

**Team:**  
**Data Pirates**

**Hack the CNN!**

Put your  
thoughts here.

**Predicting invasive ductal carcinoma in tissue slices**

Put your  
thoughts here.

Put your  
thoughts here.

# Context

Before we start with the session, let's warm up a context!

- 1 IDC – 80% of cases are of breast cancer in men and women.
- 2 It is mostly malign and hence, difficult to diagnose.
- 3 Accuracy is the need of the hour as any type-1 error can make someone at risk!

**Invasive ductal carcinoma (IDC)** is – with ~ 80 % of cases – one of the most common types of breast cancer. It's malicious and able to form metastases which makes it especially dangerous. Often a biopsy is done to remove small tissue samples. Then a pathologist has to decide whether a patient has IDC, another type of breast cancer or is healthy. In addition sick cells need to be located to find out how advanced the disease is and which grade should be assigned. This has to be done manually and is a time consuming process. Furthermore the decision depends on the expertise of the pathologist and his or her equipment. Therefore deep learning could be of great help to automatically detect and locate tumor tissue cells and to speed up the process. In order to exploit the full potential one could build a pipeline using massive amounts of tissue image data of various hospitals that were evaluated by different experts. This way one would be able to overcome the dependence on the pathologist which would be especially useful in regions where no experts are available .

Time: 5 minutes

## Our Goal

As we started with this analysis we asked ourselves if we would be able to improve the results that were presented 2014 in the paper Automatic detection of invasive ductal carcinoma in whole slide images with Convolutional Neural Networks of professor Anant Madabhushi and his group.

Many years have passed by since then and it's very likely that all methods used in the paper have already been changed, improved and that new research has already been done. Nonetheless it's a very good exercise to practice or develop own deep learning and data science skills.

# Problem Statement

Boosting accuracy of the existing models by unlocking power of CNN!

## Methodology

In the paper tissue slices of 162 patients were used all having IDC (113 used for training and 49 for validation).

One pathologist was used to determine regions of IDC given a tissue slice.

evaluation metric: F1 score and balanced accuracy

Our goal: Given a patient and a patch of a tissue slice predict whether it contains IDC or not.

3 possibilities: healthy tissue, IDC, another subtype of breast cancer.

Business case: prediction so far is done manually by pathologists and varies from expert to expert. The goal is to assist with an automatic detection of tumors (not expert dependent).

## Hit and Trials!

RANDOM STATE	TEST SIZE	ACCURACY
0	0.3	69.83%
0	0.1	89.39%
42	.05	94.9%

**ACCURACY OF 94.9% ,**

**THIS MODEL CAN PREDICT WHICH TYPE OF CANCER THAT PARTICULAR PATIENT HAVE WHETHER ITS IDC(INVASIVE DUCTAL CARCINOMA) OR NON IDC OR A HEALTHY TISSUES .**

**Are you ready?**