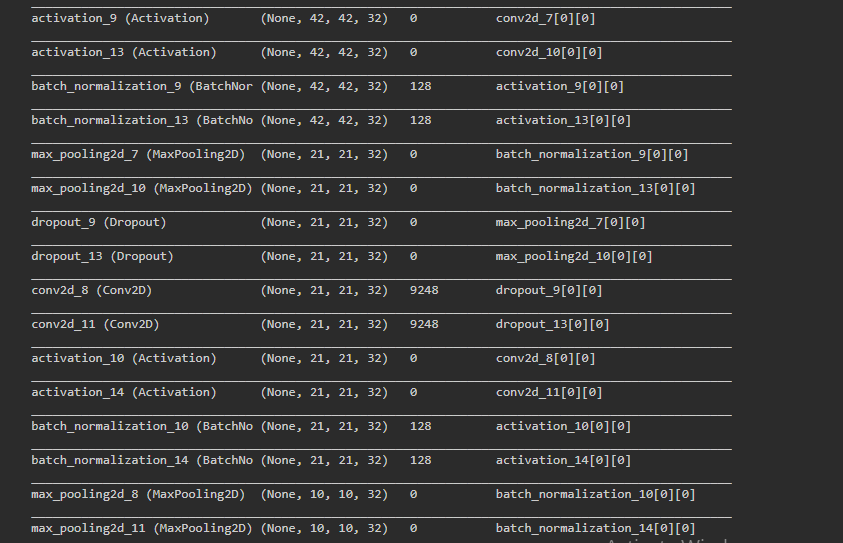


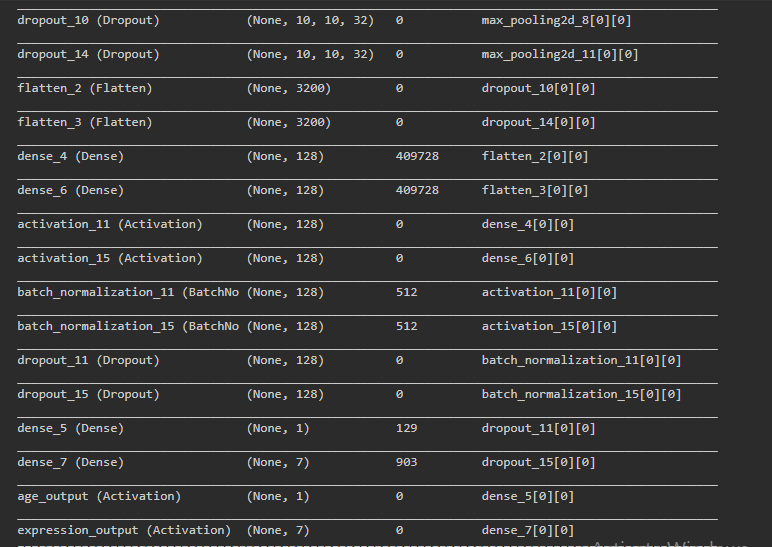
**3)**

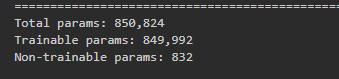


* **model.summary()**

****



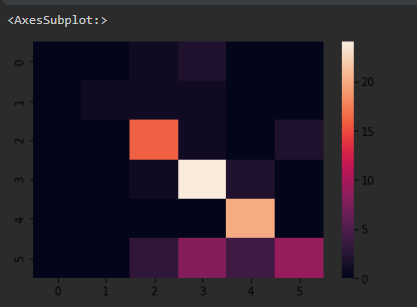




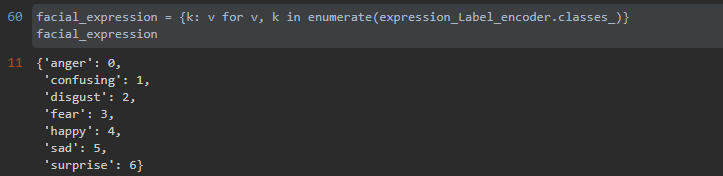
**The input of the network is an image with the following format: 128\*128\*3**

