**Breast Cancer Prediction Using Machine Learning**

**Motivation**

Breast cancer is one of the most common cancers worldwide. Early detection significantly improves treatment outcomes and survival rates. Using the **CBIS-DDSM breast cancer image dataset**, students can develop and evaluate machine learning models to classify tumors as benign or malignant. This project provides hands-on experience in **image preprocessing, CNN model training, and performance evaluation** on real medical imaging data.

**Goals**

1. Train a machine learning model, such as a Convolutional Neural Network (CNN) using transfer learning or other suitable algorithms, on the **CBIS-DDSM dataset**.
2. Evaluate model performance using metrics such as **accuracy, precision, recall, F1-score, and AUC**.
3. Compare different model architectures and hyperparameters to determine the most effective approach.
4. Document findings, analyze results, and discuss potential limitations and implications.

**Objectives**

1. **Data Preparation:**
	* Preprocess CBIS-DDSM images (e.g., resizing, normalization, augmentation).
	* Split data into **training, validation, and testing sets**.
2. **Model Development:**
	* Train one or more CNN architectures (e.g., ResNet, EfficientNet, or MobileNet) for binary classification: **benign vs malignant**.
	* Experiment with **transfer learning** to leverage pretrained models.
3. **Evaluation and Analysis:**
	* Assess model performance using standard metrics.
	* Analyze misclassified images to identify potential challenges in the dataset.
	* Visualize results and summarize insights in a report.

**Dataset**

**CBIS-DDSM (Curated Breast Imaging Subset of DDSM):**

* High-quality mammogram images with labels for benign and malignant cases.
* Available at: <https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset>