

# Deep Learning papers related to MoLAB activities

Anton Popov

UCLM, MoLAB, Spain, Ciudad Real

May 28, 2019

- **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.

## 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...*)

## 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 Foltyń, 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 Bendszust, Klaus H Maier-Heintz

Lancet Oncology ([link](#)); April 2, 2019

- **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	Heidelberg training dataset
Test	239 2034	40 532	Heidelberg testing dataset EORTC-26101 study ( <a href="#">link</a> )
Simulate	595	466	Heidelberg simulation dataset

All Heidelberg 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 → expressive ANN is needed
"single timepoint: preoperatively from initial diagnosis or early postoperatively ... or at follow-up and was specifically assembled to represent the broad phenotypic appearance of brain tumours on MRI during disease evolution"	preoperatively, postoperatively, follow-up mixed together → 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).

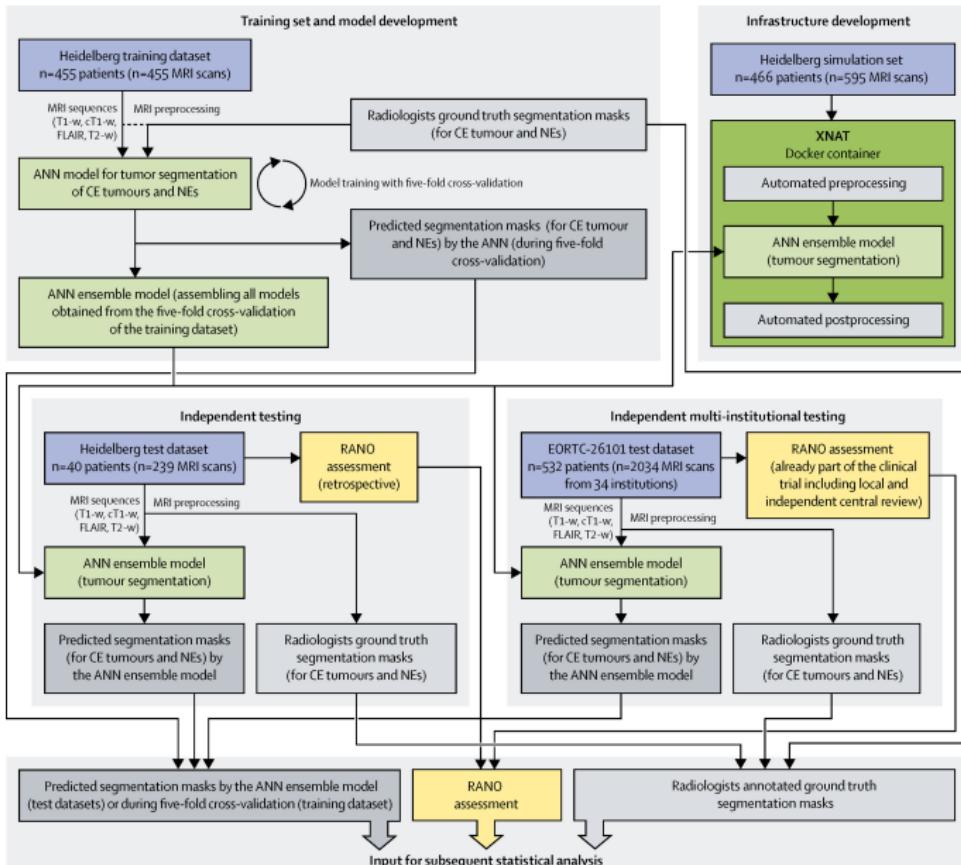
## Test

This dataset was a longitudinal dataset with preoperative and consecutive follow-up scans

