Role of Data Mining Techniques in Bioinformatics

Pushpa Singh, Delhi Technical Campus, Greater Noida, India

iD https://orcid.org/0000-0001-9796-3978

Narendra Singh, G. L. Bajaj Insitute of Management and Research, India

iD https://orcid.org/0000-0002-6760-8550

ABSTRACT

Data mining offers a highly effective technique that is useful in research and development of bioinformatics. Bioinformatics consists biological information such as DNA, RNA, and protein. Data mining tasks/techniques are classification, prediction, clustering, association, outlier detection, regression, and pattern tracking. Data mining provides important correlation, hidden patterns, and knowledge from the bioinformatics data set. This paper presents the role of data mining techniques in bioinformatics application. Classification of gene and protein structure, analyzing the gene expression, association of co-disease, outlier detection and gene selection, protein structure prediction, and drug discovery are some typical biological example that has proven data mining as a suitable technique for bioinformatics.

KEYWORDS

Bioinformatics, Classification, Clustering, Data Mining, Genes, Proteins

INTRODUCTION

Bioinformatics is the integration of biology, mathematics, statistics, medicines, information technology, and computer science. Bioinformatics is the skill of storing, retrieving and analyzing huge amounts of biological information such as DNA, RNA, and Proteins etc. (Bayat, 2002). Recent technological advancement permits the biologists to produce huge volumes of data ranging from measurements of DNA database, Protein sequence, protein structure database, Phenotype database and Genomic sequence database etc. Bioinformatics holds great potential of analysis in the different areas like genome, proteomics, drug discovery and development, protein structure, cell biology, molecular modelling, gene expression (Khan, 2018) etc. as represented in figure 1. one can analysis and extract valuable pattern in gene expression, classify protein structure, gene prediction, gene identification, diagnosing different types of disease (cancer etc.) on which genes are expressed etc. Data Mining offers capability to analysis of bioinformatics data, and useful to pattern identification, classification, prediction and genetic network induction (Mabu, 2018).

In today's world, data is the base for everything, if it is analyzeand extracted properly. In bioinformatics various types of data is available for mining as shown in figure 2.

DNA: It's the genetic code that determines all the characteristics of a living thing. DNA is heridatry material means child got his DNA from his parents. Smaller units of DNA are called as
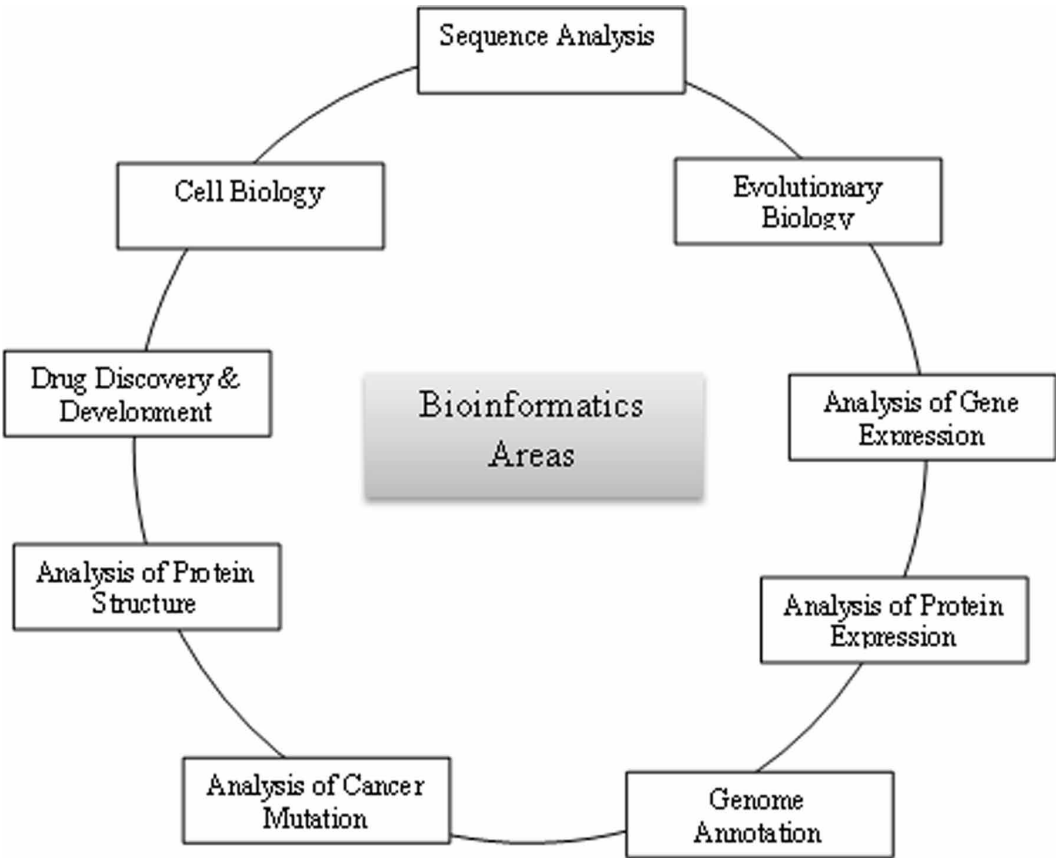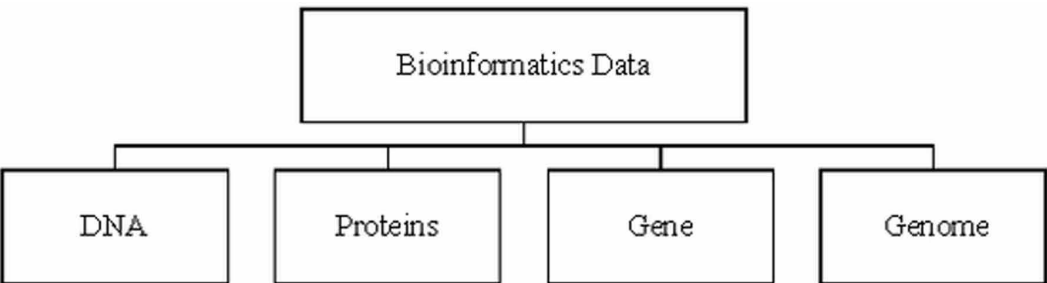
Figure 1. Bioinformatics areas

Figure 2. Types of data in bioinformatics

nucleotides. Each nucleotide entails three part nitrogen, sugar (ribose) and phosphate. There are four type of nitrogen bases are adenine (A), thymine (T), guanine (G) and cytosine (C). The order of these bases governs the genetic code (Dua & Chowriappa, 2012).

   Proteins: Proteins are huge, complex molecules that very significant for the body. Protein consists twenty different amino acids. Sequence of these amino acids regulates each protein's unique 3D structure and its precise function.

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/role-of-data-mining-techniques-in-bioinformatics/267825

Related Content

Data Graphs for Linking Clinical Phenotype and Molecular Feature Space
Andreas Heinzel, Raul Fechete, Johannes Söllner, Paul Perco, Georg Heinze, Rainer Oberbauer, Gert Mayer, Arno Lukasand Bernd Mayer (2012). *International Journal of Systems Biology and Biomedical Technologies (pp. 11-25).*
www.irma-international.org/article/data-graphs-linking-clinical-phenotype/63043

Spatial Uncertainty Analysis in Ecological Biology
Stelios Zimerasand Yiannis Matsinos (2013). *International Journal of Systems Biology and Biomedical Technologies (pp. 14-24).*
www.irma-international.org/article/spatial-uncertainty-analysis-ecological-biology/78389

Computational Models Relevant For Visual Cortex
Mitja Perušand Chu Kiong Loo (2011). *Biological and Quantum Computing for Human Vision: Holonomic Models and Applications (pp. 229-234).*
www.irma-international.org/chapter/computational-models-relevant-visual-cortex/50510

In Silico Models on Algal Cultivation and Processing: An Approach for Engineered Optimization
Lamiaa H. Hassan, Imran Ahmad, Mostafa El Sheekhand Norhayati Abdullah (2024). *Research Anthology on Bioinformatics, Genomics, and Computational Biology (pp. 989-1016).*
www.irma-international.org/chapter/silico-models-algal-cultivation-processing/342560

Genome Subsequences Assembly Using Approximate Matching Techniques in Hadoop
Govindan Rajaand U. Srinivasulu Reddy (2017). *International Journal of Knowledge Discovery in Bioinformatics (pp. 83-97).*
www.irma-international.org/article/genome-subsequences-assembly-using-approximate-matching-techniques-in-hadoop/190794