

The Data Barrier to Lightweight Drinking Detection

An Analysis of the Viability of Skeleton-Only Models on In-the-Wild Social Data

Joelle Tijssens (J.D.Tijssens@student.tudelft.nl) - TU Delft Bachelor Student CSE - Supervisors: Hayley Hung, Litian Li, Stephanie Tan

Background

- **Goal** Lightweight Drinking Detection on REAL Data
- **Reason** Medicine, Quality of Life, Human Computer Interaction, Marketing!
- **How:** Compare 2 Different Lightweight Strategies: Random forest and Temporal Convolutional Neural Net

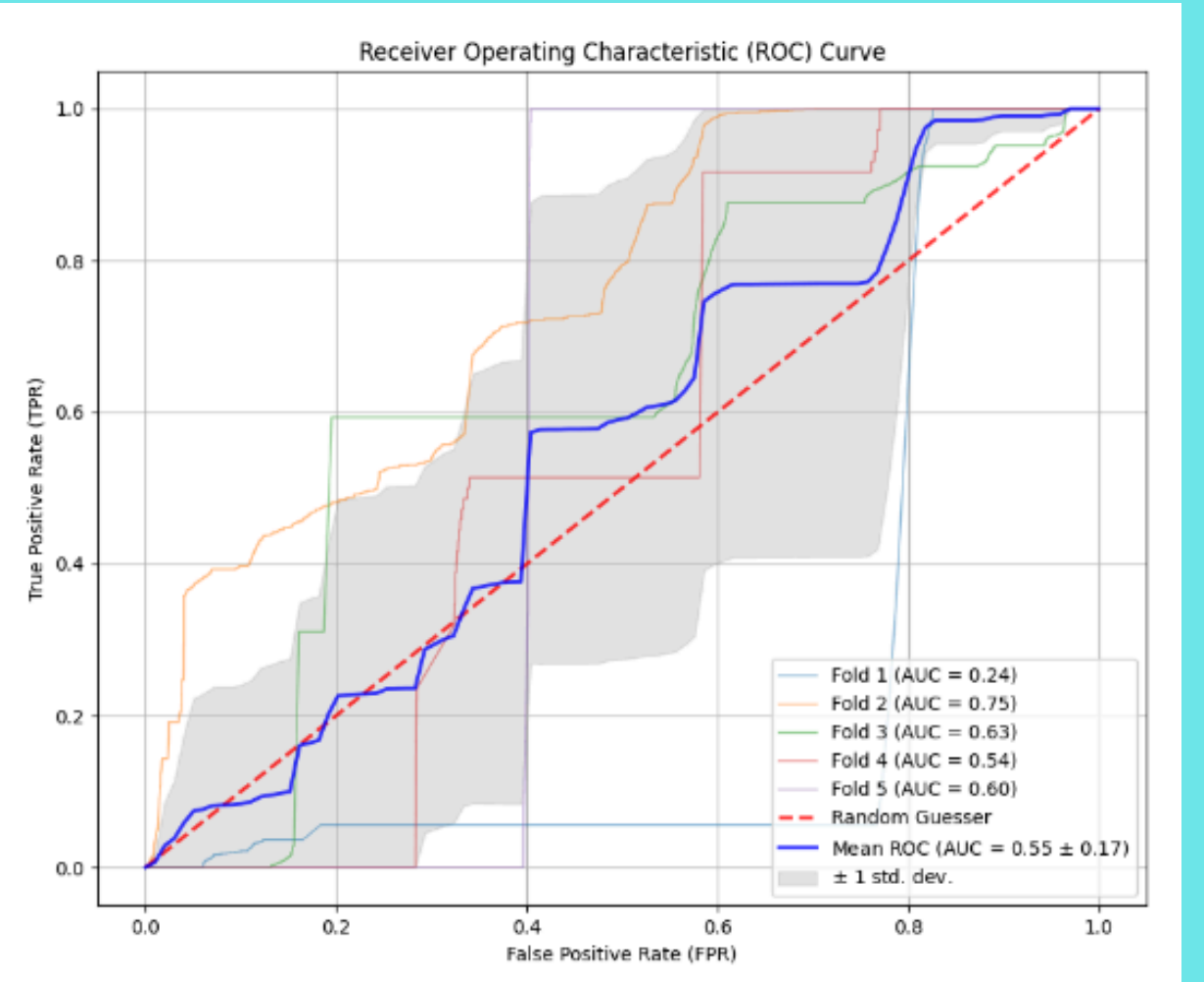
Can skeleton-only data reliably be used to **detect the drinking action in uncontrolled, human social settings in real-time** applications?

