

# Project Proposal

## BreastCancerDetect: Empowering Breast Cancer Prediction and Survival Analysis

### Team Members

- Or Tzadok
- Elon Reiner
- Eyal Keren

### Motivation

Breast cancer is a significant health concern for women worldwide, the average risk of a woman in the United States developing breast cancer sometime in her life is about 13% which means there is a 1 in 8 chance she will develop breast cancer. Early detection and accurate diagnosis are crucial for improving patient outcomes and reducing mortality rates. Our project aims to tackle this problem by developing a machine learning model that can accurately predict breast cancer diagnosis and cancer stage based on patient characteristics. We also plan to develop a survival prediction model that can predict a patient's likelihood of survival after diagnosis.

### Method

We plan to use various machine learning techniques, including logistic regression, decision trees, random forests, and support vector machines, to develop our predictive models. We will preprocess the data to handle missing values, scale the features, and encode categorical variables. We will also perform feature selection to identify the most relevant features for our models.

### Intended Experiments

We will split the dataset into training, validation, and test sets and evaluate the performance of our models using standard metrics such as accuracy, precision, recall, and F1 score. We will also compare the performance of our models to baseline models and assess the impact of different feature selection techniques on model performance.

### Dataset

We will use the Breast Cancer Wisconsin (Diagnostic) dataset, which contains information on breast cancer diagnoses and patient characteristics. The dataset is publicly available on Kaggle.

### Prior Research

Previous studies have used machine learning algorithms to predict breast cancer diagnosis and survival. For example, a study called "Breast Cancer Diagnosis by Different Machine Learning Methods Using Blood Analysis Data" by Muhammet Fatih Aslan et al. (2018) used Extreme Learning Machine (ELM) to predict breast cancer diagnosis based on patient data and achieved an accuracy of 80%. Another study "A Breakup Machine Learning Approach for Breast Cancer Prediction" by Sabari et al. (2019) used Random Forest to predict whether "M" or "B" the cancer diagnosis and got accuracy of above 90%.