

Breast Cancer Classification

The objective of this project is to build a breast cancer classifier on an IDC dataset that can accurately classify a histology image as benign or malignant.

IDC is Invasive Ductal Carcinoma; a cancer that develops in a milk duct and invades the fibrous or fatty breast tissue outside the duct; it is the most common form of breast cancer forming 80% of all breast cancer diagnoses. And histology is the study of the microscopic structure of tissues.

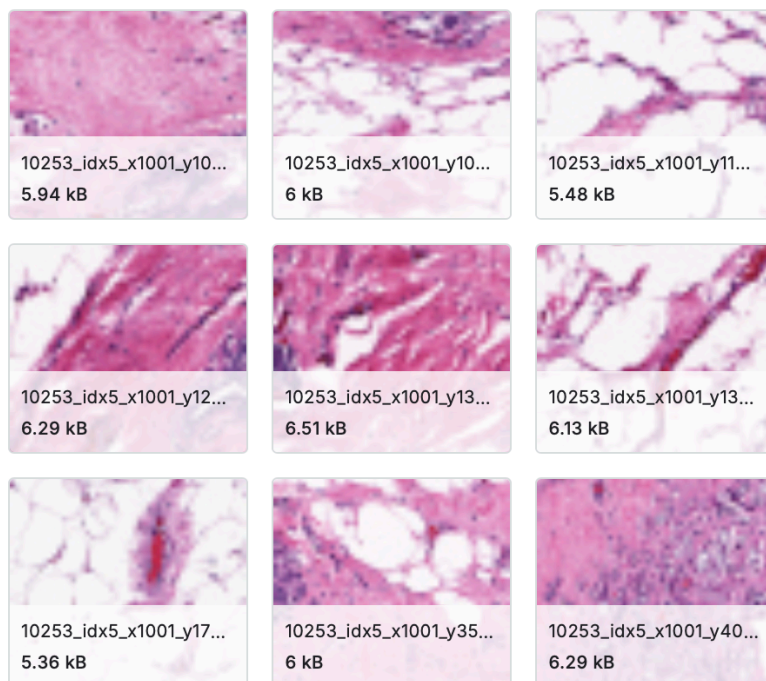
The given IDC dataset (the breast cancer histology image dataset) holds 277,524 patches of size 50×50 extracted from 162 whole-mount slide images of breast cancer specimens scanned at 40x. Of these, 198,738 test negative and 78,786 test positive with IDC.

Filenames in this dataset look like this:

8863_idx5_x451_y1451_class0

Here, 8863_idx5 is the patient ID, 451 and 1451 are the x- and y- coordinates of the crop, and 0 is the class label (0 denotes absence of IDC).

Class 0 examples:



Class 1 examples:

