

Nkululeko

a form based speech machine learning tool

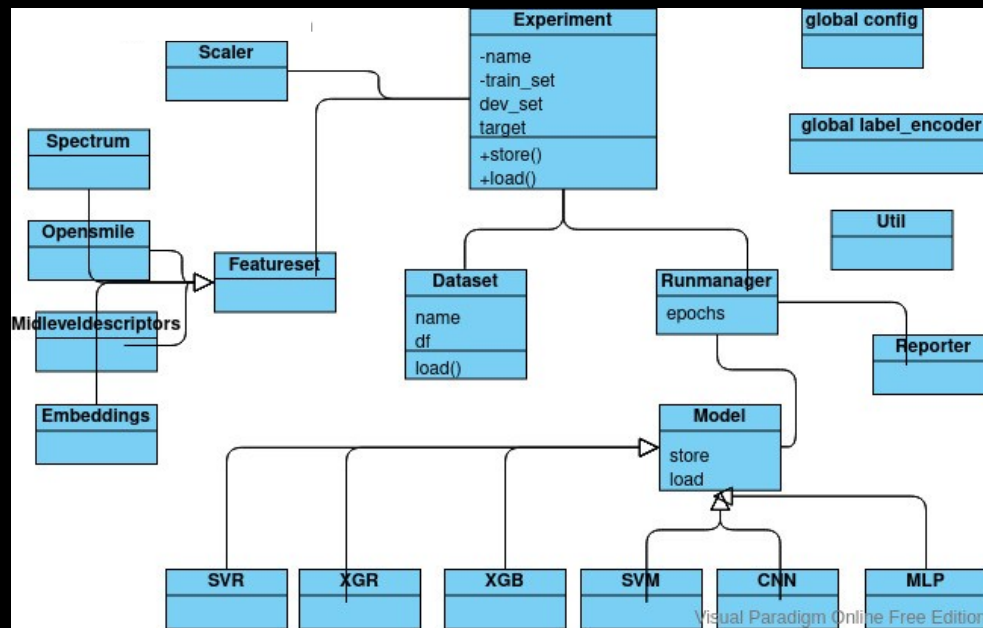
Felix Burkhardt

outline

- what is Nkululeko
- how to use it
- example experiments

what is Nkululeko?

- a software written in Python hosted on github*
- a tool to do machine learning (ML) experiments WITHOUT the need to program yourself
- focused on combinations of features and machine learners
- uses configuration file templates described in a blog**



* <https://github.com/felixbur/nkululeko>

** blog.syntheticspeech.de/?s=nkululeko

motivation

- with the success of Deep Learning, machine learning dominates science
- empiricists sometimes struggle with programming
- teaching students
- re-use of code

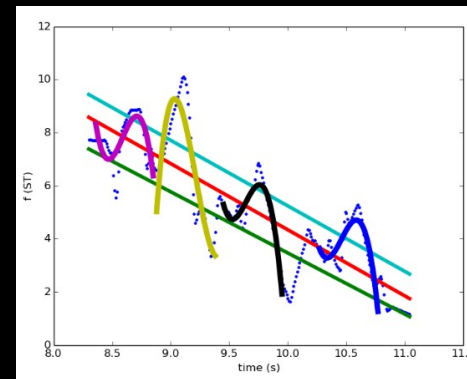
features

Three kinds:

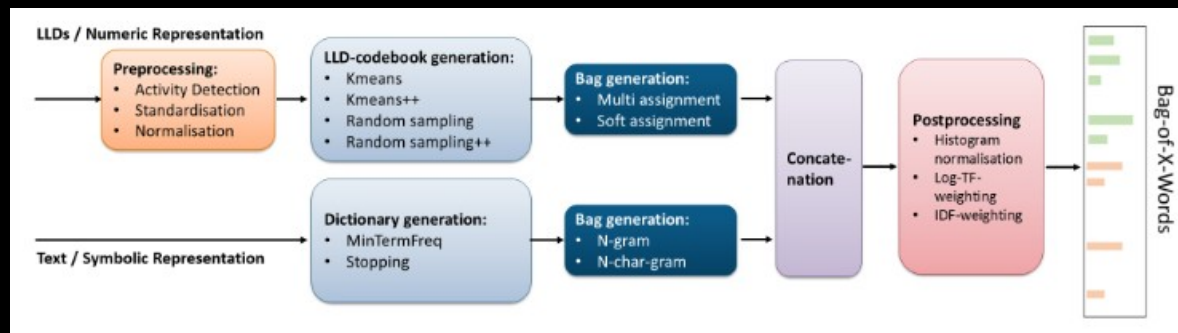
- expert features
- brute force (needs feature selection)
- learned
 - autoencoders
 - embeddings, latent space

features

- opensmile [1]
 - (e)GeMAPS – 62/88
 - Compare16 - 6,373
- mid level descriptors [2]
- openXbow [3]



img src: [2]



img src: [3]

