ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Machine learning for medical imaging

Daniel Remondini

DIFA – daniel.remondini@unibo.it

INFN Sezione di Bologna

State of the art – "Big data" challenge

Many available **public data** (BIG, i.e. *at least* Tb-sized):

Heterogeneous types of **information**
(imaging, omics, clinical)

Databases **connect and integrate** different data types
(genic & metabolic **networks**, clinical trials, in vitro experiments, catalogues of drug effects and targets)

Increased **computing** and **storage power** (HPC, GPU , Cloud)

Rapid availability and **management** of data

Public databases and repositories

Allow in silico meta-analyses and studies
Provide preliminary information before new experiments

National Human Genome Research Institute
ENCODE

GEO
Gene Expression Omnibus

RCSB PDB
PROTEIN DATA BANK

THE HUMAN PROTEIN ATLAS

Genomics of Drug Sensitivity in Cancer

EMBL-EBI

1000 Genomes
A Deep Catalog of Human Genetic Variation

Transcriptome, Epigenomics, Drugs, Clinical trials, protein structure, …

Big Data for biomedical studies

TCIA – the cancer imaging archive

TC, PET, NMR data

TCGA – The Cancer Genetic Atlas

Omics (GEP, NGS, SNP, MET)

Matched tumor & normal tissues from more than **11,000** patients, representing **33** cancer types.

TCIA-TCGA integrated imaging/omics databases

Collection	Cancer type	Modalities	#
TCGA-BLCA	Bladder Endothelial Carcinoma	CT, CR, MR, PT	106
TCGA-BRCA	Breast Cancer	MR, MG	139
TCGA-CESC	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma	MR	54
TCGA-COAD	Colon Adenocarcinoma	CT	25
TCGA-ESCA	Esophageal Carcinoma	CT	16
TCGA-GBM	Glioblastoma Multiforme	MR, CT, DX	262
TCGA-HNSC	Head and Neck Squamous Cell Carcinoma	CT, MR, PT, RTSTRUCT, RTPLAN, RTDOSE	227
TCGA-KICH	Kidney Chromophobe	CT, MR	15
TCGA-KIRC	Kidney Renal Clear Cell Carcinoma	CT, MR, CR	267
TCGA-KIRP	Kidney Renal Papillary Cell Carcinoma	CT, MR, PT	33
TCGA-LGG	Low Grade Glioma	MR, CT	199
TCGA-LIHC	Liver Hepatocellular Carcinoma	MR, CT, PT	97
TCGA-LUAD	Lung Adenocarcinoma	CT, PT, NM	69
TCGA-LUSC	Lung Squamous Cell Carcinoma	CT, NM, PT	37
TCGA-OV	Ovarian Serous Cystadenocarcinoma	CT, MR	143
TCGA-PRAD	Prostate Cancer	CT, PT, MR	14
TCGA-READ	Rectum Adenocarcinoma	CT, MR	3
TCGA-SARC	Sarcomas	CT, MR	5
TCGA-STAD	Stomach Adenocarcinoma	CT	46
TCGA-THCA	Thyroid Cancer	CT, PT	6
TCGA-UCEC	Uterine Corpus Endometrial Carcinoma	CT, CR, MR, PT	58

TCIA – public database of biomedical imaging for many tumours

For 21 tumours also omics data are available in TCGA (same samples)

Big Data for clinical studies

ADNI – Alzheimer Disease Neuroimaging Initiative
(imaging, omics, clinical, biospecimens)

ABIDE – Autism Brain Imaging Data Exchange
(imaging data on control and autistic samples)

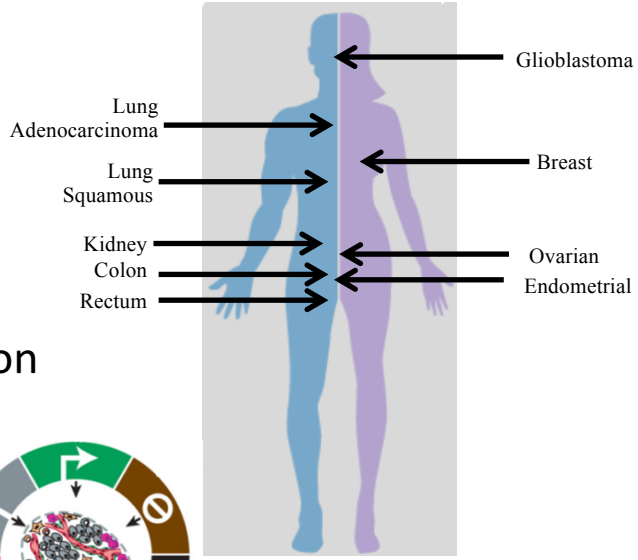
IXI – Information eXtraction from Images
(600 normal healthy subjects MRI)

