Prediction and Recommendation of Precision Medicine for Cancer using Machine Learning Techniques

Reena Lokare, Sunita Patil

Abstract— Cancer is one of the major causes of death by disease and treatment of cancer is one of the most crucial phases of oncology. Precision medicine for cancer treatment is an approach that uses the genetic profile of individual patients. Researchers have not yet discovered all the genetic changes that causes cancer to develop, grow and spread. The Neuro-Genetic model is proposed here for the prediction and recommendation of precision medicine. The proposed work attempts to recommend precision medicine to cancer patients based upon the past genomic data of patient's survival. The work will employ machine learning (ML) approaches to provide recommendations for different gene expressions. This work can be used in caner hospitals, research institutions for providing personalized treatment to the patient using precision medicine. Precision medicine can even be used to treat other complex diseases like diabetes, dentistry, cardiovascular diseases etc. Precision medicine is the kind of treatment to be offered in the near future.

Keywords— genome, oncology, neuro-genetic model, precision medicine, machine learning

I. INTRODUCTION

Cancer is a complex disease and has a property to spread and affect other parts of the body. Often, one type of cancer can lead to another type of cancer in a patient. Around 2.25 million people in the world are living with the cancer disease. Every year around 18.1 million new cancer patients are added. Cancer related deaths are around 9.6 million [1]. With the help of computer technology, lot of data related to patients is available with the hospitals. Data related to patient's history, drugs and / or therapies given, disease related information, is available in sufficient amount.

Cancer is a complex disease. Complete causes behind cancer are not yet known to the researchers. Cancer disease complexity is defined in four stages. Effective cancer prognosis techniques need to be developed for early detection of a cancer. Cancer is also called a genetic disease as lot of genetic changes occurs in the patients with this disease. Cancers caused due to genetic changes in the tumor can be treated with the Precision Medicine. Precision medicine is not a new clinical approach but lot of technological advancements is taking place in recent times.

Ms. Reena Lokare*, Ph.D. Scholar, Asst. Prof.,K.J. Somaiya Institute of Engineering & Information Technology, Sion,Mumbai Department of Computer Engineering,PHCET, Rasayani, University of Mumbai, India reena.l@somaiya.edu

Dr. Sunita Patil, Professor & Vice Principal,K.J. Somaiya Institute of Engineering & Information Technology, Sion,Mumbai University of Mumbai, Indiavice_principal@somaiya.edu

Precision medicine is an approach where patient is given treatment by understanding genomics of a patient. It is observed that in cancer, genetic changes happening in one patient's tumor are not same in other patient. This is a reason behind treatment given to one patient suffering from same type and stage of a cancer is not suited to other patient suffering from same type and stage of a cancer. There is a strong need to treat cancer patient with personalized or precision medicine and not with a generalized one. Lot of research is happening in the field of precision medicine. Precision medicine can also play a major role in prognosis of a disease.

Gene mutations take place in the cancer. A gene controls the functioning of the cell, especially how they grow and divide. When a gene mutated, cells can grow abnormally and lead to the development of the tumor. This tumor can be Benign or Malignant. Benign tumor is not cancerous and does not spread. It doesn't cause any harm to a patient. Malignant tumor is cancerous. It can grow, spread and can lead to many different types of cancers. Nearly 5% to 10 % cancers are caused due to mutations in the inherited genes from the parents. Other cancers are caused by factors such as age, gender, environmental factors such as exposure to UV radiation, lifestyle, consumption of tobacco, alcohol etc. Major deaths are observed worldwide are due to breast cancer, Lung cancer and Colorectal cancer. All cancer types are causing nearly 2 million deaths every year. Also, a particular cancer can lead to the development of another type of cancer. Idea of precision medicine is to accurately predict such genetic changes occurring in a cancer and propose a medicine which will suit a patient.

Next generation sequencing technology is one of the DNA sequencing technology available today has revolutionized the cancer related research, such as precision medicine for cancer proposed here [2]. Microarray is also one of the advanced technologies available for DNA sequencing [3].

