Course title: Introduction to Machine learning for Biotechnology			
Course code: SEC 206	**No. of credits:** 3	**L-T-P:** 34-11-0	**Learning hours:** 45
Pre-requisite course code and title (if any):			
Department: Department of Biotechnology			
Course coordinator: Dr. Priya Bhatnagar	**Course instructor:** Dr. Priya Bhatnagar and Dr. Ankit Kumar		
Contact details:			
Course type: Minor (for those opting for a minor in Data Science)/DSE (for others)	**Course offered in:** Semester 3 of FYUP/FYIPP Biotechnology		

Course description:

This course introduces undergraduate biotechnology students to the fundamental concepts and applications of Artificial Intelligence (AI) and Machine Learning (ML) in the life sciences. Designed for learners without a strong background in mathematics or programming, the course emphasizes conceptual understanding, biological relevance, and real-world applications rather than technical or computational details.

Students are introduced to the nature of biological data and the role of data-driven approaches in modern biotechnology. Core concepts of machine learning, including types of learning and commonly used models, are explained using intuitive examples from genomics, proteomics, transcriptomics, healthcare, and drug discovery. The course highlights how AI-based tools are transforming areas such as precision medicine, diagnostics, pharmaceutical research, and bioprocess optimization.

In addition to technical awareness, the course addresses ethical, legal, and societal issues associated with the use of AI in biotechnology, including data privacy, bias, and responsible decision-making. Through case studies and discussion-based learning, students develop analytical thinking, data literacy, and interdisciplinary communication skills.

Overall, the course prepares students for AI-enabled biotechnology environments and provides a strong foundation for employability in industry, healthcare, and research sectors, as well as for higher studies in bioinformatics and computational life sciences.

Course objectives:

1. Introduce fundamental concepts of Artificial Intelligence (AI) and Machine Learning (ML) in the context of biotechnology.
2. Familiarize students with different types of biological data and their applications in AI-driven analysis.
3. Demonstrate real-world applications of AI in genomics, proteomics, diagnostics, drug discovery, and precision medicine.

4. Develop student's ability to interpret AI-generated results and analyze case studies from life sciences. 5. Instill awareness of ethical, legal, and societal considerations, while preparing students for AI-related career pathways in biotechnology.				
Course contents				
S. No.	**Topic**	**L**	**T**	**P**
Module 1	**Introduction to AI, ML, and Data in Biology**	**6**	**2**	
1	**What is Artificial Intelligence? -** Definition and evolution of AI, AI vs Human Intelligence, Narrow AI vs General AI, Examples of AI in daily life and biology.	2	1	
2	**What is Machine Learning? -** Definition of Machine Learning, Difference between AI, ML, and Deep Learning (conceptual only), Why ML is useful for biological problems.	2		
3	**Types of Data in Biotechnology -** DNA, RNA, protein sequences, Images (microscopy, radiology), Clinical and epidemiological data, Structured vs unstructured data, Introduction to big data in biology.	2	1	
Module 2	**Basics of Machine Learning**	**7**	**3**	
1	**Types of Machine Learning -** Supervised learning (with biological examples), Unsupervised learning, Semi-supervised learning, Reinforcement learning (conceptual overview)	2	1	
2	**Common ML Concepts (Intuitive) -** Features and labels (using biological analogies), Training and testing data, Model and prediction, Overfitting and underfitting (visual explanation).	2	1	
3	**Common ML Algorithms (Conceptual Understanding) -** Decision trees, k-Nearest Neighbors (k-NN), Naïve Bayes, Support Vector Machines (idea only), Neural networks (biological inspiration)	3	1	
Module 3	**AI & ML in Genomics and Proteomics**	**6**	**3**	
1	**AI in Genomics -** Genome sequencing and data explosion, Gene prediction and annotation, Variant detection and disease association, AI in	2	1	

	personalized genomics			
2	**AI in Transcriptomics -** Gene expression analysis, Clustering of gene expression data, AI in single-cell RNA sequencing (scRNA-seq) – dimensionality reduction (UMAP/t-SNE concepts), automated cell-type clustering, and importance of AI in handling high-dimensional biological data (conceptual overview).	2	1	
3	**AI in Proteomics -** Protein structure prediction, Protein–protein interaction prediction, Case study: AlphaFold (conceptual impact, not technical details).	2	1	
Module 4	**AI in Drug Discovery, Healthcare, and Diagnostics**	**6**	**3**	
1	**AI in Drug Discovery -** Traditional vs AI-driven drug discovery, Target identification, Virtual screening, Drug repurposing, Time and cost reduction using AI.	2	1	
2	**AI in Medical Diagnostics, imaging, and Bedside Decision Support -** Medical imaging (X-ray, MRI, CT scans), AI in cancer detection, Pathology image analysis.POCUS, workflow assistance, and image-based biomarker discovery (conceptual overview).	2	1	
3	**AI in Precision Medicine -** Personalized treatment strategies, AI-based clinical decision support systems, Case studies from oncology and rare diseases.	2	1	
Module 5	**5: AI in Biotechnology Industry, Ethics, and Future Trends**	**9**		
1	**AI in Biotech and Pharma Industry -** AI-driven bioprocess optimization, Fermentation monitoring, Quality control and automation, Role of AI in vaccine development.	3		
2	**Ethical, Legal, and Social Issues -** Data privacy and security, Bias in AI models, Transparency and explainability, AI vs human decision-making in healthcare.	3		
3	**Future Trends and Careers -** Future of AI in biotechnology, Emerging fields like Digital biology and computational medicine, Introduction to digital health, wearable health data, and digital	3		

	biomarkers (conceptual overview). Career paths for biotechnology students, Importance of interdisciplinary learning.			
	Total	**34**	**11**	**0**

Evaluation criteria:

Assignment: 40%

End-semester exam: 40%

Case study presentation: 20%

Learning outcomes:

Students will be able to:

1. Use AI and ML terminology correctly in biological and biomedical contexts. (End-semester exam)
2. Interpret AI-based outputs used in genomics, proteomics, and healthcare. (Assignment)
3. Analyze real-life case studies involving AI in biotechnology industries. (Case study presentation)
4. Demonstrate data literacy skills relevant to biological datasets. (Assignment)
5. Evaluate ethical, legal, and societal implications of AI applications (End-semester exam)
6. Identify career pathways and skill requirements at the AI–biotechnology interface. (End-semester exam)

Pedagogical Approach:

1. Interactive Conceptual Lectures: Use biological analogies to explain AI/ML concepts.
2. Case-Based Learning: Analyze real-world applications in genomics, proteomics, diagnostics, and drug discovery.
3. Visual & Multimedia Tools: Use animations, videos, and workflow diagrams to illustrate ML models and AI applications.
4. Guided Hands-On Demonstrations: GUI based ML exercises for model building, classification, and clustering.
5. Collaborative Learning: Group discussions and presentations to develop interdisciplinary thinking and communication skills.
6. Reflective Practice: Maintain learning journals for data interpretation, case analyses, and model evaluation.
7. Ethics & Societal Awareness: Integrate discussions on bias, privacy, transparency, and responsible AI in biotechnology.
8. Future-Oriented Exposure: Introduce frontier applications like multi-omics integration, precision medicine, and computational biology.

Skill Set:

This course equips undergraduate biotechnology students with interdisciplinary data literacy, analytical thinking, and ethical awareness, enhancing their employability in AI-enabled biotechnology, healthcare, and pharmaceutical sectors.

1. Conceptual understanding of Artificial Intelligence and Machine Learning principles relevant to biotechnology

2. Ability to interpret AI-based outputs in genomics, proteomics, diagnostics, and drug discovery

3. Foundational data literacy for biological datasets, including sequence, expression, imaging, and clinical data

4. Analytical and critical thinking skills through AI-based biological case studies

5. Awareness of ethical, legal, and societal issues related to AI applications in life sciences

6. Interdisciplinary communication skills bridging biology and data-driven technologies

Employability:

1. Entry-level technical/support roles in academia and industry requiring AI/ML.

2. Assistant teacher in middle schools.

3. Junior researcher in industry & academia.

4. Entry-level data scientist.

Materials:

Suggested Readings

1. AI in Biology: A Beginner's Guide by K.N. Jha (2023).
2. Deep Learning for the Life Sciences by Bharath Ramsundar et al.
3. The AlphaFold Breakthrough: Jumper, J., et al. (2021). "Highly accurate protein structure prediction with AlphaFold." Nature. (Focus on the Results and Biological Significance sections).
4. AI in Diagnostics: Esteva, A., et al. (2017). "Dermatologist-level classification of skin cancer with deep neural networks." Nature.
5. Esteva, A., et al. “A guide to deep learning in healthcare.” Nature Medicine (2019).
6. Litjens, G., et al. “A Survey on Deep Learning in Medical Image Analysis.” Medical Image Analysis (2017).
7. Online resources:
    - Nature AI collections
    - NIH and EMBL-EBI AI resources
8. Industry case materials and technical resources
    - GRAIL – Multi-cancer early detection and AI-enabled diagnostics
    - GE HealthCare – AI-assisted imaging and clinical workflow tools
    - Butterfly Network – Portable ultrasound and AI-supported imaging
    - Illumina – Genomics and sequencing analytics workflows
    - Tempus AI – Precision medicine and clinical AI applications
    - PathAI – Digital pathology and AI-assisted pathology analysis

Additional information (if any): This course was jointly designed by Dr Ankit Kumar and Dr Priya Bhatnagar (both from EYM Life Sciences) at the request of Dr. Ramakrishnan Sitaraman. Dr. Mahesh S. Dhar (Director, Haryaksh Scientific Pvt. Ltd.) provided inputs to

effectively align the course with industry standards and student employability.

Knowledge of basic biology and statistics at high-school level recommended. *No prior programming experience required.*

Mapping table: Illustrative AI Applications in Biotechnology and Healthcare

Application Area	Data Type	Example Use Case	Reference Organization
Genomics	DNA sequencing data	Variant analysis	Illumina
Transcriptomics	scRNA-seq data	Cell-type clustering	Tempus
Diagnostics Imaging	CT/MRI/Pathology images	Cancer detection	GE HealthCare / PathAI
Portable Diagnostics	Ultrasound imaging	Bedside imaging support	Butterfly Network
Precision Medicine	Clinical + molecular data	Treatment recommendation	GRAIL
Digital Health	Wearable sensor data	Digital biomarkers	Apple Health
Assistive Healthcare Technologies	Audio and wearable sensor data	AI-assisted hearing enhancement and adaptive sound processing	Phonak

Student responsibilities:
Class attendance.
Active participation in case study discussions.
Self-study of provided reading materials.
Submission of assignment/case study presentation.

Course reviewers:

1. Dr. Pawan K Dhar, Executive Director, CVJ Centre for Synthetic Biology and Biomanufacturing, CUSAT Technology Business Incubator, 3rd Floor, CUSAT-Students Amenity Centre, Cochin University of Science and Technology, Kalamassery, Kochi 682 022 (Formerly Professor, School of Biotechnology, Jawaharlal Nehru University, New Delhi).

2. Dr. Anwar Alam, Associate Professor , Life Science, Sharda University, Plot no 32, 34, Knowledge Park III, Greater Noida, Uttar Pradesh 201310.

3. Dr. Abhinav Kaushik, Research Scientist, Harvard T.H. Chan School of Public Health, Nadeau Lab; FXB, 1st Floor, 651 Huntington Ave., Boston, MA 02115, USA.

4. Dr. Jatin Shrinet, Research Engineer, Institut Curie, 26 rue d'Ulm, 75248 Paris cedex 05, France.