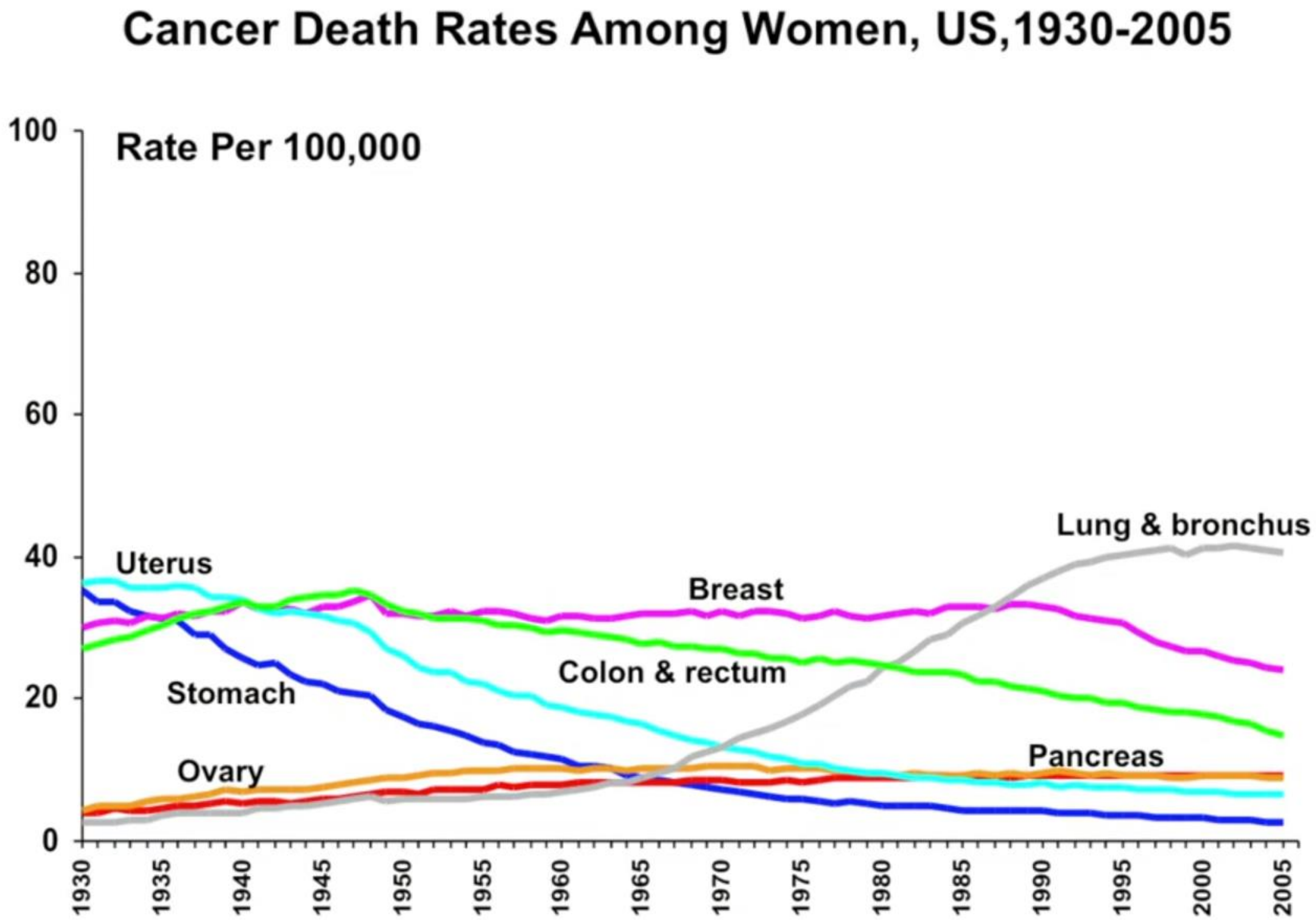
Predible

# Unsupervised and Semi-Supervised Deep Learning in Medical Imaging

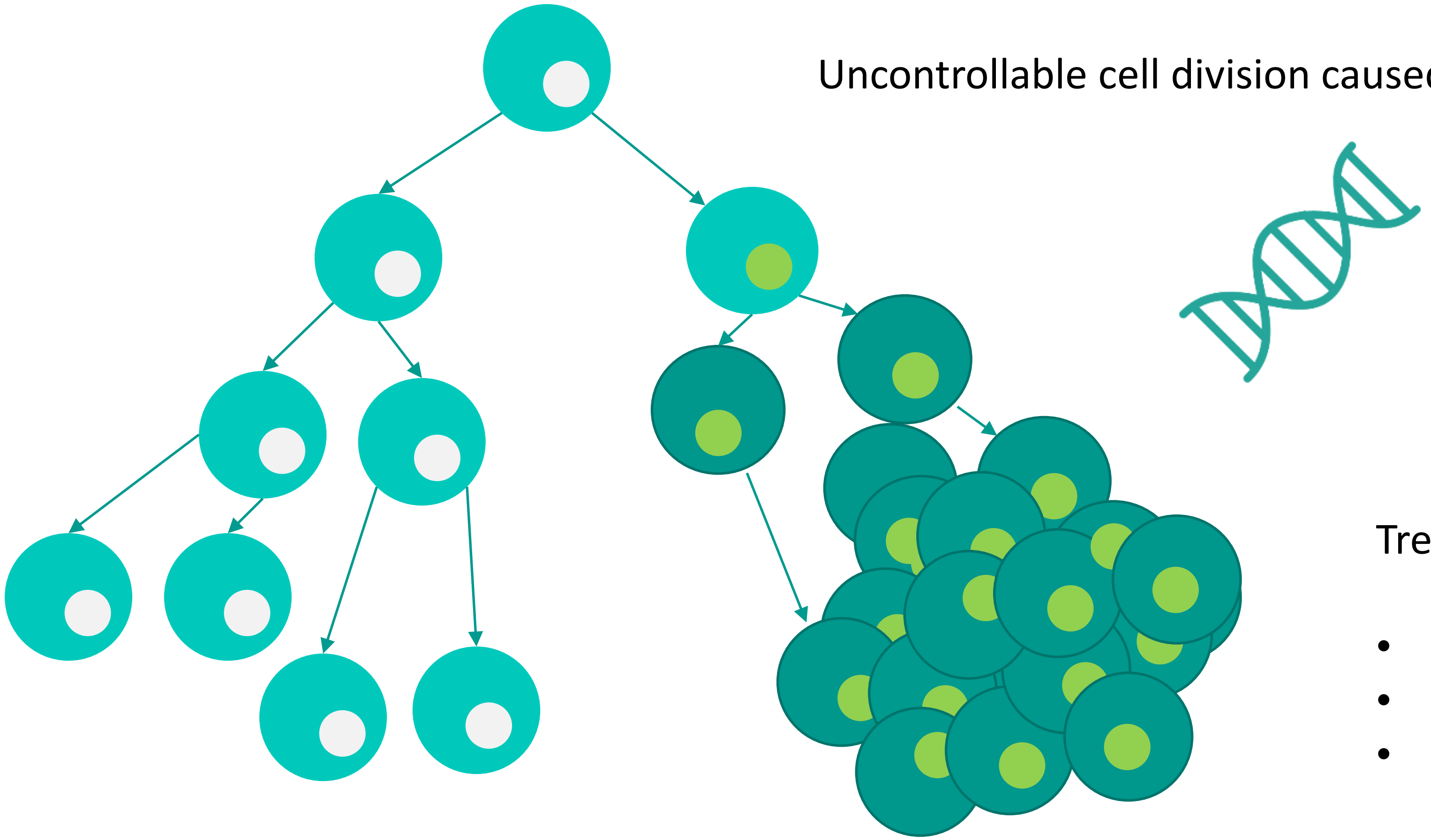
**Kiran Vaidhya**
Algorithms Researcher
kiran@predible.co

# Cancer




Source: http://www.clevelandclinicmeded.com

# How does cancer occur?



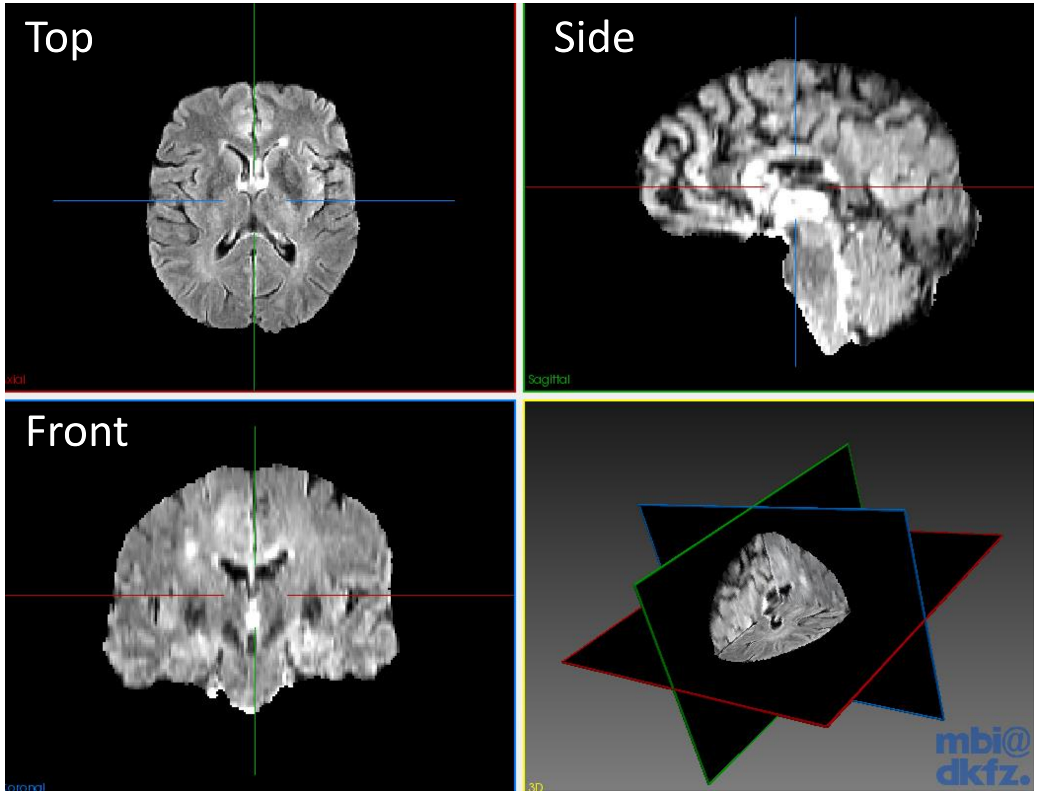
Uncontrollable cell division caused by genetic mutations

