

Department Of Bioinformatics

Revised Syllabus of I Year Diploma Program (Part Time) (UG)

1. Title of Program: **Machine Learning in Bioinformatics**
2. Year of Implementation: 2023-2024
3. Duration: One Year
4. Pattern: Semester
5. Medium of Instruction: English
6. Contact hours: 7 hours/week
7. Structure of Course:

Syllabus Structure (UG)

Year	Semester	Course No.	Course Code	Contact Hours	Credits (1Credit=15 H)	Total Marks	
1	I	CT I	DBiT 101	30	2	75	
		CL I	DBi L101	60	2	75	
	II	CT II	DBiT 202	30	2	75	
		CL II	DBi L202	60	2	75	
	Annual	CP I	DBiP101	30	1	50	
	Total				210	9	350
2	III	CT III	DBiT 303	30	2	75	
		CL III	DBi L303	60	2	75	
	IV	CT IV	DBiT 404	30	2	75	
		CL IV	DBi L404	60	2	75	
	Annual	CP II	DBiP202	30	1	50	
	Industrial and or Incubation and or Research and or Field Training				30	1	-
	Total				240	10	350
	V	CT V	DBiT 505	30	2	75	
		CLV	DBiL505	60	2	75	
	VI	CT VI	DBiT 606	30	2	75	
		CL VI	DBiL606	60	2	75	
	Annual	CP III	DBiP303	60	2	100	
	Industrial and or Incubation and or Research and or Field Training				30	1	-
	Total				270	11	400
Total				720	30	1100	

Total No. of Courses: 15 (Theory: 6, Practical: 6, Project:3)

Theory and Practical: Semester, Project: Annual

CT: Course Theory, CL: Course Lab, CP: Course Project, D: Diploma, * : First Letter Name of Subject/Department

Semester I

CT III: DBiT 101: Introduction to Machine Learning (Contact Hrs: 30 Credits: 2)

Learning Objectives: Students will be able to...,

1. get comprehensive understanding of machine learning concepts, including supervised, unsupervised, and semi-supervised learning, and their applications in various domains.
2. acquire the knowledge and skills required to apply machine learning techniques specifically in the field of bioinformatics for the analysis and interpretation of biological data

Unit I: Fundamentals of Machine Learning (15)

- Introduction to machine learning concepts
- Supervised, unsupervised, and semi-supervised learning
- Evaluation metrics in machine learning
- Introduction to neural networks
- Applications of deep learning in genomics, proteomics, and drug discovery

Unit II: Machine Learning Applications in Bioinformatics (15)

- Predictive modeling for biological data,
- Classification and regression techniques,
- Feature selection and dimensionality reduction
- Integration of Bioinformatics and Machine Learning,
- Challenges and opportunities,
- Case studies on interdisciplinary research projects, Future directions and emerging trends

Learning Outcomes:

After completion of the unit, Student should be able to..,

1. Understand the applications of fundamental machine learning concepts.
2. Apply their Insight into integrating bioinformatics and machine learning for interdisciplinary research.

Reference Books:

1. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
2. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
3. Baldi, P., & Brunak, S. (2001). *Bioinformatics: The Machine Learning Approach*. MIT Press.
4. Aluru, S. (Ed.). (2006). *Handbook of Computational Molecular Biology*. CRC Press.
5. Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C., & Collins, J. J. (2018). Next-generation machine learning for biological networks. *Cell*, 173(7), 1581-1592.

**CLIII: DBi L 101: (Practical):
(Contact Hrs: 60 Credits: 02)**

Learning Objectives: Students will be able to

1. Gain practical experience in implementing supervised learning algorithms and evaluating their performance.
2. Understand the differences between supervised, unsupervised, and semi-supervised learning algorithms through hands-on exploration.
3. Develop proficiency in applying machine learning techniques to real-world bioinformatics problems.
4. Learn to interpret and communicate the results of machine learning models effectively in the context of biological data analysis.

