Deep Learning papers related to MoLAB activities

Anton Popov

UCLM, MoLAB, Spain, Ciudad Real

May 28, 2019

Definitions

- **Tumor burden(tumor load)** - refers to the number of cancer cells, the size of a tumor, or the amount of cancer in the body.
- **Response Evaluation Criteria In Solid Tumor (RECIST)** - A standard way to measure how well a cancer patient responds to treatment. It is based on whether tumors shrink, stay the same, or get bigger. To use Response Evaluation Criteria In Solid Tumors, there must be at least one tumor that can be measured on x-rays, CT scans, or MRI scans. The types of response a patient can have are a complete response (CR), a partial response (PR), progressive disease (PD), and stable disease (SD)(*tumor-centric, not patient centric criteria*)
- **Response Assessment in Neuro-Oncology (RANO)** - criteria and requirements for a uniform protocol.

[1]

[1] https://www.cancer.gov/publications/dictionaries

Response Assessment in Neuro-Oncology Clinical Trials

Patrick Y. Wen, Susan M. Chang, Martin J. Van den Bent, Michael A. Vogelbaum, David R. Macdonald, and Eudocia Q. Lee

Journal of Clinical Oncology (link); June 22, 2017

RANO summary

- multidisciplinary international working group ... working in collaboration with government and industry to *enhance the interpretation of clinical trials related to intracranial tumors*
- born out of a workshop conducted by the Jumpstarting Brain Tumor Drug Development Coalition and the US Food and Drug Administration
- **standardized brain tumor imaging protocol now exists to reduce variability and improve reliability**

Assumption: tumors grow in spherical shapes and the 2D measurement of lesion's largest diameter on MRI is a surrogate marker of tumor volume *(relies on manual two-dimensional measurements...)*

Ways to improve?

Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study

Philipp Kickingereder, Fabian Isensee*, Irada Tursunova, Jens Petersen, Ulf Neuberger, David Bonekamp, Gianluca Brugnara, Marianne Schell, Tobias Kessler, Martha Foltyn, Inga Harting, Felix Sahm, Marcel Prager, Martha Nowosielski, Antje Wick, Marco Nolden, Alexander Radbruch, Jürgen Debus, Heinz-Peter Schlemmer, Sabine Heiland, Michael Platten, Andreas von Deimling, Martin J van den Bent, Thierry Gorlia, Wolfgang Wick, Martin Bendszus†, Klaus H Maier-Hein†*

Lancet Oncology (link); April 2, 2019

Definitons and abbreviations in the paper

- **Definition:** Given two sets X and Y:

$$DICE = \frac{2|X \cap Y|}{|X| + |Y|}$$

  where $|X|$, $|Y|$ - cardinalities of two sets.
- CE - contrast-enhancing target lesions
- NE - non-enhancing T2-signal abnormalities

Results

Segmentation results:

- median DICE on Heidelberg test dataset:
  - CE: 0.89 [95% CI 0.86 - 0.90]
  - NE: 0.93 [95% CI 0.92 - 0.94]
- median DICE on EORTC-26101 test dataset:
  - CE: 0.91 [95% CI 0.90 - 0.92]
  - NE: 0.93 [95% CI 0.93 - 0.94]

Hazard ratios:

- ANN: 2.59 [95% CI 1.86 - 3.60]
- central RANO: 2.07 [95% CI 1.46 - 2.92]; $p < 0.0001$

Yields a 36% margin over RANO ($p < 0.0001$) when comparing reliability values(i. e., agreement in the quantitative volumetrically defined time to progression [based on radiologists ground truth *vs* automated assessment with ANN] of 87% [266 of 306 with sufficient data] compared with 51% [155 of 306] with local *vs* independent central RANO assessment).

Data overview

	MRI	Patients	Comments
Train	455	455	Helidberg training dataset
Test	239	40	Heildelberg testing dataset
	2034	532	EORTC-26101 study (link)
Simulate	595	466	Heildelberg simulation dataset

All Heildelberg datasets includes:

- T1-weighted (before contrast agent)
- cT1-weighted (after contrast agent)
- fluid-attenuated inversion recovery (FLAIR)
- T2-weighted images

IMHO feeding all modalities makes sense (link).

