


# Chapter 63

## Role of Data Mining Techniques in Bioinformatics

**Pushpa Singh**

 <https://orcid.org/0000-0001-9796-3978>

*KIET Group of Institutions, Delhi-NCR, Ghaziabad, India*

**Narendra Singh**

 <https://orcid.org/0000-0002-6760-8550>

*G. L. Bajaj Institute of Management and Research, India*

### 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.*

### INTRODUCTION

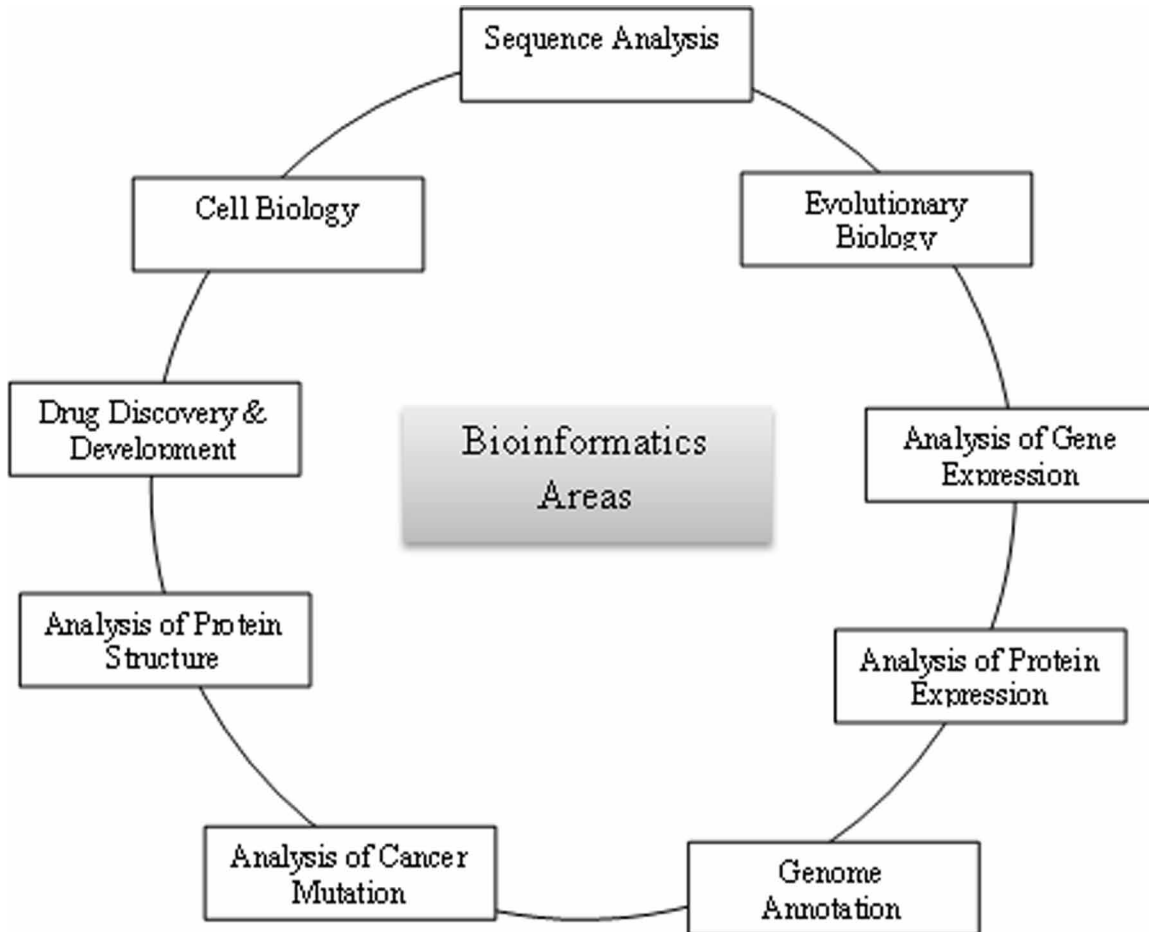
Bioinformatics is the integration of biology, mathematics, statistics, medicines, information technology, and computer science. Bioinformatics is the skill of storing, retrieving and analyzing huge amounts of biological information such as DNA, RNA, and Proteins etc. (Bayat, 2002). Recent technological 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

DOI: 10.4018/979-8-3693-3026-5.ch063

## Role of Data Mining Techniques in Bioinformatics

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).

Figure 1. Bioinformatics areas



In today's world, data is the base for everything, if it is analyzed and 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 hereditary material means child got his DNA from his parents. Smaller units of DNA are called as 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).

10 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/chapter/role-data-mining-techniques-bioinformatics/342581](http://www.igi-global.com/chapter/role-data-mining-techniques-bioinformatics/342581)

## Related Content

---

### Heuristic Principal Component Analysis-Based Unsupervised Feature Extraction and Its Application to Bioinformatics

Y-H. Taguchi, Mitsuo Iwadata, Hideaki Umeyama, Yoshiki Murakami and Akira Okamoto (2015). *Big Data Analytics in Bioinformatics and Healthcare* (pp. 138-162).

[www.irma-international.org/chapter/heuristic-principal-component-analysis-based-unsupervised-feature-extraction-and-its-application-to-bioinformatics/121456](http://www.irma-international.org/chapter/heuristic-principal-component-analysis-based-unsupervised-feature-extraction-and-its-application-to-bioinformatics/121456)

### Statistical Methods Applied in Drug Safety

Partha Chakraborty (2012). *Pharmacoinformatics and Drug Discovery Technologies: Theories and Applications* (pp. 268-279).

[www.irma-international.org/chapter/statistical-methods-applied-drug-safety/64077](http://www.irma-international.org/chapter/statistical-methods-applied-drug-safety/64077)

### Time-Aware Task Allocation for Cloud Computing Environment

Sushanta Meher, Sohan Kumar Pande and Sanjaya Kumar Panda (2017). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 1-13).

[www.irma-international.org/article/time-aware-task-allocation-for-cloud-computing-environment/178603](http://www.irma-international.org/article/time-aware-task-allocation-for-cloud-computing-environment/178603)

### Can Activated Platelet Rich Plasma Combined with Adipose-Derived Stem Cells Be Used to Treat Skin Wrinkles?: A Mechanism Study

Phuc Van Pham, Loan Thi-Tung Dang, Nhung Hai Truong and Ngoc Kim Phan (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications* (pp. 920-936).

[www.irma-international.org/chapter/can-activated-platelet-rich-plasma/76102](http://www.irma-international.org/chapter/can-activated-platelet-rich-plasma/76102)

### 3D Structural Bioinformatics of Proteins and Antibodies: State of the Art Perspectives and Challenges

Arianna Filintisi, Dimitrios Vlachakis, George Matsopoulos and Sophia Kossida (2013). *International Journal of Systems Biology and Biomedical Technologies* (pp. 67-74).

[www.irma-international.org/article/3d-structural-bioinformatics-of-proteins-and-antibodies-state-of-the-art-perspectives-and-challenges/97743](http://www.irma-international.org/article/3d-structural-bioinformatics-of-proteins-and-antibodies-state-of-the-art-perspectives-and-challenges/97743)