CS535 HW1 Report

1. To calculate the Fleiss' kappa, we first compute the P_i for each row = $\frac{1}{n(n+1)}\sum(n^2_{ij} - n)$, then average the P_i to get $\overline{P}$ by averaging all P_i. Then we compute $p_j = \frac{1}{Nn}\sum n_{ij}$ and $\overline{P}_e = \sum p_j^2$. The result of kappa score = 0.480107293222964. According to the table, it's a Moderate agreement. We verify it by using the built function from pytorch resulting in a kappa score = 0.4803 which is very close to our calculation.

2.

	Visual	Audio
Happy	Smile	Rising tone
Sad	Frown	Low voice
Angry	Widen eyes	Loud voice

3. We annotate 20 videos from using the happy clue = 'smile' and sad clue = 'low voice' and save the result in a 3*40 table. We extract the ground truth label from the filename. We propose our Hypothesis for happy and sad clue and choose $\alpha = 0.05$:
H0: there is no significant relationship between smile and happy
H1: there is a significant relationship between smile and happy
H0: there is no significant relationship between low voice and sad
H1: there is a significant relationship between low voice and sad

To perform the Chi-square test, create a contingency_table between happy or sad(Y/N) and (ANG, HAP, SAD) using scipy.stats.chi2_contingency. We perform the test on the table and we get the results as follows:

Happy:
contingency_table :
HAPPY N Y
truth
HAP 0 10
SAD 9 1

chi2 statistic: 12.929
p-value: 0.00032348
degrees of freedom: 1

expected frequencies:
[[4.5 5.5]
[4.5 5.5]]

Sad:
contingency_table :
SAD N Y
truth
HAP 10 0
SAD 1 9

chi2 statistic: 12.929
p-value: 0.00032348
degrees of freedom: 1
expected frequencies:
[[5.5 4.5]
[5.5 4.5]]

We found that both p-value for happy and sad clue = 0.00032348 which is < 0.05 and extremely close to 0. This means that we have a very high level to reject the null hypothesis that there is no significant relationship between the clues. Thus smiling is a significant clue to identify happiness and lowering voice to identify sadness