Helidberg training dataset

Feature	Comment
"histologically confirmed glioblastoma or lower-grade glioma (including diffuse astrocytic and oligodendroglial WHO grade II and III tumours)"	GB and LGG looks differently $\rightarrow$ expressive ANN is needed
"single timepoint: preoperatively from initial diagnosis or early postoperatively ... or at follow-up and was specifically ensembled to represent the broad phenotypic appearence of brain tumours on MRI during disease evolution"	preoperatively, postoperatively, follow-up mixed together $\rightarrow$ expressive ANN is needed
appropriate MRI scans were manually identified to enrich the dataset with comparatively uncommon and difficult cases on the basis of the judgement of the neuroradiologists	only patients with disease

Excellent example of following recommendations from Google (Rule 5: Test the infrastructure independently from the ML).

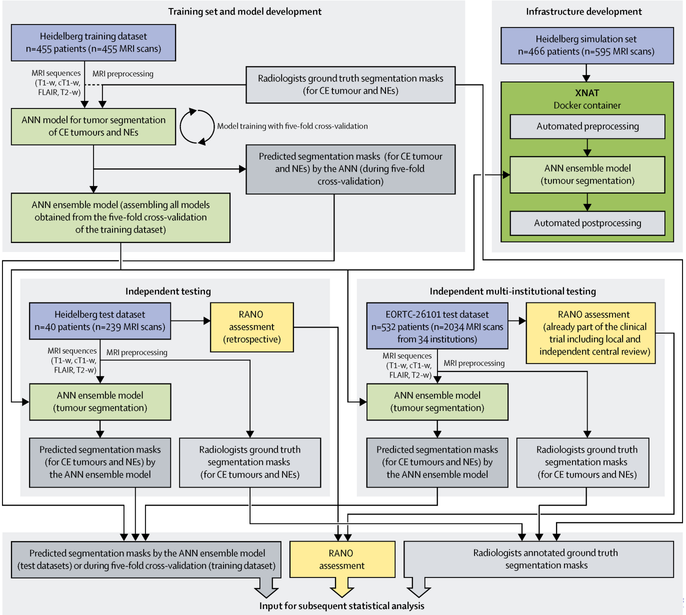
Test

This dataset was a longtitudinal dataset with preoperative and consequtive follow-up scans

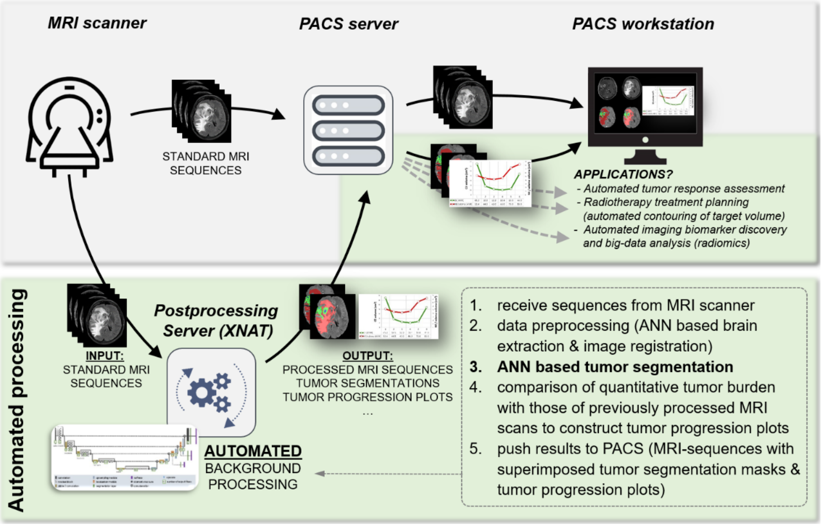
Simulate

Cohort of adult patients for testing of the developed infrastructure for automated tumour segmentation and quantitative assessment of tumour response in a simulated clinical environment

Process


Training set and model development
Heidelberg training dataset n=455 patients (n=455 MRI scans)
MRI sequences (T1-w, cT1-w, FLAIR, T2-w)
MRI preprocessing
Radiologists ground truth segmentation masks (for CE tumour and NEs)
ANN model for tumor segmentation of CE tumours and NEs
Model training with five-fold cross-validation
Predicted segmentation masks (for CE tumour and NEs) by the ANN (during five-fold cross-validation)
ANN ensemble model (assembling all models obtained from the five-fold cross-validation of the training dataset)
Infrastructure development
Heidelberg simulation set n=466 patients (n=595 MRI scans)
XNAT Docker container
Automated preprocessing
ANN ensemble model (tumour segmentation)
Automated postprocessing
Independent testing
Heidelberg test dataset n=40 patients (n=239 MRI scans)
RANO assessment (retrospective)
MRI sequences (T1-w, cT1-w, FLAIR, T2-w)
MRI preprocessing
ANN ensemble model (tumour segmentation)
Predicted segmentation masks (for CE tumours and NEs) by the ANN ensemble model
Radiologists ground truth segmentation masks (for CE tumours and NEs)
Independent multi-institutional testing
EORTC-26101 test dataset n=532 patients (n=2034 MRI scans from 34 institutions)
RANO assessment (already part of the clinical trial including local and independent central review)
MRI sequences (T1-w, cT1-w, FLAIR, T2-w)
MRI preprocessing
ANN ensemble model (tumour segmentation)
Predicted segmentation masks (for CE tumours and NEs) by the ANN ensemble model
Radiologists ground truth segmentation masks (for CE tumours and NEs)
Predicted segmentation masks by the ANN ensemble model (test datasets) or during five-fold cross-validation (training dataset)
RANO assessment
Radiologists annotated ground truth segmentation masks
Input for subsequent statistical analysis

Infrastructure



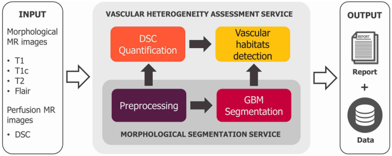


Figure: MRI → processing → output

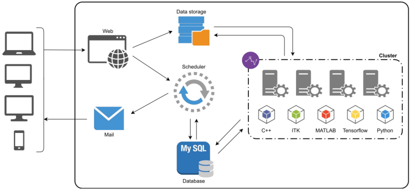
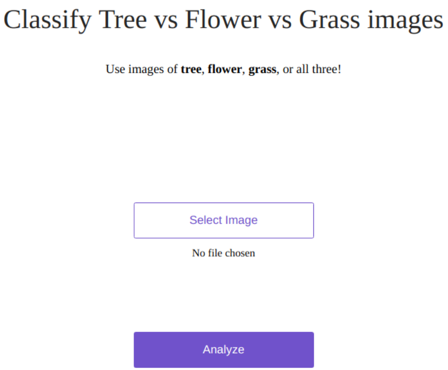


Figure: Standard Web-app

Notes:

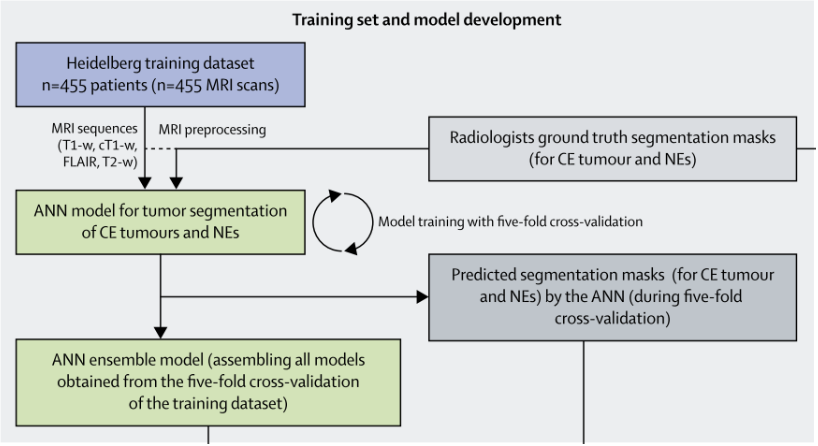
- the system is functioning
- separate storage of MRI images and other data