Treatments:

- Surgery
- Chemotherapy
- Radiation therapy

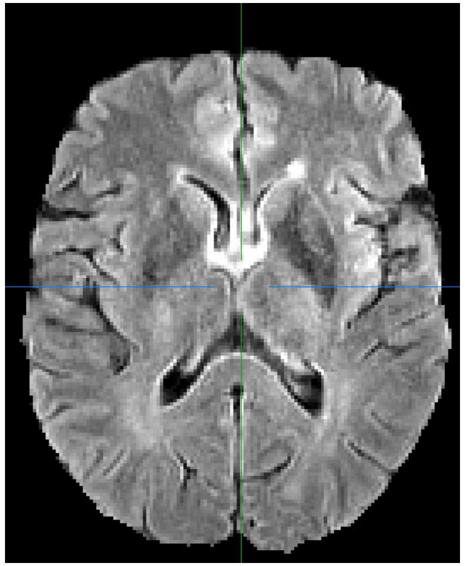
# How do we “see” inside the body?


MRI scanner

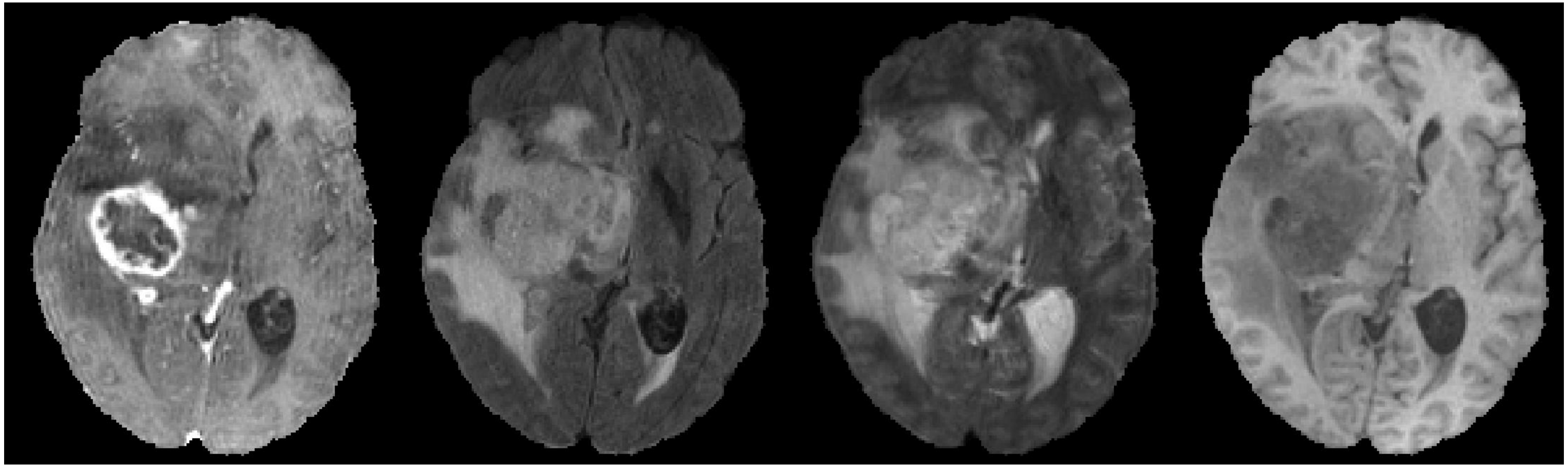



Visualization of the 3D image

# Spotting brain cancer



Normal brain

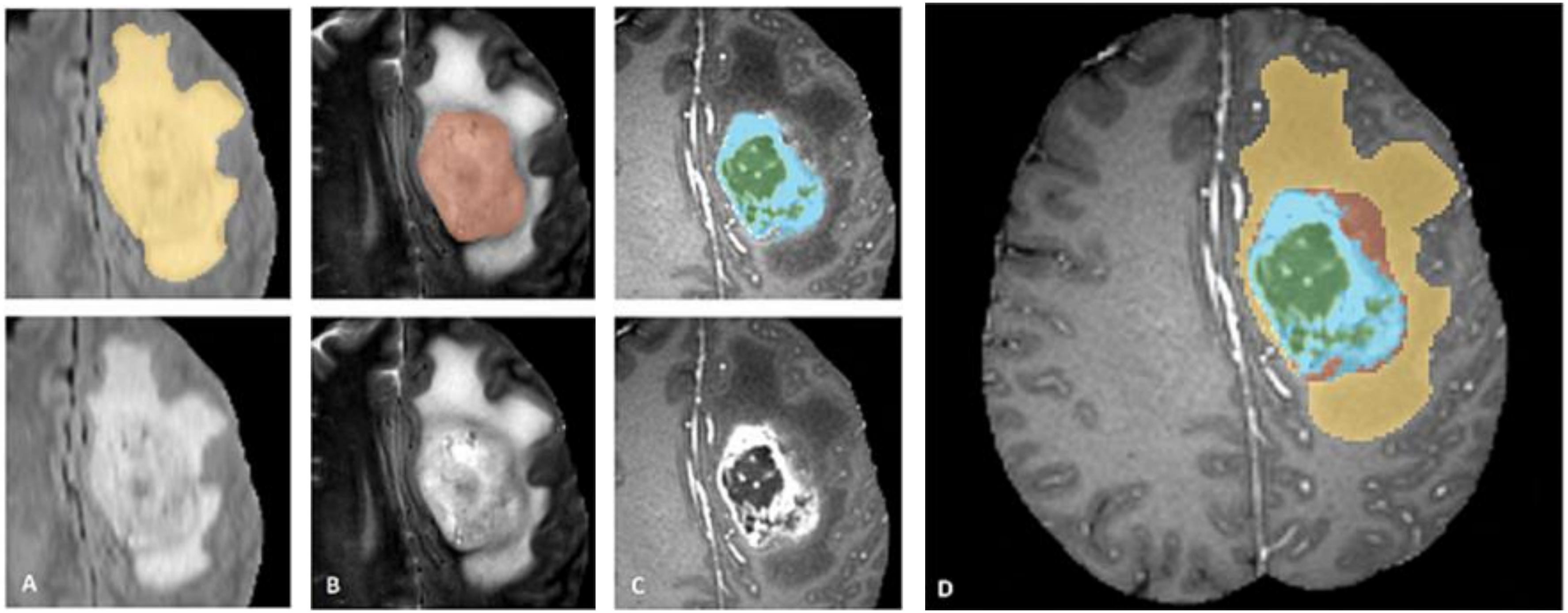


T1 contrast | Flair | T2 | T1

## What is Glioblastoma?

- Most aggressive and most malignant **brain cancer**
- Only 2% survive post treatment | Median survival of **14 months**

# Knowledge of tumor sub-types help in treatment

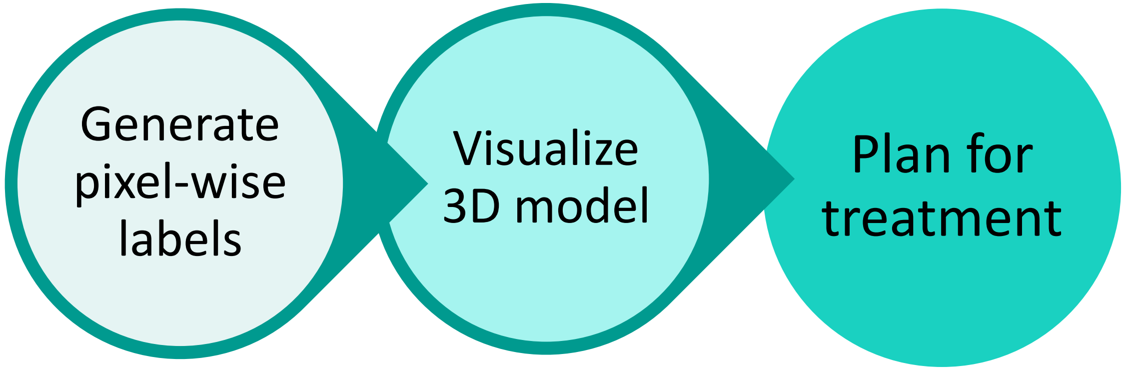


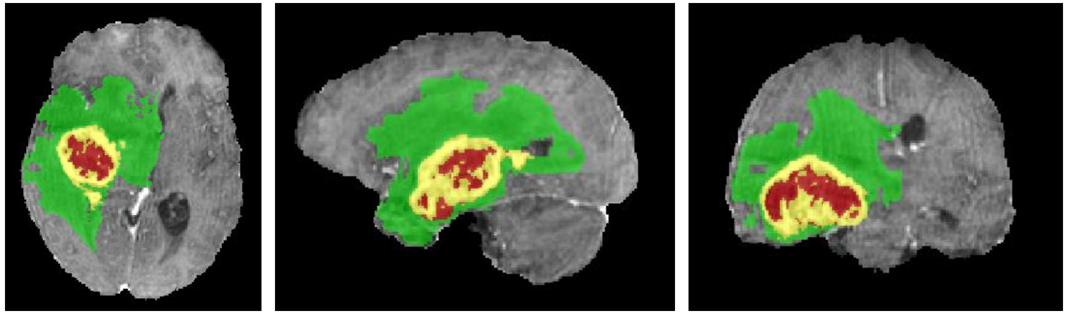

Whole tumor | Tumor core | Active tumor | Entire tumor map

Intra-tumor classification is essential to understand treatment response

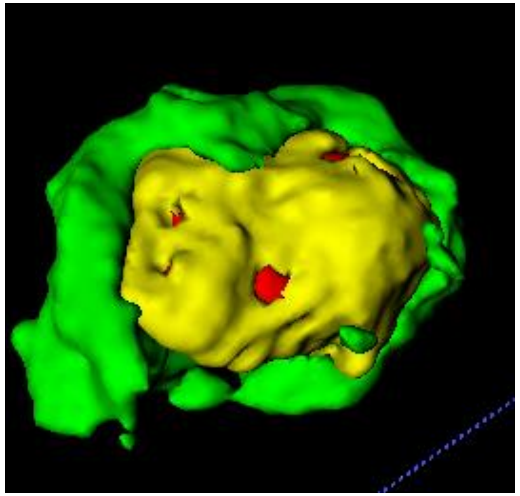
Source: http://braintumorsegmentation.org/

