Medical Information Science for Decision Support

EXECUTIVE ACADEMY

EFMD EQUIS ACCREDITED

Assoc. Prof. Dr. Andreas HOLZINGER (Med. Uni Graz)

Day 1 –Part 4 -17.4.2018

DSS: from Expert Systems to explainable AI

Overview

Day 1 - Fundamentals

01 Information Sciences meets Life Sciences

02 Data, Information and Knowledge

03 Decision Making and Decision Support

04 From Expert Systems to Explainable AI

Keywords

- Artificial intelligence
- Case based reasoning
- Computational methods in cancer detection
- Cybernetic approaches for diagnostics
- Decision support models
- Decision support system (DSS)
- Explainable AI
- Fuzzy sets
- MYCIN – Expert System
- Reasoning under uncertainty
- Radiotherapy planning

Advance Organizer (1/2)

- **Case-based reasoning (CBR)** = process of solving new problems based on the solutions of similar past problems;
- **Certainty factor model (CF)** = a method for managing uncertainty in rule-based systems;
- **CLARION =** Connectionist Learning with Adaptive Rule Induction ON-line (CLARION) is a cognitive architecture that incorporates the distinction between implicit and explicit processes and focuses on capturing the interaction between these two types of processes. By focusing on this distinction, CLARION has been used to simulate several tasks in cognitive psychology and social psychology. CLARION has also been used to implement intelligent systems in artificial intelligence applications.
- **Clinical decision support (CDS)** = process for enhancing health-related decisions and actions with pertinent, organized clinical knowledge and patient information to improve health delivery;
- **Clinical Decision Support System (CDSS)** = expert system that provides support to certain reasoning tasks, in the context of a clinical decision;
- **Collective Intelligence** = shared group (symbolic) intelligence, emerging from cooperation/competition of many individuals, e.g. for consensus decision making;
- **Crowdsourcing** = a combination of "crowd" and "outsourcing" coined by Jeff Howe (2006), and describes a distributed problem-solving model; example for crowdsourcing is a public software beta-test;
- **Decision Making** = central cognitive process in every medical activity, resulting in the selection of a final choice of action out of several alternatives;
- **Decision Support System (DSS)** = is an IS including knowledge based systems to interactively support decision-making activities, i.e. making data useful;

Advance Organizer (2/2)

- **DXplain** = a DSS from the Harvard Medical School, to assist making a diagnosis (clinical consultation), and also as an instructional instrument (education); provides a description of diseases, etiology, pathology, prognosis and up to 10 references for each disease;
- Etiology = in medicine (many) factors coming together to cause an illness (see causality)
- Explainable AI = Explainability = upcoming fundamental topic within recent AI; answering e.g. **why** a decision has been made
- **Expert-System** = emulates the decision making processes of a human expert to solve complex problems;
- **GAMUTS** in Radiology = Computer-Supported list of common/uncommon differential diagnoses;
- **ILIAD** = medical expert system, developed by the University of Utah, used as a teaching and testing tool for medical students in problem solving. Fields include Pediatrics, Internal Medicine, Oncology, Infectious Diseases, Gynecology, Pulmonology etc.
- Interpretability = there is no formal technical definition yet, but it is considered as a prerequisite for trust
- **MYCIN** = one of the early medical expert systems (Shortliffe (1970), Stanford) to identify bacteria causing severe infections, such as bacteremia and meningitis, and to recommend antibiotics, with the dosage adjusted for patient's body weight;
- **Reasoning** = cognitive (thought) processes involved in making medical decisions (clinical reasoning, medical problem solving, diagnostic reasoning;
- **Transparency** = opposite of opacity of black-box approaches, and connotes the ability to understand how a model works (that does not mean that it should always be understood, but that – in the case of necessity – it can be re-enacted

Learning Goals: At the end of this lecture you …

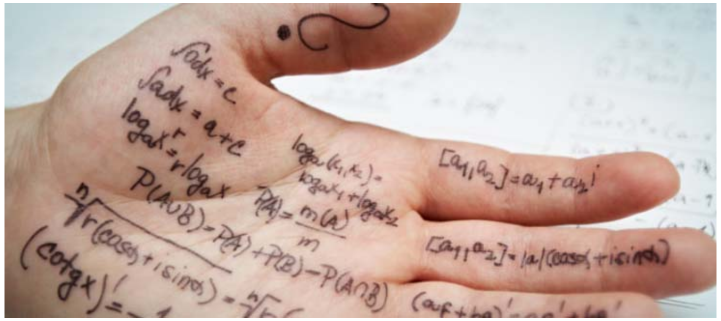
- … can apply your knowledge gained in the previous lectures to example systems of decision support;
- … have an overview about the core principles and architecture of decision support systems;
- … are familiar with the certainty factors as e.g. used in MYCIN;
- … are aware of some design principles of DSS;
- … have seen similarities between DSS and KDD on the example of computational methods in cancer detection;
- … have seen basics of CBR systems;

Agenda

- **00 Reflection – follow-up from last lecture**
- **01 Decision Support Systems (DSS)**
- **02 Computers help making better decisions?**
- **03 History of DSS = History of AI**
- **04 Example: Towards Personalized Medicine**
- **05 Example: Case Based Reasoning (CBR)**
- **06 Towards Explainable AI**

Reflection from last lecture

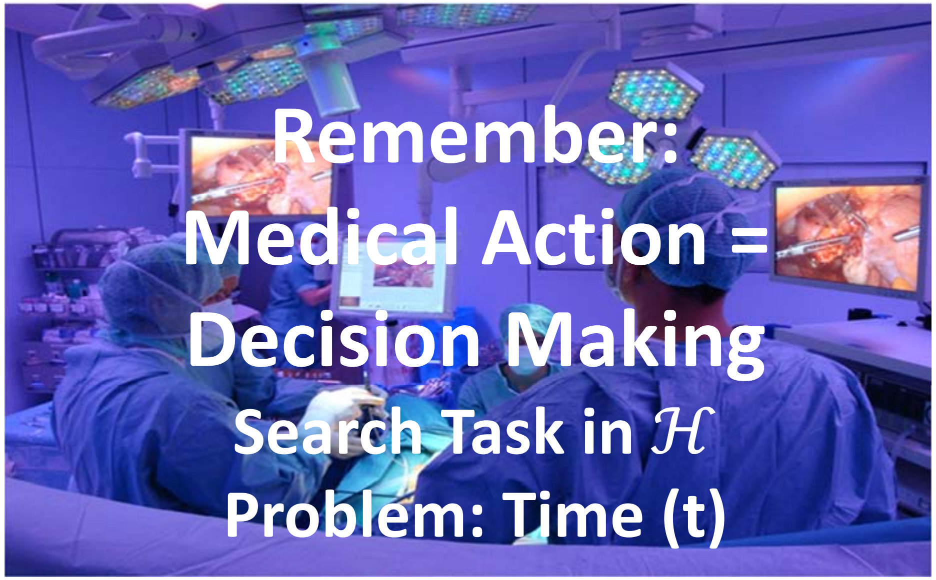
- The Quiz-Slide will be shown during the course



Key Challenges

- Remember: Medicine is an complex application domain – dealing most of the time with **probable information!**
- Some challenges include:
- (a) defining hospital system architectures in terms of generic tasks such as diagnosis, therapy planning and monitoring to be executed for (b) medical reasoning in (a);
- (c) patient information management with (d) minimum uncertainty.
- Other challenges include: (e) knowledge acquisition and encoding, (f) human-computer interface and interaction; and (g) system integration into existing clinical legacy and proprietary environments, e.g. the enterprise hospital information system; to mention only a few.

01 Decision Support Systems



Search in an arbitrarily high-dimensional space < 5 min.!

Decision Making is central in any (medical) work

Source: Cisco (2008). Cisco Health Presence Trial at Aberdeen Royal Infirmary in Scotland

The Medical Domain and Decision Making

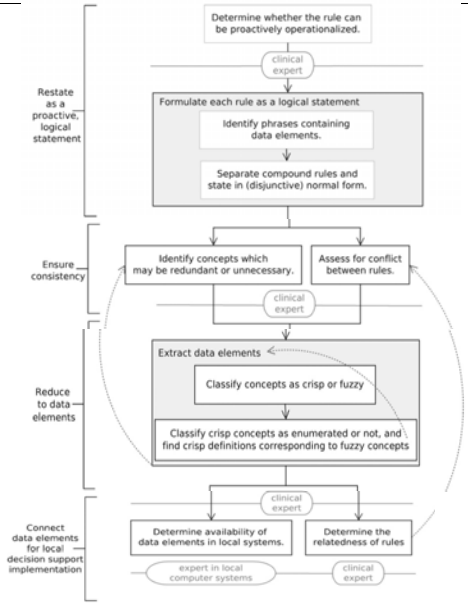
- 400 BC Hippocrates (460-370 BC), father of western medicine:
  - A medical record should accurately reflect the course of a disease
  - A medical record should indicate the probable cause of a disease
- **1890** William Osler (1849-1919), father of modern western medicine
  - **Medicine is a science of uncertainty and an art of probabilistic decision making**
- Today
  - Prediction models are based on data features, patient health status is modelled as high-dimensional feature vectors …

Digression: Clinical Guidelines as DSS & Quality Measure

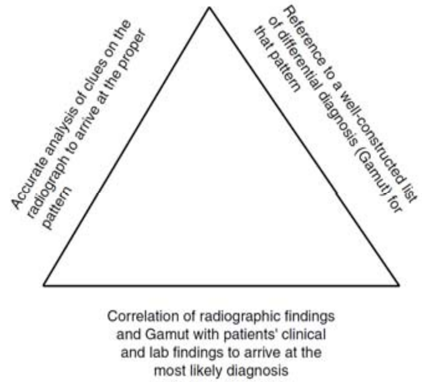
- Clinical guidelines are **systematically** developed documents to assist doctors and patient decisions about appropriate care;
- In order to build DS, based on a guideline, it is **formalized** (transformed from natural language to a logical algorithm), and
- **implemented** (using the algorithm to program a DSS);
- To increase the quality of care, they must be linked to a process of care, for example:
  - "80% of diabetic patients should have an HbA1c below 7.0" could be linked to processes such as:
  - "All diabetic patients should have an annual HbA1c test" and
  - "Patients with values over 7.0 should be rechecked within 2 months."
- **Condition-action rules** specify one or a few conditions which are linked to a specific action, in contrast to narrative guidelines which describe a series of branching or iterative decisions unfolding over time.
- Narrative guidelines and clinical rules are two ends of a continuum of clinical care standards.

Example: Clinical Guidelines

Medlock, S., Opondo, D., Eslami, S., Askari, M., Wierenga, P., de Rooij, S. E. & Abu-Hanna, A. (2011) LERM (Logical Elements Rule Method): A method for assessing and formalizing clinical rules for decision support. *International Journal of Medical Informatics, 80, 4, 286-295.*



Example: Triangulation to find diagnoses



Reeder, M. M. & Felson, B. 2003. *Reeder and Felson's gamuts in radiology: comprehensive lists of roentgen differential diagnosis, New York, Springer Verlag.*

Gamut F-137

PHRENIC NERVE PARALYSIS OR DYSFUNCTION

COMMON

1. Iatrogenic (eg, surgical injury; chest tube; therapeutic avulsion or injection; subclavian vein puncture)
2. Infection (eg, tuberculosis; fungus disease; abscess)
3. Neoplastic invasion or compression (esp. carcinoma of lung)

UNCOMMON

1. Aneurysm$_g$, aortic or other
2. Birth trauma (Erb's palsy)
3. Herpes zoster
4. Neuritis, peripheral (eg, diabetic neuropathy)
5. Neurologic disease$_g$ (eg, hemiplegia; encephalitis; polio; Guillain-Barré S.)
6. Pneumonia
7. Trauma

Reference

1. Prasad S, Athreya BH: Transient paralysis of the phrenic nerve associated with head injury. JAMA 1976;236:2532–2533

Example - Gamuts in Radiology

REEDER AND FELSON'S
GAMUTS IN RADIOLOGY

GAMUT G-25
EROSIVE GASTRITIS*

COMMON

1. Acute gastritis (eg, alcohol abuse)
2. Crohn's disease
3. Drugs (eg, aspirin ; NSAID ; steroids)
4. *Helicobacter pylori* infection
5. Idiopathic
6. [Normal areae gastricae]
7. Peptic ulcer; hyperacidity

UNCOMMON

1. Corrosive gastritis
2. *Cryptosporidium* antritis
3. [Lymphoma]
4. Opportunistic infection (eg, candidiasis {moniliasis} ; herpes simplex; cytomegalovirus)
5. Postoperative gastritis
6. Radiation therapy
7. Zollinger-Ellison S. ; multiple endocrine neoplasia (MEN) S.