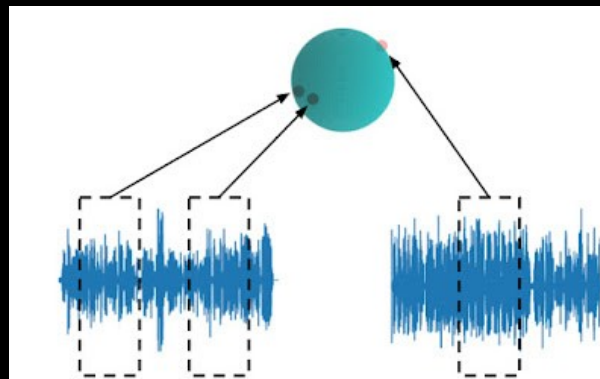
[1] Eyben, F., Wöllmer, M. and Schuller, B. (2010). opensmile – the munich versatile and fast open-source audio feature extractor.

[2] Reichel, U., Triantafyllopoulos, A., Oates, C., Huber, S., and Schuller, B. (2020). Spoken language identification by means of acoustic mid-level descriptors

[3] Schmitt, M. and Schuller, B. (2017). openxbow - introducing the passau open-source crossmodal bag-of-words toolkit.

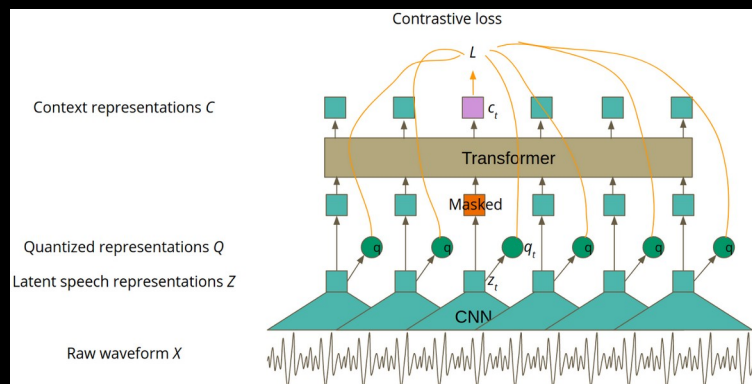
features cont.

- Logmel specs
- TRILL [1]
- Wav2vec 2.0 [2]



TRILL

img src: <https://ai.googleblog.com/2020/06/improving-speech-representations-and.html>



Wav2Vec 2.0

img src: <https://towardsdatascience.com/wav2vec-2-0-a-framework-for-self-supervised-learning-of-speech-representations-7d3728688cae>

[1] Shor, J., Jansen, A., Maor, R., Lang, O., Quitry, F., Tagliasacchi, M., Tuval, O., Shavitt, I., Emanuel, D., and Haviv, Y. (2020).

Towards learning a universal non-semantic representation of speech

[2] Baevski, A., Zhou, Y., Mohamed, A., and Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representation

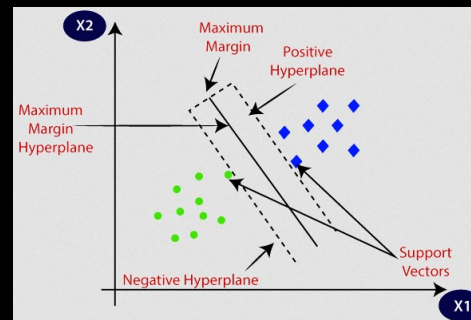
machine learners

Two distinctions

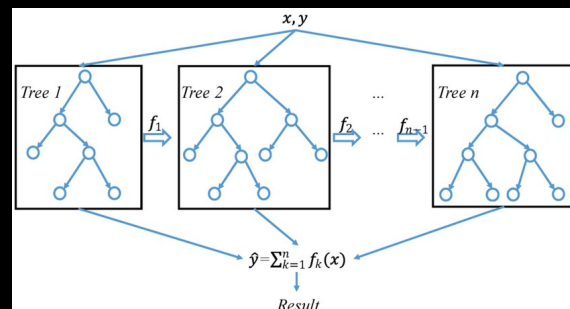
- classifiers versus regressors
 - classifier: prob. of a specific category
 - regressor: predict a scalar
- approach
 - geometric
 - decision trees
 - ANNs
 - ...

