Chapter 7
A Machine Learning Approach to Prevent Cancer

Ahan Chatterjee
https://orcid.org/0000-0001-5217-4457
The Neotia University, India

Swagatam Roy
https://orcid.org/0000-0002-8012-5529
The Neotia University, India

Rupali Shrivastava
The Neotia University, India

ABSTRACT

One of the most talked about diseases of the 21st century is none other than cancer. In this chapter, the authors take a closer look to prevent cancer through machine learning approach. At first, they ran their classifier models (e.g., decision tree, K-mean, SVM, etc.) to check which algorithm gives the best result in terms of choosing right features for further treatment. The classified results are compared, and then various feature reduction algorithm is being used to identify exactly which features affects the most. Various data mining algorithms are being used, namely rough-based theory, graph-based clustering, to extract the most important features which influence the results. In the next section they take a look in the cancer analytics part. A simulation model has been designed that can easily manage the patient flow in OPDs and a bed rotation model also have been designed to give patients an insight that how much time they will spend in the queue. Further they analyzed a risk analysis model for chemotherapy treatment, and finally, an econometric discussion has been drawn in how it affects the treatment.

INTRODUCTION

"We have two options, medically and emotionally: give up or fight like hell." – Lance Armstrong
7 Times Tour De France Winner and a Cancer Survivor

In the last year 2018, 9.6 million deaths have been recorded due to cancer. It has been estimated that every 1 out of 6 people die due to cancer. This disease is considered as heterogeneous disease as there are many subtypes of it. Early detection of this disease is much needed as it can be cured partially if detected early. Classification of cancer has moved various research organizations around the globe, and nowadays classifying the disease using Machine Learning approach is in the business. The paper is majorly divided into two sections in the first section we look into the machine learning tools to extract insights from the data by which it can be prevented and in the next section we take a closer look in the cancer analytics domain.

Machine Learning allows us to detect and extract key features or hidden patterns which are there in the complex datasets of cancer patients. Thus use of machine learning gives us an upper hand in this field detecting the disease. Classification and data mining are an effective way to classify the data. Especially in bio-medical field where based upon the classification results analyses are done and decisions are being taken. In this paper we aim to find a performance comparison between various algorithms namely, Support Vector Machine (SVM), Decision Tree, Naïve Bayes (NB) and K-Nearest Neighbor (KNN). All the algorithms are being applied on Wisconsin Breast Cancer (Original) Dataset. Then the breast cancer underwent a feature selection process to extract the key features which influences the most. All the features don't affect the result equally thus in this step we will cancel out the irrelative features to get better results. Rough Based Theory and Graph Based Clustering have been used to extract the features. PCA algorithm is also being used to reduce the dimension. The aim of this paper is to evaluate the correctness in classifying data for each algorithm with parameters namely, efficiency and effectiveness in terms of accuracy and precision. From the experiments and results we can say that use of machine learning models can improve our vision and understanding in progression of cancer cells, but we need a proper validation of this works to practice this in everyday clinical practice. In the next section we mainly focus on the analytics part of the healthcare. In easy words, we can say that how much care needed for the patients for better and stable health required it is analyzed in this step. We can say minimizing the queue time for a patient will return a better health status than being harassed by going here and there for different things. Other various factors has also been considered as the economic background to carry on the treatment, as the cancer treatment is still a costly one being one of the most dreadful disease in the modern times. The investment from the government in the treatment for cancer is also an important factor by which it influences the treatment rate or cost. In this limelight we will also analyze the waiting time and success rate for a patient undergo surgery or chemotherapy and how can we increase the success rate by risk analysis by Bayesian Modeling. A set of questions arise when we talk about the healthcare part, that is there timely detection, proper diagnosis, does the patients getting proper care from the industry specialties, in this section we will put a limelight in these sections to get an data insight which can be analyzed better performance in near future. Further we would design a patient flow model which can easily manage the flow timeliness manner, and the impact of rapid diagnosis is being measured. A patient when diagnosed with cancer it is not only the physical pain that they go through, various other factors is also closely related with those which directly or indirectly creates an impact over the health of the patient, *viz.* physical, psychological, financial, and spiritual as well plays a role. A patient can go under huge debt due to their treatment for which they go in depression which surely affects their health condition. We will go into a closer look how they affect the health and would try to frame a solution for these solutions. (Wait et al., 2017; Zhou et al., 2004)

28 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/a-machine-learning-approach-to-prevent-cancer/263317

Related Content

Automatic Moderation of User-Generated Content
Issa Annamoradnejadand Jafar Habibi (2023). *Encyclopedia of Data Science and Machine Learning (pp. 1344-1355).*
www.irma-international.org/chapter/automatic-moderation-of-user-generated-content/317543

Deep Learning for Feature Engineering-Based Improved Weather Prediction: A Predictive Modeling
Partha Sarathi Mishraand Debabrata Nandi (2021). *Handbook of Research on Automated Feature Engineering and Advanced Applications in Data Science (pp. 195-217).*
www.irma-international.org/chapter/deep-learning-for-feature-engineering-based-improved-weather-prediction/268756

Deep Learning Techniques for Smart Agriculture Applications
Ankita Mishra, Sourik Banerjeeand Brijendra Singh (2023). *Machine Learning and Deep Learning for Smart Agriculture and Applications (pp. 1-23).*
www.irma-international.org/chapter/deep-learning-techniques-for-smart-agriculture-applications/329887

Churn Prediction in a Pay-TV Company via Data Classification
Ilayda Ulku, Fadime Uney Yuksektepe, Oznur Yilmaz, Merve Ulku Aktasand Nergiz Akbalik (2021). *International Journal of Artificial Intelligence and Machine Learning (pp. 39-53).*
www.irma-international.org/article/churn-prediction-in-a-pay-tv-company-via-data-classification/266495

A Method Based on a New Word Embedding Approach for Process Model Matching
Mostefai Abdelkaderand Mekour Mansour (2021). *International Journal of Artificial Intelligence and Machine Learning (pp. 1-14).*
www.irma-international.org/article/a-method-based-on-a-new-word-embedding-approach-for-process-model-matching/266492