

**Post Graduate Diploma in Bioinformatics,  
Computational Biology and Drug Discovery**



**Dr. B. R. Ambedkar Center for Biomedical Research (ACBR)**

**University of Delhi, Delhi 110007**

## **CONTENTS**

	<b>Page</b>
<b>I. About the Department and BIF facility</b>	<b>3</b>
<b>II. Introduction to CBCS Scope &amp; Definitions</b>	<b>4</b>
<b>III. PG Diploma Programme Details</b>	<b>6-15</b>
<b>Programme scope, objectives and outcomes</b>	<b>6</b>
<b>Programme Structure</b>	<b>7</b>
<b>Eligibility for Admissions</b>	<b>10</b>
<b>Assessment of students Performance</b>	<b>10</b>
<b>Scheme of exams</b>	<b>10</b>
<b>Pass percentage and Promotion criteria</b>	<b>11</b>
<b>Semester to semester progression</b>	<b>11</b>
<b>Conversion of Marks into grades</b>	<b>12</b>
<b>Grade points</b>	<b>12</b>
<b>CGPA Calculation</b>	<b>12</b>
<b>Attendance requirement</b>	<b>12</b>
<b>Span Period</b>	<b>13</b>
<b>Guidelines for the award of Internal assessment marks</b>	<b>13</b>
<b>Programme Fee</b>	<b>14</b>
<b>IV. Course Wise Content Details for PG Diploma</b>	<b>15</b>

***Dr. B. R. Ambedkar Center for Biomedical Research, University of Delhi***

**I. About the Department**

Dr. B.R. Ambedkar Center for Biomedical Research (ACBR) came into existence in March 1991 with the foundation stone laid by the then Hon'ble Prime Minister of India Sh. Chandra Shekharji, on the occasion of the birth centenary of Baba Saheb Dr. B.R. Ambedkar. The mandate of the Centre is high quality postgraduate education and research in Biomedical Sciences. The institute also has provision for doctoral and postdoctoral training to young scientists at the start of their research career to gain the skills and insights in frontier areas of Biomedical Sciences.

Dr. B.R. Ambedkar Center for Biomedical Research (ACBR) is a unique center under the University of Delhi wherein a multi-specialty group of scientists work as a cohesive team and carryout active teaching and research. The absence of a formal departmental setup provides an excellent environment where faculties interact with each other freely that enhances better teaching and research in the complementary areas. The emphasis of research investigations is mainly on chemistry and biology and is being carried out in some of the frontline areas of basic and applied biomedical sciences such as Drug Discovery and Drug Development, Medical Biotechnology, Bioinformatics, Molecular Modeling and DNA Diagnostics, Molecular Oncology, Immunology, Genomics & Proteomics, Medicinal Chemistry, Cancer Genetics, Human Genetics and Neuropharmacology. Within a small span of time, ACBR has earned its name and fame both at the National and International level.

**Bioinformatics Centre (BIF Centre)**

Bioinformatics Facility was created at ACBR in the year 2003 with purchase of Molecular Modeling software Insight II, GCG, TSAR from Accelrys. Later on, since 2006, the facility is being funded and supported by Department of Biotechnology (DBT), Govt. of India. The primary objective of this facility is to promote biology teaching and understand its different aspects computationally through bioinformatics. It provides services to the host faculty and other neighbouring institutions of higher education in terms of retrieval of online and offline biological databases, bioinformatics software and tools for facilitating biological research particularly in the area of mining genomes and functional analysis, phylogenetic analysis, database development, molecular modeling and drug design and other fields of Bioinformatics. In addition to above, various events are periodically organized to give hands-on trainings on various aspects of application of bioinformatics tools and molecular techniques in biological research. ACBR since 2008 has consecutively organised workshops for the better understanding of different modules of drug design tools. These workshops trained approximately more than 300 participants from different universities over the period of time in the field of bioinformatics and its applications in computer aided drug design. To continue efforts in training and skill development the BIC centre at ACBR is introducing this Post Graduate diploma in Bioinformatics, Computational Biology and Drug Discovery.

## II. Introduction to CBCS (Choice Based Credit System)

### Choice Based Credit System:

The CBCS provides an opportunity for the students to choose courses from the prescribed courses comprising core, elective/minor or skill-based courses. The courses can be evaluated following the grading system, which is considered to be better than the conventional marks system. Grading system provides uniformity in the evaluation and computation of the Cumulative Grade Point Average (CGPA) based on student's performance in examinations which enables the student to move across institutions of higher learning. The uniformity in evaluation system also enable the potential employers in assessing the performance of the candidates.

### Definitions:

- (i) 'Academic Programme' means an entire course of study comprising its programme structure, course details, evaluation schemes etc. designed to be taught and evaluated in a teaching Department/ Centre or jointly under more than one such Department/ Centre
- (ii) 'Course' means a segment of a subject that is part of an Academic Programme
- (iii) 'Programme Structure' means a list of courses (Core and Elective) that makes up an Academic Programme, specifying the syllabus, Credits, hours of teaching, evaluation and examination schemes, minimum number of credits required for successful completion of the programme etc. prepared in conformity to University Rules, eligibility criteria for admission
- (iv) 'Core Course' means a course that a student admitted to a particular programme must successfully complete to receive the degree and which cannot be substituted by any other course
- (v) 'Elective Course' means an optional course to be selected by a student out of such courses offered in the same or any other Department/ Centre
- (vi) 'Credit' means the value assigned to a course which indicates the level of instruction;

One-hour lecture per week equals 1 Credit, 2 hours practical class per week equals 1 credit.

Credit for a practical could be proposed as part of a course or as a separate practical course

- (vii) 'SGPA' means Semester Grade Point Average calculated for individual semester.
- (viii) 'CGPA' is Cumulative Grade Points Average calculated for all courses completed by the students at any point of time. CGPA is calculated each year for both the semesters clubbed together.

- (ix) 'Grand CGPA' is calculated in the last year of the course by clubbing together of CGPA of two years, i.e., four semesters. Grand CGPA is being given in Transcript form. To benefit the student a formula for conversation of Grand CGPA into %age marks is given in the Transcript.

### **III. Post Graduate Diploma in Bioinformatics, Computational Biology and Drug Discovery**

#### **Programme Details:**

##### **Scope**

With the advent of new technologies, software tools, methods; Bioinformatics has evolved enormously over the past decades. It is a trans-disciplinary field of science; that combines biological sciences, chemistry, information technology, computer science, statistics and mathematics to solve research problems in biology. It is a relatively new field that aids in collecting, connecting, manipulating, and analyzing many forms of biological data to uncover new biological insights. This diploma course is aimed at providing learners with both academic and practical knowledge of bioinformatics, computational biology and drug design. After completing the course, the participant should have in-depth knowledge and a better comprehension of sequence analysis, genome annotation, comparative genomics, protein structure prediction, drug discovery, microarray profiling, advanced biological data analysis, machine learning and artificial intelligence techniques and its biomedical applications etc. The course modules have been put together to expose attendees to various elements of bioinformatics and the necessity for and the advantages of bioinformatics in biological research.

