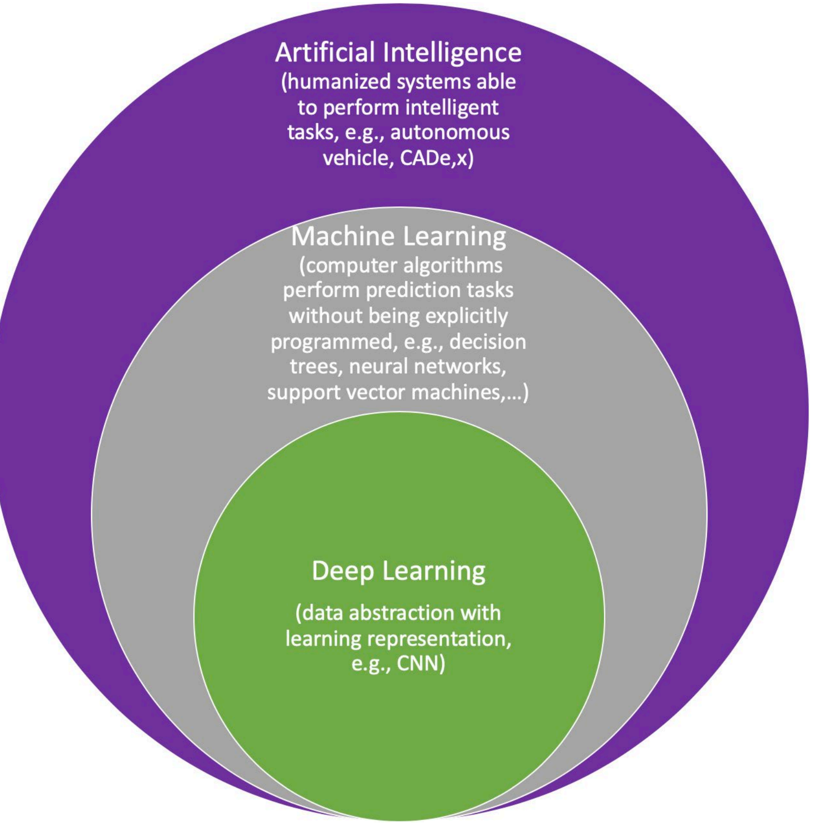
Lecture 2: Training requirements for ML

Issam El Naqa, PhD

What is AI/ML/DL?



El Naqa, BJR 125th Annv., 2020

Artificial Intelligence

Originated in the 1950s

Build machines that think like humans

Machine Learning

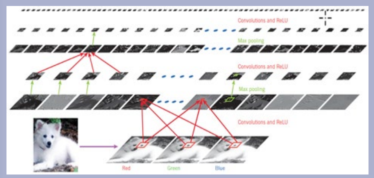
Originated in the 1960s

Computer algorithms that learn from data

Deep Learning

Originated in the 1970s

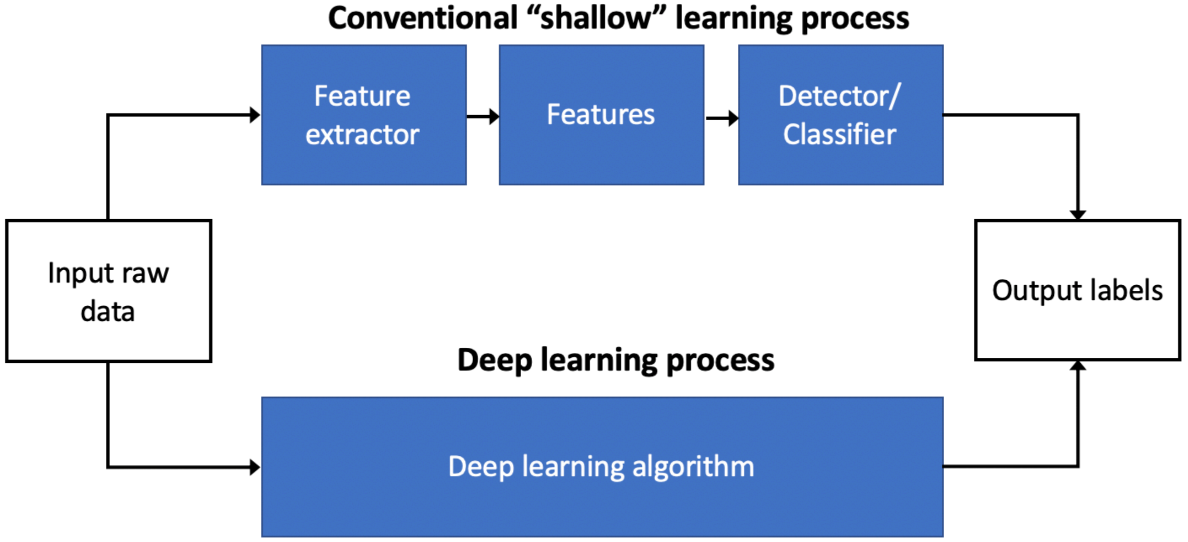
Based on neural networks that learn features

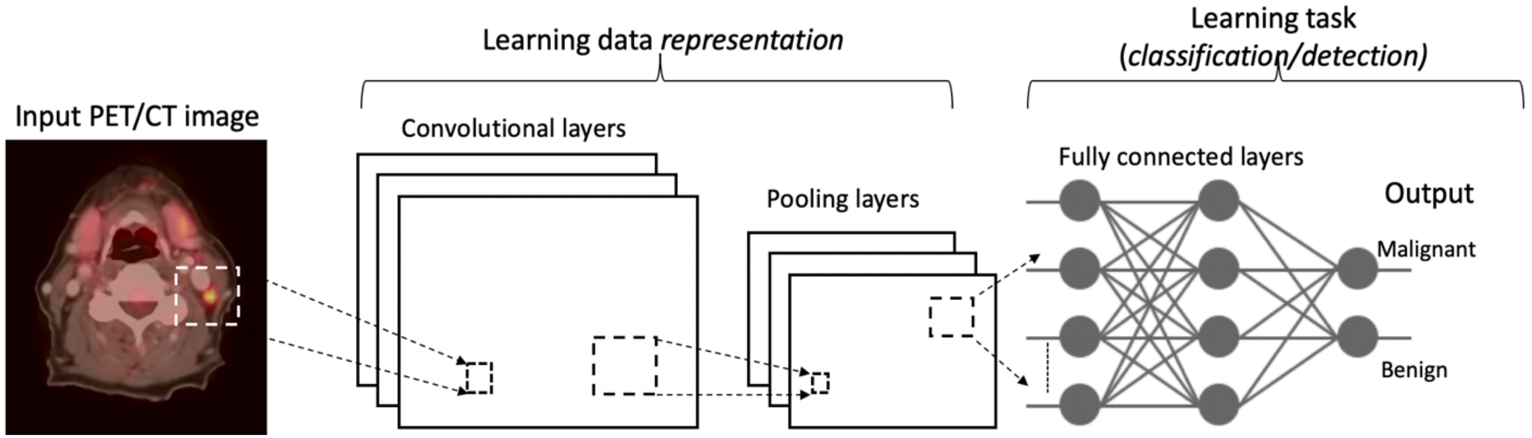


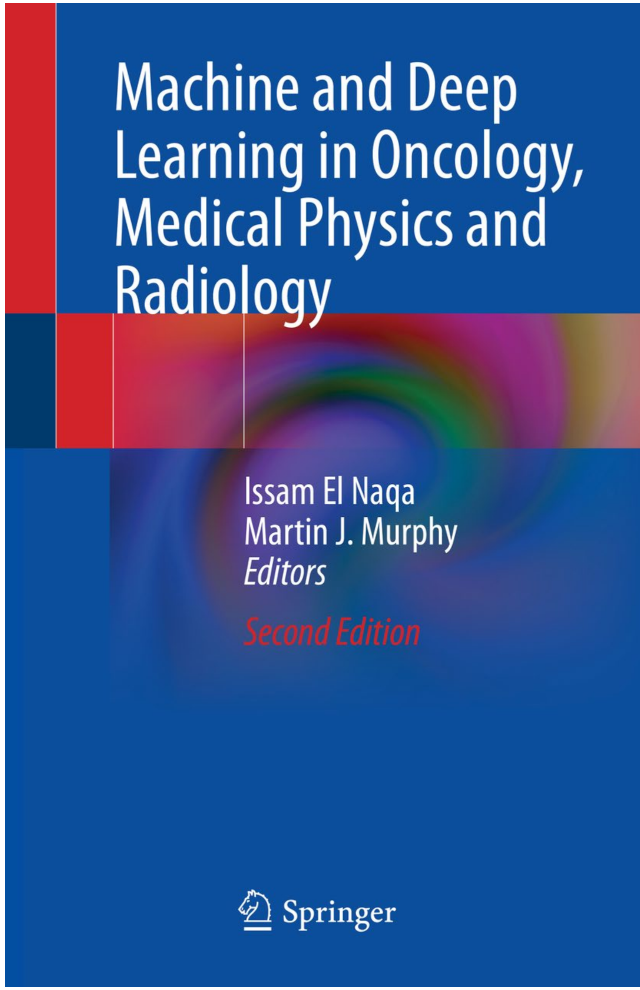
The Universe of Machine Learning (ML)

Adapted from: Brownlee, 2013

Deep vs conventional machine learning

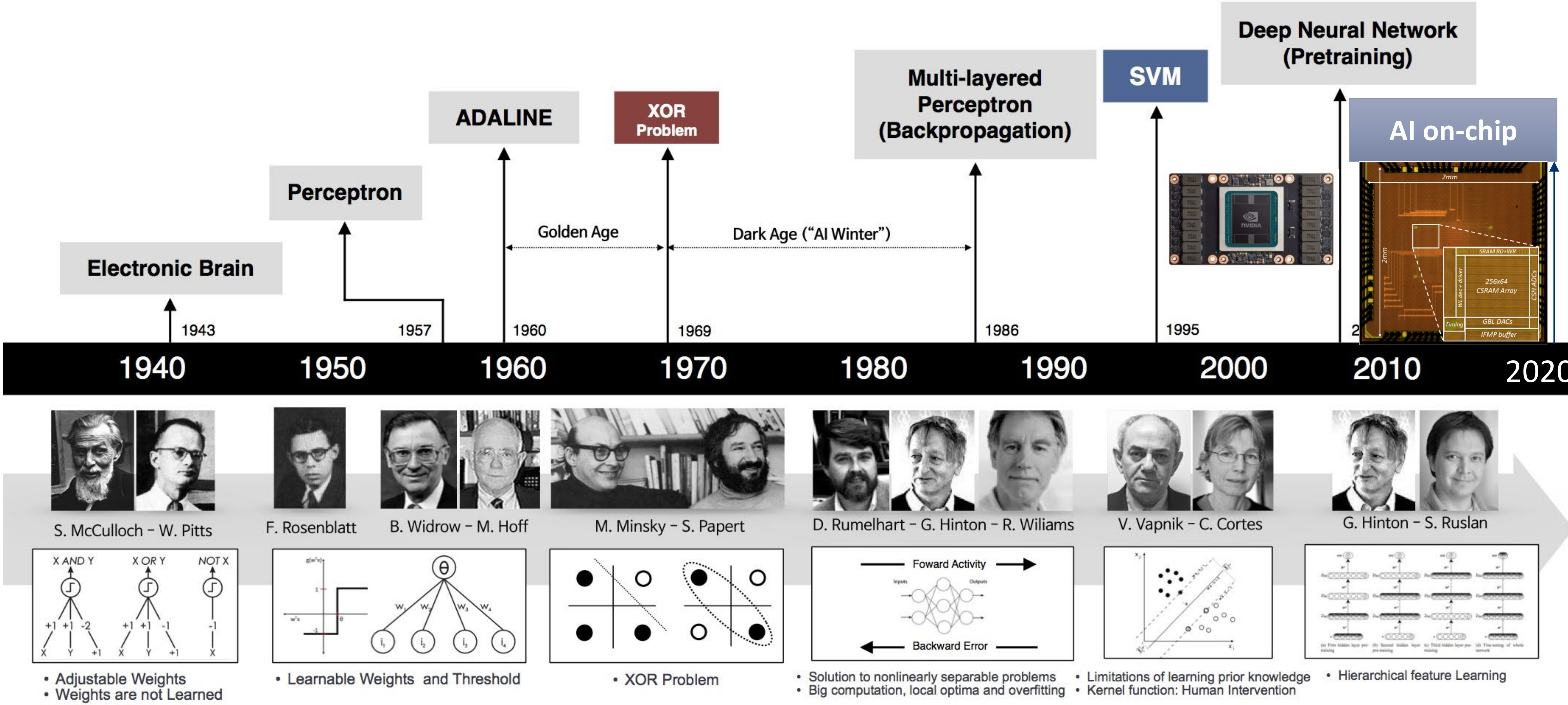






Zaidi and El Naqa, Annu. Rev. Biomed. Eng., 2021

Neural Networks Past and Present



Learning ML advice

Shane Legg
@ShaneLegg

Follow

Learn:
1. linear algebra well (e.g. matrix math)
2. calculus to an ok level (not advanced stuff)
3. prob. theory and stats to a good level
4. theoretical computer science basics
5. to code well in Python and ok in C++

Then read and implement ML papers and *play* with stuff! :-)

> **aron** @aron65900682
> @ShaneLegg Hey Shane I'm currently 17 from London England and am very passionate about AI, also learning about in-depth human needs. What would be the 5 pieces of advice and tips you would give to a young person like me?

