International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

CANCER GENE DETECTION & DIAGNOSIS

Prof. Vaishali Gedam [1], Mr. Harshal Uikey[2], Mr. Sumit Sundergiri[3], Mr. Prashant Katre[4]
Mr. Manoj Shahare[5], Mr. Aman Patil[6]

Prof. Vaishali Gedam, Computer Science and Engineering, Nagpur Institute of Technology, Nagpur
Mr. Harshal Uikey, Computer Science and Engineering, Nagpur Institute of Technology, Nagpur
Mr. Sumit Sundergiri, Computer Science and Engineering, Nagpur Institute of Technology, Nagpur
Mr. Prashant Katre, Computer Science and Engineering, Nagpur Institute of Technology, Nagpur
Mr. Manoj Shahare, Computer Science and Engineering, Nagpur Institute of Technology, Nagpur
Mr. Aman Patil, Computer Science and Engineering, Nagpur Institute of Technology, Nagpur

***

Abstract - The sequence of excrescence of cancer controls thousands of inheritable mutations. Now the most grueling task is to separate between the mutations which will further contribute to cancer growth and mutations. Interpretation of inheritable mutations is manually done presently, which consumes a lot of time and may also lead to a squishy opinion that isn't tolerable in the healthcare sector. Clinical molecular biologists have to manually review textbook substantiation of clinical exploration literature for every single inheritable mutation. Machine Learning (ML) helps in the precise and fast opinion of a complaint and leads to effective decision-making. Once the excrescence is detected we go for testing whether it's cancerous or noncancerous. However, it goes for a gene panel test which takes a much longer time which is 3- 4 weeks, So using the patient's former medical records and using a vivisection report we're detecting for which gene he/she is positive, If set up to be cancerous. Using our ML model we will help cases in early diagnosing which will also help Croaker to go with the following treatment. It takes so long time for generating the gene panel cancer report roughly 3-4 weeks.

In these 14 days period, cancer excrescence can surpass stage 3 or stage 4 which is veritably parlous and occasionally may indeed lead to death.

Using colorful machine literacy models and after resolving all business constraints this process will be done as soon as possible.

***Key Words*:** Machine Learning, Machine Literacy, Cancer, Cancer Gene Detection, Mutation

1. INTRODUCTION

Cancer Gene Detection and Diagnosis is a major project that focuses on detecting and diagnosing cancer using gene analysis techniques. Cancer is a leading cause of death worldwide, and early detection and diagnosis are crucial for the effective treatment and management of the disease. In recent years, advances in genomics and bioinformatics have revolutionized our understanding of cancer biology and enabled the identification of cancer-causing genes and mutations. The Cancer Gene Detection and Diagnosis project aims to leverage these advances to develop new tools and methods for detecting and diagnosing cancer at an early stage. The project will involve the use of advanced gene analysis techniques such as next-generation sequencing, microarray analysis, and gene expression profiling to identify the genetic alterations that are associated with cancer. The project will also involve the development of new algorithms and computational tools to analyze the large amounts of genomic data generated by these techniques. These tools will enable researchers to identify specific genes and genetic mutations that are associated with cancer and to develop personalized treatment plans for individual patients based on their unique genetic profiles. The dataset has nine different output classes, some corresponding to cancer, and some not. The input gene mutations (small changes) data must be classified into one class. A standard is defined to split the dataset in the train, CV, and test ratio. The size of the training dataset should be large as ML models learn on the train data and data points should be stable across the train, CV, and test split also called "Overlapping".

The higher the overlap, the higher the stability. The different ML algorithms are applied, and performance is evaluated on metrics such as AUC, log loss, sensitivity-specificity, precision-recall, and many more. The lower the numerical log loss value the higher the efficiency. In the ideal case, the log loss value is 1. Results vary on different disease dataset features. Consequently, standard steps are followed as data collection, data reading, pre-processing of text data, feature extraction, and then training models with the most important features to get higher prediction efficiency. Applying the emerging approach of Big Data Analytics to the healthcare sector will improve healthcare services. Overall, the Cancer Gene Detection and Diagnosis project has the potential to make a significant contribution to the field of cancer research and to improve the lives of millions of people around the world who are affected by this devastating disease.

.

2. LITERATURE REVIEW:

Related Work:

In [1], Machine learning applications in cancer prognosis and prediction authored by Konstantina Kourou, Themis P. Exarchos, Konstantinos P. Exarchos, Michalis V. Karamouzis, Dimitrios I. Fotiadis.