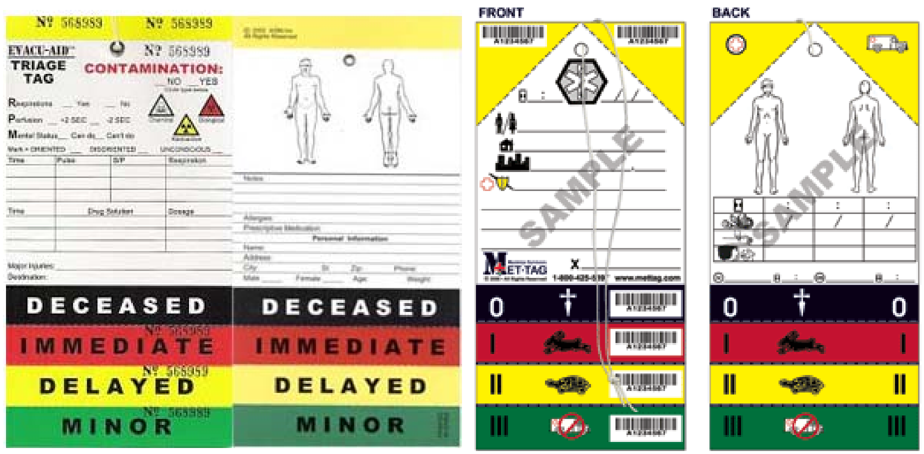
* Superficial erosions or aphthoid ulcerations seen especially with double contrast technique.

[ ] This condition does not actually cause the gamuted imaging finding, but can produce imaging changes that simulate it.

Reeder, M. M. & Felson, B. (2003) *Reeder and Felson's gamuts in radiology: comprehensive lists of roentgen differential diagnosis. New York, Springer*

http://rfs.acr.org/gamuts/data/G-25.htm

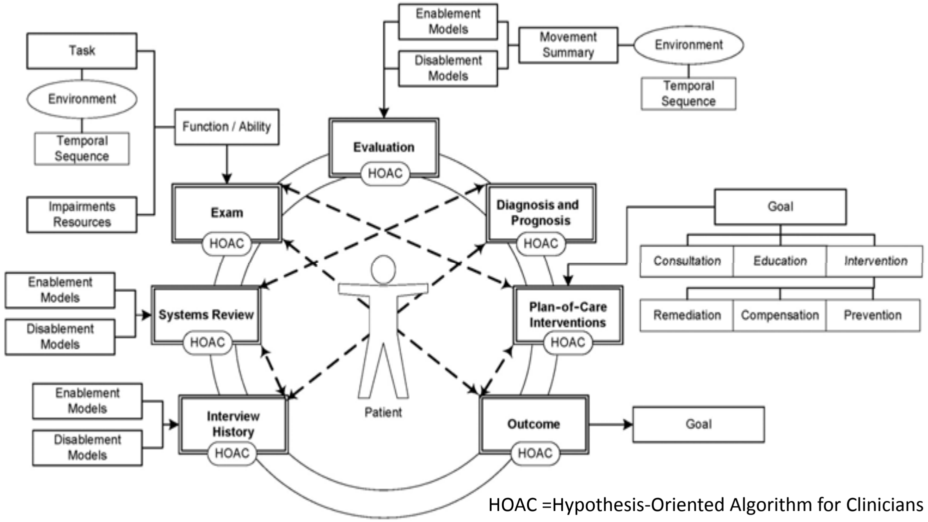
Example: Triage Tags - International Triage Tags



Iserson, K. V. & Moskop, J. C. 2007. Triage in Medicine, Part I: Concept, History, and Types. Annals of Emergency Medicine, 49, (3), 275-281.

Image Source: http://store.gomed-tech.com

Example Clinical DSS: Hypothesis-Oriented Algorithm



HOAC =Hypothesis-Oriented Algorithm for Clinicians

Schenkman, M., Deutsch, J. E. & Gill-Body, K. M. (2006) An Integrated Framework for Decision Making in Neurologic Physical Therapist Practice. *Physical Therapy, 86, 12, 1681-1702.*

Example Prediction Models > Feature Generation

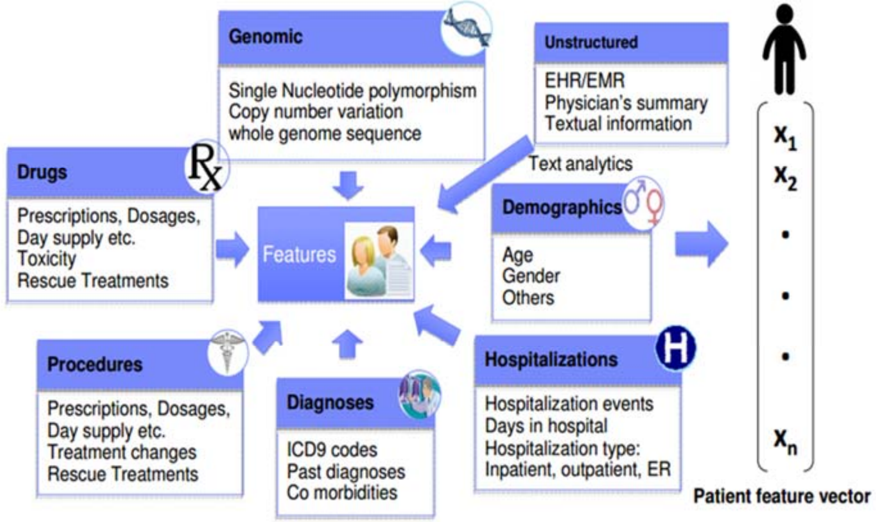
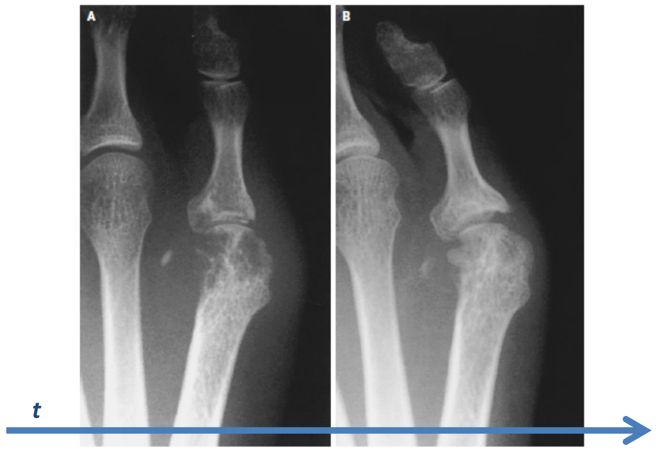


Image credit to Michal Rosen-Zvi

Example: Rheumatology

Chao, J., Parker, B. A. & Zvaifler, N. J. (2009) Accelerated Cutaneous Nodulosis Associated with Aromatase Inhibitor Therapy in a Patient with Rheumatoid Arthritis. *The Journal of Rheumatology, 36,* ***5, 1087-1088.***

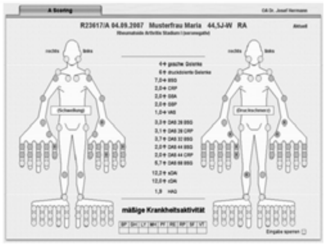
Bone Changes …



Ikari, K. & Momohara, S. (2005) Bone Changes in Rheumatoid Arthritis. *New England Journal of Medicine, 353,* ***15, e13.***

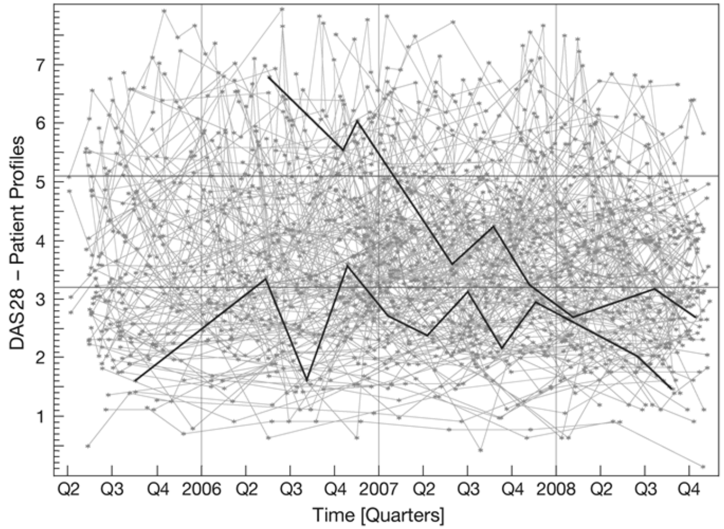
100+ clinical and functional parameter per Patient

- 50+ Patients per day ~ 5000 data points per day …
- Aggregated with specific scores (Disease Activity Score, DAS)
- Current patient status is related to previous data
- = convolution over time
- ⇒ **time-series data**



Simonic, K. M., Holzinger, A., Bloice, M. & Hermann, J. (2011). *Optimizing Long-Term Treatment of Rheumatoid Arthritis with Systematic Documentation. Pervasive Health - 5th International Conference on Pervasive Computing Technologies for Healthcare, Dublin, IEEE, 550-554.*

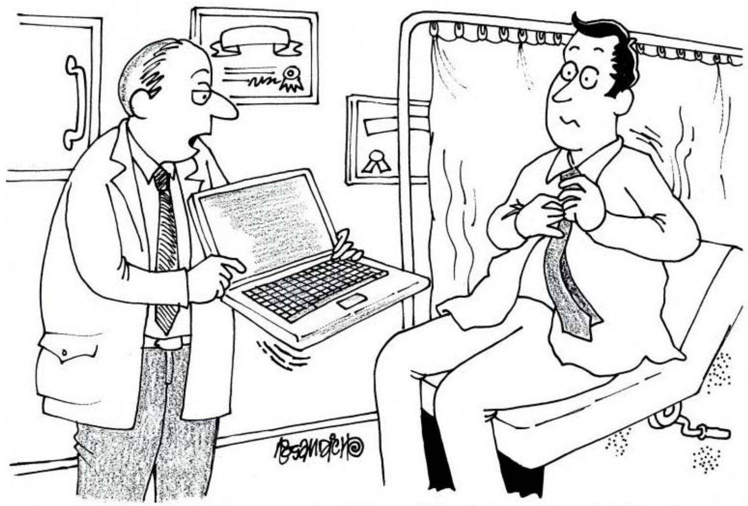
Gaining out Knowledge of time-series data



Simonic, K. M., Holzinger, A., Bloice, M. & Hermann, J. (2011). *Optimizing Long-Term Treatment of Rheumatoid Arthritis with Systematic Documentation. Pervasive Health - 5th International Conference on Pervasive Computing Technologies for Healthcare, Dublin, IEEE, 550-554.*

02 Can Computers help doctors to make better decisions?

Computers to help human doctors to make better decisions



"If you want a second opinion, I'll ask my computer."

http://biomedicalcomputationreview.org/content/clinical-decision-support-providing-quality-healthcare-help-computer

Augmenting Human Capabilities …

Slide 8-2 Two types of decisions (Diagnosis vs. Therapy)

- **Type 1 Decisions:** related to the **diagnosis**, i.e. computers are used to assist in diagnosing a disease on the basis of the individual patient data. Questions include:
  - What is the probability that this patient has a myocardial infarction on the basis of given data (patient history, ECG, …)?
  - What is the probability that this patient has acute appendices, given the signs and symptoms concerning abdominal pain?
- **Type 2 Decisions:** related to **therapy**, i.e. computers are used to select the best therapy on the basis of clinical evidence, e.g.:
  - What is the best therapy for patients of age x and risks y, if an obstruction of more than z % is seen in the left coronary artery?
  - What amount of insulin should be prescribed for a patient during the next 5 days, given the blood sugar levels and the amount of insulin taken during the recent weeks?

