A high-magnification microscopic image of breast tissue. The background consists of numerous small, rounded, pale blue-stained cells, likely normal mammary gland epithelial cells. Interspersed among them are larger, more irregularly shaped cells with darker, more granular or pleomorphic nuclei, characteristic of invasive ductal carcinoma. Some of these neoplastic cells appear to be invading through the basement membrane into the surrounding stroma, which is visible as a darker, textured area.

# **INVASIVE DUCTAL CARCINOMA CLASSIFICATION**

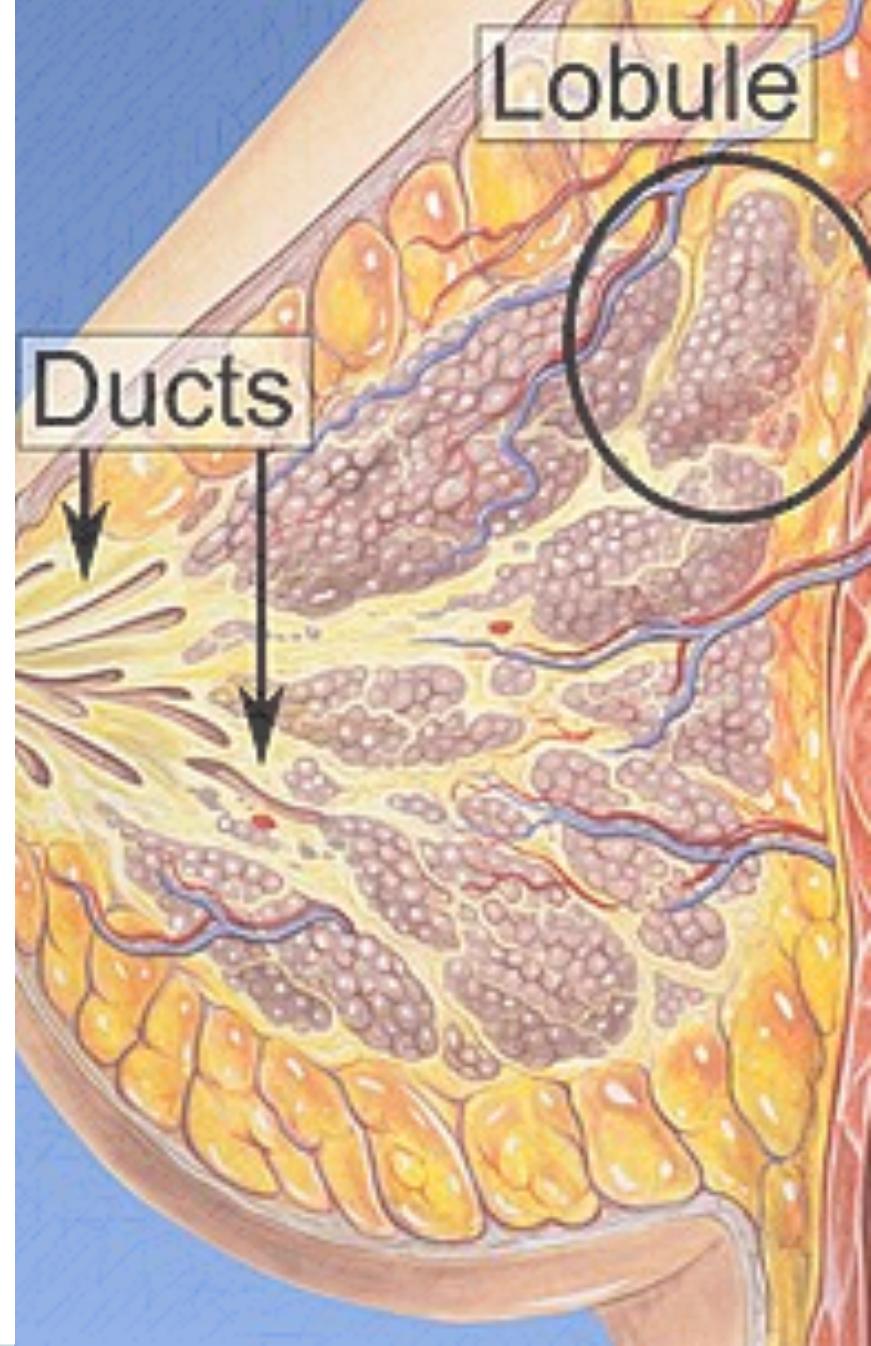
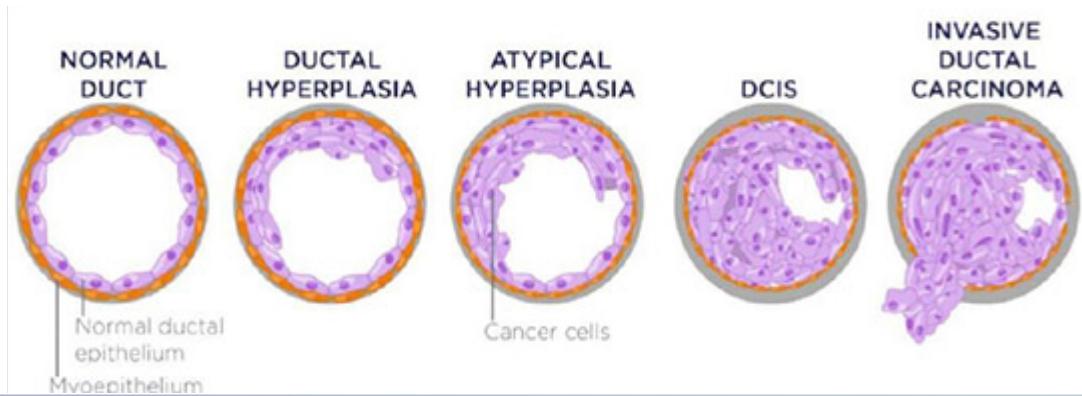
BY: DIPTA ROY

# PROJECT OVERVIEW

- Can we create an accurate model to classify if a histology slide shows Invasive Ductal Carcinoma?
- Can we create a scalable and generalizable model?
- Business Case: Predictions done by pathologists can vary from expert to expert. If we can create an accurate model we can help with the diagnosis consistency.

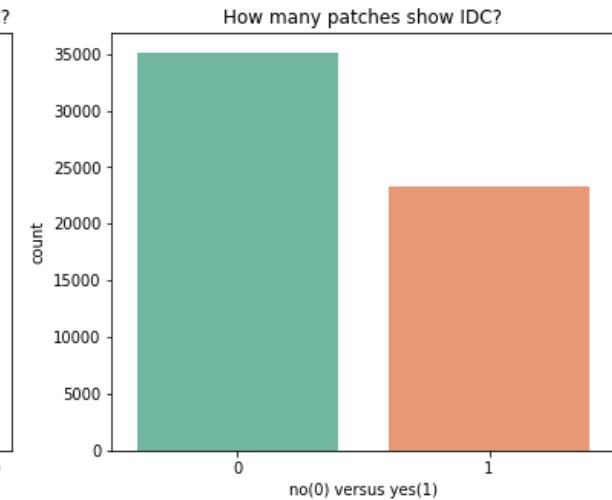
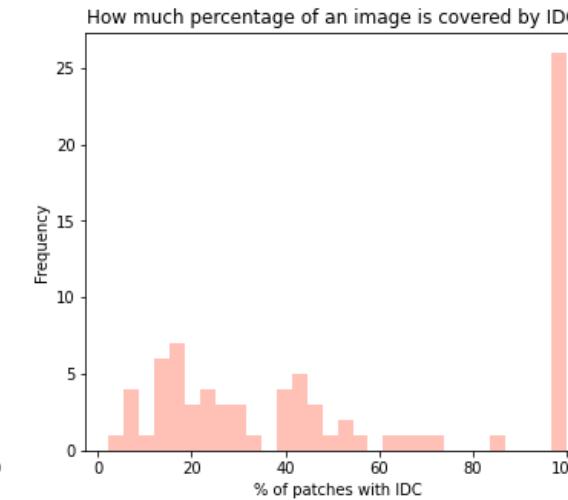
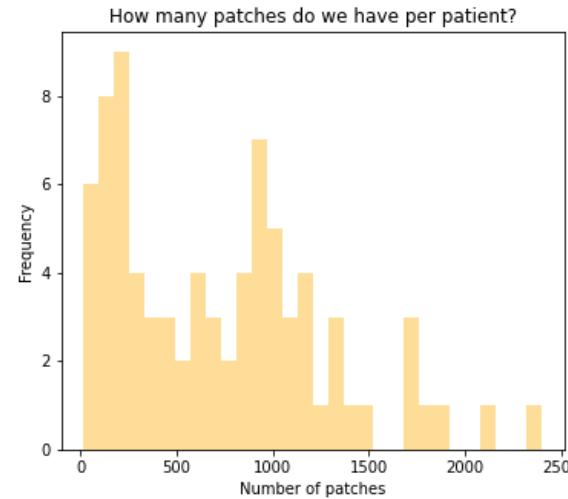
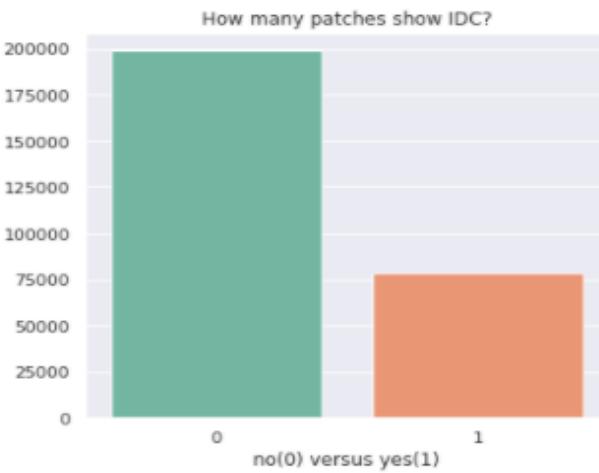
# WHAT IS INVASIVE DUCTAL CARCINOMA

- Cancer that starts in cells that line the milk ducts
- Makes up 80% of cases of breast cancers
- Able to metastasize and enter blood vessels as well as lymph nodes
- Earlier detection lead to higher survival rate