Though both technologies have its own merits and demerits, the choice of any one technology for DNA sequencing strongly depends on the application to be designed.

This paper presents application of Machine Learning techniques on cancer patient data helping doctors to predict and recommend precision medicine to a patient based on genomic data.

Prediction and Recommendation of Precision Medicine for Cancer using Machine Learning Techniques

II. RELATED WORKS

Machine learning algorithms are beingwidely used in cancer disease prediction and recommendation of precision medicine for cancer. Till date different machine learning algorithms are used to find the correlation between patient's molecular profile and drugs given. Though scientists are yet not able to find out all the genetic changes that are associated with the cancer, lot of research is in progress in the field of bioinformatics.

Chih et al. proposed a deep learning model to predict drug response based on mutation and expression profiles of cancer or tumor cell. Because of the fundamental differences between in vitro (processes performed outside living organisms) and in vivo (processes performed inside living organisms) biological systems, a translation of pharmacogenomics features derived from cells to the prediction of drug response of tumors is not yet understood [4].

Lin Eugene et al. have proposed a Deep Learning approach for predicting Antidepressant response in major depression using clinical and genetic biomarkers. The goal of the proposed work is to establish deep learning models which distinguish responders from non-responders. and to predict possible antidepressant treatment outcomes in major depressive disorder (MDD) [5].

Huang et al. proposed an open source machine learning algorithm for the prediction of precision medicine for cancer. Aim of the precision medicine is to find out optimal drug therapies based on genomic profiles of individual patient tumors. An open source platform has been introduced which employs a highly versatile support vector machine algorithm combined with a standard recursive feature elimination approach to predict personalized drug responses from gene expression profiles. The model is giving 84% accuracy when tested across NCI-60 test dataset [6].

AyushSinghal et al. developed a machine learning based method to automatically identify the mutations available in the biomedical literature related to certain disease. A tool has been designed which will be helpful in predicting a precision medicine for certain diseases. Pubmed literature has been used as a source of data. Three diseases are targeted here: prostate cancer, breast cancer, and age-related macular degeneration (AMD). The obtained result indicates that this approach will greatly benefit curation of mutation-disease databases on a mass scale [7].

III. RESULTS

Studies of the literature in the form of parameters such as validation methods used, important features, ML methods and type of data used is given in table 1.

Table 1: Machine Learning approaches in Precision Medicine

Work	ML Method	No of Patients tested	Type of data used	Result	Validation method	Important Features identified
Predicting drug response of tumors from integrated genomic profiles by deep neural networks [4]	Deep Learning Model	11,078: tumors 265:anti-cancer drugs	gene expression, gene mutation, drug-response	Mean Squared Error: 1.96 (Superior than Linear Regression and SVM)	bonferroni test	-over fitting due to model complexity -Requirement of large training data
A Deep Learning Approach for Predicting Antidepressant Response in Major Depression Using Clinical and Genetic Biomarkers [5]	Multi-layer feed forward neural network with 2 hidden layers	455	Single Nucleotide Polymorphism (SNP), age, sex, baseline Hamilton Rating Scale for Depression score, depressive episodes, marital status, and suicide attempt status of MDD patients	- area under the receiver operating characteristic curve (AUC) = 0.8228 ± 0.0571 -sensitivity = 0.7546 ± 0.0619 -specificity = 0.6922 ± 0.0765	10-fold cross-validation	phenotype antidepressant response, molecular diagnostic, prognosis tool
Open source machine-learning algorithms for the prediction of optimal cancer drug therapies [6]	Support vector machine(SVM) with recursive feature elimination (RFE)	60	gene expression profiles, drug-sensitivity profiles	Accuracy : 84%	leave-one-out cross validation	open source availability
Text mining for precision medicine: automating disease-mutation relationship extraction from biomedical literature [7]	c4.5 Decision Tree Classifier	-	Biomedical literature PubMed	precision: 0.904 recall:0.856 F-measure:0.880	10-fold cross-validation	disease-related mutations, biomedical literature

IV. PROPOSED METHOD

The aim of the proposed methodology is to devise a decision support system (DSS) which will assist doctors in predicting and prescribing suitable medicine for cancer, termed as personalized or precision medicine.