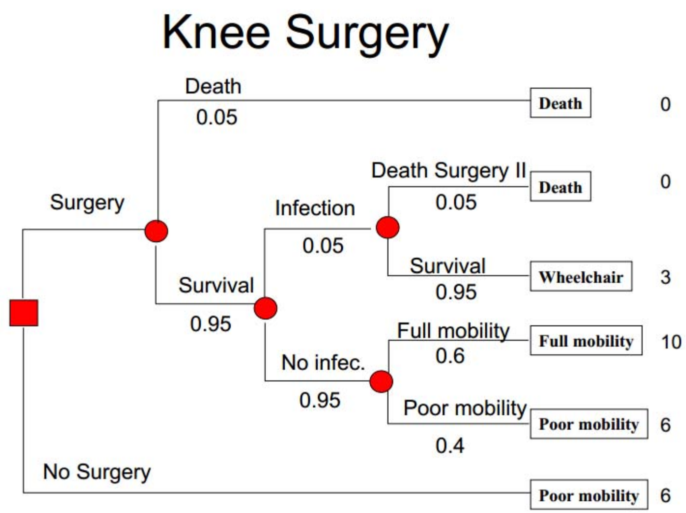
Bemmel, J. H. V. & Musen, M. A. 1997. *Handbook of Medical Informatics, Heidelberg, Springer.*

Example: Knee Surgery of a Soccer Player

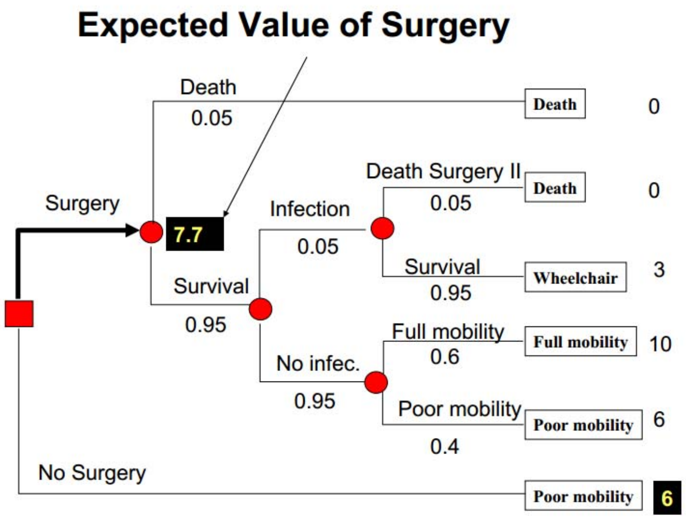
- Example of a Decision Problem
- Soccer player considering knee surgery
- Uncertainties:
- Success: recovering full mobility
- Risks: infection in surgery (if so, needs another surgery and may loose more mobility)
- Survival chances of surgery

Harvard-MIT Division of Health Sciences and Technology
HST.951J: Medical Decision Support, Fall 2005
Instructors: Professor Lucila Ohno-Machado and Professor Staal Vinterbo

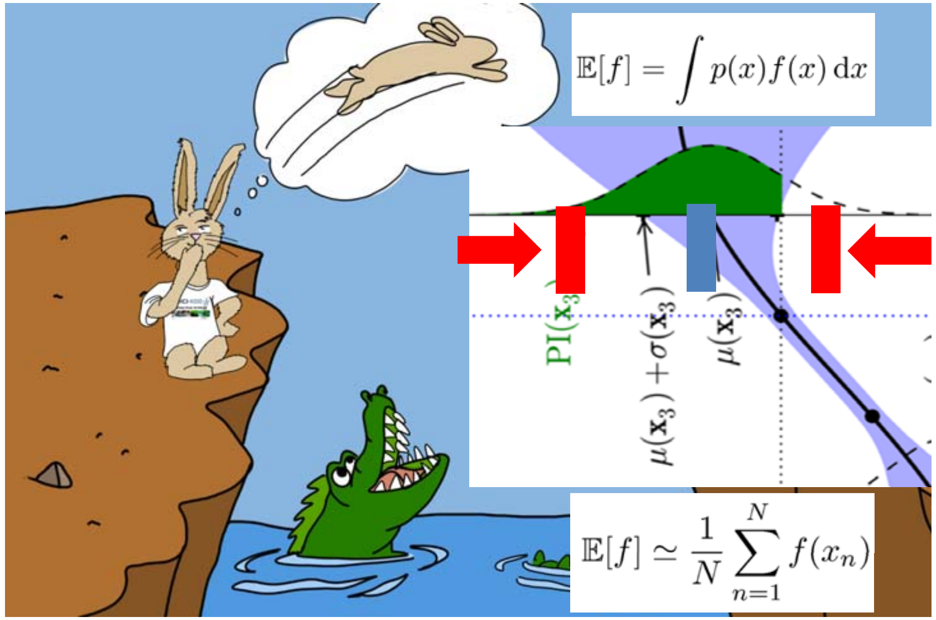
Decision Tree (this is known since Hippocrates!)



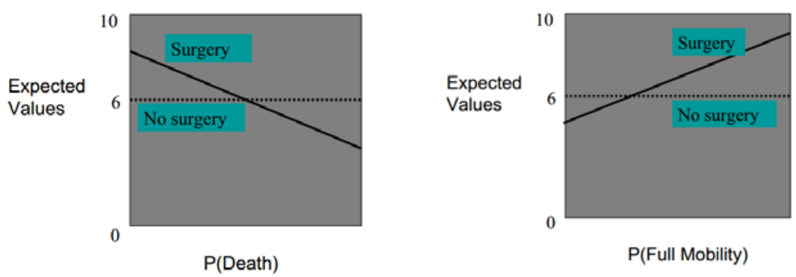
Helps to make rational decisions (risks vs. success)



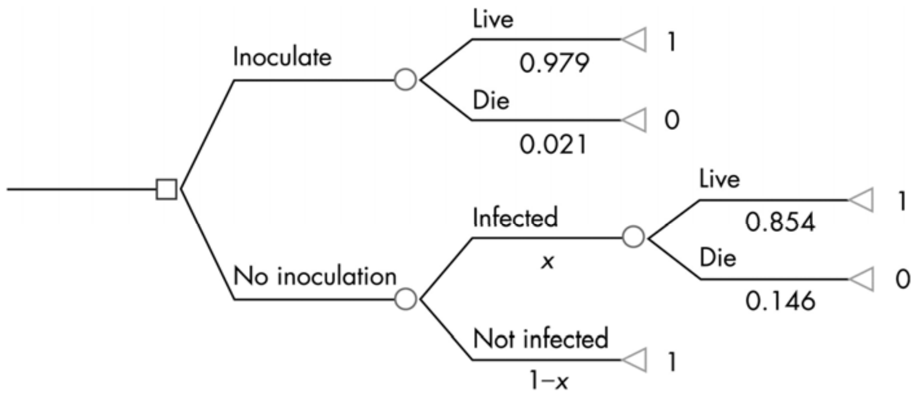
Estimate Confidence Interval: Uncertainty matters !



Effect of probabilities in the decision



Clinical Decision Tree (CDT) is still state-of-the-art



Ferrando, A., Pagano, E., Scaglione, L., Petrinco, M., Gregori, D. & Ciccone, G. (2009) A decision-tree model to estimate the impact on cost-effectiveness of a venous thromboembolism prophylaxis guideline. *Quality and Safety in Health Care, 18, 4, 309-313.*

Taxonomy of Decision Support Models

Decision Model

- Quantitative (statistical)
  - supervised
  - Bayesian
  - unsupervised
  - Fuzzy sets
  - Neural network
  - Logistic
- Qualitative (heuristic)
  - Truth tables
  - Decision trees
  - Reasoning models
  - Boolean Logic
  - Non-parametric Partitioning
  - Expert systems
  - Critiquing systems

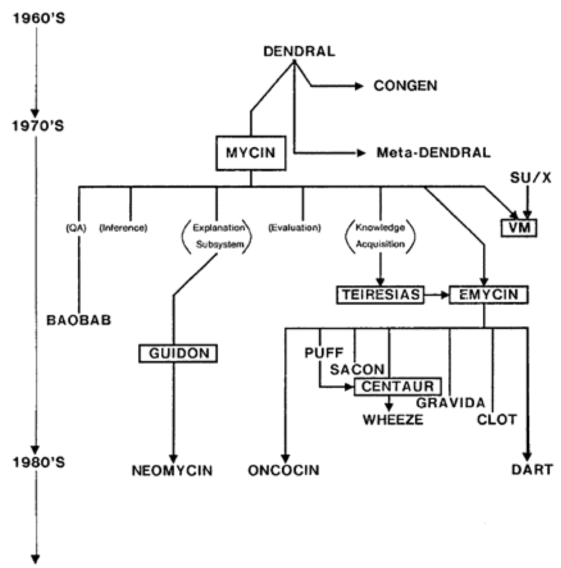
Extended by A. Holzinger after: Bemmel, J. H. v. & Musen, M. A. (1997) *Handbook of Medical Informatics. Heidelberg, Springer.*

03 History of DSS = History of AI

A ultrashort history of Early AI

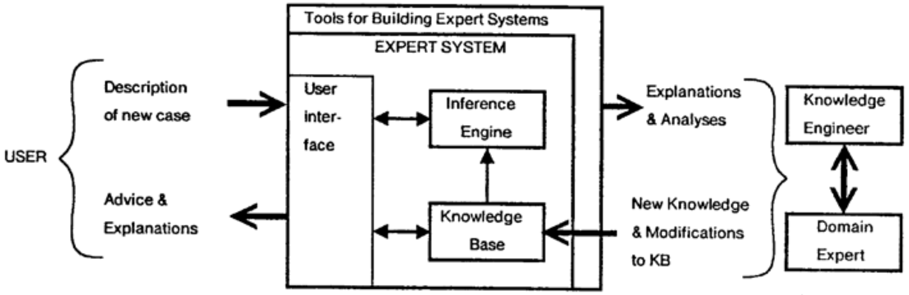
- **1943** McCulloch, W.S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biology, 5, (4), 115-133, doi:10.1007/BF02459570.
- **1950** Turing, A.M. Computing machinery and intelligence. Mind, 59, (236), 433-460.
- **1959** Samuel, A.L. Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3, (3), 210-229, doi:10.1147/rd.33.0210.
- **1975** Shortliffe, E.H. & Buchanan, B.G. 1975. A model of inexact reasoning in medicine. Mathematical biosciences, 23, (3-4), 351-379, doi:10.1016/0025-5564(75)90047-4.

