Cancer Prediction using Machine Learning

Ganta Sruthi
Department of CSE
Chandigarh University
Chandigarh, India
19BCS4634@cuchd.in

Chokkakula Likitha Ram
Department of CSE
Chandigarh University
Chandigarh, India
19BCS4633@cuchd.in

Malegam Koushik Sai
Student, B. Tech., Dept. of CSE
Chandigarh University
Chandigarh, India
19BCS4640@cuchd.in

Bhanu Pratap Singh
Student, B. Tech., Dept. of CSE
Chandigarh University
Chandigarh, India
19BCS4627@cuchd.in

Nikhil Majhotra
Student, B. Tech., Dept. of CSE
Chandigarh University
Chandigarh, India
19BCS4628@cuchd.in

Neha Sharma
Department of CSE
Chandigarh University
Chandigarh, India
Nehasharma0110@gmail.com

2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM) | 978-1-6654-6643-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIPTM54933.2022.9754059

***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 going to give different attributes for three cancer system where the user has to enter data to get result. For breast cancer we are considering attributes like clump thickness, uniform cell size, uniform cell shape etc. and the prediction result will be whether the cancer is malignant or benign. For lung cancer, we are considering smoking, yellow fingers, anxiety, peer pressure etc. In prostate cancer, we are considering are radius, texture, perimeter, area etc. and the result for both cancer is likelihood of being affected by the cancer.**

Keywords— Machine Learning, Breast Cancer, Lungs Cancer, Prostate Cancer, Support vector machine (SVM), Random Forest, Data-set

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.

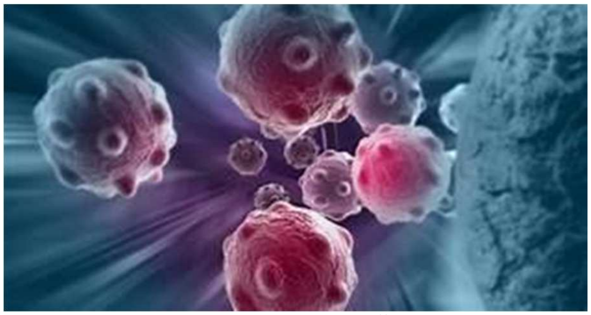


Figure 1.1 Cancer cells

Machine Learning involves creating a model that is trained on certain training data and can analyse other data to generate predictions [1]. 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.

This research paper is basically about Cancer prediction using machine learning techniques. 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. We are going to attach the images of these three cancer's data sets.

According to data provided by World health organization (WHO) two million new cases are reported and which of 626,679 were died in year 2018 [2]. At this situation, the importance of machine learning can be realized. [3]

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

978-1-6654-6643-1/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: Eindhoven University of Technology. Downloaded on September 09,2022 at 09:36:27 UTC from IEEE Xplore. Restrictions apply.

chromatin, normal nucleoli and mitosis we have plot correlation matrix of these features as shown in Figure 1.2 [4]

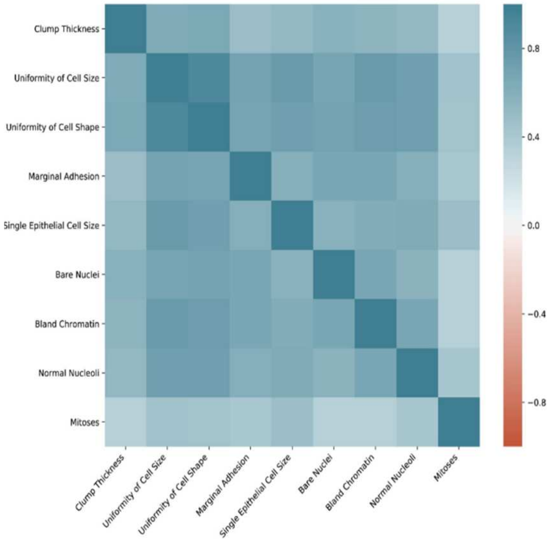


Figure 1.2 Feature Corelation of Breast cancer

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 1.3 [5].

