Project Outline & Purpose

Breast cancer is a major health concern worldwide, and early diagnosis is crucial for effective treatment and improved outcomes. In this project, we have used machine learning algorithms to develop models that can assist in the diagnosis of breast cancer based on various features extracted from diagnostic images. Specifically, we have used the breast cancer Wisconsin (diagnostic) dataset, which includes measurements of different features extracted from digitized images of fine needle aspirate (FNA) samples of breast masses. We have implemented and evaluated three machine learning algorithms: Logistic Regression, Decision Tree, and Random Forest, and compared their performance in terms of accuracy on the test dataset. Our results show that all three models achieve high accuracy, with the Random Forest model performing slightly better than the other two. Our project demonstrates the potential of machine learning algorithms in improving breast cancer diagnosis and highlights the importance of developing accurate and reliable diagnostic tools to aid in the fight against breast cancer.

Breast cancer (BC) is one of the most common cancers among women worldwide, representing the majority of new cancer cases and cancer-related deaths according to global statistics, making it a significant public health problem in today’s society. The early diagnosis of BC can improve the prognosis and chance of survival significantly, as it can promote timely clinical treatment to patients. Further accurate classification of benign tumors can prevent patients from undergoing unnecessary treatments. Thus, the correct diagnosis of BC and classification of patients into malignant or benign groups is the subject of much research. Because of its unique advantages in critical feature detection from complex BC datasets, machine learning (ML) is widely recognized as the methodology of choice in BC pattern classification and forecast modeling. Classification and data mining methods are effective ways to classify data. Especially in the medical field, where those methods are widely used in diagnosis and analysis to make decisions.