Hasso Plattner Institut
Digital Engineering · Universität Potsdam


# Medical Image Analysis by Deep Learning

Mina Rezaei

Data Management for Digital Health
Summer 2017

# Deep Learning
# Theory & Applications

- Foundations
  - Deep Learning
  - Medical Image Analysis
  - Supervised Learning

- Application Examples
  - Medical Image Segmentation
  - Medical Image Registration
  - Computer-aided Diagnosis and disease quantification
  - Medical Image Synthetization

# Traditional Programming  vs Machine learning



## Some Applications

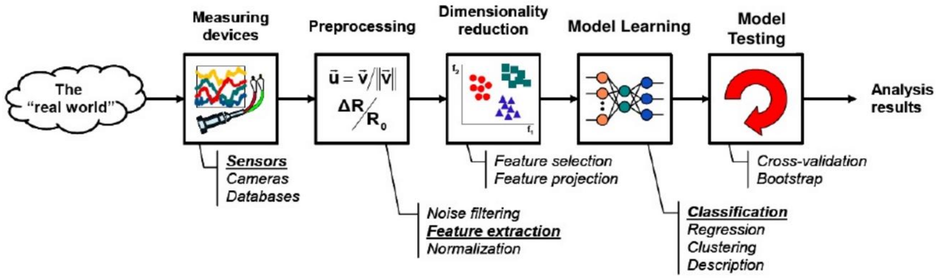
- Web search
- Computational biology
- Finance
- Robotics
- Information extraction
- Social networks
- Debugging
- *[Your favorite area]*

## Representation

- Decision trees
- Sets of rules / Logic programs
- Instances
- Graphical models (Bayes/Markov nets)
- Neural networks
- Support vector machines
- Model ensembles
- Etc.

# Machine Learning

Data → Computer → Program
Output →




"Car" → Computer → Classifier

# Key Idea Behind Neural Network

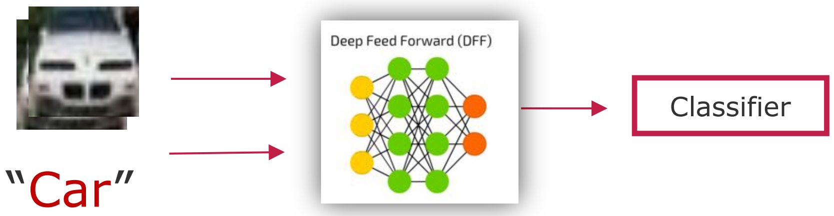
- Learn features from data
- Use differentiable functions that produce feature efficiently
- “Deep” architectures: cascade of simpler non-linear modules
- End-to-end learning: no distinction between feature extractor and classifier

?

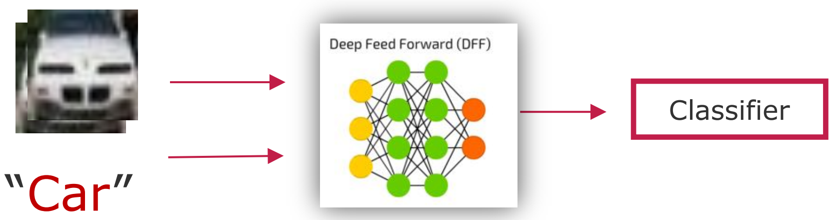
- What is the input-output mapping?
- How are parameters trained?
- How computational expensive is it?
- How well does it work?

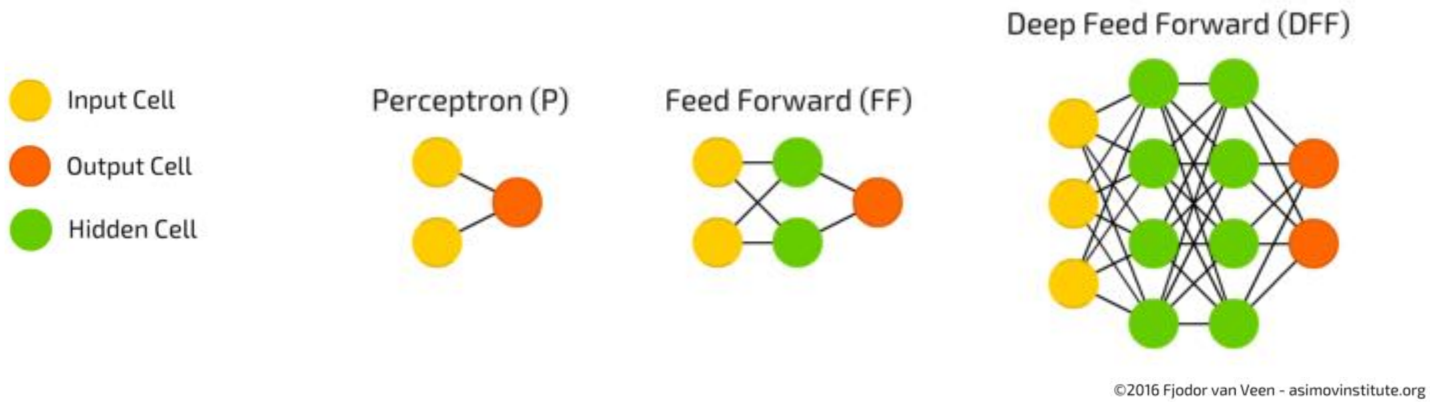
- Neural Networks for Supervised Training
  - Architecture
  - Loss function
- Unsupervised Training of Neural Networks
- Semi-supervised / multi-task / multi-modal
- Reinforcment Learning

# Artificial Neural Network (Deep Learning, Hierarchical learning)



# Artificial Neural Network (Deep Learning, Hierarchical learning)

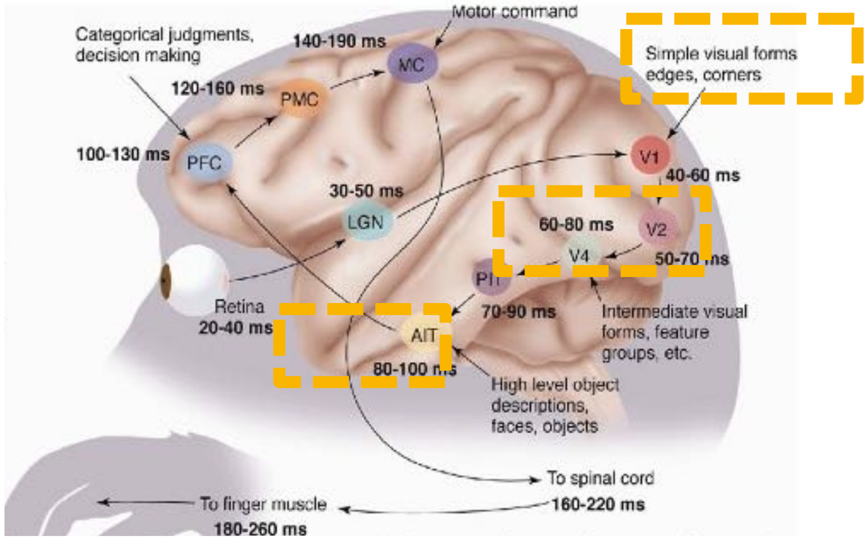


# The Human Visual Cortex Consists of Hierarchies

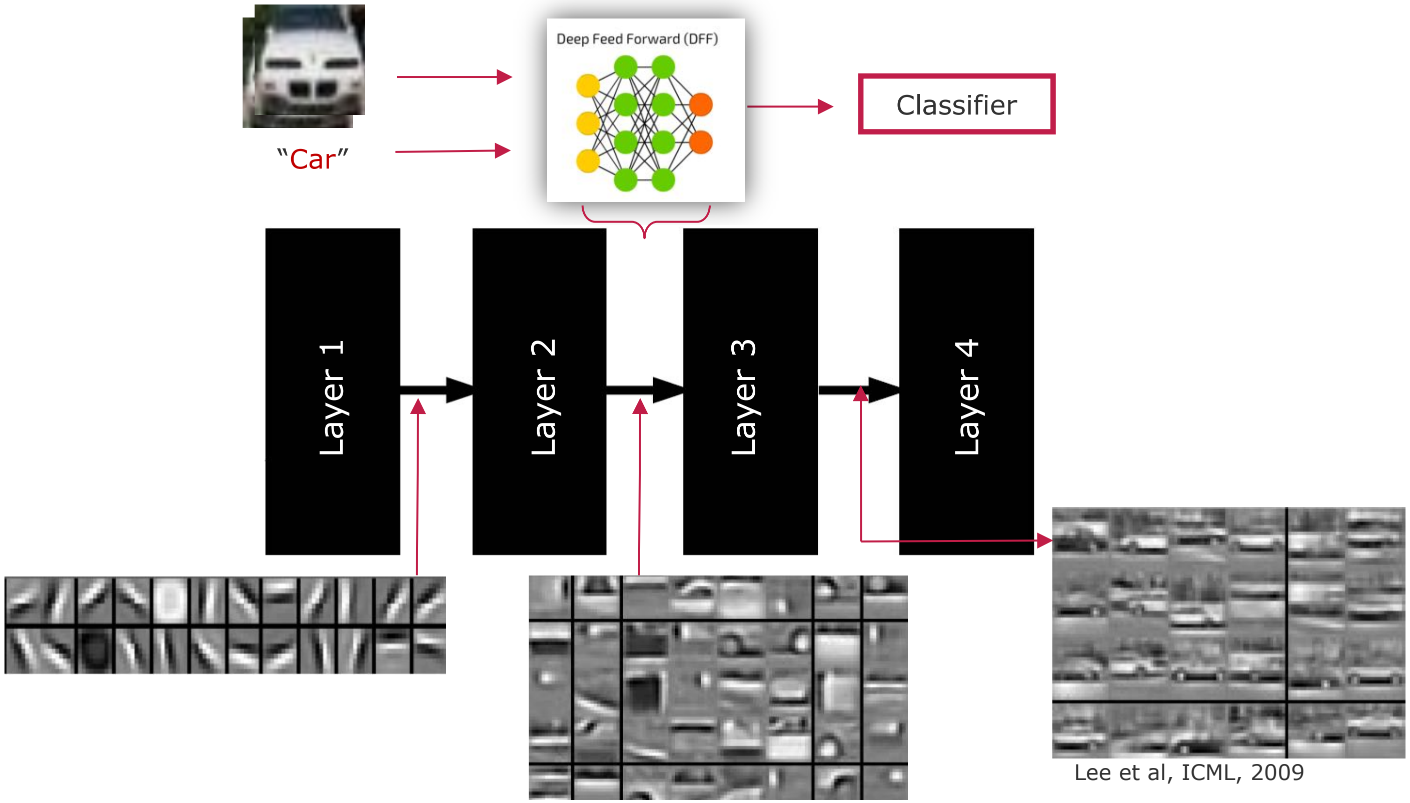
The ventral (recognition) pathway in the visual cortex has multiple stage:

- Retina - LGN - V1 - V2 - V4 - PIT - AIT …., lots of **intermediate representations**