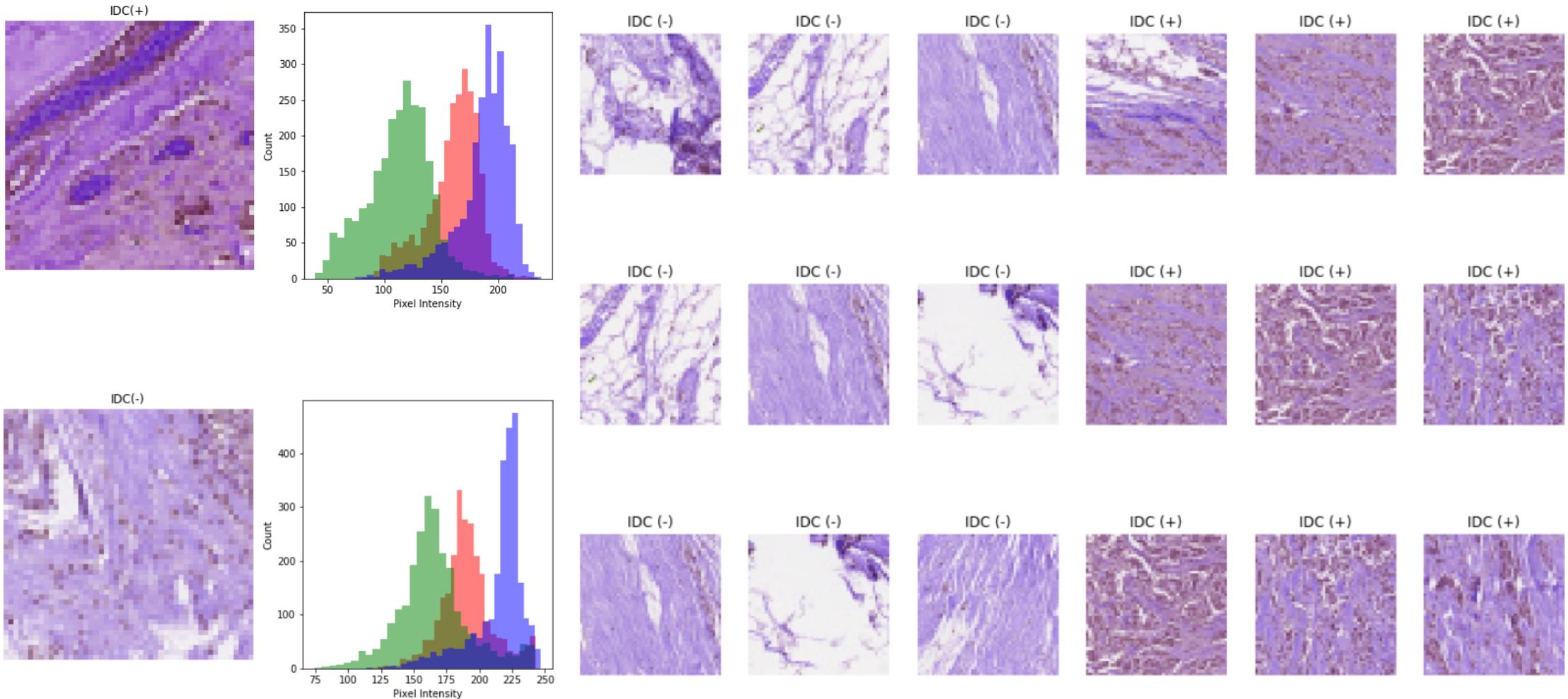


# DATASET!

- Dataset from [http://gleason.case.edu/webdata/jpi-dltutorial/IDC\\_regular\\_ps50\\_idx5.zip](http://gleason.case.edu/webdata/jpi-dltutorial/IDC_regular_ps50_idx5.zip)