Proof of concept - Classification Web App

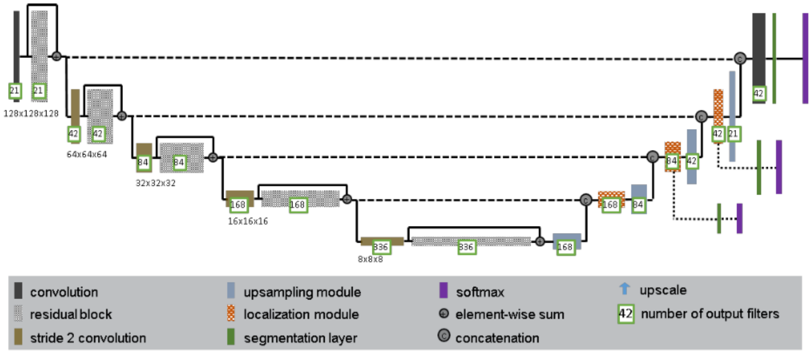


This app (tree-grass-flower.onrender.com) is running on render.com platform. After image uploading user receives class prediction.

Training set and model development

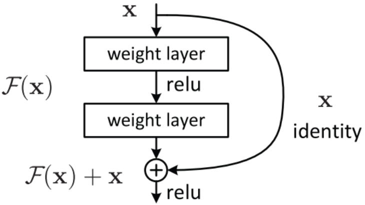


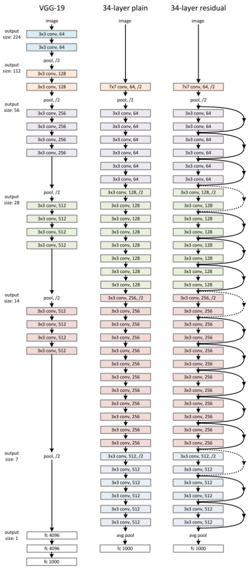
Neural Network architecture



Inspired by U-Net and utilizes residual blocks in encoder.

ResNet





- ResNet introduce **skip connection**
- solves *vanishing(exploding) gradient* problem

General notes about the paper

- Appendix contains more technical/useful information
- Links to software/technologies that were used are provided (e. g. www.xnat.org)
- Infrastructure notes:
  - Docker container utilization, Docker Swarm parallelization
  - open-source only components
  - code is available on GitHub
- Network architecture
  - Heavy encoder, light decoder
  - Large Input Patch Size ($128 \times 128 \times 128$ voxels)
  - Auxiliary Loss Layers (autograd understanding)
  - Nonlinearity and Normalization
    - instance normalization
    - leaky ReLU activation function

Artificial Intelligence in Cancer Imaging: Clinical Challenges and Applications

Wenya Linda Bi, MD[1]; Ahmed Hosny, MS[2]; Matthew B. Schabath, PhD[3]; Maryellen L. Giger, PhD[4]; Nicolai J. Birkbak, PhD[5,6]; Alireza Mehrtash, MSc[7,8]; Tavis Allison, BS[9,10]; Omar Arnaout, MD[11]; Christopher Abbosh, MD[12,13]; Ian F. Dunn, MD[14]; Raymond H. Mak, MD[15]; Rulla M. Tamimi, PhD[16]; Clare M. Tempany, MD[17]; Charles Swanton, MD, PhD[18,19]; Udo Hoffmann, MD[20]; Lawrence H. Schwartz, MD[21,22]; Robert J. Gillies, MD[23]; Raymond Y. Huang, MD, PhD[24]; Hugo J. W. L. Aerts, PhD[25,26]

CA: a cancer journal for clinicians (link); February 5, 2019
Overview for medical doctors. Covers works over lung, brain, breast and prostate cancers at detection, characterization and monitoring of tumors tasks with next key points:

- aggregation of multiple data streams(e.g. "imaging genomics")
- importance of early diagnosing
- AI has the potential to increase efficiency, reproducibility but still in its infancy

Neural Ordinary Differential Equations

Ricky T. Q. Chen*, Yulia Rubanova*, Jesse Bettencourt*, David Duvenaud
University of Toronto, Vector Institute
Toronto, Canada
{rtqichen, rubanova, jessebett, duvenaud}@cs.toronto.edu

NeuralIPS, 2018 (link)
YouTube explanation from Siraj Raval is here.

