AMERICAN INTERNATIONAL UNIVERSITY - BANGLADESH

Assign./Case Title:	**Machine Learning in Genomics and Precision Medicine**		
Assign./Case No:	02	Date of Submission:	22 September 2024
Course Title:	**Research Methodology**		
Course Code:	00652	Section:	A
Semester:	Summer 2023-24	Degree Program:	BSc [CSE]
Course Teacher:	**Prof. Dr. Md. Asraf Ali**		

Declaration and Statement of Authorship:

1. I/we hold a copy of this Assignment/Case-Study, which can be produced if the original is lost/damaged.
2. This Assignment/Case-Study is my/our original work and no part of it has been copied from any other student's work or from any other source except where due acknowledgement is made.
3. No part of this Assignment/Case-Study has been written for me/us by any other person except where such collaboration has been authorized by the concerned teacher and is clearly acknowledged in the assignment.
4. I/we have not previously submitted or currently submitting this work for any other course/unit.
5. This work may be reproduced, communicated, compared and archived for the purpose of detecting plagiarism.
6. I/we give permission for a copy of my/our marked work to be retained by the Faculty for review and comparison, including review byexternal examiners.
7. I/we understand that Plagiarism is the presentation of the work, idea or creation of another person as though it is your own. It is a form of cheating and is a very serious academic offence that may lead to expulsion from the University. Plagiarized material can be drawn from, and presented in, written, graphic and visual form, including electronic data, and oral presentations.Plagiarism occurs when the origin of the material used is not appropriately cited.
8. I/we also understand that enabling plagiarism is the act of assisting or allowing another person to plagiarize or to copy my/our work.

* *Student(s) must complete all details except the faculty use part.*
** Please submit all assignments to your course teacher or the office of the concerned teacher.

Group Name/No.: **3 (Three)**

No	Name	ID	Signature
1	Afsana Jahan Onu	21-45708-3	
2	Fatema Akter Sujana	21-45693-3	
3	Mussa Alam	21-45036-2	

Faculty use only		
FACULTY COMMENTS	**Marks Obtained**	
	Total Marks	

Machine Learning in Genomics and Precision Medicine: Breakthroughs, Obstacles, and Clinical Applications

Abstract

Objectives

This study applies machine learning techniques effectively, efficiently and accurately to precision medicine and genomics. It enhances clinical research greatly. It focuses on multi-omics data analysis. The goal is to create progressive machine learning models. These models will accelerate genetic data analysis. Treatment will be additional personalized. They will predict disease susceptibility better. The study explores AI-driven healthcare's moral ramifications. It examines the challenges of amalgamating diverse omics layers.

Research Methodology

The study analyzes genomic data using a wide range of complex learning methods, techniques, and tools. It draws from accessible data sources, which originate from genomic repositories and include sensitive information. The patients suffer from chronic illnesses such as cancer and heart conditions, as well as neurological diseases like Alzheimer's and Parkinson's. The study strictly follows confidentiality protocols, ensuring full protection of individual privacy. These regulations comply with GDPR and HIPAA standards to safeguard confidential genetic data.

Significance of the Output

Precision medicine is the goal. They offer a flexible framework for risk assessment and ailment diagnosis. The framework also allows for tailored treatment. AI integration with multi-omics data improves clinical decision-making. It helps in preliminary detection. It also helps create personalized treatment regimens.

Introduction

Motivation for the Research

Precision medicine is now potential. Precision medicine moves from conventional symptom- focused healthcare to a custom approach. This approach considers each patient's too distinct heritable profile, lifestyle and environment. Very efficacious techniques are needed to analyze data fromhigh-throughput technologies. These are high-throughput technologies. These technologies produce expansive amounts of data. Machine learning is a trustworthy solution to analyze data.

Machine learning produces insights into disease mechanisms and progression. It improves treatment outcomes.

Machine learning is a crucial subset of more extensive artificial intelligence technologies. It uses AI. It analyzes spacious datasets swiftly and precisely. It is used in diverse applications. It identifies cancer subtypes and forecasts treatment outcomes. This project builds upon current knowledge thoroughly. It advances prior studies further. It develops a framework for machine learning. The framework addresses very functional genomics problems. It enables efficacious precision medicine treatment plans.

Problem Statement

Genomic research necessitates combining varied data sources effectively nowadays globally. It requires multi-omics data integration. This task is critical but requiring. Omics data is multidimensional. Data sources are different. Very medical outcomes are difficult to predict. Current approaches have limitations. Traditional methods fail to capture omics interactions completely. Interpretations are often so wrong or partial.

It solves the problem. Challenges exist in scalability and interpretability. Moral considerations also remain, especially data privacy. This study addresses these issues. It uses developed machine learning. These techniques integrate multi-omics data. They retain very tall accuracy and scalability.

