Ilkogretim Online - Elementary Education Online, Year; Vol 20 (Issue 5): pp. 1109-1117
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.05.120

The Role of Machine Learning in Cancer Genome Analysis for Precision Medicine

Dr. Chandra J , Associate Professor, Department of Computer Science,CHRIST (Deemed to be University)
Nithya V, PG Scholar, Department of Computer Science, CHRIST (Deemed to be University)
Alwin Joseph, PG Scholar, Department of Computer Science, CHRIST (Deemed to be University)
Dr.M.Vijayakumar, Professor, Department of MCA, K.S.R.College of Engineering, Tiruchengode, Tamilnadu, India
Dr. Siddharthan S, Associate Professor, Department of Life Sciences, CHRIST (Deemed to be University)

Abstract- Genome Analysis, the process of understanding the genome has a transformative impact on healthcare. There are different treatment methods for cancer but are not effective for all individuals, some treatment plans can bring a huge negative impact on the patient's health. Thus the importance of precision medicine comes into the picture. Precision medicine is a technique that focuses on providing the patient with the appropriate treatment measures, along with genome analysis with the help of machine learning can create customized treatment plans for the patients. This intelligent method of treatment planning can create customised treatment plans for cancer patients. Here, treatment will be given by identifying the specific DNA/RNA sequence and analyse them for an individual, to avoid the negative impacts during the treatment. The information obtained from the patient genome is analysed and is utilized to prepare the customised medication for the patient. Applying machine learning techniques on the genome and different scan images help to identify the severity of cancer in the patient, in order to plan the treatment. The outcome of the analysis of cancer genome with machine learning along with different scan details is to collect the right information which helps the doctors to address the issues effectively for identifying the problems with patients and to prepare the customised personal treatment plans.

Keywords: Cancer Genome Analysis, Precision Medicine, Machine Learning, DNA, RNA sequencing, Efficient Treatment planning, Customised Medicine, Customized Treatment.

I. INTRODUCTION

Health is a major concern for the current generation; cancer is one of the major diseases which is not having an appropriate treatment. Based on the results of a survey released by the World Health Organization, cancer is the second leading disease that causes death globally. In 2018, universally around 9.6 million individuals died due to various cancers and 18.1 million new cases have been estimated. As part of the survey, WHO could identify that, one in 6 women and one in 5 men suffer due to cancer, and one in every 11 women and one out of 8 men deaths occur due to Cancer. The number of individuals alive within five years of a cancer diagnosis is estimated to be 43.8 million [1]. In India, cancer-related deaths are estimated to be around 7,84,821 and 11,57,294 new cases have been registered every year. The individuals currently diagnosed with cancer are estimated to be around 2.25 million. The statistics of India related to cancer are passing an alarm. Since there is a death of one woman occurring every 8 minutes due to cervical cancer. It is also discovered that one woman dies of every two women suffering from Breast cancer. Before the age of 75 years, the chances of dying from cancer is 7.34% in men and 6.28% in women [2]. These statistics give the importance of the precise methods that are to be implemented for the treatment of cancer.

Precision medicine is an emerging field in the health sector for treating various diseases knowing the gene variations. The main objective is to reduce the adverse reactions the individual faces during the treatment phase especially in the case of cancer. The approach of precision medicine will help physicians and researchers to discover new drugs and predict which treatment will be beneficial for each individual. In general, Cancer patients are given many treatments such as chemotherapy, radiotherapy, hormone therapy, and surgery. The usage of various prescriptions prompts numerous unfriendly responses and in most extreme events even by taking these endorsed medicines, the sickness cannot be cured. Hence precision medicine can be the solution, by knowing which specific Deoxyribo Nucleic Acid / Ribonucleic Acid (DNA/RNA) sequence is affected for which individual, precise drugs can be provided accordingly.

This is achieved through genome analysis. Recently, there has been a remarkable increase in the human genome data and this information of the human genome can be utilized for precision medicine to create a healthier environment [3]. Currently, genome sequencing is having a very big impact on different treatments especially in cancer treatments, identification of various genetic diseases, and also the information regarding the individual's likely response to treatment. The treatment planning and fast recovery from cancer can be effectively achieved through precision medicine. There is no precise treatment given to cancer patients due to which the low level in the survival rate of cancer patients, the effect of a treatment method can be forecasted for a cancer patient using machine learning techniques in genome analysis, thus the doctors can create a set of customized treatment strategy for the patient based on his physical conditions.

Genome analysis in the field of bioinformatics, that helps in discovering and comparing the genetic features and to study the variations in the gene expressions, structure, and functions with the support of modern technologies such as Artificial Intelligence, Machine Learning [4]. Genome analysis with the help of the emerging machine learning techniques can predict the possible modifications in the human genome. Genome analysis can help to identify the DNA sequence of any given plant or animal species. The output of genome sequencing provides many opportunities to understand and change the genetic structure of that species [5]. Different genome sequences can be related to understand how a particular genome has evolved. This helps the doctors determine cancer in the very early stages and treat it efficiently with precision medicine, which will lessen the reactions of the patient's experience during the malignancy treatment [6].

Machine learning is an integral part of Artificial Intelligence, which helps the system to learn automatically from the past experiences and improve [7]. Machine learning can be employed to detect and learn the mutations easily, it can be applied to classify the mutations and the treatment plans and constantly improve based on the new and past patient records. Machine Learning when integrated with genomics the processing speed of genetic data and predicting effective results from it can be done much faster. It can also be used to predict future diseases and help in discovering the new drug concerning individual genes. The main objective behind integrating machine learning and genomics is to improvise the time and reduce the cost.

Globally many countries have invested millions in genome analysis projects [8]. Those countries are the United States, England, Japan, China, Australia, Saudi Arabia, France. The main aim of this research is to create awareness about the genomics and precise medicine and its benefits to the audience of those countries where this technology is not familiar. In the present situation of the COVID-19 pandemic, genome analysis can be very helpful for inventing the vaccine and treating the patients effectively.

II. LITERATURE SURVEY

Jianting Sheng et al., [9] has proposed the drug response prediction of the cancer patient using a novel computational algorithm. Cancer patients show many adverse side effects for the anticancer drug given during the treatment. Genomic profile, as well as pharmacogenomics and drug sensitivity data of the individual cancer patients can assist with identifying the genetic variations which are mainly responsible for these adverse responses. The data of 600 cancer patients who had 50 unique types of cancer and the way they reacted to 75 drugs were gathered from the cancer database of genomics of drug sensitivity. After applying the novel computational algorithm on the data, the individual reactions to the drug were gained from this process. As the outcome, the most appropriate drug for that person's DNA/RNA was anticipated.

Yifan Zhang et al., [10] has developed the model for the breast cancer survival prediction with precision. The most widely recognized cancer in females is Breast cancer and it causes around 18.2% of deaths globally. Hence, it becomes very important to predict the survival rate of these patients. The improvisation of survival prediction of breast cancer can be done utilizing the prognostic model including the somatic mutations with gene expression. During the process 118 genes were recognized which can be utilized for building the breast cancer survival predictor model. The enrichment analysis was applied to the mutated gene and successfully identified the 15 gene ontology, 25 functional annotations which were resulting in cancer. Hence, the specified model can be used in simplifying precision medicine.

Xi Chen et al., [11] has developed Bayesian framework utilizing both tumor and normal samples from read alignment data, called BSSV. The studies from various sources show that when the paired samples were analyzed together give better recognition of Somatic Structural Variation. The tools which are accessible are not equipped for recognizing all Somatic Structural Variations, hence the precision level of BSSV is comparable to the tools which are available in the market and along with lowered false-negative rate. The above technique was applied to Cancer genome Atlas for Breast cancer the mutated genes like RAD51, BRIP, ER which cause breast cancer were effectively identified.

Y. Hsu and D. Si [12] have proposed a model of classifying RNA-sequencing data and predicting the type of cancer. Analysis of cancer has been a major topic of discussion for many researchers in recent times. By using machine learning algorithms for investigating the cancer genome atlas data can be suggested as the best solution for the queries related to cancer. By utilizing the RNA data and performing the analysis was aimed to classify cancer patients of thirty three types. The various machine learning algorithms were used in this process to make a comparative study based on their training time, accuracies, precision recall and F1 score. Linear support vector machine, polynomial support vector machine, k nearest neighbor, decision tree, artificial neural network was utilized in this model. The outcome shows that, with the accuracy rate of 95.8%, a linear support vector machine is the best classifier.

