

ADMISSION & EXAMINATION BYE-LAWS

FOR

MASTER OF SCIENCE

Computational & System Biology and Bioinformatics

M.Sc. (CSB)

Program Code: 582

***CHOICE BASED CREDIT SYSTEM (CBCS)
(W.E.F. 2021-22)***



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

School of Engineering Sciences & Technology

JAMIA HAMDARD

(DEEMED TO BE UNIVERSITY)

Hamdard Nagar, New Delhi-110 062

Ph. 011 26059688, Extn.-5858

Table of Contents

Sl. No.	Content	pages
1	Date of approval of present syllabus	3
2	Program outcomes	4
3	Admission rules	5
4	Program structure	6-12
5	Examination rules	13-16

Date of approval of present syllabus: 14.12.2021

PROGRAM OUTCOME

PO1: Communication Skills: The students will be able to demonstrate English language proficiency to perform effectively in the professional and personal life by being able to comprehend and write effectively and efficiently.

PO2: Domain knowledge: The students will be able to understand the theoretical foundations and the limits of computing in Bioinformatics.

PO3: Technical skills: The students will be to design, develop and evaluate new computer based systems for novel applications which meet the desired needs of industry and society.

PO4: Knowledge inter-disciplinary in nature: The students will be exposed to acquires sufficient knowledge of the interdisciplinary subjects and use them in developing advanced computing techniques and tools in the field of biology.

PO5: Positive attitude: The students will be able to inculcate a positive attitude through various courses.

PO6: Critical thinking and problem-solving skills: The students will be able to apply the fundamentals of computer science in various biological related problems to come up with feasible solutions.

PO7: Dynamism and team building skills: The students will be able to develop required skills to work efficiently on multidisciplinary projects and teams to accomplish a common goal.

PO8: Professional ethics and social values: The students will develop an understanding of work ethics and will have the ability to carry out any task with professional ethics and without deviating from social values

PO9: Self-awareness and emotional intelligence: The students will identify their strengths and talents and learn to establish a balance at the time of crisis.

PO10: Entrepreneurship and Innovative qualities: The students will acquire entrepreneurship and innovative qualities through various learning programs.

PO11: Responsibility towards society and environment: The students will realize their social responsibilities. The students are expected to learn tools and techniques for designing and integrating technology-based solutions for real world problems and drive scientific and societal advancement through technological innovation.

PO12: Lifelong learning: The students are expected to engage in lifelong learning for the advancement of technology and its adaptation in multi-disciplinary environments.

ADMISSION RULES
MASTER OF SCIENCE
Computational & System Biology and Bioinformatics

M.Sc(CSB) offered by the Department of Computer Science & Engineering and coordinated by BIF centre and SPER.

1. OBJECTIVE

To prepare highly skilled professionals with a strong conceptual, theoretical & practical proficiency and research ability in the field of Computer Science & Engineering and related emerging areas; such as Data Sciences, Big Data Analytics, Healthcare and Bioinformatics

2. THE PROGRAM

Highlights of the program are described in the following table:

a.	<i>Name of the Program</i>	M. Sc (CSB)
b.	<i>Nature</i>	Regular and Full Time
c.	<i>Duration</i>	Two Years (4 Semesters)
d.	<i>Total number of credits</i>	96
e.	<i>Medium of Instruction and English Examinations</i>	English
f.	<i>Eligibility Criteria</i>	A minimum of 55% marks in Bachelor's degree program in bioinformatics, or any branch of basic/applied Sciences or Technology, including Life Sciences, Medicine, Pharmacy, Engineering and Computer Science and allied disciplines.
g.	<i>Selection procedure</i>	As per the merit of the qualifying examination
h.	<i>Total Seats</i>	7-8 candidates; inclusive of seats reserved for NRI / sponsored candidates; additional seats are available for Foreign Nationals.
i.	<i>Period of Completion</i>	Not more than 04 years (8 Semesters)
j.	<i>Commencement of the Program</i>	July of every academic session

3. PROGRAM STRUCTURE

Semester-wise course structure, guidelines for teaching, practical and associated assessment of the program is described in the following tables:

Course Type	Abbreviation
Program Core	PC
Program Elective	PE
Computer Science	CS
System biology	SB
Foundation Course	FC
Ability Enhancement Course	AE
Skill Enhancement Elective	SE
Laboratory	LAB
Dissertation	DISS
Seminar	SEM

Course Type	Subject Area	Credits	Percentage (%) (Approx.)
Foundation Core (FC)	Basic Sciences (BS) including Mathematics, Computing, Biology	12	12.5
Program Core (PC)	Computer Science (CS) courses including algorithms, data analysis, programming etc	30	81.25
	System biology (SB) courses related to, Immunology, Pharmacy	20	
	Project Work, Seminar and/or Internship in Industry or elsewhere.	28	
Program Electives (PE)	Program Elective (PE) courses relevant to chosen specialization/branch	6	6.25
Total		96	100

L-T-P stands for number of contact hours as Lecture-Tutorial-Practical in a week.

1st Semester

Course Code	Course Title	Course Type	Marks			L-T-P	Credits
			Internal Assessment	Semester Exam	Total		
MCSB 101	Introduction of Bioinformatics, Biological Data & Databases	SB	25	75	100	3-1-0	4
MCSB 102	Biology for Bioinformatics	BS	25	75	100	3-1-0	4
MCSB 103	Fundamentals of computing	BS	25	75	100	3-1-0	4
MCSB 104	Essential Mathematics & Biostatistics	BS	25	75	100	3-1-0	4
MCSB 105	Programming in Perl	CS	25	75	100	2-0-0	2
	Program Elective – I	PE	25	75	100	2-0-0	2
MCSB 109	Lab – I (Based on MCSB 101)	LAB	25	75	100	0-0-2	2
MCSB 110	Lab – II (Perl Lab)	LAB	25	75	100	0-0-2	2
	Skill Enhancement- I	SE	-	-	-	2-0-0	-
Total					800	16-4-4	24

2nd Semester

Course Code	Course Title	Course Type	Marks			L-T-P	Credits
			Internal Assessment	Semester Exam	Total		
MCSB 201	Structural Bioinformatics & drug Design	SB	25	75	100	3-1-0	4
MCSB 202	Designing algorithms for Bioinformatics	CS	25	75	100	3-1-0	4
MCSB 203	Database Management System	CS	25	75	100	3-1-0	4
MCSB 204	Predictive Modelling and Analytics	CS	25	75	100	3-1-0	4
MCSB 205	Chemoinformatics ,Chemogenomics and Immunoinformatics.	SB	25	75	100	2-0-0	2
	Program Elective – II	PE	25	75	100	2-0-0	2
MCSB 210	Lab III (System Biology lab)	LAB	25	75	100	0-0-2	2
MCSB 211	Lab IV(DBMS lab)	LAB	25	75	100	0-0-2	2
	Skill Enhancement- II	SE	-	-	-	2-0-0	-
Total					800	16-4-4	24

Students be encouraged to go to Industrial Training/Internship for at least 2-3 months during semester break.

3rd Semester

Course Code	Course Title	Course Type	Marks			L-T-P	Credits
			Internal Assessment	Semester Exam	Total		
MCSB 301	NGS Data analysis – Microarray, RNA Seq, Single Cell sequencing	SB	25	75	100	3-1-0	4
MCSB 302	Programming in R and Python	CS	25	75	100	3-1-0	4
MCSB 303	Data Warehousing and Data Mining	CS	25	75	100	3-1-0	4
MCSB 304	Data Modelling and Visualization	CS	25	75	100	2-0-0	2
	Program Elective - III	PE	25	75	100	2-0-0	2
MCSB 308	Minor project ^{@#}	PROJ	25	75	100	4-0-0	4
MCSB 309	Lab V (R and Python Lab)	LAB	25	75	100	0-0-2	2
MCSB 310	Lab VI (NGS lab)	LAB	25	75	100	0-0-2	2
	Skill Enhancement- III	SE	-	-	-	2-0-0	-
Total					800	17-3-4	24

4th Semester

Course Code	Course Title	Course Type	Marks			L-T-P	Credits
			Internal Assessment	Semester Exam	Total		
MCSB 401	Project/ Dissertation [@] #	DISS	300	200	500	0-0-40	20
MCSB 402	Seminar	SEM	25	75	100	0-0-8	4
	Skill Enhancement- IV	SE	-	-	-	2-0-0	-
Total						0-0-48	24

@ Dissertation/Mini Project shall be based on latest research topics in the field of Bioinformatics, Computer Sciences.

Students are required to get approval of their title of Dissertation/Mini Project by Dissertation Assessment & Evaluation committee constituted by HOD. Supervisor of respective students must be member of the above committee. Students are required to give at least three presentations/seminars for progress monitoring & assessment purpose to their respective supervisors. Viva-voce will be held only after the submission of completion report duly signed by the supervisor of the respective student. A plagiarism report duly signed by the students are mandatory to submit in compliance with UGC (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions) Regulations, 2017 (or any such regulations notified time to time) by competent authority.