Regression methods

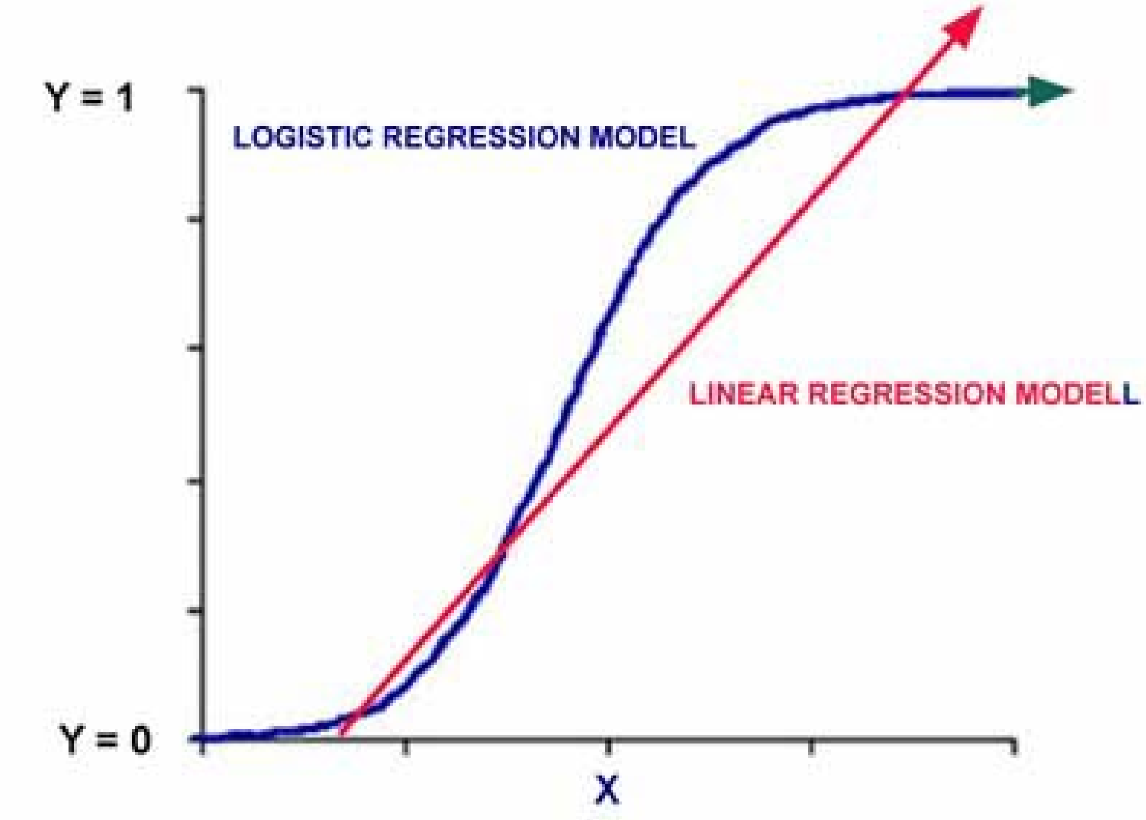
Regression model is:

$$Y \sim Model(X\beta)$$

Model: linear, logit, probit

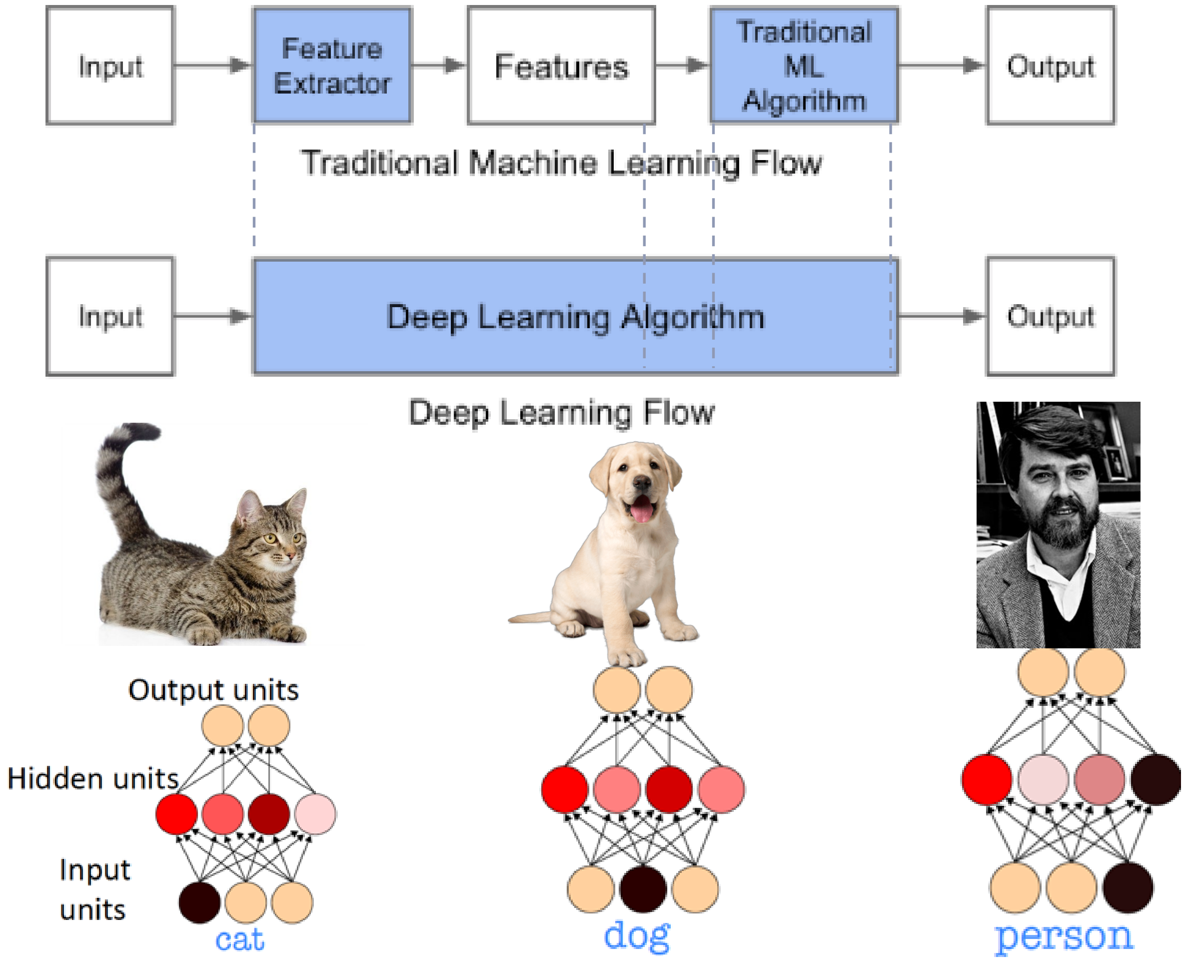
- ➢ $X_1, X_2, \ldots, X_n$ linearly additive
- ➢ Y and X have a monotonic relationship

- Ordinary estimation (no penalty)
- Regularization (shrinkage) estimation:
  - Ridge $$Y \sim Model(X\beta) + \lambda \sum_i \beta_i^2$$
  - Lasso $$Y \sim Model(X\beta) + \lambda \sum_i |\beta_i|$$
- Estimation methods: Least-squares, maximum likelihood



Steyerberg, 2008; Hastie 2015, Moiseenko 2021

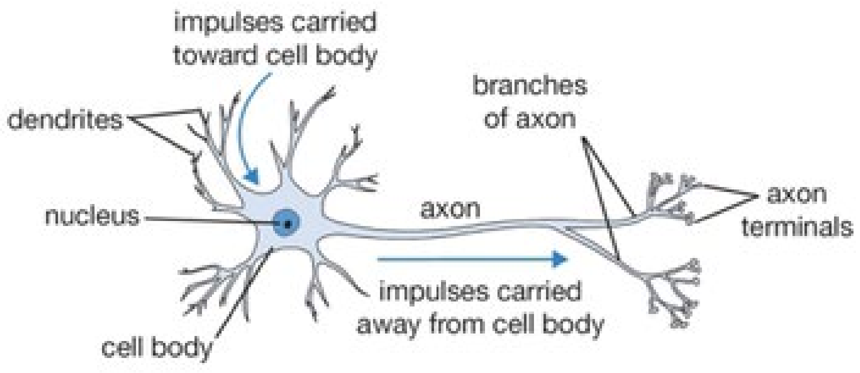
Deep learning with Neural Networks

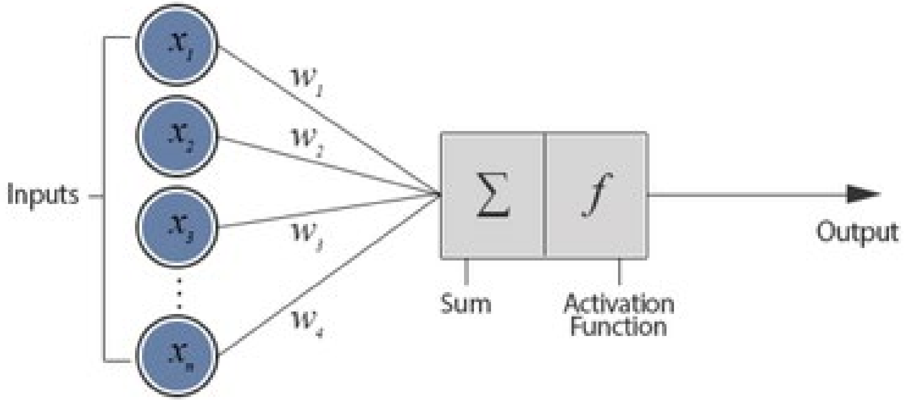


The key aspect of deep learning is that these layers of features are not designed by human engineers: they are learned from data using a general-purpose learning procedure

Rumelhart, 1986

Biological versus Artificial Neuron





Name	Plot	Equation	Derivative
Identity		$f(x) = x$	$f'(x) = 1$
Binary step		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x \neq 0 \\ ? & \text{for } x = 0 \end{cases}$
Logistic (a.k.a Soft step)		$f(x) = \frac{1}{1+e^{-x}}$	$f'(x) = f(x)(1-f(x))$
TanH		$f(x) = \tanh(x) = \frac{2}{1+e^{-2x}} - 1$	$f'(x) = 1 - f(x)^2$
ArcTan		$f(x) = \tan^{-1}(x)$	$f'(x) = \frac{1}{x^2+1}$
Rectified Linear Unit (ReLU)		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Parameteric Rectified Linear Unit (PReLU)[2]		$f(x) = \begin{cases} \alpha x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Exponential Linear Unit (ELU)[3]		$f(x) = \begin{cases} \alpha(e^x - 1) & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} f(x) + \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
SoftPlus		$f(x) = \log_e(1+e^x)$	$f'(x) = \frac{1}{1+e^{-x}}$

Artificial neural networks (ANN)

Select an architecture (problem-dependent)

- Inputs, outputs, type and number of hidden layers, activation functions

Define a cost (loss) function

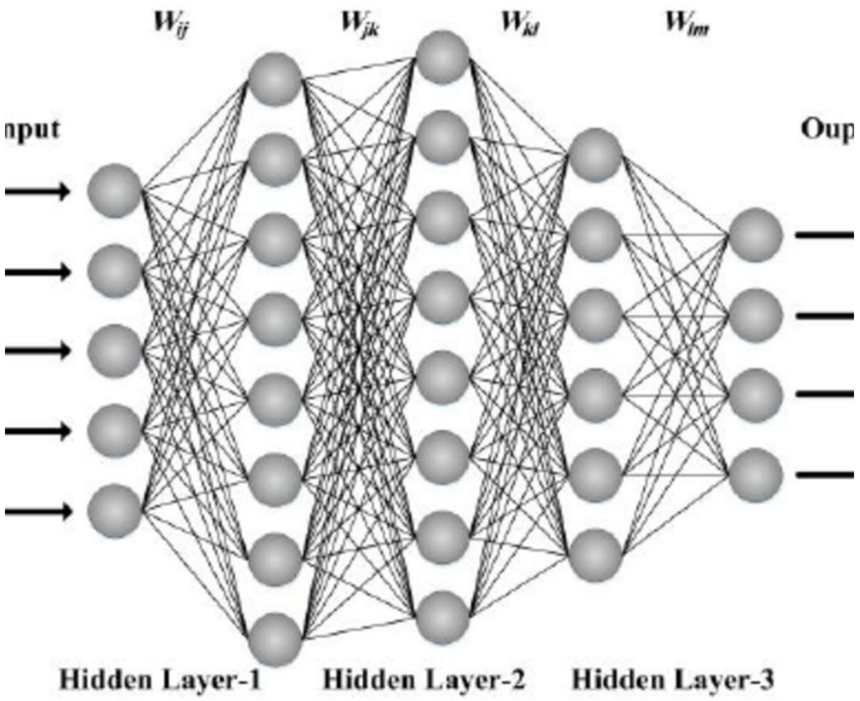
- Quadratic (Rumelhart), exponential, cross-entropy, KL divergence, etc.