Evolution of Decision Support Systems (Expert Systems)



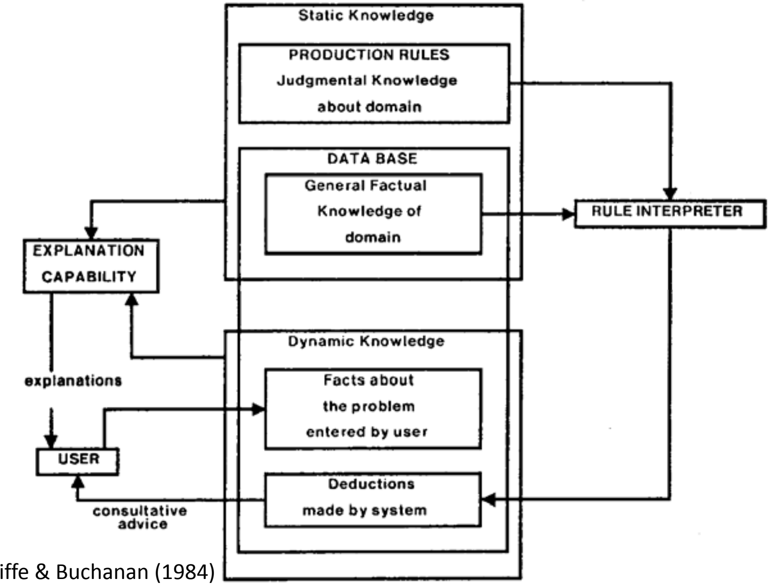
Shortliffe, E. H. & Buchanan, B. G. (1984) *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project. Addison-Wesley.*

Early Knowledge Based System Architecture



Shortliffe, T. & Davis, R. (1975) Some considerations for the implementation of knowledge-based expert systems *ACM SIGART Bulletin, 55, 9-12.*

Static Knowledge versus dynamic knowledge



Shortliffe & Buchanan (1984)

Dealing with uncertainty in the real world

- The information available to humans is often imperfect – imprecise - uncertain.
- This is especially in the medical domain the case.
- An **human agent** can cope with deficiencies.
- Classical logic permits only **exact reasoning**:
- IF A is true THEN A is non-false and
  IF B is false THEN B is non-true
- Most real-world problems do not provide this exact information, mostly it is inexact, incomplete, uncertain and/or **un-measurable!**

1967, Star Trek, I Mudd

Harcourt Fenton Mudd: Now listen, Spock, you may be a wonderful science officer but, believe me, you couldn't sell fake patents to your mother!

Spock: I fail to understand why I should care to induce my mother to purchase falsified patents.

MYCIN – rule based system - certainty factors

- MYCIN is a rule-based Expert System, which is used for therapy planning for patients with bacterial infections
- Goal oriented strategy ("Rückwärtsverkettung")
- To every rule and every entry a certainty factor (CF) is assigned, which is between 0 und 1
- Two measures are derived:
- MB: measure of belief
- MD: measure of disbelief
- Certainty factor – CF of an element is calculated by:
  $$CF[h] = MB[h] - MD[h]$$
- CF is positive, if more evidence is given for a hypothesis, otherwise CF is negative
- CF[h] = +1 -> h is 100 % true
- CF[h] = –1 -> h is 100% false

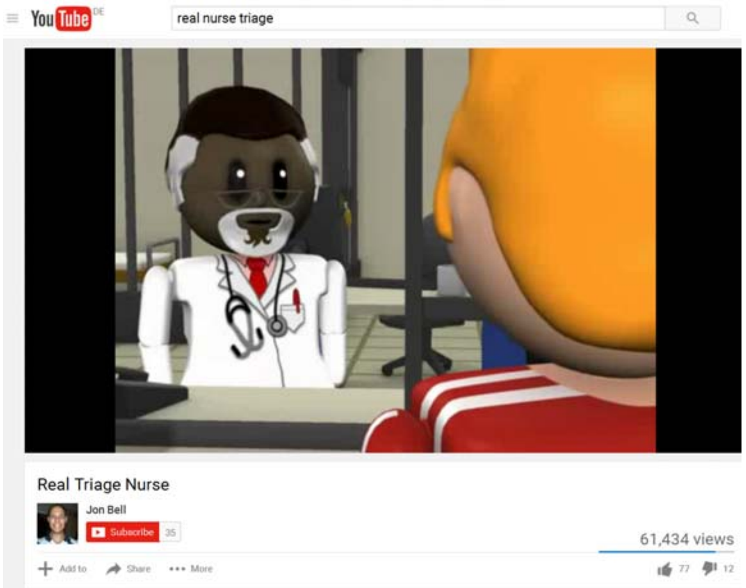
Original Example from MYCIN

h_1 = The identity of ORGANISM-1 is streptococcus
h_2 = PATIENT-1 is febrile
h_3 = The name of PATIENT-1 is John Jones

$CF[h_1,E] = .8$	:	There is strongly suggestive evidence (.8) that the identity of ORGANISM-1 is streptococcus
$CF[h_2,E] = -.3$	:	There is weakly suggestive evidence (.3) that PATIENT-1 is not febrile
$CF[h_3,E] = +1$	:	It is definite (1) that the name of PATIENT-1 is John Jones

Shortliffe, E. H. & Buchanan, B. G. (1984) *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project. Addison-Wesley.*

MYCIN was *no* success in the clinical practice



However, AI was extremely popular in the 1970ies


SIBYLLE

Für den Winter:
Kostüme
und Mäntel
Sportliche Pelze
Anoraks
SIBYLLE-Modelle:
Pullover
und Tweedröcke

4.

Was die moderne Frau von der Kybernetik wissen muß

Die Geheimnisse des Rechenautomaten

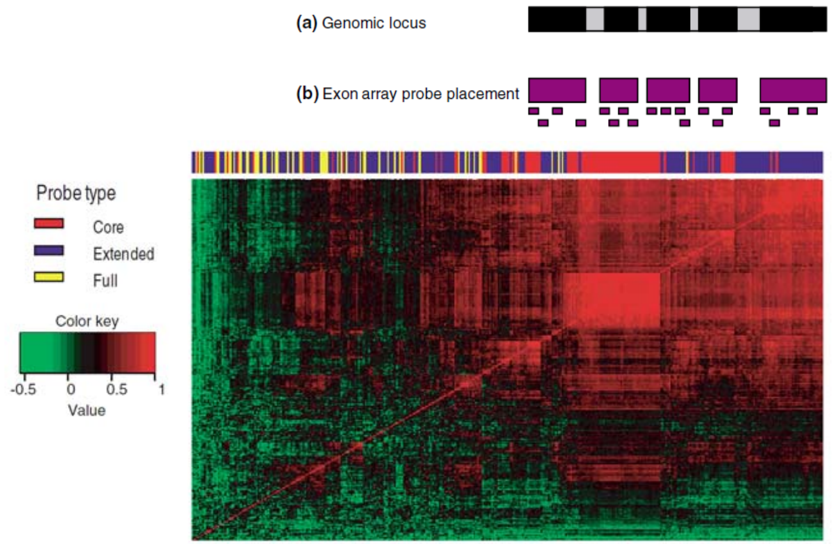
Image credit to Bernhard Schölkopf

Cybernetics was praised as the solution for everything

Image credit to Bernhard Schölkopf

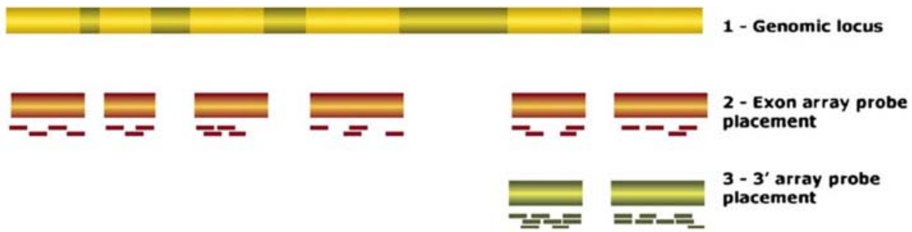
04 Towards P4-Medicine

Slide 8-22 Example: Exon Arrays



Kapur, K., Xing, Y., Ouyang, Z. & Wong, W. (2007) Exon arrays provide accurate assessments of gene expression. *Genome Biology, 8, 5, R82.*

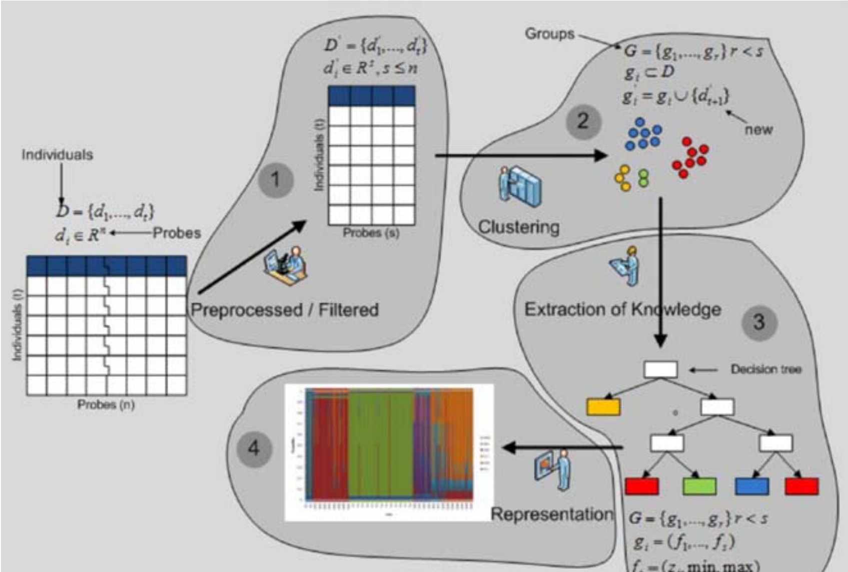
Slide 8-23 Computational leukemia cancer detection 1/6



Exon array structure. Probe design of exon arrays. (1) Exon—intron structure of a gene. Gray boxes represent introns, rest represent exons. Introns are not drawn to scale. (2) Probe design of exon arrays. Four probes target each putative exon. (3) Probe design of 30expression arrays. Probe target the 30end of mRNA sequence.

Corchado, J. M., De Paz, J. F., Rodriguez, S. & Bajo, J. (2009) Model of experts for decision support in the diagnosis of leukemia patients. *Artificial Intelligence in Medicine, 46, 3, 179-200.*

Slide 8-24 Computational leukemia cancer detection 2/6

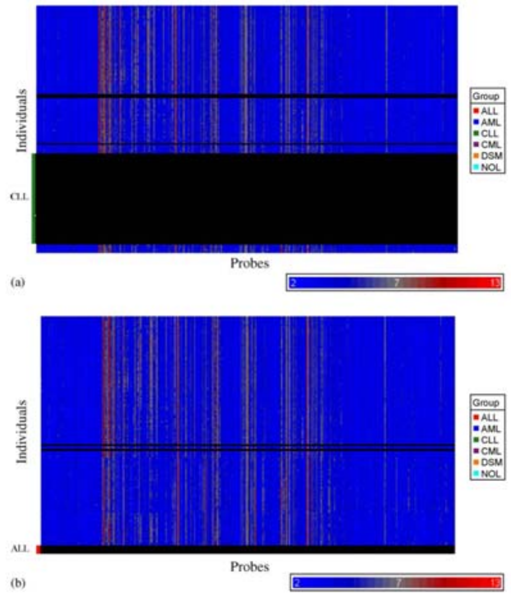


Corchado et al. (2009)

Slide 8-25 Computational leukemia cancer detection 3/6

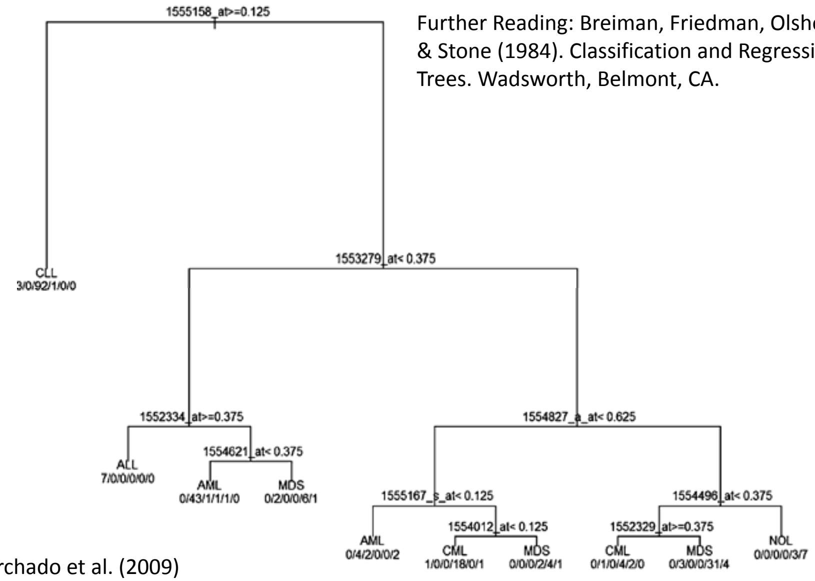
A = acute, C = chronic,
L = lymphocytic, M = myeloid

