BE700 A3: AI and Systems Biology of Disease

SPRING 2024

Tu,Thu 9-10.45AM (MORNING)

We are fully committed to deploying AI responsibly and safely in the service of social justice, inclusion, fairness, reducing social inequalities and diversity.

(This course is different from the BE700 A4: SEMINAR in AI and Biomedical Science and Engineering, taught at 3.30 – 5.15 pm (which is full but available by permission of instructor.

Instructor: Professor Simon Kasif

Time: Tu -Thu 9– 10.45AM

Approved by Petition as BME and Bioinformatics Elective. Also available to seniors with interest in AI and relevant background.

Course Outline:

The class will include many personalized group sessions with the instructor.

Prerequisites: **Senior or Graduate Student Standing** in BME, Engineering, Bioinformatics, CS, DS, Physics, Biology, Medicine, Public Health, Chemistry. **For disciplines, different from BME and Bioinformatics such as medicine, biology, chemistry, physics, engineering or computer science**

permission of instructor is required

AI and data science are playing an increasingly important role in biomedical sciences and engineering. This course will teach students to apply or develop AI and machine learning concepts to probe into the **systems biology** of disease and **personalized medicine.** The projects in previous semesters were focused on COVID-19, cancer, aging and diabetes. In the AI area, the emphasis would be on machine learning!

If you are a senior with a high GPA please write to Professor Simon Kasif (kasif@bu.edu) (prior to registration). Seniors with this profile have done exceptionally well in this class.

The course will cover computational frameworks such as **biological networks** (including metabolic, regulatory and signal transduction networks), gene expression analysis, proteomic analysis, next-generation sequencing, **AI and machine learning (ML)**, elementary genetics, **pathway modeling and analysis** and other omics technologies to focusing on clinical problems such as cancer, diabetes, inflammation, aging and personalized medicine. The emphasis is towards developing diagnostics, prognostics or drug development.

There are no exams and grading is based on **bi-weekly homework, reading research papers, class presentations and a team project.** The main aim of this course is to cover general concepts in biological computing that provide the foundation of thinking computationally about anomalous behavior in biological systems relevant to disease mechanisms, **systems biology of disease,** diagnosis, prognosis, **personalized medicine** or **network based drug design**. The course also aims to teach students to work in research teams and develop the skills to plan and coordinate a scientific

project. This semester we will have covid-19 projects in class.

During the semester we will also have guest lectures.

For information please contact:

Professor Simon Kasif: kasif@bu.edu

Goals

The main aim is to prepare students to apply and develop new concepts in integrative and systems biology of human disease and use **AI/ML methods to analyze biomedical data and design biomedical experiments.** This involves developing a familiarity with current high-throughput omics technologies, probing the complex systems biology of disease using these biotechnologies: storing, querying and manipulating massive amount of data, performing AI analysis of clinically relevant integrative data, producing models of systems across scales, capturing anomalous behavior in biological networks and making and validating predictions made by these network models and AI methods.

Tentative Syllabus

This course will cover many of the widely used techniques used for networks and analysis, systems biology and network modeling of biological systems in the context of disease anomalies, focusing on detecting dis- regulated networks in disease and identification of novel drug targets, drug repositioning and diagnostics.

The class will have a number of invited speakers covering disease biology, machine learning, AI and network analysis algorithms.

Previous semesters the class focused on cancer, aging, wellness

and diabetes and COVID-19.

No exams: reading assignments, write-ups of papers and class presentations, lab homework (applying machine learning tools, final group project.

We will accommodate students from biology without programming experience. They will be able to use widely available tools.

Thus, different homework (presentation, and write-up) criteria would be applied to graduate students, undergraduate students and students from outside engineering (e.g. biology or chemistry). We will try to form groups from different disciplines for projects.

