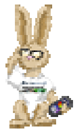
Andreas Holzinger
VO 709.049 Medical Informatics
12.10.2016 11:15-12:45



Lecture 01 Introduction
Computer Science meets Life Sciences: Challenges and Future Directions

a.holzinger@tugraz.at
Tutor: markus.plass@student.tugraz.at
http://hci-kdd.org/biomedical-informatics-big-data

Reading on Paper or on any electronic device

Holzinger, A. 2014. Biomedical Informatics: Discovering Knowledge in Big Data, New York, Springer, doi:10.1007/978-3-319-04528-3.

Slide 0-1: Overview – Roadmap trough this Course

- **01. Intro: Computer Science meets Life Sciences, challenges, future directions**
- 02. Fundamentals of Data, Information and Knowledge
- 03. Structured Data: Coding, Classification (ICD, SNOMED, MeSH, UMLS)
- 04. Biomedical Databases: Acquisition, Storage, Information Retrieval and Use
- 05. Semi structured , weakly structured data and unstructured information
- 06. Multimedia Data Mining and Knowledge Discovery
- 07. Knowledge and Decision: Cognitive Science & Human-Computer Interaction
- 08. Biomedical Decision Making: Reasoning and Decision Support
- 09. Interactive Information Visualization and Visual Analytics
- 10. Biomedical Information Systems and Medical Knowledge Management
- 11. Biomedical Data: Privacy, Safety and Security
- 12. Methodology for Info Systems: System Design, Usability & Evaluation

Keywords of Lecture 01

- Big Data (= large amounts of data sets)
- Life and Complexity
- Proteins – DNA & RNA – Cell – Tissue – Organ – Cardiovascular Systems
- Medicine – Informatics – Computer
- Personalized Medicine
- Translational Informatics
- Data Integration
- Biomarker Discovery

Learning Goals

- At the end of this first lecture you will …
- … see why the HCI-KDD approach is important
- … understand why machine learning in necessary
- … be fascinated to see our world in data;
- … be aware of the complexity of the health domain;
- … have some ideas of some future directions of Biomedical Informatics;

Advance Organizer (1/2)

- **Bioinformatics** = discipline, as part of biomedical informatics, at the interface between *bio*logy and *infor*mation science and mathema*tics*; processing of biological data;
- **Biomarker** = a characteristic (e.g. body-temperature (fever) as a biomarker for an infection, or proteins measured in the urine) as an indicator for normal or pathogenic biological processes, or pharmacologic responses to a therapeutic intervention;
- **Biomedical data** = compared with general data, it is characterized by large volumes, complex structures, high dimensionality, evolving biological concepts, and insufficient data modeling practices;
- **Biomedical Informatics** = 2011-definition: similar to medical informatics but including the optimal use of biomedical data, e.g. from genomics, proteomics, metabolomics;
- **Classical Medicine** = is both the science and the art of healing and encompasses a variety of practices to maintain and restore health;
- **Genomics** = branch of molecular biology which is concerned with the structure, function, mapping & evolution of genomes;
- **interactive Machine Learning** = defined as algorithms that can interact with both computational agents and human agents and can optimize their learning behaviour through these interactions, by bringing in a human-into-the-loop
- **Machine Learning** = addresses the question of how to design algorithms that improve automatically through experience from big data - doing it automatically (aML) without a human-in-the-loop
- **Medical Informatics** = 1970-definition: "… scientific field that deals with the storage, retrieval, and optimal use of medical information, data, and knowledge for problem solving and decision making";
- **Metabolomics** = study of chemical processes involving metabolites (e.g. enzymes). A challenge is to integrate proteomic, transcriptomic, and metabolomic information to provide a more complete understanding of living organisms;
- **Molecular Medicine** = emphasizes cellular and molecular phenomena and interventions rather than the previous conceptual and observational focus on patients and their organs;

Advance Organizer (2/2)

- **Omics data** = data from e.g. genomics, proteomics, metabolomics, etc.
- **Pervasive Computing** = similar to ubiquitous computing (Ubicomp), a post-desktop model of Human-Computer Interaction (HCI) in which information processing is integrated into every-day, miniaturized and embedded objects and activities; having some degree of "intelligence";
- **Pervasive Health** = all unobtrusive, analytical, diagnostic, supportive etc. information functions to improve health care, e.g. remote, automated patient monitoring, diagnosis, home care, self-care, independent living, etc.;
- **Proteome** = the entire complement of proteins that is expressed by a cell, tissue, or organism;
- **Proteomics** = field of molecular biology concerned with determining the proteome;
- **P-Health Model** = Preventive, Participatory, Pre-emptive, Personalized, Predictive, Pervasive (= available to anybody, anytime, anywhere);
- **Space** = a set with some added structure;
- **Technological Performance** = machine "capabilities", e.g. short response time, high throughput, high availability, etc.
- **Time** = a dimension in which events can be ordered along a time line from the past through the present into the future;
- **Translational Medicine** = based on interventional epidemiology; progress of Evidence-Based Medicine (EBM), integrates research from basic science for patient care and prevention;
- **Von-Neumann-Computer** = a 1945 architecture, which still is the predominant machine architecture of today (opp.: Non-Vons, incl. analogue, optical, quantum computers, cell processors, DNA and neural nets (in silico));

Acronyms/Abbreviations (incomplete selection)

- AI = Artificial Intelligence
- AL = Artificial Life
- CPG = Clinical Practice Guideline
- CPOE = Computerized physician order entry (medical communication about treatment of patients)
- CMV = Controlled Medical Vocabulary
- DEC = Digital Equipment Corporation (1957-1998)
- DNA = Deoxyribonucleic Acid
- EBM = Evidence Based Medicine
- EPR = Electronic Patient Record
- GBM = Genome Based Medicine
- GC = Gas Chromatography
- GPM = Genetic Polymorphism
- HCI = Human–Computer Interaction
- LC = Liquid Chromatography
- LNCS = Lecture Notes in Computer Science
- MS = Mass Spectrometry
- mRNA = Messenger RNA
- NGC = New General Catalogue of Nebulae and Star clusters in Astronomy
- NGS = Next Generation Sequencing
- NMR = Nuclear Magnetic Resonance
- PDB = Protein Data Base
- PDP = Programmable Data Processor (mainframe)
- PPI = Protein-Protein Interaction
- RFID = Radio-frequency identification device
- RNA = Ribonucleic Acid
- SNP = Single Nucleotide Polymorphism
- TNF = Tumor Necrosis Factor
- TQM = Total Quality Management

Agenda for today

- 01 What is the HCI-KDD approach?
- 02 Application Area: Health
- 03 Probabilistic Information p(x)
- 04 automatic Machine Learning (aML)
- 05 interactive Machine Learning (iML)
- 06 Key Problems in Health Informatics
- 07 Medical/Biomedical/Health Informatics
- 08 Future Challenges

01 What is the HCI-KDD approach?

- **ML is a very practical field – algorithm development is at the core – however, successful ML needs a concerted effort of various topics …**

In the focus of our ML pipeline are algorithms that learn …

01

Holzinger, A. 2014. Trends in Interactive Knowledge Discovery for Personalized Medicine: Cognitive Science meets Machine Learning. IEEE Intelligent Informatics Bulletin, 15, (1), 6-14.

http://www.bach-cantatas.com

- Cognitive Science $\rightarrow$ human intelligence
- Computer Science $\rightarrow$ computational intelligence
- Human-Computer Interaction $\rightarrow$ the bridge

"Solve intelligence – then solve everything else"

Demis Hassabis, 22 May 2015

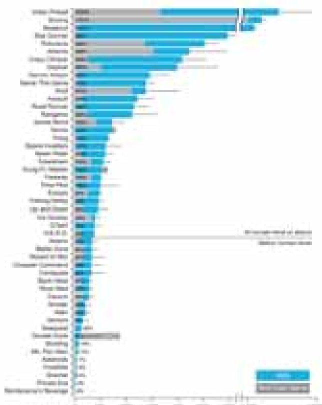
The Royal Society, Future Directions of Machine Learning Part 2

Google DeepMind

https://youtu.be/XAbLn66iHcQ?t=1h28m54s

Compare your best ML algorithm with a seven year old child ...

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. & Hassabis, D. 2015. Human-level control through deep reinforcement learning. Nature, 518, (7540), 529-533, doi:10.1038/nature14236



Today ML is enormously progressing …

- Progress is driven by the explosion in the availability of **big data** and **low-cost computation.**
- **Health is amongst the biggest challenges**

Jordan, M. I. & Mitchell, T. M. 2015. Machine learning: Trends, perspectives, and prospects. Science, 349, (6245), 255-260.

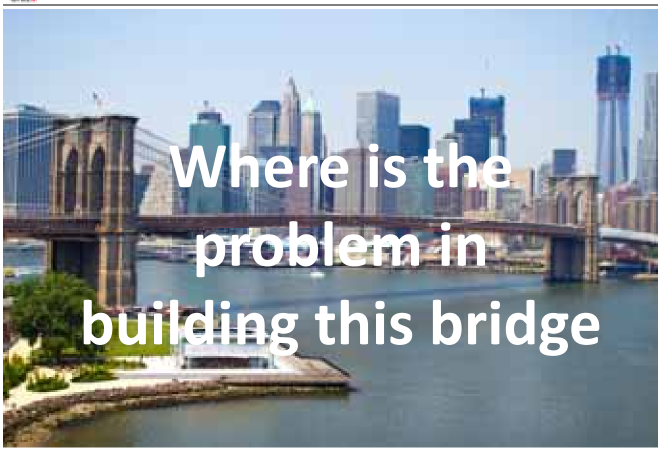
Health is a complex area

Why is this application area complex ?

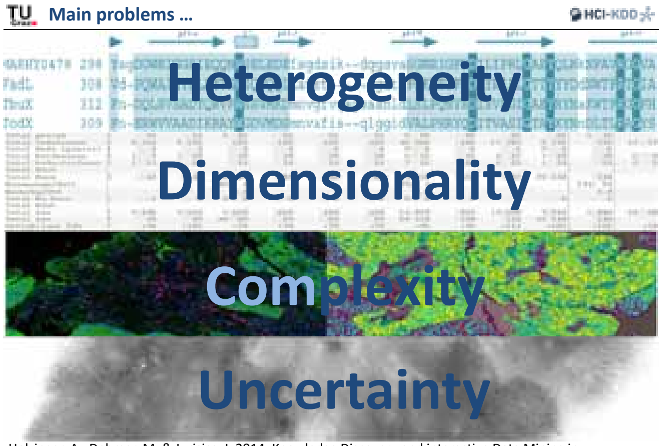
In medicine we have two different worlds …

Our central hypothesis: Information may bridge this gap

Holzinger, A. & Simonic, K.-M. (eds.) 2011. *Information Quality in e-Health. Lecture Notes in Computer Science LNCS 7058, Heidelberg, Berlin, New York: Springer.*



Main problems …



Holzinger, A., Dehmer, M. & Jurisica, I. 2014. Knowledge Discovery and interactive Data Mining in Bioinformatics - State-of-the-Art, future challenges and research directions. BMC Bioinformatics, 15, (S6), I1.

03 Probabilistic Information p(x)

The foundation for machine learning was laid in 1763 …

03

[illegible]

Richard Price
1723-1791

Bayes, T. (1763). An Essay towards solving a Problem in the Doctrine of Chances (Postum communicated by Richard Price). Philosophical Transactions, 53, 370-418.

$$p(x_i) = \sum P(x_i, y_j) \qquad p(x_i, y_j) = p(y_j|x_i)P(x_i)$$

Bayes' Rule is a corollary of the Sum Rule and Product Rule:

$$p(x_i|y_j) = \frac{p(y_j|x_i)p(x_i)}{\sum p(x_i, y_j)p(x_i)}$$

Barnard, G. A., & Bayes, T. (1958). Studies in the history of probability and statistics: IX. Thomas Bayes's essay towards solving a problem in the doctrine of chances. Biometrika, 45(3/4), 293-315.

Analogies

- **Newton, Leibniz, … developed calculus – mathematical language for describing and dealing with rates of change**
- **Bayes, Laplace, … developed probability theory - the mathematical language for describing and dealing with uncertainty**
- **Gauss generalized those ideas**

Learning and Inference

04

d … *data* $\qquad \mathcal{H}$ …$\{H_1, H_2, \ldots, H_n\}$ $\qquad \forall\, h, d$ …

h … *hypotheses*

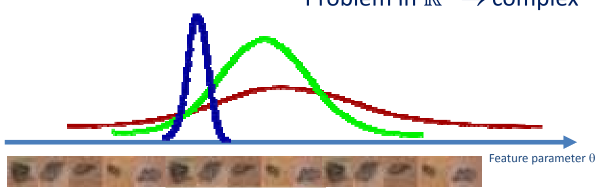
Likelihood

Prior Probability

$$p(h|d) = \frac{p(d|h) * p(h)}{\sum_{h \in H} p(d|h')\, p(h')}$$

Posterior Probability

Problem in $\mathbb{R}^n \rightarrow$ complex



Bayesian Learning from data → Generalize

05

$$\mathcal{D} = x_{1:n} = \{x_1, x_2, \ldots, x_n\} \qquad p(\mathcal{D}|\theta)$$

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta) * p(\theta)}{p(\mathcal{D})}$$

$$posterior = \frac{likelihood * prior}{evidence}$$

The inverse probability allows to learn from data, infer unknowns, and make predictions

Why is this relevant for health informatics?

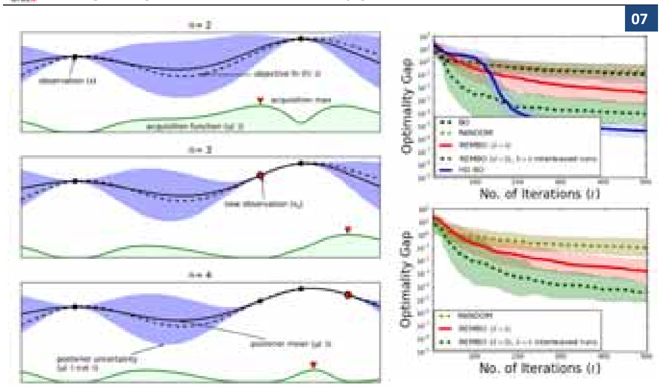
Reasoning under uncertainty

- Take patient information, e.g., observations, symptoms, test results, -omics data, etc. etc.
- Reach conclusions, and **predict** into the future, e.g. how likely will the patient be …
- Prior = belief before making a particular observation
- Posterior = belief after making the observation and is the prior for the next observation – intrinsically incremental

$$p(x_i|y_j) = \frac{p(y_j|x_i)p(x_i)}{\sum p(x_i, y_j)p(x_i)}$$

Scaling to high-dimensions is the holy grail in ML

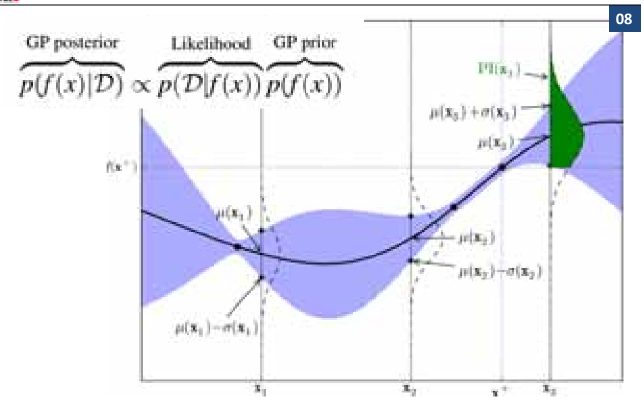
07



Wang, Z., Hutter, F., Zoghi, M., Matheson, D. & De Feitas, N. 2016. Bayesian optimization in a billion dimensions via random embeddings. Journal of Artificial Intelligence Research, 55, 361-387, doi:10.1613/jair.4806.

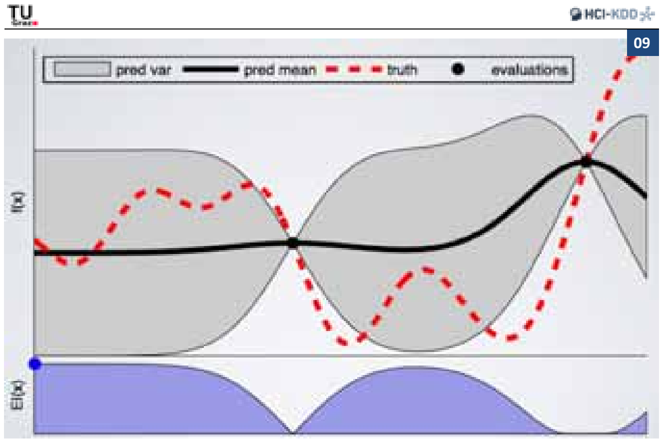
GP = distribution, observations occur in a cont. domain, e.g. t or space

08



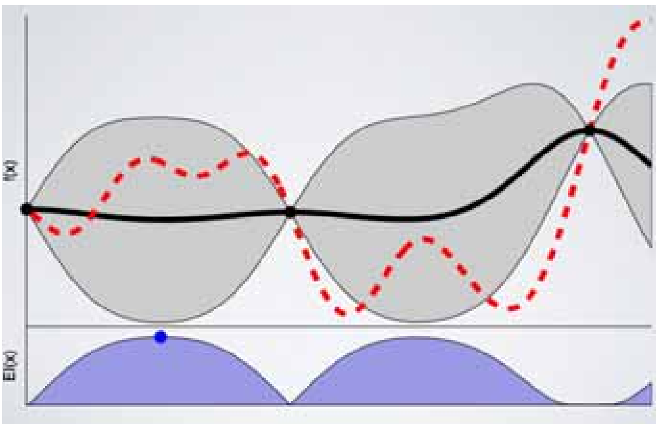
Brochu, E., Cora, V. M. & De Freitas, N. 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv:1012.2599.

09

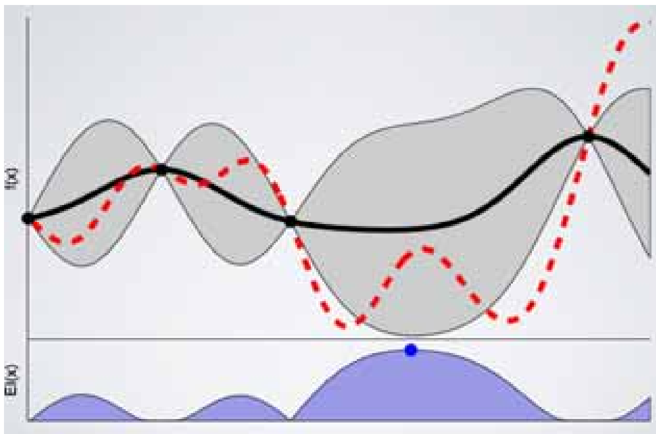


Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian optimization of machine learning algorithms. Advances in neural information processing systems, 2012. 2951-2959.

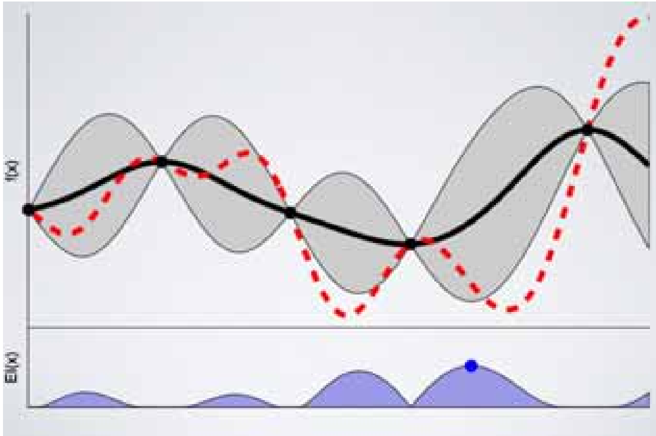
Bayesian Optimization 1



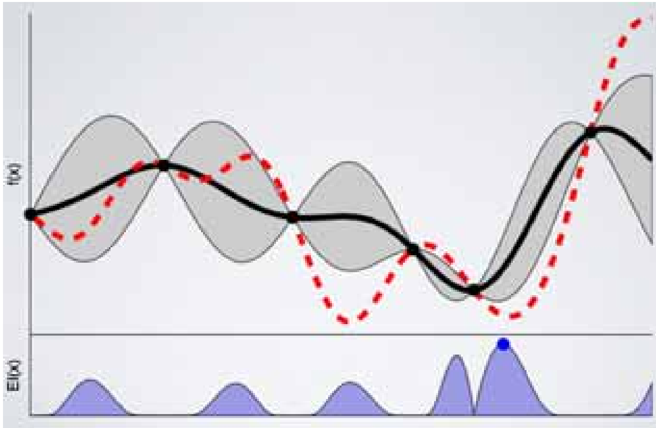
Bayesian Optimization 2



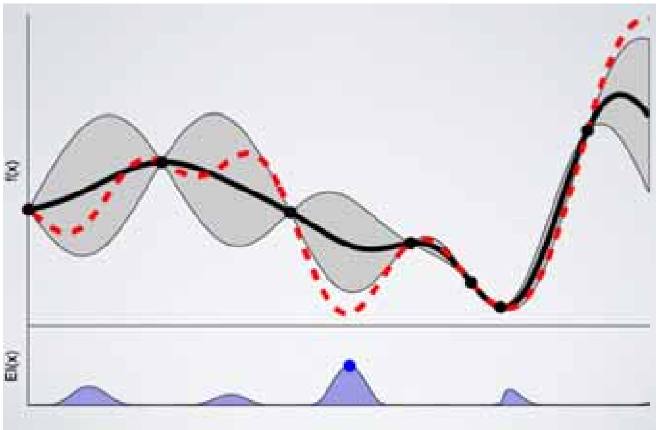
Bayesian Optimization 3



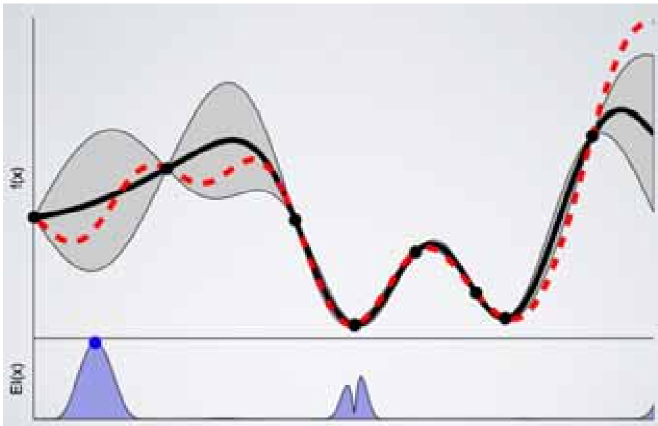
Bayesian Optimization 4



Bayesian Optimization 5

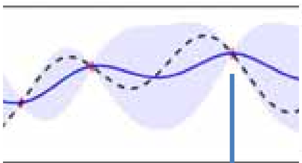


Bayesian Optimization 6



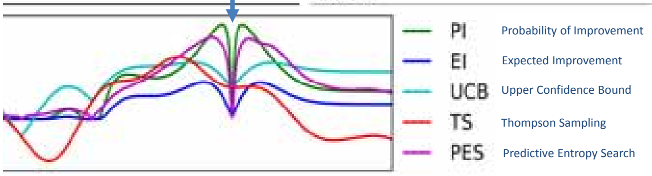
Fully automatic → Goal: Taking the human out of the loop

07



Algorithm 1 Bayesian optimization

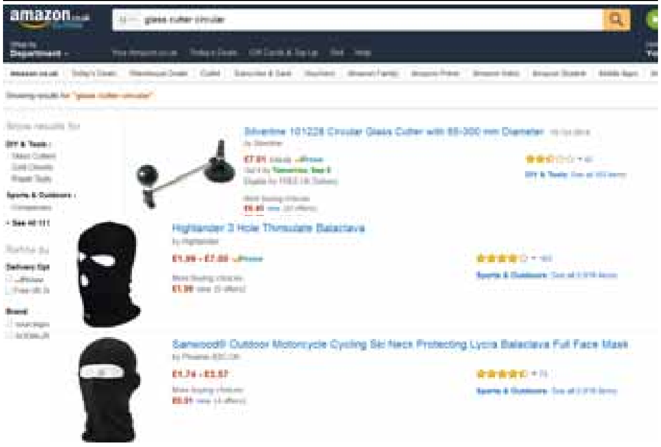
1: **for** $n = 1, 2, \ldots$ **do**
2: select new $\mathbf{x}_{n+1}$ by optimizing acquisition function α
$$\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}} \alpha(\mathbf{x}; \mathcal{D}_n)$$
3: query objective function to obtain y_{n+1}
4: augment data $\mathcal{D}_{n+1} = \{\mathcal{D}_n, (\mathbf{x}_{n+1}, y_{n+1})\}$
5: update statistical model
6: **end for**



Shahriari, B., Swersky, K., Wang, Z., Adams, R. P. & De Freitas, N. 2016. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104, (1), 148-175, doi:10.1109/JPROC.2015.2494218.

04 aML

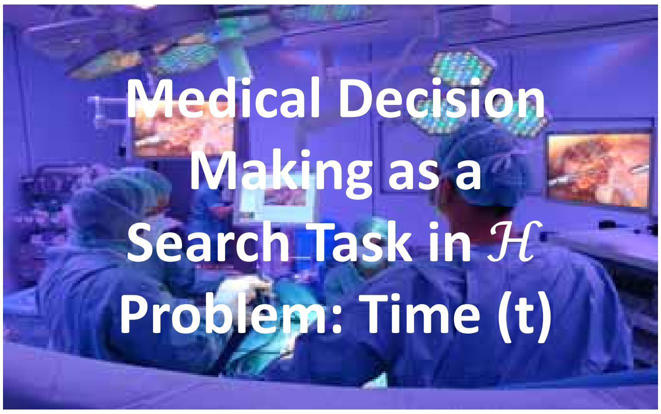
Example for aML: Recommender Systems



Fully automatic autonomous vehicles (Google car)

Dietterich, T. G. & Horvitz, E. J. 2015. Rise of concerns about AI: reflections and directions. Communications of the ACM, 58, (10), 38-40.

Does this all work here as well?



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

When does aML fail …

- Sometimes we **do not have "big data"**, where aML-algorithms benefit.
- Sometimes we have
  - **Small amount of data sets**
  - Rare Events – **no training samples**
  - **NP-hard problems, e.g.**
    - Subspace Clustering,
    - k-Anonymization,
    - Protein-Folding, …

Holzinger, A. 2016. Interactive Machine Learning for Health Informatics: When do we need the human-in-the-loop? Springer Brain Informatics (BRIN), 3, (2), 119-131, doi:10.1007/s40708-016-0042-6.

Consequently …

Sometimes we (still) need a human-in-the-loop

05 iML

Definition of iML (Holzinger – 2016)

A6

- iML := algorithms which interact with agents***)** and can optimize their learning behaviour through this interaction

***) where the agents can be human**

Holzinger, A. 2016. Interactive Machine Learning (iML). Informatik Spektrum, 39, (1), 64-68, doi:10.1007/s00287-015-0941-6.

Sometimes we need a doctor-in-the-loop

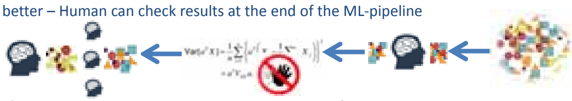
A group of experts-in-the-loop

A crowd of people-in-the-loop

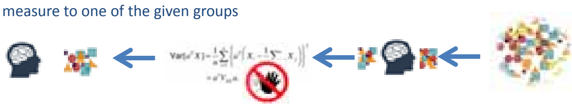
aML: taking the human-out-of-the-loop

A) Unsupervised ML: Algorithm is applied on the raw data and learns fully automatic – Human can check results at the end of the ML-pipeline

B) Supervised ML: Humans are providing the labels for the training data and/or select features to feed the algorithm to learn – the more samples the better – Human can check results at the end of the ML-pipeline



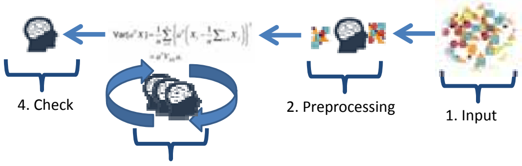
C) Semi-Supervised Machine Learning: A mixture of A and B – mixing labeled and unlabeled data so that the algorithm can find labels according to a similarity measure to one of the given groups



iML: bringing the human-in-the-loop

12

D) **Interactive Machine Learning:** Human is seen as an agent involved in the actual learning phase, step-by-step influencing measures such as distance, cost functions …



Constraints of humans: Robustness, subjectivity, transfer?
Open Questions: Evaluation, replicability, …

Holzinger, A. 2016. Interactive Machine Learning for Health Informatics: When do we need the human-in-the-loop? Brain Informatics (BRIN), 3, (2), 119-131, doi:10.1007/s40708-016-0042-6.

Three examples for the usefulness of the iML approach

- **Example 1: Subspace Clustering**
- **Example 2: k-Anonymization**
- **Example 3: Protein Design**

Hund, M., Böhm, D., Sturm, W., Sedlmair, M., Schreck, T., Ullrich, T., Keim, D. A., Majnaric, L. & Holzinger, A. 2016. Visual analytics for concept exploration in subspaces of patient groups: Making sense of complex datasets with the Doctor-in-the-loop. Brain Informatics, 1-15, doi:10.1007/s40708-016-0043-5.

Kieseberg, P., Frühwirt, P., Weippl, E. & Holzinger, A. 2015. Witnesses for the Doctor in the Loop. In: Guo, Y., Friston, K., Aldo, F., Hill, S. & Peng, H. (eds.) Lecture Notes in Artificial Intelligence LNAI 9250. Springer, pp. 369-378, doi:10.1007/978-3-319-23344-4_36.

Lee, S. & Holzinger, A. 2016. Knowledge Discovery from Complex High Dimensional Data. In: Michaelis, S., Piatkowski, N. & Stolpe, M. (eds.) Solving Large Scale Learning Tasks. Challenges and Algorithms, Lecture Notes in Artificial Intelligence LNAI 9580. Springer, pp. 148-167, doi:10.1007/978-3-319-41706-6_7.

06 Key Problems in health informatics

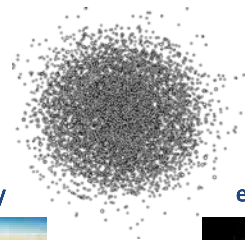
Key Problems

- **Zillions** of different biological species (humans, animals, bacteria, virus, plants, …);
- Enormous **complexity** of the medical domain [1];
- **Complex,** heterogeneous, high-dimensional, big data in the life sciences [2];
- Limited **time,** e.g. a medical doctor in a public hospital has only 5 min. to make a decision [3];
- Limited **computational power** in comparison to the complexity of life (and the natural limitations of the Von-Neumann architecture, …);

1. Patel VL, Kahol K, & Buchman T (2011) Biomedical Complexity and Error. *J. Biomed. Inform.* 44(3):387-389.
2. Holzinger A, Dehmer M, & Jurisica I (2014) Knowledge Discovery and interactive Data Mining in Bioinformatics - State-of-the-Art, future challenges and research directions. BMC Bioinformatics 15(S6):I1.
3. Gigerenzer G (2008) Gut Feelings: Short Cuts to Better Decision Making (Penguin, London).

Slide 1-1: Our World in Data (1/2) Macroscopic Structures

Two thematic mainstreams in dealing with data …



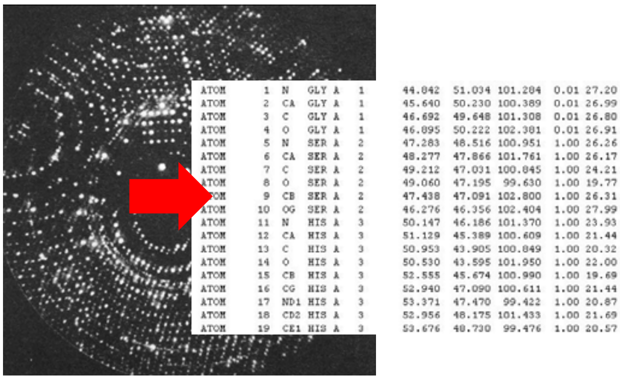
Time **Space**

e.g. Entropy **e.g. Topology**

Dali, S. (1931) The persistence of memory

Bagula & Bourke (2012) Klein-Bottle

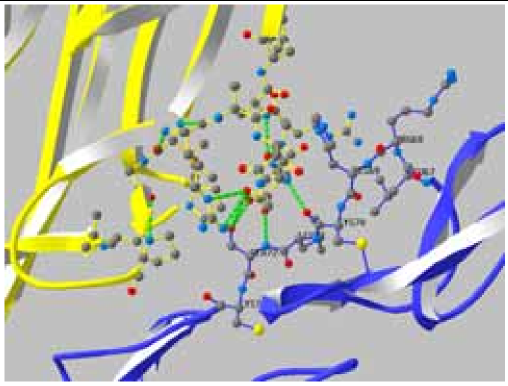
Slide 1-2: Our World in Data (2/2) – Microscopic Structures



ATOM	1	N	GLY	A	1	44.842	51.034	101.284	0.01	27.20
ATOM	2	CA	GLY	A	1	45.640	50.230	100.389	0.01	26.99
ATOM	3	C	GLY	A	1	46.692	49.648	101.308	0.01	26.80
ATOM	4	O	GLY	A	1	46.895	50.222	102.381	0.01	26.91
ATOM	5	N	SER	A	2	47.283	48.516	100.951	1.00	26.26
ATOM	6	CA	SER	A	2	48.277	47.866	101.761	1.00	26.17
ATOM	7	C	SER	A	2	49.212	47.031	100.845	1.00	24.21
ATOM	8	O	SER	A	2	49.060	47.195	99.630	1.00	19.77
ATOM	9	CB	SER	A	2	47.438	47.091	102.800	1.00	26.31
ATOM	10	OG	SER	A	2	46.276	46.356	102.404	1.00	27.99
ATOM	11	N	HIS	A	3	50.147	46.186	101.370	1.00	23.93
ATOM	12	CA	HIS	A	3	51.129	45.389	100.609	1.00	21.44
ATOM	13	C	HIS	A	3	50.953	43.905	100.849	1.00	20.32
ATOM	14	O	HIS	A	3	50.530	43.595	101.950	1.00	22.00
ATOM	15	CB	HIS	A	3	52.555	45.674	100.990	1.00	19.69
ATOM	16	CG	HIS	A	3	52.940	47.090	100.611	1.00	21.44
ATOM	17	ND1	HIS	A	3	53.371	47.470	99.422	1.00	20.87
ATOM	18	CD2	HIS	A	3	52.956	48.175	101.433	1.00	21.69
ATOM	19	CE1	HIS	A	3	53.676	48.730	99.476	1.00	20.57

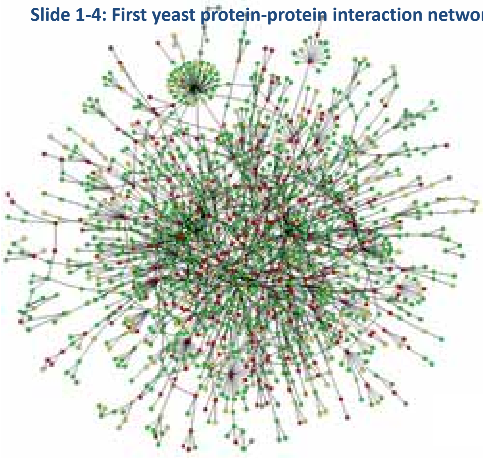
Wiltgen, M. & Holzinger, A. (2005) Visualization in Bioinformatics: Protein Structures with Physicochemical and Biological Annotations. In: *Central European Multimedia and Virtual Reality Conference. Prague, Czech Technical University (CTU), 69-74*

Slide 1-3: Knowledge Discovery from Data



Wiltgen, M., Holzinger, A. & Tilz, G. P. (2007) Interactive Analysis and Visualization of Macromolecular Interfaces Between Proteins. In: *Lecture Notes in Computer Science (LNCS 4799). Berlin, Heidelberg, New York, Springer, 199-212.*

Slide 1-4: First yeast protein-protein interaction network

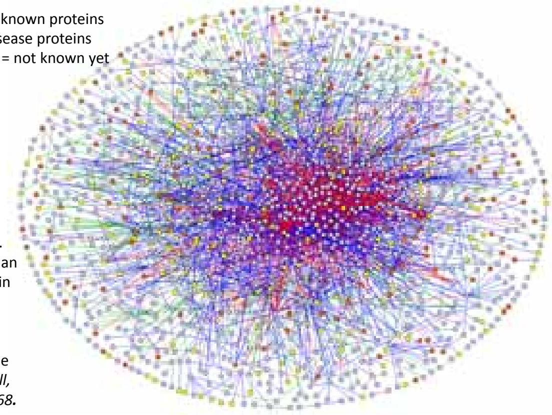


Nodes = proteins
Links = physical interactions (bindings)
Red Nodes = lethal
Green Nodes = non-lethal
Orange = slow growth
Yellow = not known

Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. (2001) Lethality and centrality in protein networks. *Nature, 411, 6833, 41-42.*

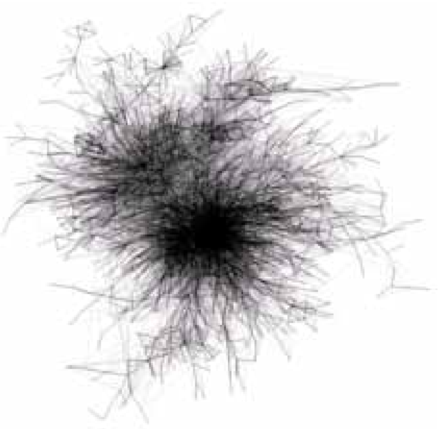
Slide 1-5: First human protein-protein interaction network

Light blue = known proteins
Orange = disease proteins
Yellow ones = not known yet



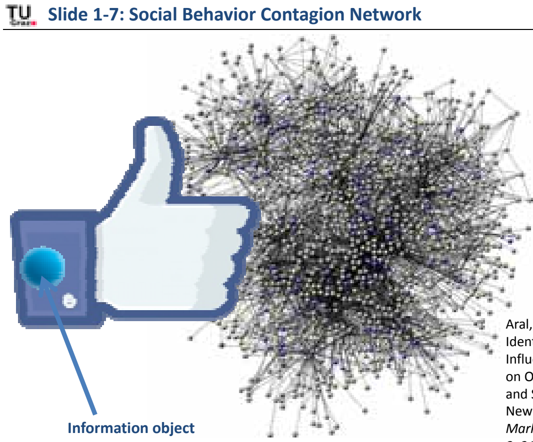
Stelzl, U. et al. (2005) A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome. *Cell, 122, 6, 957-968.*

Slide 1-6: Non-Natural Network Example: Blogosphere



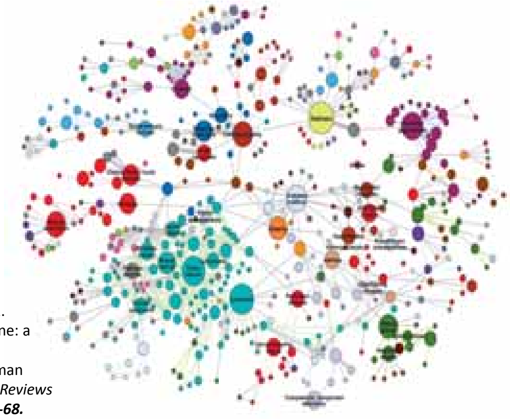
Hurst, M. (2007), Data Mining: Text Mining, Visualization and Social Media. Online available: http://datamining.typepad.com/data_mining/2007/01/the_blogosphere.html, last access: 2011-09-24

Slide 1-7: Social Behavior Contagion Network



Aral, S. (2011) Identifying Social Influence: A Comment on Opinion Leadership and Social Contagion in New Product Diffusion. *Marketing Science, 30, 2, 217-223.*

Slide 1-8: Human Disease Network -> Network Medicine

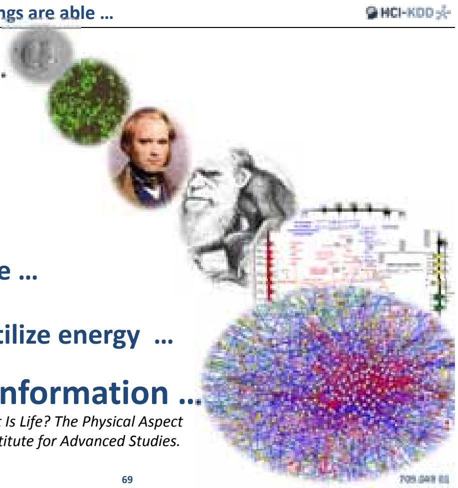


Barabási, A. L., Gulbahce, N. & Loscalzo, J. 2011. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics, 12*, ***56-68.***

What is life ?

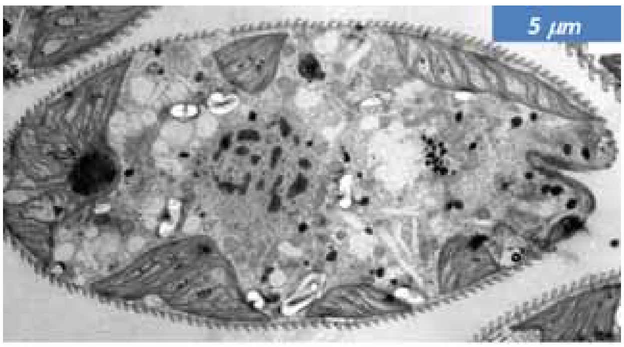
Slide 1-9 Living things are able …

- to reproduce …
- to grow …
- to evolve …
- to self-replicate …
- to generate/utilize energy …
- **to process information …**



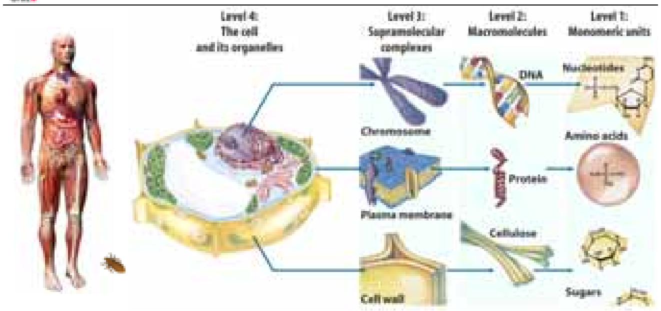
Schrödinger, E. (1944) *What Is Life? The Physical Aspect of the Living Cell. Dublin Institute for Advanced Studies.*

Slide 1-10: Life is complex information



Lane, N. & Martin, W. (2010) The energetics of genome complexity. *Nature, 467, 7318, 929-934.*

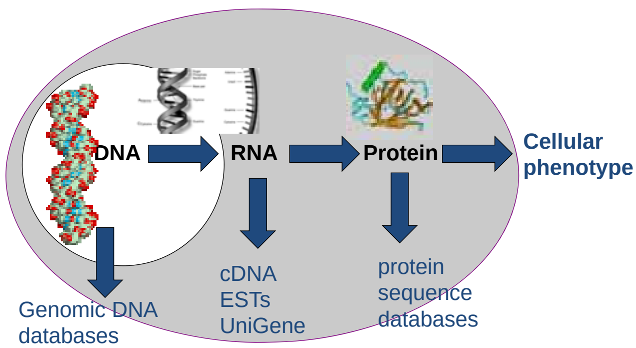
Slide 1-11 Building Blocks of Life - Overview



Human eye	Light microscope	Electron microscope	Special

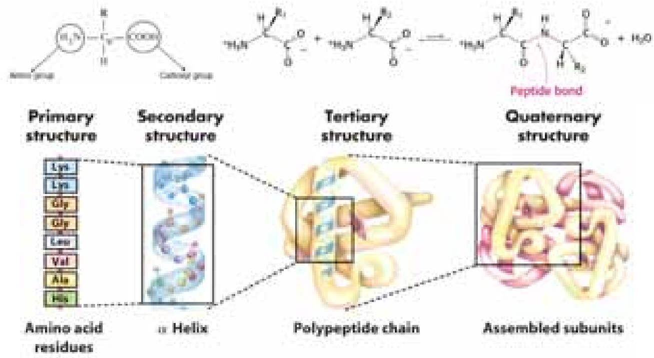
1m	1mm	1 µm	1nm	100 pm

Slide 1-12: The Dogma of Molecular Biology



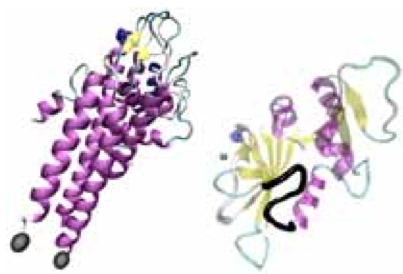
Crick, F. 1970. Central Dogma of Molecular Biology. *Nature, 227, (5258), 561-563.*

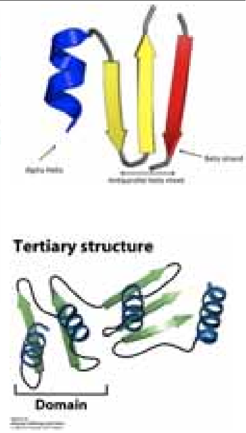
Slide 1-13 Amino-acid > Protein-chain > Protein-structure



Gromiha, M. 2010. *Protein Bioinformatics, Amsterdam, Elsevier.*

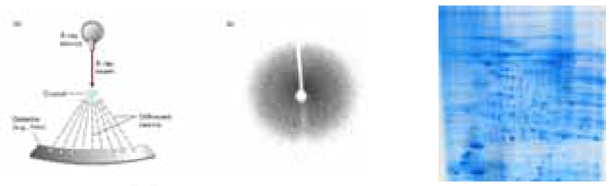
Slide 1-14 Tertiary Structure of a Protein

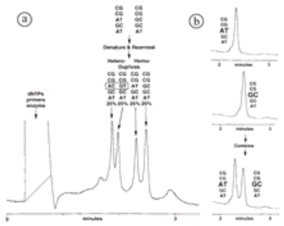




Shehu, A. & Kavraki, L. E. 2012. Modeling structures and motions of loops in protein molecules. *Entropy, 14, (2), 252-290.*

Slide 1-15 Protein Analytics





Rabilloud, et al. 2010. Two-dimensional gel electrophoresis in proteomics: past, present and future. *Journal of proteomics, 73, (11), 2064-2077.*

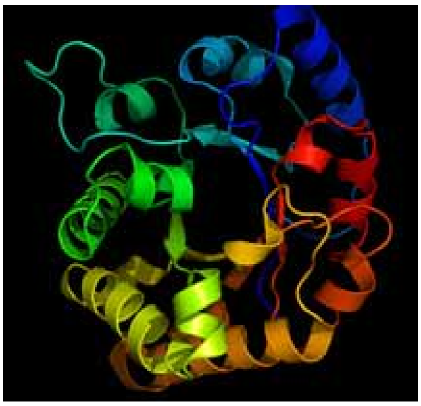
Xiao, W. Z. & Oefner, P. J. 2001. Denaturing high-performance liquid chromatography: A review. *Human Mutation, 17, (6), 439-474.*

Slide 1-16: Comparison of some current Methods

Technology	Sensitivity	Subcellular resolution	Cellular resolution	Minimally invasive?	Live cells?	Real time?
Genetically encoded nanosensors	Nanomolar to millimolar	Nanometer to millimeter	Yes	Yes	Yes	Yes
MRI	Mid-micromolar to millimolar (213)	No	Yes	Yes	Yes	Yes
PET	1–40 Bq mm^{-3} (18)	No	No	No	Yes	Yes
X-ray synchrotron	~1 mg kg^{-1} tissue (transit metals) (204)	No	Yes	No	No	No
SIMS	~1 fmol (67)	Yes	Yes	No	No	No
MALDI or TOF imaging	~1 ppm	Yes	50–100 μm (MALDI) 1–2 μm (TOF)	No	No	No
NIMS imaging	Yoctomolar (93)	No	50–100 μm	No	No	No
Mass spectrometry	Yoctomolar	No?	Yes	No	No	No
Raman	50 μM (70)	Yes	Yes	Yes	Yes	Yes

Okumoto, S., Jones, A. & Frommer, W. B. 2012. Quantitative imaging with fluorescent biosensors. *Annual review of plant biology, 63, 663-706.*

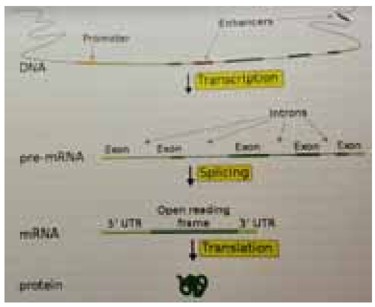
Slide 1-17 Enzymes



Klibanov, A. M. 2001. Improving enzymes by using them in organic solvents. *Nature, 409, (6817), 241-246.*

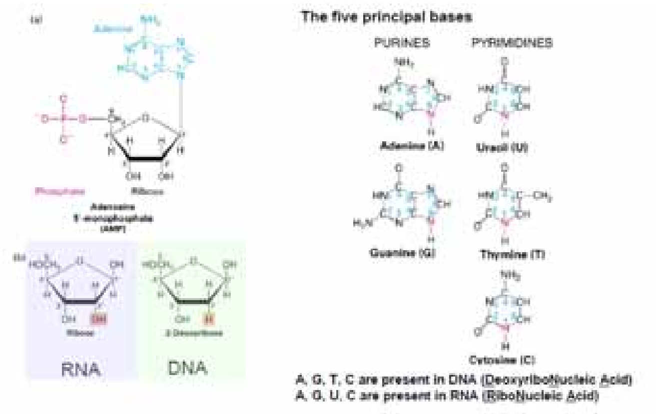
Slide 1-18 DNA-RNA-Proteins

The DNA, the RNA and the proteins are the three major macromolecules essential for all known forms of life.

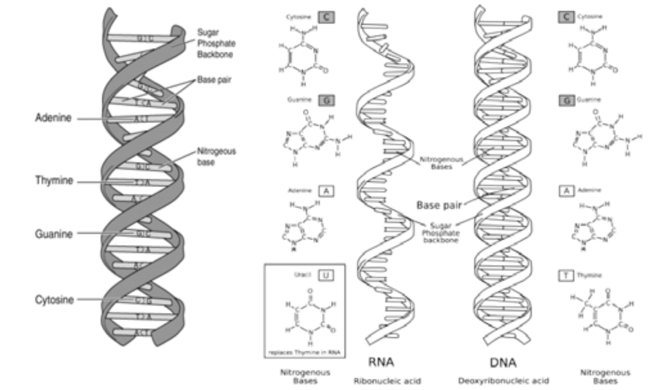


Manca, V. (2013). Infobiotics. Springer.

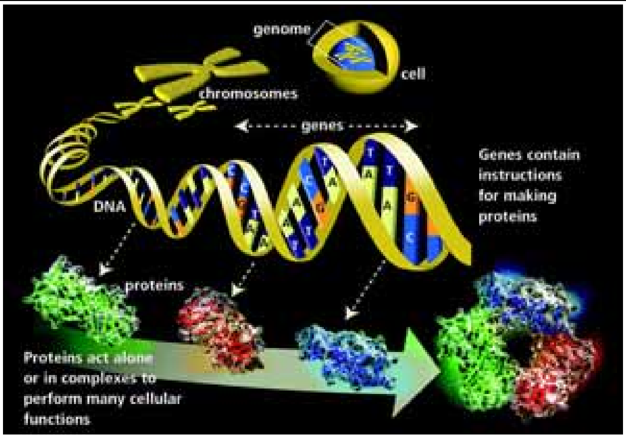
Slide 1-19 DNA & RNA – and the five principal bases



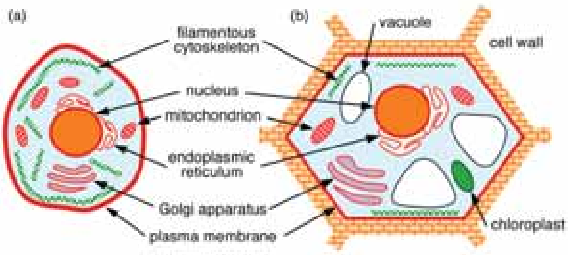
Slide 1-20: Nobel Prize 1957



Slide 1-21: Genetics – Genes – Genomics - Epigenetics

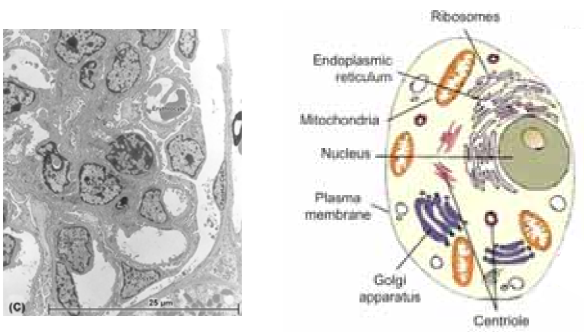


Slide 1-22: The Cell (simplified)



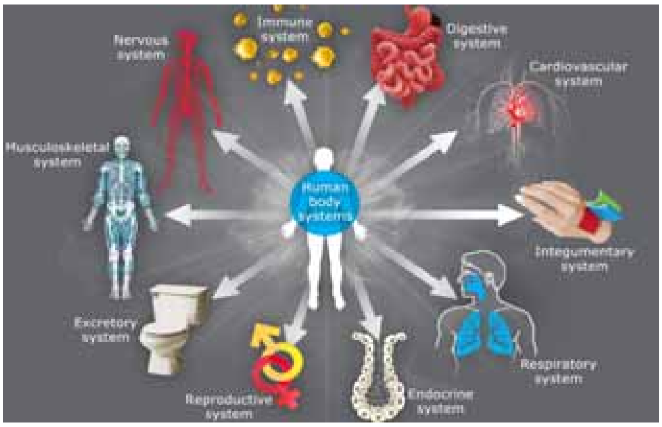
Boal, D. 2012. *Mechanics of the Cell, Cambridge University Press.*

Slide 1-23: The Cell Structure and Size

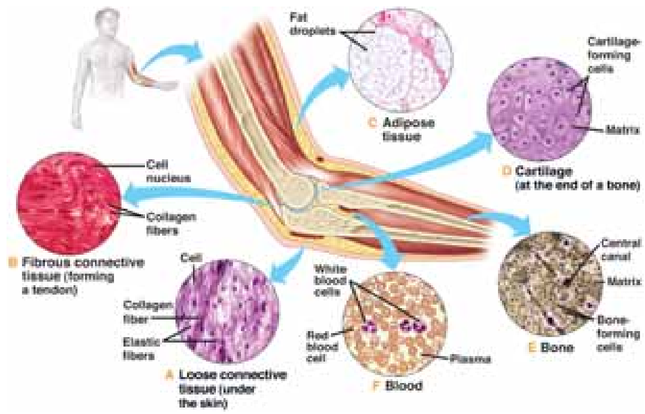


Sperelakis, N. 2012. *Cell Physiology Sourcebook: Essentials of Membrane Biophysics. Fourth Edition, Amsterdam, Elseviere.*

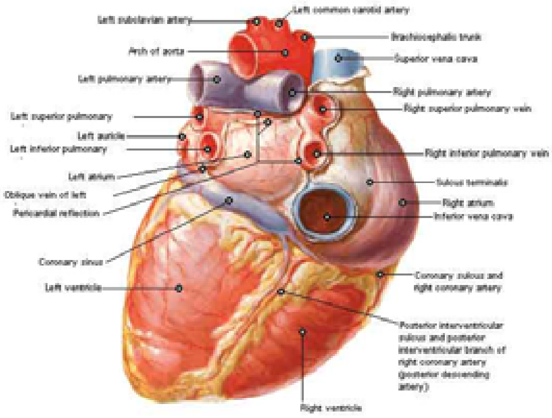
Slide 1-24: Human Organ Systems



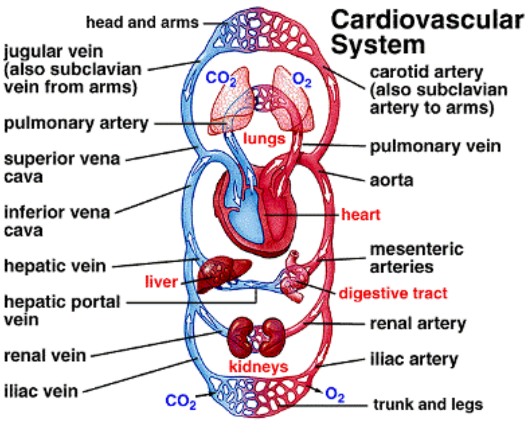
Slide 1-25: Tissue



Slide 1-26: Organ Example: Heart



Slide 1-27: Cardio-Vascular System



Slide 1-41: 4 decades from Medical to Biomedical Informatics

- 1970+ Begin of **Medical Informatics**
  - Focus on data acquisition, storage, accounting (typ. "EDV")
  - The term was first used in 1968 and the first course was set up 1978
- 1985+ Health Telematics
  - Health care networks, Telemedicine, CPOE-Systems etc.
- 1995+ Web Era
  - Web based applications, Services, EPR, etc.
- 2005+ Ambient Era
  - Pervasive & Ubiquitous Computing
- 2010+ Quality Era – **Biomedical Informatics**
  - Information Quality, Patient empowerment, individual molecular medicine

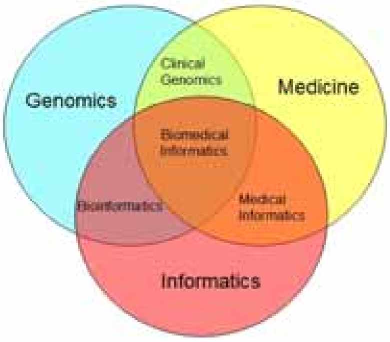
Slide 1-43: Definition of Biomedical Informatics

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

- ***Biomedical informatics*** **(BMI) is the interdisciplinary field that studies and pursues the effective use of biomedical data, information, and knowledge for scientific problem solving, and decision making, motivated by efforts to improve human health**

Shortliffe, E. H. (2011). Biomedical Informatics: Defining the Science and its Role in Health Professional Education. In A. Holzinger & K.-M. Simonic (Eds.), *Information Quality in e-Health. Lecture Notes in Computer Science LNCS 7058 (pp. 711-714). Heidelberg, New York: Springer.*

Slide 1-44: Computational Sciences meet Life Sciences



Health Informatics

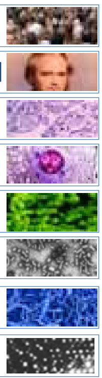
http://www.bioinformaticslaboratory.nl/twiki/bin/view/BioLab/EducationMIK1-2

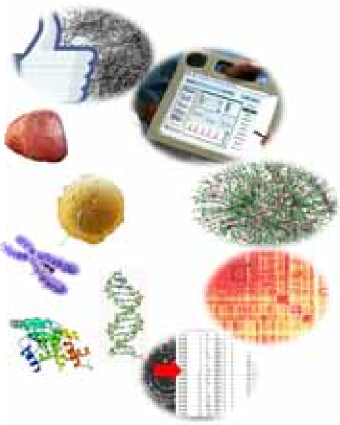
Slide 1:49 Health and Wellbeing

Holzinger, A., Dehmer, M. & Jurisica, I. 2014. Knowledge Discovery and interactive Data Mining in Bioinformatics - State-of-the-Art, future challenges and research directions. BMC Bioinformatics, 15, (S6), I1, doi:doi:10.1186/1471-2105-15-S6-I1.

- Collective
- Individual
- Tissue
- Cell
- Bacteria
- Virus
- Molecule
- Atom

10^{-12}





The next big issue …

Slide 1-52 Between Standardization and Personalization

EBM CPG

Standardized Medicine

Person-alized Medicine

GBM GPM

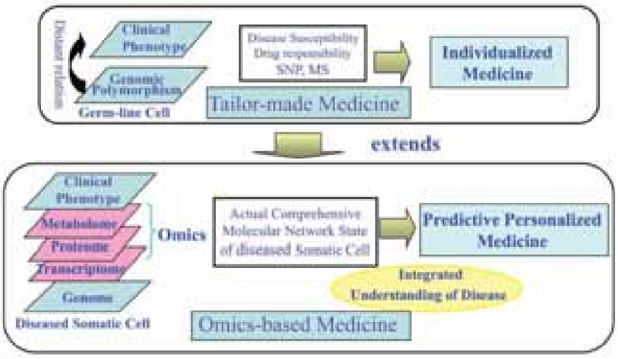
Pervasive Healthcare

Preventive Health Integration

EBM = Evidence Based Medicine
CPG = Clinical Practice Guideline
GBM = Genome Based Medicine
GPM = Genetic Polymorphism

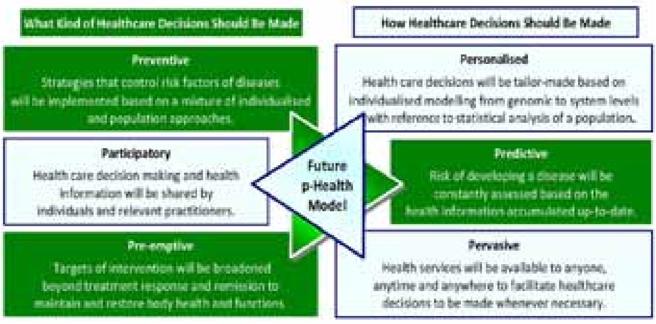
Tanaka, H. (2010)

Slide 1-53: Towards Personalized Medicine



Tanaka, H. (2010) Omics-based Medicine and Systems Pathology A New Perspective for Personalized and Predictive Medicine. *Methods of Information In Medicine, 49, 2, 173-185.*

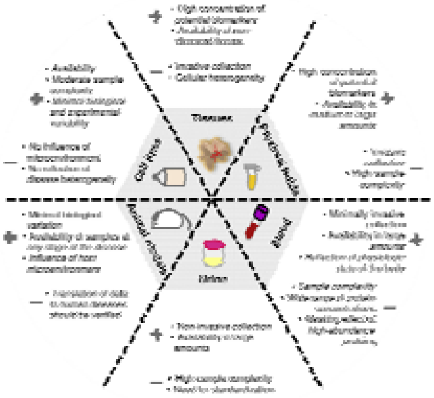
Slide 1-54: Future p-Health Model – A 6 P's paradigm



Zhang, Y. T. & Poon, C. C. Y. (2010) Editorial Note on Bio, Medical, and Health Informatics. *Information Technology in Biomedicine, IEEE Transactions on, 14, 3, 543-545.*

Slide 1-55: Proteomic Samples for Biomarker Discovery

Drabovich, A. P., Pavlou, M. P., Batruch, I. & Diamandis, E. P. 2013. Chapter 2 - Proteomic and Mass Spectrometry Technologies for Biomarker Discovery. *In: Haleem, J. I. & Timothy, D. V. (eds.) Proteomic and Metabolomic Approaches to Biomarker Discovery. Boston: Academic Press, pp. 17-37.*



Coming to the conclusion …

- 1 Heterogeneous data sources
  - need for data **integration**
- 2 Complexity – reduction of search space
  - combining the **best of Human & Computer**
- 3 What is interesting? – and relevant!
  - need of **effective** mapping $\mathbb{R}^N \rightarrow \mathbb{R}^2$
- 4 Clinical time limits "5 Minutes"
  - need of **efficient** solutions

Holzinger, A. & Jurisica, I. 2014. Knowledge Discovery and Data Mining in Biomedical Informatics: The future is in Integrative, Interactive Machine Learning Solutions In: LNCS 8401. Heidelberg, Berlin: Springer, pp. 1-18.

Questions

Sample Questions

08	Biomarkers are measured molecules which indicate the presence of an abnormal condition within a patient, and can be a gene (e.g., SNP), protein (e.g., prostate-specific antigen), or metabolite.	❒ Yes ❒ No	2 total

06	Part of the definition of Biomedical Informatics is the … ❒ … effective use of biomedical data. ❒ … motivation to improve computational capacities. ❒ … effort to expand the technological capabilities. ❒ … motivation to improve human health.	4 total

02	The Von-Neumann Architecture is the fundamental computer organization structure of nearly all of our todays computing systems (e.g. in your PC, smartphone, microwave oven, car, etc.), please roughly sketch the Von Neumann Architecture and indicate the main parts:	1-28 1 each 6 total

Quiz (1/2)

- What encompasses the HCI-KDD approach?
- Why is *understanding intelligence* a grand goal?
- What does heterogeneity of data mean?
- Why is probabilistic information so important?
- What was the groundwork done by T. Bayes?
- Why is the inverse probability important for health informatics?
- What is the big advantage of Gaussian processes?
- What is the grand goal of aML?
- Why is medical decision making so difficult?
- What is the advantage of iML?
- What are the constraints of a human-in-the-loop?

Quiz (2/2)

- Give three examples of where iML is beneficial in health informatics?
- What is the difference between Medical/Biomedical/Health Informatics?
- What are the key problems in health informatics?
- Why is both time and structure so important?
- What is life (in the sense of Erwin Schrödinger)?
- What are the building blocks of life?
- Please define BMI according to the AMIA!
- What are open problems in health informatics?
- What is personalized medicine?
- What is a biomarker? Why are biomarkers important?
- What is the famous time limit to reach a medical decision?

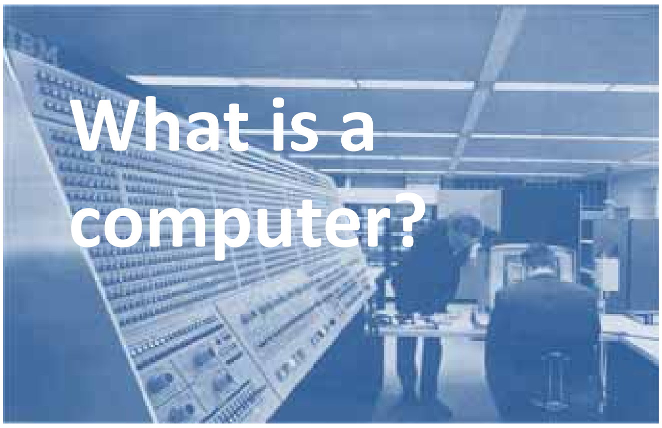
Appendix

Excursus: On the question of "what is information?"

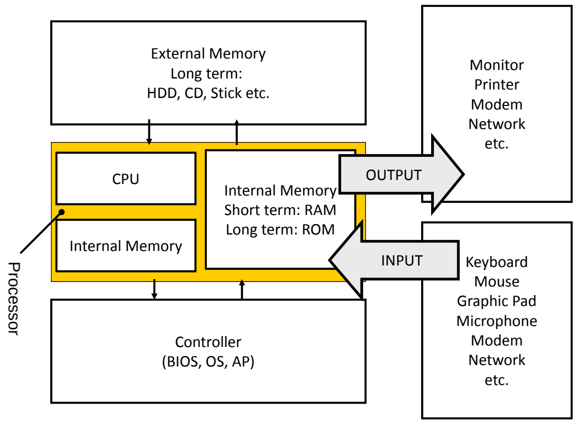
$$\left(-\frac{\hbar^2}{2m}\Delta + U(\vec{r},t)\right)\psi(\vec{r},t) = i\hbar\frac{\partial}{\partial t}\psi(\vec{r},t)$$

From Clinical Medicine to Molecular Medicine


Yapijakis, C. (2009) Hippocrates of Kos, the Father of Clinical Medicine, and Asclepiades of Bithynia, the Father of Molecular Medicine. *In Vivo, 23, 4, 507-514.*



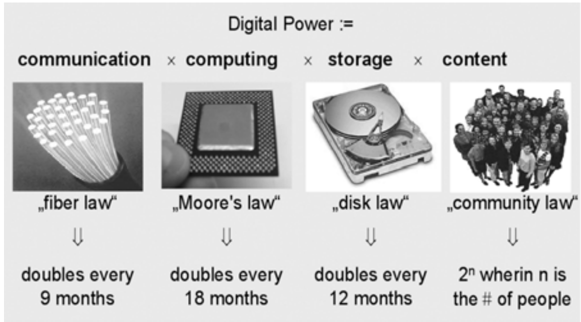
Slide 1-29: Computer: Von-Neumann Architecture



Holzinger (2002), 90-134

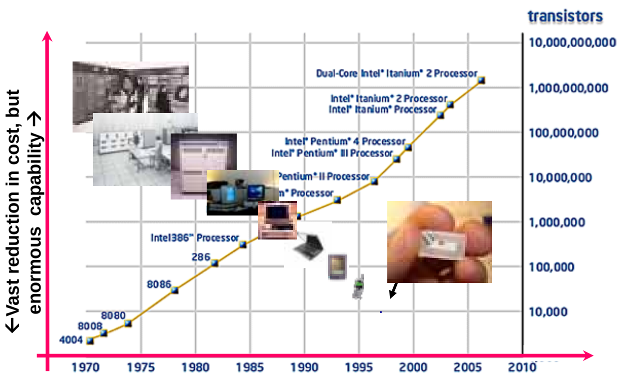
Slide 1-30: Technological Performance / Digital Power

Gordon E. Moore (1965, 1989, 1997)



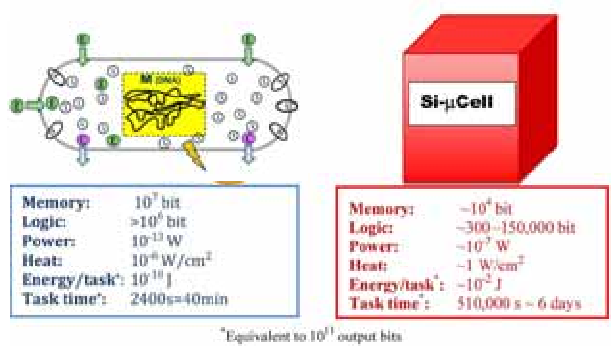
Holzinger, A. 2002. *Basiswissen IT/Informatik Band 1: Informationstechnik. Das Basiswissen für die Informationsgesellschaft des 21. Jahrhunderts, Wuerzburg, Vogel Buchverlag.*

Slide 1-31: Computer cost/size versus Performance



Cf. with Moore (1965), Holzinger (2002), Scholtz & Consolvo (2004), Intel (2007)

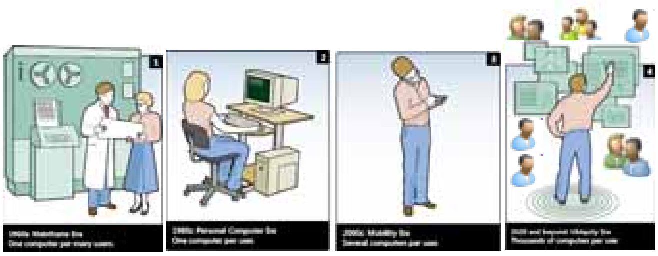
Slide 1-32: Beyond Moore's Law -> biological computing



Cavin, R., Lugli, P. & Zhirnov, V. 2012. Science and Engineering Beyond Moore's Law. *Proc. of the IEEE, 100, 1720-49* (L=Logic-Protein; S=Sensor-Protein; C=Signaling-Molecule, E=Glucose-Energy)

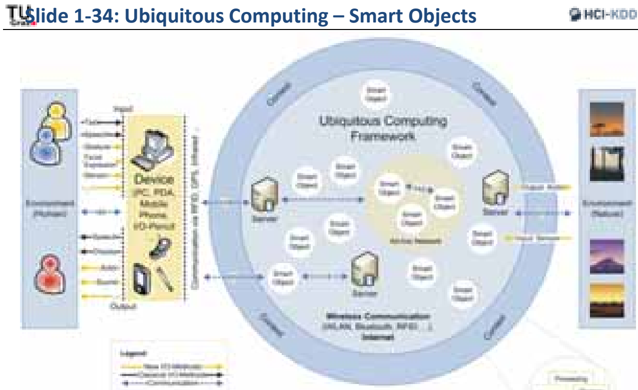
Slide 1-33 From mainframe to Ubiquitous Computing

- … using technology to augment human capabilities for structuring, retrieving and managing information



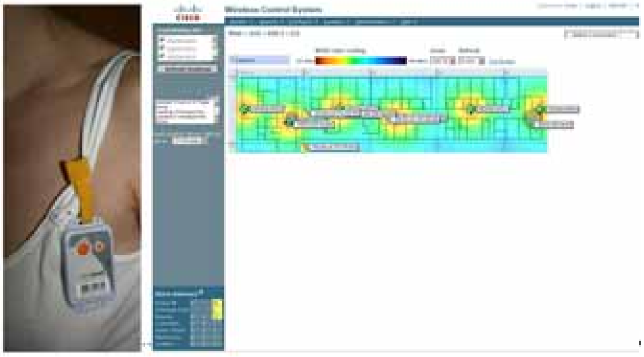
Harper, R., Rodden, T., Rogers, Y. & Sellen, A. (2008) *Being Human: Human-Computer Interaction in the Year 2020. Cambridge, Microsoft Research.*

Slide 1-34: Ubiquitous Computing – Smart Objects



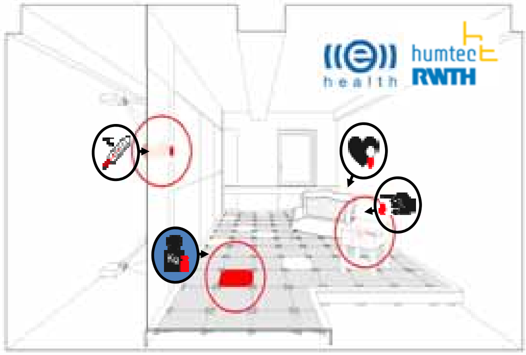
Holzinger, A., Nischelwitzer, A., Friedl, S. & Hu, B. (2010) Towards life long learning: three models for ubiquitous applications. *Wireless Communications and Mobile Computing, 10, 10, 1350-1365.*

Slide 1-35 Example: Pervasive Health Computing



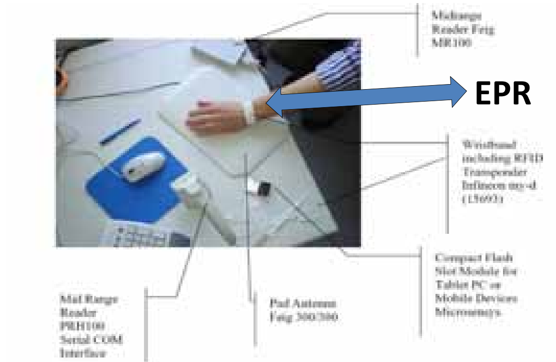
Holzinger, A., Schaupp, K. & Eder-Halbedl, W. (2008) An Investigation on Acceptance of Ubiquitous Devices for the Elderly in an Geriatric Hospital Environment: using the Example of Person Tracking In: *Lecture Notes in Computer Science (LNCS 5105). Heidelberg, Springer, 22-29.*

Slide 1-36: Ambient Assisted Living - pHealth



Alagoez, F., Valdez, A. C., Wilkowska, W., Ziefle, M., Dorner, S. & Holzinger, A. (2010) From cloud computing to mobile Internet, from user focus to culture and hedonism: The crucible of mobile health care and Wellness applications. *5th International Conference on Pervasive Computing and Applications (ICPCA). IEEE, 38-45.*

Slide 1-37: Pervasive Computing in the Hospital



Holzinger, A., Schwaberger, K. & Weitlaner, M. (2005) Ubiquitous Computing for Hospital Applications: RFID-Applications to enable research in Real-Life environments *29th Annual IEEE International Computer Software & Applications Conference (IEEE COMPSAC), 19-20.*

Slide 1-38: Smart Objects in the pathology

Holzinger et al. (2005)

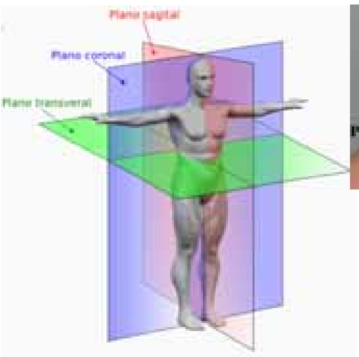
Slide 1-39 The medical world is mobile (Mocomed)

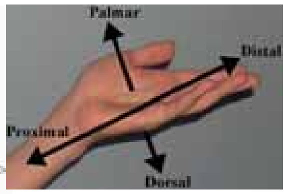
Holzinger, A., Kosec, P., Schwantzer, G., Debevc, M., Hofmann-Wellenhof, R. & Frühauf, J. 2011. Design and Development of a Mobile Computer Application to Reengineer Workflows in the Hospital and the Methodology to evaluate its Effectiveness. *Journal of Biomedical Informatics, 44*, ***968-977.***

Slide 1-40: 1970 - Turning Knowledge into Data

Photo by Institute of Medical Informatics, Graz (1970)

Slide 1-28 Anatomical Axes





Additional Reading

- **History of Probability Theory**
- Franklin, J. *The Science of Conjecture: Evidence and Probability Before Pascal.* John Hopkins University Press, 2001.
- Jaynes, E. T. *Probability Theory: The Logic of Science.* Cambridge University Press, 2003.
- **Probabilistic Reasoning**
- Gigerenzer, G., and D. J. Murray. *Cognition as Intuitive Statistics*. Hillsdale, NJ: Erlbaum, 1987.
- Gilovich, T., D. Griffin, and D. Kahneman, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, 2002.
- Kahneman, D., P. Slovic, and A. Tversky, eds. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press, 1982.
- **Bayesian Networks**
- Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufman, San Mateo, CA, 1988.
- Breese, J. S. "Construction of Belief and Decision Networks." *Computational Intelligence* 8, 4 (1992): 624–647.
- F. Bacchus, A. J. Grove, J. Y. Halpern, and D. Koller. "Statistical Foundations for Default Reasoning." *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI).* Chambery, France, August 1993, pp. 563-569.
- **Multiple-Instance Bayesian Networks**
- Pasula, H., and S. Russell. "Approximate Inference for First-order Probabilistic Languages." *IJCAI-0*1. Seattle, WA, 2001, pp. 741–748.
- Halpern, J. Y. "An Analysis of First-order Logics of Probability." *Artificial Intelligence* 46, 3 (1990): 311–350.
- D. Koller, and A. Pfeffer. "Object-Oriented Bayesian Networks." *Proceedings of the 13th Annual Conference on Uncertainty in AI (UAI)*. Providence, Rhode Island, 1997, pp. 302-313.