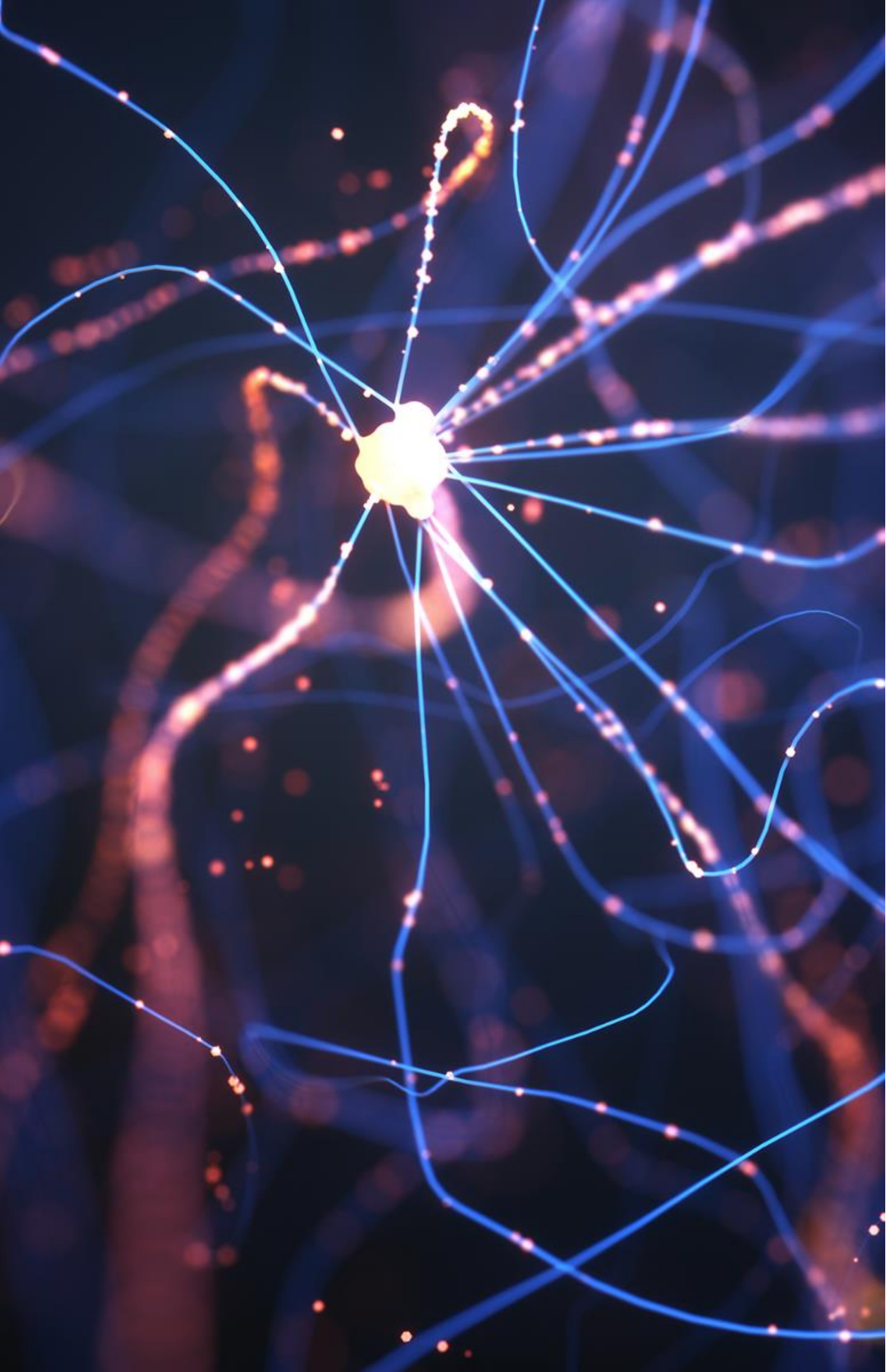


An Overview on the role of Artificial Intelligence (AI) and Deep Neural Networks (DNN) in cancers prediction, diagnosis, grading and prognosis

Hamidreza Bolhasani

Ph.D.| Data Scientist | Founder of DataBioX

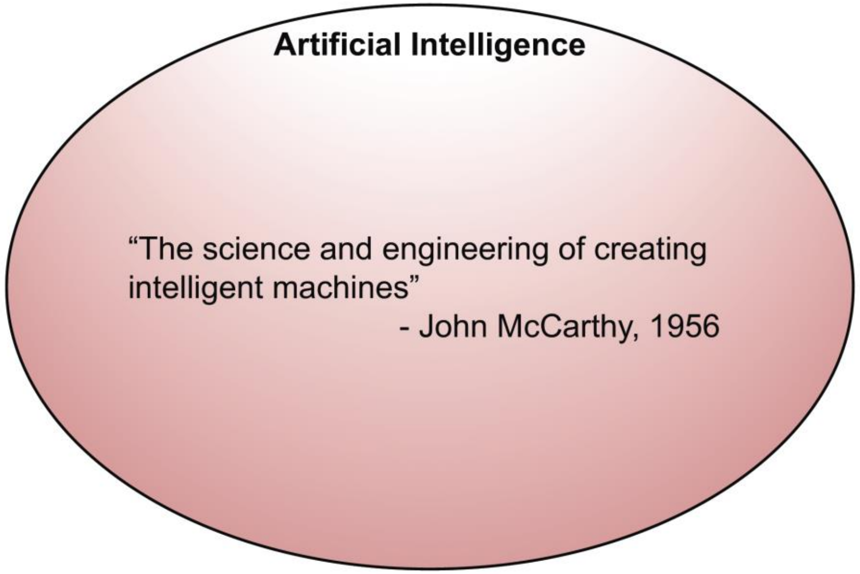
Springer Nature / Elsevier Scientific Reviewer.

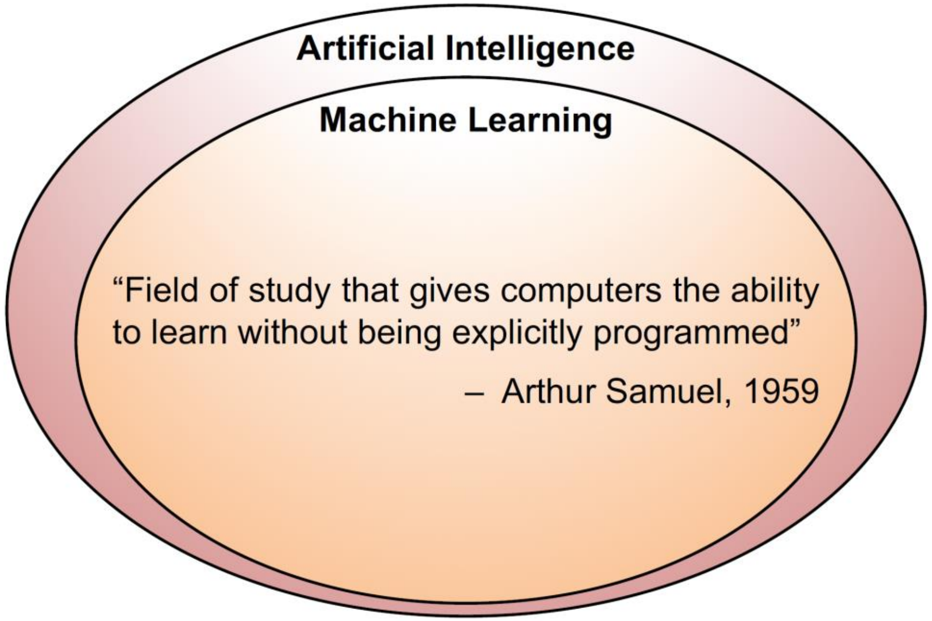
hamidreza@Bolhasani.net

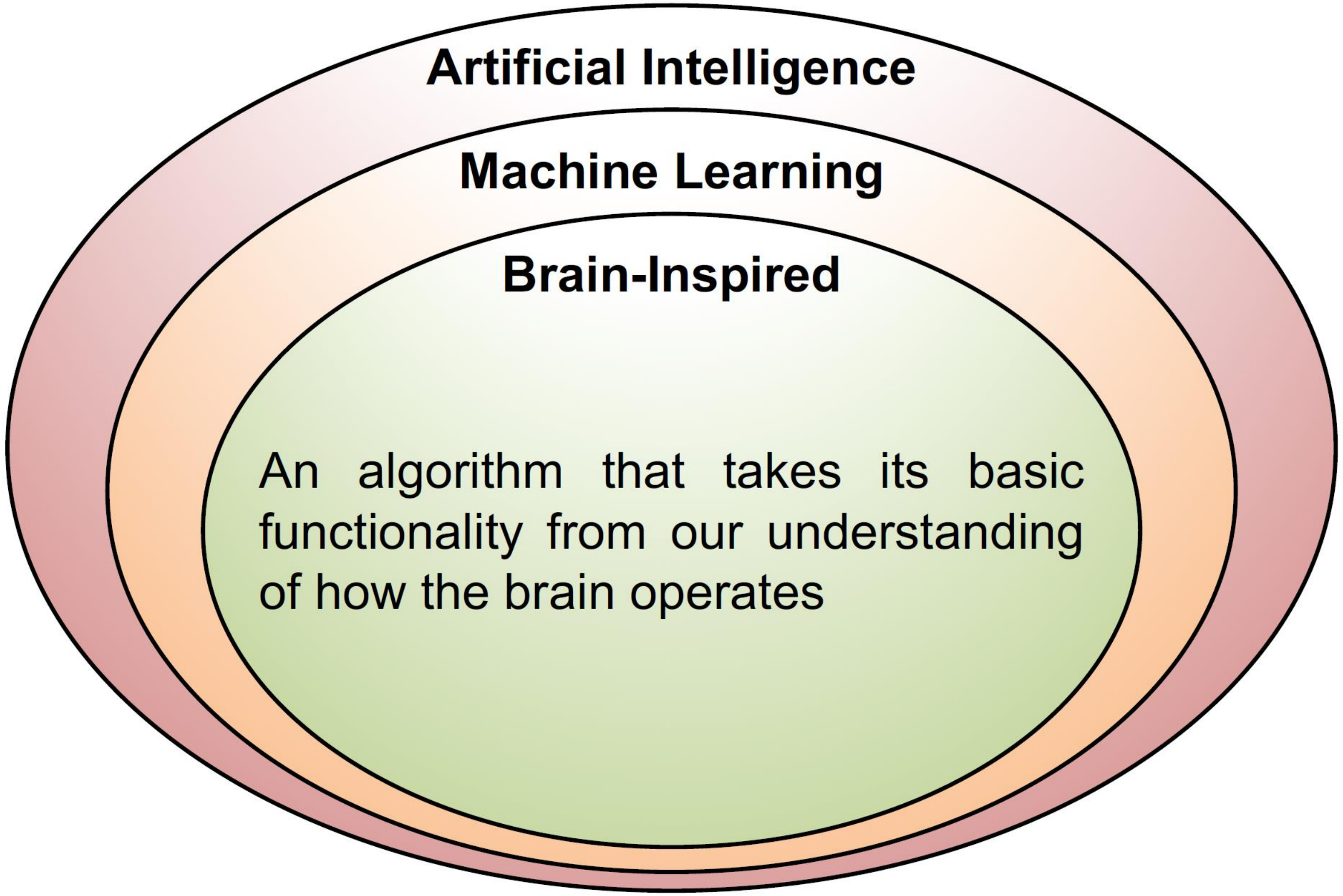
Nov 2021.

DataBioX
we give life to the data!...

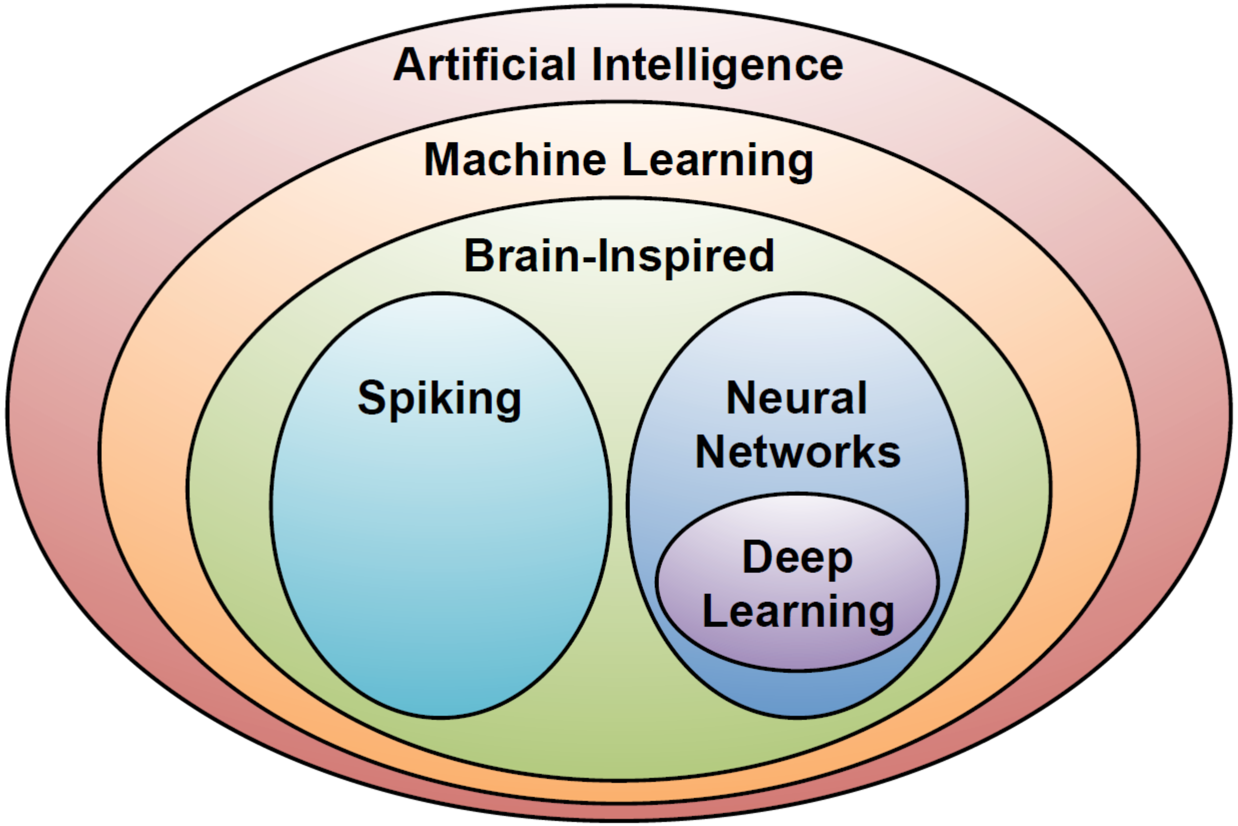
Artificial Intelligence (AI)

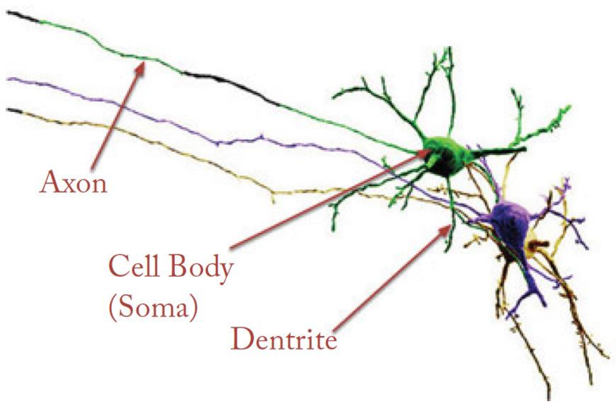


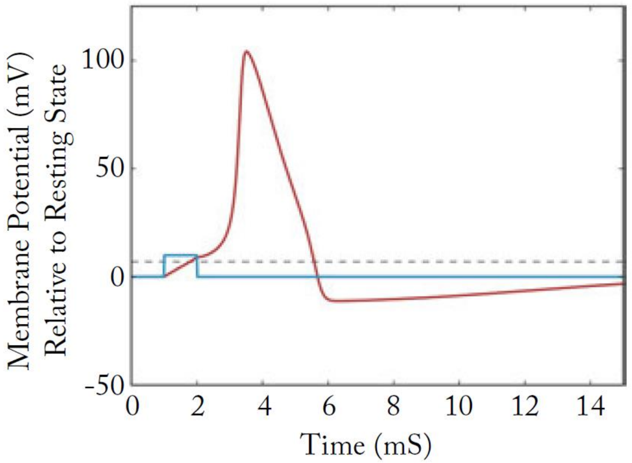


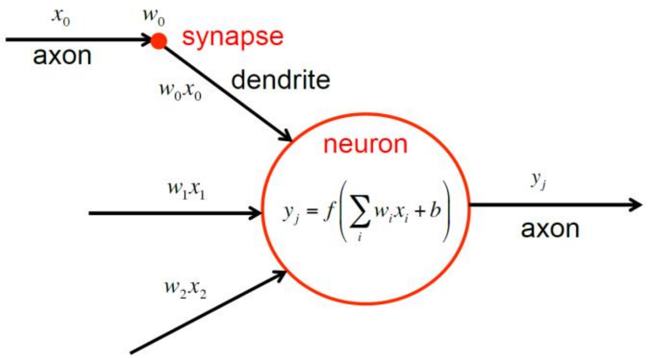


Deep Neural Networks (DNN)



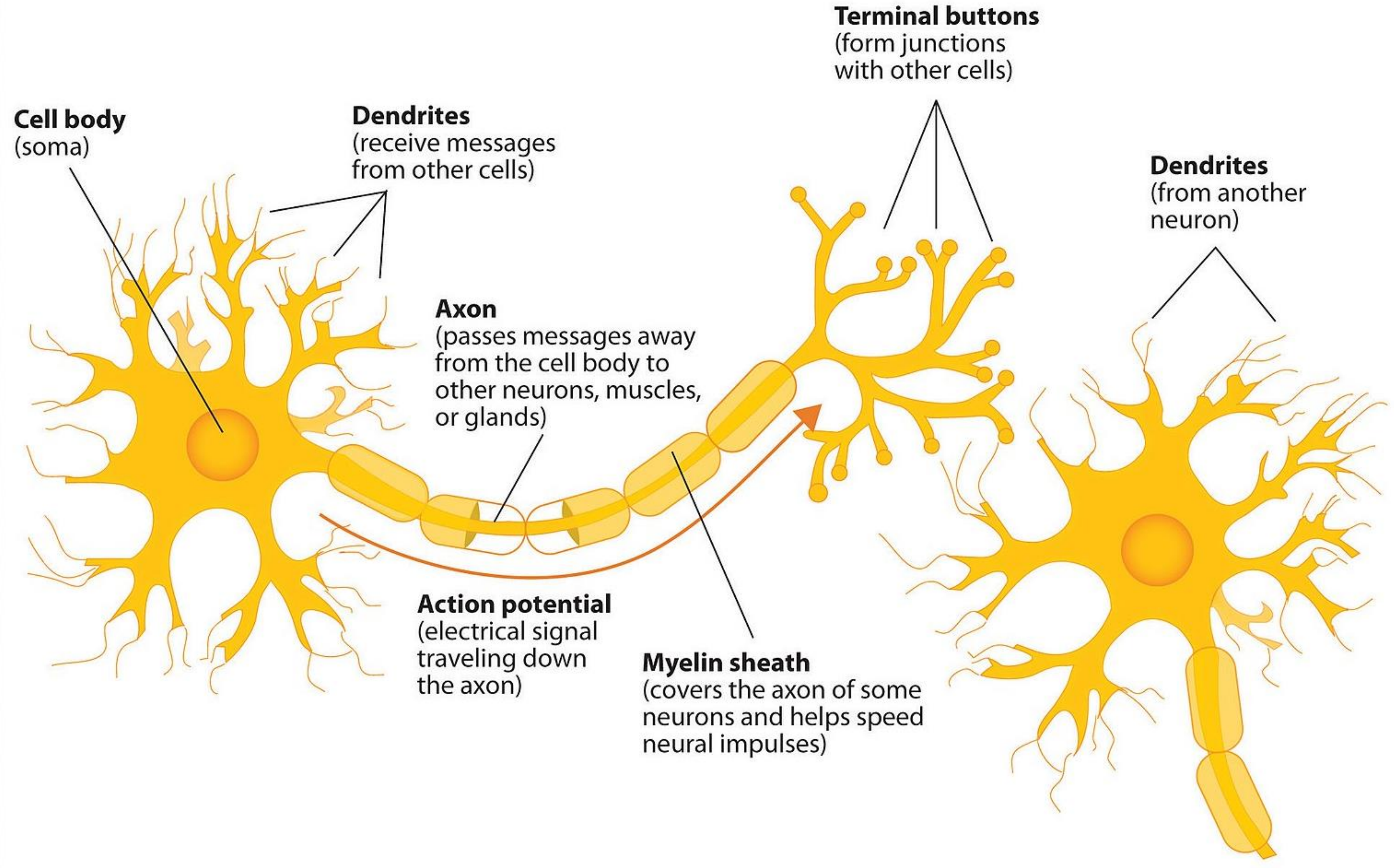






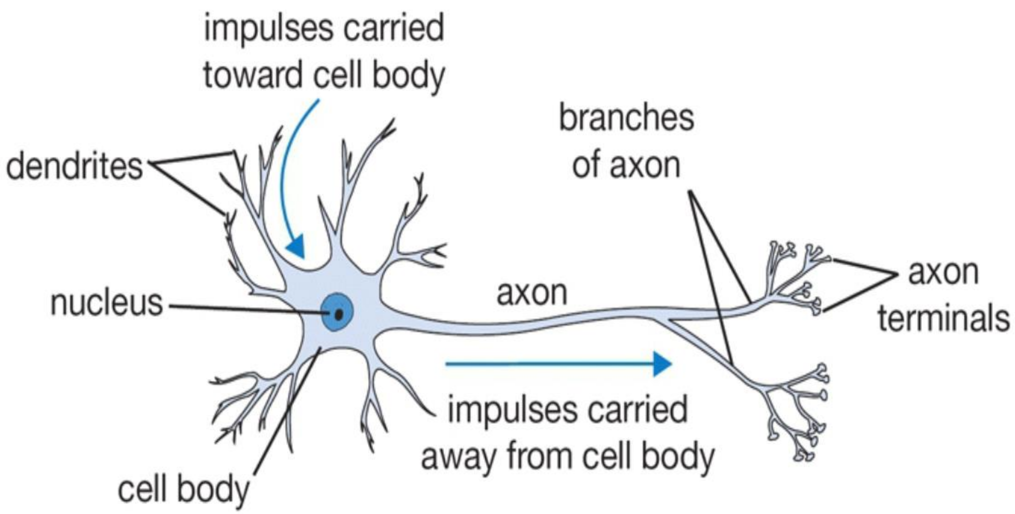
Ref. B. Reagen, R. Adolf, P. N. Whatmough, G. Wei, and D. M. Brooks, Deep Learning for Computer Architects, ser. Synthesis Lectures on Computer Architecture. Morgan & Claypool Publishers, 2017.

