DEPARTMENT OF BIOTECHNOLOGY
Ministry of Science & Technology
Government of India

A Hands-on Workshop on AI and Data Science for Advancing Translational Science

Book of Abstracts

Translational Bioinformatics Group (TBG)
International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi

11th - 15th November, 2025

सत्यमेव जयते
DEPARTMENT OF BIOTECHNOLOGY
Ministry of Science & Technology
Government of India

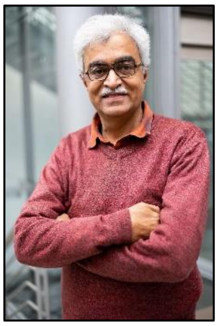
A Hands-On Workshop on AI and Data Science for Advancing Translational Science

11th November 2025 – 15th November 2025

Book of Abstracts

Translational Bioinformatics Group
International Center for Genetic Engineering and Biotechnology (ICGEB),
Aruna Asaf Ali Marg, New Delhi – 110067,
India

From the Director’s Desk



Dear Workshop Participants,

A warm welcome to ICGEB and the 5-day Workshop on **“AI and Data Science for Advancing Translational Science.”** We are delighted to have you join us for this exciting and timely workshop that brings together researchers, clinicians, and bioinformaticians at the intersection of biology, translational science, and artificial intelligence.

The **International Centre for Genetic Engineering and Biotechnology (ICGEB)** is a unique intergovernmental organization established under the United Nations system to promote research, training, and international collaboration in the field of genetic engineering and biotechnology. Headquartered in **Trieste (Italy)** with regional components in **New Delhi (India)** and **Cape Town (South Africa)**, ICGEB serves as both a **centre of excellence and a hub for capacity building** across its Member States.

ICGEB’s mission is to foster sustainable development through advanced research in molecular biology, genetics, and biotechnology, addressing challenges in **health, agriculture, energy, and the environment.** It supports scientific innovation, technology transfer, and education, particularly in developing countries, enabling them to apply modern biotechnology to meet local and global needs.

Through its extensive network of collaborations, research programs, and fellowship schemes, ICGEB continues to play a vital role in advancing science for the benefit of humanity — **bridging discovery with application** and **empowering the next generation of scientists worldwide. Recognizing the potential of AI, ICGEB has initiated training programs in AI to enhance the capabilities of researchers from developing countries.**

Artificial Intelligence (AI) and Data Science are transforming the way we understand complex biological systems, translate laboratory discoveries into practical applications, and guide research. From accelerating drug discovery and improving diagnostics to personalizing treatment strategies, AI-driven approaches are unlocking new possibilities for precision medicine and global health. Similarly, AI-driven innovations have tremendous potential for applications in agriculture, synthetic biology,

Over the course of this workshop, we look forward to engaging discussions, hands-on sessions, and collaborations that will empower participants to harness the potential of AI for impactful translational research. I wish you all a great learning experience during the next 5 days of the workshop.

Dr. Ramesh V. Sonti
Director
ICGEB, New Delhi, India

FOREWORD

It is a pleasure to present this Abstract Book for the ICGEB International Workshop on ***"AI and Data Science for Advancing Translational Science*,"** being held at the International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, from 11th to 15th November 2025.

This workshop aims to build research capacity and awareness in the rapidly evolving fields of artificial intelligence (AI) and data science, which are now driving major advances in biological and medical research. The program brings together scientists, clinicians, and technologists to explore how computational and machine learning approaches can help interpret complex biological datasets and translate discoveries into practical healthcare and life science applications.

The event is coordinated by the Translational Bioinformatics Group at ICGEB, which focuses on integrating multi-dimensional biological data, including genomics, transcriptomics, proteomics, and medical imaging, using computational and AI-based methods. The group's research focuses on converting large-scale data into meaningful biological insights that support diagnostic innovation, therapeutic discovery, and precision medicine.

Over the five days, the workshop features expert lectures, interactive discussions, and hands-on sessions designed to provide participants with both conceptual understanding and practical experience in AI-driven data analysis. By combining theoretical knowledge with real-world problem-solving, the workshop seeks to empower participants to apply these approaches effectively within their own domains.

I am grateful to **Dr. Lawrence Banks (Director-General, ICGEB)** and **Dr. Ramesh V. Sonti (Director, ICGEB, New Delhi)**, for their continued encouragement of initiatives that promote innovation and skill development in AI-enabled life sciences research. I also thank our sponsors and collaborators, including SHI Solutions India Private Limited, AWS, and Netweb Technologies, as well as Tyrone, for their contributions to this program.

I sincerely acknowledge the efforts of our invited speakers, faculty members, and the organizing team, especially the members of the Translational Bioinformatics Group, for their commitment and hard work in bringing this event together.

We hope that the ideas, interactions, and collaborations emerging from this workshop will continue to inspire new directions in translational research and strengthen the global community working at the interface of biology, computation, and data science.

Dr. Dinesh Gupta
Group Leader,
Translational Bioinformatics Group
ICGEB, New Delhi

Schedule

Workshop on “A hands-on workshop on AI and data science for advancing translational science” organized by ICGEB, New Delhi (11th Nov. 2025 - 15th Nov. 2025)

Day 1 (11.11.2025)

Timing	Title	Speaker
09:00 AM - 09:30 AM	**Registration**	
09:30 AM - 10:00 AM	Workshop Inauguration by Director ICGEB	Dr. Ramesh V. Sonti, ICGEB, New Delhi
	Welcome Speech by the Organizer	Dr. Dinesh Gupta, ICGEB, New Delhi
10:00 AM - 10:30 AM	Fundamentals and Applications of AI	Dr. Dinesh Gupta, ICGEB, New Delhi
10:30 AM - 11:00 AM	**Tea Break**	
11:00 AM - 11:45 AM	Hands-on: Basic Exercises	Dr. Dinesh Gupta, ICGEB, New Delhi
11:45 AM - 12:30 PM	Data Science in Translational Bioinformatics: Foundations and Applications	Dr. Nidhi Malhotra, Shiv Nadar University, Greater Noida, UP
12:30 PM - 01:00 PM	Hands-on: Data Science in Translational Bioinformatics: Foundations and Applications	
01:00 PM - 02:00 PM	**Lunch Break**	
02:00 PM - 03:00 PM	Introduction to CNN and Transformers	Mr. Aashish Thakur, Dr. Ankit Singhal, ICGEB, New Delhi
03:00 PM – 03:30 PM	Predicting gene expression level from promoter DNA sequences using neural networks	Dr. Divya Tej Sowpati, CCMB, Hyderabad, Telangana
03:30 PM – 04:00 PM	Hands-on: Predicting gene expression level from promoter DNA sequences using neural networks	
04:00 PM - 04:30 PM	**Tea Break**	

04:30 PM - 05:30 PM	Hands-on: Predicting gene expression level from promoter DNA sequences using neural networks	Dr. Divya Tej Sowpati, CCMB, Hyderabad, Telangana
Day 2 (12.11.2025)		
Timing	**Title**	**Speaker**
09:30 AM - 10:30 AM	MATTER: Metrics and Assessment Tools for Trustworthy, Transparent, Explainable, and Reliable AI	Dr. Priyanka Narad, ICMR, New Delhi
10:30 AM - 11:00 AM	**Tea break**	
11:00 AM - 11:30 AM	Data-driven approaches to correlate female reproductive microbiota with pregnancy complications	Dr. Abhishek Sengupta, Amity University, Noida, UP
11:30 AM - 01:00 PM	Hands-on: From Data to Decision using AI/ML in reproductive Health	
01:00 PM - 02:00 PM	**Lunch Break**	
02:00 PM - 03:00 PM	Exploring LLMs: From ideas to Applications	Ms. Dhairya, ICGEB, New Delhi
03:00 PM - 03:30 PM	Local Tools for AI: Running Models, Chat & Non-Chat Agents, and AI-Assisted Coding	Dr. Silvano Piazza, ICGEB, Italy
03:30 PM – 04:00 PM	**Tea break**	
04:00 PM - 05:30 PM	Hands-on: Local Tools for AI: Running Models, Chat & Non-Chat Agents, and AI-Assisted Coding	Dr. Silvano Piazza, ICGEB, Italy
Day 3 (13.11.2025)		
Timing	**Title**	**Speaker**
09:30 AM - 10:00 AM	Artificial Intelligence for Single-Cell Genomics Sequencing Data Analysis	Dr. Stefano Cacciatore, ICGEB, SA

10:00 AM - 10:20 AM	**Tea break**	
10:20 AM - 12.00 PM	Hands-on KODAMA: Unsupervised Learning for Single-Cell Data Analysis	Dr. Stefano Cacciatore, ICGEB, SA
12:00 PM - 01:15 PM	Introduction to Cloud computing and AI for accelerating Bio-Medical Research with AWS	Mainak Chakraborty AWS Healthcare and Life Sciences
01.15 PM - 02:15 PM	**Lunch Break**	
02:15 PM - 02:45 PM	Introduction to Machine learning for tabular data	Dr. Ishaan Gupta, IIT Delhi
02:45 PM – 04:15 PM	Hands-on: Introduction to Machine learning for tabular data	
04:15 PM - 04:30 PM	**Tea break**	
04:30 PM – 05:00 PM	A Case for Neuro-Symbolic AI : 2 Instances of Image Generation (Vision) & Learning to Build Complex Structures (Robotics)	Dr. Parag Singla, IIT Delhi
05:00 PM – 06:00 PM	Hands-on: Neuro-Symbolic AI	
Day 4 (14.11.2025)		
Timing	**Title**	**Speaker**
09:30 AM -10:00 AM	Functional insights into the Indian human microbiomes revealed through multiomics approaches	Dr. Vineet K. Sharma, IISER Bhopal, MP
10:00 AM - 10:30 AM	**Tea break**	
10:30 AM - 12:00 PM	Hands-on: Functional insights into the Indian human microbiomes revealed through multiomics approaches	Dr. Vineet K. Sharma, IISER Bhopal, MP
12:00 PM - 12:30 PM	A Ping-Pong Match in Plasma: How an AI-Guided Inertrope Reads the Game to Diagnose Cancer	Dr. Gaurav Ahuja, IIIT Delhi
12:30 PM – 01:00 PM	Hands-on: Customized tailoring of Biology-informed Drugs using Generative AI	

