



## COMP 6721 Applied Artificial Intelligence

Professor René Witte

### Group Name:

NS 01

### Group Members:

Dara Rahmat Samii (40281972)

Numan Salim Shaikh (40266934)

hahab Amrollahibioki (40292670)

### GitHub Link:

[github.com/DaraSamii/A.I.education-Analytics](https://github.com/DaraSamii/A.I.education-Analytics)

October 2023

# Chapter 1

# Dataset

## 1.1 Facial Expression Recognition(FER) 2013 Dataset

The primary dataset employed in this project is the FER2013 dataset[1], a pivotal resource in the realm of facial emotion recognition. This dataset was meticulously curated by Pierre-Luc Carrier and Aaron Courville, constituting a vital element within their ongoing research endeavor. The FER2013 dataset stands out due to its comprehensive nature and standardized structure.

Comprising grayscale images, each measuring  $48 \times 48$  pixels, the FER2013 dataset boasts an impressive total of **32,298** individual samples. This substantial dataset is a rich source for training and evaluating emotion recognition models.

Notably, the FER2013 dataset features meticulous image preprocessing. It ensures that each facial image is automatically aligned, meticulously centering the face and maintaining a consistent scale. This preprocessing greatly facilitates the extraction of essential facial features and enhances the efficiency of training emotion recognition models.

The FER2013 dataset categorizes emotions into seven distinct classes, each representing a unique spectrum of emotional expressions. These emotions encompass:

1. Angry
2. Disgust
3. Fear
4. Happy
5. Sad
6. Surprise
7. Neutral

Table 1.1: Sample of FER2013 Dataset

emotion	pixels
0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121...
0	151 150 147 155 148 133 111 140 170 174 182 15...
2	231 212 156 164 174 138 161 173 182 200 106 38...
4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1...
6	4 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84...

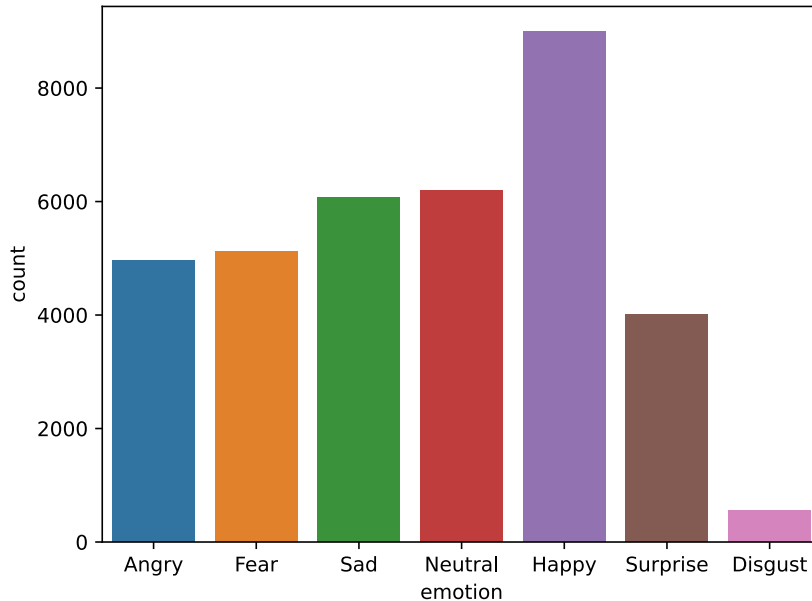


Figure 1.1: Number of images per each emotion in FER2013

## 1.2 FER+

In 2016 Emad Barsoum et al. in their paper "Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution" [2] used Crowd-sourcing and 10 taggers to label each input image of FER2013, and compared four different approaches to utilizing the multiple labels.

The dataset contained the original number of samples, the difference was that emotions mentioned in the previous section in addition to "Not Face" and "Unknown" Labels were voted by 10 taggers and the score of each label had been recorded.

We merged two datasets to create more accurate dataset which it the intensity of each emotions can be measured by the number of votes each label had received by the taggers. For the project, Neutral and Angry labels can be used directly. Based on the scores for each emotion in FER+ and the ratio two other labels of 'Bored/Tired' and 'Engaged/Focused' can be extracted which the methodology will be described in the coming sections.

In the year 2016, Emad Barsoum and colleagues, in their paper titled "Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution" [2], conducted a pioneering study that leveraged crowd-sourcing and engaged a team of ten taggers to label each input image within the FER2013 dataset.

This augmented dataset retained the original set of samples, with a notable enhancement: in addition to the emotions mentioned in the preceding section, namely "Not Face" and "Unknown," each of these emotions underwent a democratic evaluation by the ten taggers. The result was a meticulous recording of the score associated with each labeled emotion.

To enhance the dataset's accuracy, we combined the outcomes of two distinct datasets. The resulting amalgamation yielded a dataset wherein the intensity of each emotion could be precisely gauged based on the number of votes received by each label from the taggers.