Choose a training (learning) technique

- Back propagation (gradient-descent) algorithms, Newton's methods, Conjugate gradients, LM algorithm, etc.

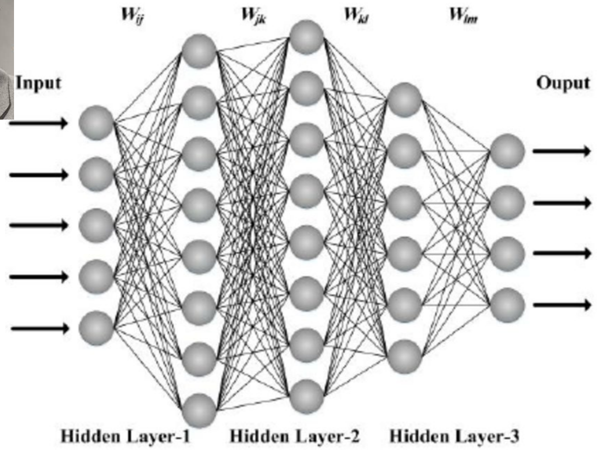
Regularization

- Norm penalty (L1, L2), early stopping, dropout, data augmentation, transfer learning, etc.

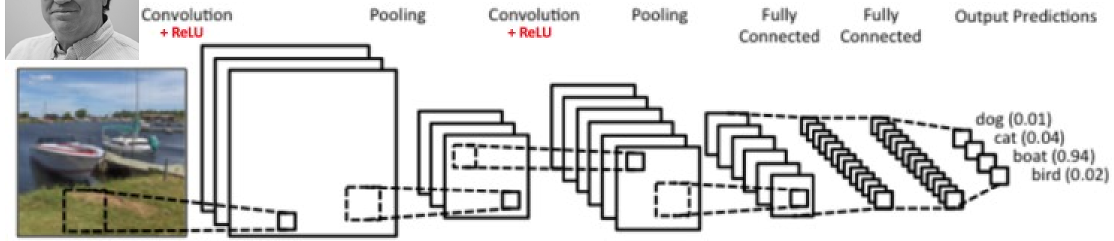


Deep Learning (NN) Architectures

Multi-layer neural perceptron (MLP)

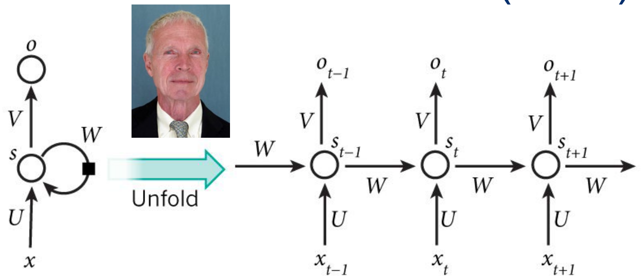


Convolutional Neural Network (CNN)



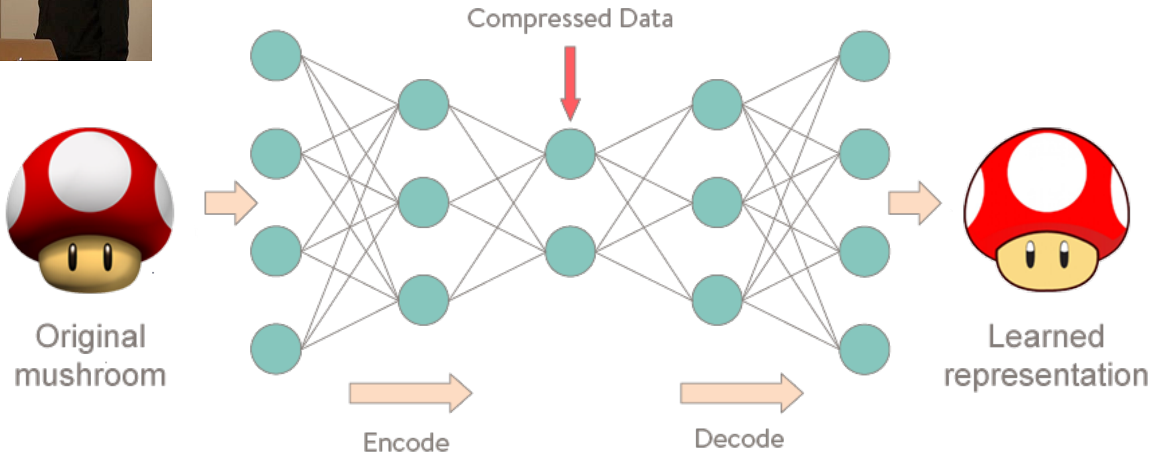
(LeNet, AlexNet, ResNet, VGG, U-net, DBN, DQN, …)

Recurrent Neural Network (RNN)

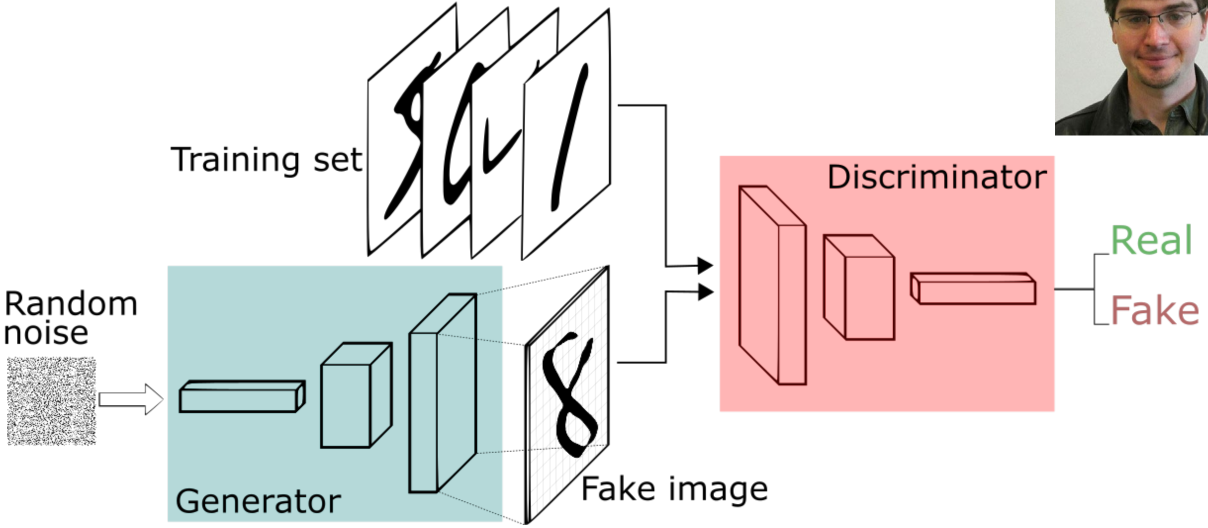


(LSTM, GRU, Transformers, …)

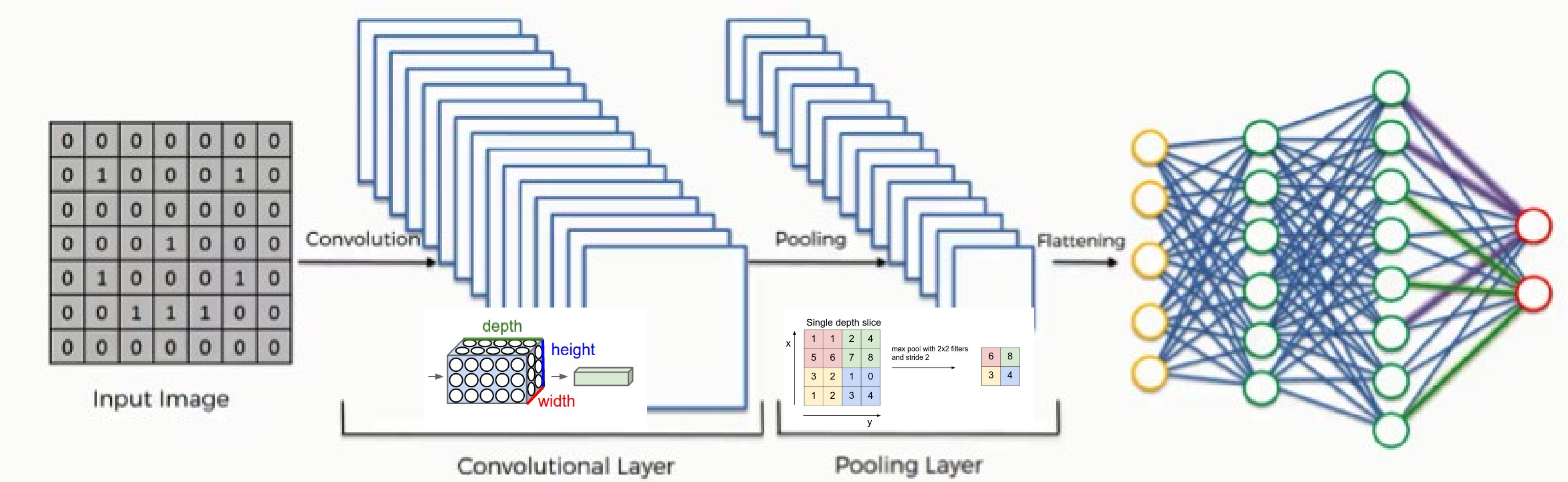
Autoencoders (AE)



Generative adversarial networks (GAN)

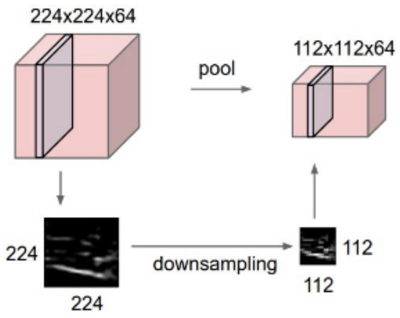


Example architecture: CNN

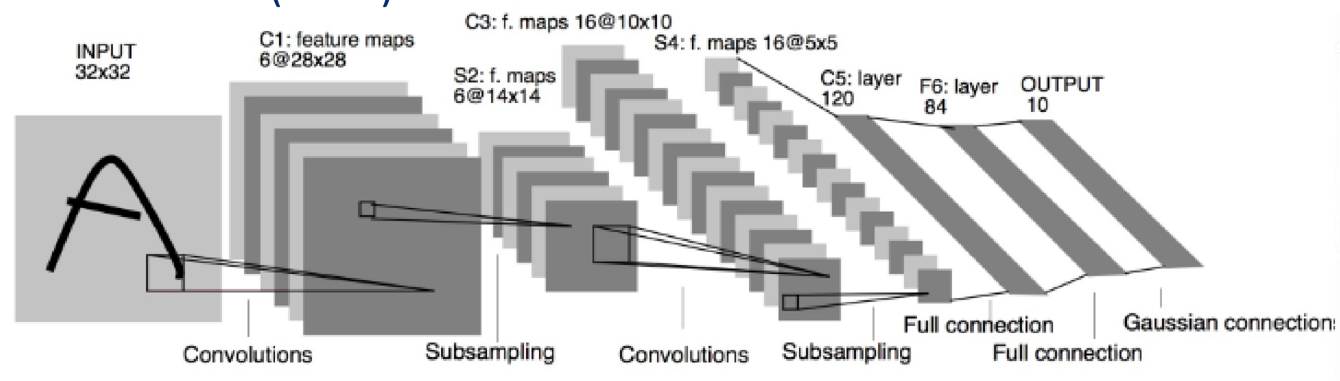


Kernel (Filter) parameters:

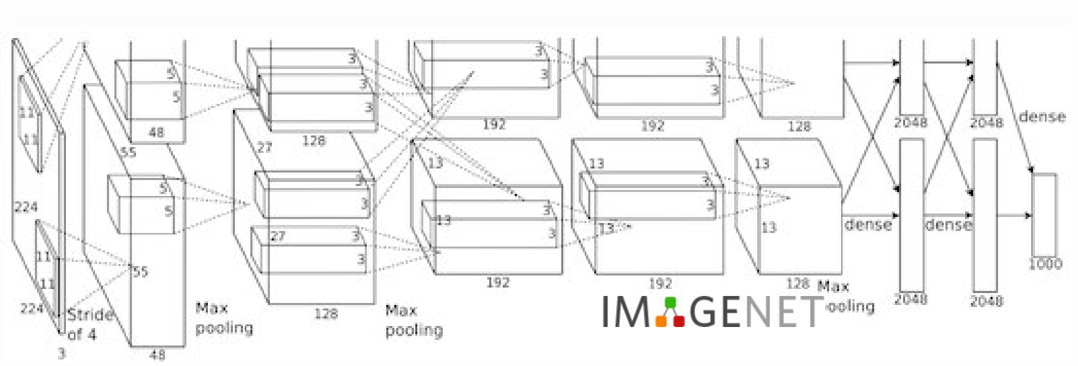
- Size (s)
- Depth: number of neurons
- Stride: amount of shift
- Zero-padding

