|  |  |
| --- | --- |
| MONO | **DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**  **NATIONAL INSTITUTE OF TECHNOLOGY PATNA**  Ashok Raj Path, PATNA 800 005 (Bihar), India |
| Phone No.: 0612 – 2372715, 2370419, 2370843, 2371929, 2371930, 2371715 Fax – 0612- 2670631 Website: [www.nitp.ac.in](http://www.nitp.ac.in/) |

***CSX443 Computational Biology***

**L-T-P-Cr: 3-0-0-3**

**Pre-requisites:** Fundamental knowledge of algorithms, prior completion of “Machine Learning” course is preferred.

**Objectives/Overview:**

* To understand necessary basic biology concepts.
* To learn common approaches for sequence analysis.
* To learn common techniques in phylogenetic analysis.
* To learn common approaches for analysing protein interaction data.

**Course Outcomes:**

At the end of the course, a student should:

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Outcome** | **Mapping to POs** |
|  | Have knowledge of fundamental concepts of biology required for appreciating computational biology problems. | PO1, PO12 |
|  | Be able to align simple sequences using dynamic programming approaches. | PO2, PO3 |
|  | Understand popular subsequence analysis approaches and feature extraction approaches. | PO3, PO4 |
|  | Understand concept of phylogenetic trees and common approaches to construct them. Should be able to construct trees from sample datasets. | PO2, PO3, PO4 |
|  | Understand common phylogenetic analysis methods and be able to apply them on sample datasets. | PO2, PO3 |
|  | Understand what are protein interaction networks and their use in protein function prediction. | PO1, PO12 |
|  | Have knowledge of common approaches for protein function prediction. | PO2, PO3, PO12 |
|  | Be familiar with common algorithms for predicting protein-protein interaction cites. | PO2, PO3 |

**UNIT I: Basic Biology Lectures: 8**

Importance of Bioinformatics, DNA, Transcription, RNA, Translation, Protein; Coding and non-coding portions of gene Biological data types, popular biological databases: NCBI, GenBank, Swiss-Prot.

**UNIT II: Sequence Analysis Lectures: 12**

Representing sequences in Computer: FASTA and PDB formats, Sequence alignment methods, local and global alignment, Smith-Waterman algorithm, Needleman-Wunsch Algorithm, BLAST, Multiple Sequence Alignment, Motifs/Domain Analysis: Introduction. Subsequence Analysis Algorithms, Feature Based Approaches: Bag-of-words model, Tf-Idf.

**UNIT III: Computational Phylogenetics Lectures: 12**

Concept of trees- Computer representation of phylogenetic trees, Popular tree formats, Deriving trees: Distance matrix methods, Character based methods. Solving UPGMA, NJ and small parsimony problems, Phylogenetic Analysis methods: Boot-Strapping, jack-knifing. Phylogenetic consensus methods: MRT, Strict consensus.

**UNIT IV: Protein Interaction Network Lectures: 8**

Introduction: Importance of Protein Interaction Networks, Existing approaches for predicting protein function from protein interaction network: Neighborhood- based Approach, Global Optimization Based Approaches, Clustering-Based Approach, Association Analysis Based Approach, Computational Methods for Prediction of protein-protein interaction site: Machine learning based techniques using – Decision trees, Support Vector Machines (SVM).

**Text/Reference Books:**

1. Bioinformatics: A Primer - P. Narayanan.
2. Bioinformatics: Sequence and Genome Analysis - D. W. Mount
3. An introduction to bioinformatics algorithms - Neil C. Jones, Pavel A. Pevzner.
4. Protein-Protein Interactions: Computational and Experimental Tools - Weibo Cai & Hao Hong.
5. Molecular Biology of the Gene – J. D. Watson.
6. Bioinformatics: The Machine Learning Approach – Pierre Baldi.
7. Pandey, Gaurav, Vipin Kumar, and Michael Steinbach. "Computational approaches for protein function prediction: A survey." *Twin Cities: Department of Computer Science and Engineering, University of Minnesota* (2006).
8. Immunological Bioinformatics**:** Ole Lund, Morten Nielsen, Soren Brunak, Claus Lundegaard, Can Kesmir, MIT Press, 2005.