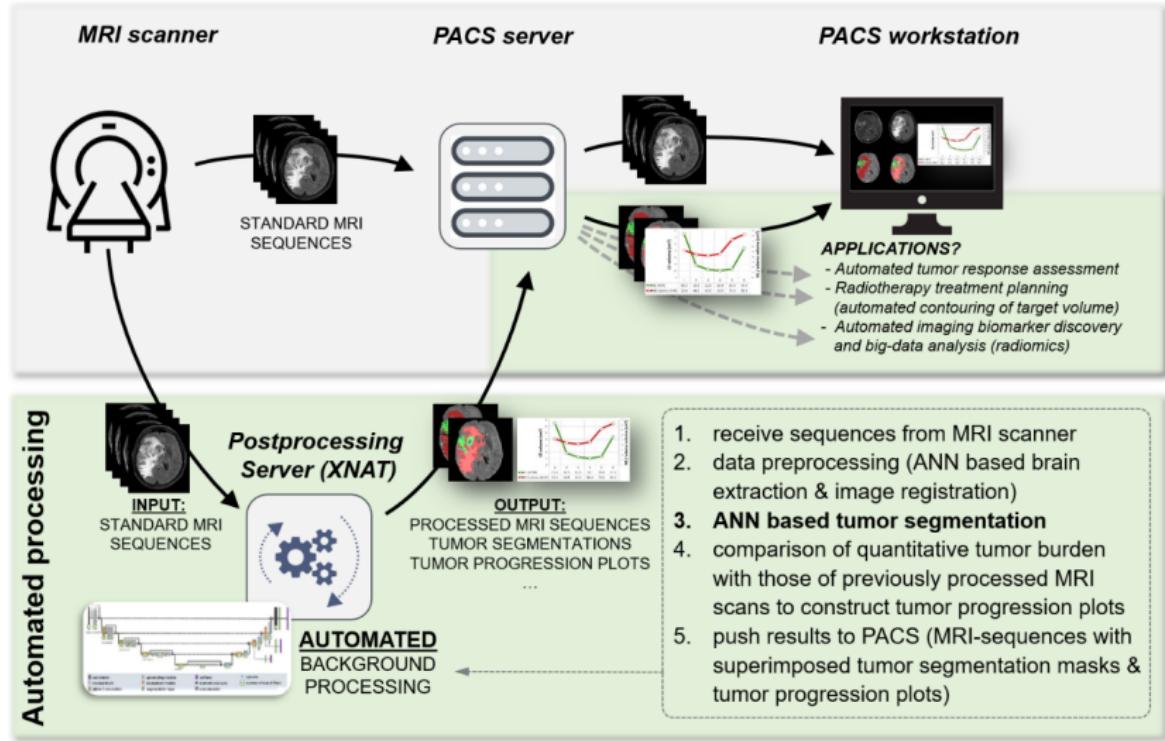
## Simulate

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



# Infrastructure



# Architecture from Elies Fuster presentation

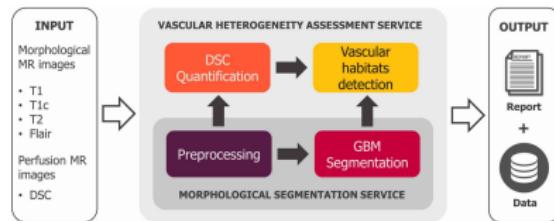


Figure: MRI → processing → output

Notes:

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

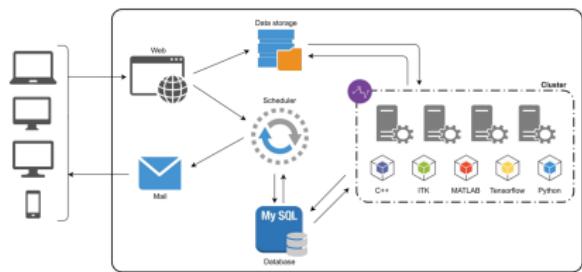


Figure: Standard Web-app

# Proof of concept - Classification Web App

## Classify Tree vs Flower vs Grass images

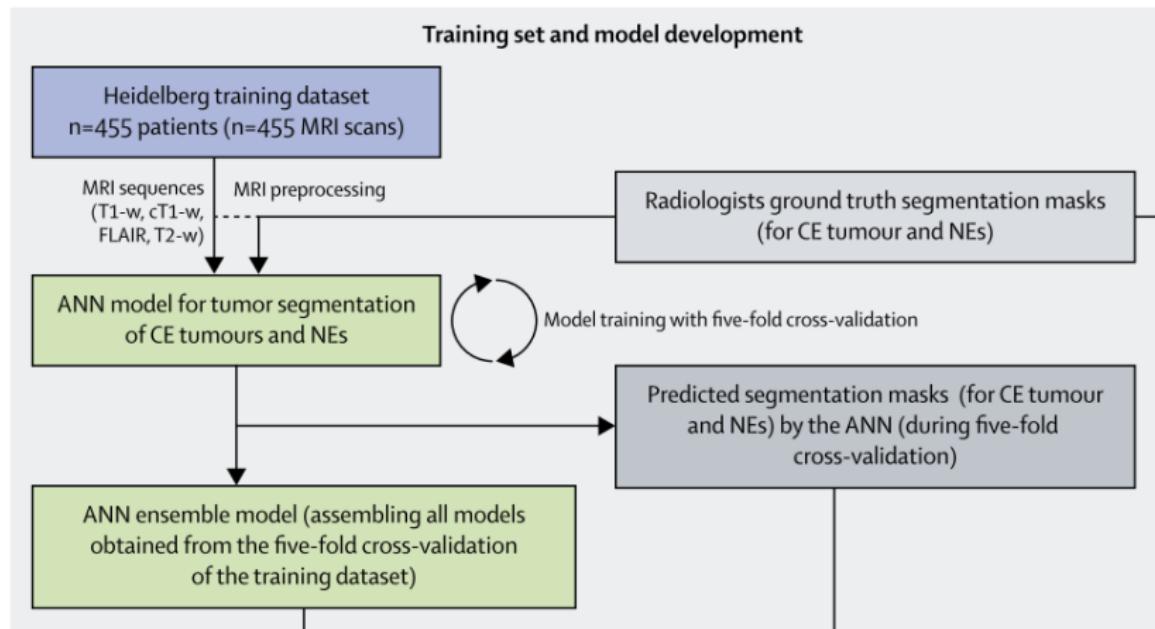
Use images of **tree, flower, grass**, or all three!

Select Image

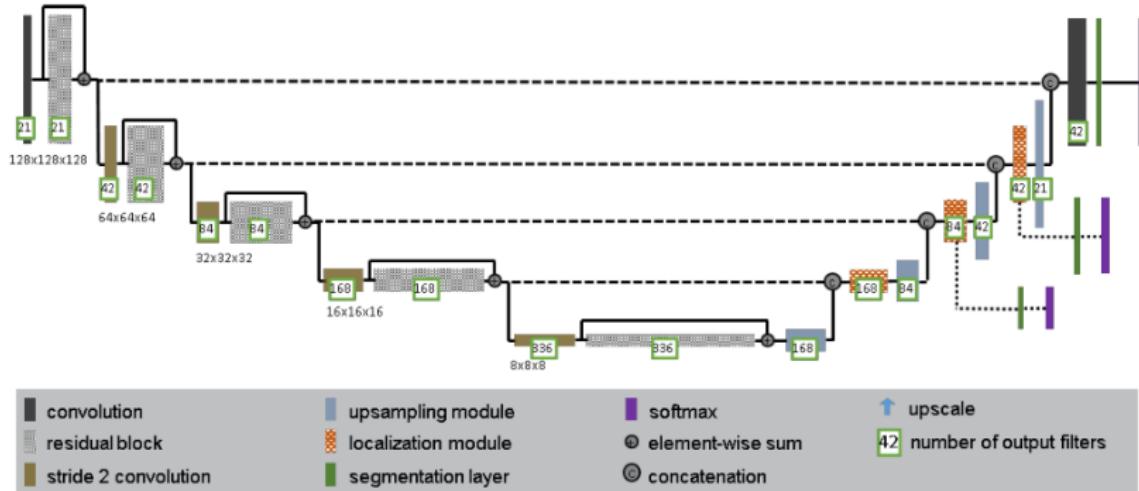
No file chosen

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

# Training set and model development

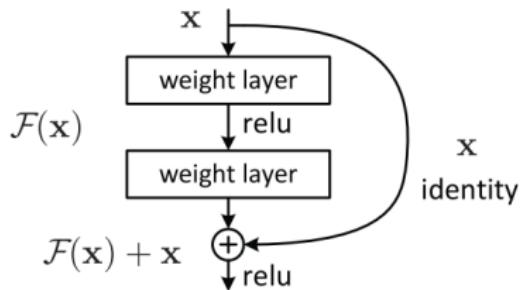


# Neural Network architecture

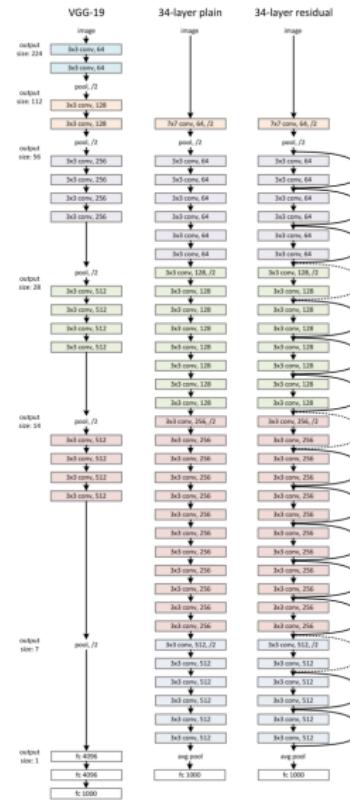


Inspired by [U-Net](#) and utilizes [residual blocks](#) in encoder.

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

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

NeurIPS, 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](#))

## 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](#))

## POLICY FORUM

### MACHINE LEARNING

# *Adversarial attacks on medical machine learning*

Emerging vulnerabilities demand new conversations

By Samuel G. Finlayson<sup>1</sup>, John D. Bowers<sup>2</sup>,  
Joichi Ito<sup>3</sup>, Jonathan L. Zittrain<sup>2</sup>, Andrew  
L. Beam<sup>4</sup>, Isaac S. Kohane<sup>1</sup>

machine learning into regulatory decisions by way of computational surrogates, 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

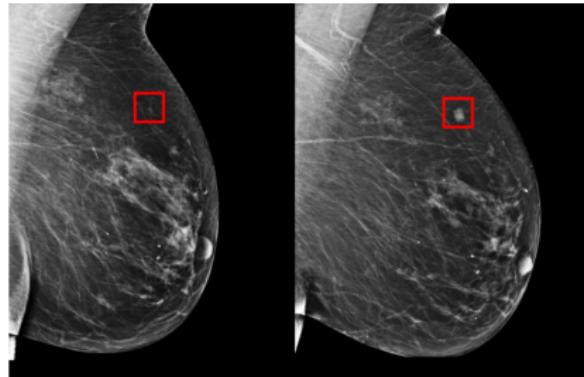
Adam Yala, MEng • Constance Lebowitz, MD, PhD • Tal Schuster, MS • Tally Parnasi, BS • Regina Barzilay, PhD

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: adam.yala@mit.edu or rbb).

Conflict of interest: no conflict exists for drugs or devices used in a study if they are not being evaluated as part of investigation.

See also the editorial by Stark and Wible in this issue.

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



Radiology, May 7, 2019 ([link](#))

- predicts breast cancer in 4 years before it actually occurs

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

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

Nature Medicine, May 20, 2019 ([link](#))

- outperforms ensemble of experienced radiologists

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

Machine learning  
students at the  
beginning of a project

vs.

