e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

CANCER PEDICTION AND ANALYSIS USING MACHINE LEARNING

Monika Parit[*1], Anuja Potdar[*2], Riya Pawar[*3], Aditi Mhetar[*4], Ishwari Nimbalkar[*5]

[*1]Lecturer, Computer Engineering, Sharad Institute Of Technology Polytechnic, Yadrav, Maharashtra, India.

[*2,3,4,5]Student, Computer Engineering, Sharad Institute Of Technology Polytechnic, Yadrav, Maharashtra, India.

DOI : https://www.doi.org/10.56726/IRJMETS45301

ABSTRACT

Machine learning is increasingly being employed in cancer detection and diagnosis. Cancer prediction will become quite easy in the future and we can predict it without the need of going to the hospitals. As we can see many technologies are being used and tested in the medical field. So, by this we can say that this will make us easier in the future to detect cancer. We are testing which algorithm will give us good result among CART, SVM AND KNN. We are making a cancer prediction using machine learning, in which we are including three types of cancer they are breast cancer, lungs cancer and prostate cancer. In breast cancer, we are using SVM algorithm and for lung and prostate we are using Random forest algorithm. We are also going to detect cancer with microscopic biopsy images. It focuses on image analysis and machine learning.

Keywords: Machine Learning, Breast Cancer, Lungs Cancer, Prostate Cancer, Support Vector Machine (SVM), Random Forest.

I. INTRODUCTION

A tumor contains cancer cells that can spread to places of the body. This is the top cause of mortality among women worldwide, and it is one of the most frequent and life threatening malignant tumors. Early discovery and treatment in any disease or any cancer can increase the survival chances as well as it decreases the chances of going under the expensive treatment for benign tumors and this early detection will help doctors to give the perfect treatments to the patients. A cancer diagnosis is a key prerequisite for such socioeconomic advantages. This detection of cancer can be done by using different methods one of them is ML based detection.

Machine Learning involves creating a model that is trained on certain training data and can analyse other data to generate predictions. For this machine learning system, several types of models have been explored and researched. There are various types of machine learning models , some of them are SVM, Random Forest, Decision tree, Logistic regression. Now-a-days, we can see the constant growth in machine learning in various fields, one such field is medical field. In this, the machine algorithms create a model based on the information to predict the result of the disease using previous instances recorded in datasets by using patterns and correlations among a large number of cases.

Detection of Cancer often involves radiological imaging. Radiological Imaging is used to check the spread of cancer and progress of treatment. It is also used to monitor cancer. Oncological imaging is continually becoming more varied and accurate. Different imaging techniques aim to find the most suitable treatment option for each patient. Imaging techniques are often used in combination to obtain sufficient information. In this project, we are going to talk about two models of machine learning they are: We have selected three types of datasets for our project. They are breast cancer, Lung's cancer, prostate cancer.

II. METHODOLOGY

To detect different cancers we are going to use different machine learning algorithms. To detect breast cancer we are going to use SVM algorithm and to detect lung cancer and prostate cancer we are going to use Random Forest algorithm. For detection of cancer using image processing we are going to use CNN (Convolution Neural Networks).

