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
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
Keywords of Lecture 01

- Big Data
- Life
- Proteins – DNA & RNA – Cell – Tissue – Organ – Cardiovascular Systems
- Medicine – Informatics – Computer
- Personalized Medicine
- Translational Informatics – Data Integration
- Open Medical Data
- Biomarker Discovery
Learning Goals

- At the end of this first lecture you will ...
- ... be fascinated to see our world in data;
- ... have a basic understanding of the building blocks of life;
- ... be familiar with some differences between Life Sciences and Computer Sciences;
- ... be aware of some possibilities and some limits of Biomedical Informatics;
- ... have some ideas of some future directions of Biomedical Informatics;
Bioinformatics = discipline, as part of biomedical informatics, at the interface between biology and information science and mathematics; 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;
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)

- **Oomics 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 in Lecture 01

- AI = Artificial Intelligence
- AL = Artificial Life
- CPG = Clinical Practice Guideline
- CPOE = Computerized physician order entry
- CMV = Controlled Medical Vocabulary
- 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
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, ...);

What is the challenge?

ESO, Atacama, Chile (2011)
Excursus: Two thematic mainstreams in dealing with data ...

- Time
  - e.g. Entropy
  - Dali, S. (1931) The persistence of memory

- Space
  - e.g. Topology
  - Bagula & Bourke (2012) Klein-Bottle
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

Slide 1-7: Social Behavior Contagion Network

Excursus: On the question of “what is information?”

\[
\left(-\frac{\hbar^2}{2m} \Delta + U(\vec{r}, t)\right) \psi(\vec{r}, t) = i\hbar \frac{\partial}{\partial t} \psi(\vec{r}, t)
\]
Living things are able to...

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

Slide 1-11 Building Blocks of Life - Overview

Human eye
Light microscope
Electron microscope
Special

1m 1mm 1μm 1nm 100 pm

### Slide 1-16: Comparison of some current Methods

<table>
<thead>
<tr>
<th>Technology</th>
<th>Sensitivity</th>
<th>Subcellular resolution</th>
<th>Cellular resolution</th>
<th>Minimally invasive?</th>
<th>Live cells?</th>
<th>Real time?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetically encoded nanosensors</td>
<td>Nanomolar to millimolar</td>
<td>Nanometer to millimeter</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>MRI</td>
<td>Mid-micromolar to millimolar (213)</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>PET</td>
<td>1–40 Bq mm⁻² (18)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>X-ray synchrotron</td>
<td>&lt;1 mg kg⁻¹ tissue (transit metals) (204)</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>SIMS</td>
<td>&lt;1 fmol (67)</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>MALDI or TOF imaging</td>
<td>&lt;1 ppm</td>
<td>Yes</td>
<td>50–300 μm (MALDI) 1–2 μm (TOF)</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>NIMS imaging</td>
<td>Yoctomolar (85)</td>
<td>No</td>
<td>50–300 μm</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Mass spectrometry</td>
<td>Yoctomolar</td>
<td>No?</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Raman</td>
<td>50 μM (70)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

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

DNA & RNA – and the five principal bases

(a) Adenosine 5'-monophosphate (AMP)

(b) RNA DNA

The five principal bases

PURINES
- Adenine (A)
- Guanine (G)

PYRIMIDINES
- Uracil (U)
- Thymine (T)
- Cytosine (C)

A, G, T, C are present in DNA (Deoxyribonucleic Acid)
A, G, U, C are present in RNA (Ribonucleic Acid)
Genes contain instructions for making proteins.

Proteins act alone or in complexes to perform many cellular functions.
Slide 1-24: Human Organ Systems
Slide 1-25: Tissue

- **A.** Loose connective tissue (under the skin)
  - cell nucleus
  - collagen fibers
  - elastic fibers

- **B.** Fibrous connective tissue (forming a tendon)
  - cell
  - collagen fibers

- **C.** Adipose tissue
  - fat droplets
  - matrix

- **D.** Cartilage (at the end of a bone)
  - cartilage-forming cells
  - matrix

- **E.** Bone
  - central canal
  - bone-forming cells

- **F.** Blood
  - white blood cells
  - red blood cells
  - plasma
Slide 1-26: Organ Example: Heart

Example:
Heart

- Left common carotid artery
- Brachiocephalic trunk
- Superior vena cava
- Right pulmonary artery
- Right superior pulmonary vein
- Right inferior pulmonary vein
- Sulcus terminalis
- Right atrium
- Inferior vena cava
- Left atrium
- Left superior pulmonary
- Left pulmonary artery
- Oblique vein of left
- Pericardial reflection
- Arch of aorta
- Left subclavian artery
- Coronary sinus
- Left ventricle
- Coronary sulcus and right coronary artery
- Posterior interventricular sulcus and posterior interventricular branch of right coronary artery (posterior descending artery)
- Right ventricle
What is biomedical informatics?
What is a computer?
Slide 1-29: Computer: Von-Neumann Architecture

- **External Memory**
  - Long term: HDD, CD, Stick etc.

- **CPU**

- **Internal Memory**
  - Short term: RAM
  - Long term: ROM

- **Controller**
  - (BIOS, OS, AP)

- **OUTPUT**
  - Monitor
  - Printer
  - Modem
  - Network etc.

- **INPUT**
  - Keyboard
  - Mouse
  - Graphic Pad
  - Microphone
  - Modem
  - Network etc.

- **Processor**

Digital Power :=

communication \times computing \times storage \times content

„fiber law“ \downarrow doubles every 9 months

„Moore's law“ \downarrow doubles every 18 months

„disk law“ \downarrow doubles every 12 months

„community law“ \downarrow 2^n wherein n is the # of people

Vast reduction in cost, but enormous capability

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

Holzinger et al. (2005)
Slide 1-40: 1970 - Turning Knowledge into Data

Photo by Institute of Medical Informatics, Graz (1970)
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, End-User Programmable Mashups
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.

Slide 1-44: Computational Sciences meet Life Sciences

http://www.bioinformaticslaboratory.nl/twiki/bin/view/BioLab/EducationMIK1-2
In medicine we have two different worlds...

Our central hypothesis:
Information bridges this gap

Where is the problem in building this bridge
Volume of Data
High Dimensional
Non-Standardized
Weakly-structured

Slide 1-48 Big Data – We need machine intelligence...

Collective
Individual
Tissue
Cell
Bacteria
Virus
Molecule
Atom

10^{-12}
Open Problems and Future Challenges
1. A unified controlled medical vocabulary (CMV);
2. A complete computer-based patient record that could serve as a regional/national/multinational resource and a format to allow exchange of records between systems;
3. The automatic coding of free-text reports, patient histories, discharge abstracts, etc.;
4. Automated analysis of medical records, yielding
   a) the expected (most common) clinical presentation and course and the degree of clinical variability for patients with a given diagnosis;
   b) the resources required in the care of patients compared by diagnosis, treatment protocol, clinical outcome, location, and physician;
5. A uniform, intuitive, anticipating user interface;
6. The human genome project;
7. A complete three-dimensional, digital representation of the body, including the brain, with graphic access to anatomic sections, etc.;
8. Techniques to ease the incorporation of new information management technologies into the infrastructure of organizations so that they can be used at the bedside or at the research bench;
9. A comprehensive, clinical decision support system.
An update of the list – 20 years later

- Grand new challenges from today’s perspective include:
  - 10. Closing the gap between Science and Practice
  - 11. Data fusion and data integration in the clinical workplace
  - 12. To provide a trade-off between Standardization and Personalization
  - 13. An intuitive, unified and universal, adaptive and adaptable user interface
  - 14. Integrated interactive Knowledge Discovery Methods particularly for the masses of still “unstructured data”
  - 15. Mobile solutions for the bedside and the clinical bench

- A consequence of 14 and 15 will be the vision of “Watson” on the Smartphone. This goal was announced by IBM for the year 2020. The problem involved are the massive unstructured clinical data sets [1]

---

The next big issue ...

Clinical Pharmacology & Therapeutics

nature
International weekly journal of science

Books and Arts

A reality check for personalized medicine

Muin J. Khoury¹, James Evans² & Wylie Burke³

Bringing genetic information into health care is welcome but its utility in the clinic needs to be rigorously reviewed, cautions Muin J. Khoury, James Evans and Wylie Burke.

PERSONALIZED MEDICINE

Genomic information: should it be treated in the same way as X-ray results?
Between Standardization and Personalization

Standardized Medicine

Person-alized Medicine

Pervasive Healthcare

Preventive Health Integration

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

Future p-Health Model – A 6 P’s paradigm

What Kind of Healthcare Decisions Should Be Made

Preventive
Strategies that control risk factors of diseases will be implemented based on a mixture of individualised and population approaches.

Participatory
Health care decision making and health information will be shared by individuals and relevant practitioners.

Pre-emptive
Targets of intervention will be broadened beyond treatment response and remission to maintain and restore body health and functions.

How Healthcare Decisions Should Be Made

Personalised
Health care decisions will be tailor-made based on individualised modelling from genomic to system levels with reference to statistical analysis of a population.

Predictive
Risk of developing a disease will be constantly assessed based on the health information accumulated up-to-date.

Pervasive
Health services will be available to anyone, anytime and anywhere to facilitate healthcare decisions to be made whenever necessary.

What price health?
Thank you!
### Sample Questions

<p>| 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 | 2 total |
| 06 | Part of the definition of Biomedical Informatics is the ... |   | 4 total |
|    |   |  - ... effective use of biomedical data. |   |
|    |   |  - ... motivation to improve computational capacities. |   |
|    |   |  - ... effort to expand the technological capabilities. |   |
|    |   |  - ... motivation to improve human health. |   |</p>
<table>
<thead>
<tr>
<th>References (1/5)</th>
</tr>
</thead>
</table>
References (2/5)

References (3/5)

References (4/5)


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.


