DEREE COLLEGE SYLLABUS FOR: BIO 4942 R for Bioinformatics	**3/2/4**
(New: Fall 2026)	**UK LEVEL: 6** **UK CREDITS: 20**
PREREQUISITES:	MA 1024 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 2623 Computational Biology Fundamentals ITC 2088 Introduction to Programming ITC 2164 Fundamentals of Artificial Intelligence ITC 2193 Operating Systems Concepts
CATALOG DESCRIPTION:	The course provides an introduction to R programming with a focus on bioinformatics applications. Students will learn core R concepts such as data manipulation and structures, data wrangling, scripting, interaction with bash, graphics visualization and statistical analysis in R. Key bioinformatics topics include sequence analysis, genomic data handling, and RNA-seq expression analysis. The course covers clustering, regression, modeling, and phylogenetic analysis techniques in R. Practical applications emphasize reproducible research and statistical modeling in bioinformatics, culminating in case studies that apply R to real-world biological data.
RATIONALE:	This course aims to equip students with practical R programming skills tailored for bioinformatics. Students will gain hands-on experience in data manipulation, statistical analysis, and visualization of biological datasets. By exploring real-world applications, such as genomic and expression data analysis, they will develop expertise in using R for sequence analysis, clustering, classification, and machine learning. Through reproducible research practices and case studies, students are exposed to the cutting-edge tools and techniques required to tackle complex biological data challenges using R.
LEARNING OUTCOMES:	*As a result of taking this course, the student should be able to:* 1. Apply R programming to manipulate and analyze biological datasets. 2. Perform statistical analyses and visualize results using R. 3. Analyze genomic and transcriptomics RNA-seq data with relevant R packages. 4. Implement clustering, regression, and modeling algorithms in R for bioinformatics

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 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.
ASSESSMENT:	Summative: 1st assessment: Project and Presentation (take-home, written report, and presentation) a) Individual or group project applying machine learning methods to a bioinformatics problem. b) Oral presentation of project findings and methodologies used. -> 40% 2nd assessment: Portfolio of student work and oral assessment -> 10% Final assessment: Comprehensive Final Exam (3 hours) Covers all topics, including deep learning, model evaluation, and advanced applications in genomics and drug discovery. -> 50% **Formative**: Online quizzes, coding exercises and take-home problems throughout the course. **0** The formative assessments and the portfolio assignments aims to prepare students for the summative assessments. The project and presentation assess the practical aspects of Learning Outcomes 1-4. The final exam is comprehensive and evaluates the theoretical aspects of Learning Outcomes 1-4. *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.*

INDICATIVE READING:	**REQUIRED READING:** In class PowerPoint presentations. The R-Book by Michael J. Crawley. ISBN13: 978-0470973929 "R Bioinformatics Cookbook: Utilize R packages for bioinformatics, genomics, data science, and machine learning" by Dan MacLean ISBN 9781837633821 (https://www.perlego.com/book/4274722/r-bioinformatics-cookbook-utilize-r-packages-for-bioinformatics-genomics-data-science-and-machine-learning) **RECOMMENDED READING:** "Bioinformatics and Functional Genomics" by Jonathan Pevsner ISBN 9781118581766 (https://www.perlego.com/book/998683/bioinformatics-and-functional-genomics) **"Data Analysis for the Life Sciences" by** Rafael A Irizarry and Michael I Love (https://leanpub.com/dataanalysisforthelifesciences)
INDICATIVE MATERIAL: *(e.g. audiovisual, digital material, etc.)*	**REQUIRED MATERIAL:** Bioinformatics Datasets (provided by the instructor)
COMMUNICATION REQUIREMENTS:	Students are expected to demonstrate proficiency in both written and oral communication skills. This includes writing clear, concise reports and project documentation, as well as delivering oral presentations that effectively communicate complex ideas. Active participation in class discussions and collaborative group work is required to foster a productive learning environment. Additionally, students should be familiar with using the course management platform for submitting assignments and engaging in discussions.
SOFTWARE REQUIREMENTS:	Students must use a computer running Windows, macOS, or Linux with MS Office and Blackboard CMS. Required installations include a stable version of R (>4.0.0) and the RStudio Integrated Development Environment (IDE). The course also requires the installation of Bioconductor, an essential collection of R packages for bioinformatics. Key R packages that will be used throughout the course include tidyverse, ggplot2, dplyr, Biostrings, edgeR, DESeq2, phytools, and caret. Additionally, Git will be required for version control and to ensure reproducible research practices.
WWW RESOURCES:	• The R Project for Statistical Computing (Official site for R, including documentation, downloads, and user guides) https://www.r-project.org • Bioconductor (A repository of tools for the analysis of genomic data using R) https://www.bioconductor.org • CRAN R Packages (Comprehensive R Archive Network listing packages and tutorials for R users) https://cran.r-project.org • RStudio (IDE for R with resources, tutorials, and extensions for bioinformatics) https://www.rstudio.com

INDICATIVE CONTENT:	• Introduction to R programming: - Basics - IDEs (RStudio) - Scripting • Data structures in R: - Vectors, matrices, lists, and data frames • Data control structures: - If / else if / else, for, while, break, return • Importing and exporting biological data - Handling formats like FASTA, tables (TXT/CSV), etc. • Data wrangling: - Cleaning and transforming biological datasets • Basic statistical analysis: - Descriptive statistics, variables, factors, hypothesis testing, and p-values • Visualization techniques: - Creating plots, graphs, and charts for biological data • Sequence analysis: - Tools for analyzing DNA, RNA, and protein sequences • Genomic data handling: - Working with Bioconductor and other genomic data packages • Expression data analysis: - RNA-seq differential expression analysis, GO term enrichment analysis • Clustering and classification: - Implementing clustering algorithms and classification models • Phylogenetic analysis: Constructing and analyzing phylogenetic trees • Statistical modeling: Linear and generalized linear models applied to biological data • Reproducible research: Using R Markdown and version control • Case studies: Real-world applications of R in bioinformatics