Machine learning  
students at the end of  
a project



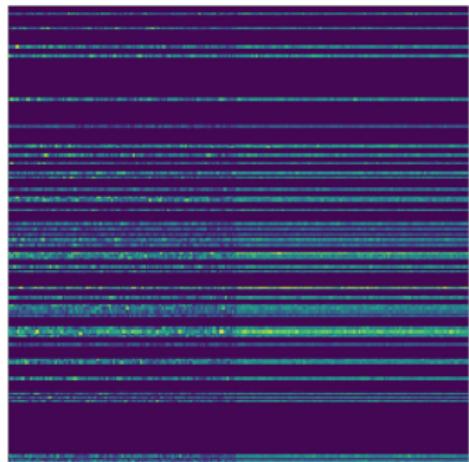
**Understand  
something  
is:**

- ➊ learn how to use it
- ➋ become familiar with it

# Example of Fast.AI library usage

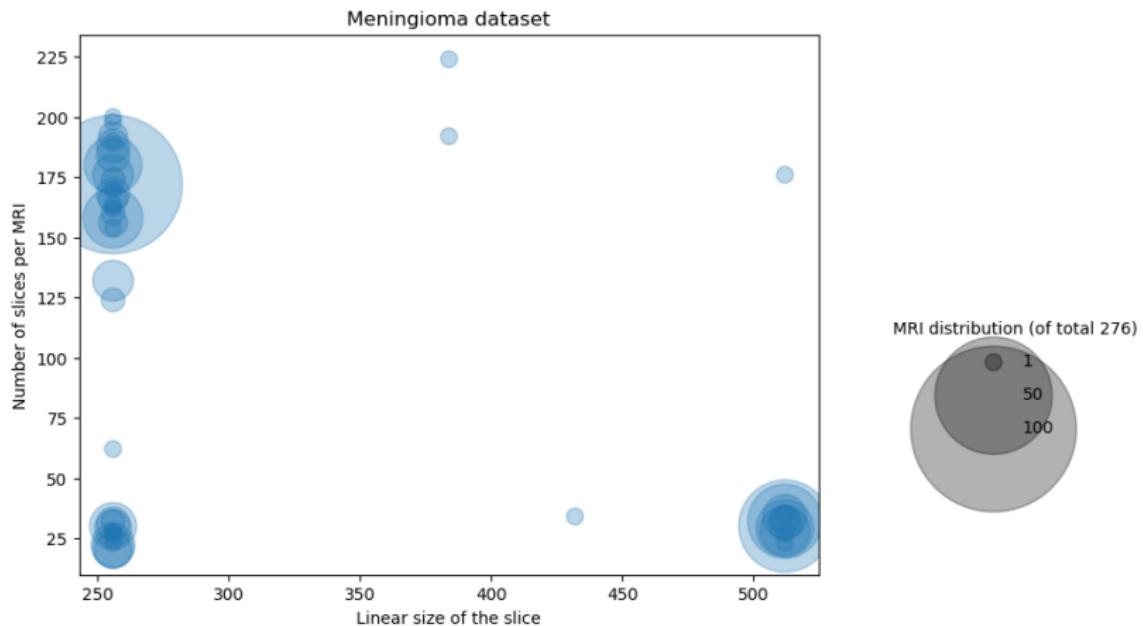
## The Mystery of the Origin—Cancer Type Classification using Fast.AI Library

 Alena Harley [Follow](#)  
Oct 30, 2018 · 6 min read

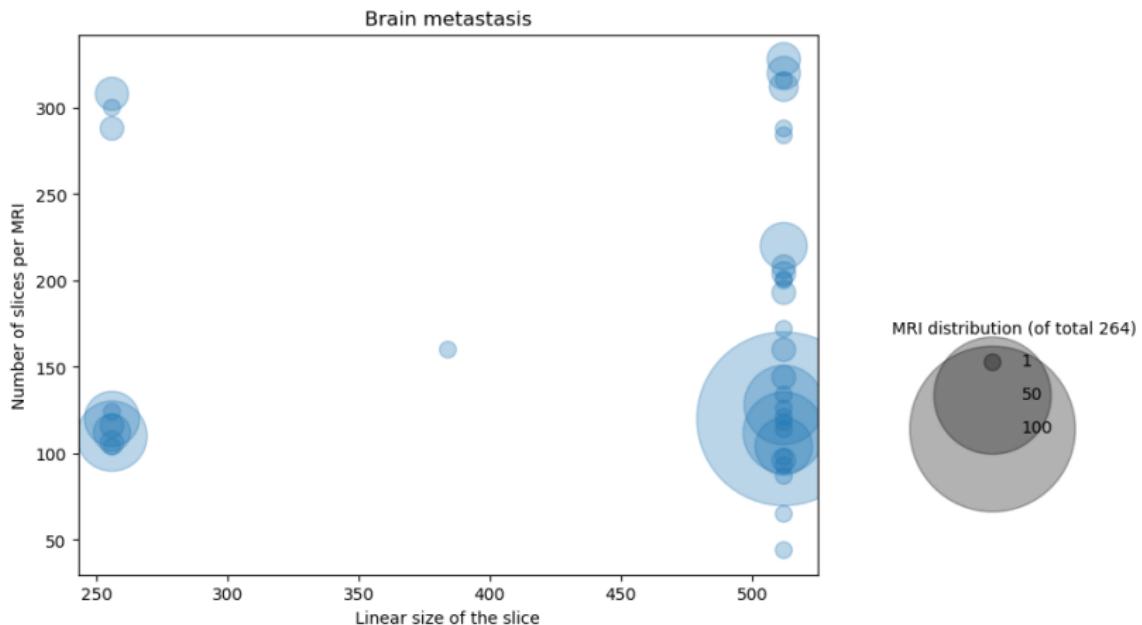


Alena Harley, October 30, 2018 ([towardsdatascience.com](https://towardsdatascience.com/the-mystery-of-the-origin-cancer-type-classification-using-fast-ai-library-3a2a2f3a2))

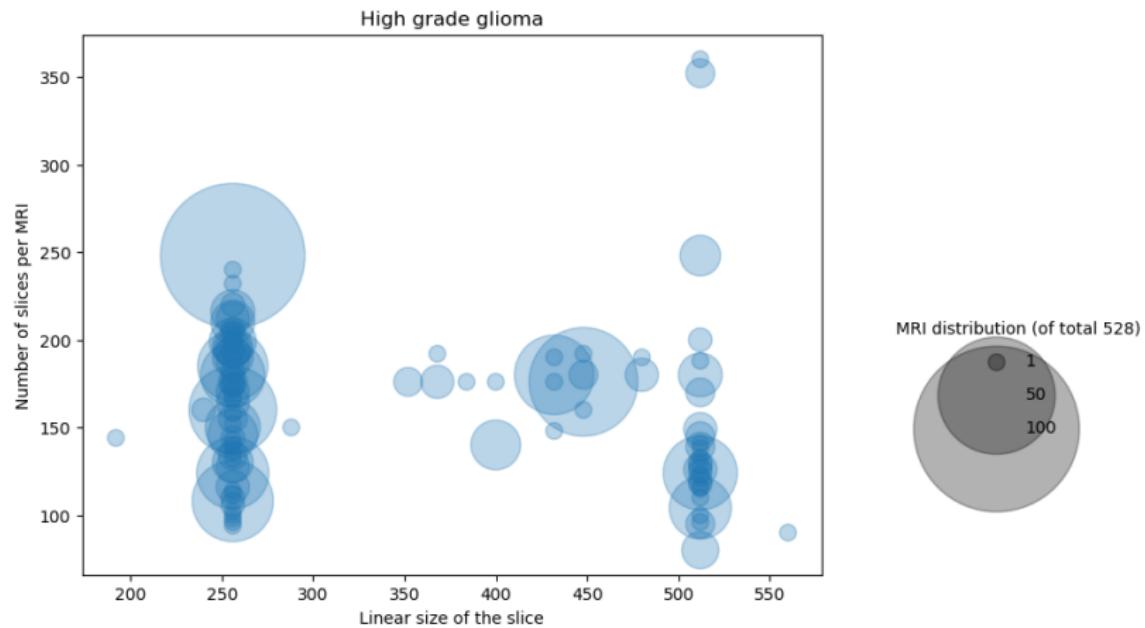
# Meningioma dataset



# 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  $224 \times 224$   
ANN architecture - ResNet34.

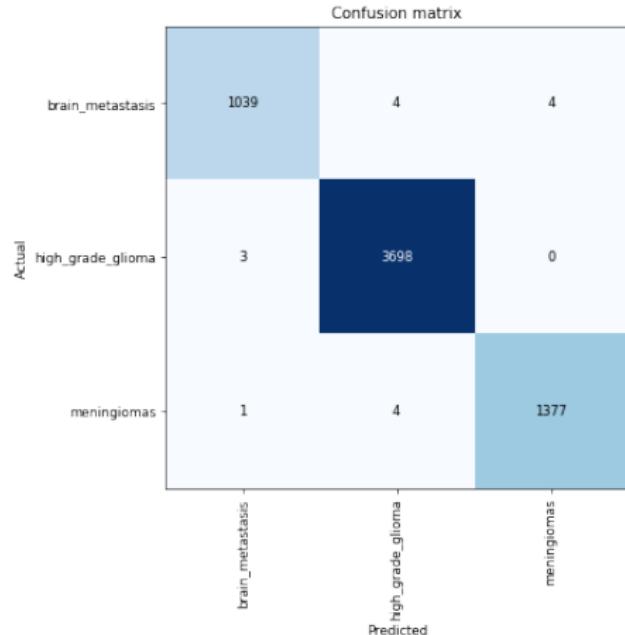
# Confusion matrix on validation data

## 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

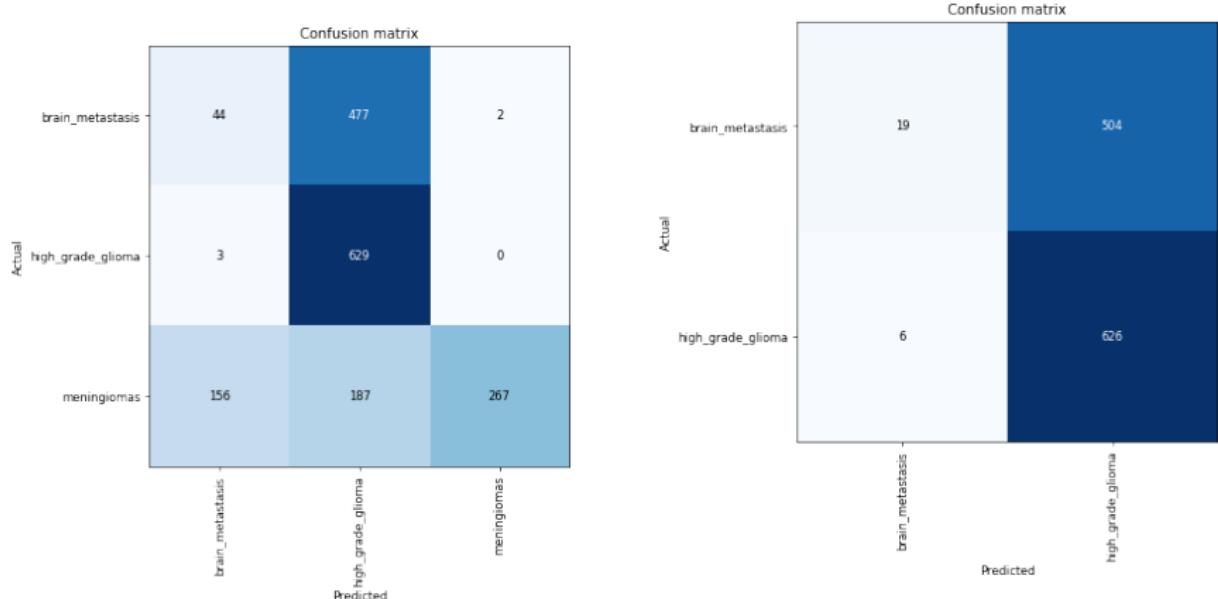