- Citations:
  - <https://www.ncbi.nlm.nih.gov/pubmed/27563488>



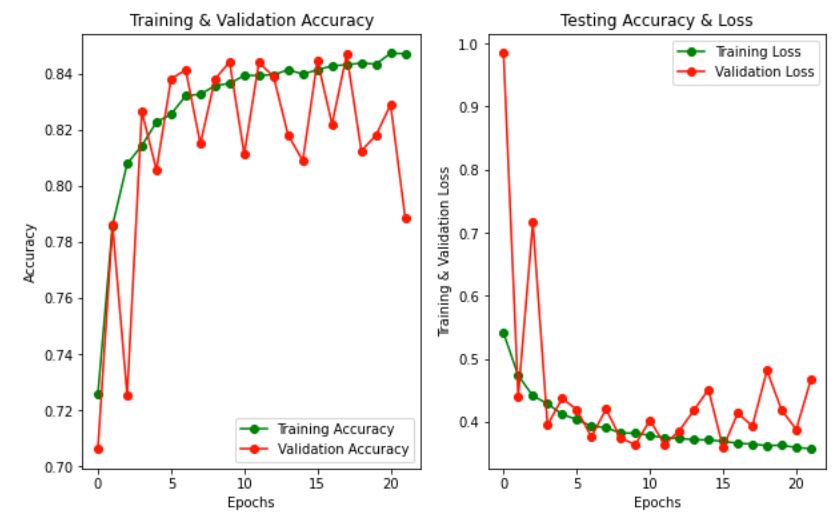
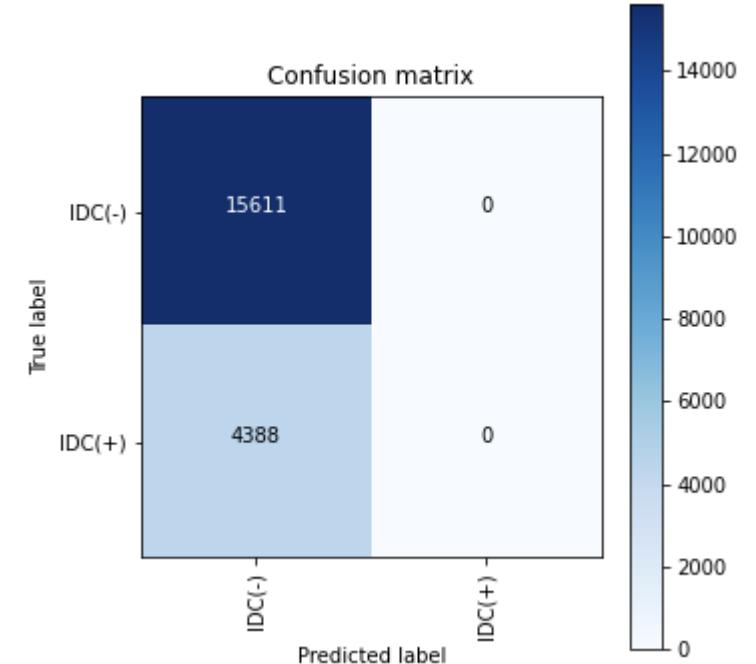
# HOW DOES IDC LOOK LIKE IN HISTOLOGY IMAGES?



