BREAST CANCER DETECTION

Date: 19-12-2022

PROJECT REPORT

Submitted by

CH.EN.U4AIE21014	G CHAITHANYA
CH.EN.U4AIE21016	Y HEMANTH
CH.EN.U4AIE21024	K KOUSHIK
CH.EN.U4AIE21034	P HANEEF
CH.EN.U4AIE21046	S PRANAV REDDY

Under the guidance of

Dr. I R Oviya

In partial fulfilment for the award of the degree

Of

BACHELOR OF TECHNOLOGY

In

INTELLIGENCE IN BIOLOGICAL SYSTEMS

AMRITA SCHOOL OF COMPUTING, CHENNAI

AMRITA VISHWA VIDYAPEETHAM

CHENNAI – 601103

DECEMBER 2022

STUDENT DECLARATION

I hereby declare that my Project Report titled **Breast Cancer Detection** is a bonafide record for the course **21BIO211 – Intelligence of Biological Systems 3** which I have submitted to Department of Computer Science and Engineering (AIE), Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai in partial fulfillment of the credit requirements for the B.Tech. degree, is my authentic work done under the guidance of **Dr. I R Oviya**. This project report has not been copied, duplicated or plagiarised from any other paper, journal, document or book and has not been submitted to any educational institute, course, department or otherwise for the award of any credit, certificate, diploma, degree or recognition. This is an authentic piece of work and in case there is any query regarding the same, I shall be held responsible for answering any queries in this regard.

Date:20-12-2022

Student name and Register no.:

CH.EN.U4AIE21014	G CHAITHANYA
CH.EN.U4AIE21016	Y HEMANTH
CH.EN.U4AIE21024	K KOUSHIK
CH.EN.U4AIE21034	P HANEEF
CH.EN.U4AIE21046	S PRANAV REDDY

Supervisor's Name:

ACKNOWLEDGEMENT

BREAST CANCER DETECTION

AMRITA SCHOOL OF ENGINEERING CHENNAI.
BY AIE DEPARTMENT
CSE AIE-A TEAM-10
Y Hemanth(21016)
G Chaithanya (21014)
K Koushik (21024)
P Haneef(21034)
S Pranav Reddy(21046)

Abstract:

Every year, 2.1 million women are affected by the most common cancer, breast cancer, which kills the most women. Early diagnosis and therapy not only aid in the treatment of cancer but also in preventing its recurrence. As a result, this system's primary focus is on the prediction of breast cancer, and it does so by using various machine learning algorithms to build models like decision trees, logistic regression, and random forests that are applied to pre-processed data and are therefore thought to have a higher prediction accuracy. With 98.6% accuracy, Random Forest Classification outperforms all other models. These methods make use of the Numpy, Pandas, and Seaborn libraries and are written in Python. **Technical Terms: Decision Tree, Logistic Regression, Random Forest Classification, Numpy, pandas, seaborn.**

1.Introduction:

Breast cancer is the second most lethal cancer after lung cancer and the most common cancer among women, according to the World Health Organization. According to studies, breast cancer claimed the lives of 627,000 women in 2018, accounting for 15% of all female cancer fatalities. People contact an oncologist if they experience any symptoms. Using breast ultrasonography, diagnostic mammography, magnetic resonance imaging (MRI), and biopsies, doctors can quickly and easily diagnose breast cancer. The doctor can suggest additional testing or treatment based on the results of these tests. Breast cancer early detection is vital. The likelihood that a patient would survive may increase if cancer risk is predicted early. Using machine learning to detect breast cancer is an alternative method.algorithms for prediction of abnormal tumor. Thus, the research is carried out for the proper diagnosis and categorization of patients into malignant and benign groups. independent document. Please do not revise any of the current designations.

What is Breast Cancer?

One type of cancer that begins in the breast is breast cancer. Both breasts can be the first place it starts. When cells start to multiply uncontrollably, cancer develops. (To discover more about the origins and progression of cancer. Even while breast cancer mostly affects women, it can also affect men. It's crucial to realise that the majority of breast lumps are benign and not cancerous (malignant). Breast tumours that are not cancerous are abnormal growths that do not spread to the exterior of the breast. While most benign breast lumps are not life threatening, some of them can raise a woman's risk of developing breast cancer. A medical professional must examine any breast lump or change to determine whether it is benign. or malignant (cancer) and if it might affect your future cancer risk.

Where breast cancer starts?

Different areas of the breast might serve as the origin of breast cancer. An organ, the breast is located over the upper ribs and chest muscles. Both the left and right breasts include mostly fatty tissue, ducts, and glands. For the purpose of feeding newborns and infants, women's breasts produce and secrete milk. Each breast's size is based on the quantity of fatty tissue .reast cancers can start from different parts of the breast. The breast is an organ that sits on top of the upper ribs and chest muscles. There is a left and right breast and each one has

mainly glands, ducts, and fatty tissue. In women, the breast makes and delivers milk to feed new borns and infants. The amount of fatty tissue in the breast determines the size of each breast.

The breast has different parts:

- **Lobules** are the glands that make breast milk. Cancers that start here are called **lobular cancers**.

- **Ducts** are small canals that come out from the lobules and carry the milk to the nipple. This is the most common place for breast cancer to start. Cancers that start here are called **ductal cancers**.

- The **nipple** is the opening in the skin of the breast where the ducts come together and turn into larger ducts so the milk can leave the breast. The nipple is surrounded by slightly darker thicker skin called the **areola**. A less common type of breast cancer called **Paget disease of the breast** can start in the nipple.

- The **fat and connective tissue (stroma)** surround the ducts and lobules and help keep them in place. A less common type of breast cancer called **phyllodes Tumor** can start in the stroma.

- **Blood vessels** and **lymph vessels** are also found in each breast. **Angiosarcoma** is a less common type of breast cancer that can start in the lining of these vessels. The lymph system is described below.
  - *A small number of cancers arise from other tissues in the breast. These cancers are

called sarcomas and lymphomas and are not actually considered breast cancer.

- The lymph (or lymphatic) system is a part of your body's immune system. It is a network of lymph nodes (small, bean-sized glands), ducts

- Or blood vessels and organs that work together to collect clear lymph and carry it through the body's tissues to the blood. The clear lymph fluid within the lymphatic vessels contains tissue by products and waste products, as well as immune system cells.

- The lymph vessels carry lymph fluid away from the breast. In the case of breast cancer, cancer cells can enter those lymph vessels and start to grow in lymph nodes. Most of the lymph vessels of the breast drain into:

- Lymph nodes under the arm (axillary lymph nodes)

- Lymph nodes inside the chest near the breastbone (internal mammary lymph nodes)

- Lymph nodes around the collar bone
  (supraclavicular [above the collar bone] and infraclavicular [below the collar bone] lymph nodes)
  - If cancer cells have spread to the lymph nodes, it is more likely that the cells have passed through the lymph system and spread (metastasized) to other parts of the body. However, not all women with cancer cells in their lymph nodes develop
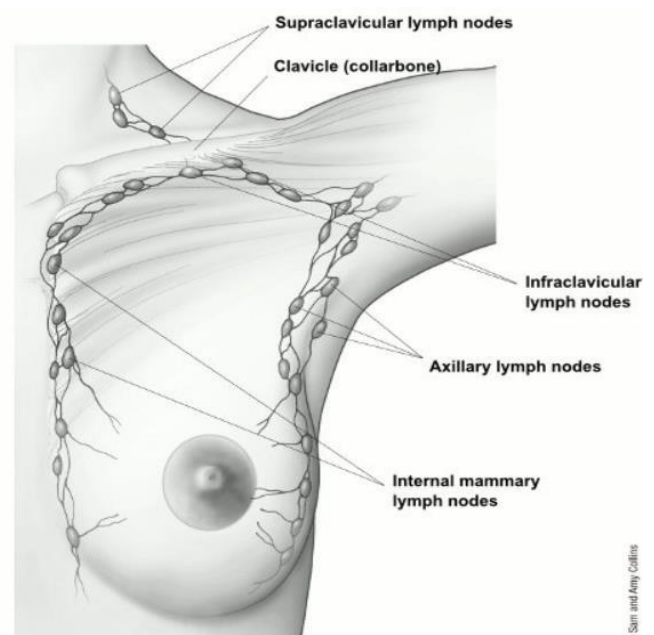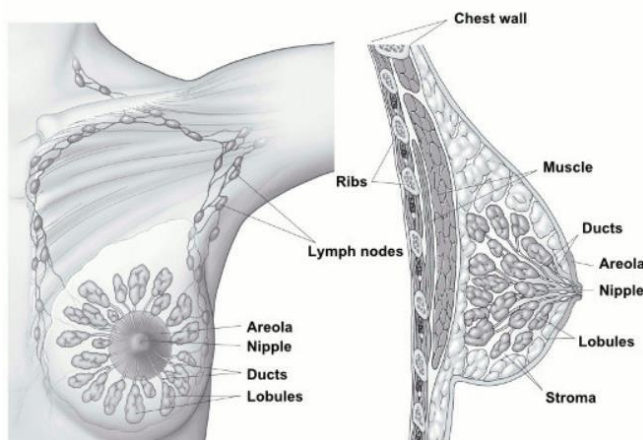
How the Breast cancer spreads?

- Breast cancer can spread when the cancer cells get into the blood or lymph system and then are carried to other parts of the body.

metastases, and some women without cancer cells in their lymph nodes may later develop metastases.
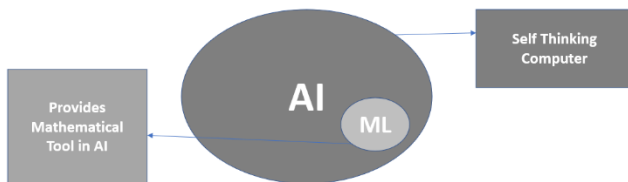
Now we are using Machine Learning Technique to detect the breast cancer
Why we are using Machine learning?
Machine Learning (ML) is a field of artificial intelligence that uses statistical techniques to give computer systems the ability to "learn" (eg.,progressively improve performance on a specific task) from data, without being explicitly programmed.

Classification Using Decision Tree
Decision trees are nonparametric supervised learning algorithms used for both classification and regression tasks. It has a hierarchical tree structure consisting of root nodes, branches, inner nodes and leaf nodes.

Example:-
- The best example is buying something on an online shopping portal where you get

multiple recommendations based on what you purchased.
- decide what to wear.
- Decide what to eat for lunch.
- choose a book to read.
- Decide what to do next.

About Data set We used
The purpose of this Data set is to observe the most predictive traits of malignant and benign cancers and identify general trends that aid in model selection and hyperparameter selection. The goal is to classify breast cancer as benign or malignant. To achieve this, we used machine learning classification methods to fit a function that can predict discrete classes of new inputs. **(The original Wisconsin-Breast Cancer (Diagnostics) dataset (WBC) from machine learning repository is a classification dataset, which records the measurements for breast cancer cases. There are two classes, benign and malignant.)**

Our dataset consists of a total of 569 samples. There are a total of 32 features characterizing the
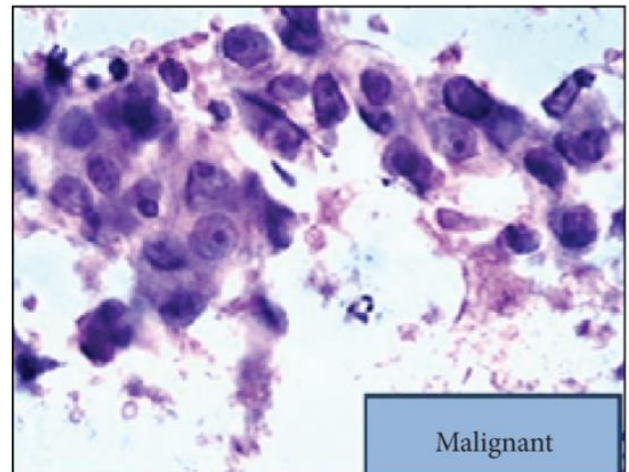
sample. The first feature is the sample's ID, the second is its class, and the remaining 30 features contain various information about the cell. The sample class designation is malignant (M) or benign (B). These are medical terms that refer to the benign and malignant tumor cells previously described. The property has no missing values. Of
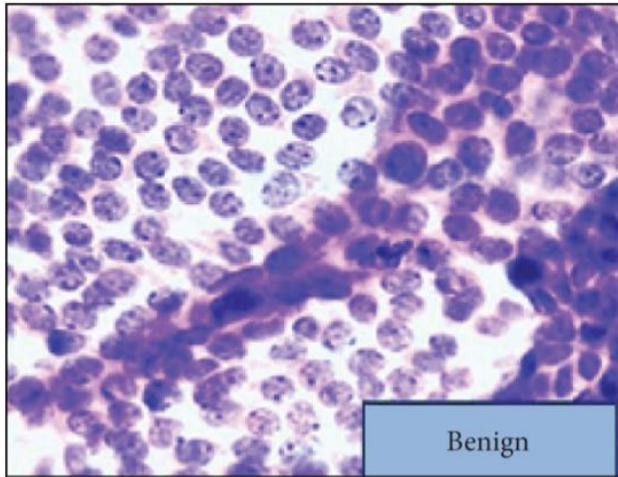
Benign

our samples, 357 are estimated to be benign and 212 to be malignant. The classification success rate increased to a value of 97.5% thanks to a 10-fold cross-validated classifier that was created correctly on the data. We correctly classified 176 patients who came before November 1995.

RESULT:-

In this project we have taken around 589 members of Breast cell dimension from the popular data set called WBCD(Wisiconsin Breast Cancer Diagnostic) for these we have found the 357 masses are benign and 212 are malignant and we have also found the 32 features of a cell to classify the cancer. And also we have used Decision tree it's a ML technique use as classifier. Now as a result we need to enter the all the 32 features of the person who need to find wheather they have cancer or not. And result will be displayed. Till now we are getting 92% accuracy in future we may use some more ML techniques as a classifier and increase the accuracy.

REFERENCES:-

WEBISTES:-

- https://www.cancer.gov/types/breast/patient/breast-screening-pdq
- https://www.cdc.gov/cancer/breast/basic_info/diagnosis.htm
- https://www.ncbi.nlm.nih.gov/books/NBK482286/
- https://www.healthdirect.gov.au/breast-cancer

- https://tmc.gov.in/index.php/en/cbe-press-release
- https://www.researchgate.net/profile/Ronak-Sumbaly-2/publication/272863357_Diagnosis_of_Breast_Cancer_using_Decision_Tree_Data_Mining_Technique/links/54f856820cf2cffe9de7b24/Diagnosis-of-Breast-Cancer-using-Decision-Tree-Data-Mining-Technique.pdf
- https://dl.acm.org/doi/abs/10.1145/3056662.3056716
- https://www.sciencedirect.com/science/article/pii/S0933365702000866

Reference papers:-

- Breast Cancer Screening in Low-Income Countries: A New Program for Downstaging Breast Cancer in Tanzania (Hindawi)
- Saving the Breast Saves the Lives of Breast Cancer Patients(Hindawi)