PROGRAM ELECTIVE (PE)

Course Code	Course Title	Marks			L-T-P	Credits
		Internal Assessment	Semester Exam	Total		
Program Elective – I						
MCSB PE106	Biomolecules	25	75	100	2-0-0	2
MCSB PE107	Bioinformatics tools and development	25	75	100	2-0-0	2
MCSB PE108	Advance Bioinformatics Applications	25	75	100	2-0-0	2
Program Elective – II						
MCSB PE 206	Research Methodology	25	75	100	2-0-0	2
MCSB PE 207	Programming in Java	25	75	100	2-0-0	2
MCSB PE 208	Intellectual Property Rights	25	75	100	2-0-0	2
MCSB PE 209	Comparative and Functional Genomics	25	75	100	2-0-0	2
Program Elective – III						
MCSB PE 305	Evolutionary bioinformatics	25	75	100	2-0-0	2
MCSB PE 306	Proteomics & Metabolomics	25	75	100	2-0-0	2
MCSB PE 307	Artificial Intelligence and Machine Learning in Bioinformatics	25	75	100	2-0-0	2

Skill Enhancement Electives (SE)

Course Code	Course Title	L-T-P
Skill Enhancement- I		
MCSB SE 111	Software testing and Quality Assurance	2-0-0
MCSB SE 112	Advanced Algorithms	2-0-0
Skill Enhancement- II		
MCSB SE 212	Neural Networks and Deep Learning	2-0-0
MCSB SE 213	Green Computing	2-0-0
Skill Enhancement- III		
MCSB SE 311	Linux and Unix Programming	2-0-0
MCSB SE 312	Information Retrieval	2-0-0
Skill Enhancement- IV		
MCSB SE 403	System optimization	2-0-0
MCSB SE 404	Speech and Natural Language Processing	2-0-0

EXAMINATION RULES

4. **MODE OF CURRICULUM DELIVERY**

Mode of curriculum delivery includes classroom teaching, assignments, test, lab work, presentations, participation in relevant events and regularity.

5. **ATTENDANCE**

- a. All students are supposed to attend every lecture and practical classes. However, the attendance requirement for appearing in the examination shall be a minimum of 75% of the classes held.
- b. Each one-period teaching shall account for one attendance unit.
- c. The concerned teacher will take a roll call in every scheduled class, maintains and consolidate the attendance record, which would be submitted to the Head of the Department at the conclusion of the semester.
- d. Attendance on account of participation (with prior permission from the Head of the Department) in the co-curricular/extra-curricular activities can be granted by the Dean on receipt of certificates or recommendations of the respective activity issued by the Head of the Department.
- e. Attendance records displayed on the Notice Board from time to time, in respect of short attendance, shall be deemed to be a proper notification and no individual notice shall be sent to the students/local guardian.
- f. In case a student is found to be continuously absent from the classes without information for a period of 30 days, the concerned teacher shall report it to the Head of the Department.
- g. Head of the Department may recommend for striking off the name of a student from rolls, after ensuring 'one month continuous absence', from all the concerned teachers.
- h. A student, whose name has been struck off on account of long absence may apply to the Dean for readmission within 15 days of the notice of striking off the name. The readmission shall be effected on payments of prescribed readmission fees.
- i. A student with less than 75% attendance in a subject shall not be allowed to appear in that subject in the semester examination. The Head of the Department shall recommend all such cases to the Dean of the School.
- j. The Dean, on the recommendation of the Head of the Department, may consider the relaxation of attendance up to 10% on account of sickness and /or any other valid reason. No application for relaxation of attendance (duly certified by a Registered Medical Practitioner/Public hospital or a competent authority) will be entertained after 15 days from the recovery from illness etc.

6. **INTERNAL ASSESSMENT**

- a. Internal assessment, to be made by concerned teachers, will be based on minor tests, quizzes, presentation, programming test, demonstrations and assignments.

- b. There will be three (3) Internal Assessment (Unit Tests) with a total of 20 marks ,and the best two (2) performances out of the three Unit tests of Internal Assessment will be counted. Other modes of assessment shall account for remaining 5 marks.
- c. Dates for unit test will be announced at the beginning of the semester, by the examination coordinator.
- d. The teacher concerned shall maintain a regular record of the marks obtained by students in unit tests and display the same in due course.
- e. The concerned teachers shall submit the compiled internal assessment marks to the Head of the Department, on the conclusion of teaching of the current semester.
- f. The Head shall display a copy of the compiled sheet, of internal assessment marks of all the papers, before forwarding it to the Controller of Examination, i.e. at the conclusion of the semester.
- g. A promoted candidate, who has to reappear in the examination of a paper, will retain internal assessment marks.
- h. In the case of re-admission, the candidates shall have to go through the internal assessment process afresh and shall retain nothing of the previous year.

7. SEMESTER EXAMINATIONS

Prescriptions for conducting semester examinations of theory and lab papers, those shall be conducted after the conclusion of each of the semesters, are presented in the following table:

S.N.	Classification	Theory	Lab
1.	Mode	Written Only	Written, Demo, Programming and viva- voce etc.
2.	Duration	03 Hours	04 Hours
3.	Total Marks	75 (Seventy Five Only)	75 (Seventy Five Only)

8. DISSERTATION/INDUSTRIAL PROJECT

- a. Each student of the final semester will have to go for a Dissertation/Industrial Project work either in the industry or in the Department under the guidance of one or two faculty members.
- b. Period of completion of Dissertation/Industrial Project work shall be full one semester.
- c. There shall normally be two supervisors-one internal and one *external (in the case of industry project form the place where the student is pursuing project-work)*.
- d. All the students, who are pursuing the Dissertation/Industrial project work, shall be continuously in touch with the internal supervisor.
- e. **There shall be a mid-term evaluation of the progress** and the internal supervisors will conduct it. However, an internal supervisor may ask the student to submit a confidential progress-report from the external supervisor (*in the case of industry project*).
- f. All the candidates shall submit **Three (03)** hard copies of the project reports that are duly approved and signed by internal as well as external (*if applicable*) supervisors.
- g. An external examiner, appointed for the purpose, shall evaluate the project report.
- h. The Head of the Department shall fix a date and time for viva-voce examinations, on receipt of the evaluation-report of the project reports from the external examiner.
- i. Head of the Department shall forward the compiled total marks (awarded in internal assessment, project Report and Viva-voce Examination), in the project-semester of each of the candidate, to the Controller of Examination.

9. EXAMINATION

- a. The performance of a student in a semester shall be evaluated through continuous class assessment and end semester examination. The continuous assessment shall be based on class tests, assignments/ tutorials, quizzes/ viva voce and attendance. The end semester examination shall be comprised of written papers, practical and viva voce, inspection of certified course work in classes and laboratories, project work, design reports or by means of any combination of these methods.
- b. The marks obtained in a subject shall consist of marks allotted in end semester theory paper, practical examination and sessional work.
- c. The minimum pass marks in each subject including sessional marks (Theory, Practical or Project etc.) shall be 40%.

10. PROMOTION SCHEME

- a. A student will be required to clear minimum **40% of his/her papers** in a semester/annual examination to be eligible **for promotion to the next**

semester/year. A student may appear in the supplementary examination after each semester/annual examination and can have a choice to appear in the backlog papers in the supplementary examination or in the subsequent regular semester/annual examination with a prescribed fee. A students detained due to shortage of attendance will repeat his/her paper in the subsequent semester concerned (even/odd).

b. A **detained** Student is not allowed to re-appear in the internal assessment (Unit test). His/her old internal assessment marks will remain same.

A student who cleared all the papers of a semester/annual examination of a programme /course will be eligible for improvement examination as per university rule.

11. THE GRADING SYSTEM

As per University Rule

12. CALCULATION OF SGPA AND CGPA OF A STUDENT IN A SEMESTER

As per University Rule

After having passed all the FOUR semesters, the students shall be eligible for the award of **Master of Science (Computational & System Biology and Bioinformatics)**, degree of JAMIA HAMDARD.

13. CLASSIFICATION OF SUCCESSFUL CANDIDATES

The result of successful candidates, who fulfill the criteria for the award of **Master of Science (Computational & System Biology and Bioinformatics)**, shall be classified at the end of last semester, on the basis of his/her final CGPA (to be calculated as per university rule).

MCSB 101	Introduction of Bioinformatics, Biological Data & Databases	SB	25	75	100	3-1-0	4
-----------------	--	----	----	----	-----	-------	---

UNIT 1

Introduction to Bioinformatics: Introduction and branches of Bioinformatics. Aim, scope and research areas of Bioinformatics. Overview of Available Bioinformatics Resources on the Web, Protein and Genome; Information Resources and Analysis Tools; Established Techniques and Methods; Sequence File Formats FASTA, GenBank, FASTQ, GTF and Structured File Formats.

UNIT 2

Biological Databases: Introduction of Biological databases and its type. Classification format of biological databases. Biological database retrieval system. Protein Sequence and Structural Databases, Nucleotide Sequence Databases, Gene Expression Database;

- Biological database and datasets (2)
- Overview of Bioinformatics resources on web (NCBI/EBI/EXPASY etc.) (2)
- Nature of biological data and formats (2)
- Biological literature databases, PubMed (2), Google scholar, Scopus
- Nucleic acid sequence databases, GenBank/EMBL/DDBJ/ RefSeq/dbSTS/dbEST (4)
 - Protein sequence databases, Uniprot, PIR, UniProtKB, UniRef, UniParc, Proteomes, NextProt (4)
- Derived databases: InterPro and constituent databases, Recent derived databases (4)
 - RNA sequence databases, SRA, GEO, TCGA, ENA, miRBase, lncRNAdb, MIT/ICBP siRNA database (3)
- Species and Biodiversity databases / resources, NCBI Taxonomy database, GBIF, Sahyadri (2)
- Specialized Databases: Pfam, SCOP, GO, Metabolic Pathways.

