International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST523103148

Prediction of Cancer Treatment Effectiveness and Patient Outcomes using Machine Learning Classification Approaches – a Review

Maneesh Ragavendra K[1], Mohammed Mutahar[2], R Aafrein[3], Mohammed Mustafa Jafer[4], Dr. R. Chinnaiyan[5]
[1]B. Tech CSE – AI ML, School of Computer Science and Engineering, Presidency University, Bangalore , Karnataka, India
[2]Professor & Head- RC, School of Computer Science and Engineering, Presidency University, Bangalore , Karnataka, India

ARTICLEINFO

Article History:

Accepted: 01 June 2023
Published: 11 June 2023

Publication Issue
Volume 10, Issue 3
May-June-2023

Page Number
801-807

ABSTRACT

This study systematically reviews the Machine Learning methods developed to help predict the patient outcome and treatment effectiveness in cancer treatment. This research paper has been drafted from several other similar papers and with the help of a few topics related websites providing information regarding the radiation toxicity, survival rate and tumor response. Which are the main classification criteria for the patients. The use of ANN, DT, SVM and BNs has proved to be very beneficial in the classification of any given dataset, the accuracy of the model will be high with the use of all these ML methods.

Keywords : Machine learning, cancer treatment, radiation toxicity, survival rate, tumor response, ANN, DT, SVM, BNs.

I. INTRODUCTION

Cancer can be defined as a disease caused by the uncontrolled division of abnormal cells in a part of the body. Doctors divide cancer into 4 types based on where they begin, they are:

A) Carcinoma:

They begin in the skin or in the tissue that covers the surface of the internal organs and glands. They are the most common type of cancer. They are usually solid form of tumors. Example: Prostate cancer, Breast cancer and Colorectal cancer.

B) Sarcoma:

They usually occur in the connective tissues of the body. They can occur in fat, muscles, nerves, tendons, joints, blood vessels, lymph vessels, cartilage or bones.

C) Leukemia:

Leukemia is a cancer of blood.

D) Lymphoma:

Lymphoma is cancer that develops in the lymphatic system of the body.

There are two types of tumors:

A) Benign tumors:

These are usually non-cancerous cells which are curable and don't spread to any other parts of the body to form new lumps or cancerous cells.

B) Malignant tumors:

As the tumor grows, the bloodstream or the lymphatic system carries the cancerous cells to other parts of the body. And the cancerous lump may grow and develop into new tumors in new parts of the body. This phenomenon is called metastasis. And such tumors are called malignant tumors.

Cancer is characterized as a heterogeneous disease that consists of many subtypes. The early diagnosisand prognosis of cancer types have become a necessityin cancer research, as they can facilitate thesesubsequent clinical management of patients. The application of ML (Machine Learning) comes into play in determining if the patient is a low risk or a high-risk patient.

Spotting clues for outcomes, Dr. Ganesh Vigneshwaran lead study said: "While cancertreatments are helping many people, we cannot predict who will respond well to the treatment."

This paper is solely based on creating a model that will help the doctors or the diagnostic department to determine the patient outcome or response to the treatment and the treatment outcome, that being which form of treatment is most suitable and most efficient to each patient.

II. LITERATURE REVIEW

S. Lee et al., 'Prediction of Cancer Patient Outcomes Based on Artificial Intelligence' is one of the paper that has helped in drafting the framework of this model, the criteria for classifying the dataset are radiation toxicity, survival rate and tumor response. "Artificial intelligence is being used to predict cancer outcomes and survival" is another paper which has helped in providing information regarding the methods that can be used to help the classification.

III. METHODS

Machine Learning, being a subset of Artificial Intelligence, can be used to find complicated patterns in big datasets. Most of the cancer patients have routine scans, biopsy, blood tests etc. These mightcontain clues or insights on the outcome or survival rates. The goal is to employ AI to extract and uncover the information to help predict treatment outcomes. And to establish which patient is more likely to respond and predict which treatment is most likely to be useful, we can improve decision-making and save the patients from ineffective treatments and side effects.

A) Data Preparation:

Data collection from multi-institutional sources:

For prediction models using supervised learning, patients' data can be obtained by analyzing the outcome and prognosis of individual patients. To compare huge amounts of data collected from various institutions and sources multi-institutional data collection is useful. That being the data of one institution can be used to verify the data gathered by the other institution.

Oncospace (http://oncospace.radonc.jhmi.edu/) is a multi-institutional big data platform that offers various data on radiation oncology.

B) Data collection from literature-based sources:

The data from previously recorded databases, research papers, articles, books etc. can be used as sources to feed the model with sufficient data for supervised learning.

C) Criteria for drafting cancer patients' outcomes:

According to OWG (Outcomes Working Group), the survival rate / the quality of life of the patient should be prioritized over the cancer outcomes. And multiple outcomes are to be considered as a single outcome cannot determine the overall patient outcome of the following cancer treatment. There are three main criteria to consider, them being: Toxicity, Response and Survival Rate.

D) Toxicity:

Acute or Chronic Toxicity is very important, where chronic toxicity is specifically critical in children. The Radiation Therapy Oncology Group (RTOG) differentiates between acute and late toxicity from the side effects of radiation therapy and provides guidelines for the clinical management of the toxicity grade for each critical organ. CTCAE (Common Terminology Criteria for Adverse Events) is a kind of scoring system and is a product of the US National Cancer Institute.

And there are 5 grades of toxicity that is used todenote a fatality occurring during the treatment, they are:

- Grade 1: Mild
- Grade 2: Moderate
- Grade 3: Severe
- Grade 4: Life-Threatening
- Grade 5: Death.

E) Response:

The assessment of a solid tumor usually consists of a bidimensional (World Health Organization criteria, WHO) or a unidimensional (response evaluation criteria in solid tumors guidelines, RECIST) measurement of tumors before and after chemotherapy.

The response of the patients can be classified into 2 categories, they are:

- A Complete Response
- A Partial Response

F) Survival rate:

The survival rates of each patient and each group of cancer varies.

Consider an example of lung cancer: the international 5- year rate of survival for patients with lung cancer varies from 5-16%. G) Survey of ML application in cancer: The most commonly used electronic databases for the prediction model were PubMes and Scopus. As these datasets were rather raw, they had to be further scrutinized in order to maintain the most relevant information/ articles. We also use Oncospace as mentioned earlier as it is a multi-dimensional big data platform offering large amounts of data on radiation oncology. H) Prediction Model: The accurate and exact prediction of the patients' outcome is a quite challenging task. Machine learning techniques have become very popular with medical researchers as these machine learning models can identify the patterns and relationships between treatment methods and outcomes easily and far more accurately than the usual/conventional methods. These ML models use complex datasets to easily predict outcomes for a specific type of cancer.

IV. Proposed Methods

The most commonly/widely used ML method is semi-supervised learning, which is a blend of supervised and unsupervised learning methods. It basically combines labeled and unlabeled data to construct a better and more accurate learning model. And this type of learning is usually used when the amount of unlabeled data is bigger than labeled data.

To make the raw data easier to use and susceptible for usage, preprocessing steps should be applied that modify the data. And a variety of techniques are used to modify the data, a few of the most important approaches are as follows:

I. Dimensionality Reduction: As the machine learning models tend to work better when the dimensionality is lower. The additional reduction of dimensionality will eliminate irrelevant features, reduce noise and can produce more robust learning models as there is involvement of fewer features.

II. Feature Selection: Feature selection is basically dimensionality reduction by selecting new features which are a subset of the old ones. There are three main approaches related to feature selection being embedded, filter and wrapper approaches.

III. Feature Extraction: In feature extraction a new set of features can be created from the set of data that was initially being used which consisted of all the significant information in a dataset. This method of creating new sets of features allows the gathering of the described benefits of dimensionality reduction.

The prime objective of the ML model is to provide a model which can be used to perform all sorts of classification, prediction, estimation or any other similar task. The most common task in a learning model is classification. A good classification model should be able to fit the training dataset accurately and classify all the instances specified. The phenomenon of overfitting occurs when the test errorrates of a model begin to increase even when thetraining error rates decrease. To reduce the errors the complexity of the model is increased. Once a classification model is drafted using one of more ML techniques, it is important to calculate the classifier's performance. The performance of the classifier is measures in terms of sensitivity, accuracy, specificity and area under the curve (AUC).

There are several methods for evaluating the performance of the classifier, the most used methods by splitting the initial labeled data into subsets are:

I. Holdout Method: The data samples are split into training and testing data sets, which in-turn generates a classification model from the training set while its performance is estimated based on the test set.

II. Random Sampling: It is similar to the holdout method. In order to increase the accuracy of the model the holdout method is repeated several times by choosing the training a d test instances randomly

III. Cross-Validation: In this method each sample is used the same number of times for training and just once for testing, which results in the original data set being covered successfully both in the training and testing sets.

IV. Bootstrap: In this method the samples are separated with replacement into training and test sets. Once the data is preprocessed and we have the kind of learning task, there are a few suitable ML methods that can be applied in the literature for the case study of cancer prediction and prognosis. They are:

I. ANN's (Artificial Neural Networks): This method can handle a variety of classification and pattern recognition problems. Even though the ANN method served as a gold standard in several classification tasks, they had a few drawbacks. The layered structure was proving to be time consuming as they had three layers (input, hidden, output), which in turn leads to poor performance and the technique is also called the black-box technique as error recognition was almost impossible to find.

II. DT's (Decision Trees): DTs are one of the oldest and highly prominent ML techniques that's been used for classification purposes. And the DTs are quick to learn and very simple to interpret. The nodes of these trees represent the input variables and the leaves being the decision outcomes.

III. SVM's (State Vector Machines): SVMs are a more recent methods of ML that are being applied in cancer prognosis and prediction. They map the input vector into a feature plane of larger dimensionality and identify the hyperplane that separates the data points into two classes. The marginal distance of the decision hyperplane and the instances that are closest to boundaries is maximized. And thus, the classifier achieves a considerable amount of generalizability and can therefore be used to attain the reliable classification of new samples.

IV. BN's (Bayesian network): They produce the probability estimation of the data rather than the conventional prediction. They are used to represent knowledge along with probabilistic dependencies among the variables of interest in the form of a directed acyclic graph.

V. Prediction of Cancer

Cancer prediction using machine learning methods involves training models on existing data to identify patterns and make predictions on new data. Here's an explanation of how four different ML methods can be used for cancer prediction:

A) Artificial Neural Networks (ANNs):

ANNs can be trained to predict radiation toxicity, survival rate, and tumor response by using a dataset that includes relevant patient information such as treatment details, genetic factors, medical history, and tumor characteristics. The ANN model can learn the complex relationships between these features and the desired outcomes (radiation toxicity, survival rate, and tumor response) through its interconnected layers of nodes. By training the ANN on a dataset with known outcomes, it can make predictions on new patientdata, estimating the likelihood of radiation toxicity, survival rate, and tumor response based on the input features.

B) Decision Trees (DTs):

Decision trees can also be used to predict radiation toxicity, survival rate, and tumor response. By constructing a decision tree based on patient features like treatment parameters, genetic markers, and tumor characteristics, the algorithm can identify the most informative features that correlate with the desired outcomes. Each leaf node of the tree represents a predicted value for the outcomes (e.g., high/low radiation toxicity, good/poor survival rate, positive/negative tumor response). By following the decision path based on patient characteristics, the decision tree can predict the level of radiation toxicity, survival rate, and tumor response.

C) Support Vector Machines (SVMs):

SVMs can be utilized to predict radiation toxicity, survival rate, and tumor response by training the model on a dataset with patient features and corresponding labels for the outcomes. For instance, the SVM can learn from data that includes treatment parameters, genetic factors, and tumor characteristics along with the corresponding radiation toxicity level, survival rate, and tumor response. By finding an optimal hyperplane in a high-dimensional feature space, SVMs can classify new patient data, determining the likelihood of different levels of radiation toxicity, survival rate, and tumor response.

D) Bayesian Networks (BNs):

Bayesian networks can also be employed to predict radiation toxicity, survival rate, and tumor response. By constructing a probabilistic graphical model that represents the relationships between patient features and outcomes, BNs can incorporate prior knowledge and dependencies between variables. For example, variables such as treatment parameters, genetic markers, and tumor characteristics can be represented as nodes, and the conditional dependencies among them can be modeled through edges. By training the Bayesian network using historical data with known outcomes, it can perform probabilistic inference to predict the probabilities of different levels of radiation toxicity, survival rate, and tumor response given new patient data. In summary, these machine learning methods (ANNs, decision trees, SVMs, and Bayesian networks) can be applied to predict cancer outcomes based on radiation toxicity, survival rate, and tumor response. By training models on relevant patient data and known outcomes, these methods can provide predictions for new patient cases, aiding in personalized treatment decisions and prognosis assessment.

VI. Conclusion

Due to the increasing sizes of the datasets and the rate at which the data is being produced is very fast and the increasing range of formats employed, big data and artificial intelligence is being used in predictive analysis extensively. The constant need to improve and the need for more accurate predictions, the implementation of AI is demanded. Outcomes such as radiation toxicity, survival rate and tumor response are important in cancer patients and physicians aswell. As ANN is sometimes superior to conventional statistical analysis in predicting the prognosis of a cancer patient. As the inclusion of heterogeneous data has been increasing over the years, the combined application of different techniques in this model will yield better results in the prediction and prognosis of cancer patient outcomes and the treatment outcomes.

VII. REFERENCES

[1]. Artificial Intelligence - Scope and Limitations. IntechOpen, Apr. 24, 2019. doi: 10.5772/intechopen.81872.

[2]. Abdulhamit Subasia, Bayader Kadasaa, Emir Kremic, “Classification of the Cardiotocogram Data for Anticipation of Fetal Risks using

Bagging Ensemble Classifier", Procedia Computer Science 168 (2020) 34–39

[3]. Alessio Petrozziello, Ivan Jordanov, Aris T.Papageorghiou, Christopher W.G. Redman, and Antoniya Georgieva," Deep Learning for ContinuousElectronic Fetal Monitoring in Labor", Preprint, Researchgate

[4]. Attallah O, Sharkas MA, Gadelkarim H. Fetal Brain Abnormality Classification from MRI Images of Different Gestational Age. Brain Sciences. 2019; 9(9):231.

[5]. Balachandar S., Chinnaiyan R. (2019) Centralized Reliability and Security Management of Data in Internet of Things (IoT)with Rule Builder. In: Smys S., Bestak R., Chen JZ., Kotuliak I. (eds) International Conference on Computer Networks and Communication Technologies. Lecture Notes on Data Engineering and Communications Technologies, vol 15. Springer, Singapore

[6]. Balachandar S., Chinnaiyan R. (2019) Reliable Digital Twin for Connected Footballer. In: Smys S., Bestak R., Chen JZ., Kotuliak I. (eds) International Conference on Computer Networks and Communication Technologies. Lecture Notes on Data Engineering and Communications Technologies, vol 15. Springer, Singapore

[7]. Comert Z., Kocamaz A. F., Subha V. (2018). Prognostic model based on image-based time-frequency features and genetic algorithm for fetal hypoxia assessment. Comput. Biol. Med. 99 85–97.

[8]. Daniel LaFreniere, Farhana Zulkernine, David Barber, Ken Martin. "Using Machine Learning to Predict Hypertension

[9]. G Sabarmathi, R Chinnaiyan (2019), Envisagation and Analysis of Mosquito Borne Fevers: A Health Monitoring System by Envisagative Computing Using Big Data Analytics, Lecture Notes on Data Engineering and Communications Technologies book series (LNDECT, volume 31), 630-636. Springer, Cham

[10]. G. Sabarmathi, R. Chinnaiyan (2016) , Big Data Analytics Research Opportunities and Challenges - A Review, International Journal of Advanced Research in Computer Science and Software Engineering, Vol.6 , Issue.10, 227-231

[11]. G. Sabarmathi, R. Chinnaiyan, Investigations on big data features research challenges and applications, IEEE Xplore Digital LibraryInternational Conference on Intelligent Computing and Control Systems (ICICCS), 782 – 786.

[12]. M Swarnamugi, R Chinnaiyan (2019), IoT Hybrid Computing Model for IntelligentTransportation System (ITS), Proceedings of the Second International Conference on Computing Methodologies and Communication (ICCMC 2018), 802-806.

[13]. M. Swarnamugi ; R. Chinnaiyan, "IoT Hybrid Computing Model for Intelligent Transportation System (ITS)", IEEE Second International Conference on Computing Methodologies and Communication (ICCMC),

15-16 Feb. 2018.

[14]. M. Swarnamugi; R. Chinnaiyan, "Cloud and Fog Computing Models for Internet of Things", International Journal for Research in Applied Science & Engineering Technology, December 2017.

[15]. R.Vani, "Weighted Deep Neural Network BasedClinical Decision Support System for the Determination of Fetal Health", International Journal of Recent Technology and Engineering (IJRTE)ISSN: 2277-3878, Volume-8 Issue-4, November 2019,8564-8569.

[16]. Ragab DA, Sharkas M, Attallah O. Breast Cancer Diagnosis Using an Efficient CAD System Based on Multiple Classifiers. Diagnostics. 2019; 9(4):165.

[17]. S. Balachandar, R. Chinnaiyan (2019), Internet of Things Based Reliable Real-Time Disease

Monitoring of Poultry Farming Imagery Analytics, Lecture Notes on Data Engineering and Communications Technologies book series (LNDECT, volume 31), 615- 620. Springer, Cham

[18]. S.Balachandar , R.Chinnaiyan (2018), A Reliable Troubleshooting Model for IoT Devices with Sensors and Voice Based Chatbot Application, International Journal for Research in Applied Science & Engineering Technology,Vol.6,Iss.2, 1406-1409.

[19]. S.Balachandar , R.Chinnaiyan (2018), Centralized Reliability and Security Management of Data in Internet of Things (IoT) with Rule Builder, Lecture Notes on Data Engineering and Communications Technologies15, 193-201.

[20]. S.Balachandar , R.Chinnaiyan (2018), Reliable Digital Twin for Connected Footballer, Lecture Notes on Data Engineering andCommunications Technologies 15, 185-191.

Cite this article as :

Maneesh Ragavendra K, Mohammed Mutahar[2], R Aafrein[3], Mohammed Mustafa Jafer[4] and Dr. R. Chinnaiyan[5],"Prediction of Cancer Treatment Effectiveness and Patient Outcomes using Machine Learning Classification Approaches - A Review", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 10 Issue 3, pp. 801-807, May-June 2023.

Available atdoi :

https://doi.org/10.32628/IJSRST523103148

Journal URL : https://ijsrst.com/IJSRST523103148