

Course on “Application of Machine Learning in Biological Systems” (ES60011)

Slot E: WED(12:00-12:55) , THURS(11:00-11:55) , FRI(09:00-09:55)

Faculty members teaching this course:

Dr. Pralay Mitra, (Dept. of Computer science & Engineering)

Dr. Amit Ghosh (School of Energy Science and Engineering)

Contents	Lectures
1. Introduction to Machine-learning (ML): Introduction to learning techniques	(2 lectures)
2. Python programming boot camp: Introduction to Python, UNIX shell and jupyter notebooks, Functions and Control Flow, Data Structures, File Manipulations and Modules, Biopython and System Calls	(4 lectures)
3. Exploratory data analysis: Plotting and visualization of biological data, Handling big-data sets in biology: existing methods and challenges	(4 lectures)
5. Single-cell RNA-seq analysis: RPKM, TPM, CPM, Isoform Inference, RNA-seq Alignment and QC, Sequencing read distribution	(4 lectures)
6. Machine learning for multiomic data: Constraint-based analysis of metabolic networks, Supervised fluxomic analysis, Unsupervised fluxomic analysis, Supervised multiomic analysis; Condition-specific constraint-based models, Elementary Flux Modes, GIMME, iMAT	(4 lectures)
7. Predictive engineering and metabolic pathway optimization: multi-gene pathway optimization (MiYA) for β -carotene production using NN, Automated Recommendation Tool (ART), Probabilistic ensemble model, Bayesian optimization for tryptophan production, Ensemble model for dodecanol production	(6 lectures)
8. Bioprocess control & optimization: Bioprocess optimization of 27 batch fermentation for xylitol production, Process control, Model Predictive Control (MPC), Policy gradient parameterized by neural network	(4 lectures)
9. Classifying cancer sub-types using clustering	(3 lectures)
10. Diagnosing breast cancer from biopsy images using NN:	(3 lectures)
11. Predicting cardiovascular risk using regression	(2 lectures)
Total	36 lectures

Books and References:

Books

1. Maurice E. Cohen, Donna L. Hudson. (1999) *Neural Networks and Artificial Intelligence for Biomedical Engineering: 3* (IEEE Press Series on Biomedical Engineering)
2. Tony J. Cleophas and Aeilko H. Zwinderman. 2015. *Machine Learning in Medicine - a Complete Overview*. Springer.
3. Stuart Russell and Peter Norvig. 2009. *Artificial Intelligence: A Modern Approach* (3rd ed.). Prentice Hall Press, Upper Saddle River, NJ, USA
4. Palsson, Bernhard O. *Systems Biology: Properties of Reconstructed Networks*. New York: Cambridge University Press, 2006.

Further reading and references

1. Li Yang, Haibin Wu, Xiaoqing Jin, Pinpin Zheng, Shiyun Hu, Xiaoling Xu, Wei Yu & Jing Yan (2020) Study of cardiovascular disease prediction model based on random forest in eastern China. *Scientific Reports*. 10: 5245
2. Kim, S. et al. (2015). A method for generating new datasets based on copy number for cancer analysis. *BioMed Research International*, 2015.
3. Geman, D. et al. (2004) Classifying gene expression profiles from pairwise mRNA comparisons. *Statistical Applications in Genetic and Molecular Biology*, 3.
4. Zhou, Y. et al. (2018). MiYA, an efficient machine-learning workflow in conjunction with the YeastFab assembly strategy for combinatorial optimization of heterologous metabolic pathways in *Saccharomyces cerevisiae*. *Metab. Eng.* 47, 294–302.
5. Zhang, J et al. (2020). Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism. *Nat. Commun.* 11, 4880.