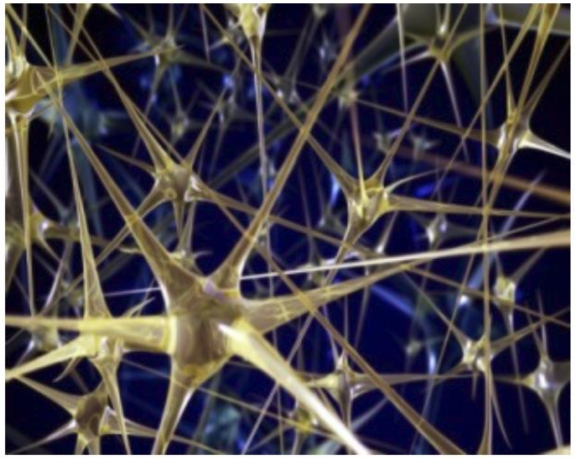
Biological Neuron



Ref. Jennifer Walinga, 2014.

Biological Neuron

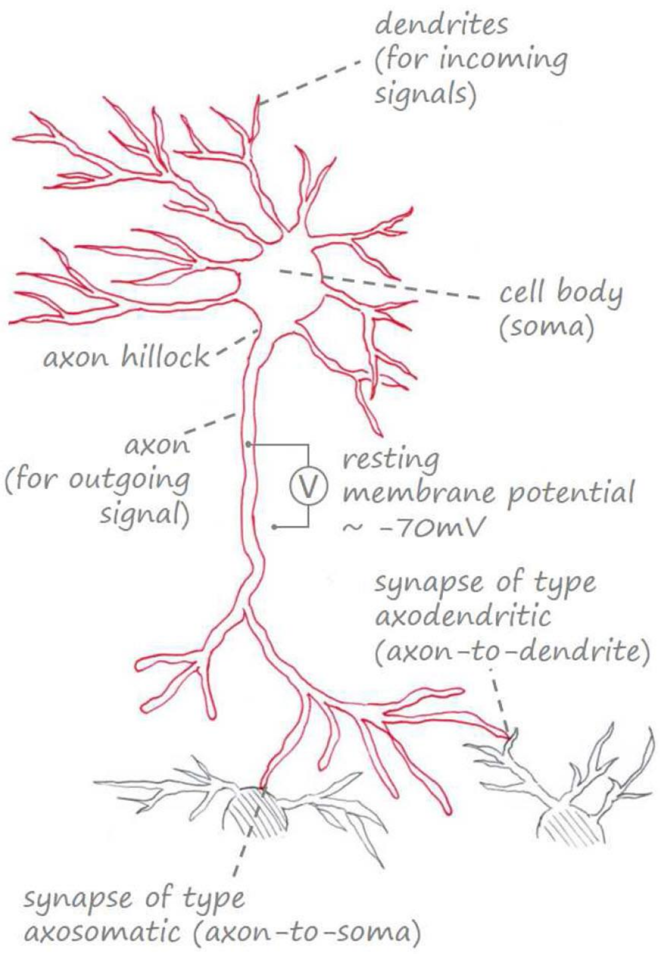


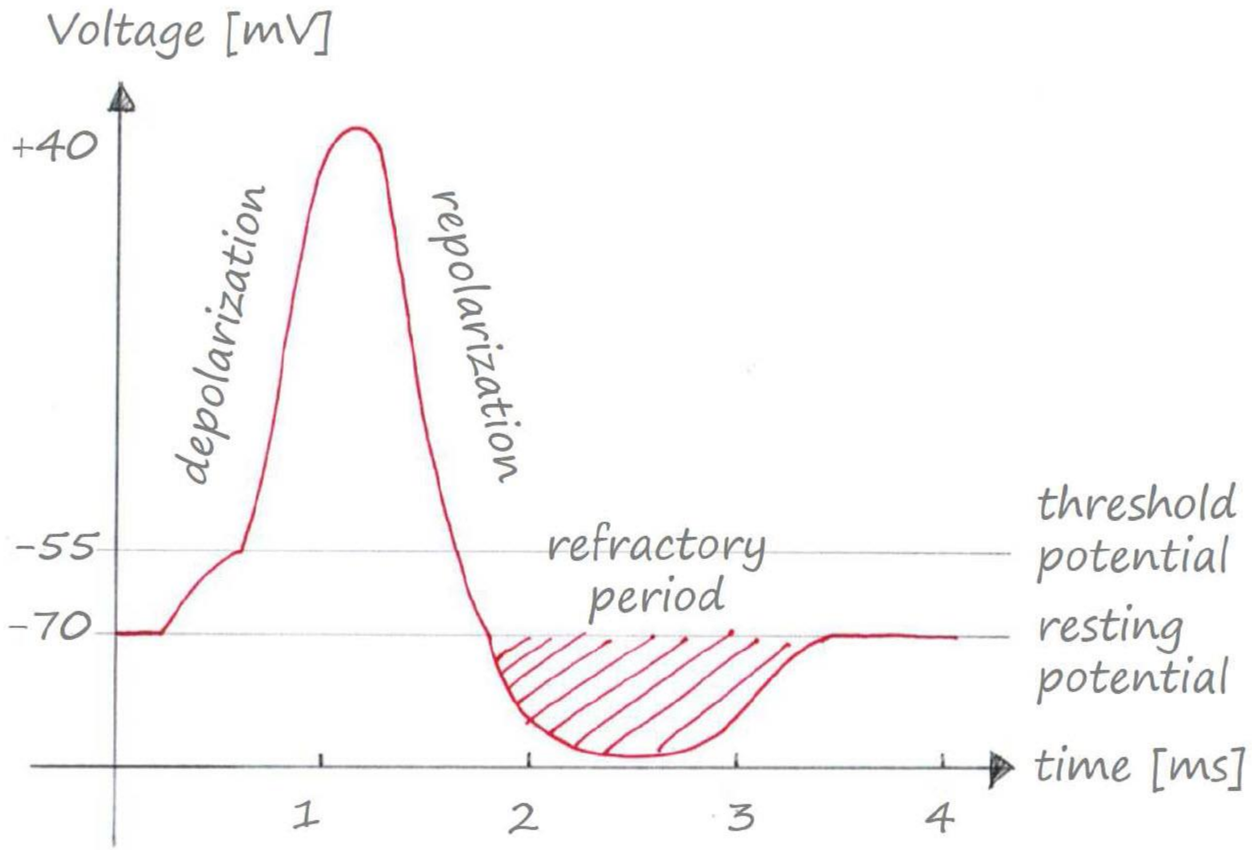


- The basic computational unit of the brain is a **neuron**
  → 86B neurons in the brain
- Neurons are connected with nearly **10^{14} – 10^{15} synapses**
- Neurons receive input signal from **dendrites** and produce output signal along **axon**, which interact with the dendrites of other neurons via **synaptic weights**
- Synaptic weights – learnable & control influence strength

Image Source: Stanford

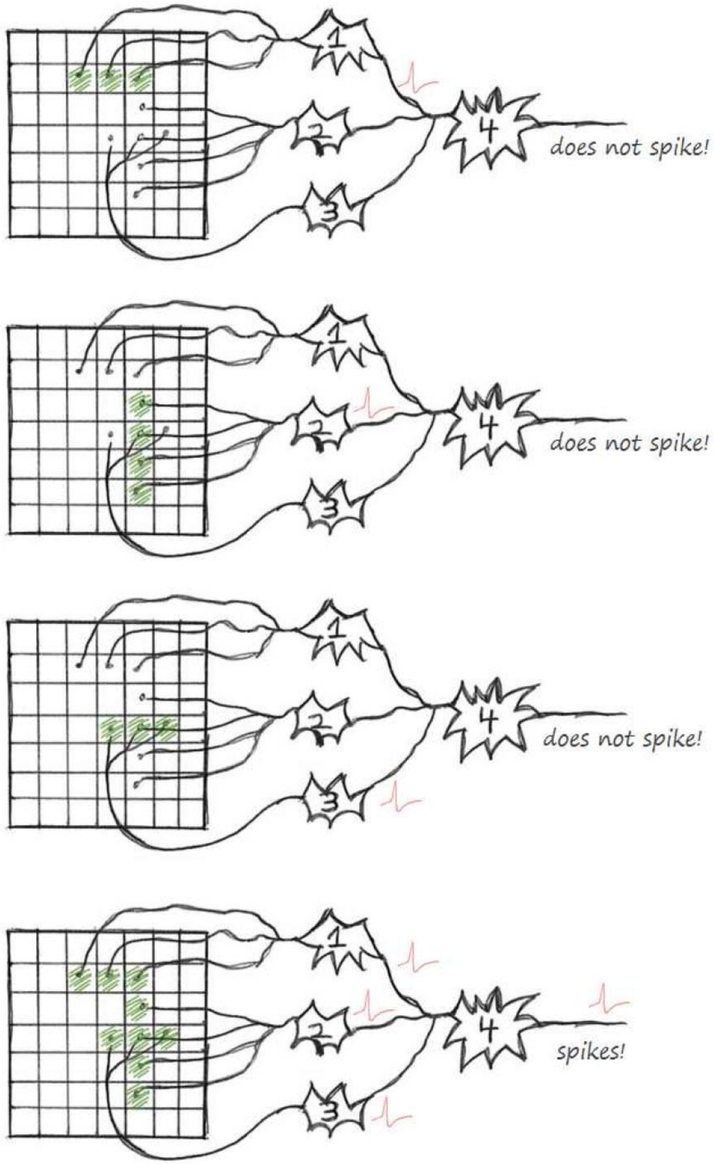
Biological Neuron

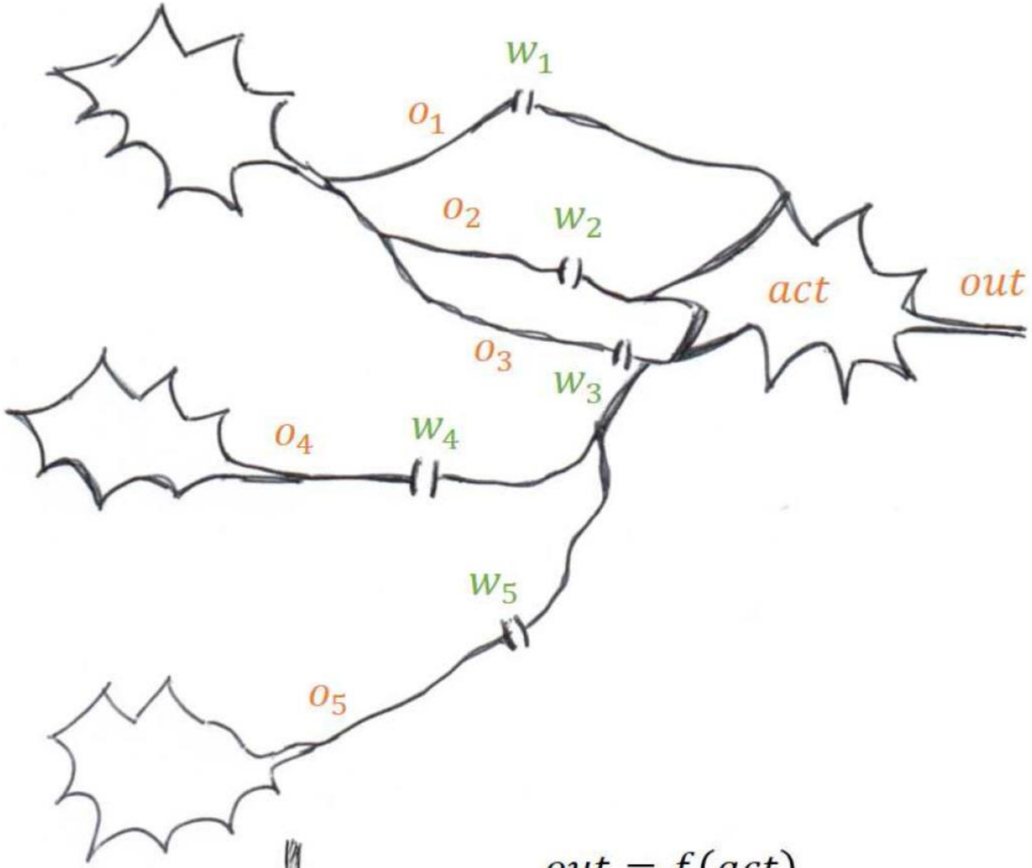




Ref. Prof. Dr. Jurgen Brauer, "Introduction to Deep Learning," 2018.

Artificial Neural Networks (ANN)

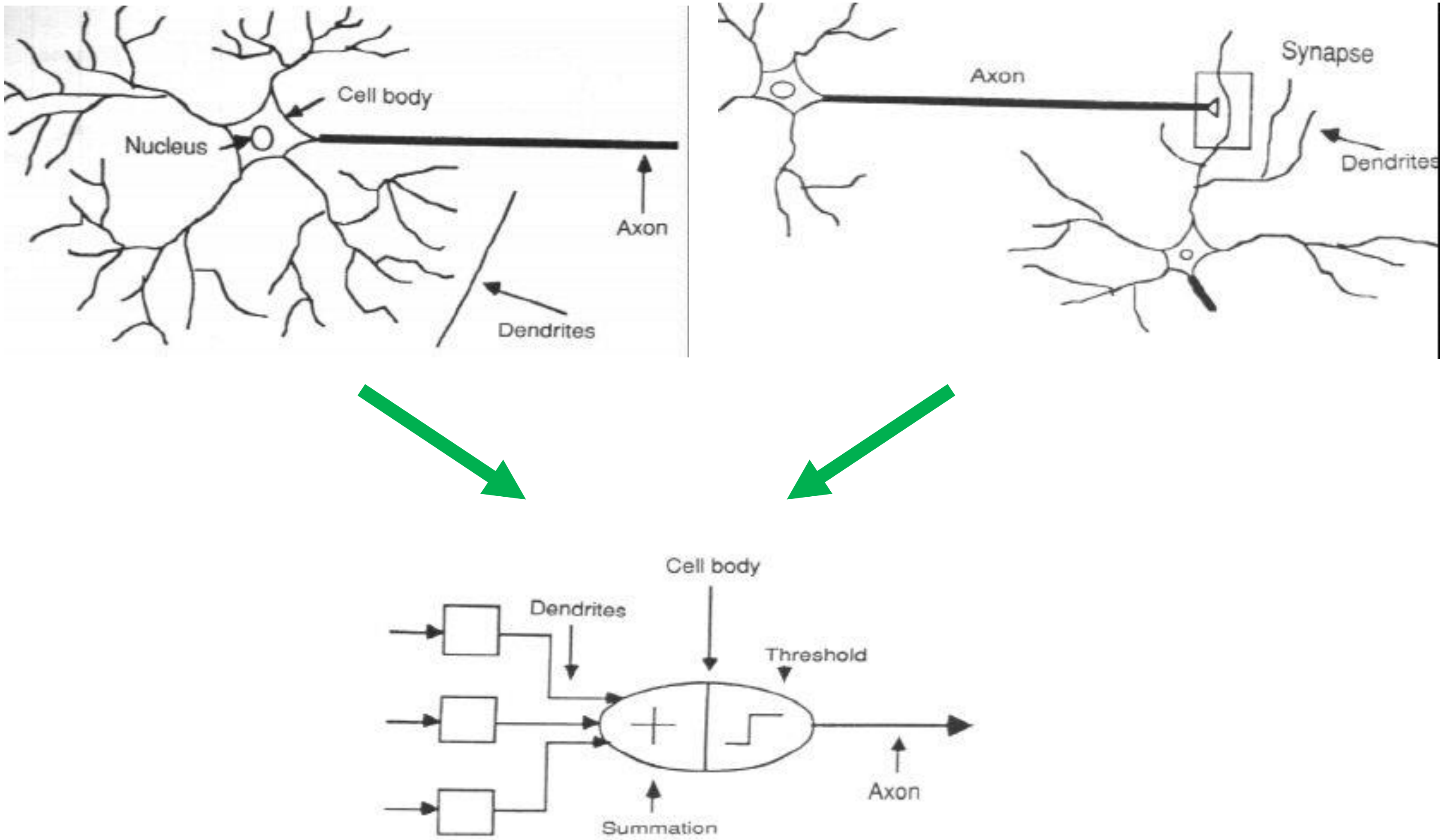




$$
\begin{aligned}
out &= f(act) \\
&= f(o_1 w_1 + o_2 w_2 + o_3 w_3 + o_4 w_4 + o_5 w_5) \\
&= f\left(\sum_{i=1}^{N} o_i w_i\right)
\end{aligned}
$$

Ref. Prof. Dr. Jurgen Brauer, "Introduction to Deep Learning," 2018.

Artificial Neural Networks (ANN)



Artificial Neural Networks (ANN)

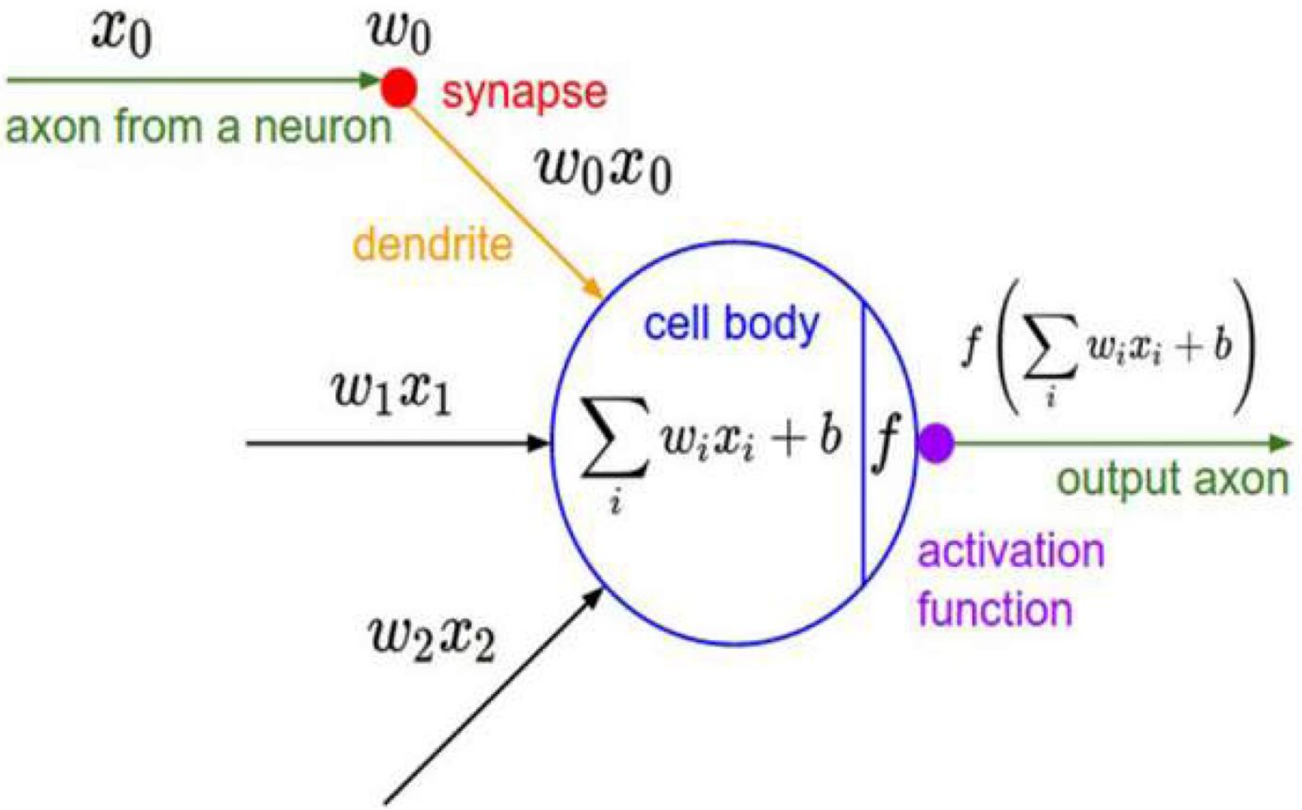
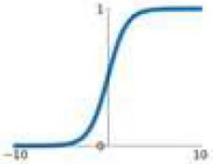


Image Source: Stanford

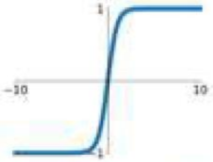
Sigmoid

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



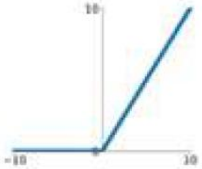
tanh

$\tanh(x)$



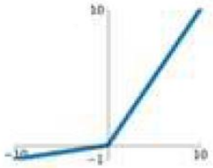
ReLU

$\max(0, x)$



Leaky ReLU

$\max(0.1x, x)$

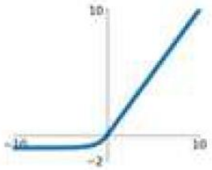


Maxout

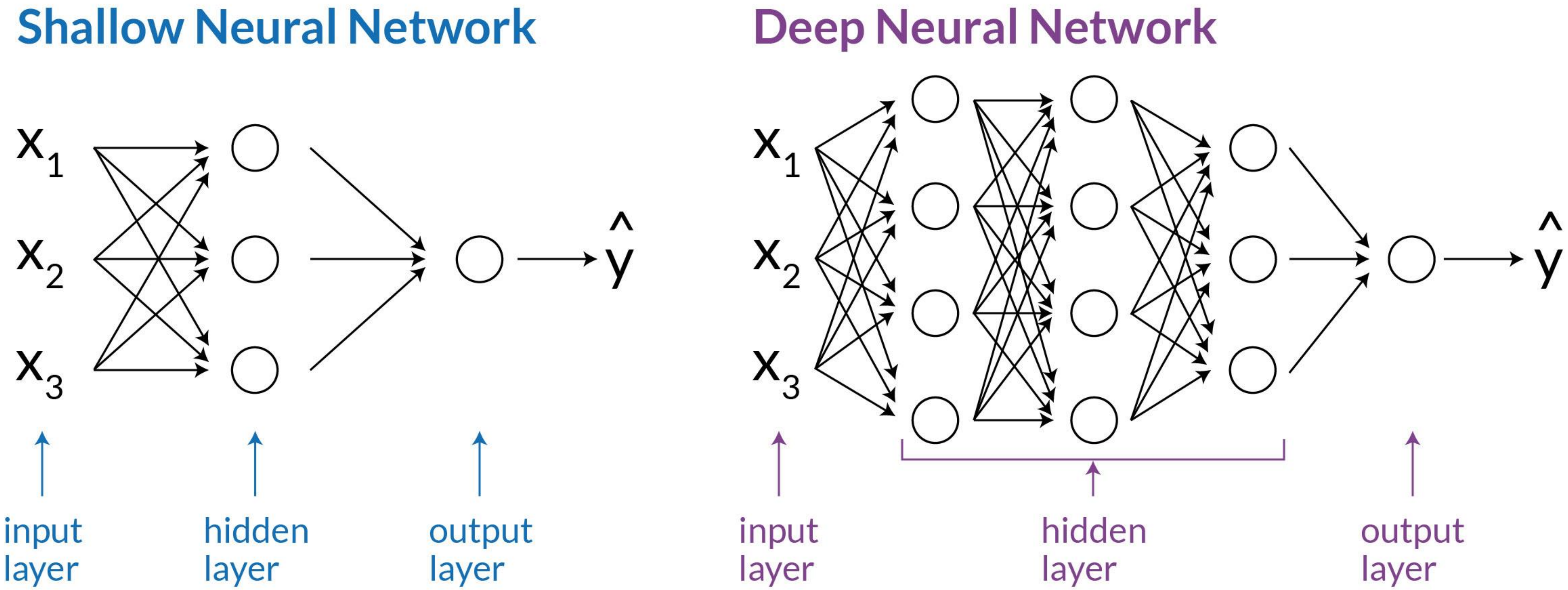
$\max(w_1^T x + b_1, w_2^T x + b_2)$

ELU

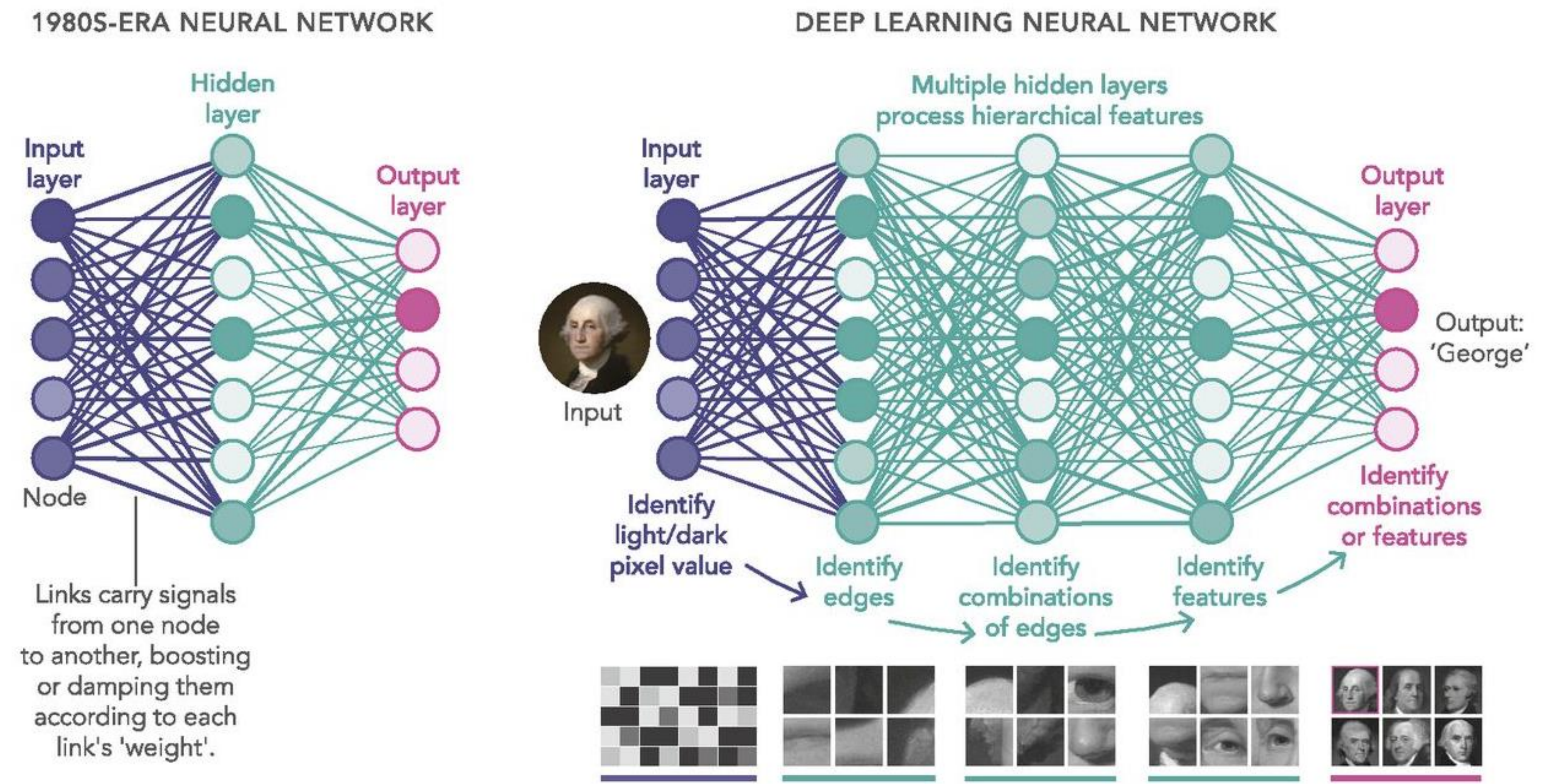
$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$



Deep Neural Networks



Deep Neural Networks



What is Deep Learning?

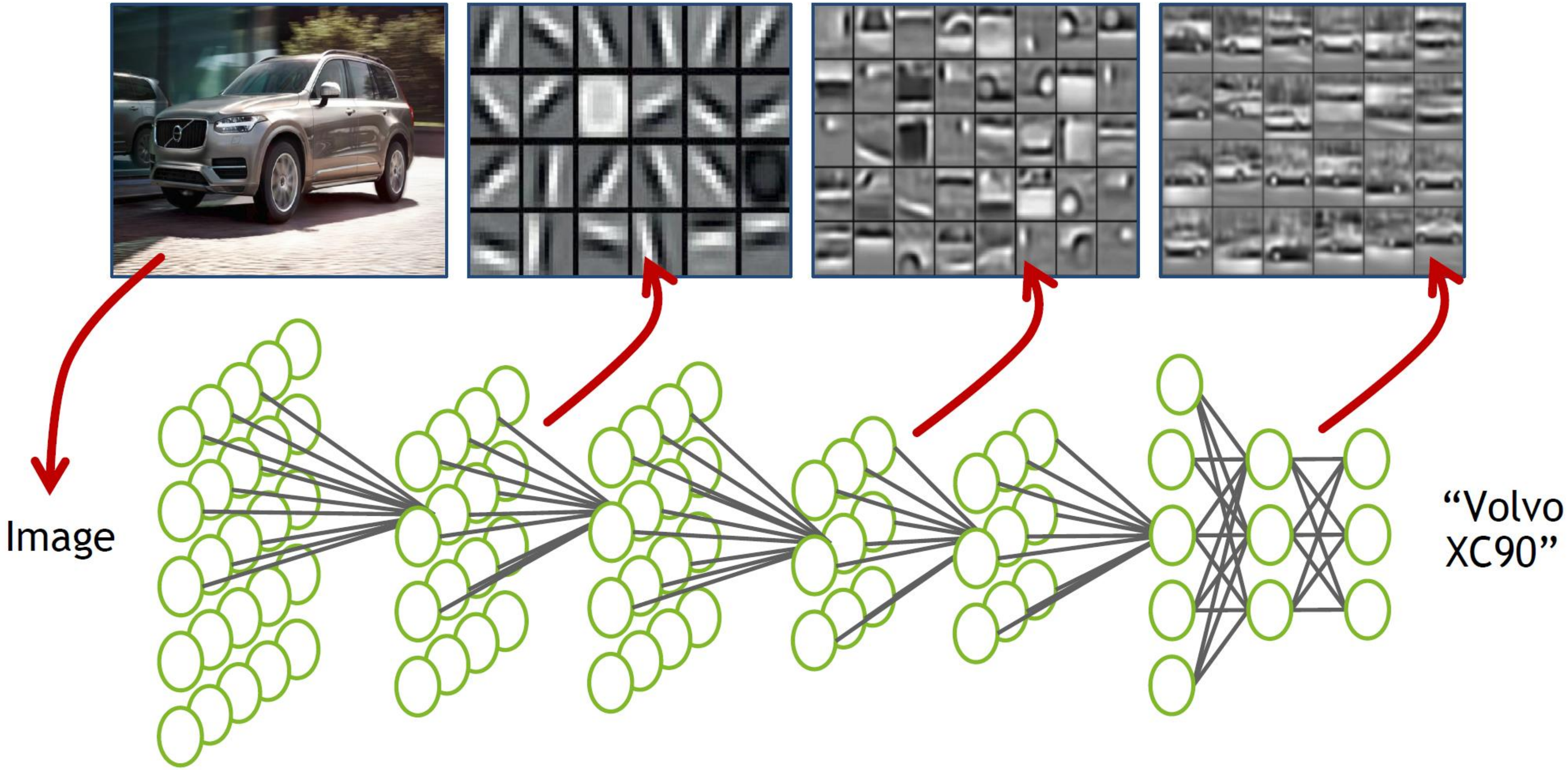


Image Source: [Lee et al., Comm. ACM 2011]

Deep Learning: Anywhere!

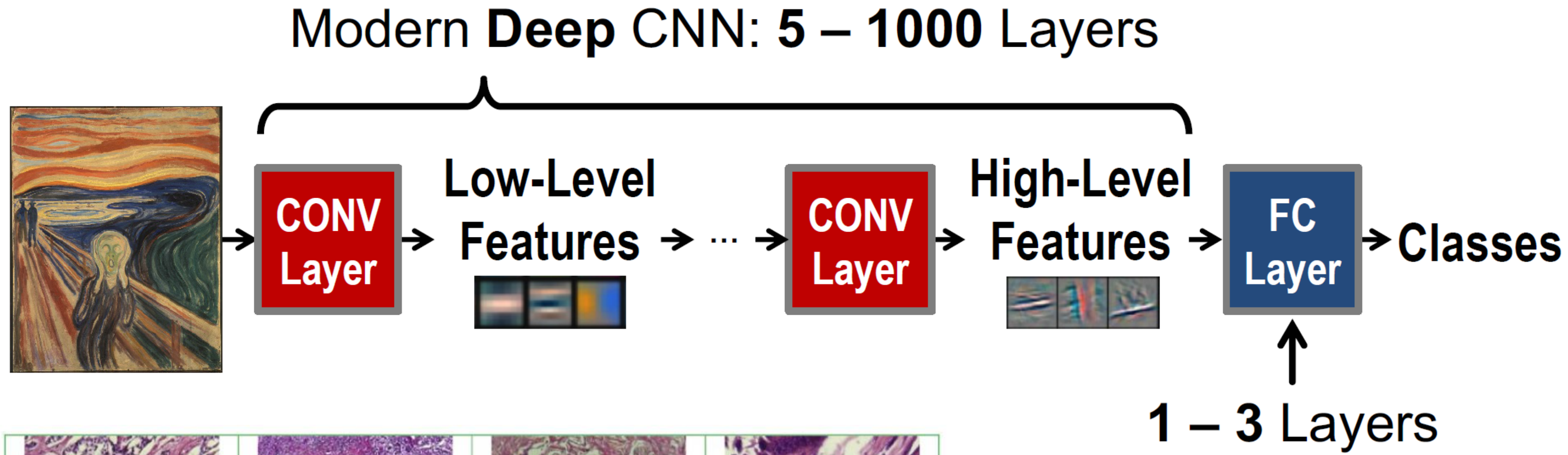
Google

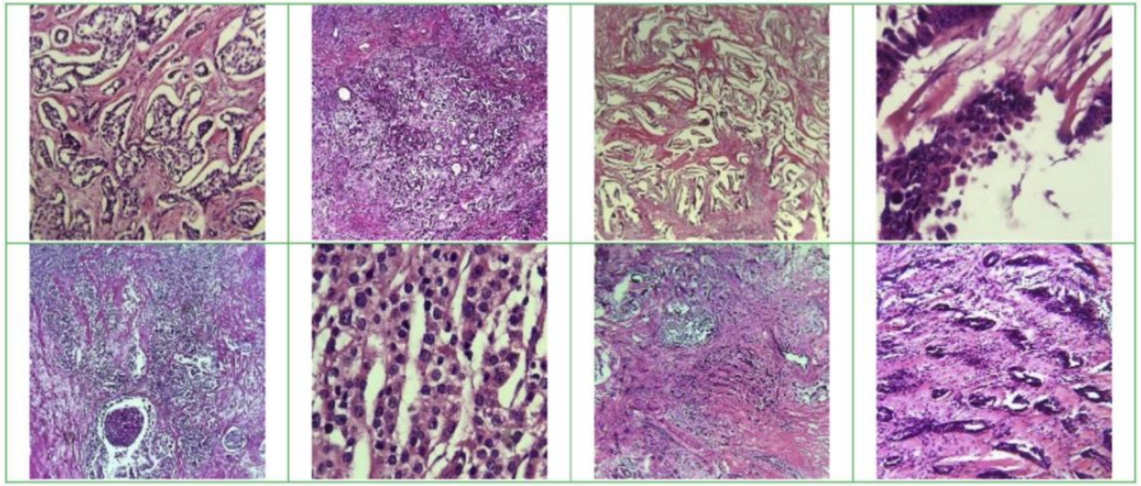
Translate

Popular Types of DNNs

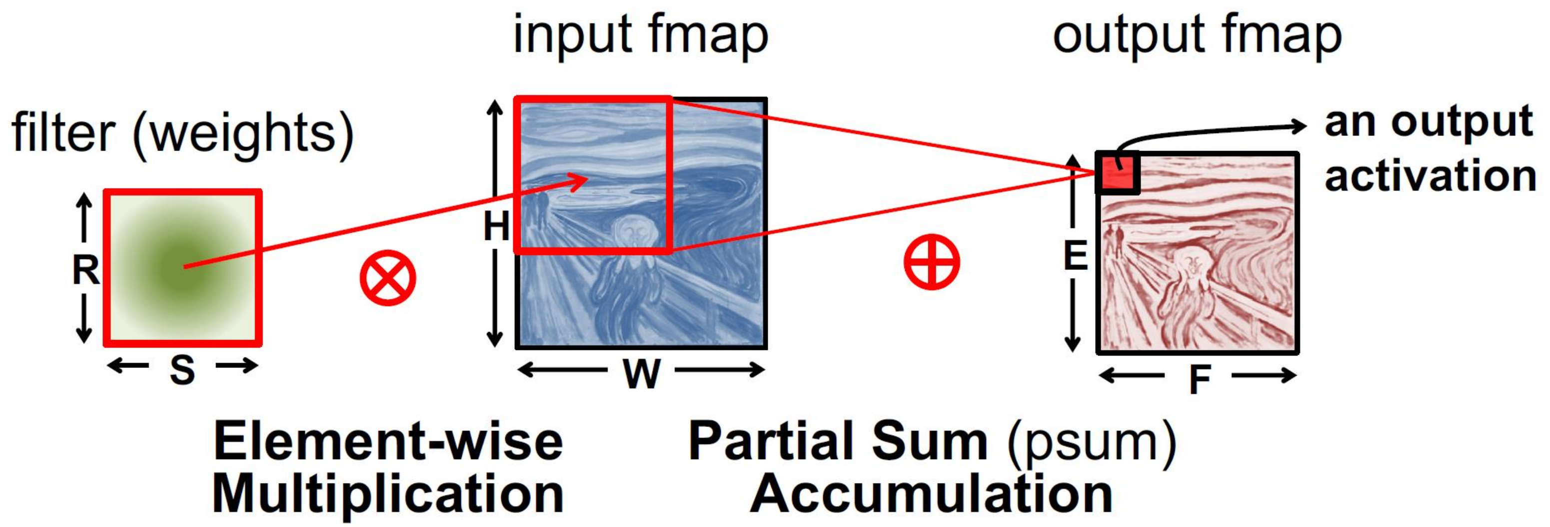
- **Fully-Connected NN**
  - feed forward, a.k.a. multilayer perceptron (MLP)
- **Convolutional NN (CNN)**
  - feed forward, sparsely-connected w/ weight sharing
- **Recurrent NN (RNN)**
  - feedback
- **Long Short-Term Memory (LSTM)**
  - feedback + storage