AI, Machine Learning, Network Analysis Methodologies

Lecture 1: Introduction to Machine Learning and AI (more generally)

Supervised Learning Methods

Lecture 2: Nearest Neighbor Methods

Lecture 3: Introduction to Decision Trees

Lecture 4: Perceptron Learning and Support Vector Machines

Lab Sessions WEKA1, WEKA2 and Reporting Accuracy (ROC, AUC, FDR, Precision-Recall)

Un-Supervised Learning Methods

Lecture 4. Clustering / Bi-clustering lecture Spectral Graph Clustering Methods, SVD, PCA (principal component analysis)

** advanced topic **: Semi-supervised learning and network learning

Lab Session (Using Gene Pattern)

Lecture 5: Using WEKA, a machine learning environment,

Lecture 6. Principles of General Algorithm Design for Learning and Optimization

Lecture 7. Ensemble Learning and Wisdom of the Crowds

Learning in Networks

Graphical Models

Lecture 1: Probabilistic network (graphical models), Mutual Information, Learning Probabilistic Networks (models for biological systems)

Lecture 2: Network Discovery Algorithms Lab Session:

Neural Network Models

Lecture 3. Backpropagation Learning in Neural Networks

Lecture 4. Principles of learning in Deep Neural Networks: stochastic gradient descent, large scale optimization and more.

Lecture 5. Hopfield Networks

Computational Systems Biology of Disease

Introductory Cancer Lectures

Lecture 1: Introduction to Cancer:

Lecture 2: Cancer expanded: One Renegade Cell or a Systemic Process?

Lecture 3a: Biological Networks from E.coli to Human Tissue

Lecture 3b: Broad Introduction to Transcriptional Regulation Networks in Cancer and Signaling Networks in Cancer

Lecture 4: Personalized Medicine: Subtypes of Cancer and Clinically Significant Cancer Phenotypes: Survival and Drug Response

Lecture 5: Genomics and Cancer: mutations, copy number variation, epigenomics.

Lecture 6: Advanced: Epigenomics of Stem Cells and Cancer Stem Cells.

Lecture 7: Diabetes

Lecture 8: Aging and Aging Genes

Biological Networks

Lecture 1: Introduction to Networks and Graphs

Lecture 2: Regulatory Networks

Lecture 3: Protein-Protein Interaction Networks (scale free networks)

Lecture 4: Signal Transduction Networks.

Lab Session: Gene Ontology and DAVID Lab Session:

Bayes Networks tools, Dynamic Network Simulations, Rudimentary Generative Learning

Personalized Medicine and Network Signatures of Disease

Introduction to Biomarkers

Lecture 1: Pathways Signatures

Lecture 2: Network Signatures

Lecture 3: Genomic Signatures and Cancer: mutations, copy number variation, epigenomics.

Lecture 4: Cancer Data Analysis Subtypes of Cancer Drug Response Signatures Survival / Metastasis Signatures

Network Signatures for Diabetes, Aging and Cancer

Lecture 5: Detailed Classification with Microarrays, RNA SEQ, Single Cell RNA seq, classification with genomic markers

Lecture 6: Introduction to TCGA (Cancer Atlas)

Lab Session: working with Cytoscape and GenePattern

Lab Session: working with Gene Network Enrichment Analysis

Personalized Medicine and Drugs

Drug Response Signatures using supervised learning NCI-60 CMAP

Project Planning and Execution (instead of exam)

Applying machine learning to current problems.

COVID-19 DRUG REPOSITIONING, AGING, CANCER, DIABETES

Optional topics:

Metabolic Diseases

Metabolic Networks & Insulin Signaling, Diabetes

Inflammation (immune response)

Innate Immune Response

Adaptive Immune Response

Lab Session: working with GNEA (detection of inflammation at the molecular network level)

Aging

Conserved Signaling Networks and Longevity

Lab Session: working with Aging Mouse Models

Additional Labs (as needed)

Probability

Bayes Law, Conditional and Joint Probabilities

Stat. (basic tests, permutation tests, ranked tests)

Gene Pattern: permutations / diagnostic genes and classification

Accuracy/AUC

DAVID and Gene Set Enrichment:

GSEA

WEKA 1 WEKA 2

Gene Network Analysis and Enrichment: GNEA

Bayes Networks / Graphical Models, Network Discovery

Algorithms (THE DREAM COMPETITION)