Fraunhofer MEVIS

DAI-Labor
Distributed Artificial Intelligence Laboratory

empaia

Funded by
Bundesministerium für Wirtschaft und Klimaschutz
01MK20002B
01MK20002C

EVALUATING GENERIC AUTOML TOOLS FOR COMPUTATIONAL PATHOLOGY

Lars Ole Schwen[1], Daniela Schacherer[1], Christian Geißler[2], André Homeyer[1]
[1]Fraunhofer MEVIS, Bremen [2]DAI-Labor, TU Berlin

MOTIVATION

- Choice of neural network architecture and training hyperparameters are key components of developing AI solutions
- AutoML approaches reduce efforts for typically exploratory search and manual optimization
- **Question:** How well does general-purpose AutoML for image classification perform for tasks in computational pathology? [6]

GENERIC AUTOML TOOLS

AUTOGLUON [1]

- runs on-premises
- open source
- used via python
- presets for some parameters
- limited hyperparameter optimization (HPO)

GOOGLE AUTOML VISION [2]

- runs in the cloud
- no access to internals
- used via web interface
- three presets + training time budget

COMPUTATIONAL PATHOLOGY USE CASES

TISSUE CLASSIFICATION [4]

- lung tissue: normal, adenocarcinoma (LUAD), and squamous cell carcinoma (LUSC)
- distinguish normal vs. tumor (binary); normal vs. LUAD vs. LUSC (ternary classification)
- assessed via area under receiver operating curve (AUROC)

MUTATION PREDICTION [5]

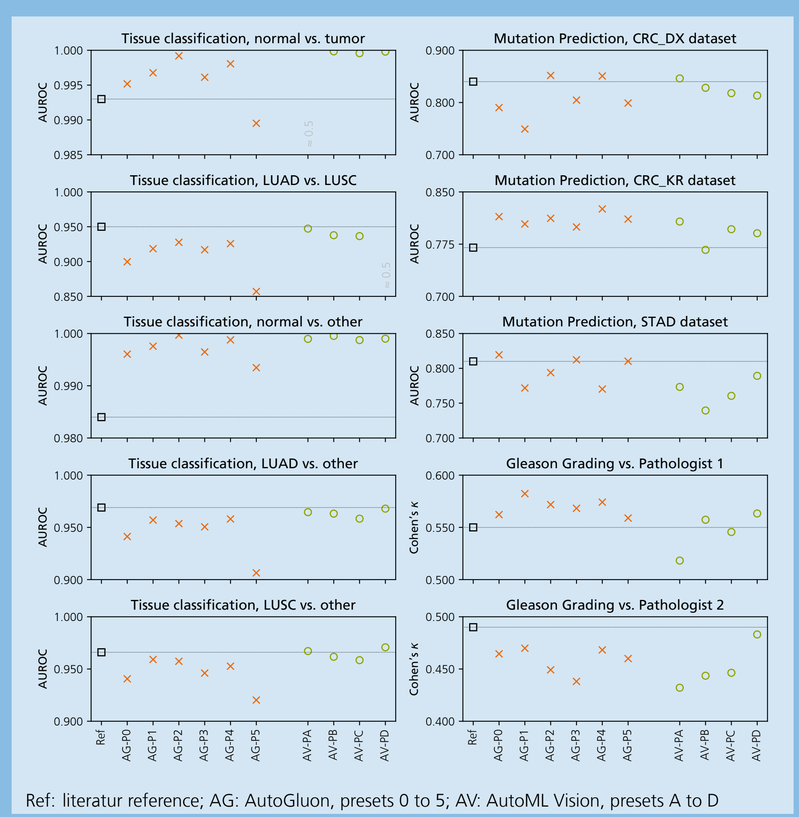
- colorectal cancer and gastric adenocarcinoma tissue
- predict microsatellite (in)stability (MSI vs. MSS): binary classification
- assessed via AUROC for three datasets

GLEASON GRADING [3]

- prostate biopsies, tissue microarray images
- predict grade 0, 3, 4, 5: quarternary classification
- assessed via Cohen's quadratically weighted kappa for annotations by two pathologists

