



## Bioinformatics: An Overview of Biomedical and Computational Science Research

Dr. Jalari Ramu<sup>1\*</sup>

*\*Associate Professor of Biochemistry, School of Medicine, WSU Teaching and Referral Hospital, Ministry of Health,  
Ethiopia.*

Corresponding Author:

\*<sup>1</sup>Email ID: [ramujalari@gmail.com](mailto:ramujalari@gmail.com)

Mobile: +91-9492633090

---

**Article History:** Received: 14.01.2022      Revised: 02.02.2022      Accepted: 19.03.2022

---

**Abstract** – Bioinformatics which is a unique subject integration with biological and computational combined research to utilize the date of biology by computational tools to produce effective outcome in biomedical engineering. It is an interdisciplinary science that develops and improves upon methods storing, retrieving, organizing, and analyzing biological data. Research implication of computational tools, which bring utilization of different kinds of scientific data like medicine and health, life sciences, computer science, chemistry, statistics, mathematics, including those acquire, store, organize, archive, analyze. Multiple approach on bioinformatics brings a miracle change in research and development of any branch. It's a big challenge for researchers to describe this infant field in a systematic scientific way and bring out the attention of its various applications and services. This article will present a broad-spectrum overview of bioinformatics, aims, applications, technologies, data produced in the biological world and how bioinformatics can organize, analyze, store, and discuss some of algorithms that can be implemented over bioinformatics data, and how to apply bioinformatics to discover, diagnosis of diseases and applications in biomedical by using computational tools.

**Keywords:** Algorithms, Applications, Bioinformatics, Biomedical, Data, Technologies.

---

**DOI:** 10.53555/ecb/2022.11.01.18

### 1.1 Introduction

Bioinformatics is young multidisciplinary field that comes out from the combination of science fields like biology, computer science, statistics, chemistry, mathematics and more [3, 6, 8, 9, 14, 15, 16, 17]. In recent years new sciences have risen up due to the demand in understanding of the world around us like Bioinformatics, Computational Biology, and Biochemistry etc. It's a big challenge for researchers and scientists to make an adequate definition for each of these newly emerged sciences [5, 9, 18]. Bioinformatics huge influence in medical field, and it can play a key role in other fields like agriculture, veterinary and even space explorations [1, 19]. Bioinformatics attracts scholars in academic field in addition an interest to those in the medical field [4, 15, 20, 21]. There were many researchers to define, but most of them agree that it is a

combination of Biology, Computer Science, Statistics and Mathematics. Each one of these disciplines is playing an important role for collecting, organizing, analyzing and digitalizing the biological data, classifying and storing it efficiently [1, 3, 12, 16, 19].

The main purpose of this paper is explored and explain Bioinformatics scientifically. Define Bioinformatics and make highlight applications of bioinformatics in the medical sector specially, and in the diagnosis of critical diseases like cancer. The race of bioinformatics research is now passing long rounds in many areas in the Biological life, so; the goal of this paper is to provide an overview summary of bioinformatics definition from different articles written in this field, what are the main implementations and aims under the skin of this science, how to understand the

data and what are the most important databases used, give a snapshot over the most common algorithms implemented in the field and how important to apply bioinformatics in the cancer research and study.

This paper will target four categories of readership who are interested in the field. (1) Students who are interested in studying this new field. (2) Instructors who would like to prepare a fundamental course to teach in bioinformatics. (3) Researchers who would like to understand more about Bioinformatics and the relationship with cancer. (4) Experts in the medical field who are interested in implementing the understanding of this field in the medical life.

The remainder of this paper will be structured as follows: Section 2 will discuss the background in methodologies applied in this paper; while Section 3 will focus on Bioinformatics definition, on the other hand section 4 will figure out the aim of studying the field. Moreover, in section 5 data, data types and databases will be presented in Bioinformatics. On the other hand, section 6 will discuss the most common Algorithms implemented in Bioinformatics. Section 7 will discuss the role of Bioinformatics in cancer research and how important to be implemented in that field. In section 8 current problems in Bioinformatics are represented, and finally section 9 will conclude this paper.

## **2. Background in methodologies**

All over the world scholars and researchers talking about Bioinformatics but all have no unified definition about Bioinformatics as a newly growing branch in science and technology, add to that there were very rare research articles methodically build a proper road for basic knowledge in Bioinformatics. From this point an effort was implemented to conduct a deep search to collect as many papers and articles discussing the historical and fundamentals of Bioinformatics in order to establish a unified basis form understanding the basics of Bioinformatics and links that with importance of applying the field in the cancer study, research and therapy. A number of books, articles, websites which are discussing about introduction of bioinformatics were collected.

A huge number of references were introducing for definition to Bioinformatics, 10 articles talked about the databases in bioinformatics, 6 discussed for algorithms used in Bioinformatics, 12 mentioned for technologies and tools, the same number discussed applications of Bioinformatics. After that grouped out the data that are relevant together from the different resources and put them together for the literature review and the findings. It was noticed that the different resources collected were not focusing on a basic knowledge of Bioinformatics, they started by defining the field then highlighting one part of

the field like databases, tools, applications, algorithms, etc...

Our contribution in this paper is to gather all the distributed fundamental information about Bioinformatics and summarize them in a systematic fundamental way. Author [1] described as storage and analysis of biological data using specific algorithms and software tool is called Bioinformatics. The authors in [2] said that bioinformatics is basically a study to model, to organize, to understand and to discover interesting information associated with the large scale molecular biological databases. In [3] the authors claimed that the use of bioinformatics to organize, manage, and analyze genomic data which is the genetic material of an organism, this new IT discipline fuses computing, mathematics, and biology to meet the many computational challenges in modern molecular biology and medical research. Chavan [4] said that biological information data regarding genomic sequences of various species changes in their protein sequences due to evolution. Such a massive data cannot be handled with ease. This requires systematic sieving of data to categorize and catalogue them. Based on this need arose the field of Bioinformatics. So, Bioinformatics can be defined as the discipline, which encompasses branches like biology, computer science, IT and mathematics. Reference in [5], said that defining the terms of bioinformatics and computational biology is not an easy task and which are broad fields involving researchers from different branches such as computer science, statistics, physics, biochemistry, and molecular biology. In [6] Zadeh defines bioinformatics as a new discipline that has emerged from the areas of biology, biochemistry, and computer science. Bioinformatics is multidisciplinary subject emerged from biochemistry and computer science. Reference [7] discussed about that the bioinformatics concerned the application and development of information sciences for analysis, modeling study of biological processes in living organism. Authors in [8] illustrate that the Bioinformatics which is a subject combination of biology and information technology focused on application of modern biotechnology. So as a result, they said that Bioinformatics is the combination of biology and information technology. It is the branch of science that deals with computer-based analysis of large biological data sets. Author in [9] talked about definition of Bioinformatics combining of many fields which are computational biology, mathematics, molecular biology, and biochemistry. So, Bioinformatics is conceptualizing biology in terms of macromolecules and then applying “informatics” tools to understand and organize the information of these molecules. Moreover, Nair in [10], explained Bioinformatics to be the application of computer sciences and allied technologies to answer the questions of Biologists, about the mysteries of life. In addition, the authors in [11] discussed that bioinformatics

is a new and rapidly evolving discipline that has emerged from the fields of experimental molecular biology and biochemistry, and from the artificial intelligence (AI), database, pattern recognition, and algorithms disciplines of computer science. Finally, in [12] the authors summarized the definition of bioinformatics as the application of computer technology to the management of biological information.

### **3. Bioinformatics Definition**

The roots of bioinformatics can be linked to Mendel's pioneering work on genetic heredity, which he unveiled in 1865.

Starting in 1953, a monumental revolution in the field of genetics unfolded, catalyzed by James Watson and Francis Crick's momentous discovery of the DNA structure [13].

In the 1960s, the landscape of bioinformatics research commenced its arduous journey, exemplified by Dayhoff's comprehensive compendium of protein sequences and the initial endeavors to model and analyze protein and RNA structures [12]. Subsequently, during the ensuing years, the concept of Bioinformatics gradually took shape and gained prominence, particularly in the 1990s, encapsulating the administration and scrutiny of DNA, RNA, and protein sequence information.

Subsequently, in the year 2000, a significant milestone was attained with the unveiling of the preliminary version of the Human Genome Sequence.

After an extensive span of dedicated research spanning from 1990 to the spring of 2003, a pivotal milestone was reached with the official proclamation of the Human Genome Sequence Project. This monumental undertaking yielded the identification of approximately 20,000 to 25,000 human genes. The acquisition and interpretation of this vast reservoir of gene-related data posed a formidable challenge, demanding intricate strategies for data access and utilization for the biologists and for this it opened the doors for a new era in modern biology with an assistant to new computerized technology or in other words the marriage between Biology and Computer Science to bear a new baby known as Bioinformatics which will play a significant role in gathering, analyzing, classifying and storing genetic data collected from the human project or at biological points in a more efficient or powerful way. From here raised the question, what is the importance of Computers in Biology? The accurate answer of this question will be resulted out from the following formula:  $\text{Biology} + \text{Computer Science} = \text{Bioinformatics}$ . So, what is Bioinformatics? What are the main problems that this field can help in?

The review of existing literature uncovers various viewpoints for defining Bioinformatics. Initially, the English Oxford Dictionary provides insight, followed by a compilation of researchers' definitions, collectively offering a well-rounded understanding of this multidimensional field.

**Bioinformatics:** (As per the Oxford English Dictionary) (Molecular) bio-informatics refers to the interpretation of biology through molecules, akin to principles of Physical

chemistry. It employs methods from disciplines like applied mathematics, computer science, and statistics—collectively termed "informatics techniques"—to comprehensively grasp and structure data linked with these molecules. Succinctly put, bioinformatics serves as an information system to oversee molecular biology, boasting numerous real-world applications. Bioinformatics emerges as an innovative hybrid science, converging biology, computer science, mathematics, statistics, and Information Technology. This discipline synergistically manages, analyzes, and organizes extensive biological data using refined algorithms, computational methods, and statistical techniques. It encompasses software tool design, construction, and theoretical frameworks to solve diverse biological challenges, facilitating data generation, storage, access, and analysis within molecular biology. Notably, the "informatics" suffix originates from "informatique," denoting computer science in French, while "Bio" signifies Biology. Refer to Figure 1 for an illustration of the constituent sciences shaping the Bioinformatics field.

The review of existing literature uncovers various viewpoints for defining Bioinformatics. Initially, the English Oxford Dictionary provides insight, followed by a compilation of researchers' definitions, collectively offering a well-rounded understanding of this multidimensional field.

**Bioinformatics:** (As per the Oxford English Dictionary) (Molecular) bio-informatics refers to the interpretation of biology through molecules, akin to principles of Physical chemistry. It employs methods from disciplines like applied mathematics, computer science, and statistics—collectively termed "informatics techniques"—to comprehensively grasp and structure data linked with these molecules. Succinctly put, bioinformatics serves as an information system to oversee molecular biology, boasting numerous real-world applications. Bioinformatics emerges as an innovative hybrid science, converging biology, computer science, mathematics, statistics, and Information Technology. This discipline synergistically manages, analyzes, and organizes extensive biological data using refined algorithms, computational methods, and statistical techniques. It encompasses software tool design, construction, and theoretical frameworks to solve diverse biological challenges, facilitating data generation, storage, access, and analysis within molecular biology. Notably, the "informatics" suffix originates from "informatique," denoting computer science in French, while "Bio" signifies Biology. Refer to Figure 1 for an illustration of the constituent sciences shaping the Bioinformatics field.

