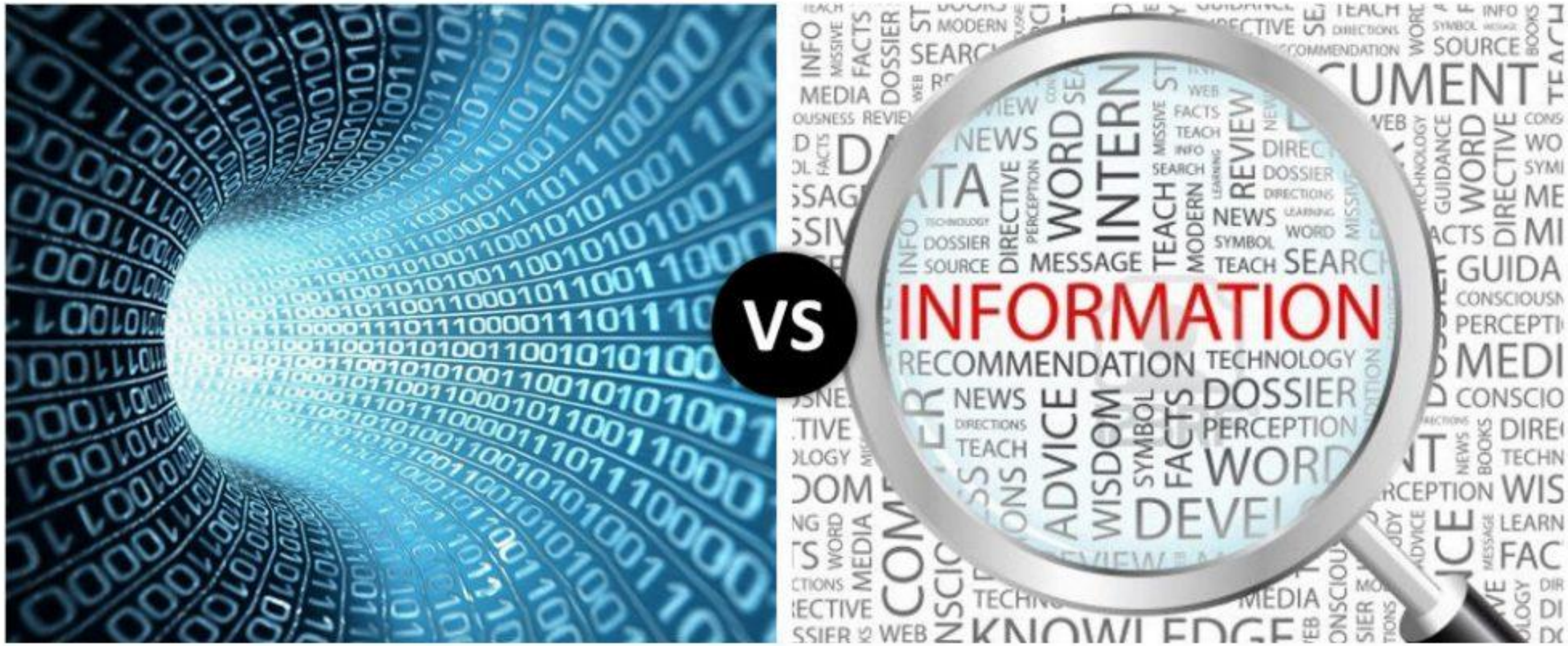
Universität
Münster

Introduction to Artificial Intelligence for Biologists

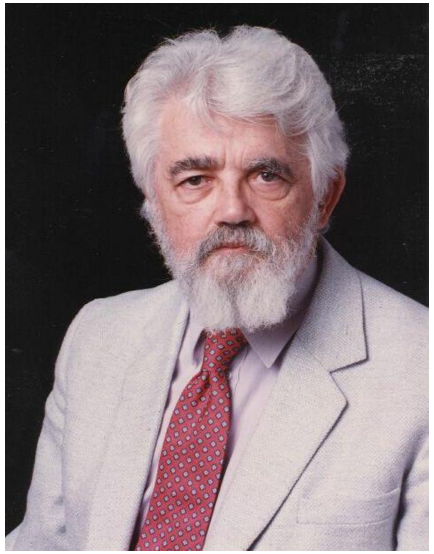
Xiaoyi Jiang
[1] Faculty of Mathematics and Computer Science
[2] Cells in Motion Interfaculty Centre (CiM)
University of Münster
Germany



Data Information

AI: “The art of creating machines that perform functions that require intelligence when performed by people”. [Kurzweil, 1990]
→ “automation of intelligent behavior” [Luger and Stubblefield, 1993]

1956: John McCarthy organized a conference and "The Dartmouth summer research project on artificial intelligence." The Dartmouth conference did bring together the founders in AI, and served to lay the groundwork for the future of AI research.



John McCarthy regarded as father of AI

IJCAI-19
August 10-16 2019, Macao, China

Home Attending Calls Grants Important Dates Program
Registration Sponsorship Committees Contact

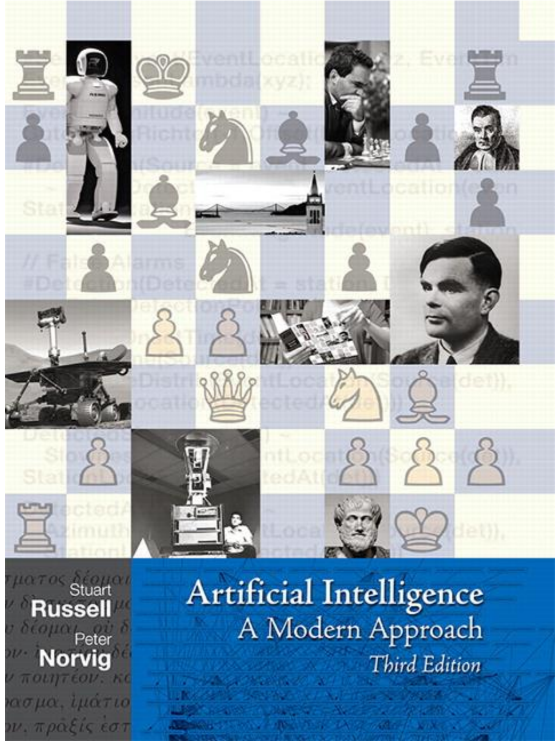
IJCAI
2019
MACAO

Topics

IJCAI-19 welcomes submissions across all areas of AI. The conference scope includes all subareas of AI, including (but not limited to) traditional topics such as machine learning, search, planning, knowledge representation, reasoning, constraint satisfaction, natural language processing, robotics and perception, and multiagent systems. We expressly encourage work that cuts across technical areas and/or integrated capabilities. We encourage all types of contributions including theoretical, engineering and applied. We also encourage papers on AI techniques in the context of novel application domains, such as security, sustainability, health care, transportation, and commerce.

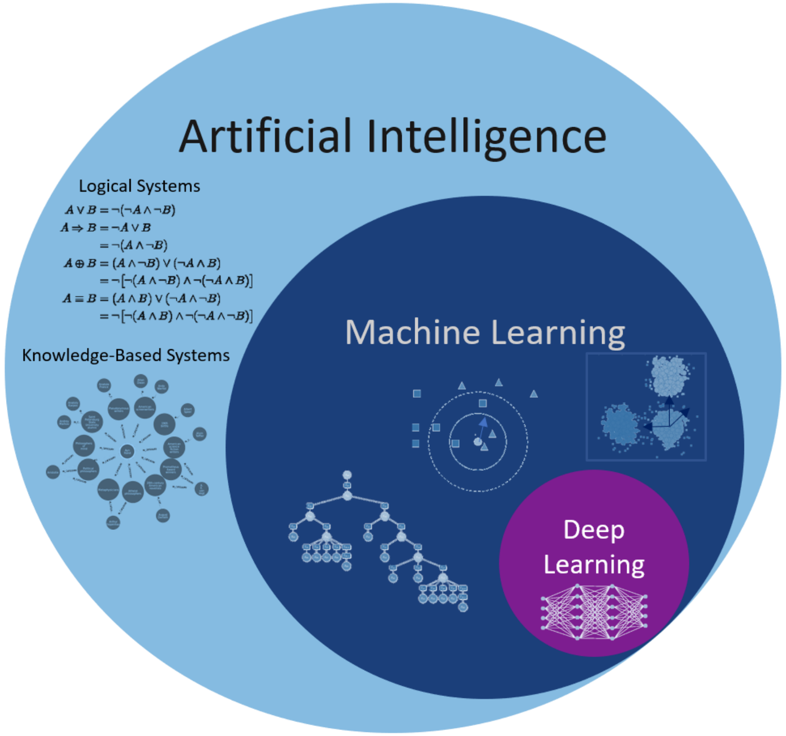
In addition there are two special tracks with a specific call for papers:

- Understanding Intelligence and Human-level AI in the New Machine Learning era
- AI for Improving Human-Well Being



AI >> Machine Learning (Deep Learning)

Media: AI = Machine Learning (Deep Learning)



Universität
Münster

Machine Learning for Biomedical Imaging

living.knowledge

Imaging:
Making the invisible visible

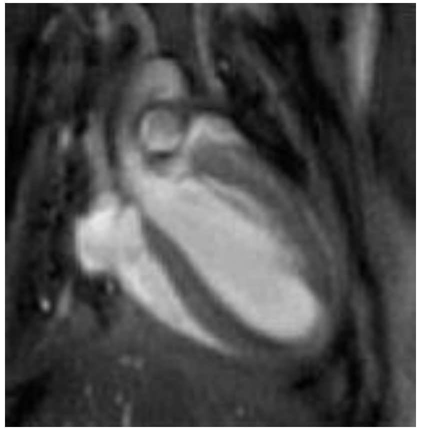
Wilhelm Conrad Röntgen (1845 – 1923)
First Nobel Prize in Physics

First medical X-ray (hand of Röntgen's wife Anna Bertha Ludwig), **1895**

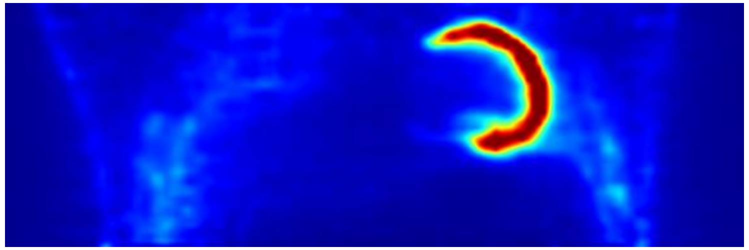
Biomedical Imaging

> 100 years

Radiograph

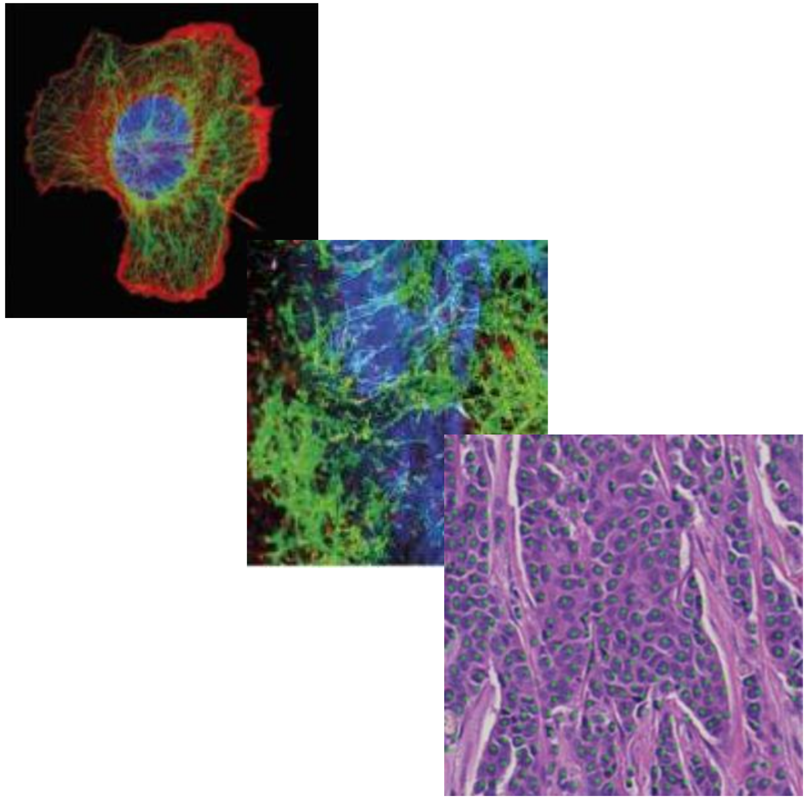


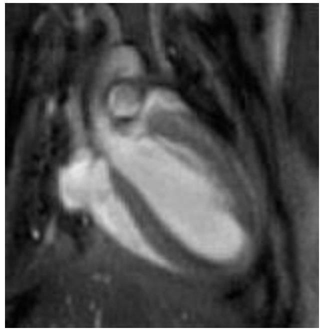
MRI

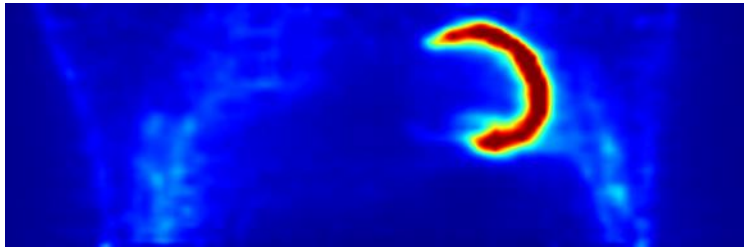


PET

Biomedical Imaging







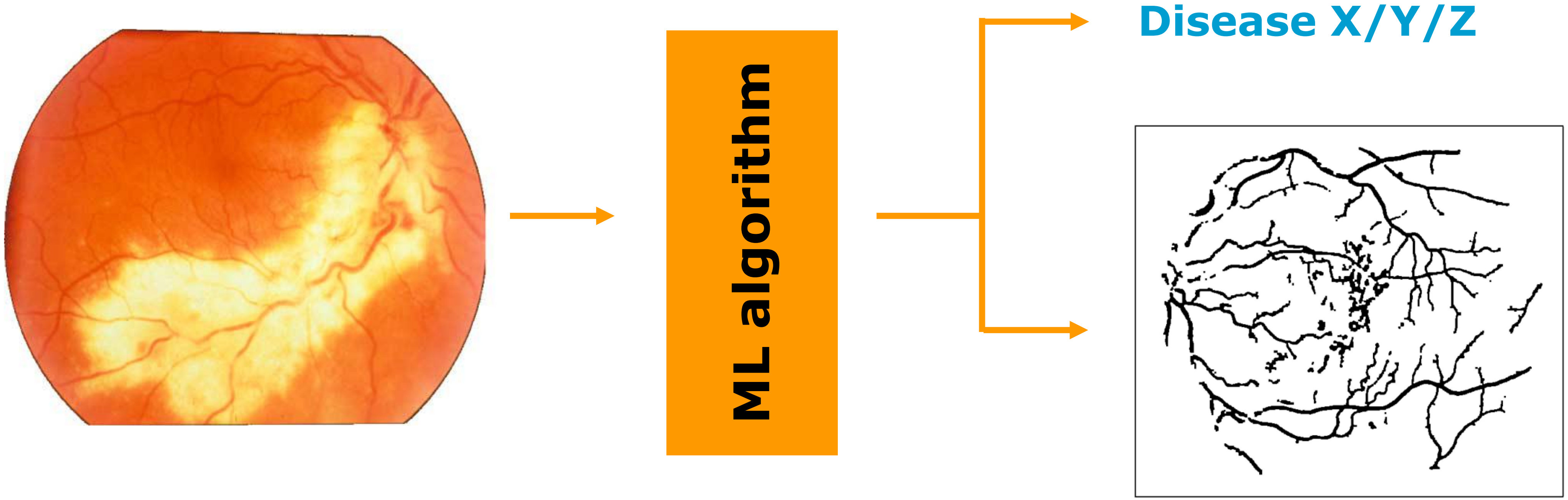
Optical Imaging | Intravital Imaging | Whole-Body Imaging

nm — Cell — µm — Tissue — mm — Organ — cm

Tasks: Knowledge extraction from image data

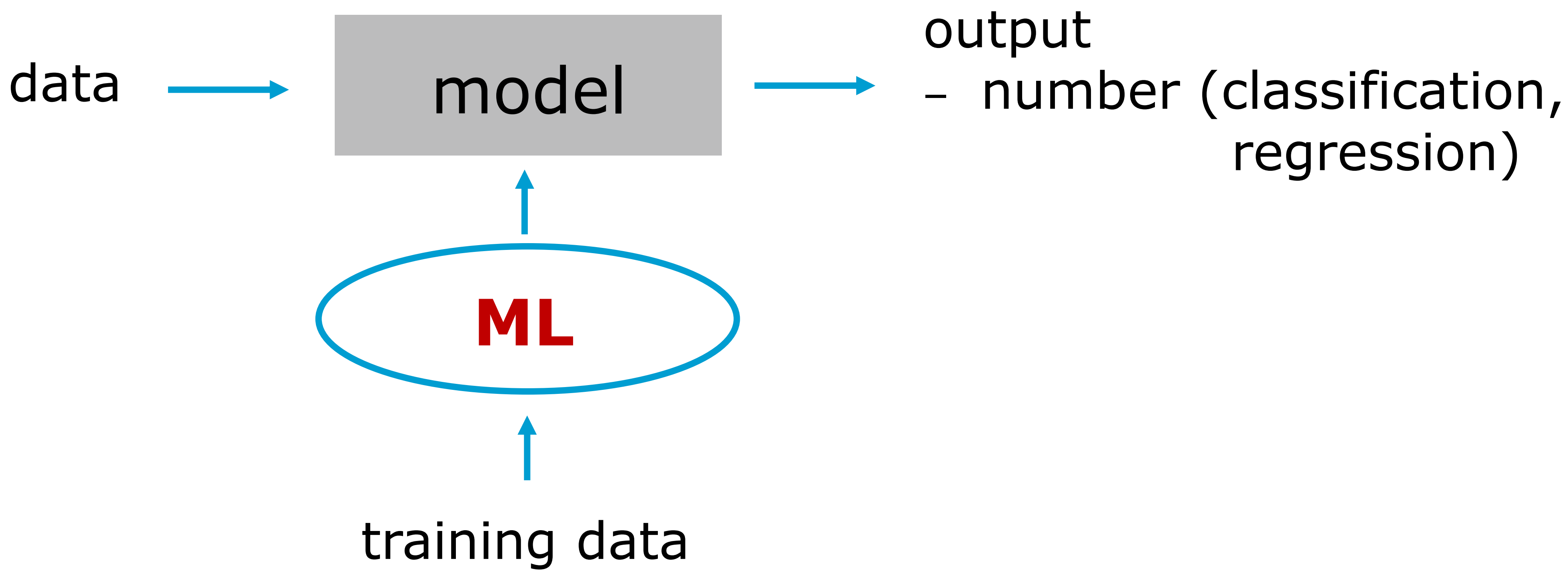
- Image enhancement
- Registration
- Motion analysis
- Segmentation (detection of anatomical structures)
- Quantification
- Decision-making (diagnosis)
- Therapy planning
- Computer-assisted surgery
- ………………

