

Course Structure: M. Sc. (Bioinformatics)

Department of Bioinformatics, Singhania University

Course Code	Course Title	Internal Assessment	Semester Exam	Total
Semester - I				
MSBI-111	Introduction of Bioinformatics	30	70	100
MSBI-121	Statistics and Probability	30	70	100
MSBI-131	Molecular Biology	30	70	100
MSBI-141	Fundamental of Computer	30	70	100
MSBI-151	Operating System (Windows, Linux)	30	70	100
MSBI-161	Computer Lab - I	50	100	150
Semester - II				
MSBI-112	Object Oriented programming in C++	30	70	100
MSBI-122	Computational Method for Sequence Analysis	30	70	100
MSBI-132	Molecular Imaging & Drug Designing	30	70	100
MSBI-142	Database Management System & Data Mining	30	70	100
MSBI-152	Genomics, Proteomics and HGP	30	70	100
MSBI-162	Computer Lab - II	50	100	150
Semester - III				
MSBI-211	Perl programming for Bioinformatics	30	70	100
MSBI-221	Computer Networks and Advanced Multimedia	30	70	100
MSBI-231	Advanced Algorithm for Computational Biology	30	70	100
MSBI-241	Gene Expression & MicroArray Analysis	30	70	100
MSBI-251	Information Theory & Stochastic Process	30	70	100
MSBI-261	Computer Lab - III	50	100	150

Semester - IV				
MSBI-212	Major Project	100	300*	400
	* (Project Report Evaluation: 150 & Comprehensive Viva-voce: 150)			
Grand Total:				2350

MSBI – 111: Introduction of Bioinformatics

The nature of chemical bonds, Introduction to Genes and Proteins, Nucleotides, Orientation, Base pairing, The central dogma, Promoter sequences, Genetic Code, ORFs, Introns and Exons, Slice variants, Protein structure, Primary, Secondary, Tertiary and Quaternary, The notation of homology. Introduction to Data Generating Techniques: Restriction Enzymes, Gel Electrophoresis, Blotting and Hybridization, Cloning, PCR. Biological databases, Search engines, Public databases: PubMed, EMBL, GenBank, PDB, Swiss-Port. Genomics and Proteomics: Prokaryotic genomes, Eukaryotic Genomes, Gene Structure, GC Content in Eukaryotic genomes, Gene Expression, Protein Classification, 2D–Electrophoresis, Mass spectrometry, Microarray technology, X–ray crystallography, NMR, Sequence and Phylogeny Analysis, Detecting ORFs, Outline of sequence alignment, Introduction to BLAST, Multiple sequence alignment, Phylogenetic analysis.

Text / Reference Books:

1. Fundamental Concepts of Bioinformatics, D.E. Krane and M.L. Raymer, Person Education, 2003.
2. Bioinformatics Computing by B. Bergeron, Prentice–Hall, 2003.
3. Bioinformatics: A practical guide to the analysis of genes and proteins, D. Baxivanis and F. Oulette, Wiley Indian Edition, 2001
4. Bioinformatics: Sequence and Genome analysis, D.W. Mount, Cold Spring Harbour Laboratory Press, 2001

MSBI – 121: Statistics and Probability

Probability, Conditional Probability, Random Variables, Expected Value. Continuous and Discrete distributions: Binomial, Geometric, Poisson, Pascal, Uniform Exponential and Normal, Poisson Process. Multidimensional random variables, Multidimensional and bivariate normal distribution, Moment generating functions. Law of large numbers, Central limit theorem, Sampling distribution, Point and Interval Estimation Testing of hypothesis, Goodness of fit, Linear regression. Principal component analysis, Discriminant analysis, Cluster analysis, Multivariate analysis of variance.

Text / Reference Books:

1. Statistical Methods in Bioinformatics: An Introduction by W. Evans & G. Grant, Springer, 2001
2. First course in Probability by Sheldon Ross, Prentice-Hall (India), 2001

MSBI – 131: Molecular Biology

Concepts in Biology: Characteristics of living organisms, Structure of cells, Energy flow; Nutrition and Metabolism; Information storage and expression, Reproduction and Inheritance. Biomolecules: Carbohydrates, Amino Acids, Proteins, Lipids, Nucleic acids. Enzymes; Properties, Activation energy, Reaction kinetics, Intercellular Communication, Flow of genetic information, DNA expression and replication. Genes; Changing size of Genes, Gene synthesis, Cloning (Plant and Animal), Mutation, Gene Targeting, Gene Therapy. Level of secondary

structure, Basic concepts of molecular dynamics, Molecular simulation studies. Protein structure determination by X-ray Diffraction, NMR, Introduction to nucleic acid structure, Genome organization, HGP.

Text / Reference Books:

1. Molecular Biology by David Friefelder, Jones & Burtlett Publishers inc, 2000
2. Molecular Cell Biology by Alberts, Bruce Bray, Watson & James, Garland Publishing, 1999.

MSBI – 141: Fundamental of Computer

Digital Circuits and Computer Architecture: Transistors – and Integrated Circuits (LSI, VLSI); Operation of processor; Number Systems and Digital Circuits; ALU; Memory Chips, Storage Devices, Memory Hierarchy; I/O Devices; Bus Systems; Computer Organization and Architecture; Microprocessor; Moore's Law, Operating Systems and DBMS: Windows and Linux; Basics of relational Databases; SQL. Programming Languages: Algorithms; Problem Solving Classification; Basics of C Language, Concepts of objects and classes, Introduction to object-oriented Programming. Computers Networks: Protocols; Wireless LAN and Mobile Computing. Introduction to Web Technology: HTML and Web pages; The Internet and Intranet; WWW; Java Script and Dynamic Web Pages; Multimedia Application; Computer Network Security, e-Commerce.

Text/Reference Books:

1. The Essential Guide to Computing: The story of Information Technology, E. Garrison Walters, PHI, 2001.
2. A Balanced Introduction to Computer science by David Reed, PE, 2000.
3. Introduction to Computing Systems: From Bits to Gates to C and Beyond. Yale N. Patt and Sanjay J. Patel. Prentice Hall India, 1999

MSBI – 151: Operating System -Windows, Linux

Introductory concepts, Operating systems, Mainframe systems, Multiprocessors, Distributed systems, Real – Time systems, Computing environments. Computer system structure, Computer system organization, I/O structure, Storage structure, Storage hierarchy, Hardware protection, Network structure. Operating system structures, System components, OS Services, System calls, System programs, System structure, Virtual machines Processes, Process concepts, Process scheduling, Operations on processes, Process states, Cooperating processes, Inter-process communication, Common OS problems. DOS and Windows, Introductory concepts, Architecture and special features, Useful DOS commands, DOS – Editor, EDIT & writing Batch Files, Different Versions and their Distinguishing features, Managing Operations with Windows. Linux: Introductory concepts, Architecture and special features, Useful Linux commands, Related operating system and their Introduction.

Text / Reference Books:

1. Operating Systems by Glavin, Addison Wessely, 2000
2. Modern Operating Systems by S. Tanenbaum, Prentice Hall (India), 1999
3. Operating Systems Concept and Design by Milan Milenkovic, Tata McGraw Hill, 2000

MSBI – 161L Computer Lab - I

Windows and Linux installation, user interface, File management, Data Security, Users management. Using Linux and Windows commands. Sequence Alignment Software: Searching for sequence databases; Protein databases, PDB, BLAST, FASTA, CLUSTALW

MSBI – 112: Object Oriented Programming in C++

Methodologies: Concepts of structured and Object-Oriented programming, Advantages of OOPs methodologies. Introduction to Object-Oriented programming: Characteristics of OOPs Languages, C++ programming Basics operators, Loops, Arrays, Pointers. Principle of Object-Oriented programming, Introduction to Abstraction, Encapsulation, Inline function, Reference variable, Access specifier, Information Hiding, Friend Functions, Friend Class. Operator overloading, Function overloading. Inheritance: Extending classes, Types of inheritance, Virtual base class, Problems with multiple inheritance, Containership. Advanced pointers, Virtual functions, Polymorphism, Pointers to Class, Array pointers to Class, Virtual Function-Abstract Class, Pure Virtual Function, Files and Streams, Building class Libraries in C++. File Handling in C++, Template class, Exception Handling.