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

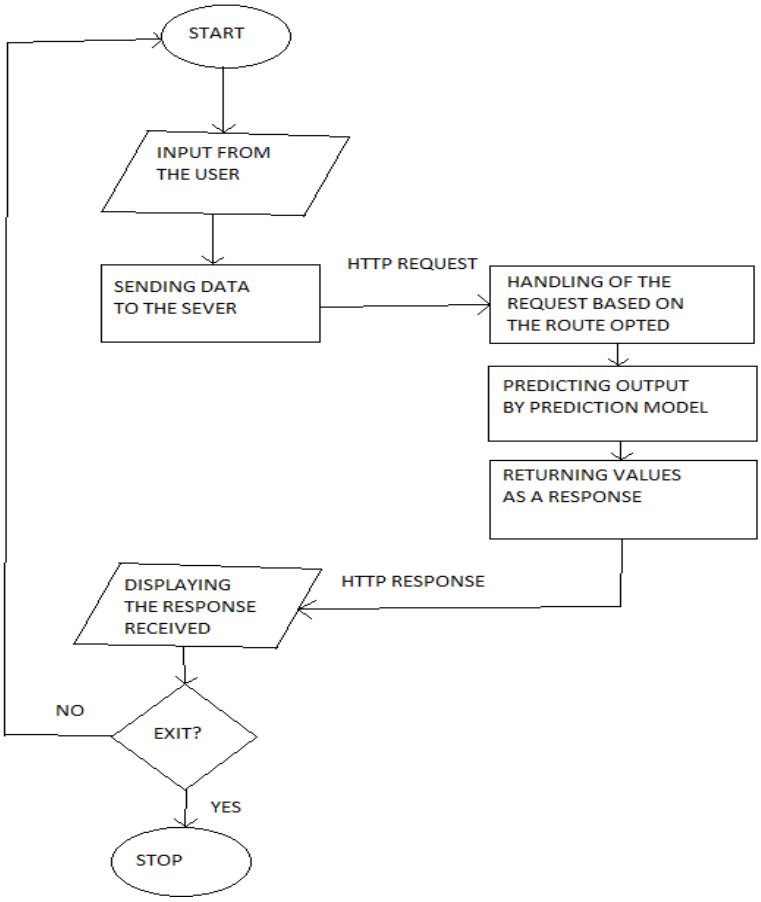

Figure 1: Flow Chart

Algorithm:

- The webpage loads, and the user is offered the option of selecting one of three cancer types.
- The user must select the cancer prediction system that he or she wishes to utilise.
- The user must next complete the form based on their symptoms.
- The data is then sent to the route that deals with machine learning models, and the result is returned on the following route.

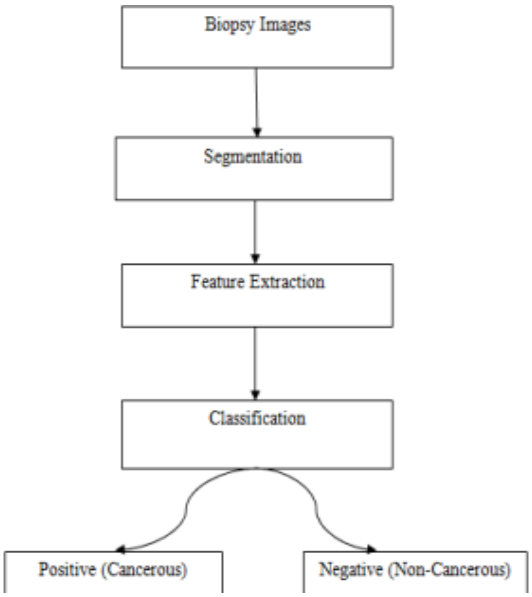

Figure 2: Architectural Diagram of cancer detection using image processing

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

Architectural diagram contains various steps:

• Microscopic tested image is taken as input after undergoing biopsy. The images are enhanced before segmentation to remove noise.

• Segmentation is done based on the input images which contains nuclei, cytoplasm and other features. They are segmented on the basis of region, threshold or a cluster and particular algorithms are applied.

• In feature extraction, various biologically interpretable and clinically notable shape and morphology based features are extracted from the segmented images which include grey level texture features, colour based features, colour grey level

III. MODELING AND ANALYSIS

1) Lung Cancer :

Lung cancer is a type of cancer which is present at the lungs. As we usually know that smoking cigars causes this lungs cancer. Not only smoking cigars but the people who doesn't smoke can also get this cancer. But the people who smoke a lot have the high chances of getting this cancer. In this cancer the cells start growing abnormally in the lungs which causes tumors and they start spreading to other parts of the body and then lung cancer happens. In this we are going to use the RANDOM FOREST model for the detection. The parameters we are going to check in this are smoking, yellow fingers, anxiety, peer pressure, chronic disease, fatigue, allergy, wheezing, alcohol consumption, coughing, shortness of breath, swallowing difficulty, chest pain as depicted in Figure 3.

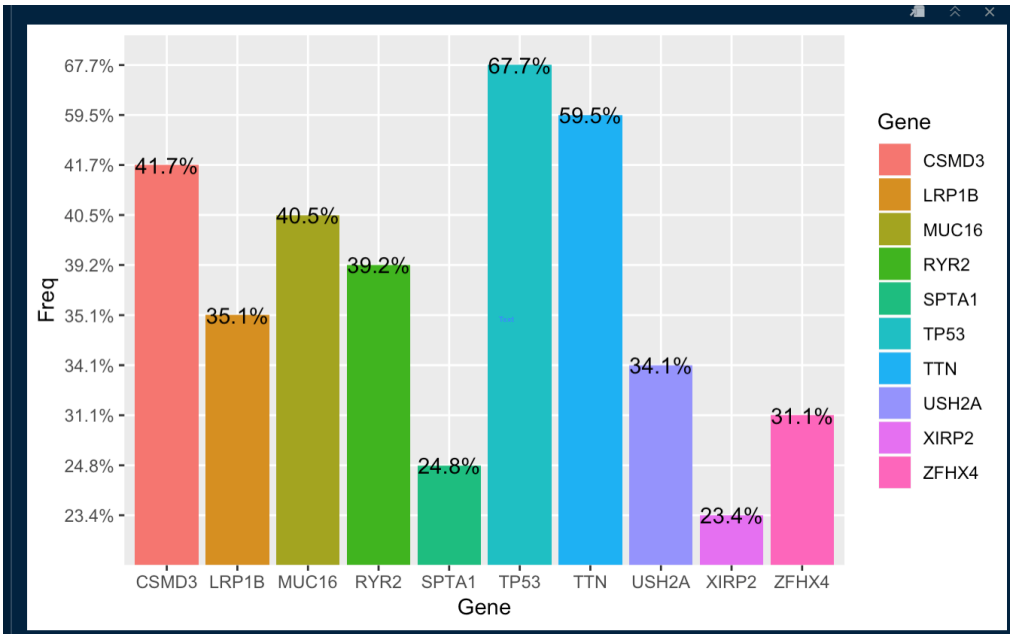

Figure 3: Lung Cancer

2) Prostate Cancer:

Prostate cancer is a type of that occurs with inside the prostate gland in men. This cancer most effectively takes place in men. Usually, this cancer grows slowly after which assaults the prostate gland. The parameters we're going to keep in mind on this prostate most cancers are radius, texture, perimeter, area, smoothness, compactness, symmetry etc.

3) Breast Cancer:

To detect this cancer, we are going to use the SVM model. In this cancer the parameters we are going to consider are clump thickness, uniform cell size, uniform cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli and mitosis we have plot correlation matrix of these features as shown in Figure.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

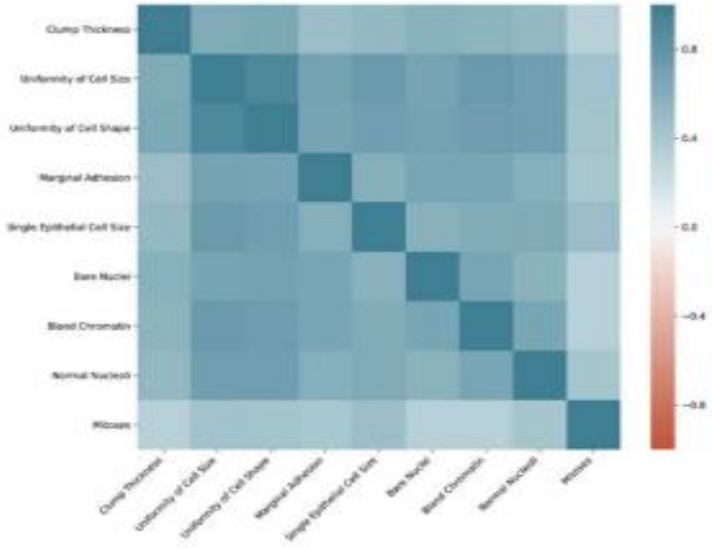

Figure 4: Breast Cancer

IV. RESULTS AND DISCUSSION

Our study has covered supervised machine learning techniques which include mainly the classification algorithms namely Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), Simple Linear Regression (SLR), Naive Bayes (NB), and K Nearest Neighbor (K-NN). This study has been implemented on two types of systems having a configuration as i3 & i5 processors with 4 and 8 GB RAM reapectively. Web page view allows end-users to give specific inputs which will be collected at the backend by forms (HTML) which will provide that data to python script through the use of Flask API. Again, python will redirect the data to the webpage for results after computing through the use of the render template method.

V. CONCLUSION

We concentrated on cancer prediction in this paper since it is an extremely serious disease that kills a lot of people all around the world. Breast cancer, lung cancer, and prostate cancer are the three types of cancer. In the field of Medicare and Biomedical, breast cancer prognosis is quite important. The goal of this work was to create a classifier that could predict the most serious malignancy, breast cancer. we developed a collaborative strategy for diagnosing this disease and providing information on the patient's condition. The breast cancer model as a classification job as is the development of the Support Vector Machine (SVM) approach to classify breast cancer as benign or malignant. Random forest classifier was employed in the lung cancer and prostate cancer prediction systems. The main aim of this paper was to develop a classifier that could predict the likelihood of a person developing lung or prostate cancer based on a set of common factors.

VI. REFERENCES

[1] S. Sayed, "Machine Learning Is The Future Of Cancer Prediction," 2018.

[2] H. Masood, "Breast Cancer Detection using Machine Learning Algorithm," International Research Journal of Engineering and Technology (IRJET), vol. 08, no. 02, 2021.

[3] V. N. Pham, "Lung cancer dataset and visualization," 22 February 2019

[4] A. D. Belsare and M. M. Mushrif, ”Histopathology Image Analysis Using Image Processing Technique”, publisher Research Gate, 2011

[5] Mahin Ghorbani and Hamed Karimi, “Role of Biotechnology in Cancer Control”, publisher Research Gate, 2015