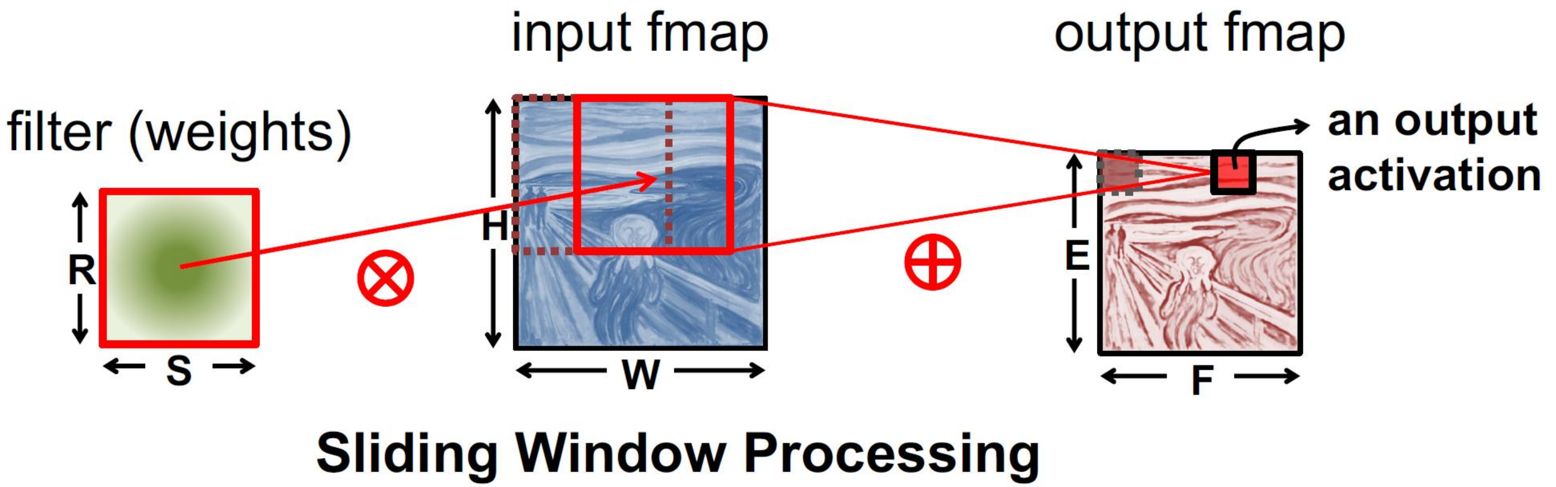
Deep Convolutional Neural Networks





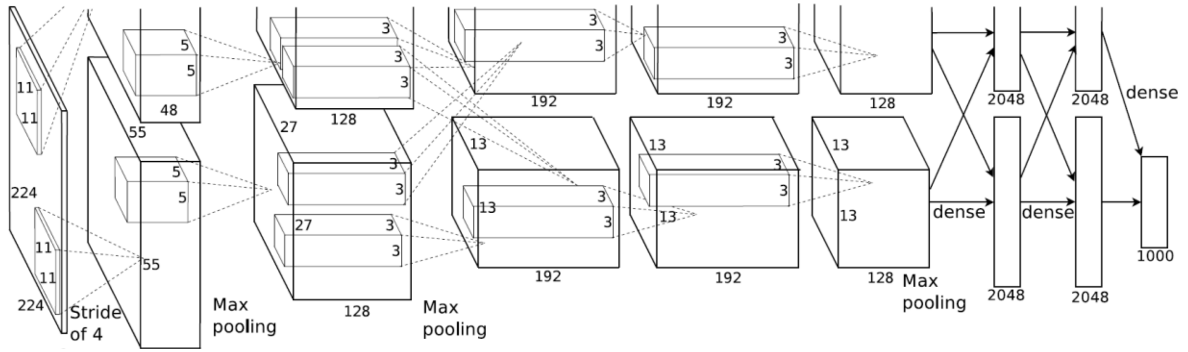
Convolutional Neural Networks (CNNs)



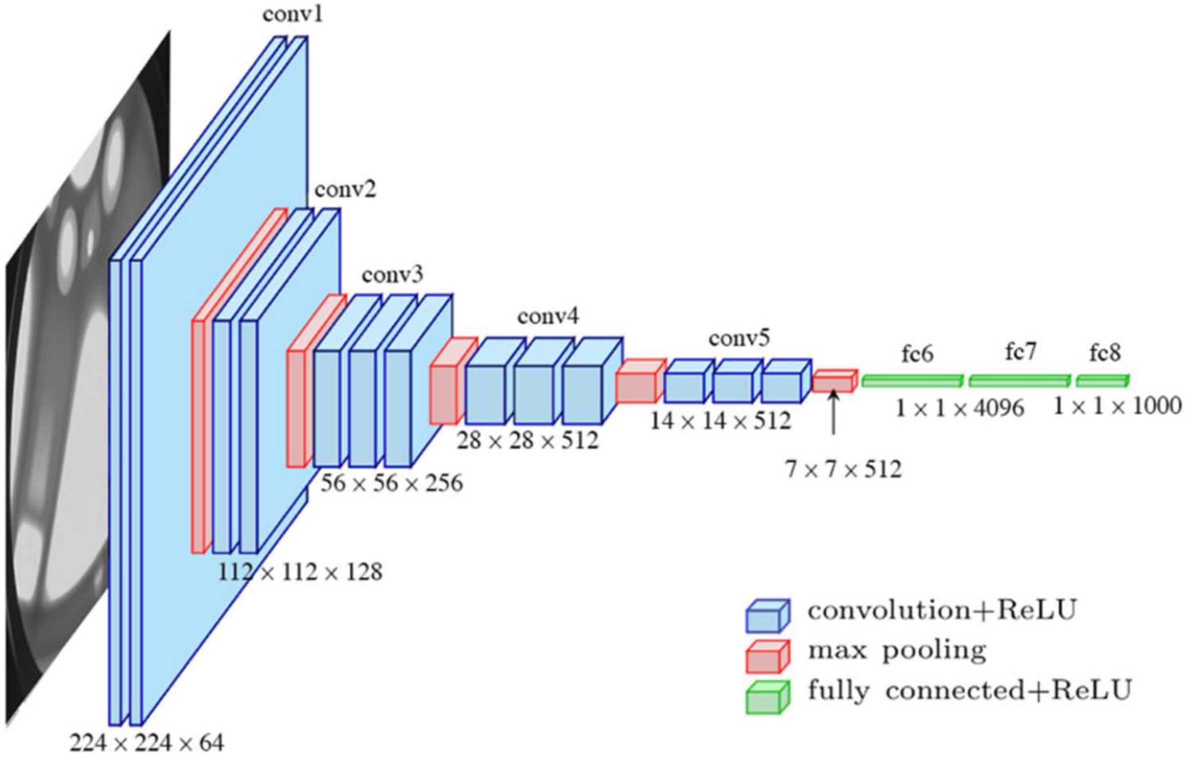


Ref. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in International Conference on Learning Representations (ICLR), 2015.

Convolutional Neural Networks (CNNs)



AlexNet

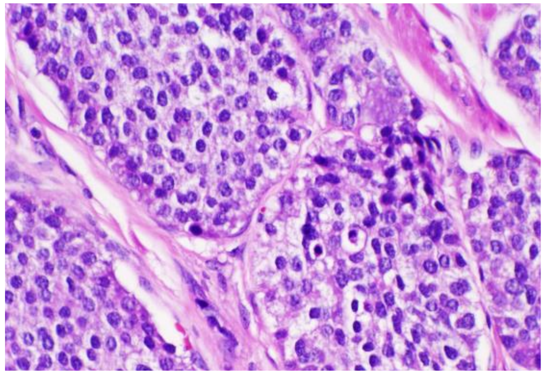


VGG16

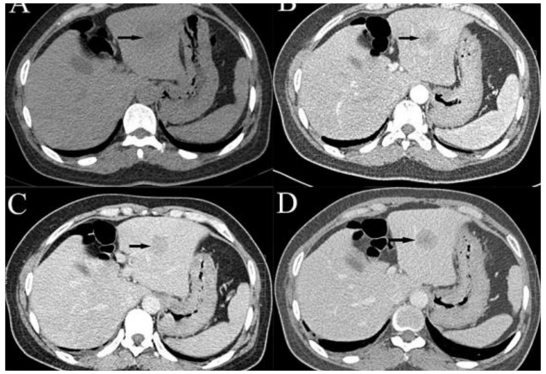
Ref. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in International Conference on Learning Representations (ICLR), 2015.

What does a Deep Neural Network needs to be Trained?

BIG DATA

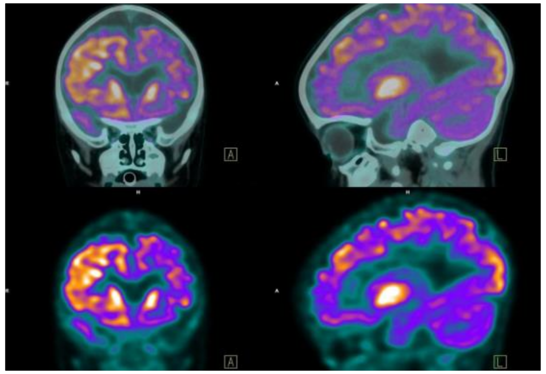


Histopathologic Images



CT-Scan

MRI



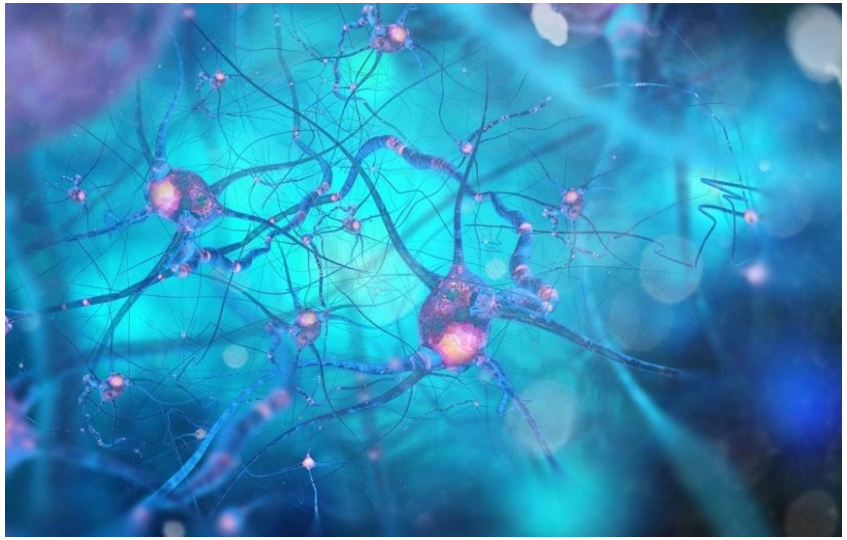
PET-Scan

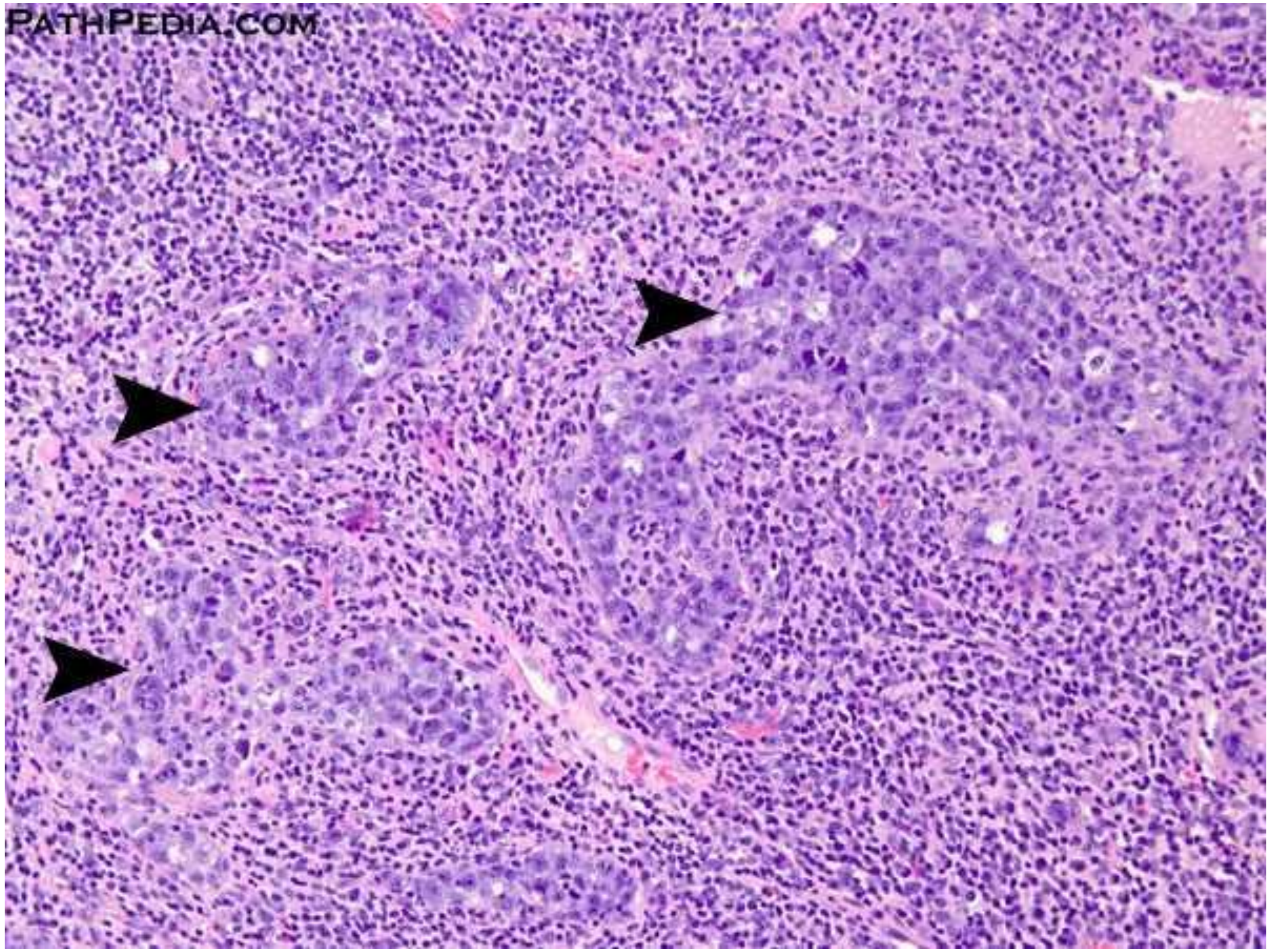
How much data?
So much that we get our desired results!

Our First Contribution

DataBioX founded in 2020 for sharing biomedical datasets.

DataBioX first dataset: Invasive Ductal Carcinoma Grading Dataset





DataBiox ..::: Data Science Solutions :::..

HOME DATASETS

Hamidreza Bolhasani

PhD Candidate | Data Scientist

BSc, Physics. University of Isfahan.

MSc, Computer Engineering, Science and Research Branch IAU.

PhD, Computer Engineering, Science and Research Branch IAU.

Dr. Somayyeh Jafarali Jassbi, PhD.

Assistant Professor, Science and Research Branch IAU

Computer Architecture, Computer Arithmetic

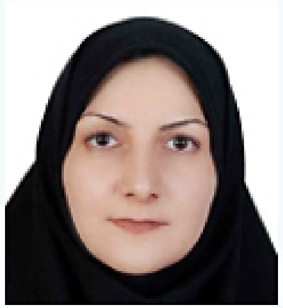
Internet of Things (IoT)

RFID Networks

Dr. Elham Amjadi, MD.

Anatomical Clinical Pathologist

Poursina Hakim Digestive Diseases Research Center, Isfahan University of Medical Sciences.



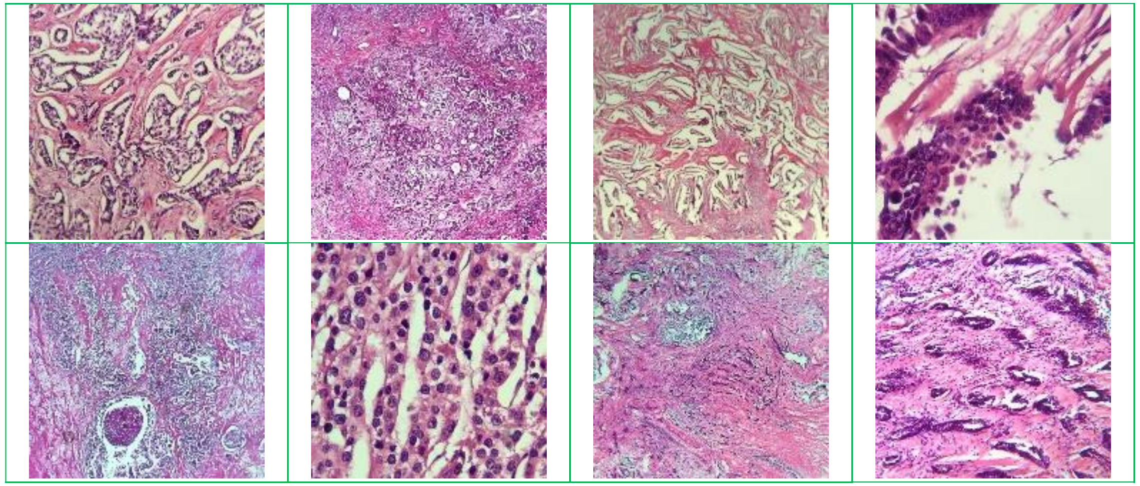
Dr. Maryam Tabatabeian, MD.

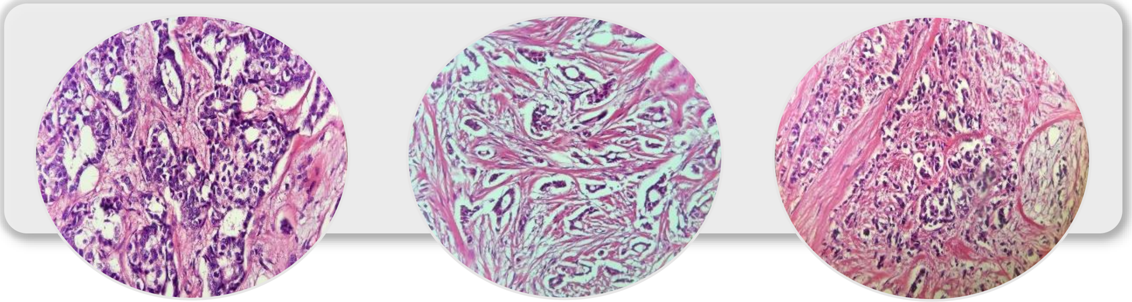
Founder, Anahid Breast Diseases Clinic.

General Surgeon.

Invasive Ductal Carcinoma Grading Dataset

DataBioX: Training dataset for Supervised Learning

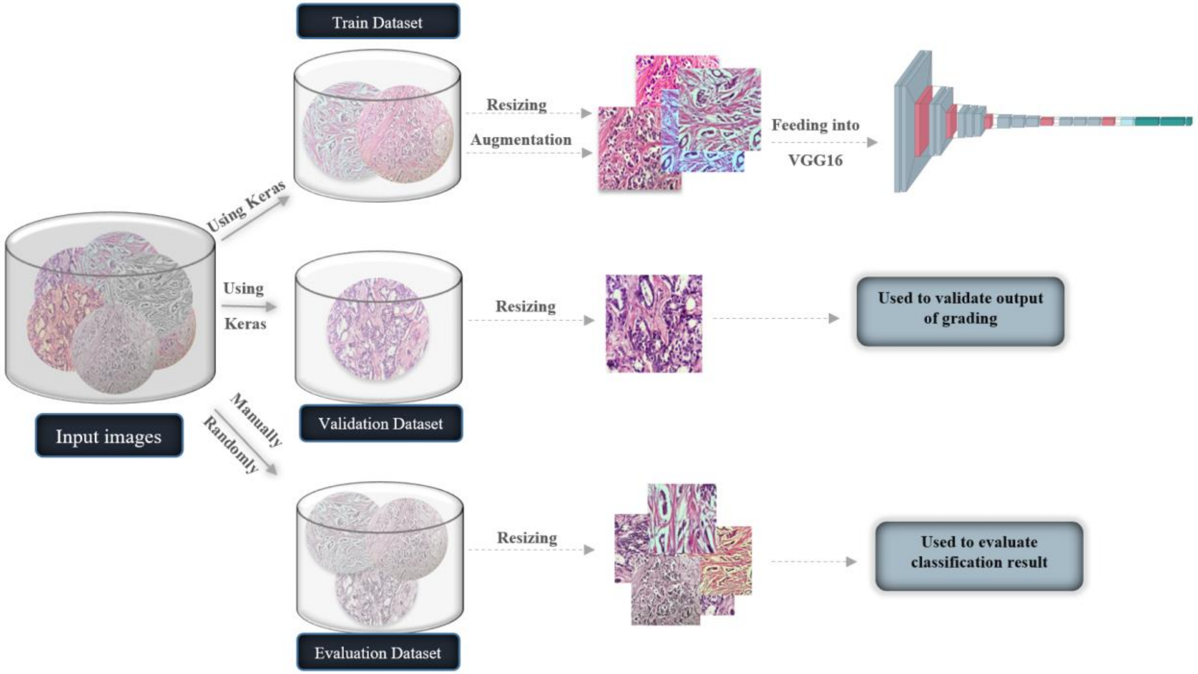




Grade I

Grade II

Grade III



Invasive Ductal Carcinoma Grading Dataset

DataBioX: Training dataset for Supervised Learning

	Numbers	4x	10x	20x	40x
Grade I	37	45	40	43	131
Grade II	43	59	64	63	180
Grade III	44	56	49	49	143
Total	124	160	153	155	454

Informatics in Medicine Unlocked 19 (2020) 100341

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

A histopathological image dataset for grading breast invasive ductal carcinomas

Hamidreza Bolhasani [a], Elham Amjadi [b], Maryam Tabatabaeian [c], Somayyeh Jafarali Jassbi [a,*]

[a] Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
[b] Poursina Hakim Digestive Diseases Research Center, Isfahan University of Medical Sciences, Isfahan, Iran
[c] Anahid Community-Based Breast Clinic, Isfahan, Iran

ARTICLE INFO

Keywords:
Breast cancer
Invasive ductal carcinoma
Histopathology
Digital pathology
Grading
Image dataset

ABSTRACT

Breast cancer is a common cancer in women, and one of the major causes of death among women around the world. Invasive ductal carcinoma (IDC) is the most widespread type of breast cancer with about 80% of all diagnosed cases. Early accurate diagnosis plays an important role in choosing the right treatment plan and improving survival rate among the patients. In recent years, efforts have been made to predict and detect all types of cancers by employing artificial intelligence. An appropriate dataset is the first essential step to achieve such a goal. This paper introduces a histopathological microscopy image dataset of 922 images related to 124 patients with IDC. The dataset has been published and is accessible through the web at: http://databiox.com. The distinctive feature of this dataset as compared to similar ones is that it contains an equal number of specimens from each of three grades of IDC, which leads to approximately 50 specimens for each grade.

1. Introduction

Cancer is a serious public health issue worldwide and the second leading cause of death in the United States [1]. According to the International Agency for Research on Cancer (IARC), about 18.1 million new cases and 9.6 million deaths caused by cancer were reported in 2018 [2].

As shown in Fig. 1, breast cancer is a common cancer and one of the major causes of death worldwide with 627,000 deaths among 2.1 million diagnosed cases in 2018 [2–6].

About 80% of all diagnosed breast cancers are invasive ductal carcinoma (IDC) which makes it the most widespread phenotypic subtype of all breast cancers [7]. It grows in milk ducts and invades fibrous tissue of the breast outside the duct. Pathologists recognize the cancer type and grade through visual investigation of tissue stained by hematoxylin and eosin (H&E) [17] as shown in Fig. 2. Accurate identification of IDC grade is essential in choosing the proper treatment plan and the outcome of the patient accordingly.

Histologic grade is a prognostic factor and an indicator of response to chemotherapy. It has been shown to be closely related to the frequency of recurrence and death due to IDC, and to disease-free interval and to longer life after mastectomy.

Several studies show that patients with a high grade IDC, and mastectomy treated, had a remarkably higher frequency of auxiliary lymph node (ALN) and mortality rate than the patients with lower grade. Histologic grade has been shown to be significantly related not only to recurrence and death due to invasive ductal carcinoma, but also to the disease-free interval and overall length of survival after mastectomy regardless of clinical stage. High-grade carcinomas result in early treatment failures, whereas later recurrences are more often observed among low-grade tumors [8,9].

In the recent years, efforts have been made to predict and detect all types of cancers, employing artificial intelligence and its sub categories like machine learning and deep learning [10–15]. The appropriate dataset is the first essential step to achieve this goal. There are some available datasets for IDC [11–13] but our findings indicate that there is no published dataset for histologic grading of IDC. Therefore, this research aimed at providing a well-organized histopathological microscopy image dataset for grading IDC. The images have been taken from breast tissues stained with H&E and are labeled based on their grade and magnification level.

This paper has been organized as follows: Related works are presented in section 2. Section 3 covers the main contribution of this research, i.e. the prepared image dataset: databiox. Methods of collecting images and doing the statistics are presented in this section.

Published Paper in 2020

Elsevier Publication

Officially requested by many prestigious Universities around the world like MIT and University of Johns Hopkins.

Name
Pawan Sinha

Address
77 Massachusetts Avenue, Cambridge, MA 02139, United States of America

Name of the University / Institution / Company
Massachusetts Institute of Technology

Department (If any)
Brain and Cognitive Sciences

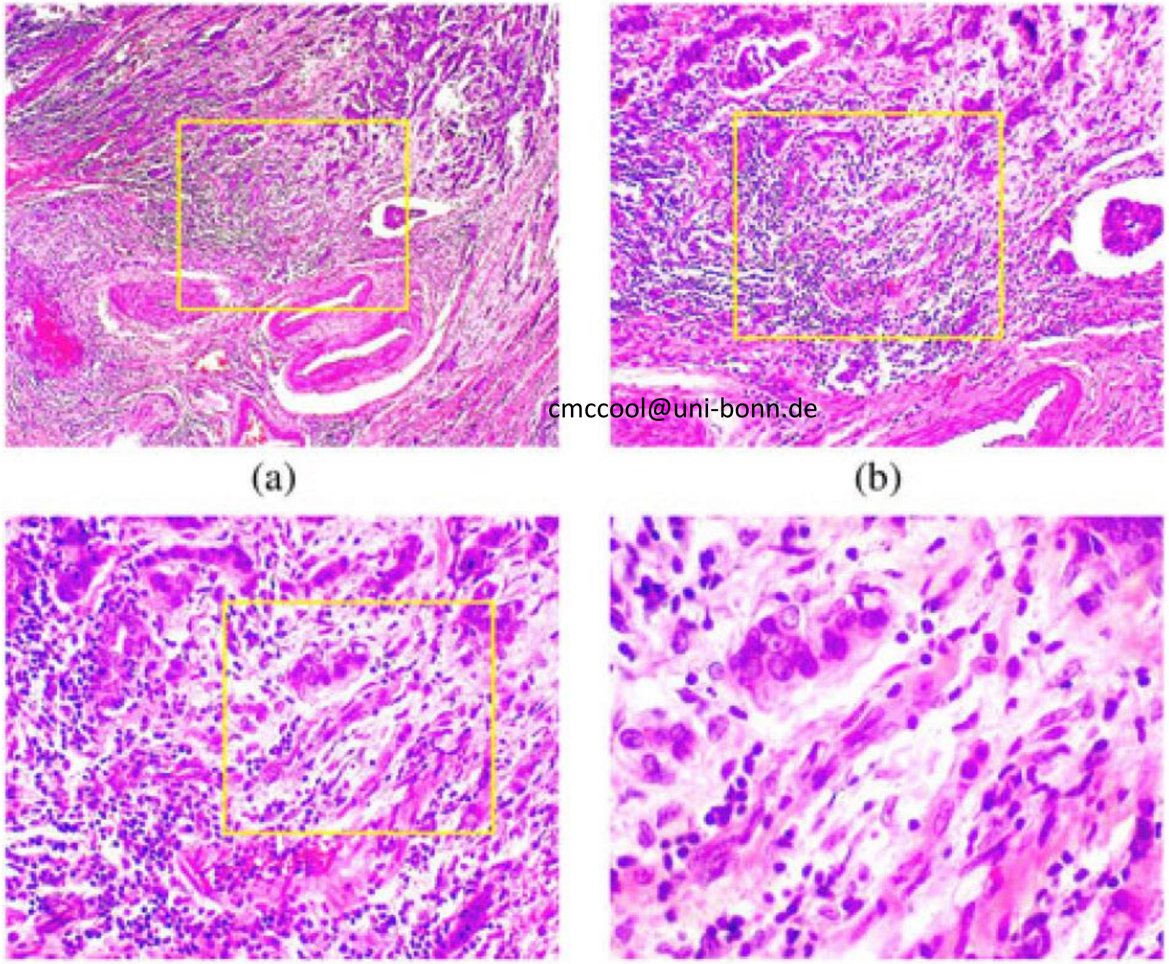
Academic Degree
PhD

Email
psinha@mit.edu

Required Dataset
Breast Cancer Grading Dataset

Breast Cancer Classification Dataset

BreaKHis: Training dataset for Supervised Learning



(a) (b) (c) (d)

cmccool@uni-bonn.de

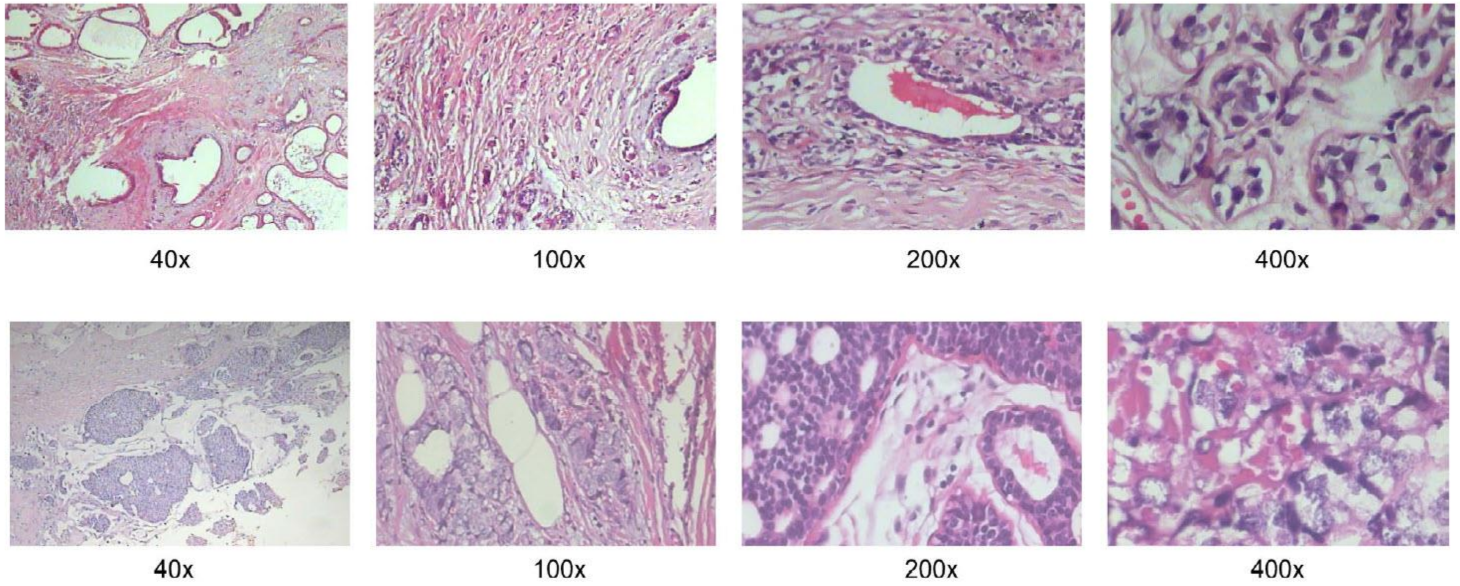
Magnification	Benign	Malignant	Total
40×	625	1370	1995
100×	644	1437	2081
200×	623	1390	2013
400×	588	1232	1820
Total	2480	5429	7909
# Patients	24	58	82

Magnification	A	F	TA	PT	Total
40×	114	253	109	149	598
100×	113	260	121	150	614
200×	111	264	108	140	594
400×	106	237	115	130	562
Total	444	1014	453	569	2368
# Patients	4	10	3	7	24

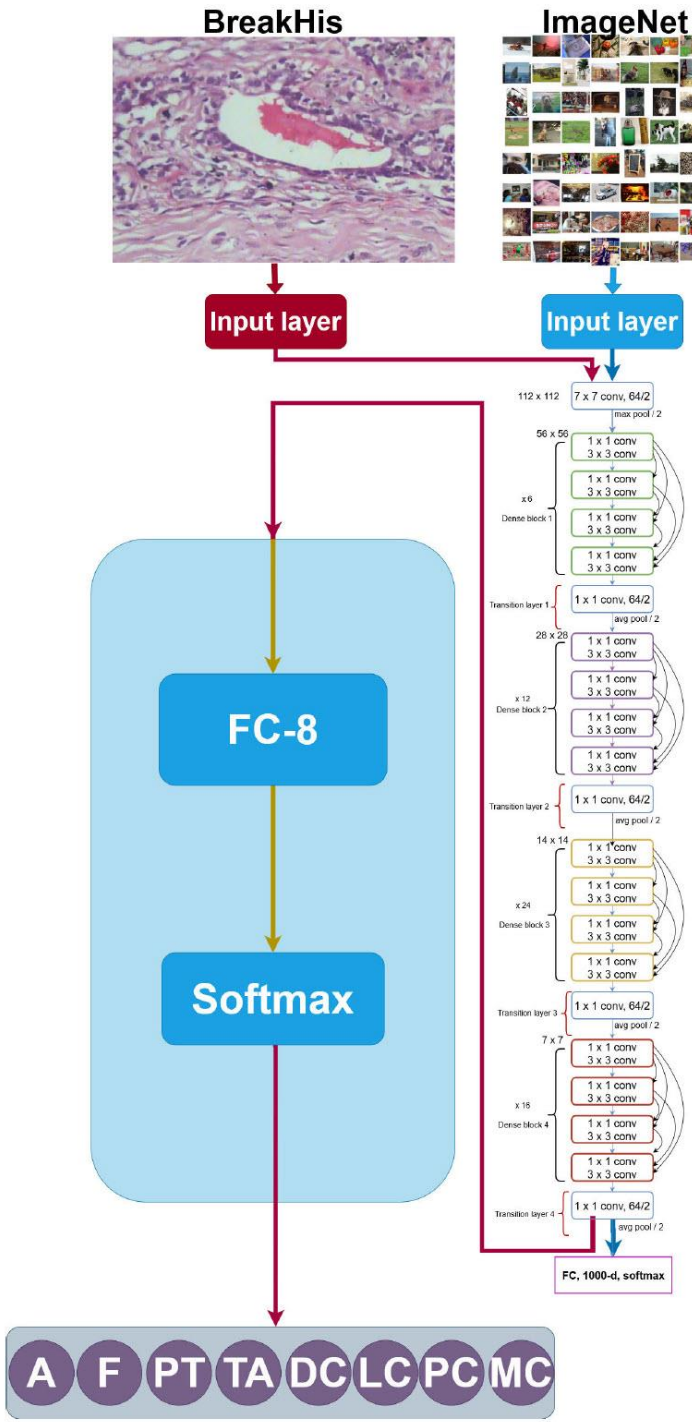
Breast Cancer Detection

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Deep Learning Applied for Histological Diagnosis of Breast Cancer (2020)

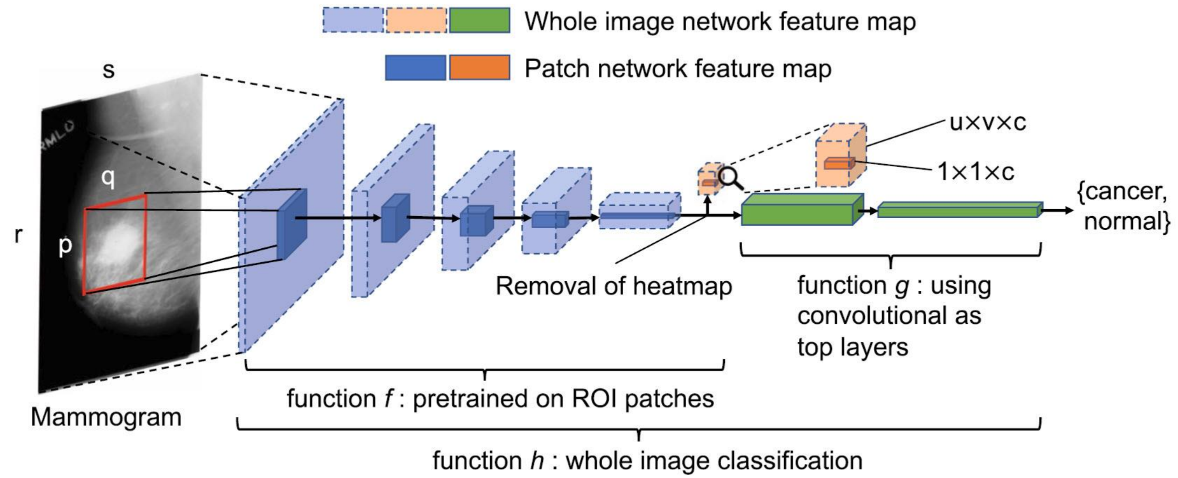


Model details	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Transfer learning	-	-	ResNet50	DenseNet121	ResNet50	Densenet121
CNN layers	2	2	40	117	40	117
setting	setting 3	setting 3	setting 3	setting 3	setting 4	setting 4
Data Augmentation	-	✓	✓	✓	✓	✓
Results						
Misclassified images	20	18	1	0	3	2
Accuracy(ILA)	79.59%	81.63%	98.98%	**100%**	96.94%	97.97%
Precision	78.20%	82.85%	98.38%	**100%**	95.31%	96.82%
Recall	95.31%	90.62%	**100%**	**100%**	**100%**	**100%**
$F1_{score}$	85.91%	86.57%	99.18%	**100%**	97.59%	98.38%

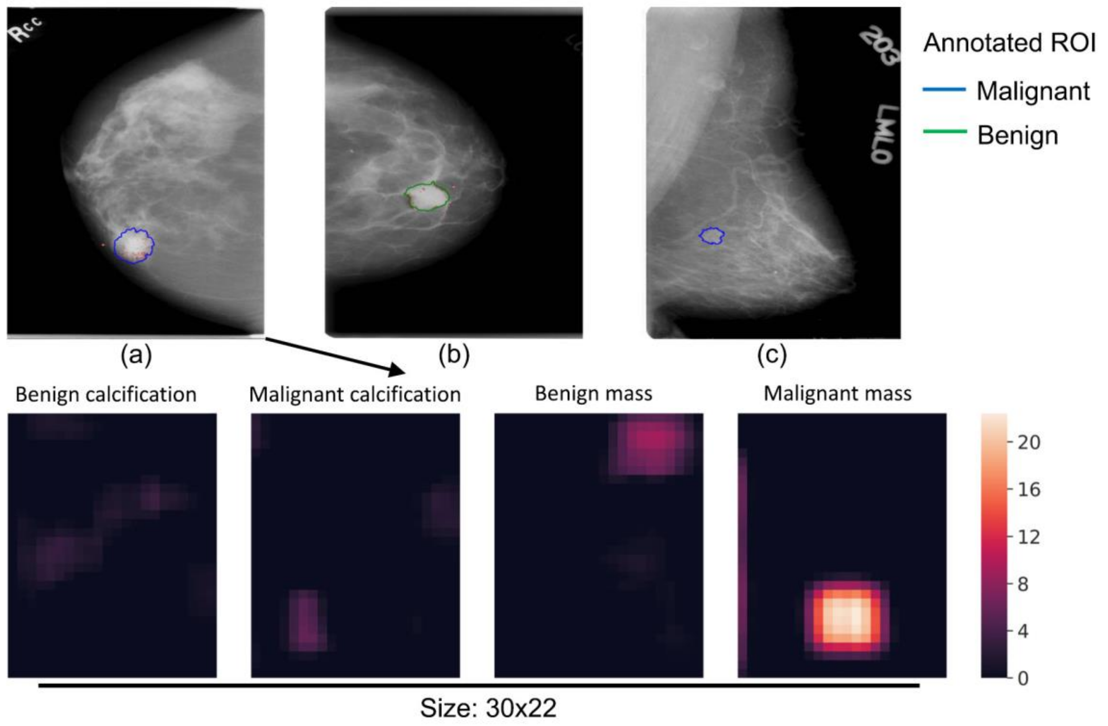


Breast Cancer Detection

Deep Learning to Improve Breast Cancer Detection on Screening Mammography (2019)



#Patients	#Images	Resnet-Resnet	Resnet-VGG	VGG-VGG	VGG-Resnet
20	79	0.92	0.88	0.87	0.89
30	117	0.93	0.94	0.93	0.90
40	159	0.93	0.95	0.93	0.93
50	199	0.94	0.95	0.94	0.93
60	239	0.95	0.95	0.95	0.94
72 (All)	280	0.95	0.95	0.95	0.95

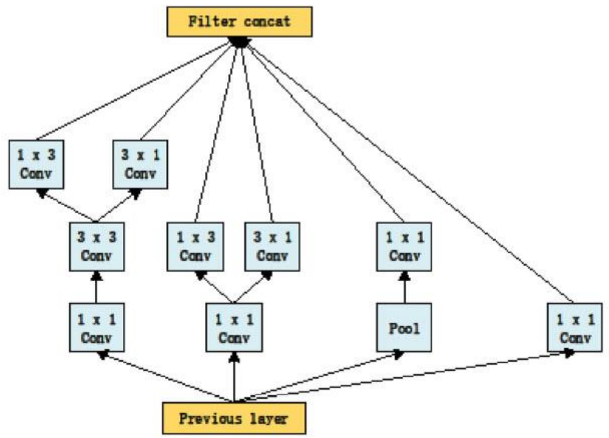


Breast Cancer Classification

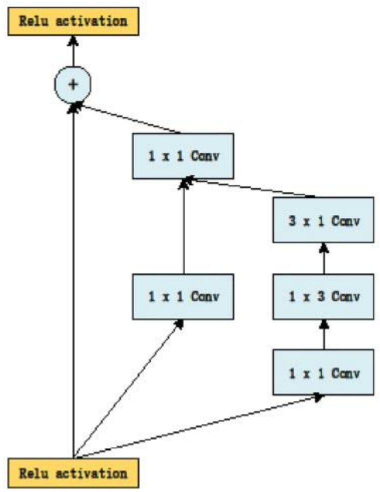
Deep Learning Based Analysis of Histopathological Images of Breast Cancer (2019)

frontiers in Genetics

Classification	Network	Criteria	Magnification factors			
			40X	100X	200X	400X
Binary	INV3	Se	98.00	98.48	**99.01**	96.41
		Sp	**94.31**	93.46	91.40	90.99
		PPV	**97.41**	96.67	95.88	95.89
		DOR	81,233	92,303	**106,700**	27,105
		ACC_IL	**96.84**	96.76	96.49	94.71
		ACC_PL	**97.74**	94.19	87.21	96.67
		F1	**97.70**	97.56	97.42	96.15
		AUC	**99.47**	99.03	99.29	97.91
		Kappa	92.64	**92.74**	91.95	87.68
	IRV2	Se	98.48	98.90	**99.13**	98.06
		Sp	**96.63**	92.95	92.80	92.10
		PPV	**98.46**	96.45	96.39	96.51
		DOR	**185,774**	118,782	147,138	58,835
		ACC_IL	**97.90**	96.88	96.98	96.98
		ACC_PL	**98.03**	97.07	82.74	88.12
		F1	**98.47**	97.66	97.74	97.28
		AUC	99.57	98.84	**99.61**	98.81
		Kappa	**95.12**	92.96	93.18	91.05
Multi-class	INV3	ACC_IL	**90.28**	85.35	83.99	82.08
		ACC_PL	**90.44**	89.05	80.63	81.08
		Macro-F1	**88.55**	82.59	79.64	77.98
		Micro-F1	**90.28**	85.35	83.99	82.08
		Kappa	**87.37**	80.26	77.91	76.39
	IRV2	ACC_IL	**92.07**	88.06	87.62	84.50
		ACC_PL	**89.11**	88.45	86.07	71.42
		Macro-F1	**90.89**	85.67	84.08	80.13
		Micro-F1	**92.07**	88.06	87.62	84.50
		Kappa	**89.74**	84.03	82.84	79.70



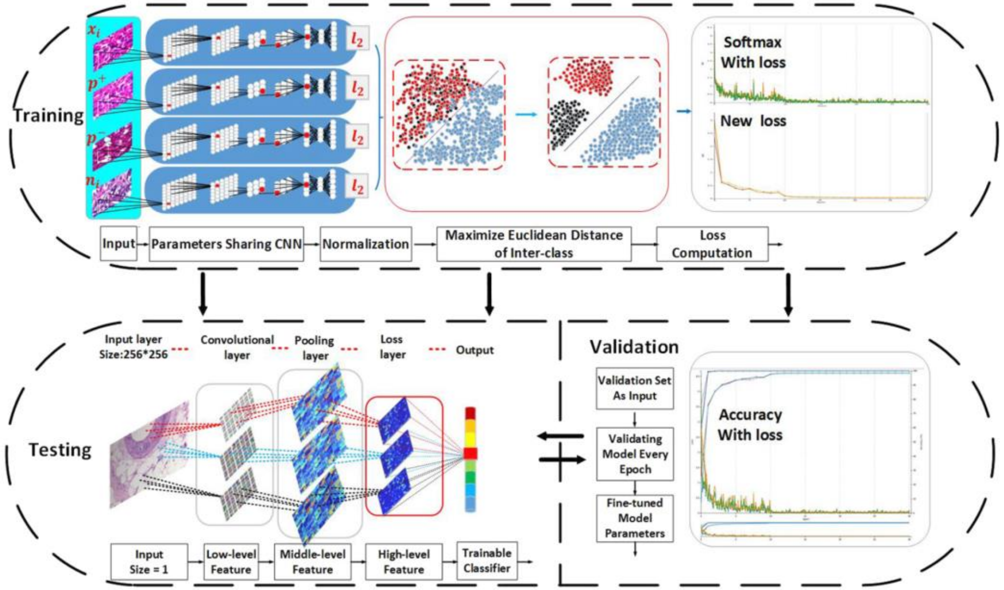
InceptionV3

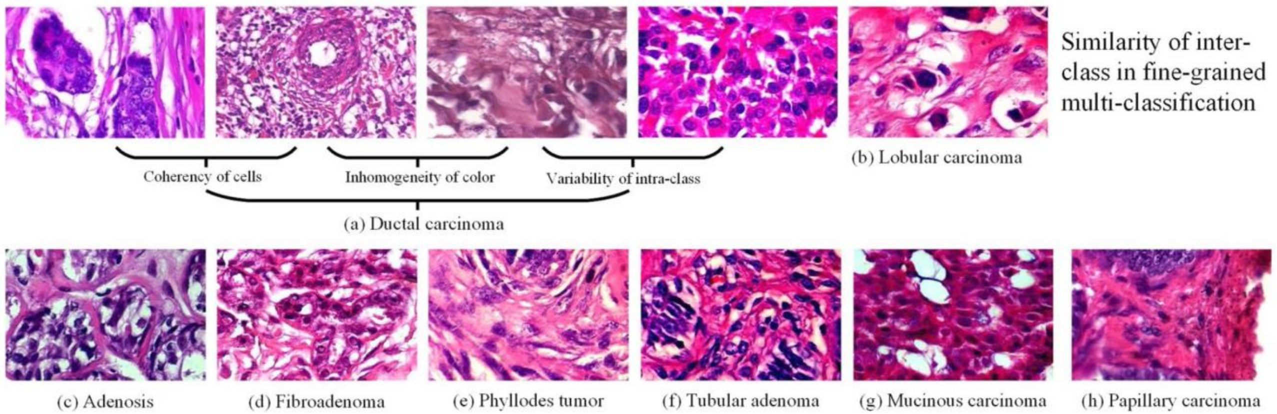


Inception_ResNetV2

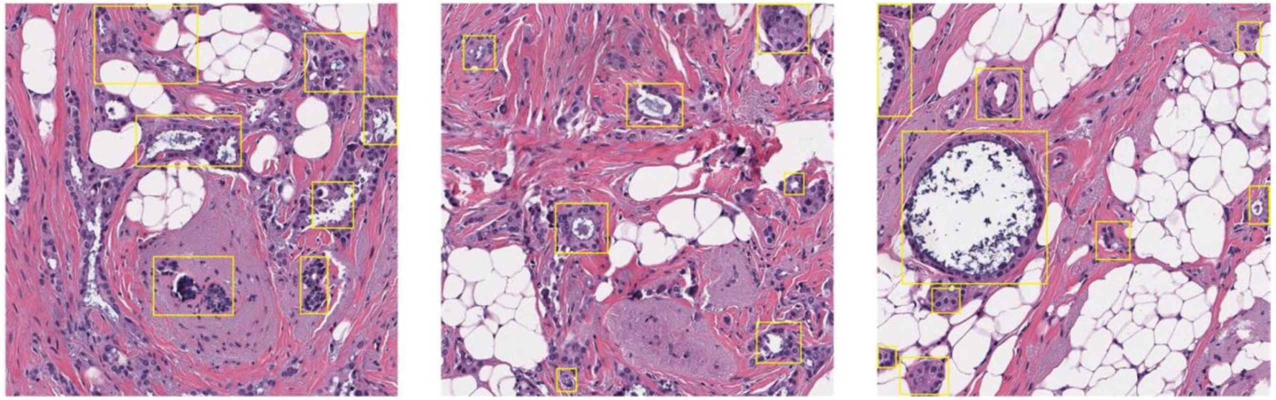
Breast Cancer Classification

Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model (2017)





Accuracy at	Methods	Magnification factors			
		40X	100X	200X	400X
Image level	AlexNet[17]	85.6±4.8	83.5±3.9	83.1±1.9	80.8±3.0
	CSDCNN	**95.8±3.1**	**96.9±1.9**	**96.7±2.0**	**94.9±2.8**
Patient level	PFTAS+QDA[12]	83.8±4.1	82.1±4.9	84.2±4.1	82.0±5.9
	PFTAS+SVM[12]	81.6±3.0	79.9±5.4	85.1±3.1	82.3±3.8
	GLCM+1-NN[12]	74.7±1.0	76.8±2.1	83.4±3.3	81.7±3.3
	PFTAS+RF[12]	81.8±2.0	81.3±2.8	83.5±2.3	81.0±3.8
	AlexNet[17]	90.0±6.7	88.4±4.8	84.6±4.2	86.1±6.2
	CSDCNN	**97.1±1.5**	**95.7±2.8**	**96.5±2.1**	**95.7±2.2**

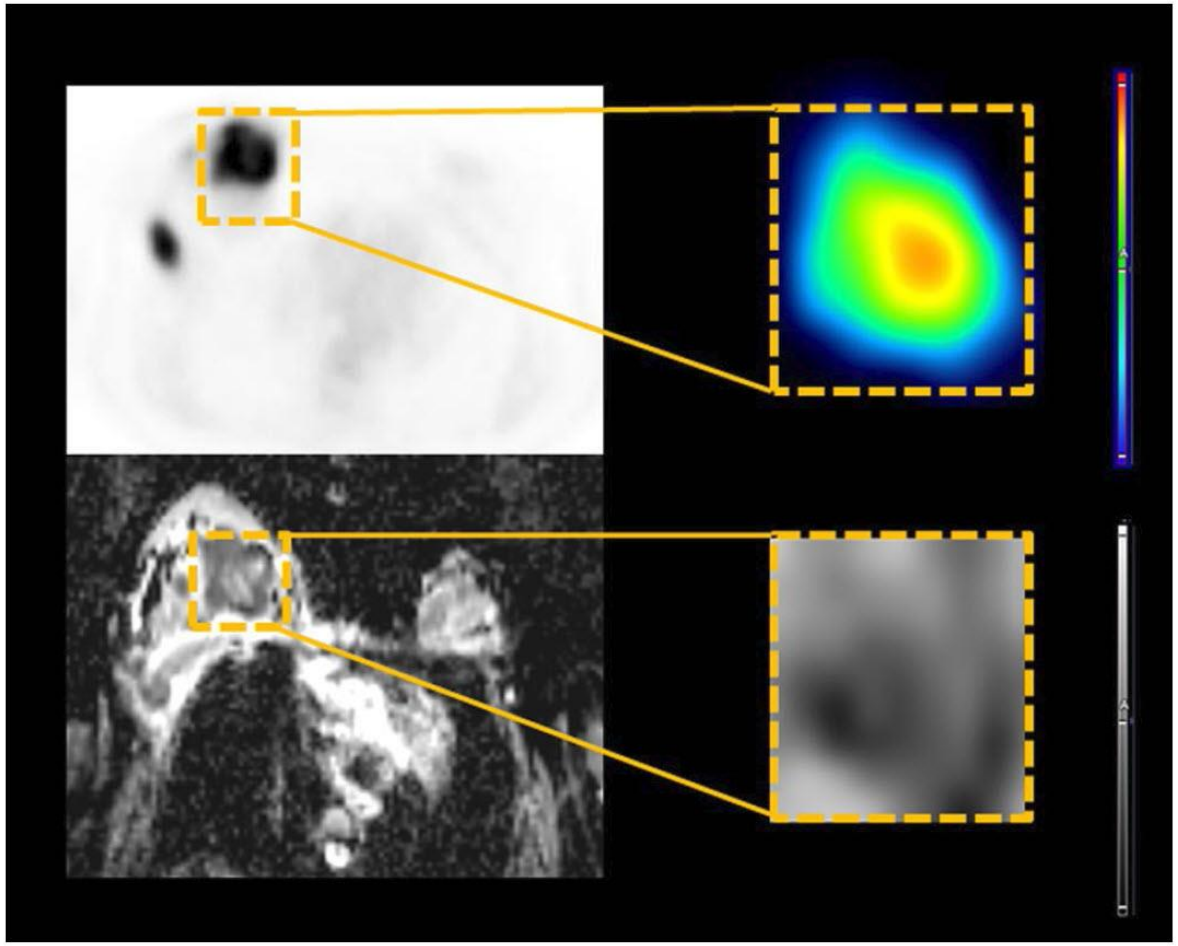


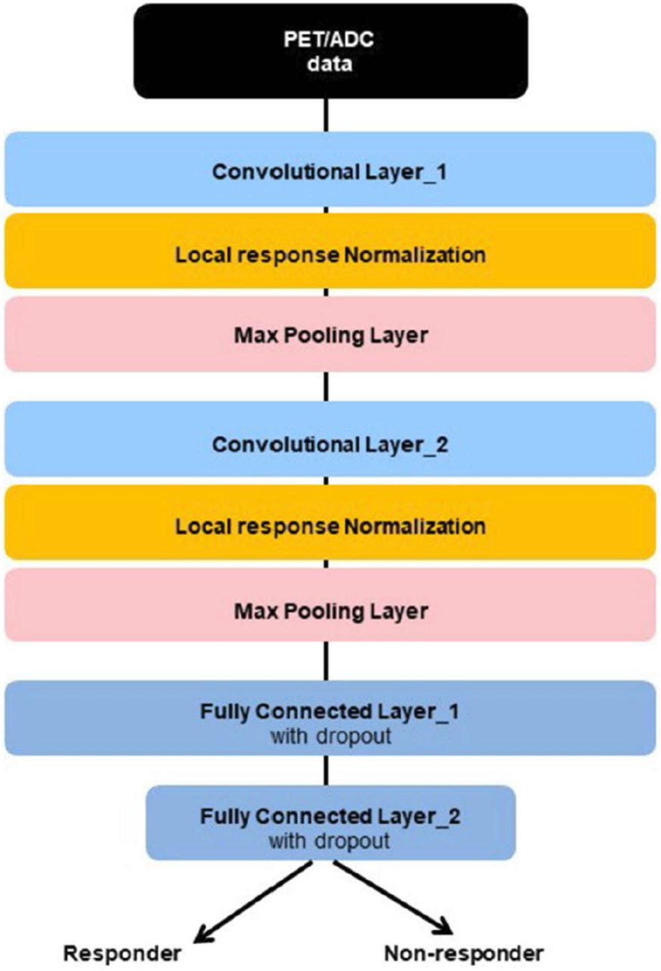
nature
SCIENTIFIC REPORTS

Breast Cancer Treatment Plan

Early prediction of neoadjuvant chemotherapy response for advanced breast cancer

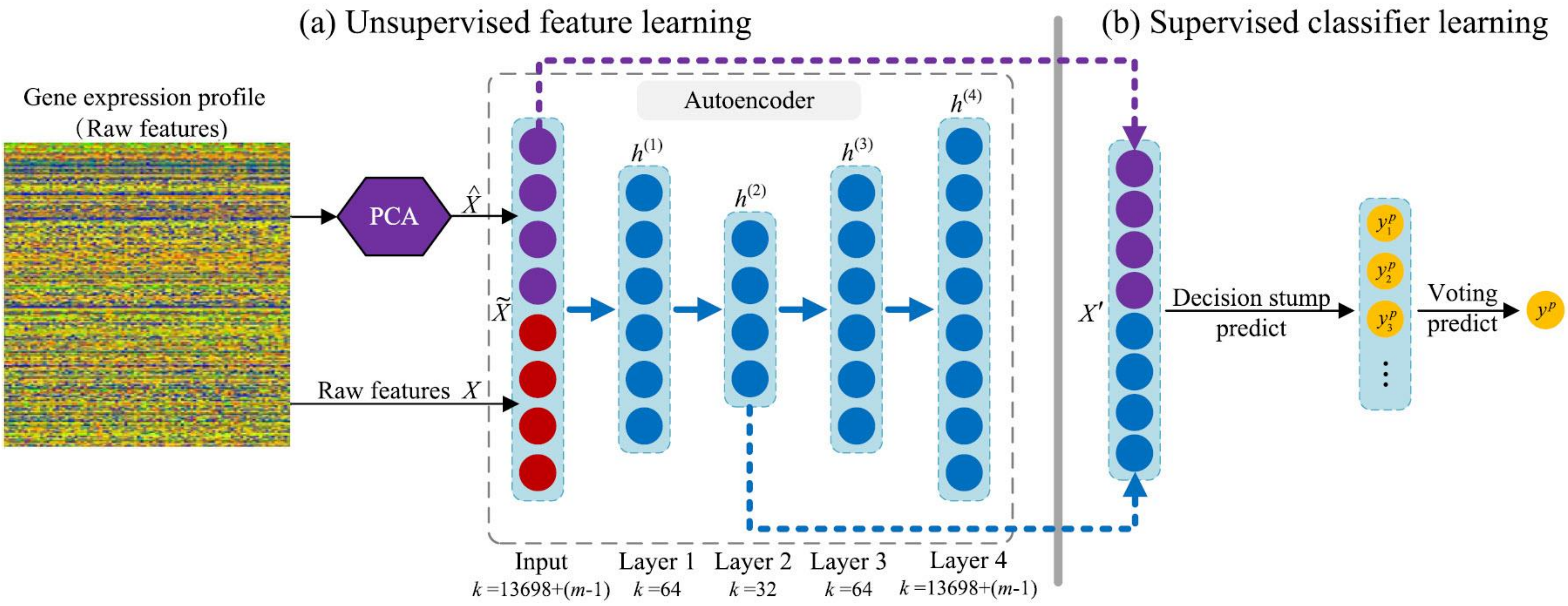
Using PET/MRI image deep learning





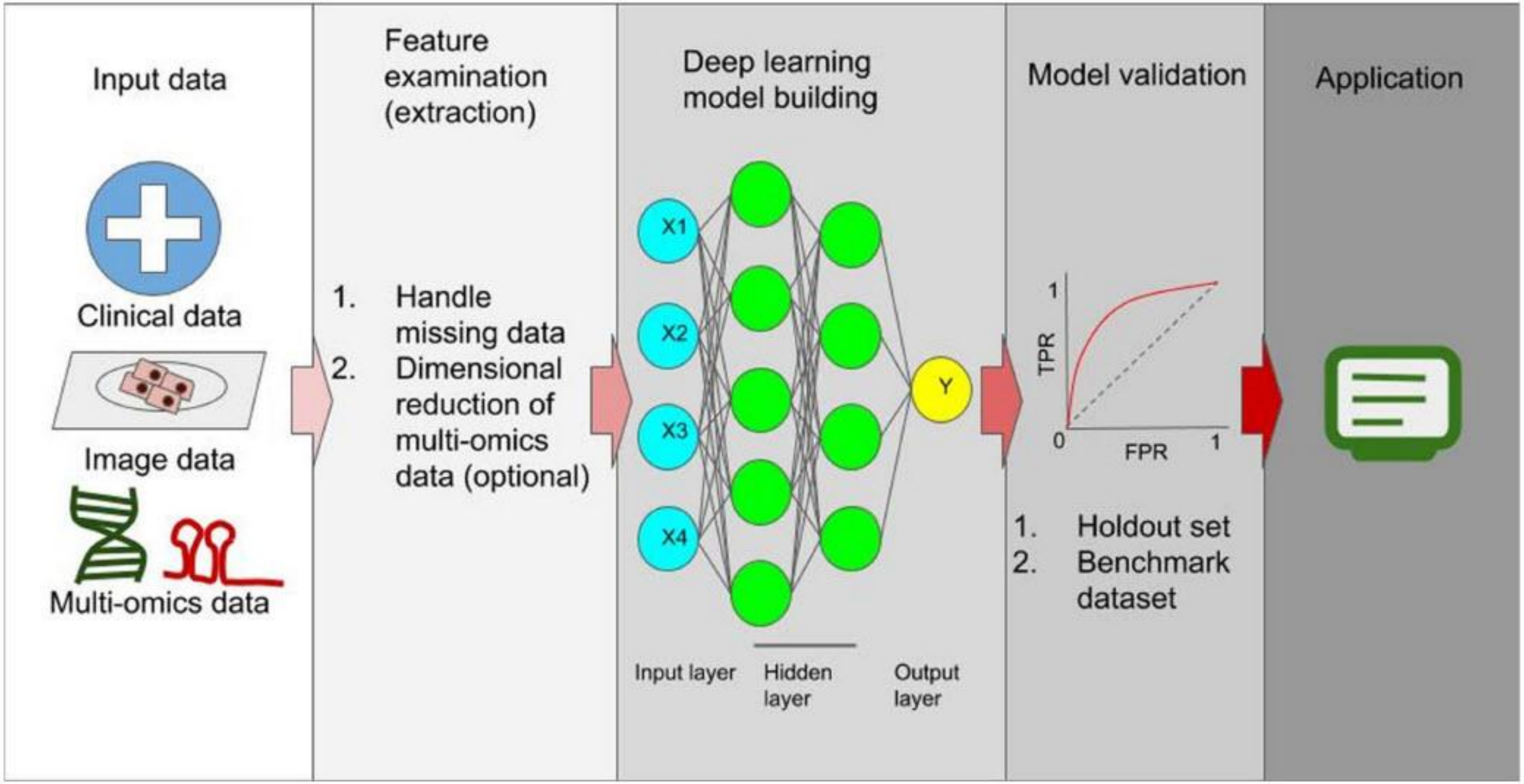
Breast Cancer Prognosis

Integrating Feature Selection and Feature Extraction Methods With Deep Learning to Predict Clinical Outcome of Breast Cancer (2018)



Breast Cancer Prognosis

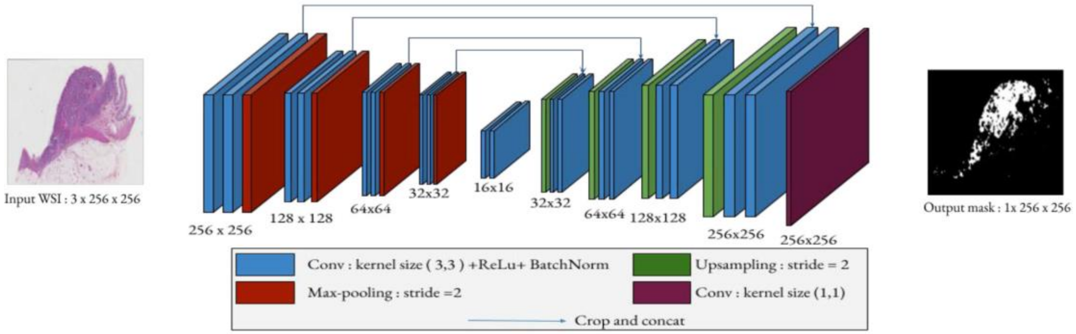
The Application of Deep Learning in Cancer Prognosis Prediction (2020)



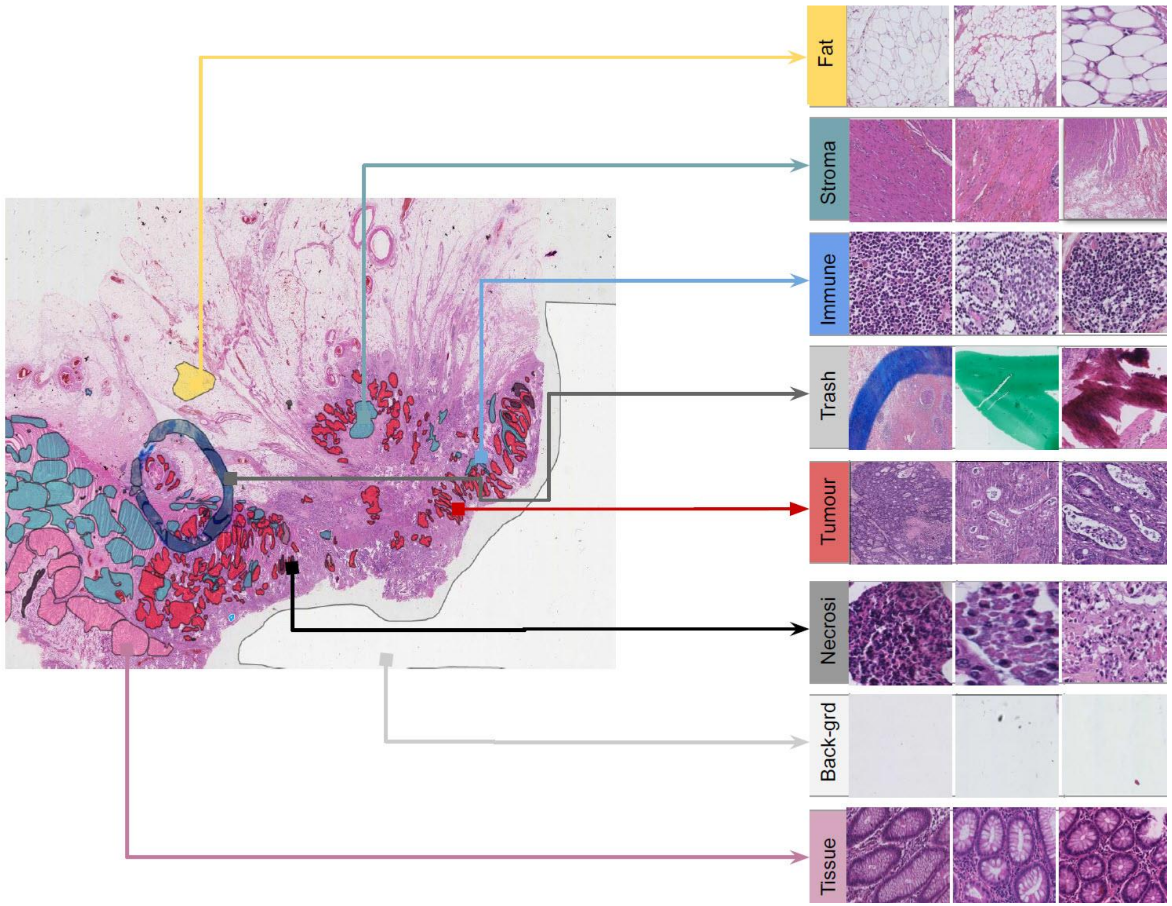
Colon Cancer Analysis

Deep learning for colon cancer histopathological images analysis (2021)

ELSEVIER

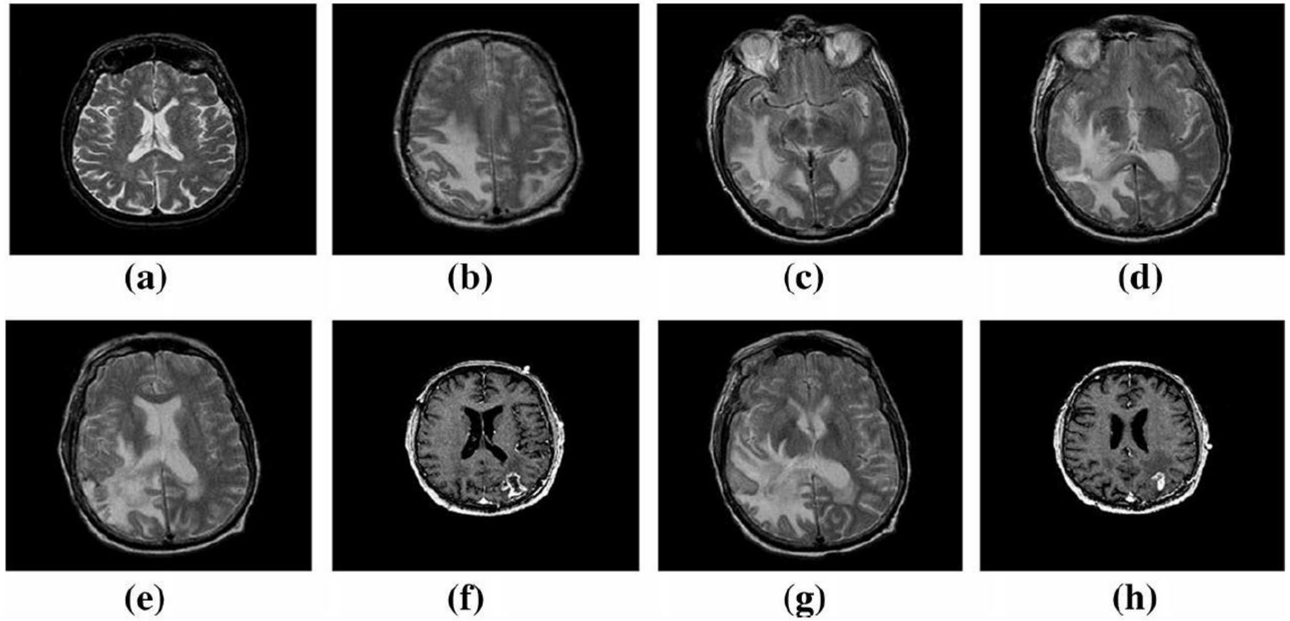


	Trash	Necrosis	Fat	Tumour	Stroma	Tissue	Immune	Backgrnd
Trash	0.9742	0	0	0	0	0	0	0
Necrosis	0	0.9151	0	0.0002	0	0	0.0015	0
Fat	0	0	1.0000	0	0	0	0	0
Tumour	0	0.0011	0	0.9900	0	0	0.0100	0
Stroma	0	0	0	0	0.9706	0.0020	0	
Tissue	0	0	0	0	0	0.9606	0	0
Immune	0	0.0071	0	0.0038	0	0.0080	0.9320	0
Backgnd	0.0023	0	0	0	0	0	0	0.9786

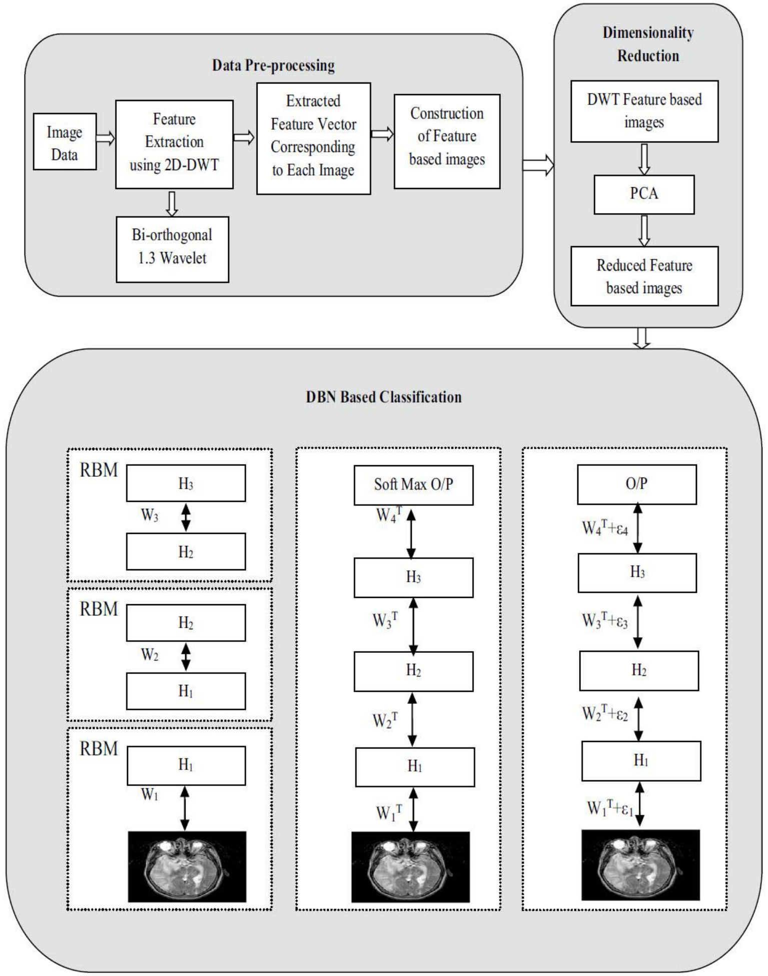


Glioblastoma Detection

Analyzing MRI scans to detect glioblastoma tumor using hybrid deep belief networks (2020)



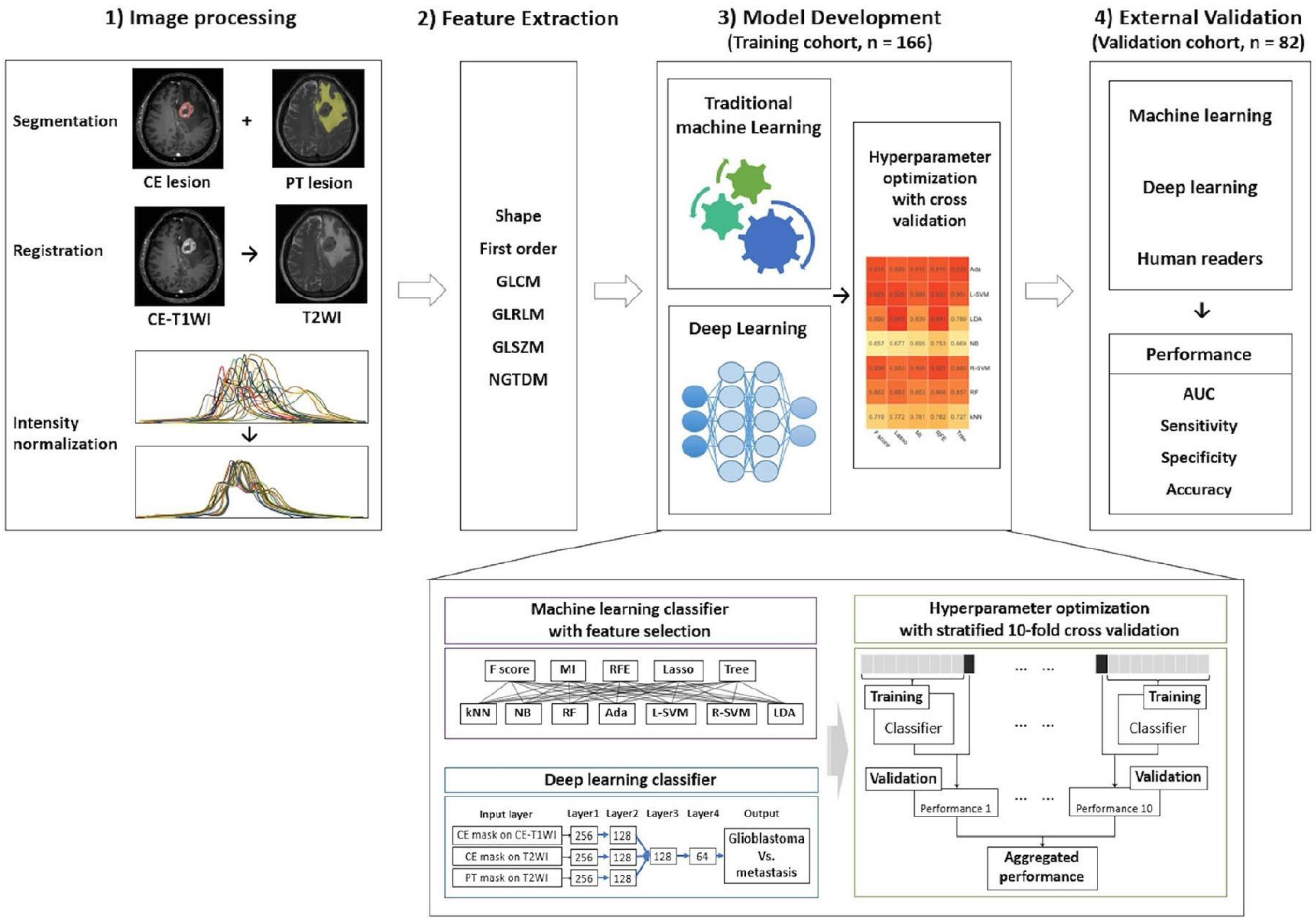
Classifiers	Overall accuracies (%)	Average accuracies (%)	Kappa statistics
DBN	91	89	0.4811
DWT-DBN	93	91	0.5732
DWT-PCA-DBN	97	95	0.6522



Glioblastoma Distinguishing

Robust performance of deep learning for distinguishing glioblastoma from single brain metastasis using radiomic features (2020)

