Course Code	Course Title	L	T	P	C
BABIT404	**Artificial Intelligence and Machine Learning in Biology**	**2**	**1**	**2**	**4**
Pre-requisite	BABIT202 - Fundamentals of Bioinformatics	**Syllabus Version**			
		1			

Course Objectives

1. Provide a foundational understanding of machine learning (ML) and artificial intelligence (AI) principles relevant to biological data.
2. Introduce various biological data types, including omics, imaging, and time-series physiological data.
3. Teach key ML/AI techniques like supervised and unsupervised learning, deep learning, and their applications in genomics, proteomics, systems biology, and disease modeling.

Course Outcomes

1. Understand the fundamentals of ML and AI and their relevance to biology
2. Analyze biological data (genomics, proteomics, clinical) using ML techniques
3. Apply supervised and unsupervised learning methods for biological pattern discovery
4. Evaluate model performance using cross-validation and biological validation metrics
5. Design ML/AI pipelines for specific biological problems such as disease prediction or drug discovery

Module:1	**Fundamentals of ML & AI in Biology**	**4 hours**

Overview of ML and AI; Scope in biological research: genomics, proteomics, systems biology; biological data types for ML and AI studies: sequences, images, expression profiles, clinical records.

Module:2	**Data Preprocessing and Feature Engineering**	**6 hours**

Data cleaning, normalization, and transformation techniques; Handling missing values and outliers; Feature selection methods: statistical, filter, wrapper, and embedded techniques; Dimensionality reduction: PCA, t-SNE, UMAP; Feature construction in omics and image-based biological data.

Module:3	**Machine Learning Algorithms for Biology**	**6 hours**

Supervised learning: Linear regression; logistic regression; decision trees; random Forests; SVM; k-NN; Neural networks (MLPS) Unsupervised learning: k-means; hierarchical clustering; DBSCAN; Autoencoders; PCA.

Module:4	**Model Evaluation**	**6 hours**

Confusion matrix; ROC-AUC; precision-recall; cross-validation; Ensemble methods and overfitting/underfitting in biological contexts; ML applications in diagnostics;

biomarker discovery; and phenotypic classification.		
Module:5	**Deep Learning and AI Tools in Biology**	**6 hours**
Deep learning for biological sequences: CNNs, RNNs, Transformers, AI in medical image analysis (e.g., histopathology, radiomics); Tools and platforms: TensorFlow, Keras, PyTorch, MXNet Bioconductor; AI-driven biosensors and wearable technologies in life sciences.		
Module:6	**Module:6 Contemporary Topics**	**2 hours**
Expert Lectures		
	Total Lecture Hours:	**30 hours**
	Tutorial Hours:	**15 hours**

Text Book(s)

Andreas Holzinger, "**Machine Learning for Health Informatics: State-of-the-Art and Future Challenges**", Springer, 1st Edition, **2020**
Bharath Ramsundar, Peter Eastman, Patrick Walters, Vijay Pande, "**Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More**", O'Reilly Media, 1st Edition, **2019**

Reference Books

Rui Jiang, Xuegong Zhang, "**Applied Machine Learning for Bioinformatics and Computational Biology**", Springer, 1st Edition, **2021**
Tiago Antão, "**Bioinformatics with Python Cookbook,**", Packt Publishing., 1st Edition, **2022**

Indicative Experiments

1. Predicting Disease Risk Based on Genomic Data Using Machine Learning using Python codes	5 hours
2. Apply classification algorithms to biological datasets to predict discrete outcomes (disease presence, classification of genes, or biological states)	5 hours
3. Use CNNs for sequence-based tasks such as DNA sequence classification, protein structure prediction, or variant classification based on genomic sequences	5 hours
4. Use RNNs for sequential data tasks like sequence alignment, sequence-to- sequence prediction, and biological time-series analysis	5 hours
5. Use Keras and TensorFlow to classify DNA sequences (promoters vs non- promoters) based on their nucleotide sequences	5 hours

6. AI-Based Biosensors for Detecting Diabetes Risk Using Blood Glucose Levels from Public Datasets	5 hours
Total Laboratory Hours:	**30 hours**
Mode of Evaluation : Continuous Assessment Test, Final Assessment Test, Oral Examination	
Recommended by Board of Studies :	26-05-2025
Approved by Academic Council : No. 78	12-06-2025