UNIT 3

Methods of Sequence Analysis: Pairwise and multiple sequence alignment for DNA and protein sequences. Local and global sequence similarity. Pairwise sequence alignment methods; Heuristic Methods; BLAST and its variants, Statistics of Sequence Alignment Score; E-Value, P-Value, scoring matrix, PAM, BLOSUM, Gap Penalty, Gonnet and Lookup tables; Multiple Sequence Alignments; ClustalW, Hidden Markov Models, HMM Based Multiple-Sequence Alignment, Sequence logos, consensus & patterns, Basic concept of sequence profiles, Derivation of profiles; applications - Gribskov's Profile Analysis method

Application tools: Primer designing.

UNIT 4

Tools for molecular mapping: QTL, minisatellites, SNP's.

Mapping techniques: JoinMap, MapQTL, LOD Score method for estimating recombination frequency.

Genome analysis and Gene identification: Sequencing, Assembly, Annotation, Sequencing pipelines and databases.

Comparative genomics: Homologs, Paralogs and orthologs; Synteny; Comparative genomics of *Arabidopsis* and *Brassica rapa* / Chimpanzees and human.

UNIT 5

Phylogenetic Analysis: Methods of Phylogeny, Software for Phylogenetic Analyses, Consistency of Molecular Phylogenetic Prediction. Distance and Character Based Methods and Software, Computing Tools for Phylogenetic Analysis, Distances, CLUSTALW, MUSCLE, DALIGN, T-Coffee, MAFFT, GROWTREE, PAUP, PHYLIP and MEGA; Construction and Visualization Phylogenetic Tree; and Application of Phylogenetic Analysis.

MCSB 102	Biology for Bioinformatics	BS	25	75	100	3-1-0	4
-----------------	-----------------------------------	----	----	----	-----	-------	---

UNIT 1

Nucleic acid: Composition, Primary and Secondary structures, Circular DNA

UNIT 2

Genome organization: Prokaryotic and eukaryotic genomes – C value paradox, repetitive and non-repetitive DNA., transposons and retroposons; Exons and introns organization of interrupted genes, one gene-many proteins concept; Gene numbers essential genes and total gene number, gene clusters, pseudogenes; Gene families globin and rDNA gene families; Organelle genome – mitochondrial and chloroplast Packaging of genome – Bacterial genome as nucleoid; Eukaryotic genome nucleosomes, chromatin, solenoids, loops, domains, scaffolds, chromosomes

UNIT 3

DNA Replication- Details of prokaryotic and eukaryotic DNA replication: DNA polymerases, initiation, elongation and termination of replication; multiple origins of replication, Regulation of replication

DNA damage, repair and recombination- DNA damaging agents physical and chemical, types of DNA damages; DNA repair systems in prokaryotes and eukaryotes Single step repair, Base excision repair, Nucleotide excision repair, Mismatch repair, Recombination repair; Recombination - homologous and non-homologous recombination

UNIT 4

Gene Expression - Transcription: Details of prokaryotic and eukaryotic transcription: RNA polymerase, promoters, initiation, elongation and termination of transcription; regulation of transcription, operons

Processing of transcripts - 5' capping, 3' polyadenylation, splicing and editing, self-splicing

Translation: Details of prokaryotic and eukaryotic translation: Protein synthesis machinery, initiation, elongation and termination, Genetic code, accuracy of translation, regulation, Post- translational modifications of proteins .

UNIT 5

Recombinant DNA technology: Enzymes, cloning strategies, vectors, screening for recombinants .

Small non-coding RNAs – micro-RNA, small nuclear RNA, small nucleolar RNA, si-RNA, piwi-RNA and Genome editing

MCSB 103	Fundamentals of computing	BS	25	75	100	3-1-0	4
-----------------	----------------------------------	----	----	----	-----	-------	---

UNIT 1

Computing Concepts: Basic Computing Systems, Layers of a Computing System, History of Computing, History of Computing Software, Stored-Program Concept and von Neumann Architecture. Fetch-Execute Cycle, RAM and ROM, Types of RAM and ROM, Secondary and Tertiary Storage Devices, Cache Memory, Memory Hierarchy, Input- Output Devices, Touch Screens.

UNIT 2

Data Representation and Number Systems: Binary Values and Computers, Data and Computers, Analog and Digital Data; Binary Representation. Number Systems: Binary, Octal, Decimal, and Hexadecimal. Conversions of Data from one Number System to another Number System. Representation of Numeric Data – Negatives and Real Data Representation. Representing Texts - ASCII and Unicode Character Sets. Binary Arithmetic – Addition and Subtraction of Numbers in Different Number Systems.

UNIT 3

Operating Systems: Introduction and Examples of Operating Systems, Roles of Operating Systems: Memory, Process, and CPU Management. Batch Processing, Categories of Operating Systems: Timesharing, Multitasking, Multithreading, and Multiprogramming Operating Systems. Process States and Process Control Block.

UNIT 4

Computer Networks: Introduction to Computer Networks. Network Topologies, Types of Networks, Internet vs. Intranet, Switch, Bridge, and Routers. Firewalls, Network Protocols, TCP/IP, Network Addresses, Domain Name System, Internet services: E-mail, FTP, ISP, etc.

UNIT 5

Linux basic concepts & installations: Basic command, file transfer, cluster handling,

MCSB 104	Essential Mathematics & Biostatistics	BS	25	75	100	3-1-0	4
----------	--	----	----	----	-----	-------	---

UNIT1

Basic Mathematics: Vectors: Vector algebra, Dot and Cross products. Matrices: Algebra of Matrices, Transpose and inverse, Diagonalization of Matrices and Characteristic roots. Basic Differentiation and Integration Techniques.

UNIT 2

Central Tendency and Dispersion: Terminologies: Population, Sample, Variable, Parameter, Primary and Secondary Data; Screening, Representation and Classification of Data; Frequency Distribution, Tabulation and Graphical Representations; Mean, Geometric Mean, Harmonic Mean, Median and Mode; Quartiles and Percentiles; Measures of Dispersion: Range, Variance, Standard Deviation, and Coefficient of Variation.

UNIT 3

Probability and Distributions: Sample space, events, Equally likely events. Probability and types; Different Approaches, Independent Events, Addition and multiplication rules, Conditional probability, Rules for Calculating Probabilities.

UNIT 4

Correlation & Regression: Methods of studying simple correlation: Scatter Diagram, Karl Pearson's Co-efficient of Correlation, Spearman's Rank Correlation. Multiple and Partial Correlation, Correlation ratio and intra-class correlation; Regression Analysis: Introduction: Meaning and Purpose of Regression. Simple and Multiple Linear Regressions using Least Square Principle, Coefficient of Determination.

UNIT 5

Analysis of Variance and Testing Hypothesis: Assumptions for Use of the ANOVA, One-Way ANOVA, F-test, Two-way ANOVA for the Randomized Block Design, Linear Model Representation of the ANOVA. Sampling, Test of Significance, Testing Hypotheses about the Variance, Testing Hypotheses about the difference of two means, Testing Hypotheses about the difference of two Proportions, Tests of Hypotheses and Sample Size. Parametric and nonparametric test. The Goodness- of-Fit Test; Chi Square Test.

MCSB 105	Programming in Perl	CS	25	75	100	2-0-0	2
-----------------	----------------------------	----	----	----	-----	-------	---

UNIT1

8 Hours

Concepts of flowcharting, algorithm development, pseudo codes etc.

UNIT 2

9 Hours

Introduction: Use of Perl in Bioinformatics, history of Perl, Availability, Support, Basic Concepts, Scalar Data: Numbers, Strings, Scalar Operators, Scalar Variables, Scalar Functions

UNIT 3

9 Hours

Basic I/O, Array operators and Functions, Scalar and List Context , Control Structures: Statement Blocks

UNIT 4

9 Hours

Hash Functions, Hash Slices, Regular Expressions, Subroutines, Miscellaneous Control Structures

UNIT 5

6 Hours

File handles and File Tests, Directory Handles, File and Directory Manipulation, Process Management: Using system and exec, Using Backquotes

Reference Books:

1. Wall, Larry, Tom Christiansen, bryan d foy, and Jon Orwant. "Programming Perl: Unmatched power for text processing and scripting, edition 4", Sebastopol, CA: O'Reilly Media, 2012.
2. Swartz, Randal L., brian d foy, and Tom Phoenix, "Learning Perl: Making Easy Things Easy and Hard Things Possible, Edition 7", 2016
3. Beginning Perl (first edition) by Simon Cozens free on-line at: <http://www.perl.org/books/beginning-perl>
4. <http://perldoc.perl.org>

PROGRAM ELECTIVE I

MCSB PE 106	Biomolecules	SB	25	75	100	2-0-0	2
------------------------	---------------------	----	----	----	-----	-------	---

UNIT 1

Physicochemical Principles of biomolecular structure organization: Basic concepts of atom structure, hybridization of atomic orbitals, valence, covalent bonds, atomic interactions and forces, formation of polymeric molecules

UNIT 2

Co-ordinate systems: Rectangular, Cylindrical and spherical coordinate systems

Experimental Methods for determination of biomolecular structures: X-ray Diffraction & NMR Spectroscopy

UNIT 3

Protein Structure : Internal Coordinates – Bond lengths, bond angles, torsional angles; peptide unit, Ramachandran Map; Calculation of dihedral angles; Fourth atom fixing Hierarchical, Organization of Protein structure - Primary, Secondary, Supersecondary, Tertiary and Quaternary structure; Membrane Protein Structures, Principles of protein folding and Energetics; Mechanism of Protein folding; methods to study protein folding.

UNIT 4

DNA and RNA Structure: Base pairing in DNA & RNA, Double Helix – Organization, types and structural features, Structural & Geometric parameters associated with DNA, Secondary structures of DNA (triple helices, quadruplex, cruciform), Sequence-structure relationships in DNA, Secondary Structures in RNA; Representations of RNA structures, Energetics of RNA structure, tRNA structure.

UNIT 5

Carbohydrates: Conformations of mono and oligosaccharides and Correlation with properties, Glycoproteins, Proteoglycans and Glycolipids: Structural aspects, Lipids & Membranes – Membrane microdomains and organization.

MCSB PE 107	Bioinformatics tools and development	BS	25	75	100	2-0-0	2
------------------------	---	----	----	----	-----	-------	---

Unit 1

10 Hours

Introduction to Matlab, Matrix calculations, Functions. ENTREZ, NCBI, Genomic databases, Organism specific databases, BLAST, FASTA, Philpp, Prosite. Visualization: Chimera, Pymol, RasMol, Cn3D, discovery studio, Meastro Visulizer etc. Aligning nucleotide sequences; designing a degenerate primer of 20 bases from nucleotide

Unit 2

9 Hours

Alignment data and calculate the level of degeneracy of this primer.Exploring NCBI Resouces, GenBank; Molecular databases (sequence databases - nucleotides, proteins, structures, genes, expression, taxonomy; Marker databases – dbSNP, dbSTS, dbEST, Unigene), genomes and maps, tools (Entrez, BLAST, nucleotide sequence analysis, protein sequence analysis, 3-D structure display and similarity searching).

Unit 3

9Hours

Literature Databases (PubMed, PubMedCentral, OMIM, Books, Citation Matcher), research at NCBI, education, FTP site.

Unit 4

10 Hours

File format conversion: FmtSeq, ReadSeq, Sequence manipulation Suite.Mutation analysis tool: Dynamut, BioEdit etc.

Protein structures: Visualizing and analysis of inter atomic distances, H-bond calculations,secondary structure analysis and salt bridge analysis of protein structures using different software. Prediction of 3D structure of protein.

Unit 5

10Hours

Searching and retrieval of nucleotide sequences from GenBank database, Retrieval of protein sequences from “SWISS-PROT” database. Searching and retrieving the protein

structure data using Entrez and protein viewing softwares. Database homology search of a nucleotide sequence using Blastn, Retrieval of related sequences in a given format.

Modeling solvers – Euler, Runge-Kutta, numerical integration.

References:

1. Jin Xiong: Essential Bioinformatics, Cambridge University Press Tramontano: Introduction to Bioinformatics, Chapman and Hall Series
2. <https://www.ncbi.nlm.nih.gov/>
3. <https://www.rcsb.org/> Course outcome: Students should be able to apply basic bioinformatic tools for DNA/RNA/protein sequence alignments, finding gene/protein homologs, designing primers, identifying mutations, reconstructing phylogenetic trees, etc.

MCSB PE 108	Advanced Bioinformatics applications	BS	25	75	100	2-0-0	2
------------------------	--	----	----	----	-----	-------	---

UNIT-I COMMERCIAL BIOINFORMATICS

10

Hours

History of Bioinformatics – Goal of bioinformatics as a separate discipline – Survey of bioinformatics companies in India and abroad – Emerging branches of Bioinformatics: Genomics, Proteomics, Systems Biology, Chemoinformatics, and Synthetic biology.

UNIT- II GENOME ANALYSIS

9

Hours

Sequence assembly and Finishing methods - Sequence assemblers – finishing and visualization programmes - Gene expression analysis - Data collection - Image processing - Measures of expression - Finding significant genes - Clustering approaches – SNP – Types - SNP discovery methods - databases and browsers - Applications of Bioinformatics in Gene Therapy.

UNIT -III PHARMACEUTICAL BIOINFORMATICS AND DRUG DISCOVERY

9

Hours

Scope of Pharmaceutical bioinformatics and drug discovery - Pharmacogenetics and individual medicine – Medical applications: Genetic basis of disease – Cancer – Drug discovery and testing – Personalized medicine - Genomics and proteomics in medicine, diagnostics, drug discovery and target finding.

UNIT -IV TRANSGENIC PLANTS AND ANIMALS

10

Hours

Transgenic plants and animals - Need for Transgenic plants and animals - Applications of transgenic animal technology - Ethical issues of transgenic plants, animals and other organisms.

UNIT-V BAR CODING

10 Hours

Bar Code of Life - Need for bar coding-Uses of bar coding - Applications of Bioinformatics in Biodiversit

Reference Books:

1. T. A. Brown, Genomes, 6th Edition, BIOS Scientific Publishers, Ltd., Oxford, UK, 2006.
2. Baxevanis D and Ouellette BFF, Bioinformatics: A practical guide to the analysis of genes and proteins (3rd Ed), John Wiley & Sons, Inc., 2005.
3. Jenson, O.N., in Proteomics. A Trends Guide (eds Black Stock, Co- and Mann), Elsevier Science, London, 1998.
4. S.R.Pennington and M.J.Dunn, Proteomics, Viva Books Pvt. Ltd., New Delhi, 2002.
5. R.C. Dubey, A Textbook of Biotechnology, S.Chand & Company, 1993.
6. Paul D. N. Hebert, Alina Cywinska, Shelley L. Ball and Jeremy R. deWaard, Biological identifications through DNA barcodes, Proc. Roy. Soc. (London) B, 270 (2003) 313 – 321.
7. Mark Y. Stoeckle and Paul D. N. Hebert, Barcode of Life, Scientific American, October 2008, 82 – 88.

2nd Semester

MCSB 201	Structural Bioinformatics & drug Design	SB	25	75	100	3-1-0	4
----------	---	----	----	----	-----	-------	---

Detailed Syllabus

UNIT 1

10 Hours

Fundamentals of macromolecular structure: Functional significance of structure, principles of protein structure: secondary structures, motifs, domains, tertiary and quaternary structures. Relationship between sequence and 3D structure of a protein. Structural implications of the peptide bond; rigid planar peptide unit; cis and trans configuration; conformations of a pair of linked peptide units; Ramachandran plot and Protein Data Bank. Conformations of biomacromolecules, Thermodynamics of protein folding, Sequence Motif Databases, Pfam, PROSITE, Protein Structure Classification; SCOP, CATH, Other Relevant Databases, KEGG, Fold Recognition, Prediction of binding pockets on protein structures, Structure-based function Prediction, Prediction of RNA structures.

UNIT 2

10 Hours

Protein Structure Alignments: Structure Superposition, RMSD, Different Structure Alignment Algorithms, DALI, and TM-align.

Protein Structure Visualization Tools: (PyMOL, Chimera, VMD), Protein Structure Comparison Tools and Methods, Protein Structural Alignment (Dali, CE, TM-Align), Protein Structure Classification and Databases (PDB, CATH, SCOPE, CDD), Drug design software, Enzyme kinetics and inhibition. Rational approaches to lead discovery based on traditional medicine, Random screening, Non-random screening, serendipitous drug discovery. Pharmacophore, Lock-Key principle and induced fit

UNIT 3

10

Hours

Structure Prediction: Protein Secondary Structure Prediction, Chou-Fasman, GOR Method, Artificial Neural Networks, Protein Secondary Structure Prediction Tools and

Severs (DSSP, STRIDE, Jpred, Psipredetc), Protein Tertiary Structure Prediction, Homology Modelling, Refinement of the homology model, Structure Validation Strategies, Comparison of various strategies in homology modelling, Prediction of protein structures by threading, Energy Minimization methods and Conformational Analysis, global conformational minima determination.

UNIT 4

10

Hours

Drug Design: Introduction to drug discovery, Current approaches and philosophies in drug design, Molecular Modelling and virtual screening techniques, lead identification, Concept of pharmacophore mapping and pharmacophore-based Screening, Analysis of the receptor, Docking: Rigid docking, flexible docking, manual docking, docking based screening, Scoring Functions, Binding energy calculations De novo drug design., Drug likeness screening, ADMET properties.

UNIT 5

10

Hours

Quantitative Structure Activity Relationship (QSAR): SAR versus QSAR, QSAR methodology, Quantum-chemical descriptors, COMFA and COMSIA. Molecular mechanisms of diseases and drug action. Pharmacokinetics: Models and Applications, Drug-receptor interactions. Pharmacodynamics, Biomolecular interactions and binding thermodynamics, Biochemical and cell-based assays. Molecular recognition and specificity. Future developments.