# Glioblastoma treatment requires pixel-wise labelling



- Tedious **slice-by-slice** labelling is carried out by doctors
- Labelling can be performed by **deep networks**



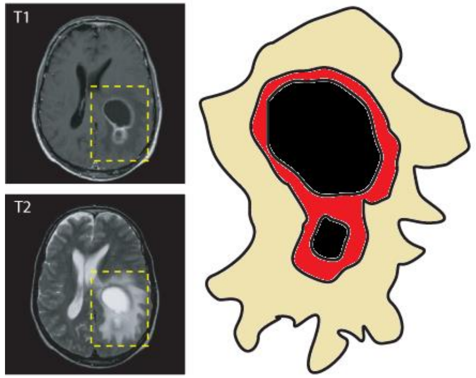
3D rendering of glioblastoma

# Glioblastoma segmentation from brain MRI is non-trivial

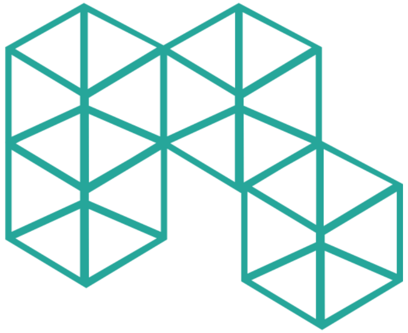
## Shortage of samples

- Limited amount of annotated data – **overfitting**
- Only **2%** of pixels contain tumor

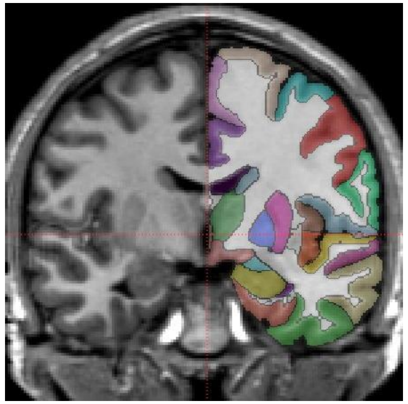
**< 300 scans**




Heterogeneity



4D data



Cost: 2449$ ~ Rs. 1,60,000

## Complexity of data

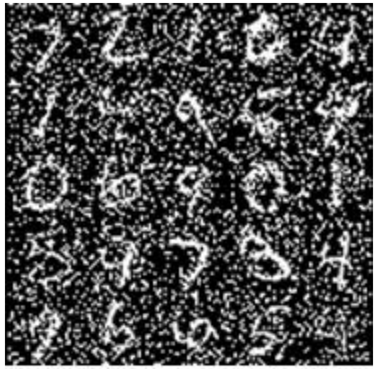
- **[4 x 155 x 240 x 240]** tensors
- Dense annotations are very expensive!

Can we leverage unsupervised feature learning?

# Deep Unsupervised Feature Extraction

# Can we use Auto-encoders to extract features?

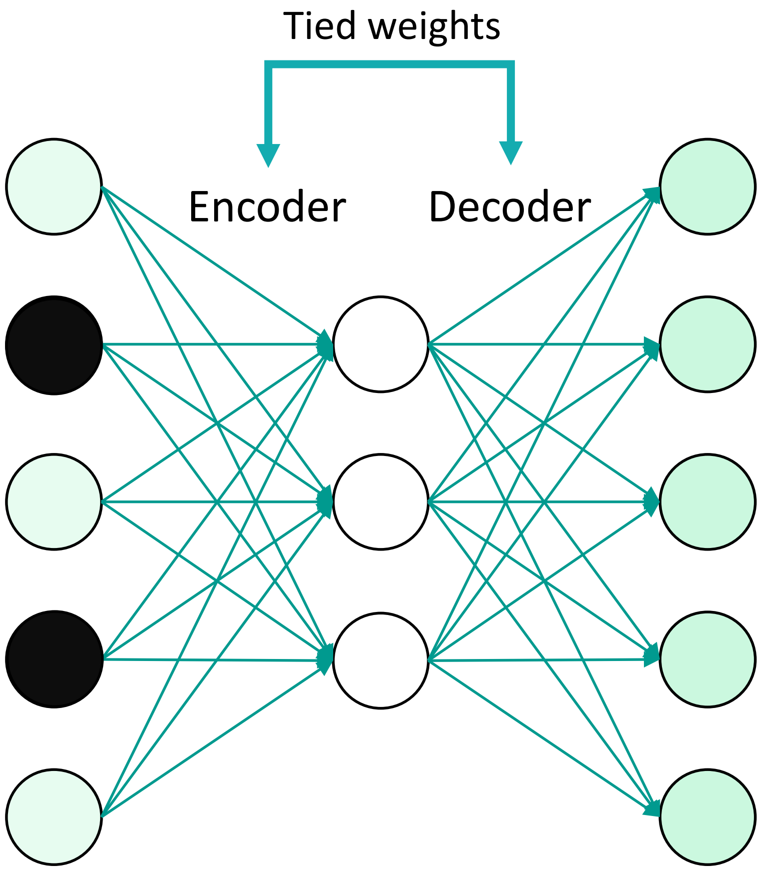
28x28



## Encode

$$\boldsymbol{h} = \boldsymbol{\sigma}(\boldsymbol{W}.\widetilde{\boldsymbol{x}})$$

Prevent identity mapping!




