Breast Cancer Diagnosis using Machine Learning

Problem Statement:

Breast cancer is one of the most prevalent and life-threatening diseases among women globally. Early detection is crucial for improving treatment outcomes and reducing mortality rates. In this project, the goal is to develop a machine learning solution that can accurately classify breast tumors as either malignant or benign based on a dataset of real-valued features extracted from cell nuclei.

Background:

- Breast cancer is characterized by the presence of abnormal cell growth in breast tissues.

- Biopsy data, including cell nucleus features, can provide valuable information for diagnosing breast cancer.

- Machine learning models can help automate the diagnosis process, providing quick and accurate results.

Challenge:

Participants will be provided with a dataset containing the following features for each biopsy sample:

1. ID number: A unique identifier for each biopsy.

2. Diagnosis (M = malignant, B = benign): The ground truth label indicating whether the tumour is malignant or benign.

3. Ten real-valued features (a to j): Computed measurements from cell nuclei, including characteristics such as radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension.

Data:

Link: https://drive.google.com/drive/folders/1Kh11aihp2nrAhn8fQq4LPdJAPNLfxwKk?usp=sharing

Evaluation:

- Participants should aim to develop a model that achieves high accuracy in distinguishing between malignant and benign tumours.

- The success of the models will be assessed using accuracy of the submitted models.

- Brownie points for team providing data analysis.

Expected Deliverables:

Participants have to submit the finals trained ML/DL model along with the training history of the ML/DL models and a readme file.

Date of submission:

8th October 2023

Prizes:

Top performers will directly be inducted in the ML Society along with a certificate of recognition.

Rules and Guidelines:

The maximum allowed team size is 4.

Please refrain from the use of pretrained models (transfer learning is allowed) as this hackathon is for learning purpose only.