The proposed architecture will consist of cancer patient's individual historical data to be used as an input to DSS and DSS will process it and provide personalized decision/recommendations in term of output.

The proposed architecture is presented in fig. 1 below. It is divided into four phases I to IV.
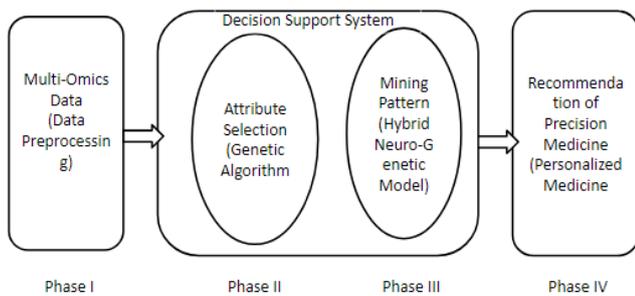

Fig. 1. Proposed Architecture for the prediction of precision medicine

Phase I: Multi-Omics Data Collection:

This phase aims at collecting data required for the prediction and recommendation of precision medicine. It is a two-step procedure as mentioned below.

- Data Collection: Data will be collected for building the proposed model using DNA Sequencing technology for cancer patients.
- Data Preprocessing: Handling of noise and missing data and encoding of categorical data will be done under this module.

Phase II: Attribute Selection:

Encoding technique will be applied on the patient's data. Genetic feature selection will be done through this module.

Phase III:Pattern Mining:

Hybrid Neuro-Genetic model will be designed and implemented for mining patterns from drug-genome datasets for the prediction and recommendation of personalized precision medicine.

Phase V: Recommendation of precision medicine:

Filtering of computed recommendations and finding strong recommendation for personalized medicine to the patients will be done in this module.

The system methodology is as given below in fig. 2. It consists of hybrid Neuro-Genetic model. It is a combination of artificial neural network (ANN), which is a good classifier; in proposed system, neural network classifies a gene-drug correlation data points to find out whether a patient responded to prescribed drug, or not.

Model also consists of genetic algorithm(GA), which is a best search and optimization algorithm. In proposed architecture, genetic algorithm has two usability:

i. To select the best inputs from available patient's dataset and ii. To decide the connection weights of the neurons so as to improve the performance of neural network are achieving more accurate prediction i.e. best suitable drugs for an individual genomic profile.

Genetic algorithm has patient's dataset as an initial population. Candidate solutions are selected from this population to undergo selection, crossover and mutation operations in order to produce a next generation of the solutions. GA then sends this new population of solutions as an input parameter to the ANN.

Finally, fitness of ANN prediction on new population is calculated. Calculating a fitness function is a challenging one as one need to decide a parameter of fitness in the given context.
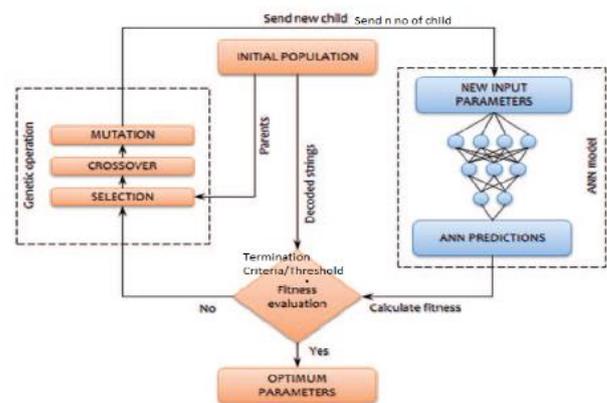

Fig. 2. System Methodology

V. FUTURE SCOPE

Precision medicine is still in its infancy in the medical field. Due to the availability of big amount of electronic data of patients, there is a strong need to propose techniques to analyze this data. Existing data mining techniques are not able to handle and analyze critical medical data successfully.

