iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** V **Month of publication:** May 2019

DOI: https://doi.org/10.22214/ijraset.2019.5112

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

Machine Learning Approach for Identification of Diseases through Gene Mapping

Shalmalee Belapurkar[1], Shivani Budhkar [2]
[1, 2]*P.E.S. Modern College of Engineering, Pune*

Abstract: ***Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. Genetic mapping is based on the use of Genetic Techniques to construct maps showing the positions of genes and other sequence features on genome. The early applications of machine learning to population genetics demonstrate that they outperform traditional approaches. Potentially important disease biomarkers have been revealed by the use of machine learning methods on gene expression data, where algorithms learn to differentiate between different disease phenotypes. Genetic testing can be considered as the perfect field for machine learning applications in many ways, considering the enormous amount of data that these programs need to contend with.***

Keywords: ***Machine Learning, Gene Mapping, Diabetes***

I. INTRODUCTION

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. It is a type of technology that develops computer programs that make the use of data that they can access, to learn for themselves. Machine learning algorithms involve the development and application of computer algorithms that improve with experience. [8] Genetic mapping is based on the use of Genetic Techniques to construct maps showing the positions of genes and other sequence features on genome. Although measuring the transcription of a single gene is not new, it is new to measure at once the transcription of all the genes in an organism. The early applications of Machine Learning to population genetics demonstrate that they outperform traditional approaches. Potentially important disease biomarkers have been revealed by the use of machine learning methods on gene expression data, where algorithms learn to differentiate between different disease phenotypes. [8] Agencies like the National Institutes of Health are documenting the many ways that machine learning and artificial intelligence contribute to better understanding of genetics and genomics.

II. CONCEPTUAL DEFINITIONS

Machine learning: Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. [9]

Gene mapping:-Gene mapping describes the methods used to identify the locus of a gene and the distances between genes. The essence of all genome mapping is to place a collection of molecular markers onto their respective positions on the genome. [10]

III.NEED AND IMPORTANCE

In the post-genomic era, it is one of the most challenging tasks is to identify disease genes from a vast amount of genetic data. Also, complex diseases present highly heterogeneous genotype, which makes biological marker identification difficult. These markers are identified widely with the help of machine learning methods. While identifying disease genes, we come across some challenges such as, extracting the location and structure of genes, identifying regulatory elements, identifying non-coding RNA genes, gene function prediction, RNA secondary structure prediction etc. These have all been tackled using machine learning approaches. Machine learning models are found to be very useful in neurological diseases. It is helpful in facilitating precision medicine and neuroscience research. The genes of an organism are expressed through the production of proteins, the building blocks of life.

The method by which the genes of an organism are expressed is through the production of proteins, the building blocks of life. A specific protein1 is encoded by each gene, and various proteins are being produced at each point in the life of a given cell. The production of specific proteins that an organism responds to environmental and biological situation, such as stress, and to different developmental stages, such as cell division, is done through it turning on and off.

It is made possible to simultaneously measure the rate at which a cell or tissue is expressing and translating into a protein each of its thousands of genes, through the use of gene-expression microarrays, commonly called as gene chips.

IV.LITERATURE SURVEY

Different types of diseases can be identified and cured with the applications of machine learning approaches through mapping. Important of them are:

Abder-Rahman Ali in his journal stated that Identifying deadline skin cancers with high accuracy rates even more than regular practitioners can be possible due to cancer classification with deep neural networks. Service based apps are being used for diagnosis due to worldwide expansion of mobile access. [1]

Harleen Kaur and Vinita Kumari, in their journal, stated that Rdata manipulation tool can be used to develop trends and detect patterns with risk factors, using machine learning techniques. Five different predictive models can be developed and analyze using Rdata manipulation tool in order to classify the patients into diabetic and non-diabetic. Supervised machine learning algorithms namely linear kernel support vector machine (SVM-linear), radial basis function (RBF) kernel support vector machine, k-nearest neighbor (k-NN), artificial neural network (ANN) and multifactor dimensionality reduction (MDR) are used for this purpose.[2]

Juan A. Botia, Sebastian Guelfi, in their journal, "G2P: Using machine learning to understand and predict genes causing rare neurological disorders", the genes that cause rare neurological disorders can be understood and predicted using Machine Learning approaches. One can get excellent demonstration of explanatory as well as predictive power of machine learning based models in neurological diseases in this research. [3]

