DEREE COLLEGE SYLLABUS FOR: BIO 4944 Machine Learning for Bioinformatics	**3/1.5/3**
(New: Fall 2026)	**UK LEVEL: 6** **UK CREDITS: 15**
PREREQUISITES:	MA1024 Algebra and Trigonometry (can be bypassed via a waiver test) **or** MA 1008 College Algebra MA 2025 Applied Statistics for Sciences **or** MA 2021 Applied Statistics BI 1000 Introduction to Biology I BI 1101 Introduction to Biology II BI 2222 Cell Biology BIO 2323 Computational Biology Fundamentals BIO 4942 R for Bioinformatics ITC 2088 Introduction to Programming ITC 2164 Fundamentals of Artificial Intelligence ITC 2193 Operating Systems Concepts
CATALOG DESCRIPTION:	The course provides an overview of machine learning techniques applied to bioinformatics. The student is introduced to core concepts, algorithms, and models used in the field, including supervised and unsupervised learning. The course emphasizes practical applications in genomics, transcriptomics, protein structure prediction, and drug discovery, along with advanced topics such as deep learning, natural language processing, and image analysis. It also covers the integration of multi-omics data, ethical considerations, and emerging trends in the discipline.
RATIONALE:	The course aims to equip students with practical skills in applying machine learning to solve complex biological problems. By focusing on genomics, protein structure prediction, and drug discovery, students gain hands-on experience with real-world datasets and tools. They learn to evaluate and interpret model results, ensuring robust, reproducible research. Exposure to advanced topics such as deep learning and data integration prepares students for cutting-edge developments in bioinformatics, fostering a deep understanding of ethical implications and future trends in the field.
LEARNING OUTCOMES:	*As a result of taking this course, the student should be able to:* 1. Explain the fundamental principles and algorithms of machine learning as they apply to bioinformatics. 2. Apply supervised and unsupervised learning techniques to analyze biological datasets. 3. Evaluate machine learning models using appropriate metrics and validation techniques. 4. Develop and implement machine learning solutions for bioinformatics problems, such as genomics, transcriptomics and protein structure prediction.
METHOD OF TEACHING AND LEARNING:	In congruence with the teaching and learning strategy of the college, the following tools are used: • Class lectures, interactive learning (class discussions, group work), and practical machine learning problems solved using biological datasets. • Homework assignments involving coding exercises, data analysis, and model evaluation to reinforce concepts learned in class. • Office hours: students are encouraged to utilize instructor

<table>
<tr><td></td><td>office hours for additional guidance, discussion of lecture materials, and review of assessments.
• Use of Blackboard course management platform to support learning with lecture notes, assignment instructions, announcements, formative quizzes, and online submission of assignments</td></tr>
<tr><td>ASSESSMENT:</td><td>Summative:

<table>
<tr><td>1st assessment:
Midterm exam (2 hour)
Includes multiple-choice questions, short-answer questions, and problem-solving exercises on fundamental machine learning concepts and techniques</td><td>30%</td></tr>
<tr><td>2nd assessment: Portfolio of student work and oral assessment</td><td>10%</td></tr>
<tr><td>Final assessment:
a) Individual or group project applying machine learning methods to a bioinformatics problem.
b) Oral presentation of project findings and methodologies used.</td><td>60%</td></tr>
</table>

Formative:
- Online quizzes and coding exercises throughout the course.
- Practice datasets for hands-on model building and evaluation.
- Feedback on draft versions of project reports and presentations.
The formative assessments are designed to prepare students for the summative assessments by reinforcing the concepts and skills needed for successful completion.

The 1st assessment tests Learning Outcomes 1-2.
The 2nd assessment assesses Learning Outcomes 2-4.
The final assessment is comprehensive and evaluates all Learning Outcomes.

The final grade for this module will be determined by averaging all summative assessment grades, based on the predetermined weights for each assessment. If students pass the comprehensive assessment that tests all Learning Outcomes for this module and the average grade for the module is 40 or higher, students are not required to resit any failed assessments.</td></tr>
</table>

INDICATIVE READING:	**REQUIRED READING:** Machine Learning by Peter Flach. ISBN13: 978-1107422223 "Introduction to Machine Learning with Python: A Guide for Data Scientists" by Andreas C. Müller and Sarah Guido ISBN: 978-1449369415 “Bioinformatics Data Skills: Reproducible and Robust Research with Open-Source Tools" by Vince Buffalo ISBN: 978-1449367374 "Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More" by Bharath Ramsundar, Peter Eastman, Patrick Walters, and Vijay Pande ISBN: 978-1492039839 **RECOMMENDED READING:** "Pattern Recognition and Machine Learning" by Christopher M. Bishop ISBN: 978-0387310732 "Python Machine Learning" by Sebastian Raschka and Vahid Mirjalili ISBN: 978-1789955750
INDICATIVE MATERIAL: *(e.g. audiovisual, digital material, etc.)*	**REQUIRED MATERIAL:** Python Programming Environment (e.g., Anaconda distribution with Jupyter Notebook) Machine Learning and Bioinformatics Datasets (provided by the instructor) **RECOMMENDED MATERIAL:**
COMMUNICATION REQUIREMENTS:	N/A
SOFTWARE REQUIREMENTS:	- Anaconda distribution (Python, Jupyter Notebook, NumPy, pandas, SciPy, scikit-learn), - TensorFlow or PyTorch - Git - R with Bioconductor

WWW RESOURCES:	• General Machine Learning Resources: - Scikit-learn documentation: https://scikit-learn.org/ (covers various machine learning algorithms in Python) - TensorFlow tutorials: https://www.tensorflow.org/tutorials (covers deep learning frameworks) - Keras documentation: https://keras.io/ (high-level neural network API, often used with TensorFlow) • Bioinformatics Resources: - Bioconductor project: https://www.bioconductor.org/ (open-source software for bioinformatics) - National Center for Biotechnology Information (NCBI): https://www.ncbi.nlm.nih.gov/ (provides access to biological databases) - European Bioinformatics Institute (EMBL-EBI): https://www.ebi.ac.uk/about (provides access to biological databases and software tools) - Open access bioinformatics journals: https://bmcbioinformatics.biomedcentral.com/ and https://www.sciencedirect.com/journal/journal-of-molecular-biology (research articles on applying machine learning in bioinformatics) • Additionally: - Kaggle (https://www.kaggle.com/) where students can find datasets related to bioinformatics and practice their machine learning skills.
INDICATIVE CONTENT:	• Foundations of Machine Learning: - Supervised vs. unsupervised learning - Classification, regression, clustering, dimensionality reduction - Evaluation metrics (accuracy, precision, recall, F1-score, ROC curves) • Bioinformatics Applications: - Genomics (gene expression prediction, SNP analysis) - Proteomics (protein structure prediction, drug discovery) - Natural Language Processing (text mining biological literature) - Image Analysis (microscopy, scans) - Multi-omics data integration • Machine Learning Techniques: - Regression (linear, logistic, SVM) - Classification (decision trees, random forests, Naive Bayes) - Clustering (K-means, hierarchical, DBSCAN) - Dimensionality reduction (PCA, t-SNE, UMAP) - Deep learning (CNNs, RNNs) • Feature Engineering and Selection: - Handling biological data (high-dimensional, missing values, categorical) - Feature selection methods (filter, wrapper, embedded) - Feature engineering techniques (normalization, standardization) • Model Evaluation and Validation: - Nesting, cross-validation - Confusion matrices - Interpretability (SHAP values, LIME) • Ethical Considerations: - Bias, fairness, transparency • Emerging Trends: - Federated learning, explainable AI