Hypothesis

Advanced machine learning models significantly improve predictions. It improves data analysis greatly. This integration makes predictions markedly more precise. Greater precise predictions facilitate tailored treatment plans in precision medicine. A combination of deep learning models and guided learning is believed to be efficient. It outperforms conventional methods. It will be especially advantageous in diseases like cancer. Cancer treatment outcomes depend on addressing with disease heterogeneity. Deep learning models can handle this crucial aspect efficiently.

Literature Review

Genomics relies on machine learning. Deep learning techniques analyze intricate genetic data. Applications include cancer variation identification. Applications include drug discovery too. These applications also include prognostic analysis tools. It assesses disease risk factors.

- **Deep Learning in Genomics:** It works very well indeed. It includes cancer subtype prediction really functional genomic data. It also includes identifying drug targets. Deep convolutional neural networks (CNNs) have been used. These tests are vital in unveiling lung cancer symptoms accurately today. They aid in diagnosis. They outperform established analytical methods in lung cancer diagnosis.

- **Multi-Omics Integration**: Machine learning integrates multi-omics data. Machine learning integrates genomics, transcriptomics and proteomics data efficiently. This integration helps understand intricate diseases superior. Machine learning models combine gene expression profiles and drug response data. It helps cancer treatment greatly.

Literature Review's Table

Author	**Title**	**Data Description**	**Technology**	**Algorithm**	**Outcome**	**Limitations**	**Comments**
Quazi, S., et al. (2022)	Artificial intelligence and machine learning in precision and genomic medicine	Review of AI/ML applications in precision and genomic medicine	AI, ML, Genomic Medicine	Various ML algorithms	Identified potential of AI in genomic medicine	Limited by current data availability and integration challenges	Highlights the need for multidisciplinary approaches in AI and medicine
Rezayi ,M., et al. (2022	Effectiveness of artificial intelligence for personalized medicine in genomics	Evaluation of AI effectiveness in genomics	AI, Genomics	Deep learning, Random Forest	Improved personalized medicine outcomes	Sample size and generalizability issues	Sample size and generalizability issues
Huang, S., Yang, J., Fong, S., & Zhao, Q. (2020)	Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges	Review of AI applications in cancer diagnosis and prognosis	AI, Cancer Research	Deep Learning ,Random Forest	Identified key opportunities in AI-driven cancer research	Ethical concerns, data privacy issues	Explores both the advantages and ethical dilemmas of AI in healthcare
Cruz, J.A., & Wishart, D. S. (2006)	Applications of machine learning in cancer prediction and prognosis	Review of ML applications in cancer prediction	ML, Cancer Research	Support Vector Machines, Neural Networks	ML in cancer prediction	Lack of longitudinal studies and potential biases	Early work in the field, providing foundational knowledge
Schork, N. J. (2019)	Artificial intelligence and personalized medicine	Discussion on AI's role inadvancing personalized medicine	AI, Personalized Medicine	Various AI algorithms	Emphasized the growing role of AI in personalizing treatment	Challenges with data integration and regulatory issues	Explores both opportunities and regulatory challenges in AI-driven medicine

Applications in Cardiovascular Diseases

Machine learning (ML) algorithms offer quite vast potential, important benefits and increased accuracy in predicting CVDs. It's a guiding worldwide killer. Random forests, SVM and brain-related networks analyze diverse data sources for formative detection. These sources include electrocardiograms, genetic data and picture-taking data for analysis. The goal is to detect formative signs of cardiovascular disease and forecast patient outcomes.

For example, atrial fibrillation detection involves intricate algorithms, machine learning and data analysis from ECG data. Deep learning models identify it. Deep learning models improve initial diagnosis and achieve elevated accuracy. Broad genetic datasets are analyzed by ml models. ML models analyze behavioral and really medical data too. They determine heart failure risk. These models predict cardiovascular risk better than analytic techniques. They combine multi-omics data with really medical variables.

Drug Discovery and Precision Medicine

Machine learning transforms the oncology field rapidly nowadays. It improves precision medicine greatly. Machine learning computational formulas forecast drug efficacy in drug discovery. Algorithms discern prospective drug targets in drug discovery. Algorithms assess patient responses to therapies in drug discovery. Models aid in cancer treatment. Patients are stratified based on genetic profiles in precision oncology. Stratification allows for customized treatment regimens in precision oncology.

For example, deep learning models effectively analyze intricate genetic data sets accurately. DNNs forecast cancer treatment outcomes. It improves curative outcomes and lowers side effect risks. Machine learning speeds up drug development processes. It identifies novel targets through genetic and protein-related data analysis. It analyzes the data sets.

Ethical Considerations in Machine Learning for Genomics

Machine learning in genomics offers very numerous important benefits and opportunities. It has several advantages. It raises ethical questions about enduring consent. It raises ethical questions about data privacy. It raises ethical questions about step-by-step judgment biases. Genomic data is highly responsive, extremely private and individual information. It requires very rigid protection. There are issues with utilizing genomic data in healthcare. There are issues with allocating genomic datain research.

Lack of diversity is the problem. This can result in unjust healthcare outcomes for Indefinite populations. To address these concerns, really powerful privacy protocols are essential. These protocols will ensure genomic data is handled accountably. AI systems should be just, responsible andtrustworthy. Models must be very really clear. They should be devoid of bias and unfairness. Advancing moral guidelines is also critical in this regard.

Research Questions

1. How can multi-omics data be efficiently integrated into machine learning algorithms to enhance disease prediction and treatment personalization?
2. How can the main ethical issues surrounding the application of machine learning in genomics be addressed?
3. How does machine learning compare to conventional statistical techniques for patient stratification and disease susceptibility prediction?

Objectives of the Research

1. **Develop ML models for multi-omics data integration:** Machine learning is the primary goal. These frameworks integrate distinct omics layers efficiently. Omics layers include transcriptomics, proteomics and genomics. This integration enhances disease prediction.
2. **Evaluate the effectiveness of ML algorithms**: It evaluates machine learning methods. These methods include profound learning and support vector machines. They also include haphazard forests and neural networks. The goal is to integrate multi-omics data. It predicts tolerant diseases. The study assesses the effectiveness of these methods.
3. **Optimize treatment personalization in precision medicine**: It uses ML models successfully. It integrates very medical data as well. It also integrates lifestyle data. This integration improves personalized treatment strategies. Treatments are more efficacious now. Mending treatments reduces secondary effects.

Methodology

Subjects

TCGA data will be used here. These datasets include transcriptome, protein and genetic information. The datasets are from diverse sick populations with contending disorders. Complex disorders include cancer and cardiovascular diseases. The research combines health, healthcare and everyday data. Study includes therapeutic information. This integration will build additional thorough models for treatment. The models will be for tailored treatment and illness prediction.

Ethical Issues

Data must be preserved confidential. This study will follow demanding data protection and moral guidelines. It will adhere to GDPR and HIPAA regulations. Informed consent will be obtained where achievable. Data will be preserved confidential.

Benefits and Risks

It helps treat diseases better. It can also speed up drug discovery. Main risks include data breaches and biases in models. These biases can cause disparate healthcare outcomes. Data protection reduces risks greatly. Datasets must represent varied populations to reduce risks.

Privacy & Confidentiality

Genomic research relies heavily on secrecy, trust and professionalism. Confidentiality is very significant. Datasets will be deidentified for privacy. Datasets will be stored in encrypted format. Authorized personnel will have access to the data. Findings will be shown briefly. This avoids distinctive very unique subjects. Data sharing policies follow elevated moral standards. Personal health information remains protected throughout research.

Significance of the Study

This study merges customary, comprehensive and cutting-edge healthcare approaches with precision medicine. It combines aged and recent methods. It uses machine learning for multi-omics data integration. The study helps precision medicine with cutting-edge models. These models predict diseases and personalize treatments. It solves important genomics problems. The study addresses moral and privacy concerns. It ensures AI-powered healthcare innovations are implemented accountably. It ensures AI-powered healthcare innovations are implemented fairly.

