**Synopsis**

**On**

**PREDICTION OF BREST CANCER**

**USING MACHINE LEARNING**

in partial fulfilment for the award of the degree of

**BACHELOR OF ENGINEERING**

IN

**COMPUTER SCIENCE AND ENGINEERING**

**Submitted by:**

Harsh Pandey (22BCS17030)

**Under the Guidance of**

**INSTITUTE – UIE**

**ACADEMIC UNIT- V**

****

**Chandigarh University**

**January 2023**

**Index**

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Content** | **Page No.** |
| **1** | Introduction | 3-4 |
| **2** | Problem statement | 4-5 |
| **3** | Key Features/Benefits | 5 |
| **4** | Software used | 6 |
| **5** | Deliverables | 7 |

1. **Introduction**

Breast Cancer is the most widely recognised type of cancer and is the main kind of malignant growth that is boundless among ladies around the world. As indicated by the report of WHO (World Health Organization) every year, 2.1 million ladies are recently affected by breast cancer.

Furthermore, there should be a proper prediction and classification about the type of cancer a patient is having so doctor may know whether surgery is required or not. Such classification will help doctor to do surgeries only when the cancer is malignant, thus reducing the unnecessary surgeries.

The project is about building a Logistic Regression Model using Sklearn to classify breast cancer as either Malignant or Benign. So, our first step is data collection. And once the data collection process is completed.

We will process this data because we cannot feed this raw data to our Machine Learning model.

So, there are some data pre-processing steps which we will cover here and once we process the data then we need to split our data into Training data and Testing data; so the purpose of this is we will train our Machine Learning model using the Training data and then we will evaluate our model using Test data and this is the only purpose of splitting the data into Training data and Testing data. After this case, we are going to use a Logistic Regression model. Logistic Regression is one of the best models when it comes to binary classification as we are going to classify the data points into two types here.

The two types are benign (1) and malignant (0).

So, once you train your Logistic Regression model with this particular Training dataset, we will have a trained Logistic Regression model and we will also do some evaluation of this model using Test data. After that we can have a new data and once you give this new data the model can tell you whether that particular tumour is **benign** or **malignant**.

1. **Problem statement**

Breast cancer is the most frequent malignancy in women worldwide and is curable

in ~70–80% of patients with early- stage, non- metastatic disease. Advanced breast cancer

with distant organ metastases is considered incurable with currently available therapies.

On the molecular level, breast cancer is a heterogeneous disease; molecular features include

activation of human epidermal growth factor receptor 2 (HER2, encoded by ERBB2), activation

of hormone receptors (oestrogen receptor and progesterone receptor) and/or BRCA mutations.

Treatment strategies differ according to molecular subtype. Management of breast cancer is

multidisciplinary ; it includes locoregional (surgery and radiation therapy) and systemic therapy

approaches. Systemic therapies include endocrine therapy for hormone receptor- positive

disease, chemotherapy , anti- HER2 therapy for HER2-positive disease, bone stabilizing agents,

poly(ADP- ribose) polymerase inhibitors for BRCA mutation carriers and, quite recently,

immunotherapy. Future therapeutic concepts in breast cancer aim at individualization of

therapy as well as at treatment de- escalation and escalation based on tumour biology and

early therapy response. Next to further treatment innovations, equal worldwide access to

therapeutic advances remains the global challenge in breast cancer care for the future

Breast cancer is the most frequent malignancy in women worldwide and is curable

in ~70–80% of patients with early- stage, non- metastatic disease. Advanced breast cancer

with distant organ metastases is considered incurable with currently available therapies.

On the molecular level, breast cancer is a heterogeneous disease; molecular features include

activation of human epidermal growth factor receptor 2 (HER2, encoded by ERBB2), activation

of hormone receptors (oestrogen receptor and progesterone receptor) and/or BRCA mutations.

Treatment strategies differ according to molecular subtype. Management of breast cancer is

multidisciplinary ; it includes locoregional (surgery and radiation therapy) and systemic therapy

approaches. Systemic therapies include endocrine therapy for hormone receptor- positive

disease, chemotherapy , anti- HER2 therapy for HER2-positive disease, bone stabilizing agents,

poly(ADP- ribose) polymerase inhibitors for BRCA mutation carriers and, quite recently,

immunotherapy. Future therapeutic concepts in breast cancer aim at individualization of

therapy as well as at treatment de- escalation and escalation based on tumour biology and

early therapy response. Next to further treatment innovations, equal worldwide access to

therapeutic advances remains the global challenge in breast cancer care for the future

Breast cancer is the most common type of cancer in women. When cancers are found early, they can often be cured. There are some devices that detects the breast cancer but many times they lead to false positives, which results in patients undergoing painful, expensive surgeries that were not even necessary. These types of cancer are called **Benign** which do not require surgeries and we can reduce these unnecessary surgeries by using Machine Learning. We take a classified dataset of the previous breast cancer patients and train the model to predict whether the cancer is **benign** or **malignant**. These classifications will help doctors to do surgeries only when the cancer is **malignant**, thus reducing the unnecessary surgeries for woman.

1. **Key Features/Benefits**

This model can be used for compiling as well as prediction of the data of **benign** and **malignant** cancer cells and their patients**.**

Doctors can easily find the record of a particular patient with the tumour using this model.

Interpretability of the function modelled by ML algorithms is only partially limited by the “black-box” nature of ML algorithms in our study because we included a limited number of well-established breast cancer risk factors.

The improvement achieved with this algorithm in accurate classification of women with and without breast cancer.

1. **Software Used**

Google Collaboratory,

Breast cancer data.csv File from kaggle

Sublime text.

* **Language and Libraries Used**

Python: language

NumPy: library for numerical calculations

Pandas: library for data manipulation and analysis

SkLearn: library which features various classification,

regression and clustering algorithms.

1. **Deliverables**

In this project, we learned to build a breast cancer classifier and predicter on the Breast cancer dataset imported from the **University of Wisconsin Hospital** for the diagnosed patient and can become helpful for the doctors to treat their disease accordingly.

This analysis is to observe which features are most helpful in predicting **malignant** or **benign** cancer and trends that may aid us in model selection and hyper parameter selection. Through this system, we can easily able to determine the past record of the patient whether she has **benign** or **malignant** tumour.

And further it can predict on the basis of past data that the new patient has or have the cancer tumour of which category whether it is **benign** or **malignant**.