**Breast Cancer Prediction and Detection through Machine Learning and Deep Learning**

David Kinney

Bellevue University

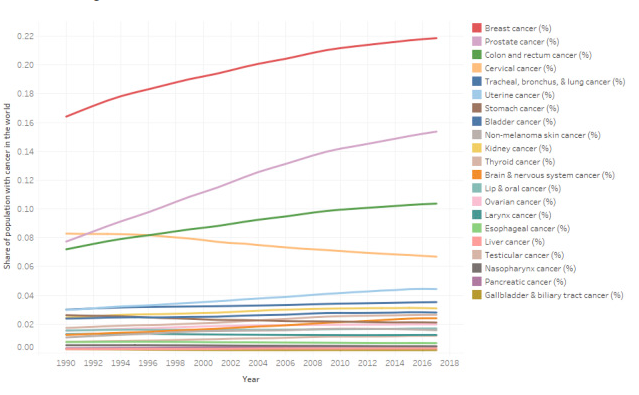
DSC 680 Applied Data Science

Professor Catherine Williams

May 30, 2021

**Introduction and Hypothesis**

Cancer is a serious public health issue worldwide and the second leading cause of death in the United States. According to the International Agency for Research on Cancer (IARC), about 18.1 million new cases and 9.6 million deaths caused by cancer were reported in 2018. [1] Breast cancer accounts for the largest share of cancer types world-wide and is on the rise.



Source: <https://www.kaggle.com/midouazerty/breast-cancer-images-classification>

**False Negative Results**

In cancer screening, a negative result means no abnormality is present. False-negative results occur when mammograms appear normal even though breast cancer is present. Overall, screening mammograms miss about 20% of breast cancers that are present at the time of screening. [2] While a false positive result may lead to undue stress and worry, the end result is no cancer. False negatives are far more alarming, as the result in this case is a woman who believes she is cancer-free when she is not.

**Hypothesis**

Great strides have been made in both Machine and Deep Learning in various medical fields regarding the prediction and detection of certain diseases. One of these areas showing promising results is breast cancer; both prediction based on observable measurements and detection regarding whether a tumor is benign or malignant. In this paper I hope to show two examples of models I trained that offer impressive results in this field.

**Method**

**Data**

I am using two datasets for this project; one is tabular, one is images. The tabular dataset (Breast Cancer Wisconsin (Diagnostic) dataset) [3] consists of 33 measurements; there are no categorical variables, therefore no encoding was required.

Attribute Information:

1) ID number  
2) Diagnosis (M = malignant, B = benign)

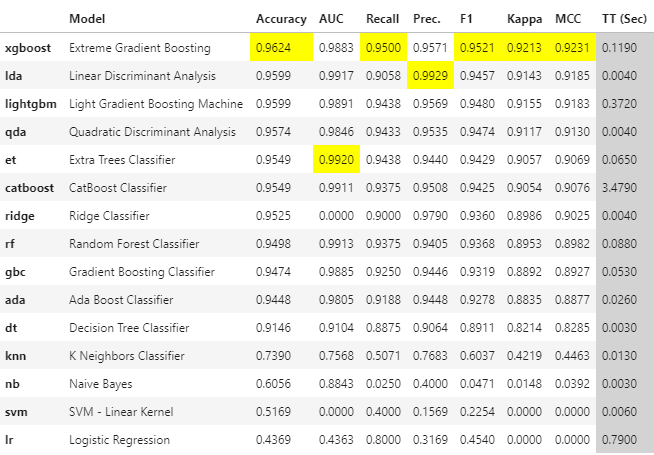
Ten real-valued features are computed for each cell nucleus:

a) radius (mean of distances from center to points on the perimeter)  
b) texture (standard deviation of gray-scale values)  
c) perimeter  
d) area  
e) smoothness (local variation in radius lengths)  
f) compactness (perimeter^2 / area - 1.0)  
g) concavity (severity of concave portions of the contour)  
h) concave points (number of concave portions of the contour)  
i) symmetry  
j) fractal dimension ("coastline approximation" - 1)

The mean, standard error, and "worst" or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features. A predictor label is provided: M for a malignant tumor, B for benign. Further, there are no missing attribute values. Class distribution: 357 benign, 212 malignant. [4]

**Machine Learning Model**

The features for the Breast Cancer Wisconsin dataset Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image. [3] The **PyCaret** library was used to select the best model for this application, which turned out to be *Extreme Gradient Boosting.*

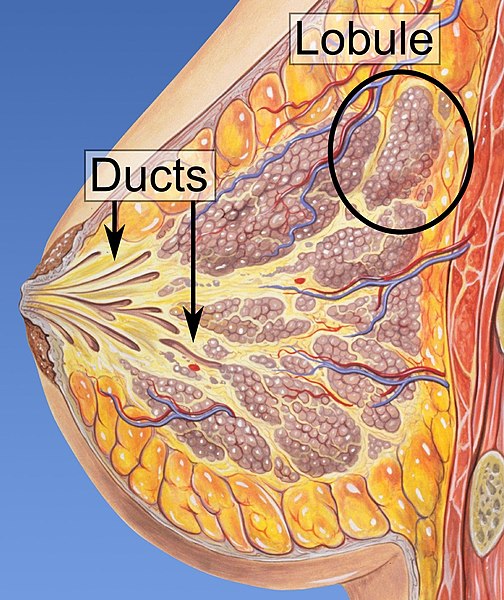


Hyperparameter tuning was then performed, followed by model prediction calculations on the holdout dataset results:



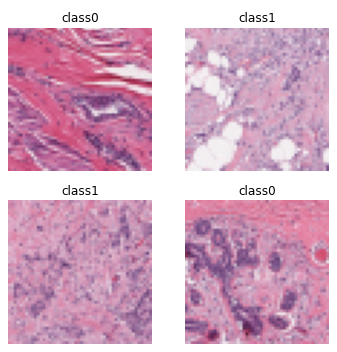
**Deep Learning Model**

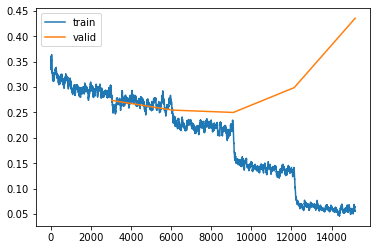
This inspiration and figure were provided by Salah Sammari on his Kaggle page, which offers a concise and informative summary regarding Abstract Breast Cancer:

“Abstract Breast cancer is a common cancer in women, and one of the major causes of death among women around the world. Invasive ductal carcinoma (IDC) is the most widespread type of breast cancer with about 80% of all diagnosed cases. Early accurate diagnosis plays an important role in choosing the right treatment plan and improving survival rate among the patients. In recent years, efforts have been made to predict and detect all types of cancers by employing artificial intelligence.” [1]

For this model, I chose the **fastai** library, which leverages **PyTorch** by wrapping many of the more tedious tasks in user-friendly wrapper methods. I first trained a *Resnet34* model to establish a baseline and trained for 5 epochs. The learning rate between the training and validation sets began to diverge after 3 epochs. For the next iteration, I trained for only 3 epochs.

**Data**

For this model I used the Breast Histopathology Images dataset, which consists of 62 whole mount slide images of Breast Cancer (BCa) specimens scanned at 40x. From that, 277,524 patches of size 50 x 50 were extracted (198,738 IDC negative and 78,786 IDC positive).

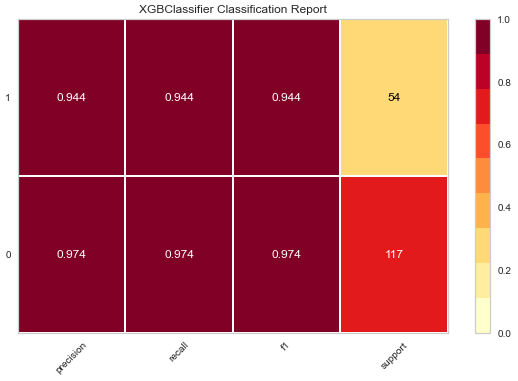
A good explanation of ResNets (shorthand for Residual Networks) and their many variants can be found here:

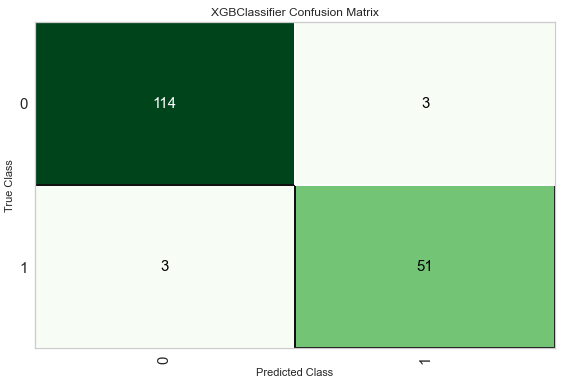
<https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035>

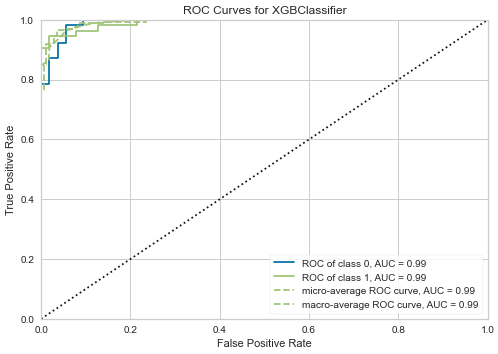
**Results**

**Machine Learning**

To reiterate, the purpose of training the Machine Learning model was to be able to predict, based on the measurement features in the dataset, whether an observation identified a benign or a malignant tumor. The **xgboost** model resulted in some impressive results.







**Deep Learning**

As previously mentioned, my focus was to train a model with the least rate of false negatives; at the very least, to do better than the human rate of 20%. By utilizing the **fastai** library, I was able to train a model with a false negative rate of about 9% (1 – Negative Predicted Value (NPV)).

Sensitivity: 0.7816111416839521

Specificity: 0.9318509575074622

PPV: 0.8196022727272727

NPV: 0.9150488234616081

**Sensitivity & Specificity**

**Sensitivity or True Positive Rate** is where the model classifies a patient has the disease given the patient actually does have the disease. Sensitivity quantifies the avoidance of false negatives

Example: A new test was tested on 10,000 patients, if the new test has a sensitivity of 90% the test will correctly detect 9,000 (True Positive) patients but will miss 1000 (False Negative) patients that have the condition but were tested as not having the condition

**Specificity or True Negative Rate** is where the model classifies a patient as not having the disease given the patient actually does not have the disease. Specificity quantifies the avoidance of false positives

Understanding and using sensitivity, specificity and predictive values is a great paper if you are interested in learning more about understanding sensitivity, specificity and predictive values.

**PPV and NPV**

Most medical testing is evaluated via PPV (Positive Predictive Value) or NPV (Negative Predictive Value).

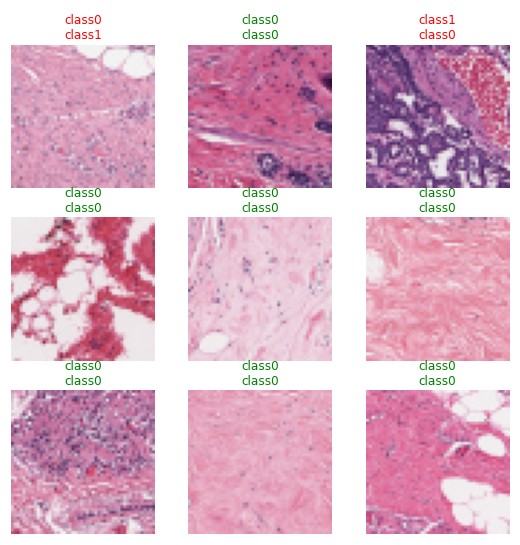
**PPV** - if the model predicts a patient has a condition what is the probability that the patient actually has the condition

**NPV** - if the model predicts a patient does not have a condition what is the probability that the patient actually does not have the condition

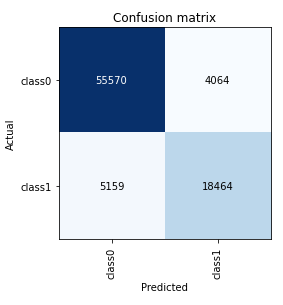
The ideal value of the PPV, with a perfect test, is 1 (100%), and the worst possible value would be zero

The ideal value of the NPV, with a perfect test, is 1 (100%), and the worst possible value would be zero

Source: <https://docs.fast.ai/tutorial.medical_imaging.html>



In a random sampling of predictions, only one resulted in a false negative.



**Research Questions**

**What, if any, benefits can deep learning bring to the detection of breast cancer?**

Leveraging the **fastai** library, I was able to train a ResNet34 model using a desktop computer on a ResNet34 model in about four hours, consuming over a quarter million images. Further, **fastai** makes use of *transfer learning*; models that have been pre-trained on millions of images. The end result was a model with an accuracy rate of just short of 90% and a False Negative rate of roughly 8%. If these results are so easily achieved on a desktop computer, imagine what can be achieved with today’s super computers.

**What, if any, benefits can machine learning bring to the prediction of breast cancer?**

I avoided countless hours of training, testing, and comparing a multitude of models by having **PyCaret** do it for me. In roughly 42 *seconds* 15 separate classifier models were trained and then tested on a holdout dataset, resulting in a “best model” accuracy score of 97%. Creating and tuning the best model took another 56 seconds. The end result was a model with an accuracy rate of 96% on unseen data. As above, this was achieved on my personal desktop computer.

**What is an acceptable level of false positives?**

More than 50% of women screened annually for 10 years in the United States will experience a false-positive result. [2] The Deep Learning model I trained had a False Positive rate of roughly 17%. Therefore, my opinion is that an acceptable level would be something less than 20%.

**What is an acceptable level of false negatives?**

As I mentioned in the Introduction, in my view False Negatives are the priority. In a perfect world, the acceptable rate is 0. But neither humans or computers are perfect. The average rate for human radiologists is somewhere around 20%. The rate for my Deep Learning model is 8%. Therefore, in my opinion, an acceptable rate is 10%.

**Which machine learning classification models provide the most promising results?**

For Deep Learning, I actually did not attempt to train anything other than *ResNet*. These models are perfect for classification as they are already pre-trained, so in many cases it’s simply a matter of training on the last layer of the neural network. For Machine Learning, PyCaret did the heavy lifting. Over multiple runs of the *compare\_model* function, the two that always rose to the top were *xgboost* and *catboost.*

**Do deep learning neural networks offer any benefits over “traditional” machine learning models?**

This is an interesting question. I did not attempt to swap the datasets to see how the Machine and Deep Learning models would fare with their roles reversed. Empirically I would say that after working with the two models I would think they would adapt. This is something I will investigate in the near future.

**For the Images Classification model, the sample size is relatively small. Was overfitting an issue?**

It was, but it was also easily rectified. I randomly selected five as the number of epochs, as it wasn’t too small, but it wasn’t so large that it would take an inordinate amount of time to train. By then plotting the loss rate, it was observed that the training and validation rates began to diverge significantly after the third epoch. At that point, I deleted the trained model and re-trained it on only 3 epochs.

**Why does this research matter?**

The key to successfully treating breast cancer is early detection. There are only so many human radiologists and in third world countries this is a huge problem. By continuing to research, test and develop better algorithmic models to detect and predict breast cancer, literally millions of lives can be saved.

**Conclusion**

In this paper I believe I’ve proven the potential efficacy of developing Machine and Deep Learning models that can detect and predict whether a benign or malignant tumor exists in observable data. While there is an existing concern in the community of radiologists regarding loss of jobs, the priority *must be on saving lives.* I was laid off from a company for which I was then laid off, again, a year later. Neither of these events were pleasant; the first time it was deeply traumatizing. But eventually I found my footing and I am now working for a caring employer and enjoy what I do. So yes: losing a job is traumatizing. But losing your life is so much worse. Artificial Intelligence, Deep Learning, and Machine Learning all have enormous potential to positively impact the reduction in breast cancer deaths, by detecting and predicting malignancy with both better accuracy and better speed. We cannot afford to *not* leverage these technologies.

**References**

[1] Sammari S. (2021) Breast Cancer Image Classification

<https://www.kaggle.com/midouazerty/breast-cancer-images-classification>

[2] National Cancer Institute (n.d.) Breast Cancer – Mammograms

<https://www.cancer.gov/types/breast/mammograms-fact-sheet>

[3] UCI Machine Learning (2016) Breast Cancer Wisconsin (Diagnostic) Data Set

<https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>

[4] Mooney, P. (2019) Breast Histopathology Images

<https://www.kaggle.com/paultimothymooney/breast-histopathology-images>