FONDAZIONE BRUNO KESSLER

The challenge of reproducibility in Deep Learning at scale

Cesare Furlanello

@furlanello

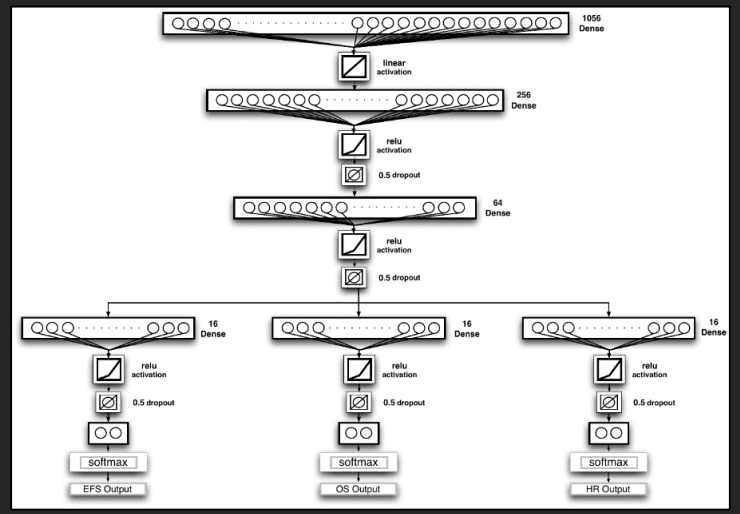
15 June 2018

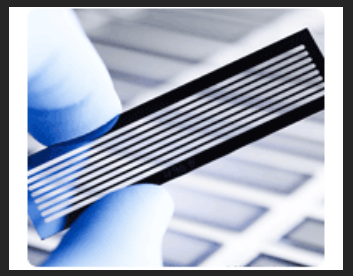
MPBA
PREDICTIVE MODELS FOR BIOMEDICINE & ENVIRONMENT

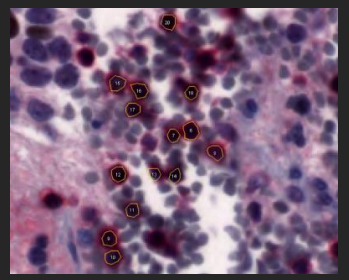
Fondazione Bruno Kessler (FBK)
Center for Information & Communication Technology
Via Sommarive 18 – 38123 Trento (Italy)

With Valerio Maggio, Stefano Fioravanzo,
Marco Cristoforetti, G. Jurman

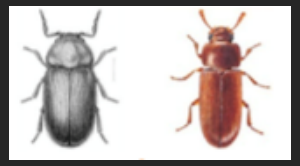
MPBA: DEEP LEARNING PER “MASSIVE DATA”





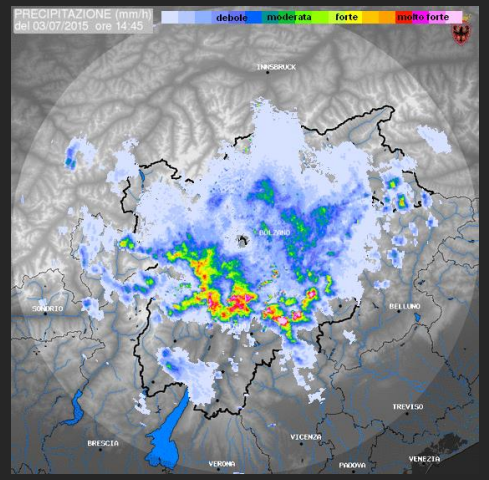


Multi-task Deep Learning per Massive Sequencing Data & Bioimages



Food safety

SEQC 2018: QC e riproducibilità di massive omics per tossicologia e Precision Medicine



Spatio-temporal NOWCASTING

AGRITEC: Stime e mappe predittive di Qualità, Maturazione, e Produzione da GIS, immagini & spettrometria low cost

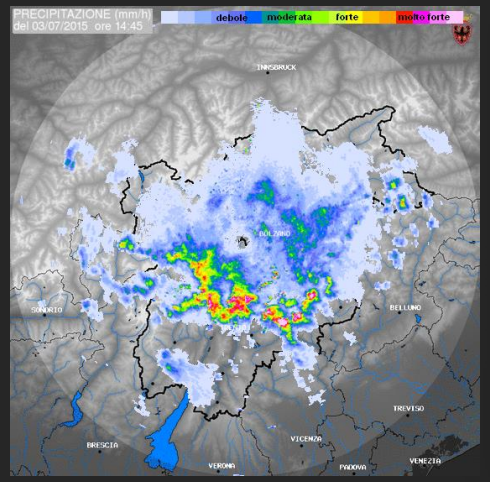
Spatio-temporal NOWCASTING (Conv-LSTM)

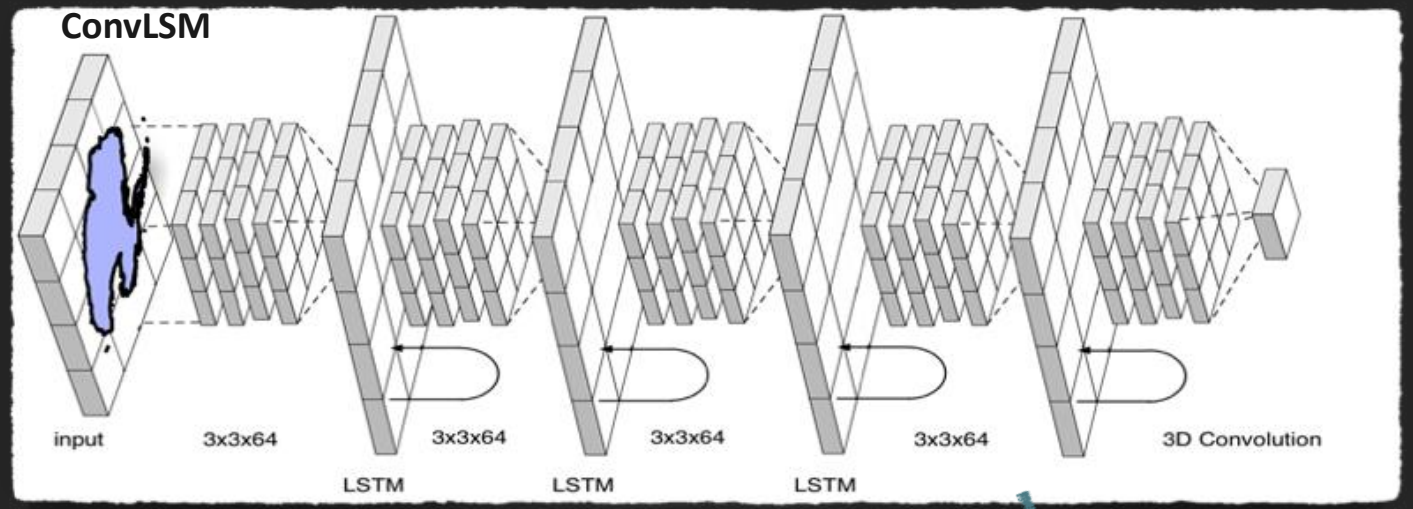
Gabriele Franch

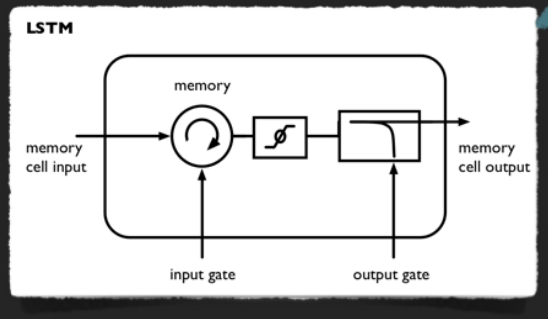
PhD Student

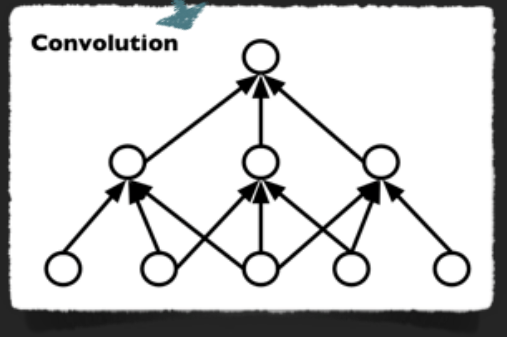
Rain & lightning nowcasting (5' - 75')

Short-time radar prediction: target for **deep learning**



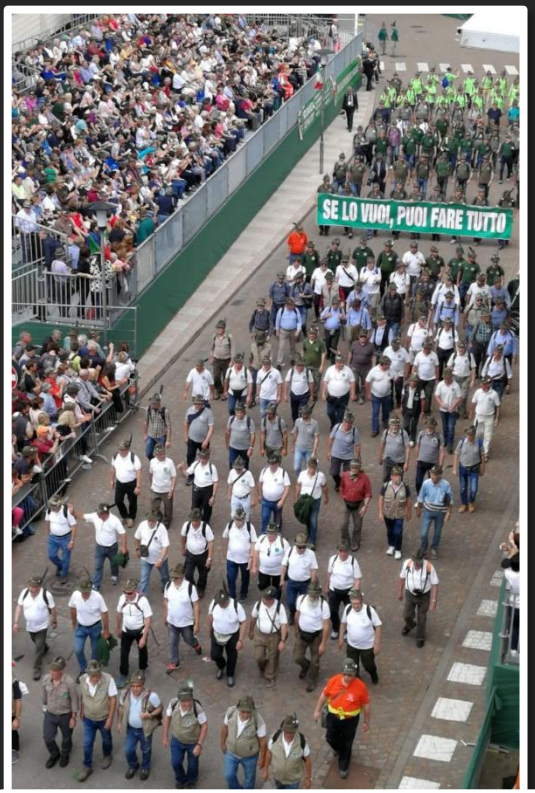


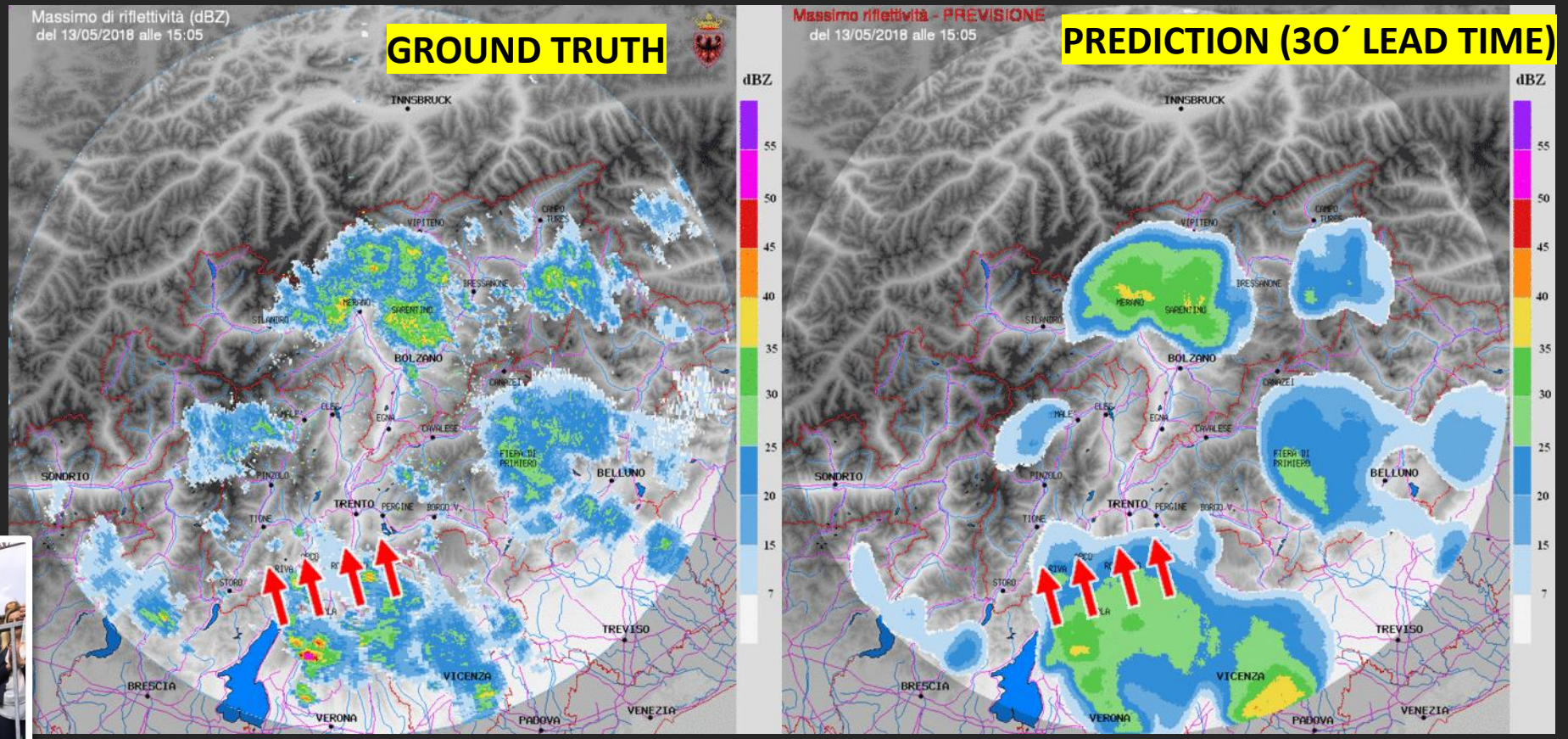




C. Furlanello, G. Franch – MPBA June 2018

Adunata Alpini (13/05/18)





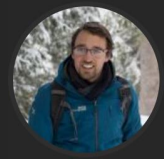
600.000 persone presenti alla parata.

Pattern di precipitazione molto complessi.

Previsione corretta (no pioggia su Trento) con 30´ di anticipo

Gabriele Franch

PhD Student



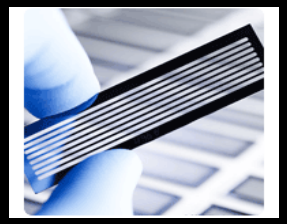
NOWCASTING: Live Cloud Architecture

Gabriele Franch

PhD Student

C. Furlanello – MPBA Dec 2017

MACHINE LEARNING FBK PER BIOMEDICINA



Next Generation Sequencing

Research

ResHosp

Pharma

Gut microbiota,
in health and disease

BIOMARKER PREDITTIVI e RIPRODUCIBILITÁ

- Nature 2014
- 4 Nature Biotech 2010-2017
- Nature Genetics 2009
- Genome Biology 2016

With the FDA: MAQC, SEQC, SEQC2

CHILDREN HOSPITAL BAMBINO GESU' - ROME:

- NAFLD: Non-Alcoholic Fatty Liver Disease
- Onco-immunology: Neuroblastoma
- Paediatric IBD/IBS
- Autism Spectrum Disorder genomics & metagenomics

Collaborazioni con Pharma

- Malattie Autoimmuni

FDA

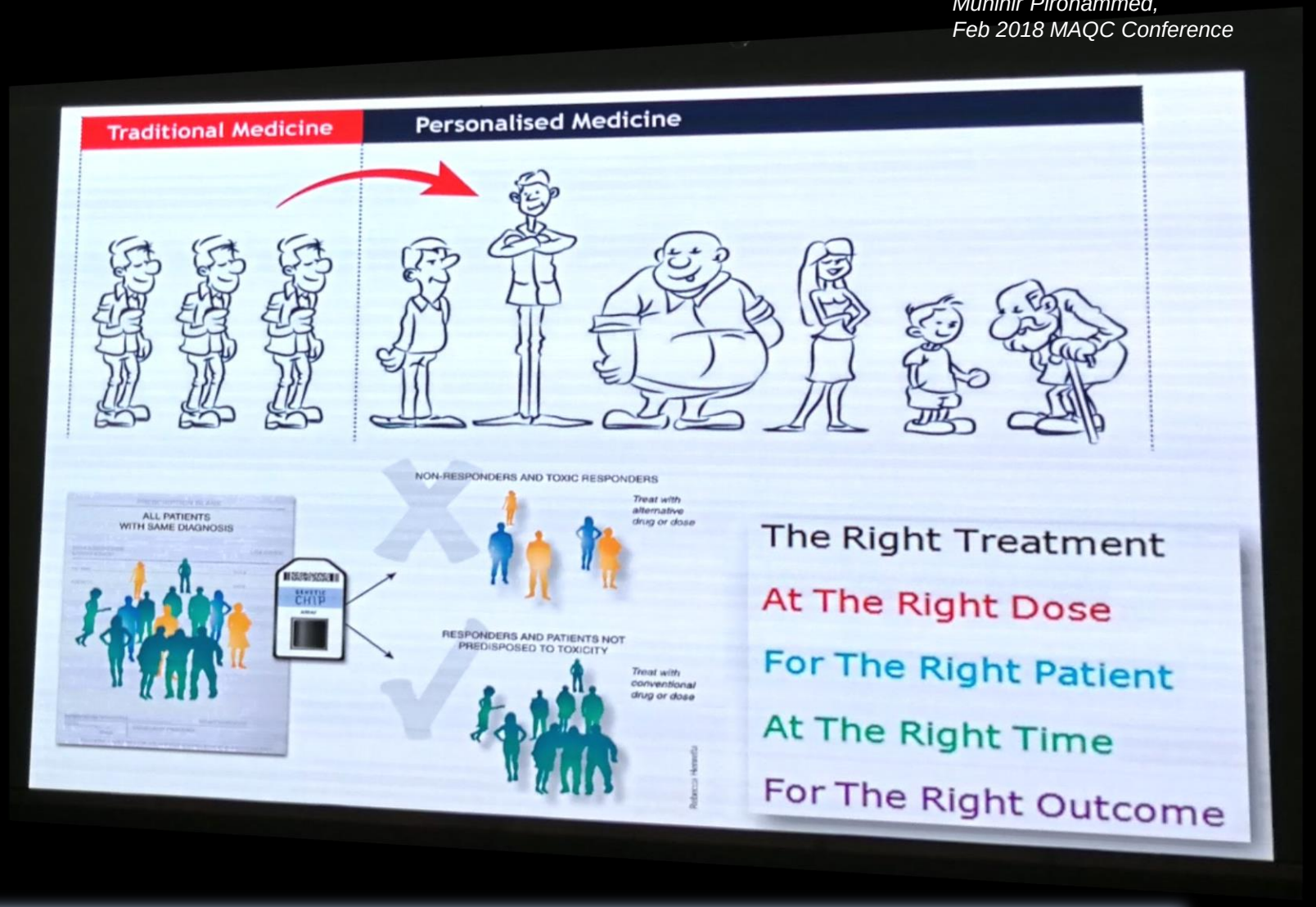
- Tossicogenomica (risposte ER alla esposizione)

- 2003-2005: Piattaforma Nazionale Bioinformatica AIRC - Italian Association for Cancer Research, con IFOM

LA GRANDE PROMESSA

- **Per diagnosi, trattamento e prevenzione, si deve adottare sistematicamente lo studio della variabilità individuale nei geni, ambiente, stile di vita**

Munihir Pirohammed, Feb 2018 MAQC Conference

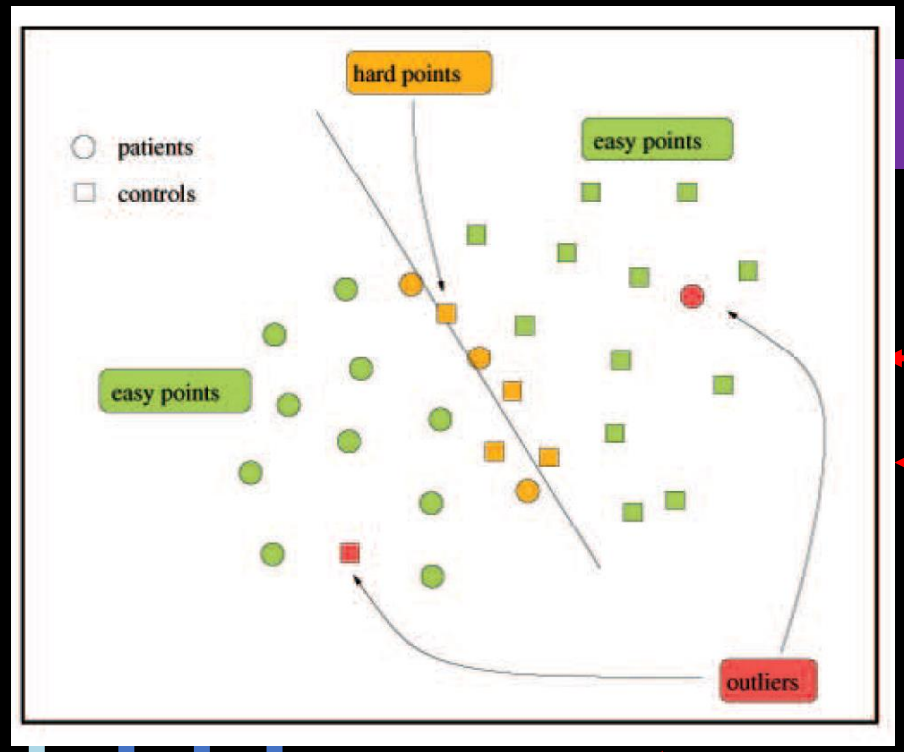


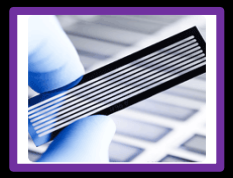
La Precision Medicine Initiative (USA 2015) è una azione visionaria per re-indirizzare la sanità e la R&D farmaceutica

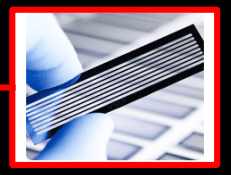
C. Furlanello – MPBA June 2018

MULTI-MODAL MACHINE LEARNING

Per la Decisione Clinica
in Biomedicina Pediatrica

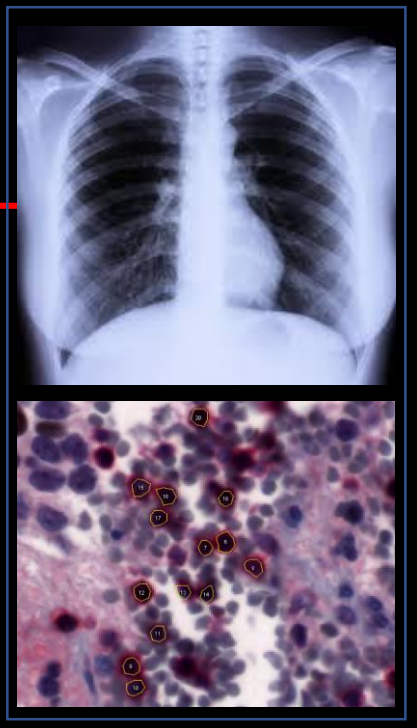






Next Generation Sequencing

Bioimaging



Radiomica

Patologia Digitale

TILs

Outcome
(fenotipo complesso)

INTEGRAZIONE E TRAIETTORIE

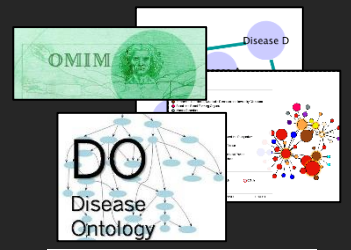
FENOTIPI COMPLESSI

Electronic Health Records (EHR)

OMICS DATA

Digital Pathology & Radiology

Phenotypes Lab, Clinical

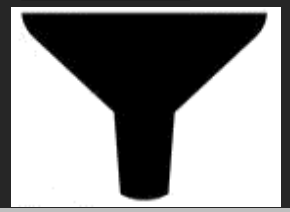


ML / DEEP LEARNING

Data Analytics Services

Endpoint Multipli
(Diagnostics, prognostic, actionable)

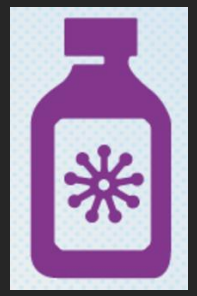
Modulati come
TRAIETTORIE DI SALUTE

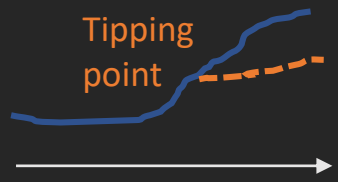


Integrated Disease Network

Bioinformatics

Precision Medicine





Aumentare significativamente il tempo di vita "libero da malattie"

C. Furlanello – MPBA June 2018

C. Furlanello – MPBA June 2018

RIPRODUCIBILITÁ

2017-2018 Sequencing Quality Control – Phase 2

Advancing Precision Medicine & Cancer Genomics

MAQC >10 years, >100 organizations (academy, industry, government) >300 participants; FBK since 2007

OBIETTIVI

(1) QUANTO SONO RIPRODUCIBILI I BIOMARKER
DA Next Generation Sequencing
whole genome sequencing (WGS) and ultra-deep targeted gene sequencing (TGS)

(2) QUALI I PARAMETRI CRITICI
per applicabilità farmacogenomica e clinica

(3) QUALI SOFTWARE E SISTEMI ESPERTI
benchmark di metodi bioinformatici e di machine learning per WGS and TGS → ricerca di protocolli standard di analisi verso attività regolatoria e medicina di precisione

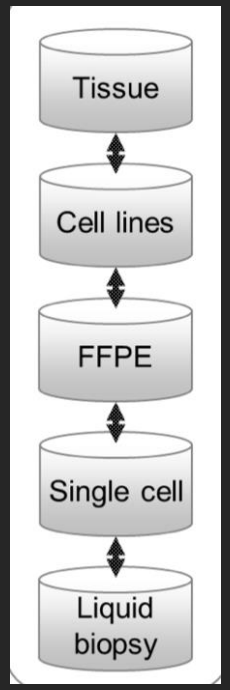
MAQC International Society

QC and analytics of massive data generated from high-throughput technologies

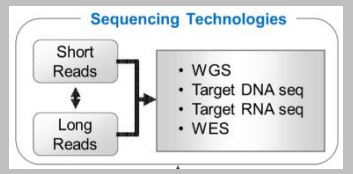
W. Tong et al 2017 Nat Bio

UPSTREAM - DOWNSTREAM

CAMPIONI



PIATTAFORME BIOTECNOLOGICHE (SEQUENCING)



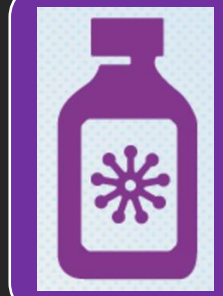
Bioinformatica

STAT/ ML
→ DEEP LEARNING

Basi di Conoscenza

Digital Pathology & Radiology

Phenotypes Lab, Clinical



Precision Medicine

FATTORI DI VARIABILITÁ (upstream) in Bioinformatics

DATA	Steps	Choices
Newly produced data (FDA) **Reference data sets** **Large cohort studies**	Preprocessing	21 tools
	Alignment to genome	68 tools 2 genome refs
	Alignment processing	2 tools (picard, GATK)
	Mutation calling	36 tools

21 x 68 x 2 x 2 x 36 =

205,632

combinations

High Performance Computing

RISCHI, LEZIONI E …

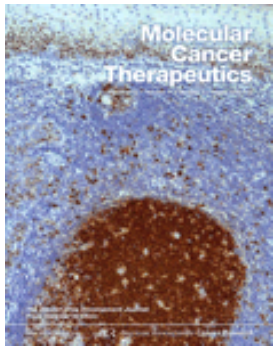
Horror stories
in Forensics Bioinformatics

Lessons Learnt
NatGen & MAQC

Forensic Bioinformatics

- **Baggerly 31/03/2011 – Inst. Of Medicine** *IOM Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials,*
- After Baggerly FGED13 17/07/2010 : "The Importance of Reproducibility in High-Throughput Biology: Case Studies in Forensic Bioinformatics" triggered by Baggerly and Coombes (2009) *Ann. App. Stat* 3(4)

- **The story: B&C could not replicate results from a series of papers by a Duke University team (on Nat Med, Lancet Oncology, JCO) for microarray signatures predicting drug response such as Cisplatin resistance. Mixing up of sample/gene labels was shown, as well as of validation data.**
- **The outcome: three Duke clinical trials suspended, 11 papers retracted within 2010-2012, legal initiatives by patients, US$ 750 000 grant revoked, PI resigned.**



FONDAZIONE BRUNO KESSLER

Oh, no!