Geometric deep learning: going beyond Euclidean data

Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, Pierre Vandergheynst

IEEE Signal Processing Magazine, 2016 (link)

Neural networks interpretation

Visualizing and Understanding Convolutional Networks

Matthew D. Zeiler and Rob Fergus

Dept. of Computer Science,
New York University, USA
{zeiler,fergus}@cs.nyu.edu

Analytical Chemistry Research, 2014 (link)

Adversarial attacks

POLICY FORUM

MACHINE LEARNING

Adversarial attacks on medical machine learning

Emerging vulnerabilities demand new conversations

By **Samuel G. Finlayson[1], John D. Bowers[2], Joichi Ito[3], Jonathan L. Zittrain[2], Andrew L. Beam[4], Isaac S. Kohane[1]**

machine learning into regulatory de
by way of computational surrogat
points and so-called "in silico clinical

Science, March 22, 2019 (link)

MIT Technology Review article

Breast cancer risk predictions

Radiology ORIGINAL RESEARCH • BREAST IMAGING

A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction

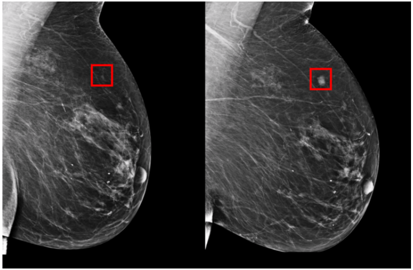
Adam Yala, MEng • Constance Lehman, MD, PhD • Tal Schuster, MS • Tally Portnoi, BS • Regina Barzilay, PhD

From the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 32 Vassar St, 32-G484, Cambridge, MA 02139 (A.Y., T.S., T.P., R.B.); and Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, Mass (C.L.). Received November 28, 2018; revision requested January 18, 2019; revision received March 14; accepted March 18. **Address correspondence to** A.Y. (e-mail: *adamyala@csail.mit.edu*).

Conflicts of interest are listed at the end of this article.

See also the editorial by Sitek and Wolfe in this issue.

Radiology 2019; 00:1–7 • https://doi.org/10.1148/radiol.2019182716 • Content code: BR



Radiology, May 7, 2019 (link)

- predicts breast cancer in 4 years before it actually occurs

nature medicine
LETTERS
https://doi.org/10.1038/s41591-019-0447-x

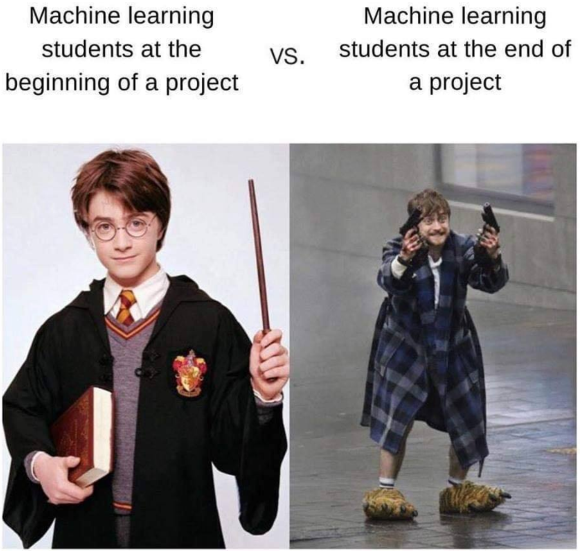
End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography

Diego Ardila[1,5], Atilla P. Kiraly[1,5], Sujeeth Bharadwaj[1,5], Bokyung Choi[1,5], Joshua J. Reicher[2], Lily Peng[1], Daniel Tse[1*], Mozziyar Etemadi[3], Wenxing Ye[1], Greg Corrado[1], David P. Naidich[4] and Shravya Shetty[1]