Text / Reference Book:

1. Object Oriented Programming with C++, E. Balagurusamy, Tata McGraw Hill
2. Object Oriented Programming in C++, Robert Lafore, Techmedia 2002
3. C++, How to program by Deitel & Deitel, Person Edition, 2002
4. The C++ Programming Languages by Stroustrup, Addison, Wesley., 2001
5. Problem Solving the object of Programming, Walter Savitch, Pearson Edition, 2003

BTBI – 122: Computational Methods for Sequence Analysis

Introduction to sequence databases: DDBJ / EMBL / GenBank; Information retrieval from databases; Analysis of DNA and Protein sequences: distribution, pattern and motif searches, Segmentation. Sequence alignment: Scoring matrices, PAM, BLOSUM, Local and global alignment concepts; Dynamic programming; Needleman - Wunch algorithm, Smith - Waterman algorithm; Statistics of alignment score; Multiple sequence alignment. Database searches for homologous sequences: FASTA, BLAST and CLUSTALW, Comparison, Artificial Neural Network concepts, Preceptron, Multilayer preceptron, Secondary Structure Prediction using ANN. Evolutionary analysis: Distances, Clustering methods, Tree representation, Genome sequence assembly; Gene finding methods; Analysis and prediction of regulatory regions. Profiles and Hidden Markov Model, Gene identification and prediction, DNA Microarrays, Gene Expression, Clustering gene expression profiles.

Text / Reference Book:

1. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, A. D. Baxevanis and B. F. F. Ouellette John Wiley and Sons (ASIA) Pvt. Ltd. 2002
2. Bioinformatics methods and applications - Genomes, Proteomics and Drug Discovery, S. C. Rastogi, N. Mendiratla and P. Rastogi, Prentice Hall (India), 2004
3. BLAST by I. Korf, M. Yandell & Joseph Bedell, O'Reilly, 2003

4. Biological Sequence Analysis – Probabilistic Models of Proteins and Nucleic acids, R. Durbin, S. Eddy, A. Krogh, G. Mitchison, Cambridge University Press, 1988

MSBI – 132: Molecular Imaging and Drug Design

Overview of Drug design, Simulation Methods, Molecular Mechanics, Monte Carlo, Semi-Impirical Method, Ab-initio Method. Analog based drug design: QSAR, Molecular Structure determination, Advance crystallographic techniques, Simulation Annealing (Autodock /Quantum), Peptide Sequencing, EST Clustering. Comparative Protein Modelling: Homology Modeling, Structural Alignment, Ramachandran plot (Side Chain Placement Refinement & Validation of Model), 3D- Pharmacophore modeling. Computer-aided drug designing: Drug design, Lead discovery, ADMET in drug design, Lipinski "rule of 5", Principle and Methods of Docking, Ligand based drug designing, Target identification, Drug Metabolism, Denovo design of drug. Structure based drug design. Introduction of Drug discovery, Pharmacogenomics, Analysis of SNP's, Drug discovery technologies and strategies.

Text / Reference Book:

1. Molecular Modelling: Principle and Applications, Andrew R. Leach, Longman Singapore Publisher, 1999
2. Bioinformatics: From Genome to Drugs (Vol. I and II), Thomas Lengauer, Wiley- VCH, 2000
3. Structural Bioinformatics, P. E. Bourne and B.H. Weissing, Wiley VCH, 1999
4. Bioinformatics Method and Applications-Genomics, Proteomics and Drug Discovery, S.C. Rastogi, N. Mendiratta and P. Rastogi, Prentice Hall (India), 2004

MSBI – 142: DataBase Management System and Data Mining

Introduction, Characteristics of Data base approach, Database users, Intended uses of databases, Implication of database approach. Database System Concepts and Architecture, Data Models, Schemas and Instances, DBMS Architecture and data independence, database languages, Classification of DBMS. Database Design: Informal Design guidelines for relation Schemas, functional Dependencies, Normal forms based on primary keys, general Definition of 2nd and 3rd Normal forms, BCNF, Need of further Normalization. Data Modeling using ER diagram, ER Model Concepts, Notation for ER Diagrams. The Relational Model, Relational Model Concepts, Relational Model Constraints, Introduction to Relational Algebra. Data Mining, Definition, Data Mining and KDD, Data Mining on relational databases, Data Warehouses, Transactional databases etc. Data Mining Functionalities, Pattern Mining, Data Mining on Biological data.