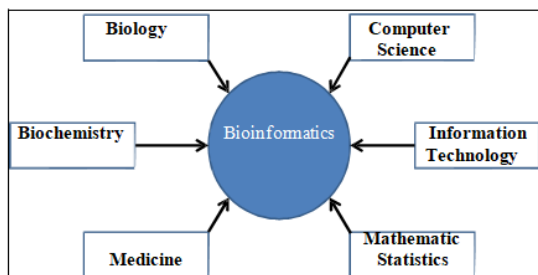


Figure 1: Bioinformatics multidisciplinary sciences

Bioinformatics has four main components: Databases, Computational Tools, Algorithms and Software. Biologists and other related people must be aware of the difference between Bioinformatics and Computational Biology and this is not an easy task, the latter is not a “field” like bioinformatics but it is “an approach” involved in using computers to study biology [9]. So, bioinformatics is concerned with information while computation biology is concerned with hypothesis [14].

#### Bioinformatics Aims

There are five main aims of Bioinformatics [12]:

1. To organize the biological data in an easy manner that helps biologists and researchers to store and access exiting information.
2. To develop and design software tools that help in the analysis and management of data.
3. To use these biological data in the analysis and interpretation of the results in a biological meaningful manner.
4. To assist researchers in the pharmaceutical industry to understand the protein structures that lead and help in the drugs industry development.
4. To Facilitating pharmaceutical researchers' comprehension of protein structures crucial for drug industry advancement.
5. To help and assist physicians in the medical fields to understand gene structures that will help in detecting and diagnosing disease like cancer.

#### 5. Biological Data, Data types and Databases

Biological data commonly exhibits vastness, with four vital data categories emerging at biological junctures [10]: DNA, RNA, Protein Sequences, and Microarray images, where the initial 3 comprise text data, while the last one adopts a digital image format. As the different biological data generated, it can be noticed that these data are represented with different types. (With the proliferation of diverse biological data, it becomes evident that this information assumes various forms and representations.)

There are four types of the data structures [13]: String to represent DNA, RNA, and protein sequences; Trees to represent protein structures; Graphs to represent metabolic and signaling pathways; and Strings (like words and phrases) are also used to express comments that reflect

meanings to researchers. Moreover, researchers and biologists are also interested in substrings, subtrees and subgraphs.

The management of extensive, intricate biological datasets necessitates efficient storage, accessibility, and manipulation methods. To address this, Bioinformatics databases emerged, encompassing categories such as sequence, microarray, genome, and protein structure databases [2].

The sequence databases represent sequence information of all the organisms. Prominent sequence databases, including GeneBank at the National Center for Biotechnology Information, EMBL DNA database at the European Molecular Biology Laboratory, DNA Data Bank Japan (DDBJ), and SWISS-PROT protein sequence database at the Swiss Institute of Bioinformatics, Geneva, stand as significant repositories. In the realm of microarray data, databases like ArrayExpress and Gene Expression Omnibus (GEO) cater to gene expression profiles across diverse biological conditions. Genome databases collect organisms' gene (DNA) sequences. Exemplary databases in this category encompass Xenbase, Corn, SEED, and RGD. There is another example of Bioinformatics databases that comes from the integration with cheminformatics which is the Drug Bank database (<http://redpoll.pharmacy.ualberta.ca/drugbank>), this database contains 4300 drug entries for and more than 6000 protein sequences which are linked to these drug entries [1].

#### 6. Common Bioinformatics Algorithms [12-13]

This section sheds the light on algorithms that are of interest to bioinformaticians and researchers. The following are some of the most important algorithmic trends in bioinformatics: Here are some of the key algorithmic trends in bioinformatics

1. Identifying similarities among strings (e.g., proteins from different organisms).
2. Recognizing specific patterns within strings (e.g., genes).
3. Identifying similarities among components of spatial structures (e.g., motifs).
4. Building trees (referred to as phylogenetic trees, depicting the evolution of organisms with known DNA or proteins).
5. Categorizing new data based on previously grouped sets of annotated data.
6. Analyzing microarray data and the associated behavior of pathways.

#### 7. Bioinformatics Applications in Cancer Research

Cancer is characterized as a genetic disorder where cells are unable to undergo the sequential phases of the

cell cycle and divide in a typical manner. That is cells will lose the control in the cell cycle and starts to divide uncontrollably and the chromosomes of the cancer cells will be arranged incorrectly, or have large pieces missing. Due to large and fast steps in the medical field research, a lot of efforts are extended in order to find a way to detect, diagnose and treat such hazardous disease. Furthermore, the unveiling of the Human Genome Project's finding in 2003 increased the demand for the application of Bioinformatics in cancer therapy. Bioinformatics is now being applied in the cancer research and therapy [21], and it is clear that experts and researchers have implemented rapid and expanded amount of research on the bioinformatics tools deemed essential in cancer therapies. One of these applications is to use the computerized models that represent biological data and information to know about the quantity of cancer cells in the body or about the biological state of the patient [22]. Such way has a positive result after the cancer therapy in which experts are now being able to monitor the tumor growth that was not possible earlier during the absence of bioinformatics. In addition, many studies have indicated that gene expression of cancer cells is imperative and this will ensure efficient results after the treatment [9, 23]. Also, bioinformatics can be applied to cancer by using the database among the cancer cells' expression and to study the drug response and tumor response also [23]. Until now bioinformatics studies show that it had succeeded in the cases of breast and ovarian cancer and future will insure the effectiveness of bioinformatics in the therapies of other cancer types [24]. Moreover, bioinformatics has made it possible for therapists to analyze immune responses that allow an understanding of the differences between controlled and uncontrolled tumors for better treatment of cancer patients. In other words, bioinformatics succeeded in explaining out the effects of the chemotherapy and the radiation therapy with the help of the mathematical models that are part of the bioinformatics discipline. It has been observed that experts and physicians utilize various databases and search engines, such as Google, to access biological data and apply bioinformatics in cancer research and treatment. However, certain organizations and experts restrict access to their work and information, limiting the benefits for other professionals. Essentially, the integration of bioinformatics databases, data types, and structures plays a crucial role in determining the future application of bioinformatics in the medical field, particularly in cancer treatment and therapies. The Human Genome Project, by providing an extensive amount of biological data and information by 2003, has significantly enriched the research community [1]. As a result, bioinformatics has found applications in

various areas, including the analysis of DNA sequence data to locate genes and the analysis of RNA sequence data to predict their structures [4, 10].

Analyzing DNA sequence data to locate genes.

- Analyzing RNA sequence data to predict their structures.
- Analyzing protein sequence data to predict their location inside the cell.
- Analyzing gene expression images.
- Understanding genetic diseases like cancer, cystic fibrosis, and sickle cell anemia.
- For gene therapy in general.
- In designing drugs for better treatment, and avoid drugs side effects and develop better drug delivery system. Moreover, NASA's experts are using Bioinformatics in their operations to explore the space and study the universe. So, NASA is also interested in Bioinformatics in their researches and discoveries.

## **8. Conclusions**

The paper tried to give an overview of this multidisciplinary field, by forming a unique clear definition that is introduced by the reaction of Biology and Computer Science in addition to some assessment factors like statistics and mathematics to result into the newly born field "Bioinformatics" after this strong reaction. In conclusion, the paper underscored the significance of employing bioinformatics in cancer research, paving the way for experts and researchers to further advance in this specialized field. The future of Bioinformatics will be bright in many biological and life areas, but one of the important issues that must be worked in for this; is the integration of the wide and huge amount of data sources and databases to unify them for better life and for a huge revolution in the biological life as will reaching the moon.

## **9. References**

- [1] Jawdat, D.; "The Era of Bioinformatics," Information and Communication Technologies, 2006. ICTTA '06. 2nd, vol.1, no., pp.1860-1865, 0-0 0
- [2] Raut, S.A.; Sathe, S.R.; Raut, A.; "Bioinformatics: Trends in gene expression analysis," Bioinformatics and Biomedical Technology (ICBBT), 2010 International Conference on, vol., no., pp.97-100, 16-18 April 2010
- [3] See-Kiong Ng; Limsoon Wong; "Accomplishments and challenges in bioinformatics," IT Professional, vol.6, no.1, pp. 44- 50, Jan.-Feb. 2004
- [4] Dr. (Mrs.) Padma R. Chavan; "Application of Bioinformatics in the Field of Cancer Research", 11th Workshop on Medical Informatics & CME on

Biomedical Communication, vol., no., 20-22 November 2008.

- [5] Ackovska, N.; Madevska-Bogdanova, A.; "Teaching Bioinformatics to Computer Science Students," Computer as a Tool, 2005. EUROCON 2005. The International Conference on Computers as a Tool, vol.1, no., pp.811-814, 21-24 Nov. 2005
- [6] Zadeh, J.; "An undergraduate program in bioinformatics," Potentials, IEEE, vol.25, no.3, pp.43-46, July-Aug. 2006
- [7] Kasabov, N.; "Bioinformatics: a knowledge engineering approach," Intelligent Systems, 2004 Proceedings. 2004 2nd International IEEE Conference, vol.1, no., pp. 19- 24 Vol.1, 22-24 June 2004
- [8] Fulekar, M.H. and J. Sharma. 2008. "Bioinformatics Applied in Bioremediation". Innovative Romanian Food Biotechnology. Vol. 2 No. 2. pp 28-36.
- [9] David Fenstermacher, Introduction to bioinformatics: Research Articles, Journal of the American Society for Information Science and Technology, v.56 n.5, p.440-446, March 2005
- [10] Achuthsankar S Nair; "Computational Biology & Bioinformatics – A gentle Overview", Communications of Computer Society of India, January 2007.
- [11] Doom, T.; Raymer, M.; Krane, D.; Garcia, O.; "Crossing the interdisciplinary barrier: a baccalaureate computer science option in bioinformatics," Education, IEEE Transactions on, vol.46, no.3, pp. 387- 393, Aug. 2003
- [12] Jana, R., Aqel, M., Srivastava, P., and Mahanti, P. K., Soft Computing Methodologies in Bioinformatics, European Journal of Scientific Research, Vol. 26, No. 2, pp. 189-203, 2009.
- [13] Jacques Cohen, Computer science and bioinformatics, Communications of the ACM, v.48 n.3, p.72-78, March 2005
- [14] DOMOKOS, A. BIOINFORMATICS AND COMPUTATIONAL BIOLOGY. Bulletin of University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca. Horticulture, North America, 6527, p. 571 – 574, 09 2008.
- [15] Poe, D.; Venkatraman, N.; Hansen, C.; Singh, G.; "Component-Based Approach for Educating Students in Bioinformatics," Education, IEEE Transactions on, vol.52, no.1, pp.1-9, Feb. 2009
- [16] Bayat A. Science, medicine, and the future: Bioinformatics. BMJ. 2002; 324:1018–1022.
- [17] National Center for Biotechnology, "Bioinformatics Factsheet," <http://www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html>, last accessed June 13, 2012.
- [18] Gavin J. Gordon, Bioinformatics in Cancer and Cancer Therapy (Cancer Drug Discovery and Development) [Kindle Edition], ISBN: 978-1-58829-753-

2 e-ISBN: 978-1-59745- 576-3, Library of Congress Control Number: 2008931368

- [19] Jacques Cohen, Computer science and bioinformatics, Communications of the ACM, v.48 n.3, p.72-78, March 2005
- [20] Umarji, M.; Seaman, C.; Koru, A.G.; Hongfang Liu; "Software Engineering Education for Bioinformatics," Software Engineering Education and Training, 2009. CSEET '09. 22nd Conference on, vol., no., pp.216-223, 17-20 Feb. 2009
- [21] Simon R. Bioinformatics in cancer therapeutics-hype or hope? Nat Clin Pract Oncol. 2005; 2:223
- [22] Goldin, L.; "Bioinformatics Integration for Cancer Research-Goal Question analysis," Information Technology: Research and Education, 2006. ITRE '06. International Conference on, vol., no., pp.248-252, 16-19 Oct. 2006
- [23] Kihara D, Yang YD, Hawkins T. Bioinformatics resources for cancer research with an emphasis on gene function and structure prediction tools. Cancer Inform. 2007; 2:25-35.
- [24] Ardekani AM, Aslani F, Lakpour N. Application of genomics and proteomics technologies to early diagnosis of reproductive organ cancers. J Reprod Infertil. 2007; 8(3):259-278.