

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/332762073>

# Bioinformatics for Healthcare Applications

Conference Paper · February 2019

DOI: 10.1109/AICAI.2019.8701277

---

CITATIONS

2

READS

5,840

3 authors, including:



Vinayak Majhi

North Eastern Hill University

12 PUBLICATIONS 5 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Development of cost effective web and mobile application for detection of Parkinson's Disease [View project](#)

# Bioinformatics for Healthcare Applications

Vinayak Majhi<sup>1</sup>, Sudip Paul<sup>2</sup>, Rachna Jain<sup>3</sup>

<sup>1,2</sup>Department of Biomedical Engineering, North-Eastern Hill University, Shillong, India

<sup>2</sup>School of Computer Sc. and Software Engg The University of Western Australia, Perth, Australia

<sup>3</sup>Amity Institute of Information Technology, Amity University, Uttar Pradesh, India

<sup>1</sup>majhi.vinayak@gmail.com, <sup>2</sup>sudip.paul.bhu@gmail.com, <sup>3</sup>rjain1@amity.edu

**Abstract:** *Bioinformatics is the combination of health information, data and knowledge. It uses computational techniques and tools to analyze the enormous biological data bases. The diseases such as metabolic disorders, urea cycle disorders, inborn errors and path-aligner can be identified at the early stage using various bioinformatics computational tools. These tools are used to process genetics and proteomics data and compare with health care data. Health care data consists of physiological data from various muscles or organs co-acquired by using various electronic components, cost reports, bill claims and surveys related to patient satisfaction. Health care analysts are helpful to provide best health care at the lower price which is good for family economic system. This paper briefly explains the history applications of bioinformatics. The brief information about various kinds of computational tools and databases with respect to the application are mentioned. The current research trends in the bioinformatics are concluded.*

**Keywords:** *Bioinformatics, Healthcare Engineering, Healthcare Informatics*

## I. INTRODUCTION

The interest for health care is not normal for the interest for most consumer items and administrations on the grounds that while the longing for consumer items and administrations originates from direct utilization, the craving for health care isn't gotten straightforwardly from the utilization of the medicinal methodology themselves; rather, it originates from the immediate estimation of enhanced health that is delivered by health care. Individuals request health in view of its job in advance and giving satisfaction. Health can be characterized along two measurements: the range of longevity and the personal satisfaction [1]. A man gets an incentive from the personal satisfaction specifically and in a roundabout way: straightforwardly on the grounds that one's dimension of health influences the delight in merchandise and relaxation and by implication since one's dimension of health upgrades efficiency. In like manner, any innovation that is custom fitted towards enhancing the healthcare framework is well on the way to appreciate great open methodology [2].

Bioinformatics is an integrative field that creates procedure and software tools for demonstrating biological information and frameworks which are ordinarily described by a lot of

information. Bioinformatics is the mix of different fields like, software engineering, computer science, statistics, informatics and engineering which assess and outline biological and genomic information. The field of Bioinformatics covers the examination of molecular data to quantifying clinical, imaging and diagnostic data for customized medication and healthcare. Bioinformatics is used for the identification of candidate genes to better understand the genetic basis of disease, unique adaptations, and differences between populations. This additionally incorporates investigation of quality articulation, frameworks biology, and understanding the transformative history of life. In basic biology, it helps in the reenactment and displaying of DNA, RNA, 3D protein structures and in addition bio molecular interactions. With rapid advances in gene therapy, genetic engineering, gene editing and drug discovery [3].

## II. HISTORY OF BIOINFORMATICS

In early stage of the 1970s, Ben Hesper and Paulien Hogeweg initiated to use the term "bioinformatics" for the research they wanted to do, defining it as "the study of informatic processes in biotic systems" [4]. In 1981 Marvin Carruthers and Leory Hood found 579 human genes mapped through situ hybridization and invented an automated DNA sequencing process. The human Genome project and human genome organization was established in 1988 [2]. In the year 2000 Huang et al, utilized these concepts for experimenting in molecular biology. These theory raised their power to demonstrate alternative trajectories of neutrophil differentiation on the basis of temporal gene expression data [4]. Bioinformatics is a combination of biological signal analysis, interpretation, and management. In order to produce enormous databases related to biological sequences and structures they established European Molecular Biology Laboratory (EMBL) and GenBank. The DNA data available from big data bases is used to match the various DNA sequences. The current research in the bioinformatics is targeted to encompasses functional details of protein/gene, protein structural analysis, and comparison of metabolic rout for various species through pre-clinical/clinical trials [5].

## III. APPLICATION OF BIOINFORMATICS

The aim of the bioinformatics is to give path to big data generated by molecular biology technology. It will produce

DNA array in large scale and tools to build positions of protein coding areas [2].

The concept of multiple gene expression measurement is used to uncover the DNA chip consists thousands of nucleotide fragments in thousands. It is alone used to identify the gene coding DNA from the sample and locate the subsets of genes [2]. In genomics and proteomics unusual feats are more useful in data mining methods. The success rate in identification of genes related disorders is depending on the discovery of novel drugs and its effectiveness. Many international, national, autonomous and charities encouraging together to store the genetic information in form of electronic health records [6]. The computation and interpretation problems associated with the molecular level can be solved by using mathematics and statistics. The computer science and information technology is used to design suitable computational tools [7]. The various applications of the bioinformatics is as follows:

#### ***A. Sequence Analysis***

The genetic basis of organism is depending on all genes of its genome. Sequence analysis is a method used to understand its structure, function, features. There are many powerful tools are available in the computer science and each tool has its own advantages and disadvantages. These tools are used to identify the sequences which are related to DNA mutations of an organism. The sequence analysis of numerous fragments can be done by using shotgun sequence technique [8].

#### ***B. Prediction of Protein Structure***

The three dimensional structure of protein from its amino acid can be predicted. It is tough to determine the secondary, tertiary or quaternary structures of proteins. For this purpose either the method of crystallography or bioinformatics tools will be used [9].

#### ***C. Genome Annotation***

The regulatory sequence and protein coding can be achieved using genome annotation. It is the process to identify the locations of all genes, coding regions and structure of genome [10].

#### ***D. Comparative Genomics***

The Genomic features such as gene, cds, mrna, trna and rrna will be compared to determine the genomic structure and function relation between various biological species. The scientists use the inter genomic maps to trace the whole path of evolution that occur in genomes of different species. The information about the point mutations and large chromosomal segments can be determined by these maps [11].

#### ***E. Health and Drug discovery***

The bioinformatics tools are used in drug discovery based on molecular disease. Based on the disease and diagnosis

management the scientist will develop the medicines and drugs which can suits more than 500 genes. The delivery of specific drug is depending on the cells to be targeted and various computational tools used [12].

### **IV. TOOL OF BIOINFORMATICS**

The aim of the computational tools is to provide the information about portrayal of qualities, phylogenetic examinations, deciding auxiliary and physiochemical properties of proteins. This information is useful to think about the behavior of biomolecule in a living cell. The different bioinformatics tools used for various applications is as follows.

#### ***A. Gene Identification and Sequence Analyses***

The comprehension and various highlights of a biomolecule such as nucleic acid or protein can be achieved through grouping investigations. This process will have started through sequences recovery of related molecules from open databases. Depending on the requirement they exposed to various tools to empower its features and highlight the findings such as capacity, structure, and homologues with an incredible precision.

Basic local alignment search tool (BLAST) is a search tool used to compare the sequences of proteins, nucleotides, DNA and RNA. "HMMER" is another free software used for identification and analysis of homologous protein sequences from the individual databases. Clustal developed the several computational techniques especially for bioinformatics. Among the available tools "Omega" has various operating systems used for multiple sequence alignment. It will produce high accurate result compared with other algorithms such as matrix based and consistency. Bioinformatics and Computational Biosciences unit (BCBU) developed the tool "Sequerome" used for profiling arrangement. "ProtParam" is a computational tool used to compute the various physicochemical properties sored in a protein. Gene production using multiple sources of evidence (JIGSAW) is used to identify the jene models through sequence alignment. It will produce better accuracy compared with other identification algorithm such as Ensembl and UCSC'S Known Gene track. Single nucleotide polymorphisms (SNPs) are the tools used to identify the genetic based diseases. Open reading frame finder (ORFF) used for various bioinformatics analyses, graphical viewing and data management. To up-stream the quality of promoter sequences prokaryotic promoter prediction toll (PPPT) will be used. Virtual Footprint is a tool used to analyze the bacterial regulon. The intrinsic transcription terminators identified in bacterial genome sequences and plasmids will be stored in "WebGeSTer DB" data base. In order to identify the complete gene structure in genomic DNA "GENSCAN" program will be used. The tools developed by the "Soft berry" mainly used for alignment and Genome comparison and regulation analysis of genome including plants and animals [13].

## B. Phylogenetic Analyses

The genetic relatedness is depending on the transformative relationship among the group of same molecules, organisms, life forms. Phylogenetic analyses are used to forecast the highlights of a molecule with unclear capacities and quality stream track. Molecular evolutionary genetics analysis (MEGA) is a user friendly and sophisticated software used to analyze the protein sequence data from classes and inhabitants. It is developed based on phylogenetic trees to work on the transformative closeness. "MOLPHY" is a program package used to examine the sub-atomic phylogenetic based on the probability technique. Phylogenetic analysis by maximum likelihood (PAML) and phylogeny inference package (PHYLIP) utilize the probability technique for phylogenetic analysis and each package contains several computational tools. The open-source library's such as Bio.js tree, and Treeview are available contains phylogenetic trees for introduction enhancement. "Jail view is the software used to edit and analysis the multiple sequence alignments [13].

## C. Predicting Protein Structure and Function

The initial stage, the protein molecules are unstructured amino acid strings. It gains three-dimensional (3D) structure as biological dynamics at the final stage. The performance of the biological functionalities is based on collapsing of protein into a right topology which is pre-imperative. Based on the above it concludes that the 3D structure of protein is important for its feature extraction and these kind of structures can be viewed by X-Ray crystallography or NMR. The comparison of forecasting structures and validity of simulation algorithm is depending on thermodynamic equilibrium physio-chemical standards, with a global minimum free energy of protein surface and minimum free energy. The various tools used to identify the protein structure and function is shown in Table 1.

Table 1. various tools used to identify the protein structure and function [13]

Name of the tool	Applications
CATH	It is a self-loader tool to arrange the association of proteins.
RAPTORX	To forecast protein structure depending on either multi-layout threading or solitary.
JPRED and APSSP2	Used to predict the auxiliary structures of proteins.
PHD	To predict the neural network structure
HMMSTR	To forecast the arrangement structure relationships in proteins.
MODELLER	To predict the 3D structure of protein based on comparative modelling
PHYRE and PHYRE2	These are the web based servers used to predict the protein structure.

## D. Drug Designing

Before inventing the bioinformatics tools, the researchers belongs to various fields such as pharmacology, clinical sciences and chemistry worked together for introducing the new compound to the world. The invention of the bioinformatics stopes the above procedure and encouraged the revelation and planning in the medical field. It is easy to break down the molecules using the software compared with various practice approaches. The advances in the software design and information technology leads to discover the computer aided drug design (CADD) to design highly effective drugs. The popular drug databases developed by various organization to facilitate the researchers and scientists who are interested to design the new drug is as follows

- Potential Drug Target Database (PDTD)
- Therapeutic Target Database (TTD)
- (TDR) Target Database
- Manually Annotated Targets and Drugs Online Resource (MATADOR)
- Drug Bank
- Tropical Disease Research
- Tuberculosis (TB) Drug Target Database

The various molecular dynamic tools such as Abalone, Ascalaph, Discovery studio, Amber, FoldX, are used to get the information related to ion transport, conformational changes of proteins, nucleic acids and complexes occurring in biological systems [13].

## V. HEALTHCARE INFORMATICS AND BIOINFORMATICS

Health informatics is combination of information and data of medical areas such as pre-clinical, clinical, post clinical and health care administration management. The advances in the information technology gave the free hands to the researchers and scientists to improve the public health, healthcare system and biomedical field. The main aim of the healthcare informatics is to provide the complete information of the patient's health to healthcare professionals. This data makes their task easy to take the decision regarding the treatment at the right time. More over by using the health care informatics the patient form the rural area can get the opinion form best health care professionals among the available [14, 15].

Bioinformatics and analytics deals with analytic and interpretation, storage development and optimization of huge biomedical data. The revolution in the computational proficiency introduces the concept of electronic health records (EHRs) which enables the development of rich data warehouse. It can be used to identify the relationship between uncover genotype and phenotype. Various computational tools are used to study and analyse the large biological databases to understand the disease and such predictions are useful to relate with the health care data. The current researchers targeted in

this discipline to invent the new procedures and algorithms for processing genomic and proteomic data used in various fields such as drug discovery, medicine and related areas [2, 16, 17].

Artificial intelligence (AI) is one of advanced technology aimed to replace the partially/fully human intelligence power. It was gained attention in research related to computational molecular biology and bioinformatics. The performance and accuracy of AI algorithms are good for DNA sequencing compared with the other computational algorithms. Shi et al explained the importance of artificial intelligence in gene selection tools to classify the microarray cancer data [18]. Serra et al gave the comprehensive review on machine learning algorithms use in field of bioinformatics for dimensionality reduction of complex data, feature selection for assigning biomarker in raw data. Single nucleotide polymorphism analysis and classification using machine learning. Drug repositioning, Classification of patient from neuroimaging data are the different application of AI in bioinformatics [19].

## VI. CONCLUSION

The development in genetic research is leading us towards customized medicine, which utilizes a patient's genetic makeup to decide the correct drug for the patient. As the field of systems biology advances with customized medicine, scientists will have the capacity to apply tools and innovation to comprehend the ailment system that advances from the molecular, cellular, tissue, and organ levels to the individual and, at long last, the populace levels. Bioinformatics discoveries can be converted into advancements that are received by the healthcare framework and biomedical industry in type of diagnostic kits, analysis programs, and so forth. With the help of AI now we can classify different types of bioinformatics data for disease identification and also helps to find the Genetical cause of a particular disease.

## REFERENCES

- [1] Executive Office of the President and Council of Economic Advisers, "Economic Report of the President," February, 2008 2008.
- [2] O. Oyelade, J. Soyemi, I. Isewon, and O. O. Obembe, "Bioinformatics, healthcare informatics and analytics: an imperative for improved healthcare system," *International Journal of Applied Information Systems*, vol. 8, pp. 1-6, 2015.
- [3] T. Braun, T. Casavant, D. Kristensen, and M. Schnieders. (2018). *Bioinformatics Track | Biomedical Engineering*. Available: <https://bme.engineering.uiowa.edu/undergraduate-program/bme-tracks/bioinformatics-trackfiles/77/bioinformatics-track.html>
- [4] P. Hogeweg, "The roots of bioinformatics in theoretical biology," *PLoS computational biology*, vol. 7, p. e1002021, 2011.
- [5] U. R. Müller and D. V. Nicolau, *Microarray technology and its applications*: Springer, 2005.
- [6] D. C. Swinney and S. Xia, "The discovery of medicines for rare diseases," *Future medicinal chemistry*, vol. 6, pp. 987-1002, 2014.
- [7] P. S. Bagga, "Development of an undergraduate bioinformatics degree program at a liberal arts college," *The Yale journal of biology and medicine*, vol. 85, pp. 309-321, 2012.
- [8] E. V. Koonin and M. Galperin, *Sequence—evolution—function: computational approaches in comparative genomics*: Springer Science & Business Media, 2013.
- [9] A. Breda, N. F. Valadares, O. N. d. Souza, and R. C. Garratt, *Protein Structure, Modelling and Applications*: National Center for Biotechnology Information (US), 2007.
- [10] M. Kellis, B. Wold, M. P. Snyder, B. E. Bernstein, A. Kundaje, G. K. Marinov, et al., "Defining functional DNA elements in the human genome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, pp. 6131-6138, 2014.
- [11] M. Pellegrini, E. M. Marcotte, M. J. Thompson, D. Eisenberg, and T. O. Yeates, "Assigning protein functions by comparative genome analysis: protein phylogenetic profiles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, pp. 4285-4288, 1999.
- [12] S. K. Gill, A. F. Christopher, V. Gupta, and P. Bansal, "Emerging role of bioinformatics tools and software in evolution of clinical research," *Perspectives in clinical research*, vol. 7, pp. 115-122, Jul-Sep 2016.
- [13] M. A. Mehmood, U. Sehar, and N. Ahmad, "Use of bioinformatics tools in different spheres of life sciences," *Journal of Data Mining in Genomics & Proteomics*, vol. 5, p. 1, 2014.
- [14] H. Nadri, B. Rahimi, T. Timpka, and S. Sedghi, "The Top 100 Articles in the Medical Informatics: a Bibliometric Analysis," *Journal of Medical Systems*, vol. 41, p. 150, 2017/08/19 2017.
- [15] L. Ohno-Machado, "Data science and informatics: when it comes to biomedical data, is there a real distinction?," *Journal of the American Medical Informatics Association : JAMIA*, vol. 20, pp. 1009-1009, 2013.
- [16] R. Chen, "On bioinformatic resources," *Genomics, proteomics & bioinformatics*, vol. 13, pp. 1-3, 2015.
- [17] X. Xia, "Bioinformatics and Drug Discovery," *Current topics in medicinal chemistry*, vol. 17, pp. 1709-1726, 2017.
- [18] T. W. Shi, W. S. Kah, M. S. Mohamad, K. Moorthy, S. Deris, M. F. Sjaugi, et al., "A review of gene selection tools in classifying cancer microarray data," *Current Bioinformatics*, vol. 12, pp. 202-212, 2017.
- [19] A. Serra, P. Galdi, and R. Tagliaferri, "Machine learning for bioinformatics and neuroimaging," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, p. e1248, 2018.