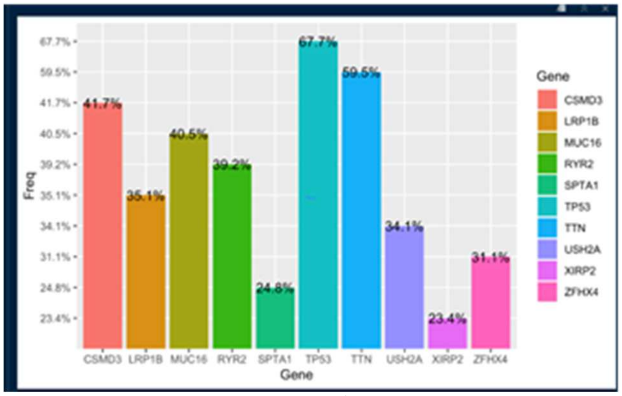


Figure 1.3 Dataset of Lung Cancer

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. we can also see the accuracy of different algorithms in Figure1. 4 [6].

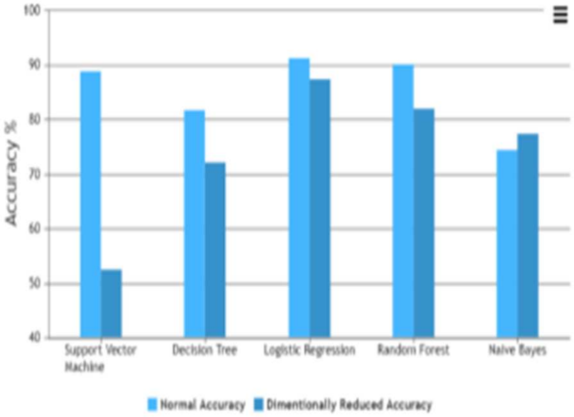


Figure 1.4 Accuracy Comparison of prostate cancer

In the analysis of massive volumes of data, machine learning plays a critical role. Without depending on a precise instruction set, computer systems use machine learning to do a given specific task based on trends and specific patterns[12]. We can easily say that machine learning is safe to be used in different fields to generate the perfect and speedier results with the given data. By using these techniques doctors can detect cancer at a much earlier stage. In this research paper, we are going to use the pictures of cancerous cells for the training and testing in order to analyse them into malignant or benign cells, to get the best result as soon as possible.

II. Requirements

A. *Existing System*

Different researchers used different methods and technologies to carry out the process of Cancer prediction system. Some of the important research paper Effectiveness of –

Data Mining-based Cancer Prediction System [7] is a system that evaluates the risk of breast, skin, and lung cancers. This model is useful because it presents multiple cancer models. It even takes into account both genetic and non-genetic data to calculate risk. However, because the Weka toolkit can only handle tiny data sets, the actual use of this model is limited due to the sophisticated numerical solutions that are required. As a result, we are attempting to resolve this issue.

Automated Melanoma Recognition in Dermoscopy Images Using Very Deep Residual Networks [8] is a paper that uses very deep CNNs to overcome the hurdles of this system by using in dermoscopy images. It is ideal because deep neural

Authorized licensed use limited to: Eindhoven University of Technology. Downloaded on September 09,2022 at 09:36:27 UTC from IEEE Xplore. Restrictions apply.

networks allow the model to learn regardless of data limitations, but they take a significant amount of processing power to train. We are trying to minimize the amount of power in our model.

TABLE I. LITERATURE REVIEW

Project Title	Year	Algorithm	Advantages	Disadvantages
Comparison of Logistic Regression and Artificial Neural Network Models in Breast Cancer Risk Estimation [7]	2010	ANN and logistic regression	They are presenting the idea of cancer on the basis of two models with different techniques and comparing them	Creating an ANN based less domain knowledge than creating a logistic regression mode
Effectiveness of Data Mining - based Cancer Prediction System [8]	2013	Data mining WEKA	It even considers factors like genetic and non- genetic information and estimate risk level	As they are using weka tool kit, it can handle only small datasets
Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks [9]	2016	Deep learning, CNN, MAT LAB	Deeper neural networks allow the model to learn irrespective of the data limitations	need lots of computational power to train networks.
Lung Cancer Detection using Deep Convolutional Networks [10]	2018	Deep learning and neural networks	help doctors make better and informed decisions when diagnosing lung cancer	Requires large amount of database
Cancer Prediction Using Machine Learning Algorithms [11]	2019	SVM	Used to classify the normal person and Tumor patient.	Not suitable when we use large data sets.