RESULTS

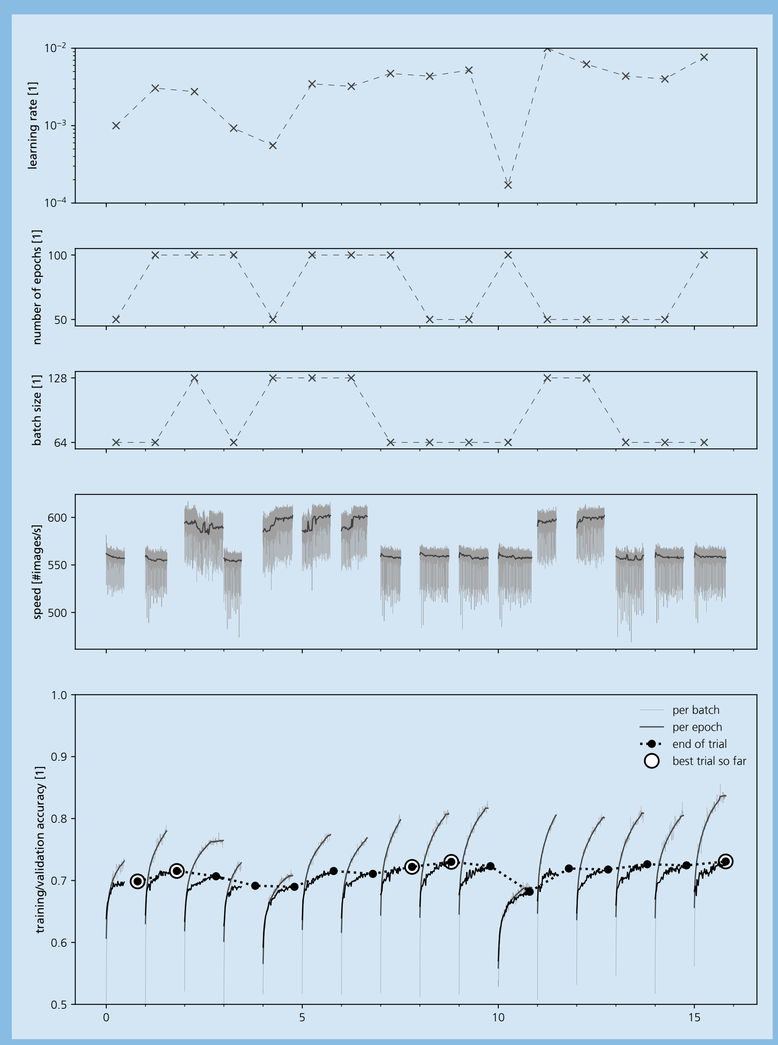
PERFORMANCE VS. REFERENCE



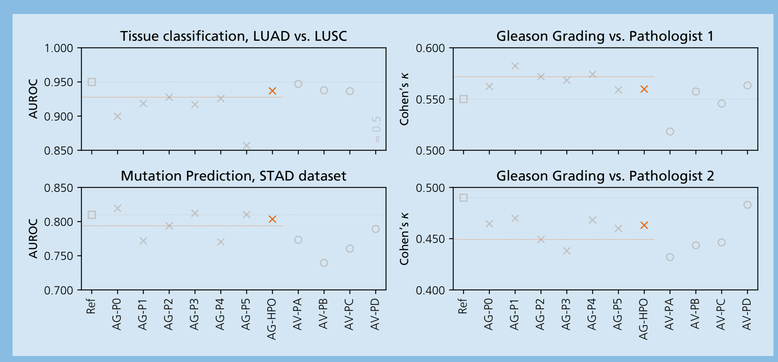
Ref: literatur reference; AG: AutoGluon, presets 0 to 5; AV: AutoML Vision, presets A to D

- generally on par with literature reference

HYPERPARAMETER OPTIMIZATION

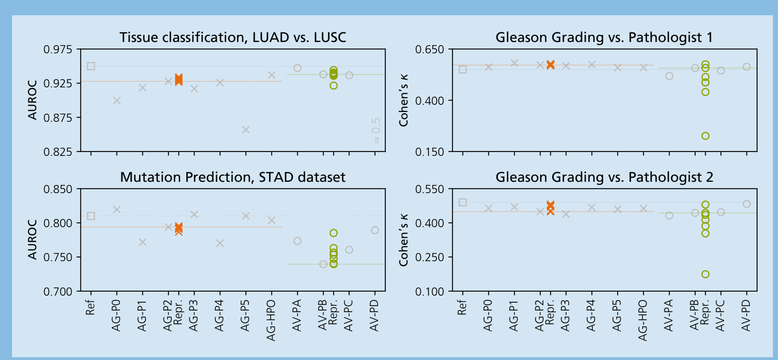


- tissue classification, LUAD vs. LUSC:
  - batch size influences speed
  - learning rate influences accuracy



- (limited) hyperparameter optimization has limited impact

RESULT VARIABILITY



- training is non-deterministic
- performance may vary substantially

REFERENCES

[1] *AutoGluon: AutoML for text, image, and tabular data*, 2021, https://auto.gluon.ai/stable/index.html.

[2] *AutoML Vision documentation*, 2021, https://cloud.google.com/vision/automl/, as available between Sep 02 and Sep 22, 2021.

[3] E. Arvaniti, K. S. Fricker, M. Moret, N. Rupp, T. Hermanns, C. Fankhauser, N. Wey, P. J. Wild, J. H. Rueschoff, and M. Claassen, *Automated Gleason grading of prostate cancer tissue microarrays via deep learning*, Scientific Reports 8 (2018), no. 1, 1–11. DOI 10.1038/s41598-018-30535-1

[4] N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, and A. Tsirigos, *Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning*, Nat. Med. 24 (2018), no. 10, 1559. DOI 10.1038/s41591-018-0177-5

[5] J. N. Kather, A. T. Pearson, N. Halama, D. Jäger, J. Krause, S. H. Loosen, A. Marx, P. Boor, F. Tacke, U. P. Neumann, H. I. Grabsch, T. Yoshikawa, H. Brenner, J. Chang-Claude, M. Hoffmeister, C. Trautwein, and T. Luedde, *Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer*, Nature Medicine 25 (2019), 1054–1056. DOI 10.1038/s41591-019-0462-y

[6] L. O. Schwen, D. Schacherer, C. Geißler, and A. Homeyer, *Evaluating generic AutoML tools for computational pathology*, Informatics in Medicine Unlocked 29 (2022), no. 100853. DOI 10.1016/j.imu.2022.100853