Dan Li et al., [13] has proposed Comprehensive Analysis of Pulmonary Adenocarcinoma in Situ (AIS) to understand lung cancer in further detail. The five year survival rate of these cancer patients is only 4%. Genome evaluation has to be performed in order to completely recognize the disease patterns, for this purpose the RNA sequencing data of normal and AIS patients are used to identify the gene expressions which make AIS patients different from normal. As the outcome 72 AIS causing genes and 41 protein-coding genes were identified. Among them twelve of protein coding genes were the lung cancer driver genes and the important features which are contributing to the early diagnosis were also identified.

Matthieu P. Schapranow et al., [14] proposed a drug response analysis using in-memory technology. As the medical field is growing which will in turn generate a huge medical data which can be called as big medical data. Many tools were used to process all these data so that it can be used by experts of health care and researchers. The main objective is to combine all genome sequencing data and its analysis in the process of precision medicine. The experimental results of drug response analysis using software tools which are built on in memory technology for processing the big data and integrating heterogeneous data sources were analyzed.

A. F. G. Taktak et al., [15] has proposed a nonlinear model role and efficiency in the decision decision-making. In few cancers the genome abnormalities play a major role, whereas in many cancers even this does provide any indication of cancer. The main difficulty is the repetition of the genome and proteome. In this process there can be a vast number of feedback loops and interconnections. In order to model these systems, there is a requirement of mature analysis and mathematical techniques. Hence neural networks can be used in this research work. By adding the Bayesian function to the stage of output can predict the survival probability. The two types of cancer were predicted through this analysis. Laryngeal carcinoma has a rate of medium survival, short risk time and ocular melanoma which has a rate of high survival and longer risk time.

Terry H. Tsai et al., [16] has proposed translational medical studies using Text mining for hypothesis and results. Text mining is usually performed on the name entity but not many tools in the market are developed for genetic studies. The complicated diseases such as cancer and diabetes are caused due to the variation in the gene expressions. To achieve the aim of translational research, gene variation must be considered. The translational medical study in the paper supports both machine learning text mining and automated interpretation for evaluating the key factors. After the process of scaling the precision level of 64% with the range of 48% to 80% were achieved.

Kiichi Fukuma et al., [17] has proposed a study on Disease Stage Classification and feature extraction for Glioma Pathology Images. For precision medicine computer aided diagnosis are very crucial. Glioma is one of the critical brain tumours and histopathological tissue images can play a vital role in evaluating the inner details and to identify the disease stages. This process used image analysis and extracted the feature and classified the image data collected from cancer genome atlas. The accuracy of the classification was checked using a support vector machine and random forests. The outcome showed the accuracy of 98.9% and 99.6 % for the specified machine learning algorithms.

Shuai Li et al., [18] has proposed a framework to map the features between gene expression profile and medical images and their correlations are evaluated. Precision medicine is one of the important strategies in the treatment, to improve the precision medicine, the medical images are converted to high dimensional data by extracting few mathematical features by the researchers. In this process, they utilized the methods of deep learning using which lung cancer was identified both at genome level as well as image level. The association was derived between the gene and tumour growth and directly the visual results were provided for clinical research. The same methodology can be used for identifying various tumor phenotypes and results.

Sheida Nabavi et al., [19] proposed a model of identifying the candidate biomarkers which are resistant to anticancer drugs in ovarian cancer using statistical machine learning. The biggest objection in treating ovarian cancer is drug resistance. In order to identify the right biomarkers for the platinum-based chemotherapy in ovarian cancer, both genome analysis and machine learning techniques are used. Utilizing regression tree and network analysis the genes were identified which had the mutation and also their gene expressions were impacting the other co regulated genes for tumours and resistance. Lastly, two genes list obtained were correlated with each other, in order to obtain the potential biomarkers of a small list which are resistant to platinum-based chemotherapy.

Vitor Teixeira et al., [20] has developed DenoisingAuto Encoders (DAE) for deriving the driver genes from the gene expression dataset of cancer. The nodes of input and final representation of the DAE were evaluated. The deep learning methods of this model were correlated with other existing methods. There are 2 main tasks in the model proposed, firstly the model was built and utilized classifiers and correlated many feature extraction methods and sampling methods, this model was able to successfully differentiate between the thyroid cancer patients and healthy people. Secondly, comprehensible description extractions of gene expression were performed. The accuracy for the first task is quite high whereas the second task is quite low.