# CNN BASELINE

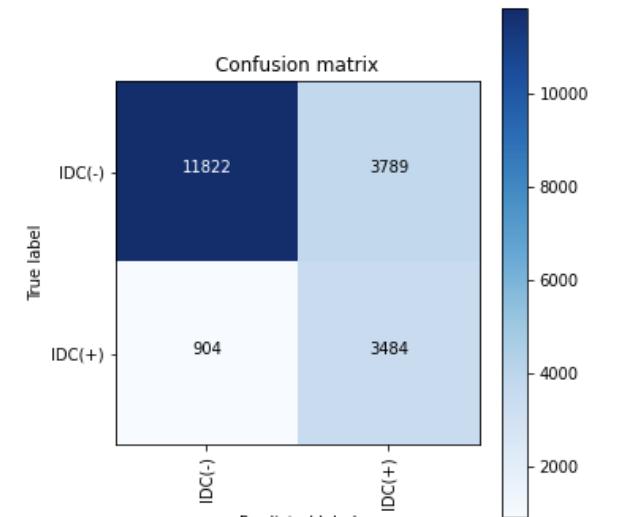
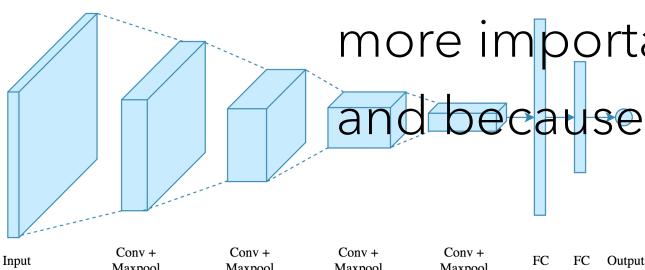
- Overfit the IDC(-) and classified everything as IDC(-)
- Can be due to imbalanced data
- 1 Convolution Layer
- 1 Dense, 1 hidden layer, 1 output

	precision	recall	f1-score	support
IDC (-)	0.78	1.00	0.88	15611
IDC (+)	0.00	0.00	0.00	4388
accuracy			0.78	19999
macro avg	0.39	0.50	0.44	19999
weighted avg	0.61	0.78	0.68	19999



# BEST MODEL: DATA AUGMENTATION + UPSAMPLING (ADDING DATA)

- Performed better than the baseline, and other hand-tweaked models
- Different techniques of data augmentation was used and best model used horizontal, vertical flips, rotation range up to 40 degrees, and translational shifts.
- 3 convolution layers
- Tuned models to give the best F1 score because put more importance into false positives and false negatives and because of class imbalance

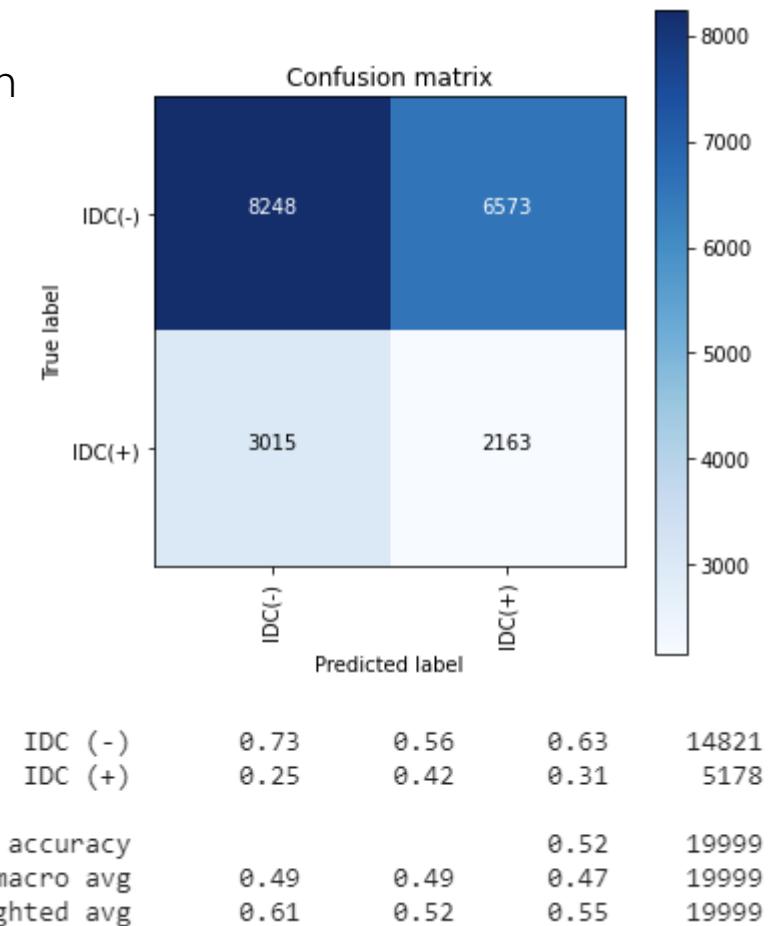
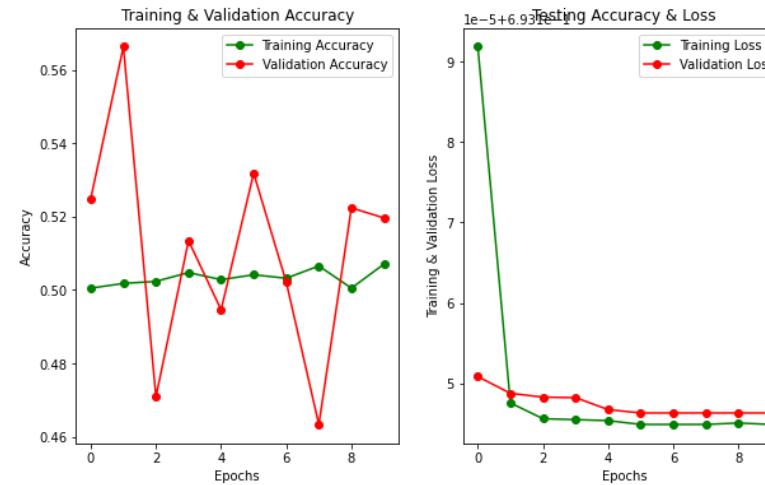


	precision	recall	f1-score	support
IDC (-)	0.93	0.76	0.83	15611
IDC (+)	0.48	0.79	0.60	4388
accuracy			0.77	19999
macro avg	0.70	0.78	0.72	19999
weighted avg	0.83	0.77	0.78	19999

# INCEPTION MODEL TRANSFER LEARNING

- Transfer model seemed to give a lot more false positives than other models.
- According to the ai.googleblog Inception is usually used for healthcare images

<https://ai.googleblog.com/2019/12/understanding-transfer-learning-for.html>



# FUTURE



Try to achieve a higher F1 and Accuracy score



Use other transfer learning models



Create a front-end site where you can upload histopathology image to be classified



Histopathology images aren't super clear, looking into SRGAN to obtain higher resolution images

# LIMITATIONS

Slides were assessed by only one pathologist

There is a big class imbalance

Low pixeled pictures



THANK YOU!