Title: “Exploring the Effect of Model Compression on Fairness in Facial Expression Recognition”

1. **Abstract**
2. **Introduction** 
   1. Motivation
   2. Goals
      1. RQ1: How effective are traditional compression method in the context of FER? Can it reduce model size while preserving reasonable accuracy?
      2. RQ2: Do compression techniques tend to amplify or mitigate biases? Or not change them at all?
      3. RQ3: How do compression techniques compare in terms of effectiveness (reducing model size and keeping accuracy high) and effect on fairness?
   3. Contributions/Deliverables
      1. **HOW DO WE EXTEND EXISTING WORK?**
3. **Background** 
   1. **Facial Expression Recognition** (FER)
   2. **Fairness** (mention the problems with fairness in facial analysis, etc.). What metrics are there?
   3. **Model Compression**
      1. Post-training quantisation
      2. Pruning
      3. Weight Clustering
      4. Impact on fairness (quote the recent studies)
4. Methodology (high-level overview of we want to do, specifics in next chapter - Implementation)
   1. **OVERVIEW:** We want to: Set up a *baseline* model performing FER. Then we want to implement compression techniques on top of it. We will then run those on a **labelled dataset** of human images – labelled in terms of the gender of the participants. And we will be interested to see the overall effect that individual compression techniques had on the baseline – did they reduce the size of the model? Did they improve/worsen overall accuracy? Did they amplify biases (as claimed by previous studies) or did they mitigate them? Was the effect of compression identical across compression techniques or did it vary across strategies? **(WE ANSWER THE RQ-S FROM SECTION 2.2)**
   2. **METRICS:** We want to compare the performance of the baseline vs the performance of the compressed models across three metrics:
      1. Accuracy – standard accuracy but also across gender (protected attribute).
         1. Overall accuracy
         2. Male vs female accuracy
         3. Balanced male accuracy vs balanced female accuracy (**WHY – DIFFERENT BALANCE 🡪 SHOW THE DATASET DISTRIBUTION ACROSS GENDERS**).
      2. Model size – how effective was the compression technique? We measure that by the size of the resulting model (after compression – i.e. how much space it would take on disk – that’s important for resource-constrained devices).
      3. Fairness – did compression the discrepancy between classifying male vs female subjects? HERE WE USE **PREDICTIVE PARITY.**
5. **Implementation**
   1. Dataset
      1. CK+ - what is it and how did you choose it?
      2. How did you prepare it?
      3. How did you label it? (Masculine vs feminine, and not male vs female – but for **brevity**, I will be referring to the subjects as male and female, referring to their appearance and not necessarily to their gender orientation).
   2. FER Classifier
      1. That is our *baseline*.
      2. Selected it from this Coursera course.
      3. Inspired by this paper ...
   3. Compression – TFLite
6. Evaluation
   1. Baseline Performance
   2. Tuning Compression Hyperparameters
   3. Performance After Compression – include the tables with the results here.