**A**

**Synopsis**

**on**

**Breast Cancer Prediction**

**Submitted by**

**Group. No.**

**BHARTI GHILDIYAL- 2100290140051**

**DIKSHA GUPTA -2100290140060**

**ANUPAM GUPTA- 2100290140033**

**AYUSH GUPTA- 2100290140044**



**Submitted to – Dr Sangeeta Arora**

**Department of Computer Applications,**

**KIET Group of Institutions,**

**Delhi-NCR, Ghaziabad**

**September’ 2022**

**ABSTRACT**

Breast Cancer is the most leading malignancy

Breast Cancer is the most leading malignancy affecting 2.1 million women each year which leads to greatest number of deaths among women. Early treatment not only helps to cure cancer but also helps in prevention of its recurrence. And hence this system mainly focuses on prediction of breast cancer where it uses K-nearest neighbors (KNN)for creatingrandom forest which

are applied on pre-processed data which suspects greater accuracy for prediction. Amongst all the models, Random Forest Classification leads to best accuracy with 98.6%. These techniques are coded in python and uses numpy, pandas, seaborn libraries.

**Table of Contents**

1. Introduction
2. Technologies / Software Requirements
3. Hardware requirement / Hardware Used
4. Modules Description
5. Reports / Outputs
6. Conclusion
7. How to serve the society
8. Gantt Chart (In terms of weeks)

**Introduction**

Breast cancer has become the most recurrent type of health issue among women especially for women in middle age. Early detection of breast cancer can help women cure this disease and death rate can be reduced. In the present-day scenario, to observe breast cancer mammograms are used and they are known be the most effective scanning technique. The leading risk factor for Breast cancer is simply being a women. Men can also get breast cancer In 2017, the American Cancer Society estimates 2,470 new case of invasive breast cancer will be diagnosed in men in US.

The early diagnosis of BC can improve the prognosis and chance of survival significantly, as it can promote timely clinical treatment to patients. Further accurate classification of benign tumors can prevent patients undergoing unnecessary treatments. Thus, the correct diagnosis of BC and classification of patients into malignant or benign groups is the subject of much research. Because of its unique advantages in critical features detection from complex BC datasets, machine learning (ML) is widely recognized as the methodology of choice in BC pattern classification and forecast modelling.

We will use the Machine Learning Repository for breast cancer dataset.

[**Give the Reference**] The dataset used in this story is publicly available and was created by Dr. William H. Wolberg, physician at the University Of Wisconsin Hospital at Madison, Wisconsin, USA. To create the dataset Dr. Wolberg used fluid samples, taken from patients with solid breast masses and an easy-to-use graphical computer program called Xcyt, which is capable of perform the analysis of cytological features based on a digital scan. The program uses a k Nearest Neighbor algorithm, to compute ten features from each one of the cells in the sample, than it calculates the mean value, extreme value and standard error of each feature for the image, returning a 30 real-valuated vector

K-nearest neighbors (KNN) is a type of supervised learning algorithm used for both regression and classification. KNN tries to predict the correct class for the test data by calculating the distance between the test data and all the training points. Then select the K number of points which is closet to the test data. The KNN algorithm calculates the probability of the test data belonging to the classes of ‘K’ training data and class holds the highest probability will be selected. In the case of regression, the value is the mean of the ‘K’ selected training points.

Attribute Information:

1. ID number 2) Diagnosis (M = malignant, B = benign) 3–32)

Ten real-valued features are computed for each cell nucleus:

1. radius (mean of distances from center to points on the perimeter)
2. texture (standard deviation of gray-scale values)
3. perimeter
4. area
5. smoothness (local variation in radius lengths)
6. compactness (perimeter² / area — 1.0)
7. concavity (severity of concave portions of the contour)

concave points (number of concave portions of the contour)

1. symmetry
2. fractal dimension (“coastline approximation” — 1)

The mean, standard error and “worst” or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features. For instance, field 3 is Mean Radius, field 13 is Radius SE, field 23 is Worst Radius.

Software Requirements

Operating System : Linux ,Windows (8,7,10,11) [Anyone]

Application Required : Visual Studio code

Technology : Python language

Library Used : NumPy, Pandas, SkLearn

Framework : Flask

Documentation : MS Word

Hardware Requirements

Processors : i3 or above

Clock Speed : 3.0ghz

RAM : 4GB or above

Hard Disk capacity : 500GB or above

Peripheral devices

Modules Description

Login Module:-

* In this module user will be able to create a account in a system and able to access the service.
* In this module user will create a account from his email and password , after successful login user will be able to analyze the breast cancer disease .
* If authentication is successful then the user is directed to the first web page .otherwise the user remains on the login page.

Machine Learning Module:-

* In this module we will upload the Dataset file will be of excel type , in this dataset there will be n number of record which will used to analyze the breast cancer of the user .
* User will interact in this module by filling the following details; Texture mean, Area mean, Concavity mean, Area SE, Concavity SE, Fractal dimension SE, Smoothness worst, Concavity worst, Symmetry worst, , Fractal dimension worst then after on predict button user will be able to know that there having breast cancer or not.
* Now user can see the pie chart by clicking on analyze button

Output

* Prediction- our project will predict the cancer in a patient is having Malignant or benign cancer is being developed or not
* Probability- The chances of tumour is cancerous or not.
* If the probability is more than 50%,the tumour is malignant and if the probability is less than 50%, it is benign
* Final output will come in the form of pie chart on the basis of parameter which is used

Conclusion

Breast Cancer represents one of the diseases that makes highest number of deaths every year. At present, only few accurate prognostic and predictive factors are used clinically for managing the patients with breast cancer. Here, by making use of Clustering with Level Set approach, high accuracy can be achieved in detection of effected cell shapes with exact marking on detected contours. The proposed system helps to enhance the performance of mammogram retrieval by selecting optimal features A decision support system for predicting breast cancer helps and assist physicians in making optimum, accurate and timely decision and reduce the overall cost of treatment.

KNN classifier has been used as it yields the highest classification accuracies, The system greatly reduces the cost of treatment and improves the quality of life by predicting breast cancer at early stage.

How to serve the Society

In India, the incidence of cancer is increasing rapidly; therefore it is important to step up cancer literacy and knowledge amongst the population.

We hope that this will lead to early detection which is important in the management and treatment of cancer and also prevention by making necessary changes in lifestyle.

Breast Cancer is rapidly rising amongst women in urban areas and it is the most common cancer in India. Early detection is essential in its management and treatment. Due to ignorance, fear, and social stigma many women present with stage III or IV cancer which requires surgery and extensive treatment.

Gantt Chart

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Feasible study | Week 1 |  |  |  |  |  |
| Requirement Analyze |  | Week 2 |  |  |  |  |
| Design |  |  | Week 3 |  |  |  |
| Coding |  |  |  | Week 4 |  |  |
| Testing |  |  |  |  | Week 5 |  |
| Maintaince |  |  |  |  |  | Week 6 |
| Final Report |  |  |  |  |  | Week 7 |