Flowchart of the process

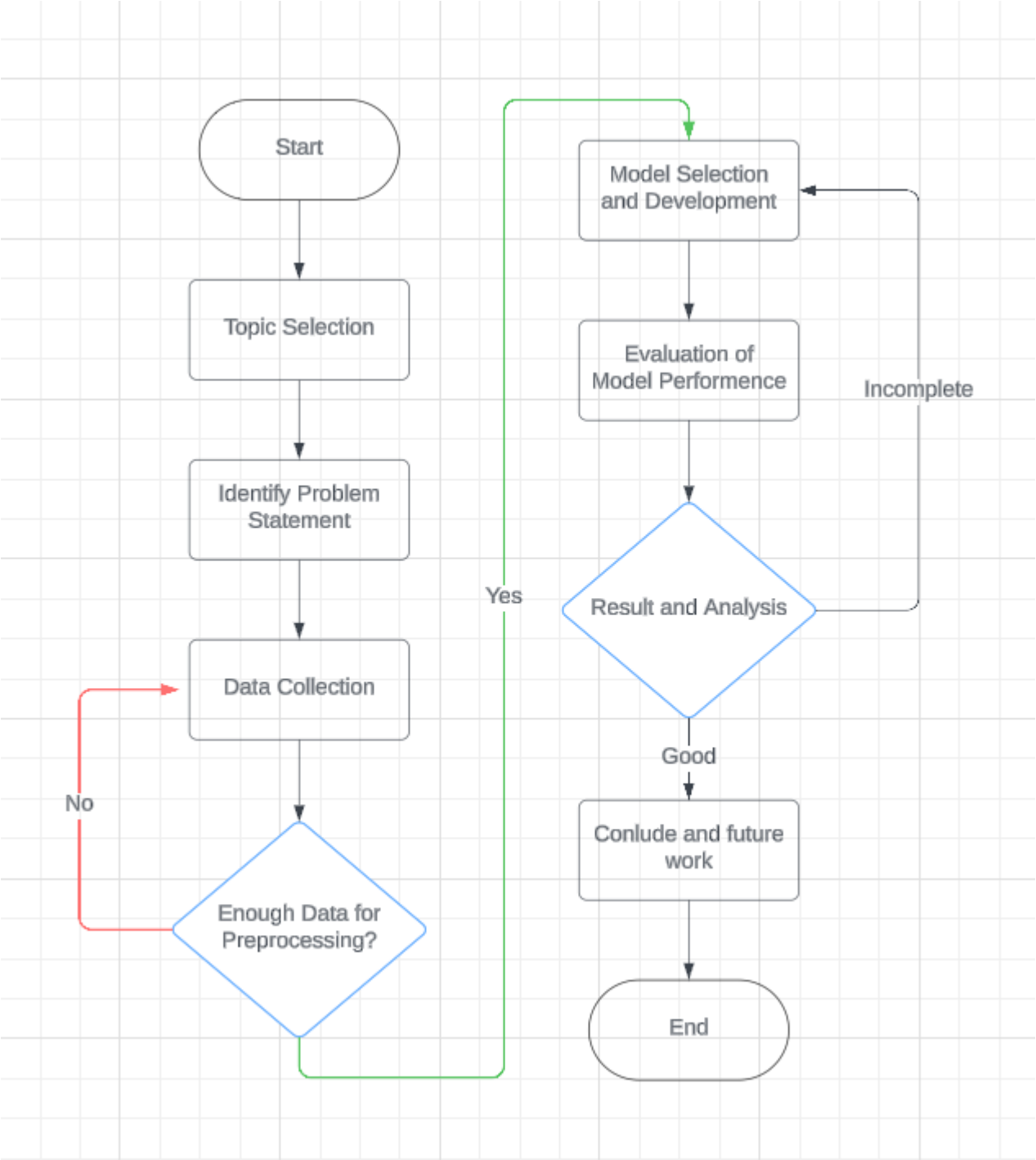

Fig 1: Flowchart of the process

Regression Testing:

For regression testing, a demo named genomics_pression_medicine_extended_dataset.csv was taken which has neutral data. After uploading and executing the file in Google Colab, our dataset was analyzed to some results. These results showed Genomic variant type, Pathogenicity, Treatment history and Response to treatment. The results could be interpreted from-

Patient_ID	Age	Gender	Disease_Status	Genomic_Variant	Variant_Type	Pathogenicity	Treatment_History	Response_to_Treatment	Drug_Name	Drug_Response_Score	recision_Medicine_Therapy_Applied	Outcome
P00001	41	Male	Healthy	KRAS	Point Mutation	Likely Pathogenic	Standard	No Response	Avastin	0.67	No	Progression-Free Survival
P00002	26	Male	Cancer	NaN	Point Mutation	Likely Benign	None	No Response	Avastin	0.73	Yes	Progression-Free Survival
P00003	70	Female	Diabetes	NaN	Point Mutation	Likely Benign	None	Complete Response	Keytruda	0.12	Yes	Progression-Free Survival
P00004	39	Male	Diabetes	TP53	Deletion	Benign	Chemotherapy	Partial Response	Nivolumab	0.87	No	Progression-Free Survival
P00005	73	Female	Healthy	BRCA1	Point Mutation	Benign	None	Complete Response	Keytruda	0.23	No	Progression-Free Survival
P00006	39	Female	Cancer	KRAS	CNV	Likely Pathogenic	Standard	Complete Response	Nivolumab	0.52	Yes	Progression-Free Survival
P00007	70	Female	Diabetes	KRAS	Point Mutation	Likely Pathogenic	Standard	No Response	Placebo	0.33	No	Progression-Free Survival
P00008	54	Male	Healthy	NaN	Point Mutation	Benign	Standard	No Response	Nivolumab	0.8	No	Progression-Free Survival
P00009	69	Male	Healthy	KRAS	Insertion	Benign	Chemotherapy	Complete Response	Avastin	0.9	No	Progression-Free Survival
P000010	75	Female	Diabetes	TP53	CNV	Likely Benign	Chemotherapy	Complete Response	Avastin	0.87	No	Overall Survival
P000011	32	Male	Healthy	TP53	Insertion	Likely Pathogenic	Chemotherapy	No Response	Nivolumab	0.74	No	Progression-Free Survival
P000012	27	Male	Cancer	TP53	Deletion	Likely Benign	None	No Response	Nivolumab	0.73	No	Overall Survival
P000013	52	Male	Diabetes	NaN	CNV	Likely Pathogenic	Chemotherapy	No Response	Avastin	0.2	Yes	Overall Survival
P000014	66	Female	Diabetes	TP53	Insertion	Likely Benign	Chemotherapy	Complete Response	Nivolumab	0.77	Yes	Progression-Free Survival
P000015	37	Male	Healthy	NaN	Point Mutation	Benign	Standard	Complete Response	Keytruda	0.38	Yes	Progression-Free Survival
P000016	37	Male	Healthy	NaN	CNV	Likely Pathogenic	Chemotherapy	Partial Response	Avastin	0.69	No	Progression-Free Survival
P000017	35	Male	Cancer	TP53	Deletion	Benign	Standard	Partial Response	Keytruda	0.85	No	Overall Survival
P000018	63	Male	Healthy	APOE	Insertion	Benign	Standard	No Response	Avastin	0.02	Yes	Overall Survival
P000019	75	Male	Cancer	BRCA1	CNV	Benign	Standard	Partial Response	Nivolumab	0.6	No	Progression-Free Survival
P000020	32	Male	Diabetes	NaN	Deletion	Benign	None	Partial Response	Keytruda	0.66	Yes	Overall Survival
P000021	68	Male	Cancer	TP53	Insertion	Benign	None	Complete Response	Nivolumab	0.01	No	Progression-Free Survival
P000022	24	Male	Diabetes	BRCA1	Point Mutation	Benign	None	Partial Response	Nivolumab	0.2	No	Progression-Free Survival
P000023	59	Male	Diabetes	KRAS	Insertion	Likely Benign	Standard	Partial Response	Nivolumab	0.08	No	Progression-Free Survival
P000024	49	Male	Healthy	BRCA1	Insertion	Benign	Standard	Partial Response	Placebo	0.53	Yes	Progression-Free Survival
P000025	73	Male	Cancer	KRAS	Insertion	Likely Benign	None	No Response	Keytruda	0.17	No	Overall Survival
P000026	35	Male	Healthy	APOE	Insertion	Likely Pathogenic	None	No Response	Placebo	0.61	Yes	Progression-Free Survival
P000027	59	Male	Diabetes	BRCA1	CNV	Likely Pathogenic	Standard	Complete Response	Placebo	0.2	No	Progression-Free Survival
P000028	76	Male	Healthy	NaN	Point Mutation	Benign	Chemotherapy	No Response	Avastin	0.1	No	Progression-Free Survival
P000029	49	Male	Cancer	NaN	Point Mutation	Benign	None	Partial Response	Keytruda	0.4	No	Progression-Free Survival
P000030	34	Female	Cancer	BRCA1	Deletion	Likely Benign	Standard	Complete Response	Nivolumab	0.95	No	Overall Survival
P000031	67	Male	Cancer	BRCA1	CNV	Likely Pathogenic	Standard	No Response	Keytruda	0.03	Yes	Overall Survival
P000032	71	Male	Healthy	APOE	Insertion	Likely Pathogenic	None	Partial Response	Nivolumab	0.08	No	Overall Survival
P000033	66	Female	Healthy	KRAS	Deletion	Likely Benign	Chemotherapy	Complete Response	Placebo	0.27	Yes	Overall Survival
P000034	64	Female	Healthy	KRAS	Point Mutation	Benign	None	Partial Response	Keytruda	0.87	Yes	Progression-Free Survival
P000035	38	Female	Diabetes	NaN	CNV	Benign	Standard	No Response	Keytruda	0.8	Yes	Overall Survival
P000036	48	Female	Cancer	TP53	Point Mutation	Likely Pathogenic	Chemotherapy	No Response	Avastin	0.95	No	Overall Survival
P000037	48	Female	Healthy	NaN	Point Mutation	Likely Benign	Chemotherapy	No Response	Nivolumab	0.76	No	Progression-Free Survival
P000038	33	Female	Cancer	NaN	Deletion	Likely Pathogenic	Chemotherapy	Complete Response	Nivolumab	0.33	No	Progression-Free Survival
P000039	22	Female	Cancer	TP53	Point Mutation	Likely Benign	Chemotherapy	Complete Response	Placebo	0.92	No	Progression-Free Survival
P000040	59	Female	Diabetes	APOE	Deletion	Benign	None	No Response	Placebo	0.76	Yes	Progression-Free Survival

Fig 2: Regression Testing

Regression Analysis:

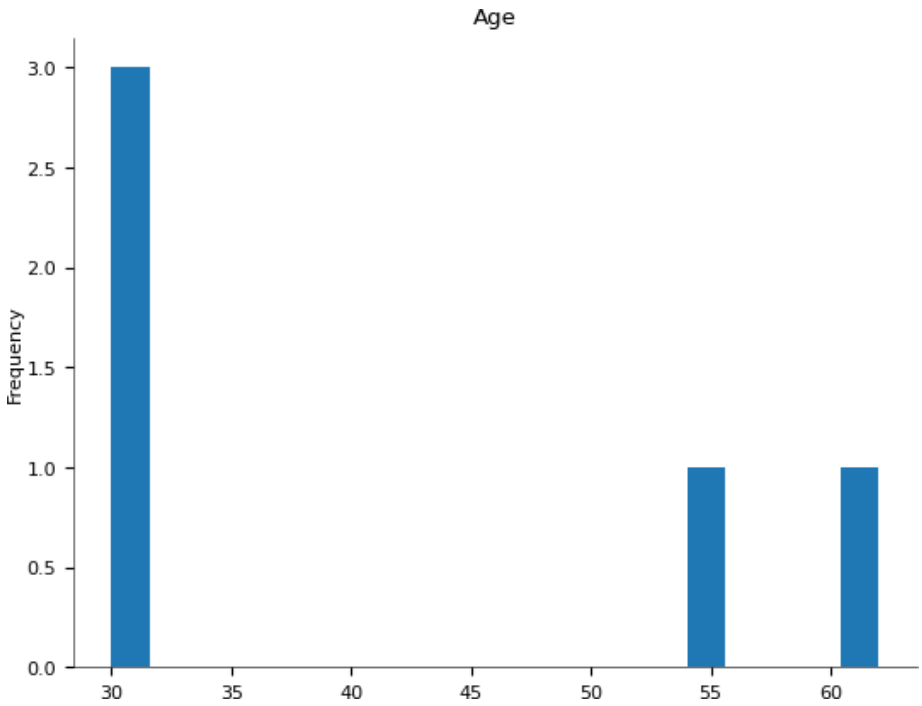

Fig 3: Age vs Disease rate

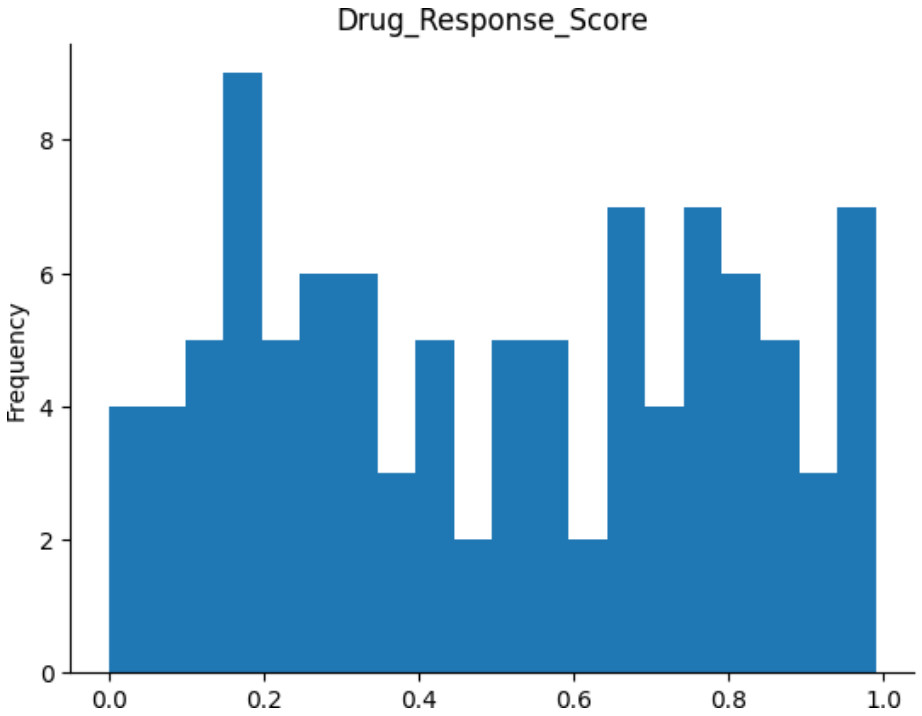

Fig 4: Response to Treatment vs Age

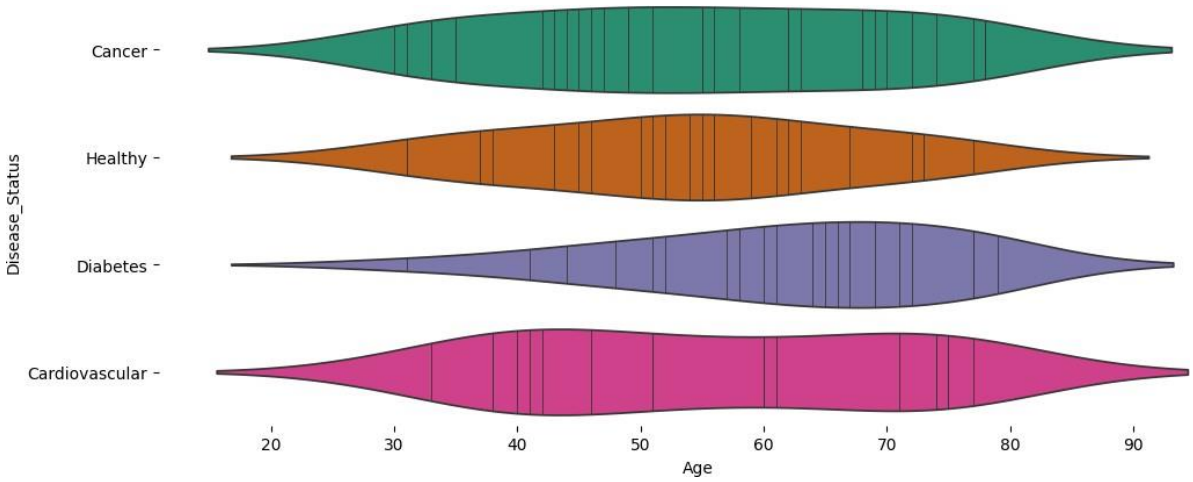

Fig 5: Disease vs Age

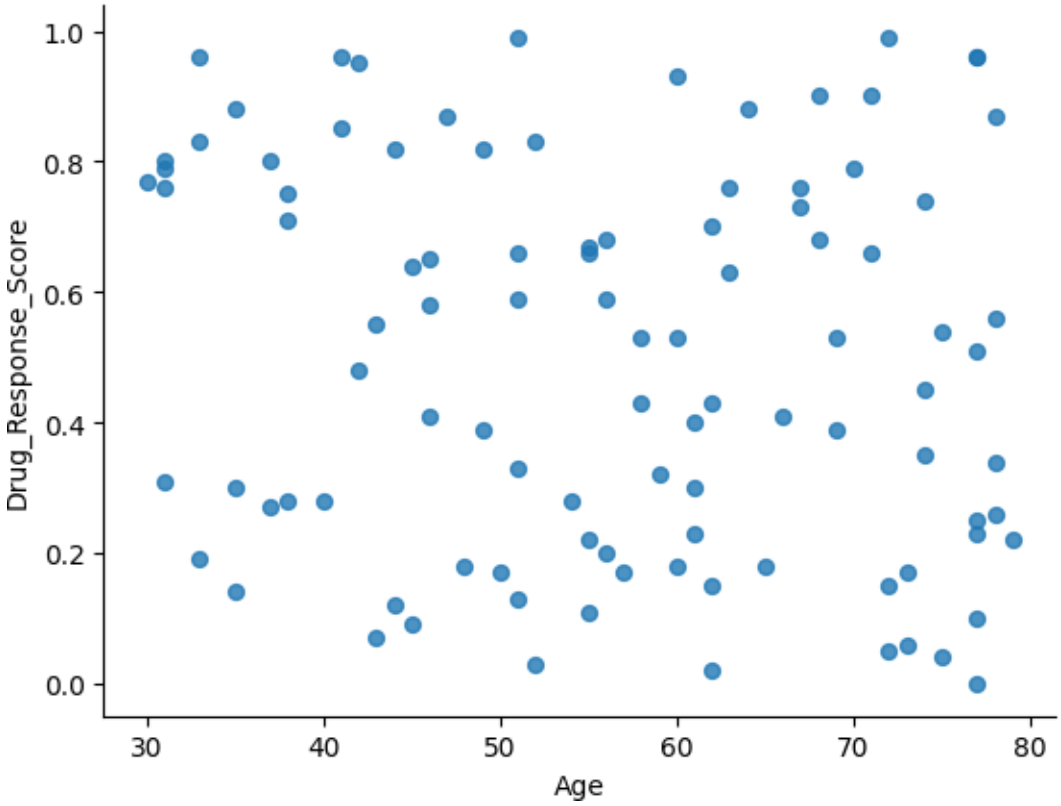

Fig 6: Drug Response Rate vs Age

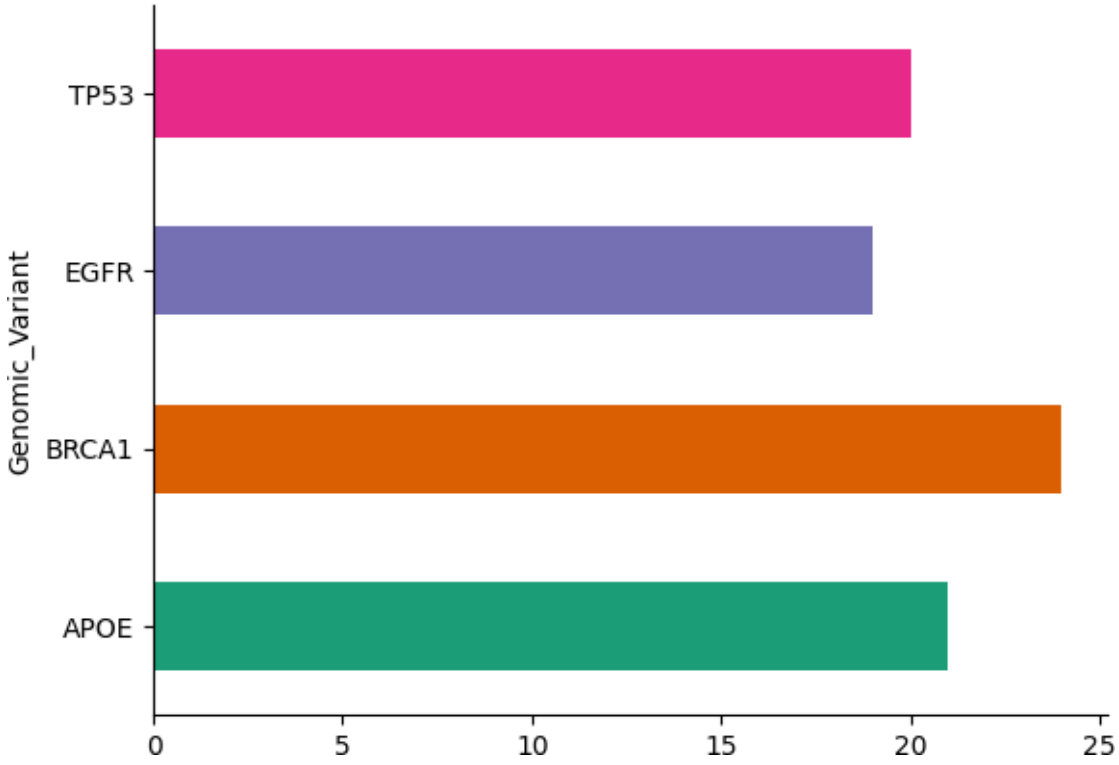

Fig 7: Age vs Genomic Variant

References

1. Quazi, S., et al. (2022). Artificial intelligence and machine learning in precision and genomic medicine. *Medical Oncology*, 39:120.
2. Rezayi, M., et al. (2022). Effectiveness of artificial intelligence for personalized medicine in genomics. *BioMed Research International*.
3. Oguz, C., et al. (2020). Machine learning applications in genomics: A review of current methods. *Nature Genetics*.
4. Alaa, A. M., et al. (2019). Cardiovascular risk prediction using machine learning models. *JAMA Cardiology*.
5. Coudray, N., et al. (2018). Classification and prediction in lung cancer using convolutional neural networks. *Nature Medicine*.
6. Huang, S., Yang, J., Fong, S., & Zhao, Q. (2020). Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. *Cancer Letters*, 471:61–71.
7. Poplin, R., et al. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*, 2(3):158–64.
8. Karczewski, K. J., & Snyder, M. P. (2018). Integrative omics for health and disease. *Nature Reviews Genetics*, 19(5):299–310.
9. Rajkomar, A., et al. (2018). Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, 1(1):1–10.
10. Sboner, A., & Elemento, O. (2016). A primer on precision medicine informatics. *Briefings in Bioinformatics*, 17(1):145–53.
11. Mesko, B. (2017). The role of artificial intelligence in precision medicine. *The Medical Futurist*.
12. Cruz, J. A., & Wishart, D. S. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics*, 2:117693510600200030.
13. Schork, N. J. (2019). Artificial intelligence and personalized medicine. *Cancer Treatment and Research*, 178:265–83.

Afsana Jahan Onu
CSE
American International University Bangladesh.
Dhaka, Bangladesh
21-45708-3@student.aiub.edu

Fatema Akter Sujana
CSE
American International University Bangladesh.
Dhaka, Bangladesh
21-45693-3@student.aiub.edu

Mussa Alam
CSE
American International University Bangladesh.
Dhaka, Bangladesh
21-45036-2@student.aiub.edu