Bioinformatics As An Emerging Field Of Data Science

Reena Hooda

Abstract: Bioinformatics is an interdisciplinary approach that originated from the interactions of different streams under one umbrella incorporating mathematics, computer science, statistics, geography, physics, biology and machine learning etc. Observing body structure & building blocks of the different species say human, plants, animals, body cells development & reactions as well as communications of different parts of body offered gargantuan approach to this emerging research area. Besides this, the study of environmental change affects, the resistance power, DNA Sequencing, proteomics, genomes, reproduction, every living thing of ecosystem & their growth stages, bacterial infection etc. under bioinformatics opens up a new domain of data science in biological research. An analysis to gain more information about living organisms, their survival in certain circumstances or devastations using advanced computational tools further accelerate the research. With range of sub-topics and huge variety of biological data, it is quite difficult to handle things manually. Thanks to development & progress of computer science & information technology that aid in fast and accurate analysis of data moreover employ this data to learn, modify, act and project more accurate outcomes. Present paper highlights the innovation of bioinformatics, its applications and advantages and tries to answer the questions that why bioinformatics is going to be most interesting area of research.

Keywords: Bioinformatics, Data Mining, Proteins, Genomes, Cell.

————————————◆————————————

1 INTRODUCTION

BIOINFORMATICS is an interdisciplinary approach that involves statistics, mathematics, computer science & softwares, biological terms like genomics and proteomics and specially the large databases of biological data. Genomic termed as complete set of DNA sequences that provide information about hereditary. Other than the genomics and proteomics, bioinformatics involve the metabolomics, transcriptomics i.e. study of transcriptome that is set of all RNA molecules of a cell [12], environment biomolecules etc. [1] Bioinformatics contain the whole ecosystem including all physical environment and all organisms or living beings and their study. The beginning of bioinformatics was recorded in 1960 by Margaret Dayhoff in studying the protein sequences and their structure. The first computer programme known as FASTP which was used to study the cancer causing viral sequence V-SIS [1]. Another research paper suggested that the term Bioinformatics firstly introduced by Paulien Hogeweg and Ben Hesper in 1970s. From a computer science point of you bioinformatics basically involvement of biology and computer science in biological data analytics, the techniques of data mining, software & tools, all considered as bioinformatics mainly for DNA Sequencing. Bioinformatics has emerged as a new area of research to answer many biological questions by the application of computer science and advance mining tools on biological data sets and turned as a new era of data science [1]. Ecosystem involve all sources and all kinds of data, like GIS, weather agriculture and air condition etc. [1]

2 BIOINFORMATICS AS AN EMERGING FIELD OF DATA SCIENCE

Bioinformatics is reduction science [2] [3] that means building descriptions and relationships of the system and it is not easy to obtain the information from them, for example the basic information like molecules composed of atoms, their structures relationship between the atoms and relationship

————————————————

• *Dr. Reena Hooda, Assistant Professor. Indira Gandhi University Meerpur (Rewari). Haryana (India), PH-08968065891. E-mail: reenah2013@gmail.com*

between the molecules are quite difficult to describe as well as obtain. Thus Bioinformatics is reductionist science that comprises all biological terms like genomics proteomics, epigenetics and other all Sciences. Study of person's genomes, interaction with the environment and the complete set of DNA are called genome. Proteomics deal with the building blocks that made body structure like organs and tissue and control all the chemical reactions and messages between cells. Genes is a unit of DNA that carries instructions and direct production of the protein in human cell. Reactions and factors influencing them are studied under epigenetics. The chemical reactions and factor influencing them are studies under the epigenetics, for example- stress and diet etc. As response to the changes a cell in any organism regulate its activity by changing the level of proteins. [3] This information about every cell, every protein, every control signal, every gene [3] for all type of living organism including humans, plants and animals make the flow of the data very huge dynamic as well as heterogeneous. Due to the development of advanced computing techniques, open sources comprising of open data and open software accelerate the popularity of bioinformatics. Bioinformatics is different from Molecular Biology, Molecular Biology that is based upon physical sciences like Physics and Chemistry whereas bioinformatics is placed with information technology. [3] It is new field of data sciences that requires the computational knowledge and biology. Medical images like DNA structure image, X-Ray image requires lots of processing to get the information about living organism. As a computer science or as a biological science, bioinformatics is the biological study through the computers that take and generate large amount of data. It is a solution of biological questions through the study of DNA, Amino acid sequence, protein etc. Biological molecules called Polymers are the chains of molecular modules called monomers which have different colors but have same thickness. They connected to each other in a similar fashion it looks same, however actually each monomers has its own set of characteristics makes them unique. Monomer can be considered as alphabet letter which makes the messages through their different arrangement and send to cell. Bioinformatics is includes the study of biological structures of genomes in different species and evolution. It is also involves the study of genetic

