

# Where AI and Big Data can help us to understand faster and better

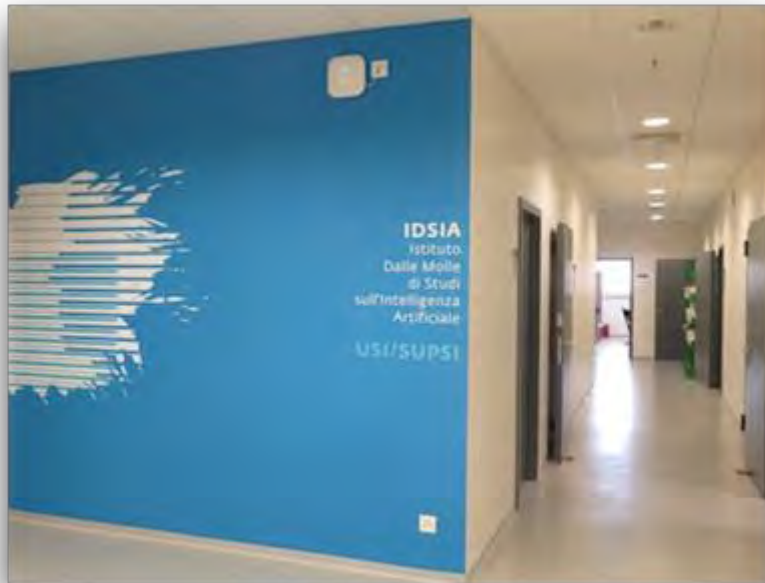
Andrea Danani



# IDSIA and Artificial Intelligence

# What is IDSIA?

IDSIA is a research institute on Artificial Intelligence <sup>[L]</sup><sub>[SEP]</sub> founded in 1988 in Lugano by the Italian philanthropist Angelo Dalle Molle (1908-2002)



IDSIA affiliated with  
USI and SUPSI since 2000



«Die Fortschritte der Wissenschaft im Allgemeinen und die der aufstrebenden Informatik im Besonderen den Menschen nicht unterwerfen, sondern ihm nützen sollten.»

# IDSIA: about 65 people

8 Professors  
26 Post Doc researchers  
18 PhD students ....

## Research Areas

- Deep Neural Networks
- Data mining and Machine Learning,
- Optimization Algorithms
- Bioinformatics, Robotics

## Projects with

- SNF-Swiss National Science Foundation
- CTI - Commission for Technology and Innovation
- European Commission, Direct Mandate

## Teaching

- Bachelor and Master in Informatics at SUPSI
- Master in AI at USI

# What is Artificial Intelligence?

AI is a science and a set of computational technologies that are inspired by the ways living systems use their nervous systems and bodies to sense, learn, reason, and take action





# Incredible progress since 2012

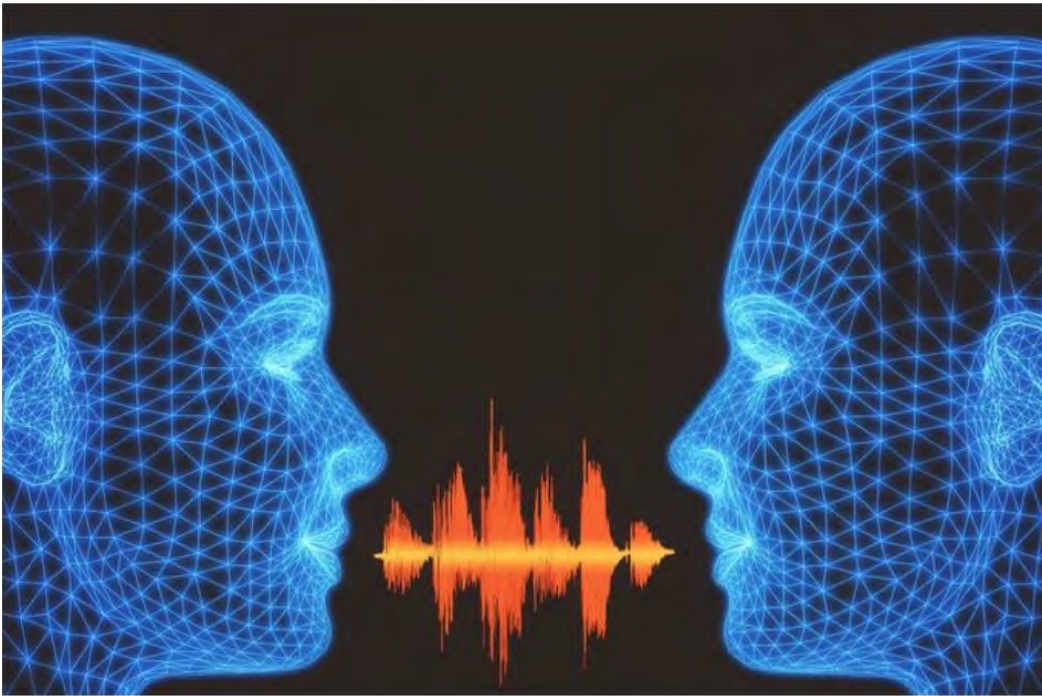


Image Recognition

Speech Recognition

Language Translation

Self-Driving Cars

Health Care

Retail and eCommerce

File Modifica Visualizza Cronologia Segnalibri Strumenti Aiuto

http://...6550 Fourtieth ... SUPSI - D... Dipartim... Servizi Web S... iCorsi2: ... DTL.C080... Competit... http://...html Google T... INFO.B.P... Scuola U... Giorgio Cora... EU Scienc... R&D proj... benef 16... Forest - It... X Priz...

www.theverge.com/2016/2/17/11032004/x-prize-ai-contest-ibm-watson-ted-2020


Più visitati Come iniziare Ultime notizie google translate iCorsi2: Login al sito

**TECH**

# X Prize and IBM announce a \$5 million artificial intelligence competition

2 COMMENTS

By Sean O'Kane on February 17, 2016 12:45 pm Email @sokanez




(IBM)

Share on Facebook (223) Tweet Share (28) Pin

The X Prize Foundation and IBM have just announced a new global X Prize competition with a focus on artificial intelligence. Teams from around the world can take part in the "IBM Watson A.I. X Prize: A Cognitive Computing Competition," as it's being called, in hopes of taking home part of a \$5 million purse to be awarded at the TED conference in 2020. Registration will open at the end of May.

In the early stages of AI development, opinions vary wildly on what it might be capable of, or what it will be useful for. That's especially true for the heads of some of the world's leading technology companies. Facebook's Mark Zuckerberg thinks it would be best suited to **turn his home into a smart house**. Tesla and SpaceX CEO Elon Musk is **afraid of**



**THE LATEST HEADLINES**

Chevy will let you pick up and drive your new car at Daytona International Speedway

Mossberg: The Nextbit Robin is a computer



# Tesla Autopilot: Behind the wheel of the world's first AI-powered driving experience

TechRepublic's Hope Reese drove an Autopilot-enabled Tesla Model S and evaluated the most impressive features as well as the common misperceptions about the technology.

By Hope Reese | October 5, 2016, 4:00 AM PST



auf Plus Swiss.  
29.-/Mt. statt 59.-  
Schweizweit unlimited.  
Nur für kurze Zeit auf salt.ch.

Mehr Infos



10    f 119    in 20    [Twitter icon]    [Menu icon]



## WHITE PAPERS, WEBCASTS, AND DOWNLOADS

White Papers // From CA Technologies

### An Enterprise Architect's Guide to API Integration for ESB and SOA (French)



Today's enterprise needs to reduce complexity, particularly at the IT integration level. The right API management solution should help you simplify integration and work across

ESB, SOA, cloud and IoT - all without writing code or adding connectors.

