

Nkululeko: Machine Learning Experiments on Speaker Characteristics Without Programming

Felix Burkhardt^{1,2}, Florian Eyben¹, Björn Schuller^{1,3,4}

¹audEERING GmbH, Germany, ²Technical University of Berlin, Germany, ³Chair EHW, Universität Augsburg, Germany, ⁴GLAM, Imperial College London, UK

Introduction & Summary

- A tool to do machine learning without the need to program
- Focused on speech, but works generally on audio
- Opensource project, 100% Python: github.com/felixbur/nkululeko
- Targets combinations of feature sets and machine learners
- Uses configuration file templates

Example config

```
[EXP]
root = ./
name = exp_emodb
[DATA]
databases = ['emodb']
emodb = ./emodb/
emodb.split_strategy = speaker_split
emodb.test_split = 30
target = emotion
labels = ['anger', 'boredom', 'neutral', 'sadness']
[FEATS]
type = ['os'] # opensmile features
[MODEL]
type = xgb # xg-boost classifier
```

Nkululeko is organized in modules:

- **explore**: data exploration
- **nkululeko**: machine learning
- **augment**: data augmentation
- **demo**: model demoing

Motivation

- Empiricists sometimes struggle with programming, teaching students
- Getting quick overview on databases

Data exploration

Sample distribution

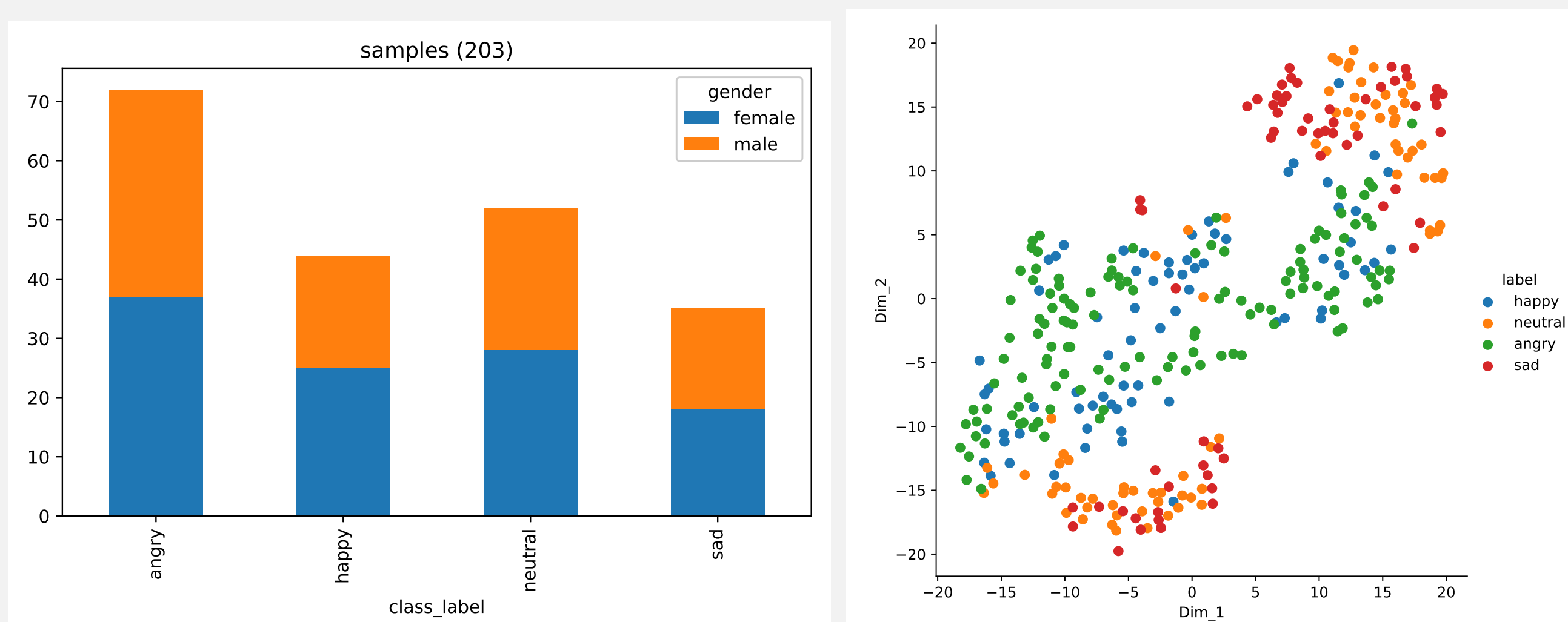


Figure: Emotion distribution per gender and t-SNE plot of features

Feature exploration

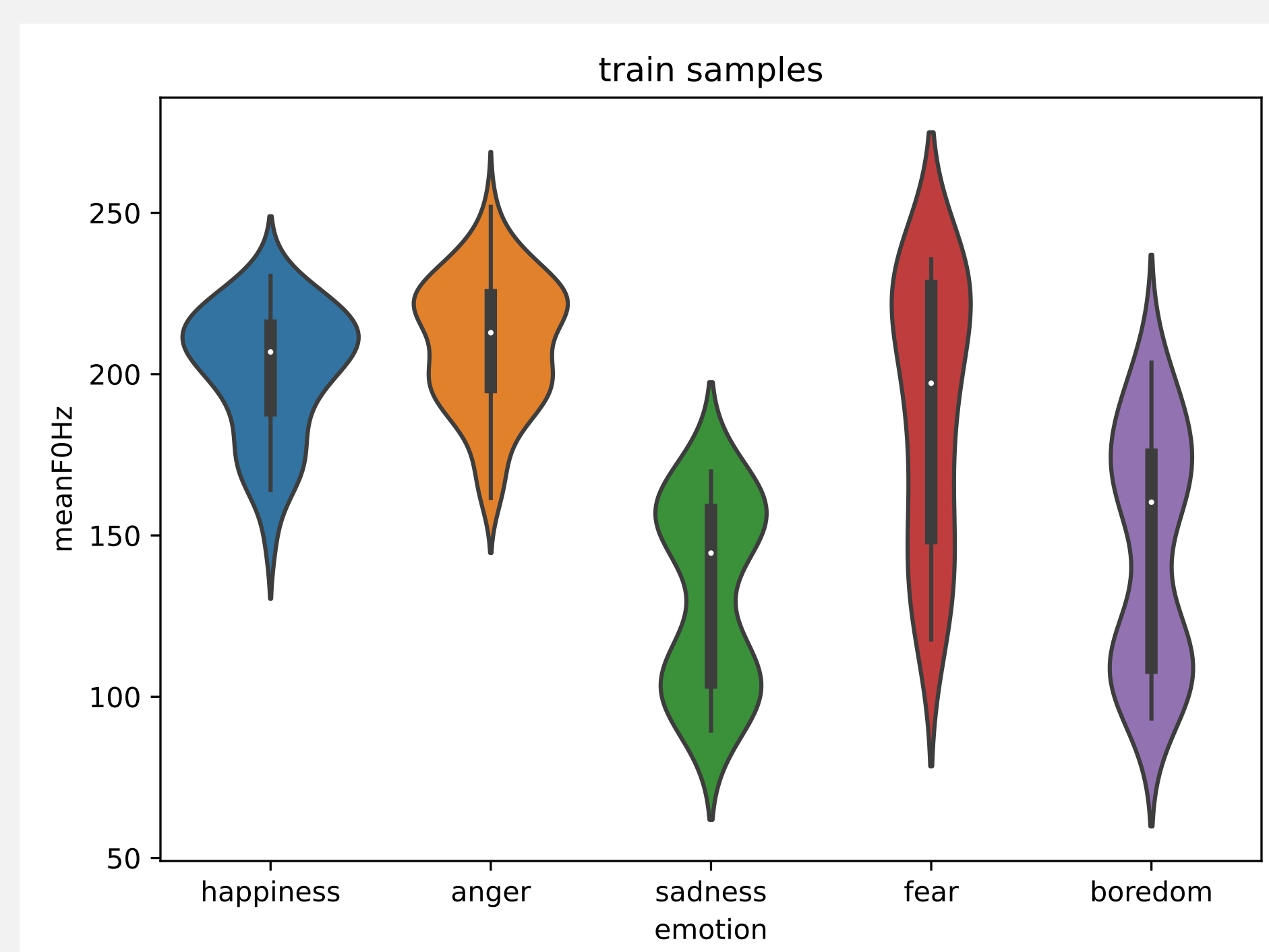


Figure: Violinplots for selected features

Learners and features

Features

Expert features: opensmile, Praat
Embeddings: Wave2vec2.0, TRILL, CLAP

Learners SVM, SVR, K-nn, Tree, XGB, XGR, GMM, MLP

Result presentation

The main format of visualizing Nkululeko experiment outcomes is the confusion matrix.