MCSB 202	Designing algorithms for Bioinformatics	CS	25	75	100	3-1-0	4
----------	---	----	----	----	-----	-------	---

UNIT I - DYNAMIC PROGRAMMING ALGORITHMS (10hours)

Introduction to Algorithms, Dynamic Programming, Sequence Alignment: Edit distance, LCS. PAM and BLOSUM Scoring Matrices. Global alignments: Needleman Wunsch Algorithm, Local Alignments: Smith Waterman Algorithm, Gap Penalties.

UNIT II- GRAPH ALGORITHMS (9 hours)

Graph Algorithms, SBH and Eulerian Paths, De-novo Peptide Sequencing: Longest Paths and Space Efficient Alignment Algorithms. Fast LCS using Table Lookup.

UNIT III- PATTERN MATCHING AND CLUSTERING (9 hours)

Exact Pattern Matching: KMP Algorithm, Keyword Trees, Aho-Corasick Algorithm. Clustering Basics: Hierarchical Clustering, Multiple Sequence Alignment: CLUSTAL, Center-based Clustering, Clustering via Cliques.

UNIT IV- EVOLUTIONARY TREES AND PHYLOGENY (9 hours)

Evolutionary Trees and Ultrametrics, Additive distance trees, Perfect Phylogeny Problem, Small Parsimony Problem, Nearest Neighbor Interchange.

UNIT V- HIDDEN MARKOV MODELS, RANDOMIZED ALGORITHMS

(11 hours)

Hidden Markov Models: Basics, Forward and Backward (Viterbi) Algorithms, Randomized algorithms and their applications.

Reference Books:

1. Neil C. Jones and Pavel A. Pevzner, “*An Introduction to Bioinformatics Algorithms*”, MIT Press, 2005.
2. Gusfields D, “*Algorithms on strings, trees and sequences: Computer Science and Computational Biology*”, Cambridge University Press, 1997.
3. Steffen Schulze-Kremer, “*Molecular Bioinformatics: Algorithms and Applications*”, Walter de Gruyter, 1996.
4. Gary Benson, Roderic Page (Eds.), “*Algorithms in Bioinformatics*”, Springer International Edition, 2004.
5. Richard Durbin, Sean R. Eddy, Anders Krogh, Graeme Mitchison. “*Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acid*”, Cambridge University Press, 1999.

MCSB 203	Database Management System	CS	25	75	100	3-1-0	4
----------	----------------------------	----	----	----	-----	-------	---

Unit – I

Introduction to DBMS

8 Hours

Basics of File Processing Systems and Database Systems, Difference between traditional file system and DBMS, Responsibilities of Database Administrator, Three level Architecture of Database System, Physical and Logical data independence.

Unit – II

Introduction to various Database Models

10 Hours

Entity Relationship Model and its importance, Introduction to various Symbols used In ERD (Entity: Types of Entities, weak Entity, Composite Entity, Strong Entity, Attribute: Types of Attribute, Relationship: Type of relationship, Connectivity, Cardinality), Conversion of ER diagram to tables, Comparative study of Network, Hierarchical and Relational Models, Codd's 12 Rules, Comparison of Object Oriented Database and Object Relational Database.

Unit – III

11 Hours

Normalization in DBMS

Normalization and its various forms(1NF, 2NF, 3NF and BCNF), Functional Dependencies, Multi- valued Dependencies, Study of various Database Integrity like Domain, Entity, Referential Integrity Constraints.

Unit – IV

10 Hours

SQL

Categories of SQL Statements, The CREATE Statement, The DROP Command, The ALTER Command, Integrity Constraints, DML Statements: The SELECT Statement, The INSERT Statement, The DELETE Statement, The UPDATE Statement, SQL Operators: Simple Selects Comparison Operators, IN and NOT IN Operators, BETWEEN Operator, The LIKE Operator Logical

Operators, IS NULL and IS NOT NULL, ANY, ALL, SQL FUNCTIONS, Joining Tables, SQL Subqueries, GROUP BY Clause, HAVING Clause

Unit-V

Advanced Database Concepts

9 Hours

Emerging database technologies and applications: Spatial databases, Mobile databases, Multimedia databases, Genome data management, Overview of Data Warehouse, OLAP, Big data, biological big data, and big data analytics

Reference Books:

1. R. Elmasri & S.B. Navathe, Fundamentals of Database Systems, Pearson Education, 6th

edition, 2010.

2. Silberschatz, H. Korth & S. Sudarshan, Database System Concepts, TMH, 5th Edition, 2010.
3. R. Ramakrishnan & J. Gehrke, Database Management Systems, 3rd edition, TMH, 2007.

MCSB 204	Predictive Modelling and Analytics	and	CS	25	75	100	3-1-0	4
----------	------------------------------------	-----	----	----	----	-----	-------	---

Unit 1:

9 Hours

Introduction to Data Mining Introduction, what is Data Mining? Concepts of Data mining, Technologies Used, Data Mining Process, KDD Process Model, CRISP – DM, Mining on various kinds of data, Applications of Data Mining, Challenges of Data Mining.

Unit 2

10 Hours

Data Understanding and Preparation Introduction, Reading data from various sources, Data visualization, Distributions and summary statistics, Relationships among variables, Extent of Missing Data. Segmentation, Outlier detection, Automated Data Preparation, Combining data files, Aggregate Data, Duplicate Removal, Sampling DATA, Data Caching, Partitioning data, Missing Values.

Unit 3.

10 Hours

Additive Models, Trees, and Boosting: Generalized additive models, Regression and classification trees, Boosting methods-exponential loss and AdaBoost, Numerical Optimization via gradient boosting, Examples (Spam data, California housing, New Zealand fish, Demographic data)

Unit 4. .

10 Hours

Neural Networks (NN), Support Vector Machines (SVM), and K-nearest Neighbor: Fitting neural networks, Back propagation, Issues in training NN, SVM for classification, Reproducing Kernels, SVM for regression, K-nearest –Neighbour classifiers (Image Scene Classification)

Unit 5.

9 Hours

Unsupervised Learning and Random forests: Association rules, Cluster analysis, Principal Components, Random forests and analysis.

Reference Books:

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning- Data Mining, Inference, and Prediction, Second Edition, Springer Verlag, 2009.
2. G. James, D. Witten, T. Hastie, R. Tibshirani- An introduction to statistical learning with applications in R, Springer, 2013
3. C.M. Bishop – Pattern Recognition and Machine Learning, Springer, 2006

MCSB 205	Chemoinformatics ,Chemogenomics and Immunoinformatics.	SB	25	75	100	2-0-0	2
----------	---	----	----	----	-----	-------	---

UNIT 1

10 Hours

Chemoinformatics: Basic concepts of Cheminformatics, Application of and modern prospective of Cheminformatics. Manipulations in 2D and 3D structures of chemical compounds, representation of chemical reactions, molecular descriptors, calculations of physical and chemical data, calculation of structural deciphers. In silico representation of chemical information (SMILES, IchI, SDF, MOL, PDB, PDBQT, etc), Computer-oriented chemical compounds nomenclature, Computer-Assisted structure elucidation. Molecular drawing and interactive visualization, Chemical name and formula searching. Pharmaceutical Chemistry.

UNIT 2

10 Hours

Molecular Structures: Cheminformatics Databases, Chemical Database Design, Bio Catalysts Database, The MOS Database, The Failed Reaction Database, Protecting Groups Database, Solid- Phase Synthesis Database, Compound Databases (ZINC, PubChem, ChEMBL, DrugBank, etc.). Chemistry & Information technology, chemical data collation, retrieval, analysis & interpretation, hypothesis generation & validation, development of structure activity/property relationships, artificial intelligence techniques in chemistry. Building molecules on a computer, quantum and molecular mechanics methods for geometry optimization, Simulation methods for molecules and materials. Representation of reactions: SMARTS, SMIRKS, Matrices. Searching chemical structures - Exact (Canonicalizing SMILES: Morgan, CANGEN), Substructure, Superstructure and Similarity search (Tanimoto, Euclidean and Tversky index), Clustering small molecules - Hierarchical and non-hierarchical methods, 3D pharmacophore-based searching: Common Pharmacophore features, building hypotheses and searching databases (5)

UNIT 3

9 Hours

Personalized Medicine: Computational Chemistry in drug design, Role of Pharmacogenomics in Drug Discovery & Therapeutics. Pharmacogenetics: Biomarkers, Promise of Personalized Medicine. Genetic drug response profiles, Effect of drug on gene expression, Pharmacogenetics, Drug Metabolism. Pharmacogenomics of drug transporters-OATP, OAT, OCT, OCTN, PepT, MRP, MDR, ABC transporter. Interethnic Differences in drug response. Drug metabolism.

UNIT 4

9 Hours

Combinatorial Chemistry: Combinatorial chemistry technologies & libraries, Solution phase synthesis, High-Throughput Synthesis and Screening, Target-oriented synthesis: Designing organic synthesis, Retrosynthetic analysis, disconnection approach, linear and convergent synthesis. Diversity-oriented synthesis: concept of forward-synthetic analysis, appendage diversity, skeletal diversity, stereo-chemical diversity, complexity and diversity. Asymmetric Synthesis: Use of chiral auxiliaries, chiron approach. Principles and use of enzymes in the synthesis of industrially important sugar / fatty acid esters, sugar nucleotide derivatives.

UNIT 5

10 Hours

Immunoinformatics overview, Reverse vaccinology, Rational Vaccine design, Evolution and escape due to variations. Immunoinformatics databases & tools. Molecular interactions, Epitope prediction algorithms:

Reference Books

- 1. Flower Darren R. Bioinformatics for Vaccinology Publisher: UK, John Wiley & Sons Inc. Latest Edition.
- 2. Flower Darren R. Immunoinformatics: Predicting Immunogenicity In Silico
3. Gasteiger Johann, Engel Thomas. Chemoinformatics: A Textbook. Publisher: Wiley-VCH; 1st edition. 2003. ISBN: 3527306811.
4. Leach Andrew R., Valerie J. Gillet. An introduction to chemoinformatics. Publisher: Kluweracademic, 2003. ISBN: 1402013477.
5. Gasteiger Johann, Handbook of Chemoinformatics: From Data to Knowledge (4 Volumes). Publisher: Wiley-VCH. 2003. ISBN: 3527306803.
6. Muthukumarasamy Karthikeyan, Renu Vyas. Practical Chemoinformatics. Publisher:
7. Springer. 2014. ISBN: 9788132217794
8. Bajorath Jürgen. Chemoinformatics and computational chemical biology. Publisher: Humana Press. 2011. ISBN: 9781607618386

PROGRAM ELECTIVE II

MCSB PE 206	Research Methodology	CS	25	75	100	2-0-0	2
-------------	----------------------	----	----	----	-----	-------	---

UNIT I –RESEARCH FORMULATION AND DESIGN 10 Hours

Motivation and objectives – Research methods vs. Methodology. Types of research – Descriptive vs. Analytical, Applied vs. Fundamental, Quantitative vs. Qualitative, Conceptual vs. Empirical, concept of applied and basic research process, criteria of good research. Defining and formulating the research problem, selecting the problem, necessity of defining the problem, importance of literature review in defining a problem, literature review-primary and secondary sources, reviews, monograph, patents, research databases, web as a source, searching the web, critical literature review, identifying gap areas from literature and research database, development of working hypothesis.

UNIT II – DATA COLLECTION AND ANALYSIS 10 Hours

Accepts of method validation, observation and collection of data, methods of data collection, sampling methods, data processing and analysis strategies and tools,data analysis with statically package (Sigma STAT,SPSS for student t-test, ANOVA, etc.), hypothesis testing.

UNIT III – SOFT COMPUTING 9 Hours

Computer and its role in research, Use of statistical software SPSS, GRETL etcin research. Introduction to evolutionary algorithms - Fundamentals of Genetic algorithms, Simulated Annealing, Neural Network based optimization, Optimization of fuzzy systems.

UNIT IV –RESEARCH ETHICS, IPR AND SCHOLARY PUBLISHING 9Hours

Ethics-ethical issues, ethical committees (human & animal); IPR- intellectual property rights and patent law, commercialization, copy right, royalty, trade related aspects of intellectual property rights (TRIPS); scholarly publishing- IMRAD concept and design of research paper, citation and acknowledgement, plagiarism, reproducibility and accountability.

UNIT V –INTERPRETATION AND REPORT WRITING 10 Hours

Meaning of Interpretation, Technique of Interpretation, Precaution in Interpretation, Significance of Report Writing, Different Steps in Writing Report, Layout of the Research Report, Types of

Reports, Oral Presentation, Mechanics of Writing a Research Report, Precautions for Writing Research Reports, Conclusions.

Reference Books

1. Garg, B.L., Karadia, R., Agarwal, F. and Agarwal, U.K., 2002.
2. An introduction to Research Methodology, RBSA Publishers. Kothari, C.R., 1990.
3. Research Methodology: Methods and Techniques. New Age International. 418p. Sinha, S.C. and Dhiman, A.K., 2002.
4. Research Methodology, Ess Ess Publications. 2 volumes. Trochim, W.M.K., 2005.
5. Research Methods: the concise knowledge base, Atomic Dog Publishing. 270p. Wadehra, B.L. 2000.
6. Law relating to patents, trade marks, copyright designs and geographical indications. Universal Law Publishing.

MCSB PE 207	Programming in Java	CS	25	75	100	2-0-0	2
-------------	---------------------	----	----	----	-----	-------	---

Unit 1 **9 Hours**

Java Basics – Importance and features of java, Modifiers, Access Controls, Data types, Expressions, Declarations, Statements & Control Structures, Program Structures, String handling, Packages, Interfaces, Working with java util Package, Garbage Collection

Unit 2 **11 Hours**

Exception Handling, I/O & JDBC – Exception Handling: built in exception, creating your own exceptions, Input Stream & Output Stream: Streams, Byte and Character stream, Predefined streams, Reading and Writing from Console and Files, Buffered Reader & Writer, Serialization, Database: JDBC Basics

Unit 3 **10 Hours**

Multithreading and Communication – Java Thread Model: Life Cycle of Thread, Thread class, Runnable interface, Inter thread Communication, Suspending, Resuming and Stopping threads, Synchronization, Scheduling and Priority of Threads.

Unit 4 **10 Hours**

AWT & Event Handling – Creating User interface with AWT, Applets, Applet Life Cycle, Simple Graphics, Fonts and Colors, Events, Listeners, Components, Containers, Working with Layouts, Event Classes, Event Listener Interfaces, Adapter and Inner Classes

Unit 5 **9 Hours**

BioJava – Installing BioJava, Symbols, Basic Sequence Manipulation (DNA to RNA, Reverse Complement, motif as regular expression), Translation (DNA to Protein, Codon to amino acid, Six frame translation), Proteomics (Calculate the mass and pI of a peptide), Sequence I/O (File Formats conversions), Locations and Features (PointLocation, RangeLocation, Feature modifications), BLAST and FASTA (Blast and FastA Parser, extract information from parsed results), Counts and Distributions, Weight Matrices and Dynamic Programming, User Interfaces.

Reference Books

1. Herbert Schildt, Java- A Beginners Guide (4th Ed.), Tata Mc-Graw-Hill publication. 2007

2. Computing Concepts with Java 2 Essentials (2nd Ed.) by Horstmann, C.S., John Wiley Publishers. 2000
3. Object Oriented Design and Applications (2nd Ed.) by Benjamin, Cummings and Booch, G., Addison Wesley Publishers. 1994

MCSB PE 208	Intellectual Property Rights	CS	25	75	100	2-0-0	2
-------------	------------------------------	----	----	----	-----	-------	---

Unit-1 Introduction to Intellectual property

Introduction, types of intellectual property, international organizations, agencies and treaties, importance of intellectual property rights.

Unit-2 Trade Marks

Purpose and function of trademarks, acquisition of trade mark rights, protectable matter, selecting, and evaluating trade mark, trade mark registration processes.

Unit-3 Law of copy rights

Law of copy rights: Fundamental of copy right law, originality of material, rights of reproduction, rights to perform the work publicly, copy right ownership issues, copy right registration, notice of copy right, international copy right law.

Law of patents: Foundation of patent law, patent searching process, ownership rights and transfer

Unit-4 Trade Secrets

Trade Secrets: Trade secrete law, determination of trade secrete status, liability for misappropriations of trade secrets, protection for submission, trade secrete litigation.

Unfair competition: Misappropriation right of publicity, false advertising.

Unit-5 New development of intellectual property

New development of intellectual property: new developments in trade mark law; copy right law, patent law, intellectual property audits. International overview on intellectual property, international – trade mark law, copy right law, international patent law, and international development in trade secrets law.

TEXT BOOKS & REFERENCES:

1. Intellectual property right, Deborah. E. Bouchoux, Cengage learning.
2. Intellectual property right – Unleashing the knowledge economy, prabuddha ganguli, Tata McGraw Hill Publishing company ltd.

MCSB PE 209	Comparative and Functional Genomics	SB	25	75	100	2-0-0	2
-------------	-------------------------------------	----	----	----	-----	-------	---

Unit 1 Comparative genomics:

10 Hours

Databases for genomics resources, Technologic principles of currently utilized methods within the field of functional genomics. Sequence alignment: global versus local, inferring gene function from relatedness to other genes, Finding Open Reading Frames, overall structure and organization of the genome in microorganisms, plants and animals. Extra-chromosomal elements: Mitochondrial genome, mitochondrial plasmids. Organization of nuclear and organellar genomes; Repetitive DNA-satellite DNAs and interspersed repeated DNAs, Transposable elements, LINES, SINES, Alu family and their application in genome Mapping.

Unit 2 Molecular Phylogenetics:

10 Hours

Fine structure of gene, split genes, pseudogenes, non-coding genes, overlapping genes and multi-gene families. Genome sequencing methods, Genome annotation at different levels, Comparative genome sequencing. Molecular Phylogenetics Origins, Applications of Molecular Phylogenetics. Phylogenomic comparisons, introduction to programs, Phylogenetic analyses: tree terminology and parsimony, Phylogenetic analyses to determine relationships and interpret character evolution. Single gene disorders-conventional and contemporary methods. Candidate gene identification; Genetic polymorphism and disease susceptibility.