01:00 PM - 02:00 PM	**Lunch Break**	
02:00 PM - 03:00 PM	Hands-on: Customized tailoring of Biology-informed Drugs using Generative AI	Dr. Gaurav Ahuja, IIIT Delhi
03:00 PM – 04:00 PM	Graph Neural Networks: Applications in Bioinformatics	Mr. Sumit Kumar, ICGEB, New Delhi
04:00 PM - 04:30 PM	**Tea break**	
04:30 PM – 05:00 PM	Power BI: Turning Data into Actionable Insights	Dr. Shahzaib Ahamad, ICGEB, New Delhi
05:00 PM – 06:00 PM	AI-Driven Translational Innovation in Healthcare Imaging: TBG Lab	Dr. Dinesh Gupta, Dr. Shweta Birla, ICGEB, New Delhi
07:00 PM onwards	**Gala Dinner**	
Day 5 (15.11.2025)		
Timing	**Title**	**Speaker**
09:30 AM - 10:00 AM (Venue: Seminar Hall, Ground Floor)	Concepts and Applications of Convolutional Neural Networks (CNN) for Radiology Image Analysis	Dr. Tanmay Basu, IISER Bhopal, MP
10:00 AM - 10:30 AM (Venue: Seminar Hall, Ground Floor)	Hands-on: Concepts and Applications of Convolutional Neural Networks (CNN) for Radiology Image Analysis	
10:30 AM – 10:45 am	**Tea break**	
10:45 AM - 01:00 PM (Venue: Seminar Hall, Ground Floor)	Hands-on: Concepts and Applications of Convolutional Neural Networks (CNN) for Radiology Image Analysis	Dr. Tanmay Basu, IISER Bhopal, MP
01:00 PM - 02:00 pm	**Lunch Break**	
02:00 PM - 03:00 pm	**Closing Session**	Dr. Dinesh Gupta, ICGEB, New Delhi

Organizer:	**Organizing Committee:**	**Volunteers:**
Dr. Dinesh Gupta	Dr. Jatin Shrinet	Mr. Sumit Kumar
	Dr. Shweta Birla Dhakonia	Ms. Prachi Priyam
	Ms. Minakshi Sharma	Mr. Aashish Thakur
	Mr. Abhishek Khatri	Dr. Sonu Kumar
	Dr. Ankit Singhal	Dr. Vikram
	Dr. Rakesh	Dr. Shruti Gupta
		Dr. Shahzaib Ahamad
		Ms. Gurupriya Takkar
		Ms. Tamseel Fatma
		Ms. Yashika Gupta

ABSTRACTS

(Day 1)

Fundamentals and Applications of AI

Dr. Dinesh Gupta*
Translational Bioinformatics Group,
International Centre for Genetic Engineering and Biotechnology,
New Delhi, India

Email: dinesh@icgeb.res.in

The lecture and the following hands-on session will introduce the workshop participants to the fundamentals and applications of Artificial Intelligence (AI). The lecture will begin with the introduction of AI, its historical development, and core concepts. Next, the talk will shift gears to emphasize how AI is transforming translational science, leveraging data science, affordable, faster computing, and the availability of large datasets. The second part of the lecture will introduce statistical concepts, data types, probability, correlation, regression, and other related topics that help establish the efficiency, robustness, interpretability, and generalizability of trained AI models. The concepts like data collection, preprocessing, exploratory data analysis, risks, advantages, and disadvantages of AI will also be discussed. The lecture will also illustrate real-world examples of how AI is accelerating translational science — from biomarker discovery, synthetic biology, and drug target identification, patient stratification, disease prediction, and precision therapeutics. The lecture will conclude with an introduction to the hands-on requirements, learning mode, and objectives of the rest of the workshop proceedings.