learners

- SVM: support vector machine
- SVR
- XGB: XG-boost
- XGR



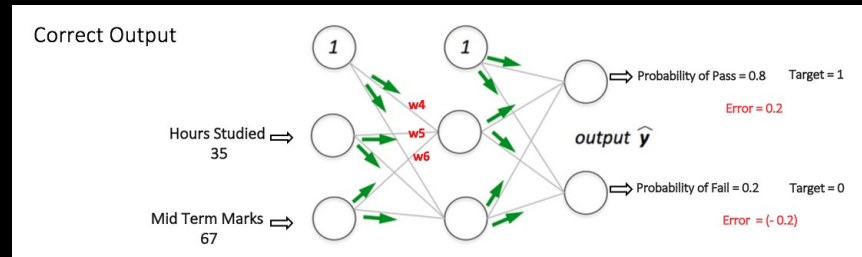
img src: <https://medium.com/@skilltohire/support-vector-machines-4d28a427ebd>



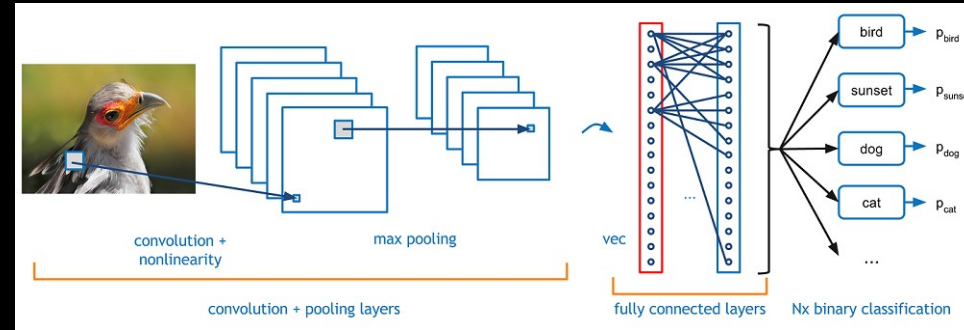
img src: Wang, Yuanchao & Pan, Z. & Zheng, J. & Qian, L. & Mingtao, Li. (2019). A hybrid ensemble method for pulsar candidate classification.

learners cont.

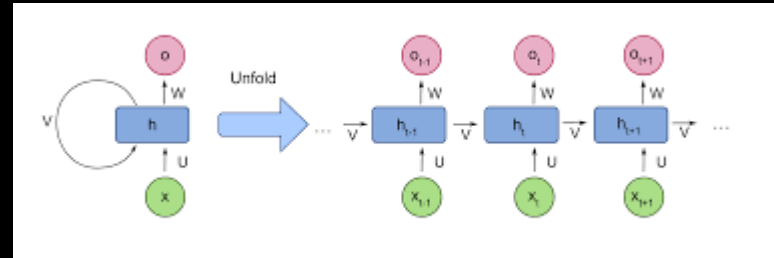
- MLP: multi layer perceptron
- CNN: convolutional neural net
- RNN: recurrent neural net



img src: <https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks/>



img src: <https://medium.datadriveninvestor.com/convolution-neural-network-22565e6d8156>



img src: Wikimedia

conf. file EXP section

- name
- type [classification | regression]
- #epochs
- #runs

https://github.com/felixbur/nkululeko/blob/main/ini_file.md

conf. file DATA section

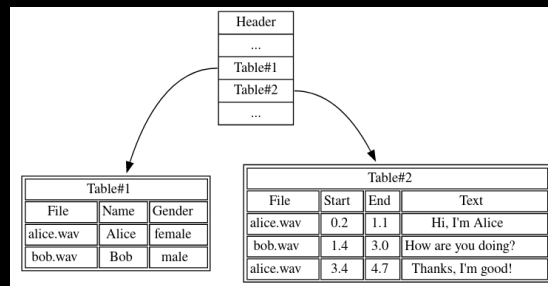
- databases
 - type: [audformat* | CSV]
 - table specifics
 - train/test splits
- type [cross corpus | train-test]
- trains, tests
- label mapping
- binning (scalar → classes)
- sex: data filter

```
x/sample.wav, s1, female, happy
```

```
...
```

```
or with age:
```

```
x/sample.wav, roger, male, 45
```



[*https://audeering.github.io/audformat/](https://audeering.github.io/audformat/)

https://github.com/felixbur/nkululeko/blob/main/ini_file.md

conf. file FEATS section

- type: [os | spectra | mld | xbow | trill | wav2vec]
- scale: [std | spkr | sex]
- model: path to model

https://github.com/felixbur/nkululeko/blob/main/ini_file.md

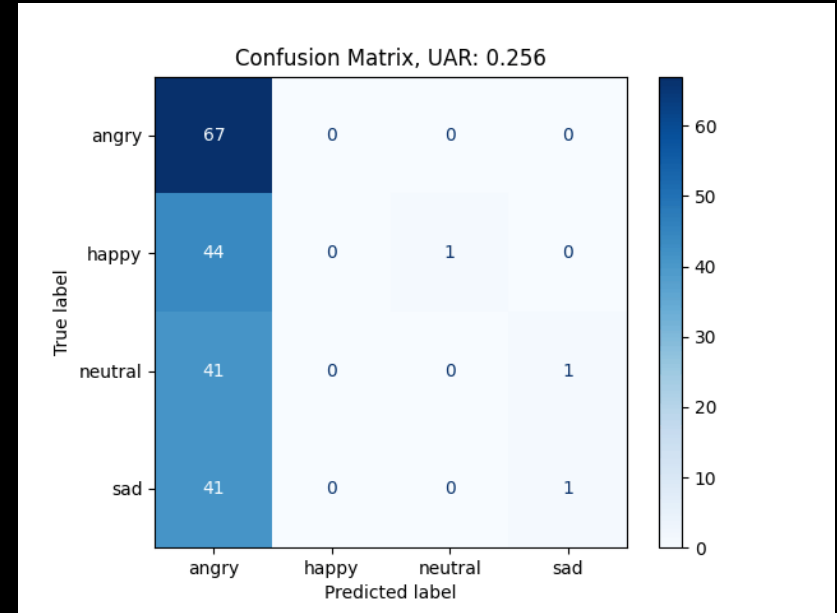
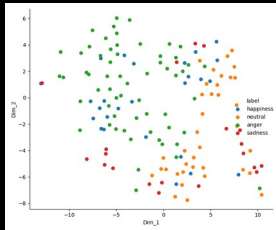
conf. file MODEL section

- type: [svm | svr | xgb | xgr
| mlp | mlp-reg | cnn]
- tuning_params: 5 fold
cross optimization
- layers, loss-function,
learning rate: ANN specs
- class_weight

https://github.com/felixbur/nkululeko/blob/main/ini_file.md

conf. file PLOT section

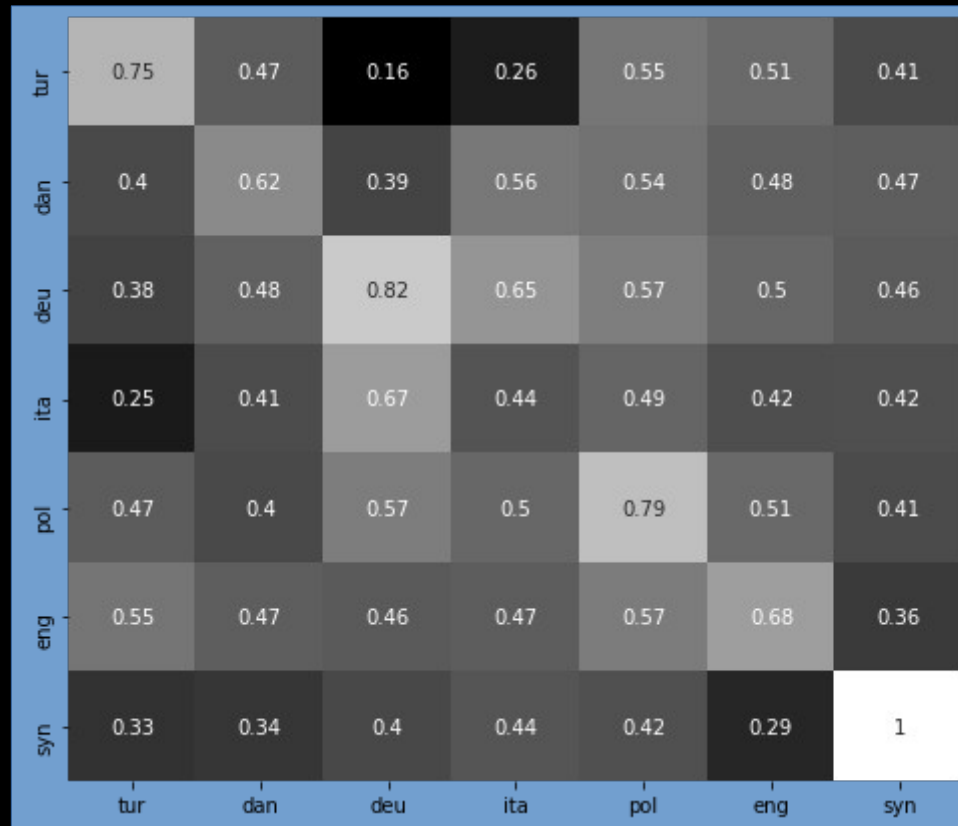
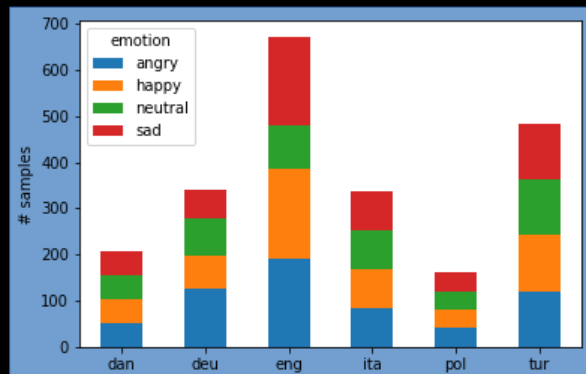
- plot_epochs
- plot_anim_progression
- plot_epoch_progression
- plot_best_model
- T- --SNE



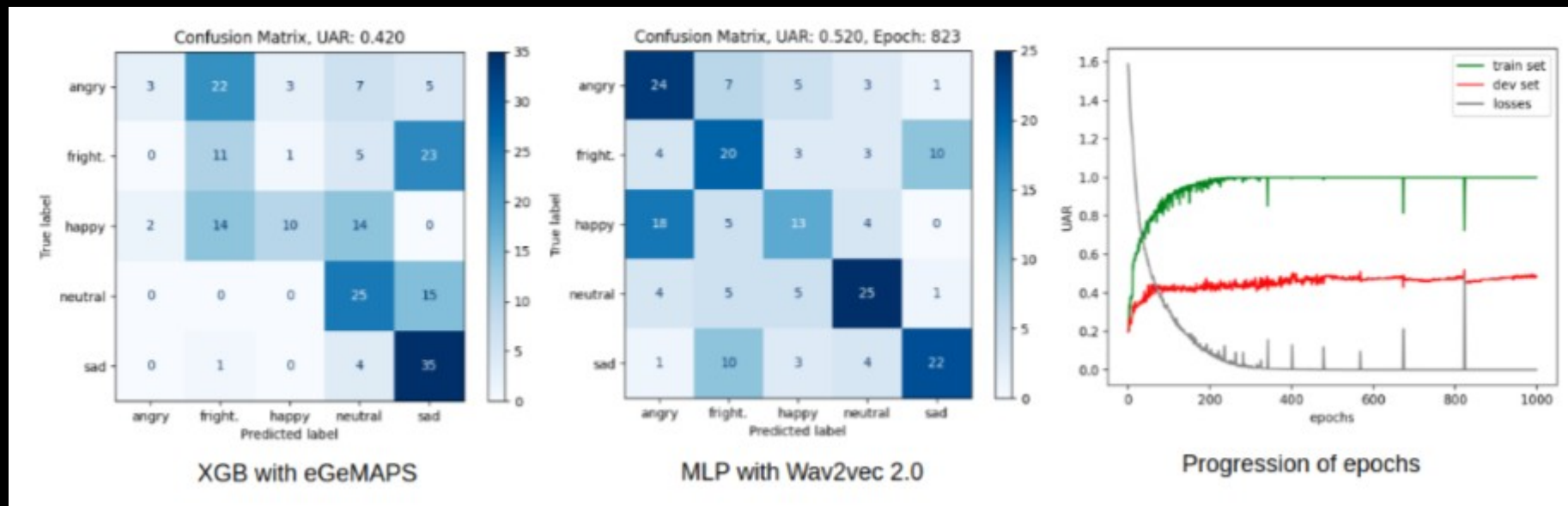
https://github.com/felixbur/nkululeko/blob/main/ini_file.md

experiments

- 6 European acted emotional databases + synthesized

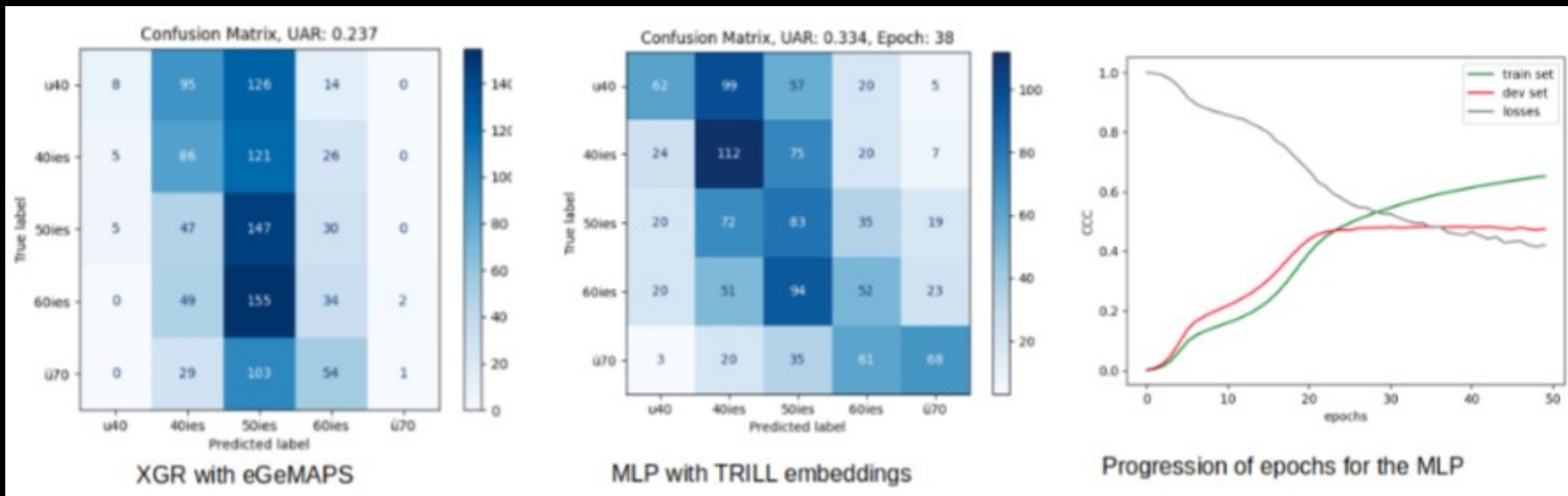


experiments cont.

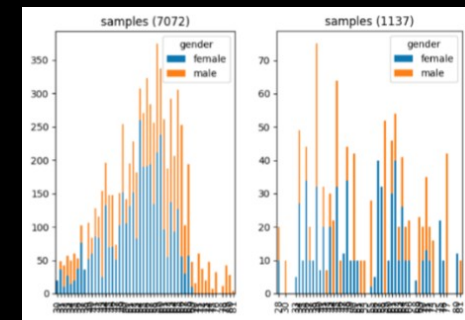


- comparing expert with learned features on cross databases acted emotion learning
- Berlin EmoDB vs. Polish data

experiments cont.



- comparing expert with learned features on age regression
- Data: German parliament data



train and test distribution

wrap up

- introduced Nkululeko
- a software to do machine learning experiments on spoken data without programming
- combines features and learners
- available open source with MIT license