Denoising Autoencoder

28x28

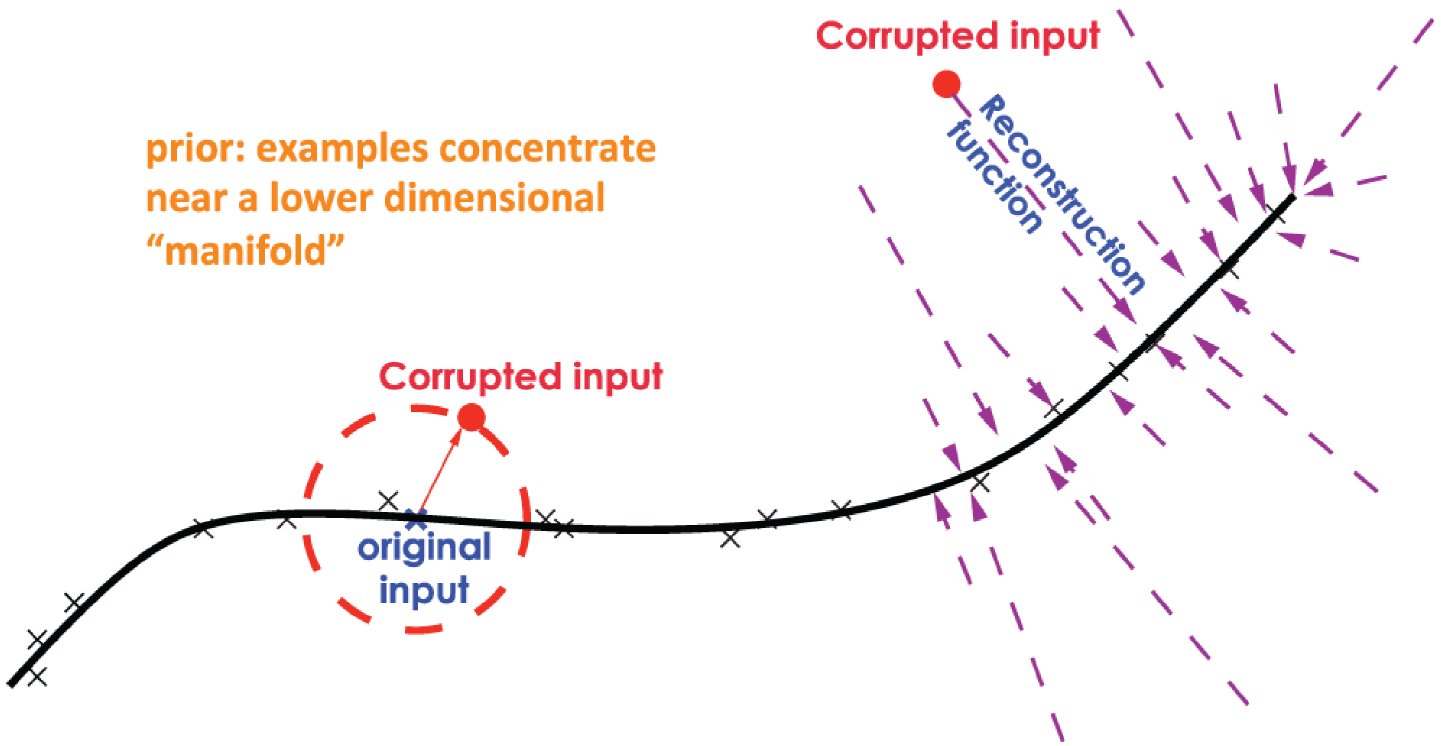
## Decode

$$\boldsymbol{f}(\boldsymbol{h}) = \boldsymbol{W}^T . \boldsymbol{h}$$

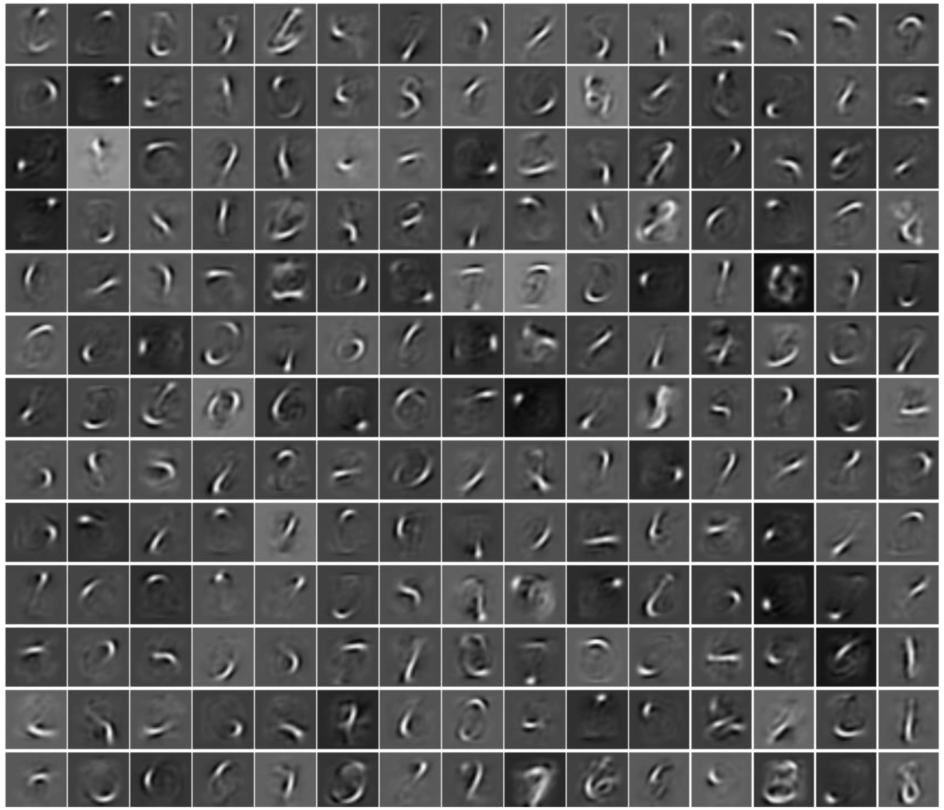
## Compute loss

$$\textbf{Loss} = \sum(\boldsymbol{f} - \boldsymbol{x})^2$$

# How do Auto-encoders learn underlying structure?



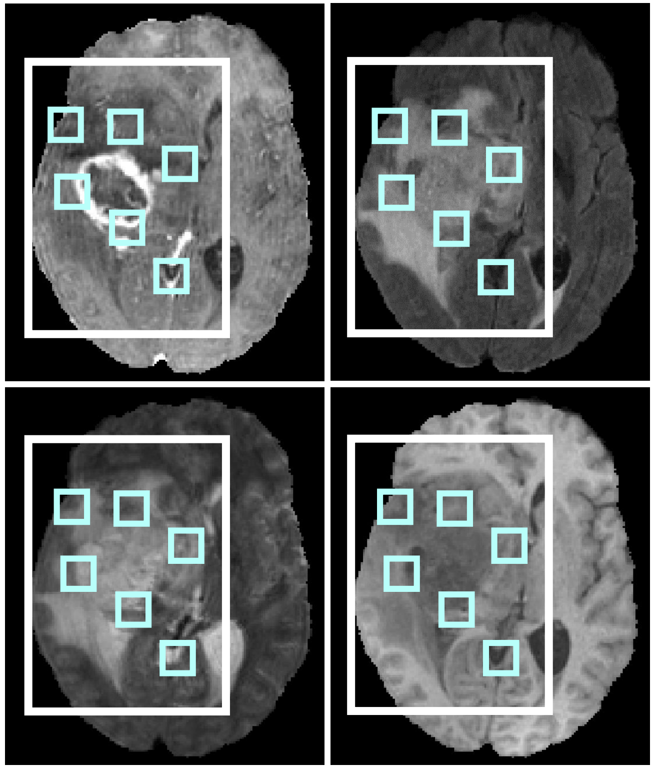

Learn Reconstruction Function



Weights of Autoencoder

Source: Stacked Denoising Autoencoder – Vincent et al (2010)

# From MNIST to brain MRI?



BRATS 2015 dataset

Patch size = [ 4 x 21 x 21 ]

Extract small patches

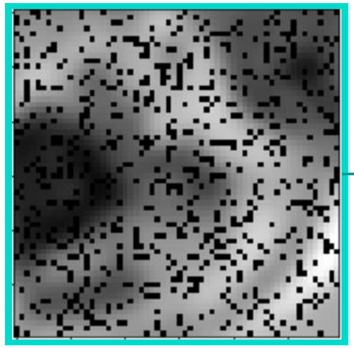
- Extract ROI around tumor
- Sample patches from the ROI