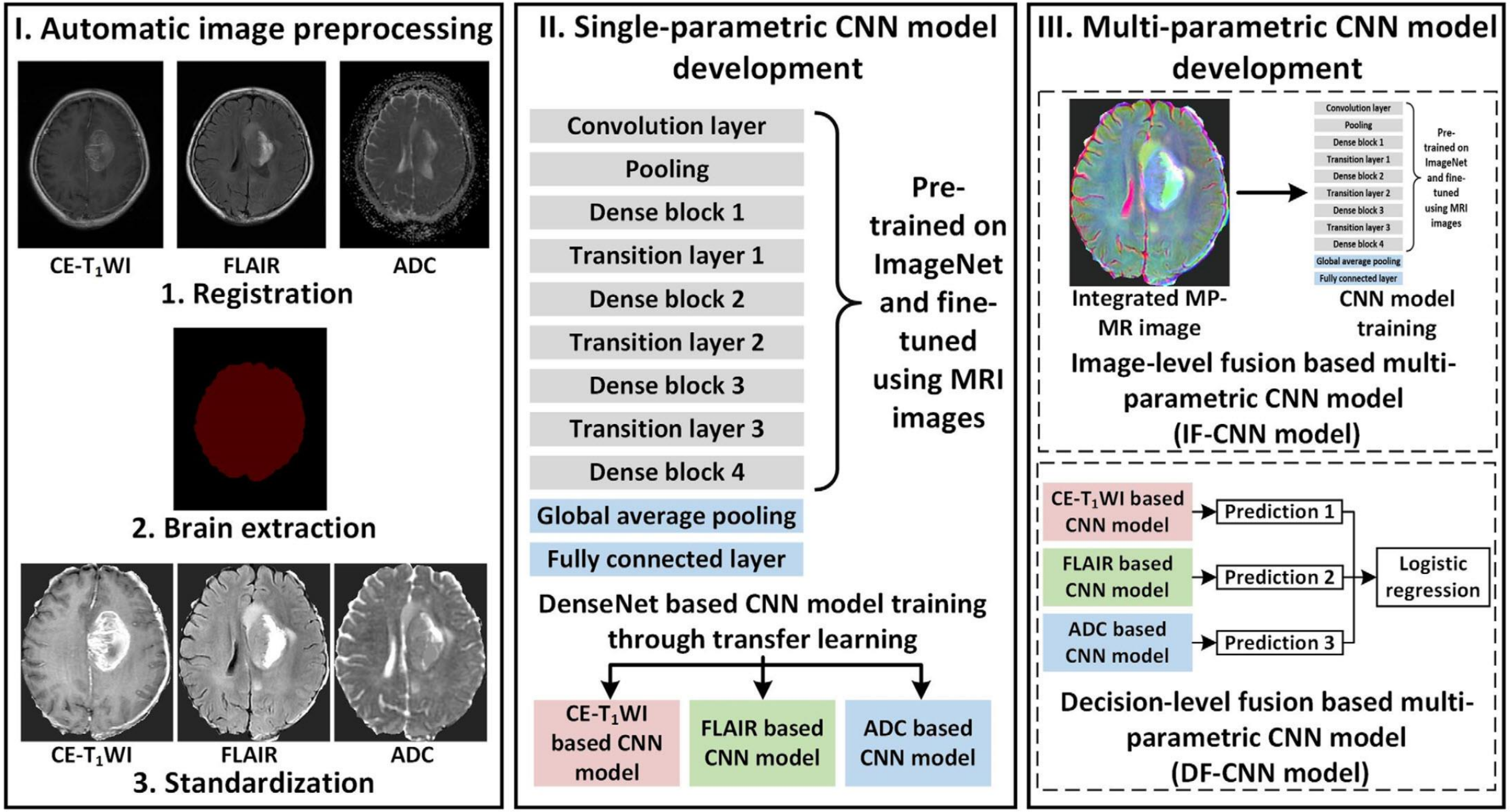
nature SCIENTIFIC REPORTS



Data	Classifier	Cross validation	External validation			
		Mean AUC	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)
Traditional machine learning						
CE mask	AdaBoost with LASSO	0.870	0.858 (0.787–0.926)	68.0 (53.3–80.5)	93.8 (79.2–99.2)	78.0 (67.5–86.4)
	L-SVM with Tree-based selection	0.875	0.833 (0.755–0.904)	62.0 (47.2–75.4)	93.8 (79.2–99.2)	74.4 (63.6–83.4)
	LDA with LASSO	0.863	0.818 (0.737–0.891)	64.0 (49.2–77.1)	87.5 (71.0–96.5)	73.2 (62.2–82.4)
PT mask	AdaBoost with Tree-based selection	0.816	0.773 (0.668–0.870)	94.0 (83.5–98.8)	34.4 (18.6–53.2)	70.7 (59.7–80.3)
	L-SVM with RFE	0.830	0.803 (0.718–0.879)	86.0 (75.5–96.5)	65.6 (46.8–81.4)	78.0 (67.2–88.8)
	LDA with MI	0.818	0.787 (0.695–0.870)	94.0 (83.5–98.8))	50.0 (31.9–68.1	76.8 (66.2–85.4)
Combined mask	AdaBoost with Tree-based selection	0.926	0.890 (0.823–0.947)	80.0 (62.3–90.0)	87.5 (71.0–94.5)	82.9 (73.0–90.3)
	L-SVM with RFE	0.932	0.886 (0.798–0.927)	80.0 (62.3–90.0)	84.4 (67.2–94.7)	81.7 (71.6–89.4)
	LDA with LASSO	0.945	0.899 (0.839–0.951)	84.0 (70.9–92.8)	78.1 (70.9–90.7)	81.7 (71.6–89.4)
Deep learning						
CE mask	DNN	0.887	0.887 (0.812–0.951)	62.5 (45.7–79.3)	96.0 (90.6–100)	82.9 (74.8–91.1)
PT mask	DNN	0.865	0.825 (0.722–0.887)	75.0 (60.2–90.1)	82.0 (71.4–92.6)	79.3 (70.5–88.0)
Combined mask	DNN	0.986	0.956 (0.918–0.990)	90.6 (80.5–100)	88.0 (79.0–97.0)	89.0 (82.3–95.8)
Human reading						
Images	Reader 1		0.774 (0.685–0.852)	97.0 (91.1–100)	50.0 (36.1–63.9)	68.7 (58.7–78.7)
Images	Reader 2		0.904 (0.852–0.951)	81.8 (68.7–95.0)	78.0 (66.5–89.5)	79.5 (70.8–88.2)

Glioblastoma Distinguishing

Deep Learning for Automatic Differential Diagnosis of Primary Central Nervous System Lymphoma And Glioblastoma: Multi-Parametric Magnetic Resonance Imaging (2021)



Glioblastoma Distinguishing

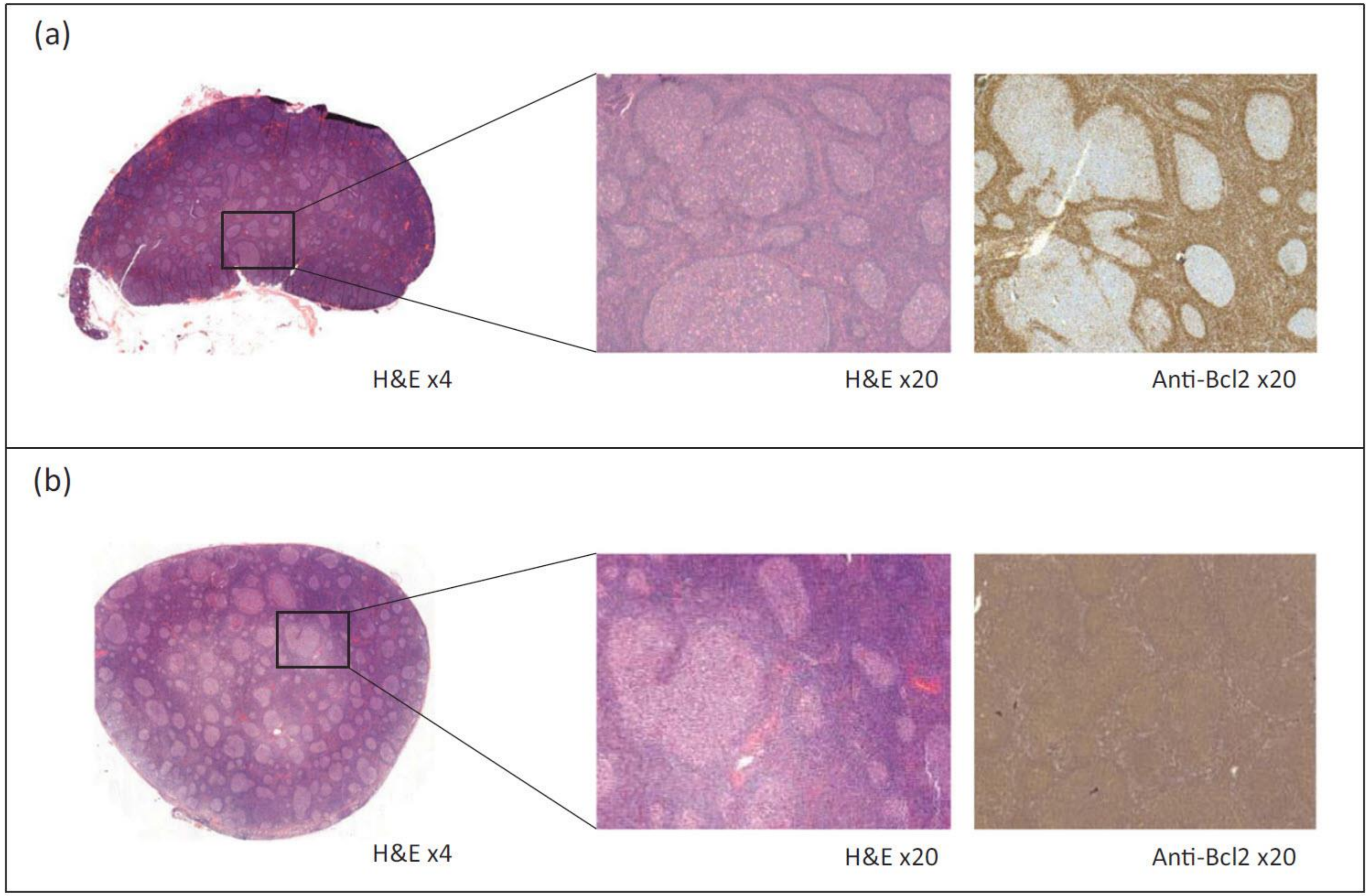
Deep Learning for Automatic Differential Diagnosis of Primary Central Nervous System Lymphoma And Glioblastoma: Multi-Parametric Magnetic Resonance Imaging (2021)

Models and Radiologists	Model Type	ACC	SEN	SPE	AUC
CNN models	CE-T1WI	0.884	0.934	0.841	0.956
	FLAIR	0.782	0.772	0.790	0.852
	ADC	0.700	0.860	0.558	0.828
	IF-CNN	0.830	0.772	0.882	0.919
	DF-CNN	0.899	0.934	0.867	0.964
Radiomics models	CE-T1WI	0.865	0.838	0.890	0.925
	FLAIR	0.771	0.757	0.784	0.867
	ADC	0.796	0.758	0.830	0.868
	IF-RADS	0.856	0.831	0.878	0.929
	DF-RADS	0.820	0.838	0.804	0.911
Junior radiologist	–	0.875	0.823	0.921	–
Intermediate-level radiologist	–	0.878	0.801	0.947	–
Senior radiologist	–	0.906	0.853	0.954	–

CNN = convolutional neural network; ACC = accuracy; SEN = sensitivity; SPE = specificity; AUC = area under the ROC curve; IF = image-level fusion; DF = decision-level fusion.

Lymphoma Accurate Diagnosis

Accurate diagnosis of lymphoma on whole-slide histopathology images using deep learning (2020)

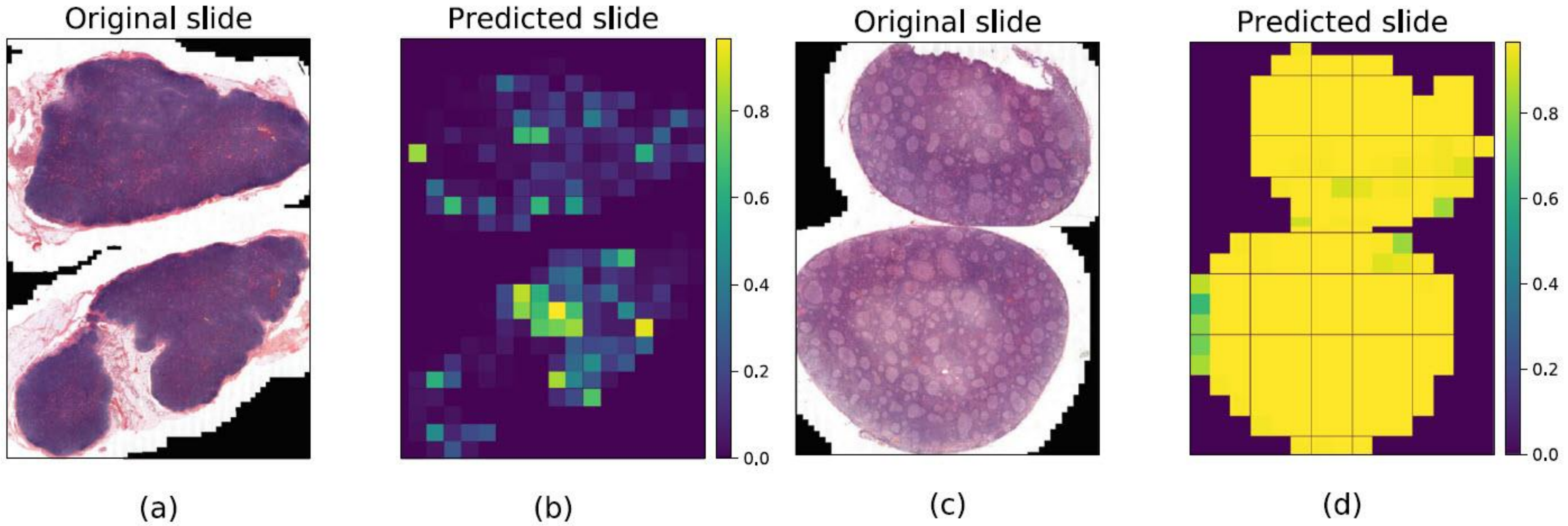


Follicular Hyperplasia (FH)

Follicular Lymphoma (FL)

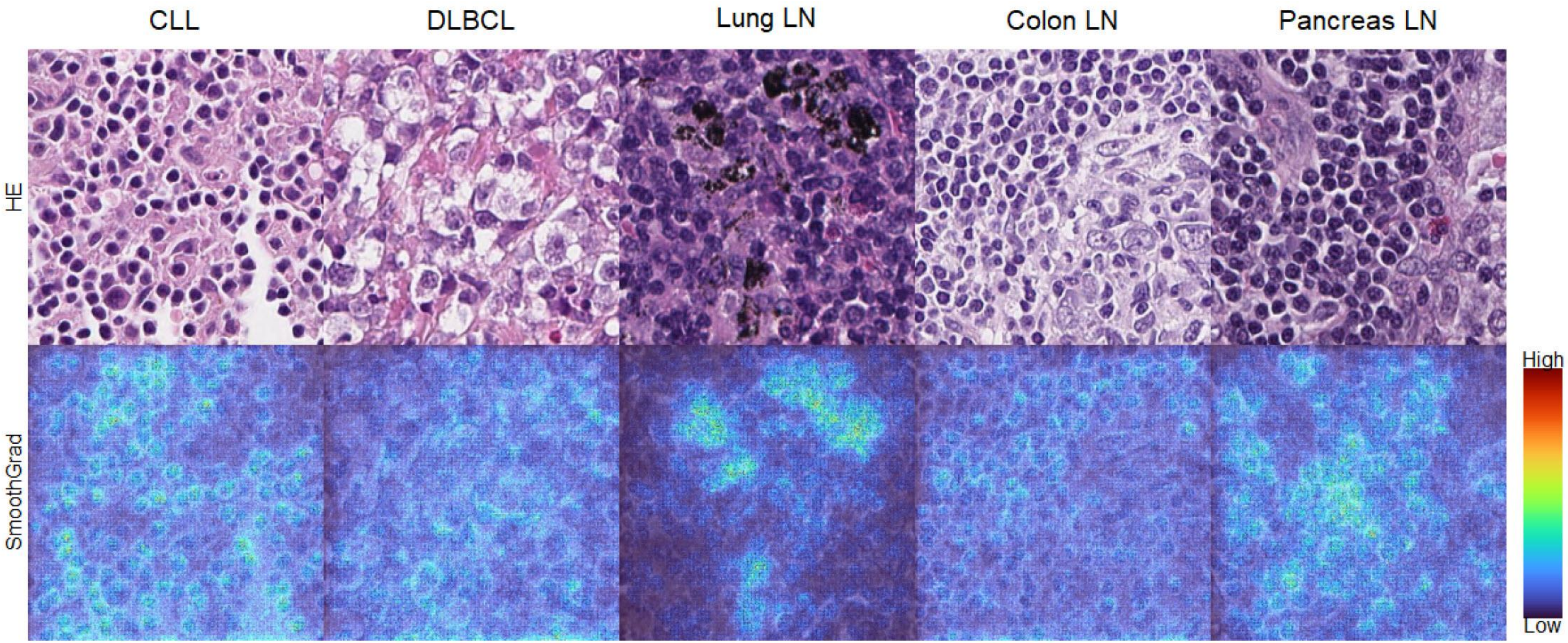
Lymphoma Accurate Diagnosis

Accurate diagnosis of lymphoma on whole-slide histopathology images using deep learning (2020)



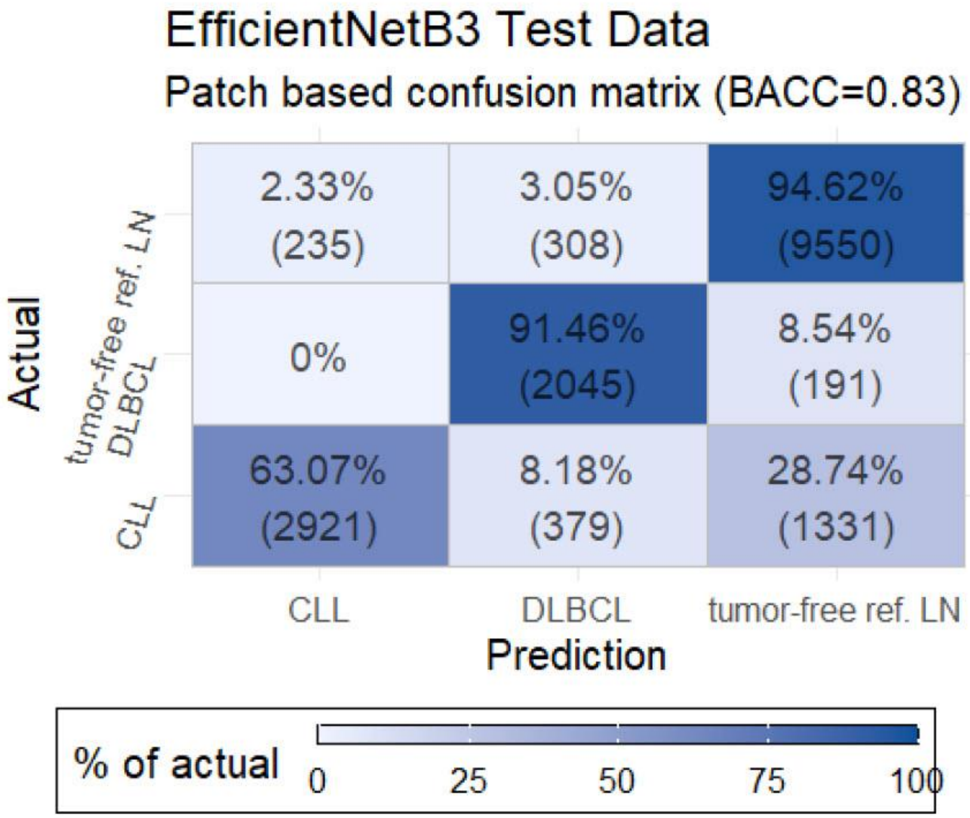
Lymphoma Non-Hodgkin Classification

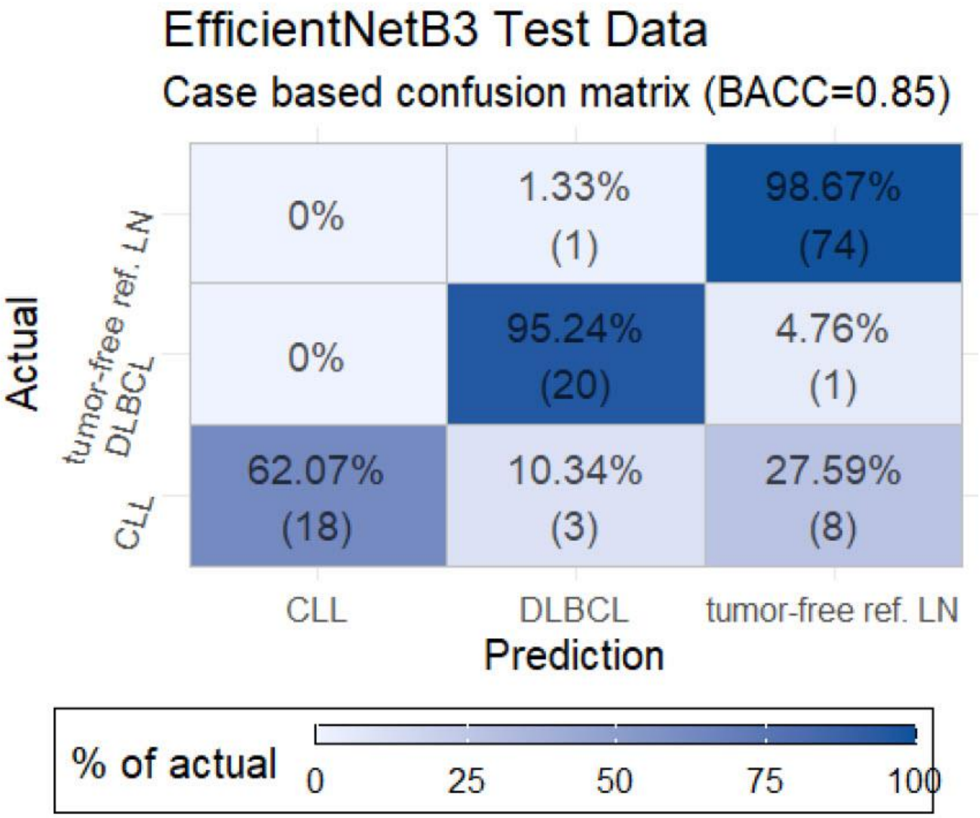
Deep Learning for the Classification of Non-Hodgkin Lymphoma on Histopathological Images (2021)



Lymphoma Non-Hodgkin Classification

Deep Learning for the Classification of Non-Hodgkin Lymphoma on Histopathological Images (2021)





Motivation & Hope…

Dream Big, Work Hard and Stay focused.

2003 - 2006

6 times Surgery

+30 times Chemotherapy and…

Hamidreza Bolhasani

Thanks!

Hamidreza Bolhasani
Hamidreza@Bolhasani.net
Nov 2021