

# राष्ट्रीय प्रौद्योगिकी संस्थान पटना / NATIONAL INSTITUE OF TECHNOLOGY PATNA संगणक विज्ञान एंव अभियांत्रिकी विभाग / DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

संगणक विज्ञान एंव अभियात्रिकी 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: <u>www.nitp.ac.in</u>

No:-	Date:

CSX4176: Bioinformatics

L-T-P-Cr: 2-0-2-3

**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 analysing protein interaction data.

## **Course Outcomes**

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

Sl. No.	Course Outcome (CO)	Program Outcome (PO)
1	Explain the fundamental concepts of molecular biology	
	relevant to bioinformatics, including the central dogma and	
	the structure and function of DNA, RNA, and proteins.	PO1, PO2
2	Analyze biological data formats like FASTA and PDB and	
	apply sequence alignment techniques (local and global)	
	using algorithms like Smith-Waterman and	
	Needleman-Wunsch. Additionally, utilize tools like BLAST	
	for sequence similarity searches.	PO1, PO2, PO5
3	Apply multiple sequence alignment methods and interpret	
	the results to identify conserved motifs and domains within	
	protein sequences.	PO2, PO4
4	Implement feature-based approaches like Bag-of-Words	
	and Tf-ldf for analyzing biological sequences and utilize	
	these techniques for tasks like protein function prediction.	PO2, PO3, PO5
5	Construct phylogenetic trees using distance matrix and	
	character-based methods (UPGMA, NJ, Parsimony) and	
	evaluate the robustness of these trees using bootstrapping	
	and jackknifing techniques.	PO2, PO4, PO5
6	Analyze protein-protein interaction networks and apply	
	computational methods for predicting protein interactions	
	and protein-protein interaction sites using machine	
	learning techniques like decision trees and support vector	
	machines (SVMs).	PO2, PO3, PO5

7	Integrate	knowledge	from	various	bioinfo	ormatics	
	techniques	(sequence	analysis,	phylogei	netics,	protein	
	interaction networks) to address biological problems.					PO2, PO3, PO11	

# **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: 16** 

- 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:14

- 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)

#### **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