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.