- 1 How well do lightweight models perform on sparse, "in-the-wild" data?
- 2 Which actions are most commonly confused with drinking?
- 3 Under which visual conditions (like occlusion and camera angle) do the models fail?

Question 1:
What performance
can we achieve?

Not enough
with **this approach**

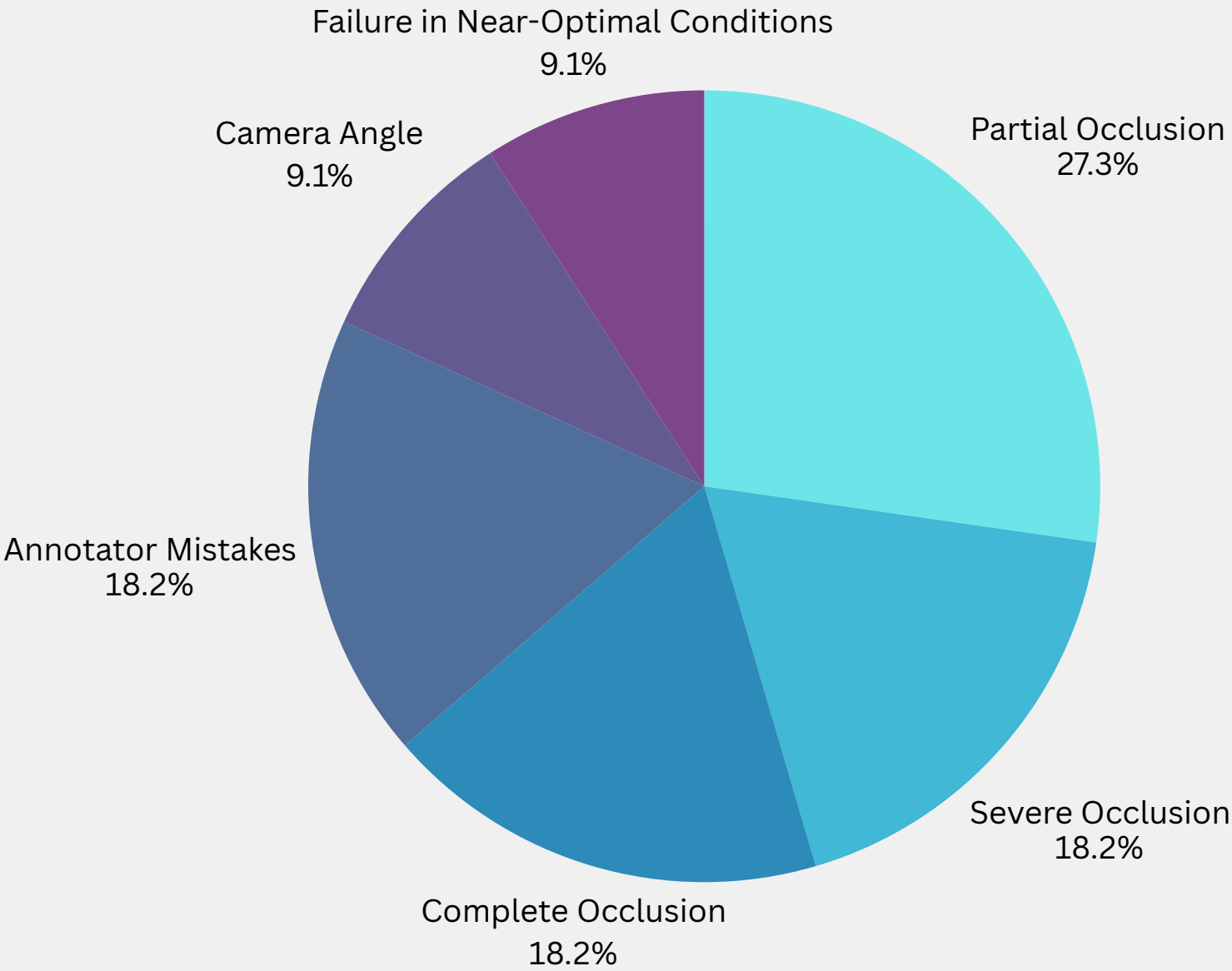
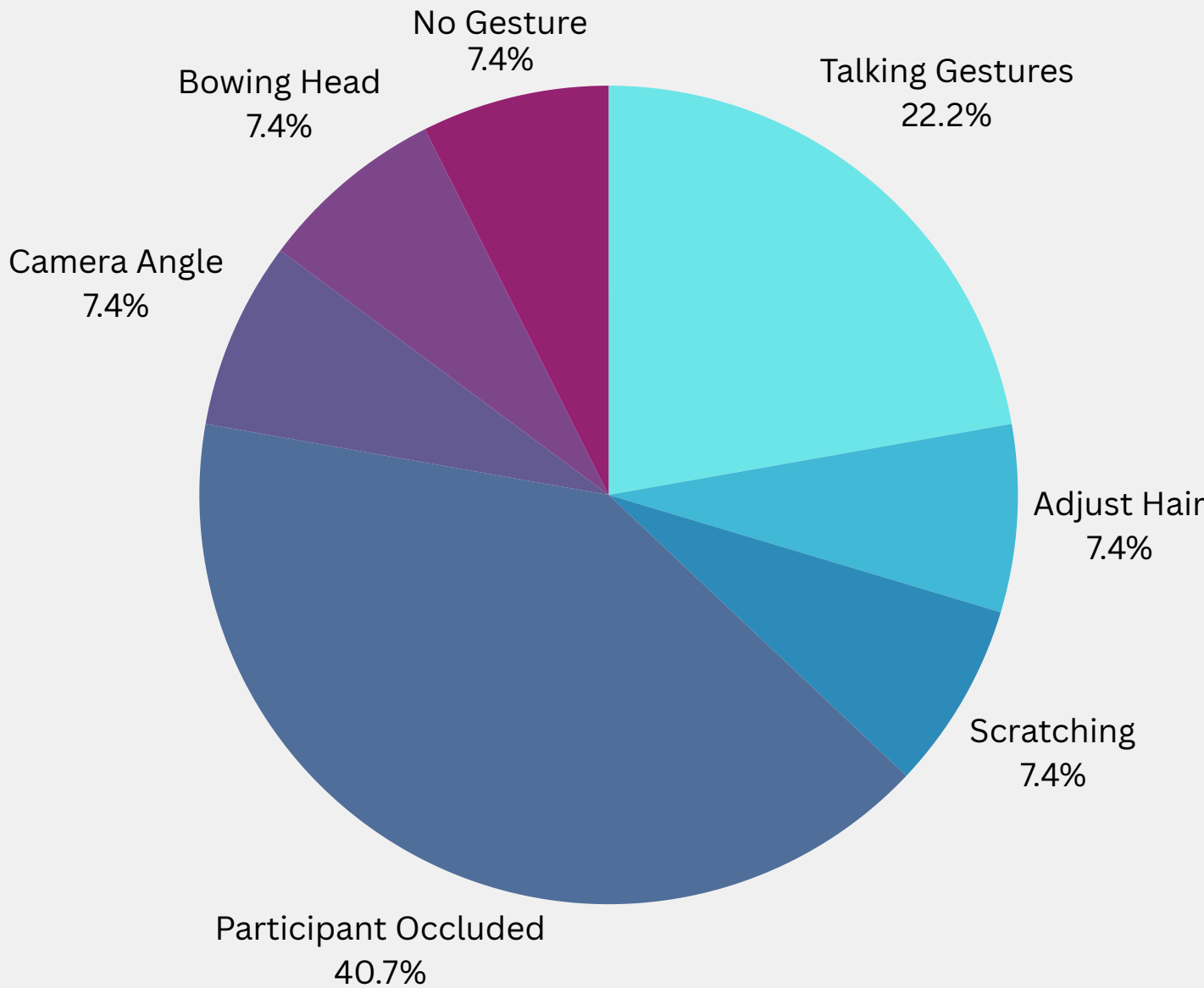
CNN
(best performing)
ROC Curve
per Fold →



Question 2:

Actions
leading to
false positives

Metric	Score
ROC AUC	0.6391
Accuracy	0.9427
Precision	0.0512
Recall	0.1730
F1-Score	0.0790



Question 3:
visual conditions
leading to
false negatives

Conclusions

- Not enough **data** (~40 instances of y=1)
- Poor performance, but all is **not** lost!
- **Better than random** for both models
- Training from scratch not viable for **sparse, messy real world data**
- highlights the **gap** between controlled environments and messy, real-world scenarios → **occlusion, camera angles**

Limitations & Future work

- Better feature extraction!
- Model performance suggests **transfer learning** should work
- e.g. MMAction2
- Cross-Dataset Validation