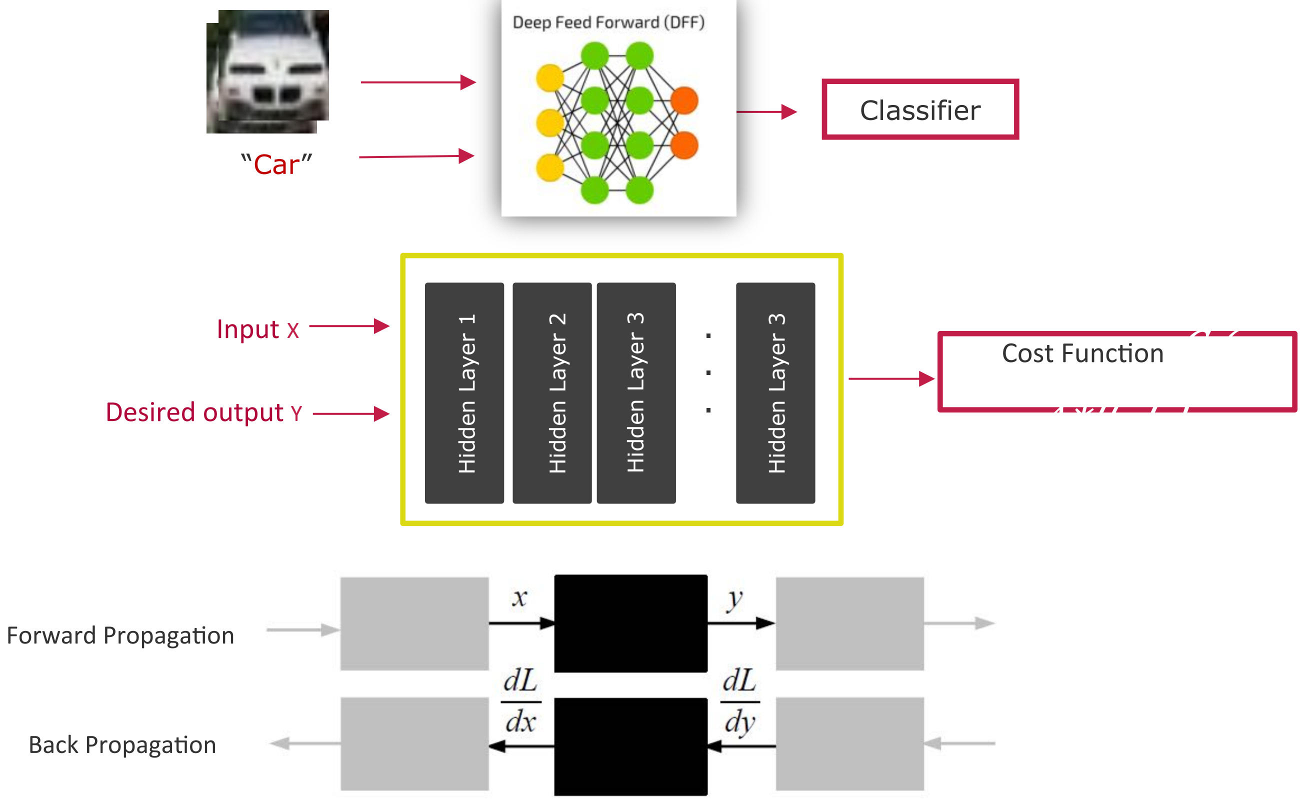



[Gallant & Van Essen]

# Hierarchical Learning



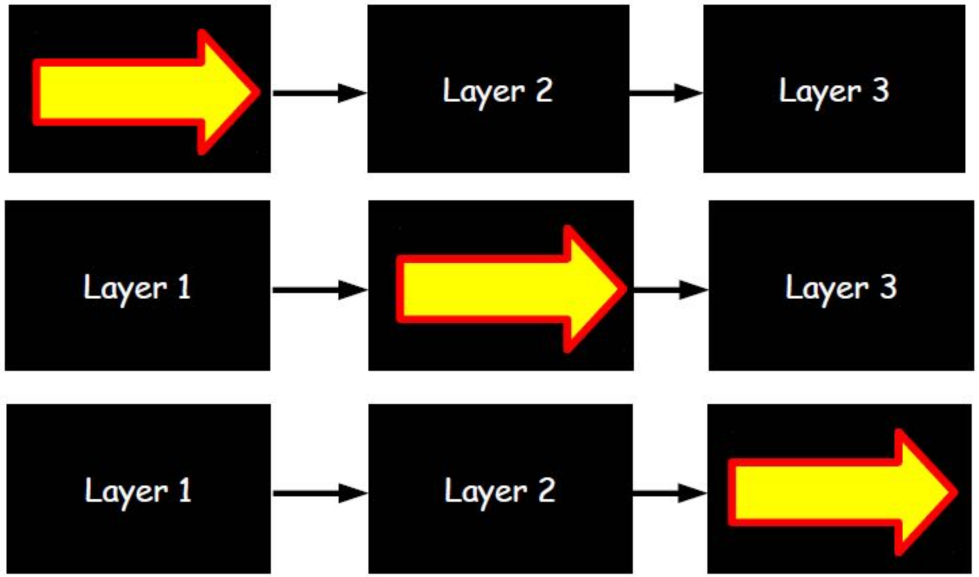
# Neural Network for Supervised Learning



# Neural Net Training (I)

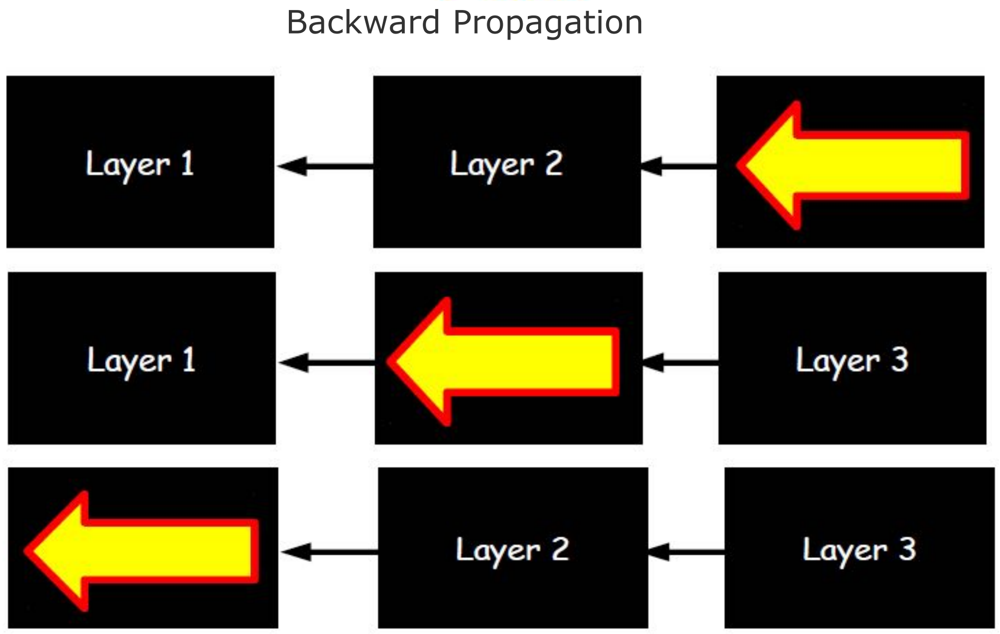
- Compute (Cost function) loss on small set of input

Forward Propagation



# Neural Net Training (II)

- Compute gradient w.r.t. parameters
- Use gradient to update parameters



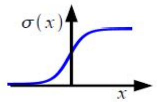
# Layer Internals

# Layer Internals
## Activation Function


a)


$$h_{j+1} = \sigma(W_{j+1}^{T} h_j + b_{j+1})$$

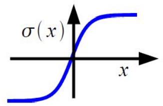
$$\sigma(x) = \frac{1}{1+e^{-x}}$$




b)


$$h_{j+1} = \sigma(W_{j+1}^{T} h_j + b_{j+1})$$

$$\sigma(x) = \tanh(x)$$



# Layer Internals
## Activation Function

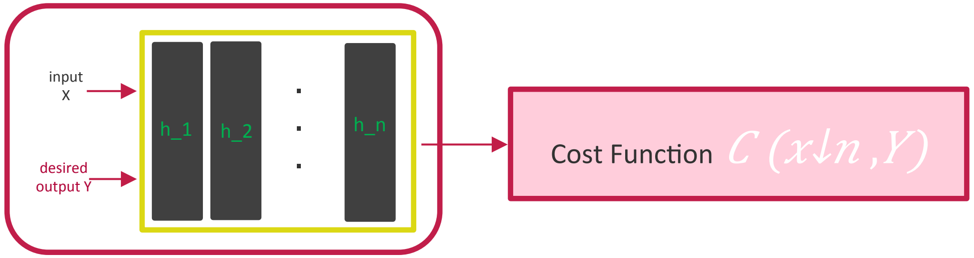


$$h_{j+1} = \sigma(W^T_{j+1} h_j + b_{j+1})$$

$$W_j \in R^{M \times N}, \; b_j \in R^N$$

$$h_j \in R^M, \; h_{j+1} \in R^N$$

Hasso Plattner Institut



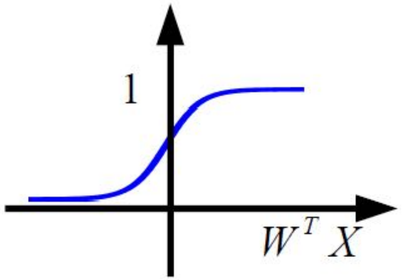

Input: $X \in R^D$

Binary label: $y$

Parameters: $W \in R^D$

Output prediction: $p(y=1|X) = \frac{1}{1+e^{-W^T X}}$

Loss: $L = -\log(p(y|X))$



# Support Vector Machine

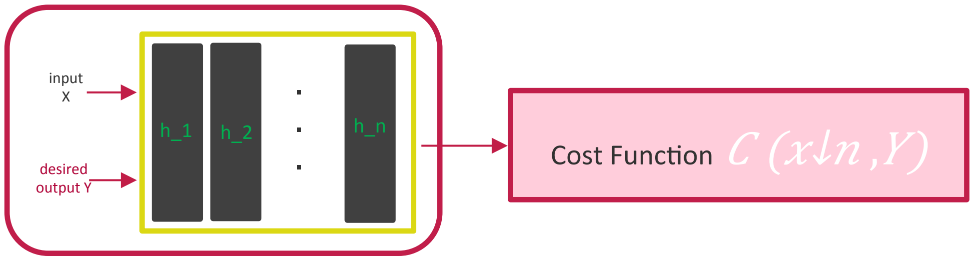
Input: $X \in R^D$

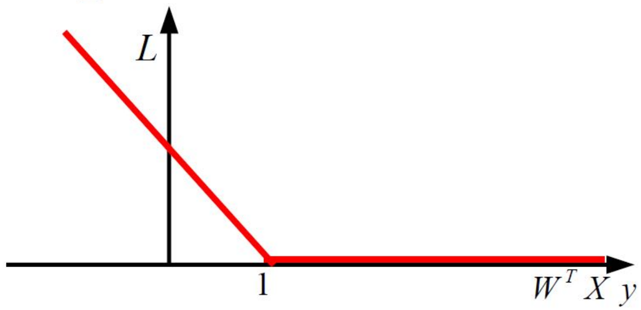
Binary label: $y$

Parameters: $W \in R^D$

Output prediction: $W^T X$

Loss: $L = \frac{1}{2} ||W||^2 + \lambda \, max[0, 1 - W^T X \, y]$



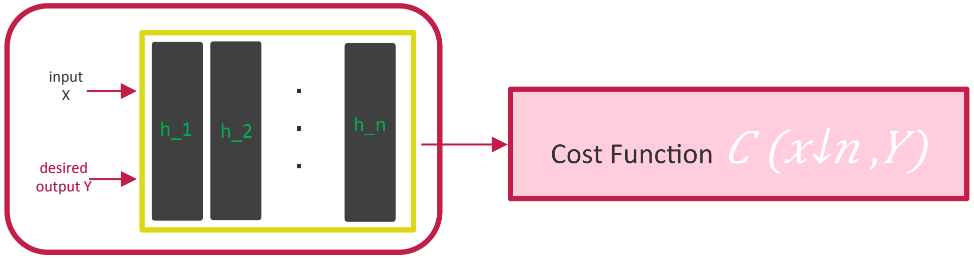




**Hinge Loss**

Ranzato

# Logistic Regression





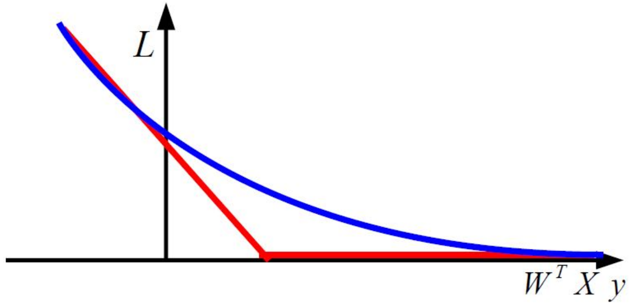

Input: $X \in R^D$

Binary label: $y$

Parameters: $W \in R^D$

Output prediction: $W^T X$

Loss: $L = \frac{1}{2} \|W\|^2 + \lambda \log(1 + \exp(-W^T X y))$




**Log Loss**

Ranzato

# Example

## Input :

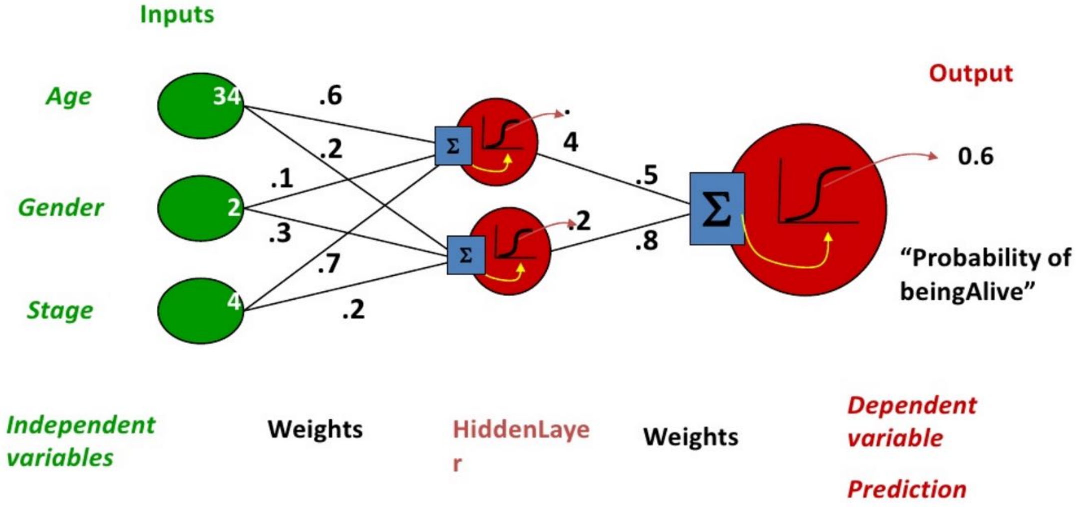
Age:34

Gender: female

Stage: IV

## Output :

Probability of being alive?



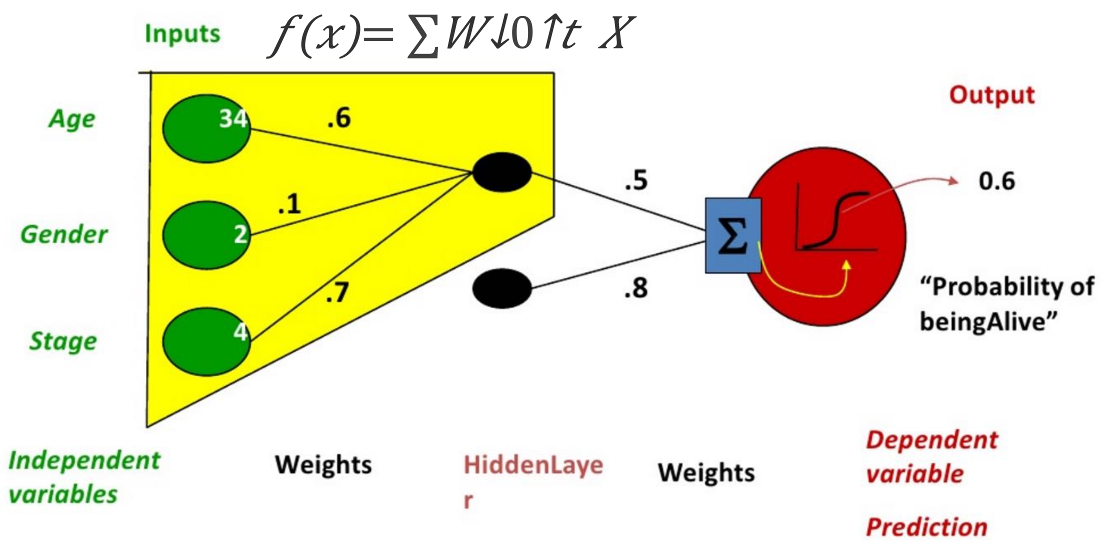

Machine Learning, Dr. Lior Rokach, Ben-Gurion University

# Example

=0.6

=0.1

=0.7

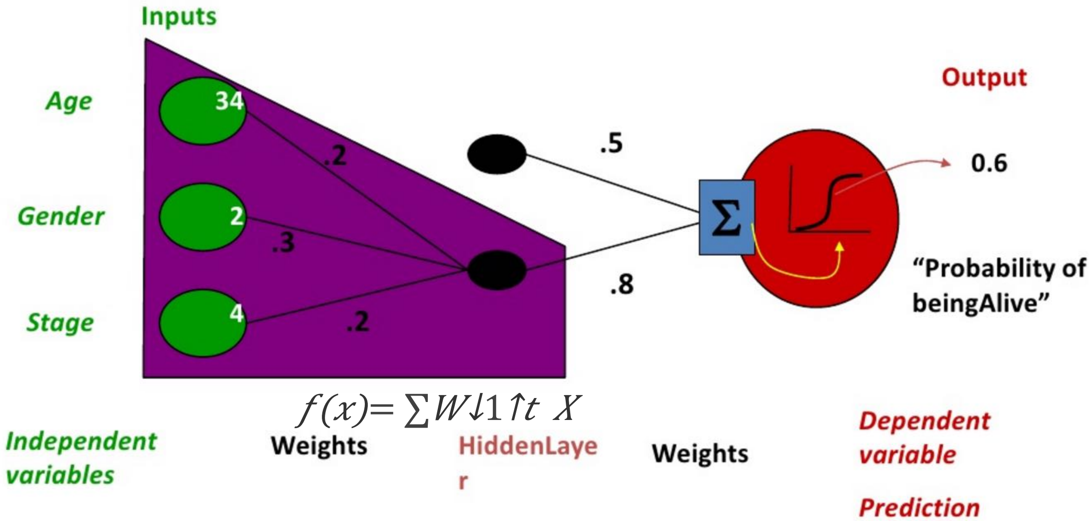


# Example

$W_1^1 = 0.2$

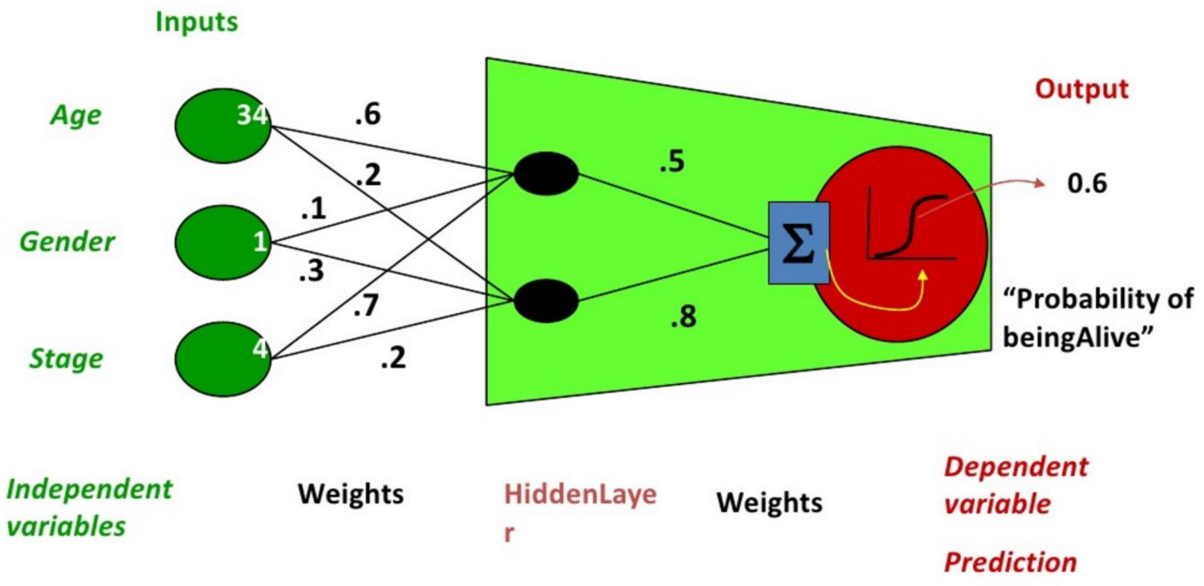
$W_1^2 = 0.3$

$W_1^3 = 0.2$

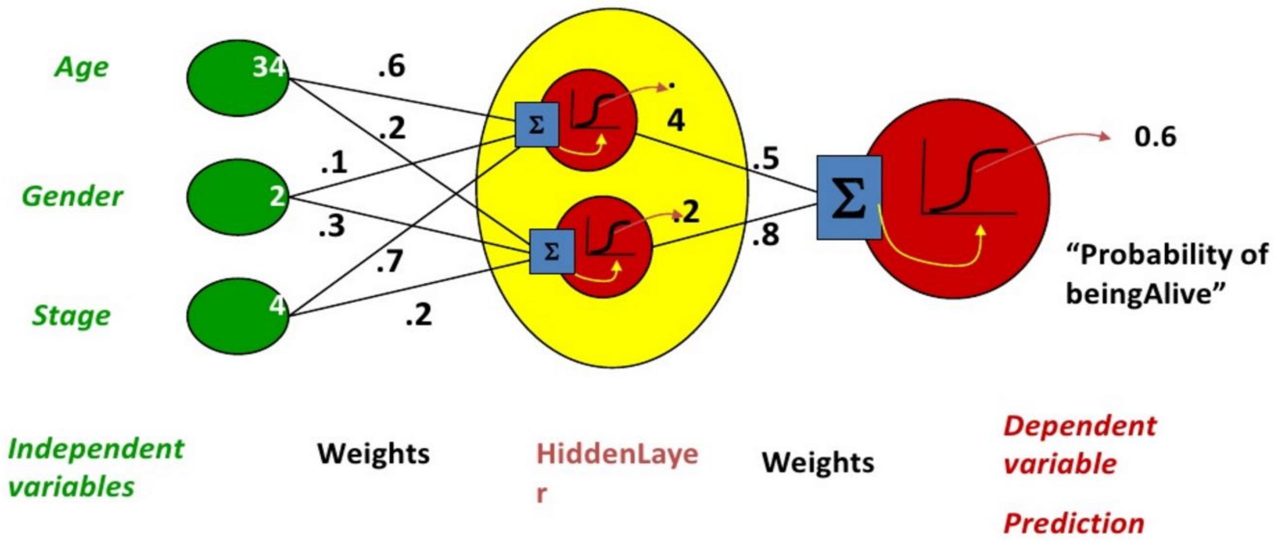



$f(x) = \sum W_1^t X$

# Example


Inputs
Age
34
Gender
1
Stage
4
.6
.2
.1
.3
.7
.2
.5
.8
Σ
Output
0.6
"Probability of beingAlive"
Independent variables
Weights
HiddenLayer
Weights
Dependent variable
Prediction

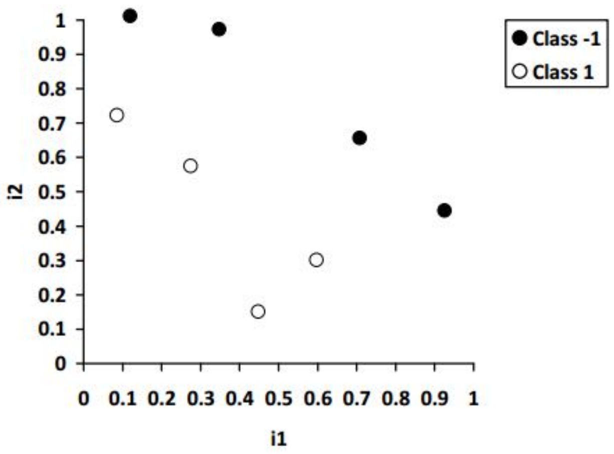
# Example




**Medical Image Analysis by Deep Learning**

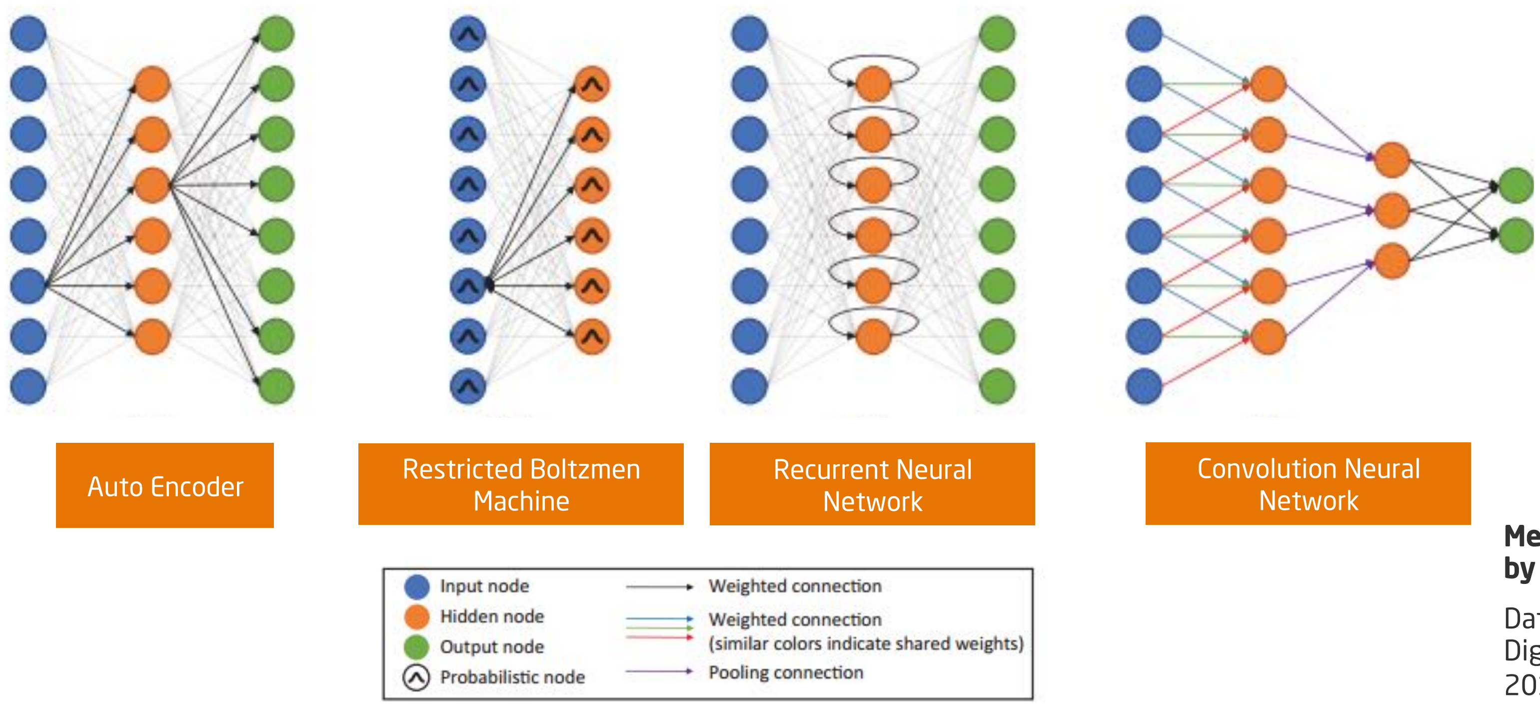
# Example

The chart below shows a set of two-dimensional input samples from two classes:




1. How many samples are misclassified for both classes?
2. Calculate the new weights from each class?
3. How many samples are missclassified with new updated weights?

# Popular Neural Network Architecture



# Deep Learning Frameworks

Hasso Plattner Institut

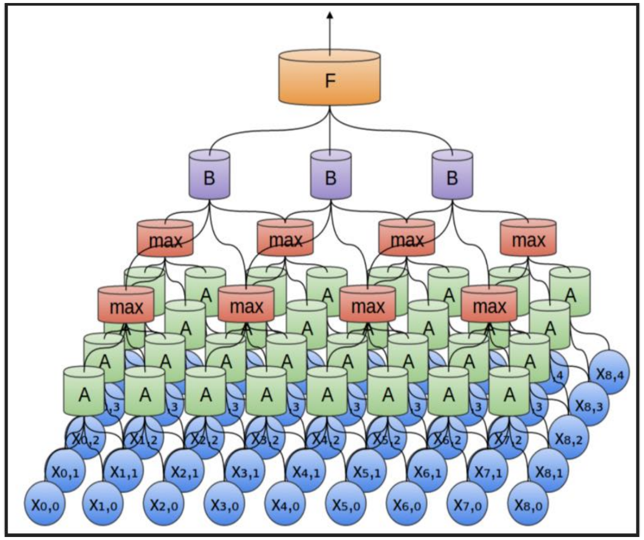
# Classification Model

```
from caffe import layers as L
from caffe import params as P

def lenet(lmdb, batch_size):
    # our version of LeNet: a series of linear and simple nonlinear transformations
    n = caffe.NetSpec()
    n.data, n.label = L.Data(batch_size=batch_size, backend=P.Data.LMDB, source=lmdb,
                             transform_param=dict(scale=1./255), ntop=2)
    n.conv1 = L.Convolution(n.data, kernel_size=5, num_output=20, weight_filler=dict(typ
e='xavier'))
    n.pool1 = L.Pooling(n.conv1, kernel_size=2, stride=2, pool=P.Pooling.MAX)
    n.conv2 = L.Convolution(n.pool1, kernel_size=5, num_output=50, weight_filler=dict(typ
e='xavier'))
    n.pool2 = L.Pooling(n.conv2, kernel_size=2, stride=2, pool=P.Pooling.MAX)
    n.ip1 = L.InnerProduct(n.pool2, num_output=500, weight_filler=dict(type='xavier'))
    n.relu1 = L.ReLU(n.ip1, in_place=True)
    n.ip2 = L.InnerProduct(n.ip1, num_output=10, weight_filler=dict(type='xavier'))
    n.loss = L.SoftmaxWithLoss(n.ip2, n.label)
    return n.to_proto()

with open('examples/mnist/lenet_auto_train.prototxt', 'w') as f:
    f.write(str(lenet('examples/mnist/mnist_train_lmdb', 64)))

with open('examples/mnist/lenet_auto_test.prototxt', 'w') as f:
    f.write(str(lenet('examples/mnist/mnist_test_lmdb', 100)))
```




Picture source: Christopher Olah's Blog

# Why Is Deep Learning So Popular?

New layer-wise training algorithm [Science 2006], i.e. train on atomic task

Big data, compared to 20 years ago

Powerful computers

- Previous algorithms may be theoretically working, but not practically
- not converged to good local minima with the previous less powerful computers

# Why Is Deep Learning So Popular?

Hasso Plattner Institut

- New layer-wise training algorithm [Science 2006], i.e. train on atomic task
- Big data, compared to 20 years ago
- Powerful computers
  - Algorithms were working but impractical
  - Algorithms did not converge to good local minima

# Deep Learning; Advantages and Disadvantages

Hasso Plattner Institut

## Advantages

- High performance on multiple domains problems, e.g. speech, language, vision, games, etc.
- Reduces the need for feature engineering
- Can be adapted to new problems with new architecture, e.g. vision, time series, language, etc. using convolutional neural networks, recurrent neural networks, long short-term memory, etc.

## Disadvantages

- Requires a large amount of data .
- Is extremely computationally expensive to train.
- No strong theoretical foundation
- Determining the topology, training method, hyper parameters for deep learning is considered as black magic

# Deep Learning Application on Medical Imaging



## Segmentation

- Hippocampus Segmentation

## Registration

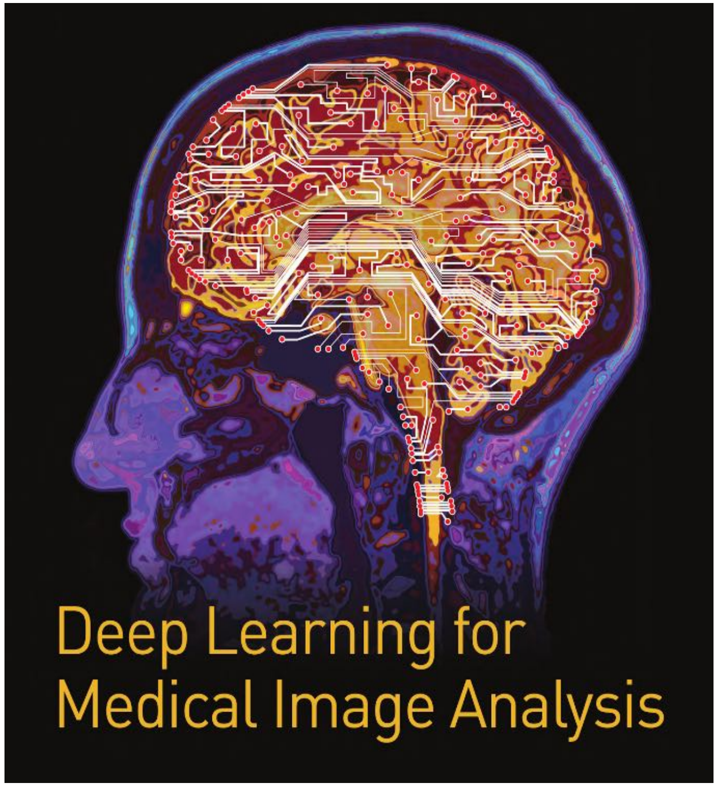
- Brain MRI Registration

## Computer-aided Diagnosis

- AD/MCI Diagnosis
- Skin Cancer Detection

## Image Synthesis

- Estimating CT from MRI

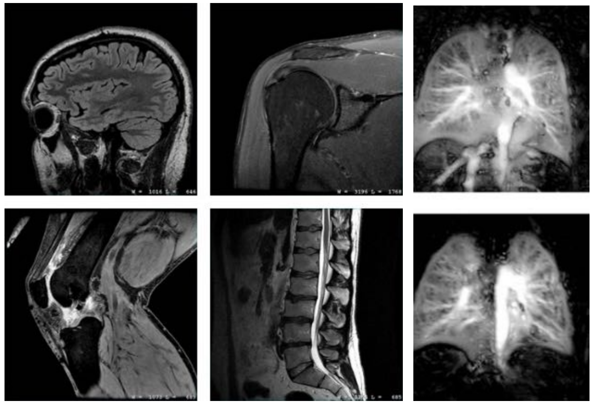


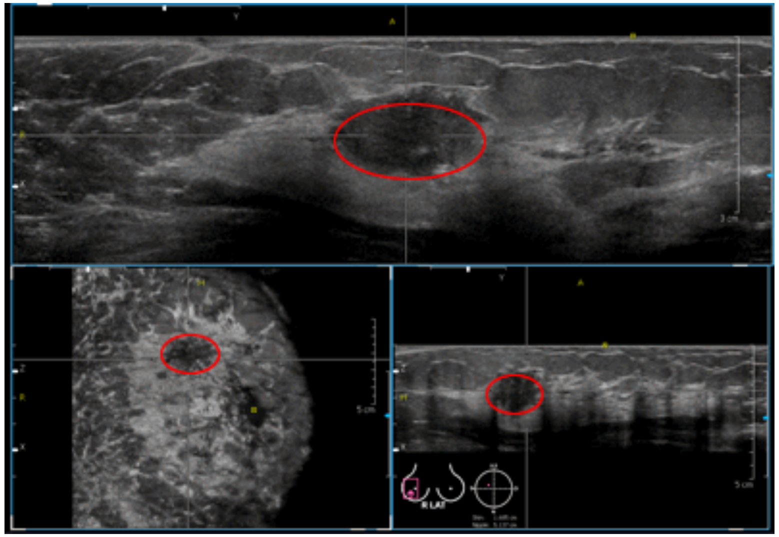
# Medical Imaging Examples

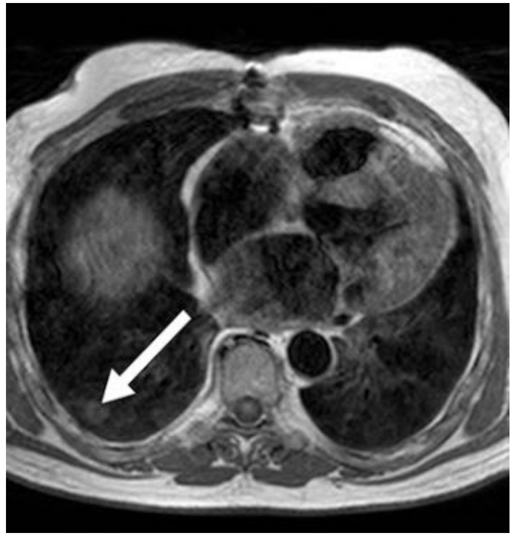
- CT scan produces up to 2000 images within 25s
- PET/CT requires review of up to 6000 images
- Breast ultrasound can create 5000 images*


*http://www.cancer.org

# Medical Imaging Examples

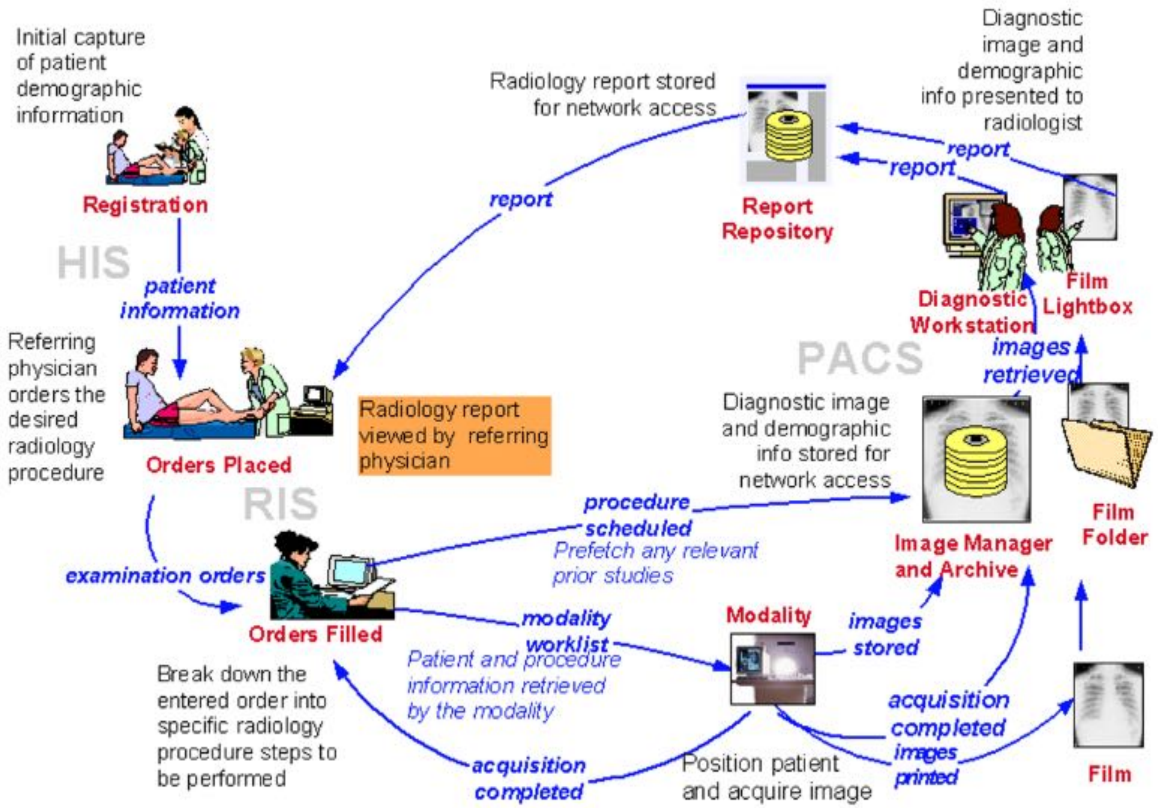






*http://www.cancer.org

# Current Radiologist's Workflow




Hasso Plattner Institut

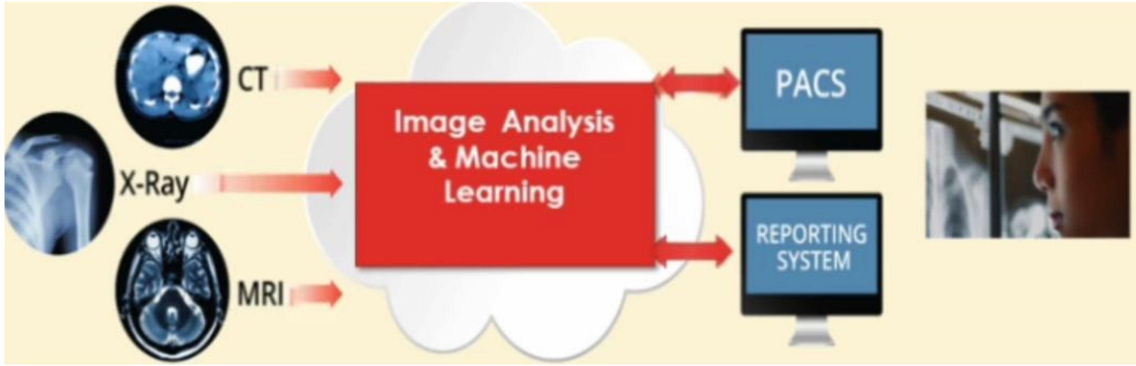
**Medical Image Analysis by Deep Learning**

Data Management for Digital Health, Summer 2017

# Image Analysis & Machine Learning

- Challenges in Image Analysis:
  - Time consuming
  - Missed findings

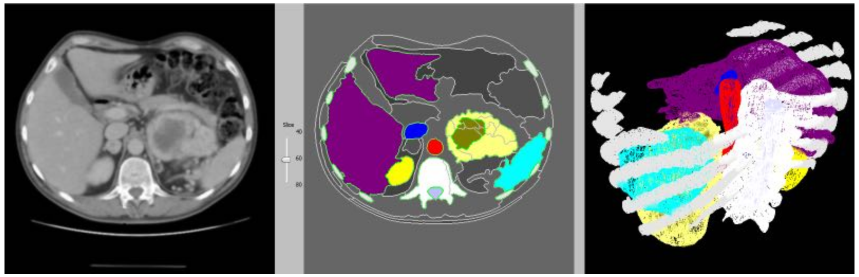
- Machine learning may assist radiologist in:
  - Formulating findings
  - Taking measurements
  - Characterization
  - Work more efficiently



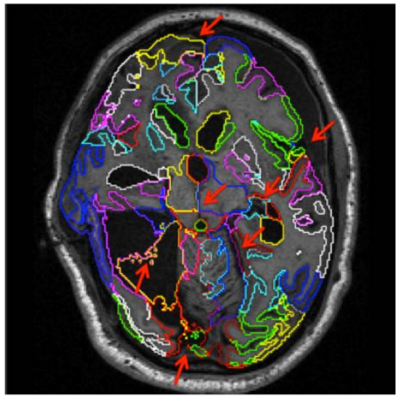
# Application Example:
# Medical Image Segmentation

Segmentation is the process of partitioning an image into different meaningful segments.

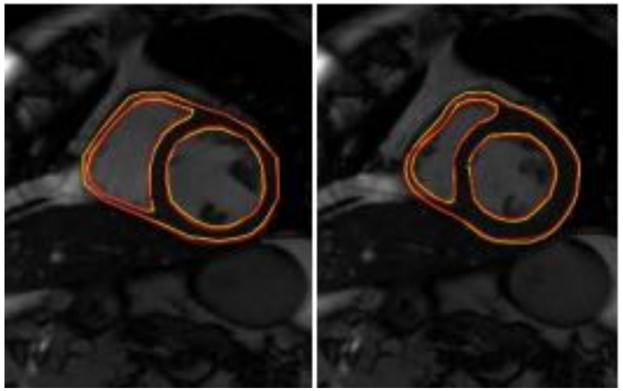
In medical imaging, these segments often correspond to different tissue classes, organs, pathologies, or other biologically relevant structures.



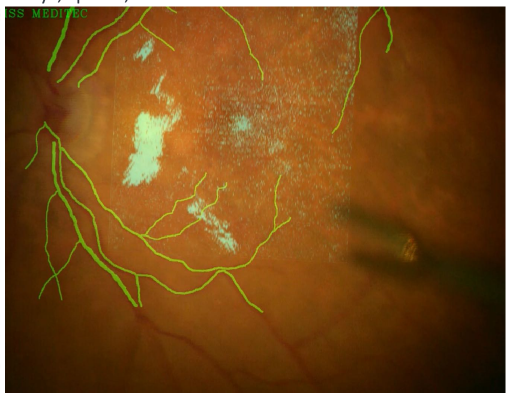
abdominal CT scan.
the rib cage, liver, kidneys, spleen, blood vessels and a renal tumor



whole-brain segmentation
C. Ledig 2015



endocardium and epicardium at end of diastole and systole. M.Rezaei 2013

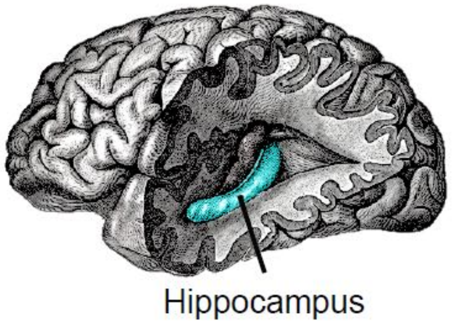


Retina segmentation
M.Rezaei 2014

# Hippocampus Segmentation

Importance

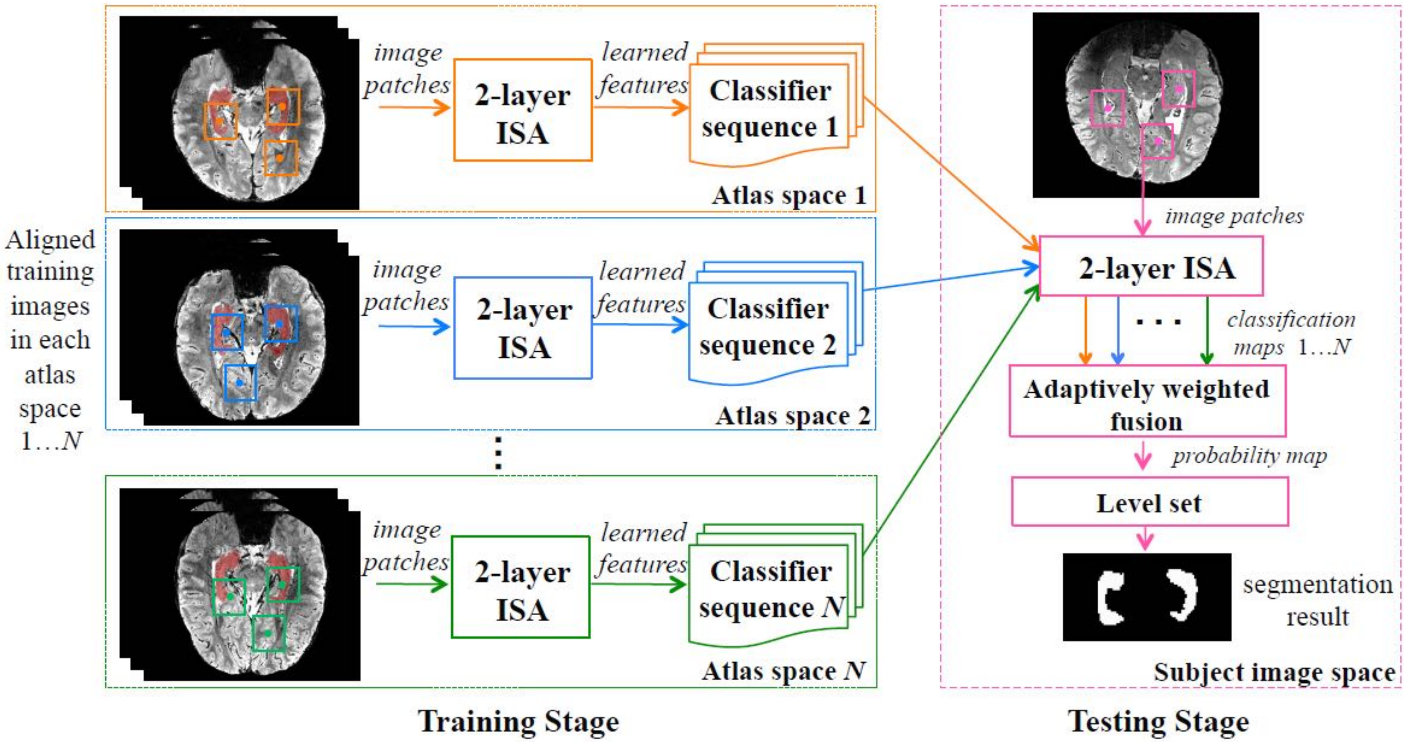
- The volume of the hippocampus is an important trait for early diagnosis of neurological diseases (e.g., Alzheimer's disease)




Challenges

- The hippocampus is small (≈35×15×7mm)
- The hippocampus is surrounded by complex structures
- Low imaging resolution (≈1×1×1mm) of 1.5T or 3T MRI scanners

# Deep Learning Solution for Hippocampus Segmentation




ISA: Independent Subspace Analysis

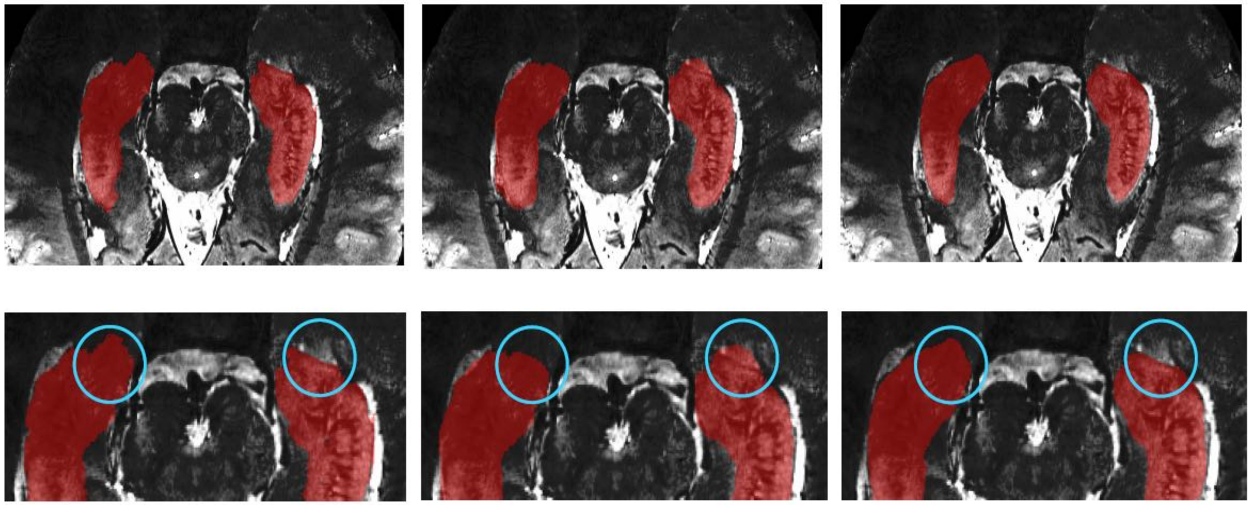
UNC, Dinggang Shen-2017

# Evaluation

| | P | R | RO | SI |
|---|---|---|---|---|
| Hand-Crafted Haar + Texture Features | 0.843 | 0.847 | 0.772 | 0.865 |
| **Hierarchical Patch Representations** | **0.883** | **0.881** | **0.819** | **0.894** |

$V(A)$ : The volume of the ground-truth (manual segmentation)

$V(B)$ : The volume of the automatic segmentation



Ground Truth | Haar + Texture Features | Hierarchical Features

Precision $P = \frac{V(A \cap B)}{V(B)}$

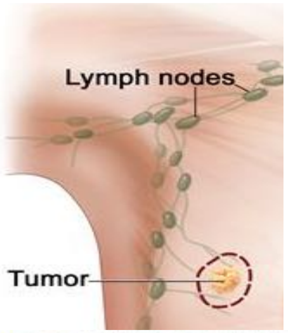
Recall $R = \frac{V(A \cap B)}{V(A)}$

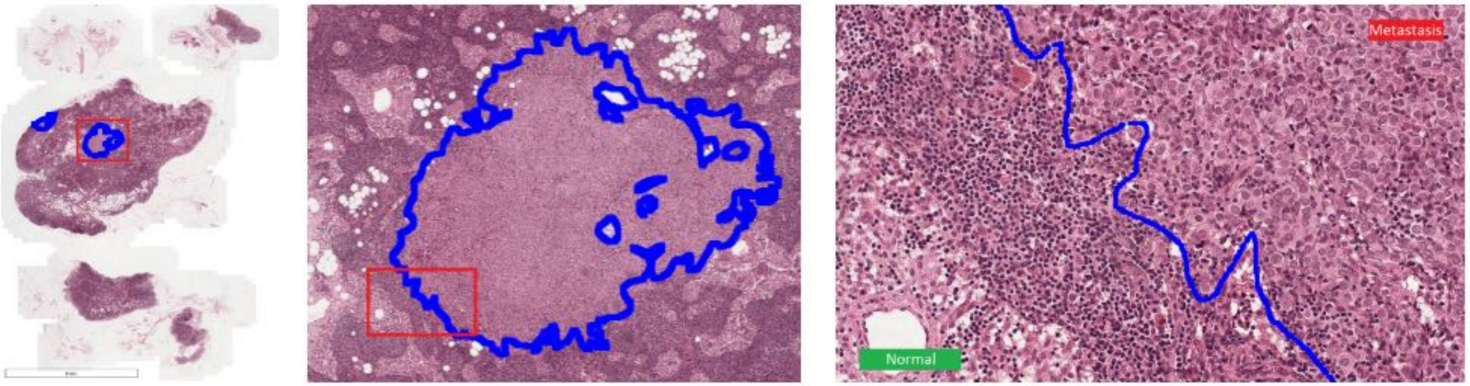
Relative overlap $RO = \frac{V(A \cap B)}{V(A \cup B)}$