Samples of [4 x 21 x 21] patches
extracted around tumor

# Training Autoencoders on brain MRI

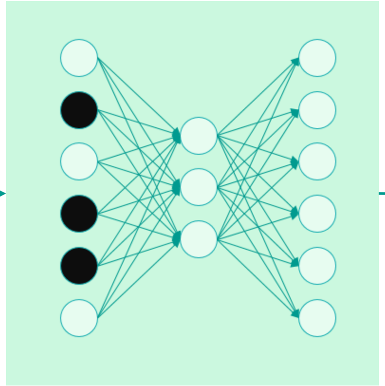
3D patch size: 4 x 21 x 21

1728 – 3500 – 1728

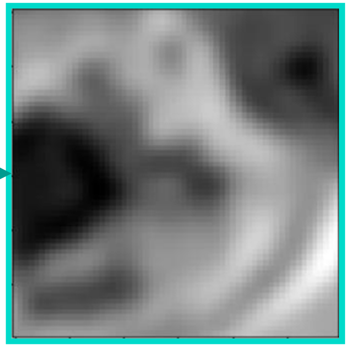


Feed noisy patches

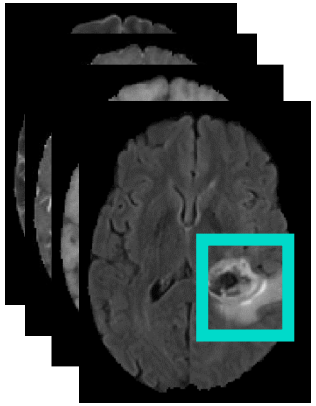
Drop 20% pixels

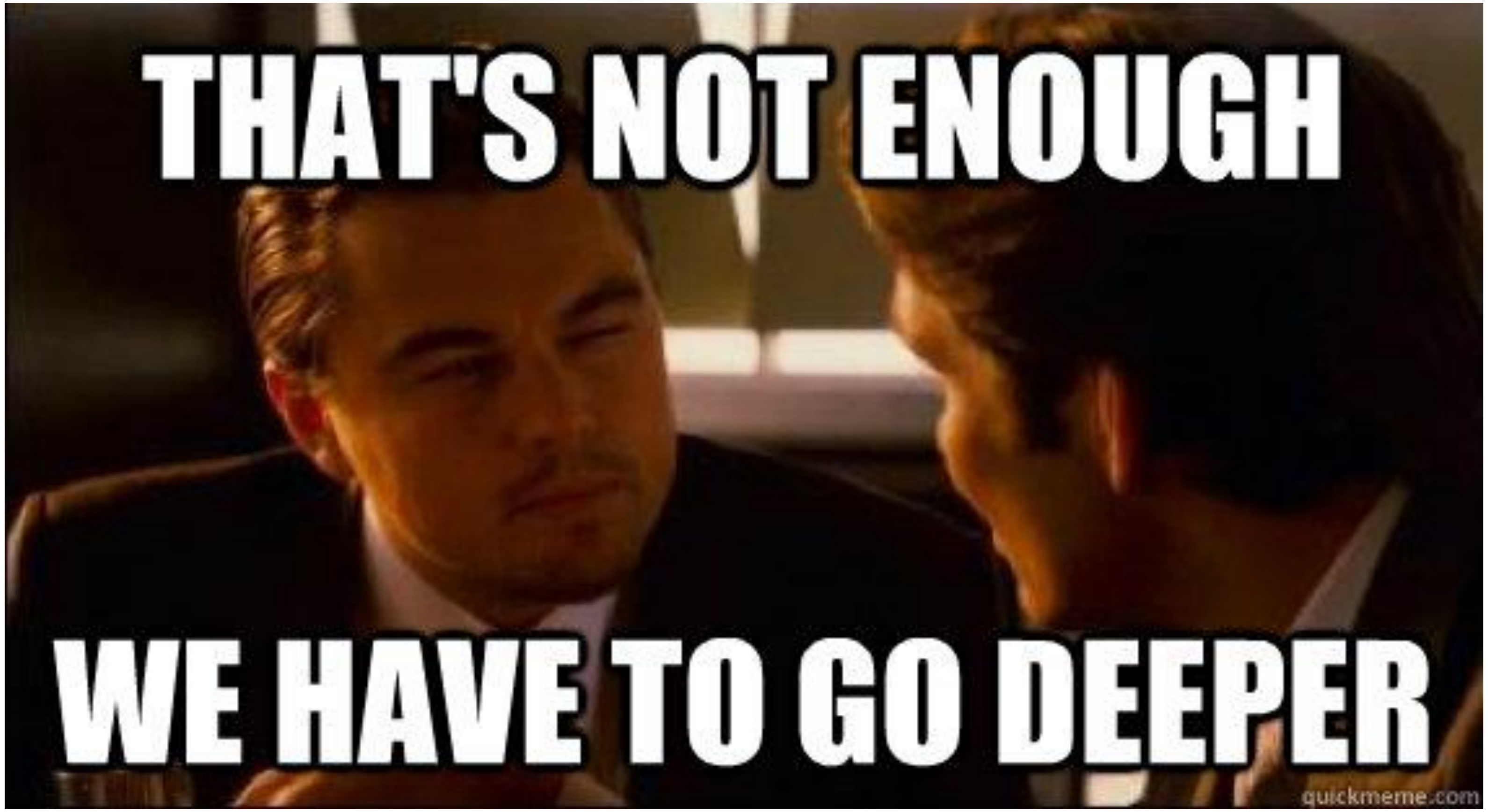


Reconstruct original patch



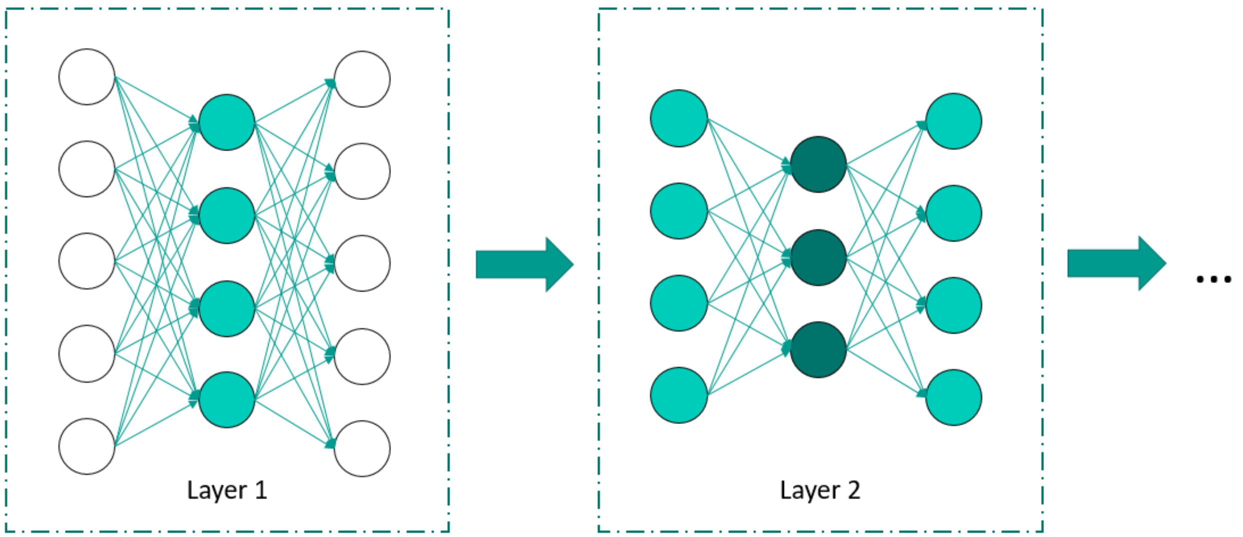
**Extract, noise, reconstruct!**



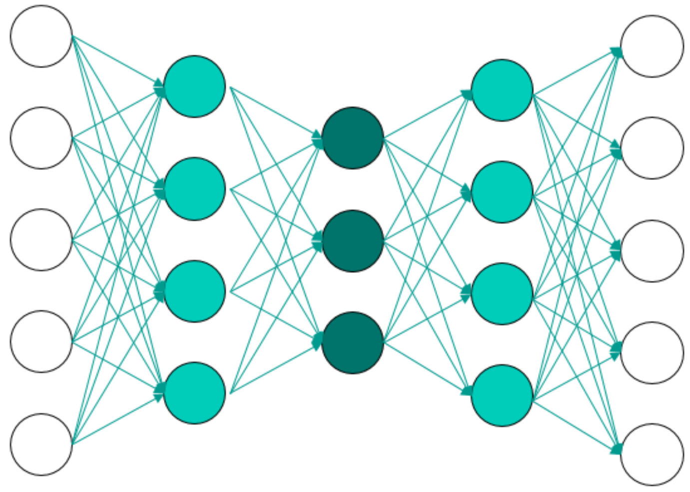

THAT'S NOT ENOUGH
WE HAVE TO GO DEEPER
quickmeme.com

# How do we train deep Auto-encoders?

- Deep layers – learn a hierarchy of features
- Vanishing and exploding gradients
- Pre-train **layer by layer**




