Artificial Intelligence and multi-omics data

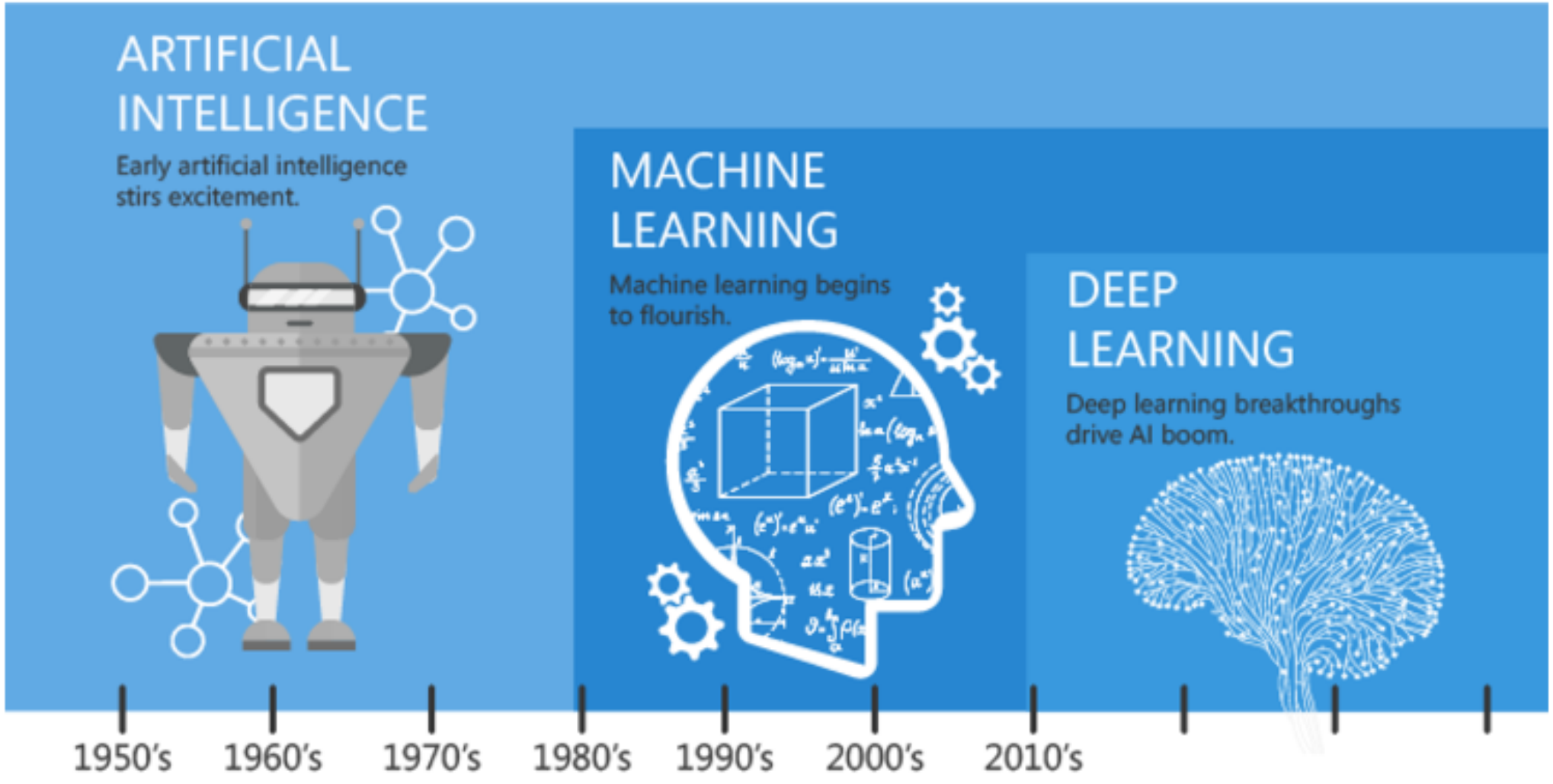
Gennady Roshchupkin

Erasmus MC Medical Center Rotterdam
Department of Epidemiology
Department of Radiology and Nuclear Medicine

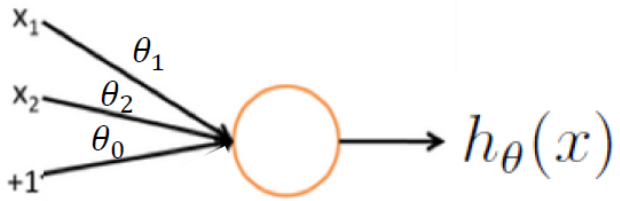
Email: g.roshchupkin@erasmusmc.nl
Twitter: @Roshchupkin

Erasmus MC
Universitair Medisch Centrum Rotterdam

BETTER
FASTER
STRONGER


ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

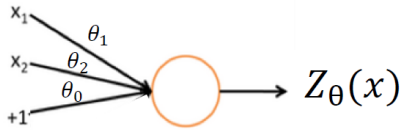
Basic principles



parameters

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 = \sum_{i=0}^{n} \theta_i x_i = \theta^T x,$$

A non-linear activation (logistic)



$$Z_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

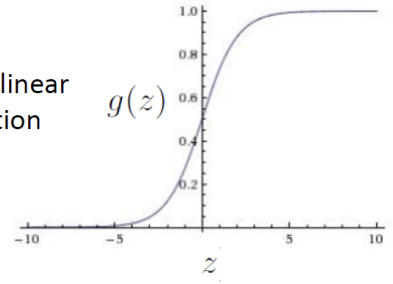
$$g(z) = \frac{1}{1+e^{-z}}$$

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}},$$

Logistic / Sigmoid

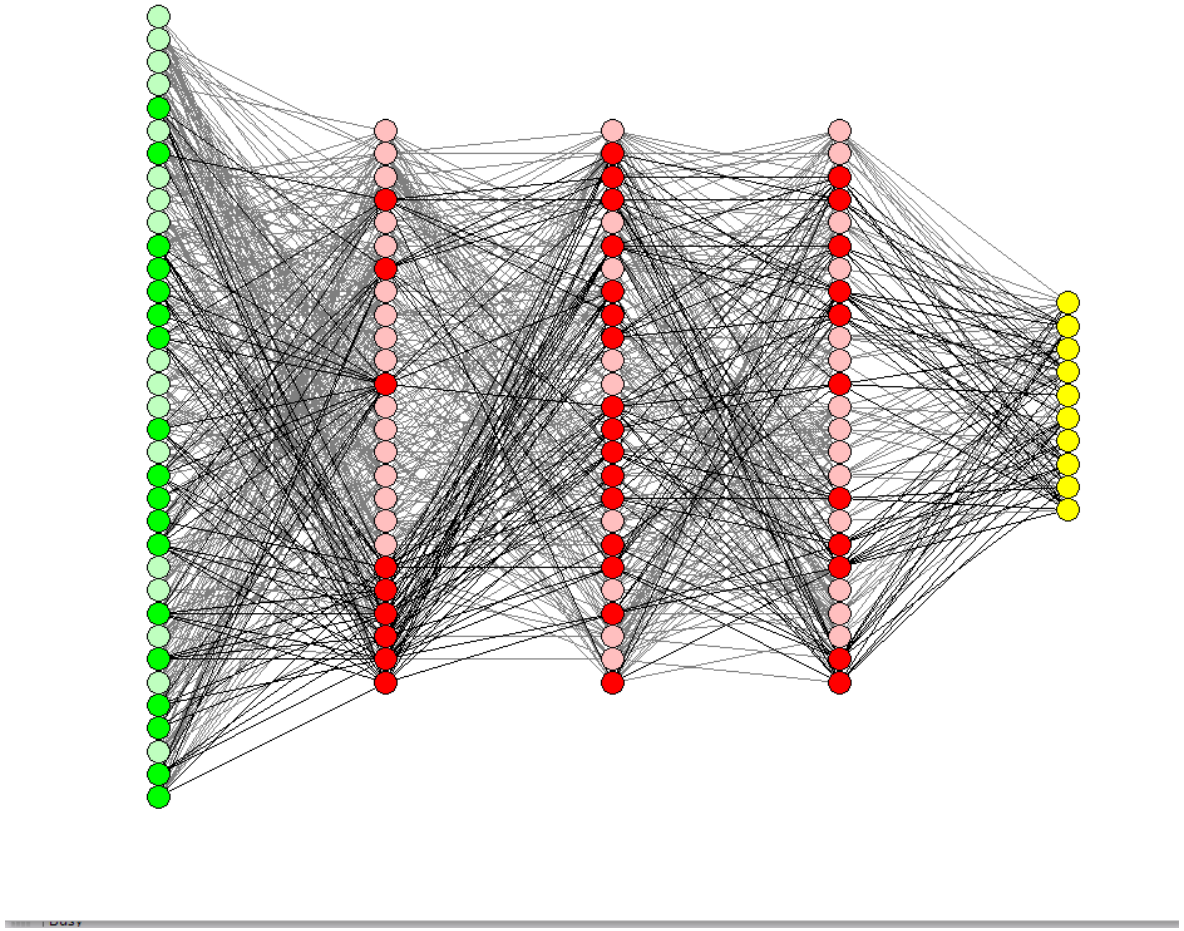
Useful for predicting probabilities

Non-linear function

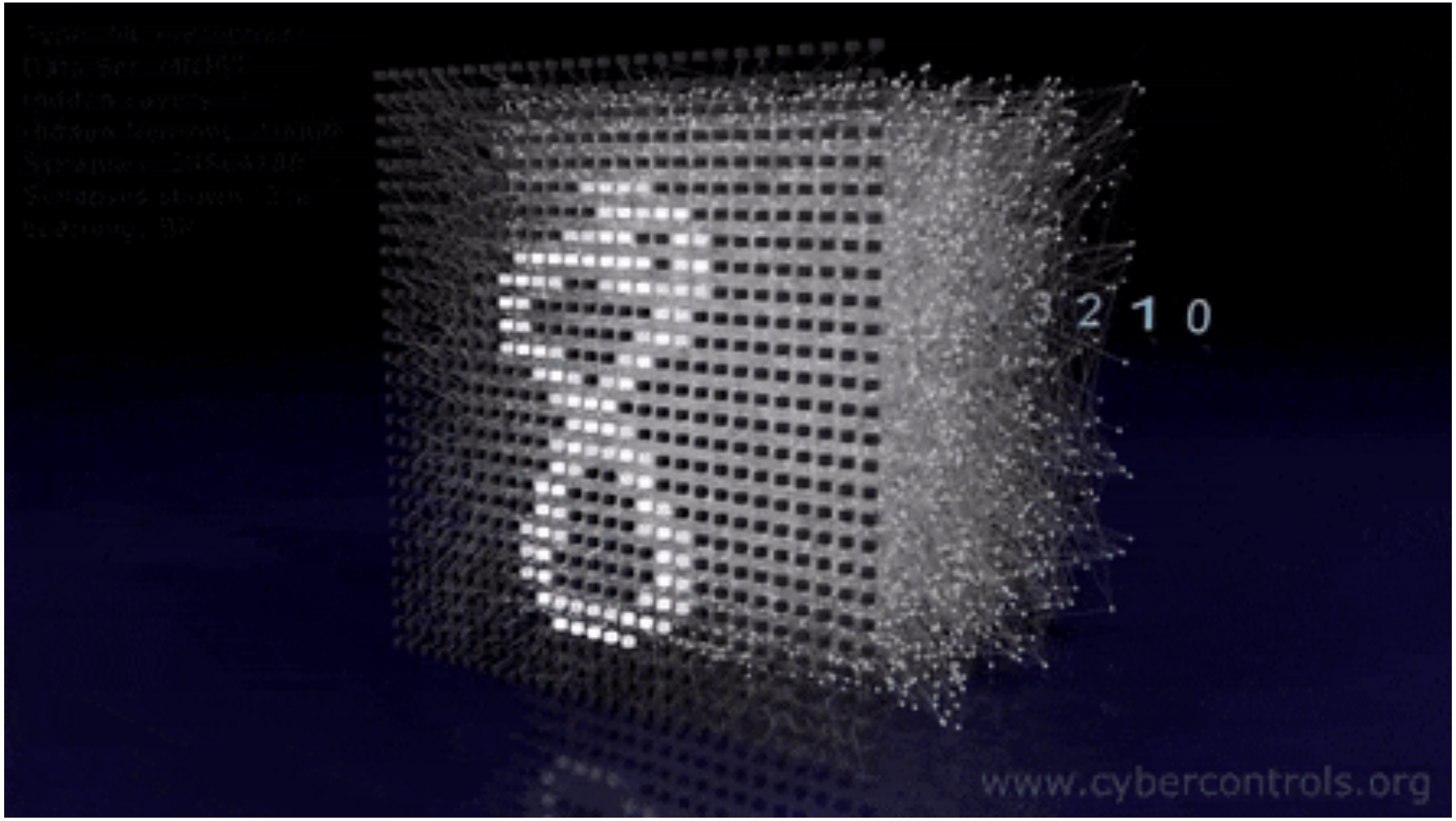


Erasmus MC

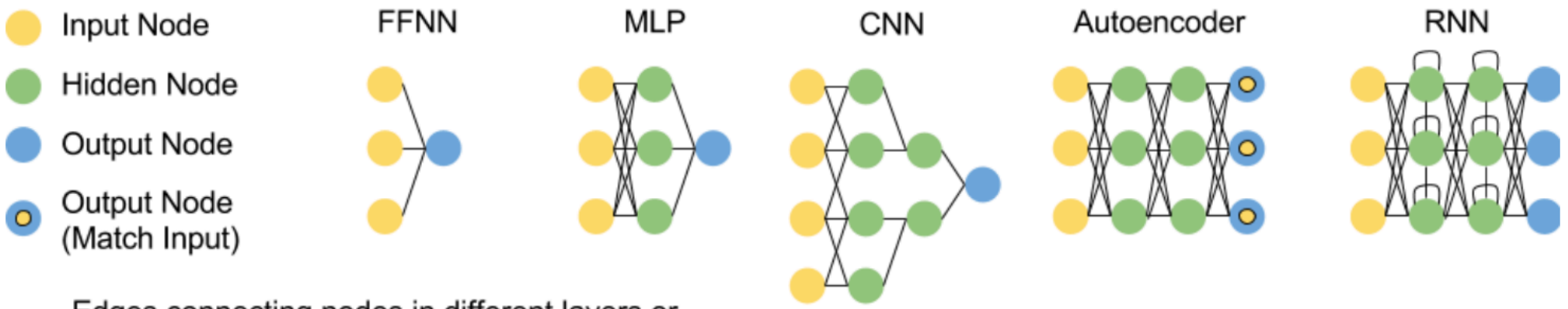
Basic principles



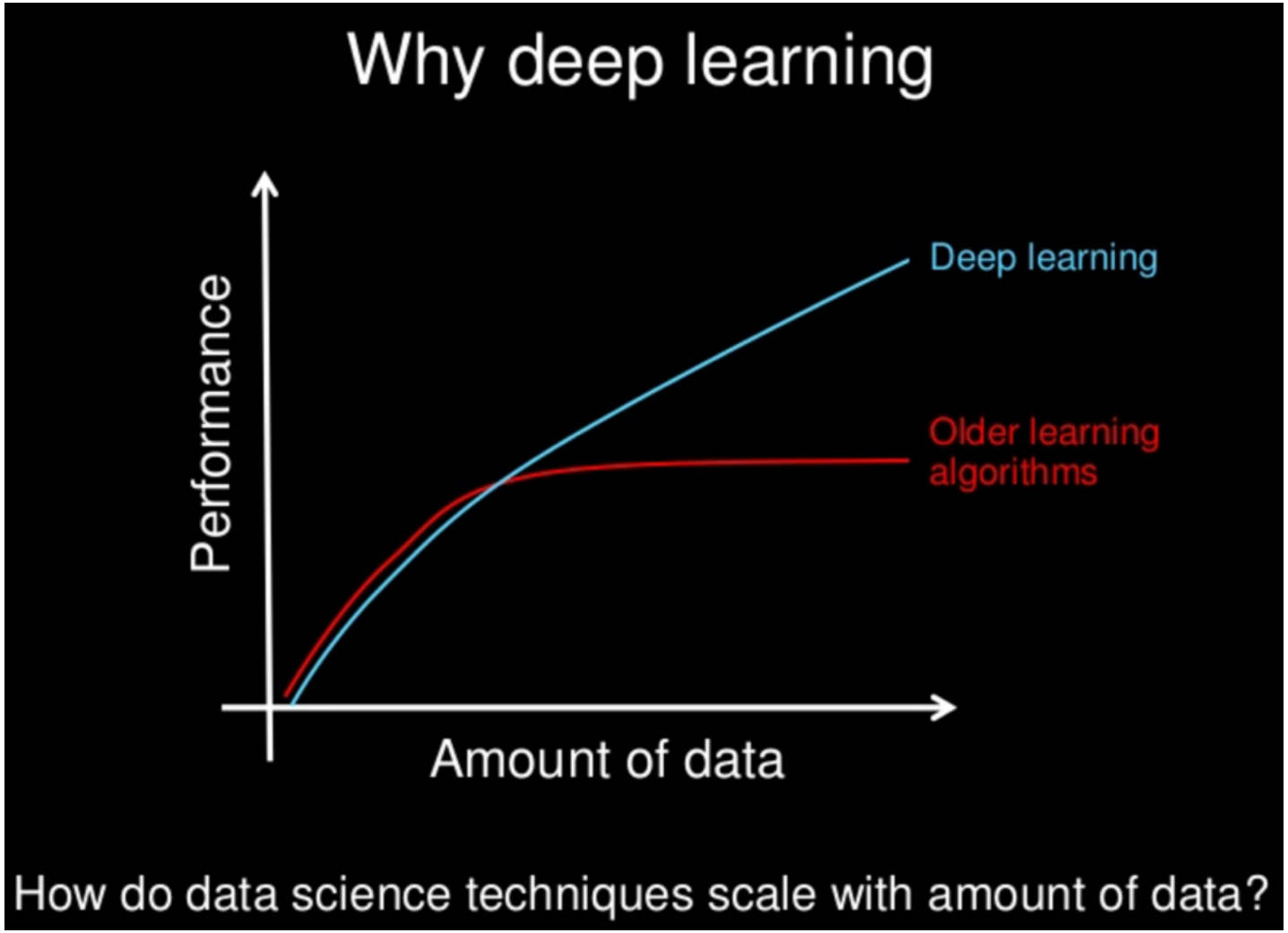
Basic principles



Basic principles



FASTER


Why deep learning
Performance
Deep learning
Older learning algorithms
Amount of data
How do data science techniques scale with amount of data?

Titan X
vs
GTX 980
vs
Tesla M40
vs
Tesla K80

Table 1 | Peer-reviewed publications of AI algorithms compared with doctors

Specialty	Images	Publication
Radiology/ neurology	CT head, acute neurological events	Titano et al. [27]
	CT head for brain hemorrhage	Arbabshirani et al.[19]
	CT head for trauma	Chilamkurthy et al.[20]
	CXR for metastatic lung nodules	Nam et al.[8]
	CXR for multiple findings	Singh et al.[7]
	Mammography for breast density	Lehman et al.[26]
	Wrist X-ray*	Lindsey et al.[9]
Pathology	Breast cancer	Ehteshami Bejnordi et al.[41]
	Lung cancer (+driver mutation)	Coudray et al.[33]
	Brain tumors (+methylation)	Capper et al.[45]
	Breast cancer metastases*	Steiner et al.[35]
	Breast cancer metastases	Liu et al.[34]
Dermatology	Skin cancers	Esteva et al.[47]
	Melanoma	Haenssle et al.[48]
	Skin lesions	Han et al.[49]
Ophthalmology	Diabetic retinopathy	Gulshan et al.[51]
	Diabetic retinopathy*	Abramoff et al.[31]
	Diabetic retinopathy*	Kanagasingam et al.[32]
	Congenital cataracts	Long et al.[38]
	Retinal diseases (OCT)	De Fauw et al.[56]
	Macular degeneration	Burlina et al.[52]
	Retinopathy of prematurity	Brown et al.[60]
	AMD and diabetic retinopathy	Kermany et al.[53]
Gastroenterology	Polyps at colonoscopy*	Mori et al.[36]
	Polyps at colonoscopy	Wang et al.[37]
Cardiology	Echocardiography	Madani et al.[23]
	Echocardiography	Zhang et al.[24]

Prospective studies are denoted with an asterisk.

Table 2 | FDA AI approvals are accelerating

Company	FDA Approval	Indication
Apple	September 2018	Atrial fibrillation detection
Aidoc	August 2018	CT brain bleed diagnosis
iCAD	August 2018	Breast density via mammography
Zebra Medical	July 2018	Coronary calcium scoring
Bay Labs	June 2018	Echocardiogram EF determination
Neural Analytics	May 2018	Device for paramedic stroke diagnosis
IDx	April 2018	Diabetic retinopathy diagnosis
Icometrix	April 2018	MRI brain interpretation
Imagen	March 2018	X-ray wrist fracture diagnosis
Viz.ai	February 2018	CT stroke diagnosis
Arterys	February 2018	Liver and lung cancer (MRI, CT) diagnosis
MaxQ-AI	January 2018	CT brain bleed diagnosis
Alivecor	November 2017	Atrial fibrillation detection via Apple Watch
Arterys	January 2017	MRI heart interpretation

Imputation

Data type	Tools name	Methods	Strengths	Limitations
Genotype	***SCDA***	Sparse convolutional denoising autoencoder	Deep learning	Hard to interpret the prediction mechanisms
Gene expression	***EnImpute***	Multi-task deep autoencoder	High scalability	Overfitting
	SAUCIE	Autoencoder-based	Constructs sub-neural networks	
	AutoImpute	Autoencoder with the ZINB loss function		
	DCA	Stochastic optimization and VAE		
	scVI	Deep neural network-based		
	DeepImpute			
Epigenomic	***Avocado***	Tensor factorization and deep neural network	Uses associations between neighbor CpGs as well as between DNA sequence patterns and methylation states	Hyperparameter settings may influence precision and recall
	SCALE	VAE and GMM		
	DeepCpG	Deep learning-based joint model		

Song, M.. Frontiers in Genetics, 2020

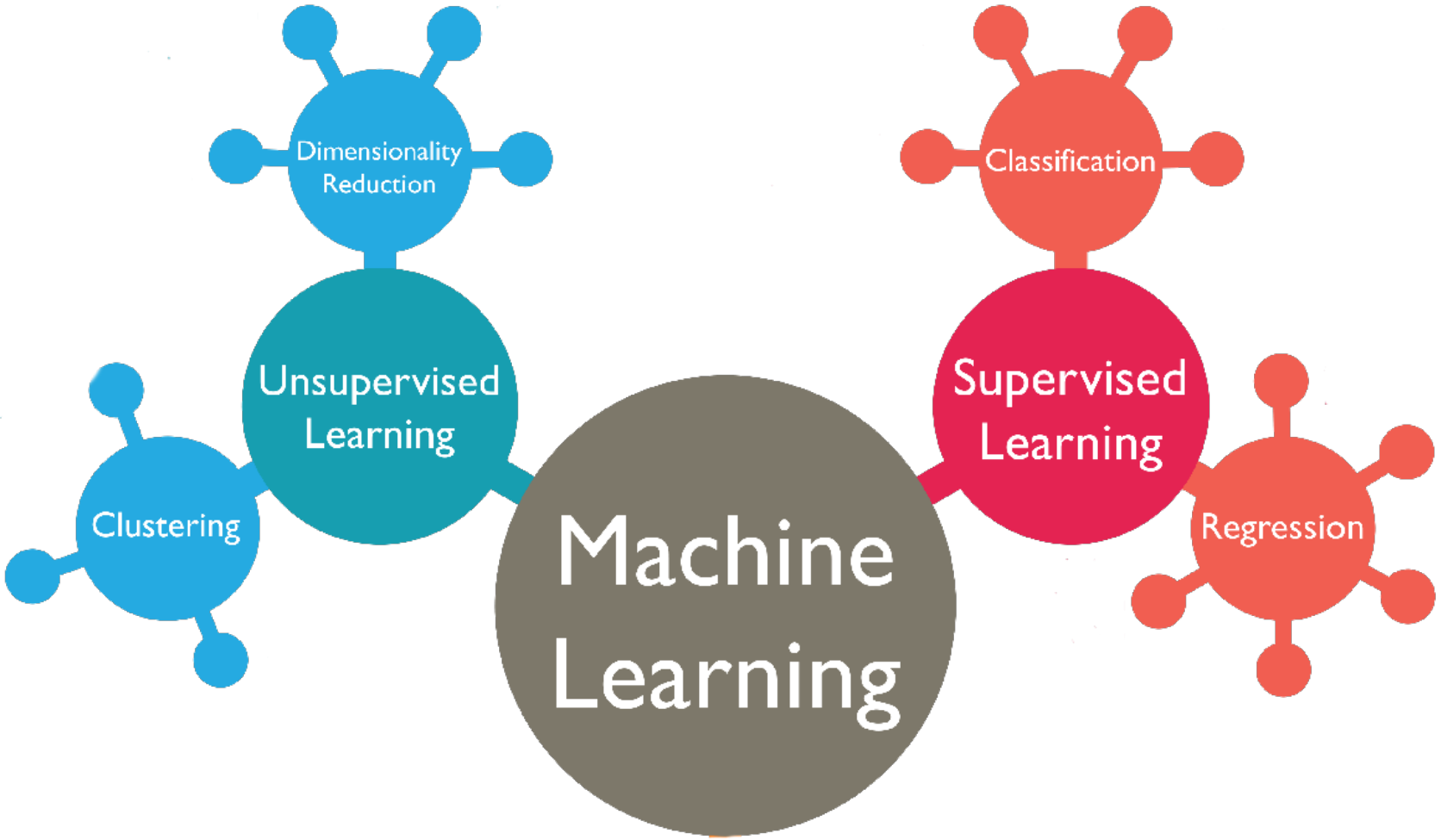
Erasmus MC

Saving computational time

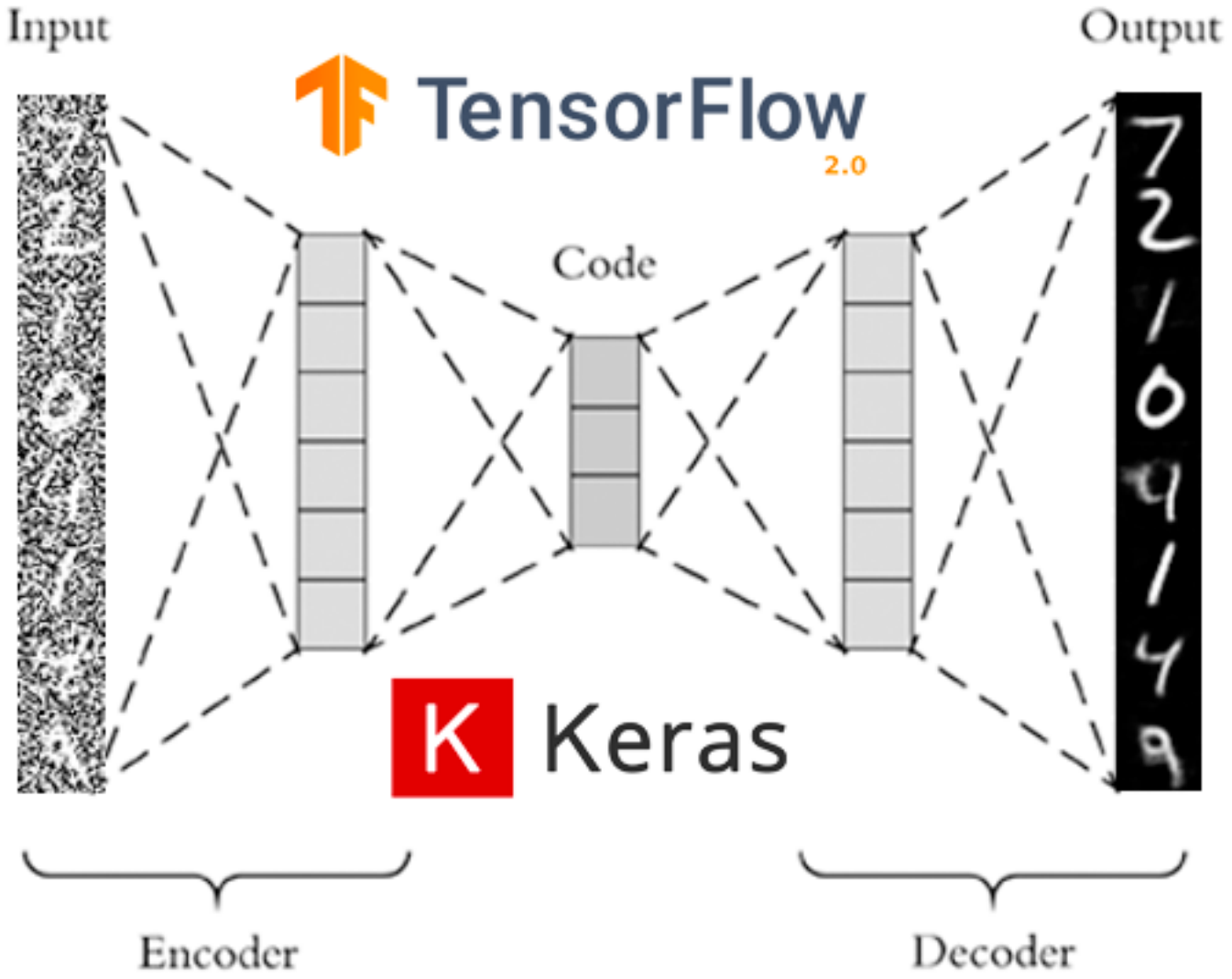
Erasmus MC

UNIVERSAL APPROXIMATION THEOREM

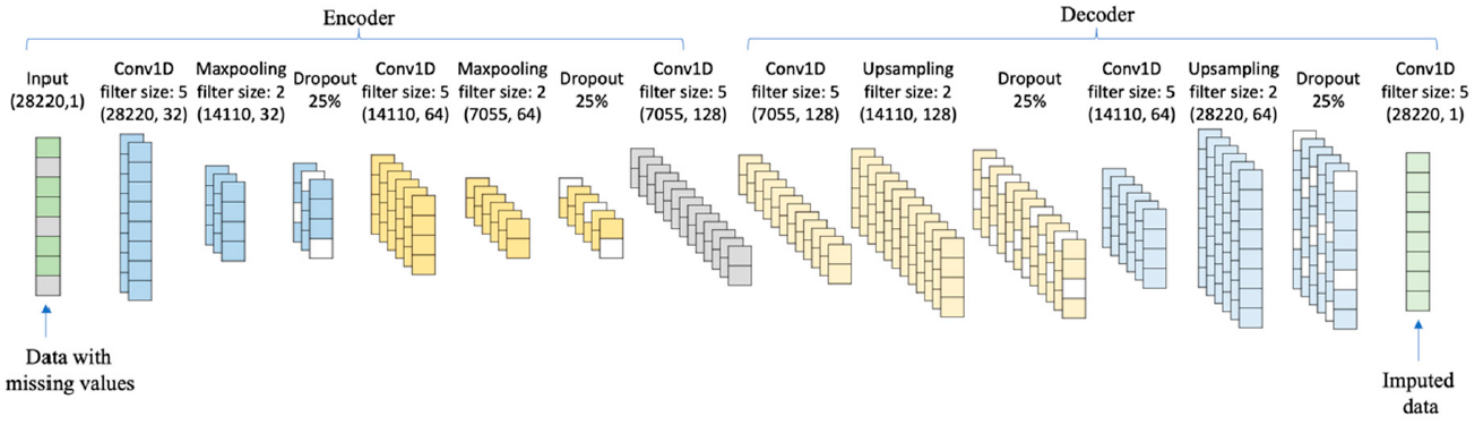
A feedforward network with a single layer is sufficient to represent any function, but the layer may be infeasibly large and may fail to learn and generalize correctly.


Dimensionality Reduction
Classification
Unsupervised Learning
Supervised Learning
Clustering
Regression
Machine Learning

Denoising autoencoders



Sparse Convolutional Denoising Autoencoders for Genotype Imputation

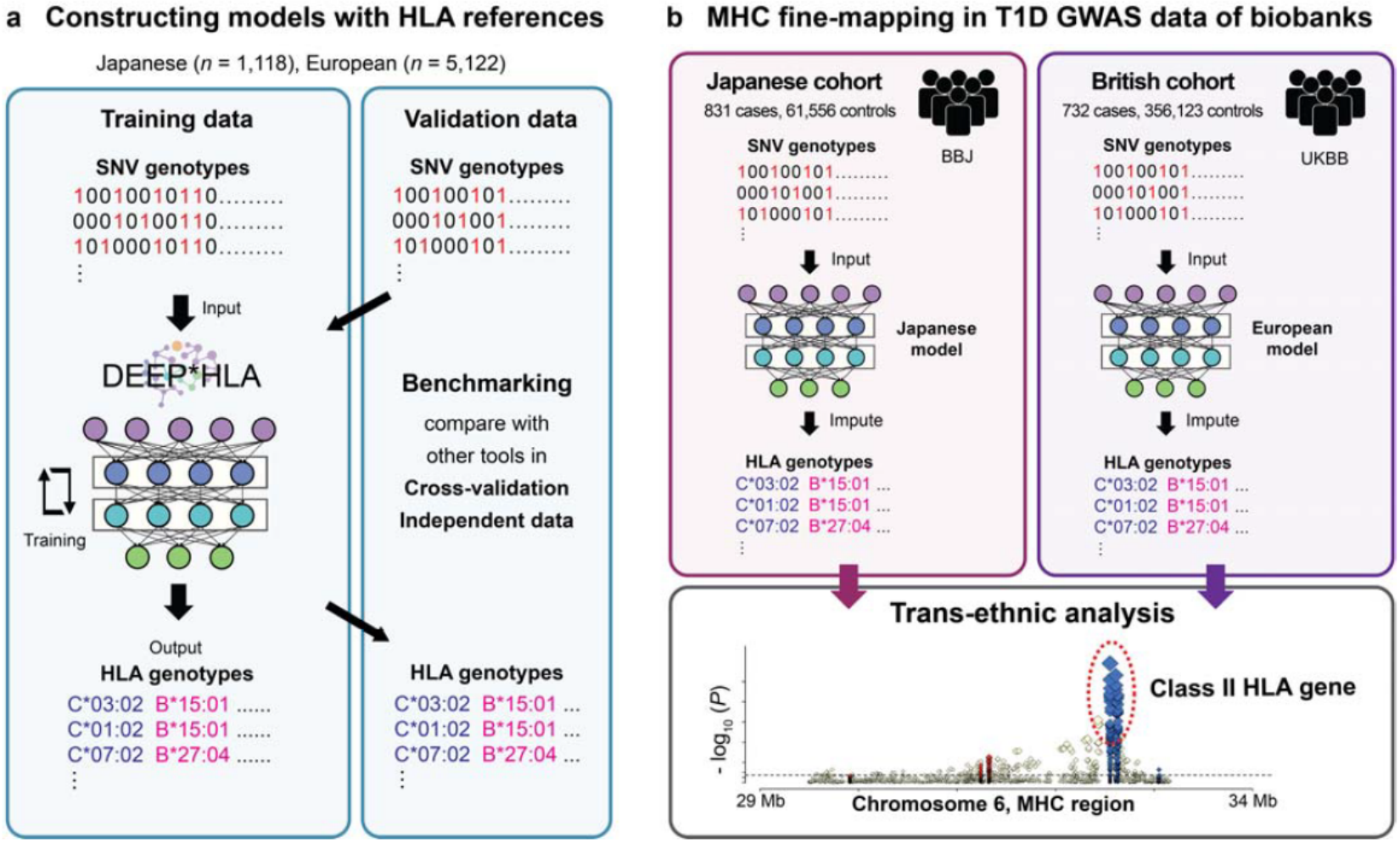


Methods	5%		10%		20%		Total	
	Average Accuracy	Standard Deviation	Average Accuracy	Standard Deviation	Average Accuracy	Standard Deviation	Average Accuracy	Standard Deviation
Average	0.9549	3.2×10^{-4}	0.9549	1.7×10^{-4}	0.9549	1.3×10^{-4}	0.9549	2.1×10^{-4}
KNN	0.9883	9.5×10^{-5}	0.9881	8.8×10^{-5}	0.9877	6.4×10^{-5}	0.9880	8.2×10^{-5}
SVD	0.9899	1.0×10^{-4}	0.9899	9.0×10^{-5}	0.9898	6.9×10^{-5}	0.9899	8.6×10^{-5}
SCDA	0.9975	6.0×10^{-5}	0.9952	1.4×10^{-4}	0.9900	4.2×10^{-4}	0.9942	2.1×10^{-4}

J. Chen et al, Genes 2019

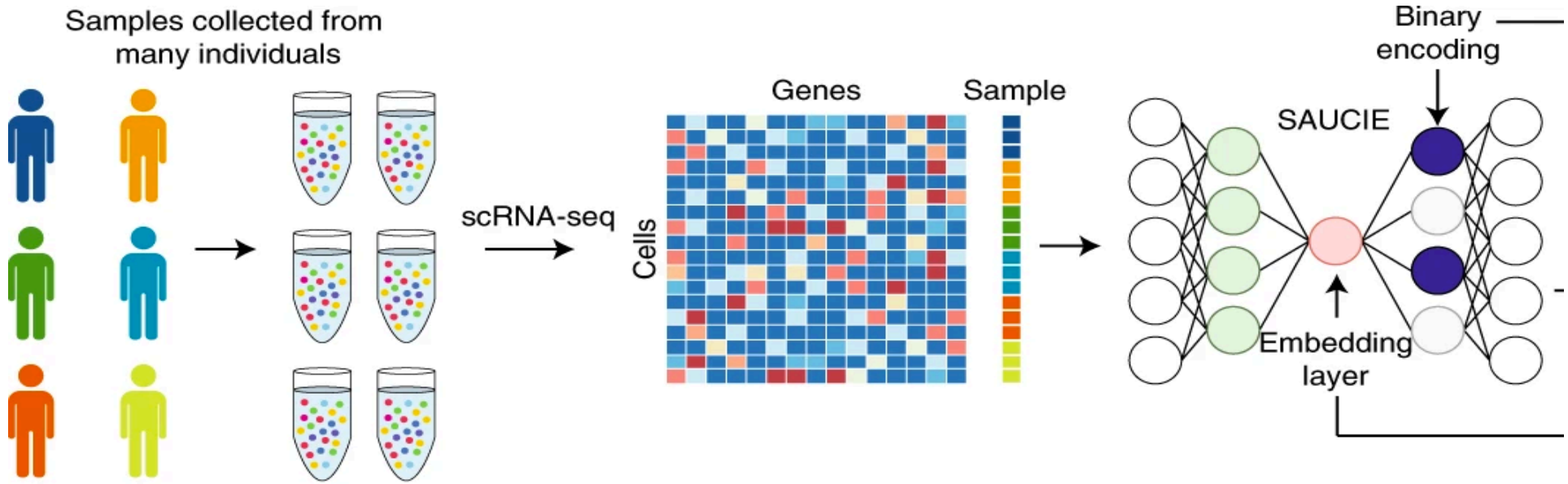
Erasmus MC

HLA allelic imputation and its application to trans-ethnic MHC fine-mapping of type 1 diabetes.



T.Naito et al, BioRxiv 2020

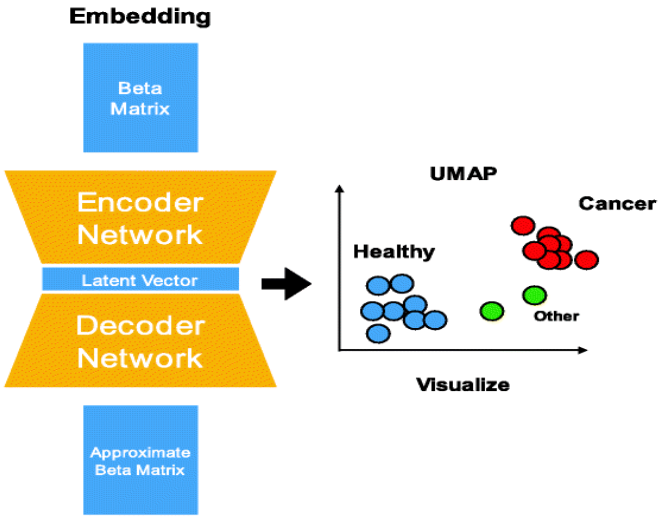
Exploring single-cell data with deep multitasking neural networks

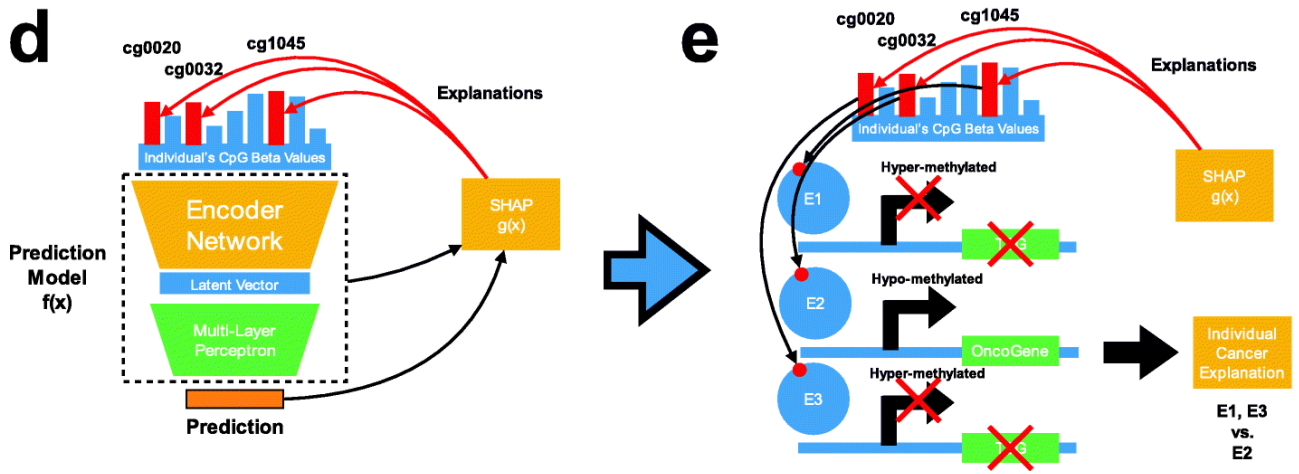


M. Amodio et al, Nature Methods 2019

Erasmus MC

MethylNet: an automated and modular deep learning approach for DNA methylation analysis

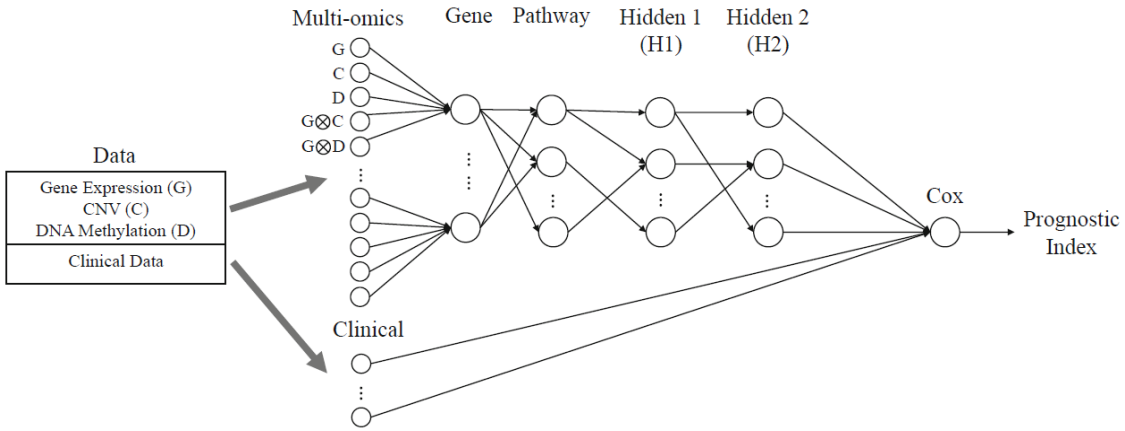




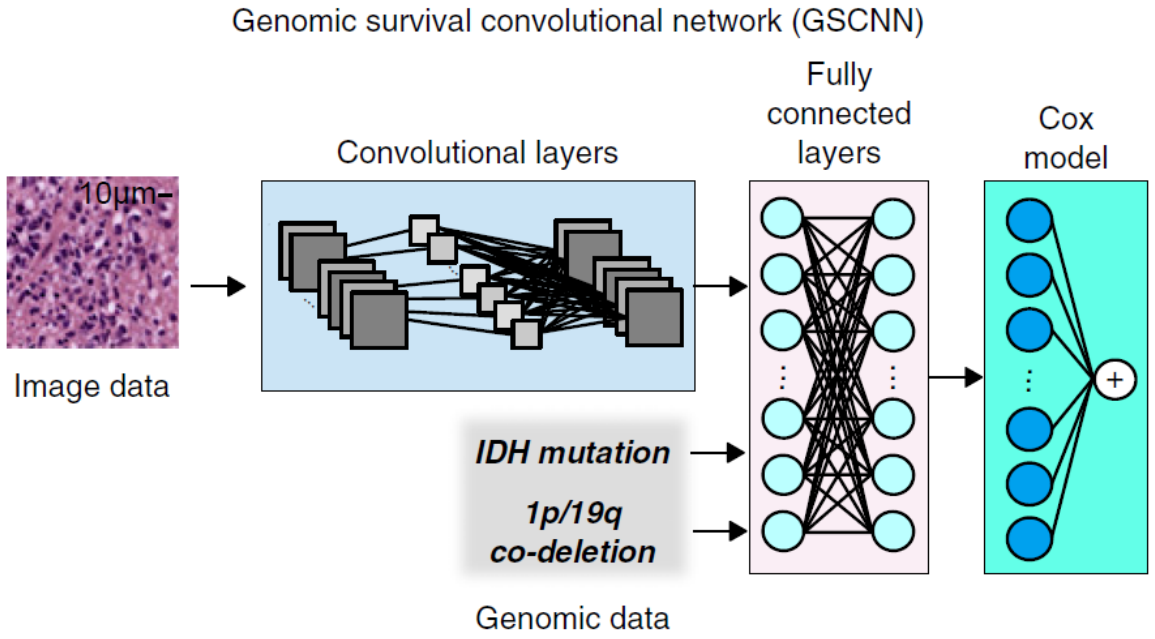
J.J. Levy et al, BMC Bioinformatics 2020

Erasmus MC

Survival Convolutional Network



J. Hao et al, 2019

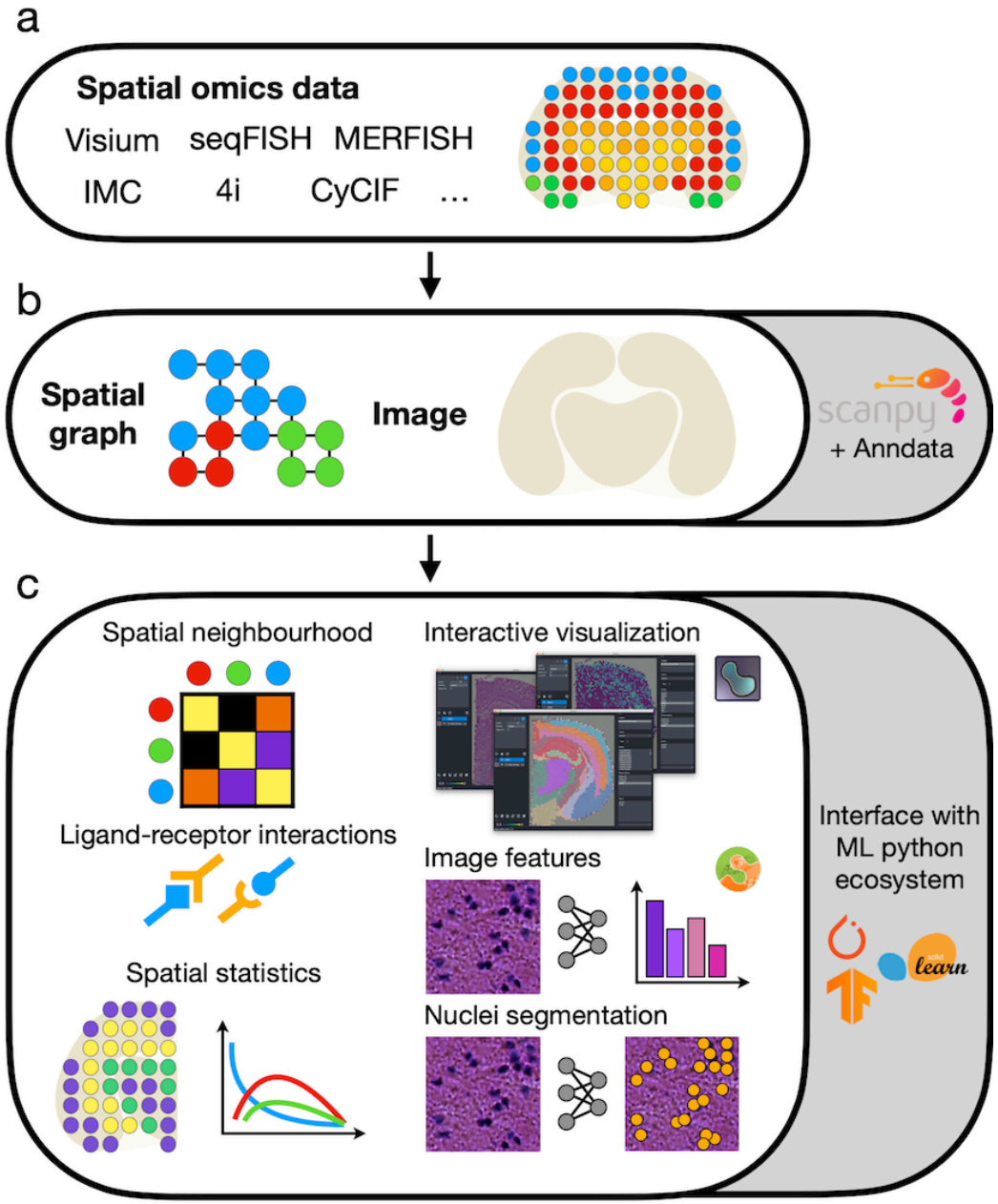


P. Mobadersany et al, PNAS 2018

Erasmus MC

squidpy

Spatial Single Cell Analysis



iMAP: integration of multiple single-cell datasets by adversarial paired transfer networks

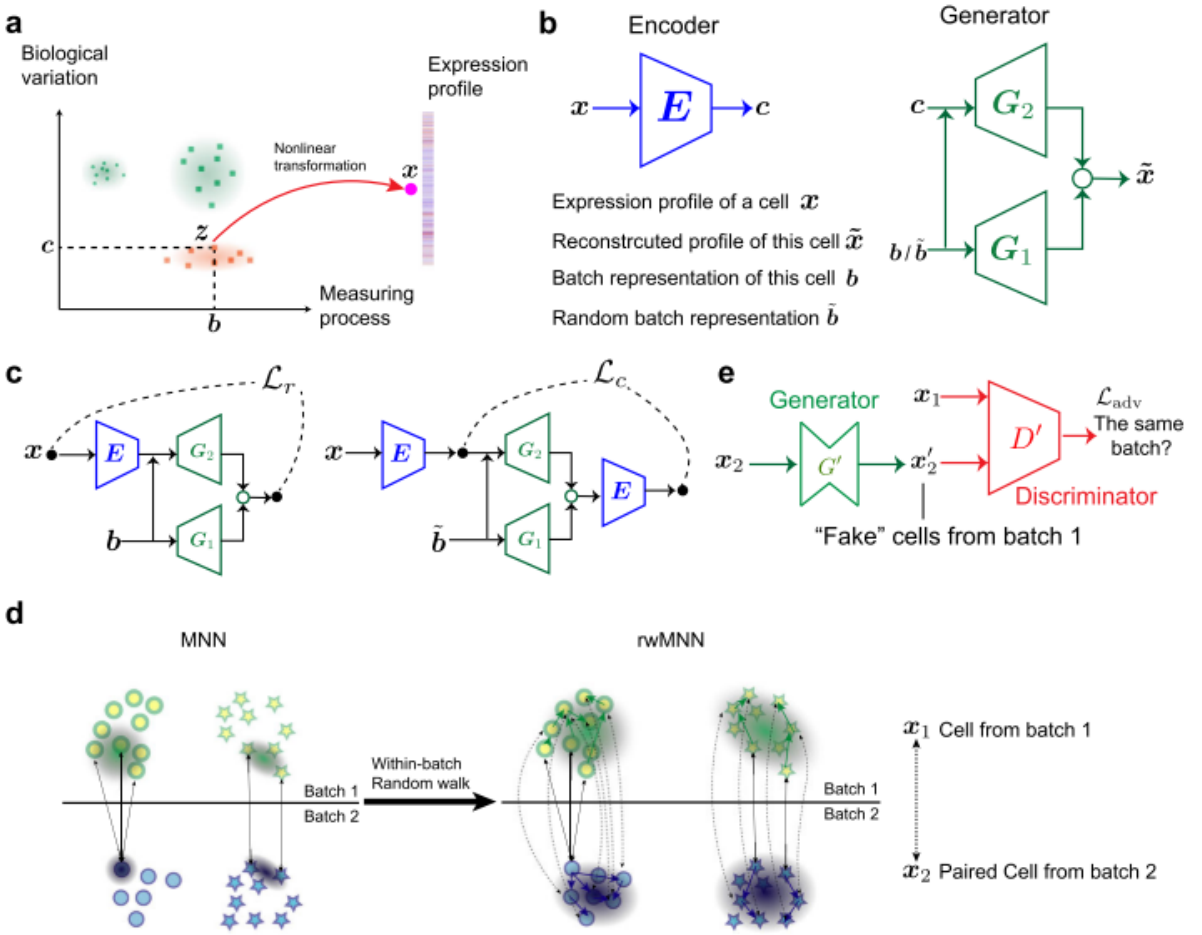
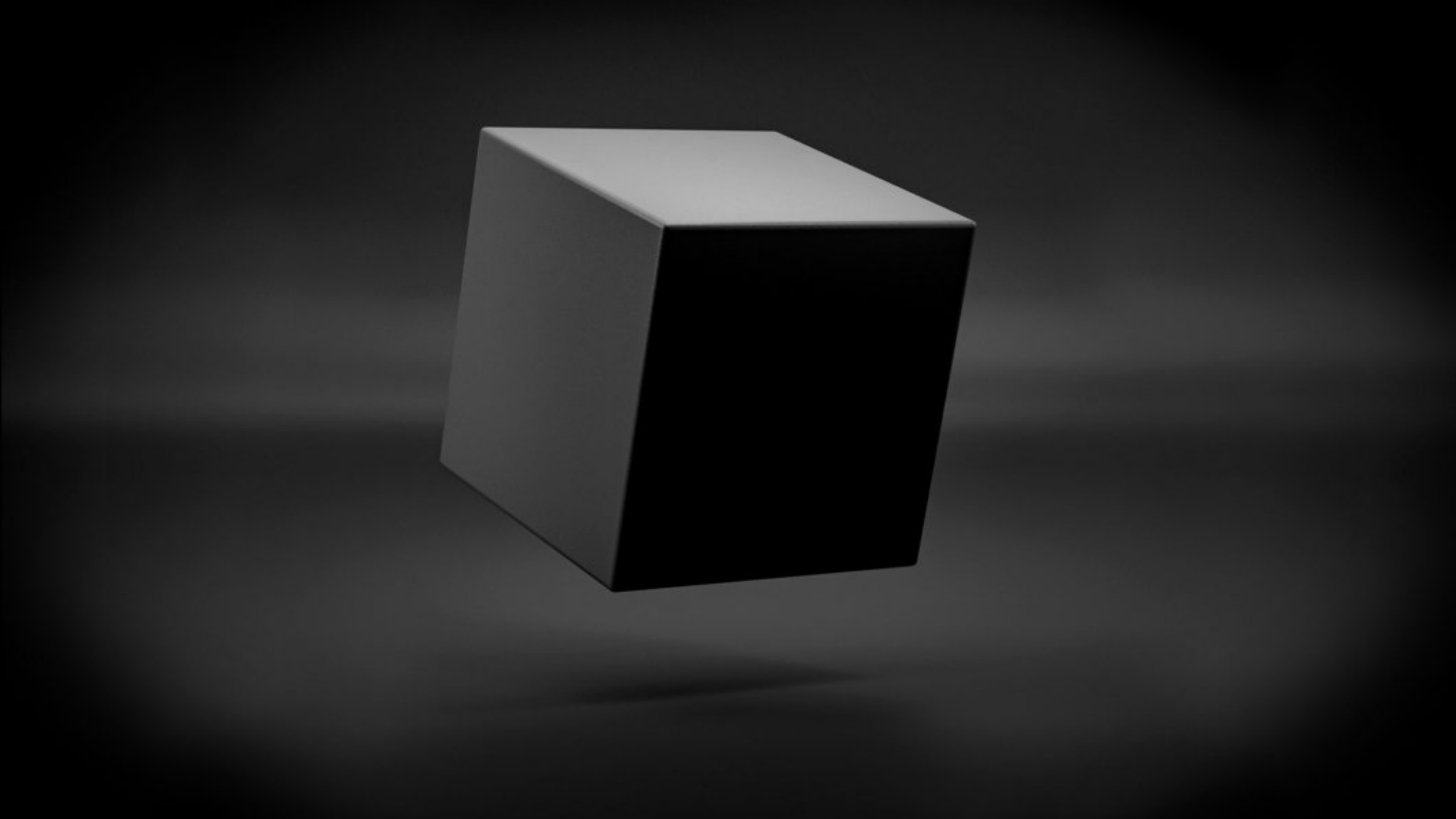
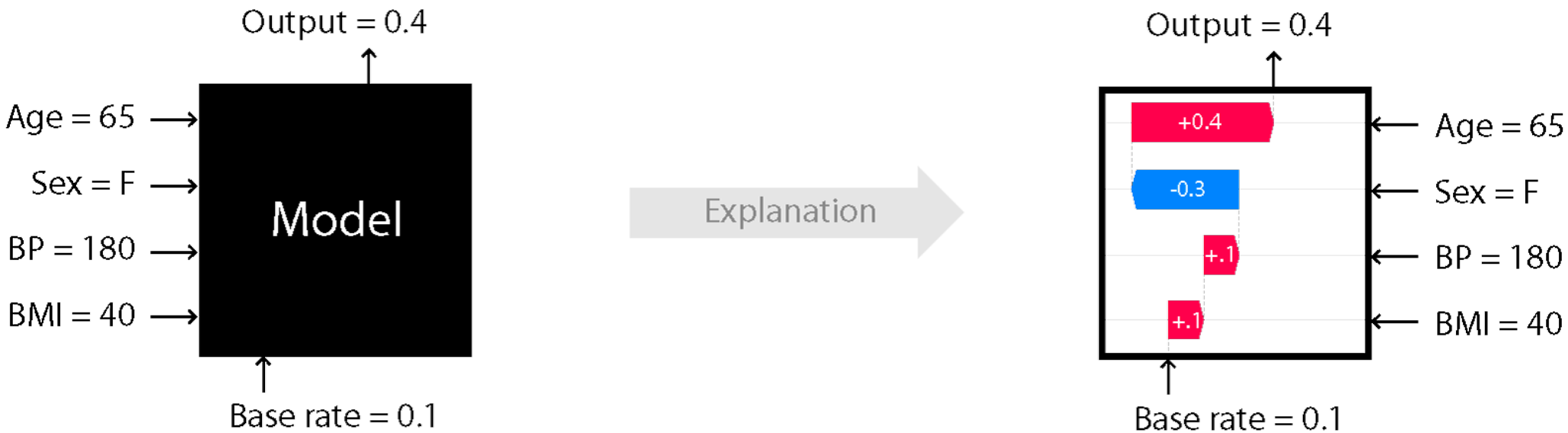


Fig. 1 Overview of the iMAP algorithm. **a** The expression profiles modeled by a function of two independent factors: biological variations and measuring processes. **b** Three feed-forward neural networks deployed for the first stage of iMAP. **c** Information flows and two losses functions used in the first stage. L_r, reconstruction loss; L_c, content loss. **d** A within-batch random walk-based procedure adopted to extend the MNN pairs. These extended MNN pairs are called rwMNN pairs. **e** A GAN structure used to remove the batch effects based on the rwMNN pairs

Erasmus MC

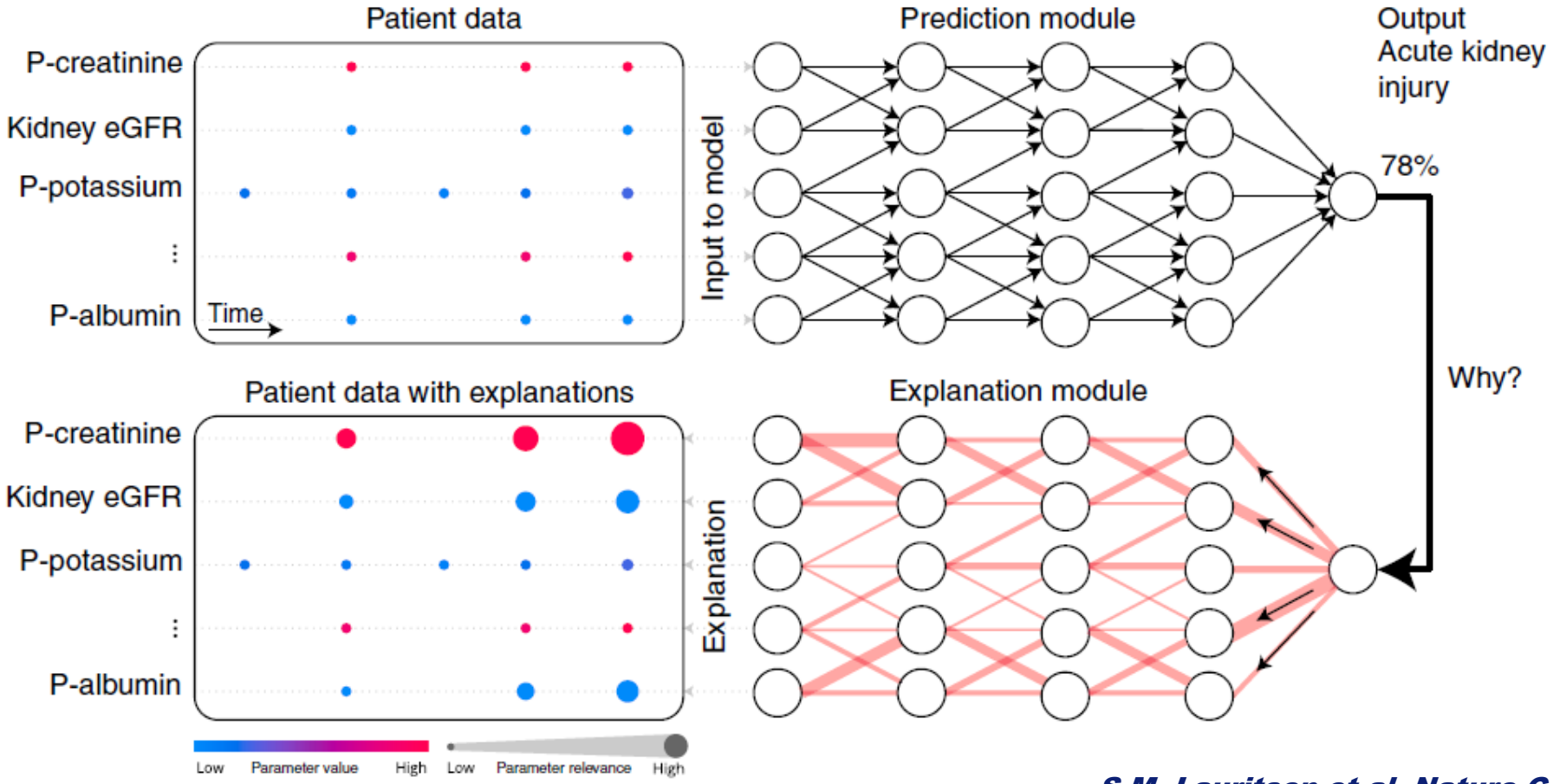


SHAP



https://github.com/slundberg/shap

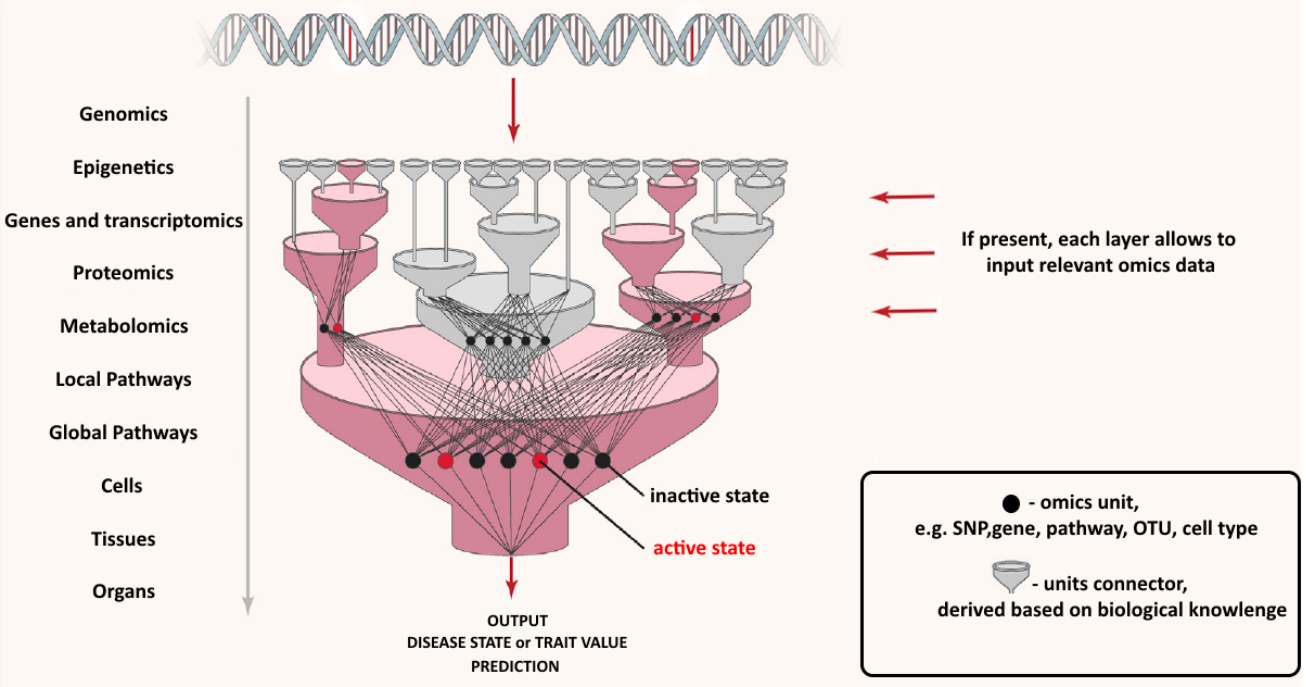
Explainable artificial intelligence model to predict acute critical illness from electronic health records



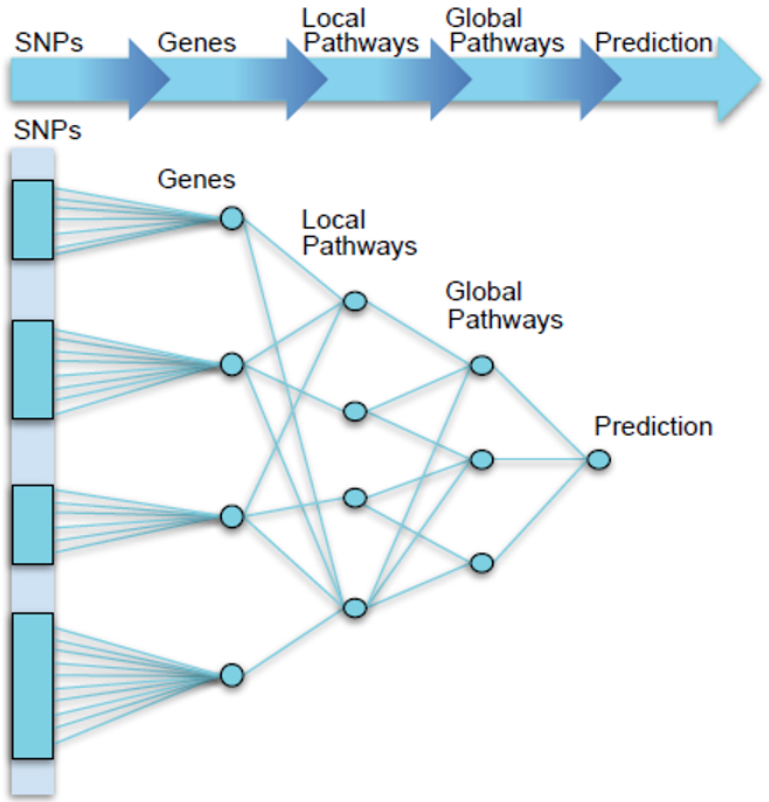
S.M. Lauritsen et al, Nature Communications 2020

Erasmus MC

Interpretable Neural Networks



M.K.Yu et al, Cell 2018



A. Hilten et al, BioRxiv 2020

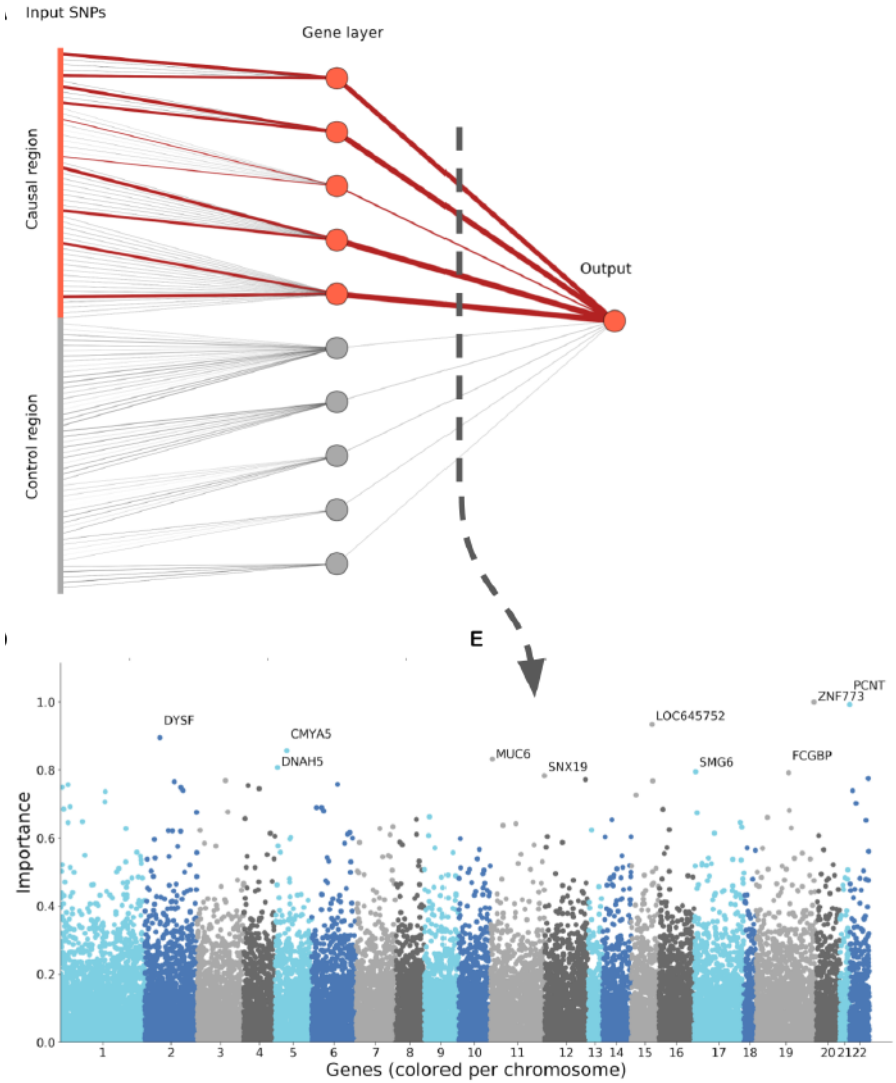
Erasmus MC

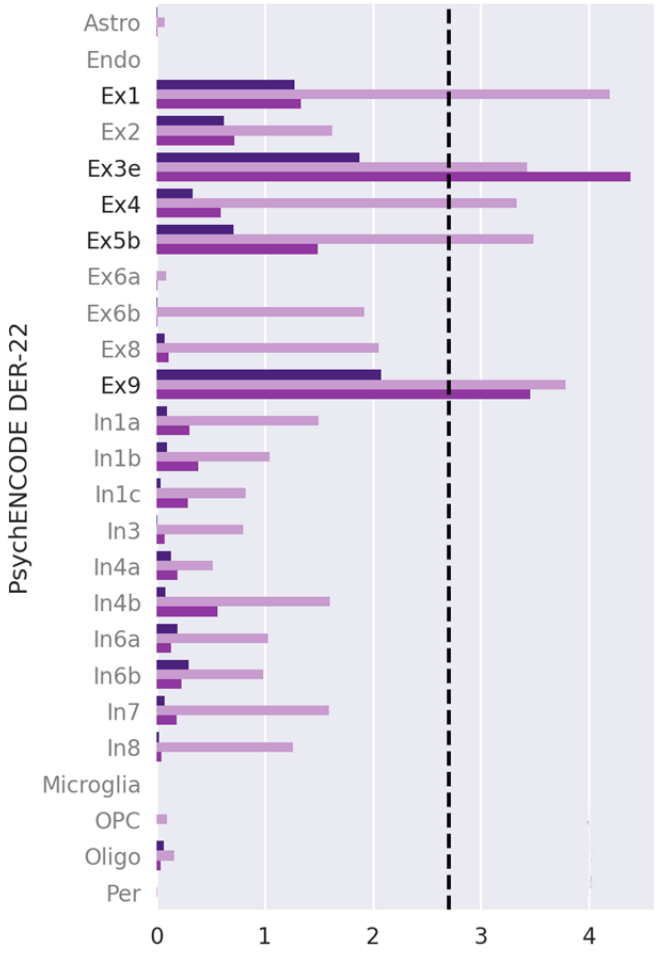
Biological Databases

Connections can be defined by:

- Gene annotations (Annovar)
- Pathway annotations (KEGG, Reactome)
- Cell type expression (Allen Brain Atlas)
- Tissue expression GTEx
- eQTL data
- mQTL dat
- …..
- Any other prior knowledge that groups data

Interpretable Neural Networks

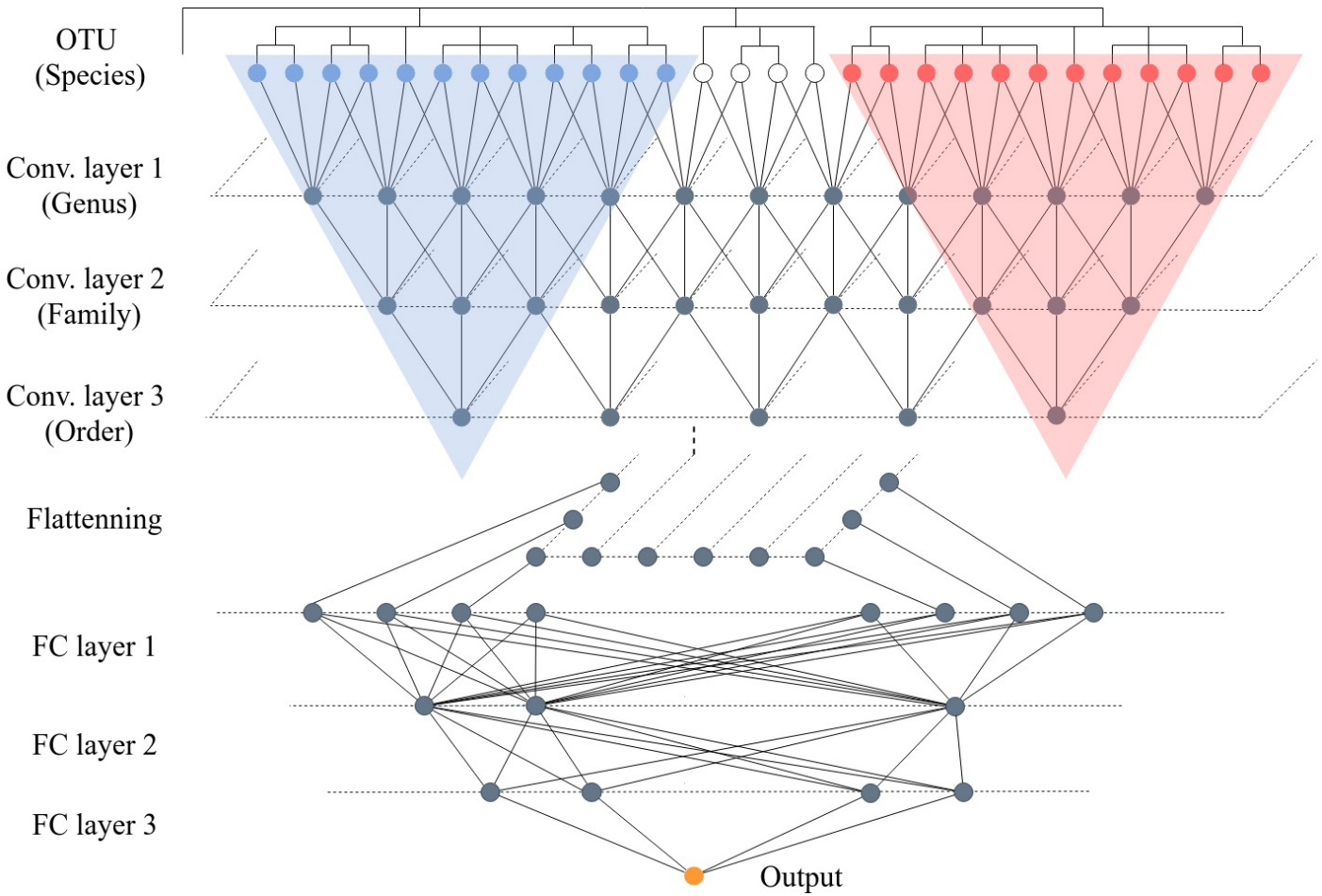




A. Hilten et al, BioRxiv 2020

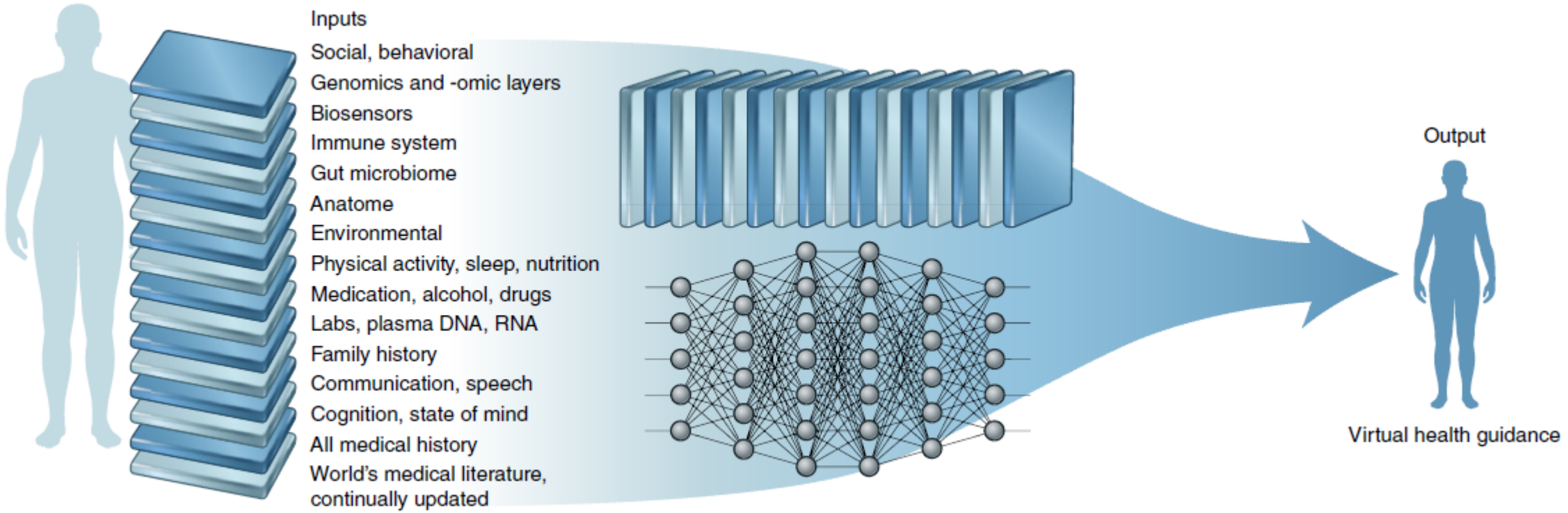
Erasmus MC

Microbiome Based Deep Learning Method

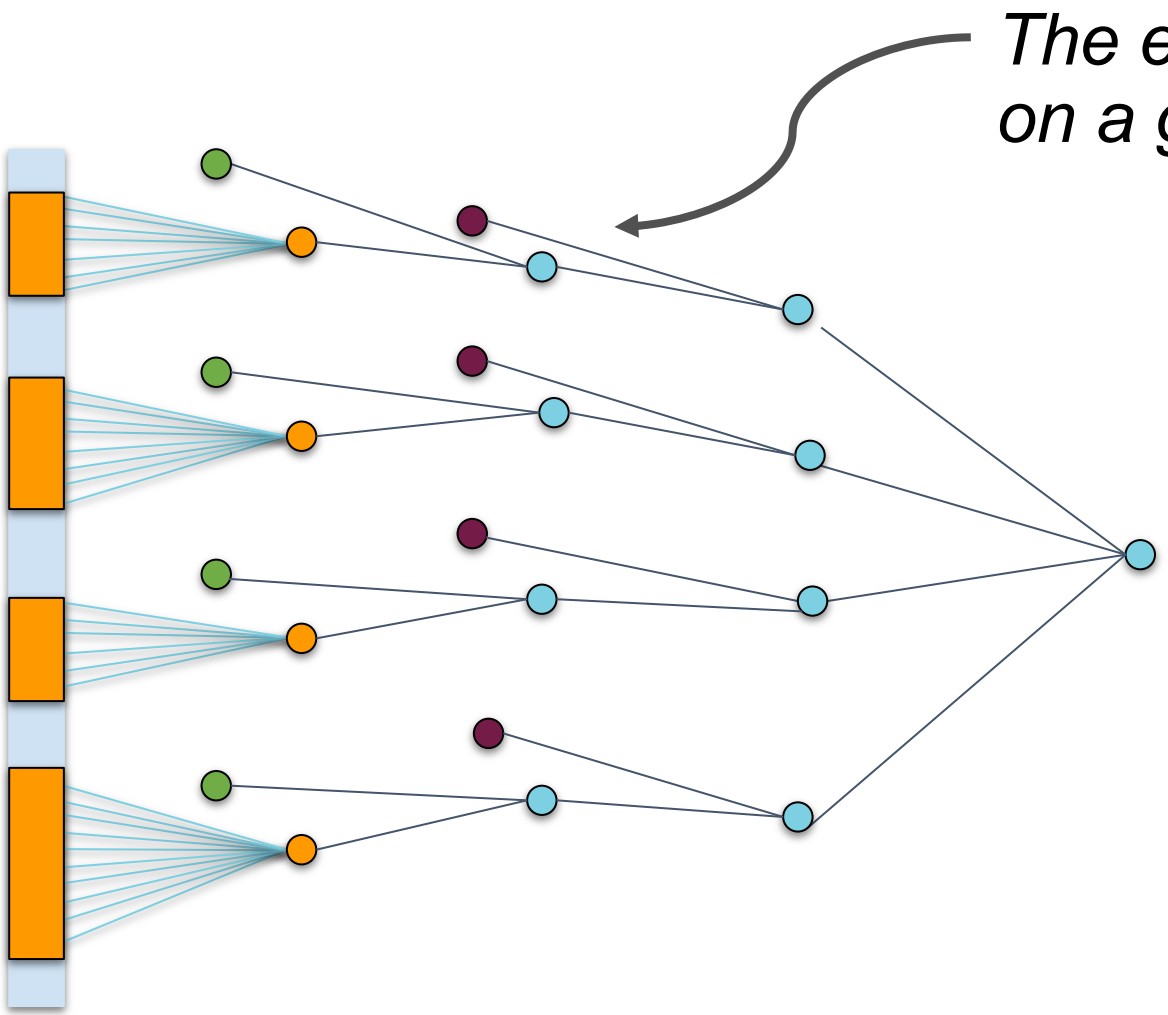


Y. Wang et al, Briefing in Bioinformatics 2020

Multi-Omics


Inputs
Social, behavioral
Genomics and -omic layers
Biosensors
Immune system
Gut microbiome
Anatome
Environmental
Physical activity, sleep, nutrition
Medication, alcohol, drugs
Labs, plasma DNA, RNA
Family history
Communication, speech
Cognition, state of mind
All medical history
World's medical literature, continually updated
Output
Virtual health guidance