DOWNLOAD NOW

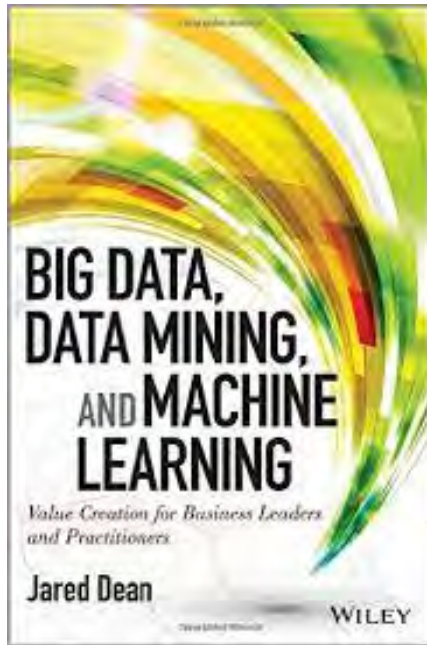
White Papers // From CA Technologies

### How do I Balance Robust Security with a Frictionless Online Shopping Experience for Cardholders?

DOWNLOAD NOW

White Papers // From CA Technologies

# Where do we apply AI?

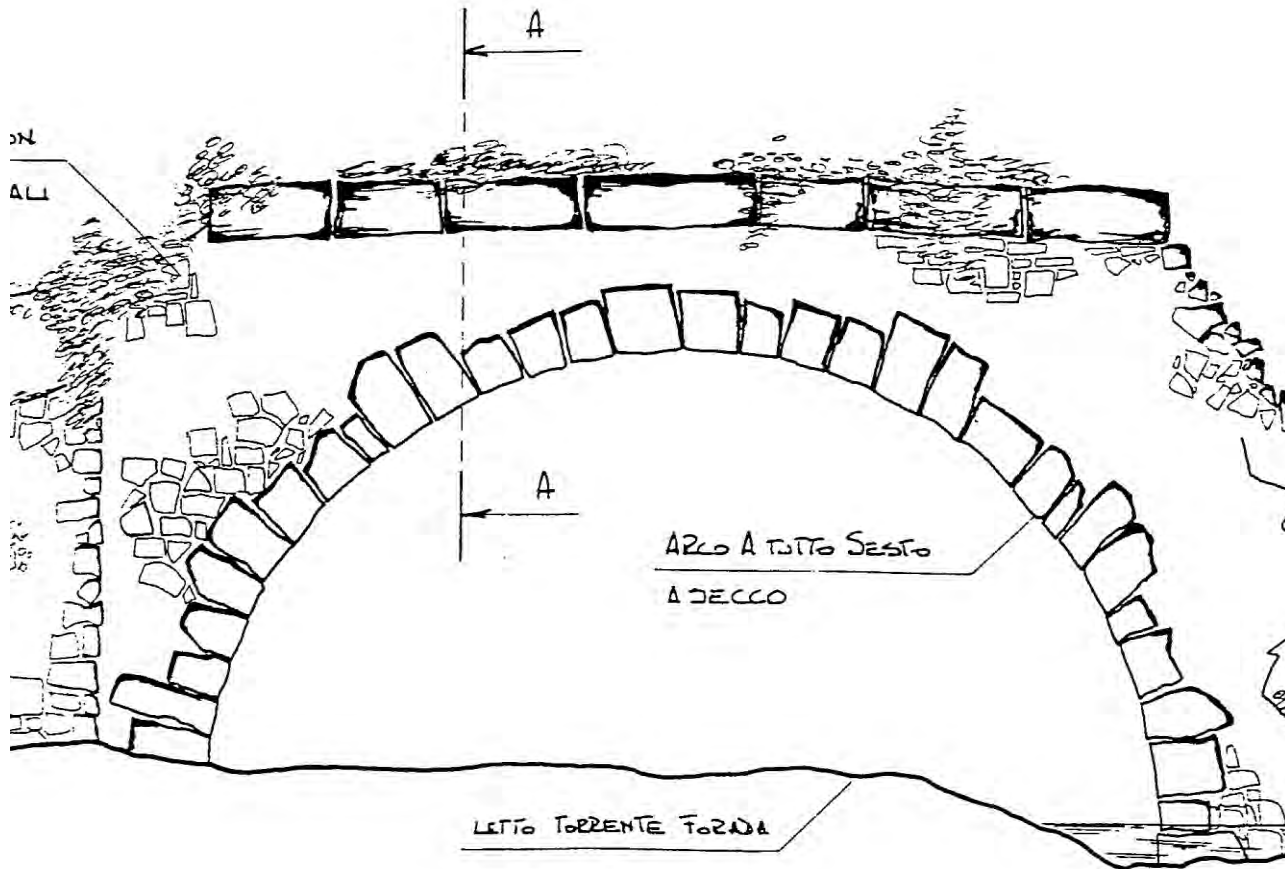


Industry  
4.0



robotics

# IDSIA: from basic to applied research



Basic Research

Applied Research

# MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21st century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

## MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

## DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative



## PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing packages, e.g., R
- ☆ Databases: SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

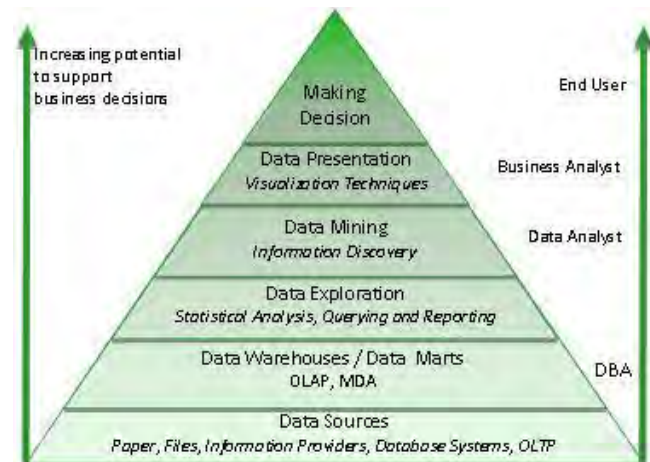
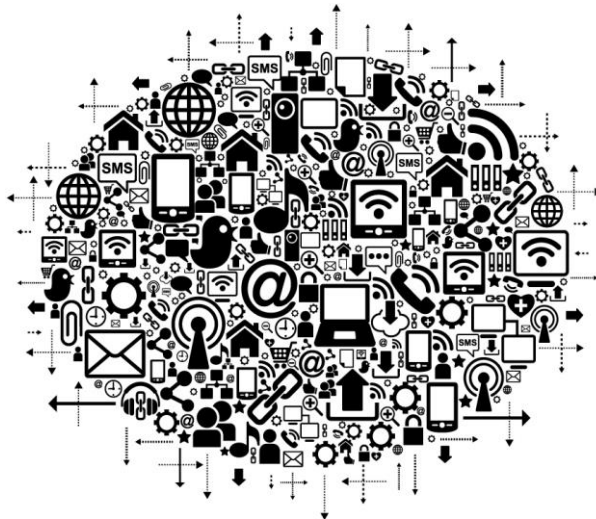
## COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

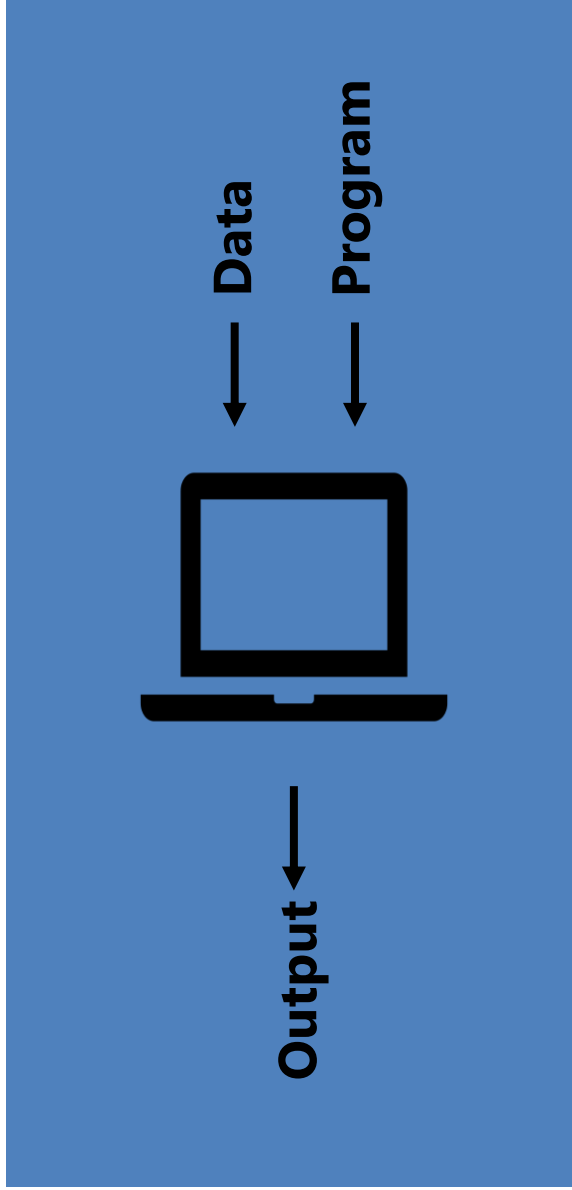
# Machine learning e data mining: Learn from data and experts



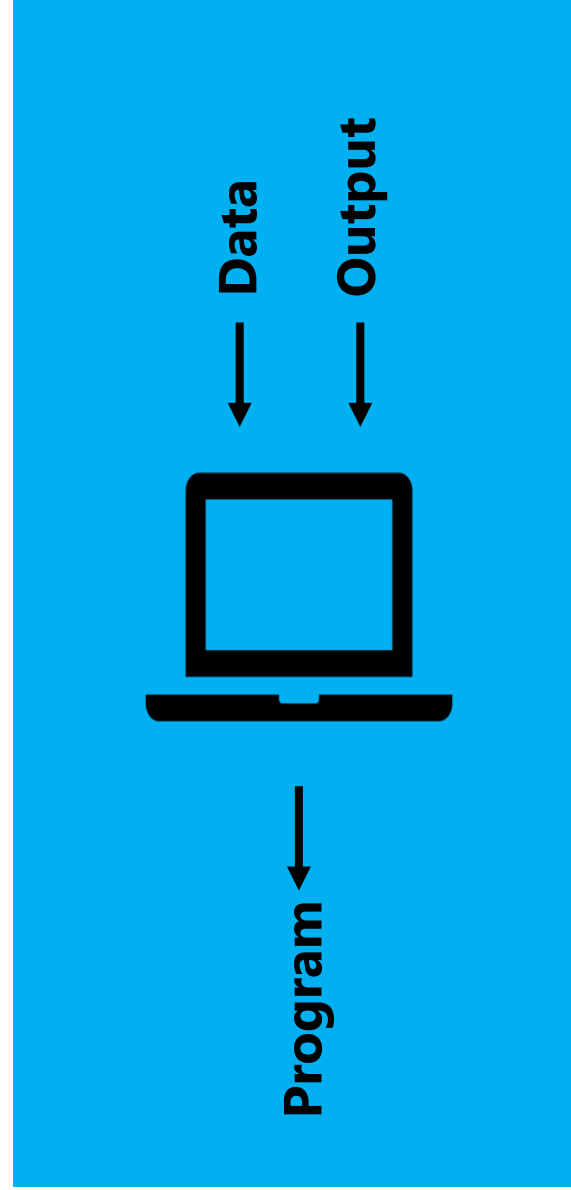
- Numbers
- Characters/texts
- Sounds
- Images/video
- Graphs
- DNA
- Non structured data