- **ALL** = cancer of the blood AND bone marrow caused by an abnormal proliferation of lymphocytes.
- **AML** = cancer in the bone marrow characterized by the proliferation of myeloblasts, red blood cells or abnormal platelets.
- **CLL** = cancer characterized by a proliferation of lymphocytes in the bone marrow.
- **CML** = caused by a proliferation of white blood cells in the bone marrow.
- **MDS** (Myelodysplastic Syndromes) = a group of diseases of the blood and bone marrow in which the bone marrow does not produce a sufficient amount of healthy cells.
- **NOL** (Normal) = No leukemias



Corchado et al. (2009)

8-26 Computational leukemia cancer detection 4/6

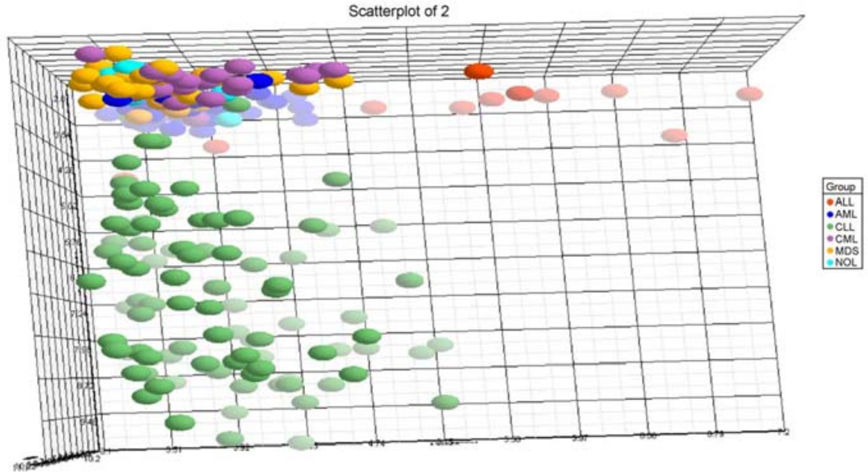


Further Reading: Breiman, Friedman, Olshen, & Stone (1984). Classification and Regression Trees. Wadsworth, Belmont, CA.

Corchado et al. (2009)

8-27 Computational leukemia cancer detection 5/6

Classification CLL—ALL. Representation of the probes of the decision tree which classify the CLL and ALL to 1555158_at, 1553279_at and 1552334_at



Corchado et al. (2009)

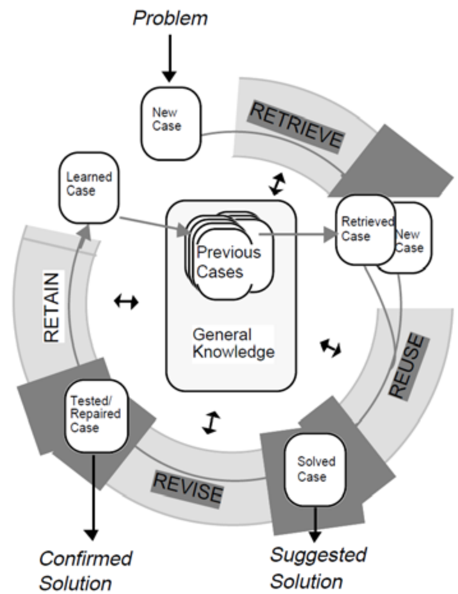
Slide 8-28 Computational leukemia cancer detection 6/6

- The model of Corchado et al. (2009) combines:
- 1) methods to **reduce the dimensionality** of the original data set;
- 2) pre-processing and data filtering techniques;
- 3) a clustering method to classify patients; and
- 4) extraction of knowledge techniques
- The system reflects how human experts work in a lab, but
- 1) **reduces the time** for making predictions;
- 2) **reduces the rate of human error;** and
- 3) **works with high-dimensional data** from exon arrays

05 Example: Case Based Reasoning (CBR)

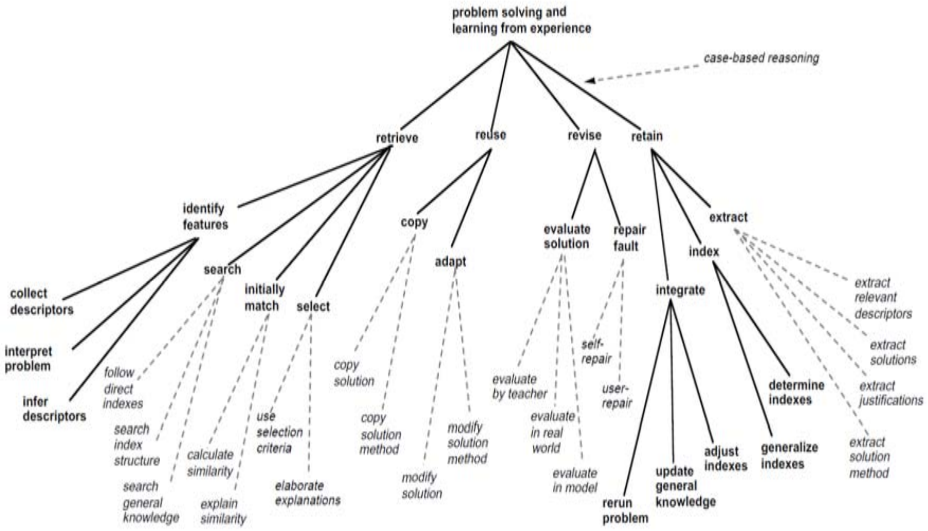
Slide 8-29 Thinking – Reasoning – Deciding – Acting

Slide 8-30 Case Based Reasoning (CBR) Basic principle



Aamodt, A. & Plaza, E. (1994) Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications, 7, 1, 39-59.*

Slide 8-31 The task-method decomposition of CBR

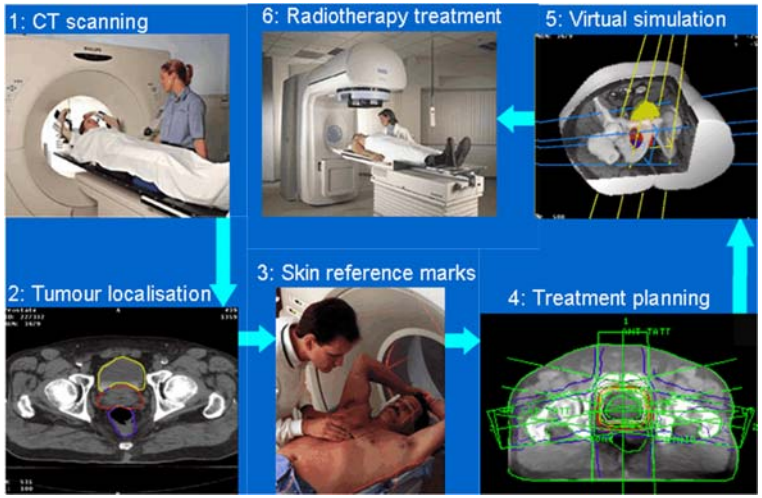


Aamodt & Plaza (1994)

Slide 8-32 CBR Example: Radiotherapy Planning 1/6

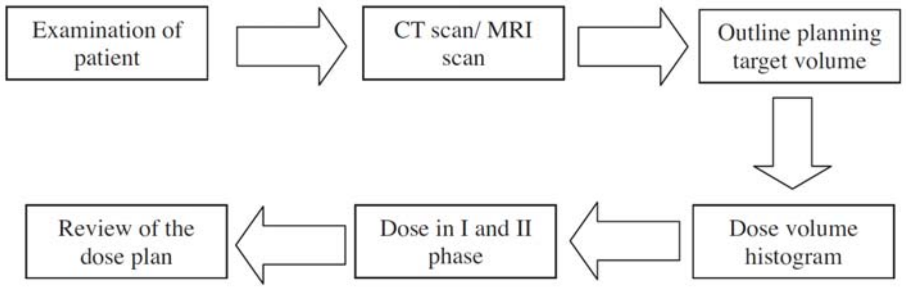
Source: http://www.teachingmedicalphysics.org.uk

Slide 8-33 CBR Example: Radiotherapy Planning 2/6



Source: Imaging Performance Assessment of CT Scanners Group, http://www.impactscan.org

Slide 8-34 CBR Example: Radiotherapy Planning 3/6

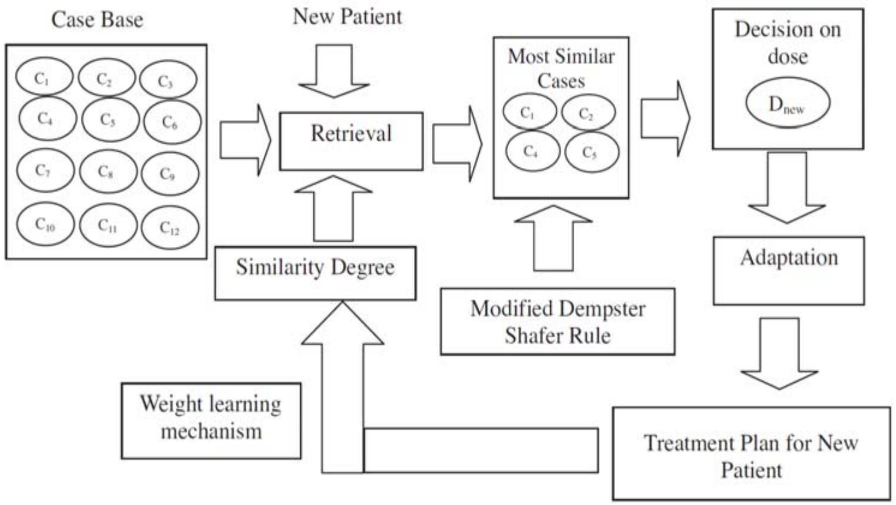


Measures:

1) Clinical Stage = a labelling system
2) Gleason Score = grade of prostate cancer = integer between 1 to 10; and
3) Prostate Specific Antigen (PSA) value between 1 to 40
4) Dose Volume Histogram (DVH) = pot. risk to the rectum (66, 50, 25, 10 %)

Petrovic, S., Mishra, N. & Sundar, S. (2011) A novel case based reasoning approach to radiotherapy planning. *Expert Systems With Applications, 38, 9, 10759-10769.*

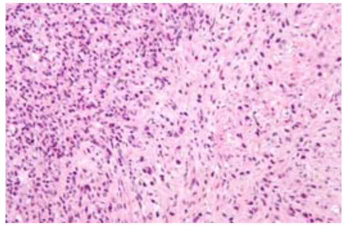
Slide 8-35 CBR System Architecture 4/6

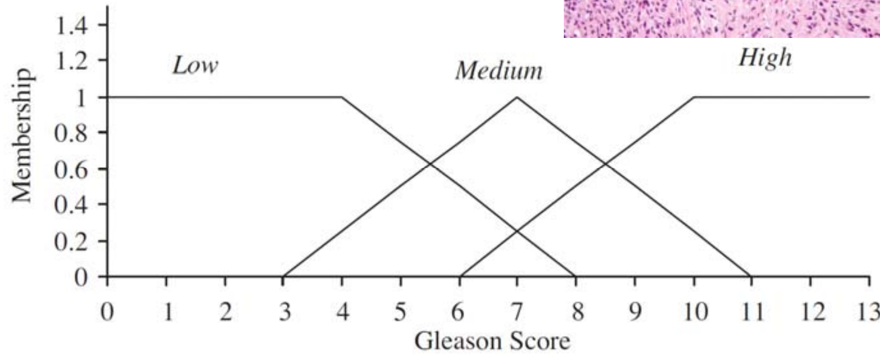


Petrovic, S., Mishra, N. & Sundar, S. (2011) A novel case based reasoning approach to radiotherapy planning. *Expert Systems With Applications, 38, 9, 10759-10769.*