Machine Learning in Biomedical Imaging

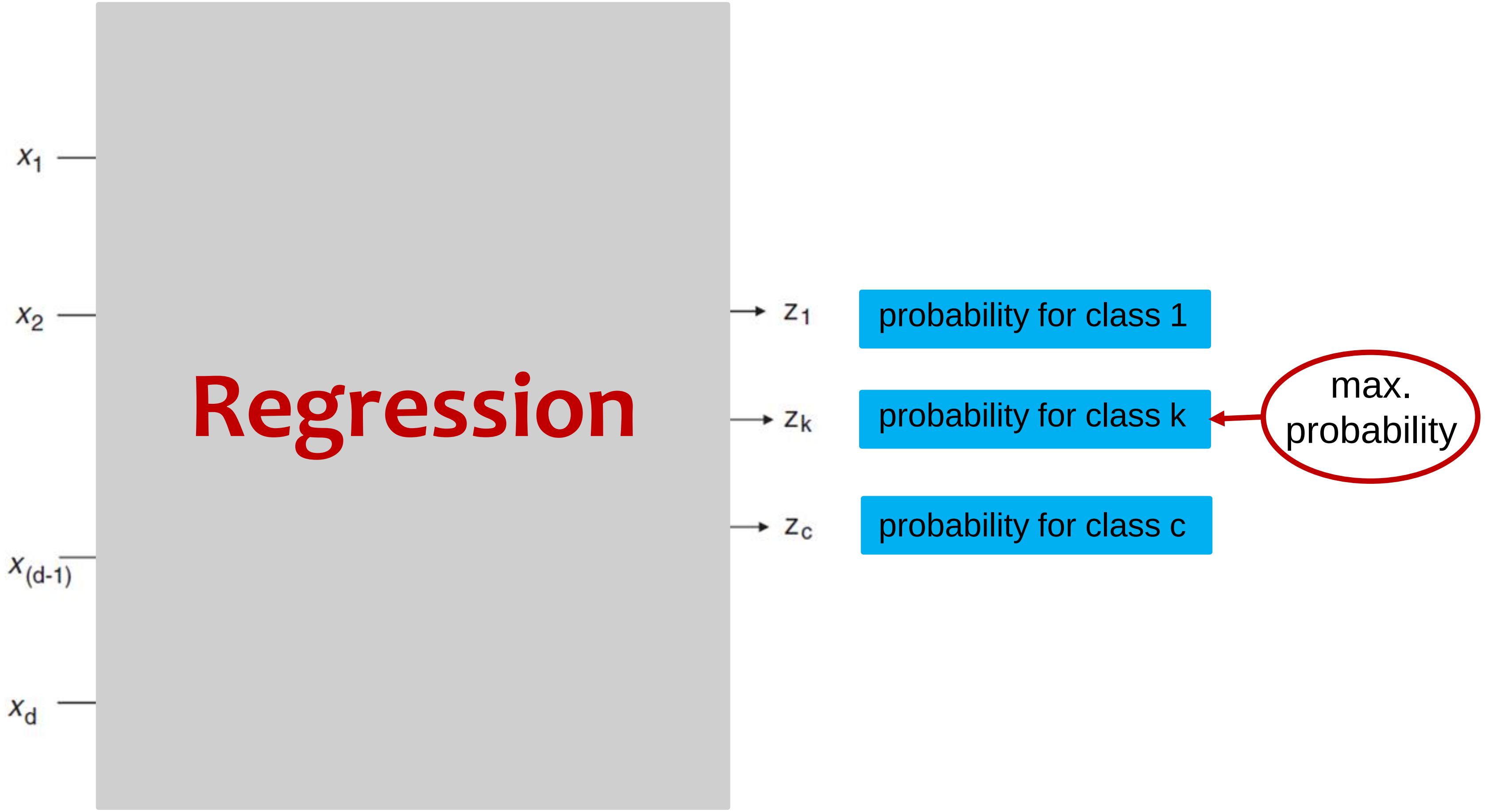


- **Classification:** Globally classify the whole image (e.g. diagnosis)
- **Segmentation:** Locally classify each pixel into vessel yes/no
- **Quantification:** e.g. ratio artery / vein (after classification artery vs. vein)

Machine Learning is about prediction



Learn a function: probability of class i = $F(\mathbf{x})$, i = 1, 2, ..., c



$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{x}$$

P(cat) = 0.9

P(dog) = 0.1

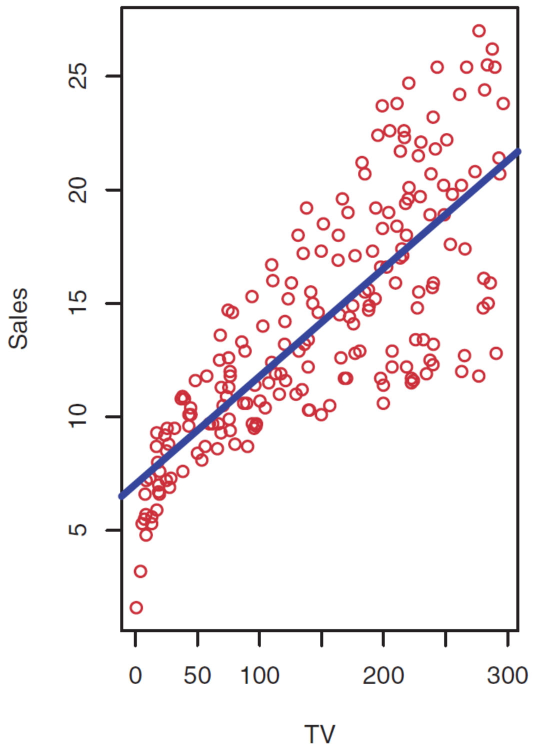
$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{x}$$

P(cat) = 0.2

P(dog) = 0.8

Classification is done by regression (computation of probabilities)!

Machine Learning: Linear Function



Input: N function values (t_1, t_2, …, t_N) at positions (x_1, x_2, …, x_N)

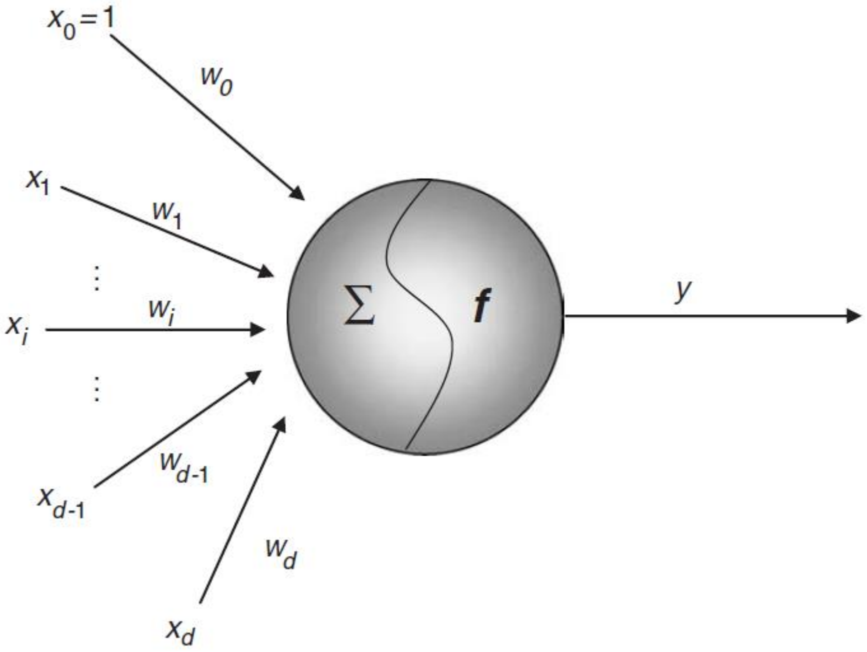
Output: determine (w_0, w_1) for $t = w_0 + w_1 x$

Linear dependence of sales on advertising budget:

$$sales = a \cdot TV + b = \boxed{F(TV)}$$

We need more complex function models!

Neural network as function approximator:



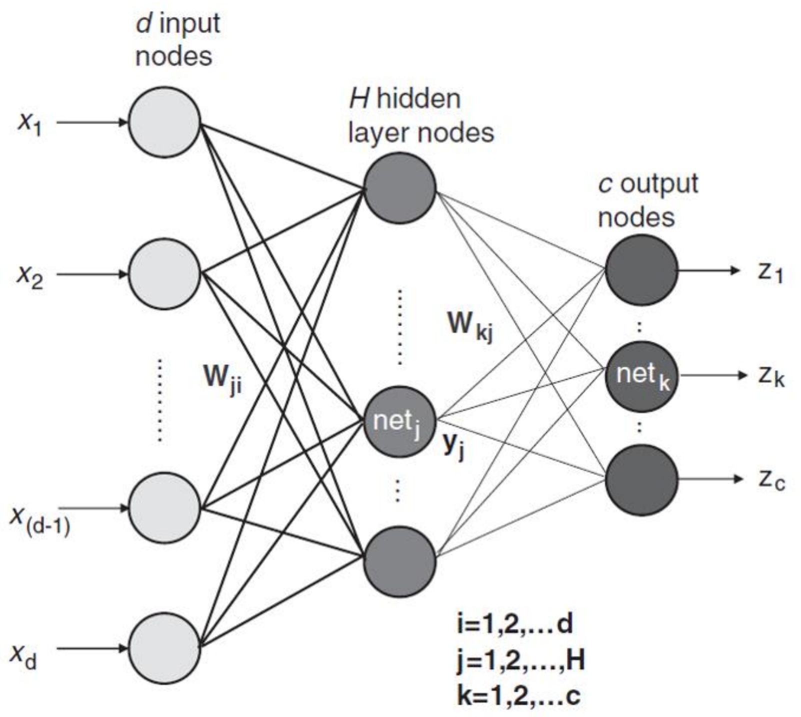
simple model of neuron

$$\begin{aligned} net &= \sum_{i=0}^{d} w_i x_i \\ &= w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_d x_d \end{aligned}$$

Activation function:

$$f(net) = \begin{cases} 1, & \text{if } net \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Multilayer feedforward neural network:



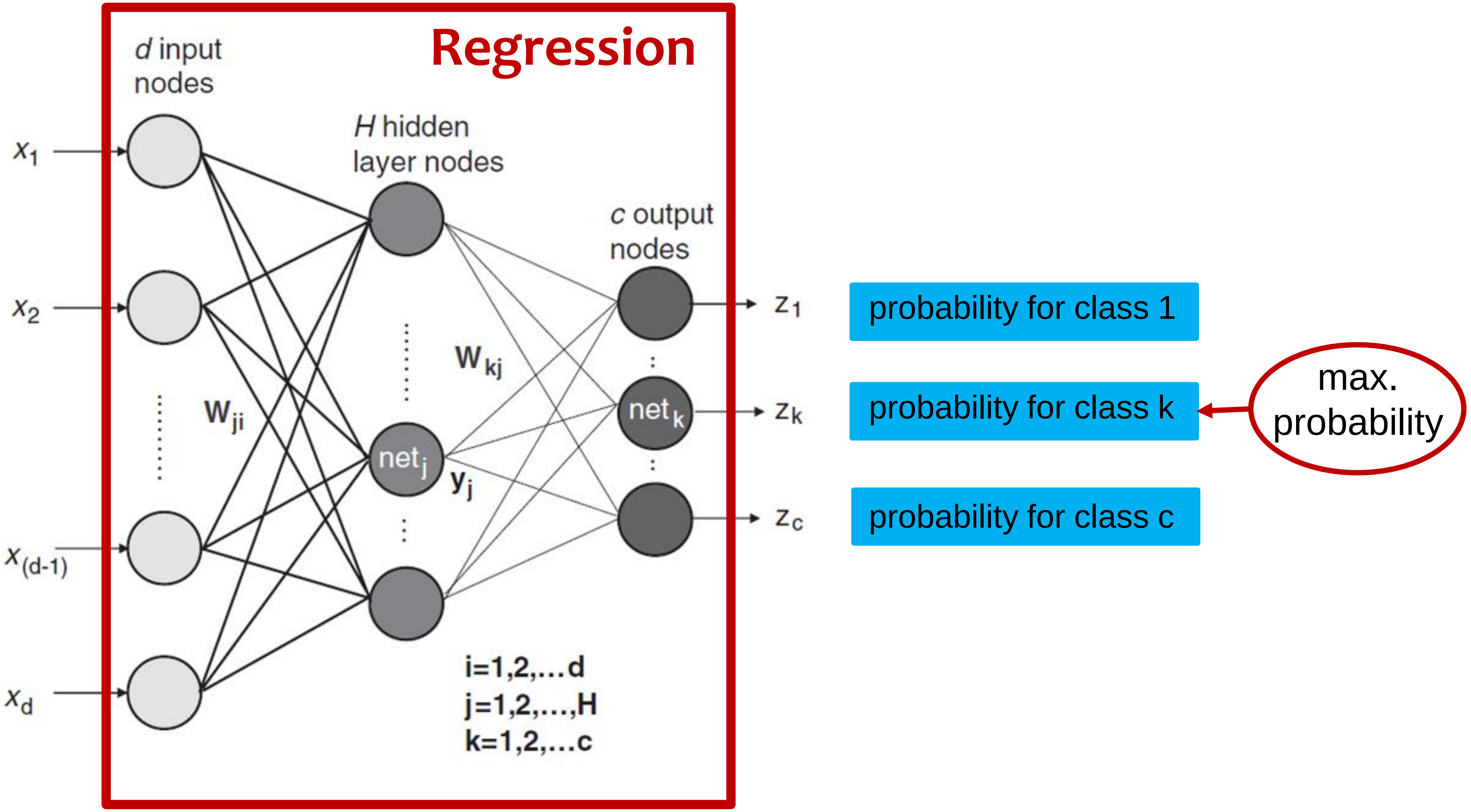
$$z_k = f\left(\sum_{j=1}^{H} w_{kj} \cdot f\left(\sum_{i=1}^{d} w_{ji} \cdot x_i\right)\right)$$

output parameters input

Theorem (multilayer networks are universal approximators): Multilayer networks with one hidden layer can approximate any function to any desired degree of accuracy.

Machine Learning: Neural Network

Learn a function: probability of class i = $F(\mathbf{x})$, i = 1, 2, …, c

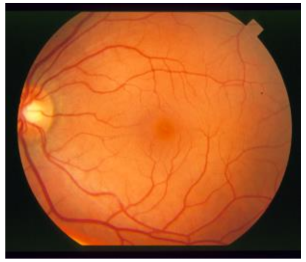


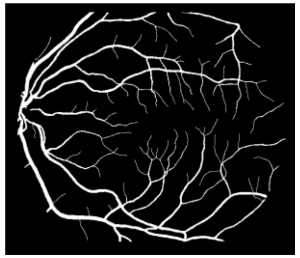
Machine Learning: Neural Network

Learning models from data for prediction

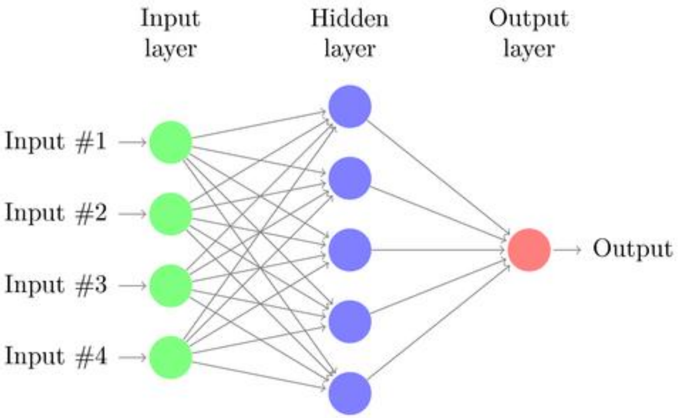
data → model → result

ML → model





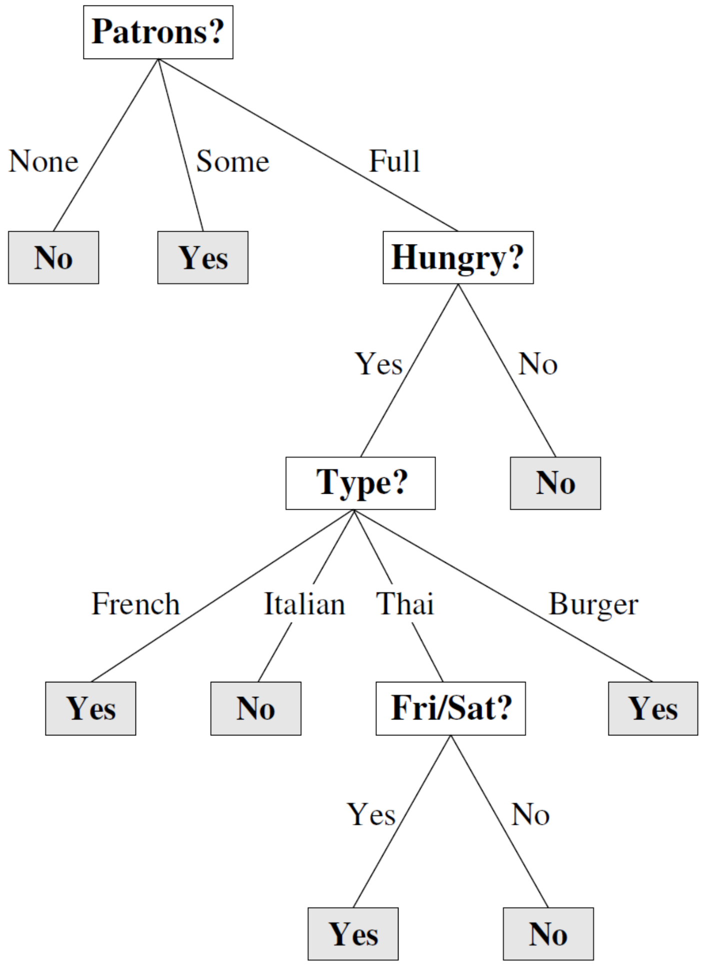
ML algorithm



Machine Learning: Tree-based Decision Models

Learning from tabular data towards decision trees

Example	Attributes										Goal
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	Wait
X_1	Yes	No	No	Yes	Some	\$\$\$	No	Yes	French	0-10	Yes
X_2	Yes	No	No	Yes	Full	\$	No	No	Thai	30-60	No
X_3	No	Yes	No	No	Some	\$	No	No	Burger	0-10	Yes
X_4	Yes	No	Yes	Yes	Full	\$	No	No	Thai	10-30	Yes
X_5	Yes	No	Yes	No	Full	\$\$\$	No	Yes	French	>60	No
X_6	No	Yes	No	Yes	Some	\$\$	Yes	Yes	Italian	0-10	Yes
X_7	No	Yes	No	No	None	\$	Yes	No	Burger	0-10	No
X_8	No	No	No	Yes	Some	\$\$	Yes	Yes	Thai	0-10	Yes
X_9	No	Yes	Yes	No	Full	\$	Yes	No	Burger	>60	No
X_{10}	Yes	Yes	Yes	Yes	Full	\$\$\$	No	Yes	Italian	10-30	No
X_{11}	No	No	No	No	None	\$	No	No	Thai	0-10	No
X_{12}	Yes	Yes	Yes	Yes	Full	\$	No	No	Burger	30-60	Yes

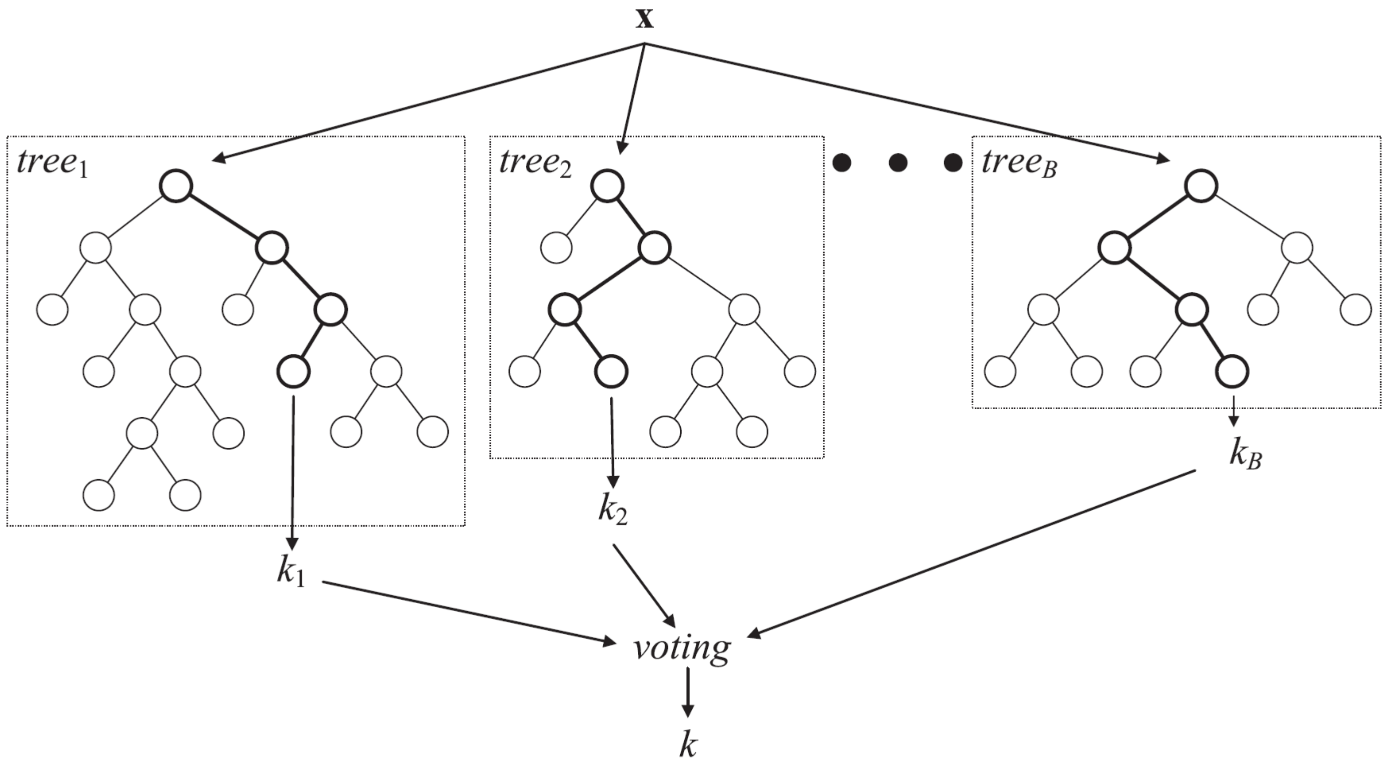


Universität Münster

Machine Learning: Tree-based Decision Models

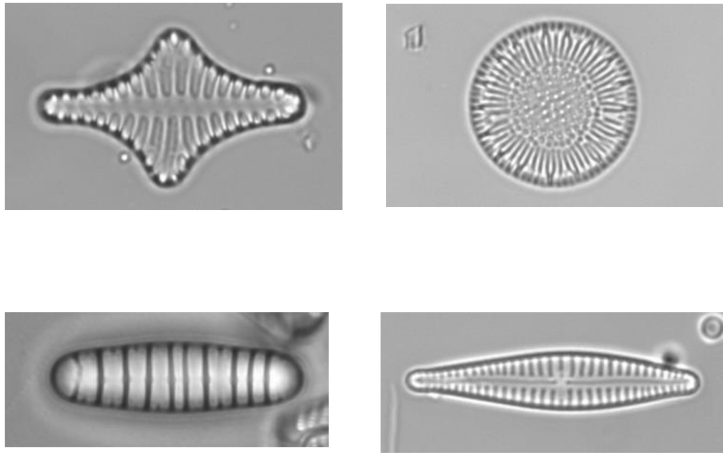
Machine Learning: Tree-based Decision Models

Powerful ensemble tree models
(bagging, random forest, gradient boosted trees)



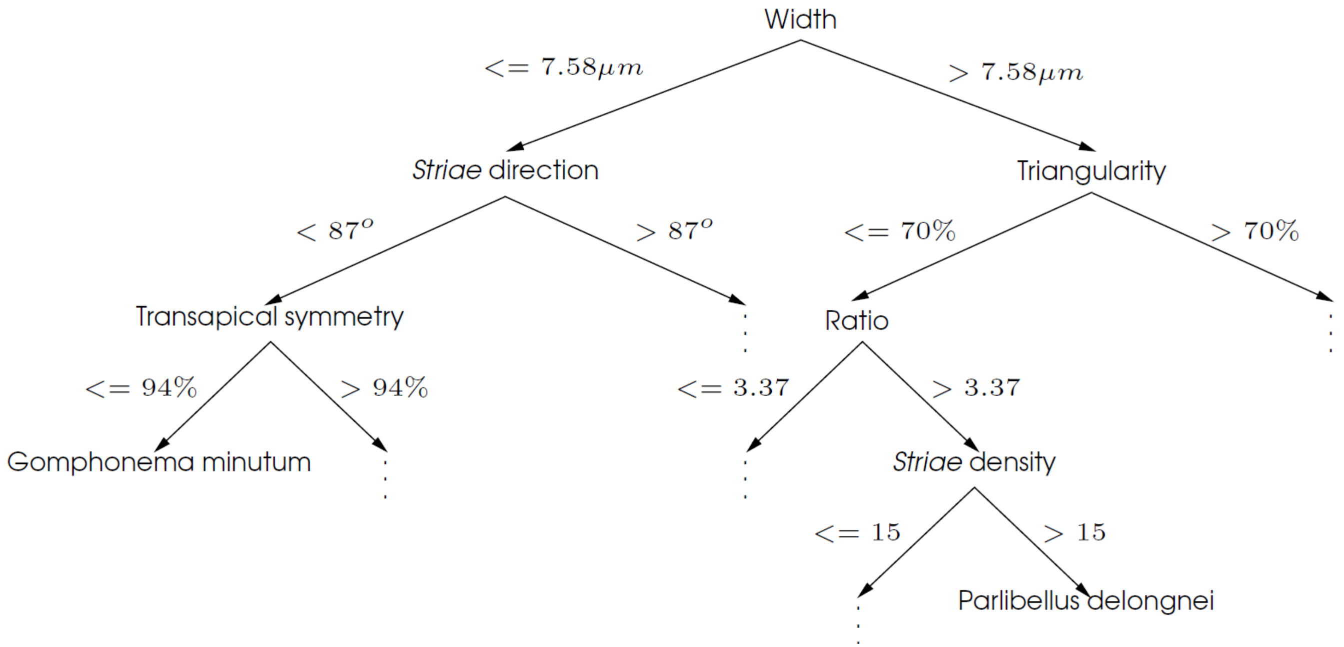
Machine Learning: Tree-based Decision Models

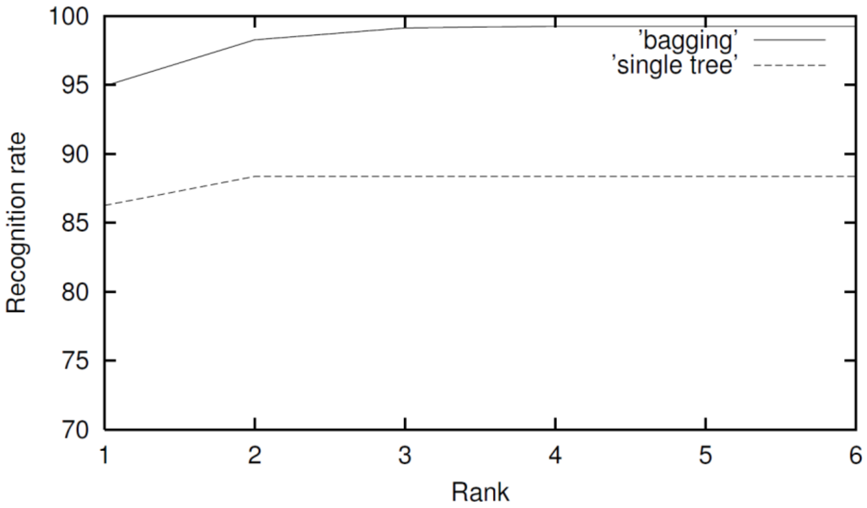
Diatom classification: Application in biology, climate research, forensic medicine



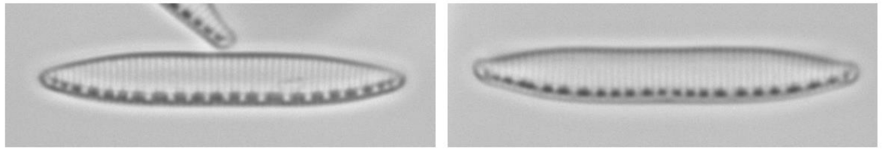
Group	Feature
Symmetry	class of symmetry (1)
Shape descriptors	rectangularity, circularity, ellipticity, triangularity compactness (5)
Shape properties	global shape properties (1) shape of the end points (10)
Geometric properties	length, width, length/width-ratio, size (4)
Diatom specific features	striae density, direction, changeover point (3) axial area width (1) costae density (1) horizontal frequency (1)
Other features	moment invariants (7) Fourier descriptors (126)

Machine Learning: Tree-based Decision Models





From the root maximal 11 tests; this maximal number is reached by two very similar patterns



$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{x}$$

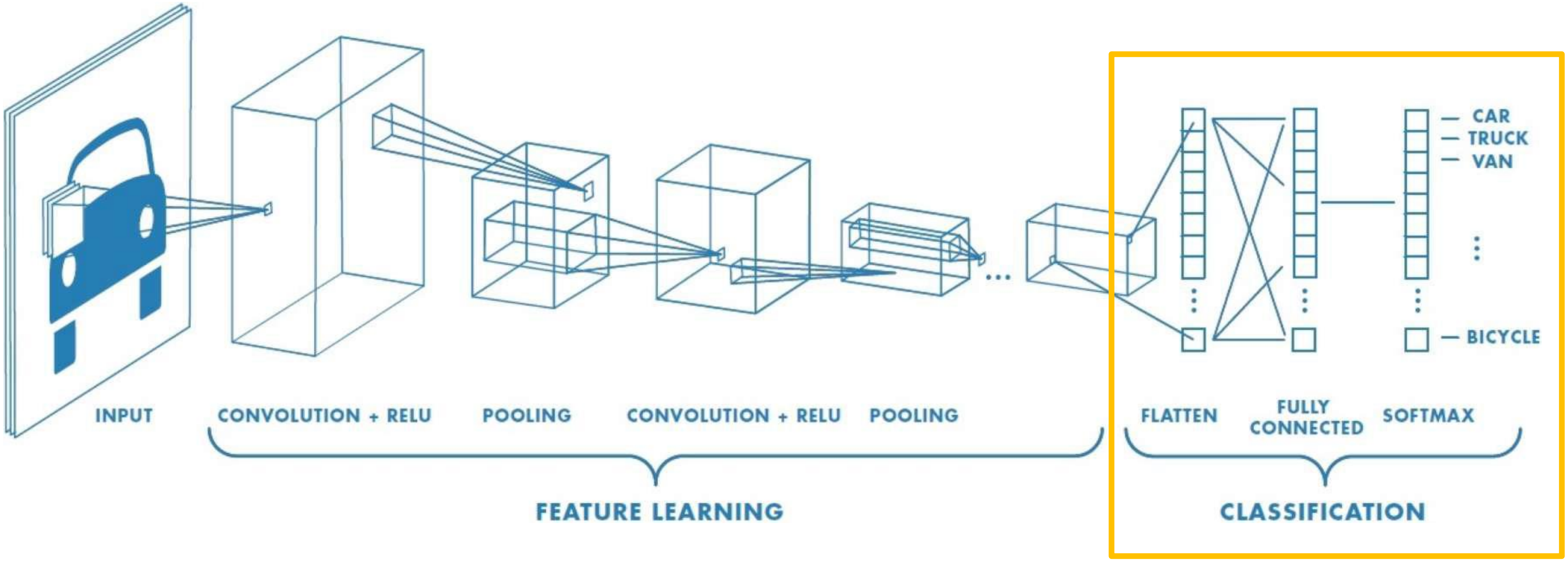
?

cat

Features are more important than classifiers!

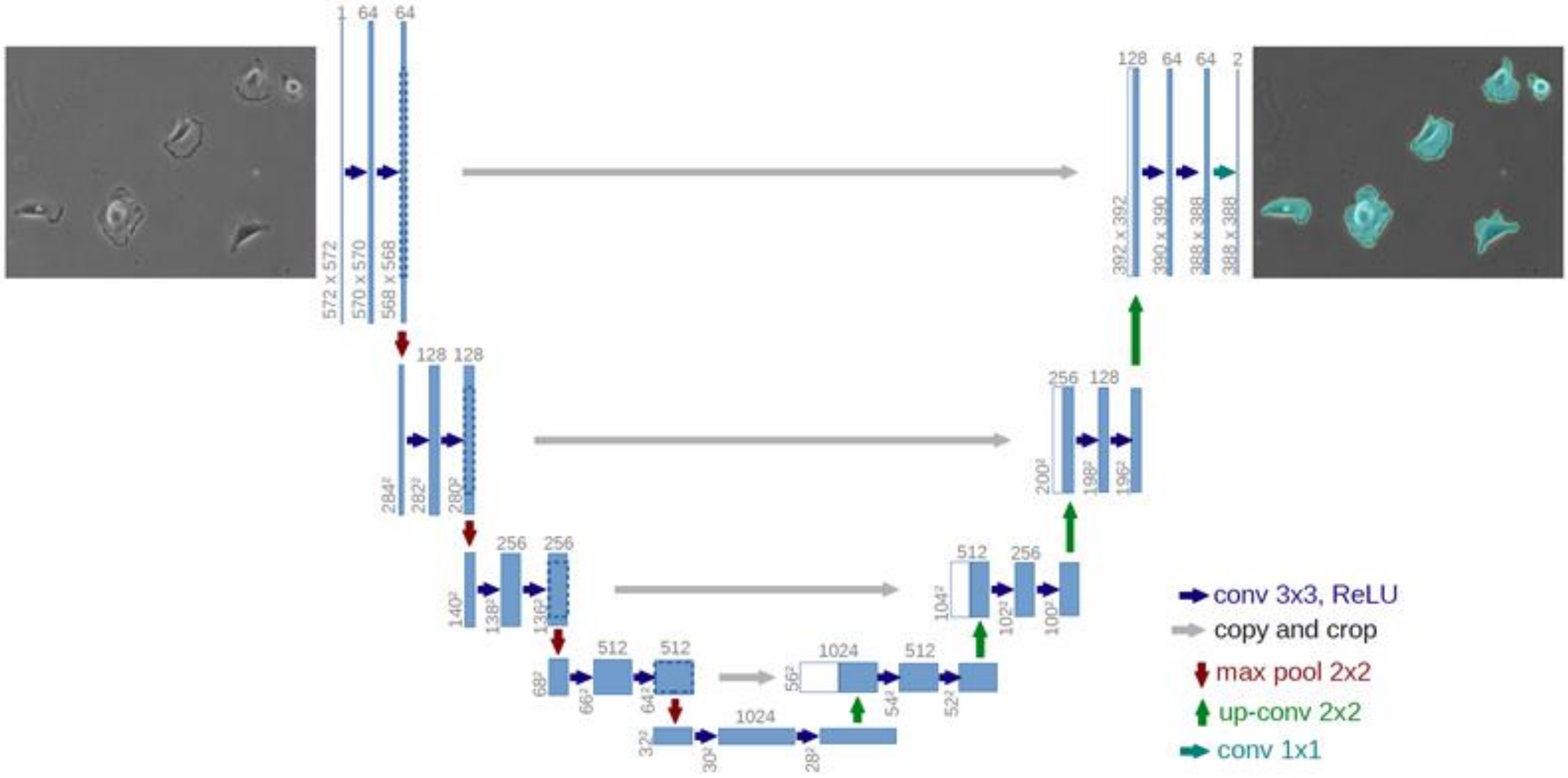
- Manual design by engineers
- **Automatic feature learning from data (Deep Learning)**

Deep Learning



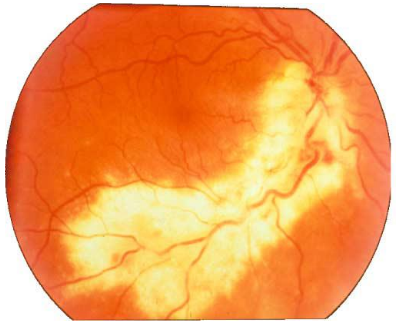
Classification: The second part of this architecture is often fully-connected neural network. After CNN training this classification part could be replaced by any classifier; such a replacement may even result in improved classification performance.

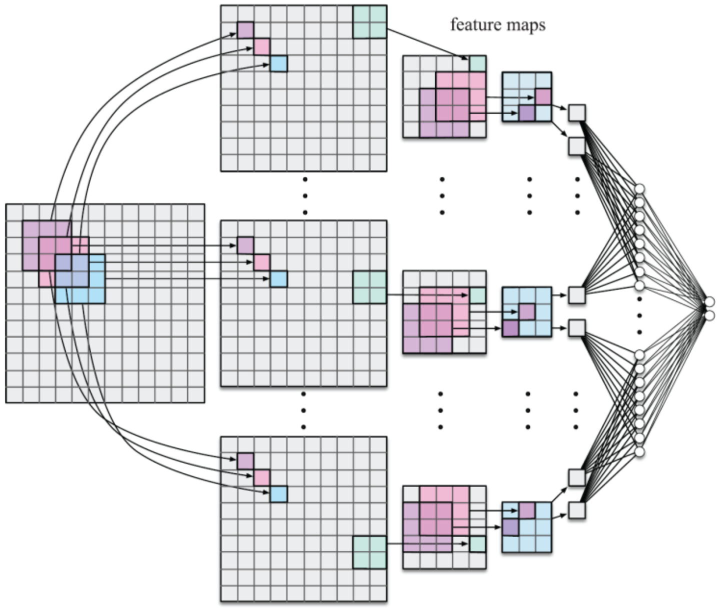
Deep Learning: Cell Segmentation



Ronneberger, 2015

Deep Learning: Vessel Segmentation





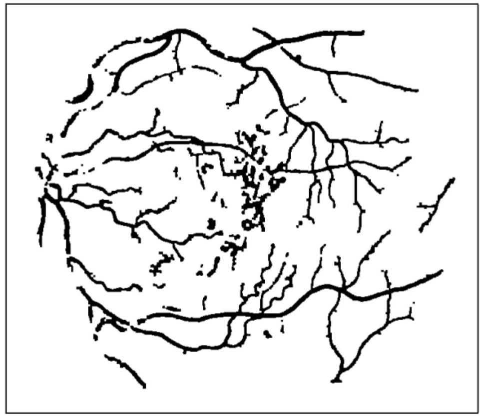
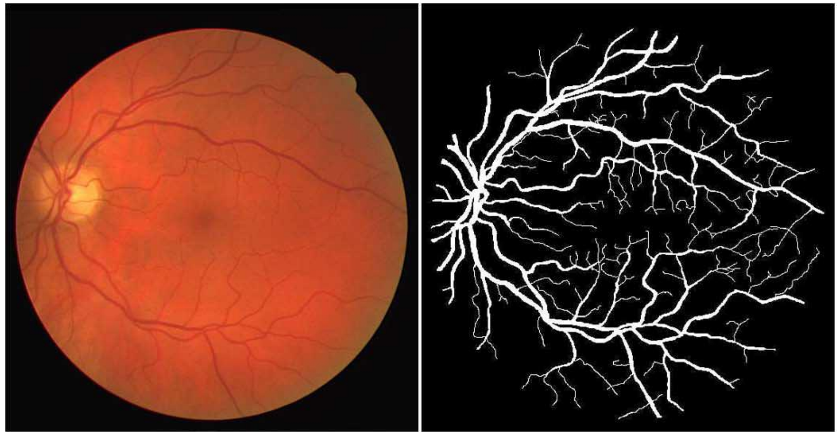


TABLE III
PERFORMANCE OF MODELS (POINT ESTIMATES FOR DRIVE, AVERAGES WITH .95 CONFIDENCE INTERVALS FOR STARE)

	DRIVE						STARE					
	AUC	Acc	Acc*	Kappa	Sens	Spec	AUC	Acc	Acc*	Kappa	Sens	Spec
PLAIN	.9683	.9479	.9473	.7653	.7417	.9804	.9767±.0053	.9559±.0071	.9551±.0072	.7477±.0451	.7495±.0721	.9788±.0081
GCN	.9708	.9487	.9475	.7708	.7550	.9792	.9787±.0049	.9571±.0064	.9572±.0064	.7573±.0394	.7620±.0656	.9789±.0072
ZCA	.9719	.9485	.9472	.7756	.7819	.9748	.9783±.0062	.9563±.0064	.9562±.0066	.7598±.0317	.7718±.0490	.9783±.0055
AUGMENT	.9663	.9466	.9453	.7610	.7447	.9784	.9744±.0048	.9527±.0068	.9512±.0069	.7306±.0431	.7376±.0720	.9769±.0086
BALANCED	.9738	.9230	.9251	.7193	.9160	.9241	.9820±.0045	.9309±.0107	.9620±.0051	.7021±.0305	.9307±.0274	.9304±.0133
NO-POOL	.9720	.9495	.9486	.7781	.7763	.9768	.9785±.0066	.9566±.0082	.9568±.0081	.7622±.0415	.7867±.0698	.9754±.0099

Liskowski, 2016

Deep Learning: Vessel Segmentation

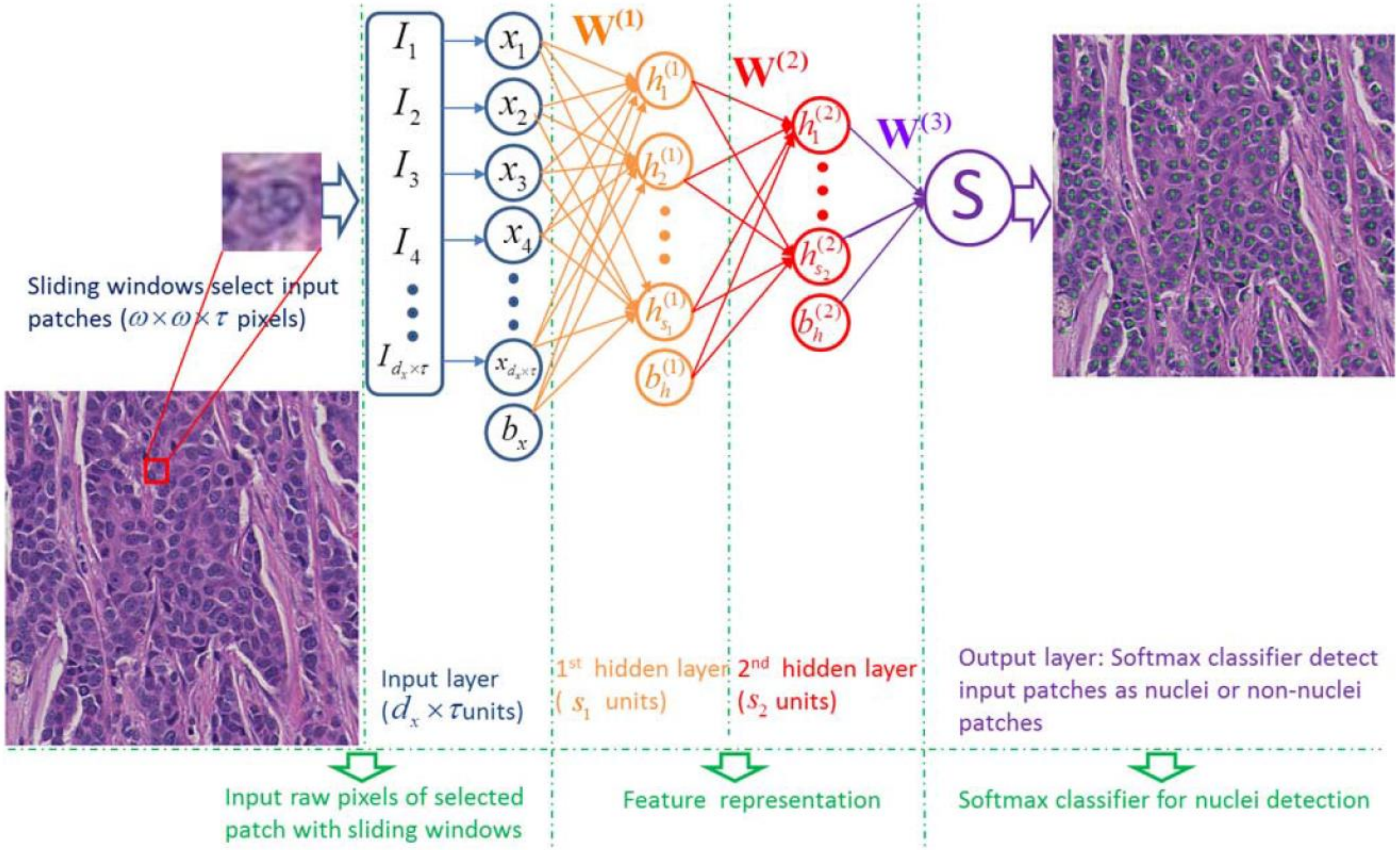


2D retinal image + ground truth (STARE, DRIVE, etc.)

3D OCT angiography

Deep Learning: Nuclei Detection

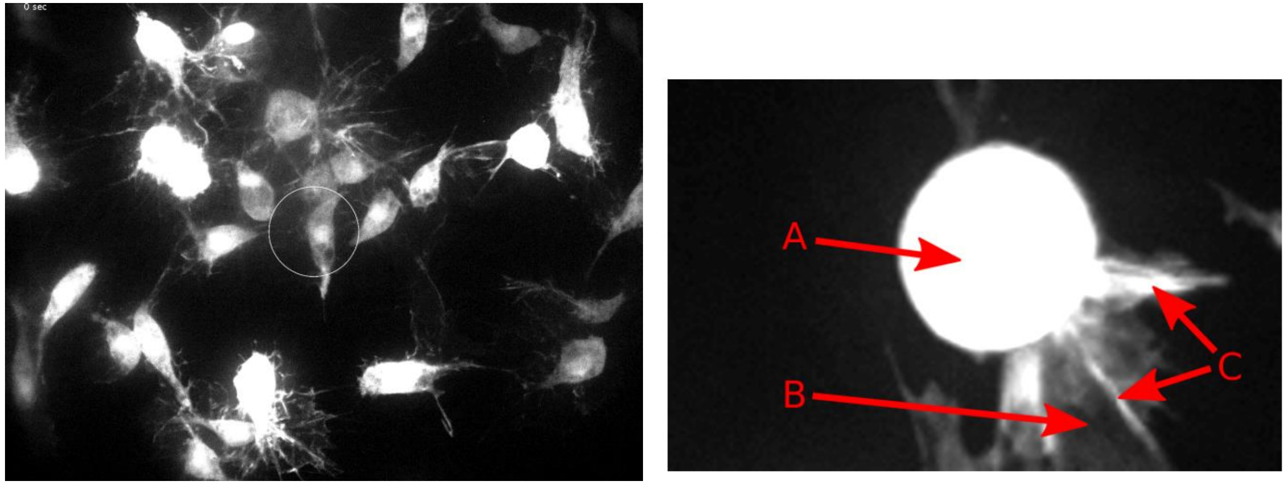
Nuclei detection in breast cancer histopathology:



Xu, 2016

Deep Learning: Image Segmentation

Segmentation of Drosophila macrophages in in-vivo fluorescence microscopy images:



(a) (b)

Fig. 1: Examples of the input images. **(a)** SDCM image of *in-vivo* wildtype Drosophila macrophages which have been marked using GFP. **(b)** Magnified illustration of a single cell. Brightness and contrast have been enhanced for better clarity. Arrows exemplarily mark the three foreground classes relevant for the segmentation: (A) cell body, (B) lamellipodium, and (C) filopodia.

Deep Learning: Image Segmentation

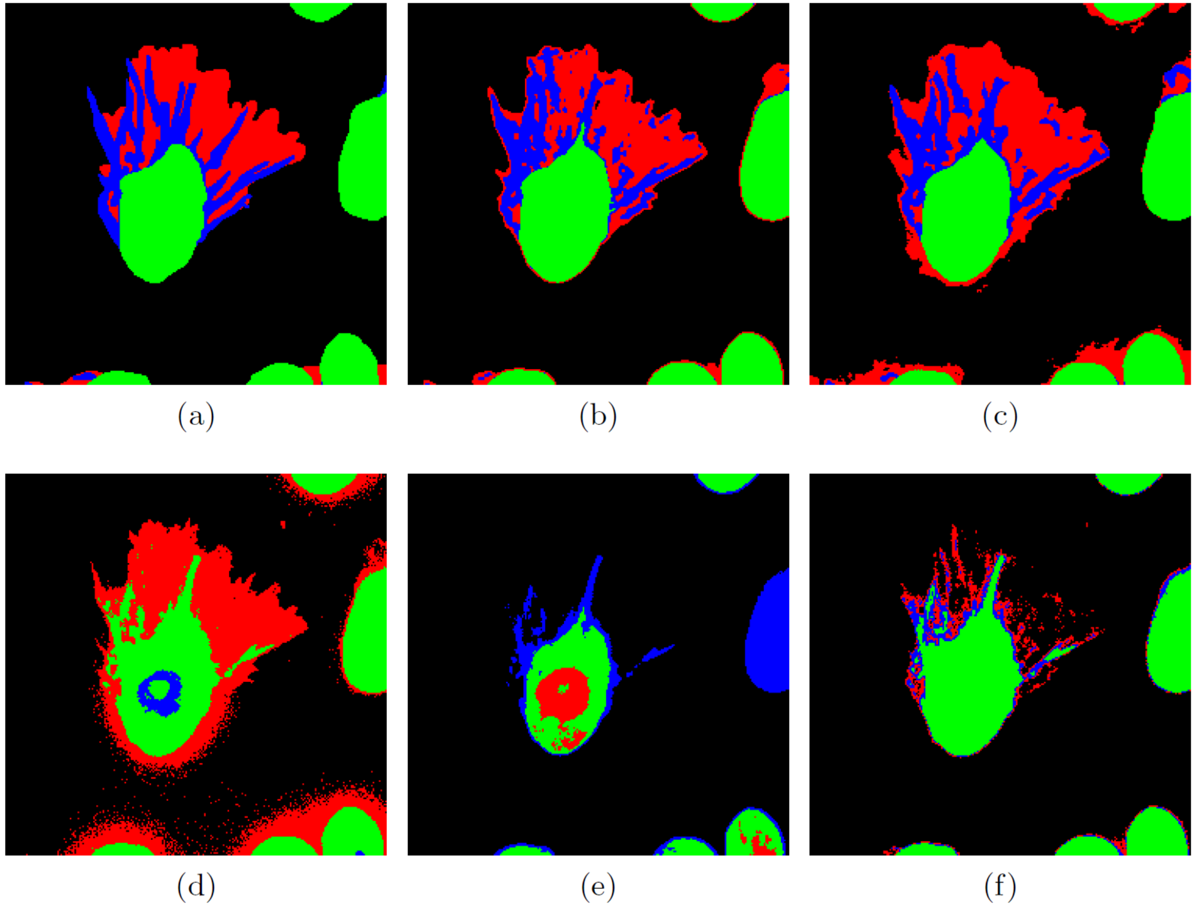
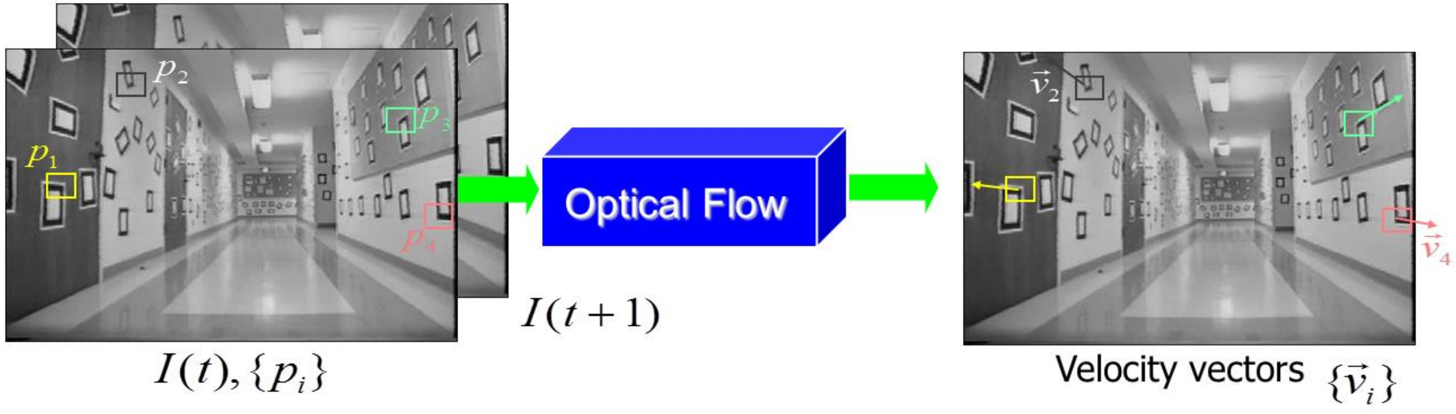
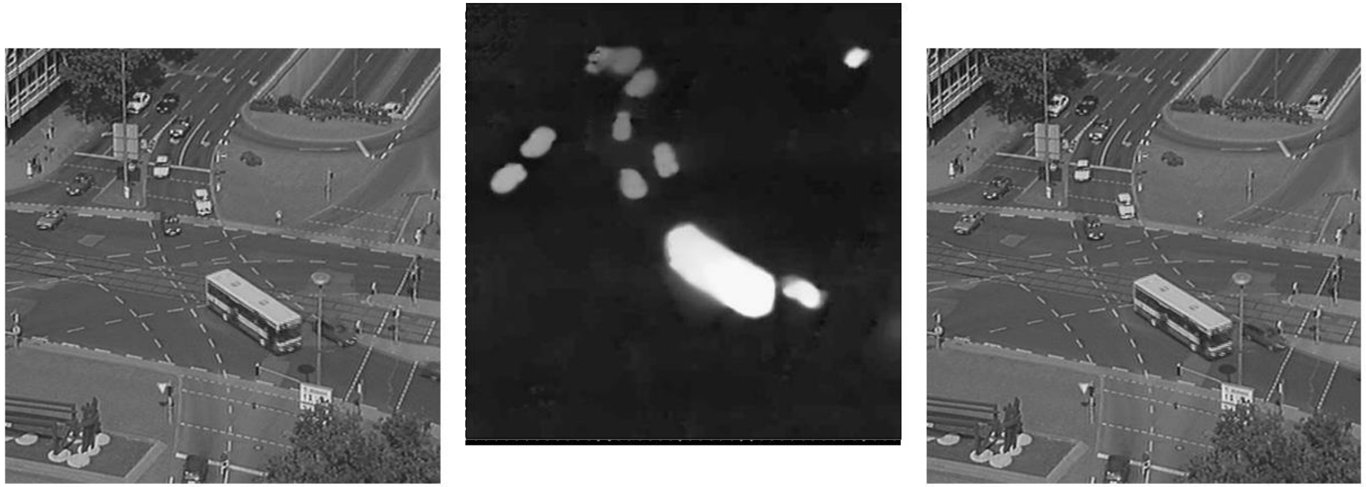


Fig. 4: Exemplary results for a single image tile from the validation set. **(a)** Ground truth labeling. **(b)** U-net with F_1-measure loss function. **(c)** U-net with weighted cross entropy. **(d)** GMM. **(e)** K-means. **(f)** Otsu.

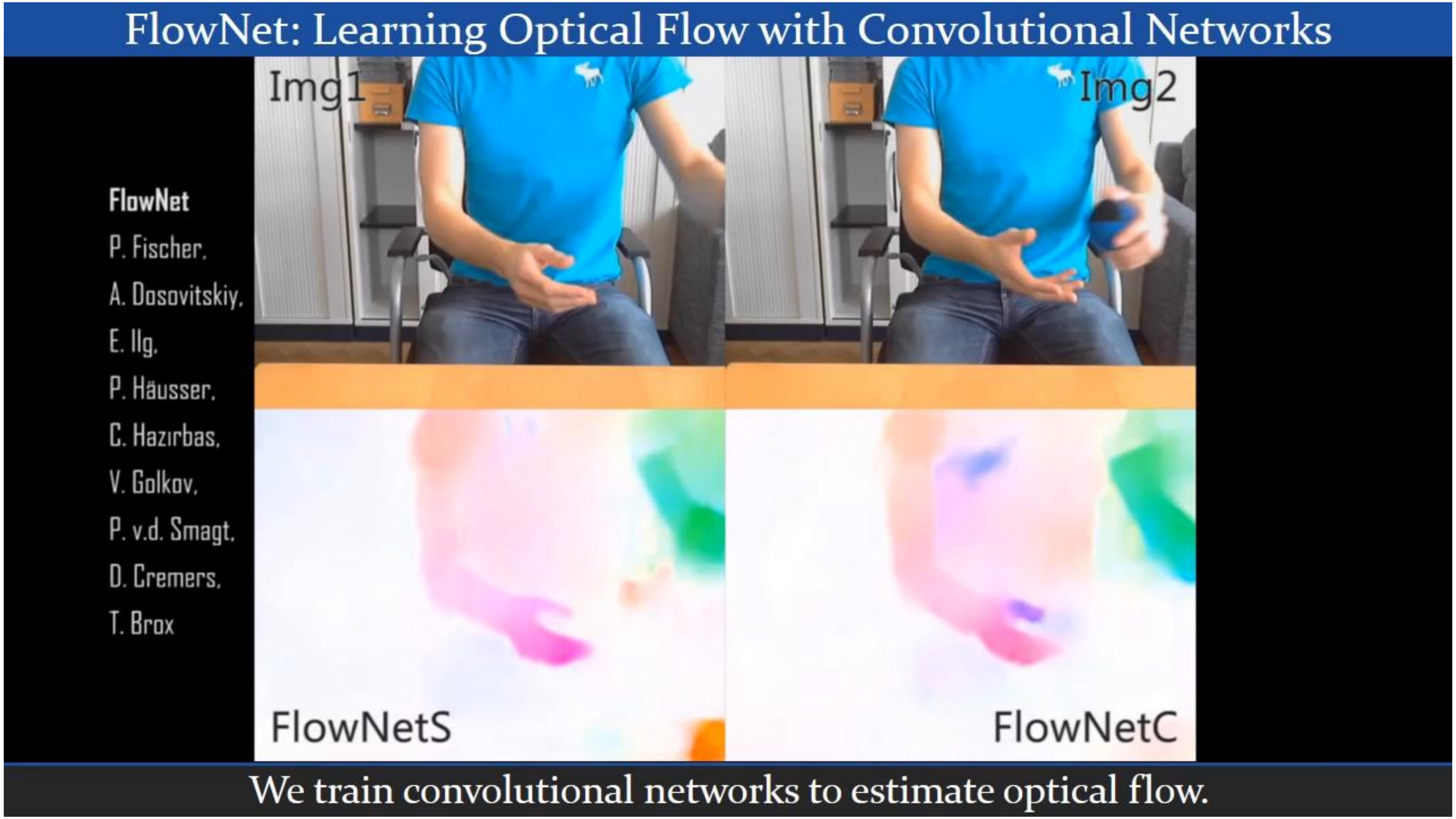
Deep Learning: Optical Flow Computation

How can we describe the motion between the two frames?





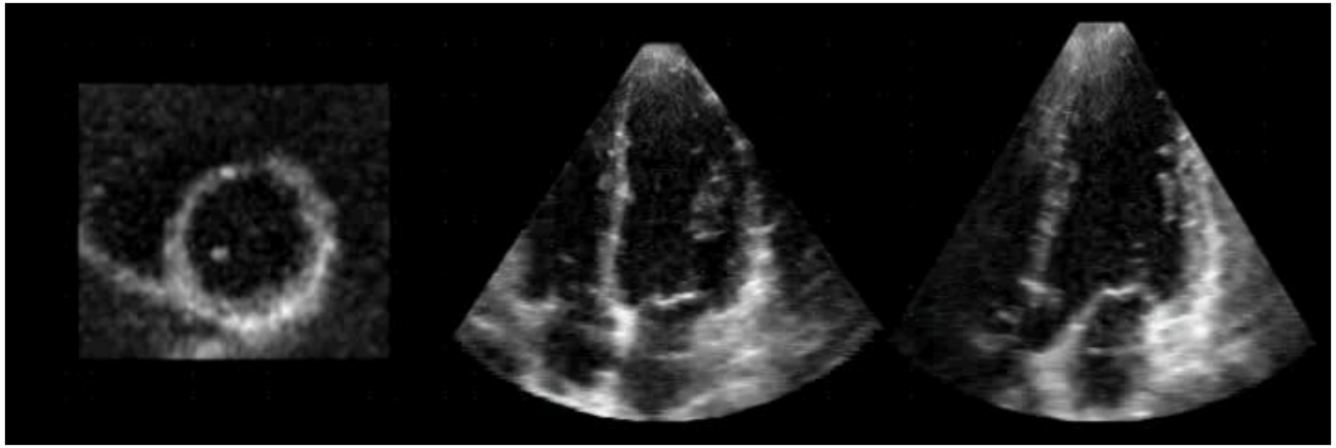
Deep Learning: Optical Flow Computation



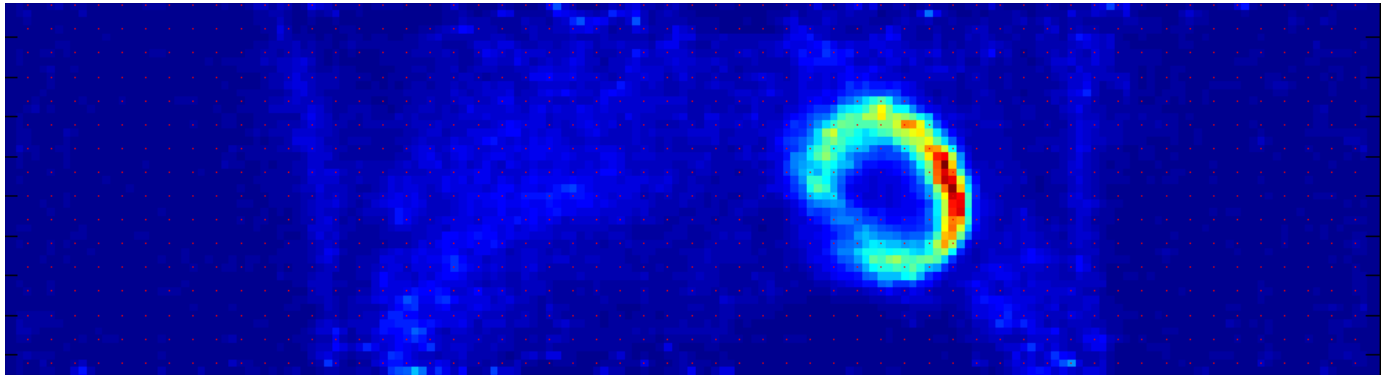
Dosovitskiy, 2015

Universität Münster

Deep Learning: Optical Flow Computation



Ultrasound: Heart movement, 2D optical flow vectors



PET: Heart movement, 3D optical flow vectors

Deep Learning: Diagnosis

Turing Award laureate 2018

"We should stop training radiologists right now, in 5 years deep learning will have better performance" (2016)

Barriers slow AI/DL's move to the mainstream in medical imaging

- Many technical challenges
- Integration of AI/DL tools into medical workflows
- Challenging regulatory process and legal issues

Deep Learning: Diagnosis

nature

Article | Published: 01 January 2020

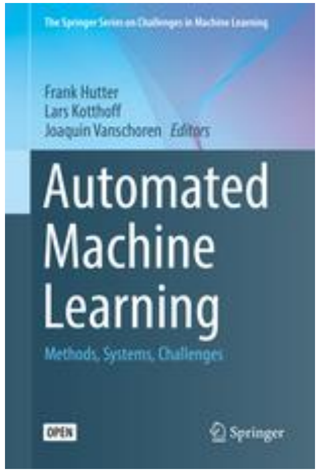
International evaluation of an AI system for breast cancer screening

Scott Mayer McKinney, Marcin Sieniek, [...] Shravya Shetty

Nature **577**, 89–94(2020) | Cite this article

Screening mammography aims to identify breast cancer at earlier stages of the disease, when treatment can be more successful[1]. Despite the existence of screening programmes worldwide, the interpretation of mammograms is affected by high rates of false positives and false negatives[2]. Here we present an artificial intelligence (AI) system that is capable of surpassing human experts in breast cancer prediction. To assess its performance in the clinical setting, we curated a large representative dataset from the UK and a large enriched dataset from the USA. We show an absolute reduction of 5.7% and 1.2% (USA and UK) in false positives and 9.4% and 2.7% in false negatives. We provide evidence of the ability of the system to generalize from the UK to the USA. In an independent study of six radiologists, the AI system outperformed all of the human readers: the area under the receiver operating characteristic curve (AUC-ROC) for the AI system was greater than the AUC-ROC for the average radiologist by an absolute margin of 11.5%. We ran a simulation in which the AI system participated in the double-reading process that is used in the UK, and found that the AI system maintained non-inferior performance and reduced the workload of the second reader by 88%. This robust assessment of the AI system paves the way for clinical trials to improve the accuracy and efficiency of breast cancer screening.