## Fine-tuning losses

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

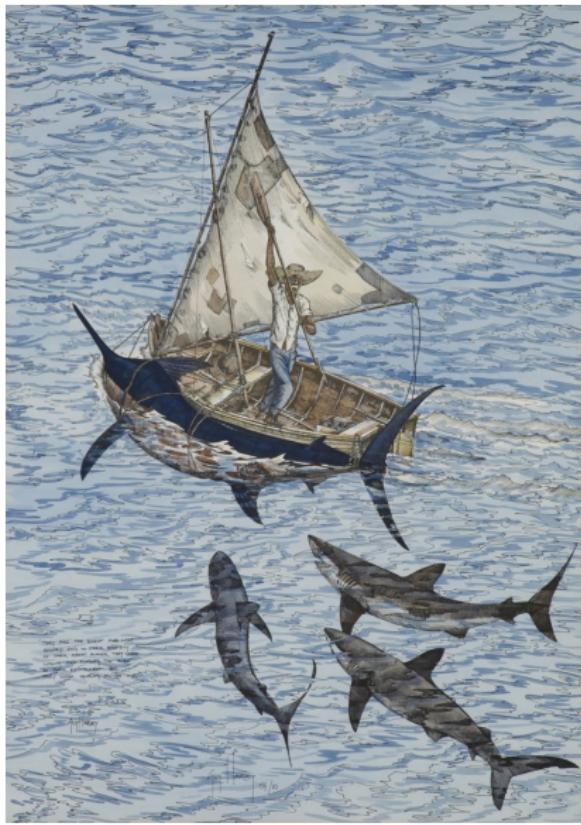


*So we see that classifier does not work. Meningioma is distributed over all classes. Brain metastasis are put in the HGG class.*

# 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!