# Traditional Programming

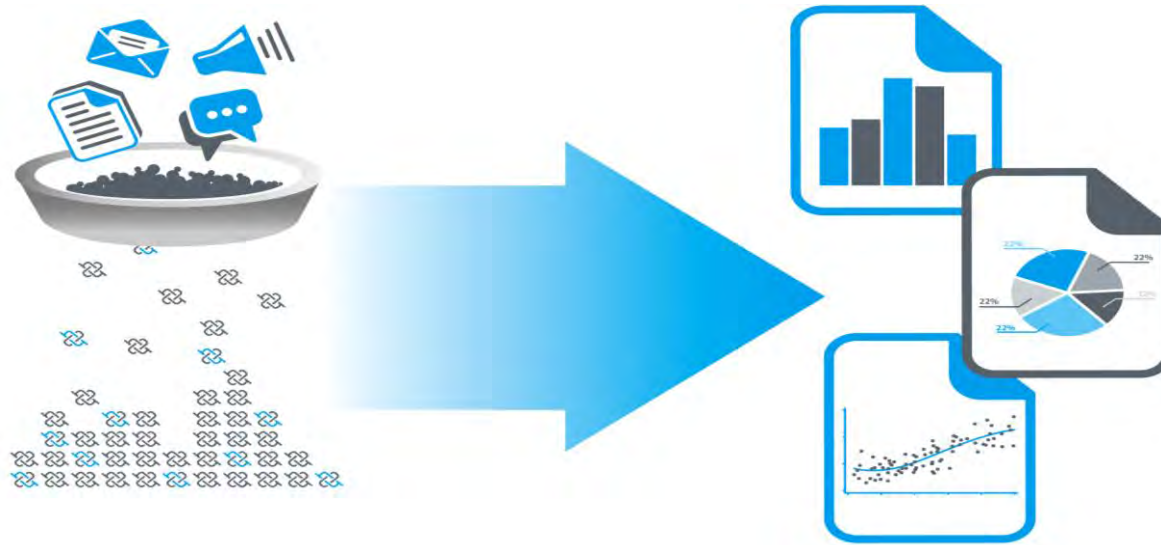


# Machine Learning



# Learning from (BIG) data

- Search for patterns in data



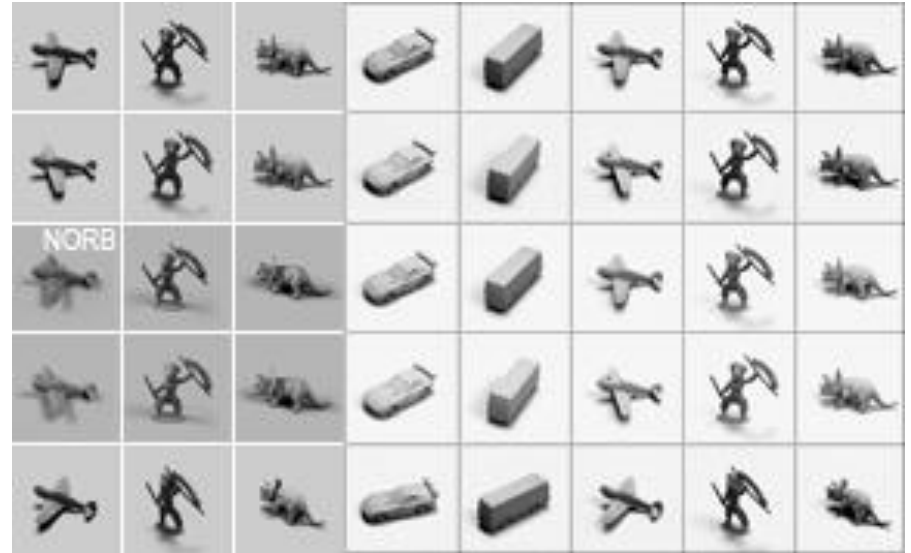
- set of patterns = a **model**
- allows us to structure information
- Models **can be queried**
  - for prediction, diagnosis, recognition, ...
- (Sort of) Domain independent
  - the meaning of data is not always needed



# Artificial vision & neural nets

## ➤ Artificial vision

- is fundamental in many practical applications
- future systems will be based more on images than on texts
- robots have to understand their environment
- automatic classification for medical, navigation, recognition



## ➤ Neural networks

- inspired by the human brain (tiny compared to it)
- universal function approximators
- **deep = sort of “big” nets = sort of resurrection of these models**
- often thanks to hardware speedup



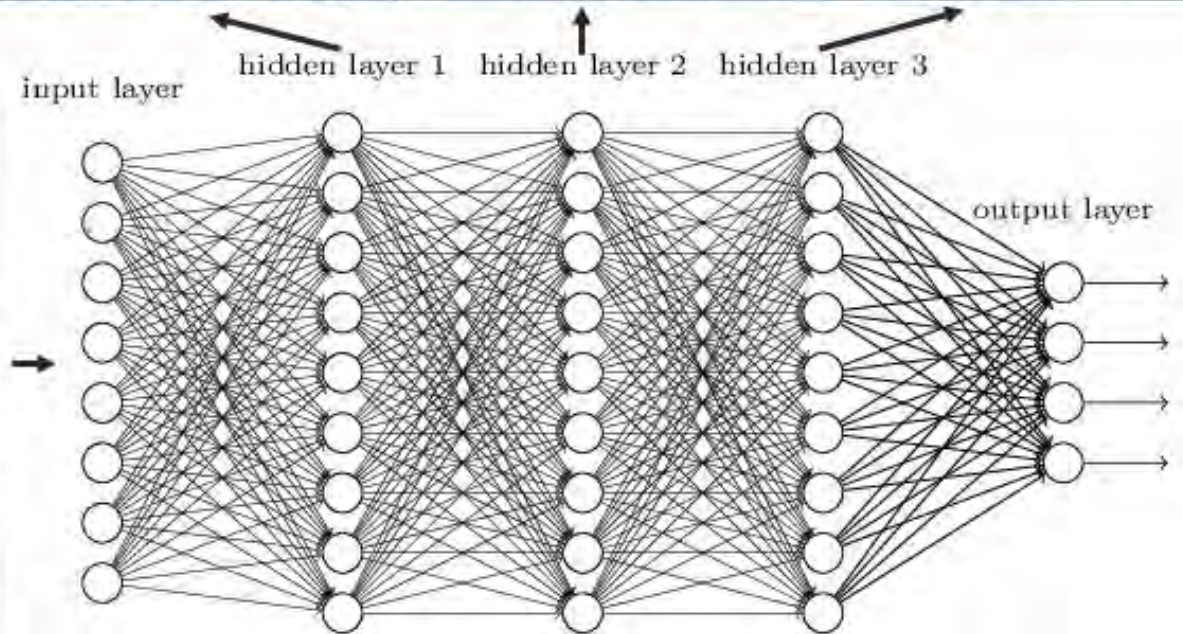


# Artificial vision & neural nets



# Deep Learning : hierarchy of abstractions

Deep neural networks learn hierarchical feature representations

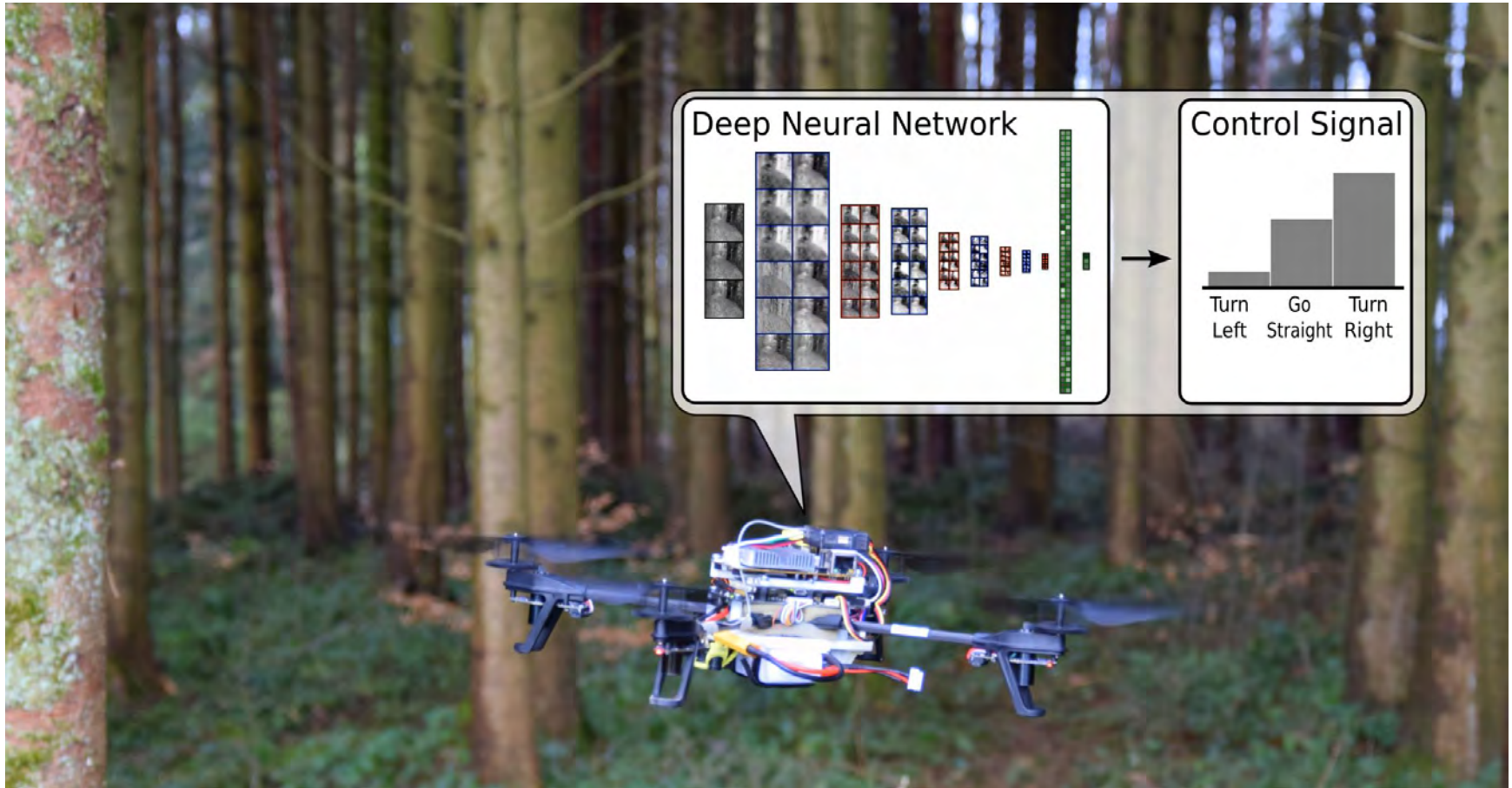


In Deep Learning, you have a lot of relatively simple layers. You increase learning capabilities by increasing the number of layers, as opposed to increased complexity of layers.