- **Retraction #7. 03 Oct 2011, Garman et al. A genomic approach to colon cancer risk stratification yields biologic insights into therapeutic opportunities. PNAS 2008.**
- *The authors … "We wish to retract this article because we have been unable to reproduce certain key experiments described in the paper regarding validation and use of the colon cancer prognostic signature.* ***This includes the validation performed with dataset E-MEXP-1224,*** *… Because these results are fundamental to the conclusions of the paper, the authors formally retract the paper. We deeply regret the impact of this action on the work of other investigators."*

From: Cesare Furlanello
Sent: Wed, Nov 23, 2011 at 4:12 PM
To: <colleague in Barcelona>
Cc: more colleagues in Barcelona and Trento …
Subject: Retraction on CRC paper (Potti's 7th)

Dear <> This includes E-MEXP-1224 !!
// cesare

Retraction Watch
Tracking retractions as a window into the scientific process
New in PNAS: Potti retraction number seven, and a Potti correction

From: <colleague in Barcelona>
Sent: Wed, Nov 23, 2011 at 4:46 PM
To: Cesare Furlanello
Cc: more colleagues in Barcelona and Trento …
Subject: Re: Retraction on CRC paper (Potti's 7th)

Cesare

Thanks for pointing this out. We will revise the manuscript accordingly

FONDAZIONE BRUNO KESSLER

Letter to NCI Director, 2010

Re: Concerns about prediction models used in Duke clinical trials

Published and peer-reviewed re-analyses of the work done by … revealed serious errors that questioned the validity of the prediction models upon which these ongoing clinical trials are based.

We strongly urge that the clinical trials in question (NCT00509366, NCT00545948, NCT00636441) be suspended until a fully independent review is conducted of both the clinical trials and of the evidence and predictive models being used to make cancer treatment decisions. **For this to happen …**

A. **Sufficiently detailed data and annotation must be made available for review.**

B. **The data should be sufficiently documented for provenance** to be assessed (as both gene and sample mislabeling have been documented in these data)

C. **The computer code used to predict which drugs are suitable for particular patients must be made available** to allow an independent group of expert genomic data analysts to assess its validity and reproducibility using the data supplied.

FONDAZIONE BRUNO KESSLER

nature genetics

Replication of analyses

Repeatability of published microarray gene expression analyses

John P A Ioannidis[1–3], David B Allison[4], Catherine A Ball[5], Issa Coulibaly[4], Xiangqin Cui[4], Aedín C Culhane[6,7], Mario Falchi[8,9], Cesare Furlanello[10], Laurence Game[11], Giuseppe Jurman[10], Jon Mangion[11], Tapan Mehta[4], Michael Nitzberg[5], Grier P Page[4,12], Enrico Petretto[11,13] & Vera van Noort[14]

Given the complexity of microarray-based gene expression studies, guidelines encourage transparent design and public data availability. Several journals require public data deposition and several public databases exist. However, not all data are publicly available, and even when available, it is unknown whether the published results are reproducible by independent scientists. Here we evaluated the replication of data analyses in 18 articles on microarray-based gene expression profiling published in *Nature Genetics* in 2005–2006. One table or figure from each article was independently evaluated by two teams of analysts. We reproduced two analyses in principle and six partially or with some discrepancies; ten could not be reproduced. The main reason for failure to reproduce was data unavailability, and discrepancies were mostly due to incomplete data annotation or specification of data processing and analysis. Repeatability of published microarray studies is apparently limited. More strict publication rules enforcing public data availability and explicit description of data processing and analysis should be considered.

Microarray-based research is a prolific scientific field[1] where extensive data are generated and published. The field has been sensitized to the research, the Uniform Guidelines of the International Committee of Medical Journal Editors state that authors should "identify the methods, apparatus and procedures in sufficient detail to allow other workers to reproduce the results"[12]. Making primary data publicly available has many challenges but also many benefits[13]. Public data availability allows other investigators to confirm the results of the original authors, exactly replicate these results in other studies and try alternative analyses to see whether results are robust and to learn new things. Journals such as *Nature Genetics* require public data deposition as a prerequisite for publication for microarray-based research. Yet, the extent to which data are indeed made fully and accurately publicly available and permit confirmation of originally reported findings in many areas, including gene expression microarray research, is unknown.

In this project, we aimed to evaluate the repeatability of published microarrays studies. We focused specifically on the ability to repeat the published analyses and get the same results. This is one important component in the wider family of replication and reproducibility issues. We evaluated 18 articles published in *Nature Genetics* in 2005 or 2006 that presented data from comparative analyses of microarrays experiments that had not been previously published elsewhere. Detailed eligibility criteria and search strategies are presented in the Methods section. Of 20 initially selected articles[14–33], 2 were excluded[21,26] when

NATURE GENETICS | VOLUME 41 | NUMBER 2 | FEBRUARY 2009 149

FONDAZIONE BRUNO KESSLER

Results from the NG study

- Reproducibility of scientific results and the need for replication: on a leading journal, a multi-institution study scans papers about gene expression profiling:
- Inability to reproduce the analysis > 50%
- Partial reproduction in 1/3
- Perfect reproduction in 11%

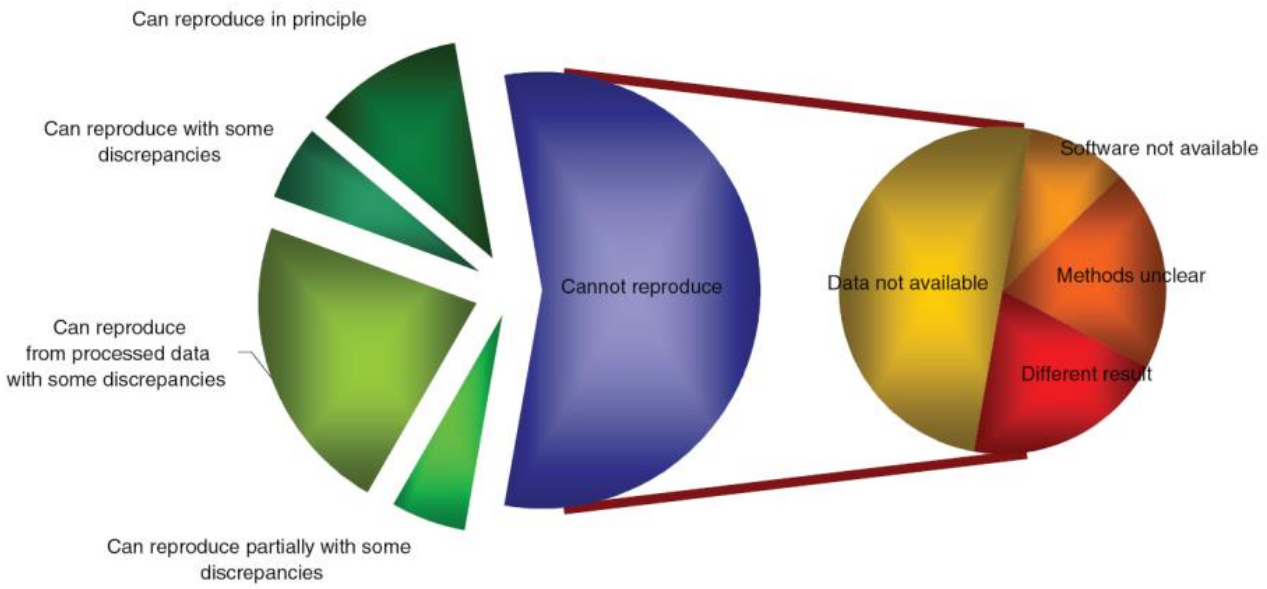


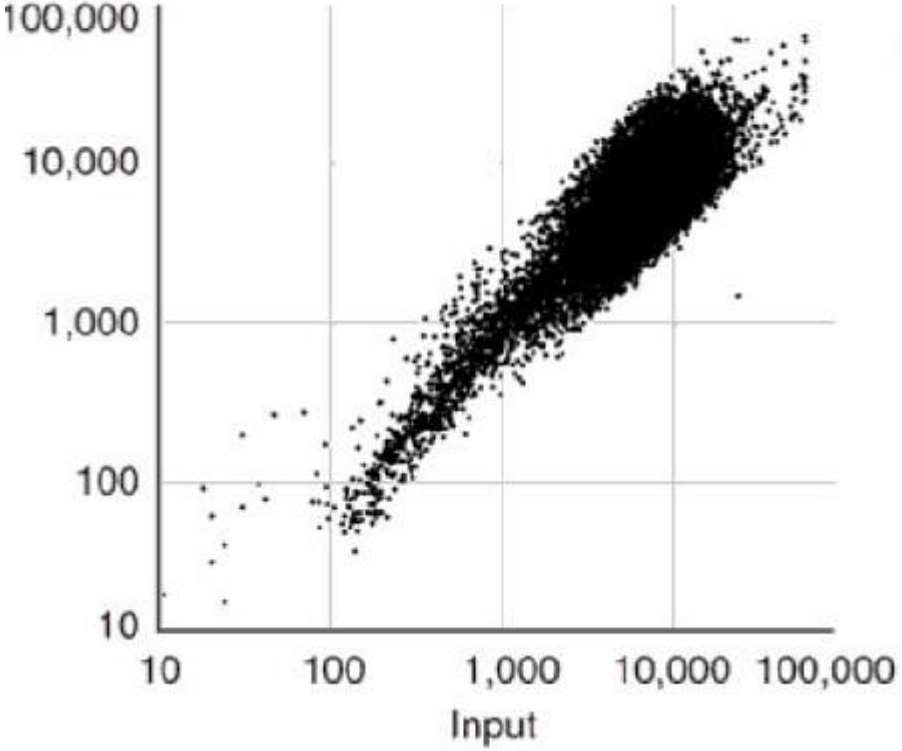
Figure 1 Summary of the efforts to replicate the published analyses.

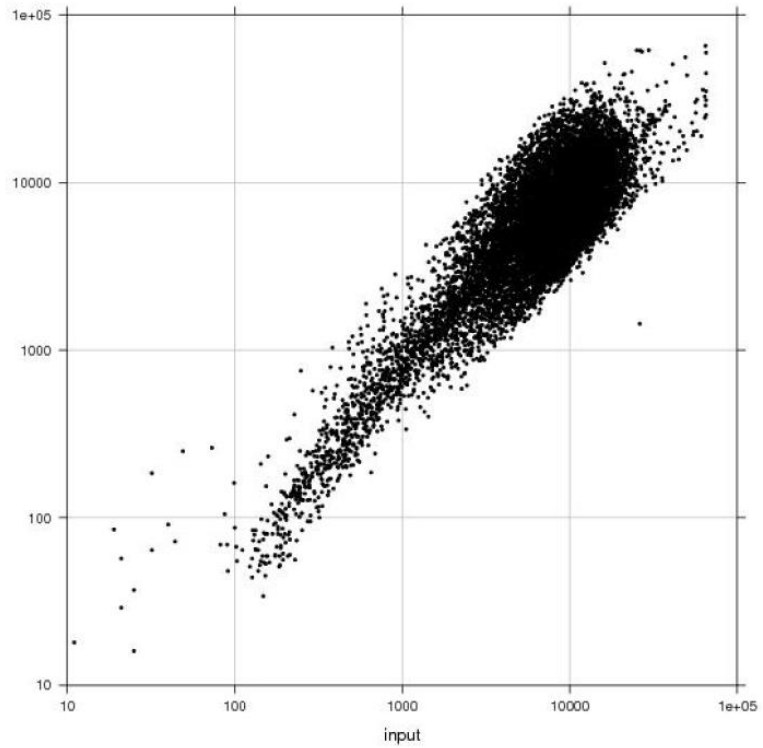
Editorial: "four teams of analysts treated the findings of a number of **microarray papers** published in the journal in 2005–2006 as their gold standard and attempted to replicate a sample of the analyses conducted on each of them, with frankly dismal results.", *Nature Genetics, Feb 2009*

FONDAZIONE BRUNO KESSLER

Repeatability: data

A. EXAMPLE: Nearly perfect replication





FONDAZIONE BRUNO KESSLER

Repeatability: (bad) markers

B. **EXAMPLE: for one article, in the attempt to use the authors' criteria, we found 120 eligible transcripts instead of the 162 published in the paper. Of those, we found 22 instead of the 33 published with an adjusted P value <0.01 and twofold enrichment.**

Seq ID	ID_REF	change in gene expression in original article	reproduced change in gene expression	confirmed change in gene expression	reproduced significance	confirmed significance
AF075436	1451086_s_at	1.7	1.1	no	0.073	no
AF075436	1452027_a_at	1.7	1.6	no	0.116	no
AF075436	1459581_at	1.7	1.7	yes	0.189	no
AF075436	1418158_at	1.7	1.9	no	0.001	yes
AF075436	1451876_a_at	1.7	2.1	no	0.139	no
AK003705	1453218_at	–310.8	-1012.3	no	0.001	yes
AK014360	1452166_a_at	NS	-1.0	-	0.552	yes
BC003828	1451970_at	NS	1.1	-	0.633	yes
BC003828	1423734_at	NS	1.4	-	0.009	no
BC011074	1423935_x_at	2.4	1.8	no	0.003	yes
BC011074	1460347_at	2.4	2.3	no	0.004	yes
NM_007700	1428210_s_at	NS	-1.2	-	0.390	yes
NM_007700	1451383_a_at	NS	-1.2	-	0.244	yes
NM_007700	1417091_at	NS	-1.1	-	0.124	yes
NM_008473	1422481_at	NS	-1.1	-	0.144	yes
NM_008508	1448745_s_at	–7.3	-8.2	no	0.002	yes

FONDAZIONE BRUNO KESSLER

Positive results

1. Better reproducibility if guidelines are followed:

Microarray analyses can potentially be reproduced if the data are available, adequately annotated and the analytic steps and parameters are sufficiently described (MIAME guidelines and journal policies address public data availability)

2. Reproducibility "pays" in Reputation !!

Analysis: the number of citations catalogued by ISI as of the end of August 2008 for articles where some reproduction of at least part of the results was feasible (in principle or with some discrepancies) vs those where we could not reproduce the selected analyses.