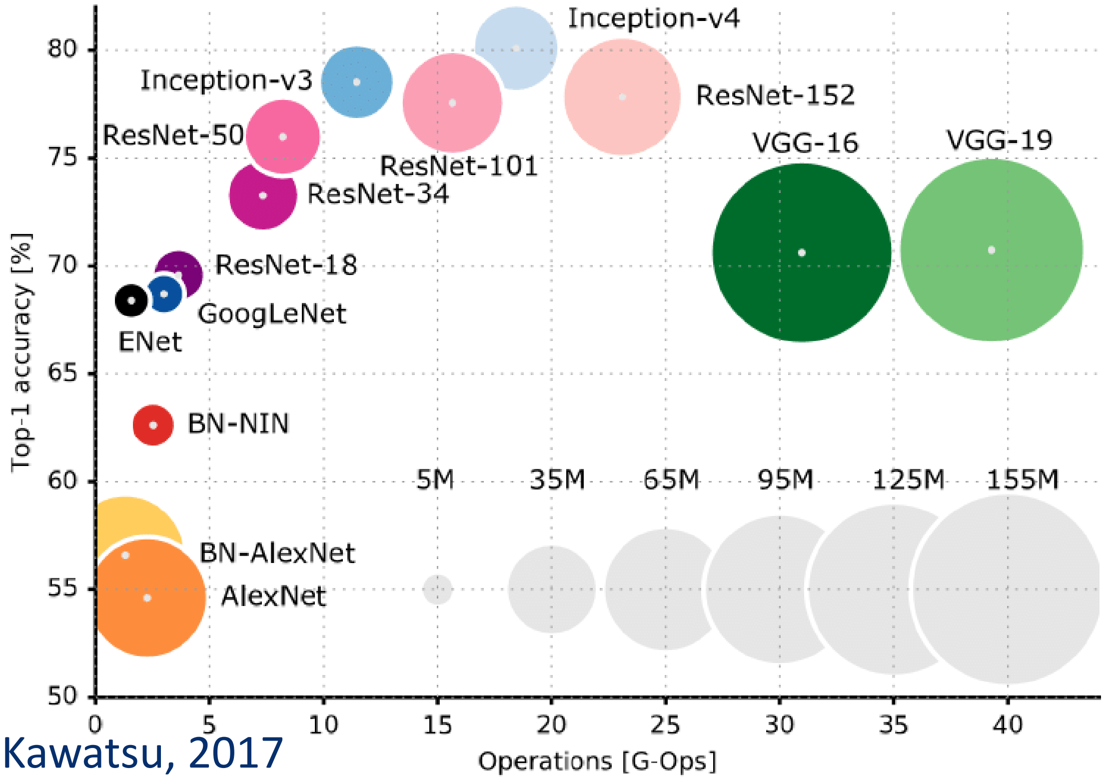


CNN- LeNet5 (1998)



Alex Krizhevsky et al, 2012

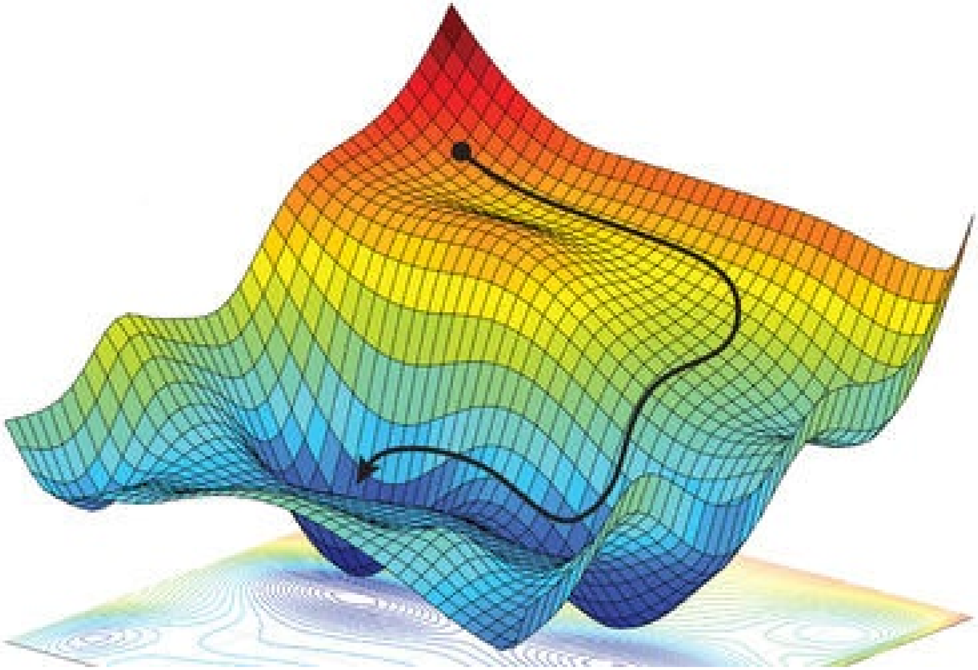




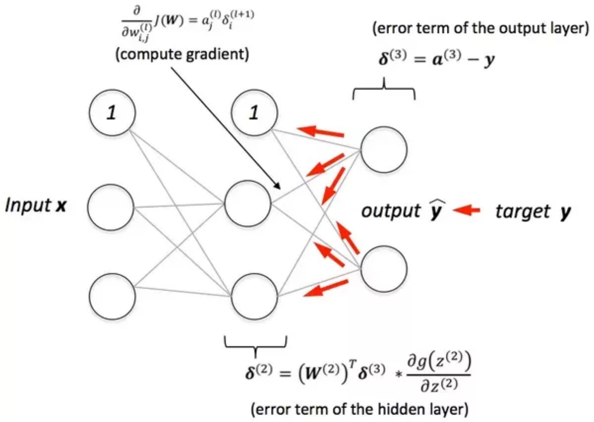
Kawatsu, 2017

Backpropagation algorithm: Gradient descent

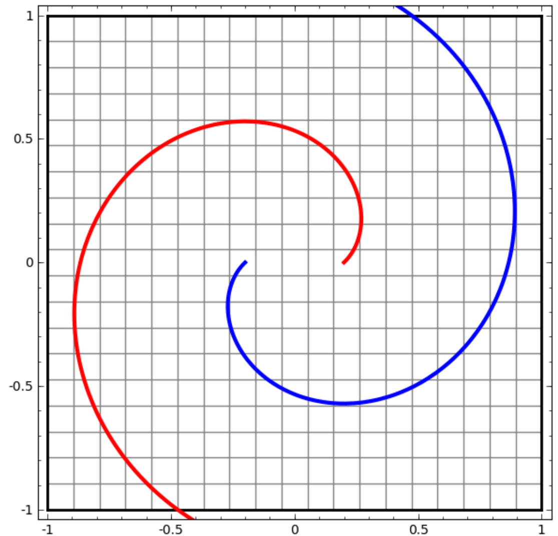
- Basic gradient descent
  - $w_{i+1}=w_i-\eta\cdot\nabla C(w; (X,y))$
- Stochastic gradient descent (SGD)
  - $w_{i+1}=w_i-\eta\cdot\nabla C(w; (X_{batch},y_{batch}))$
- Other modifications
  - Momentums, accelerations, adaptive gradient (Adagrad), adaptive momentum (Adam), etc.



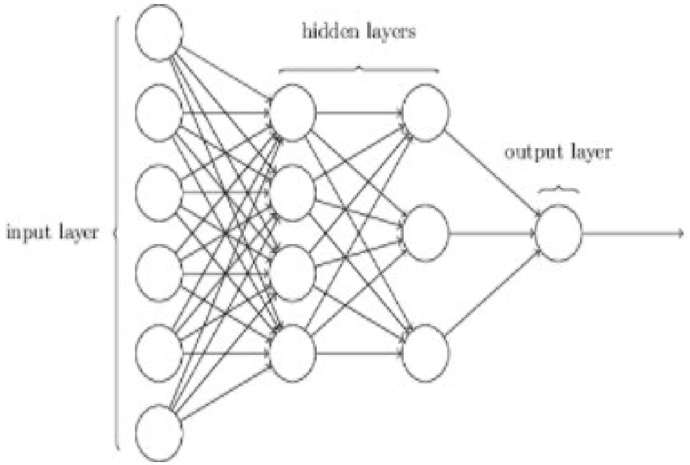
1. **Input** x**:** Set the corresponding activation a^1 for the input layer.
2. **Feedforward:** For each $l = 2, 3, \dots, L$ compute $z^l = w^l a^{l-1} + b^l$ and $a^l = \sigma(z^l)$.
3. **Output error** δ^L**:** Compute the vector $\delta^L = \nabla_a C \odot \sigma'(z^L)$.
4. **Backpropagate the error:** For each $l = L-1, L-2, \dots, 2$ compute $\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$.
5. **Output:** The gradient of the cost function is given by $\frac{\partial C}{\partial w^l_{jk}} = a^{l-1}_k \delta^l_j$ and $\frac{\partial C}{\partial b^l_j} = \delta^l_j$.