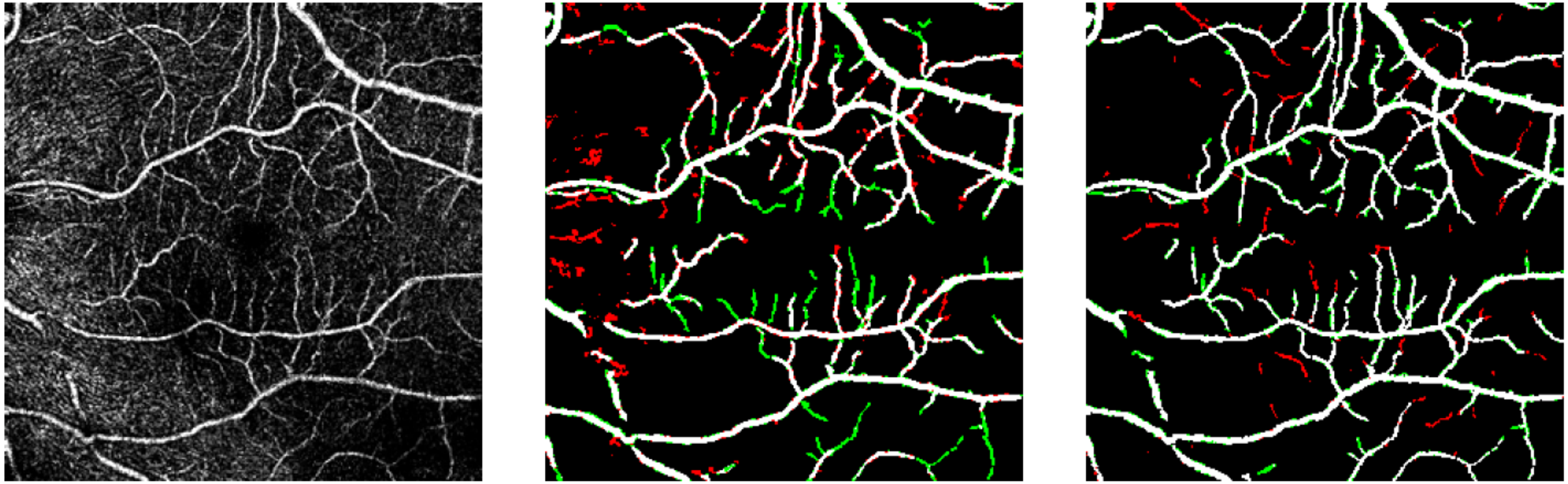
AutoML



Automated machine learning (AutoML): make the numerous decisions of system design in an objective and automated way → ***democratization*** of machine learning: with AutoML, customized state-of-the-art machine learning is at everyone's fingertips

	Input Data			Data Preprocessing						
	Spreadsheets	Images, Text	Others (2*)	Removing Duplicates	Datatype Detection	Numerical	Categorical	Datetime	Time-Series	Others
AutoGluon	●	●	●	●	●	●	●	●		
Ludwig	●	●	●		●	●	●	●	●	●
TransmogrifAI	●				●	●	●	●	●	●
H2O-AutoML	●			●	●	●	●	●	●	
Auto-Keras	●	●		?		●	●(3*)			
Auto_ML	●			?		●	●			
Auto-Weka	●			●		●	●			
Auto-sklearn	●					●	●			
TPOT	●					●	●(3*)			
Google AutoML(*)	●	●	●	●	●	●	●	●	●	
Azure ML(*)	●	●		●	●	●	●	●	●	
Darwin (*)	●				●	●	●	●	●	
H20-Driverless AI (*)	●	●	●	●	●	●	●	●	●	●

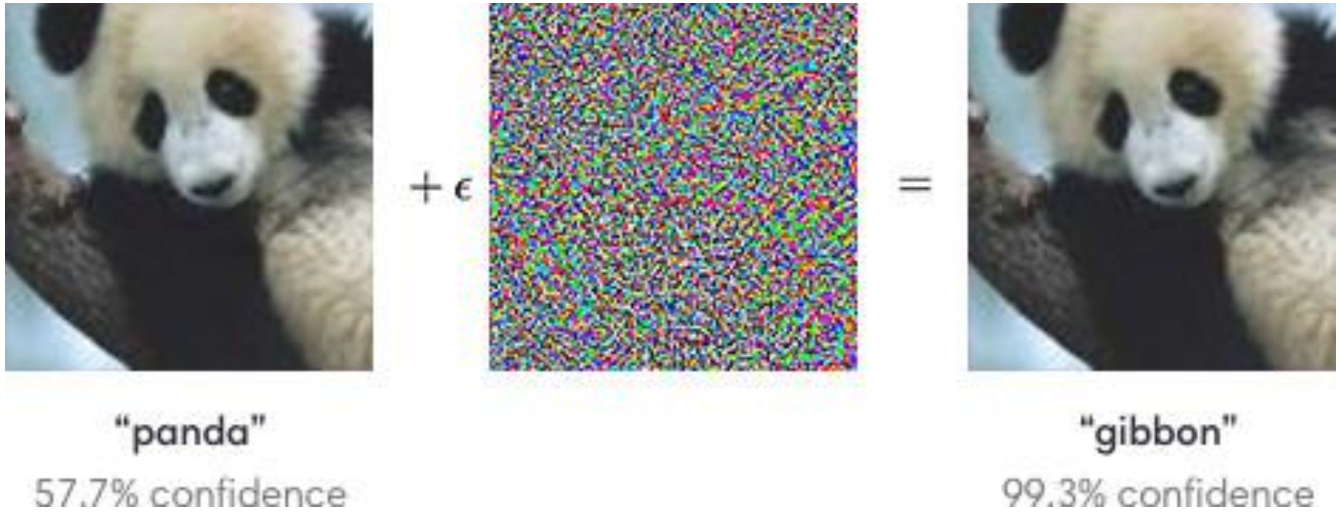
Table 2: Comparison of AutoML frameworks accepted data and data preprocessing steps. "?" represents that no indicators were found that the feature is included in that framework. An empty field indicates that the feature is not present in the selected framework and a ● represents that it is included. (*): commercial frameworks; (2*): including object tracking in videos, and video and audio classification; (3*): must be already converted into numeric values

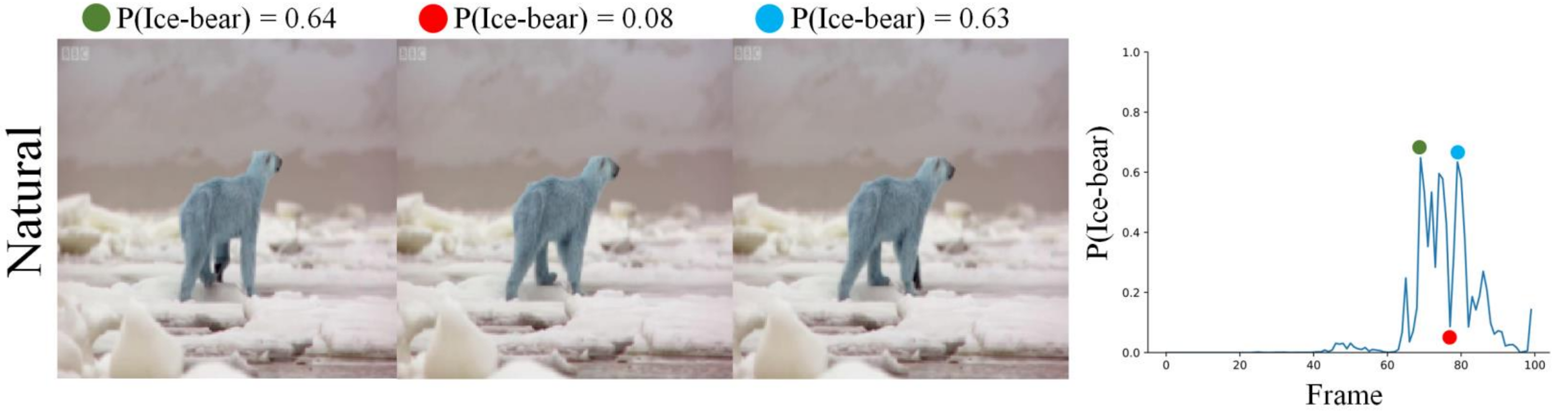


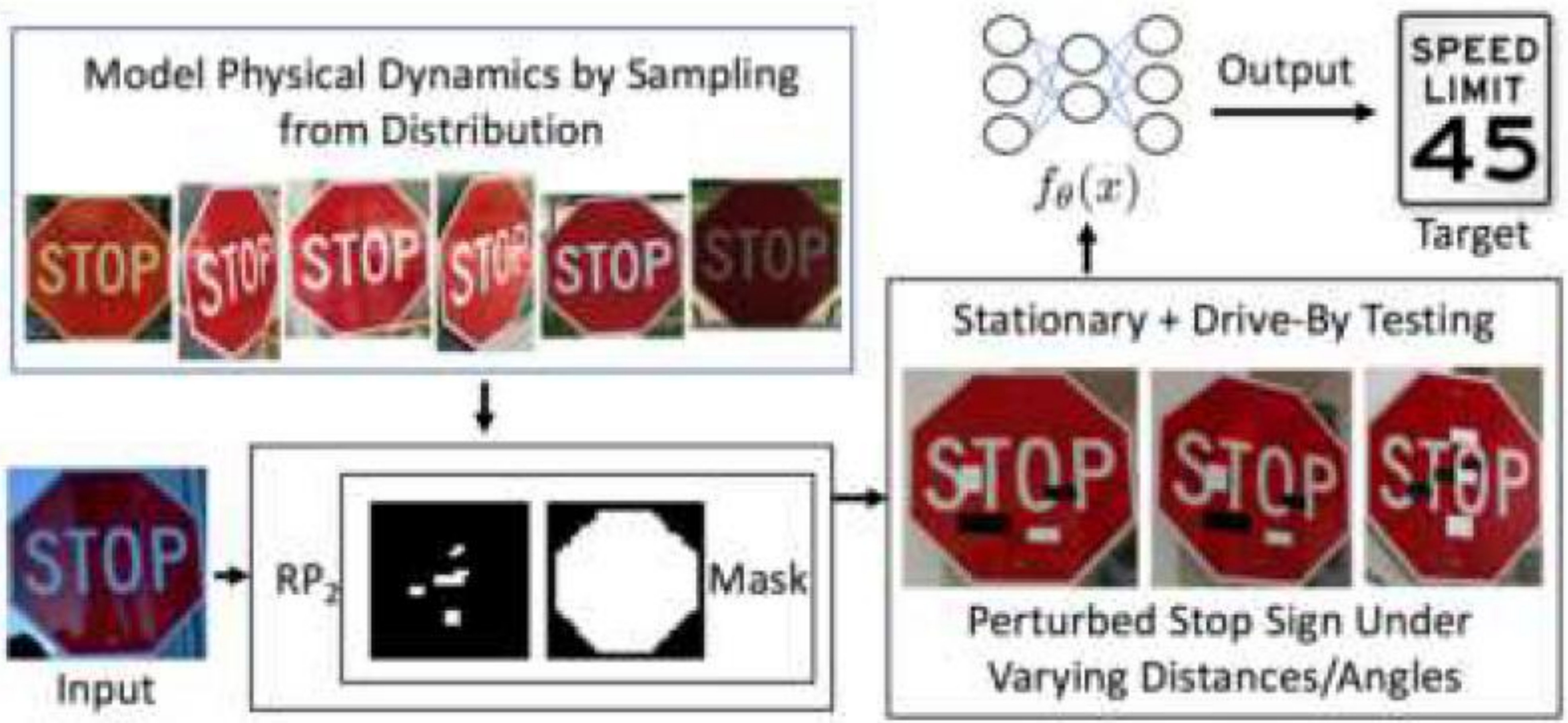
(a) Original OCTA image (b) Zana Vesselness (c) nnUnet

Figure 16: Exemplary results of the application of nnUnet [19] and a Zana vesselness measure-based [42] for retinal vessel segmentation. The left figure (a) shows a slice from the original OCTA volume. The results for the same slice for Zana (b) and nnUnet (c) encode true background and foreground detections as black and white, respectively. False background detections are colored green. False foreground detections are highlighted in in red.

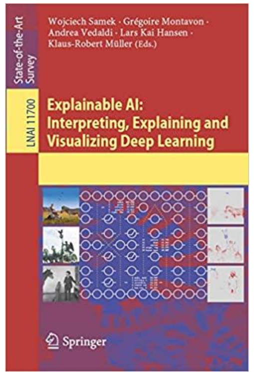
Explainable AI







Eykholt, 2018



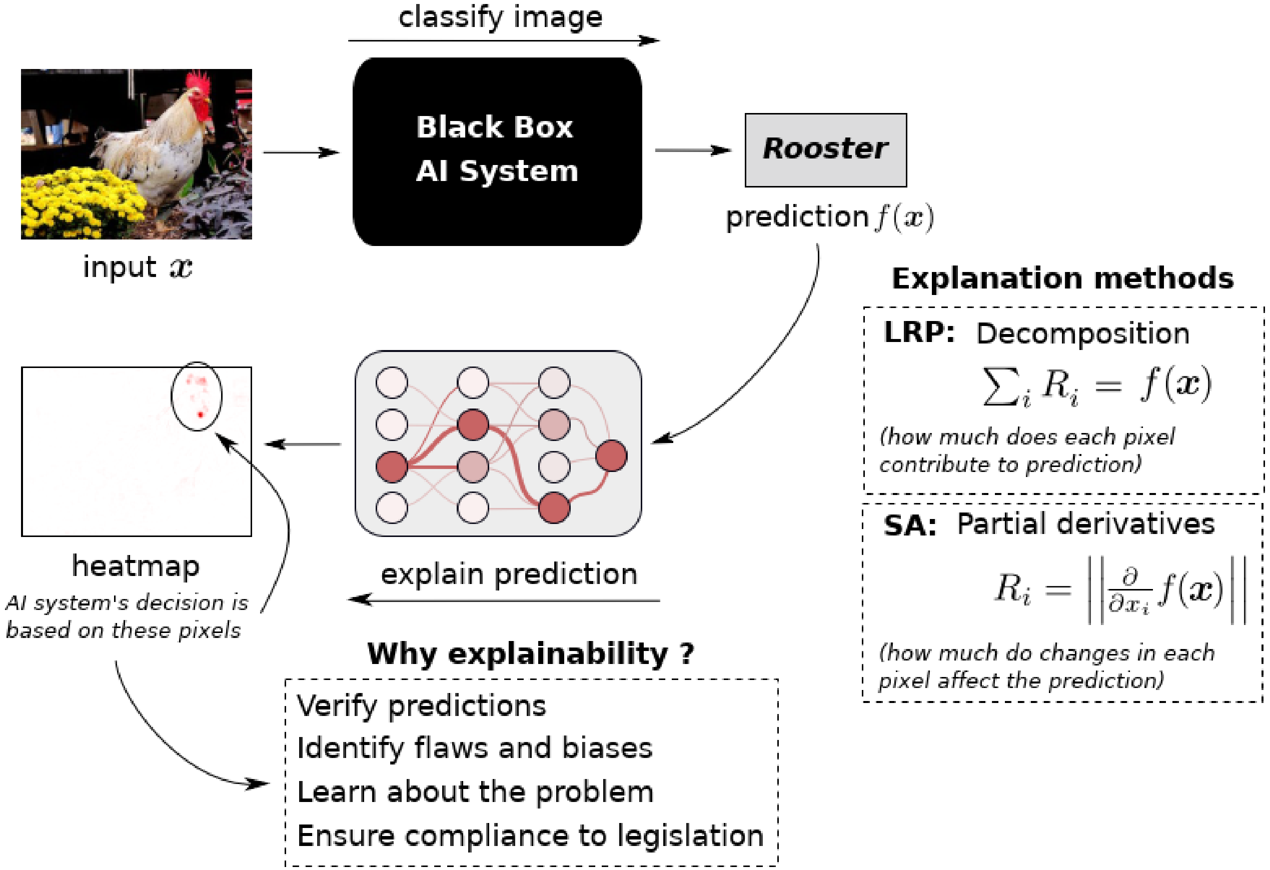
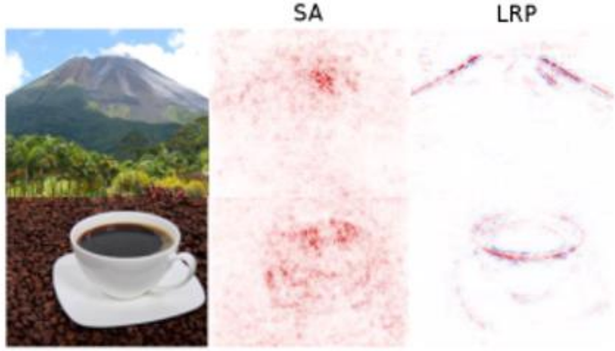


Fig. 1. Explaining predictions of an AI system. The input image is correctly classified as "rooster". In order to understand why the system has arrived at this decision, explanation methods such as SA or LRP are applied. The result of this explanation is an image, the heatmap, which visualizes the importance of each pixel for the prediction. In this example the rooster's red comb and wattle are the basis for the AI system's decision. With the heatmap one can verify that the AI system works as intended.

