Automatic Classification of Breast Tumors from MRIs using a Convolution Neural Network

Muneeza Azmat

Department of Computational Mathematics, Science and Engineering

Abstract

Due to the challenges in manual classification, computerized breast tumor classification and scoring is an important problem in medical imaging. In this project a fully convolution neural network (CNN) architecture is used to classify breast tumors, as benign or malignant, from magnetic resonance images (MRI). This project also combines patient data in the form of shallow features to improve on the image-only methods.

Data Analysis

The data is very diverse in terms of shape and size of the breast, presence of non-tumorous lumps, location of the tumor, and the quality of image. This makes it extremely difficult to accurately classify the MRIs. The data is also severely class imbalanced with 92.2% of the data being tumorous and only 7% of the data being non-tumorous. In addition to the MRI data, we also have other data (patient age, indication etc.) which is used as *'shallow features'* to improve the predictions of the CNN model.

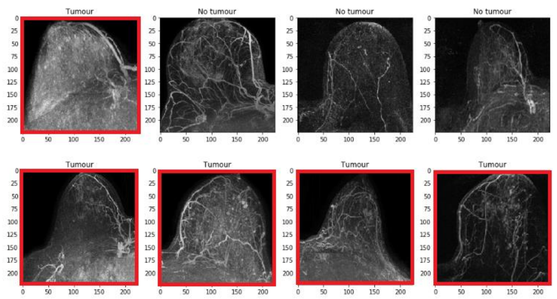
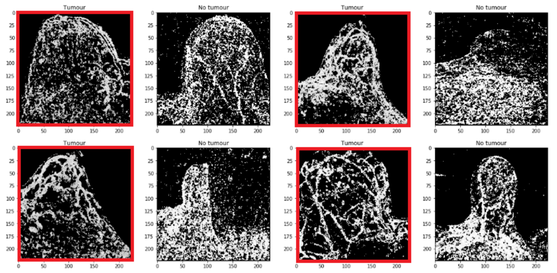


Image Augmentation

To make the model more robust, images are augmented and fed to the model together with the original images. One augmentation technique used in this project is *'high entropy thresholding'*. The idea is that tumorous regions typically have more boundaries and consequently have a high entropy. For each pixel in the input image the entropy filter outputs the entropy value of a $d \times d$ neighborhood around the corresponding pixel, the entropy is calculated using:

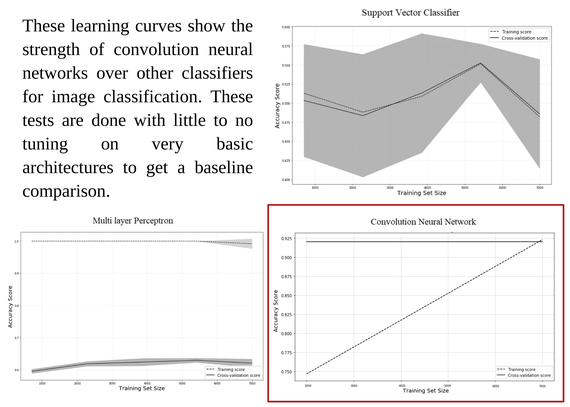
$$S = -\sum_{i} P_i \log_2(P_i)$$

where Pi is the probability that the difference between the values of two adjacent pixels is i.



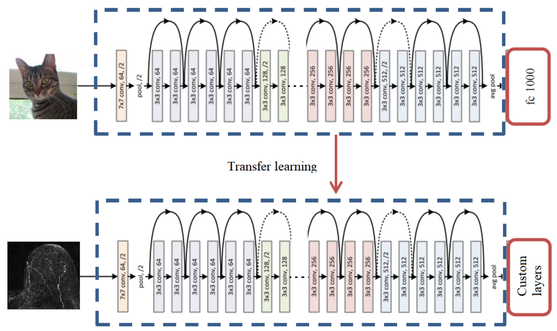
Methodology

These learning curves show the strength of convolution neural networks over other classifiers for image classification. These tests are done with little to no tuning on very basic architectures to get a baseline comparison.

