राष्ट्रीय प्रौद्योगिकी संस्थान पटना / NATIONAL INSTITUE OF TECHNOLOGY PATNA
संगणक विज्ञान एंव अभियांत्रिकी विभाग / DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
अशोक राजपथ, पटना-८००००५, बिहार / ASHOK RAJPATH, PATNA-800005, BIHAR
Phone No.: 0612-2372715, 2370419, 2370843, 2371929 Ext- 200, 202 Fax-0612-2670631 Website: www.nitp.ac.in

No:- **Date:**

CS041902: ***Bioinformatics*** **L-T-P-Cr: 3-0-0-3**

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

Objectives:
- To understand common biology concepts in Bioinformatics.
- To learn common approaches for sequence analysis.
- To learn common techniques in phylogenetic analysis.
- To learn common approaches for analyzing protein interaction data.

At the end of the course, a student should:

Sl. No	Outcome	Mapping to POs
1.	Understand fundamentals concepts of Biology	PO1, PO2
2.	Able to understand and implement the algorithms of sequence analysis	PO3, PO4, PO9
3.	Understand the methods of phylogenetic analysis	PO1, PO2
4.	Able to understand protein interaction data	PO2, PO4
5.	Able to apply machine learning models on biological problems	PO12, PO3, PO4, PO5

UNIT I: Biology in Bioinformatics **Lectures: 08**

Importance of Bioinformatics; Central Dogma of Molecular Biology; DNA, Transcription, RNA, Translation, Protein; Coding and non-coding portions of gene: Exons and Introns. Biological data types, some important biological databases: NCBI, GenBank, DDBJ, EMBL, Swiss-Prot, RCSB-PDB.

UNIT II: Sequence analysis **Lectures: 10**

A) Representing sequences in Computer: Sequence file formats: FASTA and PDB formats.
B) Sequence alignment methods, local and global alignment, Dynamic programming based approaches: Smith-Waterman algorithm, Needleman-Wunsch Algorithm, Details of BLAST, Multiple Sequence Alignment: methods, tools, and techniques.
C) Motifs/Domain Analysis: Introduction. Subsequence Analysis Algorithms: Gibbs Algorithm, Expectation Maximization Algorithm, Kernel Methods.
D) Feature Based Approaches: Bag-of-words model, Tf-Idf.

E) Case Study: Protein function prediction based on protein sequences

UNIT III: Computational Phylogenetics **Lectures: 10**

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

UNIT IV: Protein Interaction Network **Lectures: 10**

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

UNIT V: Systems Biology **Lectures: 4**

Gene Systems, Protein Systems, Metabolic System, Signaling Systems, Design of Biological System

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
9. Systems Biology: A Text Book, Edda Klipp et al.