Confusion matrix

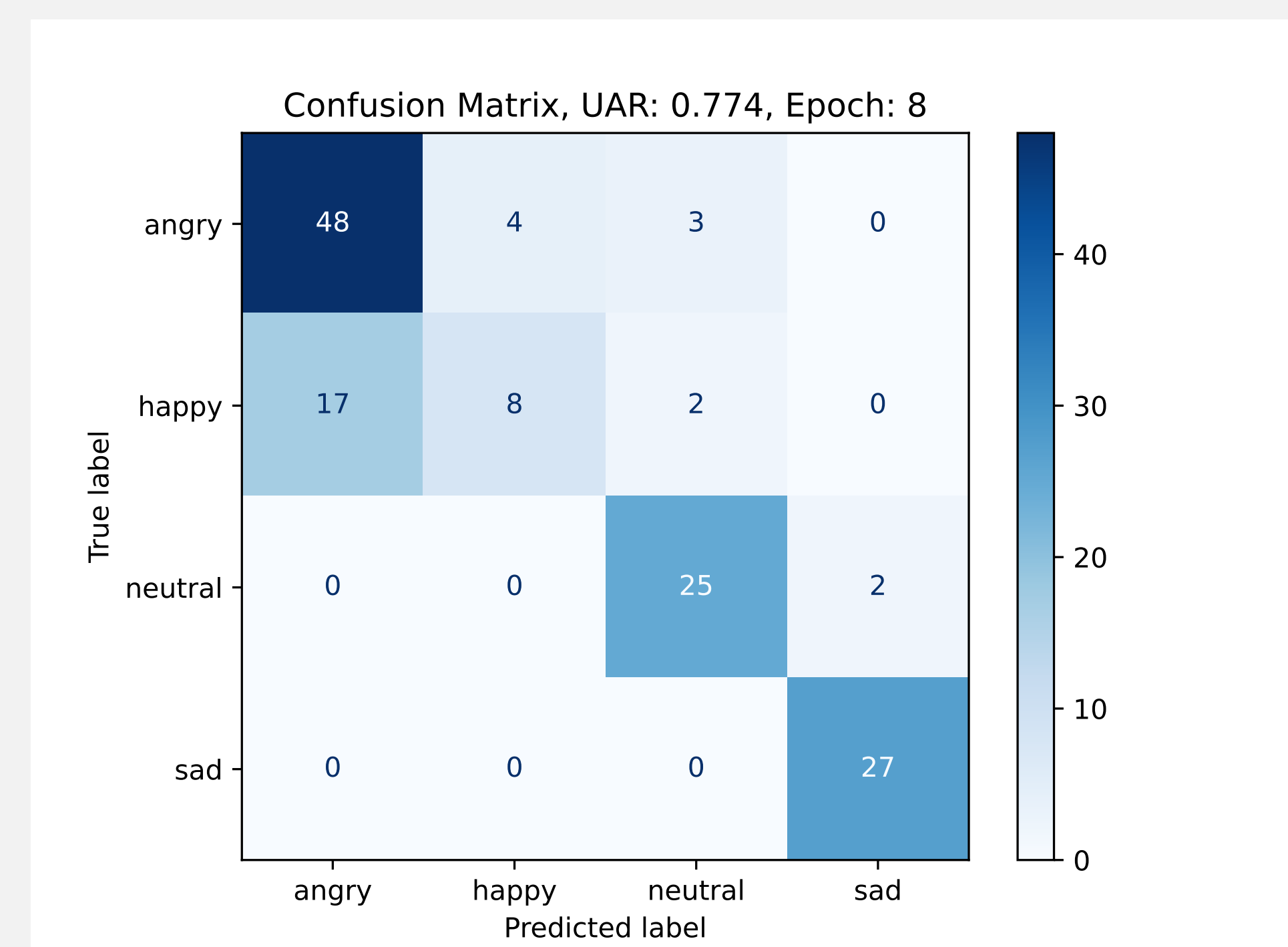


Figure: Confusion matrix per emotion

Epoch progression

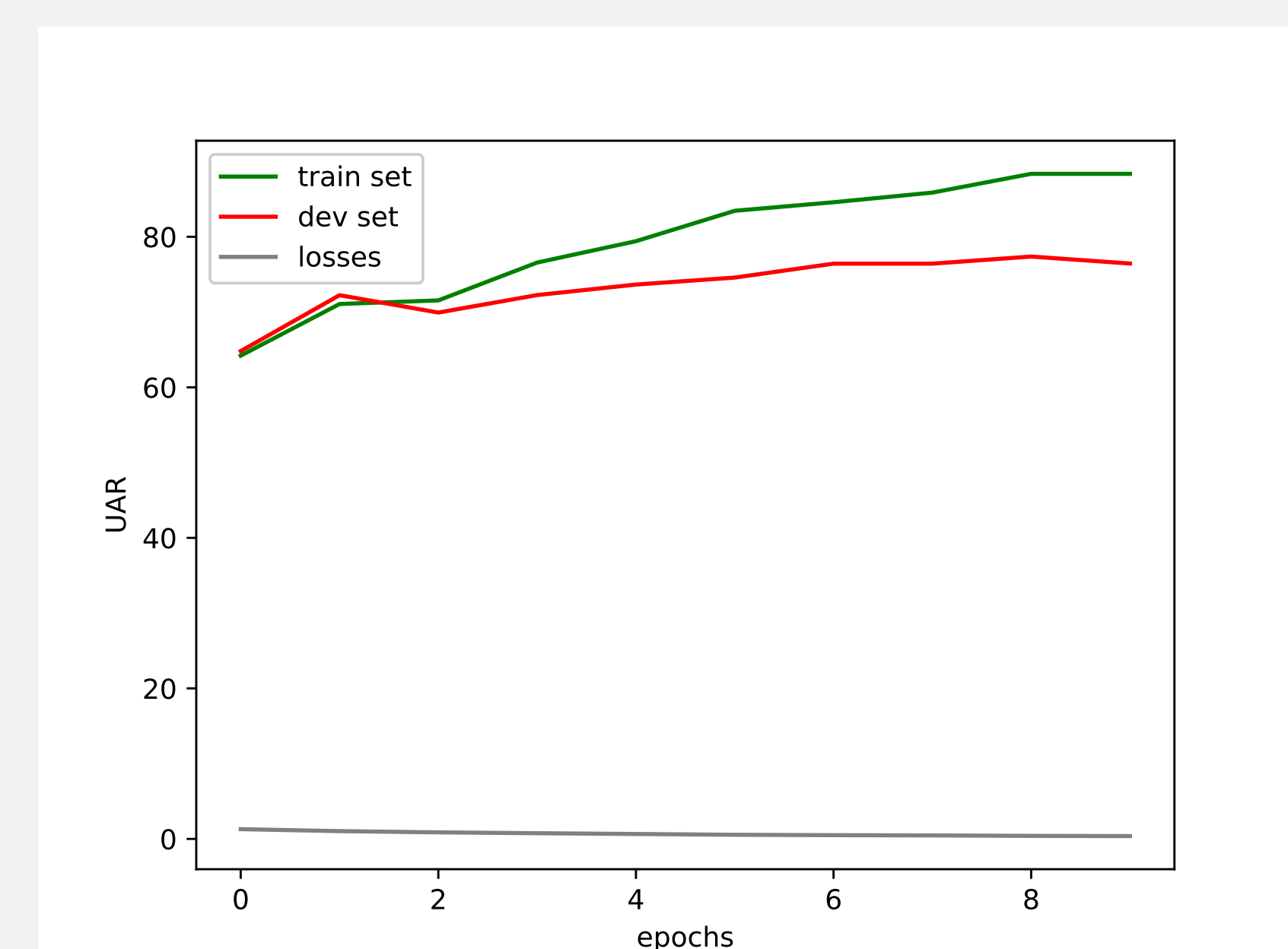


Figure: Progression of unweighted average recall per epoch

Data import

Data can be imported with a simple csv format:

```
file, speaker, age, emotion
f1.wav, mark, 34, happy
...
f123.wav, lisa, 21, sad
```

Acknowledgements

This research has been partly funded by the European EASIER (Intelligent Automatic Sign Language Translation) project (Grant Agreement number: 101016982), the European SHIFT (Metamorphosis of cultural Heritage Into augmented hypermedia assets For enhanced accessibility and inclusion) project (Grant Agreement number: 101060660) as well as the European MARVEL (Multimodal Extreme Scale Data Analytics for Smart Cities Environments) project (Grant Agreement ID: 975337).

