# BREAST CANCER DETECTION USING DEEP LEARNING TECHNIQUES

**By –**

Pushkar Dhabe (Dhabe.p@northeastern.edu)

Aditya Agarwal (agarwal.adi@northeastern.edu)

Himani Thakker (thakker.hi@northeastern.edu)

## PROJECT OBJECTIVE:

The second leading cause of death for women is breast cancer. Early detection and treatment reduce breast cancer mortality. Mammography is crucial for breast cancer screening because it can find early breast lumps or calcification areas. One drawback of breast mammography is that breast cancer masses are harder to find in unusually dense breast tissue.

The objective is to examine various deep-learning techniques that can be applied to create a system that learns how to recognize breast cancer in mammograms and forecast classification results.

With this project, we will be able to help physicians to detect breast cancer at its early stages, maximizing the patient's survival rate and prevent unnecessary or late treatment of patients.

## LITERATURE REVIEW:

The Computer Science and Artificial Intelligence Laboratory (CSAIL) at MIT has developed a predictive analytics model that can predict a patient's risk of developing breast cancer over a number of time periods. But this model is presently being created. This technology for detection was invested in by doctors from Apollo in India, Novant Health in North Carolina, and Barretos in Brazil. This model evaluates the need for additional testing and how frequently women should be screened by taking into account a number of factors, including age, hormones, genetics, and breast density.

According to pertinent published data, the precision of risk models used in clinical practice remains constrained despite decades of research and work. Although other Al methods and predictive analytics have shown promise in predicting cancer risk, they frequently perform poorly in new patient populations, and it can be challenging to find masses in dense breast tissue.

## DATASET:

The dataset to be used in this project is put together by Mendeley Data. The dataset contains Mammographic Imaging Analysis Society (MIAS) database and INbreast dataset. The mammography dataset includes both benign and malignant masses. In order to create the images

for this dataset, 106 masses from the INbreast dataset, 53 masses from the MIAS dataset, and 2188 masses from the DDSM dataset were first extracted.
INbreast dataset has 7632 images, MIAS dataset has 3816 images, and DDSM dataset has 13128 images after data augmentation. Additionally, we combine DDSM, MIAS, and INbreast. The size of each image was changed to 227*227 pixels.

Dataset link - https://data.mendeley.com/datasets/ywsbh3ndr8/2

**METHODOLOGY:**

To find the solution for our problem, we will follow the below approach –

1. Pre-process the data
   A. Apply Image processing – Filtering and Normalization techniques
   B. Extract feature vectors from the image datasets using 2D Convolution technique
   C. Apply Max pooling, Flatten the image data
2. Split the dataset into Training, Testing and Validation dataset.
3. Apply Data Augmentation
4. Train the model and fit training data
5. Predict and show the results in form of Visualizations.