Pre-training

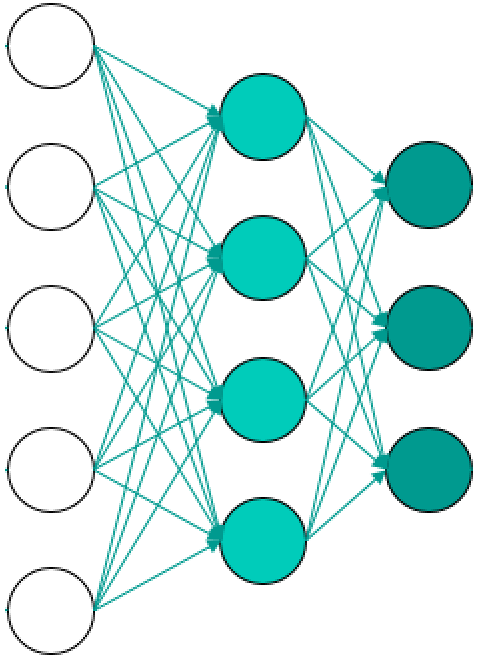


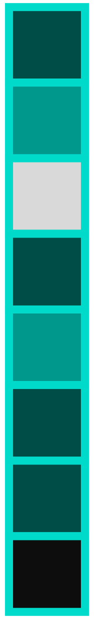
Stacking

# Learn rich latent representations

**Don’t** noise during inference

[4x21x21] patch



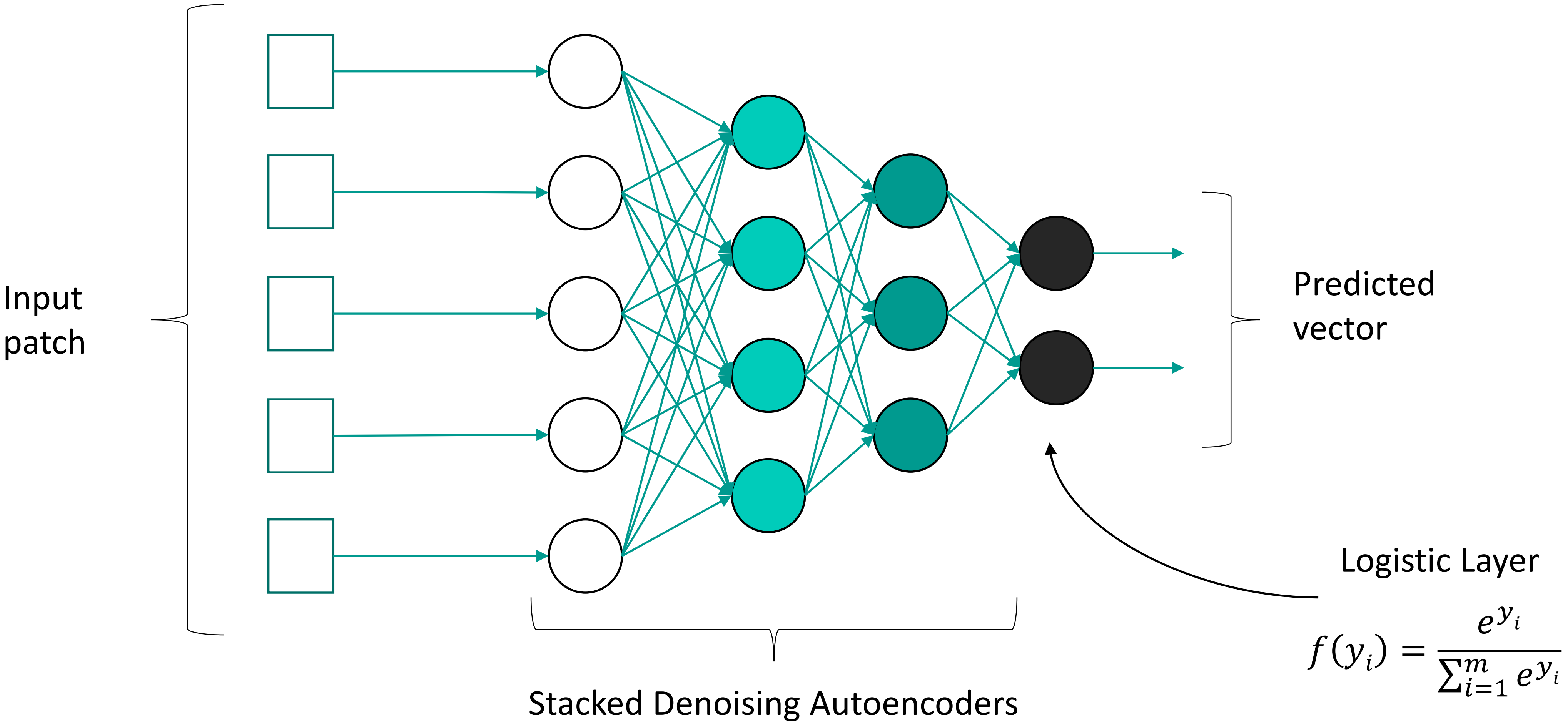


Extracted feature representation

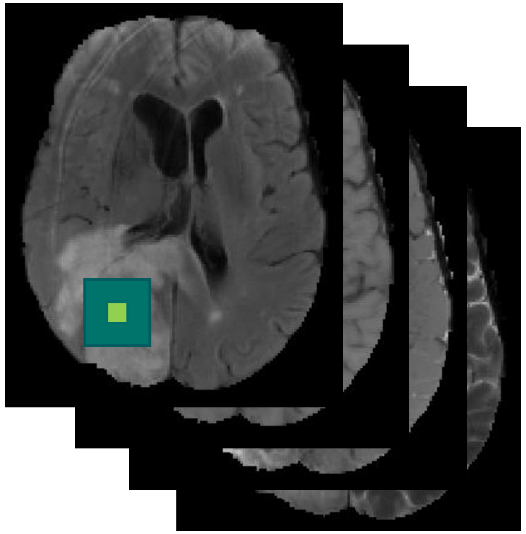
 

# Fine-tuning for Classification

# How do we use Autoencoders for classification?

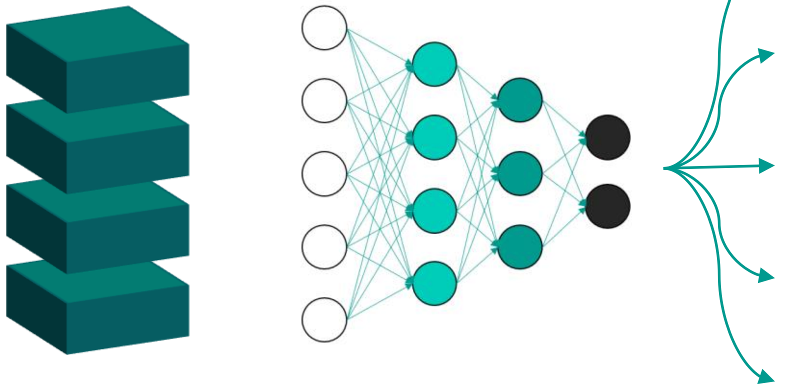


# Fine-tune the network for classification



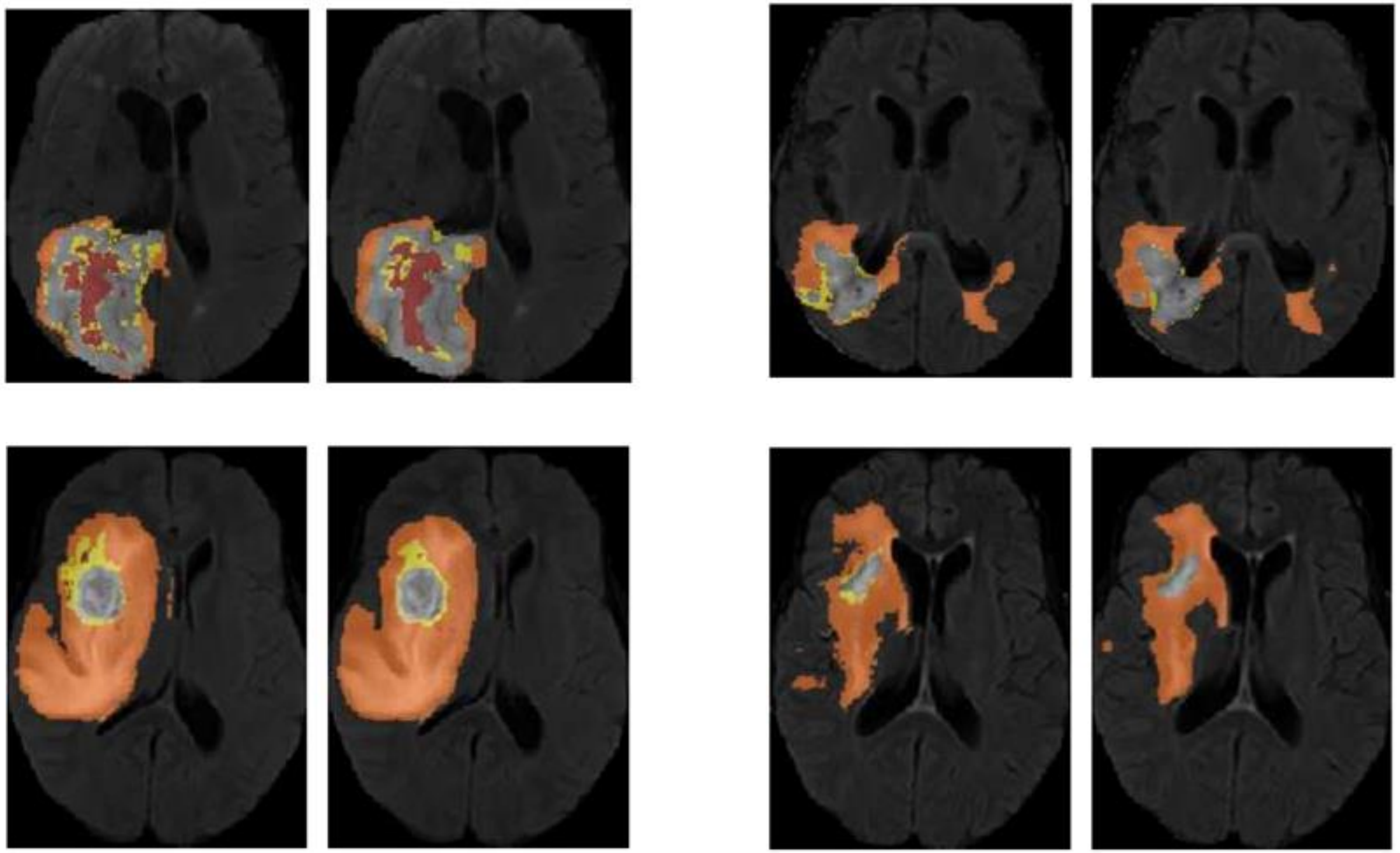
4 sequences of MRI

| Architecture | | | | |
|---|---|---|---|---|
| 1764 | 3500 | 2000 | 1000 | 5 |



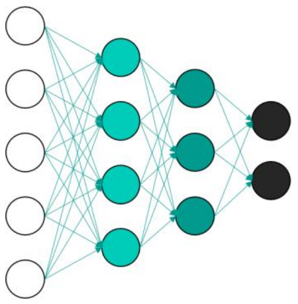

**Extract, classify, stride, repeat**

# Segmentation results



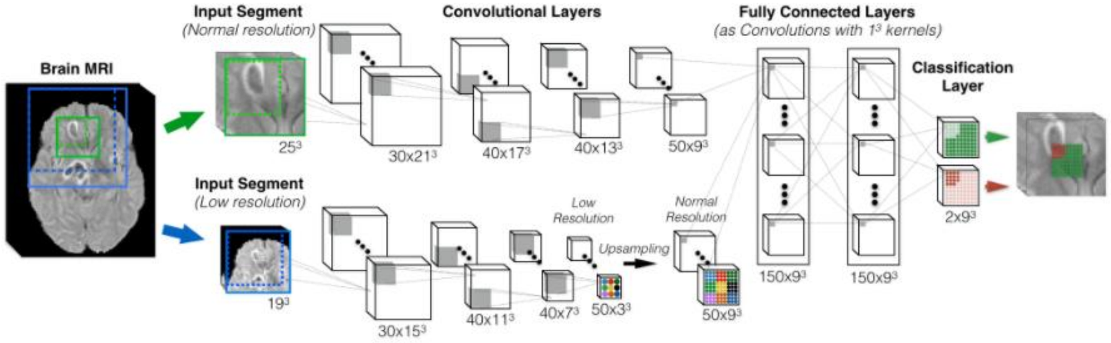
# Performance of semi-supervised learning?

Our Model: SDAE



| Architecture | | | | |
|---|---|---|---|---|
| 1764 | 3500 | 2000 | 1000 | 5 |

State-of-the-art: DeepMedic




$$\text{Dice} = 2 * \frac{A \cap B}{A + B}$$

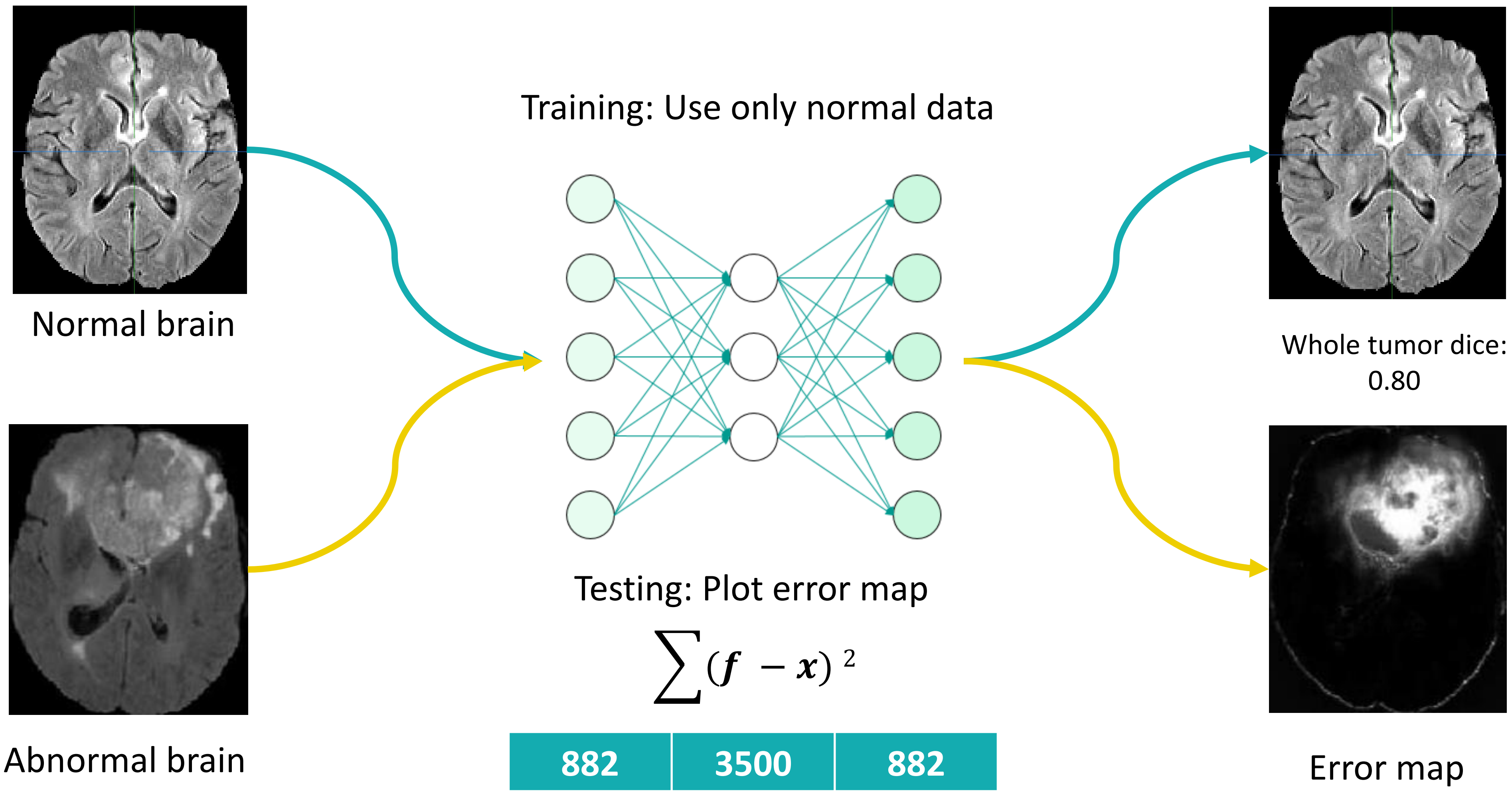
11 layers + 3D Convolutions + 2 Pathways + Fully supervised

| | Scans | Whole Tumor | Tumor Core | Active Tumor |
|---|---|---|---|---|
| DeepMedic | 220 | 0.90 | 0.75 | 0.72 |
| SDAE | 135 | 0.85 | 0.78 | 0.73 |
| SDAE | 20 (Pre-trained on 135) | 0.84 | 0.72 | 0.74 |

DeepMedic - https://arxiv.org/pdf/1603.05959.pdf

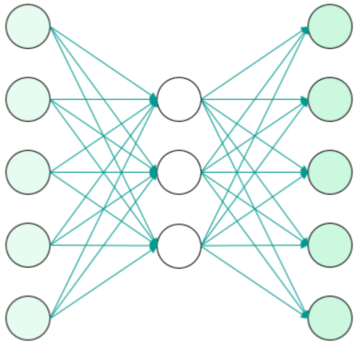
# Unsupervised classification

# Can we perform classification with just **unlabelled** data?

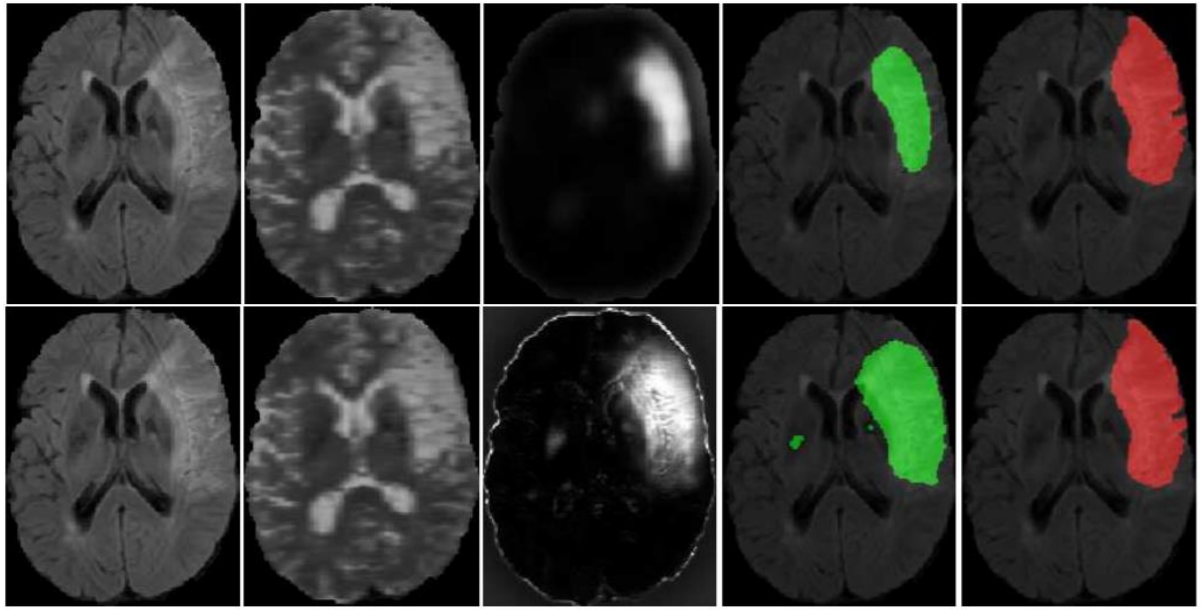


# How good is novelty detection?

Novelty Detector



Trained on BRATS



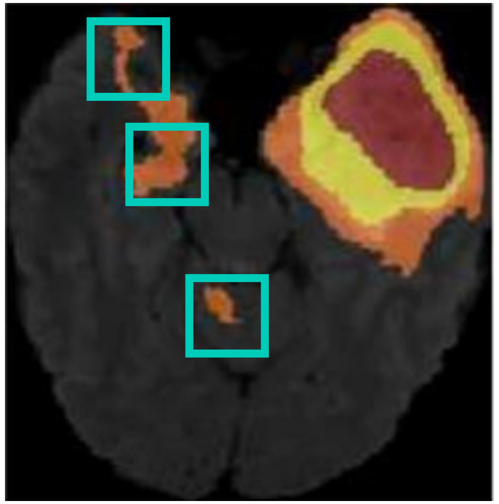
ISLES dataset
Stroke lesion segmentation

| | Scans | Lesion Dice |
|---|---|---|
| **Novelty detector (Unsupervised)** | **28** | **0.64** |
| **DeepMedic (Fully supervised)** | **28** | **0.66** |

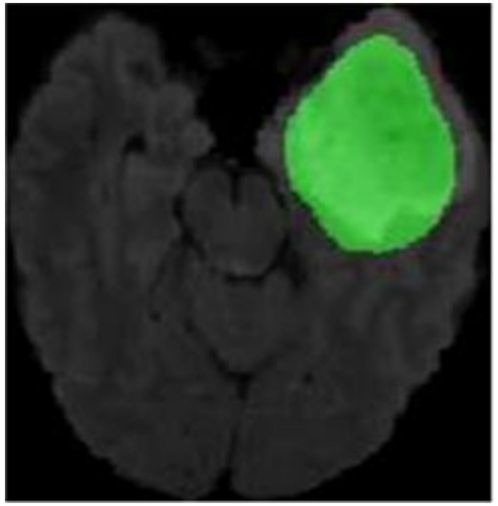
DeepMedic - https://arxiv.org/pdf/1603.05959.pdf

# More false positive reduction

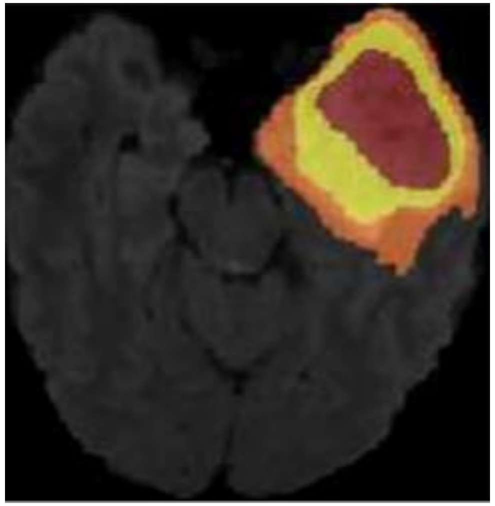
Use novelty detector to reject false positives



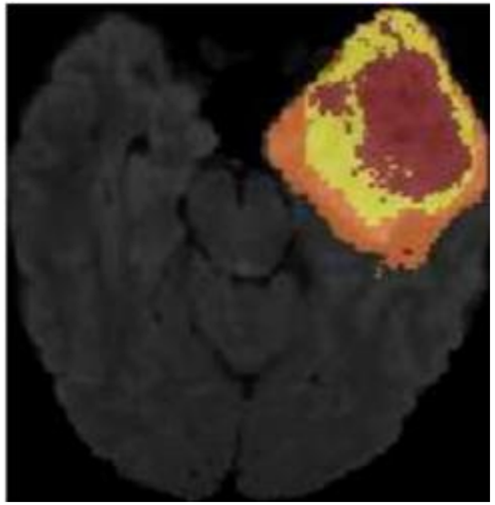
Raw prediction



Novelty detector’s mask

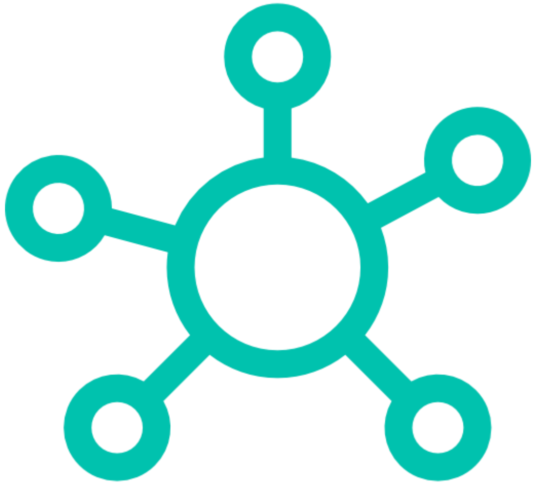


Post-processed



Ground Truth

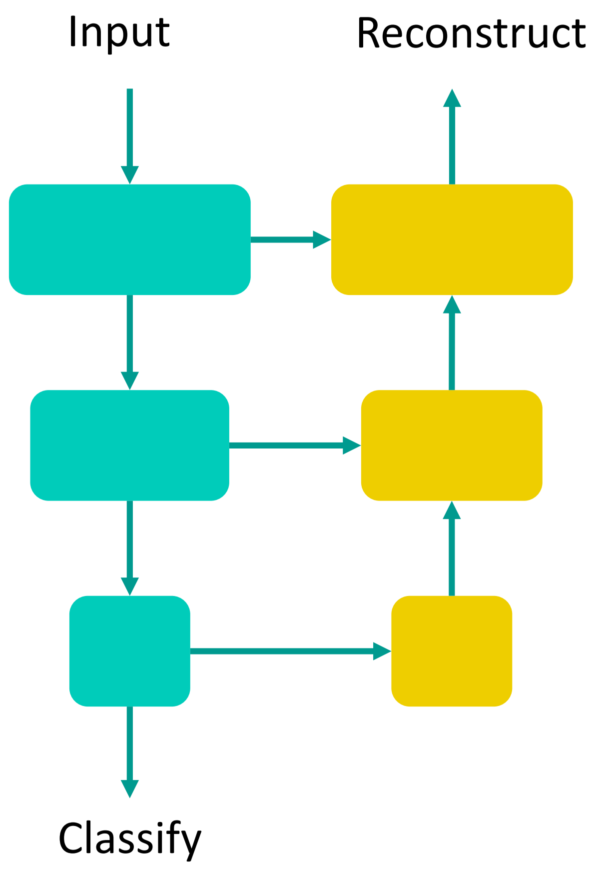
Semi-Supervised learning in Brain Tumor Segmentation - https://arxiv.org/pdf/1611.08664.pdf



# Hybrid architectures

# Can we do joint training on labelled and unlabelled data?

- Joint training
  - Reconstruct on unlabelled data
  - Reconstruct and classify on labelled data
- Add skip connections to fuse features




Ladder networks

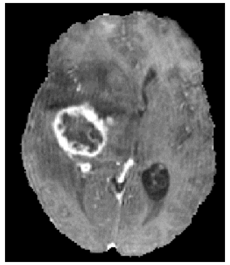
Reconstruction loss

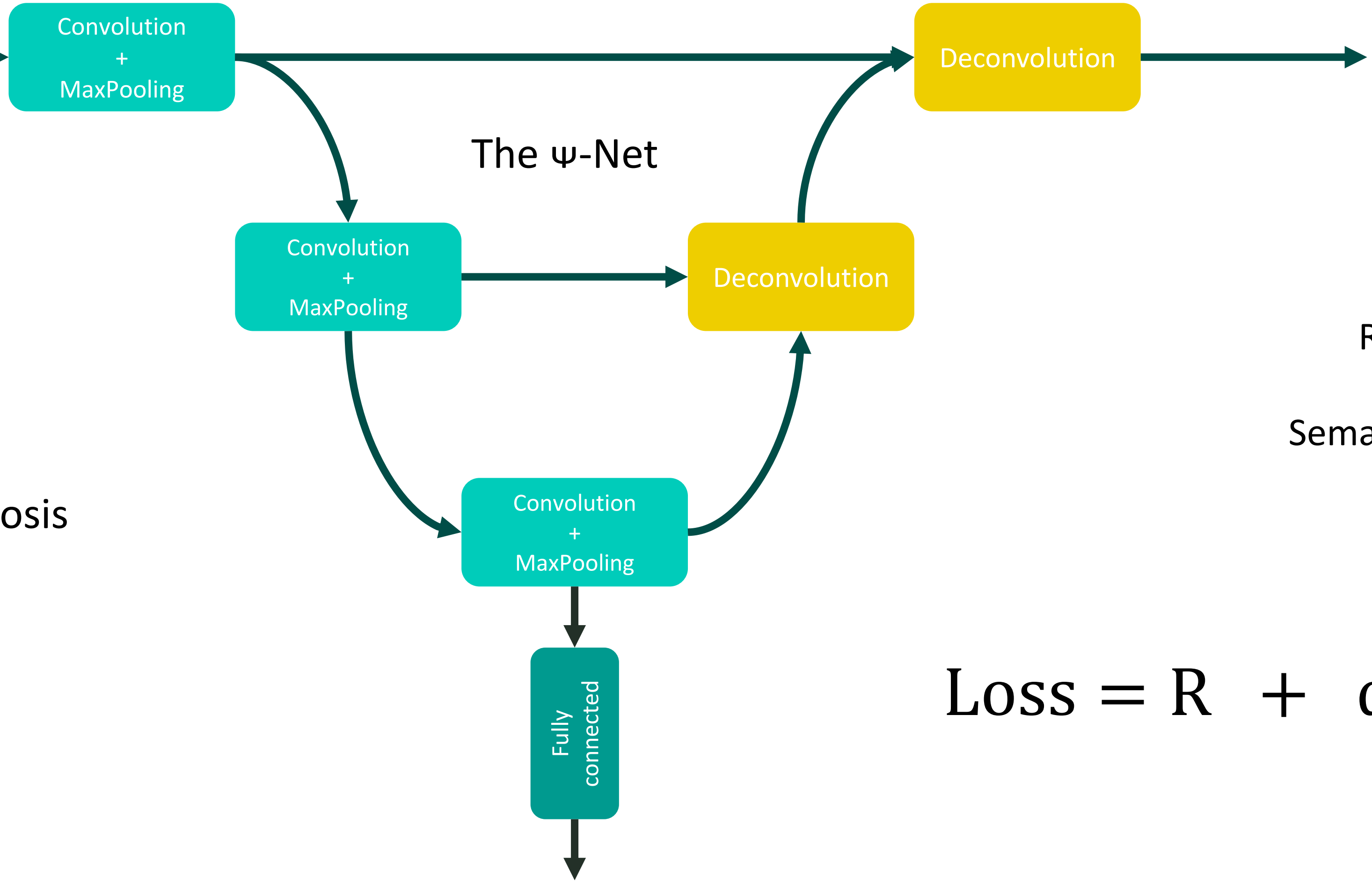
$$\text{Loss} = R + \propto C$$

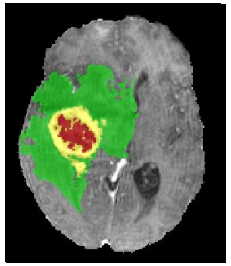
Classification loss

Source - https://arxiv.org/pdf/1507.02672.pdf

# Fully convolutional ladder networks





BRATS 2017 Challenge

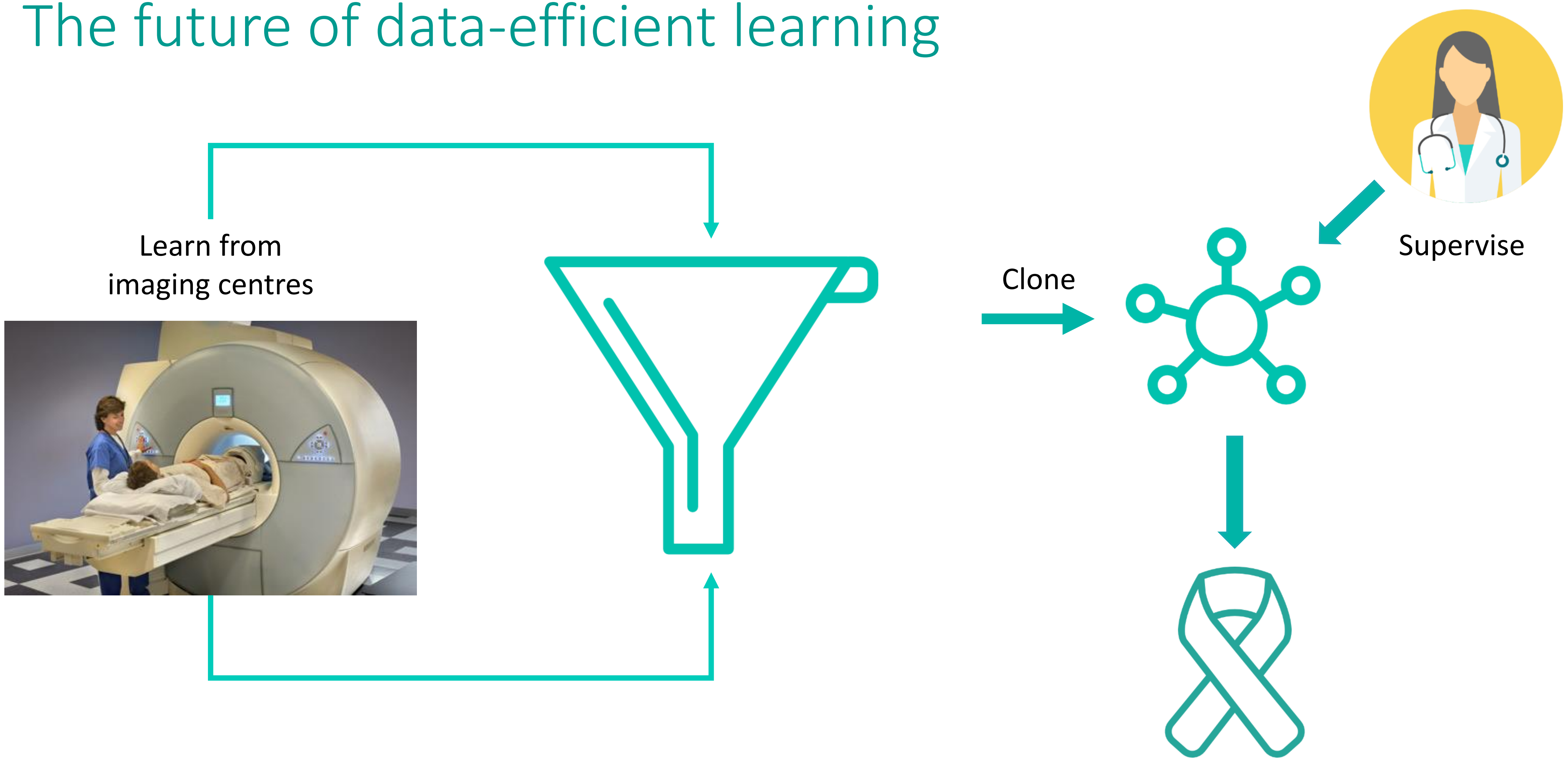
- Segment tumor
- Predict patient prognosis

Classification/Regression

Reconstruction
Or
Semantic Segmentation

$$\text{Loss} = R + \alpha C + \beta S$$

# The future of data-efficient learning



# > sudo kill cancer

Predible

**Kiran Vaidhya**
Algorithms Researcher
kiran@predible.co

**Acknowledgments:**
Varghese Alex – co-author
Subramaniam Thirunavukkarasu - co-author
Dr. Ganapathy Krishnamurthi - Assistant Professor, IIT Madras
Dr. C. Kesavdas - Professor, SCTIMST