Bioinformatics

Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data, in particular when the data sets are large and complex. As an interdisciplinary field of science, bioinformatics combines *biology, computer science, information engineering, mathematics and statistics* to analyse and interpret the biological data.

To study how normal cellular activities are altered in different disease states, the biological data must be combined to form a comprehensive picture of these activities. Therefore, the field of bioinformatics has evolved such that the most pressing task now involves the analysis and interpretation of various types of data. This also includes nucleotide and amino acid sequences, protein domains, and protein structures. The actual process of analysing and interpreting data is referred to as computational biology.

Important sub-disciplines within bioinformatics and computational biology include:

- Development and implementation of computer programs that enable efficient access to, management and use of, various types of information.
- Development of new algorithms (mathematical formulas) and statistical measures that assess relationships among members of large data sets. For example, there are methods to locate a gene within a sequence, to predict protein structure and/or function, and to cluster protein sequences into families of related sequences.

Major research efforts in the field include sequence alignment, gene finding, genome assembly, drug design, drug discovery, protein structure alignment, protein structure prediction, prediction of gene expression and protein–protein interactions, genome-wide association studies, the modelling of evolution and cell division/mitosis.

Gene Ontology is a major bioinformatics initiative to unify the representation of gene and gene product attributes across all species. More specifically, the project aims to:

- maintain and develop its controlled vocabulary of gene and gene product attributes
- annotate genes and gene products and assimilate and disseminate annotation data
- offer tools for easy access to all aspects of the data provided by the project

Applications of Neural Networks in Bioinformatics

With the exponential growth of biological data, one needs to pay attention to the efficient storage and management of information, also to extract relevant information from this data. There are several biological domains where machine learning tools can be utilized for extracting the information from data, following are applications of neural network in bioinformatics:

1. In the recognition of coding region of genes.
2. In the identification of genes problems
3. Identification and analysis of signals generated from regulatory sites
4. Sequence, classification, and features detection
5. Expression of genetic and genomic data
6. Image and signal processing

Compiled By: Rohan Khurana

Categorization of deep learning applied research in bioinformatics

	Omics	Biomedical imaging	Biomedical signal processing
Deep neural networks	Protein structure Gene expression regulation Protein Classification Anomaly classification	Anomaly classification Segmentation Recognition Brain decoding	Brain decoding Anomaly classification
Convolutional neural networks	Gene expression regulation	Anomaly classification Segmentation Recognition	Brain decoding Anomaly classification
Recurrent neural networks	Protein structure Gene expression regulation Protein Classification		Brain decoding Anomaly classification
Emergent architectures	Protein structure	Segmentation	Brain decoding

*Emergent architectures: (DST-NNs - Deep spatio-temporal neural networks, MD-RNNs – Multi - Dimensional Recurrent Neural networks and CAEs – Convolutional Auto-Encoders)

- In Omics research, genetic information such as genome, transcriptome and proteome data are used to approach problems in bioinformatics. Some of the most common input data in omics are raw biological sequences (i.e., DNA, RNA, amino acid sequences) which have become relatively affordable and easy to obtain with next generation sequencing technology.

- Biomedical imaging is another an actively researched domain with a wide application of deep learning in general image related tasks. Most biomedical images used for clinical treatment of patients—magnetic resonance imaging (MRI), radiographic imaging, positron emission tomography (PET) and histopathology imaging - have been used as input data for deep learning algorithms.

- Biomedical signal processing is a domain where researchers use recorded electrical activity from the human body to solve problems in bioinformatics. Various data from EEG, electrocorticography (ECoG), electrocardiography (ECG), electromyography (EMG) and electrooculography (EOG)have been used.

References:

1) Seonwoo Min, Byunghan Lee, Sungroh Yoon, Deep learning in bioinformatics, *Briefings in Bioinformatics*, Volume 18, Issue 5, September 2017, Pages 851–869, https://doi.org/10.1093/bib/bbw068

2) Bioinformatic Analyses and Gene Distributions https://bio.libretexts.org/Bookshelves/Microbiology/Book%3A_Microbiology_(Boundless)/7%3A_Microbial_Genetics/7.02%3A_Prokaryotic_Genomes/7.2D%3A_Bioinformatic_Analyses_and_Gene_Distributions

Compiled By: Rohan Khurana