Unit 3 Databases and Software for Genome Biology:

10 Hours

Gene Expression Analysis using Microarrays and RNA-Seq, Application of DNA microarrays for the analysis of gene expression, protein-DNA binding, chromatin structure, chromatin modifying complexes, and RNA polymerase occupancy. Error models and data normalization techniques for high-resolution array technologies. Clustering genes into sets and discovering gene set features that can be used for diagnostic purposes. Importance of chromatin structure in contemporary modelling, relationship between chromatin structure and transcriptional regulation.

Unit 4 Functional Genomics-I:

10 Hours

Determining the Functions of Individual Genes, Pathway and GO annotation systems, Non-coding DNA, gene prediction methods and tools. Genome Evolution, Genome and Transcriptome Assembly Algorithms, Profile HMMs for Protein Family Modelling, Human Genome Project, Genomics and Personalized medicine, Sex-linked inheritance: Conceptual basis, sex influenced traits, mechanism of sex determination. Mutation variability across the genome. Mutation detection, diagnosis and therapy. Genes and Environment - heritability, penetrance and expressivity.

Unit 5 Functional Genomics-II:

10 Hours

Integration of genomic, epigenetic and external effects in the context of understanding gene and genome function, Genome regulation: Introduction, levels of regulation, evidence and experimental designs/methodologies, role of genetic analysis in understanding gene function and regulation. Transcriptional control - DNA sequence elements and transcription factors. Mutation – Classification, mechanism, repair, role in genetic analysis and evolution. Post- transcriptional regulation – Alternative RNA

Reference

- 1.Hartl & Ruvolo. **Genetics: Analysis of Genes and Genomes.** Jones & Bartlett Learning Mount.
- 2.Bioinformatics sequence and Genome analysis. Cold Spring Harbor Laboratory Press Deonier, Tavaré, & Waterman. Computational Genome Analysis: An Introduction. Springer

3RD SEMESTER

MCSB 301	NGS Data Analysis- Microarray, RNA Seq, Single cell sequencing	SB	25	75	100	2-0-0	2
----------	---	----	----	----	-----	-------	---

UNIT I:

9 Hours

History & evolution of sequencing: First-generation technologies – Sanger dideoxy sequencing, Maxam-Gilbert sequencing. Technologies used in Human Genome Project. Shotgun sequencing. Advantages and disadvantages of first generation sequencing. Next (second)-generation sequencing – difference between first and next generation sequencing, NGS platforms – Roche 454, ABI SOLiD, Ion torrent, Illumina. Advantages and disadvantages of second generation sequencing.

UNIT II:

9 Hours

Third-generation sequencing: PacBio, Oxford nanopore. Advantages and disadvantages of third generation sequencing. Comparison of output, accuracy and types of errors of first, second and third generation sequencing technologies. Introduction to NGS technologies: DNA-seq, RNA-seq, ChIP-seq, Hi-C, Metagenomics, Single cell sequencing.

UNIT III:

10 Hours

Preparation for sequencing: Different sample preparation methods for different type of NGS (DNaseq, RNASeq, ChIPseq, Metagenomics, Single cell), Adaptors, Index, Barcode. Library preparation methods - Bridge amplification, Emulsion PCR. Sequencing methods – sequencing by synthesis, ion semiconductor, SMRT, nanopore.

UNIT IV:

10 Hours

NGS Data formats & Pre-processing: Data formats overview – FASTQ, subreads, nanopore data, single cell data. Single-end, Paired-end, Mate-pair. NGS Data sources – NCBI SRA, EBI-ENA, DDBJ-SRA, GEO; Retrieving data from data sources - SRA toolkit; Aspera connect. Sequence quality measures – Phred quality score. Quality check – tool – FASTQC, Pre-processing: Trimmomatic, Fastx-toolkit.

UNIT V:

10 Hours

Introduction to NGS Data Analysis: Assembly (overview) – principles, output file formats – contigs, scaffolds, assembly quality assessment – N50, total length, no. of contigs/scaffolds; Mapping (overview) – Principles, tools – BWA, Bowtie, SAMtools, output file formats – BAM, SAM, mapping alignment assessment – no. of reads mapped, concordant reads; Visualisation tools – IGV

Reference:

1. Dr. Michal Janitz. (2008), Next Generation Genome Sequencing: Towards Personalized Medicine, Wiley.
2. Stuart M Brown. (2013), Next-Generation DNA Sequencing Informatics, Cold Spring Harbor Laboratory Press.
3. Sara El-Metwally, Osama M. Ouda & Mohamed Helmy. (2014), Next Generation Sequencing Technologies and Challenges in Sequence Assembly, Springer, New York, NY.
4. Gaurav Sablok, Sunil Kumar, Saneyoshi Ueno, Jimmy Kuo & Claudio Varotto. (2015), Advances in the Understanding of Biological Sciences Using Next Generation Sequencing (NGS) Approaches, Springer International Publishing.
5. Xinkun Wang. (2016), Next-generation sequencing data analysis, CRC Press.

6. Urszula Demkow and Rafał Płoski. (2016), *Clinical Applications for Next Generation Sequencing*, Academic Press.
7. Masoudi-Nejad, Ali, Narimani, Zahra, Hosseinkhan & Nazanin. (2013), *Next Generation Sequencing and Sequence Assembly - Methodologies and Algorithms*, Springer, New York.
8. Ion Măndoiu & Alexander Zelikovsky. (2016), *Computational Methods for Next Generation Sequencing Data Analysis*, Wiley.
9. Jerzy K. Kulski (2016), *Next Generation Sequencing - Advances, Applications and Challenges*, InTech
10. Emmanuel A. Kornyo. (2017), *A Guide to Bioethics*, CRC Press.
11. Richard McCombie W, Elaine R. Mardis, James A. Knowles & John D. McPherson. (2019), *Next-Generation Sequencing in Medicine*, Cold Spring Harbor Laboratory Press.
12. Wing-Kin Sung. (2020), *Algorithms for Next-Generation Sequencing*, Chapman and Hall/

MCSB 302	Programming in R and Python	CS	25	75	100	2-0-0	2
----------	-----------------------------	----	----	----	-----	-------	---

Unit 1: R Basics

10 Hours

Basic operations in R, Math operations in R, Vector , working with null values, Import & Export files in R, Data-frame, Joins, One-way and Two way tables, Vectors, Matrices

Unit II: Python Basics

10 Hours

Objects and Functions, Identifiers, Variables and Datatypes, Operators, Python Flow, Function Arguments, Recursive functions ,Lambda, Exception Handling , Iterators, Generators and Decoders.

Unit III: Mathematics for Data science

10 Hours

Probability, Statistics, Linear Algebra, Gradient Descent, Calculus for data science, ANOVA, Hypothesis testing.

Unit IV: Numpy and Pandas Numpy:

10 Hours

Arrays, Vectorization, Boolean Indexing, Matrix multiplication, Tuple, Join/Merge data, Unicode strings etc. Pandas: Data Structure, Data frame, Reading data, Handling missing data.

Unit V: Exploratory Data Analysis

9 Hours

Data Visualization using GGLOT2 and Matplotlib, Data Pre-processing, Data Transformation, Data Reduction, Feature Extraction. Univariate and Multi-variate analysis.

Reference Books:

1. Business Analytics (The science of data driven decision making) Kumar
2. [https://dataflair.training/blogs/data-- manipulationinr/](https://dataflair.training/blogs/data--manipulationinr/) U Dinesh
3. Machine Learning For Dummies, John Paul Mueller, Luca Massaron
4. Python for Dummies, Aahz Maruch and Stef Maruch
5. Introduction to Machine Learning with Python Muller Andreas

MCSB 303	Data Warehousing and Data Mining	CS	25	75	100	2-0-0	2
----------	----------------------------------	----	----	----	-----	-------	---

Unit I

Overview, Motivation(for Data Mining),Data Mining-Definition & Functionalities, Data Processing, Form of Data Preprocess-ing, Data Cleaning: Missing Values, Noisy Data, (Binning,Clustering, Regression, Computer and Human inspection),Inconsistent Data, Data Integration and Transformation. Data Reduction:-Data Cube Aggregation, Dimensionality reduction, Data 35 Com-pression, Numerosity Reduction, Clustering, Discretization and Concept hierarchy generation

Unit II

Concept Description:- Definition, Data Generalization, Analytical Characterization, Analysis of attribute relevance, Mining Class comparisons, Statistical measures in large Databases. Measuring Central Tendency, Measuring Dispersion of Data, Graph Dis-plays of Basic Statistical class Description, Mining Association Rules in Large Databases, Association rule mining,mining Single-Dimensional Boolean Association rules from Transactional Databases– Apriori Algorithm, Mining Multilevel Association rules from Transaction Databases and Mining Multi-Dimensional Association rules from Relational Databases

Unit III

Classification and Predictions: What is Classification & Prediction, Issues regarding Classification and prediction, Decision tree, Bayesian Classification, Classification by Back propagation, Multilayer feed-forward Neural Network, Back propagation Algo-rithm, Classification methods K-nearest neighbor classifiers, Genetic Algorithm. Cluster Analysis: Data types in cluster analysis, Categories of clustering methods, Partitioning methods. Hierarchical Clustering- CURE and Chameleon, Density Based Methods-DBSCAN, OPTICS, Grid Based Methods- STING, CLIQUE, Model Based Method –Statistical Approach, Neural Network approach, Outlier Analysis

Unit IV

Data Warehousing: Overview, Definition, Delivery Process, Difference between Database System and Data Warehouse, Multi Dimensional Data Model, Data Cubes, Stars, Snow Flakes, Fact Constellations, Concept hierarchy, Process Architecture, 3 Tier Architecture, Data Marting.