Nature Medicine, May 20, 2019 (link)

- outperforms ensemble of experienced radiologists

Deep Learning is hard... tribute to Fast.AI



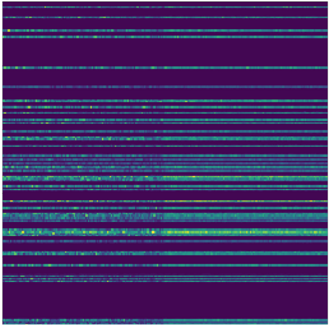
Understand something is:

1. learn how to use it
2. become familiar with it

Example of Fast.AI library usage

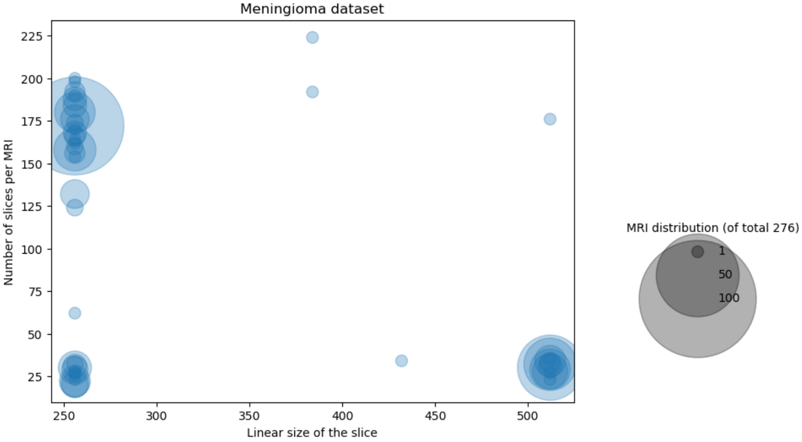
The Mystery of the Origin — Cancer Type Classification using Fast.AI Library

Alena Harley Follow
Oct 30, 2018 · 6 min read

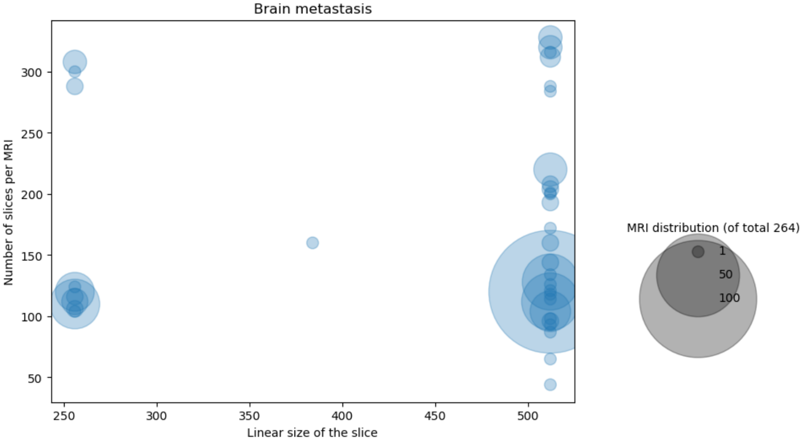


Alena Harley, October 30, 2018 (towardsdatascience.com)

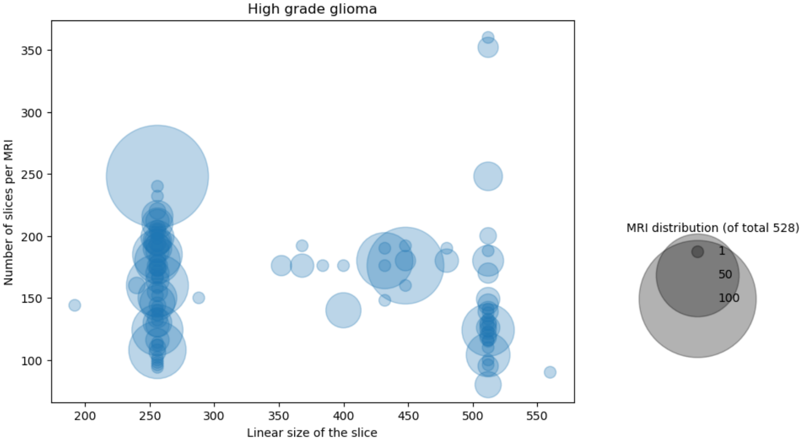
Meningioma dataset



Brain metastasis dataset



High grade glioma dataset



Classification data splitting

Number of non-empty slices:

- Gliomat - 16594
- Meningioma - 7607
- Metastasis - 5806

Dataset splitting:

- Train - 70%
- Validate - 20%
- Test - 10%

Data was splitted by patient, i.e. slices from the same MRI can not be in more than one subgroup.

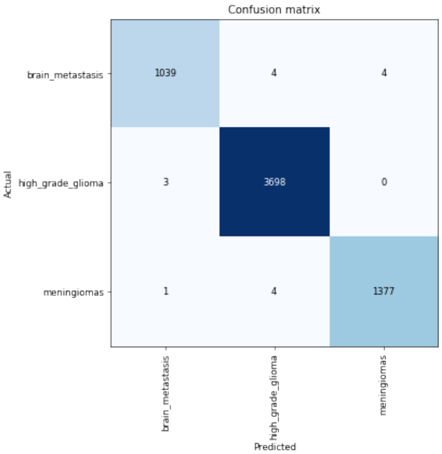
Images were resized to 224×224
ANN architecture - ResNet34.

Training losses

epoch	train_loss	valid_loss	error_rate	time
0	0.202105	0.093568	0.034584	00:42
1	0.084514	0.022100	0.007504	00:41
2	0.048037	0.014657	0.005873	00:42
3	0.033046	0.011079	0.003589	00:42

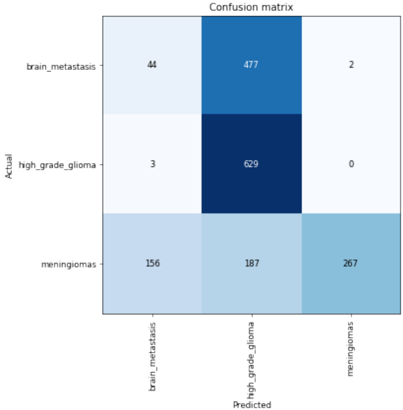
Fine-tuning losses

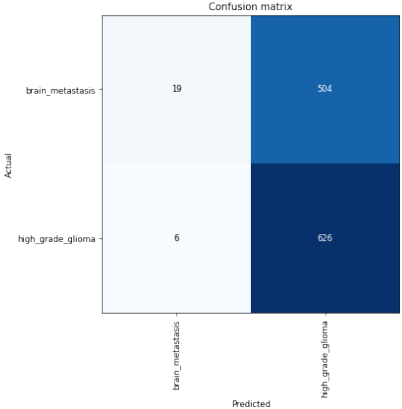
epoch	train_loss	valid_loss	error_rate	time
0	0.027540	0.009791	0.003426	00:52
1	0.022103	0.008221	0.002610	00:52



Validation data is used during the training process for classifier quality evaluation. Confusion matrix above shows how ANN approximated train dataset with respect to validation dataset criterion.

Confusion matrices on the unseen test data





So we see that classifier does not work. Meningioma is distributed over all classes. Brain metastasis are put in the HGG class.

Ideas for improvement

- Super-resolution like pre-processing;
- Generative Adversarial Networks generated datasets;
- "Dumb" 3D synthetic dataset formed with ellipsoids, MRI structure and noise (Gaussian, Rice etc.). Probably I will run it on Google Colab platform
- Use Google's Open Images V5 dataset for 2D segmentation pretration.
- Use federated learning - centralized model with decentralized data (details are in Google comics here)

Competitions

Current competitions:

- PAIP2019
- ...

Aim is to reach TOP 10%. People there usually know what are they doing. More medical related ML competition can be found here.

¡Gracias!