Multi-Omics



Gene-expression

Methylation

Gene node

Covariates

BBMRI

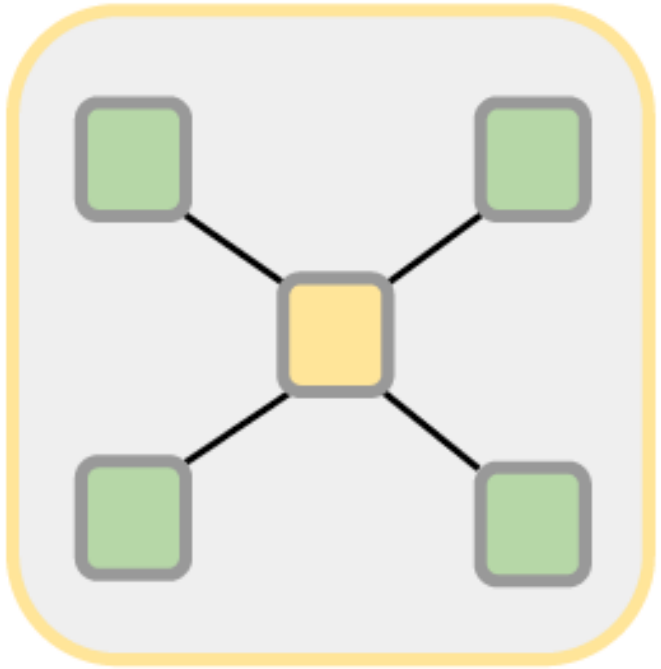
More accurate

Capturing complex patterns

Interpretable

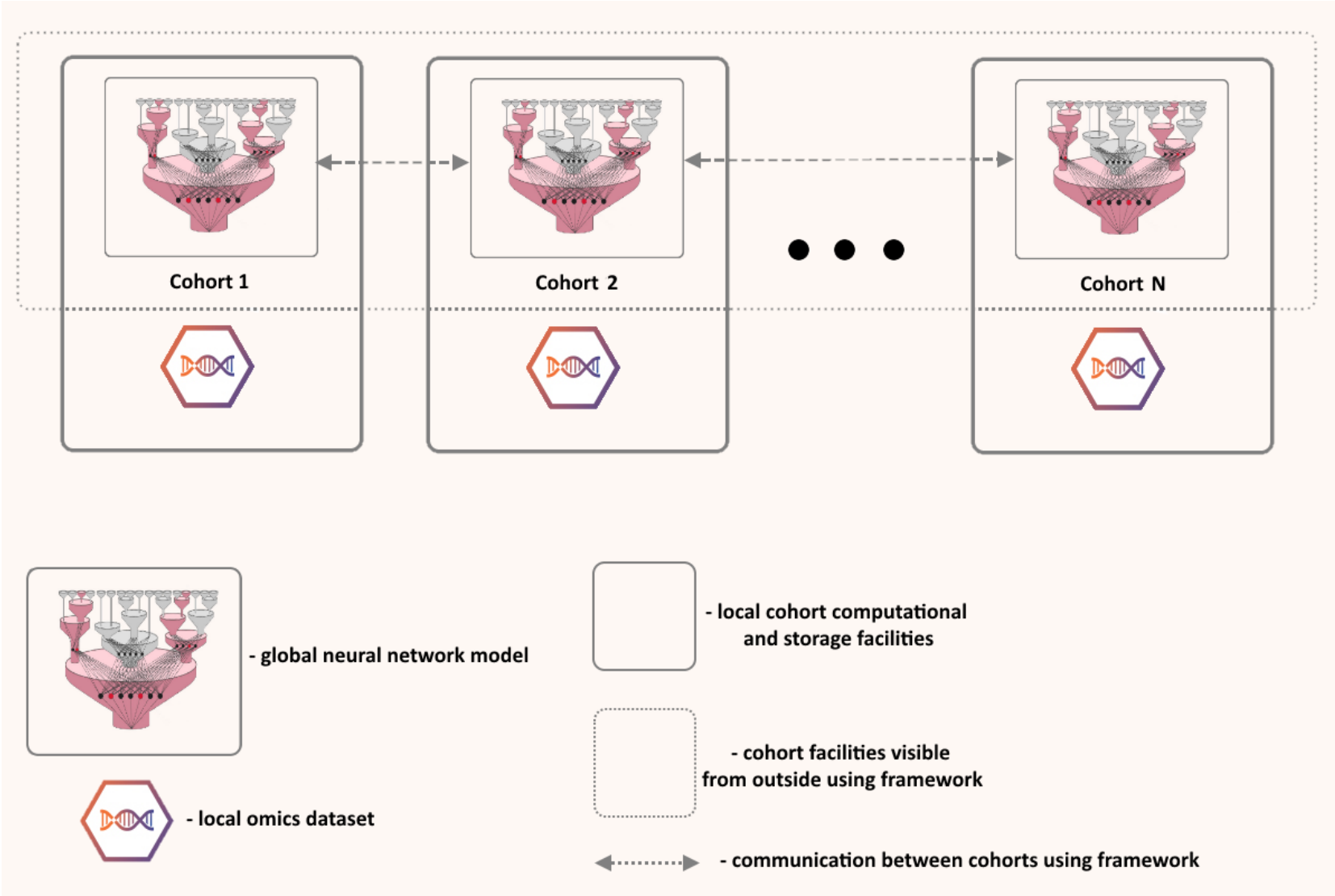
Erasmus MC

Federated Learning

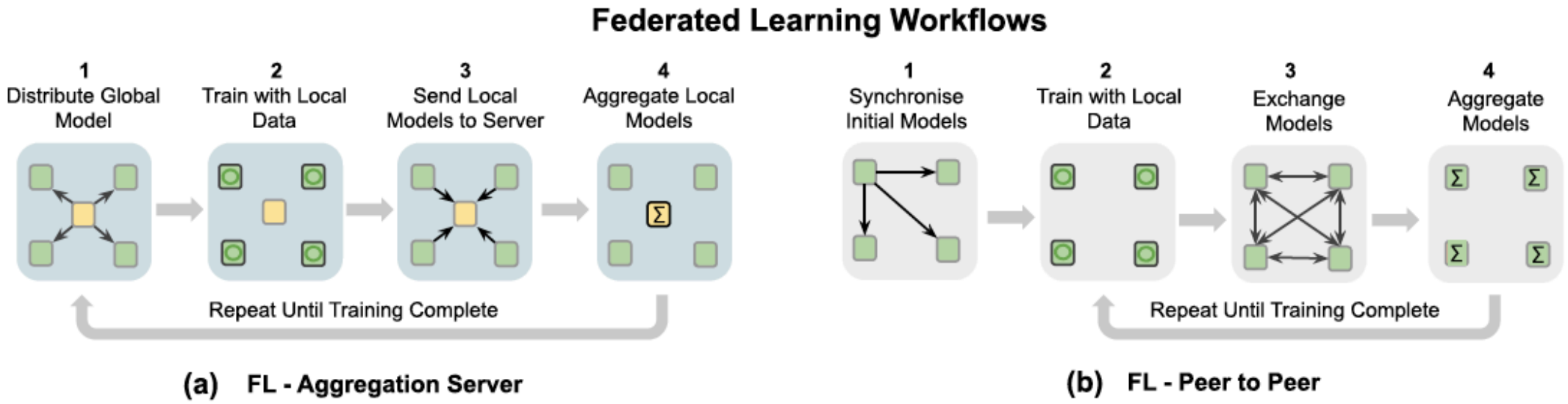


GWAS Meta-Analysis

Federated Learning

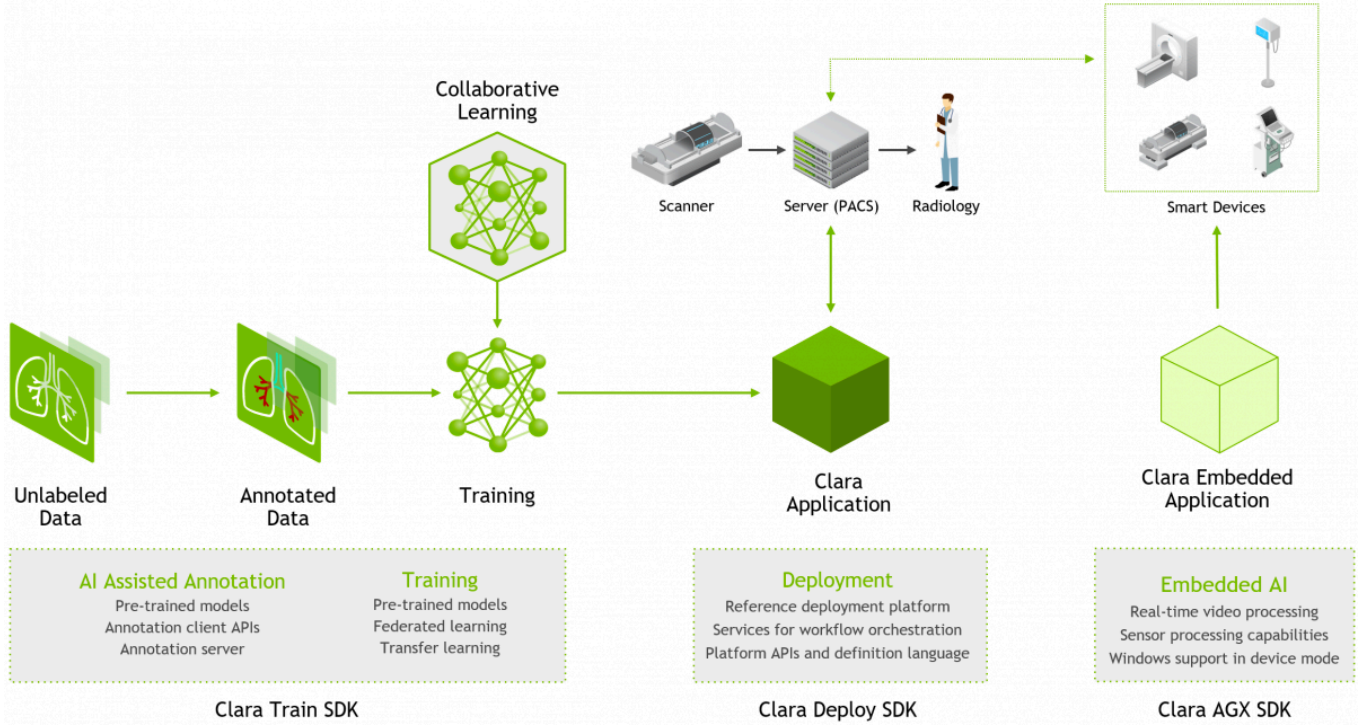


Federated Learning



MONAI+

Syft



Erasmus MC

Application of new methods to accelerate research

Developing of new methods based on expertise and data

Distributed learning

Software

TENSORFLOW

- Developed by : Google Brain team
- TensorBoard for effective data visualization
- Written in C++ and Python
- Need more programming experience
- Static computation graphs
- Has a larger community base
- Plethora of learning resources

PYTORCH

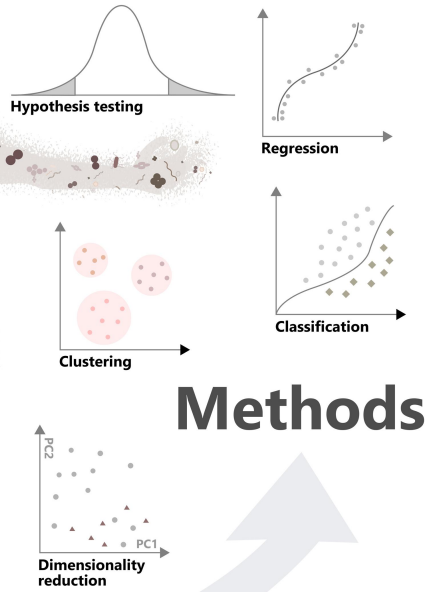
- Developed by : Facebook's AI research group
- Dynamic Computation graphs
- Easy to use as compared to TensorFlow
- Written in C and Python
- More popular amongst researchers
- Not regularly used as a production framework

KERAS

- Developed by : François Chollet, Google engineer
- High level library and hence it enables fast experimentation
- Uses TensorFlow at the backend
- Easiest framework to start your deep learning journey

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

ML4 MICROBIOME



Data

Methods

COST Action CA18139

GEMSTONE

Genomics of Musculoskeletal Traits Translational Network

CHARGE CONSORTIUM

COHORTS FOR HEART & AGING RESEARCH IN GENOMIC EPIDEMIOLOGY

MORENO-INDIAS, et .al., Frontiers in Microbiology, 2021

Erasmus MC

Links

- https://playground.tensorflow.org/
- https://juniorrojas.com/ff-net/
- https://www.cs.ryerson.ca/~aharley/vis/conv/
- https://poloclub.github.io/cnn-explainer/

Thank you!

g.roshchupkin@erasmusmc.nl

Twitter: @Roshchupkin

Erasmus MC