Course: Informatics and Data Science
Year: 4th, 2nd semester
Credits: 4 (24 x 2 hours = 48 hours)

Course Overview
This course introduces the fields of informatics and data science, emphasizing their relevance to clinical practice, medical research, and personalized medicine. The course will cover fundamental concepts and tools of bioinformatics, data science and artificial intelligence (AI), and their applications in the clinics.

Learning Objectives
By the end of this course, students will be able to:

1. Understand the relevance of informatics and data science in modern healthcare and medical research.
2. Utilize public biological databases to extract relevant clinical and molecular information.
3. Interpret bioinformatics data, focusing on personalized medicine applications.
4. Understand concepts and techniques in data science, machine learning and AI techniques for analyzing clinical and molecular data.
5. Perform data cleaning and visualization in the biomedical context.
6. Understand how medical informatics systems, such as clinical decision support systems, are integrated into clinical workflows.
7. Address ethical considerations in the use of bioinformatics and patient data, including privacy, consent, and data security.
8. Appreciate the applications and the future role of data science in clinical care and biomedical research.

Course Outline
<u>Module 1: Bioinformatics (6 lessons, Charlotte Ng)</u>
Lesson 1: Course introduction & introduction to bioinformatics
- Definition, historical development and scope of bioinformatics
- Relevance of bioinformatics to medical practice and research
- Key databases: NCBI, Uniprot, ClinVar etc

Lesson 2: Genomics
- Applications of genomics
- Overview of steps in genomics sequencing data analysis
- Useful web-based tools

Lesson 3: Transcriptomics
- RNA-seq and applications
- Overview of steps in RNA-seq analysis
- Downstream analysis: differential expression analysis, class discovery, pathway analysis

Lesson 4: Epigenomics
- Applications of epigenomics
- Omics technologies to study the epigenome
- Analysis methods to define DNA methylation and chromatin modifications

Lesson 5: Single-cell and spatial omics
- Bulk vs single-cell RNA-seq vs spatial transcriptomics
- Applications of single-cell and spatial omics
- Fundamentals of data analysis and data interpretation

Lesson 6: Structural bioinformatics
- Protein and RNA structure & structure prediction
- Protein-protein interactions, molecular dynamics and drug discovery

Module 2: Data science, machine learning and AI (8 lessons, Giuseppe Jurman)

Lesson 7: Introduction to data science in healthcare
- Overview of data science principles
- Data science and its relevance to medicine
- Key components: data handling, exploration, visualization, and AI applications in medicine

Lesson 8: Data wrangling and cleaning in healthcare
- Importance of clean data in medical research
- Handling missing data and outliers in clinical datasets
- Simple data cleaning and interpretation

Lesson 9: Data visualization in healthcare
- Best practices for visualizing medical data
- Tools for creating visual representations of clinical data
- Interpreting visual data for patient insights

Lesson 10: Introduction to machine learning concepts
- Types of machine learning: supervised vs. unsupervised
- Applications of machine learning in healthcare
- Reviewing case studies on AI applications

Lesson 11: Supervised learning for medical data
- Regression and classification in medical datasets
- Applications in disease prediction and diagnostics
- Using ML models to analyze clinical data

Lesson 12: Unsupervised learning and clustering
- Clustering techniques (e.g., k-means) for patient stratification
- Use cases in healthcare (e.g., grouping patient populations)
- Interpreting clustering results in medical contexts
- The concept of dimensionality reduction

Lesson 13: Introduction to deep learning in medicine
- Neural networks and their applications in radiology and genomics
- Overview of tools and platforms (without coding)
- AI in medical imaging diagnostics
- Hints of generative AI & foundation models

Lesson 14: Problems in AI/ML and synthetic data
- Explainability, reproducibility, dataset shift, data leakage etc
- Data poverty and the concept of synthetic data
- Synthetic data in Biomedical Imaging, Omics and EHRs

Module 3: Informatics in the Clinic (4 lessons, guest lecturers)
Lesson 15: Imaging informatics (Politi)
- AI in clinical practice: from image acquisition to automated reporting
- Application of ML and CNN to medical images
- Implementation and integration of AI in the clinical practice

Lesson 16: Telemedicine and digital health tools (Greco)
- Telemedicine technologies and their impact on healthcare
- Wearable devices and data-driven health monitoring
- Case studies/examples: Telemedicine in practice

Lesson 17: Digital pathology and AI in the clinical practice (Di Tommaso)
- Digital pathology: tools for image analysis and disease detection
- AI in pathology: how and when
- in the clinical practice: AI tool for screening prostatic and gastric biopsies

Lesson 18: Clinical decision support systems (CDSS, Hassan)
- Overview of CDSS and applications in healthcare
- Supporting diagnostics and treatment decisions
- Case studies/examples on CDSS in practice

Module 4: Informatics and data science in healthcare and research (6 lessons, Charlotte Ng)
Lesson 19: Big data in healthcare and research
- Diverse collections of structured and unstructured data in healthcare
- Challenges and opportunities with large-scale medical datasets
- Examples of big data projects in medicine

Lesson 20: Personalized medicine and pharmacogenomics
- How bioinformatics and data science are transforming personalized medicine
- Examples of pharmacogenomics in clinical practice

Lesson 21: Precision oncology
- The role of genomic testing on patient management and clinical trial design
- Case study: non-small cell lung cancer

Lesson 22: Integrating data science into clinical workflows
- Challenges in integrating (bio)informatics and data science into daily medical practice
- Examples of real-world clinical implementations

Lesson 23: Data privacy and security
- Privacy concerns with genomic data
- Data security and patient consent

- Ethical dilemmas in AI-based medical systems

Lesson 24: Future directions in bioinformatics and data science
- Emerging trends in AI, genomics, and digital health
- Course wrap-up

Assessment
Written exam of 32 multiple choice questions. Each correct answer is worth 1 point. No penalties will be applied for incorrect answers. Passing score is 18 points. Cum laude is 30 points or more. Time limit: 45 minutes