III. PROPOSED METHODOLOGY

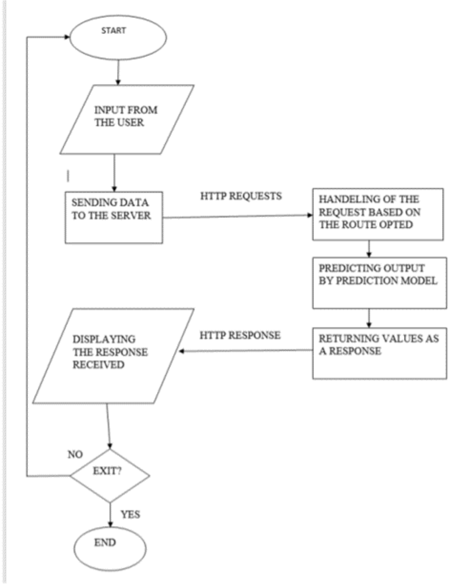

START
INPUT FROM THE USER
SENDING DATA TO THE SERVER
HTTP REQUESTS
HANDELING OF THE REQUEST BASED ON THE ROUTE OPTED
PREDICTING OUTPUT BY PREDICTION MODEL
RETURNING VALUES AS A RESPONSE
HTTP RESPONSE
DISPLAYING THE RESPONSE RECEIVED
NO
EXIT?
YES
END

Figure 3.1 FLOW CHART

Algorithm

1. The webpage loads, and the user is offered the option of selecting one of three cancer types.

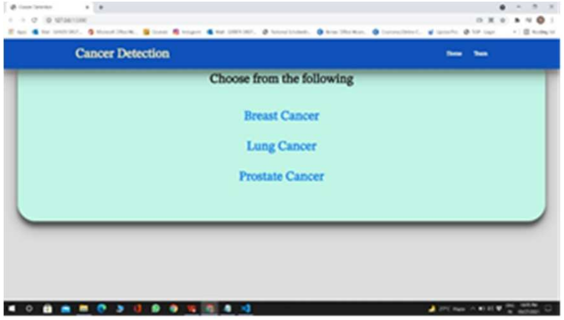
2. The user must select the cancer prediction system that he or she wishes to utilise.

3. The user must next complete the form based on their symptoms.

4. The data is then sent to the route that deals with machine learning models, and the result is returned on the following route.

- *Webpages of the project*

This is the homepage of the website in which we can see breast cancer, lung cancer and prostate cancer.


Cancer Detection

Choose from the following

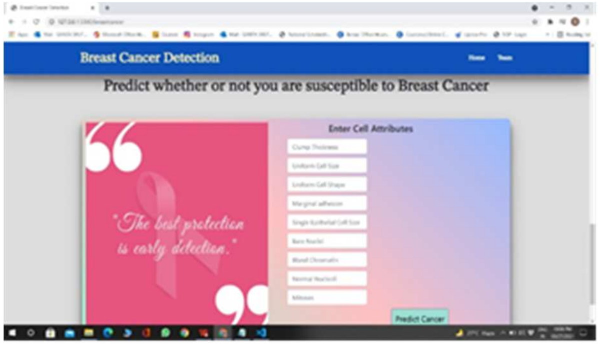
Breast Cancer

Lung Cancer

Prostate Cancer

Figure 3.2 Home Page

This webpage is of breast cancer. As we know that pink colour ribbon indicates breast cancer. In this webpage, we can see the parameters which are used to check whether the cancer is malignant or benign.


Breast Cancer Detection

Predict whether or not you are susceptible to Breast Cancer

"The best protection is early detection."

Enter Cell Attributes

Clump Thickness

Uniform Cell Size

Uniform Cell Shape

Marginal adhesion

Single Epithelial Cell Size

Bare Nuclei

Bland Chromatin

Normal Nucleoli

Mitoses

Predict Cancer

Figure 3.3 Breast cancer page

This webpage is of Lung cancer. As we know that grey colour ribbon indicates the lung cancer. In this webpage we can see the parameters which are used to predict whether the person is suspectable to lung cancer or not.

Authorized licensed use limited to: Eindhoven University of Technology. Downloaded on September 09,2022 at 09:36:27 UTC from IEEE Xplore. Restrictions apply.

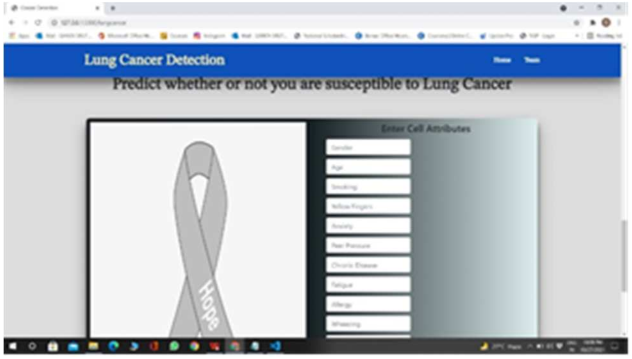


Figure 3.4 Lung Cancer page

This webpage is of prostate cancer. As we know that blue colour ribbon indicates the lung cancer. In this webpage we can see the parameters which are used to predict whether the person is suspectable to prostate cancer or not.

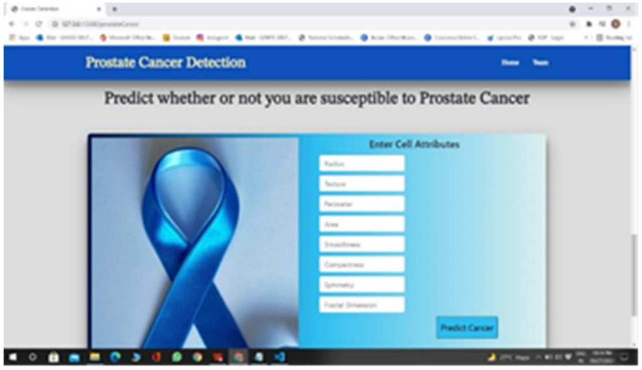
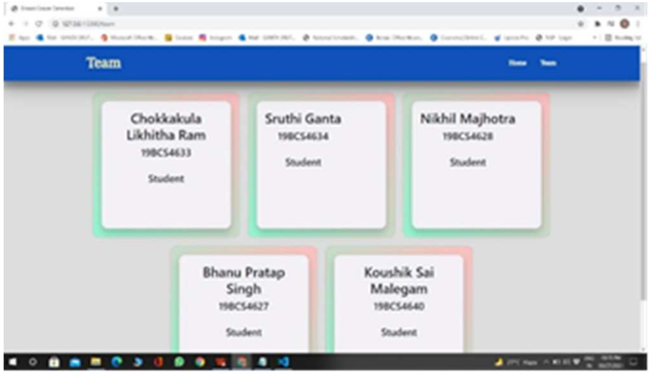


Figure 3.5 Prostate cancer page



Chokkakula Likhitha Ram
19BCS4633
Student

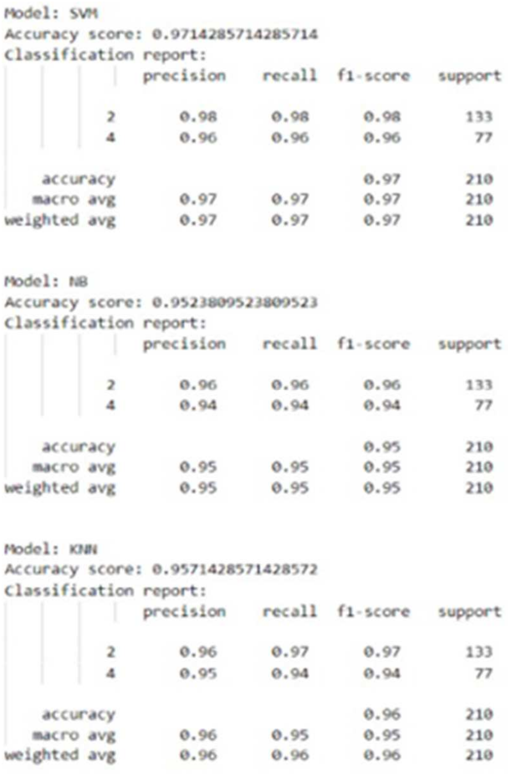
Sruthi Ganta
19BCS4634
Student

Nikhil Majhotra
19BCS4628
Student

Bhanu Pratap Singh
19BCS4627
Student

Koushik Sai Malegam
19BCS4640
Student

Figure 3.6 Team Page



```
Model: SVM
Accuracy score: 0.9714285714285714
Classification report:
              precision    recall  f1-score   support

           2       0.98      0.98      0.98       133
           4       0.96      0.96      0.96        77

    accuracy                           0.97       210
   macro avg       0.97      0.97      0.97       210
weighted avg       0.97      0.97      0.97       210

Model: NB
Accuracy score: 0.9523809523809523
Classification report:
              precision    recall  f1-score   support

           2       0.96      0.96      0.96       133
           4       0.94      0.94      0.94        77

    accuracy                           0.95       210
   macro avg       0.95      0.95      0.95       210
weighted avg       0.95      0.95      0.95       210

Model: KNN
Accuracy score: 0.9571428571428572
Classification report:
              precision    recall  f1-score   support

           2       0.96      0.97      0.97       133
           4       0.95      0.94      0.94        77

    accuracy                           0.96       210
   macro avg       0.96      0.95      0.95       210
weighted avg       0.96      0.96      0.96       210
```

Figure 3.7 Testing result of Algorithm

IV. 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.

V. LIMITATION AND FUTURE SCOPE

We have some project limitations that will need to be addressed in the future, as listed below. Because the entire project is currently offline, the project's scope is constrained. More cancer prediction models can be introduced as time goes on. There is no database to keep track of the values entered by the user.

Authorized licensed use limited to: Eindhoven University of Technology. Downloaded on September 09,2022 at 09:36:27 UTC from IEEE Xplore. Restrictions apply.

In the future, we will be able to improve our project. The project's scope can be hosted to expand the project's scope. It is possible to predict cancer using image or X-ray data. A user database can be kept up to date so that they can learn and improve.

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] Md. Milon Islam, Md. Rezwanul Haque, Hasib Iqbal, Md. Munirul Hasan, Mahmudul Hasan and Muhammad Nomani Kabir , "Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques," 01 September 2020.

[4] A. Jain, "Exploring Breast Cancer Data set," 29 April 2018.

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

[6] Georgina Cosma, Stéphanie E. McArdle, Stephen Reeder, Gemma A. Foulds and Simon Hood, "Identifying the Presence of Prostate Cancer in Individuals Using Computational Data Extraction Analysis of High Dimensional Peripheral Blood Flow Cytometric Phenotyping Data," 18 December 2017.

[7] A.Priyanga and S.Prakasam, "Effectiveness of Data Mining - based Cancer Prediction System," International Journal of Computer Applications, vol. 83, 2013.

[8] lequan hu, hao chen, Qi dou and jing qin, "Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks," IEEE Transactions on Medical Imaging , 2016.

[9] Turgay Ayer, Jagpreet Chhatwal, Oguzhan Alagoz, PhD Charles E. Kahn, MS, Ryan W. Woods and Elizabeth S. Burnside, "Comparison of Logistic Regression and Artificial Neural Network Models in Breast Cancer Risk Estimation," 2010.

[10] J. salomon, "Lung Cancer Detection using Deep Convolutional Networks," 2018.

[11] M. Agrawal, " Cancer Prediction Using Machine Learning Algorithms," International Journal of Science and Research, 2018.

[12] Sharma, S., Mishra, V.M. Development of Sleep Apnea Device by detection of blood pressure and heart rate measurement. *Int J Syst Assur Eng Manag* **12,** 145–153 (2021). https://doi.org/10.1007/s13198-020-01041-3

Authorized licensed use limited to: Eindhoven University of Technology. Downloaded on September 09,2022 at 09:36:27 UTC from IEEE Xplore. Restrictions apply.