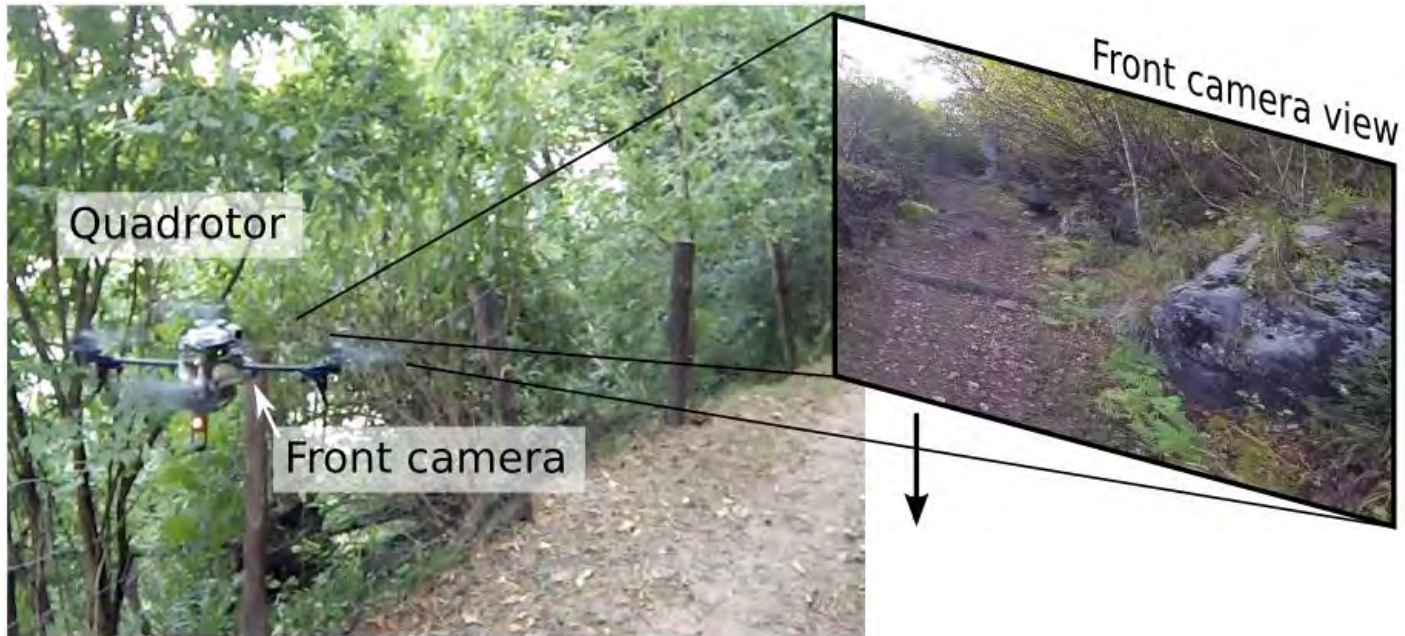
# Some Examples

# Visual Perception of Forest Trails

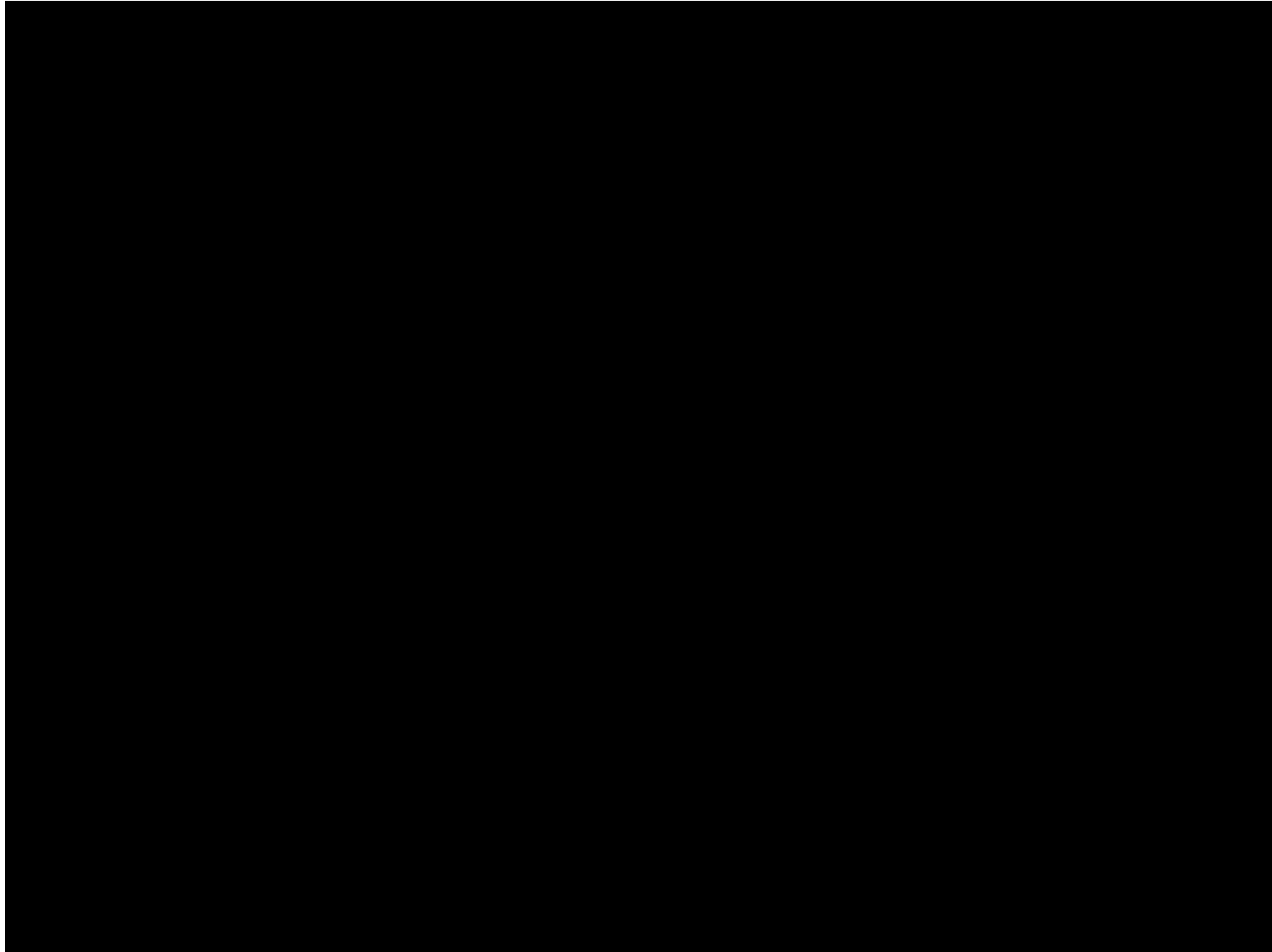


# Drone: visual perception based navigation in Forest Trails

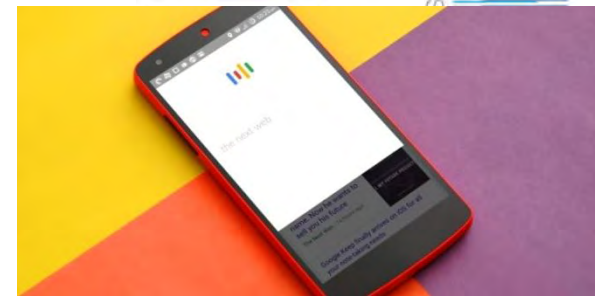
- Drone autonomously follows a forest trail
- Applications in search and rescue



# Neural Autopilot for Drone in the forest



# We give voice to Google !!



## Google Research Blog

The latest news from Research at Google

### Google voice search: faster and more accurate

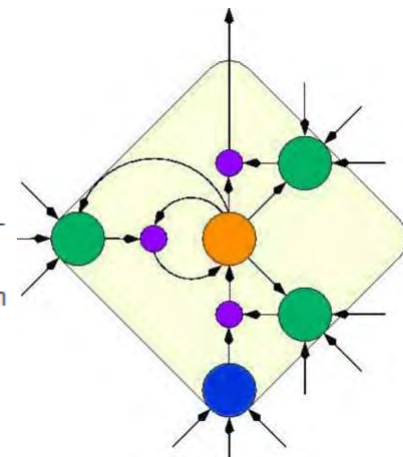
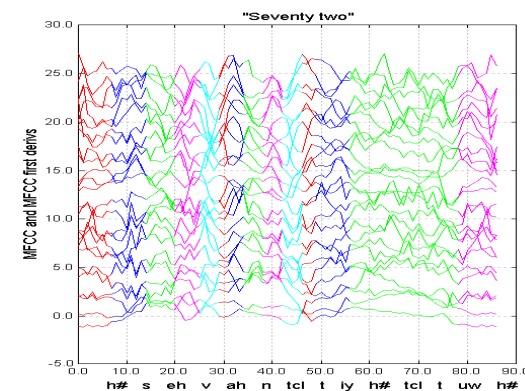
Posted: Thursday, September 24, 2015



Posted by Haşim Sak, Andrew Senior, Kanishka Rao, Françoise Beaufays and Johan Schalkwyk – Google Speech Team

Back in 2012, we announced that Google voice search had taken a new turn by adopting [Deep Neural Networks \(DNNs\)](#) as the core technology used to model the sounds of a language. These replaced the 30-year old standard in the industry: the Gaussian Mixture Model (GMM). DNNs were better able to assess which sound a user is producing at every instant in time, and with this they delivered greatly increased speech recognition accuracy.

Our improved acoustic models rely on [Recurrent Neural Networks \(RNN\)](#). RNNs have feedback loops in their topology, allowing them to model temporal dependencies: when the user speaks /u/ in the previous example, their articulatory apparatus is coming from a /j/ sound and from an /m/ sound before. Try saying it out loud - "museum" - it flows very naturally in one breath, and RNNs can capture that. The type of RNN used here is a [Long Short-Term Memory \(LSTM\)](#) RNN which, through memory cells and a sophisticated gating mechanism, memorizes information better than other RNNs. Adopting such models [already improved the quality](#) of our recognizer significantly.



# We give people to Google

## Google to acquire artificial intelligence company Deep Mind

Monday, Jan 27 2014, 10:57 GMT

Google is reportedly close to acquiring artificial intelligence company Deep Mind.

The web giant has agreed to pay \$500 million (£302m) for the London-based startup

DeepMind is a cutting edge artificial intelligence company. We combine the best techniques from machine learning and systems neuroscience to build powerful general-purpose learning algorithms.



LUGANO - Si chiama Shane Legg, ha conseguito il suo dottorato di ricerca presso l'Istituto Dalle Molle di studi sull'intelligenza artificiale ed è uno dei tre fondatori di DeepMind ...

## SUPSI

[People directory](#) [AZ Index](#) [Reserved area](#)

- > SUPSI
- > Bachelor, Diploma and Master
- > Continuing education
- > Research

### Profession and Passion

Celine Cometti, Student, Engineering and Management. Flexibility, versatility, inter-disciplinarity and technology are future of tomorrow.



### People directory

NEW SEARCH

Person

Name, Surname or phone number

Unit

All



### Search details

Contacts »

Publications

Research Projects »

Name and Surname

Shane Legg

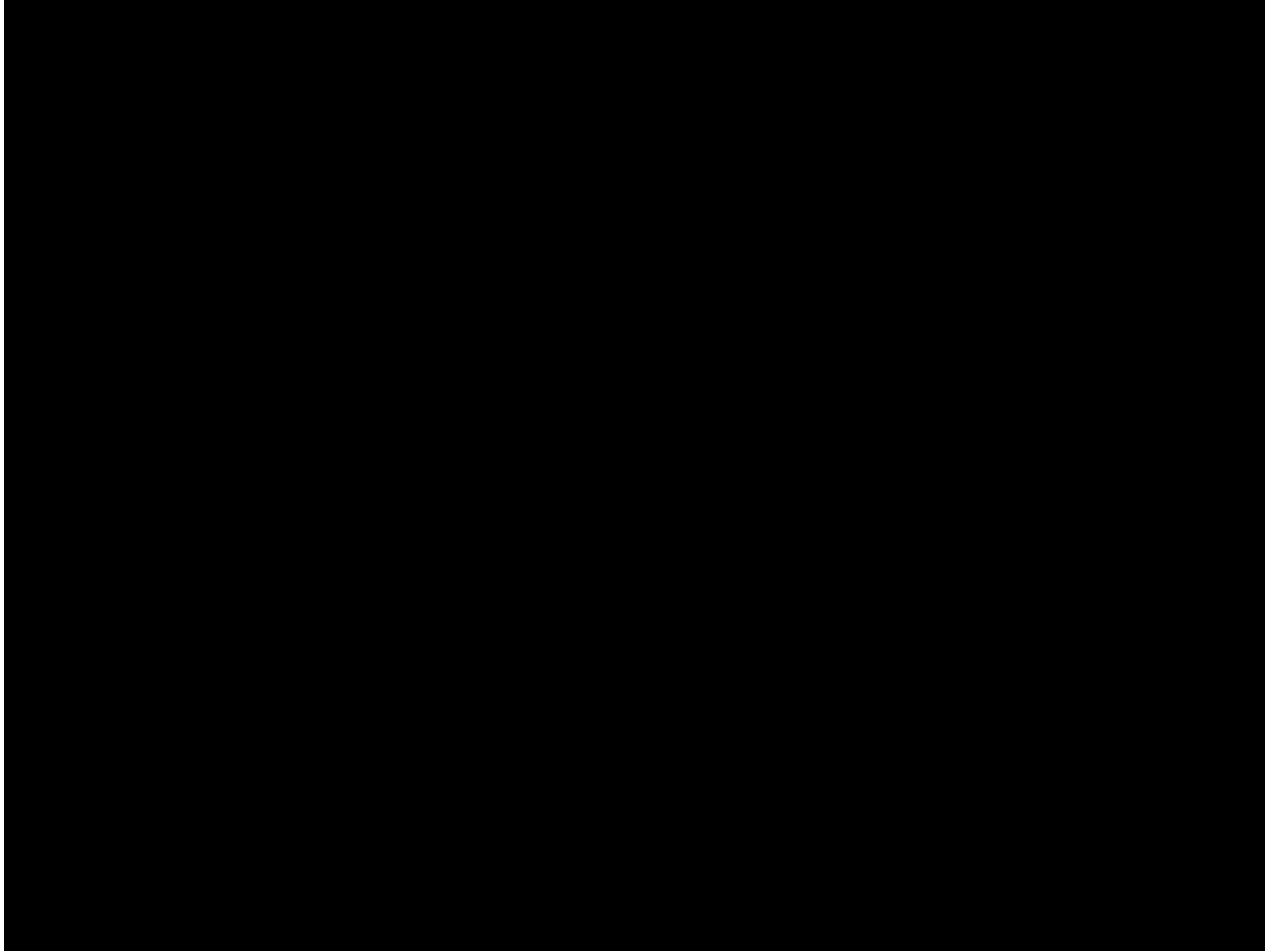
Department/School/Unit

Dalle Molle Institute for Artificial Intelligence

Send card by email



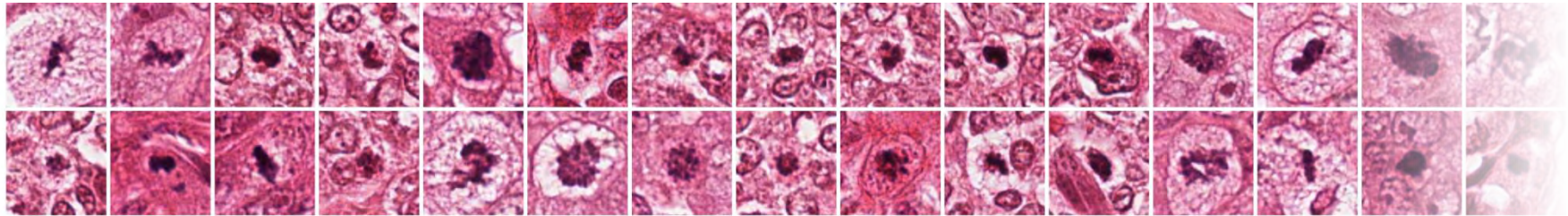
# DeepMind - Nature



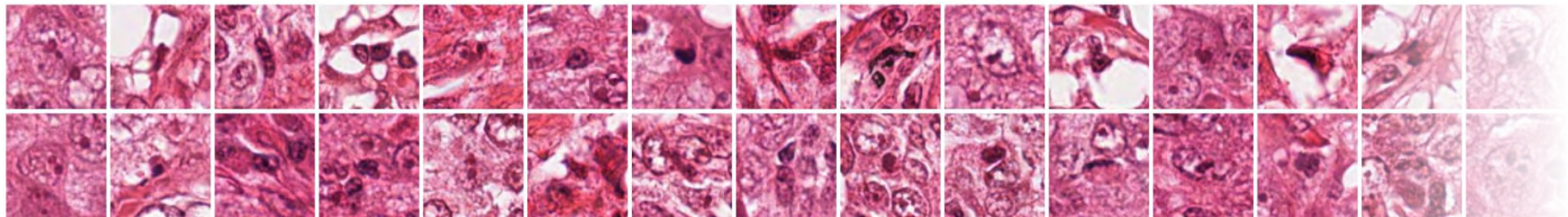
# Detection of mitotic nuclei in breast cancer histology images

Big Data / Deep Learning techniques applied to Biomed Imaging:  
huge, GPU-trained neural nets learn to solve challenging pattern  
recognition problems from labeled training datasets

mitosis (C1)



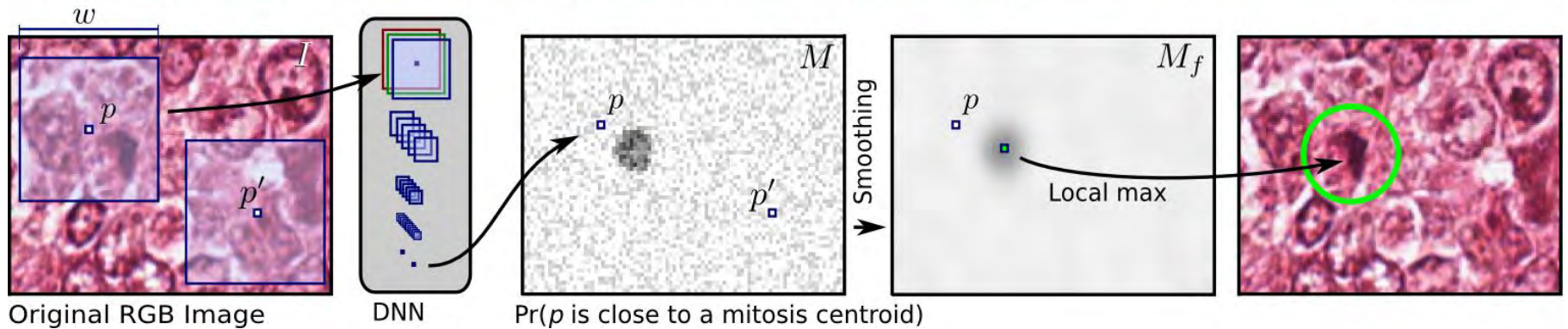
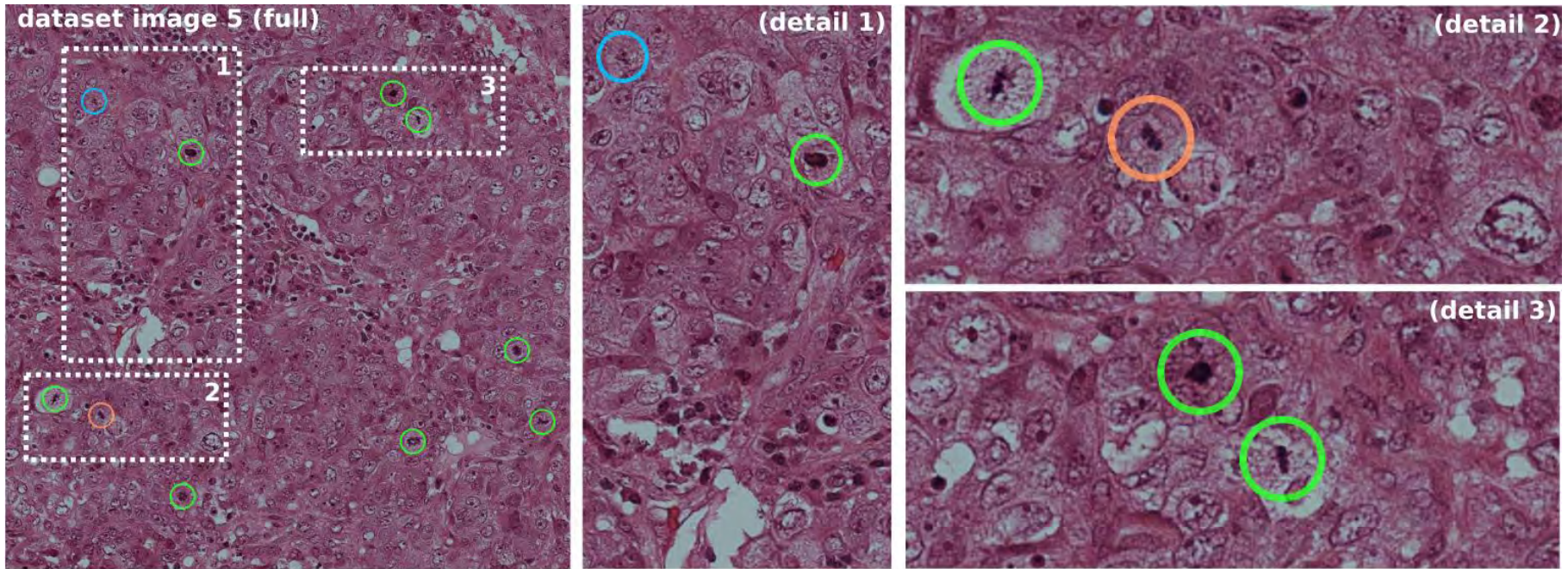
nonmitosis (C0)



## Selected Publications by IDSIA Researchers

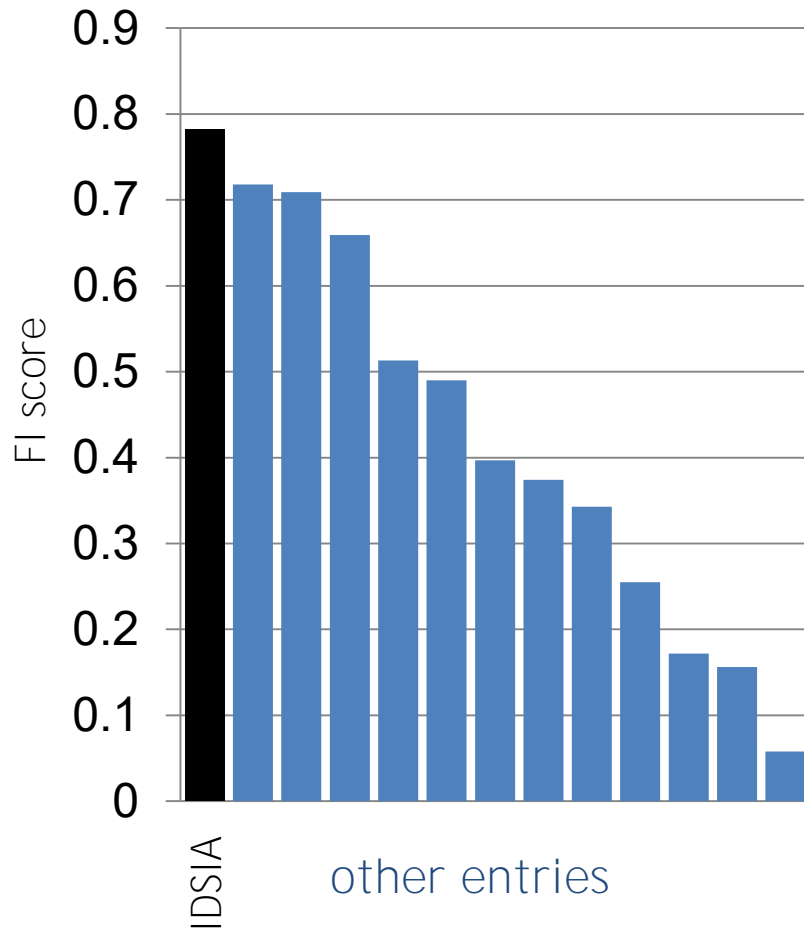
- Veta et al., Medical Image Analysis 2016
- Giusti et al., ISBI 2015
- Ciresan et al., MICCAI 2013

# Detection of mitotic nuclei in breast cancer histology images

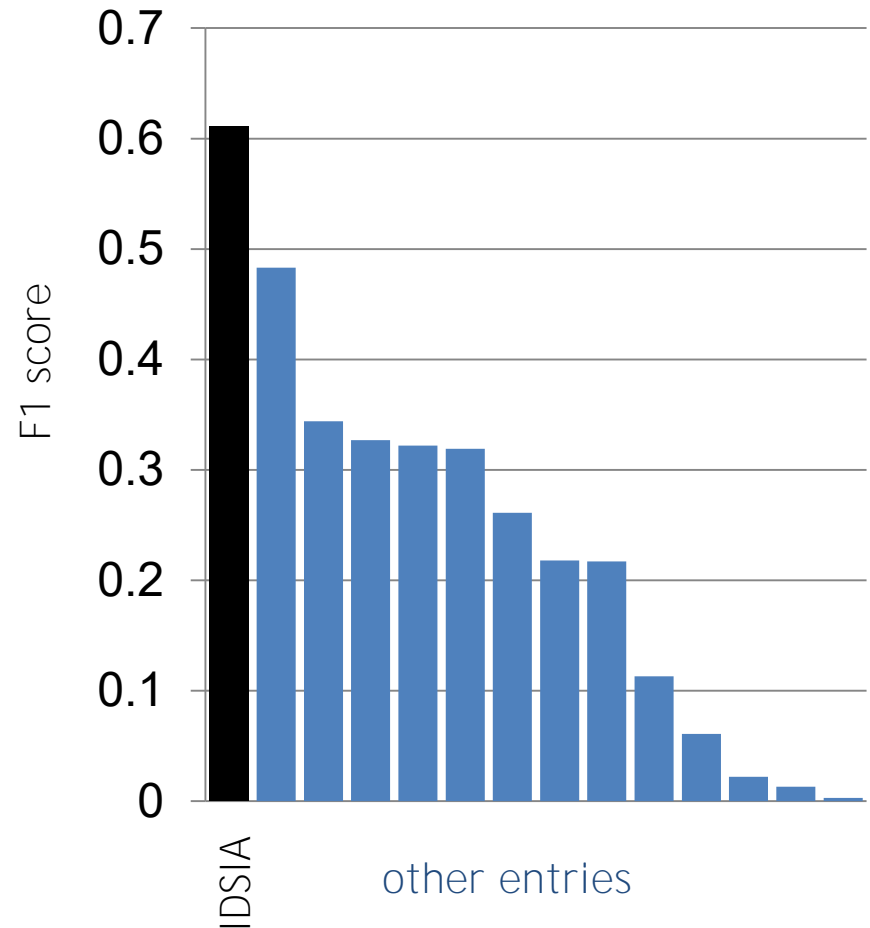


# Results of Mitosis Detection Competitions

ICPR 2012 International Competition  
50 images, 300 mitosis

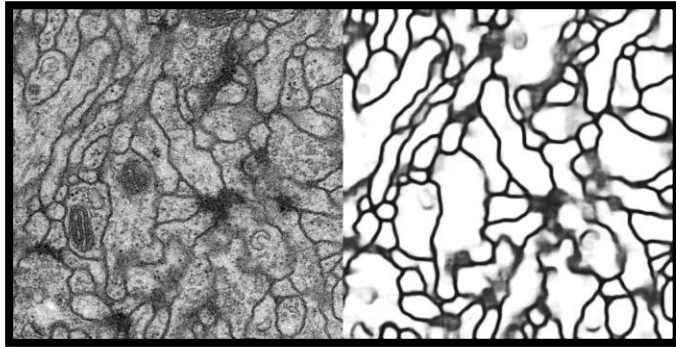


MICCAI 2013 International Comp.  
600 images, 1157 mitosis

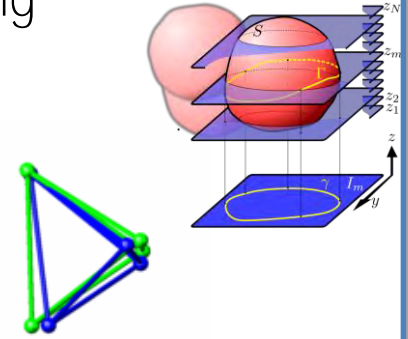
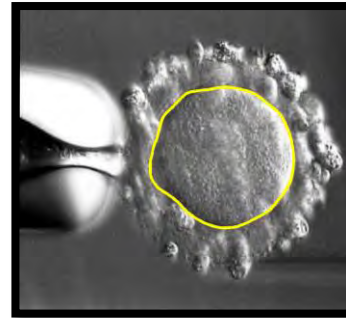


# Microscope Image Analysis with Deep Learning Techniques

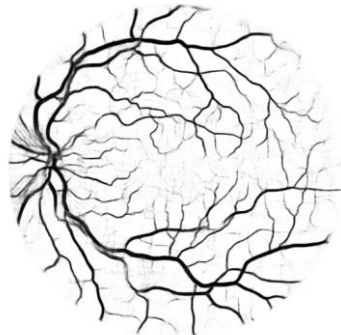
Segmentation of neural membranes in EM



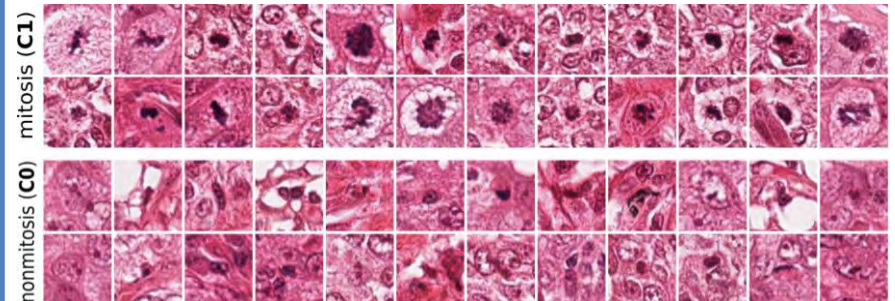
Zygote / Embryo Assessment and Scoring



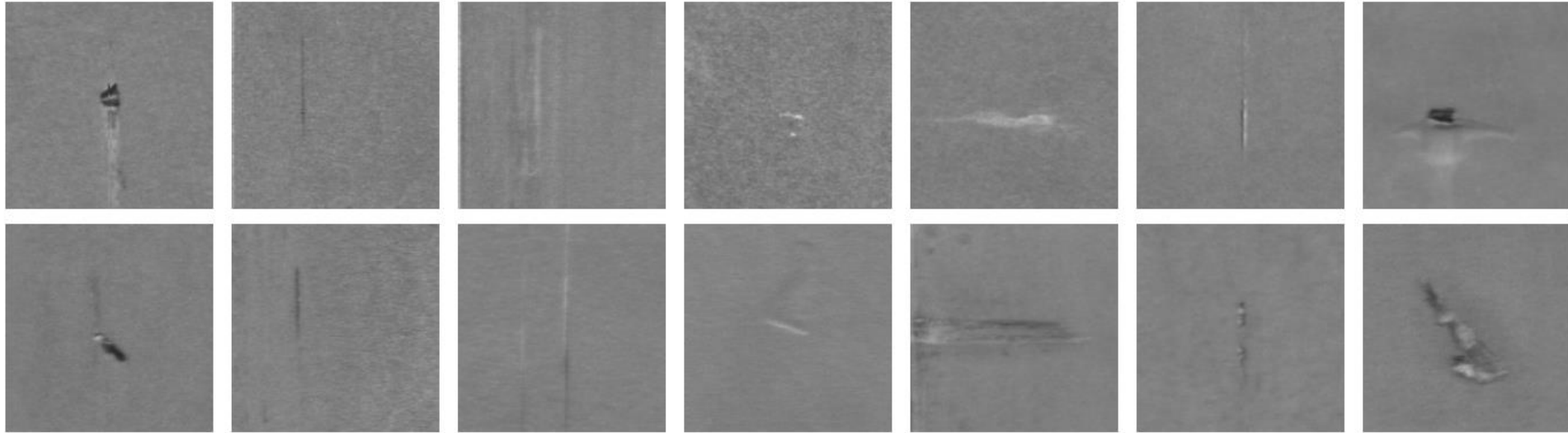
Segmentation of retinal vessels



Mitosis Detection in Whole-Slide Histology Images



# Classification of Steel Defects (with Arcelor Mittal)



	0	1	2	3	4	5	6
0	0.91	0.01	0.01				0.07
1	0.09	0.85				0.07	
2	0.01		0.99				
3	0.04			0.94	0.02		
4	0.02			0.02	0.97		
5		0.05	0.11			0.84	
6	0.14						0.86

	0	1	2	3	4	5	6
0	0.80	0.01	0.01	0.08	0.02	0.01	0.08
1	0.11	0.89					
2	0.03	0.02	0.94			0.01	
3	0.14		0.02	0.83	0.01		
4				0.09	0.91		
5	0.05	0.19				0.76	
6	0.25				0.01		0.74

	0	1	2	3	4	5	6
0	0.85	0.01		0.03	0.01	0.01	0.11
1	0.12	0.88					
2	0.01		0.98			0.01	
3	0.09		0.01	0.88	0.02	0.01	
4			0.02		0.98		
5	0.05	0.05		0.14		0.76	
6	0.20						0.80

Fig. 5. Confusion matrices for the best classifiers. Left: MPCNN, middle: PHOG, right: PHOG + MONO-LBP committee. Only on defect number 2 the classical features obtained a better result than our MPCNN. Also note the non marginal improvement of a committee w.r.t. the single best classifier.

# Pharmaceuticals: formulation optimization for new natural compounds



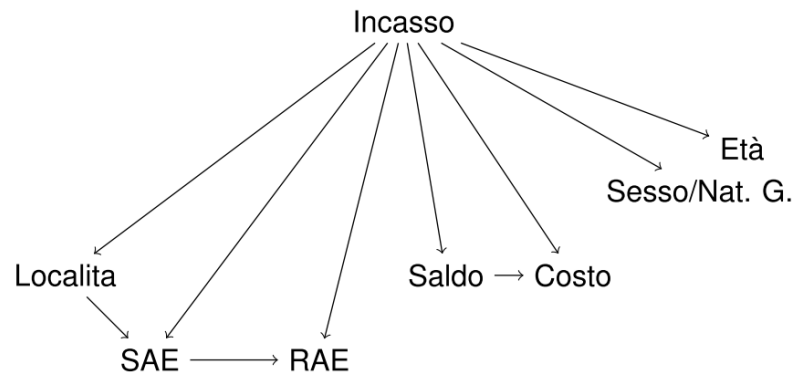
In Silico screening to support in vitro experiments

# Portfolio of loans in litigation

Learning from historical data the value of new portfolios

## Rete credale

- Modello (grafico) delle relazioni fra i diversi attributi (del debitore)



- Permette anche di valutare l'importanza dei diversi attributi  
(es. Saldo > Età > Sesso)



# What is a non-performing loan?

You buy something: a house, car, television **or cash ... and for** some reason you cannot pay the installments anymore

... after some months your creditor asks you for his money back

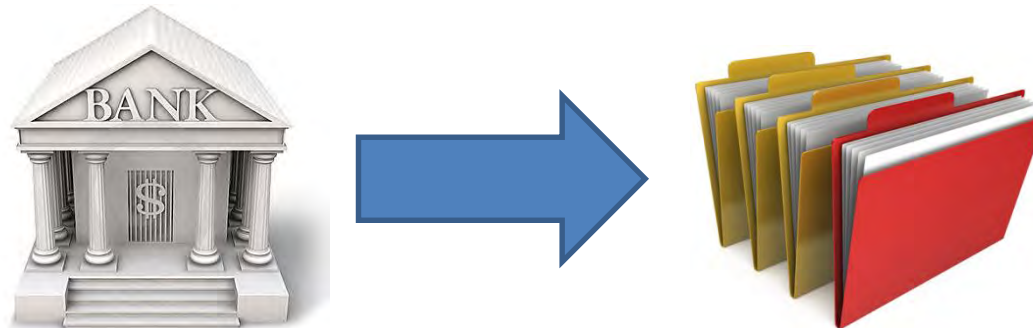
If you do not pay ... **the** forced recovery activity starts ... **and** your claim becomes a non-performing loan



# Non-performing loans

NPLs are problematic for banks ... **a large** amount of them means bad balance sheets ...

Banks can decide to sell NPL portfolios to investors that will perform the recovery activity better, faster and cheaper



... but ... **what is the selling price?**

QBT assesses this price for NPL portfolios

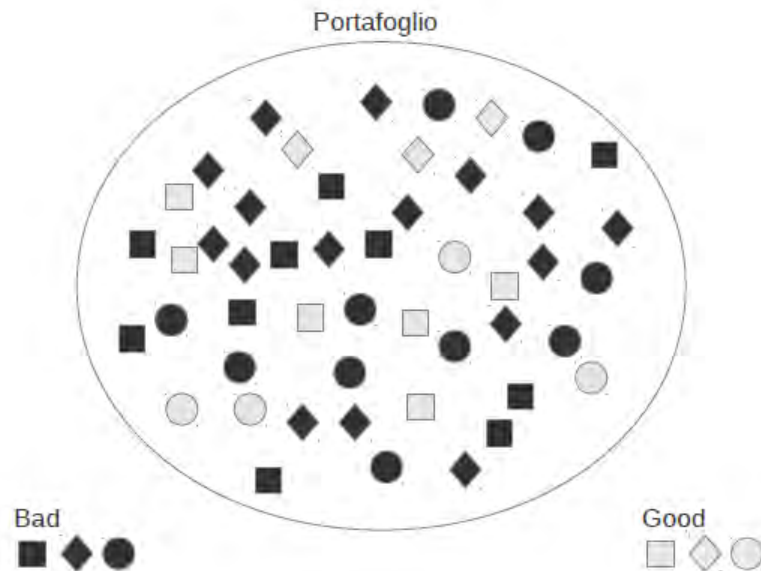
# Unsecured tool (Utool)

- The traditional statistical approach failed after the 2008 economic crisis:  
  
errors in cash flow prediction **bigger then 30%**
- We addressed the problem by **Machine Learning**

**KTI/CTI**

# Unsecured Tool

QBT-PF2012: database  
period range 2009-2012  
(22.000 borrowers)



Research steps:

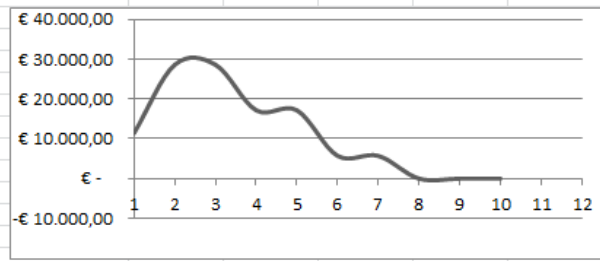
- Preliminary data analysis
- Model definition (training)
- Error estimations (testing)

# Unsecured Tool

UTOOL PREVISIONE	Incasso	Incasso%Saldo
Incasso Previsto Worst	78.253,34	5,07%
Incasso Previsto Average	114.417,03	7,41%
Incasso Previsto Best	295.492,90	19,13%
Saldo	1.544.360,07	

Linee: 30	Incasso Previsto Average	Saldo	Incassi lordi%Saldo
PREVISIONE	€ 114.417,03	€ 1.544.360,07	7,41%

AVERAGE CASE		
Incassi	€ 114.417,03	
Costi di gestione	€ 11.441,70	
Asset fee	€ 17.162,55	
NET RECOVERY	€ 85.812,77	
Saldo	€ 1.544.360,07	
Rendimento atteso	PRICING	Pricing%Saldo
18%	€ 52.694,28	3,412%
25%	€ 44.653,27	2,891%
30%	€ 39.955,99	2,587%



```

Building classifier...
Model saved.

Processing classifier #9
Building regressor...
Building classifier...
Model saved.

Processing classifier #10: building classifier...
Building regressor...
Building classifier...
Model saved.

Saving output datasets... done!
Saving data... done!
Total income:                0,0000
Worst case:                   34273,9813
Average case:                 40112,2843
Best case:                    93852,8466

Saving results...done!
Completed.