Within the context of our project, the "Neutral" and "Angry" labels from this enhanced dataset can be directly employed. Moreover, by considering the scores attributed to each emotion within the FER+ dataset, we can derive two additional labels, specifically "Bored/Tired" and "Engaged/Focused." The subsequent sections will expound upon the methodology underpinning this extraction process in greater detail.

Table 1.2: Sample of FER+ merged with FER2013

emotion	pixels	neutral	happiness	surprise	sadness	anger	disgust	fear	contempt	unknown	NF
Angry	70 80 ...	4	0	0	1	3	2	0	0	0	0
Angry	151 150 ...	6	0	1	1	0	0	0	0	2	0
Fear	231 212 8 ...	5	0	0	3	1	0	0	0	1	0
Sad	24 32 ...	4	0	0	4	1	0	0	0	1	0

## Chapter 2

# Data Cleaning

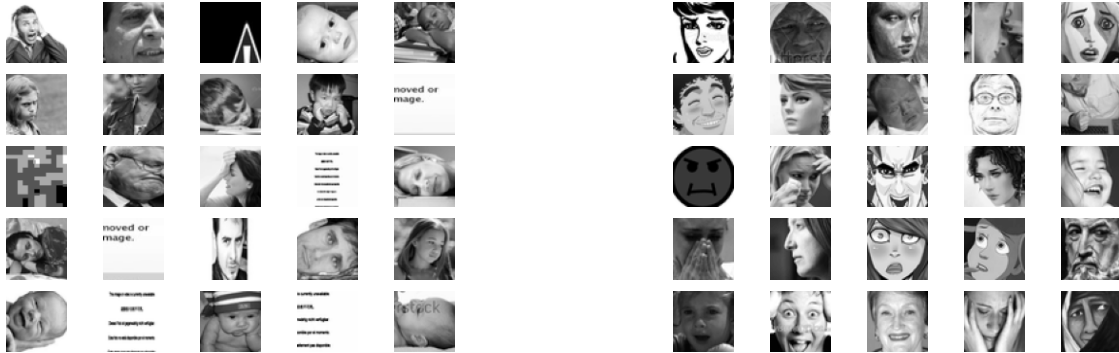
### 2.1 Eliminating Images Labeled as 'Not Face'

In the enhanced dataset, a subset of images was labeled as 'Not Face.' These images received various scores from the taggers, and their distribution is as follows:

- The number of images with a score  $NF = 10$  is 176.
- The number of images with a score  $NF = 4$  is 2.
- The number of images with a score  $NF = 2$  is 4.
- The number of images with a score  $NF = 1$  is 167.

A visual representation of some images labeled as 'Not Face' by the taggers can be observed in Fig. 2.1 .

Given that the total number of images with a non-zero NF score is relatively small in comparison to the overall dataset size, it was deemed prudent to remove all images with a non-zero NF score from the dataset.



(a) Images with  $NF=10$

(b) Images with  $NF=2$

Figure 2.1: Images with "Not Face" score of non-zero

### 2.2 Eliminating Images Labeled as 'Unknown'

A similar curation process must also be applied to images labeled as 'Unknown.' The distribution of 'unknown' scores for these images is detailed below:

- number of images with score unknown=8 is 3.
- number of images with score unknown=7 is 3.
- number of images with score unknown=6 is 18.
- number of images with score unknown=5 is 55.
- number of images with score unknown=4 is 224.
- number of images with score unknown=3 is 751.
- number of images with score unknown=2 is 2526.
- number of images with score unknown=1 is 8220.



(a) Images with unknown=6



(b) Images with unknown=5

Figure 2.2: Images with "unknown" score of non-zero

In view of the substantial number of images with non-zero 'Unknown' scores, a decision was reached after careful consideration. It was determined that images with a 'Unknown' score of 5 or greater should be excluded from the dataset. This action aims to maintain the dataset's integrity and coherence for subsequent analysis and machine learning model development.

For a visual representation of the images with non-zero 'unknown' scores, please refer to Figure 2.2, which provides insight into the selection criteria employed for image removal.

## Chapter 3

# Labeling

In the context of this project, it is imperative to classify images into four distinct labels: Anger, Neutral, Bored, and Focused. While the original dataset inherently includes the emotions Anger and Neutral, it lacks explicit representations for Bored and Focused emotions. This chapter elucidates the methodology employed to extract these target labels from the original emotion scores.

### 3.1 Anger

The process of labeling an image as 'Angry' based on the emotion scores is relatively straightforward. An image is labeled as 'Angry' if the score for anger surpasses the scores of all other emotions, signifying it as the predominant emotion. However, to account for rare cases where all emotion scores are equal, an additional criterion is established: the score for anger must be at least 2, ensuring a minimal threshold for labeling an image as 'Angry.'

### 3.2 Neutral

The labeling of an image as 'Neutral' hinges on the emotion scores, with 'Neutral' being assigned if the score for neutrality holds the highest value among all the emotions.

### 3.3 Focused

The criteria for labeling an image as 'Focused' were derived from a comprehensive analysis of facial expressions that convey focus. Notably, these expressions exhibit a lack of sadness, and the neutrality score should outweigh the scores of other emotions.