Text/Reference Books:

1. Fundamentals of Database System by Elmasri and Navathe, Prentice Hall (India), 2001
2. Data Mining Concepts and techniques- Jiawei Han, Micheline Kamber, Morgan Kaufmann publisher, 2001

MSBI – 152: Genomics, Proteomics and HGP

Primer Selection Phylogeny, Protein Sequence Analysis, SAGE, Primer Chain Reaction (PCR). Genomics, Fragmenting the genomes, Need of markers, Marker Sequencing (RFLP, AFLP & SNP, etc); Genome sequencing: Basic Sanger sequencing, Automated sequence, Gene Identification, MUMmer , Blast2. Proteomics, Microarray: The technique of microarray design, 2-D gel electrophoresis, Peptide sequencing, Mass spectroscopy. Metabolomics, Metabolic pathway databases: KEGG, COG, WIT; Drug target identification. Functional genomics: Application of sequence based and structure based approach of gene function eg: Sequence comparison, Structure analysis (active site, binding site) and comparison; Approaches for determining gene expression patterns and functions. What is HGP: Goals, Progress & History of HGP.

Text / Reference Book:

1. Handbook of Comparative Genomics: Principles and Methodology by T.B. Kitano, Graziano Pesole, 2003.
2. Computational Molecular Biology by P. Clate & R. Backofen, Willy Publication, 1998 Proteomics,
3. Timothy Palkill, kluwer Academic Publisher, 2002
4. Introduction to Proteomics: Tools for the new Biology, Daniel C. Liebler, Humana Press, 2001
5. Foundation to systems Biology (Ed) by Hiroaki Kitano, MIT Press, 2001

MSBI – 162: Computer Lab - II

Problem solving using C++ programming language. Sequence Alignment Software: Searching for sequence databases; Protein databases, PDB, BLAST, FASTA, CLUSTW, Comparing and aligning sequences; Phylogenetic trees for sequences. Creating database for Biological projects including Schema, E-R Diagram, and Normalization. Developing database applications: SQL queries using Oracle. Clustering and classification software such as SOM, SVM, PCA; using visualization software such as TreeView, Sammon Mapping.

MSBI – 211: Perl Programming for Bioinformatics

Getting started with PERL, Perl's Benefits, Versions of Perl, Downloading and installing Perl in Linux/Windows environment, How to run perl programs in Linux/Windows environment, Error Message, Debugging, Perl interpreters, Perl scripts. Scalar values and scalar variables, Strings and numbers, Assignment statements, Blocks, Arrays, Hashes, Operators, Binding operators, Loops, IO, Input from file, Input from SDIN, Input from file named in command line, Regular expression, Pattern matching, Meta symbols, Pattern modifiers, Subroutines and modules, Built-in functions. Application of Perl in Bioinformatics: Representing strings and sequenced data in Perl, Program to store a DNA sequence, Concatenating DNA fragments, DNA to RNA transcription, Reading proteins from files, Finding motifs, Counting nucleotides, Exploding strings into arrays, Operating on strings. A program to simulate DNA mutation, Generating random DNA, Analyzing DNA, Translating DNA into proteins, GenBank sequence and

annotation, Parsing annotation, Indexing GenBank with database. PDB files, Parsing PDB files, Parsing BLAST output, Introduction to BioPerl, CPAN, and Python.

