IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Prediction Of Cancer Using Machine Learning Models

Robin Chauhan*

Vivek Prajapati**

Bennett University, Greater Noida

Abstract: A common disease that affected the larger section of the society is Cancer. It is of many types like breast cancer, Lung cancer etc. This paper involves the Machine Learning in the prediction of cancer at initial stages. Cancer is a large set of diseases that can affect various body parts. It starts when cells in the body start to grow uncontrollable manner and spread to other parts. These abnormal behaviors of cells are termed as cancerous cells which ultimately lead to cancer. In this Disease a cell divides continuously and abnormally and it is a Deadly Disease. Aim of this document is to generate the Machine Learning (ML) models which can predict the cancer in the initial stages so that we can improve the chances of survival. Patients who are suffering from this disease can only be saved if it is identified at initial stages like stage I and stage II. If it is detected in stages like stage III and stage IV also known as advance stages, then the chance of survival of patient will be minimal. ML and the new technologies evolve around us can be helpful to tackle this kind of healthcare problem. Cancer symptoms depending on its type, its location, and its stages. Some general symptoms that may indicate the presence of cancer like Unexplained weight loss, Fatigue, Persistent cough, Unexplained pain, Changes in skin moles or growths, Persistent indigestion or difficulty swallowing etc. Different Model provides some statistics and understanding which can be useful in the assessment of model and provide a better way for the analysis.

Keywords: Machine Learning, model, Cancer, Prediction, Cancerous, Deadly Disease, stages, healthcare, Symptoms, abnormal, analysis.

Introduction

Cancer is a disease having a feature like uncontrolled growth and spreading of the abnormal cells in the various part of body [1]. These cells can destroy normal tissue, causing the malfunctioning of organs and systems in the body [9]. There are various types of cancer, each with its unique behavior and nature, but they all share on thing common that is the abnormal cell growth [2]. Cancer is having various symptoms such as tumor, abnormal blood loss, persistent cough, abnormal loss of weight etc. [1]. In medical field and academics cancer is one of the interesting and challenging field which fascinate the research scholars. For the provision of better healthcare system there is a need of model which can automatically predict the various types of cancers. Because it takes millions of lives every year, cancer has a significant impact on world health. 2018 saw it rank as the second most common reason of death globally, accounting for an estimated 9.6 million deaths. Disparities in cancer care are caused by unequal access to healthcare, with low- and middle-income nations having higher death rates as a result of scarce resources for diagnosis, treatment, and prevention.

To counter this worldwide problem, a multi-layered strategy is needed. It is very important to support public health programs that encourage healthy habits, such as quitting smoking and get vaccinated against viruses that cause cancer. Cancer can be Identified by Two ways One is Traditional Method and Other one is Modern world Internet of things (IOT) and machine Learning (ML) Techniques which can be used in the Early Prediction and hence can be helpful to save many lives. This Paper is Divided into Four parts. First One is Traditional Methods and their Drawbacks. Second is need Machine Learning in the Present Era [2]. Datasets Used in the ML models and fourth one is Result which we get from those models like Accuracy.

1. Traditional method for detecting the cancer in the patients has some serious drawbacks like

Insensitivity: Many conventional cancer detection methods, such as biopsies or surgical procedures, can be insensitive and may pose risks to the patient, including bleeding, infection etc. [1]

Limited Sensitivity: Some screening tests, such as PSA (prostate-specific antigen) tests or even mammograms, may produce faulty results, leading to unnecessary anxiety and further invasive procedures.[5]

Exposure to Radiation: Imaging such as X-rays, CT scans, and PET scans expose patients to ionizing radiation, which carries potential risks, particularly with continuous or excessive exposure.[8]

Cost and Accessibility: Many conventional cancer detection methods, such as imaging tests or laboratory screenings, can be very costly, making them inaccessible to individuals who are from poor financial background or without optimal insurance policy.

Limited Ability for Early Detection: Some conventional screening methods may only detect cancer at advance stages when it has already significantly, minimizing the chances of successful cure.

Patient Discomfort and Anxiety: Some screening procedures, such as colonoscopies or mammograms, can be uncomfortable or anxiety-provoking for patients, leading to reluctance or avoidance of screening.

2 Need of Machine Learning as a tool to find the possibilities.

Machine Learning can be helpful and provide very insightful way to predict the cancer in early stages by analyzing complex and large datasets also known as big data which include the features or we can say that the data of patient's symptoms. We can analyze these features and include various important features and can avoid less import features which is technically known as feature extraction.[3]

Early Detection: ML algorithms can be used to analyze voluminous patient data, including genetic information, medical images, and clinical data, to detect patterns that are the indicator of the early-stage cancer.[2]

Risk Stratification: ML algorithms can assess an individual's chance of developing cancer based on various factors such as genetic features, environmental exposures, lifestyle habits, and demographic characteristics and conditions.

Personalized Medicine: ML enables the development of personalized cancer prediction models that take into account an individual's genetic makeup, molecular profile, and clinical history.

Prognostic Assessment: By analyzing data of cancer patients, ML algorithms can identify prognostic biomarkers, treatment response predictors, and factors associated with disease progression, helping Doctors make firm decisions about treatment strategies and follow-up care.

Treatment Response Prediction: ML algorithms can predict how individual patients are responded to specific cancer treatments, such as chemotherapy, immunotherapy, or targeted therapy.

Monitoring: ML can contribute to ongoing surveillance of cancer patients by analyzing longitudinal data and detecting signs of disease.

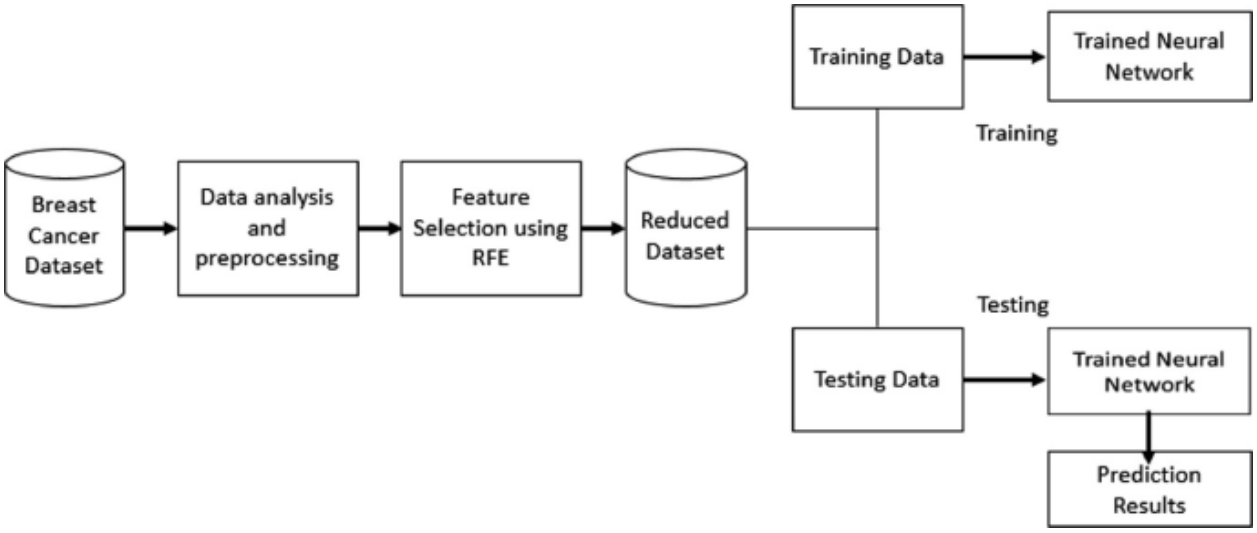

Fig.2 Data Flow System in Machine Learning
[https://images.app.goo.gl/GiSoQp2xE9EXTQZM8]

3 Datasets

Wisconsin Breast Cancer Dataset: This dataset is widely acceptable and used for many applications because it has a large number of attributes which is free from any kind of noise-free [3].

4. Machine Learning and the Detection or Prediction of Cancer

Several machine learning algorithms have shown potential in predicting cancer, with each having its strengths and weaknesses with much higher accuracy and low error rate which can be helpful for the medical practitioners and researchers whose are working in the medical field. Here are some prominent examples:

Support Vector Machines (SVMs): These are used to identifying patterns in high-dimensional data, making them suitable for analyzing complex medical data like gene expression profiles or imaging scans. They are known for good interpretability, allowing researchers to understand the reasoning behind predictions. [10-11]

Random Forest: These are ensembles of decision trees, where multiple decision trees vote on the final prediction, improving accuracy and robustness compared to single trees [11]. They are versatile and handle various data types well.

Deep Learning: Artificial neural networks with multiple layers, copying the structure and functioning of the human brain. Convolutional Neural Networks (CNNs) are particularly effective in analyzing medical images, automatically extracting features for prediction, while Recurrent Neural Networks (RNNs) can handle sequential data like genetic sequences.

XGBoost: This is a gradient boosting algorithm known for its efficiency and accuracy. It works by combining multiple weaker models (decision trees) into a stronger one, iteratively improving performance.

K-Nearest Neighbors (KNN): This simple algorithm classifies data points based on the similarity to their "k" nearest neighbors in the training data [12]. It can be effective for specific tasks but may not perform as well for complex datasets.

Table 1. Features for Breast Cancer Detection

References	Datasets	Feature Available	Number of Samples
[3]	Wisconsin Breast Cancer Dataset	Diagnosis, radius mean, texture mean, perimeter mean etc.	569

Table 11. Comparison of Different Prediction Model

Dataset	Prediction Model	Accuracy	Precision	Recall	F1 Score
Wisconsin Breast Cancer Dataset	Logistic Regression	97%	98%	95%	96%
Wisconsin Breast Cancer Dataset	Random Forest Classifier	96%	98%	93%	95%
Wisconsin Breast Cancer Dataset	Support Vector Machine	96.41%	95%	95%	95%
Wisconsin Breast Cancer Dataset	Decision Tree	95%	93%	93%	93%
Wisconsin Breast Cancer Dataset	Multi-Layer Perceptron	97%	98%	95%	96%

Conclusion: ML is a latest technology that has potential to provide a new and better way of cancer prediction through earlier diagnosis, better treatment outcomes, and medicine. Not with standing, it is imperative to navigate the data challenges, ethical considerations, and patient permission. It will be very important to address these problems in order to guarantee the proper development and application of this technology [10]. With further research and resolution of these issues, machine learning for cancer prediction has potential to improve patient health outcomes in the future.

REFERENCES

[1] Harini Shree M. S., Aditya C. R.*, Sachin D. N. "Detection of Breast Cancer using Machine Learning Algorithms – A Survey," in Proceedings of the Fifth International Conference on Computing Methodologies and Communication.

[2] Jovan Andjelkovic, Branimir Ljubic, Ameen Abdel Hai, Marija Stanojevic, Martin Pavlovski, Wilson Diaz, Zoran Obradovic, "Sequential machine learning in prediction of common cancers".

[3] Md. Mehedi Hassan, Md. Mahedi Hassan, Farhana Yasmin, Md. Asif Rakib Khan, Sadika Zaman, Galibuzzaman, Khan Kamrul Islam, Anupam Kumar Bairagi, "A comparative assessment of machine learning algorithms with the Least Absolute Shrinkage and Selection Operator for breast cancer detection and prediction.

[4] V.Nanda Gopal a,* , Fadi Al-Turjman b , R. Kumar c , L. Anand d , M. Rajesh e, "Feature selection and classification in breast cancer prediction using IoT and machine learning".

[5] M. Vosooghifard and H. Ebrahimpour, "Applying grey wolf optimizer-based decision tree classifier for cancer classification on gene expression data," in 5th International Conference on Computer and Knowledge Engineering (ICCKE), Mashhad, 2015, pp. 147–151.

[6] S. Wang, F. Chen, J. Gu, and J. Fang, "Cancer classification using collaborative representation classifier based on non-convex lp-norm and novel decision rule," in Seventh International Conference on Advanced Computational Intelligence (ICACI), Wuyi, 2015, pp. 189–194.

[7] Chen, Yukun, et al., "Classification of cancer primary sites using machine learning and somatic mutations," BioMed Research International, 2015.

[8] K. Machhale, H. B. Nandpuru, V. Kapur, and L. Kosta, "MRI brain cancer classification using hybrid classifier (SVM-KNN)," in 2015 International Conference on Industrial Instrumentation and Control (ICIC), Pune, 2015, pp. 60–65.

[9] Chakraborty, Debasis, and Ujjwal Maulik, "Identifying cancer biomarkers from microarray data using feature selection and semi supervised learning," IEEE Journal of Translational Engineering in Health and Medicine, vol. 2, pp. 1–11, 2014.

[10] Kar, Subhajit, Kaushik Das Sharma, and Madhubanti Maitra, "Diagnostic prediction of multi-class cancer using SVM and nearest neighbor classifier," in International Conference on Control, Instrumentation, Energy and Communication (CIEC), IEEE, 2014.

[11] S. Dash and A. Dash, "A correlation based multilayer perceptron algorithm for cancer classification with gene-expression dataset," in 14th International Conference on Hybrid Intelligent Systems (HIS), Kuwait, 2014, pp. 158–163.

[12] H. Elouedi, W. Meliani, Z. Elouedi, and N. Ben Amor, "A hybrid approach based on decision trees and clustering for breast cancer classification," in 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR), Tunis, 2014, pp. 226–231.