We bandied the categorization of cancer as a miscellaneous complaint conforming to many different subtypes. The early opinion and prognostic of a cancer type have come a necessity in cancer exploration, as it can grease the posterior clinical operation of patients. The significance of classifying cancer cases into high or low-risk groups has led numerous exploration brigades, from the biomedical and the bioinformatics field, to study the operation of Machine Learning (ML) styles. thus, these ways have been employed as an end to model the progression and treatment of cancerous conditions. In addition, the capability of ML tools to descry crucial features from complex datasets reveals their significance. A variety of these ways, including Artificial Neural Networks (ANNs), Bayesian Networks (BNs), Support Vector Machines (SVMs), and Decision Trees(DTs) have been extensively applied in cancer exploration for the development of prophetic models, performing in effective and accurate decision timber.

In [2], Applications of Machine Learning in Cancer Prediction and Prognosis by Joseph A. Cruz, David S. Wishart.

We studied a number of trends noted in cancer vaticination, including a growing dependence on protein biomarkers and microarray data, a strong bias towards operations in prostate and bone cancer, and a heavy reliance on " aged " technologies similar to artificial neural networks (ANNs) rather of more lately developed or more fluently interpretable machine literacy styles. A number of published studies also appear to warrant an applicable position of confirmation or testing. Among the better-designed and validated studies, it's clear that machine literacy styles can be used to mainly (15- 25) ameliorate the delicacy of prognosticating cancer vulnerability, rush, and mortality. At a more abecedarian position, it's also evident that machine literacy is also helping to ameliorate our introductory understanding of cancer development and progression.

In [3], Technologies for deriving primary tumor cells for use in personalized cancer therapy Authored by MitraA1, Mishra L, Li S.

This review focuses on our current understanding and the pros and cons of different styles for primary excrescence cell culture. likewise, colorful culture matrices similar to biomimetic pulpits and chemically defined media supplemented with essential nutrients, have been prepared for different tissues. These well-characterized primary excrescence cells review cancer curatives with high translational applicability.

In [4], Breast Cancer Prediction and Detection using Data Mining Classification Algorithms a Comparative Study.

This paper aims to prognosticate and descry bone cancer beforehand with non-invasive and effortless styles that use data mining algorithms.

3. ADVANTAGES & CONCLUSIONS

- **Improving early detection:**

  By identifying genetic mutations associated with cancer, a project focused on cancer gene detection and diagnosis could help improve early detection of the disease. Early detection is critical for effective treatment and improving patient outcomes.

- **Personalizing treatment:**

  Genetic testing can help identify specific gene mutations or biomarkers that may indicate a patient's response to certain treatments. By personalizing treatment plans based on genetic information, patients may experience improved outcomes and fewer side effects.

- **Developing targeted therapies:**

  Genetic testing can also identify potential targets for new cancer therapies. By targeting specific gene mutations or biomarkers, researchers may be able to develop more effective and less toxic treatments.

- **Identifying new cancer genes and biomarkers:**

  As genomic sequencing technologies continue to advance, a project focused on cancer gene detection and diagnosis could help identify new cancer genes and biomarkers. This information could lead to the development of new diagnostic tools and treatments.

Overall, a project focused on cancer gene detection and diagnosis has the potential to make significant contributions to cancer research and healthcare, ultimately leading to improved patient outcomes and better overall cancer care.

ACKNOWLEDGEMENT

We hereby grateful to our project Guide
Prof. Vaishali Gedam Ma'am, for kind support and guidance in sharing technical Knowledge and mentoring us at every stage.

REFERENCES

1. Ng PK, Li J, Jeong KJ, et al. Systematic functional annotation of somatic mutations in cancer. Cancer Cell. 2018;33(3):450- 462.e10. doi: 10.1016/j.ccell.2018.01.021

2. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-421. doi: 10.1038/nature12477

3. Bailey MH, Tokheim C, Porta- Pardo E, et al. Comprehensive characterization of cancer driver genes and mutations. Cell. 2018;174(4):1034-1035. doi: 10.1016/j.cell.2018.07.034

4. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324

5. Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, Liang Y, Rivkin E, Wang J, Whitty B, Wong-Erasmus M, Yao L, Kasprzyk A. International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data. Database (Oxford). 2011;2011:bar026. doi: 10.1093/database/bar026

6. Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: an integrative exploration of genetic variation and genome annotations. PLoS Comput Biol. 2013;9(7): e1003153. doi: 10.1371/journal.pcbi.1003153

7. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16): e164. doi: 10.1093/nar/gkq603

8. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31(3):213-219. doi: 10.1038/nbt.2514

9. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-26. doi: 10.1038/nbt.1754

10. Griffith M, Spies NC, Krysiak K, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet. 2017;49(2):170-174. doi: 10.1038/ng.3774