##### **Programme Objectives (POs):**

With this in mind the objectives of the **Post Graduate Diploma in Bioinformatics, Computational Biology and Drug Discovery** to provide training to cater to the modern industries working in the relevant subject. The diploma holders will be able to address problems related to modern genomics applications, big data analytics and drug discovery, especially using ML and DL based techniques. The training programme will provide opportunities to work in collaboration with Industries for practical training purpose, as per the recent surge of trained professionals in this field.

##### **Programme Specific Outcome**

After completing the course the students are expected to-

- Develop problem-solving and critical thinking skills.
- Demonstrate the uses of information and skillsets learned in the course and at work
- Solve real-world working issues by analysing the work environment and conditions and selecting relevant skill sets learned in the course

**Programme Structure:**

The PG diploma programme in Bioinformatics, Computational Biology & Drug Discovery is a one-year course divided into two semesters. A student is required to complete 48 credits for completing the course and the award of diploma.

**Course credit Scheme**

Semester	Core Courses			Elective Course			Dissertation #	Dissertation (credits)	Total Credits
	No. of papers	Credits (L+ T/P)	Total Credits	No. of papers	Credits (L+ T/P)	Total Credits			
I	5	20 L + 4 P	24	-	-	-	-	-	24
II	2	8 L + 2 P	10	1	4 L + 0 P	4	10 D <sup>#</sup>	10	24
Total no of credits for the course	7	28 L + 6 P	34	1	4L	4	10 D <sup>#</sup>	10	48

L = Theory; P = Practical; D = Dissertation

# D=Duration of dissertation = 6-8 weeks

**Semester wise details of the PG diploma in Bioinformatics, Computational biology & Drug Discovery courses**

Semester I/II

Semester I				
Number of core courses (5)	Credits in each core course			
Course	Theory	Practical	Tutorial	Credits
CC1	4	1	0	5
CC2	4	0	0	4
CC3	4	1	0	5
CC4	4	1	0	5
CC5	4	1	0	5
Semester II				
No. of core courses (2)				
No. of elective course = (1)				
Dissertation = (1)				
CC6	4	1	0	5
CC7	4	1	0	5
EC*	4	0	0	4
Dissertation (D)**	0	10	0	10
				48

\*ONE ELECTIVE COURSE IS TO BE TAKEN OUT OF OPTIONAL COURSES OFFERED.

\*\*ONE DISSERTATION OF 6-8 WEEKS DURATION

**SEMESTER 1:**

**A. CORE COURSES INCLUDING PRACTICAL (24 CREDITS)**

- CC1: Introduction to Bioinformatics, Biological Databases and Data Generation
- CC2: Biostatistics, Biomolecules, cell biology and molecular biology
- CC3: Structural Bioinformatics and Proteome based Technology
- CC4: Computational Molecular Biology and Computational tools (Genomics based analysis)
- CC5: Molecular modelling and drug design

**SEMESTER 2:**

**A. CORE COURSES INCLUDING PRACTICAL (10 CREDITS)**

- CC6: Bioprogramming and Bioinformatics Software; various languages (BioPython, Perl etc.)
- CC7: Introduction to Big Data Analytics; Basics of ML, DL methods & algorithms

**B. ELECTIVE COURSES (4 CREDITS)**

**(Students are required to choose any one of the following courses)**

- EC1: Deep Learning and Artificial Intelligence (DL & AI); or
- EC2: The Use of Big Data Analytics in the Pharmaceutical, Healthcare Industries and framing public health policies

**C. PROJECT/ DISSERTATION/ INTERNSHIP (10 CREDITS) (6-8 WEEKS)**

**Teaching:**

The faculty of the Center is primarily responsible for organizing lecture work for the Post Graduate Diploma in Bioinformatics, computational biology and drug discovery. Faculty from some other Departments and constituent colleges are also associated with lecture and tutorial work in the Department. Faculty from IIT, Delhi will also be participating in teaching of the course.

There shall be 90 instructional days excluding examination in a semester.

The dissertation/internship projects are mainly carried out at ACBR under the mentorship of the faculty of ACBR itself in the core area of research of the faculty. The students will also be placed in Industries for the training. In addition, students can also join laboratories in nearby Institutes or departments such as V. Patel chest, INMAS, DIPAS, IGIB.

During the dissertation, students are trained to design the experiments, carry out the research by performing experiments independently, analyse the data and interpret the results. With the help of the mentor, student is asked to compile the project report and finally defend their research work in an open viva voce examination.

**Eligibility for Admissions:**

***Mode of admission***

**ONLY THROUGH ENTRANCE** conducted by the University of Delhi.

***Minimum Eligibility Criteria:***

The candidate should be Graduate in the appropriate field of science - B.Tech./B.Sc. in Biomedical Sciences/ Microbiology/ Life Sciences/ Botany/ Zoology/ Chemistry/ Food Science/ Food Technology/ Biotechnology/ BE/ B.Pharma/ MBBS/ BDS/ BHS/ BUMS/ BAMS or any other related discipline.

**Assessment of Students' Performance and Scheme of Examinations:**

1. English shall be the medium of instruction and examination.
2. Examinations shall be conducted at the end of each semester as per the Academic Calendar notified by the University of Delhi.

1. **E x a m i n a t i o n :** The system of evaluation shall be as follows:

a) Each four credit course shall be evaluated for 100 marks and will have two components:

- |   |          |
|---|----------|
| <b>i. Internal Assessment:</b>          | 30 marks |
| a. Attendance:                          | 05 marks |
| b. Test /Assignments/Seminar:           | 25 marks |
| <b>ii. End of Semester Examination:</b> | 70 marks |

b) ***Practical Evaluation:***

- A total of 4 credit practical examination will be conducted at the end of the 1<sup>st</sup> semester for 100 marks.
- A total of 2 credit practical will be conducted at the end of 2<sup>nd</sup> semester for 50 marks

d) ***Dissertation Evaluation:*** The scheme of evaluation for the candidate opting for a dissertation project/internship (II Semester) shall be as follows:

Project work will begin from end of 1<sup>st</sup> Semester and will be completed at the end of 2<sup>nd</sup> semester.

The evaluation criteria for project work will be for 250 marks (10 credits) which includes internal evaluation by the concerned supervisor based on general performance in the laboratory (for 75 Marks) and student seminar, Project work and viva-voce evaluated by a board of examiners constituted by committee of courses of ACBR as per Delhi University Rules (175 Marks) as detailed below:

Attendance, experimental ability, notebook record keeping- 75 Marks (Internal assessment)

Data analysis and Project report 100 Marks

Presentation of work and Viva voce examination 75 Marks

5. Examinations for courses shall be conducted only in the respective odd and even Semester as per the Scheme of Examinations. Regular as well as Ex-students shall be permitted to appear / re-appear / improve in courses of Odd Semesters only at the end of Odd Semesters and courses of Even-Semesters only at the end of Even Semesters.

### **Pass Percentage**

Minimum marks for passing the examination in each semester shall be 40% in each paper and 45% in aggregate of a Semester. Details as per University of Delhi rules.

However, a candidate who has secured the minimum marks to pass in each paper but has not secured the minimum marks to pass in aggregate may reappear in any of the paper/s of choice in the concerned semester in order to be able to secure the minimum marks prescribed to pass the Semester in aggregate.

No student would be allowed to avail more than two chances to pass a paper inclusive of the first attempt.

### **Promotion Criteria**

#### **Semester to Semester:**

Students shall be required to fulfill the semester to semester Promotion Criteria. Students shall be promoted from a semester to the next Semester, provided she/he has passed at least three out of the five theory examinations of the current Semester. However, passing in practical is mandatory for promotion from one semester to the next. There shall not be any repeat/improvement allowed for practical examination.

**Conversion of Marks into Grades:**

As per University Examination rules.

**Grade Points:**

Grade point table as per University Examination rules.

**CGPA Calculation:**

As per University Examination rules.

**SGPA Calculation:**

As per University Examination rules.

**Grand SGPA Calculation:**

As per University Examination rules.

**Conversion of Grand CGPA into Marks**

As notified by competent authority the formula for conversion of Grand CGPA into marks is: Final %age of marks = CGPA based on all four semesters  $\times$  9.5

**Attendance Requirement:**

No student shall be considered to have pursued a regular course of study unless he/she is certified by the Head /Director of the Dr. B.R. Ambedkar Center for Biomedical Research, University of Delhi, to have attended 75% of the total number of lectures and seminars conducted in each

Semester, during his/her course of study. Provided that he/she fulfils other conditions, the Head/Director, Dr. B. R. Ambedkar Center for Biomedical Research may permit a student to the next Semester who falls short of the required percentage of attendance by not more than 10% of the lectures and seminars conducted during the Semester.

**Span Period:**

No student shall be admitted as a candidate for the examination for any of the Semesters after the lapse of **four** years from the date of admission to the Semester-I of the PG diploma programme.

**Guidelines for the Award of Internal Assessment Marks PG diploma programme (Semester Wise)**

Internal assessment will comprise 30% of the maximum marks in a specified paper. Internal assessment will have several components like attendance marks, marks for mid-term written test / assignment / seminar, marks for discipline and interaction in class, etc. as outlined below semester wise. Marks for mid-term written test / assignment / seminar and attendance will be discussed with the students and copies for written test / assignment will be returned to students appropriately marked. Attendance will be marked separately for each theory and practical papers as well as for Dissertation as per the prevailing rules of the University of Delhi.

**Theory paper (4 credits, 100 marks):** Attendance marks will be included as part of 30% internal assessment in each paper and marked in three tiers. Internal assessment components for theory papers (4 credits, 100 marks) in Semester I/II:

Attendance = 5 marks;

Written test/ Assignment/ Seminar/ Viva-voce (with documented evidence) = 25 marks;

Total = 30 marks.

**Internal assessment component of Practical paper (4 credits, 100 Marks) in semester I:**

Attendance /Class interaction/ discipline = 10 marks;

Notebook record keeping = 10 marks;

Experimental ability = 10 marks;

Total = 30 marks.

Rest 70 marks will be evaluated during end semester practical examination

**Internal assessment components for Practical papers (2 credits, 50 marks) in Semester II:**

Attendance /Class interaction/ discipline = 5 marks;

Notebook record keeping = 5 marks;

Experimental ability = 5 marks;

Total = 15 marks.

Rest 35 marks will be evaluated during end semester practical examination

**Internal assessment components (75 marks) for Dissertation (10 credits, 250 marks) in Semester II:**

Attendance/ Lab interactions/Discipline in lab = 25 marks

Experimental ability & Notebook record keeping = 50 marks

Total = 75 marks.

**NOTE:**

The promotion/passing/attendance/other rules are subject to change from time to time by the University, and the rules prevailing at that time will be applicable.

**Programme Fee**

**Rs. 20000/- (Annual fee) for the diploma programme.**

**Number of intake per year: 10**

Reservation rules will be as per the University of Delhi rules

**Future career prospects of the field**

The students will be trained for theoretical as well as practical aspects of the specialized diploma programme. By the end of the programme they will have hands-on experience in computational biology, bioinformatics and drug designing. They will get the chance to work in laboratories/Industries across India to gain the latest experience, which will have demand in modern Biotech, pharmaceutical, and IT companies requiring manpower in the field.

## **PG Diploma in Bioinformatics, Computational Biology and Drug Discovery**

### **Detailed contents**

#### **CORE COURSE 1 (CC1): Introduction to Bioinformatics; Biological Databases and Data Generation (60 Hours)**

**Learning outcome-** After completing the course, the students are expected to-

- get familiarized with various biological databases (online as well as standalone versions)
- know the need and methods to retrieve data from these databases using various search engines
- acquire basic knowledge of Linux commands, shell scripting
- understand pairwise and multiple alignment algorithms and their uses
- know different software for phylogenetic analysis and reconstruction of phylogenetic trees
- get exposed to different NGS platforms, genome assembly algorithms, sequence alignment formats, tools for conversion from one format to another, etc.

#### **Unit 1. Introduction and classification of different biological databases (primary and secondary). (11 - 12 Hours)**

A list of web-based bioinformatics resources, including NCBI, EXPASY, and others. Biological literature databases such as Pubmed, Nucleic acid databases like GenBank, EMBL, DDBJ, Refseq, etc. and Protein databases, e.g., PDB, Uniprot, Swiss-Prot, etc.). Databases of RNA sequences such as (miRBase, IncRNAdb, siRNA database, etc.), Species and Biodiversity databases (NCBI Taxonomy database, etc.)

#### **Unit 2. Analysis of Biological sequences (11 - 12 Hours)**

Pairwise sequence alignment algorithms (Needleman & Wunch; Smith & Waterman), Multiple sequence alignment algorithms (CLUSTALW, MUSCLE, DALIGN, T-coffee). Database similarity searches (FASTA, NCBI).

#### **Unit 3. Scoring matrices and sequence logos- (11 - 12 Hours)**

BLAST score, substitution of amino acids, matrices, s-value and e-value, calculating the alignment score and significance of e and p value. Scoring matrices for protein and nucleotide sequences (PAM, BLOSUM). Sequence logos, consensus and patterns.

#### **Unit 4. Molecular phylogenetics (11 - 12 Hours)**

Concept and overview. Distance-based methods (UPGMA & NJ), character-based methods (Maximum Parsimony). Phylogenetic analysis algorithms and their tools, cladistics. Reliability of trees (bootstrap, jackknife, etc.). Difference between phylogenetic tree and dendrogram. Phylogenetic trees and their composition: Various types of phylogenetic trees. Trees to trees distances, similarity. Construction of phylogenetic trees (PHYLIP) and identifying homologs.

**Unit 5. Basics of Next Generation Sequencing and data analysis**

**(11 - 12 Hours)**

Various NGS platforms, different genome assembly algorithms, sequence alignment formats (SAM, BAM), tools for conversion (SAMtools, etc.), alignment viewers (IGV, MGAviwer, etc.), NGS frameworks (Galaxy, ), Biological applications of NGS.

**Practical : 1 credit**

**(13-14 sessions)**

**Hands-on -working on following software/modules Practicals:**

Unit 1.

- NCBI (Nucleotide, Gene, Protein, Pubmed, PubChem, etc.)
- Expasy: UniprotKB/Swissprot, PROSITE
- Introduction to Linux command line

Unit 2.

- NCBI BLAST(different types) , N-W, S-W
- EBI: BLAST, ClustalW

Unit 3-4

- Phylip
- MEGA

Unit 5

- NGS framework (e.g., Galaxy) and use on any ChiP-seq and RNAseq sample datasets

**References**

- Rigden DJ, Fernández XM. The 2021 Nucleic Acids Research database issue and the online molecular biology database collection. *Nucleic Acids Res.* 2021 Jan 8;49(D1): D1-D9. doi: 10.1093/nar/gkaa1216. PMID: 33396976; PMCID: PMC7778882
- Baxevanis, A. D., & Davison, D. B. (2021). *Current Protocols in Bioinformatics*. John Wiley & Sons.
- Lesk, A. (2019). *Introduction to Bioinformatics* (5th ed.). Oxford University Press.
- Rastogi, S., Mendiratta, N. and Rastogi, P., 2013. *Bioinformatics methods and applications*. Dehli, India: PHI Learning Private Limited.
- Mandoiu, I., & Zelikovsky, A. (2016). *Computational Methods for Next Generation Sequencing Data Analysis*. Wiley.

**CORE COURSE 2 (CC2): A. Biomolecules, cell biology and molecular biology**

**B. Basic Mathematics and Biostatistics**

**(60 Hours)**

**Learning outcomes-** After completing the course, the students are expected to-

- get introduced to the basic statistical techniques utilized to solve biological problems
- become familiar with the building blocks of the living system (Bio-molecules such as DNA, RNA, protein, carbohydrate, lipids, etc.) and their functions
- understand the flow of information in the biological systems, i.e., central dogma (replication, transcription and translation), concepts of heredity, cell cycles, etc.

**A. Biomolecules, cell biology and molecular biology**

**Unit 1. Biomolecules**

**(5 -6 Hours)**

- Carbohydrates: structures and functions
- Proteins -definition, amino acids, structure, types, conjugate, derived proteins. Proteins-primary, secondary, tertiary, quaternary structure and function
- Nucleic acids- nucleotide, phosphodiester bond, DNA, RNA, functions
- Lipids - definition, esters, fatty acids, unsaturated and saturated fatty acids. Lipids- types-simple, derived and compound lipids, functions
- Detection and analysis of Carbohydrates, Proteins & Lipids
- Reducing and Non- reducing sugar, bonds, Hydrophilic, Hydrophobic

**Unit 2. Cell Biology**

**(3 - 4 Hours)**

- Introduction to Cell Biology, Cells and its components, cell Organelles and functions, Plant and Animal tissues.

**Unit 3. Molecular biology**

**(9 -10 Hours)**

- Structure/Biochemistry/function of DNA and protein, DNA replication, DNA Transcription, RNA Translation, Mendelian and modern concepts of heredity, Mitosis & Meiosis, Genes, and proteins, Mutation, Genomics, Transposable elements, and Chromosomes.

**B. Basic Mathematics, Probability & Statistics**

**Unit 1. Linear Algebra-** Vectors, Matrices, determinants, matrix addition and multiplication, PCA (Principal component analysis), Eigenvalues, Eigenvectors, Singular Value Decomposition.

**(7- 8 Hours)**

**Unit 2. Calculus-** Limits, continuity, linear functions, non-linear functions, local maximum, global maximum, etc.

**(7-8 Hours)**

**Unit 3. Probability-** Different types of theories: probability theory, decision theory, Information theory; variables (binary, multinomial, random); Binomial distribution, Poisson distribution, Geometric distribution, Normal distribution and extreme value distribution; conditional probability, marginal probability; Bayes' theorem, Monte Carlo approximation.

**(9 - 10 Hours)**

**Unit 4. Biostatistics**

**(13-14 Hours)**

Concept of sampling-Methods of sampling distributions, Central tendency measures including mean, median, and mode, as well as frequency distribution. Standard deviation, variance, coefficient of variation, regression, coefficient of regression, Correlation, Coefficient of correlation, Rank correlation. measures of skewness and kurtosis. Hypothesis testing; Type I, Type II errors; statistical tests such as chi-square test, t-test and f-test; Goodness of fit.

**References**

- Voet, V. J. D. G. (2021). Fundamentals of Biochemistry: Life at the Molecular Level. John Wiley & Sons Inc.
  - Tymoczko, J. L., Berg, J. M., Jr., G. G. J., & Stryer, L. (2018). Biochemistry: A Short Course (Fourth ed.). W. H. Freeman.
  - Nelson, David L., and Michael M. Cox. 2017. Lehninger Principles of Biochemistry. 7th Edition. New York, NY: W.H. Freeman.
  - Watson James D., Baker Tania A., Bell Stephen P., Alexander Gann, Levine, Michael Losick Richard 2008. Molecular Biology of the Gene 6th Edition. Publisher: New York, Cold Spring Harbor Laboratory Press. 2008. ISBN: 9780321507815
  - Benjamin Lewin, Jocelyn E. Krebs, Elliott S. Goldstein, Stephen T. Genes XIth Edition. Publisher: Kilpatrick Jones & Bartlett Publishers, 2014.
  - Looney, S., 2002. Biostatistical methods. Totowa, N.J.: Humana Press.
  - Le, C. T., & Eberly, L. E. (2016). Introductory biostatistics. Wiley.
-

**CORE COURSE 3 (CC3): Structural Bioinformatics and Proteome based Technology** **(60 Hours)**

**Learning outcome-** After completing the course, the students are expected to-

- describe different protein structures and their functions
- get familiar with various computational methods to predict tertiary and secondary structures of protein, tools to visualize protein structures in 3-D form
- understands the basic principles behind the three most often used structural biology approaches- NMR, X-ray crystallography and Cryo-EM)
- comprehend the PDB file format and parsing, view a molecule in different forms, detect interacting residues (protein-ligand interactions), calculate atomic distances, etc.
- Analysis, interpretation of protein sequences and structures to predict protein function