Similarity index $SI = \frac{V(A \cap B)}{\{(V(A) + V(B))/2\}}$

# Open Challenge:
# Breast Cancer Metastase Segmentation

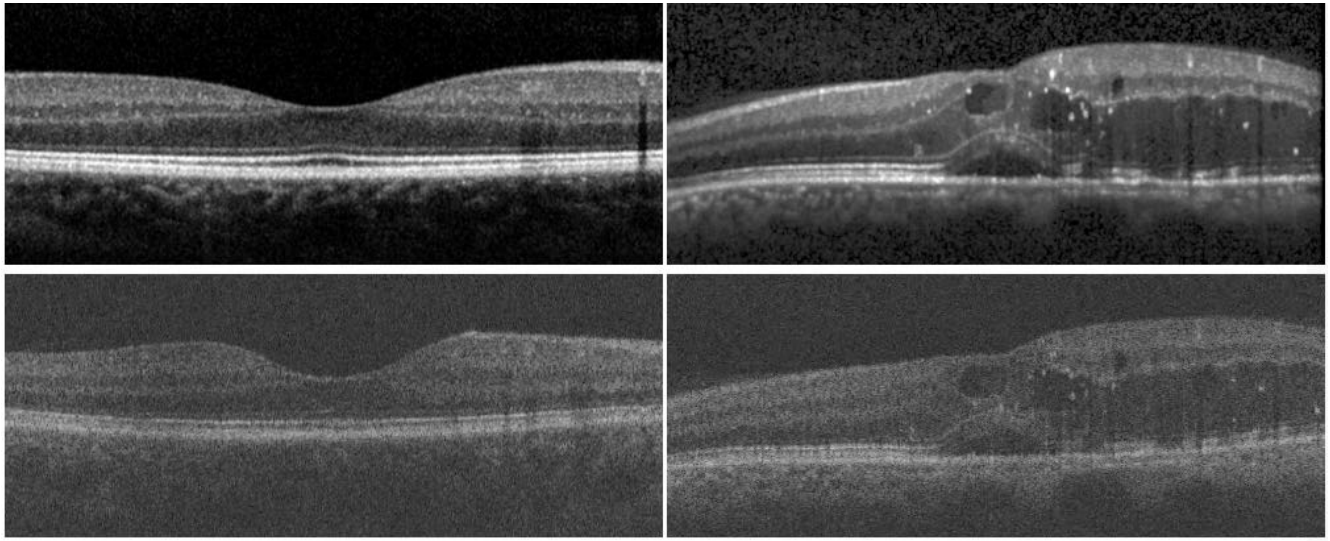
- CAMELYON17 is the second grand challenge in pathology



# Open Challenges:
## Retinal OCT Fluid

RETOUCH is Retinal OCT Fluid Challenge, to segment  variety of retinal fluid lesions 2013-2017



Normal retina imaged with OCT from the three manufacturers.
The three slices come from three different subjects.

Retina with macular edema imaged with OCT from the three manufacturers.
The slices are of the same patient and approximately at the same anatomical position.

# Open Challenges:
# Brain Tumor Segmentation

Hasso Plattner Institut

- BraTS : BraTs Challegnes is Brain Tumor Segmentation Challenges(2012-2017)

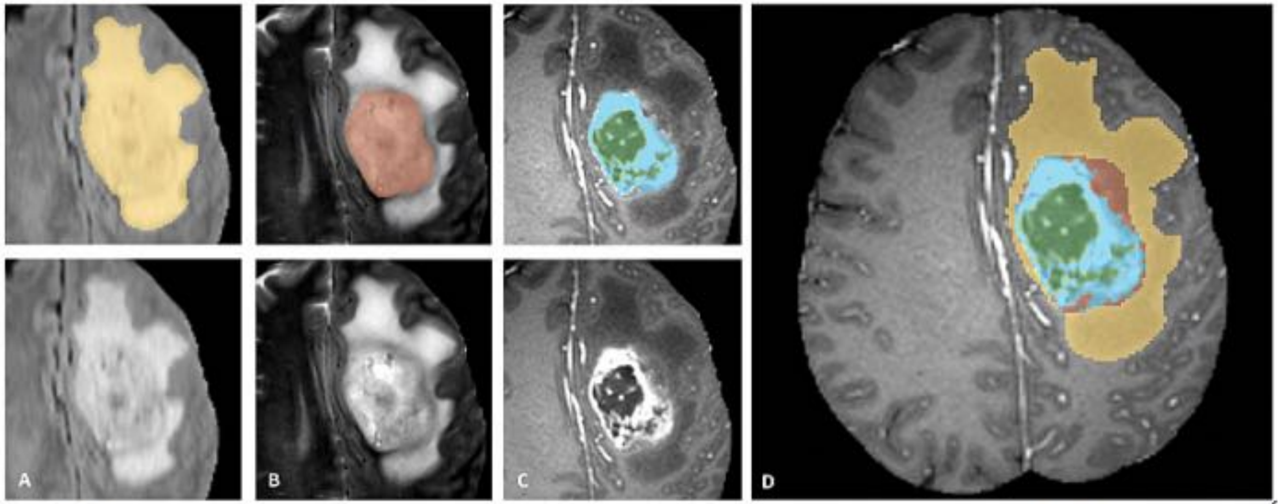
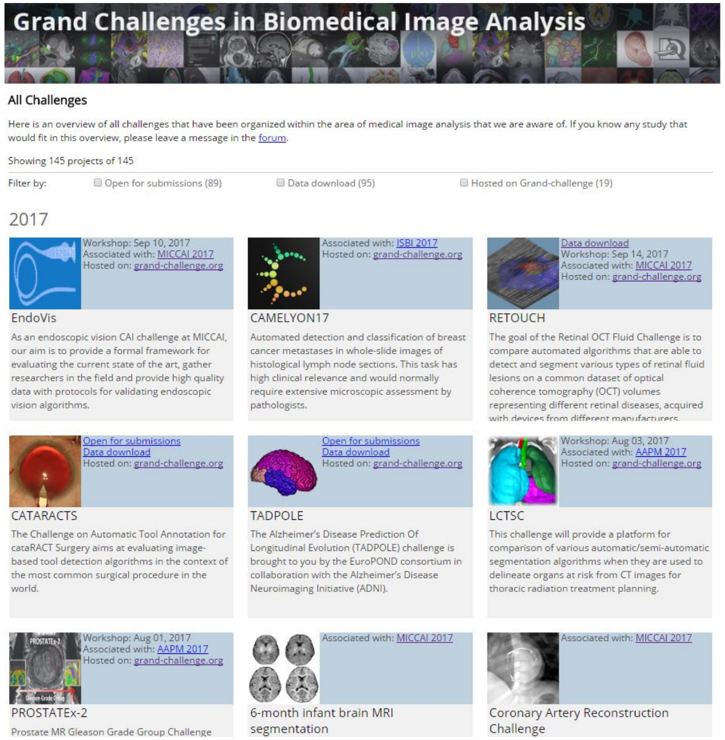



**Fig.1: Glioma sub-regions.** Shown are image patches with the tumor sub-regions that are annotated in the different modalities (top left) and the final labels for the whole dataset (right). The image patches show from left to right: the whole tumor (yellow) visible in T2-FLAIR (Fig.A), the tumor core (red) visible in T2 (Fig.B), the enhancing tumor structures (light blue) visible in T1Gd, surrounding the cystic/necrotic components of the core (green) (Fig. C). The segmentations are combined to generate the final labels of the tumor sub-regions (Fig.D): edema (yellow), non-enhancing solid core (red), necrotic/cystic core (green), enhancing core (blue). (Figure taken from the BraTS IEEE TMI paper.)

# Other Challenges

- https://grand-challenge.org/all_challenges/
- Medical Image Computing and Computer Assisted Interventions Conference-MICCAI
- IEEE Symposium on Bio Medical Imaging (ISBI)
- Artificial Intelligence in Medicine (AIME)



# Application Example:
# Medical Image Registration

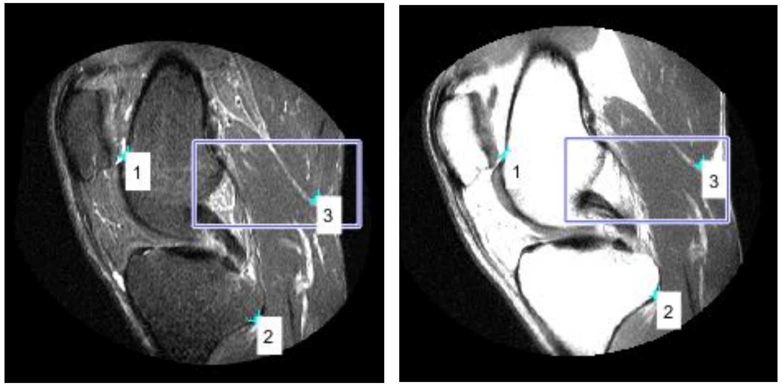
- Studying temporal changes. (time series study e.g. cognitive processes, heart deformations)
- Combining complementary information from different imaging modalities
- Characterizing a population of subjects.

HOW?

  - Transformation model (rigid, affine, deformable)
  - similarity metric (sum of squared distance, Correlation coefficient)

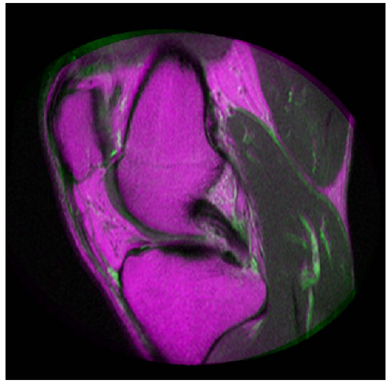

Data augmentation by registration M.Rezaei 2017

# Affine registration on Knee (MRI-Flair & T1c) –(II)




Unregistered Images

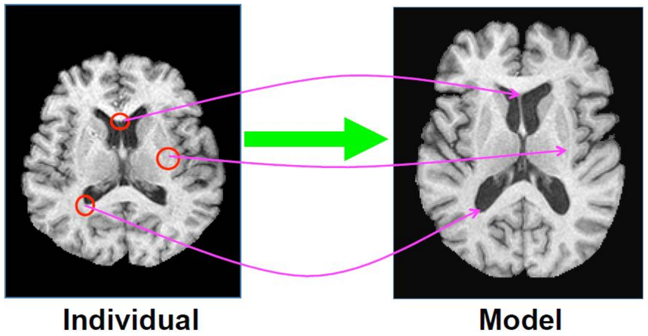
registered Images

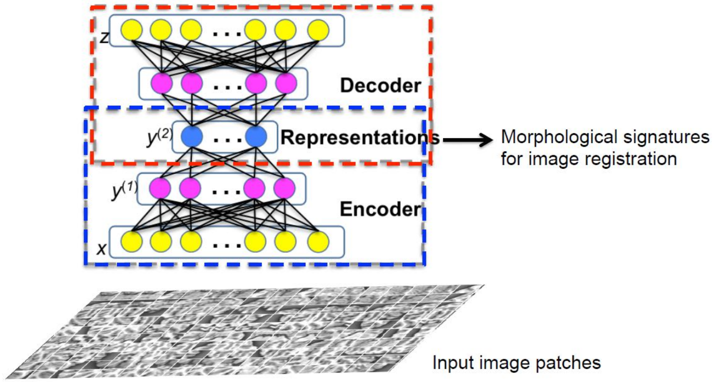


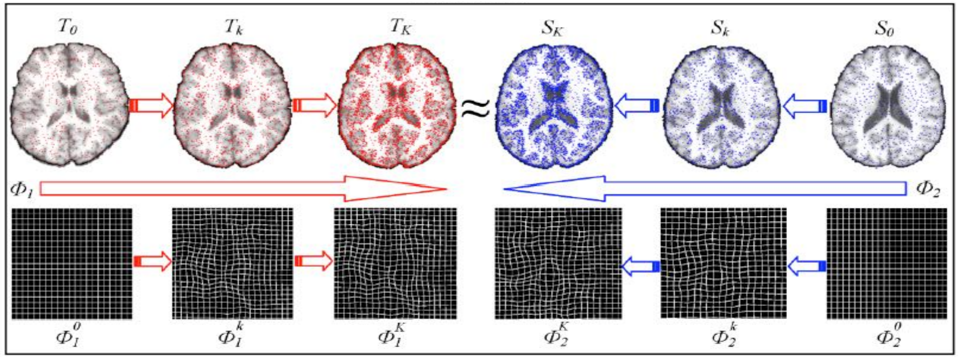
Reflection, Scale, Rotate, Shear, Identity

Mathwork-2014

# Alzheimer Diseases Detection by Auto-encoder Registration



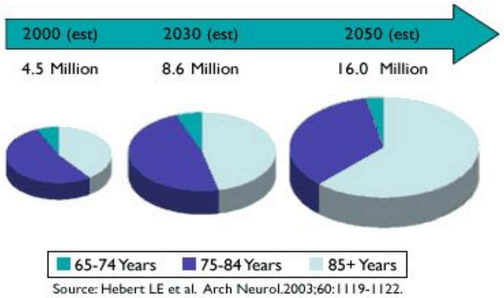



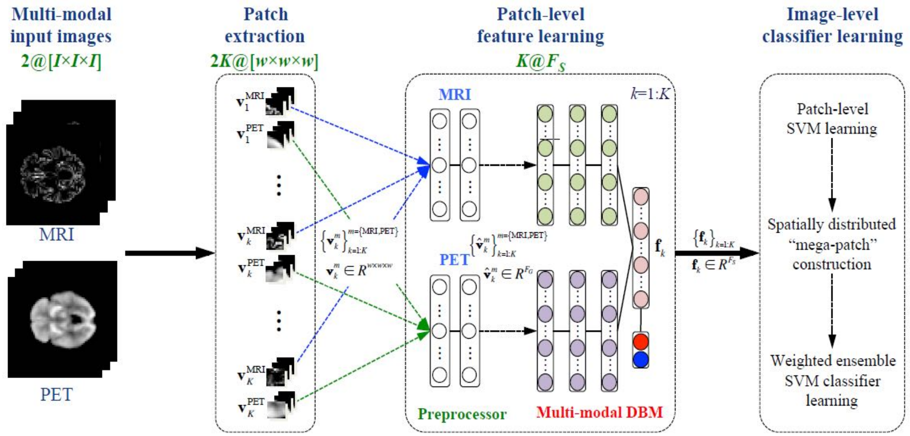

UNC, Dinggang Shen-2017

# Brain and Alzheimer Diseases

15,000 leading cause-of-death in Deutschland

6th leading cause-of-death in US

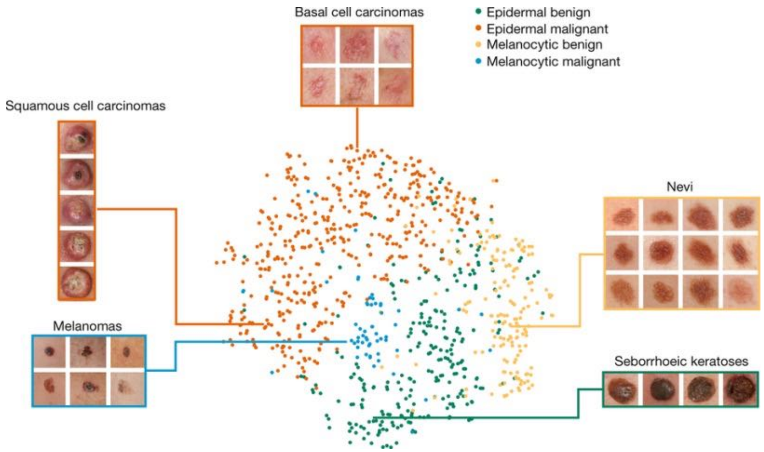



*I*: image size, *w*: patch size, *K*: # of selected patches, *m*: modality index,
$F_G$: # of hidden units in Gaussian restricted Boltzmann machine,
$F_S$: # of hidden units in the top-layer of multi-modal Deep Boltzmann Machine (DBM)

Esteva et al. “Dermatologist-level classification of skin cancer with deep neural networks”, in ***Nature***, 25 January 2017

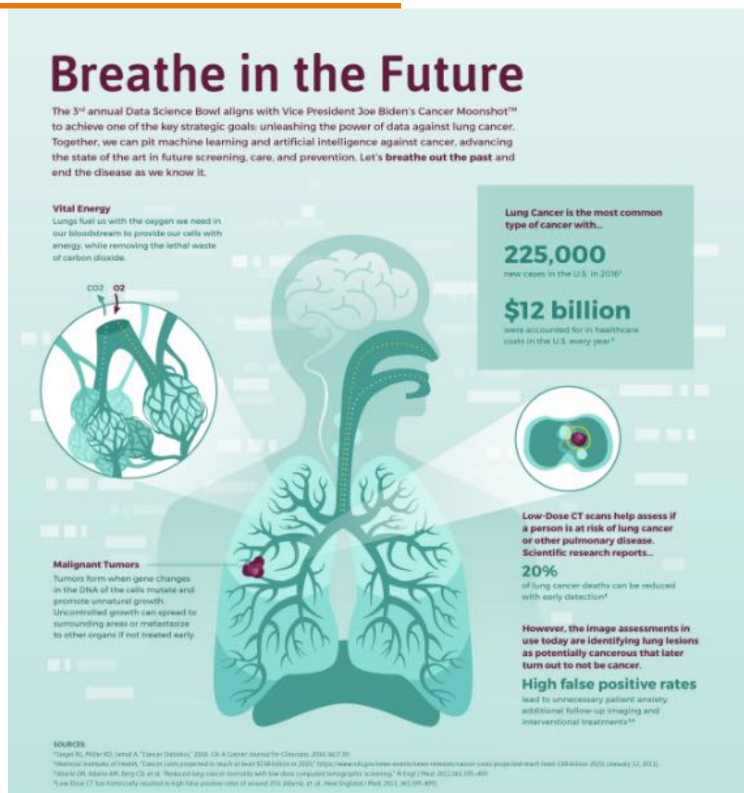
- Demonstrates capabilities of artificial intelligence in classifying skin cancer with a level of competence comparable to dermatologists



# Lung Cancer Detection

- Strikes approx. 225,000 people every year just in USA
- Causes approx. USD 12 billion in health care costs.

- Data Science Bowl 2017 Challenges
  (USD 1,000,000)
- Lung Nodule Detection (2015-2017,ISBI)

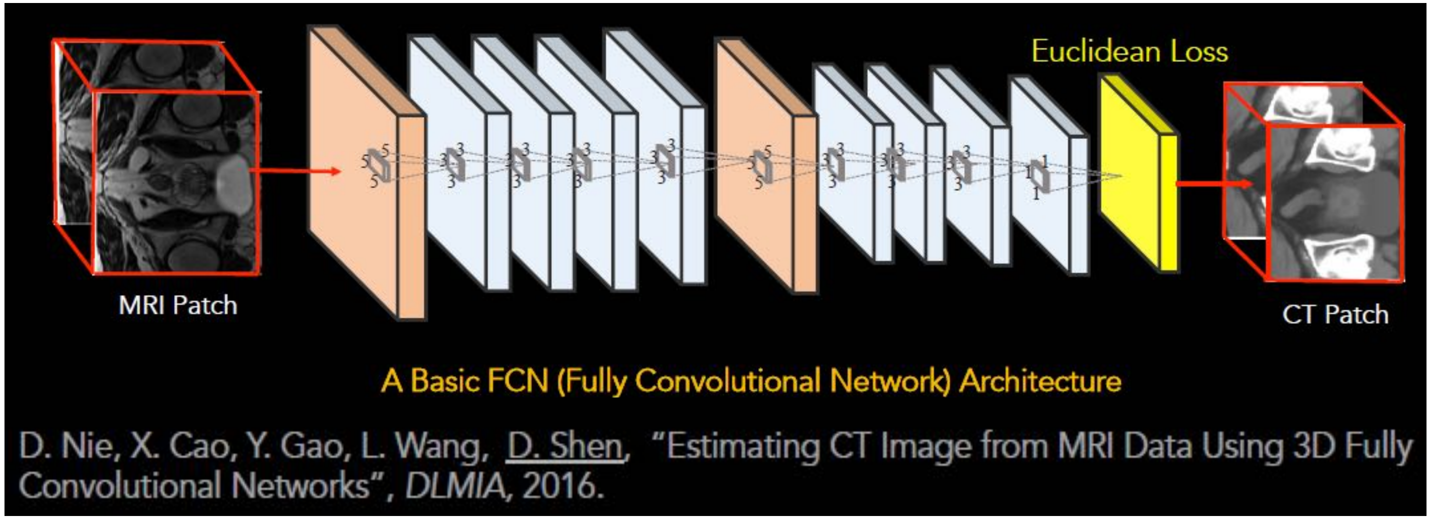


## Estimating CT from MRI with Fully Convolution Neural Network

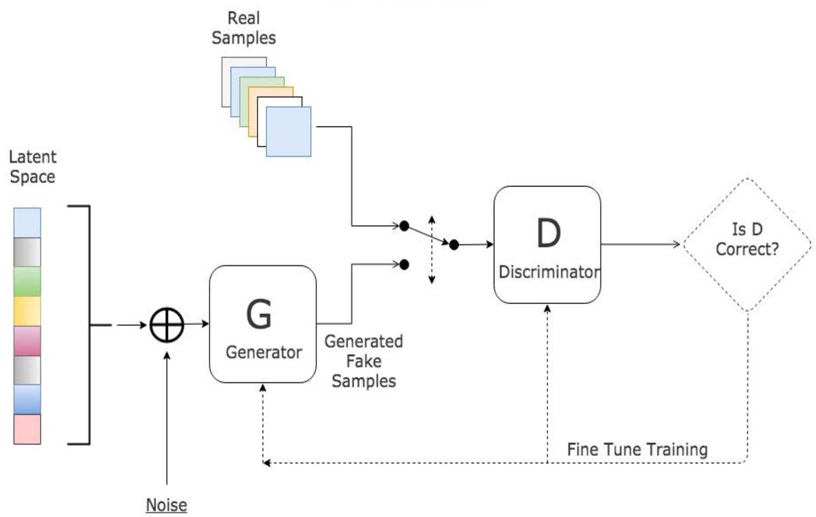
(+) Bone injuries

(+) Organs in the pelvis, chest and abdomen

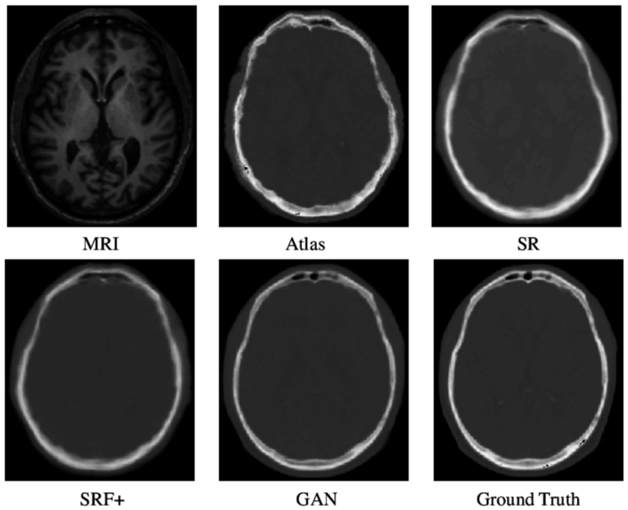
(-) CT images use Radiation



# Unsupervised Learning for Estimating CT from MRI




(Goodfellow 2016)




UNC, Dinggang Shen-2017

# Popular Neural Network Architecture in Medical Imaging