**It outputs a probability for each of the 7 following emotions : anger, sad, happy, disgust, fear, and surprise.**

**The results on the training set are the following:**

****

**Where:**

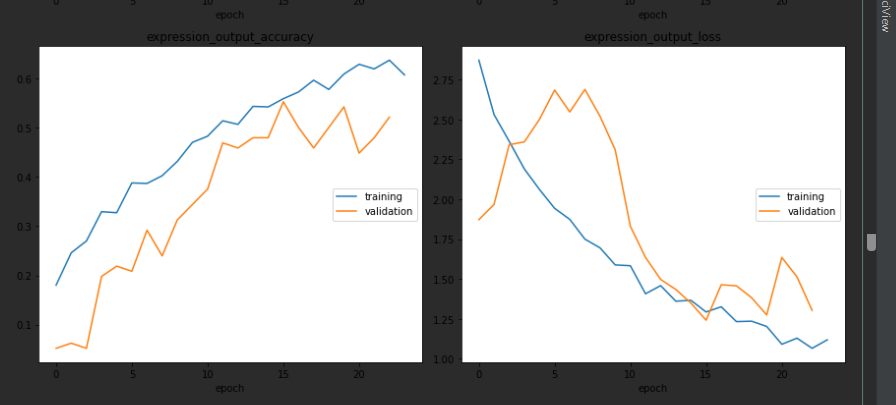
****

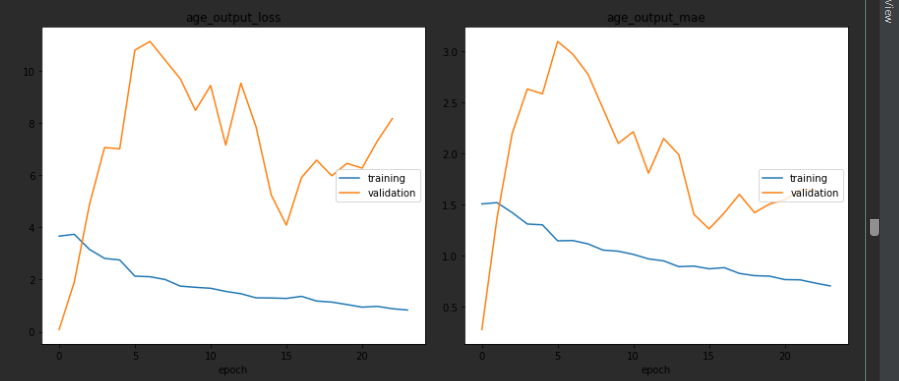
**4.3 Experimental framework and results**

**As discussed earlier, CNN requires large database to making it practical for different applications. This bottleneck is usually.**

**Transfer learning is a machine learning approach that focuses on ability to apply relevant knowledge from previous learning experiences to a different but related problem. We have used**

**transfer learning approach to built framework for expression recognition using proposed database as the size of our database is not sufficiently large to robustly train all layers of CNN from the very beginning.**



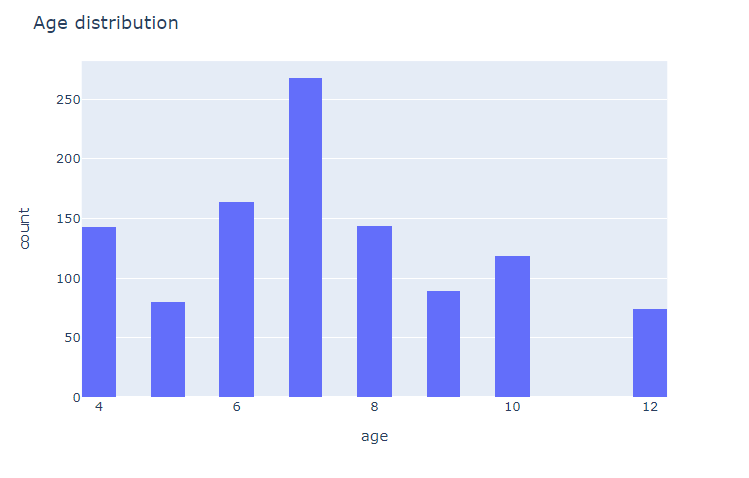


**4.4 Graphs**

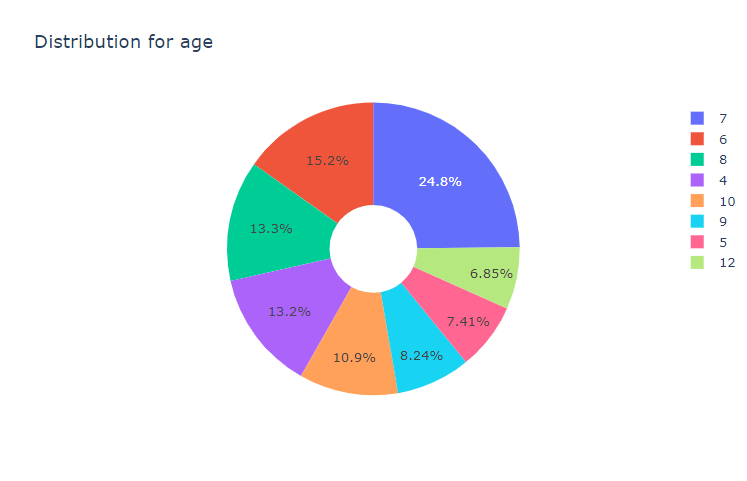
**One of the main goals in our project is to display the results in a meaningful way. Thus, we created graph of emotions and ages as function of time for each child.**

**The graph data is based on our saved sequences for each child.**

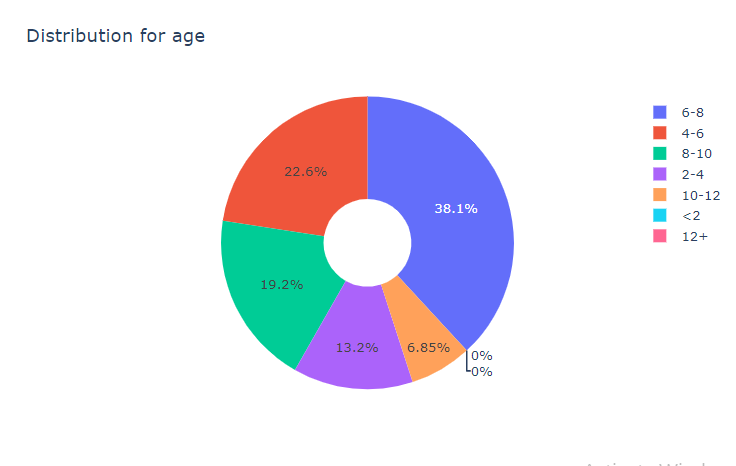
**At first, we created histogram graphs that look like the following graph:**



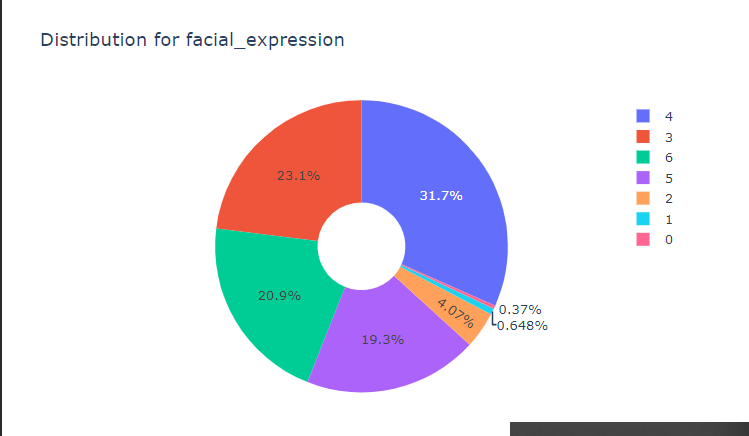
**Secondly, we decided to change the plotting display and create graphs that look like the following:**



**Also:**



**For emotions we have this graph:**



**Report**

**Conclusion**

**The goal of the project is to create an algorithm capable of detecting children's feelings in a video, We were able to achieve this goal, or part of it**

**We succeeded in recognizing the children in the video with an accuracy of 60%. Despite the lack of data and immaturity..**

**The success of emotion recognition is difficult to measure because it is so subjective. In fact, even humans recognize feelings with an accuracy of 70%.**

**But we do have some ideas about possible optimization solutions.**

**Future work**

**(Implementation)**

**In the coming period:**

* **We will search or collect a lot of data about children**
* **We will work on better algorithms and train more than other models**
* **Study the matter further**
* **Analyze the available data to extract information from it**