Sanvesh Srivastava et al., [21] proposed a Hierarchical Bayesian Relevance Vector Machines (H-RVM) for combining the multiple platform genome data. In H-RVM Hierarchical Bayesian generalization is used, along with the Hierarchic kernel learning framework. Glioblastoma data from the cancer genome atlas was tested on the H-RVM, to predict the tumor volume based on the imaging and also by combining the miRNA data and gene expressions. The outcome shows that H-RVM is the best option for prediction when related to other methods in the market.

Gianluca Bontempi [22] has proposed a blocking strategy to improve feature selection for classification of expressed gene data. The aim of this research work is to select the relevant gene subsets the one which can return best generalization in classification. The blocking strategy is used to be very confident in the level of accuracy that the result is because of the actual differences and not for the effects of noise. This strategy is basically for improving the feature selection. The results for the set of 16 available cancer gene expression data sets using this blocking strategy improves the performance of the forward selection.

Gabriela Alexe et al., [23] has proposed diagnosis of cancer from the gene expressed data using a robust meta-classification strategy. The main objective of this paper is to develop a classification model which can be independent of analyzing techniques and which can integrate the data from multiple sources such as laboratories. The proposed strategy is a Meta classification approach which follows the multi gene selection procedure and combines the outcomes of several tools of machine learning which are trained on raw and pattern data. When this method was applied to 2 separate datasets to differentiate between B-cell lymphoma from follicular lymphoma. The results showed the predictive accuracies were higher using meta classification technique rather than training the individual classifier on the same dataset.

Ahmed T. Soliman and Mei-Ling Shyu [24] developed a multi model-based driver missense identification using machine learning techniques. Is there increase in sequencing techniques of DNA and protein and this has promoted the accessibility of the mutations data, hence there is great need of tools which can automatically predict the driver mutations. These mutations not only help us to understand the cancer better but also to measure the progression of the cancer over time, along with this, it can also suggest the target treatment and also proper analysis of the cancer can be performed. The model consists of independent parallel classifiers and where a single set of features are handled by the single classifier. Lastly a fusion module integrates all these classifiers and the output of the final mutation label is produced. The model is trained and also validated with the feature set.

Ka Yee Yeung [25] has proposed a signature discovery method for analyzing personalized medicine. As the data generated from the genome analysis is very huge, hence this prompts the challenges of computation on this data. Many tools are developed to address this issue. This model proposes a signature discovery along with machine learning algorithms in which the model is built with the subsets of variables. This research paper focuses on two main applications such as evaluating the signature gene and predicting the disease which has occurred and the networks of interpretation. This information obtained can be very helpful in treating the disease with precision medicine.

The analysis of the human genome can help in identifying mutations which are responsible for the cancer, which are called as signature genes and driver genes. The genome analysis with the help of the emerging machine learning techniques can predict the possible modifications in the Human genome. This helps the doctors determine cancer and various genetic disorders and treat them accordingly with precision medicine, which will lessen the reactions the patient's experience during the malignancy treatment.

III. METHODOLOGY

Precision medicine is an advancement in the health sector which can help to identify the fingerprints and evolution of any disease such as cancer through the complete analysis of the genome. The information obtained from the gene expression data can help in identifying the characteristics of that gene. The identification of characteristics has numerous benefits such as grading the stage of cancer, planning the treatment, predicting the survival period and the appropriate drug which can be used in the treatment, in order to avoid the adverse reactions caused by other treatments. This will in turn reduce the hardships the cancer patients usually undergo during the malignancy treatment.

The following flow diagram is the representation of overall workflow of the genome analysis, which is used to identify the genes and process the DNA structure of an individual, which is a major part of creating precise medicine for individuals.