Slide 8-36 Membership funct. of fuzzy sets Gleason score 5/6

Gleason score evaluates the grade of prostate cancer. Values: integer within the range

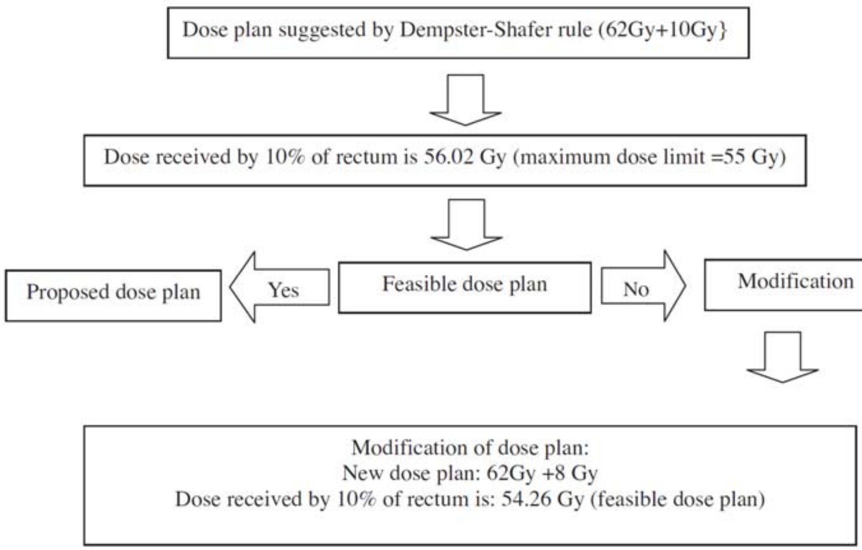




Petrovic, S., Mishra, N. & Sundar, S. (2011) A novel case based reasoning approach to radiotherapy planning. *Expert Systems With Applications, 38, 9, 10759-10769.*

Slide 8-37 Case Based Reasoning 6/6

Petrovic et al. (2011)



06 Towards Explainable AI

Mastering the game of Go without human knowledge

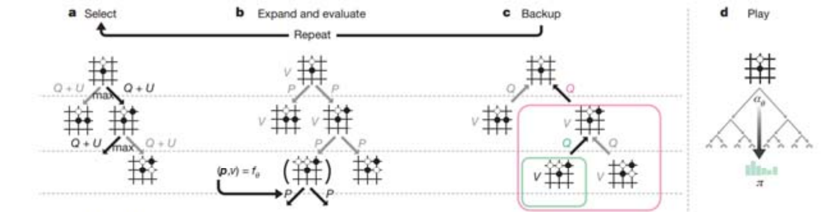
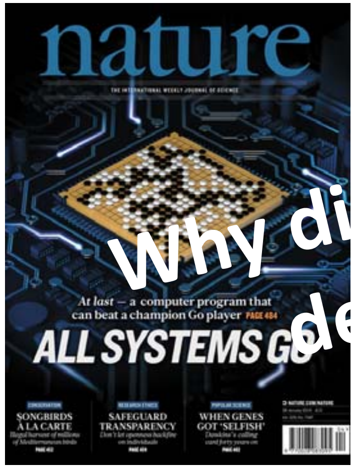


Figure 2 | MCTS in AlphaGo Zero. a, Each simulation traverses the tree by selecting the edge with maximum action value Q, plus an upper confidence bound U that depends on a stored prior probability P and visit count N for that edge (which is incremented once traversed). **b,** The leaf node is expanded and the associated position s is evaluated by the neural network $(P(s,\cdot),V(s)) = f_\theta(s)$; the vector of P values are stored in the outgoing edges from s. **c,** Action value Q is updated to track the mean of all evaluations V in the subtree below that action. **d,** Once the search is complete, search probabilities π are returned, proportional to $N^{1/\tau}$, where N is the visit count of each move from the root state and τ is a parameter controlling temperature.

19 OCTOBER 2017 | VOL 550 | NATURE | 355

$$(\boldsymbol{p}, v) = f_\theta(s) \quad \text{and} \quad l = (z - v)^2 - \boldsymbol{\pi}^{\mathrm{T}} \log \boldsymbol{p} + c\|\theta\|^2$$

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George Van Den Driessche, Thore Graepel & Demis Hassabis 2017. Mastering the game of go without human knowledge. Nature, 550, (7676), 354-359, doi:doi:10.1038/nature24270.



Why did it make this decision???

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel & Demis Hassabis 2016. Mastering the game of Go with deep neural networks and tree search. Nature, 529, (7587), 484-489, doi:10.1038/nature16961.

Deep Learning Context recognition state-of-the-art

a woman riding a horse on a dirt road

an airplane is parked on the tarmac at an airport

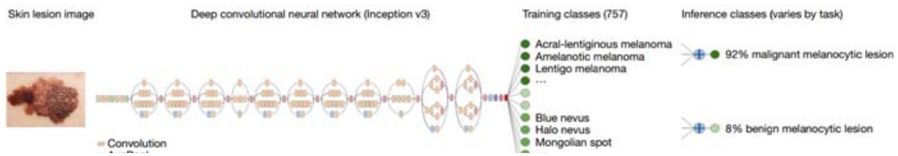
a group of people standing on top of a beach

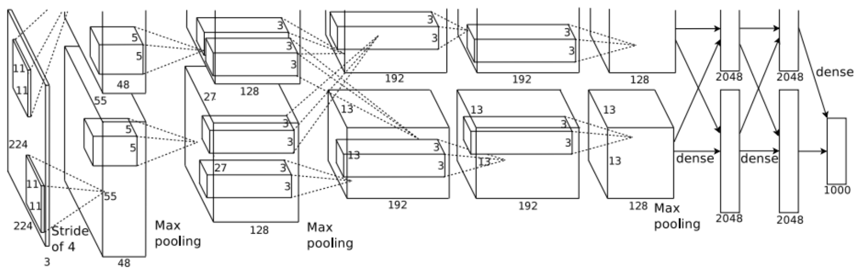
Andrej Karpathy & Li Fei-Fei. Deep visual-semantic alignments for generating image descriptions. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. 3128-3137.

Image Captions by deep learning : github.com/karpathy/neuraltalk2

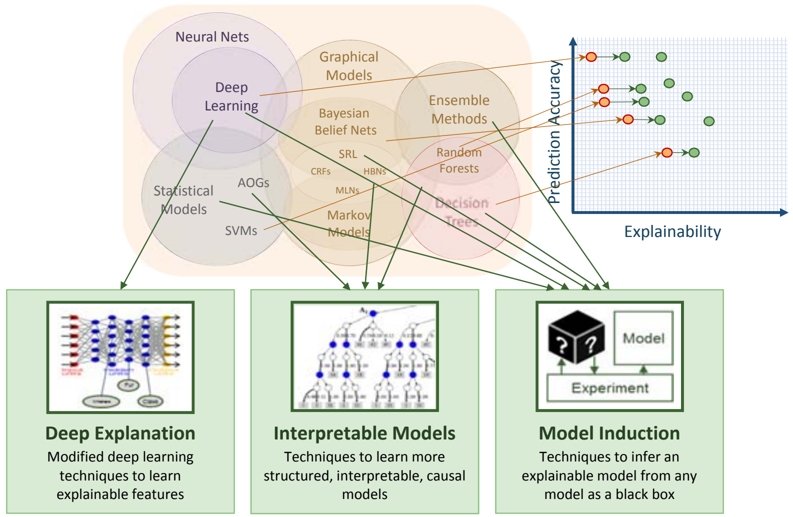
Image Source: Gabriel Villena Fernandez; Agence France-Press, Dave Martin (left to right)

Deep Convolutional Neural Network Pipeline





Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In: Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q., eds. Advances in neural information processing systems (NIPS 2012), 2012 Lake Tahoe. 1097-1105.

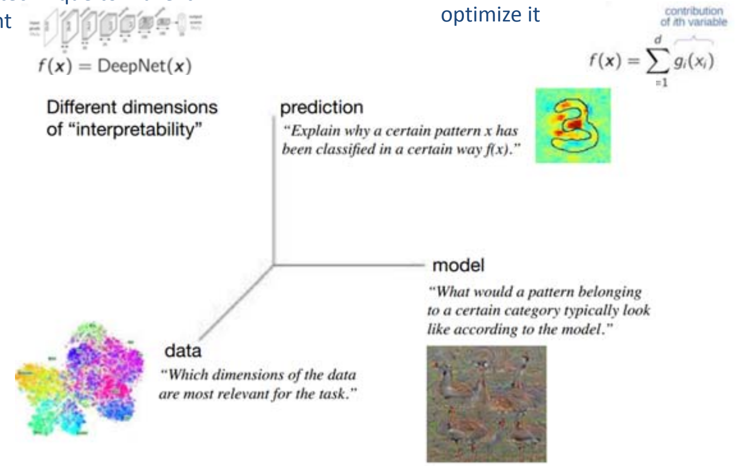


David Gunning 2016. Explainable artificial intelligence (XAI): Technical Report Defense Advanced Research Projects Agency DARPA-BAA-16-53, Arlington, USA, DARPA.

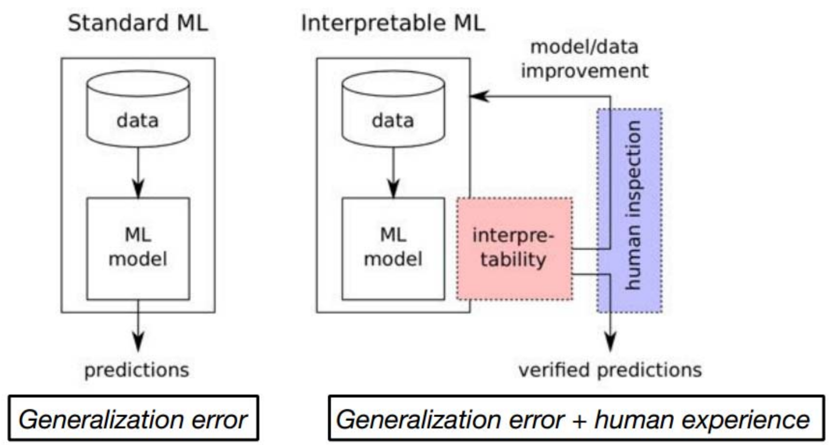
Post-hoc vs. Ante-hoc

Post-hoc: Select a model and develop a technique to make it transparent

Ante-hoc: Select a model that is already transparent and optimize it



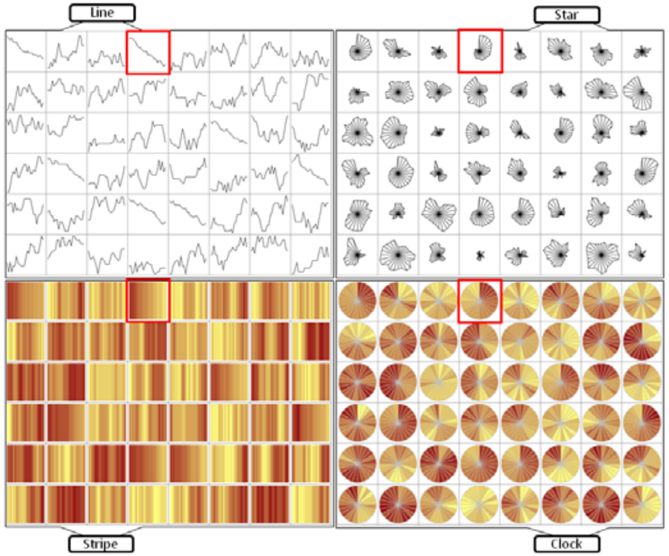
Montavon, G., Samek, W. & Müller, K.-R. 2017. Methods for interpreting and understanding deep neural networks. arXiv:1706.07979.



