

# CS5131 Programming Assignment 2

Jed Lim, Karimi Zayan, Matthew Kan, Makuyh Das

12 October, 2022

## 1 Background

Breast cancer is the number one most common cancer amongst women in Singapore. Early and accurate diagnosis of breast cancer is important for breast-saving and life-saving treatment.

The gold standard for the diagnosis of breast cancer is by surgically removing the breast lump with a complete microscopic examination of the breast tissue to look for cancer cells.

Fine needle aspiration is an alternative that allows the doctor to take out a small amount of tissue from the breast lump, without the need for surgery to remove the entire breast lump. By examining the characteristics of the cells, doctors have been able to diagnose breast cancer with variable success. Increasing the success of fine needle aspiration allows for diagnosis of breast cancer without the need for a woman to undergo surgery to remove the breast lump.

To resolve this, this project uses a fuzzy decision tree to classify breast tumor cells into malignant cancer cells or benign non-cancerous cells.

## 2 Dataset

The dataset used was the Breast Cancer Wisconsin (Diagnostic) Data Set from the University of Irvine (UCI) Machine Learning Repository.

The dataset contained 569 instances, with no missing data. 357 instances were benign (not cancerous) and 212 were malignant (cancerous).

The features were computed from digitalized images of fine needle aspirates of breast tumors.

The features described 10 characteristics of the cell nuclei present in the images:

- The radius of an individual nucleus
- The nuclear perimeter
- The nuclear area
- Compactness of the nucleus
- The smoothness of the contour of the nucleus
- The number of contour concavities
- The symmetry of the nuclear contour
- The texture of the cell nucleus