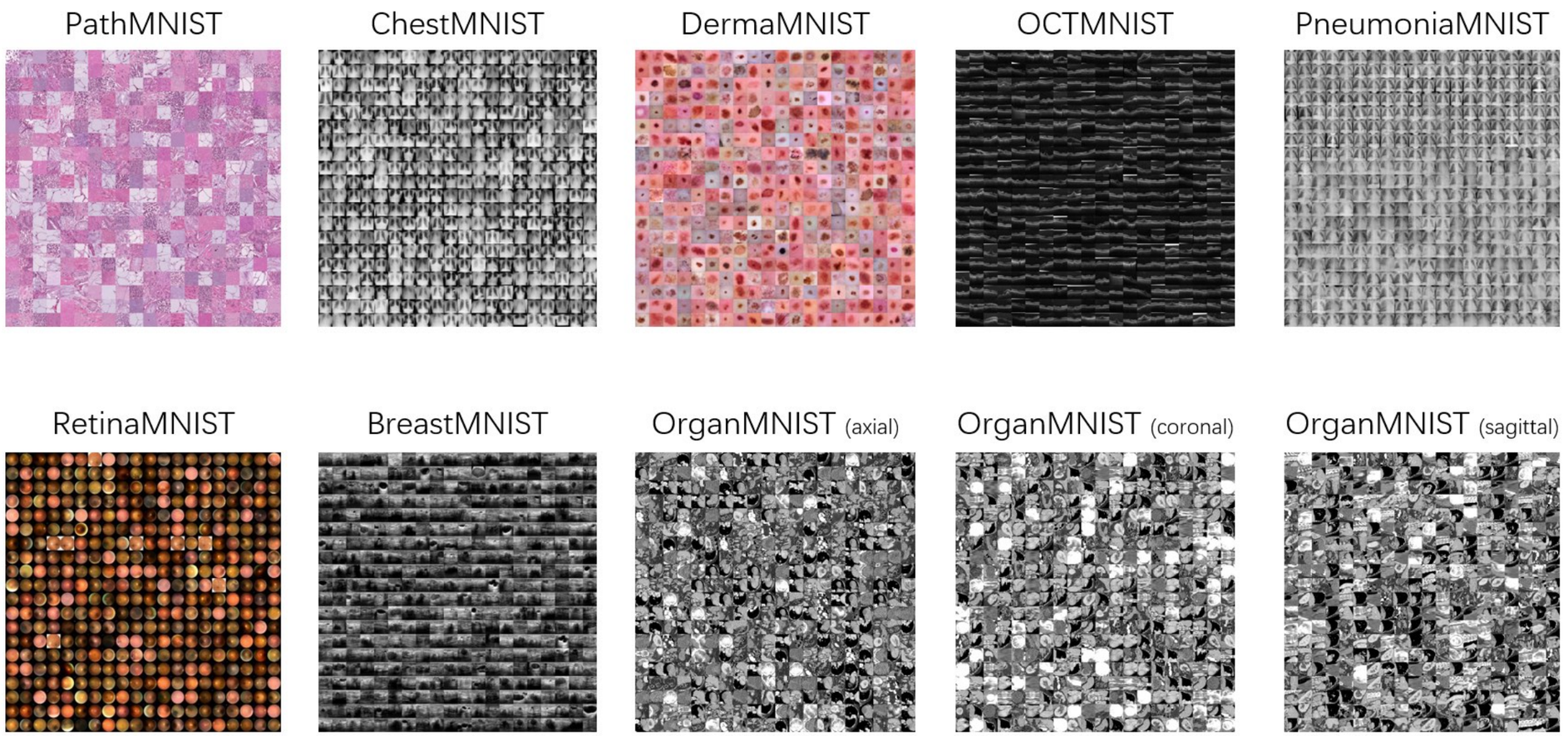
MedMNIST Classification Decathlon: A Lightweight AutoML Benchmark for Medical Image Analysis

[Jiancheng Yang](#) Rui Shi [Bingbing Ni](#) [Bilian Ke](#)
Shanghai Jiao Tong University, Shanghai, China

Paper [ISBI'21] Code [Github] Dataset [Download]



Abstract

We present MedMNIST, a collection of 10 pre-processed medical open datasets. MedMNIST is standardized to perform classification tasks on lightweight 28 * 28 images, which requires no background knowledge. Covering the primary data modalities in medical image analysis, it is diverse on data scale (from 100 to 100,000) and tasks (binary/multi-class, ordinal regression and multi-label). MedMNIST could be used for educational purpose, rapid prototyping, multi-modal machine learning or AutoML in medical image analysis. Moreover, MedMNIST Classification Decathlon is designed to benchmark AutoML algorithms on all 10 datasets; We have compared several baseline methods, including open-source or commercial AutoML tools.

Key Features

- **Educational**: Our multi-modal data, from multiple open medical image datasets with Creative Commons (CC) Licenses, is easy to use for educational purpose.
- **Standardized**: Data is pre-processed into same format, which requires no background knowledge for users.
- **Diverse**: The multi-modal datasets covers diverse data scales (from 100 to 100,000) and tasks (binary/multiclass, ordinal regression and multi-label).
- **Lightweight**: The small size of 28 × 28 is friendly for rapid prototyping and experimenting multi-modal machine learning and AutoML algorithms.

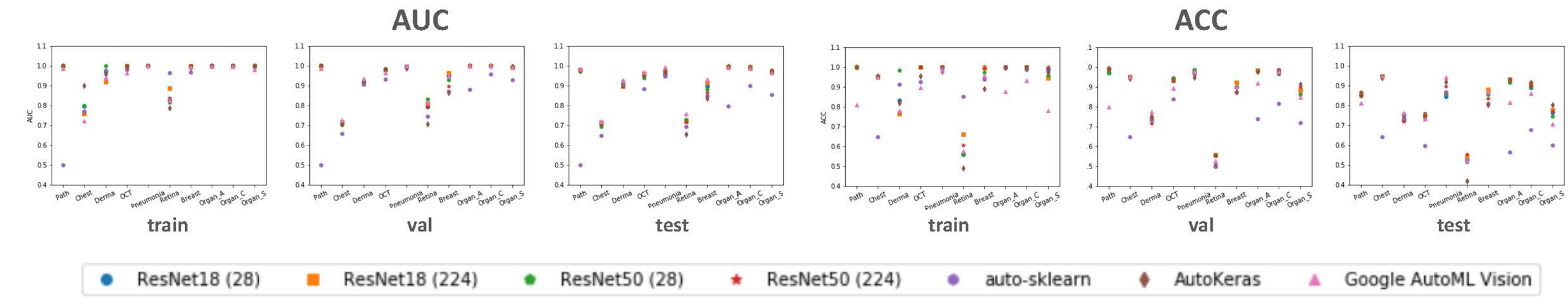
  Please note that this dataset is **NOT** intended for clinical use.

Materials

An Overview of MedMNIST Dataset

Name	Data Modality	Tasks (# Classes/Labels)	# Training	# Validation	# Test
PathMNIST	Pathology	Multi-Class (9)	89,996	10,004	7,180
ChestMNIST	Chest X-ray	Multi-Label (14) Binary-Class (2)	78,468	11,219	22,433
DermaMNIST	Dermatoscope	Multi-Class (7)	7,007	1,003	2,005
OCTMNIST	OCT	Multi-Class (4)	97,477	10,832	1,000
PneumoniaMNIST	Chest X-ray	Binary-Class (2)	4,708	524	624
RetinaMNIST	Fundus Camera	Ordinal Regression (5)	1,080	120	400
BreastMNIST	Breast Ultrasound	Binary-Class (2)	546	78	156
OrganMNIST_Axial	Abdominal CT	Multi-Class (11)	34,581	6,491	17,778
OragnMNIST_Coronal	Abdominal CT	Multi-Class (11)	13,000	2,392	8,268
OrganMNIST_Sagittal	Abdominal CT	Multi-Class (11)	13,940	2,452	8,829

Performance Analysis



Download

You could download the dataset(s) via the following free accesses:

- **zenodo.org** (recommended): You could also use our code to download the datasets from zenodo.org automatically.
- **Google Drive**
- **百度网盘** (code: gx6i)

Citation and Licenses

If you find this project useful, please cite our paper as:
Jiancheng Yang, Rui Shi, Bingbing Ni. "MedMNIST Classification Decathlon: A Lightweight AutoML Benchmark for Medical Image Analysis," arXiv preprint arXiv:2010.14925, 2020.

or using bibtex:
@article{medmnist,
title={MedMNIST Classification Decathlon: A Lightweight AutoML Benchmark for Medical Image Analysis},

author={Yang, Jiancheng and Shi, Rui and Ni, Bingbing},
journal={arXiv preprint arXiv:2010.14925},
year={2020}
}

Besides, please cite the corresponding paper if you use any subset of MedMNIST. Each subset uses the **same license** as that of the source dataset.

PathMNIST

Jakob Nikolas Kather, Johannes Krisam, et al., "Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study," PLOS Medicine, vol. 16, no. 1, pp. 1–22, 01 2019.

License: *CC BY 4.0*

ChestMNIST

Xiaosong Wang, Yifan Peng, et al., "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in CVPR, 2017, pp. 3462–3471.

License: *CC0 1.0*

DermaMNIST

Philipp Tschandl, Cliff Rosendahl, and Harald Kittler, "The ham10000 dataset, a large collection of multisource dermatoscopic images of common pigmented skin lesions," Scientific data, vol. 5, pp. 180161, 2018.

License: *CC BY-NC 4.0*

OCTMNIST/PneumoniaMNIST

Daniel S. Kermany, Michael Goldbaum, et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," Cell, vol. 172, no. 5, pp. 1122 – 1131.e9, 2018.

License: *CC BY 4.0*

RetinaMNIST

DeepDR Diabetic Retinopathy Image Dataset (DeepDRiD), "The 2nd diabetic retinopathy – grading and image quality estimation challenge," https://isbi.deepdr.org/data.html, 2020.

License: *CC BY 4.0*

BreastMNIST

Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy, "Dataset of breast ultrasound images," Data in Brief, vol. 28, pp. 104863, 2020.

License: *CC BY 4.0*

OrganMNIST_{Axial,Coronal,Sagittal}

Patrick Bilic, Patrick Ferdinand Christ, et al., "The liver tumor segmentation benchmark (lits)," arXiv preprint arXiv:1901.04056, 2019.

Xuanang Xu, Fugen Zhou, et al., "Efficient multiple organ localization in ct image using 3d region proposal network," IEEE Transactions on Medical Imaging, vol. 38, no. 8, pp. 1885–1898, 2019.

License: *CC BY 4.0*

/