**Unit 1.** Protein structure folding, flexibility prediction; engineering and designing of the same. Secondary structure prediction- 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> generation methods to identify secondary structural elements. **(11 - 12 Hours)**

**Unit 2.** Structure visualization using PyMol and VMD, active site determination, Cavity analysis using CASTP or ACSITE, or similar tools. **(11 - 12 Hours)**

**Unit 3.** Secondary structure prediction of protein using PHD and PSI-PRED methods; X-ray, Cryo-EM and NMR techniques: Basics and application for protein structure determination. **(15 - 16 Hours)**

**Unit 4.** Tertiary Structure prediction of protein using ab-initio methods, Homology modeling and fold recognition. **(11 - 12 Hours)**

**Unit 5.** Study of oligomeric protein structures and their interaction interfaces. **(5- 6 Hours)**

**Practical: 1 Credit** **(13-14 sessions)**

Unit 1 & 2: Secondary structure prediction tools in Expasy

Unit 3.

- PyMol
- VMD

Unit 4

- PHD
- PSI-PRED

Unit 5

- MODELLER
- I-TASSER
- Rosetta

Unit 6

- OCAM webservice
- osFP webservice

## References

- Gáspári, Z. (2020). Structural Bioinformatics: Methods and Protocols (Methods in Molecular Biology, 2112) (1st ed. 2020 ed.). Springer.
  - Molecular Modeling: Basic Principles and Applications by Hans-Dieter (2008–01-16). (2021). Wiley VCH.
  - Mohan, G. C. (2019). Structural Bioinformatics: Applications in Preclinical Drug Discovery Process (Challenges and Advances in Computational Chemistry and Physics, 27) (1st ed. 2019 ed.). Springer.
  - Koča, J., Vařeková, S. R., Pravda, L., Berka, K., Geidl, S., Sehnal, D., & Otyepka, M. (2017). Structural Bioinformatics Tools for Drug Design: Extraction of Biologically Relevant Information from Structural Databases (Springer Briefs in Biochemistry and Molecular Biology) (1st ed. 2016 ed.). Springer.
-

**CORE COURSE 4 (CC4): Computational Molecular Biology and Computational tools (Genomics based analysis) (60 Hours)**

**Learning outcome-** After completing the course, the students are expected to-

- exhibit a knowledge base in genomics, transcriptomics, DNA methylation, etc.
- identify in detail the protein components of the nucleosome and comprehend the interactions between the DNA double helix and the nucleosome and the necessary changes that it entails
- show understanding of common and advanced laboratory procedures for fundamental genomic and transcriptomic sequence processing methods, as well as skills in the usage of commonly used software tools in this field
- know different types of sequencing techniques (DNaseq, RNAseq, chipseq, etc.)
- understand different file formats of sequencing data and their uses
- perform comparative analyses of genomic data of different species

**Unit 1. (14 -15 Hours)**

- Genomics- Structural genomics, functional genomics
- Transcriptomics- Microarrays and RNAseq
- Methylation- Genomic methylation, protein methylation and analyses
- Transcription factors analysis- Transcription factor binding analysis and their associated techniques
- RNA binding protein analysis studies

**Unit 2. (14 -15 Hours)**

- Techniques to study chromatin accessibility (e.g. DNase-seq, single-cell DNase-seq (scDNase-seq), ATAC-seq and single-cell ATAC-seq (scATAC-seq))
- Techniques to study chromatin structure (e.g., CHIP-seq, CHIP-chip, HiC, ChIA-PET, etc. )

**Unit 3. (14 -15 Hours)**

- History of sequencing technologies
- Next generation sequencing platforms
- Advanced library preparation steps for specialized assays, including DNase I footprinting, CHIP-seq, ATAC-seq, HiC, eCLIP, etc.

**Unit 4. (15-16 Hours)**

- Understanding the basics of computer networks: Graph Theory, its use in Biology and Integration.
- Topology of biological networks: Random vs Scale-Free networks. Metabolic networks, Signal transduction networks, Gene regulatory networks. Databases on metabolic & signaling pathways.
- Introduction to computational tools for analysis (Network analysis & clustering) of high throughput data from genomics (NGS), transcriptomics (Microarray/RNASeq), proteomics & metabolomics.

**Practical: 1 credit**

*(13-14 sessions)*

**Practical**

- Use bioinformatics tools for sequence analysis,
- transcription factor analysis,
- RNA binding protein analysis,
- For any NGS framework (e.g., NEAT) and any ChiP-seq and RNA-seq sample datasets.
- Creating networks and analysis, Cytoscape and Pajek, STRING, KEGGS and other annotation tools.

**References**

- Kelley, S. T., & Didulo, D. (2018). Computational Biology: A Hypertextbook (ASM Books) (1st ed.). ASM Press.
  - Head, S. R., Ordoukhanian, P., & Salomon, D. R. (2017). Next Generation Sequencing: Methods and Protocols (Methods in Molecular Biology, 1712) (1st ed. 2018 ed.). Humana.
  - Low, L., & Tammi, M. (2017). Bioinformatics: A Practical Handbook of Next Generation Sequencing and Its Applications. World Scientific Publishing Co.
-

**CORE COURSE 5 (CC5): Molecular modelling and drug design****(60 Hours)**

**Learning outcome-** After completing the course, the students are expected to-

- be efficient in installing and using various drug designing software.
- have an understanding of structure activity relationship (SAR) of molecules important in the drug development process.
- understand the process of drug discovery and development
- competent in performing different types of molecular docking of compounds/molecules against protein structure
- know different types of scoring functions
- be able to describe the difficulties encountered at each stage of the drug development process
- be able to perform energy minimization, molecule preparation, molecular dynamics simulation and trajectory analysis.

**Unit 1: Molecular modelling****(9-10 Hours)**

Basics method, energy minimizations, force fields, Introduction to MM, QM, Semi-Empirical based methods, conformational analysis.

**Unit 2: Molecular Dynamics Simulation****(11- 12 Hours)**

Foundation of molecular dynamics, Radial distribution functions, Pair Correlation functions, and Newtonian dynamics are examples of fundamental types; Integrator algorithms (Leapfrog and Verlet); Potential truncation and shifted-force potentials, implicit and explicit solvation models, periodic boundary conditions. Controlling the temperature and pressure in molecular dynamics simulations.

**Unit 3. Introduction to QSAR****(15- 16 Hours)**

QSAR, QSPR; various descriptors used in QSARs: topological, geometrical, electronics; thermodynamics, constitutional; Chemical-based descriptors. Regression analysis, the relevance and viability of QSAR Regression equations, Multi Linear Regression analysis, Partial Least Squares (PLS) analysis. Use of Principal Component Analysis (PCA), Genetic Algorithms and Neural Networks in the QSAR equations.

**Unit 4. Drug design using case studies****(17-18 Hours)**

The procedure for developing a drug. Identification and validation of targets, as well as lead optimization and validation. Analog-based drug design-methods and tools in computer-aided molecular design- Pharmacophores (3D database searches, conformation searches, constructing and utilising 3D Pharmacophores, restricted systematic search, Genetic Algorithm, clique detection approaches, maximum likelihood method);

structure-based drug design methods such as docking, De Novo Drug Design (Fragment Placements, Connection Methods, Sequential Grow), and Virtual Screening, etc.

**Practical: 1 credit**

*(13-14 sessions)*

**Unit 1.** MODELLER software, Swiss PDB Viewer, PyMol, Rasmol, Discovery Studio, Cresset

**Unit 2.** GROMACS, VMD, NAMD, CHARMM

**Unit 3.** Building QSAR models using Weka, KNIME, Discovery Studio, Cresset

**Unit 4.** Drug design case studies using Autodock, Discovery Studio, Cresset

**References**

- K, A. S. (2015). Molecular Modelling and Drug Design (1st ed.). MJP Publishers.
- Herrmann, E. C., & Franke, R. (2012). Computer Aided Drug Design in Industrial Research (Ernst Schering Foundation Symposium Proceedings, 15) (Softcover reprint of the original 1st ed. 1995 ed.). Springer.
- Koča, J., Vařeková, S. R., Pravda, L., Berka, K., Geidl, S., Sehnal, D., & Otyepka, M. (2017). Structural Bioinformatics Tools for Drug Design: Extraction of Biologically Relevant Information from Structural Databases (Springer Briefs in Biochemistry and Molecular Biology) (1st ed. 2016 ed.). Springer.
- Loging, W. (2018). Bioinformatics and Computational Biology in Drug Discovery and Development (1st ed.). Cambridge University Press.
- Ouyang, D., Smith, S. C., Douroumis, D., Fahr, A., Siepmann, J., Snowden, M. J., & Torchilin, V. (2015). Computational Pharmaceutics: Application of Molecular Modeling in Drug Delivery (Advances in Pharmaceutical Technology) (1st ed.). Wiley.

**CORE COURSE 6 (CC6): Bioprogramming and Bioinformatics Software; various languages (Python, R, Bioperl etc.) ( 60 Hours)**

**Learning outcome-** After completing the course, the students are expected to-

- be familiar with procedural statements such as assignments, conditionals, loops, method calls, and arrays
- develop competency in the Python programming language and Python libraries such as Pandas, Numpy, and Scipy.
- efficient in performing exploratory data analysis using Python.
- be able to call routine and modules and their implementation using Biopython.
- have confidence in using R to import data, convert it, produce high-quality images for publishing, and run standard statistical analyses.

**Unit 1: Introduction to Python** – Python versions, installation of python, basic programming concepts, comparison with other languages, flavours of python, editors used for python, syntax rules. **(3 - 4 Hours)**

**Unit 2: Python data Structure, Python conditional loops** – Data types- Basic data types used in python- python numbers (integers, floating numbers, complex numbers) Strings, Lists, Tuples, Dictionaries, Sets, Regular expressions. Conditional executions & Conditional loops- Conditional executions such as Boolean expressions (TRUE or FALSE), logical operators (AND, OR and NOT). Conditional loops such as IF, IF ELSE, for loop, while loop, nested loops, break, iterations, exception handling. **(7 - 8 Hours)**

**Unit 3: Python strings, Python file handling (e.g., CSV, binary, normal files, etc.)** Python strings- string notation (single quotes, double quotes), methods associated with strings (replace, count, find, index, split, join, strings comparison, string manipulation (lower, upper). Python file handling- Opening, processing (reading or writing), closing the files. Open () function with different modes, close () function, write () function, read () function. **(3- 4 Hours)**

CSV (comma separated values) files, CSV Module and its associated functions, reading & writing using CSV Module. File handling using os Module. Important functions on file handling- getcwd(), chdir (path), listdir (dir), remove (file), rename (source, destination), mkdir (path) etc. **(3-4 Hours)**

Binary files- Examples of binary files, opening, reading and writing in a binary file. Pickling using dump () function, unpickling using load () function. **(3- 4 Hours)**

**Unit 4: Python data Handling-** Data handling using python libraries such as NumPy (numerical python), pandas. Creating arrays and their manipulation using NumPy. Creating data frames and manipulating them using pandas. **(7 - 8 Hours)**

**Unit 5: Python implementation-** Various Python implementations including CPython (default), IronPython (.NET Python), Jython (Java Python), PyPy, and Stackless Python. **(7 - 8 Hours)**

**Unit 6: Introduction to R-** Fundamentals of R programming, downloading and its installation, R data types, data frames, conditional loops, data analysis using R (ggplot2, scatterplots, boxplot, Bar chart, Histogram, T-test, etc.). *(7 - 8 Hours)*

**Unit 7: Introduction to Perl/BioPerl-** History, installation, basic constructs (numbers, scalar operators and variables, array, functions, conditions, and loops), advance constructs (matching operators, regular expressions, user and system functions, split and join function, substitutions, file handling), object-oriented Perl. Perl application in Bioinformatics, BioPerl introduction (overview of different modules). *(7 - 8 Hours)*

**Practical:** All the Units listed above will be taught as practical/practice sessions

- Self-practice online platforms for python such as Dataquest, HackerRank, CodingGame, Edabit, R and Perl. *(13-14 sessions)*

### References

- Mitchell L Model, Bioinformatics programming using Python. Publisher: O'Reilly Media. ISBN: 9788184048988
  - Rocha, M., & Ferreira, P. G. (2018). Bioinformatics Algorithms: Design and Implementation in Python (1st ed.). Academic Press.
  - Wesley J Chun, Core Python Applications Programming. Publisher: Pearson Education ISBN: 8131791343
  - Sebastian Bassi, Python for Bioinformatics. Publisher: CRC Press. ISBN: 9781584889298
  - Jason Kinser, Python for Bioinformatics. Publisher: Jones and Bartlet. ISBN: 0763751863
-

**CORE COURSE 7 (CC7): Introduction to Big Data Analytics; Basics of ML, DL methods algorithms  
(60 Hours)**

**Learning outcome-** After completing the course the students are expected to-

- familiarize with basic mathematics and statistics required for machine learning
- differentiate between machine learning and deep learning
- comprehend and differentiate between various machine learning and deep learning approaches, as well as their application to biological issues utilizing Bioinformatics
- have the capability to identify and select suitable model parameters for different machine learning techniques
- the capacity to choose and implement appropriate machine learning methods and computing environments for the applications at hand
- know how to evaluate data-driven models
- have a fundamental knowledge of different applications of AI techniques (ANN, CNN, etc.)

**Unit 1. Simple Introduction to ML-** History of machine learning; its goals and applications. Types of machine learning- Supervised, Unsupervised Reinforcement learning. Types of supervised learning (classification and regression). Clustering: k-means, Outlier analysis, the curse of dimensionality. **(7 - 8 Hours)**

**Unit 2. Probability and estimation-** Random variable independence; Bayes rule; Estimators, Maximum likelihood estimator (MLE), Maximum a posteriori estimator (MAP). **(7 - 8 Hours)**

**Unit 3. Naive Bayes -** Bayes Optimal Classifiers, Conditional Independence, Naive Bayes. Learning for Naive Bayes, Gaussian Naive Bayes, Naive Bayes use case: the Bag of Words model. **(7 - 8 Hours)**

**Unit 4. Logistic Regression (LR) -** Generative vs discriminative classifiers, Classification using the logistics function, Gradient methods to solve LR: gradient descent, stochastic gradient descent, MLE and MAP estimates for LR. **(7 - 8 Hours)**

**Unit 5. Generalization and Evaluation -** Training error and generalization error, Hypothesis space, model capacity, Generalization, overfitting, underfitting, bias-variance trade-off, Regularization, model selection, cross-validation. **(7 - 8 Hours)**

**Unit 6. Basics of Model Learning-** Training, testing and validation datasets. Criteria to divide these datasets. Techniques for cross-validation- N-fold cross-validation, holdout cross validation, leave one out cross-validation. Hyperparameters tuning. Performance measures for classification and regression (accuracy, precision, sensitivity, specificity, ROC, AUC, etc.). Overfitting, underfitting and bias variance trade-off. **(9 -10 Hours)**

**Unit 7. Image Analysis with Convolutional Neural Networks-** Uses of CNN in image recognition, object detection and segmentation. Three layers of CNN. Importing of images using matplotlib, storing colored images in 3-D arrays. **(9 - 10 Hours)**

**Reading List/References**

- Brownlee, J. (2018). Basics of Linear Algebra for Machine Learning. Machine Learning Mastery.
- Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists. " O'Reilly Media, Inc."
- Rifaioğlu, Ahmet Sureyya, et al. "Recent Applications of Deep Learning and Machine Intelligence on in Silico Drug Discovery: Methods, Tools and Databases." Briefings in Bioinformatics, no. 5, Oxford University Press (OUP), July 2018, pp. 1878–912. Crossref, doi:10.1093/bib/bby061.
- Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence (Vol. 3, No. 22, pp. 41-46).
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398). John Wiley & Sons.
- Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. Proceedings of the National Academy of Sciences, 116(32), 15849-15854.
- Kubben, P., Dumontier, M., & Dekker, A. (2019). Fundamentals of clinical data science.
- Pattern Recognition and Machine Learning, Bishop Christopher M., (2006) Springer Science + Business Media, LLC, ISBN-10: 0-387-31073-8.
- Anwar, Syed Muhammad, et al. "Medical Image Analysis Using Convolutional Neural Networks: A Review." Journal of Medical Systems, no. 11, Springer Science and Business Media LLC, Oct. 2018. Crossref, doi:10.1007/s10916-018-1088-1.

**ELECTIVE COURSE 1 (EC1): Deep Learning and Artificial Intelligence (DL & AI) (60 Hours)**

- Unit 1. Basics of deep learning-** Layers in deep learning (Input, Hidden and Output layers). Types of deep learning networks (ANN, Feed forward neural network, RNN, CNN), Autoencoders. Tensors. Backpropagation algorithm. Limitations of deep learning. Examples of deep learning. **(6 - 7 Hours)**
- Unit 2. Recurrent Neural Networks for Natural Language Processing** – Techniques to transform natural language to the mathematical framework- Tokenization, Stemming, Lemmatization, N-grams (1-gram, 2-gram). Methods used for transformation- TF-IDF, One-Hot encodings, Word Embeddings. Three types of architectures of RNN- Many to One, One to Many, and Many to Many. **(6 - 7 Hours)**
- Unit 3. The Transformer Network for Natural Language Processing-** Encoders and Decoders. Encoders components- self-attention layer (Z), feed-forward neural network. Four learning parameters of self-attention layer (query matrix, key matrix, value matrix, output matrix). Softmax function. Decoder components- self attention layer, feed forward network and Encoder Decoder attention. **(6 - 7 Hours)**
- Unit 4. Introduction to Reinforcement Learning (RL)-** Terms used in Reinforcement learning (Agent, Environment, Reward, state, Policy, Value, Value function etc.) Types of reinforcement learning methods (positive and negative). Reward hypothesis. Different ways to implement reinforcement learning algorithm- model based, value based and policy based. Important models (Markov decision learning process and Q-learning). Uses of Reinforcement learning. **(6-7 Hours)**
- Unit 5. Construct deep learning models with Keras library-** Loading of data using numpy. Defining the model (layers (number and types), number of neurons in each layer etc.) Keras model compilation (loss function, optimizer, metrics), training and assessment. **(6 - 7 Hours)**
- Unit 6. Fine tuning Keras models-** Important steps to fine tune the model- removing the connected nodes, replacing the old nodes with new ones, freezing the layers, creating the new model, training. **(6 - 7 Hours)**
- Unit 7. Implementation of Deep learning with examples using research papers.** Research papers will be followed to study how above mentioned Deep Learning algorithms/methods were employed. **(6 - 7 Hours)**
- Unit 8. Applications of Deep Learning and AI in Biomedical Sciences:**  
 An overview of deep learning in Biology: applications of deep learning to regulatory genomics, variant scoring and population genetics; pharmacogenomics; electronic health records and medical imaging data. **(5 -6 Hours)**  
 An overview of deep learning in drug discovery: Target identification and prediction of structure, prediction of bioactivity, predicting ADMET/Toxicity, Biomolecular interactions, drug design and drug repurposing etc. **(5 -6 Hours)**

**Practical: 1 credit**

**(13-14 sessions)**

**Practical/small projects on following examples (any two):**

- Fake news detection
- Breast cancer prediction
- Sales forecasting
- Iris classification
- Movie ticket price prediction
- Stock price prediction

**Practical/small projects on following (any two)**

- Image Classification
- Face detection
- Applying AI to 2-D medical imaging data (Pneumonia Detection with chest X-ray images etc.)
- Information contents of EHR

**References for Unit 1-6:**

- Brownlee, J. (2018). Basics of Linear Algebra for Machine Learning. Machine Learning Mastery.
- Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (2nd ed.). O'Reilly Media.
- Moolayil, J. (2018). Learn Keras for Deep Neural Networks: A Fast-Track Approach to Modern Deep Learning with Python (1st ed.). Apress.
- Li, B. T. (2018). Applied Natural Language Processing with Python: Implementing Machine Learning and Deep Learning Algorithms for Natural Language Processing (1st ed.). Apress.
- Deitel, P., & Deitel, H. (2019). Intro to Python for Computer Science and Data Science: Learning to Program with AI, Big Data and The Cloud (1st ed.). Pearson.
- Ponteves, H. D. (2019). AI Crash Course: A fun and hands-on introduction to machine learning, reinforcement learning, deep learning, and artificial intelligence with Python. Packt Publishing.
- Alpaydin, E. (2020). Introduction to Machine Learning, fourth edition (Adaptive Computation and Machine Learning series) (fourth edition). The MIT Press.
- Russell, R. (2018). Machine Learning: Step-by-Step Guide To Implement Machine Learning Algorithms with Python. CreateSpace Independent Publishing Platform.