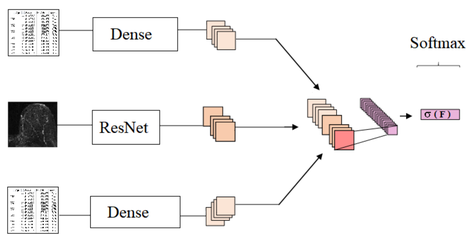


The convolutional architecture used in this project is a residual neural network ResNet, pre-trained on 1.2 million images in the ImageNet database. These pre-trained weights are kept for the most part (excluding the last layer) and custom layers are added to the 50-layer ResNet to include the shallow (non-image) features and make predictions based on both image and non-image patient data.

Network Architecture



The custom layers take multiple features from the ResNet and the shallow neural nets. It then concatenates all these features and passes them through a dense layer with a SoftMax function to get class probabilities.

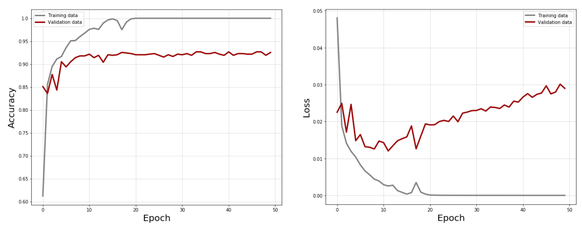


Training

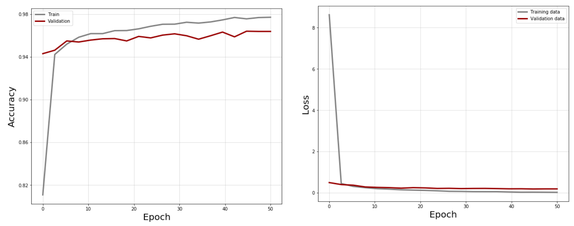
During the training phase, the weights of the first 49 layers of ResNet are frozen and only the last two fully connected layers are trained along with the shallow neural nets with age and indication information. L1 and L2 regularization is used and the training is done using Adam optimizer with a starting learning rate of 1e-4.

To prevent overfitting, the training data is further split into training and validation data and the cross validation is done using this validation data. If he validation loss does not improve after a fixed number of consecutive epochs, the training is stopped.

The learning curves for training done with an unweighted cross-entropy loss is shown below:

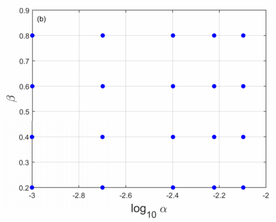


Since the problem is extremely class-imbalanced and there are very few tumorous cases (7%) the model labels all breasts as tumorous and the validation accuracy stops increasing after 92% which is the percentage of tumorous breasts in the data. To remedy this, the loss function is weighted in a way which penalizes false positives more than false negatives. With the new weighted cross-entropy loss the learning curves are :



Hyperparameter Tuning

The coefficients of L1 and L2 regularizers are tuned using grid search.



L1 coefficient	0.001
L2 coefficient	0.2

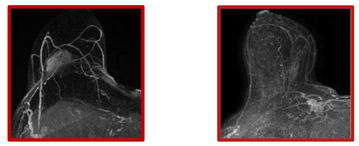
Results

The performance of this classifier on test data is summarized in the confusion matrix below

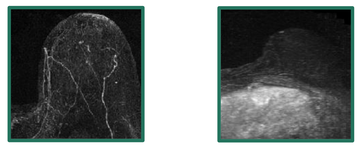
		Predicted	
		Absent	Present
Actual	Absent	102	109
	Present	0	2514

Accuracy	0.96
Sensitivity	1.0
Specificity	0.48

The main challenge for this classifier is dealing with the extreme class imbalance even though the image augmentation, shallow features and tuning greatly improved results, it still has a high false positive (FP) rate. Some of the breasts incorrectly classified as tumorous are shown below:



Looking at the false positive test images, an interesting observation can be made. Most of the breasts which are incorrectly labelled as tumorous are extremely nodular or have node like structures present. If we look at the images that are correctly classified as non-tumorous (below) it is very likely that the classifier is associating the nodular structures with tumors.



Acknowledgements

This project extends the work done by Carina Pereira as part of her master's thesis with Prof. Adam Alessio, her prior research has been quite helpful for this project.

I am also thankful for the help of my advisor Adam Alessio throughout this project.