### 3.4 Bored

The criteria for labeling an image as 'Bored' are rooted in the observation that when individuals are bored, they typically do not exhibit happiness. Furthermore, boredom is unaccompanied by fear. In cases where a person may appear slightly angry, the criterion is that the level of anger should be lower than the expressed level of sadness in the facial expression.

Table 3.1: Summary of criterion's for labeling images based on emotion's scores

Label	criteria
Anger	<ul style="list-style-type: none"> <li>• Anger score <math>&gt;</math> other emotion's score</li> <li>• Anger score <math>&gt; 2</math></li> </ul>
Neutral	<ul style="list-style-type: none"> <li>• Neutral score <math>&gt;</math> other emotion's score</li> </ul>
Focused	<ul style="list-style-type: none"> <li>• sadness score <math>== 0</math></li> <li>• neutral score <math>&gt;</math> other emotion's score</li> </ul>
Bored	<ul style="list-style-type: none"> <li>• happiness score <math>== 0</math></li> <li>• fear score <math>== 0</math></li> <li>• anger score <math>&lt;</math> sadness score</li> </ul>



## Chapter 4

# Dataset Visualization

After meticulous data cleaning, removal of irrelevant data, and the extraction of the desired emotion labels, the dataset is now composed of the following:

- Number of samples labeled as 'Neutral': 6184
- Number of samples labeled as 'Angry': 3954
- Number of samples labeled as 'Focused': 5453
- Number of samples labeled as 'Bored': 3168

The bar-plot of the plot's per emotion is shown in Fig. [4.1](#)

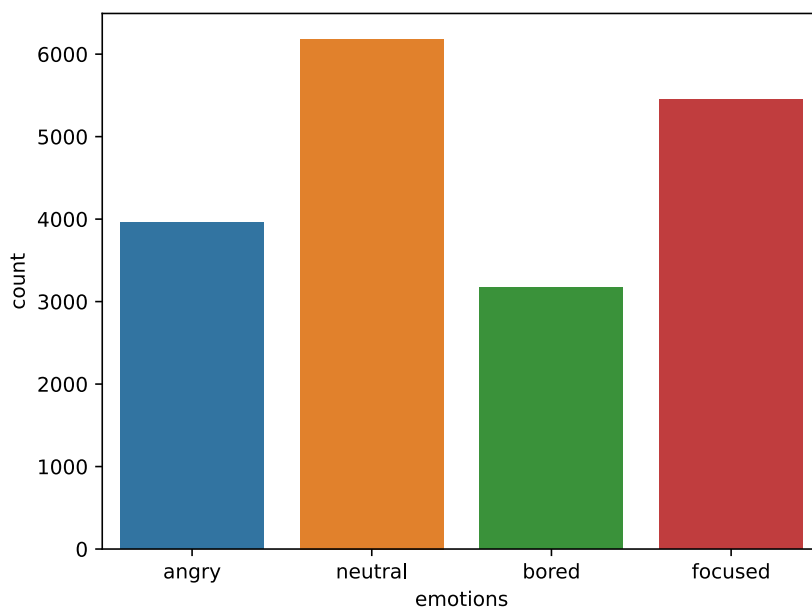
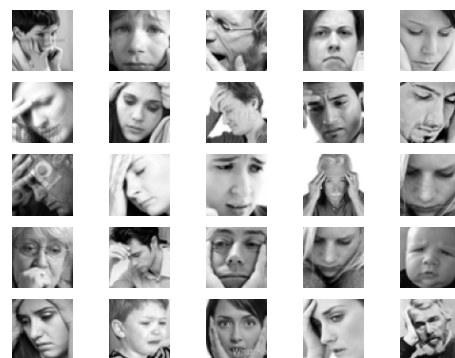


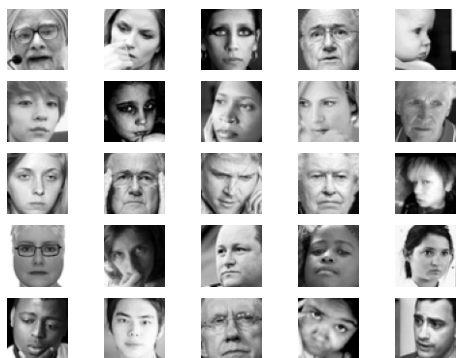
Figure 4.1: Bar-plot of images per desired emotion



(a) Samples of images labeled as 'Angry'



(b) Samples of images labeled as 'Bored'



(c) Samples of images labeled as 'Neutral'



(d) Samples of images labeled as 'Focused'

Figure 4.2: Samples of the cleaned dataset

# Bibliography

- [1] W. C. Y. B. Dumitru, Ian Goodfellow, “Challenges in representation learning: Facial expression recognition challenge,” 2013.
- [2] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, “Training deep networks for facial expression recognition with crowd-sourced label distribution,” in *Proceedings of the 18th ACM international conference on multimodal interaction*, pp. 279–283, 2016.