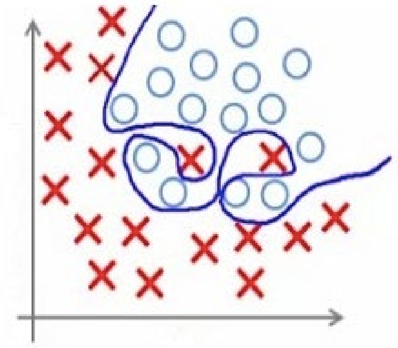


Modified from Nielsen, 2018



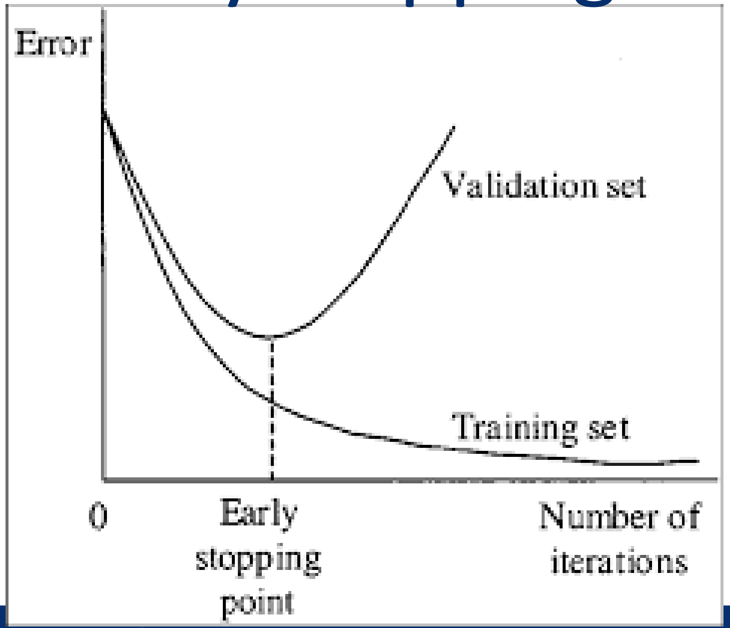
Learning Regularization



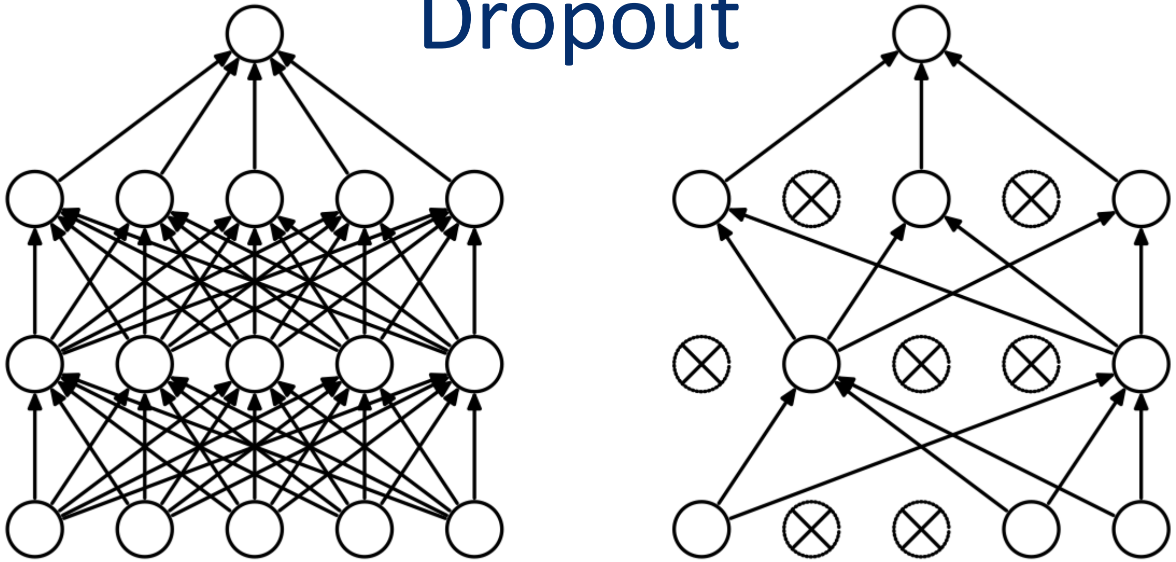


Over-fitting

Early stopping



Dropout



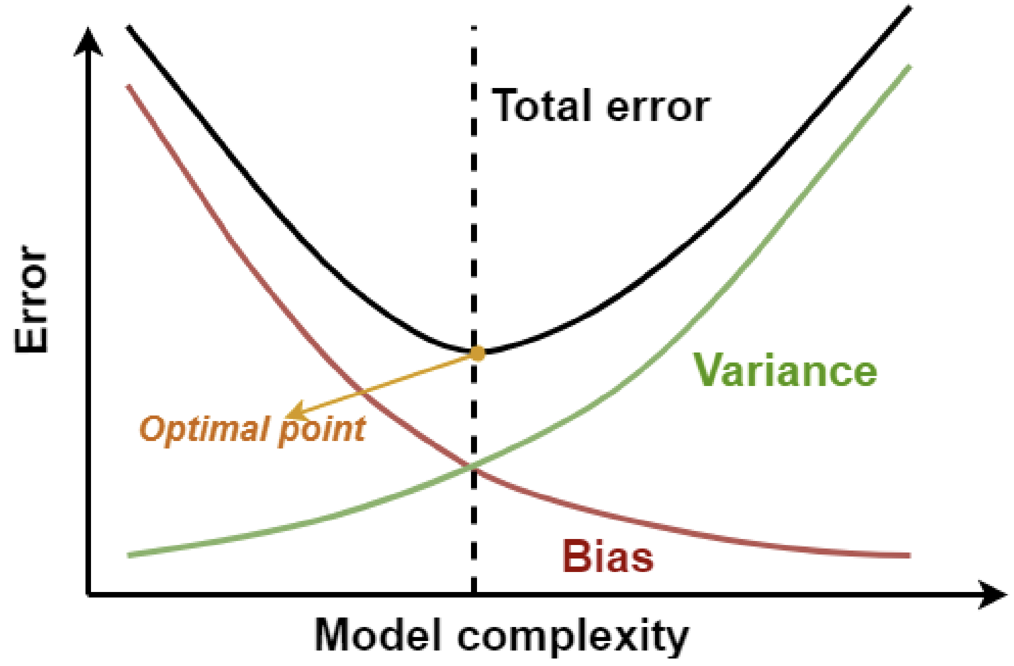
(a) Standard Neural Net

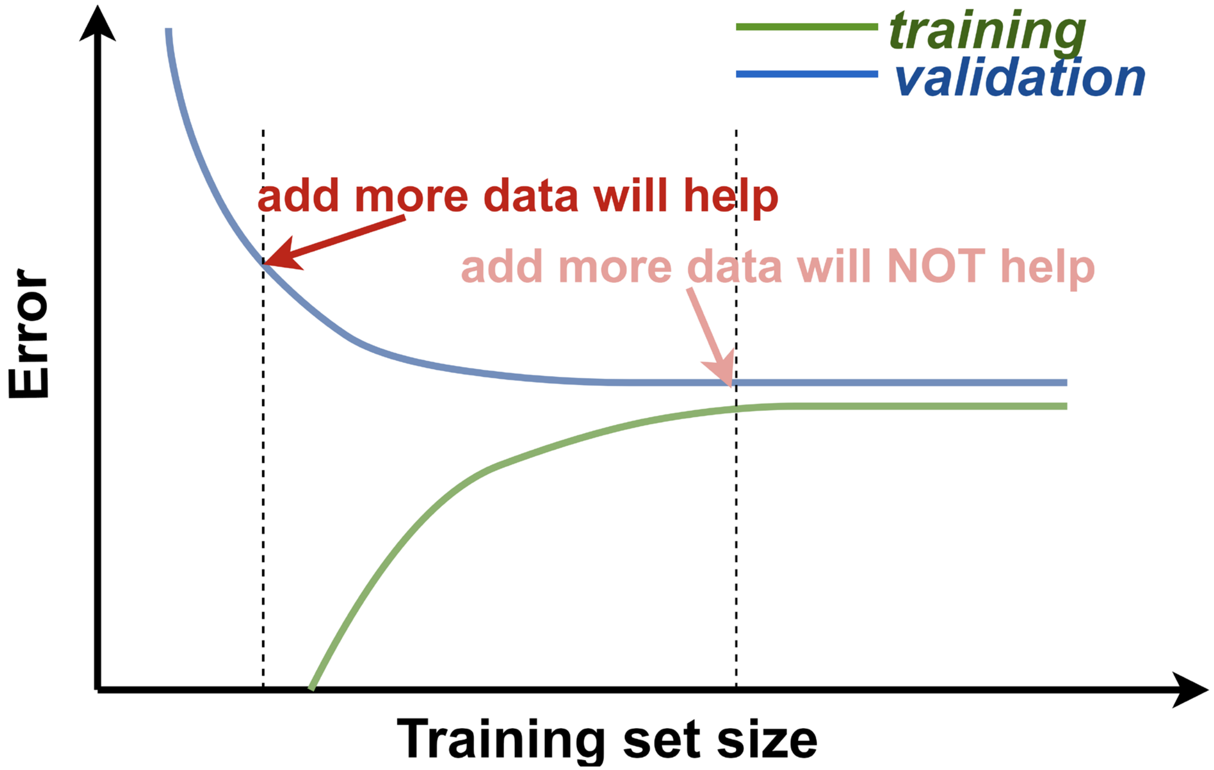
(b) After applying dropout.

What training sample size is required?

Introduction to Machine and Deep Learning for Medical Physicists

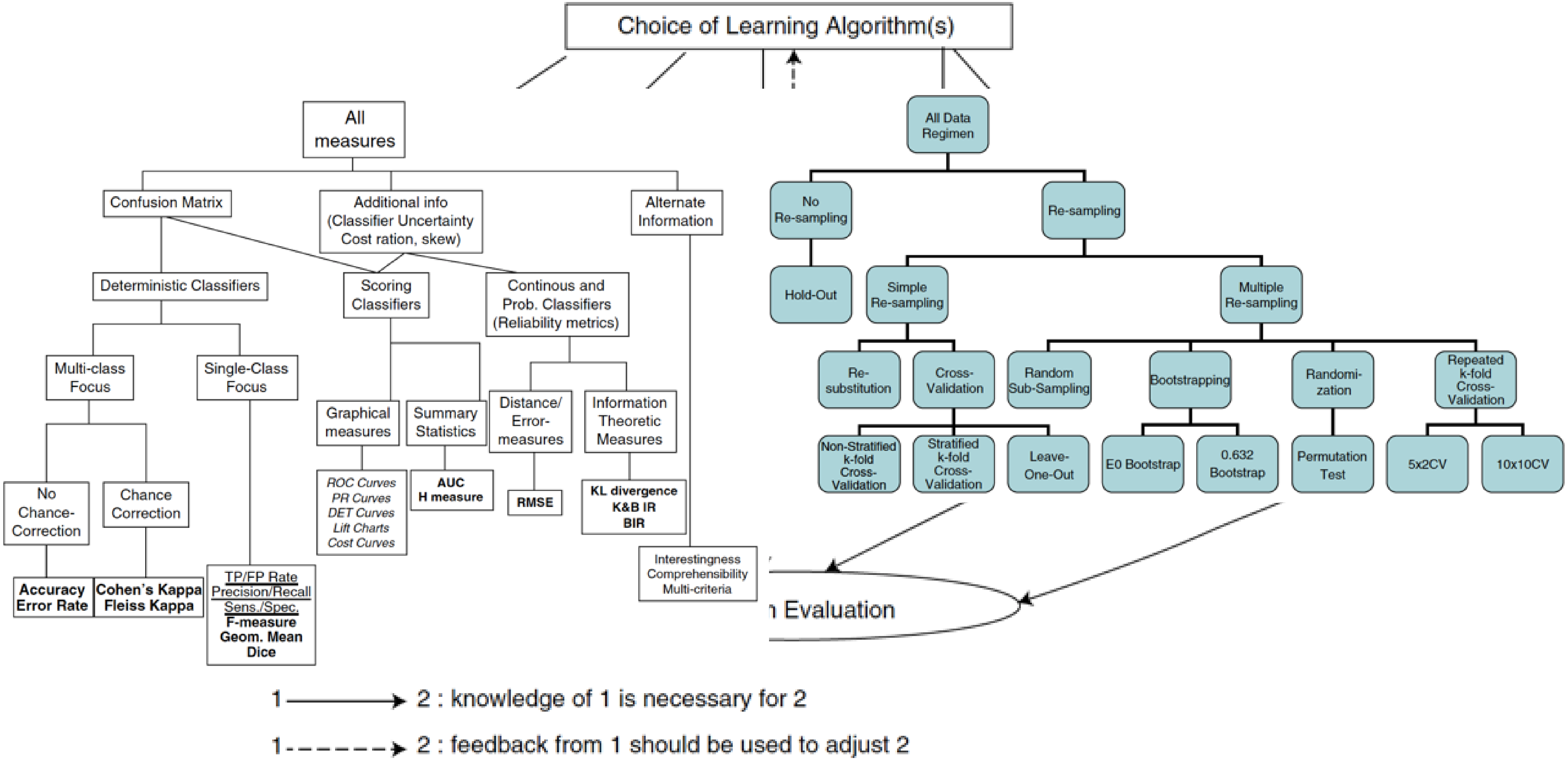
Sunan Cui,* Huan-Hsin Tseng,† Julia Pakela,‡ Randall K. Ten Haken,§ and Issam El Naqa¶
Department of Radiation Oncology, University of Michigan, Ann Arbor, MI 48103, USA





Cui, Medical Physics, 2020

What evaluation plan for ML?



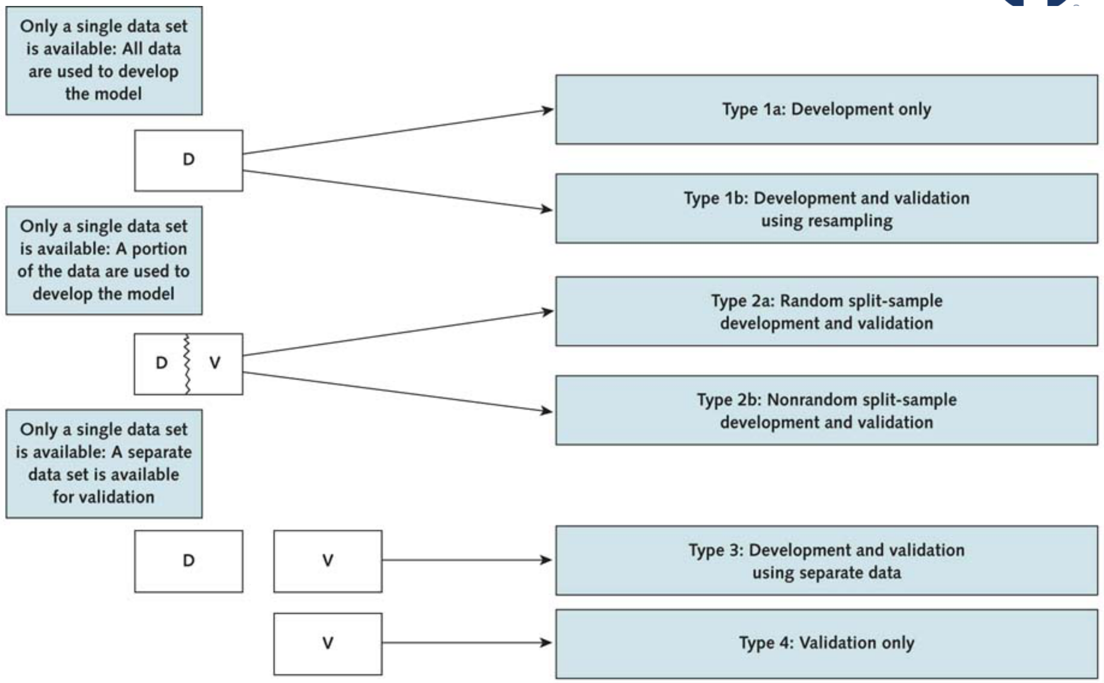
Japkowicz and Shah, 2015

AI/ML Validation

Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD)

Depending on the level of evidence

- Selection appropriate learning algorithms
- Validation and evaluation (TRIPOD criteria)
    - Internally (cross-validation schemes)
    - Externally (independent datasets)
- Checklists: CLAIM, MI-, CLAIM, CONSORT-AI, CLAMP, …
- Provide interpretation of machine learning prediction



Analysis Type	Description
Type 1a	Development of a prediction model where predictive performance is then directly evaluated using exactly the same data (apparent performance).
Type 1b	Development of a prediction model using the entire data set, but then using resampling (e.g., bootstrapping or cross-validation) techniques to evaluate the performance and optimism of the developed model. Resampling techniques, generally referred to as “internal validation”, are recommended as a prerequisite for prediction model development, particularly if data are limited (6, 14, 15).
Type 2a	The data are randomly split into 2 groups: one to develop the prediction model, and one to evaluate its predictive performance. This design is generally not recommended or better than type 1b, particularly in case of limited data, because it leads to lack of power during model development and validation (14, 15, 16).
Type 2b	The data are nonrandomly split (e.g., by location or time) into 2 groups: one to develop the prediction model and one to evaluate its predictive performance. Type 2b is a stronger design for evaluating model performance than type 2a, because allows for nonrandom variation between the 2 data sets (6, 13, 17).
Type 3	Development of a prediction model using 1 data set and an evaluation of its performance on separate data (e.g., from a different study).
Type 4	The evaluation of the predictive performance of an existing (published) prediction model on separate data (13).

Types 3 and 4 are commonly referred to as “external validation studies.” Arguably type 2b is as well, although it may be considered an intermediary between internal and external validation.

Quality assurance for AI/ML application in the clinic

Acceptance Testing

- To ensure that the ML tool meets all applicable safety and performance standards (prediction) and that it meets contractual specifications
- Manufacturer includes an acceptance test procedure with the ML tool
  - Selection of evaluation endpoint and definition of performance criteria (e.g., AUC);
  - Selection of a benchmark data

Commissioning

- The process whereby the needed tool-specific data/parameters are acquired and operational procedures are defined
- May include:
  - Training data collection
  - Developing procedures
  - User training before first use

Quality Assurance (QA)

- Effort to ensure treatments are given accurately, safely and efficiently according to established tests and evaluations

Continuing Quality Improvement (CQI)

- Effort that seeks to make treatments and operations better by recognizing current weaknesses in the program, anticipating problems before they happen, streamlining tasks and responding to changes in practice

312 THE MODERN TECHNOLOGY OF RADIATION ONCOLOGY

Table 10.1
Contemporary QA considerations for the current state of machine learning applications

TYPE OF MACHINE LEARNING APPLICATION	QA CONSIDERATIONS FOR THE CURRENT STATE			
	PERFORMED BY REVIEWED BY	COMMISSIONING	ROUTINE QA	RISK BEING MITIGATED
ML replaces human tasks: linear accelerator QA	Confirm functionality with sample QA data (Ritter et al. 2018)	• Evaluate ML against current clinic standards (Klein et al. 2009) • Test limits of analytics such as by inserting errors into delivery tests or datasets for analysis, e.g., intentional leaf offset present in the measurement result but missing in the delivery file • Document situations where the software passes and fails • Document situations where results differ by >5%	• Frequency: monthly • Monitor software settings for analysis • Repeat analysis of a subset of the commissioning dataset (e.g., dynamic leaf gap) including one at the limit • Expect identical results unless the software has changed. • If software has changed, determine if a new baseline is needed • Evaluate against a subset of the manual analysis for software update • Review trends	• Confirm that the analysis is performed correctly to avoid the hazards of expectation bias
ML supplemental to human tasks: treatment planning	• Confirm functionality with vendor-supplied treatment plans • Define scope of ML for planning	• Evaluate behavior against appropriate portions of original TPS commissioning results (if available) (Fraass et al. 1998) • Are clinical goals met? Is the agreement within ±5% for key metrics, such as mean dose for targets and max dose to a volume (e.g., 1 cc)? • Evaluate ML tools for a range of body sites and have site-specific rollout of techniques for at least a limited number of body sites • Evaluate permissions of different user types for applying ML techniques (e.g., physicist vs. dosimetrist) • Have different users perform the same test case—results within 5%? • Establish procedures for quality control steps post-application of ML, e.g., MD and physicist review of final dose distribution	• Repeat analysis of a subset of the commissioning dataset (e.g., dynamic leaf gap) including one at the limit • Monitor key dosimetric results from ML techniques using Big Data Analytical tools where available by body site: e.g. target coverage and maximum dose to a volume (e.g., 1cc) for OARs (Mayo et al., 2017) • Add extra scrutiny on key metrics for the first 5 patients per body site	• Monitor for any unintentional shift in clinical practice due to settings in the ML algorithm • Maintain evaluation of plan against MD-provided goals (planning objectives) (Evans et al. 2016; Marks et al. 2013)

(continued next page)

CHAPTER 10: MACHINE LEARNING IN RADIATION ONCOLOGY 313

Table 10.1 *(continued)*
Contemporary QA considerations for the current state of machine learning applications

TYPE OF MACHINE LEARNING APPLICATION	QA CONSIDERATIONS FOR THE CURRENT STATE			
	PERFORMED BY REVIEWED BY	COMMISSIONING	ROUTINE QA	RISK BEING MITIGATED
ML/AI enhances human tasks: patient workflow, such as preparation for optimization	Confirm functionality and understand the scope of what is automated	• Define if ML tools will be applied and implemented for all patients or by body site • Create a commissioning dataset which includes manual preparation of the plan for optimization and automated preparation • Confirm reasonably concordant results between human and automated creation • Inspect the overlay of human vs. automated volumes to confirm expansions are correct • Verify volumes for optimization are within 5% or 2 cc (for optic and other small structures)	• Repeat a subset of the commissioning dataset • Confirm derivative structures such as optimization structures are consistent with those by humans (monthly) • Confirm that quality control steps post-application remain in place, such as review of the final dose distribution by MD and physicist	• Risk being mitigated is an incorrect expansion from target or OAR volumes to create optimization structures for dose coverage or sparing, respectively • Maintain evaluation of plan against MD provided goals (planning directive) (Evans et al., 2016; Marks et al. 2013)
ML additive: decision-making (El Naqa et al. 2018a)	• Evaluate with vendor-supplied dataset • Define size of training and testing dataset	• Partner with physicians to determine which disease types and staging are appropriate for the algorithm • Assess baseline variation in clinical practice among physicians within a practice, within a registry, or via publications before implementation • Assess sensitivity of the output of algorithms with training sets across the spectrum of limited variability to significant variability • Is the algorithm supporting implementation of a national practice standard? • Is the algorithm being used to apply new science in a clinical trial?	• Confirm that the input and expected output are consistent with the intent of the practice • Assess the frequency of patient type to determine how often the training dataset should be updated • Monitor the relationship between decisions with prior practice using Big Data Analytical tools where available by body site	

MEDICAL PHYSICS
The International Journal of Medical Physics Research and Practice

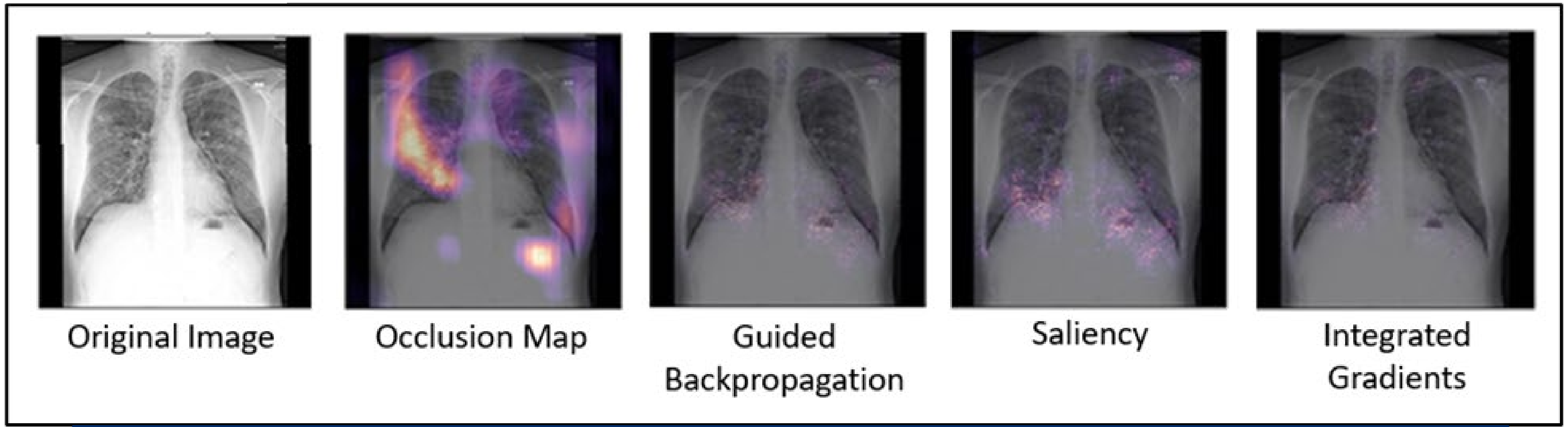
REVIEW ARTICLE | Free Access

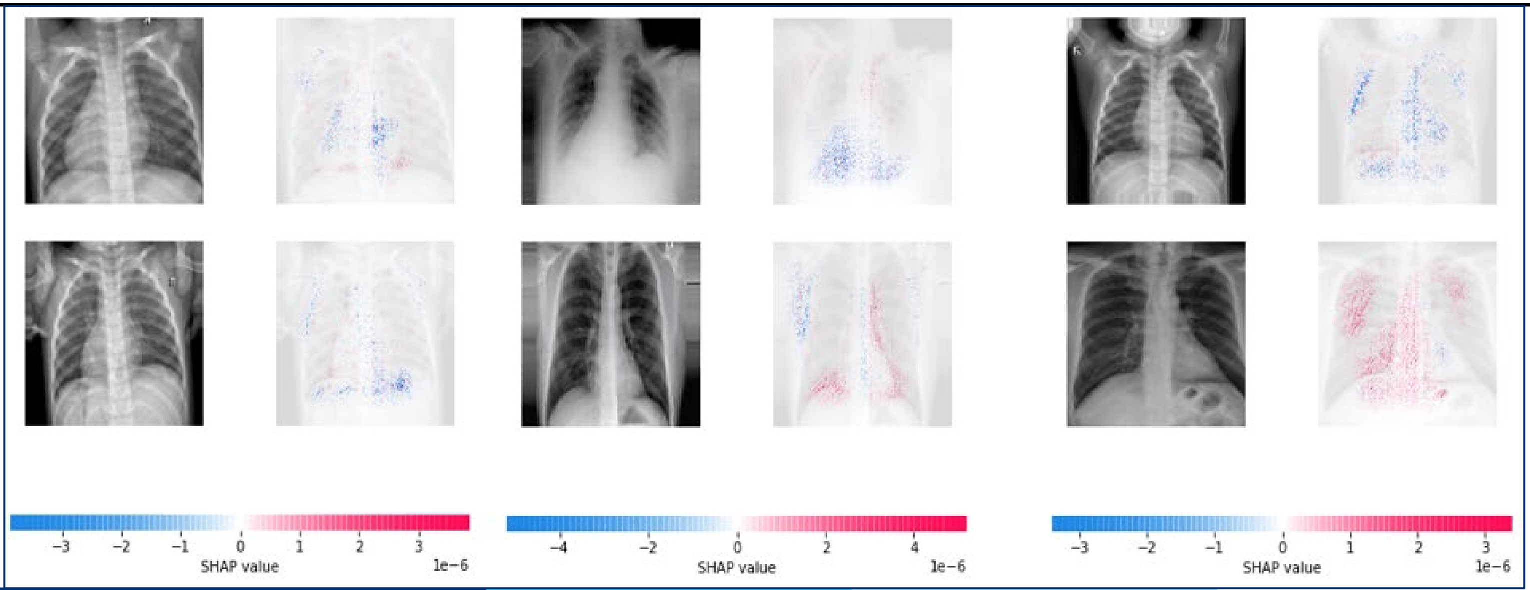
A review of explainable and interpretable AI with applications in COVID-19 imaging

Jordan D. Fuhrman, Naveena Gorre, Qiyuan Hu, Hui Li, Issam El Naqa, Maryellen L. Giger

First published: 18 November 2021 | https://doi.org/10.1002/mp.15359

Senior author: Maryellen L. Giger m-giger@uchicago.edu





ML Accuracy versus interpretability

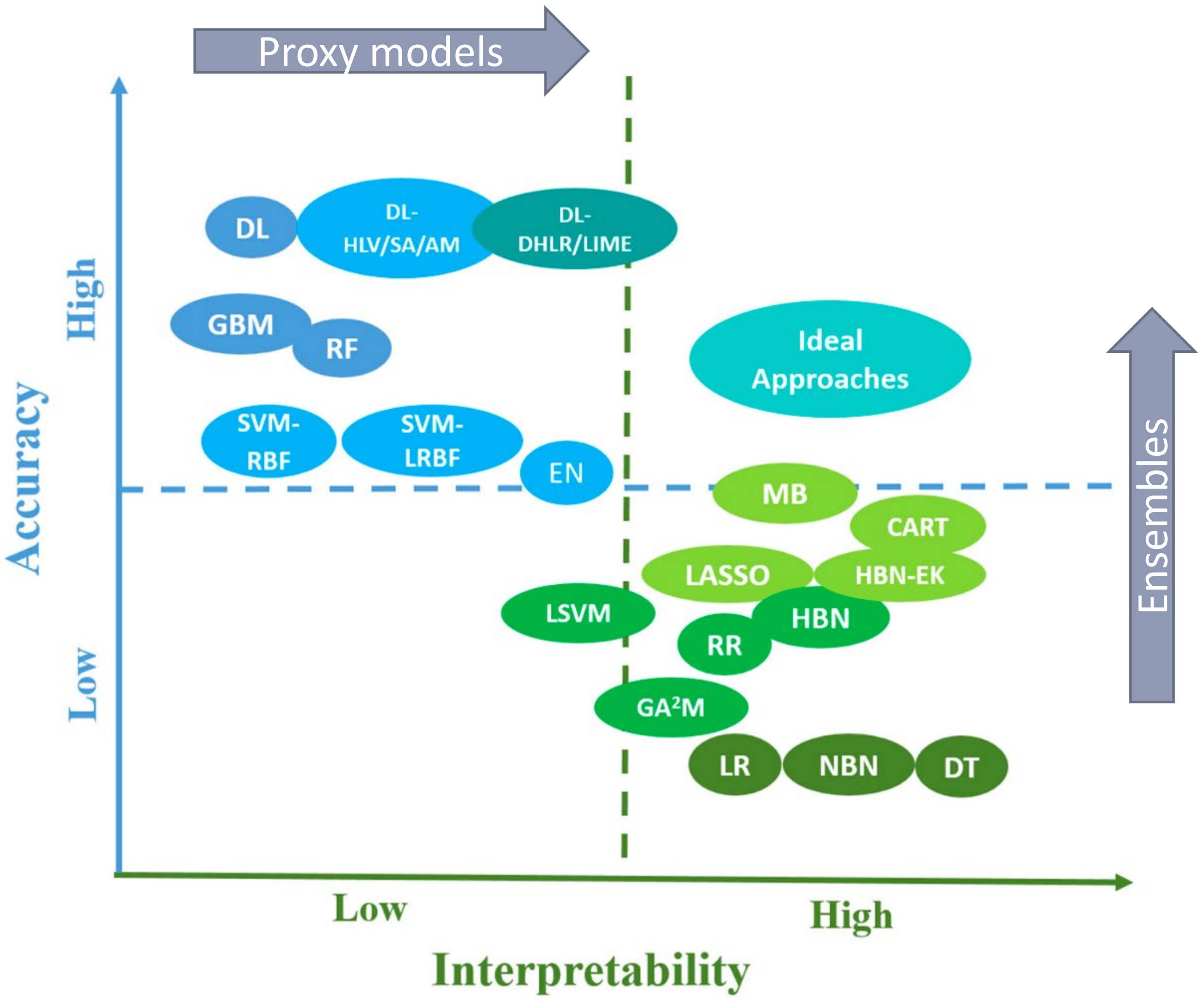


Table 1. The evaluation of the accuracy (**A**), interpretability (**I**) and explainability (**E**) of ML approaches in radiation outcomes prediction

