

# Genomics and Machine Learning

Velmurugan R., Phong Thanh Nguyen, E. Laxmi Lydia, K. Shankar, Wahidah Hashim, Andino Maseleno

**Abstract:** Genomics is one of the most focused area for studying and helps to understand the nature of disease and it is an area where genetics can be deeply studied and research conclusion can be obtained. Genomics is different from genetics as genetics is the composition of only single gene but on the opposite side the genomics contains all gens and also keep track of their collectively growth during the development process of an organism. Here the datasets of DNA on the organism is called Genomic data. This datasets are further used in bioinformatics for doing experiments on collect and process for research. For this purpose a very large storage space and specifically-built computer program is required to analyze. Genomic is also different from the proteomics because in proteomics only focuses on the proteins present in the cell. The Genomics research involves many scientific factors, which leads to identify many diseases symptoms such as heart related disease, diabetic, cancer etc. Here in this approach genomics is useful because somewhere and somehow the genetic and the external factors are causing such diseases. The purpose of deep learning with genomics is to identify the disease and learning the development structure of disease. Such research may help in treating diseases in a better way.

**Keywords:** Genomics Study, Genomics Applications, Machine Learning Techniques, Genomics and Research.

## I. INTRODUCTION

The Genomics research includes numerous logical elements, which prompts recognize numerous infections side effects, for example, heart related illness, diabetic, malignant growth and so forth [1-3]. Here in this methodology genomics is valuable since some place and by one way or another hereditary and the outer components are causing such infections [4-6]. The reason for profound learning with genomics is to distinguish the ailment and learning the advancement structure of infection. Such research may help in treating illnesses in a superior manner. Genomics is one of the most engaged zone for considering and comprehends the idea of sickness and it is a territory where hereditary qualities can be profoundly examined and investigate end can be gotten.

Genomics is not the same as hereditary qualities as hereditary qualities is the creation of just single quality however on the contrary side the genomics contains all gens and furthermore monitor their all in all development during the advancement procedure of a life form [7-9]. Here the datasets of DNA on the living being is called Genomic information. This datasets are additionally utilized in bioinformatics for doing investigates gather and procedure for research. For this reason a huge extra room and explicitly assembled PC program is required to examine. Genomic is likewise not the same as the proteomics in light of the fact that in proteomics just spotlights on the proteins present in the cell [10-13].

## II. MACHINE LEARNING

Machine learning is one of the most important application of artificial intelligence which have ability to detect, analysis and produce results related to disease [14-17]. The beauty of machine learning process is deep learning. The program learns new things every day on the basis of datasets provided as in input. Figure 1 shows machine learning supporting the clinical DSS.

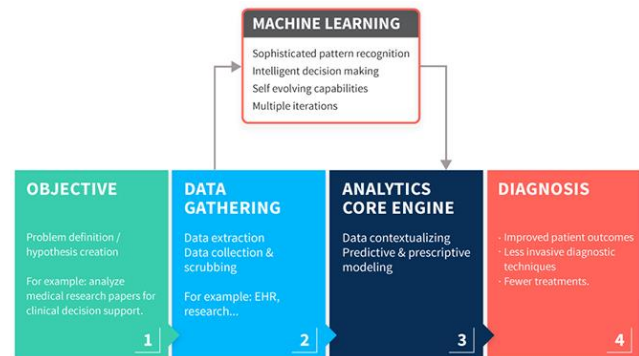


Figure 1. Machine Learning Supporting the Clinical DSS

In other words it can be said that machine learning is the process of continuously improvement and development of computer software which access datasets or patient data and use it for themselves for further learning. A large amount of datasets is used for machine learning. Although it helps to solve a problem in a faster way but there is also an disadvantage that it is very difficult to cross verify the datasets weather they are compatible for the machine learning software or not. Notwithstanding figuring out how to perceive designs in DNA groupings, AI can take as info information created by other genomic examines, for example, microarray or RNA-seq articulation information, chromatin availability tests, for example, DNase-seq, MNase-seq, and FAIRE, or histone alteration, translation factor (TF) restricting ChIP-seq information, and so forth.

Manuscript published on 30 July 2019.

\* Correspondence Author (s)

Velmurugan R., Department of Computer Science, Kristu Jayanti College, India.

Phong Thanh Nguyen, Department of Project Management, Ho Chi Minh City Open University, Vietnam.

E. Laxmi Lydia, Vignan's Institute of Information Technology (A), Department of Computer Science and Engineering, Visakhapatnam, Andhra Pradesh, India.

K. Shankar, Department of Computer Applications, Alagappa University, India.

Wahidah Hashim, Institute of Informatics and Computing Energy, Universiti Tenaga Nasional, Malaysia.

Andino Maseleno, Institute of Informatics and Computing Energy, Universiti Tenaga Nasional, Malaysia. E-mail: andino@uniten.edu.my

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Quality articulation information can be utilized to figure out how to recognize distinctive ailment phenotypes and, all the while, to distinguish possibly profitable malady biomarkers (Section 6). Chromatin information can be utilized, for instance, to explain the genome in an "unaided" style, along these lines conceivably taking into account the distinguishing proof of novel classes of practical components.

AI has likewise been utilized broadly to dole out utilitarian comments to qualities. Such comments most every now and again appear as Gene Ontology term assignments. Prescient calculations can take as information any at least one of a wide assortment of information types, including the genomic succession, quality articulation profiles crosswise over different test conditions or phenotypes, protein-protein connections, manufactured lethality information, open chromatin information, histone adjustment or TF restricting ChIP-seq information, and so forth. As a choice to GO term expectation, a few indicators rather recognize co-practical connections; i.e., the AI strategy yields a system where qualities are hubs and an edge between qualities A and B shows that the two qualities share a typical capacity .

Examine such factors are too much time consuming. Some of the advantages and Disadvantages of Machine learning for Genomics are given below:

### Advantages:

- Faster Identifies the common trends & Patterns
- Minimum human interaction required
- Deep Learning and Improvement in Code
- Faster Solution of Genetic Diseases
- Handling large amount & variety of data
- Web Applications

### Disadvantages:

- Always possibility of High Error
- Data Acquisition
- Results production and interpretation in human understandable format
- Time Consuming in Preparing Datasets
- Resource Availability

### III. TECHNIQUES DONE IN GENOMICS

As ML calculations increase involvement, they continue improving in exactness and effectiveness. This gives them a chance to settle on better choices. Let's assume you have to make a climate gauge model. As the measure of information you have continues developing, your calculations figure out how to make progressively exact expectations quicker. Although there are many techniques are available for Genomics. But below the most important eight techniques are given.



**Figure 2.** Genomics Applications: (1) Genomic DNA Isolation, (2) Separation of DNA, (3) Putting and Joining of DNA, (4) Cloning and Vectors, (5) Detection of Gene of Interest, (6) Recombinant DNA and Cloning, (7) Production of Multiple Copies of DNA Using Polymerase Chain Reaction (PCR), and (8) DNA Sequencing.

The techniques are not limited to the names given above. But these techniques are the most common techniques used by almost every research group.

IV. APPLICATIONS OF GENOMICS

There are many clinical applications of Genomic techniques are available. Such clinical applications are

vast and helps to improve the research in healthcare and provide support to the base of medical specialities. Here some of the applications are explained below:

<b>Quality revelation and analysis of uncommon monogenic issue</b>	Genomic advances can be utilized by clinicians from all specialities to analyze their patients who have high hazard hereditary mistakes causing ailment. Analysts are utilizing these strategies to recognize new qualities which cause hereditary sickness at a surprising rate - more than 4000 illnesses presently have a known single hereditary reason, contrasted with around 50 of every 1990.
<b>Distinguishing proof and conclusion of hereditary variables adding to basic ailment</b>	Genomic advances are progressively being utilized to comprehend the commitment of both uncommon and regular hereditary variables to the improvement of normal maladies, for example, hypertension, diabetes and disease.
<b>Pharmacogenetics and focused on treatment</b>	Hereditary data might be utilized to anticipate whether an individual will react to a specific medication, how well they will react to that medication and whether they are probably going to get any symptoms from the utilization of a particular medication. This permits their getting group settle on individualized choices about the correct medication treatment. At times, for example, malignancy, we can recognize the hereditary drivers of malady and afterward give drugs which explicitly focus on that pathway. This is known as focused treatment.
<b>Pre-birth determination and testing</b>	Hereditary illnesses are frequently destroying and may cause critical inability and even passing in adolescence. Pre-birth determination of hereditary sicknesses enables guardians to settle on choices about whether to proceed with the pregnancy or to permit early conclusion and conceivable treatment in utero or during childbirth. While past ways to deal with pre-birth finding could put the pregnancy in danger, new techniques utilizing genomic innovation can take a gander at the DNA of the embryo from a maternal blood test, without expanding the danger of unnatural birth cycle - this is known as non-obtrusive pre-birth testing. The utilization of NGS and cluster innovation in pre-birth tests is likewise on the expansion to improve demonstrative yields in a pregnancy.
<b>Irresistible maladies</b>	Sequencing the genomes of microorganisms which cause human disease can recognize the precise creature causing side effects, help to follow the reason for irresistible flare-ups, and give data with respect to which anti-toxins are well on the way to be viable in treatment.
<b>Customized drug</b>	As the definite DNA grouping of the genome of every human is novel to them, we will all have special ailment susceptibilities and treatment reactions. Customized prescription portrays the utilization of our hereditary data to tailor medicinal services mediation to our very own individual need.
<b>Quality treatment</b>	Quality treatment includes the organization of DNA or RNA, so as to address a hereditary anomaly, or adjust the declaration of qualities.
<b>Genome altering</b>	Genome altering utilizes sub-atomic strategies to adjust the genome - genome altering can include, cut out, or supplant areas of the DNA succession.

V. CONCLUSION

The Genomics research incorporates various legitimate components, which prompts perceive various contaminations symptoms, for instance, heart related disease, diabetic, dangerous development, etc. As Machine Learning computations increment association, they keep improving in precision and adequacy. This allows them to choose better decisions. How about we expect user need to make an atmosphere check model. As the proportion of data user has keeps building up, their estimations make sense of how to make continuously correct desires speedier. Despite the fact that there are numerous methods are accessible for Genomics. In any case, underneath the most significant

eight strategies are given. Here in this technique genomics is profitable since some spot and by somehow inherited and the external parts are causing such contaminations. The explanation behind significant learning with genomics is to recognize the infirmity and learning the headway structure of disease. Such research may help in treating sicknesses in an unrivaled way. Genomics is one of the most connected with zone for considering and fathoms the possibility of infection and it is where inherited characteristics can be significantly inspected and explore end can be gotten.

Genomics isn't equivalent to innate characteristics as inherited characteristics is the formation of simply single quality anyway despite what might be expected side the genomics contains all gens and moreover screen their with everything taken into account improvement during the headway methodology of a living thing. Here the datasets of DNA on the living being is called Genomic data. This datasets are moreover used in bioinformatics for doing explores assemble and strategy for research. Therefore an enormous additional room and expressly collected PC program is required to look at. Genomic is in like manner not equivalent to the proteomics in light of the way that in proteomics just spotlights on the proteins present in the cell. The conclusion of this study can be summarized as Machine Learning is the base of Medical and Genomics. Continuously research and improvements in the results may provide the faster treatment of many serious disease.

### References

1. Niu, B., Liang, C., Lu, Y., Zhao, M., Chen, Q., Zhang, Y., Zheng, L. and Chou, K.C., 2019. Glioma stages prediction based on machine learning algorithm combined with protein-protein interaction networks. *Genomics*.
2. Cordier, T., Lanzén, A., Apothéloz-Perret-Gentil, L., Stoeck, T., & Pawlowski, J. (2018). Embracing environmental genomics and machine learning for routine biomonitoring. *Trends in microbiology*.
3. Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., & Telenti, A. (2018). A primer on deep learning in genomics. *Nature genetics*, 1.
4. Eraslan, G., Avsec, Ž., Gagneur, J. and Theis, F.J., 2019. Deep learning: new computational modelling techniques for genomics. *Nature Reviews Genetics*, p.1.
5. Schrider, D.R. and Kern, A.D., 2018. Supervised machine learning for population genetics: a new paradigm. *Trends in Genetics*, 34(4), pp.301-312.
6. Beam, A.L. and Kohane, I.S., 2018. Big data and machine learning in health care. *Jama*, 319(13), pp.1317-1318.
7. Ohler W, Liao C, Niemann H, Rubin GM. Computational investigation of center advertisers in the drosophila genome. *Genome Biology*. 2002;3
8. Degroev S, Baets BD, de Peer YV, Rouz P. Highlight subset determination for join site forecast. *Bioinformatics*. 2002;18:S75–S83.
9. Bucher P. Weight network portrayal of four eukaryotic RNA polymerase II advertiser components got from 502 inconsequential advertiser arrangements. *Diary of Molecular Biology*. 1990;4:563–578.
10. Heintzman N, et al. Particular and prescient chromatin marks of transcriptional advertisers and enhancers in the human genome. *Nature Genetics*. 2007;39:311–318.
11. Segal E, et al. A genomic code for nucleosome situating. *Nature*. 2006;44:772–778. [PMC free article]
12. Picardi E, Pesole G. Computational strategies for stomach muscle initio and near quality finding. *Strategies in Molecular Biology*. 2010;609:269–284.
13. Gene Ontology Consortium Gene cosmology: apparatus for the unification of science. *Nature Genetics*. 2000;25:25–29.
14. Mandl, K.D. and Manrai, A.K., 2019. Potential excessive testing at scale: biomarkers, genomics, and machine learning. *Jama*, 321(8), pp.739-740.
15. Hathaway, Q.A., Roth, S.M., Pinti, M.V., Sprando, D.C., Kunovac, A., Durr, A.J., Cook, C.C., Fink, G.K., Chevront, T.B., Grossman, J.H. and Aljahli, G.A., 2019. Machine-learning to stratify diabetic patients using novel cardiac biomarkers and integrative genomics. *Cardiovascular diabetology*, 18(1), p.78.
16. Eraslan, G., Avsec, Ž., Gagneur, J. and Theis, F.J., 2019. Deep learning: new computational modelling techniques for genomics. *Nature Reviews Genetics*, p.1.
17. Trakadis, Y.J., Sardaar, S., Chen, A., Fulginiti, V. and Krishnan, A., 2019. Machine learning in schizophrenia genomics, a case-control study using 5,090 exomes. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 180(2), pp.103-112.