The available ML methods are showing good performance as a computer technique but lags clinical testing. Due to which, model performance cannot be analyzed to its best and hence lefts no chance in improving model performance. This can be handled by making DNA sequencing mandatory to cancer patients and create awareness among doctors to use DSS while prescribing treatment to the patient.

VI. CONCLUSION

Precision medicine is tried in large no of diseases in the recent years like pediatric oncology, psychiatric disorders, cardiovascular diseases, diabetes treatment, severe asthma, dentistry, cognitive ageing to name a few. Majority of these methodologies are based on patient's molecular profile such as genomics, proteomics etc. and are using machine learning techniques.

Prediction and Recommendation of Precision Medicine for Cancer using Machine Learning Techniques

This paper has proposed a prediction and recommendation of precision medicine for cancer patients based on genomic data using hybrid Neuro-Genetic model. The proposed model can give better accuracy compared to existing methods.

Genomic profiles of patients are an important parameter in the personalized treatment. Once genetic variations of an individual are completely understood, most appropriate and personalized treatment can be given to patients on diseases which alter genomic profiles in a patient such as cancer, diabetes etc.

The biggest challenge in using ML techniques to analyze genomic data is the interpretation of results. The results obtained by ML techniques need to be interpreted by a medical person in the correct context. The proposed work integrates the biology and computer science techniques together which demands the collaboration between computer scientists and medical professionals.

REFERENCES

1. National Cancer Institute, https://www.cancer.gov/about-cancer/understanding/statistics, Accessed 19 November, 2019, 1.33 pm.
2. Behjati S, Tarpey PS. What is next generation sequencing?. *Arch Dis Child Educ Pract Ed*. 2013;98(6):236–238. doi:10.1136/archdischild-2013-304340
3. Govindarajan R, Duraiyan J, Kaliyappan K, Palanisamy M. Microarray and its applications. *J Pharm Bioallied Sci*. 2012;4(Suppl 2):S310–S312. doi:10.4103/0975-7406.100283
4. Chiu, Y., Chen, H.H., Zhang, T. *et al*. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med Genomics* **12,** 18 (2019) doi:10.1186/s12920-018-0460-9
5. Lin Eugene, Kuo Po-Hsiu, Liu Yu-Li, Yu Younger W.-Y., Yang Albert C., Tsai Shih-Jen (2018) A Deep Learning Approach for Predicting Antidepressant Response in Major Depression Using Clinical and Genetic Biomarkers, Frontiers in Psychiatry,Volume=9 ,DOI=10.3389/fpsyt.2018.00290.
6. Huang C, Mezencev R, McDonald JF, Vannberg F (2017) Open source machine-learning algorithms for the prediction of optimal cancer drug therapies. PLoS ONE 12(10): e0186906. https://doi.org/10.1371/journal.pone.0186906.
7. Ayush Singhal, Michael Simmons, Zhiyong Lu, Text mining for precision medicine: automating disease-mutation relationship extraction from biomedical literature, Journal of the American Medical Informatics Association, Volume 23, Issue 4, July 2016, Pages 766–772, https://doi.org/10.1093/jamia/ocw041.

AUTHORS PROFILE

Ms. ReenaLokare is Ph.D. scholar in the university of Mumbai and an assistant professor in the department of Information Technology, KJSIEIT, university of Mumbai. She has nearly 16 years of teaching experience. She has publications in International and national conferences.

Dr. Sunita R. Patil is a Vice Principal, Professors at KJSIEIT, Mumbai, University of Mumbai (UoM), India. She is a member, Boardof Studies in Computer Engineering, UoM. She received Ph.D. in Computer Engineering in the domain Data Mining, Big Data & Data Science. She is having around 20 years of teaching & administrativeexperience. She has publications her research work in various recognizedNational/International Journals and conferences. She has visited variousinternational universities and organizations for attending conferences, knowledge sharing & exchange of information. Her passion is to bring in various outcome based reforms in the field of academics contributing to the growth of society, nation and world at large.

Retrieval Number: B3727129219/2019©BEIESP
DOI: 10.35940/ijeat.B3727.129219
Journal Website: www.ijeat.org

795

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication