Deep Learning as a Cancer Diagnosis Tool

Alicia Bilbao Martinez[1], Allison J. Dusek[1,2], Alexander Claman[1], Daniel Bilbao Cortes[3], Vanessa Aguiar-Pulido[1]

[1]Department of Computer Science, University of Miami, Coral Gables, FL, USA

[2]Department of Mathematics and Computer Science, Samford University, Birmingham, AL, USA

[3]Department of Pathology and Laboratory Medicine, Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL, USA

Introduction

In 2021, breast cancer became the most common cancer globally with more than 3.8 million women reporting a history of breast cancer. About 43,250 women in the U.S. are expected to die in 2022 from breast cancer, yet the overall deaths have decreased by 1% between 2013 and 2018 as a result of advancements in treatment and early diagnosis through screening.[1] Evaluation of histopathologic stained tissue is a vital contributor to the broadening and progress of treatment options. The digitalization of tissue slides has made such evaluation more feasible thanks to advances in slide scanning technology and the affordability of digital storage in recent years. Digital slides can be analyzed using deep learning, which has shown promising results as a diagnosis tool.

Objective: Use deep learning to classify tumors as benign or malignant (Fig. 1).

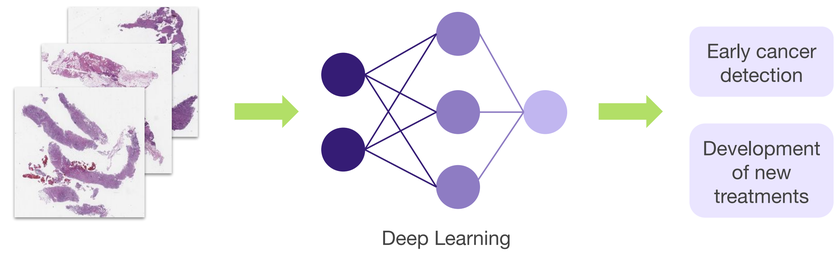


Figure 1 | **General approach for slide analysis using deep learning**

Methods

BreakHis Dataset[2]
4,960 images of breast tumor tissue slides collected from 82 patients in different resolutions (4X, 10X, 20X, 40X) subsequently divided as benign and malignant.

Stratified k-Fold Cross-Validation
Cross-validation uses different portions of the data as training set and validation set for each iteration (Fig. 2). Stratified sampling ensures all classes are present in the same proportion in all sets.

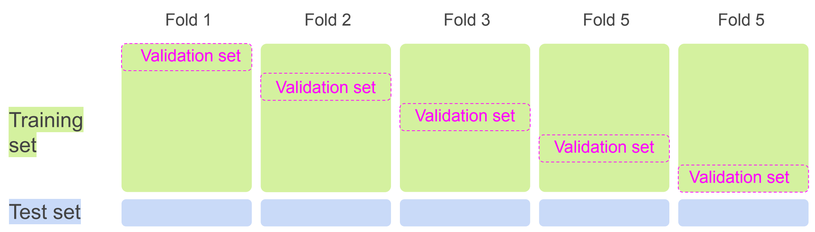


Figure 2 | **5-Fold Cross Validation**

Convolutional Neural Networks (CNNs)
A type of artificial neural networks used to analyze spatial data with local dependencies, such as images.

- Convolutional layers consist of filters that scan the image and produce a value which quantifies the match between the filter and the image.
- Pooling layers apply pooling operations to reduce the image and coarsen the signal.
- The output of convolutional layers can be used as input to fully connected layers to make the final prediction.

Transfer Learning
Incorporating prior knowledge by initializing a model with the majority of parameters from an already trained model (Fig.3).

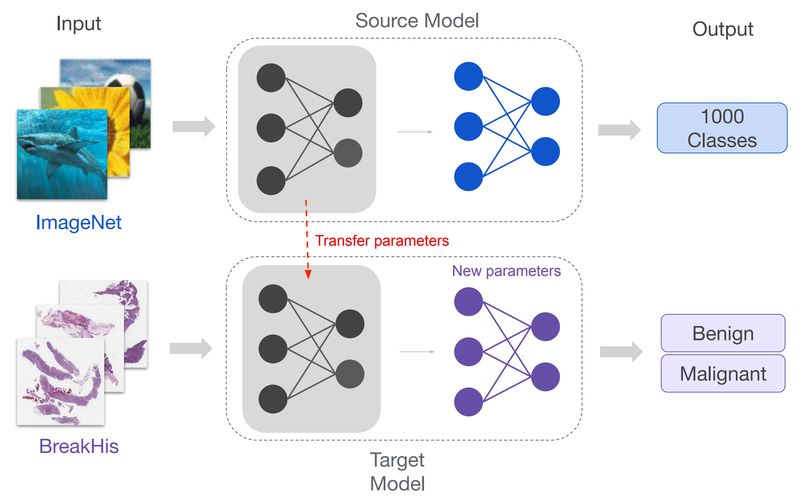
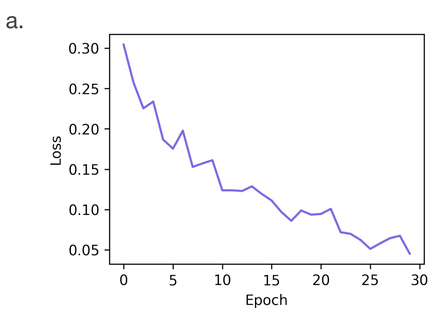


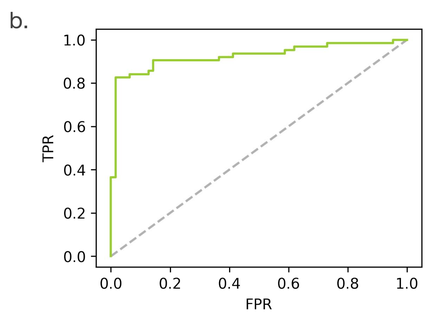
Figure 3 | **Transfer learning with VGG16 trained on ImageNet**

In this work:

- We first divided the full dataset into train (90%) and test (10%). We used stratified 5-fold cross-validation on the training set to select the best model and further tested it on the hold-out (test) set.
- We used VGG16[2], a convolutional neural network with 13 convolutional layers, 5 max-pooling layers, and 3 fully connected layers, resulting in a total of 16 layers. It uses a 3x3 filter with a stride of 1.
- We used transfer learning with the VGG16 model trained on the ImageNet dataset as our source model.

Results





c.

Metric	CV	Test
Accuracy	0.7214	0.8730
Precision	0.7297	0.8924
Recall	0.7036	0.8730
AUROC	0.7989	0.9247
F1	0.7164	0.8571

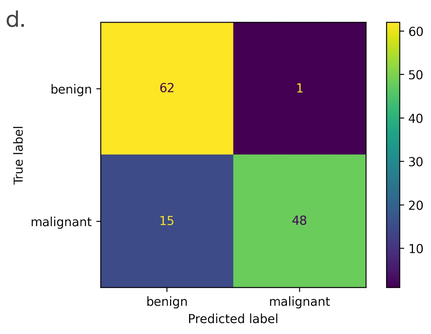


Figure 4 | **Results of applying VGG16 to the 40X magnification**
a. | Loss/Error curve **b.** | ROC curve **c.** | Performance metrics table **d.** | Confusion Matrix for the test (hold-out) set

Conclusions

We built a model to classify a tumor sample slide as benign or malignant using:

- Stratified 5-fold cross validation for hyperparameter optimization.
- A VGG16 architecture trained on the ImageNet dataset as the source model for transfer learning.

Our model achieved 87% accuracy with 89% precision and 87% recall when tested on the test (hold-out) set.

Future Work

Future work concerns deeper analysis of the dataset using different resolutions and architectures, as well as the analysis of other datasets. Some of the ideas include:

- Employing data augmentation approaches to increase the size of the dataset and optimizing the training of our model.
- Using different datasets to test our model to evaluate its performance.
- Building a model for each resolution and comparing the results. This could be interesting as we might get insight into whether having more information about the entire tumor sample, as opposed to a specific subsection, is a deterministic factor in the performance of the model.
- Similarly, different model architectures could be trained on the same data and their results compared to evaluate which one, if any, performs better. We could compare the performances of the different architectures on this dataset with the performances of said architectures trained on different datasets and investigate whether certain architectures perform better for a specific dataset (ex. cancer images vs faces).

Acknowledgements

Support for the program REU Site: Scientific Computing for Structure in Big or Complex Datasets is by the NSF grant CNS-1949972, program solicitation NSF 13-582, Research Experiences for Undergraduates (REU).

References

1. Breast Cancer Facts and Statistics. (2022). Breast Cancer. https://www.breastcancer.org/facts-statistics. Last accessed: 07/22/22.
2. Spanhol F.A. et al. (2015) A dataset for breast cancer histopathological image classification. IEEE Trans. Biomed. Eng., 63, 1455–1462.
3. K. Simonyan and A. Zisserman (2015). Very deep convolutional networks for large-scale image recognition. Proc. Int. Conf. Learn. Represent.