Results: more reproducible articles had received more citations,

(median **29.8 per year** (range 7.5–86.6) versus **12.4 per year** (range 5.7–29.4), P = 0.038)

after adjustment for the time of publication

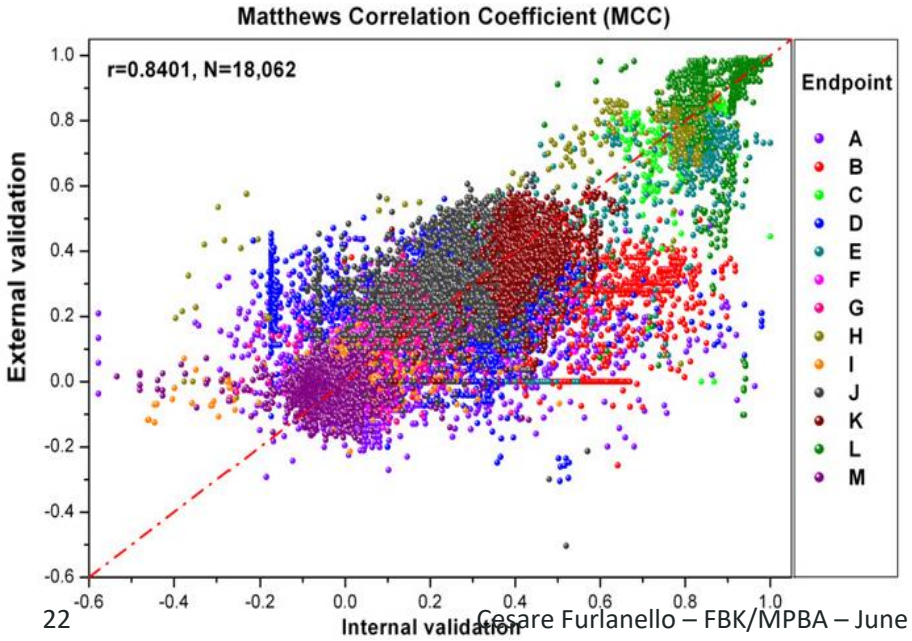
Lessons Learnt: MAQC2

The MicroArray Quality Control (MAQC) Consortium. *The MAQC-II Project: A comprehensive study of common practices for the development and validation of microarray-based predictive models.* Nature Biotechnology, 28(8):827-838, 2010

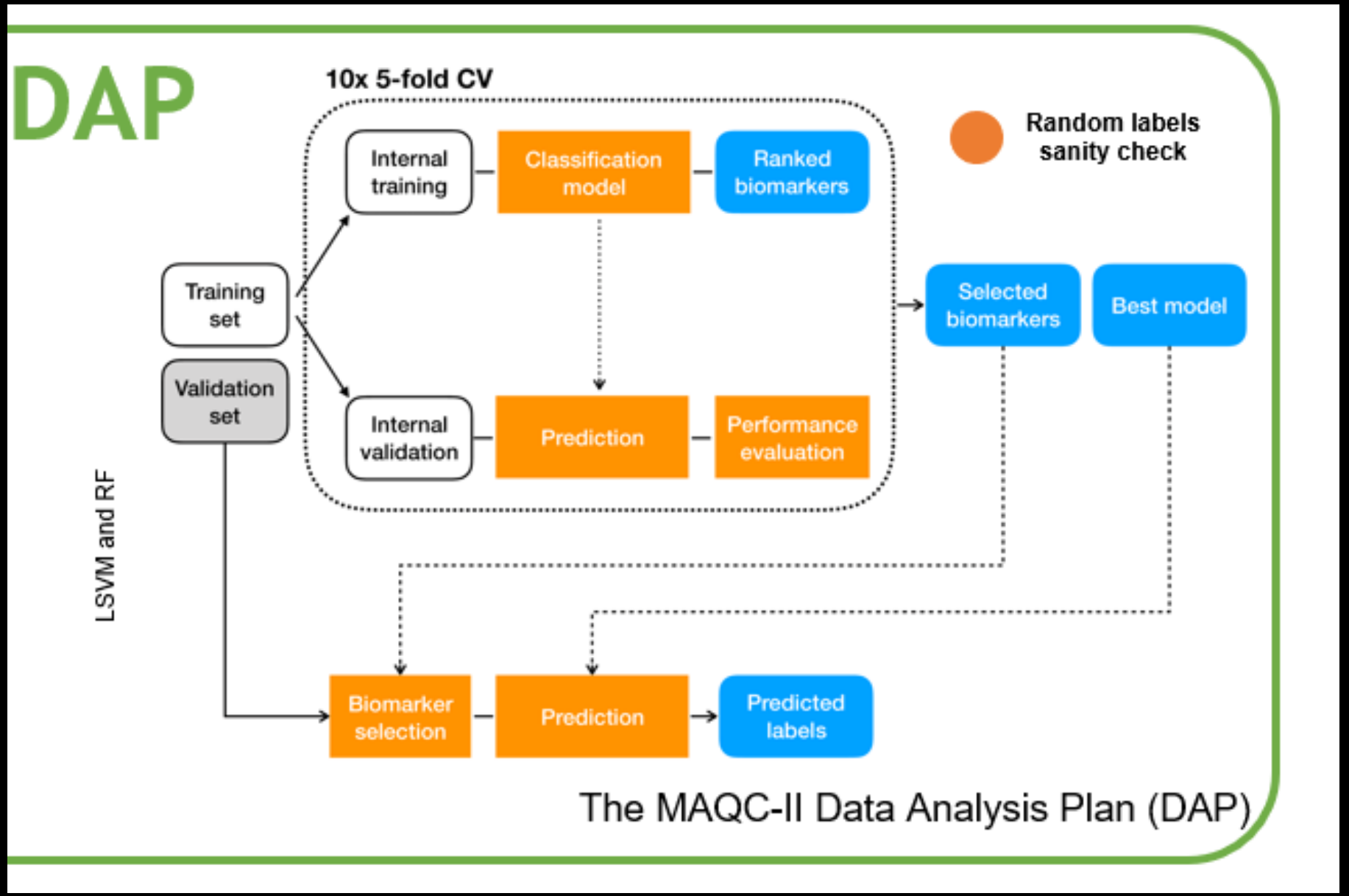
1. **Predictive models can be derived from microarray data,**
2. **But they need to be carefully developed and independently tested**

3. Reproducibility requires substantial effort.

- Appreciable differences in prediction performance by models of diverse teams on same endpoint



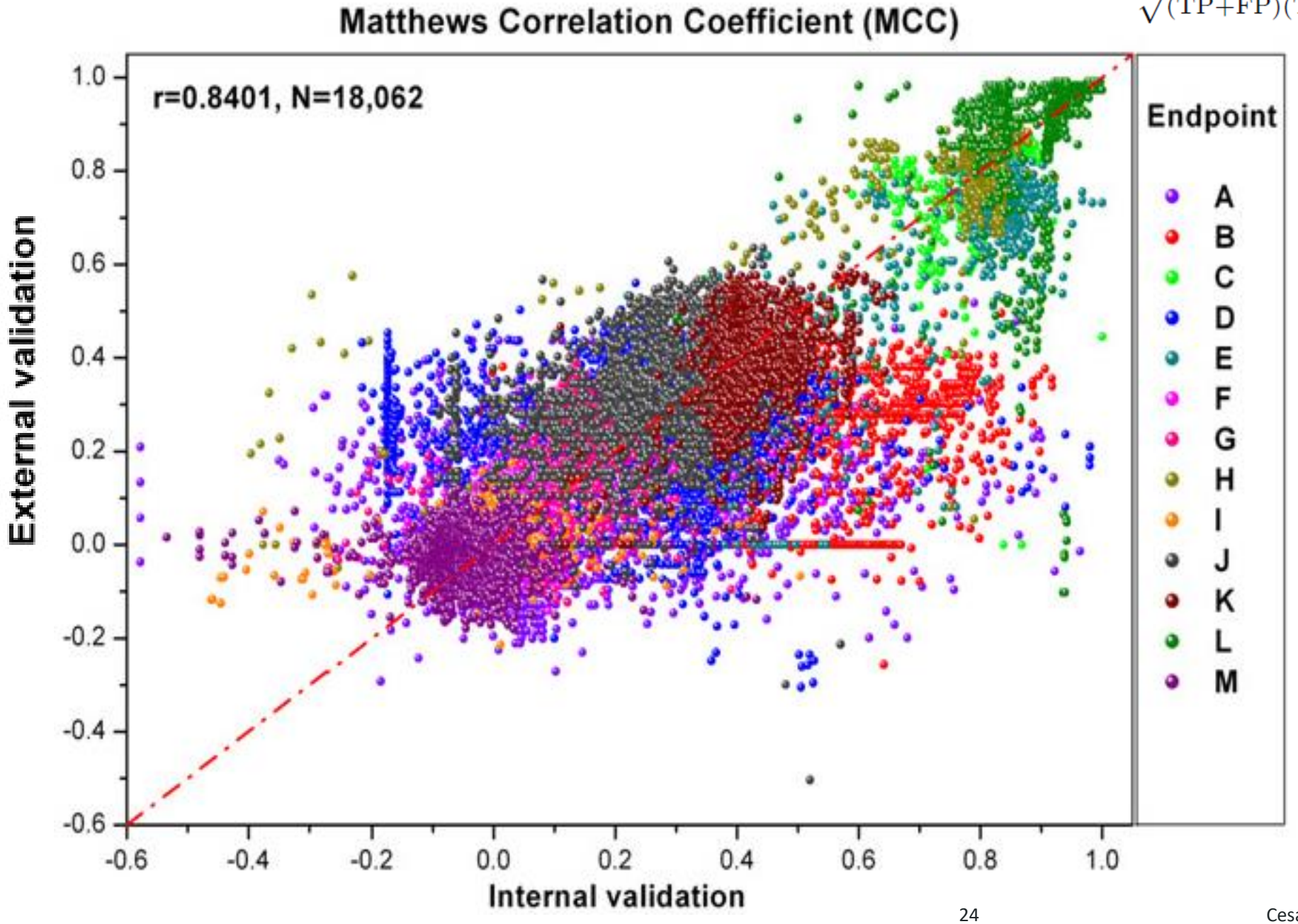
- **13 candidate models chosen based on their Data Analysis Plans had higher correlation** between internal and external performance than all models
- **Dominant factors:** mainly difficulty of the endpoint. Smaller contributions: normalization, different tuning of algorithms, feature selection methods, **overall: modeling approaches can distort the internal performance estimate**.



C. Furlanello – MPBA June 2018

C. Furlanello – MPBA June 2018

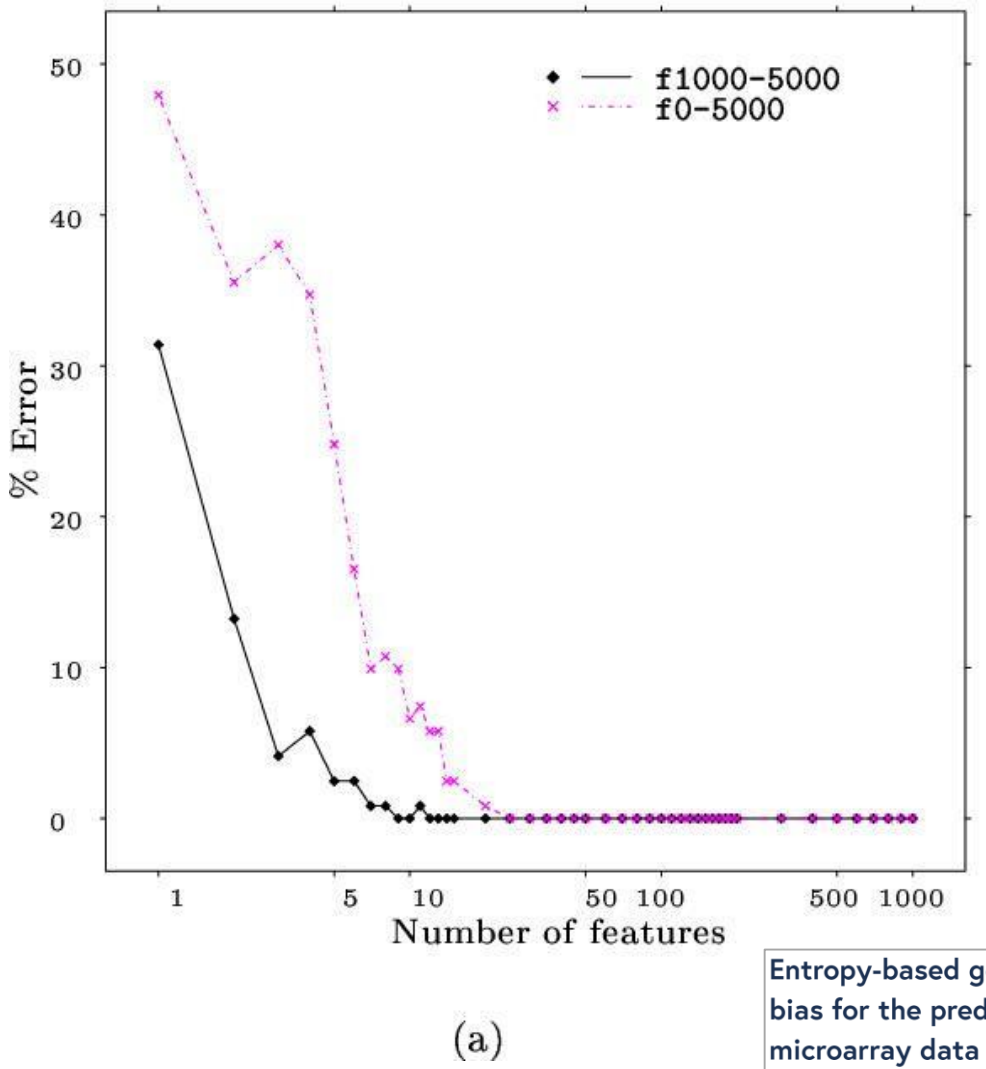
$$\mathrm{MCC} = \frac{\mathrm{TP}\cdot\mathrm{TN} - \mathrm{FP}\cdot\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$$

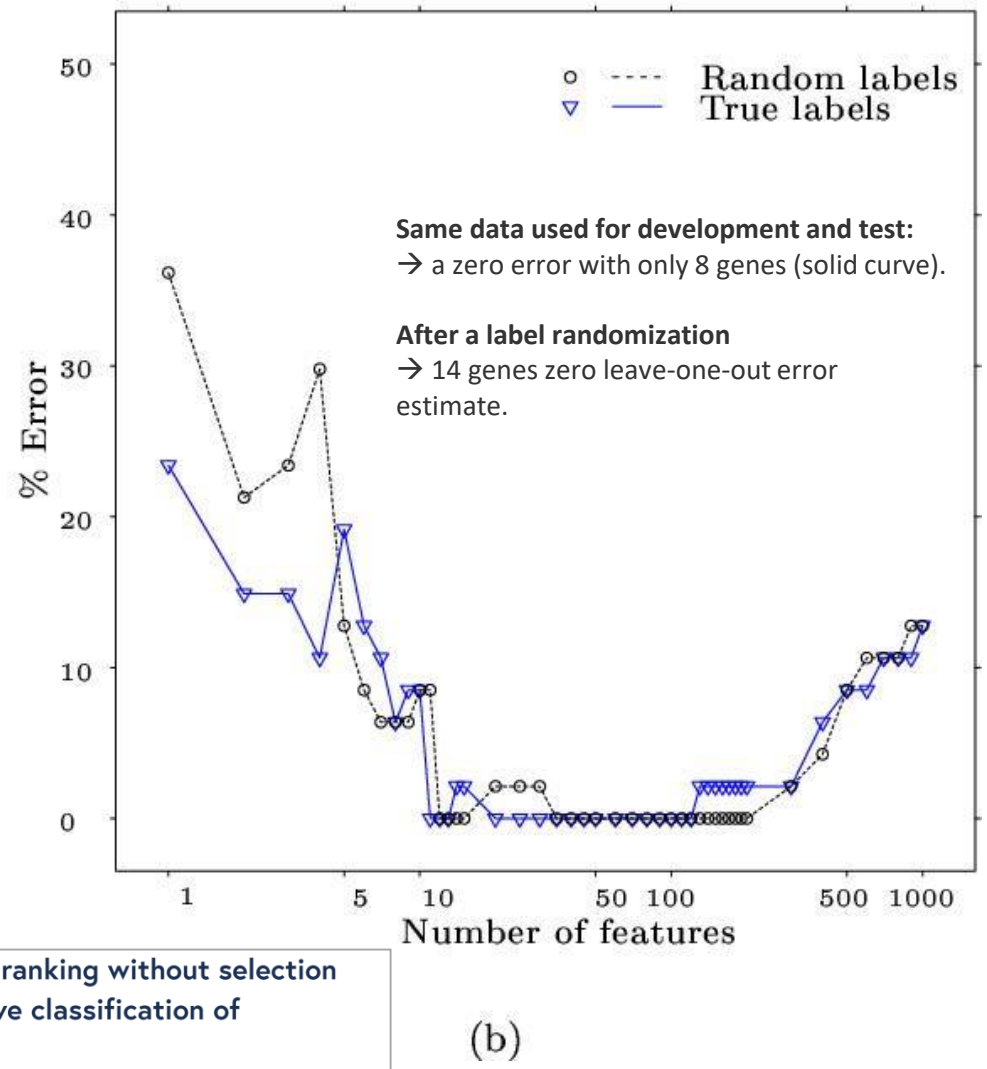


f1000–5000: 100 samples by 5000 features, of which 1000 are significant (i.e. generated by 1000 Gaussian distribution centered in 1 and -1, with standard deviation uniformly ranging between 1 and 5), and the remaining are uniform noise in the range
f0-5000 all uniform noise features

Rank with SVM RFE, then retrain SVM in 10-fold cv for increasing feature set sizes

Colon cancer microarray data set: expression of 2000 genes from 62 tissues (22 normal and 40 tumor cases, Affimetrix oligonucleotide arrays).
RFE error curves estimated by leave-one-out cross-validation for models trained on feature subsets of increasing size, after a feature ranking performed on all the available data.





Same data used for development and test:
→ a zero error with only 8 genes (solid curve).

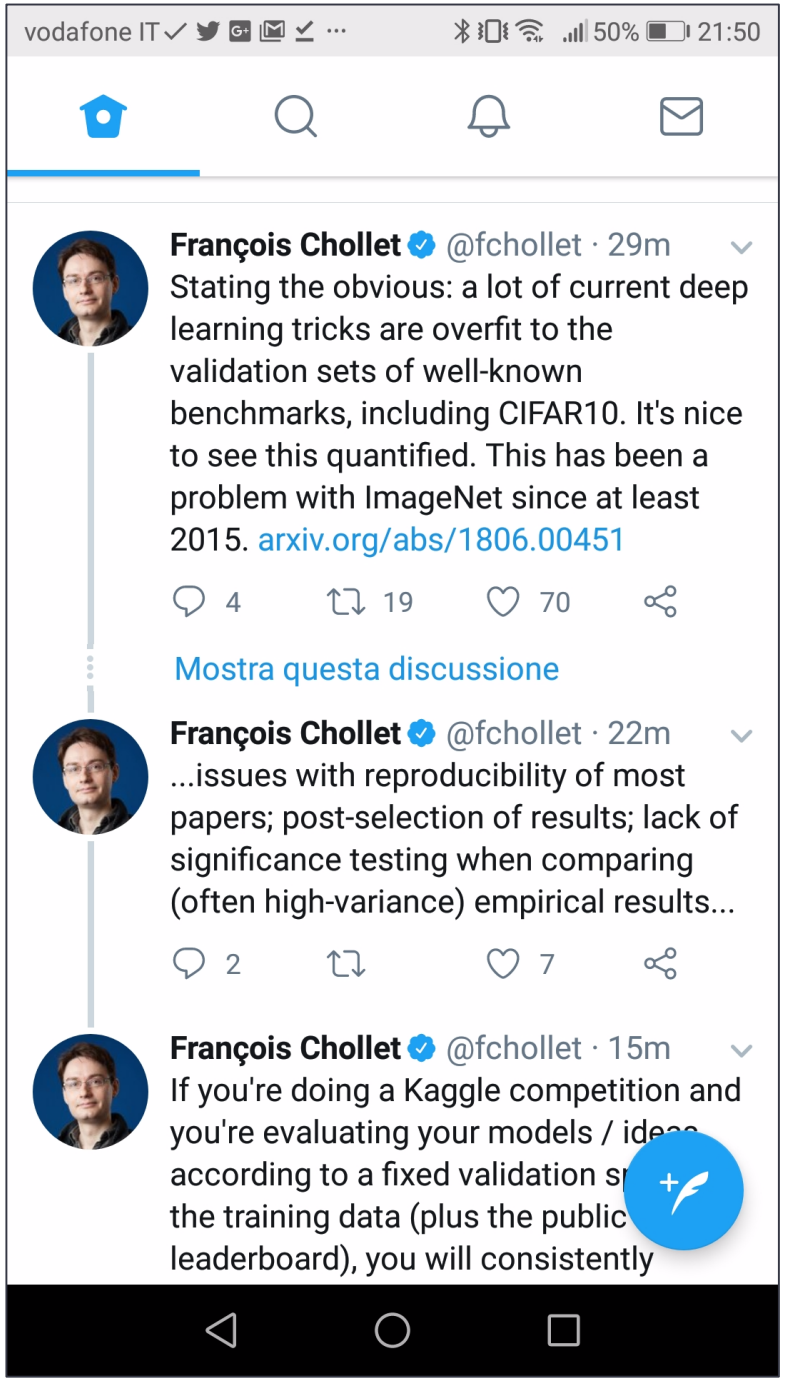
After a label randomization
→ 14 genes zero leave-one-out error estimate.

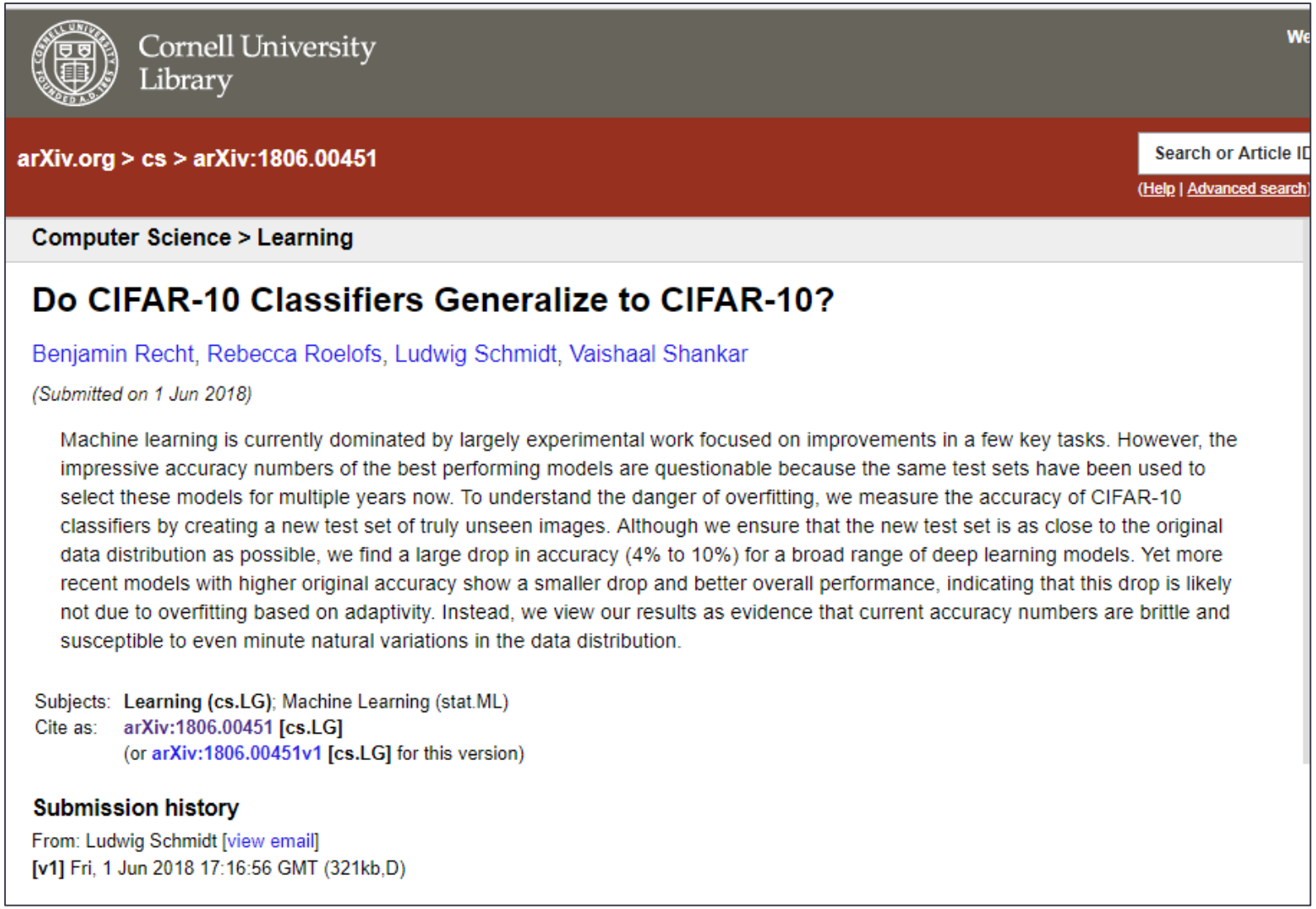
2013

Entropy-based gene ranking without selection bias for the predictive classification of microarray data

Cesare Furlanello, Maria Serafini, Stefano Merler and Giuseppe Jurman

BMC Bioinformatics 2003 **4**:54
https://doi.org/10.1186/1471-2105-4-54
Received: 16 April 2003 | Accepted: 06 November 2003 | Published: 06 November 2003


vodafone IT
50% 21:50
François Chollet @fchollet · 29m
Stating the obvious: a lot of current deep learning tricks are overfit to the validation sets of well-known benchmarks, including CIFAR10. It's nice to see this quantified. This has been a problem with ImageNet since at least 2015. arxiv.org/abs/1806.00451
4 19 70
Mostra questa discussione
François Chollet @fchollet · 22m
...issues with reproducibility of most papers; post-selection of results; lack of significance testing when comparing (often high-variance) empirical results...
2 7
François Chollet @fchollet · 15m
If you're doing a Kaggle competition and you're evaluating your models / ide
according to a fixed validation s
the training data (plus the public
leaderboard), you will consistently


Cornell University Library
arXiv.org > cs > arXiv:1806.00451
Search or Article ID
(Help | Advanced search)
Computer Science > Learning
Do CIFAR-10 Classifiers Generalize to CIFAR-10?
Benjamin Recht, Rebecca Roelofs, Ludwig Schmidt, Vaishaal Shankar
(Submitted on 1 Jun 2018)
Machine learning is currently dominated by largely experimental work focused on improvements in a few key tasks. However, the impressive accuracy numbers of the best performing models are questionable because the same test sets have been used to select these models for multiple years now. To understand the danger of overfitting, we measure the accuracy of CIFAR-10 classifiers by creating a new test set of truly unseen images. Although we ensure that the new test set is as close to the original data distribution as possible, we find a large drop in accuracy (4% to 10%) for a broad range of deep learning models. Yet more recent models with higher original accuracy show a smaller drop and better overall performance, indicating that this drop is likely not due to overfitting based on adaptivity. Instead, we view our results as evidence that current accuracy numbers are brittle and susceptible to even minute natural variations in the data distribution.
Subjects: Learning (cs.LG); Machine Learning (stat.ML)
Cite as: arXiv:1806.00451 [cs.LG]
(or arXiv:1806.00451v1 [cs.LG] for this version)
Submission history
From: Ludwig Schmidt [view email]
[v1] Fri, 1 Jun 2018 17:16:56 GMT (321kb,D)

New Results

Previous

Evaluating reproducibility of AI algorithms in digital pathology with DAPPER

Posted June 6, 2018.

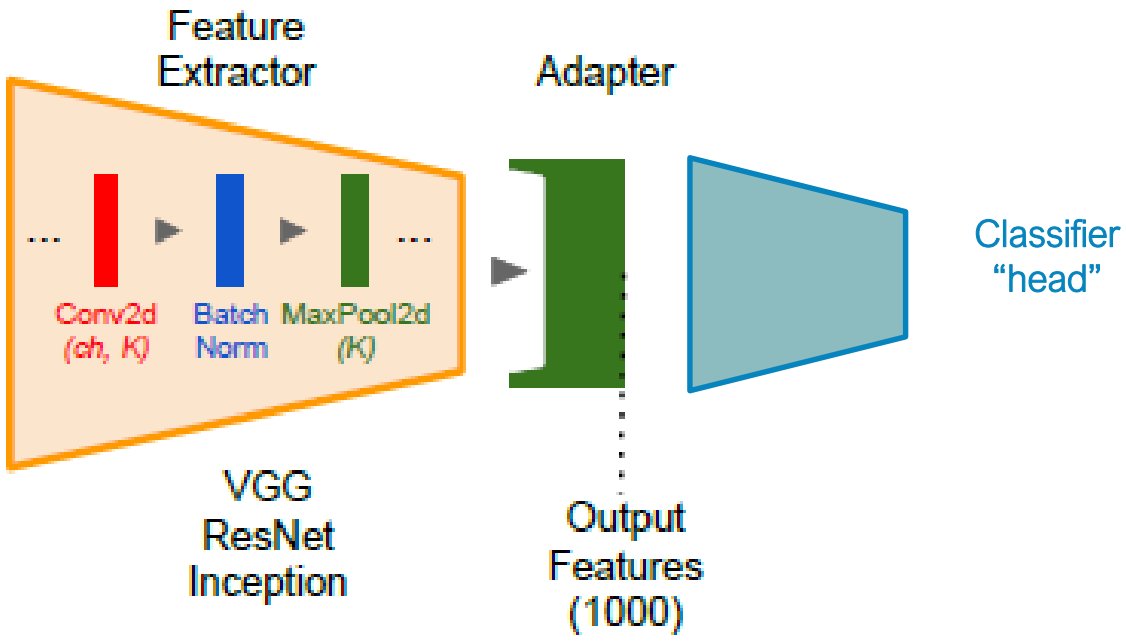
Andrea Bizzego, Nicole Bussola, Marco Chierici, Marco Cristoforetti, Margherita Francescatto, Valerio Maggio, Giuseppe Jurman, Cesare Furlanello

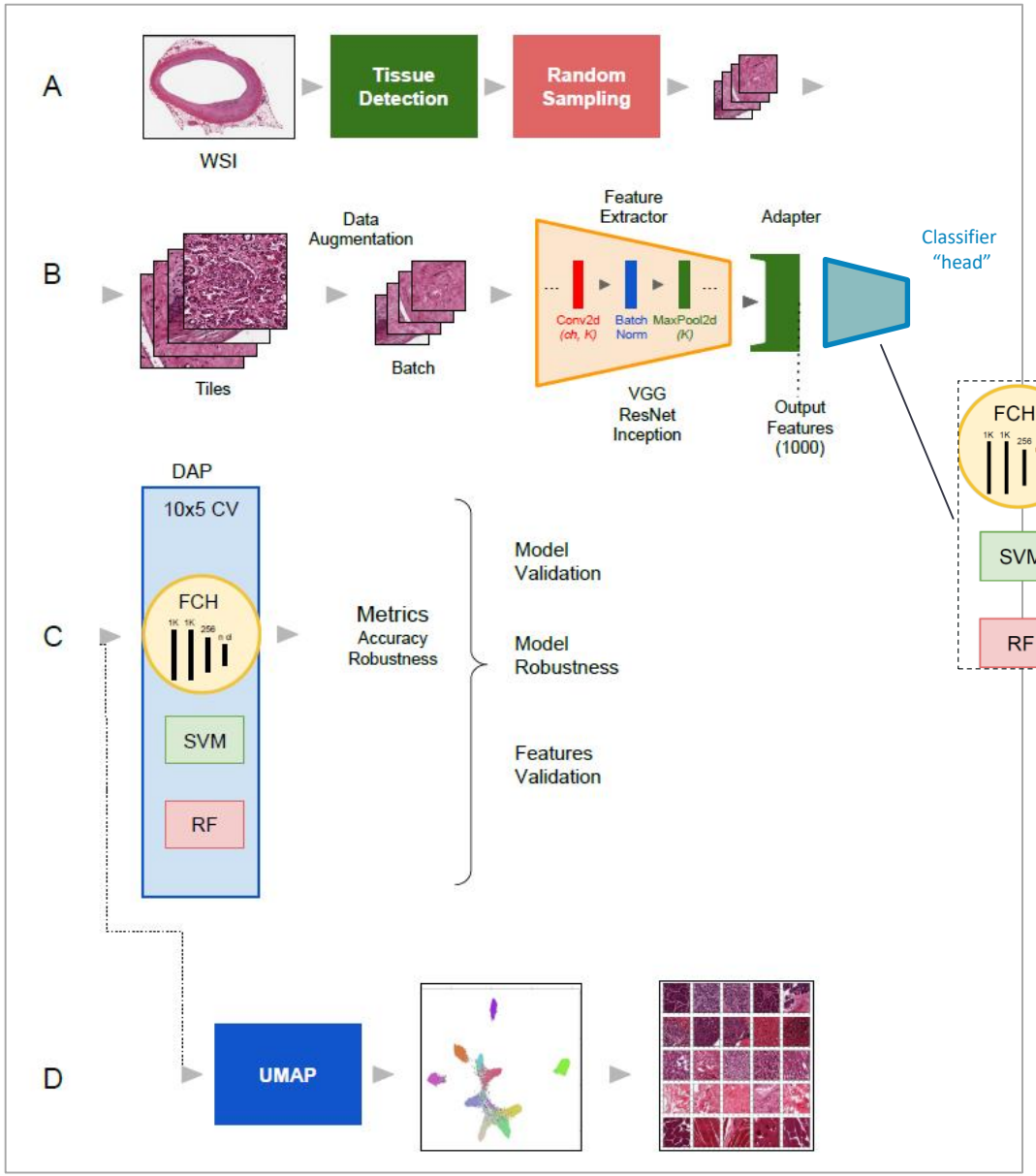
doi: https://doi.org/10.1101/340646

1. **The reliable estimation on a given training dataset of predictive accuracy and stability of deep learning models (or of deep features used by external models) is still a gray area.**

2. **The underlying risk is that of overfitting the training data, or worse to overfit the validation data if the labels are visible, which is typical when datasets are fully released at the end of a MI data science challenge.**

Fig 1. The DAPPER environment. Components: A) The WSI preprocessing pipeline; B) the deep learning backbone, to extract deep features; C) the Data Analysis Plan (DAP) for the machine learning models; and D) the UMAP module and other modules for unsupervised analysis.





OPPORTUNITÀ

IDEE:

- «DISTILLARE» ALGORITMI DIAGNOSTICI ESISTENTI PER MIGLIORARE MODELLI DI PROGNOSI
- EMBEDDING IN MULTITASK

Quantitative Biology > Quantitative Methods

A multiobjective deep learning approach for predictive classification in Neuroblastoma

Valerio Maggio, Marco Chierici, Giuseppe Jurman, Cesare Furlanello

(Submitted on 22 Nov 2017 (v1), last revised 22 Feb 2018 (this version, v3))

Neuroblastoma is a strongly heterogeneous cancer with very diverse clinical courses that may vary from spontaneous regression to fatal progression; an accurate patient's risk estimation at diagnosis is essential to design appropriate tumor treatment strategies. Neuroblastoma is a paradigm disease where different diagnostic and prognostic endpoints should be predicted from common molecular and clinical information, with increasing complexity, as shown in the FDA MAQC-II study. Here we introduce the novel multiobjective deep learning architecture CDRP (Concatenated Diagnostic Relapse Prognostic) composed by 8 layers to obtain a combined diagnostic and prognostic prediction from high-throughput transcriptomics data. Two distinct loss functions are optimized for the Event Free Survival (EFS) and Overall Survival (OS) prognosis, respectively. We use the High-Risk (HR) diagnostic information as an additional input generated by an autoencoder embedding. The latter is used as network regulariser, based on a clinical algorithm commonly adopted for stratifying patients from cancer stage, age at insurgence of disease, and MYCN, the specific molecular marker. The architecture was applied to Illumina HiSeq2000 RNA-Seq for 498 neuroblastoma patients (176 at high risk) from the Sequencing Quality Control (SEQC) study, obtaining state-of-art on the diagnostic endpoint and improving prediction of prognosis over the HR cohort.

Comments: NIPS ML4H workshop 2017 & MAQC 2018
Subjects: **Quantitative Methods (q-bio.QM)**; Learning (cs.LG)
Cite as: **arXiv:1711.08198 [q-bio.QM]**
(or **arXiv:1711.08198v3 [q-bio.QM]** for this version)

Submission history

From: Giuseppe Jurman [view email]
[v1] Wed, 22 Nov 2017 09:54:48 GMT (56kb,D)
[v2] Fri, 1 Dec 2017 13:38:22 GMT (61kb,D)
[v3] Thu, 22 Feb 2018 18:43:29 GMT (185kb,D)

Improved prognostic profiling in high-risk neuroblastoma by multi-task deep learning with distillation of the clinical diagnostic algorithm

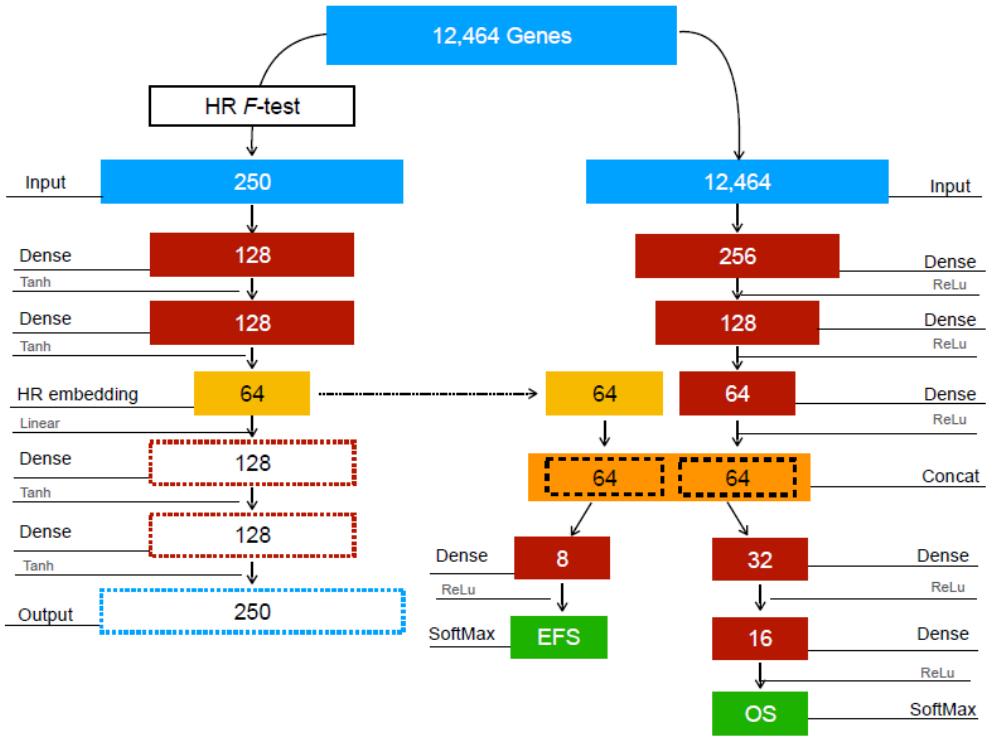


Fig 1. Deep learning architecture. The layer/node structure of the CDRP deep learning architecture. On the left side: the CDPR-A autoencoder; on the right side: the CDRP-N component, with two branches. Blocks indicate net layers, with the input dimensions for the SEQC-NB dataset.

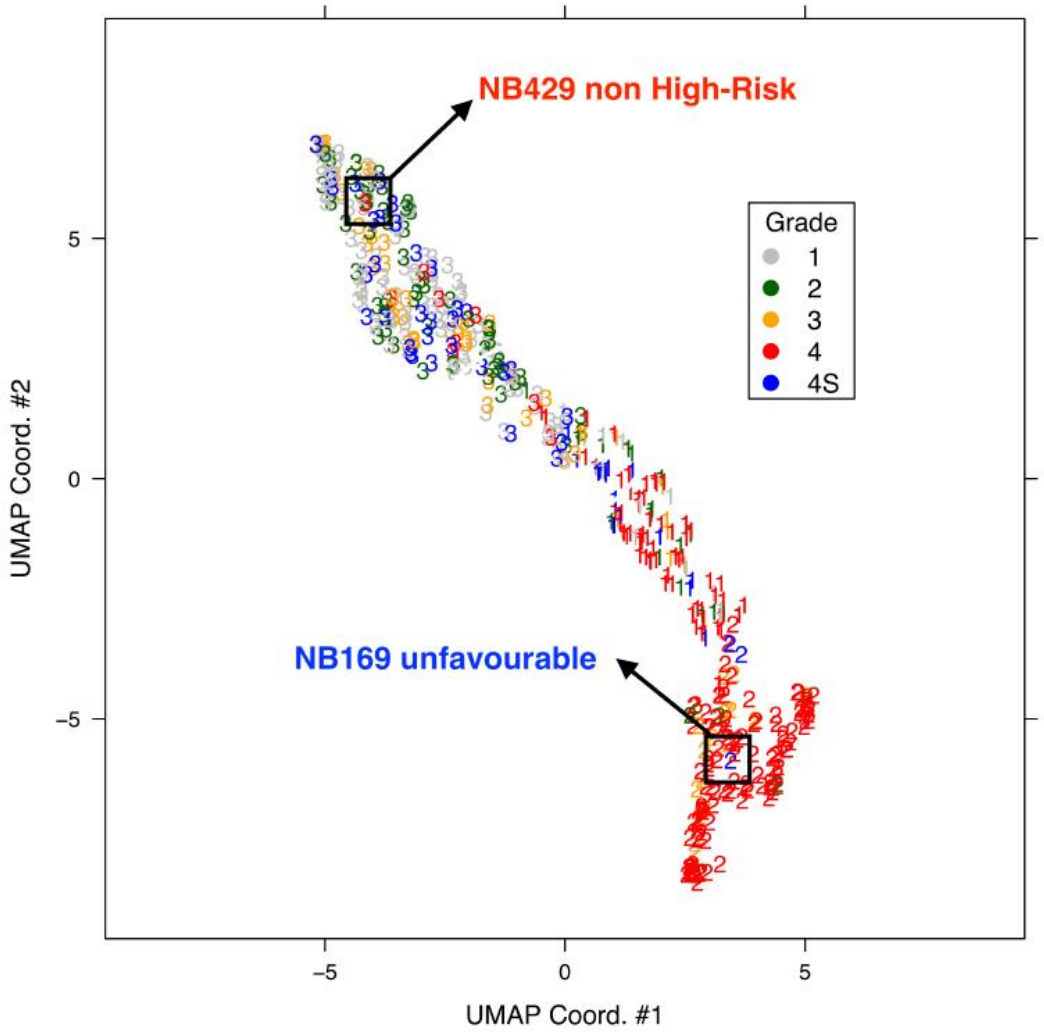
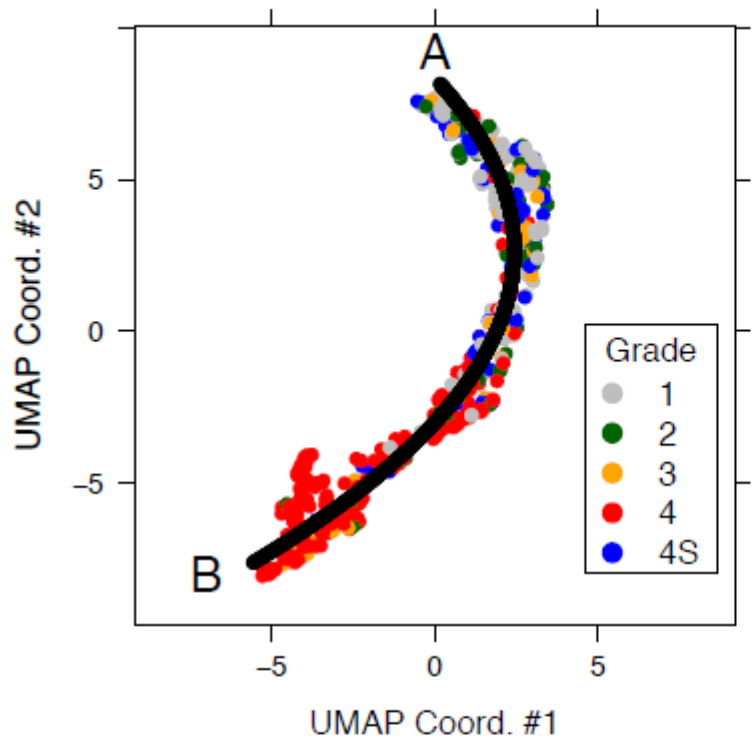


Fig 7. UMAP projection of the 1000 deep features of SEQC-NB samples on the hidden Overall Survival layer with 32 nodes. Colors indicate tumor grade, while numbers correspond to the hierarchical clusters of Fig. 6, 1:gray, 2: yellow, 3:blue. Two outlier samples are highlighted.

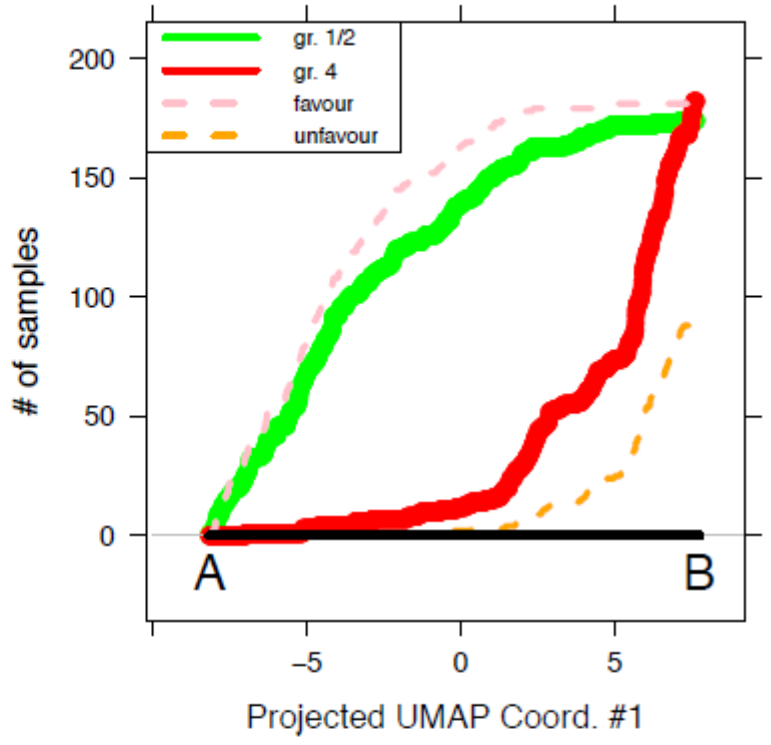
V. Maggio et al, submitted 2018

C. Furlanello – MPBA June 2018

Improved prognostic profiling in high-risk neuroblastoma by multi-task deep learning with distillation of the clinical diagnostic algorithm



(a)



(b)

Fig 8. Manifold approximation of UMAP projection. (a) Colors indicate tumor grade and the black line is the approximating parabola; (b) Cumulative sum of severe (red line) and less severe (green) cases while traversing the linearly projected manifold from point A to point B. Samples with low grading and favorable prognosis concentrate close to point A, while patients with more severe condition or unfavorable prognosis are grouping towards point B.

V. Maggio et al, submitted 2018

C. Furlanello – MPBA June 2018

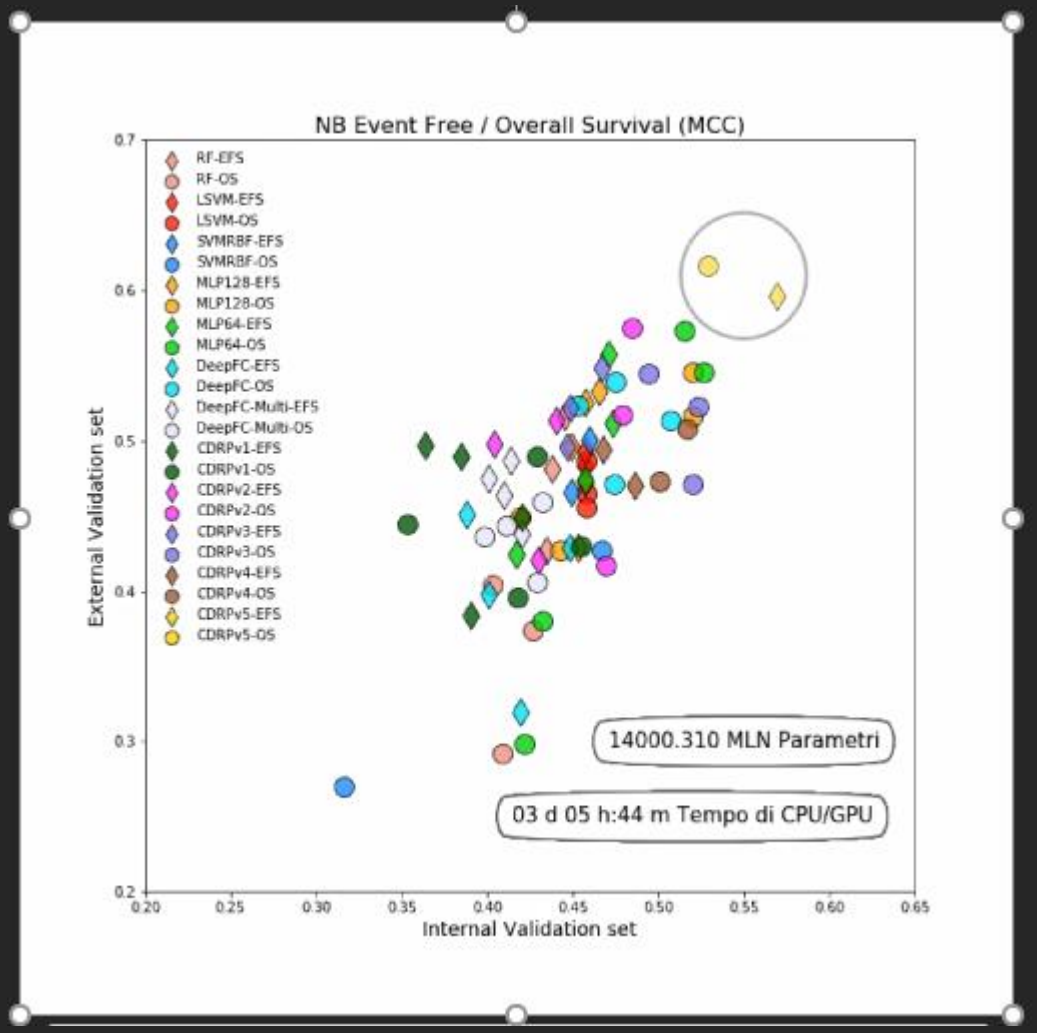
UNDER THE HOOD

"It's ML, not magic" Credits: @smerity

- The Data Science stack (Python, R) and resources (arXiv.org)
- Keras with TensorFlow backend
- PyTorch
- Fast, well tuned baselines
- Model Selection: human intuition in Deep Learning is bad
- Ability to accurately measure progress over time

OK, but ... do I need a GPU Armada?

MODEL SELECTION

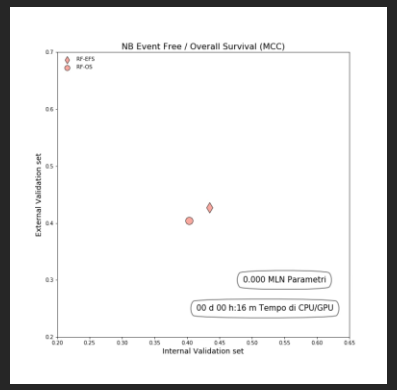


- 200 runs for each model
- 120 M parameters for inputs only
- Science or … ?