Unit V

Aggregation, Historical information, Query Facility, OLAP function and Tools. OLAP Servers, ROLAP, MOLAP, HOLAP, Data Mining interface, Security, Backup and Recovery, Tuning Data Warehouse, Testing Data Warehouse.

Reference Books:

1. H.Dunham,"Data Mining:Introductory and Advanced Topics" Pearson Education.
2. Sam Anahory, Dennis Murray, "Data Warehousing in the Real World : A Practical Guide for Building Decision Support Systems, Pearson Education.
3. Jiawei Han, Micheline Kamber, "Data Mining Concepts & Techniques" Elsevier.
4. Mallach,"Data Warehousing System",McGraw –Hill.

MCSB 304	Data Modelling and Visualization	CS	25	75	100	2-0-0	2
----------	----------------------------------	----	----	----	-----	-------	---

UNIT - I Data-Analytic Thinking

6 Hours

Knowing your data, Data preprocessing, Storytelling with data

UNIT - II Data Modeling

11 Hours

[Linear regression](#), Logistic regression, K-nearest neighbors, K-means clustering, Performance measure, Implementation of some modeling algorithms using R

UNIT - III Data Visualization

11 Hours

Value of Visualization – What is Visualization and Why do it: External representation – Interactivity – Difficulty in Validation. Data Abstraction: Dataset types – Attribute types – Semantics. Task Abstraction – Analyze, Produce, Search, Query. Four levels of validation – Validation approaches – Validation examples. Marks and Channels

UNIT - IV Rules of thumb

10 Hours

Categorical regions – Spatial axis orientation – Spatial layout density. Arrange spatial data: Geometry – Scalar fields – Vector fields – Tensor fields. Arrange networks and trees: Connections, Matrix views – Containment. Map color: Color theory, Color maps and other channels.

UNIT - V Manipulate view

10 Hours

Change view over time – Select elements – Changing viewpoint – Reducing attributes. Facet into multiple views: Juxtapose and Coordinate views – Partition into views – Static and Dynamic layers – Reduce items and attributes: Filter – Aggregate. Focus and context: Elide – Superimpose – Distort – Case studies.

Reference Books:

1. Tamara Munzner, Visualization Analysis and Design, A K Peters Visualization Series, CRC Press, 2014.
2. Scott Murray, Interactive Data Visualization for the Web, O'Reilly, 2013.
3. Alberto Cairo, The Functional Art: An Introduction to Information Graphics and Visualization, New Riders, 2012
4. Nathan Yau, Visualize This: The FlowingData Guide to Design, Visualization and Statistics, John Wiley & Sons, 2011.

PROGRAM ELECTIVE III

MCSB PE 305	Evolutionary Bioinformatics	SB	25	75	100	2-0-0	2
-------------	--------------------------------	----	----	----	-----	-------	---

Unit 1 Evolution of DNA and proteins **9 Hours**

origin of the genetic code. Hardy-Weinberg equilibrium; Evolutionary changes by mutation, gene flow, natural selection and genetic drift.

Unit 2 molecular evolution **9Hours**

Role of transitions and transversions; chromosomal deletions and insertions in evolution. Role of pseudogenes, repetitive DNA, transposable elements and junk DNA in evolution.

Unit 3 Theories **10Hours**

Neutral theory (Kimura) and nearly neutral theory (Ohta) of molecular evolution. Phylogenetic tree. Reconstruction of phylogenetic trees using distance matrix methods, the Maximum Parsimony method, Maximum likelihood and Bayesian inference. Estimation of selection at the molecular level.

Unit 4 Molecular Clock **10 Hours**

The concept of the Molecular Clock. Calibration. Limitation of molecular clock models. Human molecular clock: deducing evolutionary histories through mitochondrial DNA and Y chromosome.

Unit 5 Evolution of the genome: **10 Hours**

Human Genome Project, ENCODE, Genome duplication (Ohno's hypothesis), Exon Shuffling, Concerted evolution. Evolutionary Medicine.

Reference books:

1. An Introduction to Molecular Evolution and Phylogenetics by Lindell Bromham, 2016, Oxford University Press.
2. Molecular Evolution by Wen Hsiung-Li, 1997, Sinauer Associates, Sunderland, MA.
3. Molecular Evolution and Phylogenetics by Masatoshi Nei and Sudhir Kumar, 2000, Oxford University Press.
4. Neutral Theory of Molecular Evolution by Motoo Kimura, 1985, Cambridge University Press.
5. Bioinformatics and Molecular Evolution by Paul G. Higgs and Teresa K. Attwood, 2013, Willey-Blackwell.

MCSB PE 306	Proteomics and metabolomics	SB	25	75	100	2-0-0	2
-------------	-----------------------------	----	----	----	-----	-------	---

UNIT -I INTRODUCTION

10 Hours

Proteomics introduction – Protein sequencing – Protein Digestion Techniques – Mass Spectrometers for Protein and Peptide Analysis – Protein Identification by Peptide Mass Fingerprinting – Software Tools for Peptide Mass Fingerprinting: Finding the Matches – Peptide Sequence Analysis and Protein Identification with Tandem Mass Spectrometry

UNIT -II PROTEOME DATABASES

10 Hours

Proteome databases – Comparative proteomics methods – 2D gel databases – Protein interaction data bases – Metabolic pathway databases – resources for interaction prediction – network and pathway visualization tools – Protein network analysis

UNIT- III FUNCTIONAL PROTEOMICS

10 Hours

Functional proteomics – protein phenotypes – Protein-Protein Interaction Mapping: Experimental – Yeast two-hybrid system – phage display – protein fragment complementation assays – Computational approach

UNIT -IV METABOLOME INFORMATICS

9Hours

Introduction to the ARM Database- The Genome-Based E-CELL Modeling (GEM) System- Large-Scale Simulation of Metabolism-Metabolomics and Medical Sciences

UNIT-V INTRODUCTION TO METABOLIC ENGINEERING

9Hours

Importance of metabolic engineering-comprehensive models for cellular reactions-material balances & data consistency- metabolic pathway synthesis.

References books:

- 1.M. Tomita, T. Nishioka, Metabolomics- The Frontier of Systems Biology, Springer Publications, 2003.
2. Gregory N. Stephanopoulos, Metabolic Engineering- Principles and Methodologies, Academic press, First Edition, 1998.
3. Baxevanis D and Ouellette BFF, Bioinformatics: A practical guide to the analysis of genes and proteins (2nd Edn.), John Wiley & Sons, Inc. 2002.
4. Brown TA, Genomes (2nd Edn.), BIOS Scientific Publishers, Oxford, UK, 2002.
5. Sensen CW, Essentials of Genomics and Bioinformatics, Wiley–VCH. 2002.

MCSB PE 307	Artificial Intelligence and Machine Learning in Bioinformatics	CS	25	75	100	2-0-0	2
-------------	--	----	----	----	-----	-------	---

Detailed Syllabus

Unit 1

10 Hours

Introduction – Definition – Future of Artificial Intelligence – Characteristics of Intelligent Agents – Typical Intelligent Agents – Problem Solving Approach to Typical AI Problems.

Unit 2

10 Hours

Problem Solving Methods – Search Strategies – Uninformed – Informed – Heuristics – Local Search Algorithms and Optimization Problems – Searching with Partial Observations – Constraint Satisfaction Problems – Constraint Propagation – Backtracking Search – Game Playing – Optimal Decisions in Games – Alpha – Beta Pruning – Stochastic Games.

UNIT 3

10 Hours

AI Applications – Language Models – Information Retrieval – Information Extraction – Natural Language Processing – Machine Translation – Speech Recognition – Robot – Hardware – Perception – Planning - Moving.

UNIT 4

10 Hours

Algorithmic models of learning, Learning classifiers, functions, relations, grammars, probabilistic models, value functions, behaviors and programs for experience. Bayesian, maximum some posterior, and minimum description length frameworks.

UNIT 5

10 Hours

Parameter Estimation, sufficient statistics, decision trees, neural networks, support vector machines, Bayesian networks, bag of words classifiers, N-gram models; Markov and Hidden Markov models, probabilistic relational models, association rules, nearest neighbor classifiers, locally weighted regression, ensemble classifiers.

Reference Books

1. Bishop, C. (2006). Pattern Recognition and Machine Learning. Berlin: Springer-Verlag.
2. S. Russell and P. Norvig, —Artificial Intelligence: A Modern Approach||, Prentice Hall, Third Edition, 2009.
3. Bratko, - Prolog: Programming for Artificial Intelligence, Fourth Edition, Addison-Wesley Educational Publishers Inc., 2011.