**NAME: OBIRIJA JEREMIAH ONUABUCHI**

**Reg No: 2022/HND/35634/CS**

**Assignment on COM 423 (ES/ML)**

**application of machine learning in healthcare, specifically in breast cancer**

**Problem Statement**

Breast cancer is one of the leading causes of cancer-related deaths among women worldwide. Early detection significantly increases the chances of successful treatment and survival. However, traditional diagnostic methods can be time-consuming, subjective, and sometimes inaccurate. The goal is to develop a machine learning model that can accurately and efficiently detect breast cancer at an early stage using diagnostic data.

**Solution**

A machine learning model can be trained to classify whether a patient has breast cancer based on features extracted from medical imaging and other diagnostic tests. By analyzing these features, the model can help in providing early detection with a high degree of accuracy, potentially reducing the need for invasive procedures and allowing for earlier intervention.

**Machine Learning Algorithms**

Several machine learning algorithms can be used for this task, including:

**1. Support Vector Machines (SVM):** Effective for high-dimensional spaces and commonly used for classification tasks.

**2. Random Forest:** An ensemble method that combines multiple decision trees to improve prediction accuracy and control over-fitting.

**3. K-Nearest Neighbors (KNN):** A simple, instance-based learning algorithm effective for small datasets.

4. **Neural Networks (Deep Learning):** Particularly useful when dealing with image data, as in the case of mammograms.

**Dataset**

A well-known dataset for this problem is the \*\*Wisconsin Diagnostic Breast Cancer (WDBC) dataset\*\*, available from the UCI Machine Learning Repository. This dataset includes features computed from a digitized image of a fine needle aspirate (FNA) of a breast mass, describing characteristics of the cell nuclei present in the image.

- **Features**: The dataset contains 30 features, such as radius, texture, perimeter, area, smoothness, compactness, etc., all of which are computed from the images.

- **Labels**: The output is binary, indicating whether the tumor is malignant or benign.