

FOR REVIEW

Learning Purpose - Omics and Radiomics Data Analysis Using Deep learning

Nazmul Ahasan Maruf*

*Corresponding author. Email: nmaruf0001@stu.kau.edu.sa

Abstract

Omics data is accessible as genome, proteome, transcriptome, and metabolome in the age of big data. Aside from single omics data, integrative omics, also known as multi-omics, and omics imaging data, also known as radiomics, are being employed for predictive analysis. To process, standardize, integrate, and analyze omics data, numerous computational methodologies such as data mining, machine learning, deep learning, statistical methods, and metaheuristic techniques have gained traction.

1. Background Overview Of Omics Data

The omics research started with genomics, which aids in the study of genetic variations associated with a complicated and hereditary illness. Genomic research focuses on the whole genome, as opposed to genetic research, which focuses on single genes or particular variations. The omics discipline is growing in relevance as technological advances allow for the efficient study of biological molecules. Omics data is classified into four types: genome, proteome, transcriptome, and metabolome. A genome is used in a cell to represent DNA molecules, whereas a proteome is used to represent proteins. The transcriptome is used to represent RNA molecules, whereas the metabolome is used to represent metabolites.

1.1 Genome

A genome contains all of an organism's DNA information. The primary goal of genomics in medical research is to find genetic variations associated with disease, therapy response, and patient prognosis. Whole-genome sequencing (DNA-seq) is another approach for interrogating DNA information that is used to assemble and detect genetic variations in a previously sequenced organism.

Database

Both the Sequence Read Archive (SRA) and the Gene Expression Omnibus (GEO) include publically accessible genomic data.

1.2 Proteome

The proteome describes the whole universe of proteins in a cell. Mass-spectrometry is the platform utilized for proteome profiling. MS-based approaches are utilized for protein analysis and quantification, as well as for analyzing hundreds of proteins in bodily fluids or cells.

Database

Proteome profiling is kept in the databases ProteomeXchange, PRIDE, and ProteomicsDB.

1.3 Transcriptome

The transcriptome is the collection of all RNA molecules in a cell. Transcriptomics examines RNA levels throughout the genome in both qualitative (presence of transcripts, discovery of new splice sites) and quantitative (expression value of each transcript) ways.

Database

The GEO database and SRA are the public databases that provide transcriptional profiling.

1.4 Metabolome

The metabolome is a full collection of small-molecule types found in an organism, such as amino acids, carbohydrates, and fatty acids. Metabolome quantification, like proteome quantification, is found using Mass-Spectrometry (MS) technology.

Database

There are extremely few databases that include metabolome data. MetaboLights is one of the databases that includes metabolome studies.

2. Radiomics

Radiomics is the extraction of mineable data from medical imaging and has been used in cancer to enhance diagnosis, prognosis, and clinical decision support with the objective of providing precision medicine. In the new era of precision medicine, there is a growing interest in investigating possible biomarkers encoded in various pictures. With the advancement of powerful machine and deep learning algorithms, it is now possible to extract the form and texture of tissues of interest from multimodality images such as X-ray, computed tomography (CT), magnetic resonance (MR), positron emission tomography (PET), and ultrasound (US). Radiomics is a new area of research that combines computational and analytical tools for medical pictures.

3. Analysis

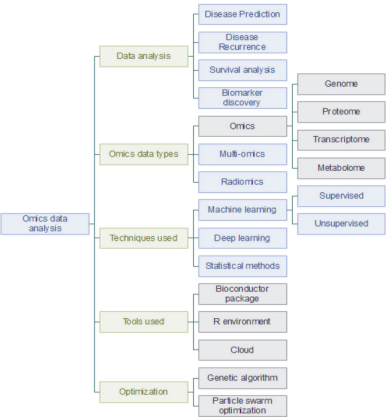


Fig. 5 Taxonomy of omics data analysis

Figure 1. Taxonomy of Omics Data Analysis