06/14

CLASSIFICATION OF BREAST CANCER IMAGES

DEEP LEARNING - NEURAL NETWORKS

Ben Richman
Neville Fernandes
Tyler Young
Vinit Nair

PROBLEM STATEMENT

1 in 8 US women will develop invasive breast cancer in their lifetime.

The lifetime risk of breast cancer for US men is 1 in 1000.

PROBLEM

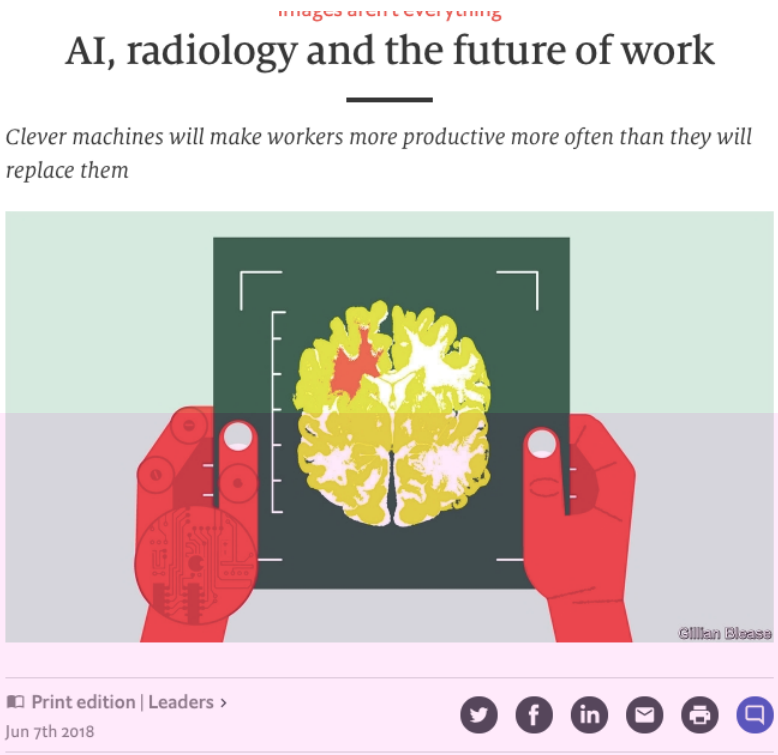
- Detecting the incidence and extent of cancer currently performed by manually looking at images.
- Painstaking, long, inefficient and error-filled process.

GOALS

- Train a model to classify images with invasive ductal carcinoma.
- Assign an aggressiveness grade the sample, and determine the exact regions of IDC in the sample.

HEALTHCARE

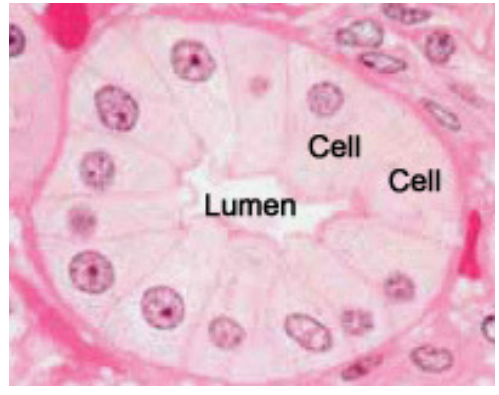
- Increased confidence in a flagging system that can detect malignant cells in a sample
- Faster classification and aggressiveness grade assignment
- Earlier detection, save lives
- Efficiency can lower healthcare costs
- Allows doctors to spend more time interacting with and seeing more patients



To read more about machines empowering doctors:
https://www.economist.com/leaders/2018/06/07/ai-radiology-and-the-future-of-work

IMAGE DATA - MANIPULATION

- Data sourced from Kaggle, originally from research by Anant Madabhushi at Case Western
- H&E staining of breast cancer tissue
  pink - cytoplasm, purple - nucleus & some organelles



162 whole mount slide color images
Each slide scanned at 40x zoom
Broken down to 50x50 px images

→

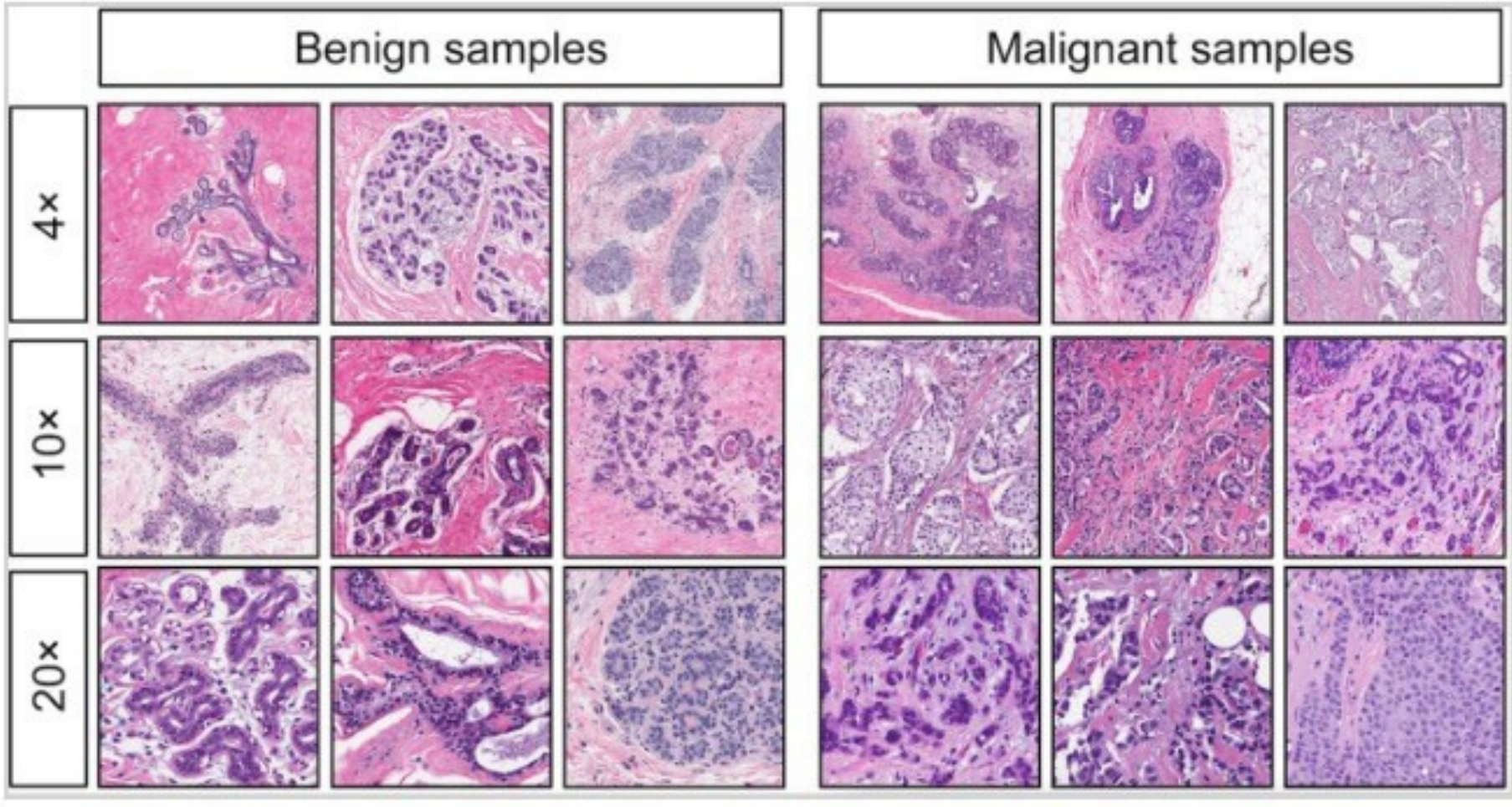
277,524 images subset
28.4% IDC positive
71.6% IDC negative

Each pixel in 50x50 image (2D)
encoded in red, green and blue

→ normalize by dividing by 255

50x50x3 array (1D)
Each pixel is a 3x1 vector
with values $\in S[0,1]$

BREAST HISTOPATHOLOGY IMAGES

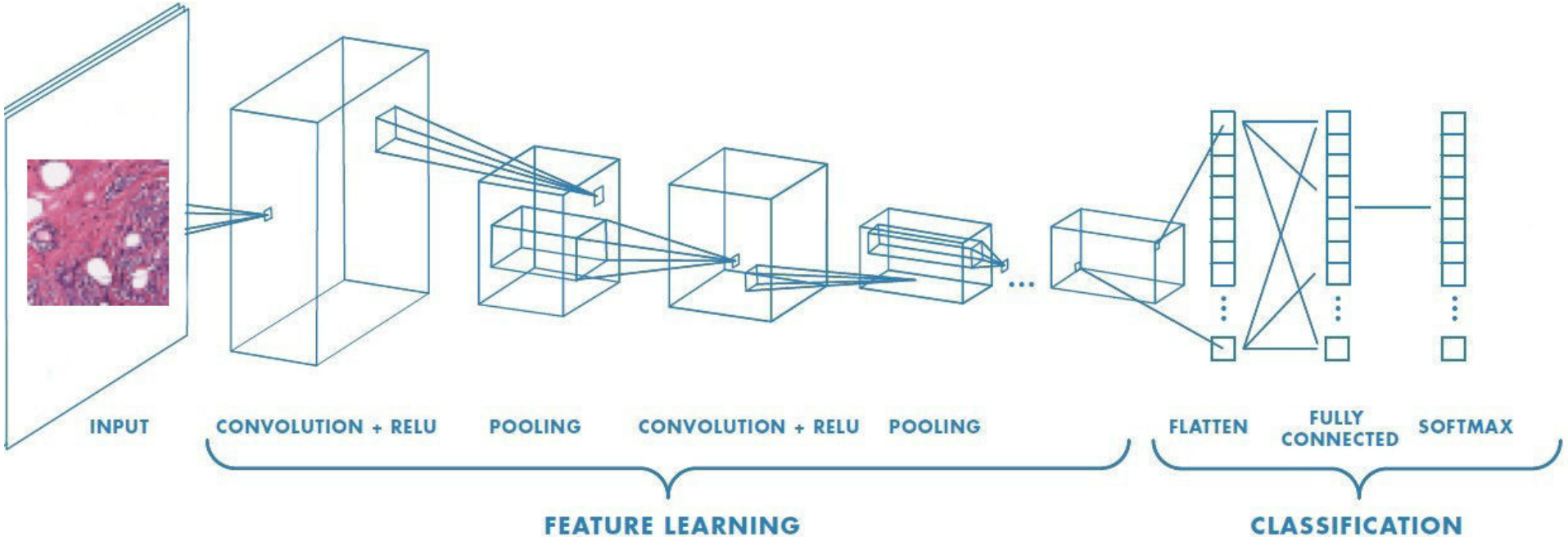


Examples of benign (left) and malignant (right) breast specimens stained with hematoxylin and eosin, at different magnifications. The

OUR TECHNIQUE: CONVOLUTIONAL NEURAL NETWORKS (CNN)

- Special architecture of artificial neural networks that replicates some features of the visual cortex.
- Utilizes layers of neurons dedicated to identifying specific features in a dataset
- Most popular use - image classification
- Eg: automatic tagging algorithms used by Facebook, Instagram, Google to annotate and search through users' photos.

BASIC MODEL



Output channels - 32
Maxpooling - pool size 2 x 2
Flattened layer
Dense layer - 100 nodes
Optimizer - sgd; Loss - crossentropy

ACCESSORIZING

ALTERING THE ARCHITECTURE

- Change Conv2D layers
- Add 'dropout' layers to avoid overfitting
- Batch normalization
- Padding (adding zeroes around the input matrix to preserve information)
- Change optimizers (sgd, RMS, Adadelta etc.)
- Change loss functions (MSE, cross entropy etc.)

DATA AUGMENTATION

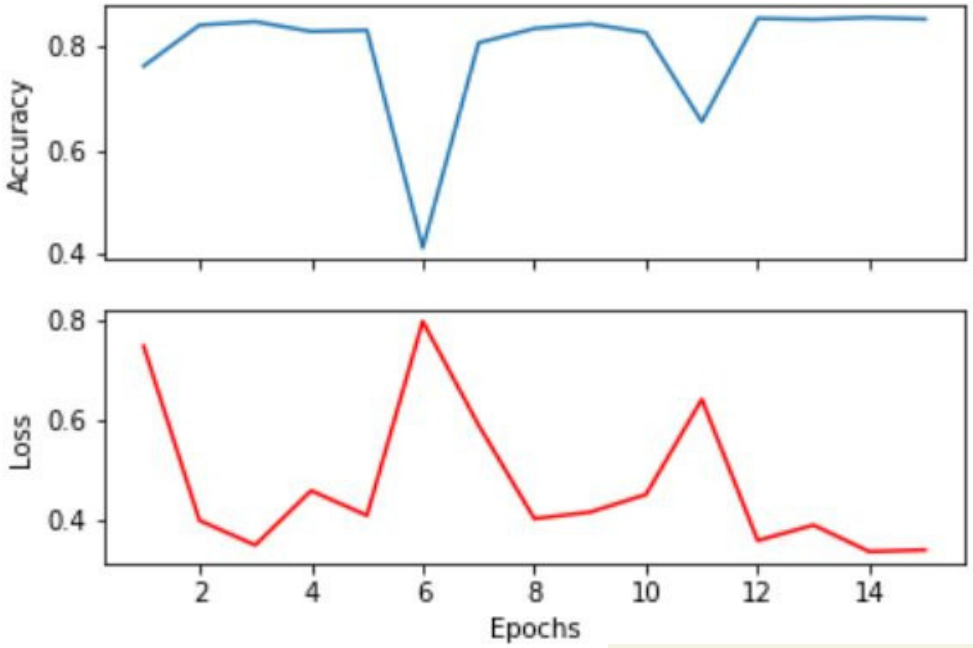
- Rotating and flipping the images
- Shifting images horizontally/vertically

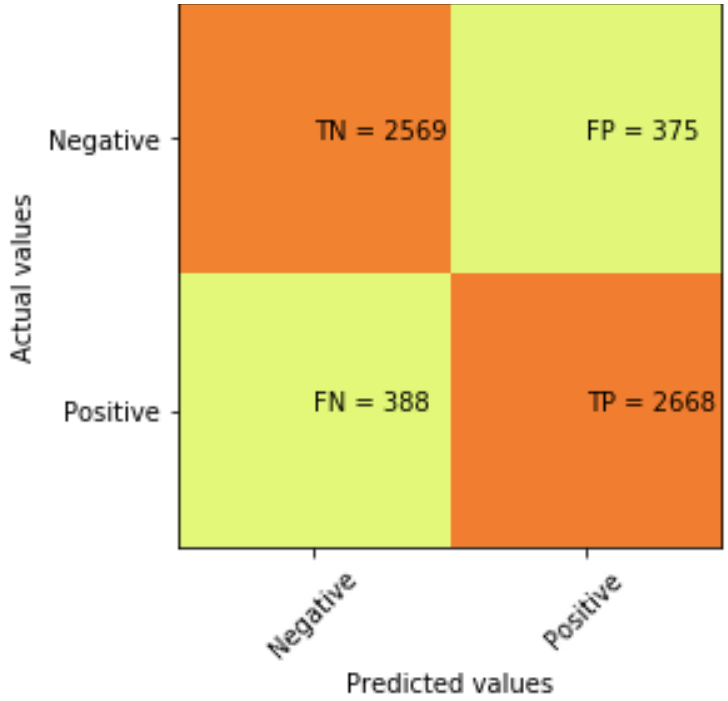
FINAL MODEL

4 convolution layers
Output channels: 32 & 64
Padding
Maxpooling - pool size 2 x 2
Dropout - 0.25
Dense layer - 512 nodes
Optimizer - Adadelta
Loss - crossentropy
Data augmentation

```
model5 = Sequential()

model5.add(Conv2D(32, kernel_size=(3, 3), strides=(1, 1),
                  padding='same',
                  activation='relu',
                  input_shape=X_train[0].shape))
model5.add(Conv2D(32, (3, 3), activation='relu'))
model5.add(MaxPooling2D(pool_size=(2, 2)))
model5.add(Dropout(0.25))

model5.add(Conv2D(64, kernel_size=(3, 3), strides=(1, 1),
                  padding='same',
                  activation='relu',
                  input_shape=X_train[0].shape))
model5.add(Conv2D(64, (3, 3), activation='relu'))
model5.add(MaxPooling2D(pool_size=(2, 2)))
model5.add(Dropout(0.25))

model5.add(Flatten())
model5.add(Dense(512, activation='relu'))
model5.add(Dropout(0.5))
model5.add(Dense(num_classes, activation='softmax'))

optimizer = Adadelta(lr=1.0, rho=0.95, epsilon=None, decay=0.0)

model5.compile(loss='categorical_crossentropy',
               optimizer=optimizer,
               metrics=['accuracy'])
```

RESULT





	precision	recall	f1-score	support
benign	0.87	0.87	0.87	2944
malignant	0.88	0.87	0.87	3056
avg / total	0.87	0.87	0.87	6000

- Accuracy > 87% over 15 epochs; loss < 0.3
- Around 87% precision and recall scores for benign and malignant samples
- Research shows accuracy of 84% and F-1 score of 76%

Janowczyk, Madabhushi. Deep learning for digital pathology image analysis. (2016)

NEXT STEP: INTERFACE

Heat Map: 2D/3D visual showing extent to which tissue is cancerous as well as the distribution of the cancerous cells

Embedded in software: Patients are selected and their images classified in real time in different measures

- highlighting cancerous tissue
- percent of tissue that is cancerous
- stage of the cancer

THANK YOU!