Valerio Maggio

GPU: NVIDIA Titan K80 Maxwell K80 - 12GB RAM

SEQC_NB - Neuroblastoma 498 patients
RNA-Seq, 248 training, 248 validation;



DEEP LEARNING IN AZURE CLOUD

7 PostDoc / Ricercatori

10 PhD - Master

WebValley (Data Science)

Microsoft Azure

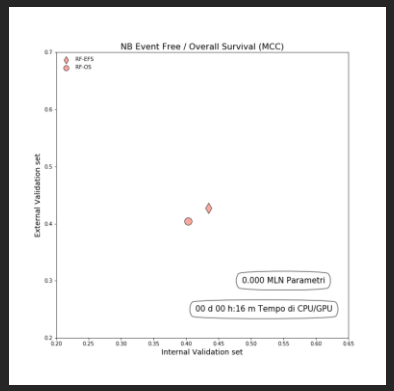
FONDI FBK
→ DEEP LEARNING

AZURE RESEARCH GRANT
→ DEEP LEARNING

GRANT MPBA
→ DEEP LEARNING

WebValley

DEEP LEARNING IN AZURE CLOUD



DATA SCIENTIST

PROVISIONING SINGOLA RISORSA

PROVISIONING con DAP su multiple risorse

Valerio Maggio

Microsoft Azure

DEEP LEARNING IN AZURE CLOUD

ToxicoGenomics

- 64 vCPUS
- 2x 2TB SSD
- 432 GB RAM

Running

Medical Images

- 24 vCPUS
- 4x NVIDIA TITAN K80 (12 GB each(
- 224 GB RAM

Running

Now-Casting

- 12 vCPUS
- 2x NVIDIA TITAN K80 (12 GB each(
- 112 GB RAM

Running

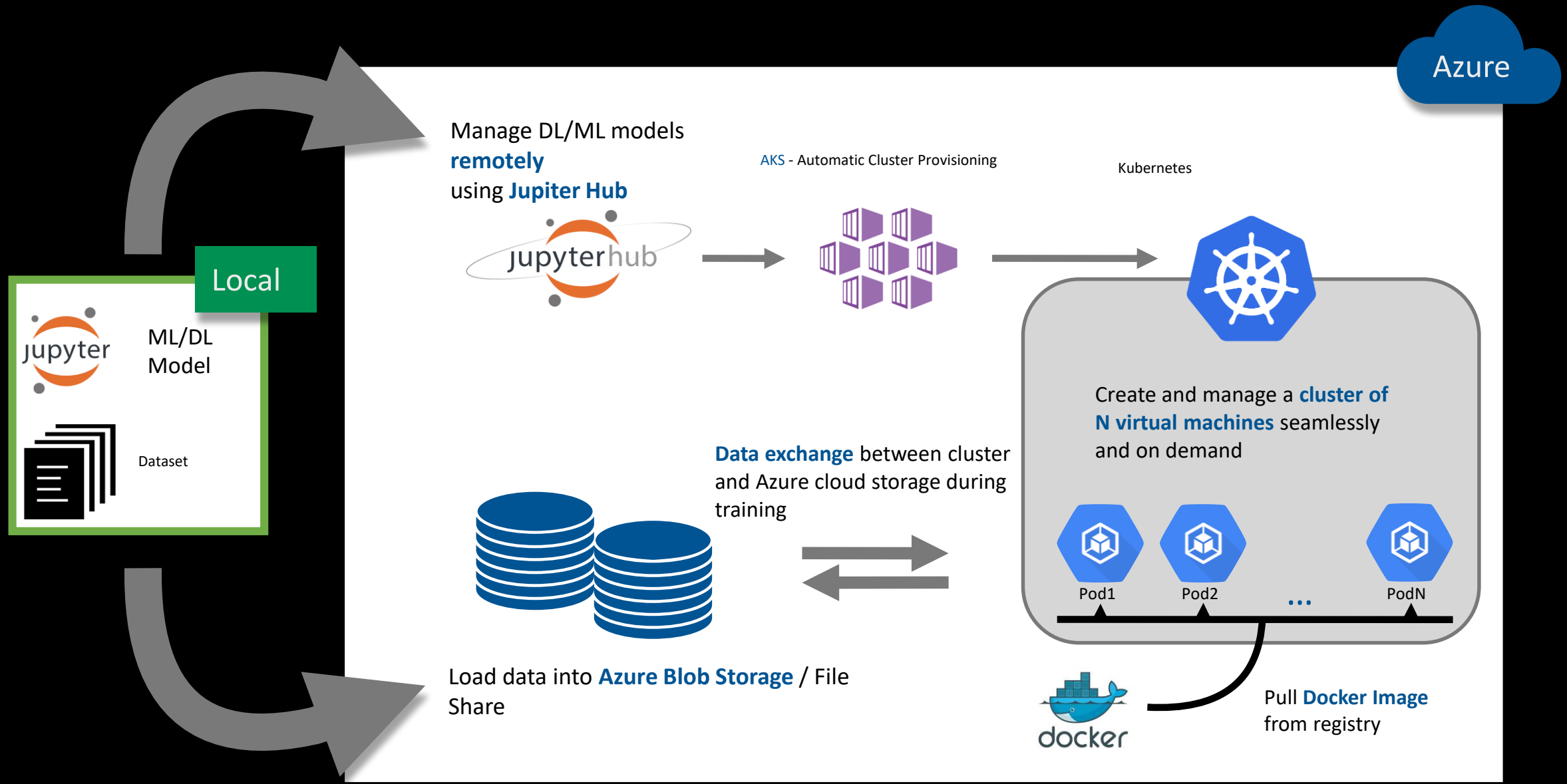
Pediatric Oncology

- 24 vCPUS
- 4x NVIDIA TITAN K80 (12 GB each(
- 224 GB RAM

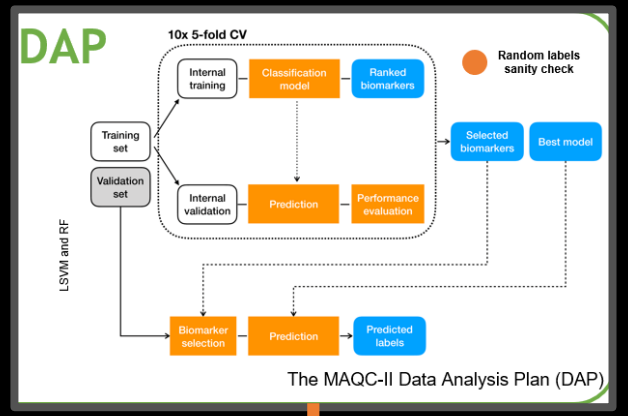
Running

Valerio Maggio

DISTRIBUTED COMPUTATION: WORKFLOW ON AZURE



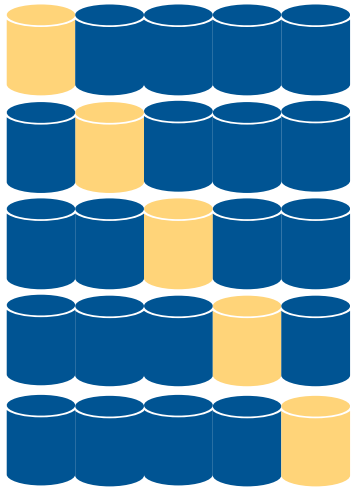
DAP-DATA Manager



Local

Dataset locale. Suddivisione iterativa dei dati in fold

500x

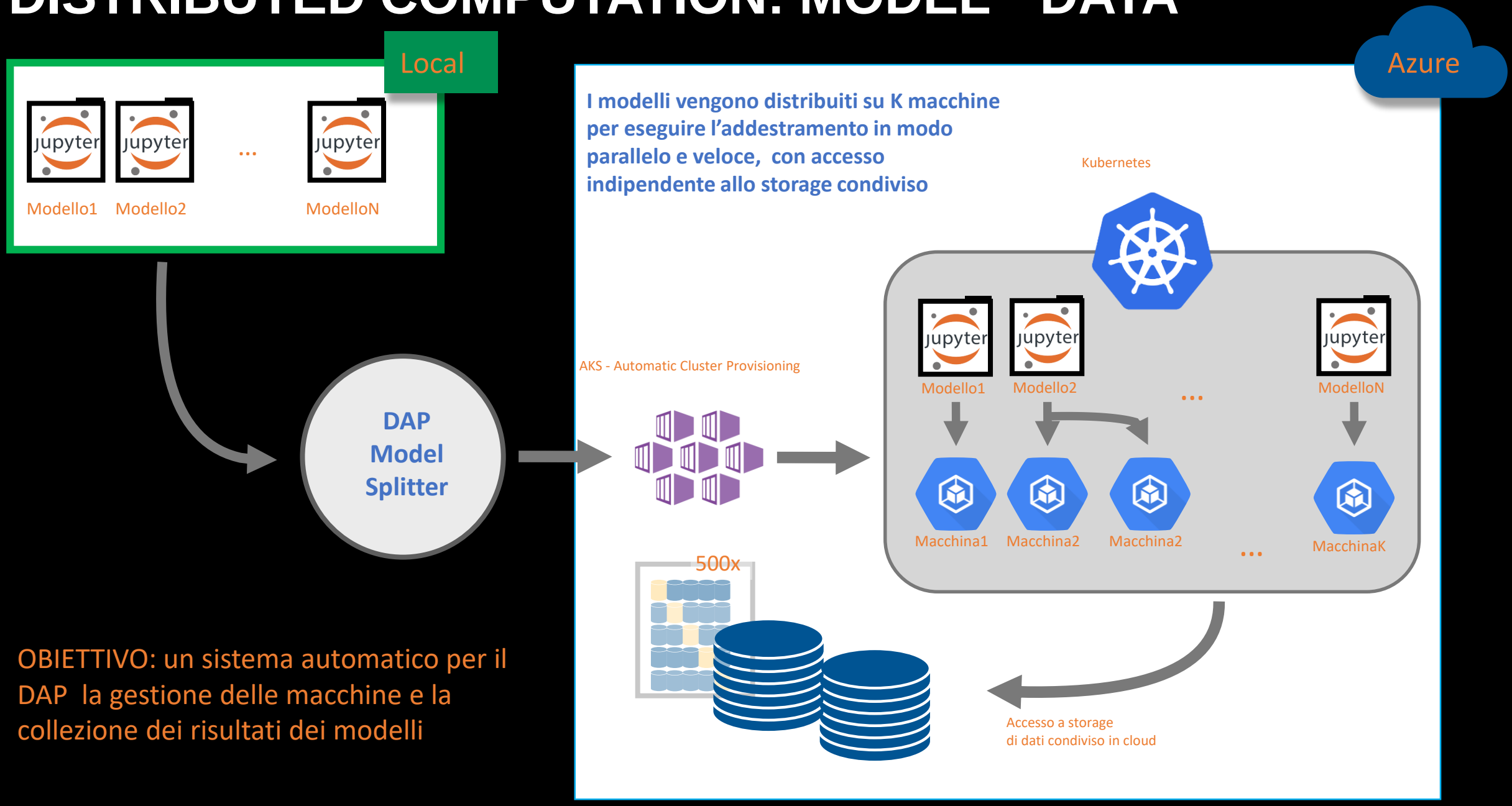


DAP
Data Loader

Azure

Dati in
Azure Blob Storage

DISTRIBUTED COMPUTATION: MODEL * DATA



OBIETTIVO: un sistema automatico per il DAP la gestione delle macchine e la collezione dei risultati dei modelli

WebValley is the FBK summer school for data science and interdisciplinary research: close to 350 students from around the world (17-19y old) have attended the WebValley camps since its first edition in 2001.

In 2016 and 2017, the team developed a new Deep Learning solution for fruit quality control based on portable low cost spectrometry and imaging

Agritech as an accelerator of Precision Medicine: Deep Learning, cloud infrastructure (MS Azure), local GPU boxes, blockchain

START FAST, START EARLY

Thanks to G. Jurman, M. Chierici, C. Dolci, M. Cristoforetti, V. Maggior, Ernesto Arbitrio, G. Franch

Deep Learning Projects: omics e immagini

WebValley2018 Giugno-Luglio Bambino Gesù Roma: radiomica per oncologia pediatrica

(Locatelli, Mastronardi, Vinci, Colafati, Tomà)

Predictive models and privacy-by-design in Healthcare:

- Deep Learning (radiologia + omics + clinica)
- Blockchain
- Omics biomarkers

Bambino Gesù, neuroblastoma: digital pathology (TILs) + panel immunologico GE

(Fruci, Melaiu, Locatelli)

Deep Learning for Electronic Health Records (EHR) – Mount Sinai 2018

Tecniche di Artificial Intelligence in Colonoscopia diagnostica-terapeutica» Gastroenterologia TN

Medicina di precisione nel Carcinoma Renale – Urologia TN

Deep Learning per Tumori ai polmoni Radiomica CT+PET: Medicina Nucleare BZ

+ "biopsia liquida": Wistar Institute

Metagenomica

Modelli di prevenzione e terapia nella disbiosi intestinale (DL + Complex Systems) – Lyon; in Autismo - OPBG e Sc. Cognitive UniTN.

REPRODUCIBILITY AS ACCOUNTABILITY. EVERYWHERE

FDA's MAQC/SEQC → WebValley 2018

- Upscale DL within Data Analysis Plans,
- Ledger for all steps in training, auditing of contact (automated or not) by predictors, check for hidden prototypes shadowed in models
- Cover both upstream feature extraction and downstream analytics

FONDAZIONE
BRUNO KESSLER

Cesare Furlanello

@furlanello

GRAZIE !

The challenge of reproducibility in Deep Learning at scale

DATASCIENCE // MPBA

15 June 2018

MPBA
PREDICTIVE MODELS
FOR BIOMEDICINE & ENVIRONMENT

Fondazione Bruno Kessler (FBK)
Center for Information & Communication Technology
Via Sommarive 18 – 38123 Trento (Italy)

With Valerio Maggio, Stefano Fioravanzo,
Marco Cristoforetti, Giuseppe Jurman,
Nicole Bussola, Margherita Francescatto,
Andrea Bizzego, Marco Chierici