Data Science in Translational Biology: Foundations and Applications

Dr. Nidhi Malhotra*
Department of Chemistry,
School of Natural Sciences, Shiv Nadar Institution of Eminence,
Gautam Buddha Nagar, UP

Email: nidhi.malhotra@snu.edu.in

The integration of data science and artificial intelligence (AI) with omics-driven biology is reshaping translational research and accelerating the drug discovery and development pipeline. Rentosertib exemplifies this shift, becoming the first fully AI-designed drug to advance to Phase IIa clinical trials in 2025. Data science enables systematic analysis of complex biological datasets, transforming them into actionable insights for understanding health and disease. This session will introduce participants to core data science concepts and their applications in biology, with emphasis on high-throughput approaches such as differential gene expression analysis and drug sensitivity profiling. Key steps including data preprocessing, quality control, visualization, and statistical analysis will be highlighted to demonstrate how robust analytical practices enhance reproducibility and can help shorten the drug discovery process. The session will also include a practical hands-on component, providing participants with experience in accessible workflows. By bridging biological questions with computational methods, the workshop aims to equip participants with both a conceptual foundation and practical skills that can be applied across diverse translational science contexts.

Introduction to CNN and Transformers

Aashish Thakur*

Translational Bioinformatics Group,
International Centre for Genetic Engineering and Biotechnology,
New Delhi, India

Email: aashish@icgeb.res.in

Artificial Intelligence has witnessed remarkable progress through the development of deep learning architectures, among which Convolutional Neural Networks (CNNs) and Transformers have emerged as two foundational paradigms. This lecture seeks to impart a conceptual basis of these models, detailing their main principles, origin, and importance in contemporary artificial intelligence solutions. CNNs are essentially meant for processing grid-structured information like images, with their spatial hierarchies and local relations captured using convolutional layers efficiently. Their hierarchical feature learning mechanism has rendered them invaluable for computer vision applications like image classification, object detection, and analyzing medical images. Conversely, Transformers transformed the discipline of Natural Language Processing with the idea of self-attention, allowing models to learn long-range dependencies and contextual relations without the use of recurrence or convolution. Scalability and flexibility of the transformer architecture have led to its application extending beyond text to vision and multimodal learning. Through this talk, we discuss how both architectures differ in philosophy in design but complement one another in pushing the boundaries of deep learning. By knowing their underlying mechanisms and relative strengths, researchers and practitioners can better appreciate their revolutionary impact in enabling intelligent systems with perception, reasoning, and decision-making abilities over various data modalities.

Predicting Gene Expression from DNA Sequences Using Neural Networks

Dr. Divya Tej Sowpati*
Centre for Cellular & Molecular Biology,
Habsiguda, Uppal Road, Hyderabad

Email: tej@ccmb.res.in

The genome contains a complex regulatory code written in the language of DNA, which dictates the precise expression of thousands of genes. Deciphering this code is fundamental to understanding biology. However, this remains a key challenge due to the underlying complexity. Artificial intelligence offers a powerful new toolkit for identifying and learning subtle patterns within sequences that are linked to specific biological functions.

This talk and the subsequent session will explore the core concept of treating gene regulation as a predictive problem: can we learn a function that maps a DNA sequence to its functional output, such as gene expression? We will discuss how computational models, specifically neural networks, act as powerful pattern recognizers. By analyzing thousands of examples, these systems can learn to associate specific features in a promoter sequence with high or low gene expression.

The goal is to provide an intuitive understanding of this new data-driven paradigm. While we will not be developing a "state-of-the-art" predictive model, we will delve into the foundational concepts required to understand and eventually build more complex and powerful models.

ABSTRACTS

(Day 2)

MATTER: Metrics and Assessment Tools for Trustworthy, Transparent, Explainable, and Reliable AI

Dr. Priyanka Narad*

Division of Development Research,
Indian Council of Medical Research,
New Delhi
Email: pnarad.hq@icmr.gov.in

Artificial Iintelligence’s rapid deployment in domains such as healthcare, finance, and administration necessitates rigorous frameworks for trust, transparency, and explainability. **MATTER**—*Metrics and Assessment Tools for Trustworthy, Transparent, Explainable, and Reliable AI*—brings together cutting-edge benchmarks, interpretability methods, and governance practices for systematic evaluation. Benchmarks like **TrustLLM** and **AIR-Bench 2024** utilize 30+ datasets across 18 evaluation dimensions, including robustness to adversarial inputs, hallucination detection, privacy leakage analysis, fairness across subpopulations, and ethical reasoning. Technical advances in explainability are equally vital. Tools such as **LIME**, **SHAP**, and saliency mapping, embedded in **IBM’s AI Explainability 360**, provide model-agnostic, human-understandable insights into decision pipelines. Governance metrics now integrate data lineage tracking, bias monitoring, and *adaptive trust calibration*, bridging technical evaluation with societal accountability. Despite progress, challenges remain in chain-of-thought reliability, dynamic bias detection, and ensuring fairness under domain shift—underscoring the need for holistic frameworks like MATTER.

From an Indian perspective, the **Indian Council of Medical Research (ICMR)** has pioneered AI ethics in healthcare through its *2023 Guidelines on AI in Biomedical Research and Healthcare*. These emphasize algorithmic transparency, human-in-the-loop validation, India-specific datasets, and task-based risk evaluation. ICMR is also investing in AI-driven clinical trial management, digital health registries, and context-sensitive bias assessments, aligning India’s healthcare AI landscape with global best practices. In this talk, I will present the MATTER framework as a unifying architecture that bridges international benchmarks with ICMR’s initiatives, enabling the development of AI systems that are technically robust, ethically sound, and societally aligned.

Keywords: MATTER framework, Trustworthy AI, Explainable AI (XAI), AI governance, Transparency, Reliability

Microbiomics Integrated with Machine Learning Driving Precision in Female Reproductive Health

Dr. Abhishek Sengupta*

Systems Biology and Data Analytics Research Lab,
Centre for Computational Biology and Bioinformatics,
Amity Institute of Biotechnology, Amity University,
Noida, Uttar Pradesh

Email: asengupta@amity.edu

The dynamic interplay between the vaginal microbiome and reproductive physiology has emerged as a critical determinant of maternal health outcomes. Recent advances in bioinformatics and artificial intelligence (AI) are enabling a deeper, systems-level understanding of how microbial communities influence gynaecological and obstetric health. This talk presents an integrated overview of the theoretical underpinnings and real-world applications of microbiome-informed AI approaches in reproductive medicine. We will explore how next-generation sequencing and metagenomic pipelines are leveraged to decode microbial signatures, followed by the application of machine learning algorithms to predict adverse reproductive conditions. Emphasis will be placed on microbial community transitions, immune modulation, and host–microbe crosstalk, interpreted through the lens of computational modelling. Case studies will demonstrate how these techniques are used to stratify risk, identify biomarkers, and support precision diagnostics. The presentation also discusses emerging trends—such as AI, network-based interpretation, and integration with other omics technologies—that are shaping the future of microbiome research. Together, these approaches offer a promising framework for transforming reproductive healthcare from reactive to predictive, with profound implications for early detection and personalized intervention.

Local Tools for AI: Running Models, Chat & Non-Chat Agents, and AI-Assisted Coding

Dr. Silvano Piazza*
International Centre for Genetic Engineering and Biotechnology,
Trieste, Italy
Email: Silvano.Piazza@icgeb.org

This session will explore the growing ecosystem of **local AI tools** that allow researchers and developers to run and experiment with AI models directly on their own machines — without relying exclusively on cloud APIs. I'll start with an overview of *local AI models*: what they are, why they matter (privacy, reproducibility, cost efficiency), and the distinction between **chat-based** and **non-chat** models.We'll then look at practical tools for deploying and using these models locally, including **Hugging Face Transformers** (for downloading and running open models), **llama.cpp** and **LocalAI** (for efficient CPU/GPU inference), and **text-generation-webui** for quick experimentation.The second part of the session will focus on **AI-assisted coding** using local or open-source tools such as **Windsurf**, and **Tabby** showing how they integrate with local models or Hugging Face checkpoints to provide coding assistance similar to Copilot.During the hands-on, participants will run a small local model (e.g. a LLaMA or Mistral variant), connect it to VS Code or Jupyter, and test code completion, summarization, and prompt engineering in both chat and script modes.

ABSTRACTS

(Day 3)

Artificial Intelligence for Single-Cell Genomics Sequencing Data Analysis

Dr. Stefano Cacciatore*
International Centre for Genetic Engineering and Biotechnology,
Cape Town, South Africa

Email: Stefano.Cacciatore@icgeb.org

Artificial intelligence (AI) has become a cornerstone of single-cell genomics, enabling the discovery of subtle cellular phenotypes and trajectories that traditional analytical methods often miss. The explosion of single-cell RNA-sequencing (scRNA-seq) demands approaches capable of managing non-linearity, noise, and high dimensionality.

This talk will present an overview of how AI and machine-learning methods have beed used for single-cell analysis—from data preprocessing and batch correction to cell-type annotation and lineage inference. It will highlight representative algorithms such as manifold learning approaches that capture complex cellular relationships.

Particular attention will be given to **unsupervised and weakly supervised learning**, including methods like **KODAMA**, which iteratively maximizes sample discrimination to reveal biologically meaningful substructures without prior labels. Case studies will illustrate how AI can uncover hidden cell states, predict differentiation pathways, and identify new biomarkers in cancer and immunology.

The session aims to equip attendees with a conceptual understanding of how AI empowers hypothesis generation and data-driven discovery in single-cell biology.

Introduction to Cloud computing and AI for accelerating Bio-Medical Research with AWS

Mainak Chakraborty*
Senior Solutions Architect,
AWS Healthcare and Life Sciences

Email: mainakc@amazon.com

Research is an innovation practice – with researchers needing access to tools that enable rapid experimentation. Amazon Web Services (AWS) works with prominent research labs and researchers around the world to offer cost-effective, scalable, and secure compute, storage, and database capabilities to accelerate time to science. With AWS, researchers can quickly analyze massive data pipelines, store petabytes of data, and advance research using transformative technologies like artificial intelligence (AI), machine learning (ML), and quantum – all while securely sharing their results with collaborators around the world. AWS also provides researchers with access to open datasets, funding, and training to accelerate the pace of innovation. Join me in this introductory session to understand what cloud computing is and how you can use it accelerate your bio-medical research.

Introduction to Machine learning for tabular data

Dr. Ishaan Gupta*
Department of Biochemical Engineering and Biotechnology,
Indian Institute of Technology Delhi,
New Delhi, India

Email: Ishaan.Gupta@dbeb.iitd.ac.in

Machine learning (ML) methods have become indispensable for discovering patterns and predictive relationships in biomedical and translational datasets. Unlike images or sequences, most clinical and experimental data—such as gene expression matrices, proteomic profiles, patient metadata, and electronic health records—are inherently tabular in structure. This talk will introduce key ML frameworks designed for tabular data, including classical algorithms such as SVM, random forests, gradient boosting (XGBoost, LightGBM, CatBoost), and emerging deep learning models tailored to structured inputs. Emphasis will be placed on how these approaches handle heterogeneous variables, missing values, and class imbalance commonly encountered in biological datasets. Through translational applications, we will explore how models built on tabular data can predict patient outcomes, drug responses, or biomarker signatures. The session will also highlight practical strategies for feature selection, data preprocessing, model evaluation, and interpretability—critical elements for ensuring biological and clinical relevance. By the end of the session, participants will understand how to structure, analyze, and interpret tabular data using ML tools, enabling them to translate heterogeneous biomedical information into actionable knowledge for precision medicine and disease research.

A Case for Neuro-Symbolic AI : 2 Instances of Image Generation (Vision) & Learning to Build Complex Structures (Robotics)

Dr. Parag Singla*
Department of Computer Science and Engineering,
Indian Instiute of Technology,
Hauz Khas, New Delhi
Email: parags@cse.iitd.ac.in

The last decade has seen phenomenal growth in application of neural models to a variety of problems, including those in Computer Vision, NLP and Robotics & Sciences, among other domains. One of the recent research directions in the deep learning world has been around the problem of incorporating symbolic & object-centric representation in neural networks, to exploit the underlying problem structure, and enable them to be more accurate by exploiting domain knowledge and attain effective generalisation while also being more interpretable. In this talk, we will present two different problem instances from 2 different application domains, viz., (a) image generation (b) learning to construct grounded spatial concepts. We will also discuss corresponding solutions that exploit the underlying problem structure for building effective neuro-symbolic architectures to solve these problems. First, we deal with the problem of text-conditioned image generation. Our approach reposes the problem of image generation, as one of generation followed by (multi-step) editing, where the editing specifically focuses on objects & their properties which do not align with original instruction, identified via an MLLM. In the second instance, we look at the problem of learning complex structures from data (e.g., a staircase is a sequence of towers of increasing height), and explore how these can be learned by a robotic agent from human demonstrations, in the form of executable (python) programs. Our approach combines MCTS for guiding program search, with LLMs to aid generalisation of learned programs. For each of these problem instances, we will have a discussion on potential application of the ideas presented to translational science.

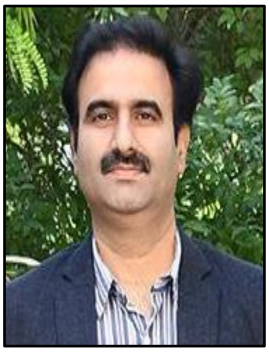
ABSTRACTS

(Day 4)

Functional insights into the Indian human microbiomes revealed through multiomics approaches

Dr. Vineet K. Sharma*
MetaBioSys Group,
Indian Institute of Science Education and Research,
Bhopal, Madhya Pradesh, India

Email: vineetks@iiserb.ac.in



The Indian human microbiome has several unique aspects such as the human gut that is primarily enriched in Prevotella species with high diversity and compositional variations in both inter-individual and inter-population comparisons. To investigate the composition, diversity, and functional roles of the key driver species in human microbiomes, we carried out a population-wide analysis on healthy samples belonging to western and non-western populations, and generated gut microbiome data for a large Indian cohort. A higher abundance and diversity of *Prevotella copri* species enriched in complex plant polysaccharides metabolizing enzymes, particularly pullulanase containing polysaccharide-utilization-loci (PUL), were found in Indian and non-western populations. A higher diversity of oral inflammations-associated *Prevotella* species and an enrichment of virulence factors and antibiotic resistance genes in the gut microbiome of western populations speculates an existence of a mouth-gut axis. We also included the Inflammatory Bowel Disease samples from western populations to examine the association of Prevotella with inflammatory bowel disease (IBD), including ulcerative colitis and Crohn's disease. The results were further compared with IBD cohorts from US, Netherlands, and Spain. This study provided new insights into the role of diversity, composition, and function of key bacterial species in gut microbiome and their impact on human health. The study also revealed the landscape composition of human gut microbiome and its impact on health in western and non-western populations, particularly in the Indian context. We also examined the differential abundance of Flavanoid Metabolizing Enzymes in the gut microbiomes of healthy and diseased individuals from Western and non-Western populations with distinct dietary habits. Furthermore, using a hybrid of AI, machine learning and chemoinformatics approaches, we have developed novel AIML based computational approaches to predict the metabolic enzymes and bacterial species that can potentially carry out the promiscuous biotransformation of xenobiotic molecules.

A Ping-Pong Match in Plasma: How an AI-Guided Inertrope Reads the Game to Diagnose Cancer

Dr. Gaurav Ahuja*
Department of Computational Biology,
Indraprastha Institute of Information Technology,
Delhi, New Delhi, India

Email: gaurav.ahuja@iiitd.ac.in

This talk introduces a revolutionary liquid biopsy that redefines cancer diagnostics. Instead of asking what is in the blood, we ask how the blood behaves as a complete thermodynamic system by playing a game of 'thermodynamic ping-pong' in just a few drops of patient plasma. Our star player is Inertrope, a unique biologically inert probe discovered and guided by AI. It doesn’t hunt for a specific biomarker; instead, it ‘pings’ the plasma, and we read the energetic rally that follows. This interaction generates a rich calorimetric fingerprint, a holistic heat signature of the patient's health. When fed into a machine learning model, these fingerprints allowed us to classify cancer, benign, and healthy states with over 95% AUC-ROC in a 176-patient cohort. Mechanistic insights link the role of hydrogen peroxide in plasma with that of the observed pathology-specific calorimetric fingerprints. Our work establishes a powerful 'diagnostic-by-thermodynamics' paradigm, transforming chemical "dark matter" into potent diagnostic probes and paving the way for a new generation of non-invasive cancer detection. Please note: The hand-on session will be conducted in building the model using ITC, and shall include data acquisition, pre-processing, and then model building and validation.

Graph Neural Networks - Applications in Bioinformatics

Mr. Sumit Kumar*
Translational Bioinformatics Group,
International Centre for Genetic Engineering and Biotechnology,
New Delhi, India

Email: sumit@icgeb.res.in

Graph Neural Networks (GNNs) have emerged as powerful tools for modeling complex, non-Euclidean data structures, making them especially valuable in the life sciences. With the rise of high-throughput technologies, biological data is increasingly represented as graphs—ranging from drug compounds and protein–protein interaction networks to gene regulatory networks. GNNs enable the extraction of meaningful patterns from such data, offering new opportunities in bioinformatics and computational biology. This workshop introduces participants to the foundational principles of GNNs and their applications in life sciences. We will cover the core architecture and mechanisms of GNNs, illustrating how they address three major research challenges: node-level prediction, edge-level prediction, and graph-level prediction. The session will include a case study followed by a hands-on practical segment, where participants will build GNN models for real-world biological data, including steps for graph construction, model training, and evaluation. We will also discuss key considerations such as interpretability challenges, best practices in designing biological graph models, and future directions at the intersection of network biology and machine learning.

References

1. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. arXiv [cs.LG] (2016).
2. Stokes, J. M. et al. A deep learning approach to antibiotic discovery. Cell 180, 688-702.e13 (2020).
3. Huang, K. et al. A foundation model for clinician-centered drug repurposing. Nat. Med. 30, 3601–3613 (2024).
4. Li, S., Hua, H. & Chen, S. Graph neural networks for single-cell omics data: a review of approaches and applications. Brief. Bioinform. 26, (2025).

Power BI: Turning Data into Actionable Insights

Dr. Shahzaib Ahamad*
Translational Bioinformatics Group,
International Centre for Genetic Engineering and Biotechnology,
New Delhi, India

Email: shah.bioinfo@gmail.com

Power BI is a data visualization and analysis tool that allows users to connect, clean, and explore data through interactive dashboards. In bioinformatics and healthcare research, large datasets such as genomic profiles, clinical records, biomarkers, and transcriptomic data require clear visual interpretation to support meaningful conclusions. The workflow includes data import, preprocessing using Power Query, model building, and interactive plots using DAX (Data Analysis Expressions) functions. Power BI enhances bioinformatics research by enabling real-time data exploration, easy result sharing, and improved decision-making. Overall, this session highlights how Power BI can support data-driven research in life sciences and healthcare, among other areas.

Keywords: Power BI, Data Visualization, Biomedical Data Analysis, Dashboard Analytics, Data Integration, Power Query.

In vivo identification of angle dysgenesis and its relation to genetic markers associated with glaucoma using artificial intelligence

Dr. Dinesh Gupta*
Translational Bioinformatics Group,
International Centre for Genetic Engineering and Biotechnology,
New Delhi, India

Email: dinesh@icgeb.res.in

Purpose: To predict the presence of angle dysgenesis on anterior-segment optical coherence tomography (ADoA) by using deep learning (DL) and to correlate ADoA with mutations in known glaucoma genes.

Participants: In total, 800 high-definition anterior-segment optical coherence tomography (AS-OCT) images were included, of which 340 images were used to build the machine learning (ML) model. Images used to build the ML model included 170 scans of primary congenital glaucoma (16 patients), juvenile-onset open-angle glaucoma (62 patients), and adult-onset primary open-angle glaucoma eyes (37 patients); the rest were controls (n = 85). The genetic validation dataset consisted of another 393 images of patients with known mutations that were compared with 320 images of healthy controls.

Methods: ADoA was defined as the absence of Schlemm's canal, the presence of hyperreflectivity over the region of the trabecular meshwork, or a hyperreflective membrane. DL was used to classify a given AS-OCT image as either having angle dysgenesis or not. ADoA was then specifically looked for on AS-OCT images of patients with mutations in the known genes for glaucoma.

Results: The final prediction, which was a consensus-based outcome from the three optimized DL models, had an accuracy of >95%, a specificity of >97%, and a sensitivity of >96% in detecting ADoA in the internal test dataset. Among the patients with known gene mutations, (MYOC, CYP1B1, FOXC1, and LTBP2) ADoA was observed among all the patients in the majority of the images, compared to only 5% of the healthy controls.

Conclusion: ADoA can be objectively identified using models built with DL.

A Novel Three-Stage AI-Assisted Approach for Accurate Differential Diagnosis and classification of NIFTP and Thyroid Neoplasms

Dr. Shweta Birla*
Translational Bioinformatics Group,
International Centre for Genetic Engineering and Biotechnology,
New Delhi, India
Email: sbirla@icgeb.res.in

The recent introduction of the term non-invasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) marked a pivotal shift in the classification of encapsulated follicular variants of papillary thyroid carcinoma (EFVPTC) lacking invasive features. While its reclassification from the “malignant” to “low-risk neoplasm” category significantly reduced overtreatment, its histopathological diagnosis remains challenging due to overlapping features with other thyroid lesions and inter-observer variability. Artificial intelligence (AI) overcomes such key limitations of histopathological evaluation, ensuring a robust and efficient diagnostic process. While preliminary studies are promising, AI models capable of efficiently distinguishing NIFTP from other benign and malignant thyroid entities are yet to be developed. We devised an innovative AI-based three-stage hierarchical pipeline that systematically evaluates architectural patterns and nuclear features. The prioritized models were trained using 154,498 patches, derived from 134 sections prepared from 125 thyroid nodules, representing follicular nodular disease (FND), follicular adenoma, dominant nodule in FND, invasive EFVPTC (IEFVPTC), and classic and infiltrative follicular subtypes of PTC. External validation revealed good accuracy at the overall, patient-wise, and classwise levels. However, it showed limitations in the differential diagnosis of NIFTP from IEFVPTC—an expected challenge due to overlapping nuclear features and the absence of incorporating the assessment of the tumor capsule for invasive characteristics. While the novel approach and the algorithm show promise in transforming histopathological NIFTP diagnostics, further improvements and rigorous validations are necessary before considering its application in real-world clinical settings.

ABSTRACTS

(Day 5)

Concepts and Applications of Convolutional Neural Networks (CNN) for Radiology Image Analysis for Better Treatment Planning

Dr. Tanmay Basu*

Department of Data Science and Engineering,
Indian Institute of Science Education and
Research,
Bhopal, Madhya Pradesh, India

Email: tanmay@iiserb.ac.in

With the advent of artificial intelligence, the perspective of radiology diagnostics is being changed in terms of automatization towards quick and effective treatment planning. Deep learning techniques, a typical artificial intelligence technology, have become prevalent in medical images analysis in this decade. Convolutional Neural Networks (CNN) are prominent deep learning models that have been extensively used in many aspects of radiology image analysis, including almost all kinds of modalities like X-Ray, CT, MRI etc. to improve the precision and accuracy of computer-aided diagnosis. It is designed to learn spatial hierarchies of features through backpropagation over multiple layers, such as convolution, pooling, and fully connected layers to interpret complex but latent information underlying in the images. The objective of this tutorial cum hands-on session will be to explain the basics of different building blocks of various kinds of CNNs with illustrations through different use cases, so that it will be useful to understand their power and potential. The hands-on will deliver the coding skills required to implement CNN architecture for chest X-Ray and brain hemorrhage classification in CT scans followed by brain tumor contouring using the U-Net model, another kind of convolutional architecture for image segmentation. Eventually the candidates is expected to be skilled in implementing CNN for various other kinds of translational health sciences problems like gene network inference, proteomics analysis etc.

Translational Bioinformatics Group,
International Center for Genetic Engineering and Biotechnology (ICGEB),
New Delhi –110067, India

@2025