Margaret A. Shipp, et al, tells that machine learning can be effectively applied in **Diagnosis of tumor**. A supervised learning prediction method is applied to identify cured versus fatal or refractory diseases by analysing the expression of various genes in diagnostic tumor specimens from the DLBCL patients. [4]

Muhammad,Hugo F. M. et al, states that improvement in identification of genes involved in complex diseases is possible by gene ontology. In his research, he built various types of machine classifiers on quantitative semantic similarity matrices of ASD and non ASD genes. Through gene functional similarities ASD classifiers are reported. [5]

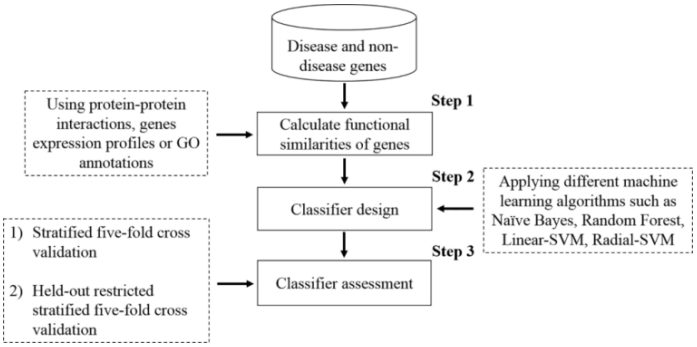


Fig.2: Steps in identification [5]

According to Paul W.C. Hsu and Hsien-Da Huang, the expression profiles of the known mi RNA's cross-species comparisons, gene annotations and cross- links to other biological databases can be provided by mi RNA maps. Here for retrieval of data from miRNA maps, textual as well as graphical data is provided. [6]

Stig Nikolaj Blomberg introduced that machine learning framework can be trained to recognize cardiac arrest from the recorded calls. Sensitivity, specificity, and positive predictive value for recognizing out-of-hospital cardiac arrest can be calculated. The performance of the machine learning framework is then compared to the actual recognition and time-to recognition of cardiac arrest by medical dispatchers. Machine learning may play an important role as a decision support tool for emergency medical dispatchers. [7]

Input data	Task
DNA sequence	Identify transcription start sites, splice sites, exons, etc.
DNA sequence	Identify TF binding sites
DNA sequence	Identify genes
Gene expression	Predict regulatory relationships

Input data	Task
Gene expression data	Identify biomarkers for a disease
Histone and TF ChIP-seq data	Partition and label the genome with chromatin state annotation
DNA sequence + gene expression + …	Predict gene function
DNA sequence + histone mods + …	Predict gene expression
DNA sequence + histone mods + …	Predict variant deleteriousness
Sequence variants + gene expression + …	Predict disease phenotype or prognosis

Table 1: Selected applications of machine learning [7]

