University of Nottingham
UK | CHINA | MALAYSIA

Benchmarking Framework for Machine Learning with fNIRS

Johann Benerradi, Jeremie Clos, Aleksandra Landowska, Michel F. Valstar, Max L. Wilson

@HanBnrd

Benchmarking Framework for Machine Learning with fNIRS

Contents

Background

Background: fNIRS

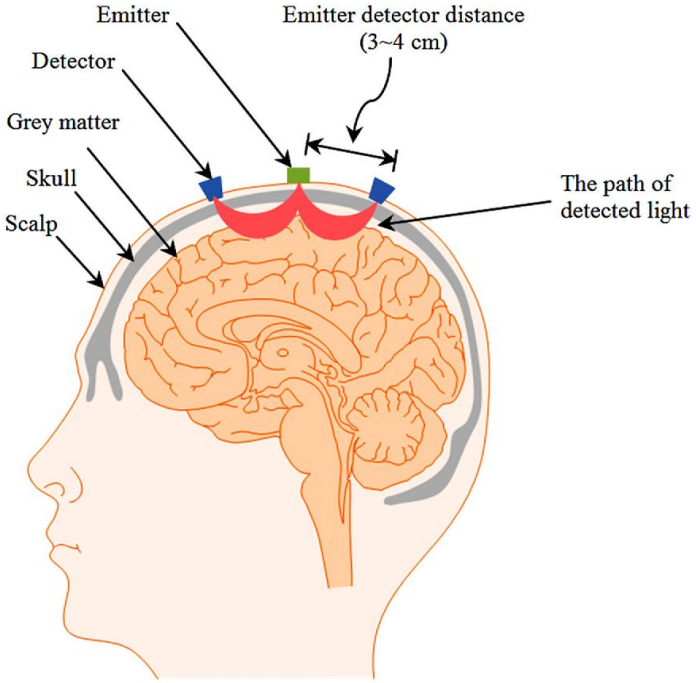
Functional near-infrared spectroscopy (fNIRS)

Attenuation of light in the cerebral blood ➔ measure changes in hemoglobin concentrations

Hemoglobin carries oxygen to the brain ➔ indication on brain activity in neocortex

Measure 2 types: oxy & deoxyhemoglobin

Classification ➔ BCI, offline analysis, …



Kumar et al., 2017

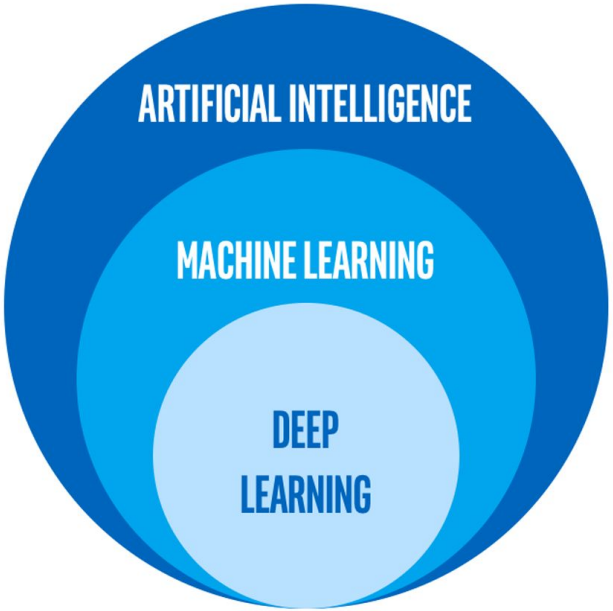
Background: machine learning

Machine learning: techniques that enable a computer learn from examples

Deep learning: neural networks with multiple layers

Supervised: each example has a label associated

Hyperparameters: parameters external to the model that are set and control the learning



intel.com

Motivation

Motivation: current limitations

Easy to learn machine learning, hard to learn the pitfalls

Potential mistakes, lack of robustness, lack of details in some publications

- **Methodology**: hyperparameters, cross-validation, testing, stats, class imbalance, ...

- **Reporting**: data, hyperparameters, cross-validation, chance level, ...

Reproducibility impossible in most cases, possible **bias** in the results

Framework

Framework: datasets

Dataset	Classes	Nb participants	Nb examples/class
Herff et al. 2014 n-back	1-back; 2-back; 3-back	10	100
Shin et al. 2018 n-back	0-back; 2-back; 3-back	26	234
Bak et al. 2019 motor execution	right hand; left hand; foot	30	750
Shin et al. 2018 word generation	rest; word generation	26	780
Shin et al. 2016 mental arithmetic	rest; mental arithmetic	29	870

Framework: data processing

- Signal processing (*NIRSimple*[1], *MNE*[2]):
  - ➔ conversion of raw signal into hemoglobin concentration changes
  - ➔ region of interest averaging
  - ➔ filtering (low-pass, high-pass, band-pass)
  - ➔ baseline correction

- Feature extraction:
  - ➔ mean
  - ➔ standard deviation
  - ➔ slope of linear regression

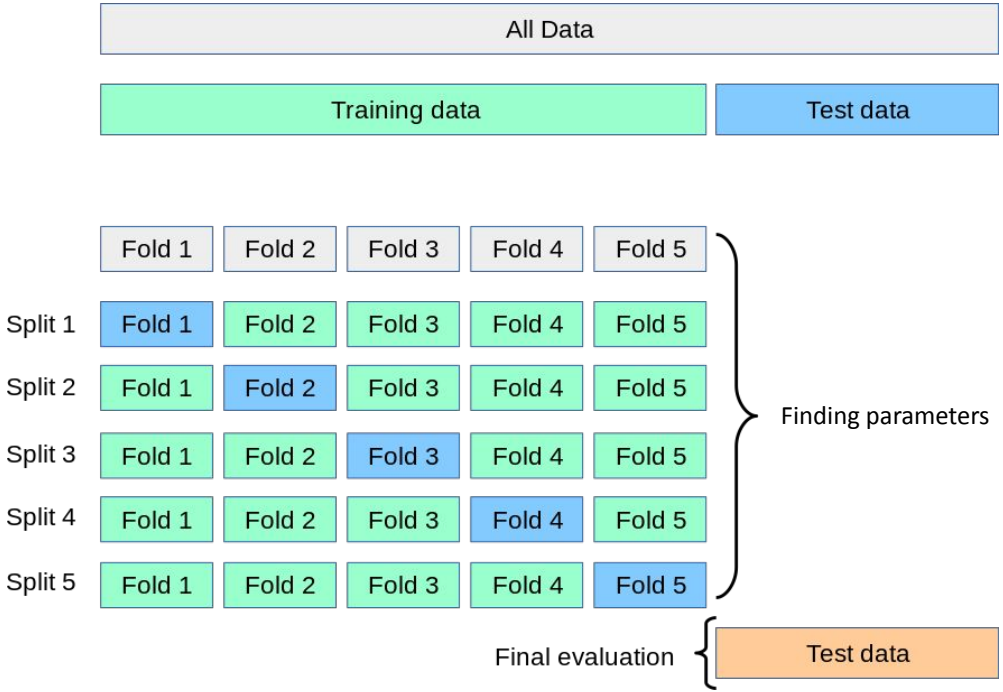
1. https://github.com/HanBnrd/NIRSimple
2. https://mne.tools/stable/index.html

Framework: cross-validation

Nested cross-validation
(also called double cross-validation):

➜ outer cross-validation for evaluation on **test set**

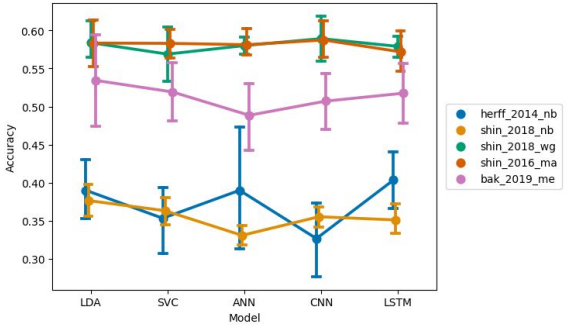
➜ inner cross-validation for hyperparameter tuning on **validation set**
(eg. grid search)

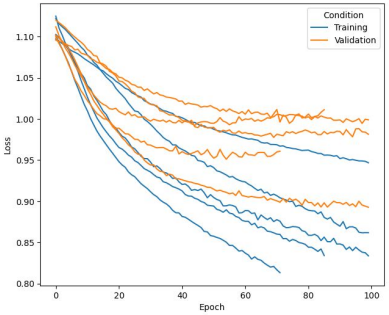


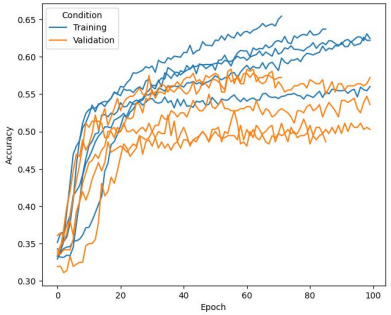
scikit-learn.org

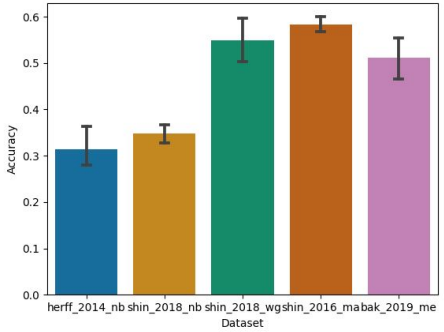
Framework: metrics and figures

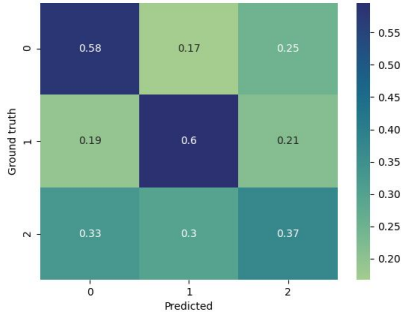
- Accuracy
- Precision
- Recall
- F1 score

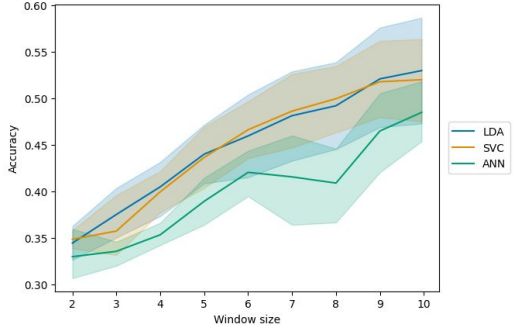












Framework: stats

- **Comparison to chance level**: t-test, Wilcoxon test
- **Influence of the model on the performance**: ANOVA, Kruskal-Wallis test
- **Comparison of models to each other**: pairwise t-test, pairwise Wilcoxon test
- **Correlation of a metric to a variable**: Pearson test, Spearman test

It is very important to **verify the assumptions** of statistical tests (independent measures, normality, homogeneity of variance, ...)

Early results

Early results: generalised approach

Dataset	Chance	LDA	SVC	ANN	CNN	LSTM
Herff 2014 n-back	0.33	0.39	0.35	0.39	0.33	**0.40** *
Shin 2018 n-back	0.33	**0.38** *	**0.36** *	0.33	**0.36** *	0.35
Bak 2019 motor exec.	0.33	**0.53** *	**0.52** *	**0.49** *	**0.51** *	**0.52** *
Shin 2018 word gen.	0.50	0.58	**0.57** *	**0.58** *	**0.59** *	**0.58** *
Shin 2016 mental arithm.	0.50	0.58	**0.58** *	**0.58** *	**0.59** *	**0.57** *

* significantly different than chance level

Conclusion

Conclusion

- Framework called ***BenchNIRS*** will be open sourced (paper under review)

- Use cross-validation (nested)
  ➜ hyperparameter tuning on validation set
  ➜ do not use the test set for anything else than evaluation (no optimisation)

- Report all the details for reproducibility

- Justify choices

- Publish data and code

Contact details

johann.benerradi@nottingham.ac.uk

Webpage: *hanbnrd.gitlab.io*

@HanBnrd

Thank you for your attention