Predicting invasive ductal carcinoma in tissue slices.

IDC → Invasive ductal carcinoma with 80% of cases - one of the most common types of breast cancer. Pathologist has to decide whether a patient has IDC, another type of breast cancer or is healthy. Deep learning could be a great help to automatically detect and locate tumour tissue cells and to speed up the process.

Collecting information:

In paper tissue slices of 280 patients are used which have IDC [about 265 are used for training and 15 for validation].

Evaluation Metric: F1_score - and balanced accuracy

Our goal: given a patients and a patch of tissue slice predict wheater it contains IDC or not.

Possibilities → healthy tissues
→ IDC
→ Any other subtype of breast cancer.

Mainly we have used libraries such as numpy, pandas, matplotlib → visualization and torch for CNN.

Initially we analyse the patches by taking a example of one patient having ID number "13616", then we devide the dataset for training and testing, we fit the training dataset to form vision models.

We have used pytorch for image datasets and we took for augmentations. Then transfer learning with a pretrained CNN. So this helps us to extract feature from previous layers and improve the upcoming layers, A common loss for this promblem is - binary cross entropy function - for single output neuron to make prediction.

Selection of evaluation metric is f1-score and precision and recall. We made many trials for dataset starting with random state = 0 test size = 0.3 with 69.83%, then modified test size = 1 and got accuracy 89.39% and then changed random state = 42 and test size to 0.05 to get an accuracy for 94.9%.

— Team :
Data - Pirates