Omics multivariate analysis: examples

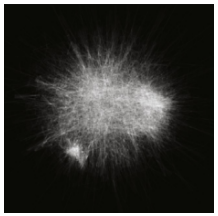
TCGA database

11 Cancer types

> 2000 samples



Drug **repurposing**
Target identification



BioPlex
Protein-Protein Interaction Network

Ontocancro
Genes annotated in cancer-related pathways

nature COMMUNICATIONS 2018

DOI: 10.1038/s41467-018-06992-7 OPEN

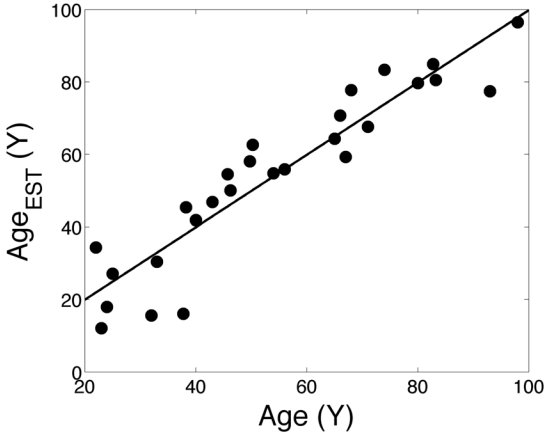
Network integration of multi-tumour omics data suggests novel targeting strategies

Ítalo Faria do Valle[1,2], Giulia Menichetti[3], Giorgia Simonetti[4], Samantha Bruno[4], Isabella Zironi[1], Danielle Fernandes Durso[4,5], José C.M. Mombach[6], Giovanni Martinelli[4,7], Gastone Castellani[1] & Daniel Remondini[1]

Identification of a T cell gene expression clock obtained by exploiting a MZ twin design

Daniel Remondini[1,2], Nathan Intrator[3], Claudia Sala[1], Michela Pierini[4,6], Paolo Garagnani[2,4], Isabella Zironi[1], Claudio Franceschi[5], Stefano Salvioli[2,4] & Gastone Castellani[1,2]

SCIENTIFIC REPORTS 2017



125-gene **signature** to predict sample age from blood cells

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

Machine Learning

Huge amount of data: ***data-driven*** approaches

Machine Learning: usage of **advanced algorithms** for data analysis (e.g. image analysis)

1) ***Unsupervised methods***:

   a) data **clustering** (e.g. ROI segmentation)

   b) **feature extraction** (e.g. texture features)

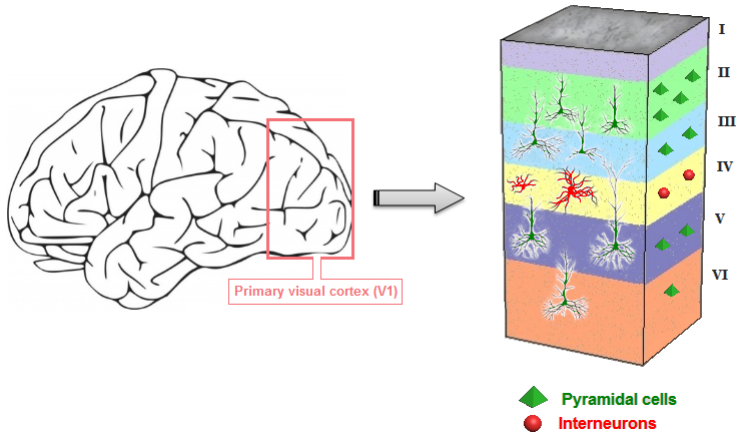
2) ***Supervised methods***: the algorithm uses known information (e.g. reference samples, standards) for

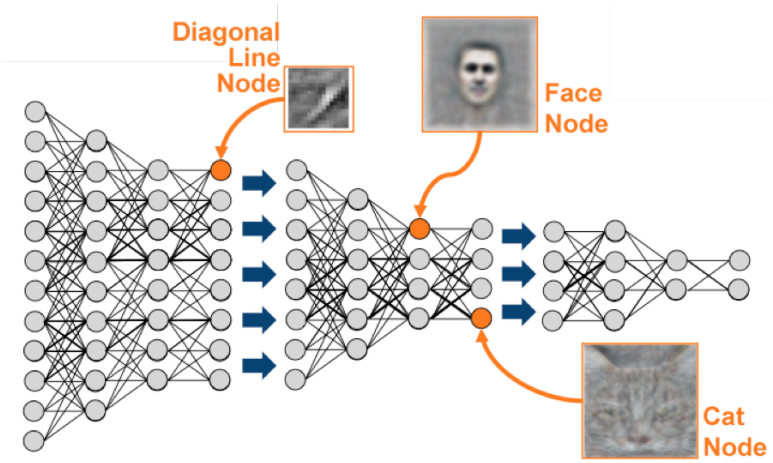
   a) sample **classification**

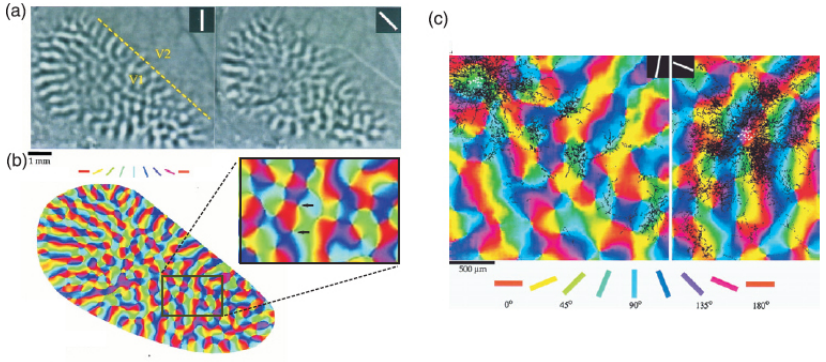
   b) parameter **regression** (e.g. risk score, age)

Artificial Intelligence and Deep Learning

Some machine learning techniques take inspiration from **anatomical** and **functional** structures of the **brain** (i.e. visual cortex, known since the '60s)





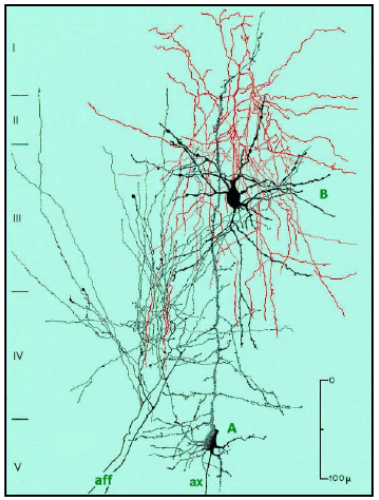


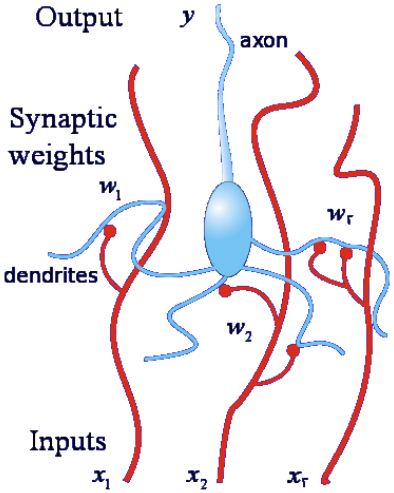
Layered (modular, organized)
Hierarchical (from contours to shapes)
Somatotopic (organization)

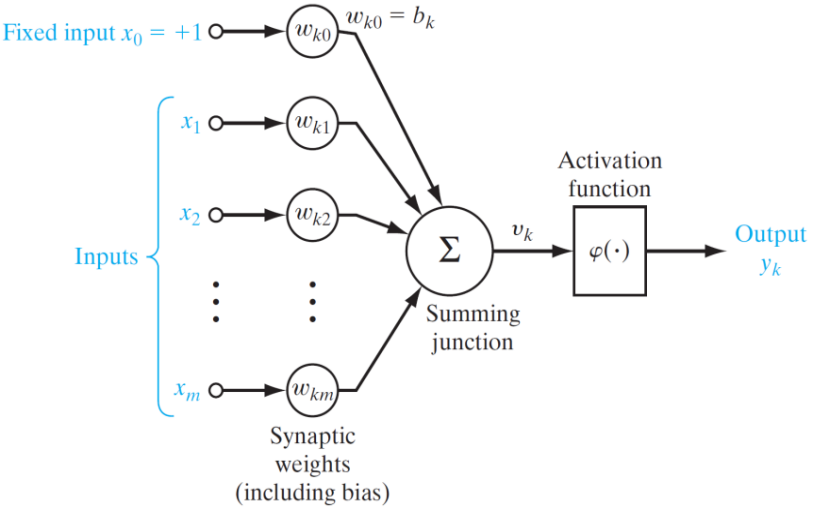
Hubel & Wiesel J Physiol 160, 1962

Artificial Intelligence and Deep Learning

The functional units of Neural Networks take inspiration from neurons (since 1957 Rosenblatt's *perceptron*)

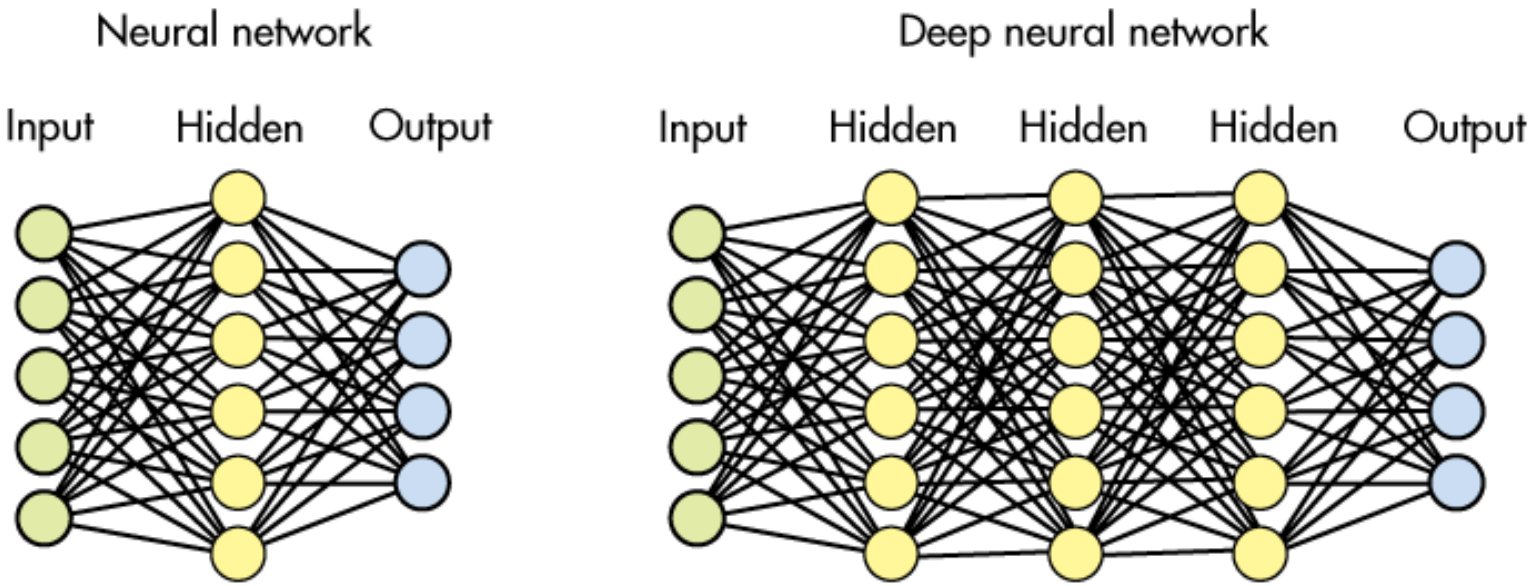






Deep Neural Network for Machine Learning - supervised

Supervised methods (classification, regression):
Feedforward Deep Networks FDN
FDN is **trained** with examples, and **generalizes** to unseen data

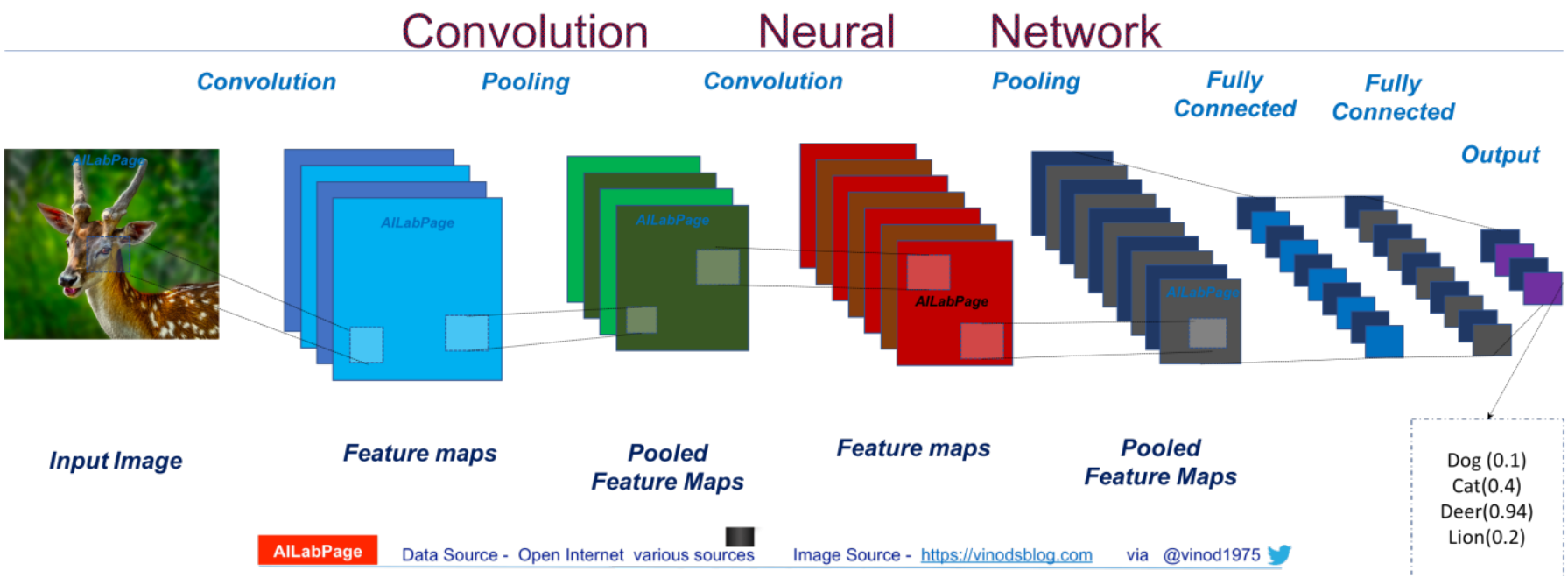


My thesis (1996): 4 neurons, 386 Intel CPU, 20 Mb HD
Actual FDN: 10^5-10^6 neurons, multi-core CPU & GPU, Tb HD (& RAM)

Deep Neural Network for Machine Learning - unsupervised

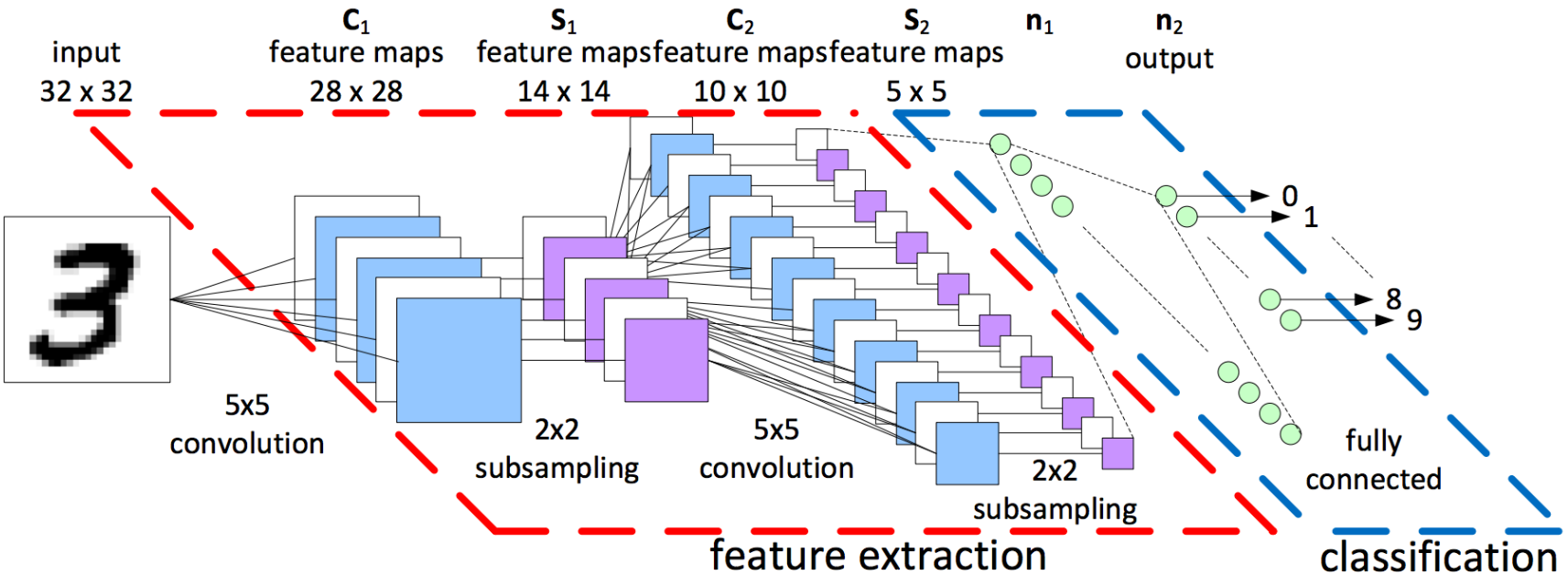
Unsupervised feature extraction: Convolutional Neural Networks CNN

CNN layers extract a **hierarchy** of **features** (from contours to shapes)



Deep Neural Network for Machine Learning

Typical DNN architecture: encoding (CNN) + task (FDN)



Machine Learning for NMR data: examples

Data processing:

1) NMR fingerprinting

2) QSM processing

Image analysis:

1) Automated segmentation

2) Quantification (feature & texture analysis)

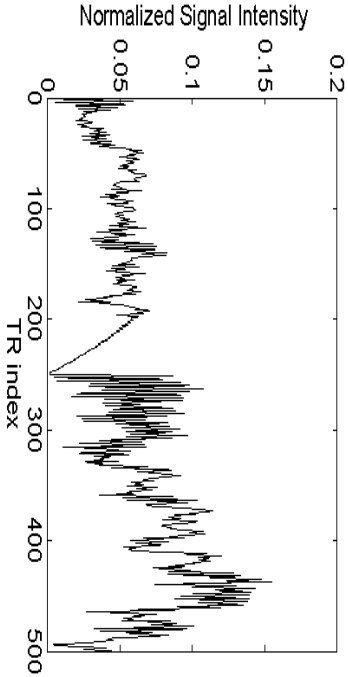
3) Image quality enhancement (super-resolution)

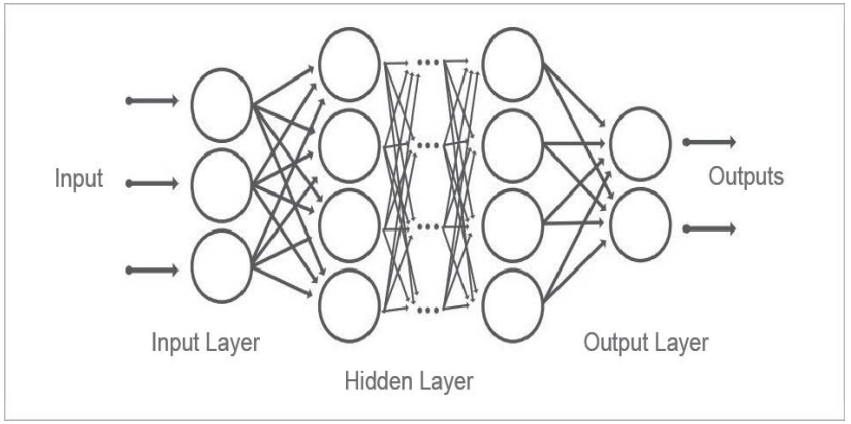
Fingerprinting

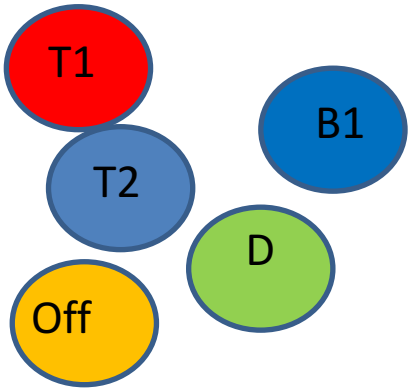
Associate vectors of features for each MRI "pixel" (training set) with specific values (B, T1, T2, …)

Original strategy [*Ma et al, Nature 495, 2013*]: define a "dictionary" of feature/value associations

Our strategy: train a DL network to discriminate the feature vectors and reliably associate the physical NMR parameters





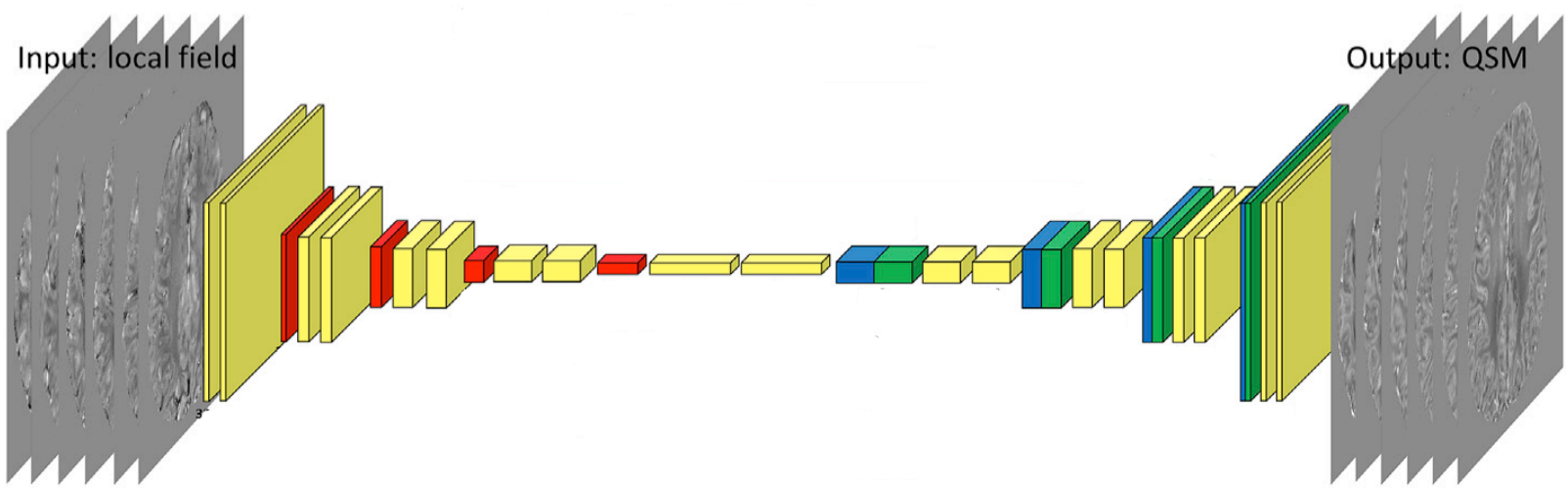


Barbieri, … Remondini arXiv:1811.11477

Quantitative Susceptibility Mapping

Transform Phase data into Magnetic susceptibility data χ
Analytical function has singularities ("magic angle")

Deep Learning can learn the transform from "good"examples (our case, reconstructions obtained with other approaches) and **overcome** singularities



Cristiana Fiscone's Tomorrow C15 talk

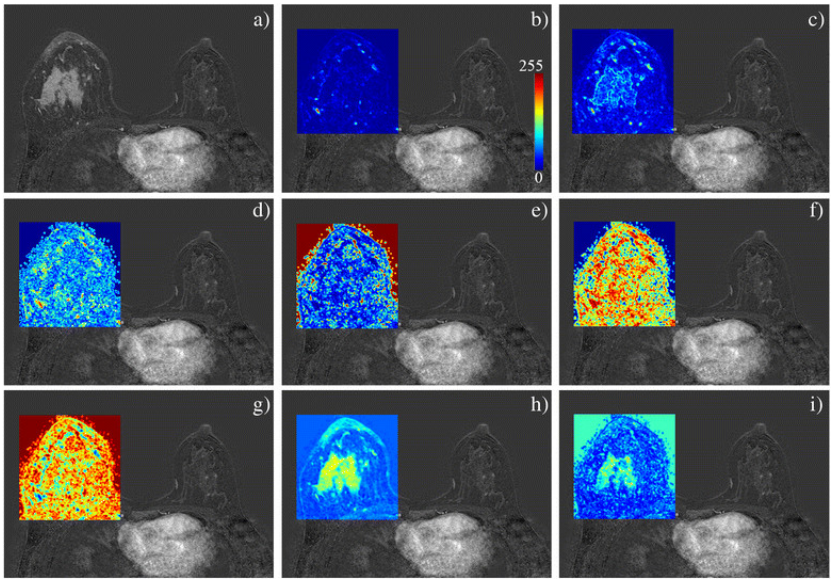
In collaboration with Prof. R. Bowtell
Sir Peter Mansfield Institute, Nottingham UK

Feature extraction & analysis

Many observables can be extracted from single pixels (or larger patches) of MRI images

- Graylevel histogram
- Texture features (based on spatial and intensity proximity)
- Segmented region parameters (eccentricity, complexity, fractal dim, …)

Each sample is mapped into a high-dimensional feature space



$N = 10^2\text{-}10^3$

Feature vectors can be used for

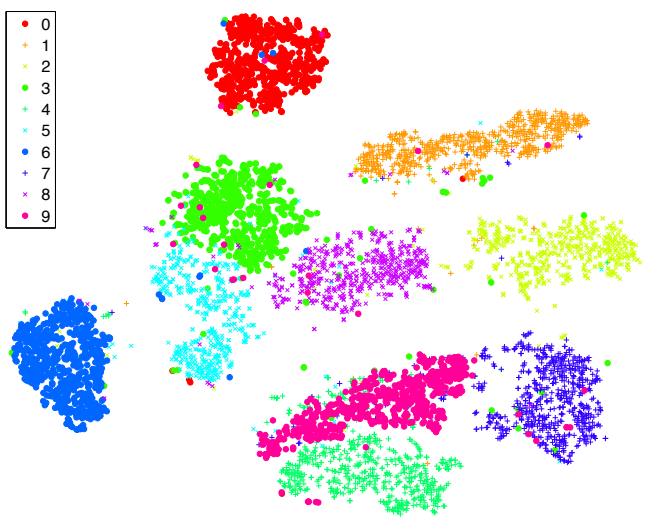
- low-dimensional visualization
- Supervised and unsupervised machine learning

Visualization

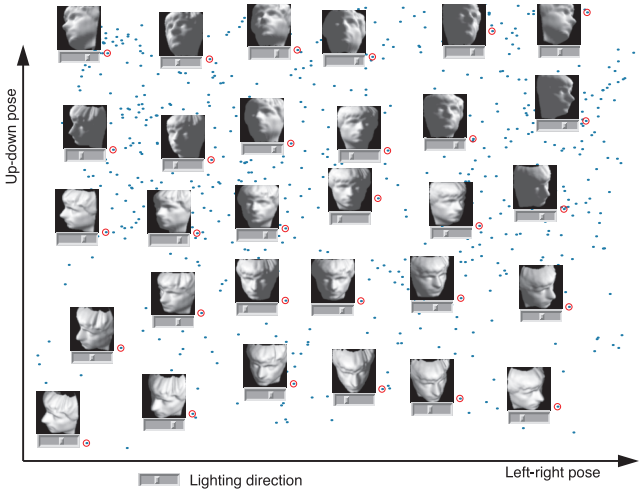
Several tecniques can be used for low-dimensional reduction and visualization (in 2-3 dim)

- PCA/SVD "family" of methods
- ISOMAP (network-geodetics)
- SNE (Stochastic Neighbor Embedding)

SNE 2-d representation of hand-written digit images



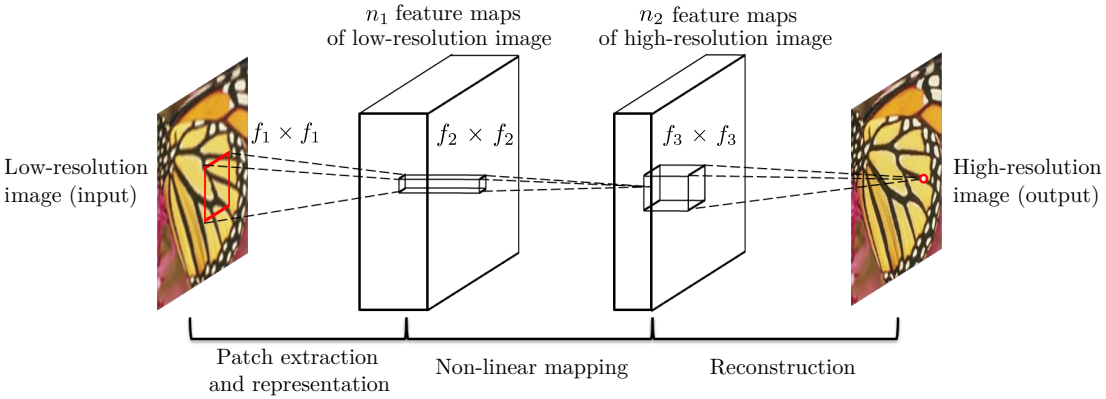
ISOMAP 2-d representation of face images

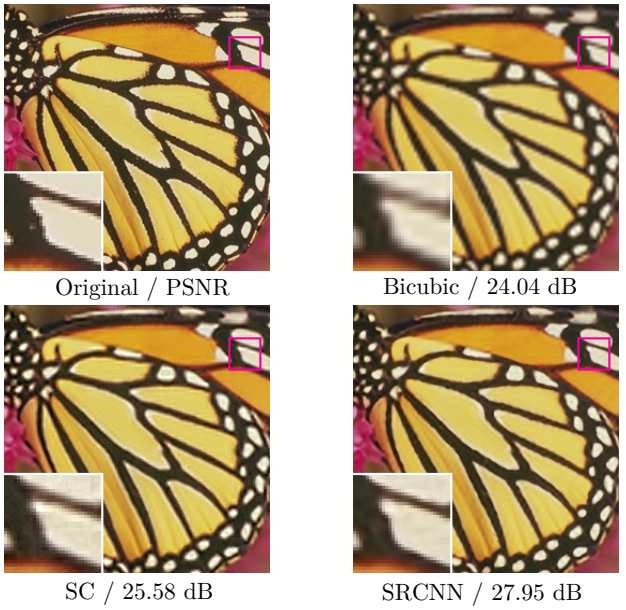


Super resolution

A DNN can learn to "improve" image quality (resolution) from an adequate training set

Results of a DNN trained on natural images





Original image 0887 from DIV2K

HR (PSNR/SSIM) | Bicubic (23.93 dB/0.729)

WDSR (25.31 dB/0.788) | EDSR (25.37 dB/0.791)

Dong et al., arXiv:1501.00092v3

Classification & regression

Machine learning techniques (including DL) can be used to **classify** samples or to **regress** parameters (e.g. tumour grade, age):

- Partial Least Squares
- Support Vector Machine
- Discriminant Analysis
- K-Nearest Neighbour
- Ridge regression
- LASSO
- …

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Daniel Remondini

Department of Physics and Astronomy – DIFA
INFN Sezione di Bologna – AIM initiative (Artificial Intelligence in Medicine)

daniel.remondini@unibo.it

Per quanto concerne i moderatori, relatori, formatori, tutor, docenti è richiesta dall'Accordo Stato-Regioni vigente apposita dichiarazione esplicita dell'interessato, di trasparenza delle fonti di finanziamento e dei rapporti con soggetti portatori di interessi commerciali relativi agli ultimi due anni dalla data dell'evento.
La documentazione deve essere disponibile presso il Provider e conservata per almeno 5 anni.

Dichiarazione sul Conflitto di Interessi

Il sottoscritto DANIEL REMONDINI in qualità di:

□ relatore

dell'evento "X CONGRESSO AIRMM - RISONANZA MAGNETICA IN MEDICINA 2019:
DALLA RICERCA TECNOLOGICA AVANZATA ALLA PRATICA CLINICA"
Milano, 28-29 marzo 2019

da tenersi per conto di **Biomedia srl Provider n. 148,**

ai sensi dell'Accordo Stato-Regione in materia di formazione continua nel settore "Salute" (Formazione ECM) vigente,

Dichiara

X che negli ultimi due anni NON ha avuto rapporti anche di finanziamento con soggetti portatori di interessi commerciali in campo sanitario

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA