INDIAN BIOLOGICAL SCIENCES AND RESEARCH INSTITUTE
भारतीय जैविक विज्ञान एवं अनुसंधान संस्थान
An ISO 9001:2015 Certified Institute
Knowledge Leads Innovation

CURRICULUM DOCUMENT

Artificial Intelligence & Machine Learning in Biology

6-Month Diploma Program

Admissions Close — 20th September 2026

- Bioinformatics & Data Science • Genomics & Proteomics • Deep Learning
- Drug Discovery • Systems Biology • Capstone Project

Duration	6 Months (24 Weeks)	**Mode**	Online / Offline / Hybrid
Schedule	2 days/week • 2 hrs/class	**Level**	Diploma
Course Code	IBRI-AIML-BIO-06	**Certification**	IBRI Diploma Certificate
Admissions Close	20th September	**Career Support**	LinkedIn, Resume & Placement Assistance

About the Programme

This diploma equips learners with a working, project-ready foundation in Artificial Intelligence and Machine Learning as applied to biological and biomedical data. Over six months, students progress from core programming, data handling, and biological literacy through classical machine learning and deep learning, into applied genomics, proteomics, drug discovery, and systems biology culminating in a mentor-guided capstone project modelled on real research and industry workflows.

Duration	6 Months (24 Weeks)	**Mode**	Online / Offline / Hybrid
Eligibility	B.Sc./B.Tech/M.Sc. in Life Sciences, Biotech, Bioinformatics, CS or allied fields	**Prerequisites**	Basic biology + willingness to code (no prior programming required)
Weekly Schedule	2 days/week • 2 hours/class • 4 hours/week contact	**Certification**	IBRI Diploma Certificate on completion of capstone

Programme Learning Outcomes

- Apply core machine learning and deep learning techniques to biological and biomedical datasets.
- Clean, wrangle, and visualize real biological data to a publication-ready standard.
- Design and execute end-to-end computational pipelines for genomics, proteomics, and drug discovery problems.
- Use industry-standard tools (Python, R, AutoDock, Cytoscape) confidently in a research or industry setting.
- Critically evaluate model performance, bias, and reproducibility in biomedical AI applications.
- Independently plan, execute, and present a mentor-guided applied research project.

Programme Structure at a Glance

Month	Theme	Focus
Month 1	Foundations & Data Science	AI/ML orientation, Python & R, core biology, data wrangling, visualization & EDA
Month 2	Machine Learning	Supervised & unsupervised learning, model evaluation, dimensionality reduction
Month 3	Deep Learning	Neural networks, CNNs, RNNs, sequence models
Month 4	Multi-Omics & Language Models	Statistics for genomics, preprocessing at scale, GWAS, proteomics, biological language models
Month 5	Applied AI — Drug Discovery	Docking, QSAR, ADMET, drug-target & repurposing prediction
Month 6	Systems Biology & Capstone	Networks, biomedical NLP, research methodology, mentor-guided project

Software & Tools Stack

- **Languages:** Python, R, Bash/Linux shell
- **ML/DL Frameworks:** scikit-learn, PyTorch, TensorFlow/Keras
- **Bio Tools:** BLAST, ClustalW, AutoDock Vina / PyRx, PyMOL, Cytoscape
- **Data & Notebooks:** Pandas, NumPy, Matplotlib, Seaborn, Jupyter, Google Colab, GitHub
- **Databases:** NCBI, UniProt, PDB, GEO, KEGG, STRING

Month 1 — Foundations: AI, ML, Biology & Data Science Bridge

Environment setup, programming basics, data wrangling, visualization, and the biological literacy needed for computational work.

Week	Topics Covered
Week 1	Orientation: landscape of AI/ML in biology, career pathways IDE setup — Jupyter, Google Colab, VS Code Python fundamentals: variables, control flow, functions Setting up a GitHub-based lab workflow and version control basics
Week 2	NumPy & Pandas for biological data handling Introduction to R and RStudio Data import/cleaning: CSV, FASTA, Excel-based biological datasets
Week 3	Biology refresher: central dogma, genomics/transcriptomics/proteomics overview Common bioinformatics file formats: FASTA, FASTQ, PDB, VCF, GenBank Data wrangling: filtering, grouping, and merging gene/protein/clinical datasets
Week 4	Data visualization for biological insight: Matplotlib, Seaborn, ggplot2, publication-quality plotting Case studies: AlphaFold, cancer diagnostics, AI-driven drug pipelines Exploratory data analysis (EDA) on a public biological dataset

Learning Outcomes

- Navigate Python/R environments and manipulate tabular and sequence-based biological data.
- Explain the central dogma and identify major omics data types and file formats.
- Independently clean, wrangle, and visualize a messy biological dataset to publication quality.
- Produce a clean, reproducible EDA workflow using version control.

> **Month-End Mini Project:** Exploratory Data Analysis report with publication-quality visualizations on a public gene-expression or clinical dataset, submitted via GitHub.

Month 2 — Machine Learning in Bioinformatics

Classical ML algorithms applied to gene expression, clinical, and diagnostic datasets.

Week	Topics Covered
Week 1	Supervised learning: linear & logistic regression, decision trees, k-NN; applied case: gene-expression-based disease classification
Week 2	Advanced supervised learning: SVM, Random Forest, XGBoost; model evaluation — accuracy, precision/recall, F1-score, ROC-AUC, cross-validation
Week 3	Unsupervised learning: k-means, hierarchical clustering, DBSCAN; applied case: microbiome clustering and patient subgroup discovery
Week 4	Dimensionality reduction: PCA, t-SNE, UMAP; feature selection strategies for high-dimensional omics data; mini-project build week

Learning Outcomes

- Select and justify an appropriate ML algorithm for a given biological problem.
- Evaluate model performance using standard metrics and cross-validation.
- Apply dimensionality reduction and feature selection to high-dimensional omics data.

> **Month-End Mini Project:** Classify cancer subtypes from microarray / RNA-seq expression data using at least two ML algorithms, with a model comparison report.

Month 3 — Deep Learning Foundations

Neural network foundations through convolutional and sequence models.

Week	Topics Covered
Week 1	Neural network foundations: perceptrons, backpropagation, activation & loss functions; introduction to PyTorch/TensorFlow
Week 2	Convolutional Neural Networks (CNNs); applications: histopathology image classification, medical imaging diagnostics
Week 3	Sequence models: RNNs, LSTMs, attention mechanisms; applications: DNA/protein sequence classification, promoter prediction
Week 4	Applied sequence-model practice and architecture selection (CNN vs. RNN vs. structured data); mini-project build week

Learning Outcomes

- Build, train, and evaluate a neural network for a biological classification task.
- Distinguish appropriate architectures (CNN vs. RNN) for image, sequence, and structured biological data.

> **Month-End Mini Project:** Build a CNN-based classifier for histopathology images (e.g., cancer vs. non-cancer tissue) and report accuracy/loss curves.

Month 4 — Data Engineering, Multi-Omics & Biological Language Models

Scaling preprocessing pipelines, statistical foundations for genomics, and integrating multiple omics layers.

Week	Topics Covered
Week 1	Data preprocessing at scale: cleaning, normalization, batch effect correction, handling missing and imbalanced biological data Statistical foundations for genomics: hypothesis testing, multiple-testing correction, effect size
Week 2	Applications in genomics: variant calling basics, GWAS with ML, SNP-disease association
Week 3	Applications in proteomics & structural biology: protein structure/stability prediction (ddG / free-energy change), homology modelling basics, AI-assisted docking overview; multi-omics integration and network-based integration (STRING, KEGG pathway enrichment)
Week 4	Transformers and biological language models (ESM, DNABERT, AlphaFold concepts); transfer learning with pretrained biological models; mini-project build week

Learning Outcomes

- Design a preprocessing pipeline that handles noisy, missing, or imbalanced omics data.
- Perform a basic GWAS-style association analysis with correct statistical controls and interpret pathway enrichment results.
- Integrate at least two omics data layers into a single analytical workflow.
- Apply transfer learning using a pretrained biological language model.

> **Month-End Mini Project:** End-to-end SNP-disease association pipeline: preprocessing → statistical/ML association test → pathway enrichment interpretation, with a short transfer-learning exercise using a pretrained biological language model.

Month 5 — Applied AI: Drug Discovery

From virtual screening to drug-target and combination-therapy prediction.

Week	Topics Covered
Week 1	AI in drug discovery: virtual screening, QSAR modelling, ADMET prediction basics
Week 2	Molecular docking with AutoDock Vina / PyRx
Week 3	Drug-target and drug-drug interaction prediction with ML
Week 4	Drug repurposing and drug-synergy/combination-therapy prediction (deep learning); mini-project build week

Learning Outcomes

- Run and interpret a molecular docking-based virtual screening workflow.
- Apply ML models to predict drug-target interactions and repurposing candidates.

> **Month-End Mini Project:** Docking-based virtual screening for a chosen protein target, ranked by binding affinity with interaction analysis.

Month 6 — Systems Biology, Capstone Project & Professional Skills

Network biology, biomedical NLP, research methodology, and the mentor-guided capstone.

Week	Topics Covered
Week 1	Biological networks and systems biology: construction & analysis (Cytoscape, NetworkX), centrality measures, motif detection, pathway enrichment NLP & LLMs in biomedicine: literature mining, biomedical text classification, responsible use of LLMs Research methodology: problem definition, literature review strategy, research proposal writing Data ethics, reproducibility, and bias in biomedical AI
Week 2-3	Mentor-guided capstone project execution (choose one track — see below); weekly progress reviews and troubleshooting sessions with mentor
Week 4	Scientific writing and poster/paper drafting; final project presentation and defense; portfolio building, GitHub/resume polishing, mock interviews

Learning Outcomes

- Construct and analyze a biological network to identify key nodes/pathways.
- Apply NLP/LLM tools responsibly to accelerate literature-based research.
- Independently scope, plan, and execute an applied AI/ML-in-biology research project.
- Communicate technical findings clearly through written reports and oral presentation.

> **Month-End Mini Project:** Full Capstone Project, submitted with report, code repository, and a live presentation/defense.

Capstone Project Tracks

In Month 6, each learner selects one mentor-guided capstone track below. Tracks may be adapted based on learner interest and mentor availability.

Track	Project Focus
Oncology & Diagnostics	Cancer subtype classification / drug-resistance prediction using ML/DL on expression or imaging data
Infectious Disease & Immunology	AI-based vaccine/epitope design or antimicrobial-resistance prediction
Structural Bioinformatics	Protein structure or stability (ddG / free-energy change) prediction using deep learning models
Drug Discovery	Drug-target interaction / drug-repurposing pipeline combining ML and docking
Single-Cell & Transcriptomics	Single-cell RNA-seq clustering and biomarker discovery

Assessment & Certification

- Weekly lab assignments — hands-on exercises reinforcing each week's topics (30%)
- Monthly mini-projects (Months 2–5) — applied deliverables graded on correctness, methodology, and reporting (30%)
- Capstone project (Month 6) — written report, code repository, and live presentation/defense (30%)
- Participation & professional readiness — attendance, portfolio quality, engagement (10%)

A minimum of 75% attendance and successful completion of the capstone project (with defense) are required to receive the IBRI Diploma in AI & ML in Biology.

Who Should Apply

- Life sciences, biotechnology, bioinformatics, pharmacy, and allied-field graduates/postgraduates seeking computational skills.
- Computer science / data science graduates seeking to specialize in biomedical and life-sciences applications.
- Working professionals in biotech, pharma, or healthcare aiming to upskill in AI/ML.

Career Support & Perks

Every learner receives dedicated career-readiness support alongside the technical curriculum, at no additional cost:

- **Free LinkedIn Profile Creation & Enhancement:** One-on-one profile build-out or optimization to help you get discovered by recruiters in AI/biology roles.
- **Free Resume Building:** A recruiter-ready resume built around your capstone project, tools stack, and diploma credentials.
- **Placement Assistance:** Support connecting graduates with hiring partners and referrals in bioinformatics, biotech, and AI-driven research roles.

⏳ Admissions close on 20th September. Enrol before the deadline to secure your seat and lock in the Early Bird fee.

PROGRAMME FEE

₹ 50,000

Early Bird: 10% off → ₹ 45,000 | Limited Seats Available

⏳ Reserve Your Seat — Only 20 Seats Available!

Register with just ₹1,000 to lock in your seat. This amount will be deducted from your total programme fee at the time of final enrolment.

Seats are limited and allotted on a first-come, first-served basis — register early to avoid missing out.

REGISTER HERE ⬇

https://docs.google.com/forms/d/e/1FAIpQLSd3sRewxsBa8vj-KcJaCVdEO3vqI1HZTmeTJFSJCjLnMQZGzQ/viewform?usp=publish-editor

Still got questions or doubts?

Please reach out to us at

✉ info@ibri.org.in

☎ +91 99995 09892

🌐 ibri.org.in

☎ +91 99714 41910

VISIT US

Indian Biological Sciences and Research Institute (IBRI)
C-50, 2nd Floor, Sector 2, Noida, Uttar Pradesh, India