International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Disease Free Genetic Editing using AI-ML

[1]Sampada Bhongale, [2]Prof. Pritesh Patil

[1,2]AISSMS Institute of Information Technology, Pune

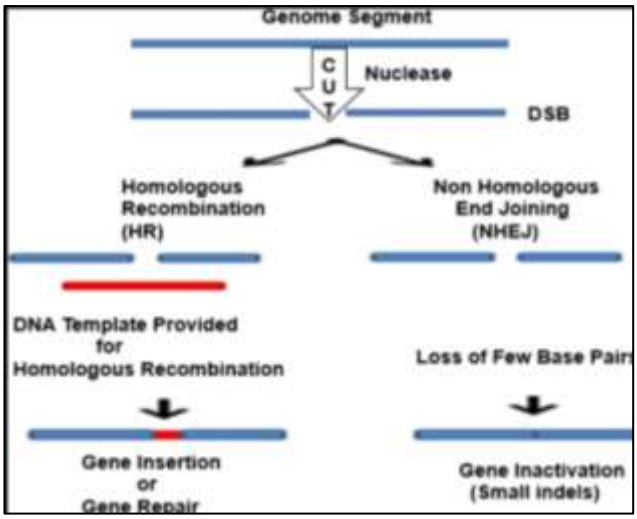
Abstract-

Bioinformatics, as related to genetics and genomics, is used to collect, store and analyze DNA and amino acid sequences or annotations about those sequences. Genome editing (also called gene editing) is a group of technologies that give scientists the ability to change an organism's DNA. These technologies allow genetic material to be added, removed, or altered at particular locations in the genome. This technology will help us to predict disease-free genetic codes without any hereditary gene .AI_ML will help in Identifying disease-causing genomic variants . Machine learning can be used to help find patterns in how small variations in genes and regulatory regions result in phenotypic changes (traits, wellness, and health) in a more automated fashion.

Keywords- Bioinformatics, Gene, CRISPR-Cas9

Introduction-

Genome editing, also known as gene editing, refers to a range of scientific techniques that enable the modification of an organism's DNA. The prevention and treatment of human diseases is a major area of focus for genome editing. Genome editing is currently employed in research facilities to study diseases in cells and animal models. Researchers are currently figuring out whether this method is secure and efficient for usage in people. For a wide range of illnesses, including single-gene diseases like cystic fibrosis, haemophilia, and sickle cell disease, it is being investigated in research and clinical trials. Additionally, it shows promise in the management and avoidance of more complicated illnesses like cancer, heart disease, mental illness, and HIV infection. From which CRISPR-Cas9 was derived.



Methods-

Results of genome editing. Double-strand breaks are produced by genome editing nucleases (DSBs). There are two methods for repairing breaks: homologous recombination (HR) in the presence of a donor template, or non-homologous end joining (NHEJ) in the absence of a donor template. Few base insertions or deletions, which result in an indel, or frameshifts, which disrupt genes, are produced by the NHEJ.

John Holland, father of genetic algorithms-

Five phases are considered in a genetic algorithm:

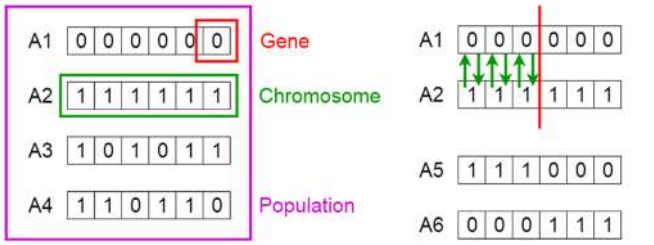
A. Initial population

B. Fitness function

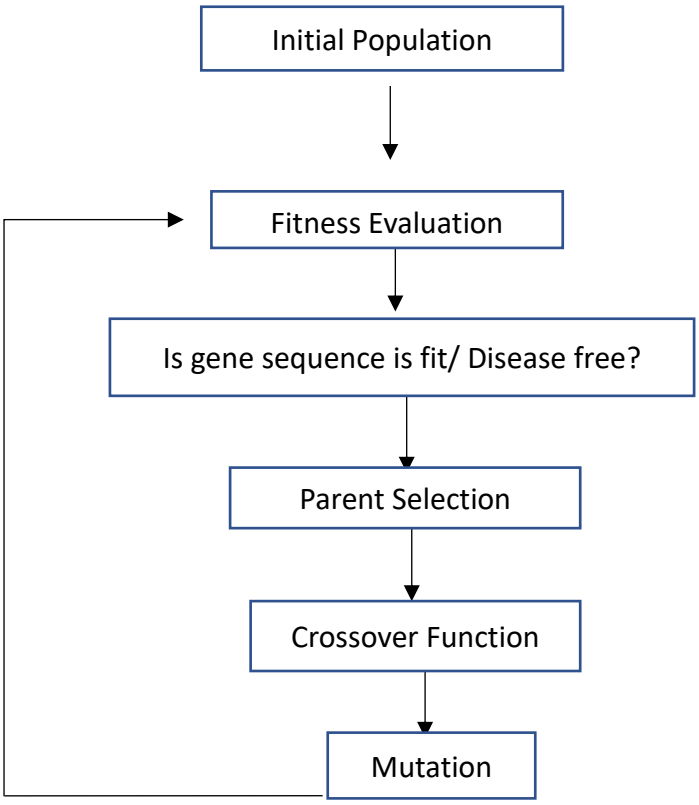
C. Selection

D. Crossover

E. Mutation



1. Steps of Gene Algorithm



Methods

A. Crispr-Cas9

- CRISPR-Cas9 technology caused a major stir in the field of genome engineering when it was first described as an editing tool in 2012, including functional genomics, diagnostics, DNA imaging, and therapeutics.
- Genome Editing Techniques: The Tools That Enable Scientists to Alter the Genetic Code (synthego.com)

- One critical system supporting precision genome editing is CRISPR-Cas9, which provides scalable targeted cleavage of DNA through guide RNA programming. Thus, the prediction and minimizing of off-target cleavage effects of CRISPR-Cas9 is a crucial research area to improve the efficacy of genome editing technology.
- After applying several machine learning algorithms and a simple deep neural network (DNN), we identified a relatively high-performing support vector machine (SVM) model with a 64% recall rate.
- CRISPR -Cas9

B. The Genome Sequence Data Base (GSDB)

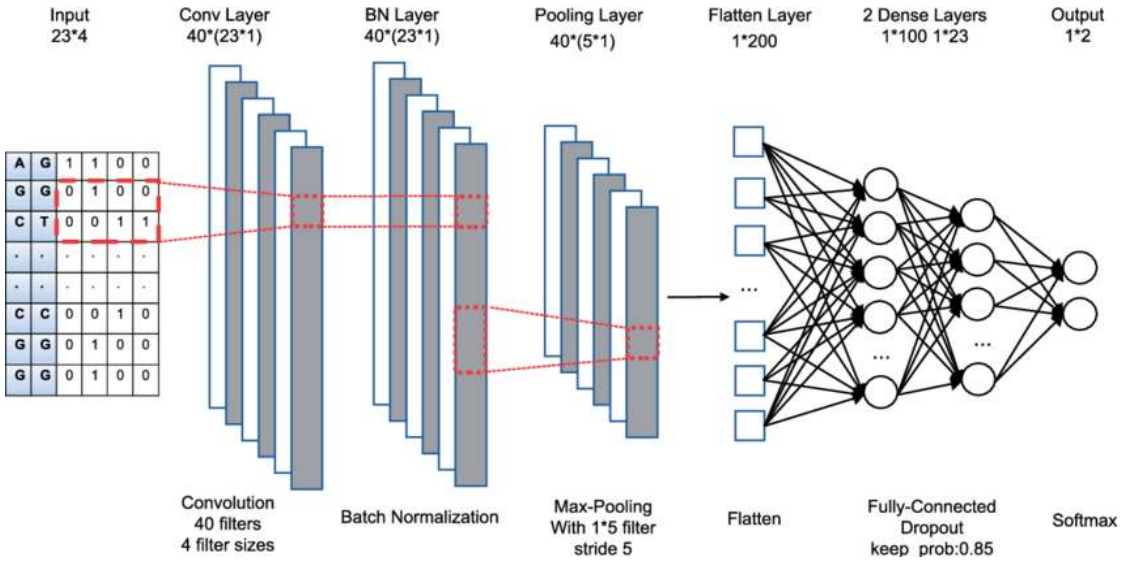
- All publicly available DNA sequences and associated annotations can be found in the Genome Sequence Data Base (GSDB). DNA sequences and relevant biological annotations make up GSDB entries.
- The four nucleotides adenine, cytosine, guanine, and thymidine are represented by the letters a, c, g, and t in the sequence itself. The annotation generally contains details indicating the biological use of specific sequence sections.https://www.ncbi.nlm.nih.gov/genbank/

➢ ***Drawback***

- Unintended cleavage and mutations at untargeted genomic locations with a similar but not identical sequence to the target site are examples of off-target effects.
- As a result, there are drastic changes that result in accidental disease-related deaths.