V. LIMITATIONS

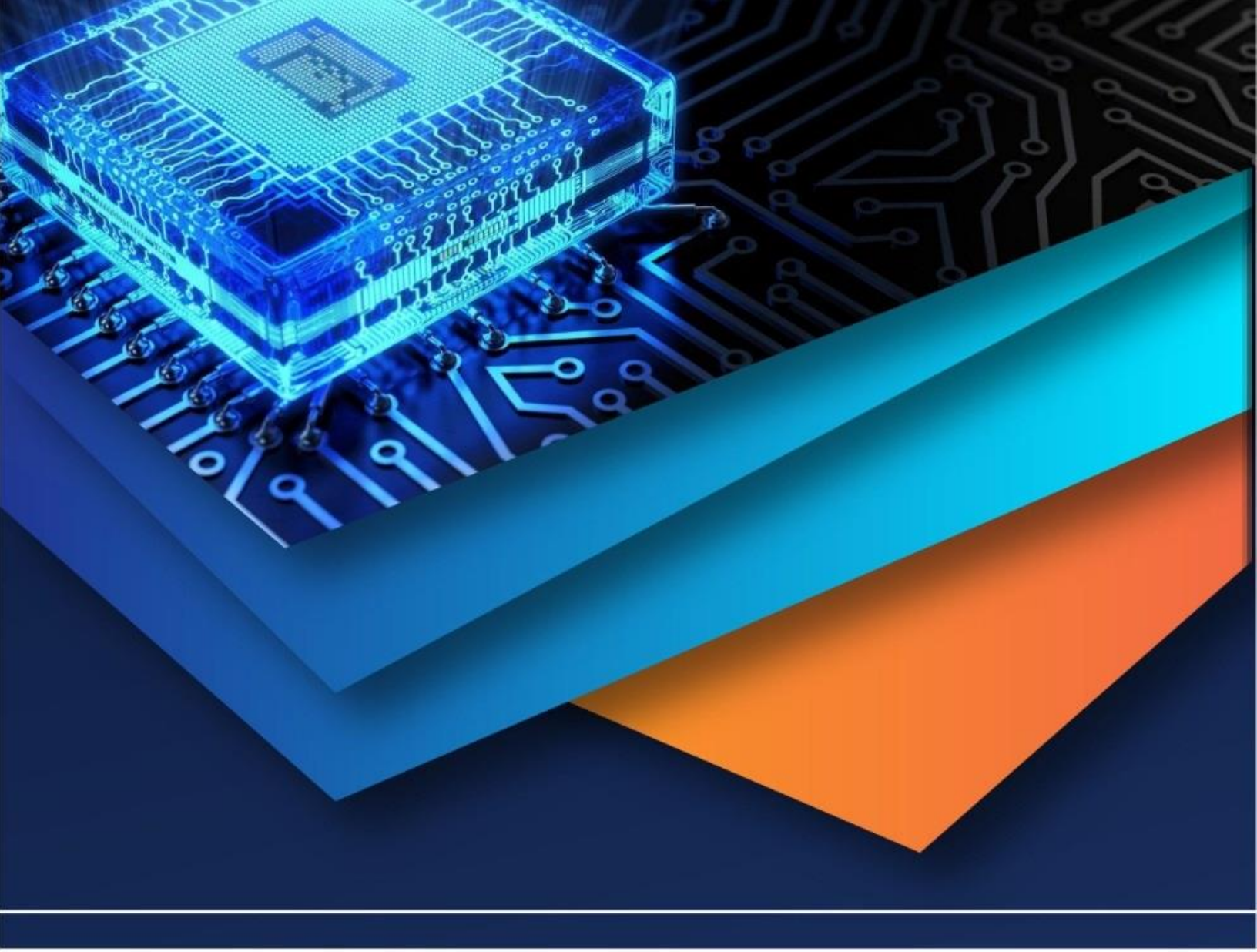
A. Machine learning algorithms require massive stores of training data.
B. Machines cannot explain themselves.
C. When machine learning algorithms are deployed, there may be more instances in which potential bias finds its way into algorithms and data sets.
D. Machine learning is susceptible to errors.

VI. CONCLUSIONS

The utilization of machine learning algorithms and techniques in genomics is increasingly extensive. It is an alternative to the traditional genome-wide association studies (GWAS). Recent successful application includes cancer research, where crucial information regarding patient genotypes, gene-expression-related phenotypes, and patient outcomes has been revealed. The applications of machine learning are nearly endless. The scientists have taken the help of machine learning in activities like analyzing DNA, decoding the human genome, assessing disease phenotypes, understanding gene expression, processes such as gene editing (a process which splices the DNA into an organism's genetic code).

REFERENCES

[1] Abder-Rahman Ali, Deep Learning in Oncology – Applications in Fighting Cancer, Business Intelligence and Analytics, February 19, 2019
[2] Harleen Kaur, Vinita Kumari, Predictive modeling and analytics for diabetes using machine learning approach, ScienceDirect, 19 December 2018
[3] Juan A. Botia, Sebastian Guelfi , G2P: Using machine learning to understand and predict genes causing rare neurological disorders,
[4] Margaret A. Shipp, et al, Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning, nature, Vol 8, Pages 68–74, 01 January 2002.
[5] Muhammad, Hugo F. M. et al, Identifying disease genes using machine learning and gene functional similarities, assessed through Gene Ontology, December 10, 2018
[6] Paul W.C. Hsu and Hsien-Da Huang, miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes, Nucleic Acids Research, 2006, Vol. 34, October 27, 2005, Pages D135-D139.
[7] Stig Nikolaj Blomberg, FredrikFolke, Machine learning as a supportive tool to recognize cardiac arrest in emergency calls, ScienceDirect, 18 January 2019.
[8] William Noble, Maxwell W. Libbrecht, Machine learning applications in genetics and genomics, Nature Review Genetics, May 2015.
[9] https://www.expertsystem.com/machine-learning-definition/
[10] https://en.wikipedia.org/wiki/Gene_mapping



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY