I A R E

INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)
Dundigal - 500 043, Hyderabad, Telangana

COURSE CONTENT

<table>
<tr><th colspan="10">BIO-INFORMATICS</th></tr>
<tr><td colspan="10">VII Semester: CSE | IT</td></tr>
<tr><th rowspan="2">Course Code</th><th rowspan="2">Category</th><th colspan="3">Hours/Week</th><th>Credits</th><th colspan="3">Maximum Marks</th></tr>
<tr><th>L</th><th>T</th><th>P</th><th>C</th><th>CIA</th><th>SEE</th><th>Total</th></tr>
<tr><td>ACSD52</td><td>Elective</td><td>3</td><td>0</td><td>0</td><td>3</td><td>40</td><td>60</td><td>100</td></tr>
<tr><td>Contact Classes: 48</td><td>Tutorial Classes: Nil</td><td colspan="4">Practical Classes: Nil</td><td colspan="3">Total Classes: 48</td></tr>
<tr><td colspan="10">Prerequisite: Programming for Problem Solving</td></tr>
</table>

I. COURSE OVERVIEW:

The course introduces the fundamental concepts of bio-informatics with emphasis on computational methods used to store, analyze, and interpret biological data. Topics covered include biological databases, sequence analysis, genomics, proteomics, phylogenetics, and applications of machine learning in biology. The main objective of the course is to teach students how computational techniques can be used to understand biological problems and develop solutions in areas such as drug discovery, genome analysis, and personalized medicine.

II. COURSE OBJECTIVES:

The students will try to learn:

I The basic principles of molecular biology, biological data types, and their importance in computational analysis.

II The skills to use biological databases and tools for sequence analysis, alignment, and annotation.

III Techniques used in genomics and proteomics including sequence alignment, genome mapping, and protein structure prediction.

IV Computational algorithms such as dynamic programming, Hidden Markov Models, phylogenetic methods, and machine learning approaches.

V Applications of bioinformatics in drug design, systems biology, and personalized medicine essential for future research and development in computational biology.

III. COURSE OUTCOMES:

At the end of the course students should be able to:

CO1 Describe various biological databases and retrieve genomic and proteomic information.

CO2 Apply sequence alignment techniques for DNA/protein sequence comparison.

CO3 Analyze genomic and proteomic data using computational tools and algorithms.

CO4 Construct phylogenetic trees and interpret evolutionary relationships.

CO5 Utilize bioinformatics methods for drug discovery, docking, and disease analysis.

CO6 Compare various bioinformatics algorithms in terms of implementation, accuracy, and computational performance.

IV. COURSE CONTENT:

MODULE – I: INTRODUCTION TO BIO-INFORMATICS (09)

Introduction to bioinformatics, scope and applications; Molecular biology fundamentals: DNA, RNA, proteins, gene expression; Biological data types – genomic, proteomic, metabolomic; Biological file formats – FASTA, GenBank, PDB; Biological databases – NCBI, EMBL, DDBJ, UniProt, PDB, KEGG; Introduction to sequence analysis and data retrieval; Role of computer science in biology.

MODULE - II: SEQUENCE ALIGNMENT AND ANALYSIS (09)

Sequence similarity, identity, homology; Pairwise alignment – global (Needleman–Wunsch) and local (Smith–Waterman); Scoring matrices – PAM, BLOSUM; Multiple sequence alignment – ClustalW, MUSCLE, MAFFT; Gene and protein annotation; BLAST, FASTA tools; Applications of sequence alignment in genome analysis.

MODULE - III: GENOMICS AND PROTEOMICS (09)

Genomics: Genome sequencing technologies – Sanger, NGS, Illumina, Nanopore; Genome assembly – de novo and reference-based assembly; Gene prediction algorithms; Comparative genomics.

Proteomics: Protein structure – primary to quaternary; Structure prediction – homology modeling, threading, ab initio; Protein modeling tools – SWISS-MODEL, MODELLER; Protein-protein interaction networks; Structural bioinformatics.

MODULE – IV: COMPUTATIONAL METHODS & PHYLOGENETICS (09)

Dynamic programming principles in bioinformatics; Hidden Markov Models – gene finding, sequence labeling; Phylogenetic analysis – distance-based, character-based methods; Phylogenetic tree construction – UPGMA, Neighbor-Joining, Maximum Parsimony, Maximum Likelihood; Evolutionary models; Biological data mining; Machine learning applications in genomics and proteomics.

MODULE – V: APPLICATIONS OF BIO-INFORMATICS (09)

Drug discovery and design – docking basics, scoring functions, pharmacophore modeling; Systems biology – networks, pathways; RNA-Seq and microarray analysis; Personalized medicine & precision genomics; Tools and platforms – PyMOL, Galaxy, HMMER, GROMACS; Ethical issues, data privacy in bioinformatics; Latest trends – AI in protein prediction (AlphaFold), CRISPR informatics.

V. TEXT BOOKS:

1. Arthur M. Lesk, “Introduction to Bioinformatics”, Oxford University Press.
2. Jin Xiong, “Essential Bioinformatics”, Cambridge University Press.

VI. REFERENCE BOOKS:

1. Dan E. Krane & Michael L. Raymer, “Fundamental Concepts of Bioinformatics”, Pearson.
2. David Mount, “Bioinformatics: Sequence and Genome Analysis”, CSHL Press.
3. Andreas D. Baxevanis & Francis Ouellette, “Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins”, Wiley.

VII. ELECTRONICS RESOURCES:

1. https://www.ncbi.nlm.nih.gov
2. https://www.uniprot.org
3. https://www.ebi.ac.uk
4. https://www.kegg.jp
5. https://www.rcsb.org

VIII. MATERIAL ONLINE:

1. Course template
2. Tech-talk topics
3. Assignments
4. Definition and terminology
5. Tutorial question bank
6. Model question paper – I

7. Model question paper – II
8. Lecture notes
9. Early lecture readiness videos (ELRV)