Explainable AI

(A) Image classification

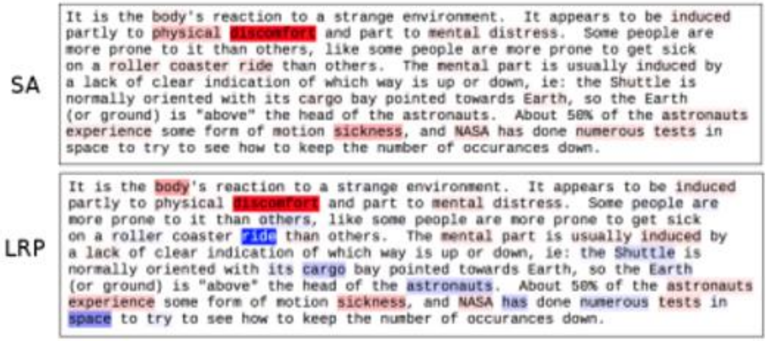
Explaining predictions: "Volcano", "Coffe Cup"



Quantitave comparison of SA and LRP

(B) Text document classification

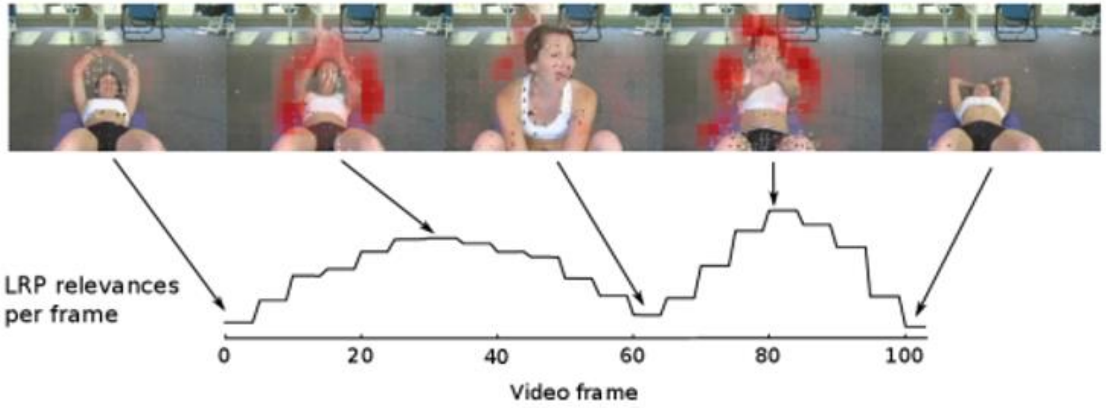
Explaining prediction: "sci.med"



Quantitave comparison of SA and LRP

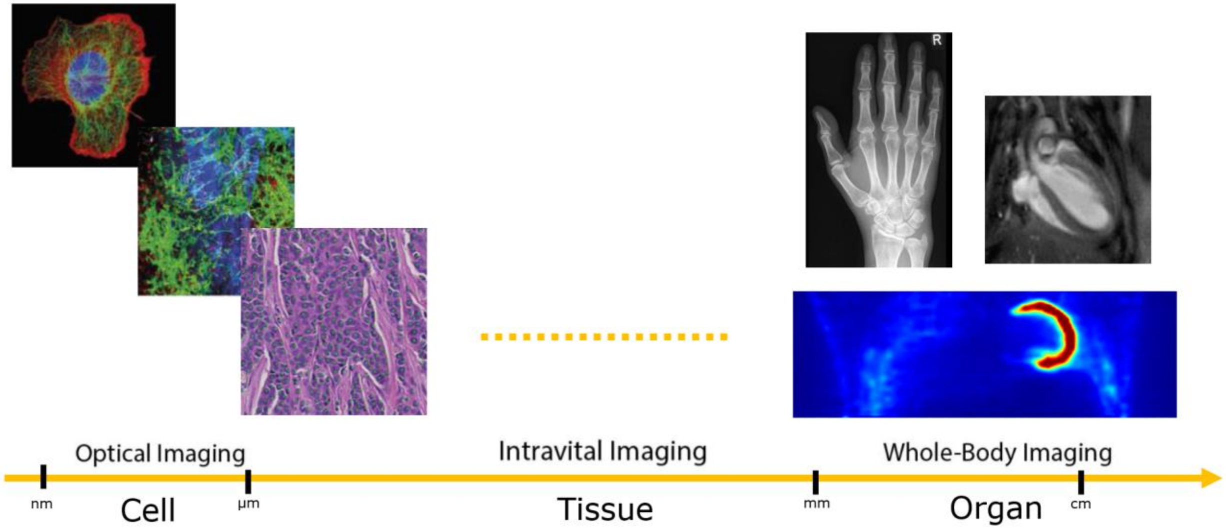
(C) Human action recognition in videos

Explaining prediction: "sit-up"



Biomedical Imaging

> 100 years



+ Powerful (ML-based) image analysis algorithms & tools

There is no modern biology and medicine without imaging

Universität
Münster

Machine Learning for Bioinformatics

Machine Learning for Bioinformatics

OXFORD

Briefings in Bioinformatics, 18(5), 2017, 851–869

doi: 10.1093/bib/bbw068
Advance Access Publication Date: 25 July 2016
Paper

Deep learning in bioinformatics

Seonwoo Min, Byunghan Lee and Sungroh Yoon

Corresponding author: Sungroh Yoon, Department of Electrical and Computer Engineering, Seoul National University, Seoul 08826, Korea. Tel.: +82-2-880-1401; Fax: +82-2-871-5974; E-mail: sryoon@snu.ac.kr

Deep learning in bioinformatics: introduction, application, and perspective in big data era

Yu Li
KAUST
CBRC
CEMSE

Chao Huang
NICT
CAS

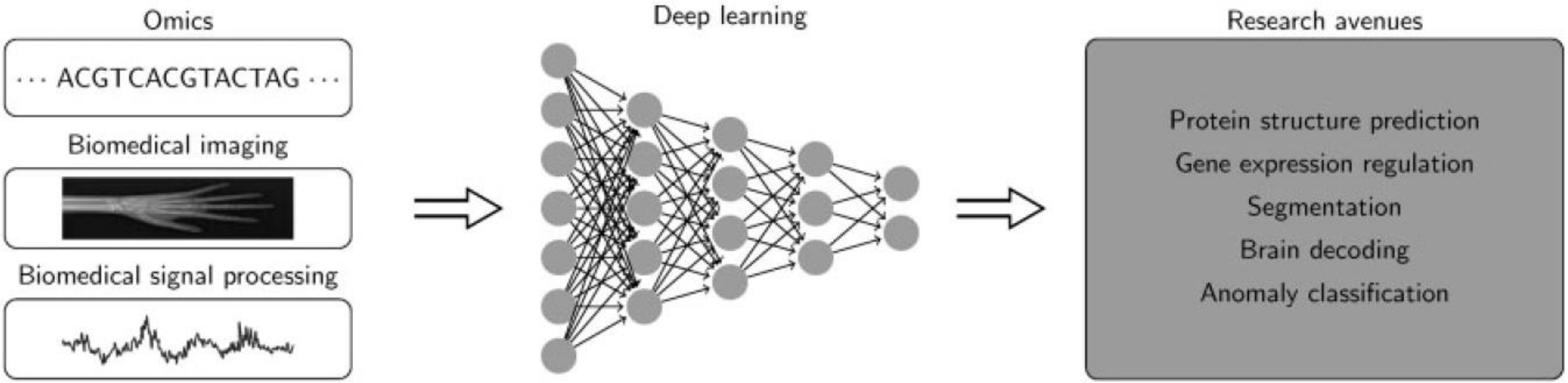
Lizhong Ding
IIAI

Zhongxiao Li
KAUST
CBRC
CEMSE

Recent Advances of Deep Learning in Bioinformatics and Computational Biology

Binhua Tang [1,2†], Zixiang Pan [1†], Kang Yin [1] and Asif Khateeb [1]*

Machine Learning for Bioinformatics



Min, 2017

Machine Learning for Bioinformatics

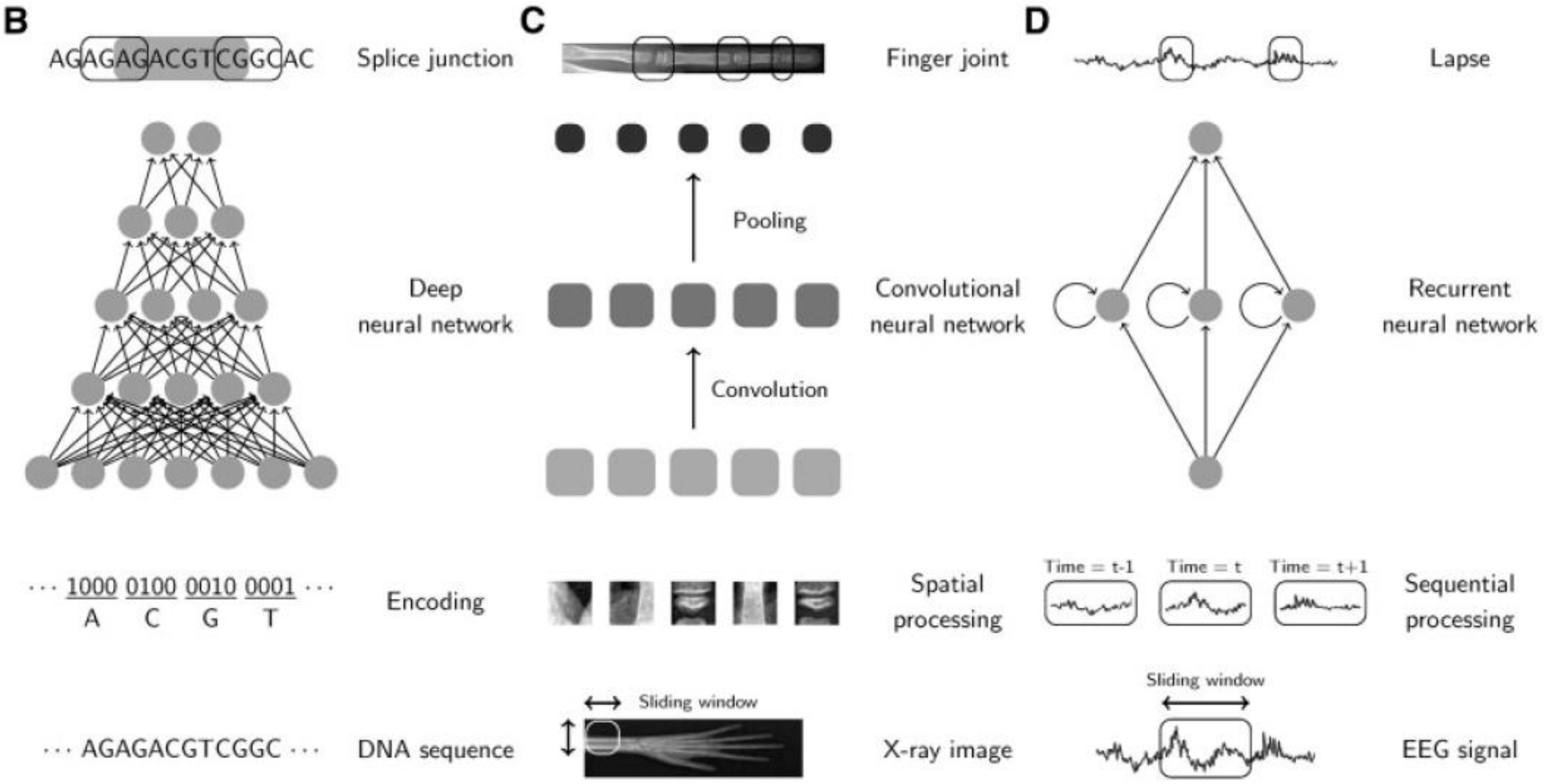


Figure 2. Application of deep learning in bioinformatics research. (A) Overview diagram with input data and research objectives. (B) A research example in the omics domain. Prediction of splice junctions in DNA sequence data with a deep neural network [94]. (C) A research example in biomedical imaging. Finger joint detection from X-ray images with a convolutional neural network [145]. (D) A research example in biomedical signal processing. Lapse detection from EEG signal with a recurrent neural network [178].

Machine Learning for Bioinformatics

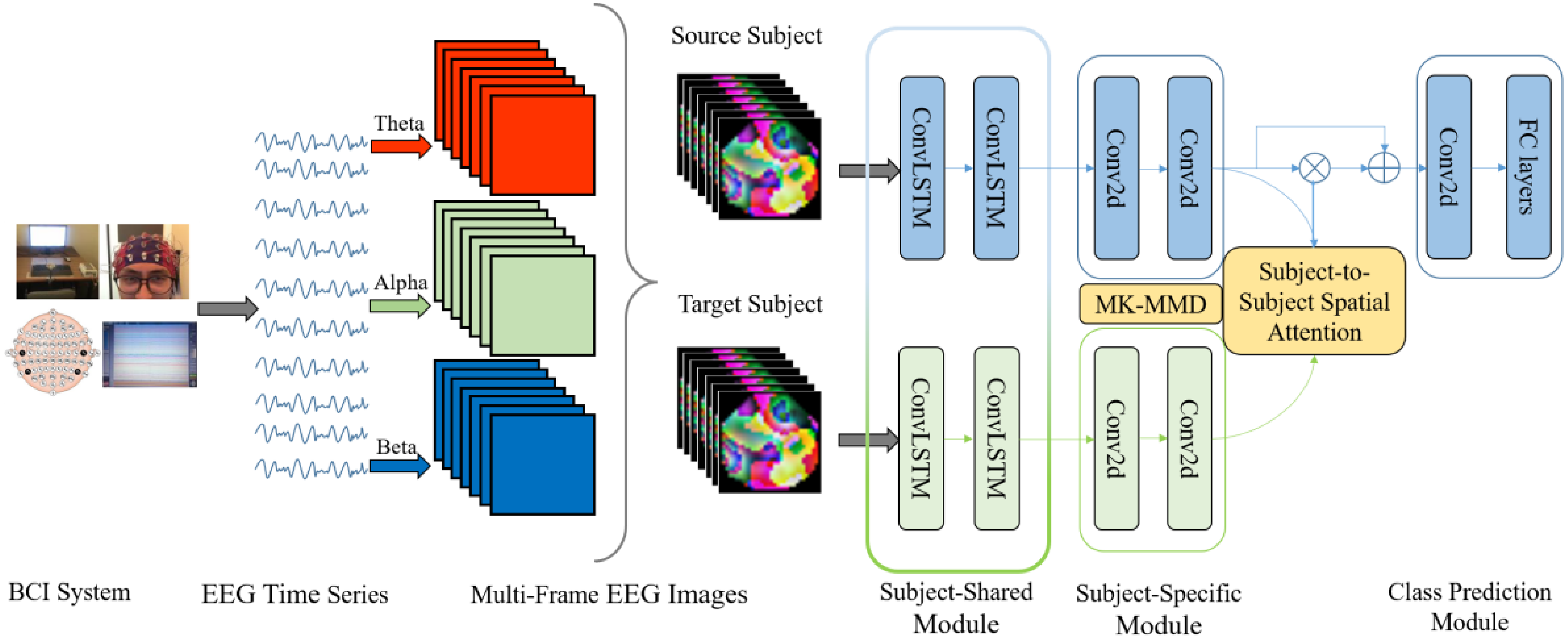
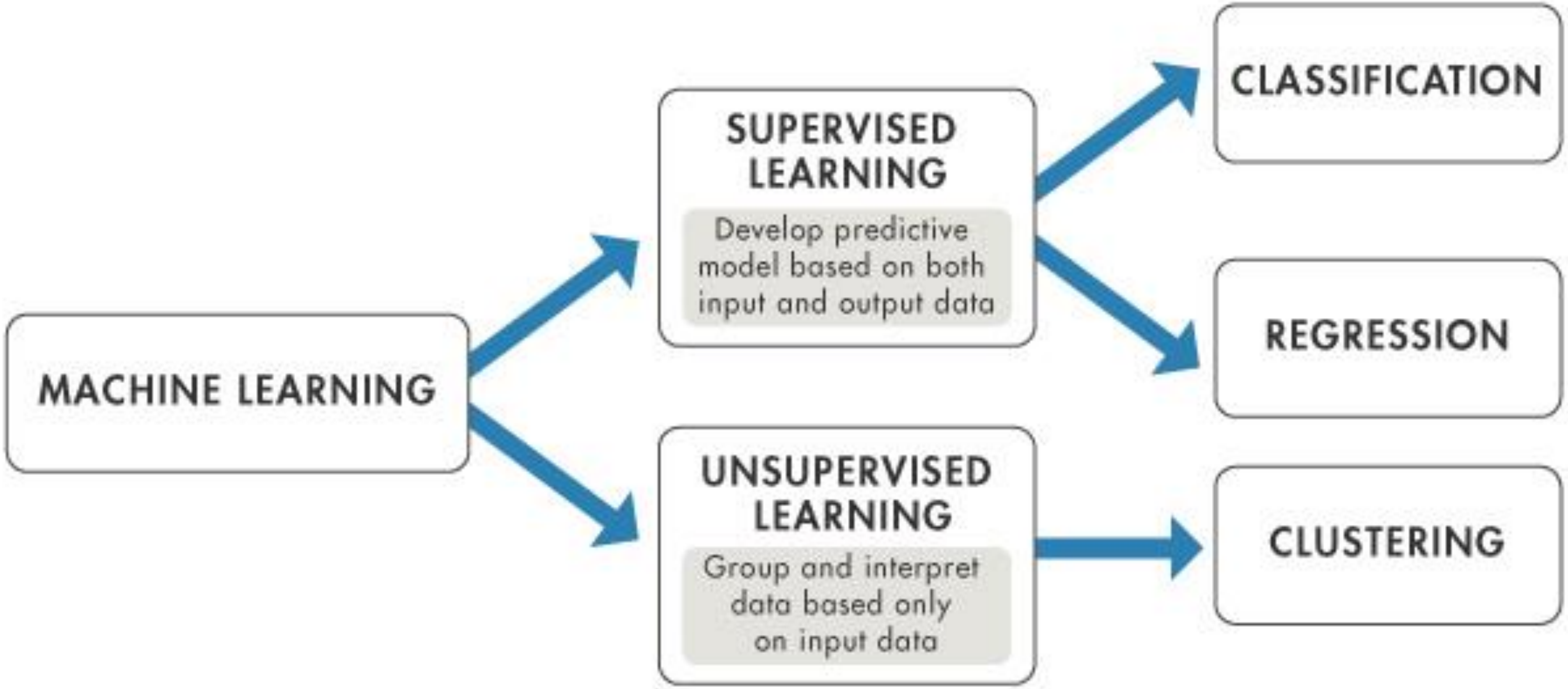


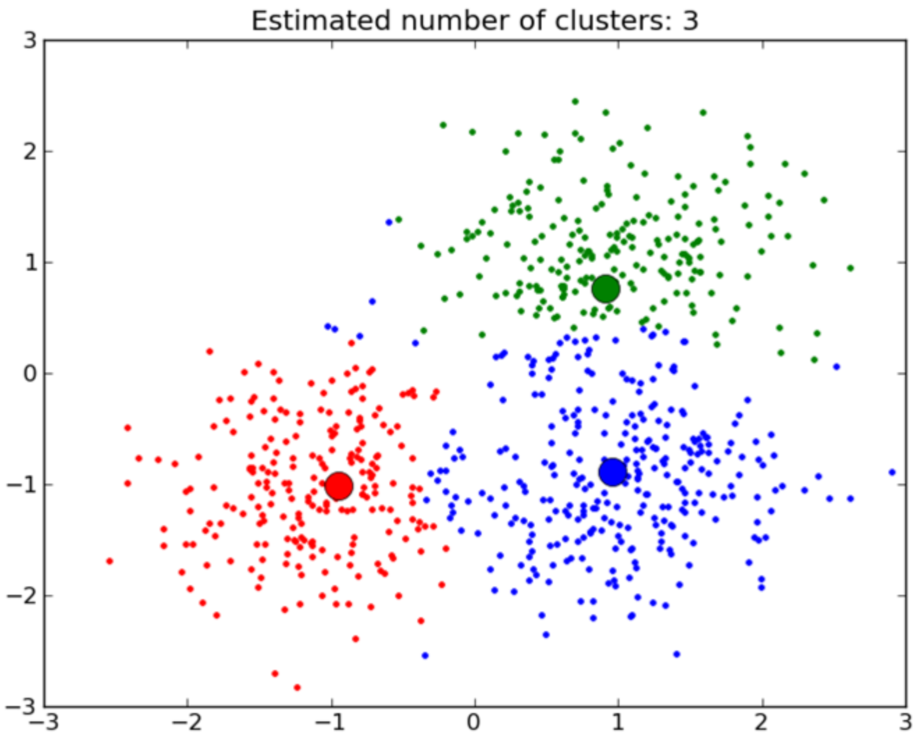
Fig. 2. Pipeline overview of the proposed cross-subject transfer model: (1) EEG time series are divided into 7 slices of 0.5-second windows and spectral power within the three frequency bands (theta (4-7Hz), alpha (8-13Hz), and beta (13-30 Hz)) are extracted by Fast Fourier Transform (FFT); (2) A 2-D space position map of EEG electrodes formed by the azimuthal equidistant projection is then utilized to make 7-frame EEG images with three spectral channels; (3) Then, a pair of source and target EEG images are fed into the Subject-Shared module. This Subject-Shared module has been pretrained only by the source data and is frozen in the training phase; (4) Subsequently, the output feature pair is input in the Subject-Specific module. In this module, the features from different domains are further fed into the same layer architecture but with different network parameters; (5) Finally, the Class-Prediction module processes the output features by the last module and makes a decision for the level of WM load.

Chen, 2022

Machine Learning for Bioinformatics



Clustering

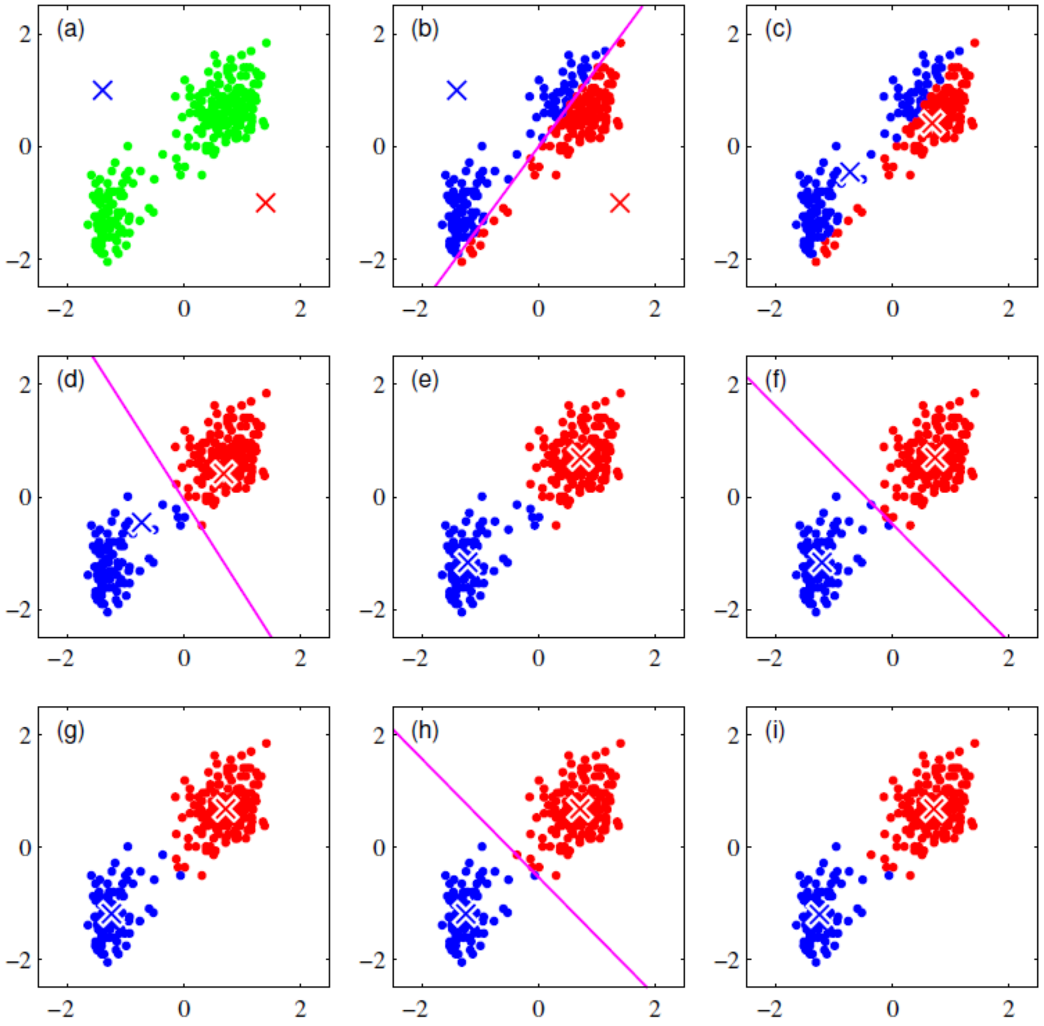


Group data into clusters such that there is

- high intra-class similarity (→ compact clusters)
- low inter-class similarity (→ good separability)

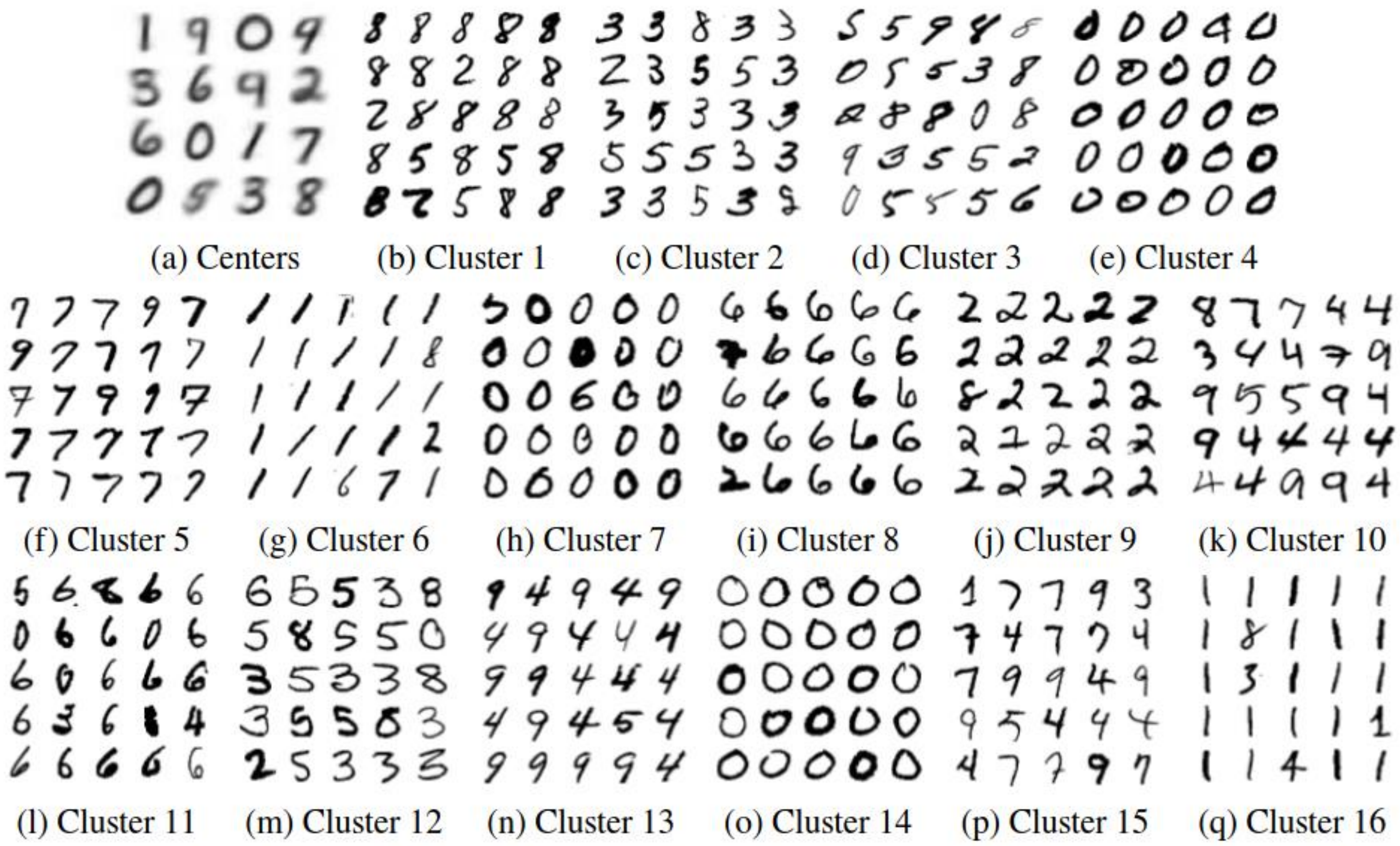
Clustering: K-Means

Example: k=2



Clustering: K-Means

Example: K-Means for MNIST digits data; k=16. (a) The 16 cluster centers. (b-q) 25 data examples for each of the 16 clusters. The clusters roughly grab digits with similar stroke patterns.



A **main limitation of one-way clustering** algorithms when applied to high-dimensional molecular data for disease subtyping is that cluster assignment of samples is based on the assumption that all molecular features are relevant to the sample groups or disease subtypes.

Biclustering (or two-way clustering, co-clustering, two-mode clustering) is a popular statistical method for simultaneously identifying groups of samples (rows) and groups of variables (columns) characterizing different sample groups.

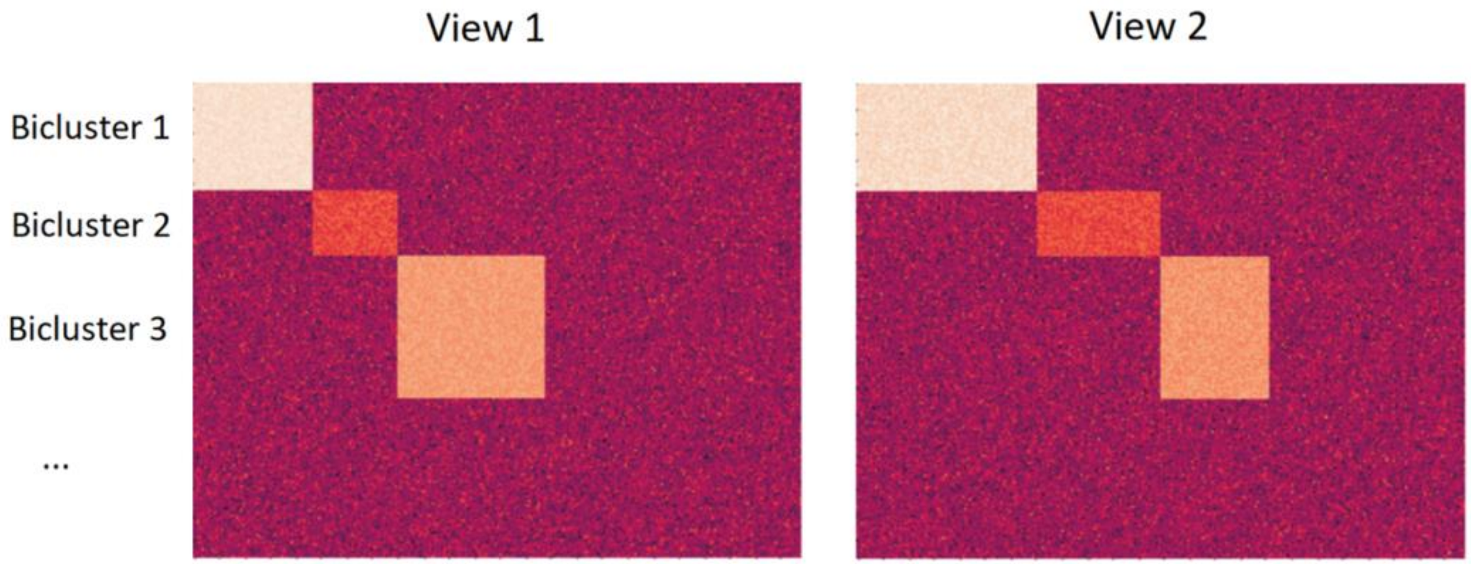


Figure 1: Pictorial illustration of integrative biclustering with two views. A bicluster is comprised of rows (or samples) and columns (or variables) from each view, with the samples in view 1 being the same as the samples in view 2. The samples in each bicluster can overlap or not. Similarly, the variables in each bicluster within a view can overlap or not. In this figure, the samples and variables in each bicluster do not overlap.

Padilha and Campello *BMC Bioinformatics* (2017) 18:55
DOI 10.1186/s12859-017-1487-1

BMC Bioinformatics

RESEARCH ARTICLE Open Access

CrossMark

A systematic comparative evaluation of biclustering techniques

Victor A. Padilha[1*] and Ricardo J. G. B. Campello[2,1]

OXFORD

Briefings in Bioinformatics, 20(4), 2019, 1449–1464

doi: 10.1093/bib/bby014
Advance Access Publication Date: 27 February 2018
Paper

It is time to apply biclustering: a comprehensive review of biclustering applications in biological and biomedical data

Juan Xie, Anjun Ma, Anne Fennell, Qin Ma and Jing Zhao

Machine Learning provides powerful tools for bioinformatics

- omics data analysis
- biomedical imaging
- biomedical signal processing