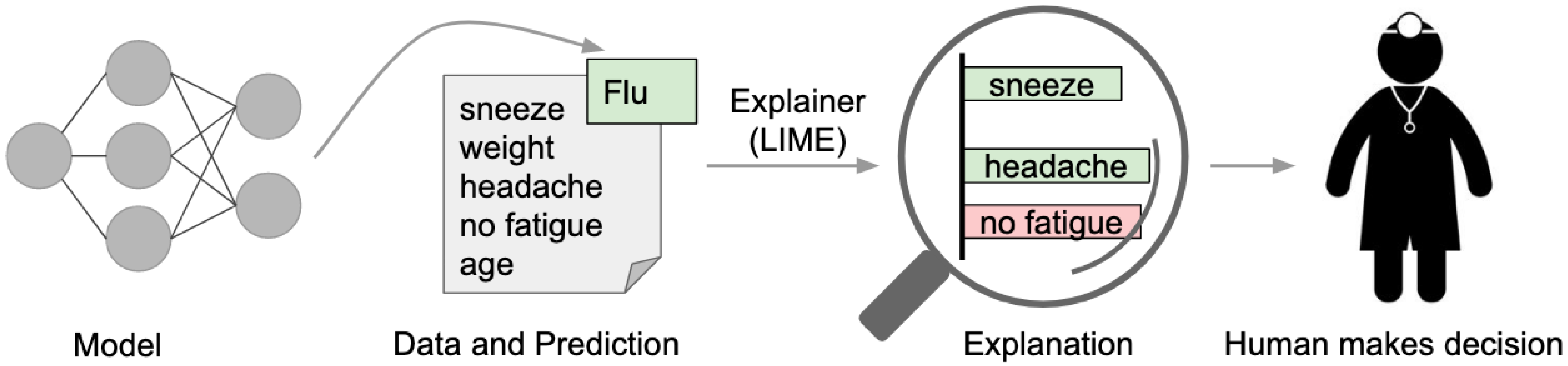
Basic ML	Type	A	I	E	Improved ML	Type	A	I	E
Logistic regression[20,21]	IP	*	****	***	GA^{2}M[68]	IP	**	***	**
					Ridge Regression[22]	IP	**	**	*
					LASSO[23]	IP	**	***	**
					Elastic Net[9,24]	IP	***	**	*
Decision tree[24,30,31]	IP	**	*****	*****	CART[32]	IP	***	****	*****
					Random Forests[7]	NIP	****	*	NA
					GBM[9,33]	NIP	****	*	NA
					MediBoost[9,34]	IP	****	**	*
Naïve BN[35,37]	IP	*	****	****	HBN[38,40]	IP	**	***	**
					HBN-EK[41]	IP	**	****	***
Linear SVM[24]	NIP	**	**	*	SVM-RBF[43]	NIP	***	*	NA
					SVM-LRBF[44]	NIP	***	**	*
Deep learning[49,50]	NIP	****	*	NA	DL-HLV[48,55,56]	NIP	*****	**	NA
					DL-SA[52,57] /AM[59,60]	NIP	*****	**	NA
					DL-DHLR[61–63]	NIP	*****	***	NA
					DL-LIME[69]	NIP	*****	***	NA

BN, Bayesian network; CART, classification and regression tree; DHLR, disentangled hidden layer representation; DL-AM, deep learning withattention mechanisms; DL-HLV, deep learning withcombination of handcrafted features and latent variables; GBM, gradient boosting machine; HBN, hierarchical Bayesian network; HBN-EK, hierarchical Bayesiannetwork with expert knowledge; HLV, handcrafted features and latent variables; IP, interpretable; LASSO, least absolute shrinkage and selection operator; LIME, local interpretable model-agnostic explanation; ML, machine learning; NIP, non-interpretable; SVM, support vector machine.

Luo, BJR-O, 2019

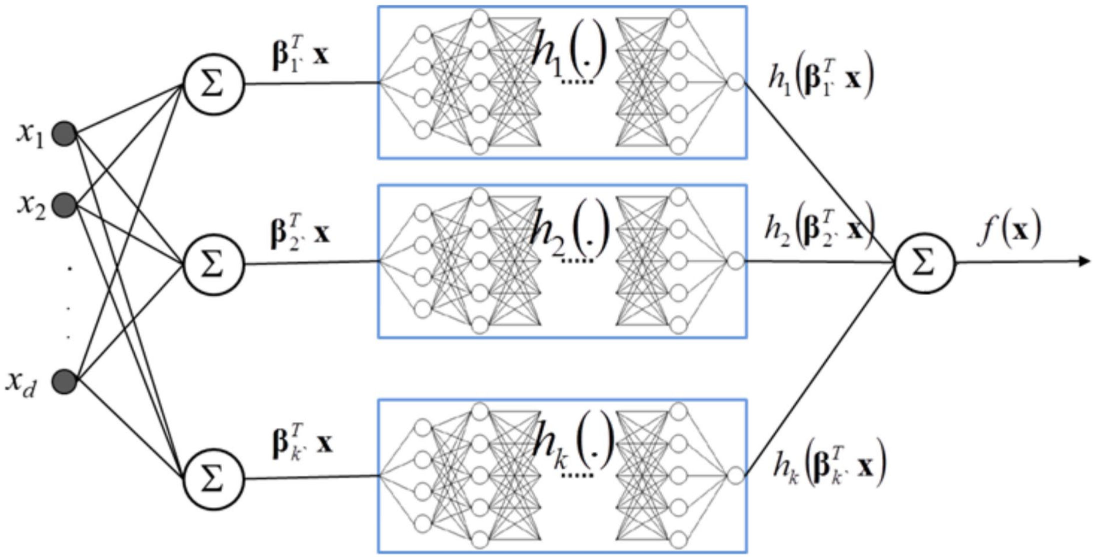
Deep learning interpretability approaches

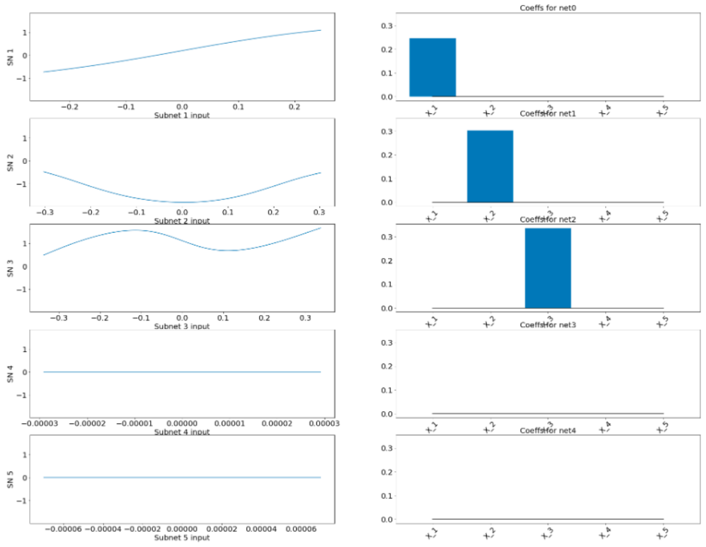
Post-modeling proxy Models



Interpretable Components

Explainable Neural Networks (xNN) based on Additive Index Models





Why interpretability important?

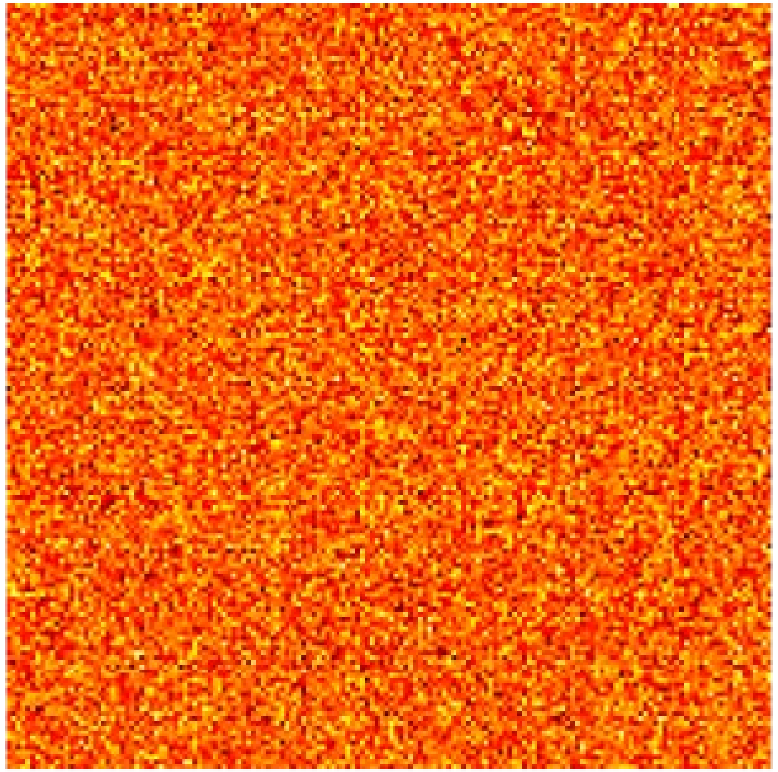
Adversarial Attacks

Original image

Sylvester Stallone

Noise perturbation

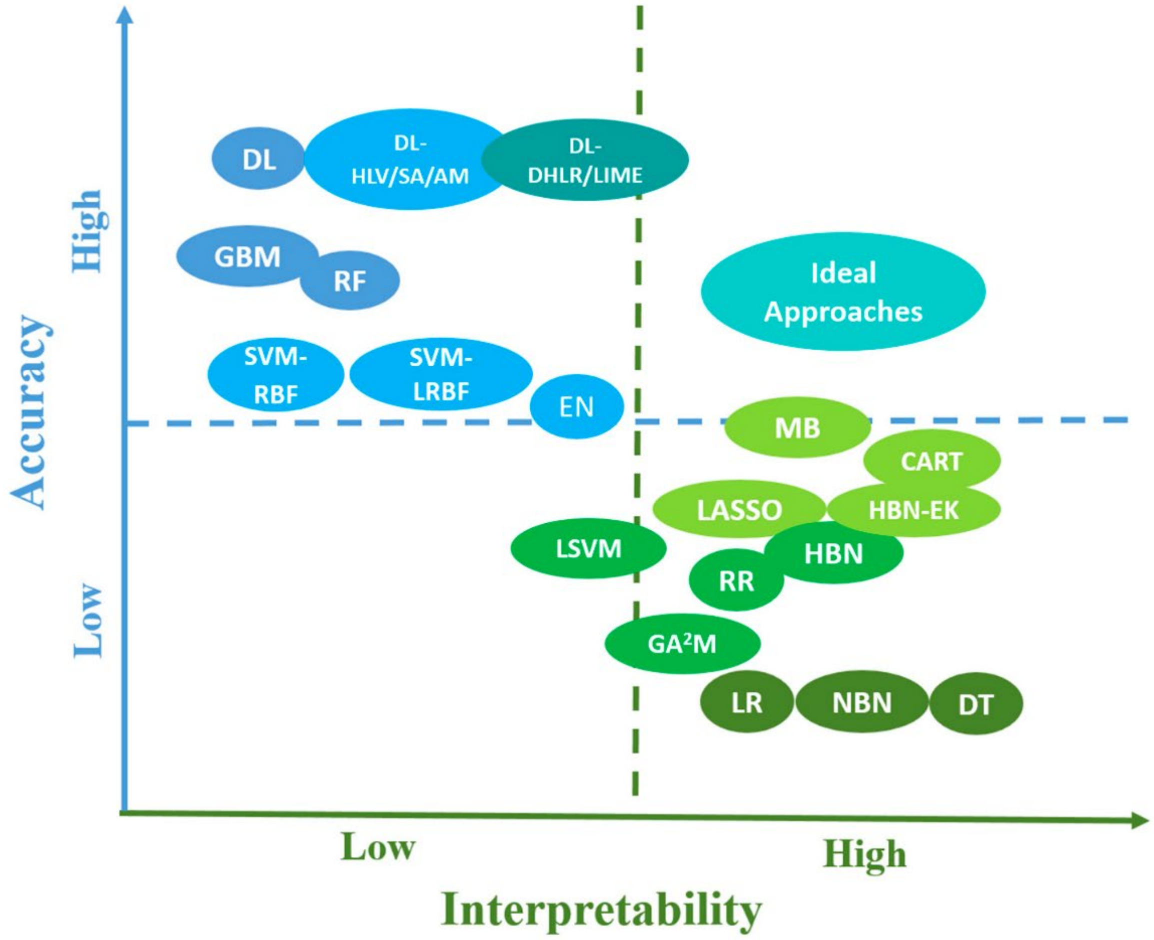
Adversarial noise



Neural network Classification

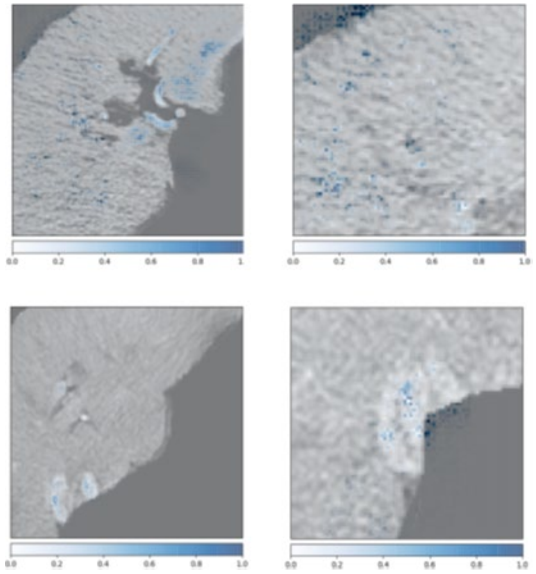
Keanu Reeves

ML/DL Interpretability



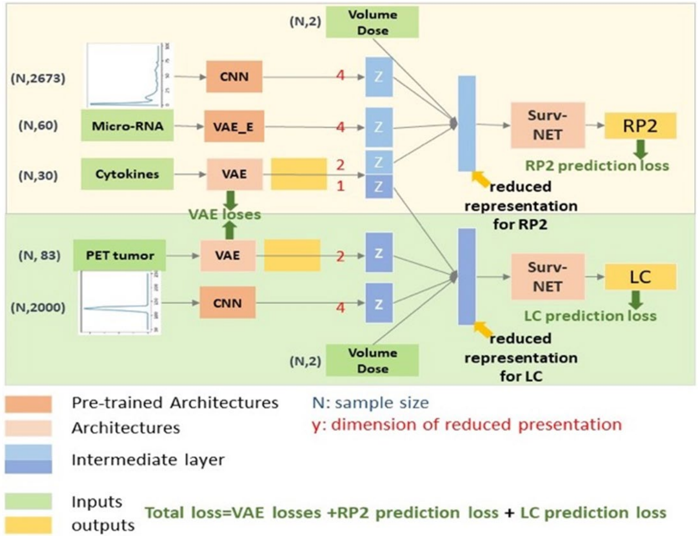
Luo, BJR-O, 2019

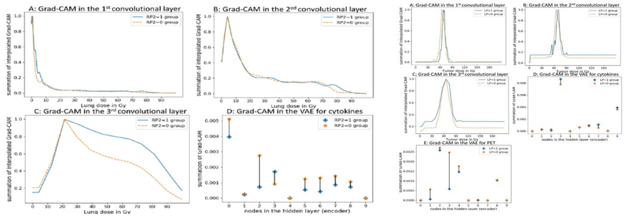
Radiomics Interpretability for Liver Cancer (Grad-CAM)



Wei et al, Physica Medica, 2021

Multi-omics interpretability for Lung Cancer





Cui et al, IJROBP, 2021

Some Popular ML/DL platforms

ML tools
in Java

WEKA
The University
of Waikato

Google
(py/C)

TensorFlow

Interface
(python)

Library
(python)

scikit
learn

Microsoft
Cognitive
Toolkit

theano

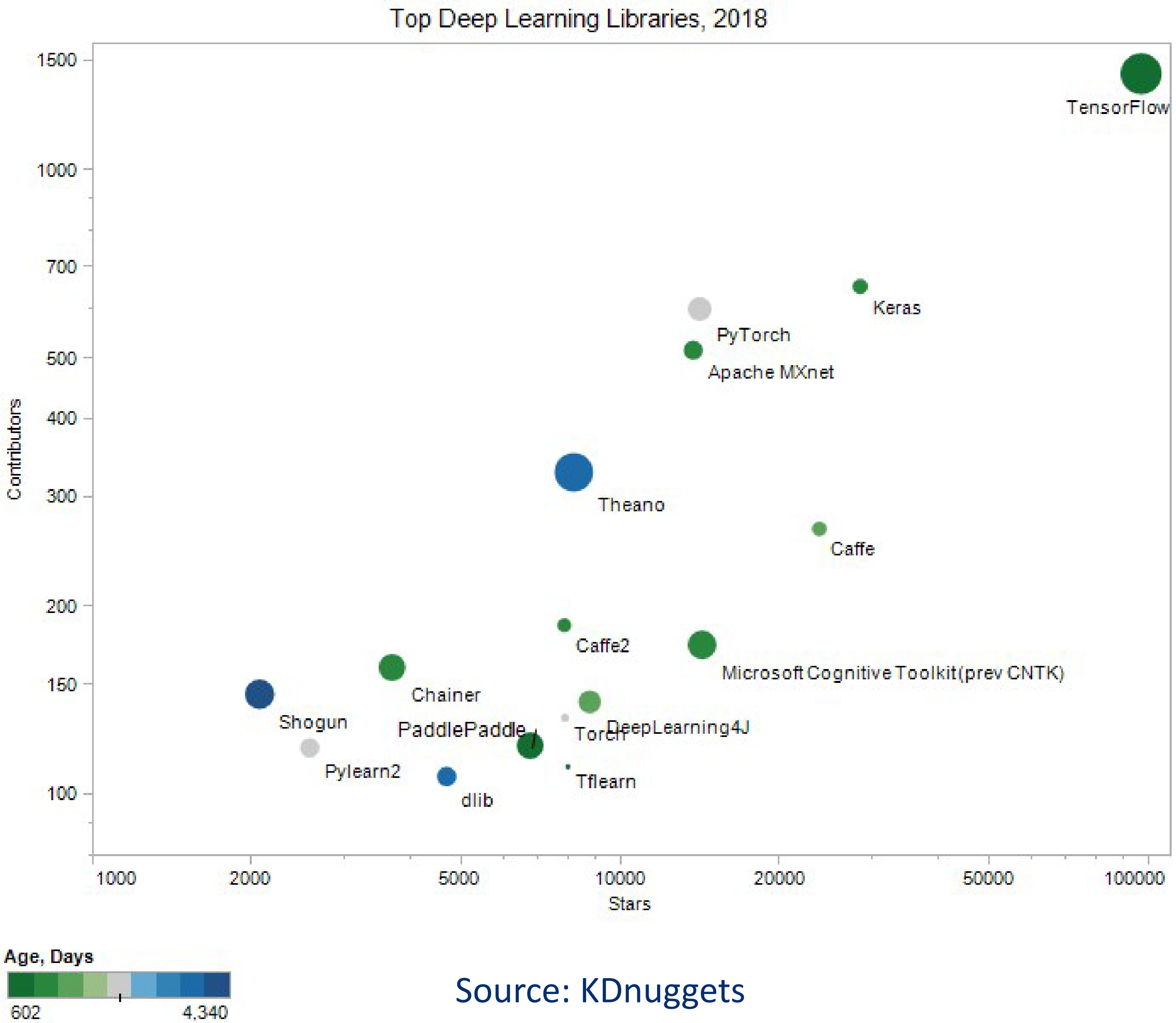
UdM
(python)

Caffe

Berkeley
(py/C)

Torch
(py/C)

PyTorch
Facebook



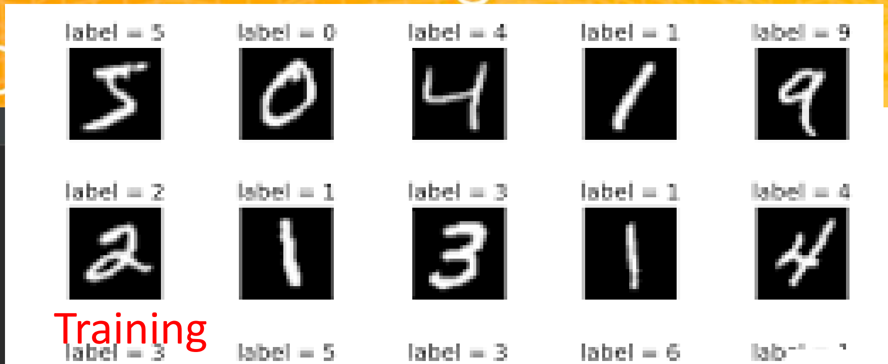
Source: KDnuggets

TensorFlow ™ Install Learn API Resources Community

An open source machine learning framework for everyo

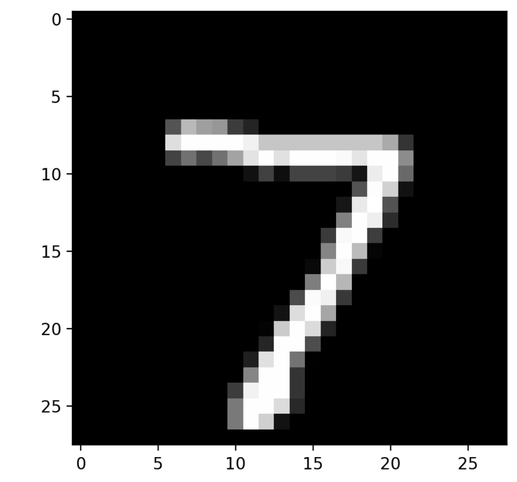
TF1.py

```python
import tensorflow as tf
import numpy as np
import matplotlib.pyplot as plt

mnist = tf.keras.datasets.mnist

(x_train, y_train),(x_test, y_test) = mnist.load_data()
x_train, x_test = x_train / 255.0, x_test / 255.0


#plt.imshow(x_train[0], cmap=plt.cm.binary)
#print(x_train[0])
#plt.show()
model = tf.keras.models.Sequential([
  tf.keras.layers.Flatten(),
  tf.keras.layers.Dense(512, activation=tf.nn.relu),
  tf.keras.layers.Dropout(0.2),
  tf.keras.layers.Dense(10, activation=tf.nn.softmax)
])
model.compile(optimizer='adam',
              loss="sparse_categorical_crossentropy",
              metrics=['accuracy'])

model.fit(x_train, y_train, epochs=5)
model.evaluate(x_test, y_test)
val_loss, val_acc= model.evaluate(x_test, y_test)
print(val_loss, val_acc)
prediction=model.predict([x_test])
print(prediction)
print(np.argmax(prediction[0]))
print(x_test[0])
plt.figure()
plt.imshow(x_test[0], cmap='gray')
plt.show()
```



Training

```
60000/60000 [==============================   0.1997 - acc: 0.9412
Epoch 2/5
60000/60000 [==============================   0.0800 - acc: 0.9754
Epoch 3/5
60000/60000 [==============================   0.0513 - acc: 0.9833
Epoch 4/5
60000/60000 [==============================   0.0361 - acc: 0.9887
Epoch 5/5
60000/60000 [==============================   0.0263 - acc: 0.9914
```



Testing

```
10000/10000 [==============================
10000/10000 [==============================] - 0s 39us/step
0.06758721045646235 0.9795
```

Take home Messages

- There are different classes of ML/DL algorithms with varying accuracy and interpretability levels and the choice of the proper algorithm(s) is a problem and data dependent
- Once an ML/DL algorithm is identified, a proper plan for training (sample size), evaluation (statistics), and validation (testing) should be developed to assure generalizability (out-of-sample)
- Acceptance and commissioning of ML/DL for medicine is in its infancy, however, efforts are being made by public and private sectors towards safe ML/DL implementation

References

Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again, Basic Books, 2019

Machine Learning in Radiation Oncology: Theory and Applications. El Naqa, Li, Murphy (editors), Springer, 2015.

Emerging Developments and Practices in Oncology: El Naqa (Editor), IGI Global, Hershey, PA, 2018.

A Guide to Outcome Modeling In Radiotherapy and Oncology: Listening to the Data, El Naqa (Editor): CRC press: Taylor and Francis Group, Boca Raton, FL, USA, 2018.

Machine Learning with Radiation Oncology Big Data (eBook): Deng, El Naqa, Xing (Editors), Frontier in Oncology, Lausanne, Switzerland, 2019.

Machine and Deep Learning in Oncology, Medical Physics and Radiology: El Naqa, Murphy, Springer Nature, 2022.

Useful ML/DL websites:

Tensorflow. CNTK.
https://www.kaggle.com/. Scikit
Coursera: machine-learning

COVER STORY

GUEST EDITORIAL

Moffitt Cancer Center: Why we are building the first machine learning department in oncology

By Issam El Naqa and Dana Rollison

THANK YOU!