```

# Unsecured Tool

To confirm the appropriateness of our approach we performed a series of elaborations based on data from the past to obtain testable predictions

Below are the results for a portfolio of 5,000 loans with a series of historical data for 20,000 borrowers

<b>GBV</b>	<b>Pedicted income</b>	<b>Pred. Inc Vs GBV</b>	<b>Real income</b>	<b>Real Inc. Vs GBV</b>	<b>Delta %</b>
312.538.905,06	28.287.401,20	9,05%	30.090.831,50	9,63%	0,58%

Average error after Utool: 6%

# Who's working with Utool?

**PRIMUS CAPITAL**  
CREDIT MANAGEMENT

**DCM & PARTNERS**  
Distressed Credit Management

**BAYVIEW**  
ASSET MANAGEMENT

  
**UNIONE ARTIGIANI**  
*della Provincia di Milano*

**SENSALE  
CAPITAL**

**ARC**

  
**EAGLE & WISE**  
SERVICE

  
**REAG**

**DUFF & PHELPS**  
Real Estate Advisory Group

**AT**   
Advancing Trade spa  
a WCMG COMPANY

**CNF**  
FINANZIARIA  
società per azioni

  
**ITALIA SERVICE S.r.l.**

**samex.net**  
circuito di credito commerciale

**deepview**  
capital sa

**J Invest**



**Efficient**



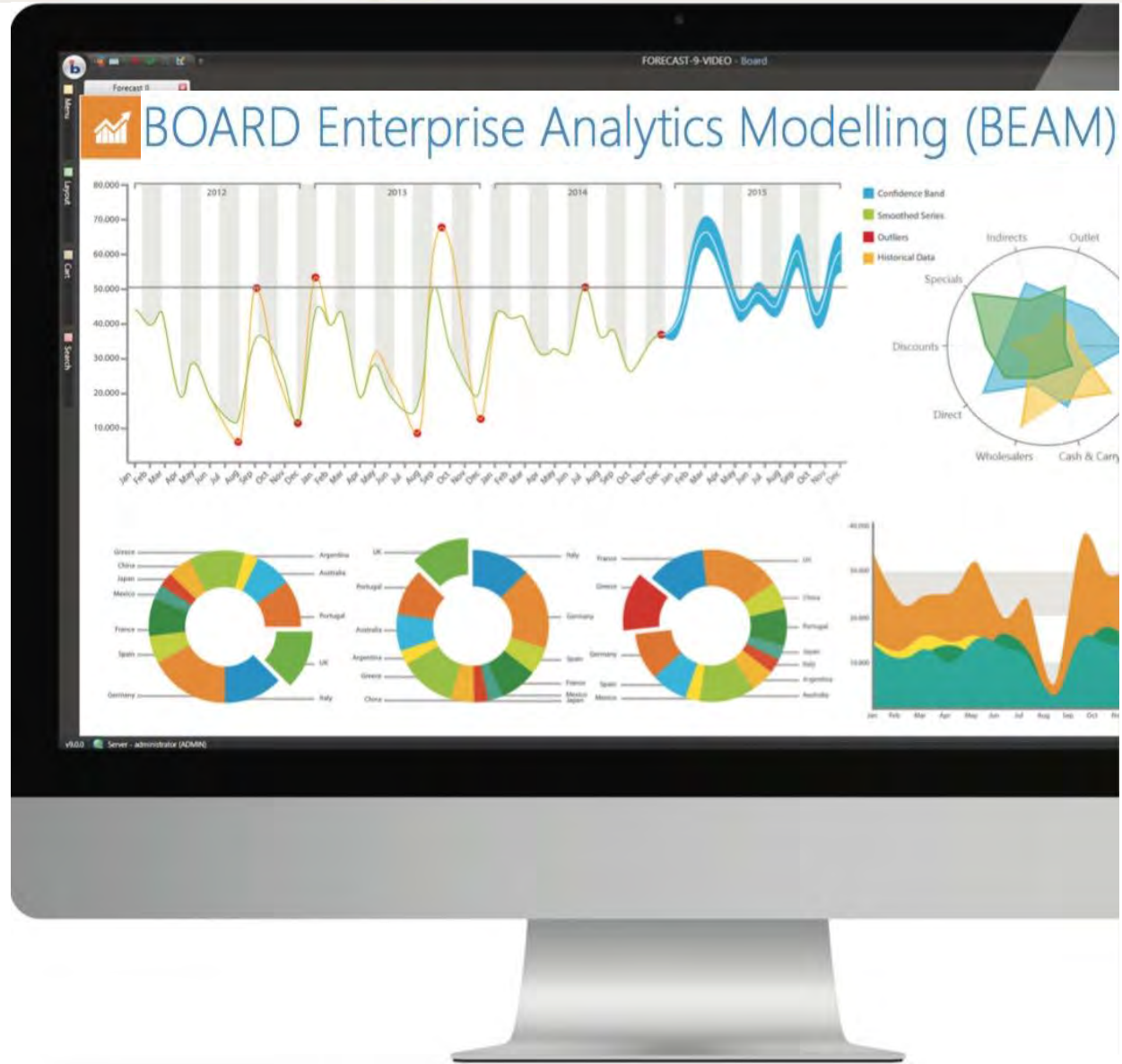
**Automated**



**Scalable**

**KTI/CTI**

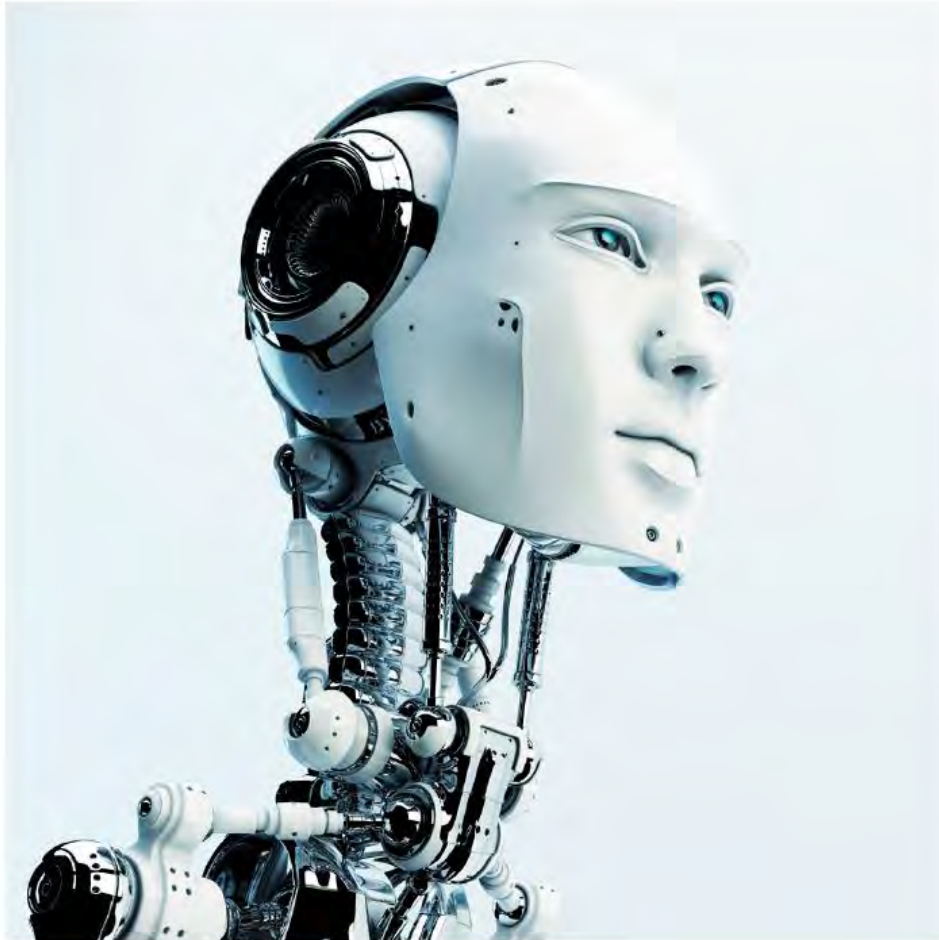
USI/SUPSI  
 Istituito  
 Dalle Molle  
 di studi  
 sull'intelligenza  
 artificiale  
 IDSIA





# Intelligent Automation

A UBS Group Innovation **White Paper**



## CORRIERE DEL TICINO

**Tecnologie, UBS punta sul Ticino**  
La banca svizzera si muove in corso di tempo per l'intero gruppo svizzero con circa 180 dipendenti nel settore di tecnologia e servizi

**Con l'ospedale arriva la scienza**

**Sotto tiro la libera circolazione**

**Il nuovo corso di studi per la laurea triennale in...**

**Il nuovo corso di studi per la laurea triennale in...**

**Il nuovo corso di studi per la laurea triennale in...**

# PORTFOLIO MANAGEMENT — UNDER — STRESS

A BAYESIAN-NET APPROACH TO  
COHERENT ASSET ALLOCATION



RICCARDO REBONATO  
AND ALEXANDER DENEV

CAMBRIDGE

# MANAGING UNCERTAINTY, MITIGATING RISK

Tackling the Unknown in Financial  
Risk Assessment and Decision Making

NICK FIROOZYE  
FAUZIAH ARIFF

## ... And other projects

**DufEnergy Trading**  
Dufenco GROUP

**KTI/CTI**

- Energy trading and eolic plants optimization

**hoosh**

**KTI/CTI**

- Visibility index from Google search results

**Medigest**

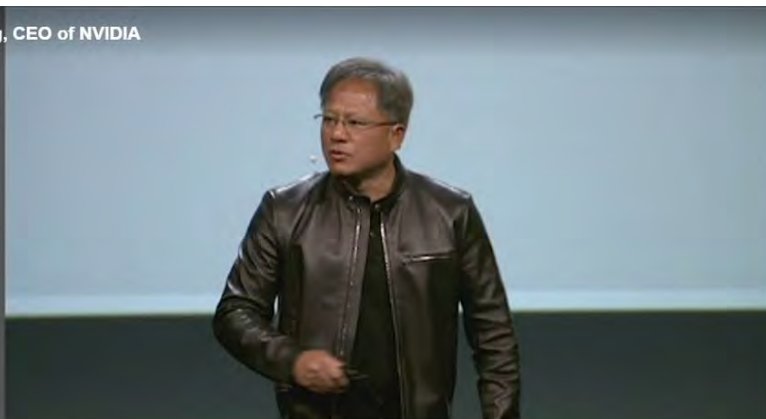
**KTI/CTI**

- Business analytics

 **IOR**  
Institute of Oncology Research

- Genetic analysis on lymfomes
- Medical data analysis and strategical planning

  
Ente Ospedaliero Cantonale



- Frameworks for Multi-GPU Pascal
- Large-scale Deep Learning
- Reinforcement Learning
- Unsupervised and Transfer Learning
- Natural Language Understanding
- Autonomous Driving
- Medical Applications

# PIONEERS IN AI RESEARCH

GTC keynote on Tuesday April 5, 2016 Jensen H Huang, CEO of NVIDIA

The prize is the new NVIDIA DGX-1 Deep Learning Supercomputer valued at **130'000\$**

# Swiss Special ICT Award 2016



One of the best international bio-inspired AI institutes and Swiss companies directly benefit from the work of its researchers

# robotics+

Swiss National  
Centre of Competence  
in Research



# H U G O B O S S

TINEXT ▶

ALPIQ

  
ASTES  
next technology  
board

HUPAC  
moving together

Medigest

 BOSCH

  
ArcelorMittal

 Lifeware

board

+GF+

inspire

hoosh  
meaningful numbers

EMPA   
Materials Science & Technology

JPSI  
SIA

# Concluding remarks

- AI is in the process of deeply changing our Society
  - An unprecedented opportunity in history
  
- The AI wave is here to stay
  
- Machine learning is a very big part of it
  - But machine learning is craft not science (yet)
  - People (well) before algorithms



Thank you for your attention