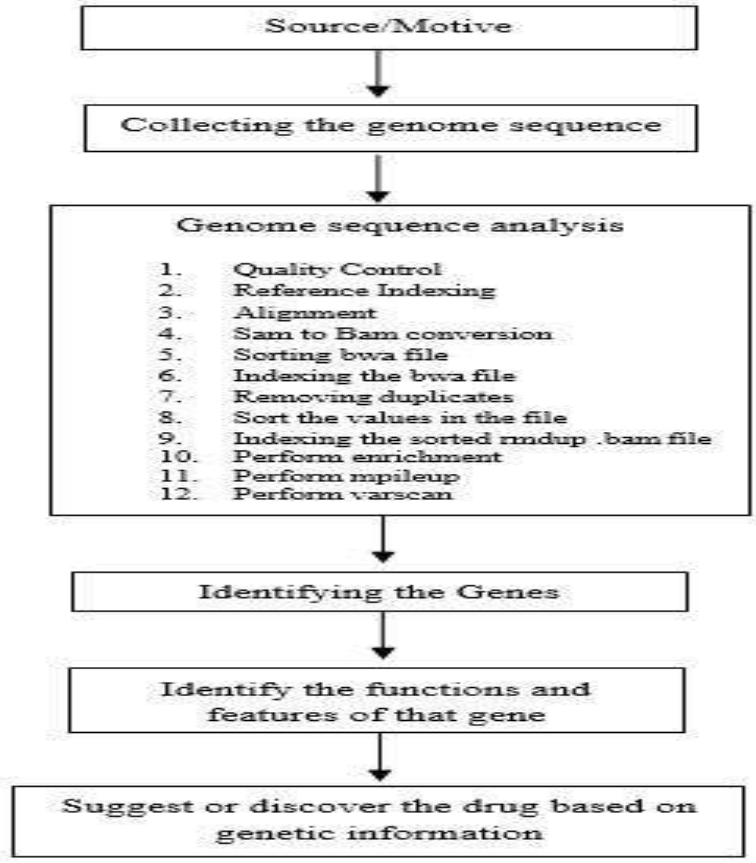

Fig. 1. Workflow of Genome Analysis

Fig 1, depicts the workflow of the genome analysis in a simple manner. The process begins with the source or motive for which the genome is being sequenced, for example, to identify the gene which is causing cancer can be one of the motives. After knowing the main source or motive the appropriate genome sequence is collected to perform further analysis. After obtaining the genome sequence, the analysis is performed by utilizing various tools. The final step of the analysis provides the gene id, through which the particular gene which is responsible for any diseases which the individual is suffering from is identified and further the function and features of that particular gene are studied, and finally the drug is suggested or discovered for the patients. The steps involved in genome data processing are quality control, reference indexing, alignment, enrichment and varscan.

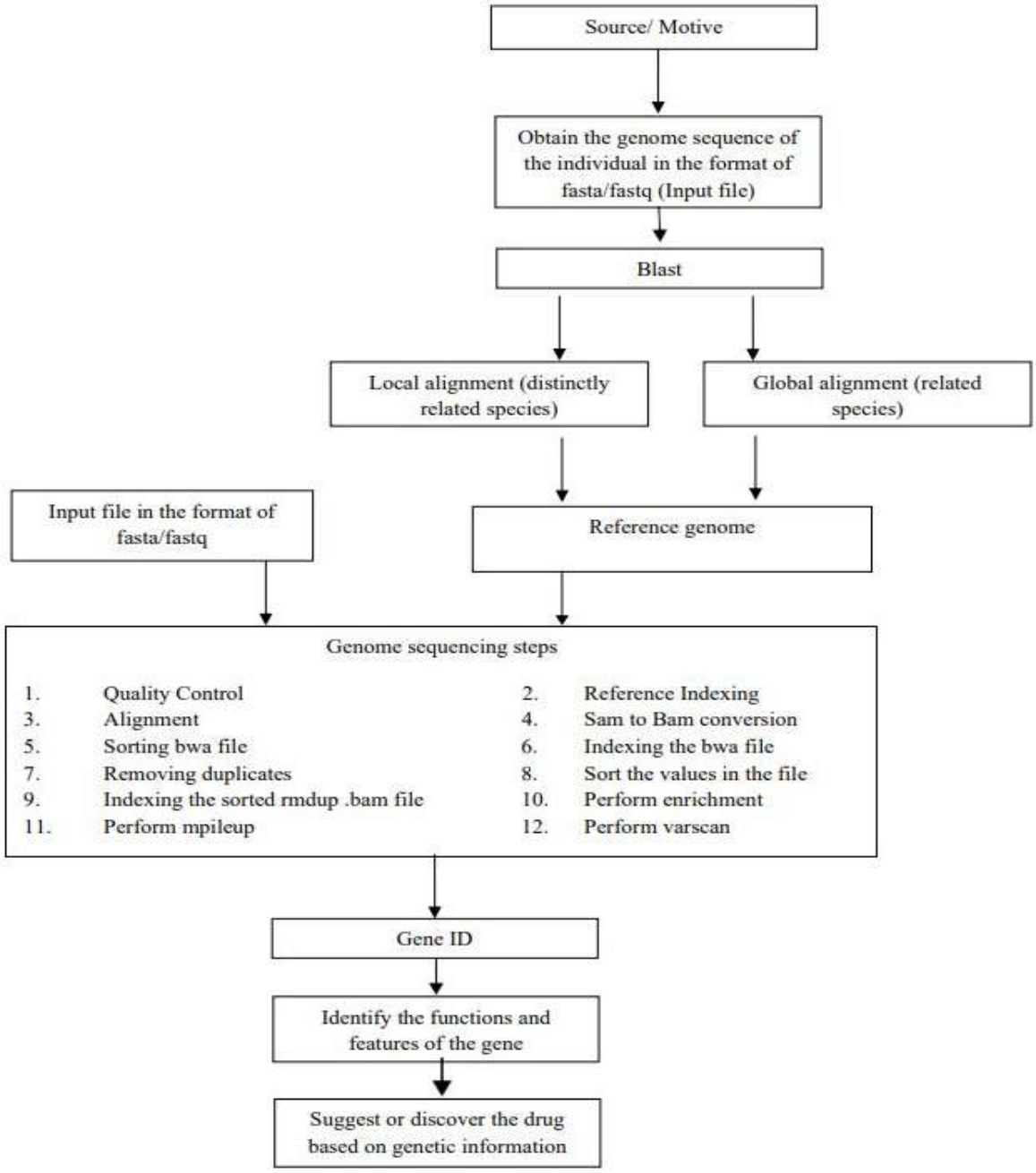

Fig. 2. Flow of Machine Learning based Cancer Genome Analysis for Precise Medicine.

The proposed flow of the solution is shown in Fig 2. The model mainly concentrates on improving the health of the cancer patients by providing precision medicine which reduces the stress and side effects the cancer patient has to undergo. This model can also be used to identify genetic disorders, various other diseases, and provide precise treatment. Based on the cause, the input fasta/fastq file of an individual is taken and blast is performed. Blast is the search tool that compares the sequence of our input fasta/fastq

file with the available sequences across genomes , to identify the right reference genome to proceed. In Blast, there are two type's local alignment for distinctly related species and global alignment for related species. After Blast, the Reference Genome obtained after performing the local alignment or global alignment along with the Input fasta/fastq file goes through many genome sequencing steps, the genome sequencing steps requires various stages and tools, which finally produce the output of gene Id. Based on the Gene Id the features and functions of the gene are studied. And lastly, the most appropriate drug is suggested or discovered for the particular individual.

Machine Learning techniques can be used to simplify the sequencing process [26]. After obtaining the gene ids, each gene ids feature and functions are studied individually, which is the tedious task. This process can be automated by creating the search tool with the help of Machine Learning and different other artificial intelligence tools, which can identify and provide all the features and functions of that gene. Finally, when the gene functions are known, the drug can be suggested based on the information obtained. The accuracy of the suggested drug can be gathered using any appropriate machine learning technique [27].

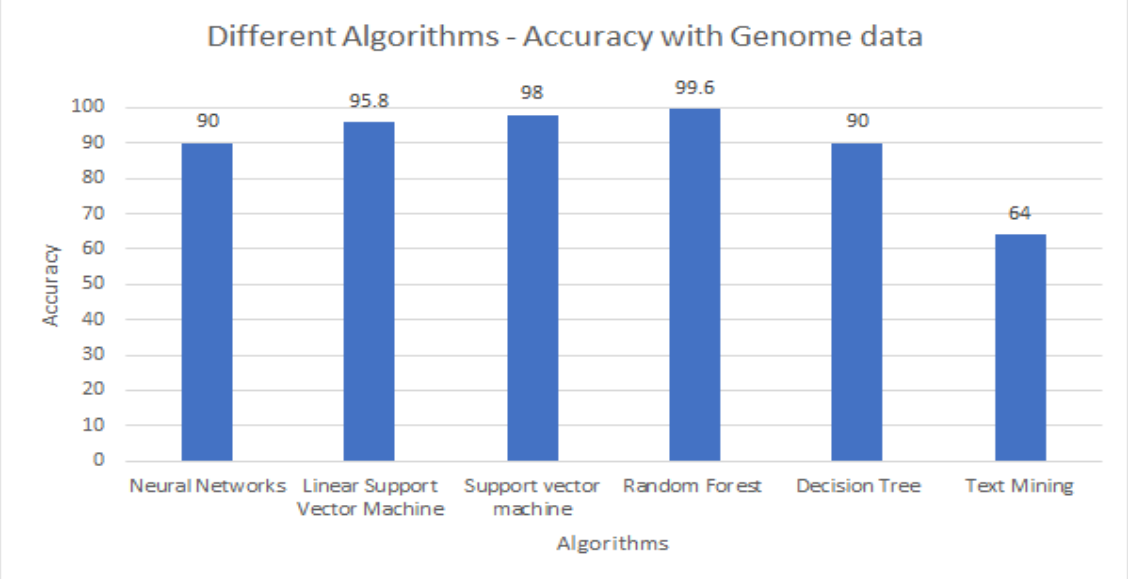

Fig. 3. Algorithms comparison chart

The Fig.2 depicts the accuracy level of various Machine Learning algorithms with the genome data.The accuracy level varies for each Algorithm as shown in the figure.There are various Machine Learning Algorithms used in the process of classification and prediction. The most widely used algorithms are Support Vector Machine [28], Hidden Markov Model [29], Neural networks, Decision tree and Text mining can also be used for classification and prediction. Based on the survey, we could identify that The Random Forest algorithm [30], works well for large data, hence the most appropriate algorithm for the huge data like genomics. This algorithm can perform the required operation more efficiently, hence can be deployed in the genome data processing for identifying and preparing the precise treatment methods.

IV. CONCLUSION

There are not many studies in the field of cancer genome analysis for planning precise medicine using machine learning techniques. There is no effective existing model for predicting the variations in the gene and suggest the drug based on the genetic information. By using machine learning this process can be performed faster. The study of the genome analysis using machine learning can support precision medicine, to plan the appropriate treatment for the individual based on the genetic information. The modification in the gene expression can help the researchers and doctors to gain more knowledge regarding that particular gene. By using the precision medicine in the process of treating the cancer patients will reduce the health hazards associated with them. Thus, cancer genome analysis for precision medicine can be considered as one of the important areas of research, which can be integrated using machine learning to obtain the effective results.

The planning and determining of precise medical treatment can be effective when we apply the genome information as an input for the analysis of the patient. The genome data can be processed with the genome analysis methods and then applied as input to the machine learning techniques to provide quick insights into the analysis.

REFERENCES

1. WHO, 2018, “Latest global cancer data”, https://www.who.int/cancer/PRGlobocanFinal.pdf.
2. ICMR, NICPR, June 4, 2020,” Cancer Statistics”, http://cancerindia.org.in/cancer-statistics/.
3. W. H. Freeman, 2000, NBK21783, “An Introduction to Genetic Analysis”, https://www.ncbi.nlm.nih.gov/books/NBK21783/.
4. NIH, November 7, 2019 “A brief guide to genomics”, https://www.genome.gov/about-genomics/fact-sheets/A-Brief-Guide-to-Genomics.
5. Anthony J.F. Griffiths, Nov 13, 2018**,** “Genomics”, https://www.britannica.com/science/genomics.
6. Kelley Harris, Nature Genetics 51, 1306-1307, https://www.nature.com/subjects/genomic-analysis.
7. Machine learning, 7 March 2017, “What is Machine learning? A definition”, https://expertsystem.com/machine-learning-definition/.
8. Alex Philippidis, August 30, 2018, “10 Countries in 100K Genome Club”,https://www.clinicalomics.com/topics/translationalresearch/biomarkers-topic/biobanking/10-countries-in-100k-genome-club/.
9. J. Sheng, F. Li and S. T. C. Wong, "Optimal Drug Prediction from Personal Genomics Profiles," in IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 4, pp. 1264-1270, July 2015.
10. Y. Zhang et al., "Toward precision breast cancer survival prediction utilizing combined whole genome-wide expression and somatic mutation analysis," 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, 2017, pp. 830-835.
11. X. Chen, X. Shi, A. N. Shajahan, L. Hilakivi-Clarke, R. Clarke and J. Xuan, "BSSV: Bayesian based somatic structural variation identification with whole genome DNA-seq data," 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, 2014, pp. 3937-3940.
12. Y. Hsu and D. Si, "Cancer Type Prediction and Classification Based on RNA-sequencing Data," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, 2018, pp. 5374-5377.
13. D. Li et al., "Comprehensive analysis of pulmonary adenocarcinoma in situ (AIS) revealed new insights into lung cancer progression," 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, 2017, pp. 792-797.
14. M. Schapranow, K. Klinghammer, C. Fähnrich and H. Plattner, "In-memory technology enables interactive drug response analysis," 2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom), Natal, 2014, pp. 417-422.
15. A. F. G. Taktak, A. C. Fisher, A. S. Jones and B. E. Damato, "The role and efficacy of non-linear models in decision making and prognostication," The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Francisco, CA, 2004, pp. 411-414.
16. T. H. Tsai, N. Kasch, C. Pfeifer and T. Oates, "Text Mining for Hypotheses and Results in Translational Medicine Studies," 2014 IEEE International Conference on Data Mining Workshop, Shenzhen, 2014, pp. 127-132.
17. K. Fukuma, H. Kawanaka, S. Prasath, B. J. Aronow and H. Takase, "Feature extraction and disease stage classification for Glioma histopathology images," 2015 17th International Conference on E-health Networking, Application & Services (HealthCom), Boston, MA, 2015, pp. 598-599.
18. S. Li, H. Han, D. Sui, A. Hao and H. Qin, "A Novel Radiogenomics Framework for Genomic and Image Feature Correlation using Deep Learning," 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 2018, pp. 899-906.
19. S. Nabavi, M. Maitituoheti, M. Yamada and P. Tonellato, "Application of statistical machine learning in identifying candidate biomarkers of resistant to anti-cancer drugs in ovarian cancer," 2014 40th Annual Northeast Bioengineering Conference (NEBEC), Boston, MA, 2014, pp. 1-2.
20. V. Teixeira, R. Camacho and P. G. Ferreira, "Learning influential genes on cancer gene expression data with stacked denoising autoencoders," 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, 2017, pp. 1201-1205.

21. S. Srivastava, W. Wang, P. O. Zinn, R. R. Colen and V. Baladandayuthapani, "Integrating multi-platform genomic data using hierarchical Bayesian relevance vector machines," Proceedings 2012 IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS), Washington, DC, 2012, pp. 18-21.
22. G. Bontempi, "A Blocking Strategy to Improve Gene Selection for Classification of Gene Expression Data," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 4, no. 2, pp. 293-300, April-June 2007.
23. G. Alexe et al., "A robust meta-classification strategy for cancer diagnosis from gene expression data," 2005 IEEE Computational Systems Bioinformatics Conference (CSB'05), Stanford, CA, 2005, pp. 322-325.
24. A. Soliman and M. Shyu, "A Multi-Model Based Approach for Driver Missense Identification," 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, 2018, pp. 419-425.
25. K. Y. Yeung, "Signature discovery for personalized medicine," 2013 IEEE International Conference on Intelligence and Security Informatics, Seattle, WA, 2013, pp. 333-338.
26. A. Jamthe and D. P. Agrawal, "Harnessing Big Data for Wireless Body Area Network Applications," 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, 2015, pp. 868-875, doi: 10.1109/CICN.2015.172.
27. M. Qu Yang and J. Y. Yang, "An investigation into the feasibility of detecting microscopic disease using machine learning," 2007 IEEE International Conference on Bioinformatics and Biomedicine Workshops, Fremont, CA, 2007, pp. 1-2, doi: 10.1109/BIBMW.2007.4425392.
28. Carter, R J et al. “A computational approach to identify genes for functional RNAs in genomic sequences.” Nucleic acids research vol. 29,19 (2001): 3928-38. doi:10.1093/nar/29.19.3928.
29. Xi Chen, Hemant Ishwaran, “Random forests for genomic data analysis, “Genomics, Volume 99, Issue 6,2012, Pages 323-329, ISSN 0888-7543.
30. Meyer, Irmtraud M, and Richard Durbin. “Gene structure conservation aids similarity-based gene prediction.” Nucleic acids research vol. 32,2 776-83. 4 Feb. 2004, doi:10.1093/nar/gkh211.