List of Practical's (15)

1. Simple Supervised Learning Implementation using linear regression, k-nearest neighbors using Python libraries like scikit-learn.
2. Learning Algorithm Comparison using decision trees, k-means clustering, self-training on a dataset using Python.
3. Evaluation Metric Interpretation: Calculate and interpret evaluation metrics (e.g., accuracy, precision, recall, F1-score) for different machine learning models applied to classification tasks.
4. Basic Neural Network Classification : Build a basic neural network model using Tensor Flow or PyTorch for a classification task and analyze its performance.
5. Genomic/Proteomic Deep Learning Exploration using convolutional neural networks, recurrent neural networks for tasks such as sequence classification or drug target prediction.
6. Use machine learning algorithms to predict biological properties or outcomes (e.g., protein function, disease risk) based on relevant datasets.
7. Implement classification (e.g., support vector machines) and regression (e.g., random forest) techniques to solve bioinformatics problems such as gene expression prediction or protein structure prediction.
8. Apply feature selection methods (e.g., mutual information, feature importance) and dimensionality reduction techniques (e.g., principal component analysis) to preprocess biological datasets and improve model performance.

9. Interdisciplinary Research Question Identification

10. Integration Challenges Discussion on integrating bioinformatics and machine learning, such as data heterogeneity and interpretability issues, and brainstorm potential solutions.

11. Analyze case studies of successful interdisciplinary research projects that have utilized machine learning techniques for solving bioinformatics problems, and present findings and insights.

12. Future Trends Exploration- Explore emerging trends in bioinformatics and machine learning, such as single-cell analysis, deep learning architectures, and personalized medicine, and discuss their potential impact on future research directions

13. Conduct hyperparameter tuning for machine learning models using techniques like grid search or random search to optimize model performance on a given dataset.

14. Explore techniques for interpreting machine learning models, such as feature importance analysis, SHAP values, or LIME (Local Interpretable Model-agnostic Explanations) to gain insights into the decision-making process of the models.

15. Transfer Learning Application: Implement transfer learning approaches in machine learning models applied to bioinformatics tasks, leveraging pre-trained models (e.g., from ImageNet) and fine-tuning them on domain-specific biological data to improve performance.

Learning Outcomes:

After completion of the unit, Student should be able to

1. Demonstrate their ability to implement and compare various machine learning algorithms for classification and regression tasks.
2. Evaluating machine learning models using appropriate metrics and interpreting their results for biological data analysis.
3. Apply skills in preprocessing biological datasets through feature selection, dimensionality reduction, and other techniques to enhance model performance.
4. Execute interdisciplinary collaboration between bioinformatics and machine learning, to identify the challenges.

Reference Books:

1. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
2. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

3. Baldi, P., & Brunak, S. (2001). *Bioinformatics: The Machine Learning Approach*. MIT Press.
4. Aluru, S. (Ed.). (2006). *Handbook of Computational Molecular Biology*. CRC Press.
5. Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C., & Collins, J. J. (2018). Next-generation machine learning for biological networks. *Cell*, 173(7), 1581-1592.

Semester II**CT IV: DBiT 202: Machine Learning Foundations
(Contact Hrs: 30 Credits: 2)**

Learning Objectives: Students will be able to

1. Comprehend the foundational principles and applications of deep learning in bioinformatics, encompassing neural network architectures and their roles in genomic, proteomic, and drug discovery domains.
2. Summarize the essential techniques of data preprocessing and feature engineering, aiming to prepare raw data effectively for deep learning models in bioinformatics applications.

Unit I: Deep Learning in Bioinformatics (15)

- Introduction to neural networks: perceptrons, activation functions, feedforward networks
- Deep learning architectures: convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their variants
- Applications of deep learning in genomics: sequence analysis, gene expression prediction
- Applications of deep learning in proteomics: protein structure prediction, protein-protein interaction prediction
- Applications of deep learning in drug discovery: molecular property prediction, virtual screening
- Case studies on deep learning models in bioinformatics with hands-on projects

Unit II: Fundamentals of Data Preprocessing and Feature Engineering (15)**1.Data Preprocessing:**

- Address missing data and outliers through appropriate imputation and treatment methods.
- Normalize numerical features and encode categorical variables for compatibility with machine learning algorithms.

2.Feature Engineering:

- Create informative features by transforming existing variables and generating new ones.
- Utilize techniques such as one-hot encoding, feature scaling, and dimensionality reduction to enhance model performance.

Learning Outcomes:

After completion of the unit, Student should be able to

1. Demonstrate a comprehensive understanding of neural network fundamentals and their diverse applications in bioinformatics, evidenced by their ability to explain the workings of various architectures and their relevance to genomics, proteomics, and drug discovery.
2. Acquire proficiency in data preprocessing and feature engineering techniques, enabling them to prepare complex bioinformatics datasets for deep learning analysis effectively.

Reference Books:

1. Alipanahi, B., Delong, A., Weirauch, M. T., & Frey, B. J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature Biotechnology*, 33(8), 831–838. [doi:10.1038/nbt.3300](https://doi.org/10.1038/nbt.3300)
2. Angermueller, C., Pärnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for computational biology. *Molecular Systems Biology*, 12(7), 878. [doi:10.15252/msb.20156651](https://doi.org/10.15252/msb.20156651)
3. Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., ... & Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*, 15(141), 20170387. [doi:10.1098/rsif.2017.0387](https://doi.org/10.1098/rsif.2017.0387)
4. Huang, C., Sun, Y., Shrivastava, A., & Sukthankar, R. (2017). FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2392-2400). [doi:10.1109/cvpr.2017.258](https://doi.org/10.1109/cvpr.2017.258)
5. Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning* (pp. 807-814). [doi:10.1145/3104322.3104425](https://doi.org/10.1145/3104322.3104425)
6. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. [Scikit-learn website](https://scikit-learn.org/stable/)

CL VI: DBiL 202: (Practical):
(Contact Hrs: 60 Credits: 02)

Course Objectives: Students will be able to..

1. Understand neural network fundamentals and their implementation in Python.
2. Explore advanced neural network architectures for bioinformatics tasks.
3. Apply deep learning techniques to genomic analysis and prediction tasks.
4. Gain proficiency in preprocessing biological data for machine learning models.

List of Practical's (15)

1. Learn about basic neural network components like perceptrons and activation functions, and implement them using Python libraries.
2. Application of different activation functions affect model performance.
3. Explore convolutional neural networks (CNNs) and recurrent neural networks (RNNs) and apply them to tasks like genomic sequence analysis and protein structure prediction.
4. Demonstrate the structure and functionality of CNNs and RNNs for bioinformatics tasks.
5. Use deep learning to predict DNA or RNA binding sites and analyze gene functions or regulatory elements.
6. Evaluate deep learning models for their effectiveness in genomic analysis tasks.
7. Apply deep learning for protein structure prediction and protein-protein interaction prediction.
8. Implement deep learning models for protein sequence analysis and evaluate their performance.
9. Learn methods to deal with missing data and assess their impact on model performance.
10. Apply missing data imputation techniques and evaluate their effectiveness.
11. Identify outliers in datasets and apply appropriate treatment methods.
12. Assess the impact of outlier treatment on model stability.
13. Encode categorical variables for machine learning models using techniques like one-hot encoding.
14. Compare the performance of models trained with different encoding methods.
15. Scale numerical features and apply transformations like log transformation.

Course Outcomes:

1. Implement basic neural network components and evaluate activation functions' impact on model performance.
2. Utilize CNNs and RNNs for protein structure prediction and sequence analysis.

3. Apply deep learning to predict biological phenomena like DNA/RNA binding sites and gene functions.
4. Master data preprocessing techniques such as handling missing data, outlier detection, and feature scaling for bioinformatics tasks.

Reference Books:

1. Alipanahi, B., Delong, A., Weirauch, M. T., & Frey, B. J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature Biotechnology*, 33(8), 831–838. [doi:10.1038/nbt.3300](https://doi.org/10.1038/nbt.3300)
2. Angermueller, C., Pärnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for computational biology. *Molecular Systems Biology*, 12(7), 878. [doi:10.15252/msb.20156651](https://doi.org/10.15252/msb.20156651)
3. Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., ... & Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*, 15(141), 20170387. [doi:10.1098/rsif.2017.0387](https://doi.org/10.1098/rsif.2017.0387)
4. Huang, C., Sun, Y., Shrivastava, A., & Sukthankar, R. (2017). FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2392-2400). [doi:10.1109/cvpr.2017.258](https://doi.org/10.1109/cvpr.2017.258)
5. Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning* (pp. 807-814). [doi:10.1145/3104322.3104425](https://doi.org/10.1145/3104322.3104425)
6. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. [Scikit-learn website](https://scikit-learn.org/stable/)

CP II: DBiP202 (Project):
(Contact Hrs. 60, Credits: 2)

Industrial and or Incubation and or and or Field Training is compulsory

BOS Sub-Committee

- 1.Dr.N.N.Bendre Chairman
- 2.Ms.S.N.Sanglikar, Member
3. Mr.P.M.Bhosale. Member

Expert Committee

1. Name of Academic Expert: Dr.R.A.Jamdade
2. Name of Industrial Expert: Dr.Bajrang Kumbhar