What is interpretable for humans?

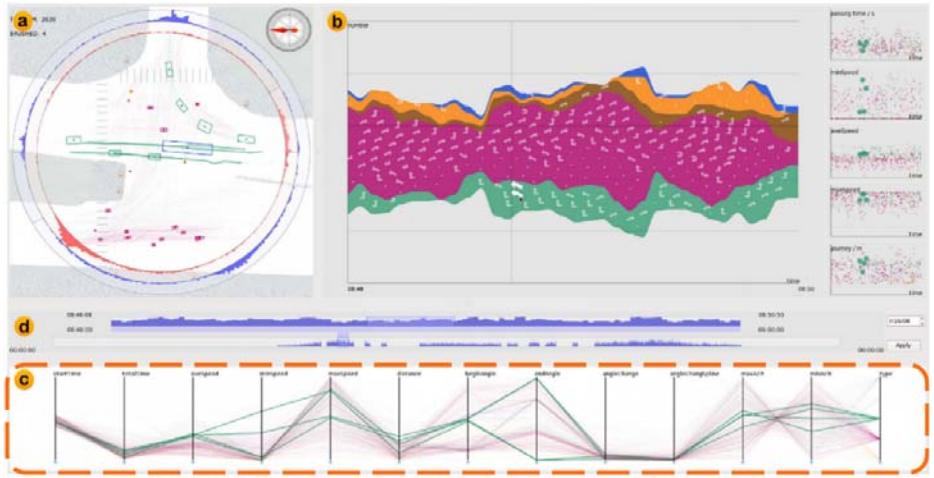
1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.953	0.894	0.620	0.699	0.629	0.546	0.540	1.000	0.526	1.000	0.522	0.483	0.471	1.000	0.522	0.576	0.658
1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.722	0.638	1.000	0.785	0.743	0.792	0.801	0.875	0.712	1.000	0.444	0.947	0.431	1.000	0.793	1.000	0.635
1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.658	0.633	0.569	0.561	0.589	0.640	0.659	0.845	0.932	0.512	0.575	0.941	1.000	0.991	1.000	0.892
1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.932	0.639	0.575	0.544	0.501	0.489	0.470	0.454	0.576	0.576	0.581	0.707	0.992	1.000	1.000	1.000	1.000	1.000
1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.711	0.644	0.569	0.541	0.461	0.430	0.425	0.381	0.364	0.437	0.562	0.509	0.528	0.678	1.000	0.991	1.000	1.000	1.000
1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.680	0.594	0.579	0.513	0.490	0.429	0.405	0.425	0.381	0.401	0.387	0.367	0.484	0.428	0.483	0.659	0.936	1.000	1.000	1.000
1.000	1.000	1.000	1.000	1.000	0.761	0.677	0.610	0.565	0.511	0.498	0.457	0.416	0.396	0.388	0.369	0.355	0.359	0.468	0.392	0.380	0.487	[illegible]	[illegible]	0.744	1.000	0.485
1.000	1.000	1.000	1.000	0.861	0.640	0.579	0.560	0.542	0.476	0.470	0.441	0.405	0.389	0.392	0.396	0.436	0.355	0.327	0.394	0.407	[illegible]	[illegible]	[illegible]	0.766	0.676	0.437
1.000	1.000	1.000	0.827	0.646	0.579	0.556	0.545	0.489	0.505	0.489	0.478	0.411	0.387	0.404	0.401	0.391	0.452	0.352	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.510	0.578	0.538
0.909	1.000	0.860	0.675	0.598	0.528	0.535	0.500	0.497	0.517	0.468	0.520	0.623	0.619	0.507	0.472	0.385	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.354	0.318	0.462	0.491	0.426
1.000	0.989	0.693	0.561	0.546	0.523	0.532	0.452	0.441	0.461	0.649	0.659	0.695	0.686	0.632	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.358	0.295	0.310	0.336	0.363	0.418	0.458
0.969	0.849	0.606	0.530	0.521	0.494	0.437	0.396	0.421	0.626	0.698	0.741	0.737	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.565	0.506	0.435	0.358	0.311	0.299	0.313	0.402	0.488
1.000	1.000	0.590	0.509	0.486	0.445	0.411	0.372	[illegible]	0.675	0.732	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.650	0.622	0.573	0.467	0.405	0.286	0.274	0.358	0.419	0.445
1.000	0.924	0.554	0.517	0.450	0.416	0.449	[illegible]	0.585	[illegible]	0.727	[illegible]	[illegible]	0.785	0.740	0.700	0.653	0.626	[illegible]	0.502	0.431	0.338	0.279	0.295	0.330	0.446	
1.000	1.000	0.557	0.517	[illegible]	[illegible]	[illegible]	[illegible]	0.635	[illegible]	[illegible]	[illegible]	0.751	0.757	0.792	0.764	0.714	0.694	0.642	0.597	[illegible]	0.419	0.341	0.289	0.291	0.326	0.380
1.000	1.000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.623	[illegible]	0.670	0.711	0.748	0.771	0.775	0.772	0.724	[illegible]	[illegible]	0.434	0.378	0.354	0.414	0.307	0.282	0.278	0.402
0.763	1.000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.590	0.646	0.687	0.718	0.724	0.748	0.717	0.659	[illegible]	[illegible]	[illegible]	[illegible]	0.483	0.499	0.472	0.273	0.234	0.279	0.306
1.000	1.000	0.750	[illegible]	[illegible]	[illegible]	0.344	0.328	0.490	0.550	0.623	0.593	[illegible]	0.521	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.519	0.500	0.566	0.521	0.286	0.249	0.234	0.258
0.754	0.830	1.000	[illegible]	0.435	0.326	0.327	0.489	0.474	0.421	0.388	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.594	0.627	0.590	0.613	0.585	0.529	0.438	0.328	0.487	0.200
0.929	0.672	0.503	0.654	0.388	0.335	0.306	0.475	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.616	0.550	0.649	0.686	0.658	0.667	0.587	0.564	0.486	0.416	0.546	0.263
1.000	0.758	0.639	0.726	0.931	0.330	0.299	0.398	[illegible]	[illegible]	[illegible]	[illegible]	0.646	0.644	0.517	0.605	0.517	0.546	0.616	0.714	0.683	0.609	0.578	0.563	0.478	0.314	0.252
1.000	0.790	0.907	0.701	0.897	0.382	0.296	0.358	[illegible]	[illegible]	0.674	0.683	0.666	0.605	0.526	0.620	0.527	0.514	0.616	0.666	0.670	0.628	0.549	0.512	0.262	0.321	0.254
0.760	0.587	0.639	0.557	0.681	0.593	0.397	0.340	0.575	0.574	0.647	0.691	0.666	0.620	0.506	0.614	0.550	0.532	0.487	0.589	0.610	0.616	0.504	0.482	0.310	0.271	0.237
0.577	0.599	0.443	0.561	0.657	0.363	0.914	0.626	0.482	0.553	0.631	0.678	0.722	0.561	0.523	0.639	0.634	0.510	0.481	0.558	0.533	0.597	0.570	0.509	0.342	0.263	0.243
0.639	0.615	0.748	0.639	0.911	0.796	0.647	0.614	0.529	0.553	0.588	0.651	0.644	0.585	0.433	0.606	0.588	0.467	0.313	0.363	0.349	0.415	0.578	0.512	0.305	0.274	0.256
0.569	0.661	0.486	0.605	0.448	0.494	0.705	0.730	0.579	0.532	0.526	0.623	0.518	0.387	0.310	0.338	0.466	0.378	0.559	0.479	0.444	0.430	0.494	0.465	0.232	0.248	0.237
0.493	0.522	0.508	0.553	0.458	0.457	0.435	0.742	0.636	0.434	0.553	0.578	0.369	0.394	0.502	0.539	0.532	0.555	0.601	0.582	0.548	0.498	0.328	0.237	0.242	0.252	0.273
0.891	0.817	0.441	0.445	0.473	0.452	0.720	0.423	0.700	0.492	0.525	0.509	0.463	0.614	0.466	0.477	0.603	0.615	0.509	0.517	0.563	0.405	0.224	0.258	0.234	0.211	0.228
0.543	0.548	0.598	0.433	0.386	0.627	0.482	0.345	0.835	0.751	0.581	0.502	0.482	0.610	0.531	0.524	0.615	0.625	0.562	0.481	0.566	0.306	0.266	0.407	0.366	0.243	0.252
0.762	0.720	0.506	0.496	0.495	0.698	0.396	0.627	0.555	0.317	0.491	0.294	0.382	0.393	0.572	0.449	0.405	0.407	0.357	0.567	0.518	0.243	0.255	0.465	0.415	0.323	0.248
0.472	0.437	0.618	0.547	0.500	0.439	0.580	0.579	0.474	0.406	0.320	0.302	0.233	0.262	0.387	0.622	0.556	0.499	0.580	0.558	0.378	0.214	0.364	0.502	0.413	0.311	0.269
0.461	0.503	0.513	0.432	0.537	0.537	0.467	0.530	0.387	0.504	0.353	0.362	0.456	0.222	0.241	0.342	0.510	0.622	0.454	0.441	0.285	0.218	0.545	0.502	0.445	0.508	0.623
0.529	0.464	0.455	0.824	0.476	0.411	0.498	0.405	0.408	0.400	0.382	0.387	0.482	0.422	0.210	0.242	0.281	0.309	0.295	0.241	0.213	0.549	0.569	0.522	0.500	0.493	0.529
0.383	0.458	0.482	0.370	0.384	0.361	0.400	0.391	0.320	0.319	0.425	0.377	0.433	0.528	0.497	0.285	0.247	0.198	0.226	0.410	0.570	0.597	0.576	0.588	0.531	0.493	0.546
0.459	0.476	0.391	0.431	0.563	0.321	0.364	0.382	0.365	0.368	0.405	0.287	0.263	0.509	0.606	0.569	0.509	0.554	0.551	0.591	0.622	0.647	0.612	0.648	0.594	0.537	0.546

What is understandable, interpretable, intelligible?

