Nkululeko: A Python Package to Predict Speaker Characteristics With a High-level Interface

Felix Burkhardt[1], Bagus Tris Atmaja[2], Florian Eyben[1], Björn Schuller[1,3,4]
[1]audEERING GmbH, Germany, [2]NAIST, Japan, [3]CHI, TUM, Germany, [4]GLAM, Imperial College, UK

Introduction & Summary

Nkululeko is software that detects speaker characteristics through machine-learning experiments, with a high-level interface. e.g.:

- combine acoustic/linguistic features and machine learning models (including feature selection and features concatenation, ensemble learning, parameter tuning, ...),
- perform data exploration, intelligent data splitting, and visualization,
- explore feature performance, e.g. acoustic biomarkers,
- pre-processing (cleaning), re-sampling, segmenting, over-sampling and augmentation of databases.

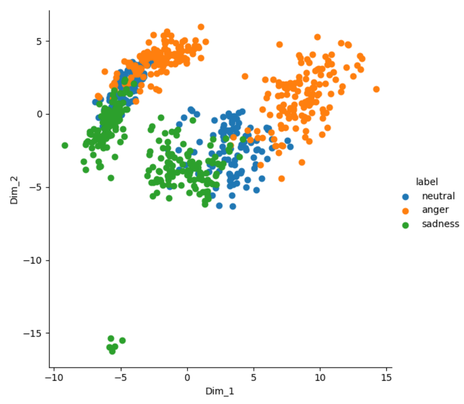
The Input Format

Nkululeko is configured by a text file (no GUI):

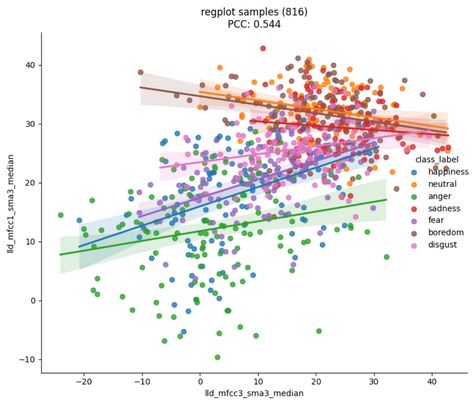
```
[EXP]
name = my-exp
[DATA]
databases = ['androids']
androids = /data/androids/androids.csv
target = depression
[FEATS]
kind = ['wav2vec2']
[MODEL]
type = gmm
[PREDICT]
targets = ['pesq', 'mos', 'text', 'age']
[EXPL]
value_counts = [['gender'], ['age']]
```

Exploration: Feature Importance

Acoustic and linguistic features can be visualized with respect to their importance in distinguishing categories.



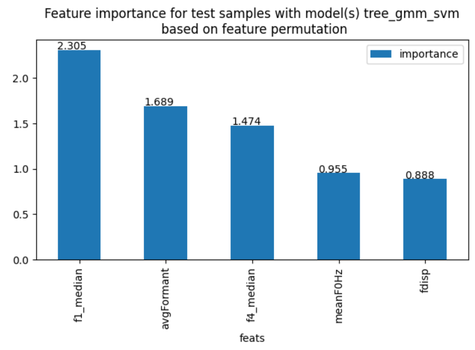
(a) Dimension reduction of feature space with Umap, PCA or TSNE



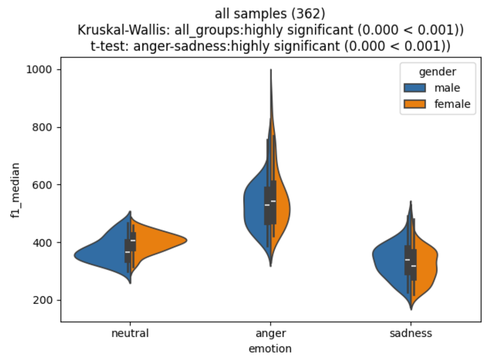
(b) Compare correlation of two acoustic features per emotion

Figure: Visualize feature/target explanation and two feature correlations depending on the target.

Nkululeko can be used to estimate and visualize the importance of specific features and how they are influenced by the target variable.



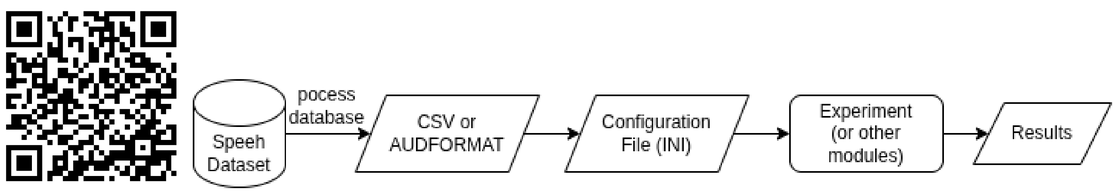
(a) Feature importance calculated



(b) Influence of target on specific feature

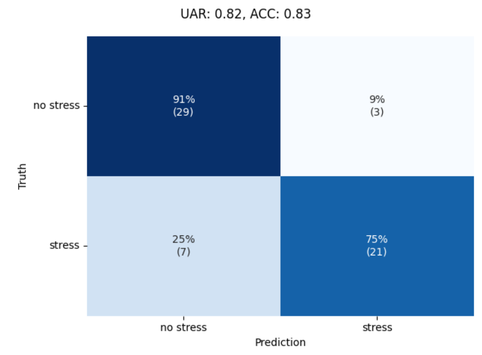
Figure: Visualizing the correlation of acoustic feature and target.

https://github.com/felixbur/nkululeko

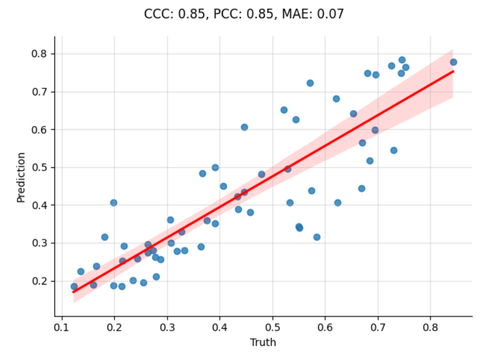


Machine Learning Experiments

Nkululeko performs machine learning experiments as a combination of acoustic features (expert, like Praat or opensmile) and data-driven, like Wav2vec2 or WavML) and learners (all sklearn and PyTorch: e.g., SVM, MLP, CNN, ...).

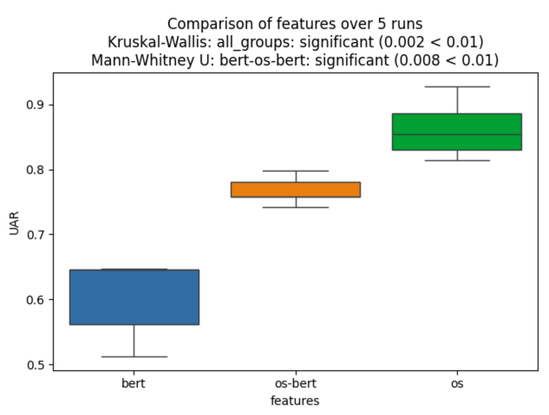


(a) Automatic binning also for non-categorical experiments

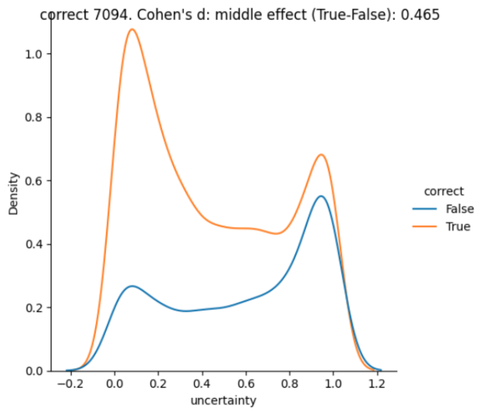


(b) Scatter plot for regression experiments

Figure: Confusion matrix and scatter plot as a result of a machine learning experiment.



(a) Estimating feature differences by random weight initialization

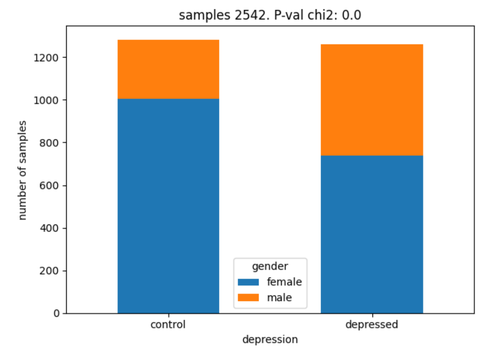


(b) Estimating uncertainty by entropy over the logits

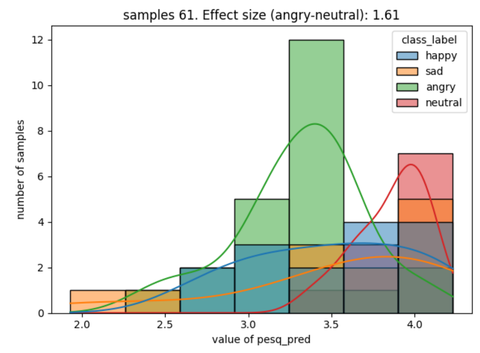
Figure: Additional examples of Nkululeko features.

Prediction and Confounders

Nkululeko integrates many Hugging Face AI models to automatically predict Speech segments, Speech quality (MOS, SNR, PESQ), Speaker characteristics (age, gender), Speaker ID, Spoken text (ASR), and Text translations.
Estimation of the influence of confounder variables is possible.



(a) Effect of gender on target variable



(b) Emotion and estimated PESQ (Perceived Evaluation of Speech Quality)

Figure: Distribution of confounder and target variable.

Conclusion

Nkululeko is a Python command-line tool that uses INI configuration files to specify experiments. Data are imported from a CSV file (file path, speaker ID, gender, task labels) or audformat. The functionality is encapsulated in software modules invoked on the command line. Nkululeko has been used in several research projects since 2022.
https://joss.theoj.org/papers/10.21105/joss.08049.pdf

Acknowledgements

We acknowledge support from: European SHIFT project (Grant 101060660); European EASIER project (Grant 101016982); Project JPNP20006 (NEDO, Japan); Project 24K02967 (JSPS). We thank audEERING GmbH for partial funding.