Office of Academic Research

Details of the Research Scholar					
Name	**Dhivya G**			Register No.	**24PDT0002**
Programme	**Ph. D (Deep Tech)**	School	**SBST**	Category	**IFT**
Topic of Research	**A study on the detection and validation of Protein-Protein Interaction Inhibitors for Cystic Fibrosis treatment using Machine Learning-driven virtual screening approach**				

Details of Special Elective (Self-Study Course/ Guide Course)				
COURSE TITLE:	**Python Programming for Bio-informatics**			
Credit Structure (Common to all Special Elective Courses)	L	T	P	C
	0	3	0	3

Module 1

Introduction to Python and Basic Programming
Variables and Data types: Text data type, Numeric data type, Sequence data type, Mapping type.
Expressions, Type conversion, Keywords, Operators, Input statement, Conditional statements: if statement, if…else statement. Loops: Definite loop, Indefinite loop in Python, Nested loops.

Module 2

String Handling, Modular Programming, and Data Structures
String Handling: Accessing characters, string slicing, Handling DNA/RNA/protein sequences, String methods, Useful in sequence parsing and mutation handling.
Data Structures: Lists, Tuples, Dictionaries, and Sets essential for storing sequences, annotations, and metadata.
Modular Programming: Writing functions and modules, encourages reusable code in analysis pipelines.

Module 3

File Handling and Object-Oriented Programming
Handling Biological Files: CSV files for gene/protein data, FASTA files, standard format for nucleotide/protein sequences. Working with Directories, sequence datasets.
Object-Oriented Programming: Creation of custom classes for managing sequences, structures, or biological models.

Module 4

Module 4: Biopython Module
Sequence Objects: Seq, SeqRecord, and related operations, Sequence Access from FASTA, GenBank
Database Access: Fetching proteins from NCBI, UniProt, etc. Structure Analysis: Bio.PDB, parse PDB/mmCIF files, extract atoms, compute distances, Integration with PyMOL for 3D visualization.

Module 5

Data Handling and Visualization in Bioinformatics
NumPy: Fast numeric array operations, genome matrices, similarity scores, Pandas: Handling gene expression datasets, mutation tables, annotations, Matplotlib: Plotting gene expression, clustering outputs, and interaction networks, Bar plot: gene expression comparison, Histogram: sequence length distribution, Scatter plot: PCA, t-SNE, or clustering visualization.

Module 6
Practical Assignment
Machine Learning with scikit-learn: Principal component analysis, dimensionality reduction in gene/protein data, Clustering, classification of expression data or molecular fingerprints.
Mini Projects: Cluster Analysis of gene expression / sequence similarity, Application of ML in drug discovery to predict bioactivity of small molecules.
References
1. Shahnaz Verma, Parul Fatima. (2024). Python For Bioinformatics: using machine learning for drug discovery, cluster analysis, and phylogenetics. First edition. BPB PUBLICATIONS.
2. Youens-Clark, K. (2021). Mastering Python for Bioinformatics. First edition. O'Reilly Media, United states.
3. Idris, I. (2014). Python data analysis: learn how to apply powerful data analysis techniques with popular open-source Python modules. First edition. Packt Pub.
Mode of Evaluation: CAT / Assignment / Quiz / Seminar / Tutorial /FAT

Approval			
S.No	**Name of the Member**	**Role**	**Signature**
1	Dr. ABILASH V G	Dean Nominee	
2	Dr. RM. VIDHYAVATHI	External Member	
3	Dr. C. JAYAPRAKASH	External Member	
4	Dr. KUMAR K	Internal Member	
5	Dr. MANOOV R	Guide	