messages at different stages of disease, clinical side effect of the drugs etc. It is leveraging the laboratory Biology to cross species. Computational study of molecules, organization of biological information and comparison between dataset acquired name Bioinformatics. Biological data produced daily is turn into explosion of data in petabytes that required high performance computing to assemble varieties of genomic datasets and their comparisons. Bioinformatics in overall based on Molecular Biology in terms of Physics and Chemistry to understand the structure of molecules, their arrangements, message passing and reaction to change. The interesting part in bioinformatics is study of complete structure of genomes to find coding and non-coding region that are active or inactive for message passing. [7] It is not just study of biological system for a particular purpose or for a particular problem solving, it involves everything that can be processed and expressed to lead new discoveries by means of super computation powers. It involves the study of organisms that have the several copies of a gene of different species and got the similar property inherited in evolution having similar structure and sequencing. Bioinformatics opens new insights and world of innovations in biological components, study variety of bacteria & bacterial infections, their causes of resistance, and their survival producing variety of data inculcating vast knowledge. The core aim of bioinformatics is the analysis of existing biological structures to get the information, put this information back to the system to learn more and gain knowledge for the further processing. It is a continuous learning computational process that leads to the building of neural networks models and development of more sophisticated computational tools. [7] The field Bioinformatics requires a sound knowledge of computing + biological terms like DNA molecules, amino acids, proteins etc. Even a computer specialist may find it difficult to understand if he or she is not having the science background. The study of organisms in different environments, cells cycles, multiple data as data is more sequential than the structural, genomics, typical algorithms and repeated measurements, correlations between patterns, mappings make bioinformatics diverse, complex, at learning & training stage, in search of the meaningful and useful data. [1] [2] [7] [8]

3 APPLICATIONS OF BIOINFORMATICS

Bioinformatics is amalgamation of computer science physics, chemistry, biology, mathematics and information technology for data warehousing and mining the large biological data and DNA Sequencing. In data warehousing, data that is extracted, transformed and loaded is biological data for the purpose of achieving outputs or target values from multidimensional view. Following are common applications of bioinformatics [2]:

1. Bio-weapon development also known as germ weapon that involve the use of the biological disease producing or infectious agents like virus, fungi, rickettsia to harm humans, plants and animals. [8] It is more dangerous than nuclear or chemical weapons.
2. Forensic investigations used in crime investigations to find the clues to determine the cause of death or criminal detection.
3. Crops investigation & development and Insect resistance to study and develop crops that are capable to resist insect damage after genetic

engineering and no need of pesticides or chemical insecticides. [9]
4. Vetinary science to study effect of different medicines or chemical reactions on them, study of disease, injuries, causes & treatment of animals and effect of climatic changes on them.
5. Antibiotic resistance antibiotics are drugs or medicines taken to treat certain infections in humans or animals through killing or stopping further spread of bacteria causing infections, but these antibiotics has the side effects. In long run or taking too much antibiotics lead to antibiotic resistance in which bacteria cannot be killed or affected by the same drugs and it will multiply. [10] As per WHO it is the biggest threat to all living beings, food security, and effect to every age, resulted in poor health, long hospital stays and stop the positive effect of treatments. [11]
6. Viral infections like flu that cannot be treated by antibiotics. [10]
7. Climate change, health study & health improvement and effect of various climatic changes on living organisms those lead to become extinct.
8. Comparisons between sequencing & structures of cells in different species or different bodies
9. Recycling of wastage to protect ecosystem.
10. Nutrition and Proteins study that involve the analysis of optimal percentage requirement of nutrition in human body and everything that affects human or other living things, to improve their health conditions and their reproductions. Nutritious food may involve proteins, fats, vitamins, water, fiber carbohydrates which must be taken in right percentage to make body strong enough to fight against diseases.
11. Study of change in biological development of organisms

In short and common sense, Bioinformatics include all facts and figures, recall as play with nature through rigorous study of biological and environmental dynamics that generate large amount of data involving everything, every science, every technology and every sophisticated tool to help in this difficult and the crucial studies what are and what are not effect a living organism or components of an ecosystem.

4 ADVANTAGES OF BIOINFORMATICS

Employment of data mining techniques in knowledge Discovery make the faster analysis and accuracy of the research results, [2] even in predictions for finding unknown facts those may be important in new findings. Machine learning algorithms are created on the basis of change in inputs or target values or fluctuations in data, due to environmental change and their reactions, machine learning algorithms reproduce and adjust their weights of input variables, acting more accurately. Accuracy and copying the living beings is the main purpose of the machine learning to develop neural models which are simulating and act in the same way as a living react in a particular circumstances or circumferences. Advantages of bioinformatics [2] are given below.

1. More accuracy and faster analysis in getting results.

2. Study of historical data help in better learning and time series analysis, predictions and machine learning.
3. Selection of optimized methods for faster detections of outputs.
4. Study of heterogeneous, complex and large amount of data that is dynamic.
5. Less cost & time in preparation of the results and various reports helps in accurate selection of right cause of disease, taking right percentage of dose at right time and required level of changes in surroundings to recover faster from a disease.
6. Unbiased analysis and forecasting with maximum output with User-friendly interface.
7. Selection of outliers or microbes and tissues which are not affected by any climatic change, drug or bio weapon.
8. Findings of attributes which are missing or can be ignored in analysis while maintaining the accuracy of results.
9. Accuracy of Sampling is must for health information to improve the patient's health conditions. All tools like storage of the blood cells, their study & analysis, medical knowledge and computer knowledge is must as a background. Further inclusion of all main streams of the sciences, Humanities and Social Sciences for example: oncologist, pathologist, data automation, data specialist or expert makes it a complete domain of knowledge and skills. All physics or chemistry labs tools, instruments used in their labs and lab technicians, dietitian, nutritionist physicians, physiotherapist etc. have their own role in the bioinformatics. For these reasons, Bioinformatics become a separate field of study and research.

5 CONCLUSIONS AND FUTURE SCOPE OF WORK

Data science is not limited to study the current data to better answer the questions; it involves the investigations of new data dispersed at independent isolated locations, discovering unknown questions and attempts to explore the hidden data sources. Learning few basic biological concepts like cell, genomes, nutrition, DNA and proteins etc., is not a single step process rather it opens new insights of research that leads gradually in enhanced knowledge about this emerging field of data science. It is an interactive process of learning & knowledge acquisition through computer applications and information technology. It is more fruitful to get useful information, understand the lifecycle of living organisms, environmental impact on their health, drug selection, and right percentage of right drugs to avoid antibiotic resistance, minimize bio-weapons as well as making possible developments of healthy genes. Thus the bioinformatics and its sub-modules offered a wider scope of analysis and developments.

REFERENCES

[1] Joannefox (2006). "What Is Bioinformatics?". The Science Creative Quarterly. Retrieved from: Https://Www.Scq.Ubc.Ca/What-Is-Bioinformatics/
[2] A.L. Lapidus (n.d.). "Bioinformatics and its applications". Retrieved from: http://bioinformaticsinstitute.ru/sites/default/files/lapidus_1_0.pdf
[3] M. Ragan (2013). "Explainer: what is bioinformatics?". Retrieved from: https://theconversation.com/explainer-what-is-bioinformatics-9911
[4] https://www.bioinformatics.org/wiki/Bioinformatics
[5] D. Gusfield (2019). "What is Bioinformatics?". Retrieved from: https://web.cs.ucdavis.edu/~gusfield/extramaster.old/node1.html
[6] P. Bickerton [2018]. "What is bioinformatics?". Retrieved from: http://www.earlham.ac.uk/articles/what-is-bioinformatics
[7] N.M. Luscombe, D. Greenbaum, & M. Gerstein (2001). "What is bioinformatics? An introduction and overview". Retrieved from: https://www.researchgate.net/publication/2330725_What_is_bioinformatics_An_introduction_and_overview
[8] B. R. Schneider (n.d.). "Bioogical Weapon". Retrievd from: https://www.britannica.com/technology/biological-weapon
[9] http://sbc.ucdavis.edu/Biotech_for_Sustain_pages/Insect_resistance/
[10] Espanol (n.d). "Antibiotic Resistance Questions and Answers". Retrieved from: https://www.cdc.gov/antibiotic-use/community/about/antibiotic-resistance-faqs.html
[11] WHO (2018). "Antibiotic resistance". Retrieved from: https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance.
[12] Wikipedia, the free encyclopedia (2019). "Transcriptomics technologies". Retrieved from: https://en.wikipedia.org/wiki/Transcriptomics_technologies