**Reading List/References for learning examples in Unit 7**

- Chan et al. Deep Learning in Medical Image Analysis. *Adv Exp Med Biol.* 2020. (DOI: 10.1007/978-3-030-33128-3\_1).
- Lee et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nat Biomed Eng.* 2019. (DOI: 10.1038/s41551-018-0324-9)
- Zaharchuk et al. Deep Learning in Neuroradiology. *AJNR Am J Neuroradiol.* 2018
- Kalinin et al. Deep learning in pharmacogenomics: from gene regulation to patient stratification. *Pharmacogenomics.* 2018. (DOI: 10.2217/pgs-2018-0008).

**Reading List/Research articles for Unit 8 (Applications of Deep learning and AI)**

- Alhazmi et al. Application of artificial intelligence and machine learning for prediction of oral cancer risk. *J Oral Pathol Med.* May 2021. (DOI: 10.1111/jop.13157)
  - Senior et al. Improved protein structure prediction using potentials from deep learning. *Nature* 2020. (DOI: 10.1038/s41586-019-1923-7)
  - Stokes et al. A Deep Learning Approach to Antibiotic Discovery. *Cell.* 2020. (DOI: 10.1016/j.cell.2020.01.021)
  - Sakellaropoulos et al. A Deep Learning Framework for Predicting Response to Therapy in Cancer. *Cell Rep.* 2019. (DOI: 10.1016/j.celrep.2019.11.017)
  - Wang et al. Using Electronic Health Records and Machine Learning to Predict Postpartum Depression. *Stud Health Technol Inform.* 2019. (DOI: 10.3233/SHTI190351)
  - Currie et al. Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. *J Med Imaging Radiat Sci.* October 2019. (DOI: 10.1016/j.jmir.2019.09.005)
  - Shickel et al. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *IEEE J Biomed Health Inform.* 2018. (DOI: 10.1109/JBHI.2017.2767063).
-

**ELECTIVE COURSE 2 (EC 2): The Use of Big Data Analytics in the Pharmaceutical, Healthcare Industries and framing public health policies** **(60 Hours)**

**Learning outcome-** After completing the course the students are expected to-

- Understand the terms "Big Data" analytics, descriptive data analytics, and predictive data analytics as they apply to healthcare data innovation.
- Understand the theoretical and practical aspects of healthcare data collecting, sharing, and organizing - Know how to locate, download/extract, clean, and perform descriptive analysis on various healthcare datasets.
- Recognize the current challenges in healthcare administration and the benefits that data analytics may bring. Possibilities for improving quality, lowering costs, and advancing results.
- Identification and referencing of sources of public health data and information
- Develop fundamental abilities in the use of popular data analysis applications.

**Unit 1: Basics of Big data in health care**

**(11 -12 Hours)**

- ‘V’ of big data: Volume, velocity, variety, veracity and value
- Health-Care constituents: care settings, stakeholders, finance, public health, regulatory/research
- Challenges and opportunities: The triple objective (care, health and cost), patient care quality. The healthcare service's experience and accessibility.
- Systems approach: A system approach including evidence-based medicine, quality improvement, and value-based reimbursements.
- Trends in health care: Consumerism/personalized medicine, demographics/population health, and emerging health care trends

**Unit 2: Health-care Informatics basics**

**(11 -12 Hours)**

- Outline of health IT: How does health informatics work? What part of the triple aim does health informatics play?
- Components of health IT: Enterprise vs. best of breed, structured vs. unstructured data, EMR/EHR modules and associated data systems
- EHR Adoption: EHR regulations, barriers to EHR, adoption interoperability, and HIT standards: health IT standards, data exchange, clinical Decision Support, HIPAA security, public health IT, and consumer engagements

**Unit 3: Data Analytics overview**

**(11 -12 Hours)**

- Data terms and concepts: Goal of data analytics. Considerations before analysis, data terminology, big data terminology, getting data ready for analysis. A virtuous cycle in analytics: Integrating data across data sets
- Patient data control, privacy, and security: Data governance in an organization including patient identification, regulatory issues, and data security.
- Analysis using Artificial Intelligence (AI): Analysis prediction of diseases using machine learning and natural language processing using patient’s data.

- Making Data availability to Others: Preparing data for analysis, and communicating data

**Unit 4: An outline to Visualizations**

*(7 - 8 Hours)*

- Visualization value: Visualization best practices: What to do and what not to do based on the use case.

**Unit 5: Application of Big Data Analysis**

*(15 -16 Hours)*

- **Implementation of Big Data analytics in medical imaging:**
  - Uses of big data analytics and artificial intelligence in mental health.
  - Breast cancer prediction using a kernel support vector machine and the Gray Wolf Optimization technique based on big data.
  - Employing convolutional neural networks and long short-term memory as a hybrid approach to diagnose cardiac disorders.

**Practical: 1 credit**

*(13 – 14 sessions)*

- Big data-based breast cancer prediction using SVM
- Obtaining data from EHR for further analysis
- Prediction of cardiac disorders using CNN

**Reading List/References**

- Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, 6(1), 1-25.
- Sun, J., & Reddy, C. K. (2013, August). Big data analytics for healthcare. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1525-1525).
- Wyatt, J. C., & Liu, J. L. (2002). Basic concepts in medical informatics. *Journal of Epidemiology & Community Health*, 56(11), 808-812.
- Menachemi, N., & Collum, T. H. (2011). Benefits and drawbacks of electronic health record systems. *Risk management and healthcare policy*, 4, 47.
- Menachemi, N., & Collum, T. H. (2011). Benefits and drawbacks of electronic health record systems. *Risk management and healthcare policy*, 4, 47.
- Bagga, S., Gupta, S., & Sharma, D. K. (2021). Big Data analytics in medical imaging. In Applications of Big Data in Healthcare (pp. 113-136). Academic Press.
- Alhazmi et al. Application of artificial intelligence and machine learning for prediction of oral cancer risk. *J Oral Pathol Med*. May 2021. (DOI: 10.1111/jop.13157)

---

**(Dissertation C): Internship/project/Industrial training of 6 to 8 weeks duration**