B. Deep learning-based off-target predictions for CRISPR-Cas9 gene editing

- Using typical deep learning models and 2D arrays of genetic code, CNN is utilised to anticipate off-target effects.
- The conventional deep CNN architecture for off-target prediction. The 23 base pair long encoded sgRNA-DNA sequence serves as the deep neural network's input.
- The convolutional layer has 40 filers in it. In order to accelerate learning and prevent over-fitting, the BN layer is employed to normalise the convolutional layer's output.
- The max-pooling layer's outputs are combined into one.



1	000	001	010	011	100	101	110	111
000	TTT	TTC	TCT	TCC	CTT	CTC	CCT	CCC
001	TTA	TTG	TCA	TCG	CTA	CTG	CCA	CCG
010	TAT	TAC	TGT	TGC	CAT	CAC	CGT	CGC
011	TAA	TAG	TGA	TGG	CAA	CAG	CGA	CGG
100	ATT	ATC	ACT	ACC	GTT	GTC	GCT	GCC
101	ATA	ATG	ACA	ACG	GTA	GTG	GCA	GCG
110	AAT	AAC	AGT	AGC	GAT	GAC	GGT	GGC
111	AAA	AAG	AGA	AGG	GAA	GAG	GGA	GGG

Conclusion :

- A novel binary representation of the canonical genetic code is proposed, in which the bits correspond to the physicochemical characteristics of the amino acids they code for and have a physical significance.
- Surprisingly, I discovered that CNN with label encoding beats the other models while having poor testing accuracies. The optimum technique for achieving high testing and validation accuracy is K-mer encoding.
- Compared to other state-of-the-art models, CNNCrispr performs better in classification and regression. CNNCrispr can forecast sgRNA's tendency to bind to particular DNA pieces off-target. CNNCrispr automatically develops the sgRNA-DNA pair's sequence characteristics.
- Genetic engineering can treat hereditary illnesses like AIDS, COVID-19, blood problems, and muscular dystrophy.

References:

- Jinghui Zhong, Liang Feng, Yew-Soon Ong, "Gene Expression Programming: A Survey [Review Article]", *IEEE Computational Intelligence Magazine*, vol.12, no.3, pp.54-72, 2017.
- Qianyun Zhang, Kaveh Barri, Pengcheng Jiao, Hadi Salehi, Amir H. Alavi, "Genetic programming in civil engineering: advent, applications and future trends", *Artificial Intelligence Review*, vol.54, no.3, pp.1863, 2021.
- Adaptive Control of Ill-Defined Systems, 1984, Volume 16 ISBN : 978-1-4684-8943-9 John H. Holland
- JOUR ,Thai, Khac-Minh,Tapak, Leili,Afshar, Saeid ,Afrasiabi, Mahlagha,Ghasemi, Mohammad Kazem,Alirezaei, Pedram,2021 ,2021/09/08,TI - Application of Genetic Algorithm-Based Support Vector Machine in Identification of Gene Expression Signatures for Psoriasis Classification: A Hybrid Model
- The Bibliometric Landscape of Gene Editing Innovation and Regulation in the Worldwide cells,PY - 2022 ,VL - 11 ,SN - 2073-4409
- Journal of Genetic Engineering and Biotechnology, 2020, Volume 18, Number 1, Page 1 Ahmad M. Khalil
- GGA-MLP: A Greedy Genetic Algorithm to Optimize Weights and Biases in Multilayer Perceptron, SP - 4036035 ,VL - 2022 UR - https://doi.org/10.1155/2022/4036035
- Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell. 2014 Jun 5;157(6):1262-1278. doi: 10.1016/j.cell.2014.05.010. PMID: 24906146; PMCID: PMC4343198.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. Nucleic Acids Res. 2005 Jan 1;33(Database issue):D34-8. doi: 10.1093/nar/gki063. PMID: 15608212; PMCID: PMC540017.