Text / Reference Book:

1. James D.Tisdall, "Beginning Perl for Bioinformatics", O'Rilley and Associates, 2001.
2. Cynthia Gibas & Per Jamesbeck, "Developing ioinformatics Compuer skills", O' Rilley & Associates, 2000
3. Rex A.Dawyer, "Genomic Perl", Cambridge University Press.
4. Learning Perl, 3rd Edition: Randal L.Schawrtz and Tom Phoneix, O'Rilley

MSBI – 221: Computer Networks and Advanced Multimedia

Basic concepts of Computer Networks, Networks Hardware, Networks Software, ISO-OSI reference models, Protocols. The Physical and Data Link Layer, Different transmission media, Data Link Layer Design Issues, Error Detection and Correction. The Network Layer, Network Layer Design Issues, Shortest Path Routing, IP Protocols, IP Address. The Transport Layer, The Transport Service, Services Providing to the Upper Layers, Transport Services Primitives, Internet Programming Tools- HTML, DHTML, JavaScript: Designing and developing web pages using HTML/DHTML. Overview of Multimedia: Different Views and Perspectives; Multimedia Applications: Video-on-demand, Interactive television, Video conferencing, Hypermedia courseware, Groupware, World Wide Web, Digital libraries; Media and its Types: Static, Continuous, Time-Based.; Visual Communication and Human Perception: visual and auditory; Multimedia Devices and Systems: Audio/Video boards, Support for Media, CMS services; User Interface Design: Need, Significance, Guidelines etc. Multimedia Development: Phases, Stages, Team etc, Multimedia Building Blocks, Text: Value and significance, Formats, style, use in MM Networking and Applications, Streaming Audio and Video, Best Effort Service, Protocols for Real Time Applications, RTP, RTCP, SIP etc.; Graphics: Significance, different Image Formats, Scheduling Policing Mechanism, Integrated services and Differentiated services

Text / Reference Books:

1. Computer Networks by A.S. Tanenbaum, 4th Edition, Pearson Education, 2000
2. Multimedia – An Introduction by Villmil:, Prentice Hall (India), 1999
3. Computer Networking by J.F. Kurose and Keith W. Ross, 2nd Edition, Pearson Education, 2001

MSBI – 231: Advanced Algorithm for Computational Biology

Algorithms and Complexity, Algorithm Design Techniques, Tractable versus Intractable Problems. Restriction Mapping, Graphs, Measuring Fragment Size, Restriction Mapping Algorithms, Double Digest Problem, Classifying multiple solutions, NP-Completeness, Regulatory motifs in DNA sequences, Profiles, Motif finding Greedy algorithms, Divide and Conquer Algorithms. Markov Chains, Hidden Markov Models, CG-Islands, Forward & Backward Algorithm, HMM Parameter estimation, Profile HMM, Building viterbi Algorithm, Multiple Sequence Alignment by profile HMM's. DNA Sequencing, Shortest Superstring Problem, DNA Arrays as an Alternative Sequencing Technique, Sequencing by Hybridization,

SBH as a Hamiltonian Path Problem, SBH as an Eulerian Path Problem, Fragment assembly in DNA Sequencing. Phylogenetic Reconstruction, Gene Expression Analysis, Hierarchical Clustering, *k*-Mean Clustering, Distance Based Tree Reconstruction, Additive Metrics, Metrics on Tree, Ultrametric Trees, Character-Based Tree Reconstruction, Parsimony Method, Evolutionary Methods, Maximum Likelihood Method, Model Comparison. Protein Folding: Levels of Proteins Structures, Prediction by Profile HMMs, Molecular Modeling.

Text / Reference Book:

1. An Introduction to Bioinformatics Algorithms, Neil C Jones & Pavel A. Pevzner, Ane Books, 2005.
2. Introduction to Computational Molecular Biology, Setubal and Meidanis, Thomson, 2003.
3. Biological Sequence Analysis Probabilistic Models of Proteins and Nucleic Acid, R. Durbin, Eddy, A. Krogh and Mitchison, Cambridge University Press, 1998.
4. Introduction to Mathematical Method in Bioinformatics, Alexcender Isaev, Springer Universitext, 2004

MSBI – 241: Gene Expression & MicroArray Analysis

DNA MicroArray: The Technical Foundations, Why are MicroArray Important? What is a DNA MicroArray. Designing a MicroArray Experiment: The Basic steps, Types of MicroArray, NCBI and MicroArray Data Management, GEO (Gene Expression Omnibus), MAML, The benefits of GEO and MAML, The Promise of MicroArray Technology in Treating Disease. MicroArray Data, Preprocessing the Data, Measuring Dissimilarity of Expression Pattern, Distance Motifs and Dissimilarity measures, Visualizing MicroArray Data; Principal Component Analysis, PCA and MicroArray Data. *K*-Means Clustering, Hierarchical Clustering, Self Organization Maps (SOM). Identifying Genes: Expressed usually in a sample, Expressed significantly in population, Expressed differently in two populations. Classifying Samples from two populations using Multilayer Perceptron, Support Vector Machines and their applications, Using genetic algorithm and perceptron for feature selection and supervised classification.

Text/Reference Books:

1. Microarray Data Analysis and Visualization by Arun Jogota, Bioinformatics, The Bay Press, 2001
2. NCBI Science Premier on MicroArrays.
3. Bioinformatics Methods and Application-Genomics, Proteomics and Drug Discovery by S. C. Rostogi, N. Mendiratla and P. Rostogi, Prentice Hall (India), 2004

MSBI- 251: Information Theory & Stochastic Process

Overview: Probabilities defined on events, Random variable, Binomial, Geometric, Poisson variates, Uniform, Exponential, Gamma, Normal Variates, Expectation. Introduction to stochastic process, Stationary process Poisson process, Non-homogeneous Poisson process, Brownian motion process. Markov Chain: Chapman-Kolmogorov equation, Classification of states, Limiting Probabilities, Time Reversible Markov Chain, Markov Chain Monte Carlo Method, Continuous Time Markov Chain, Birth and Death Process, Time reversibility, Uniformization. Information and Entropy, Joint and Conditional Entropies, Mutual information, Chain rules for

entropy, Relative entropy. Entropy rules of stochastic process, Entropy rate, Hidden Markov Models. Channel Capacity, Noiseless binary channel, Symmetric, Gaussian channel.

Text / Reference Book:

1. Element of Information Theory by Cover and Thomas, New York: Wiley, 1991.
 2. A First Course in Stochastic Process by Samuel Karlin and Howard M. Taylor, Academic Press Inc. (London) Ltd., 1988.
 3. Stochastic Processes by Sheldon M. Ross, John Wiley & Sons, 2000.
- An Introduction to Stochastic Processes by Edward P. C. Kao, Duxbury, 1999

MSBI – 261: Computer Lab - III

Perl programming for Motif search, Open reading frames, Parsing, etc. Signing and Creating Web pages using: HTML, Java Script. Multimedia: Introduction to Photoshop, Developing multimedia application using Flash. Gene expression and MicroArray: Pre-processing of datasets; using clustering and classification software such as SOM, SVM, PCA using visualization Software such as Tree View, Sammon Mapping

MSBI – 212: Major Project:

Project Guideline

- Students must submit his/her project synopsis to the concern supervisor along with his/her Name, Father's Name, ID Number, Title of Project, etc and get it approved by them.
- Project must be done by individually or in group of maximum three students.
- Approval by the supervisor is based on the directions given by the Department.
- After scrutinizing the same by the supervisor, students must complete their project and submit the final report within the stipulated time in the form of soft and hard copy to the Department.
- The Department will evaluate project.
- Students must demonstrate their project in their Department to the concern persons on the date fixed by the learning center.

Objectives of the project proposed:

1. To provide the scope to apply the concept of Bioinformatics, computer programming language, database used and application of computer science in biology.
2. To analyze biological problems and provides good exposure to development & implementation of software database by following the software engineering concepts.

Marks Distribution:

- **Internal Assessment:** 100 Marks

Project Supervisor:	40
Presentation & Viva Voce Internal	40 + 20 = 60
.....	
Total:	100

• **Semester Assessment:** 300 Marks

Presentation & Viva Voce - Final:	150 + 50 = 200
Project Report:	100
.....	
Total:	400