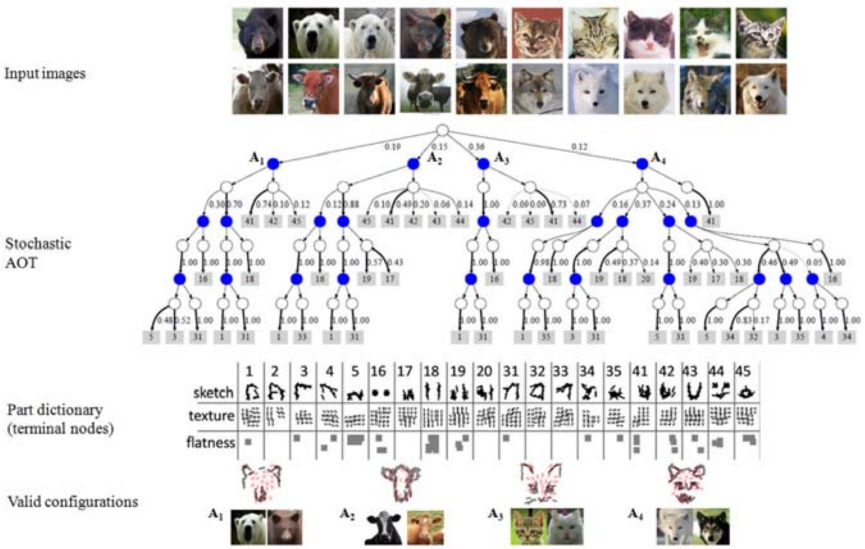


https://www.vis.uni-konstanz.de/en/members/fuchs/

Explainable AI is a huge challenge for visualization

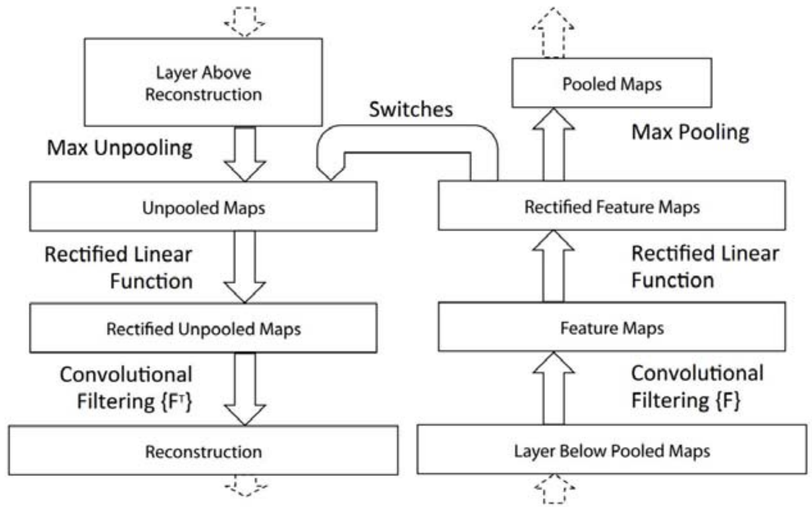


AHC: Stochastic And-Or-Templates



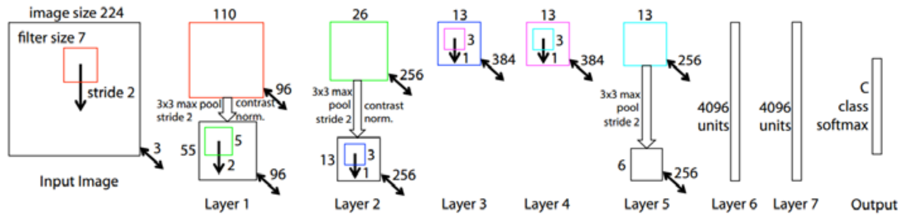
Zhangzhang Si & Song-Chun Zhu 2013. Learning and-or templates for object recognition and detection. IEEE transactions on pattern analysis and machine intelligence, 35, (9), 2189-2205, doi:10.1109/TPAMI.2013.35.

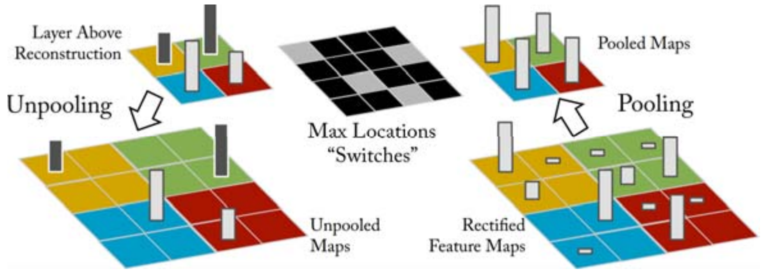
Example: Interpretable Deep Learning Model



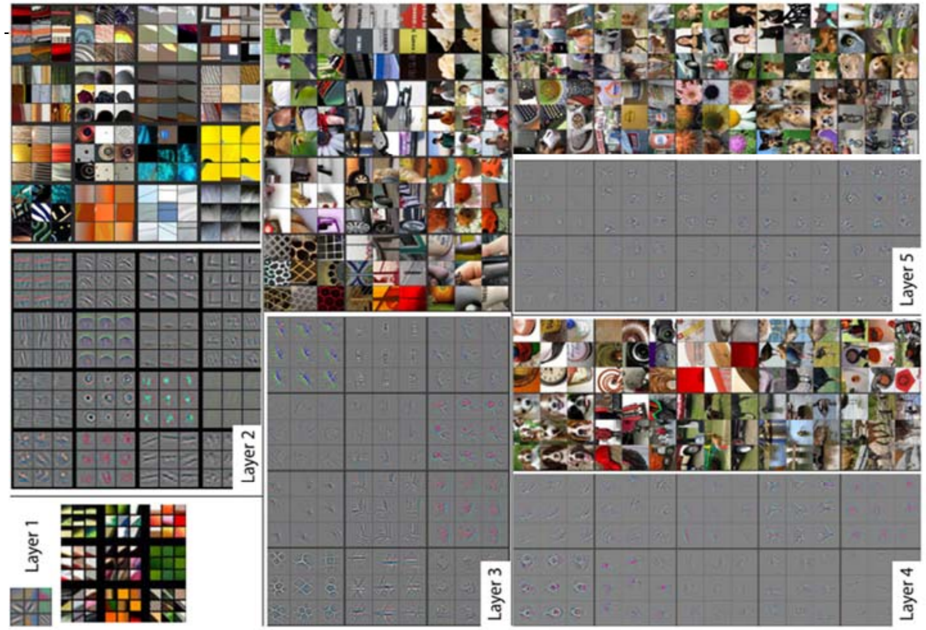
Matthew D. Zeiler & Rob Fergus 2013. Visualizing and Understanding Convolutional Networks. arXiv:1311.2901.

Visualizing a Conv Net with a De-Conv Net



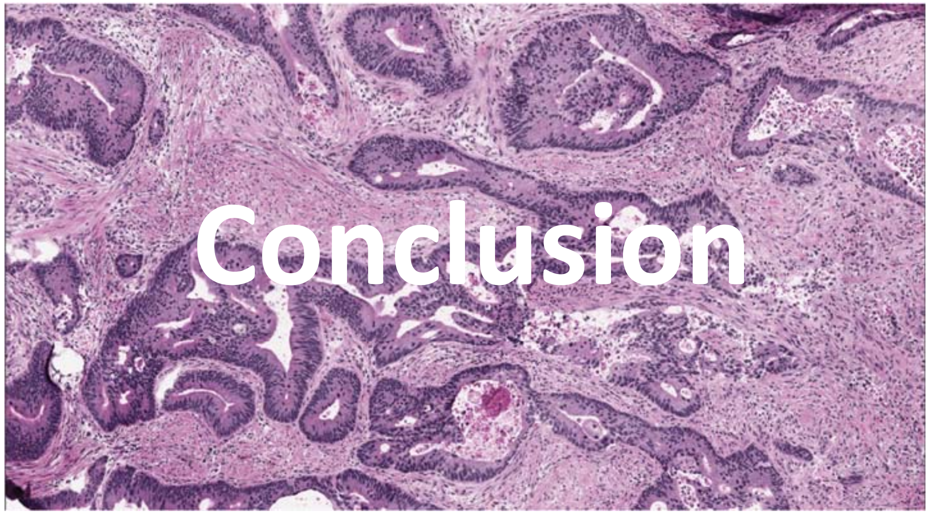


Matthew D. Zeiler & Rob Fergus 2014. Visualizing and understanding convolutional networks. In: D., Fleet, T., Pajdla, B., Schiele & T., Tuytelaars (eds.) ECCV, Lecture Notes in Computer Science LNCS 8689. Cham: Springer, pp. 818-833, doi:10.1007/978-3-319-10590-1_53.



Matthew D. Zeiler & Rob Fergus 2013. Visualizing and Understanding Convolutional Networks. arXiv:1311.2901.

What is interesting? What is relevant?



This is compatible to interactive machine learning

- Computational approaches can find in R^n what no human is able to see
- However, still there are many hard problems where a human expert in R^2 can understand the **context** and bring in experience, expertise, knowledge, intuition, …
- Black box approaches can not explain **WHY** a decision has been made …

The fist wave of AI (1943 – 1975): Handcrafted Knowledge

- Engineers create a set of logical rules to represent knowledge (Rule based Expert Systems)
- Advantage: works well in narrowly defined problems of well-defined domains
- Disadvantage: No adaptive learning behaviour and poor handling of p(x)

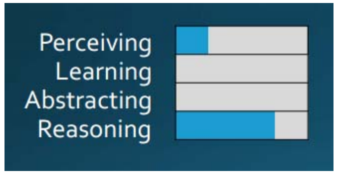


Image credit to John Launchbury

The second wave of AI (1975 –): Statistical Learning

- Engineers create learning models for specific tasks and train them with big data (e.g. Deep Learning)
- Advantage: works well for standard classification tasks and has prediction capabilities
- Disadvantage: No contextual capabilities and minimal reasoning abilities

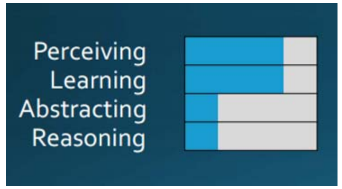


Image credit to John Launchbury

The third wave of AI (?): Adaptive Context Understanding

- A contextual model can perceive learn and understand and abstract and reason
- Advantage: can use transfer learning for adaptation on unknown unknowns
- Disadvantage: Superintelligence,

Image credit to John Launchbury

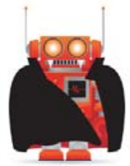
3 selected dangers of AI and superintelligence

Myth:
Superintelligence by 2100 is inevitable

Myth:
Superintelligence by 2100 is impossible

Fact:
It may happen in decades, centuries or never: AI experts disagree & we simply don't know

Myth:
Robots are the main concern



Fact:
Misaligned intelligence is the main concern: it needs no body, only an internet connection

Myth:
AI can't control humans

Fact:
Intelligence enables control: we control tigers by being smarter

https://futureoflife.org/background/benefits-risks-of-artificial-intelligence

Contextual adaptations can explain decisions

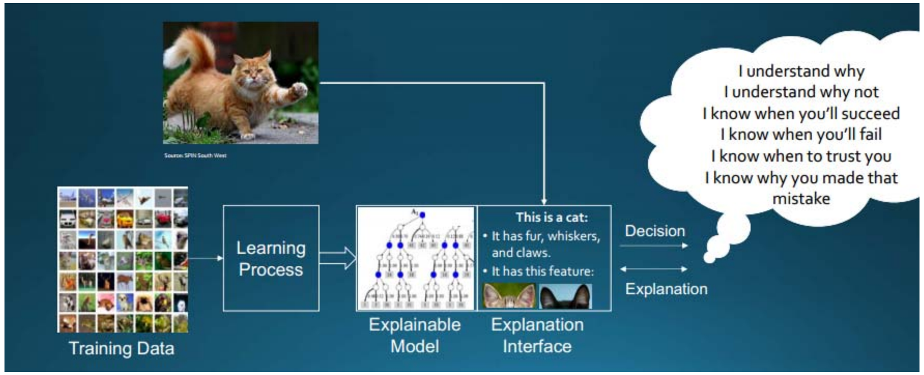


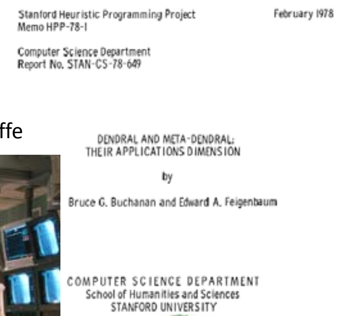
Image credit to John Launchbury

Thank you!

Appendix

History of DSS is a history of artificial intelligence

E. Feigenbaum, J. Lederberg, B. Buchanan, E. Shortliffe



Stanford Heuristic Programming Project
Memo HPP-78-1
February 1978

Computer Science Department
Report No. STAN-CS-78-649

DENDRAL AND META-DENDRAL:
THEIR APPLICATIONS DIMENSION

by

Bruce G. Buchanan and Edward A. Feigenbaum

COMPUTER SCIENCE DEPARTMENT
School of Humanities and Sciences
STANFORD UNIVERSITY

Rheingold, H. (1985) *Tools for thought: the history and future of mind-expanding technology. New York, Simon & Schuster.*

Buchanan, B. G. & Feigenbaum, E. A. (1978) DENDRAL and META-DENDRAL: their applications domain. *Artificial Intelligence, 11, 1978, 5-24.*