Wavelet and Deep Learning-based detection of Breast Cancer from Histopathology Images

Chris Light, Joshua Orozco, Brandon Cramer

November 19, 2022

Introduction and Motivation

We plan to explore the use of Wavelets and the Discrete Wavelet Transformation (DWT) in conjunction with a Deep Convolutional Neural Network (DCNN) for detections of Breast Cancer from Histopathology images. The motivation for this particular application is due to the significance of breast cancer in the modern day, making up nearly 1/3rd of all new cancer cases for women in the US.

Our acquired data set [1] contains images for four types of invasive breast cancer including ductal, lobular, mucinous, and papillary carcinoma. In addition, the dataset includes four types of benign tumors including adenosis, fibroadenoma, phyllodes, and tubular adenoma. All tumor types contain tissue slide images at 40X, 100X, 200X, and 400X magnification for various patients. The goal of our project is pre-process these images using the DWT and feeding the output into a DCNN to classify the image as having either a benign or malignant result. These results will be compared to the DCNN output without DWT pre-processing to see if the DWT can improve the classification results. Some consideration are whether a single DWT can be applied across all images or if we will have to change the DWT based on the magnification of the image.

Related Work

We are planning to follow the work of A.K. Verma et al [2], where the DWT and a DCNN is applied to detect SARS-nCoV from X-ray images. Their work explores two-level DWT with various types of mother wavelets and DCNN architectures.

In [3], *Araujo et al.* explore the use of a DCNN as well as Support Vector Machine (SVM) classifier to distinguish different type of malignant cells. They classify images as containing either normal tissue, benign lesion, *in situ* carcinoma, and invasive carcinoma. The differentiation between the two types of carcinoma is of large interest when getting a second specialist opinion. The DCNN and SVM architecture is designed to extract features from images which have magnifications ranging from nuclei size to overall tissue structure. Nuclei level images help distinguish if the cells are malignant while the larger images allow the network to distinguish between *in situ* and invasive carcinoma. Their results show that using a binary classification method of carcinoma being present was more accurate than using the four classifications listed above. Our dataset is similar in that it has slides from multiple magnification levels and their work shows that using images with different magnifications can aid in the detection of malignant cells.

Approach

We plan to experiment with more complex wavelet packages (e.g. second-level decompositions of more than just the low-pass/low-pass output) to try to improve upon the work in [2]. We will investigate differences in performance between first-level and second-level decompositions. We will explore different choices of mother wavelet function, different wavelet types (orthogonal vs. biorthogonal), and varying filter lengths. Specifically, we will look at the benefits and trade-offs of using more complex wavelet transforms (i.e. higher-level, longer filters, etc.) to look at how the transform can be optimized for our particular use-case. This trade-off study will also include a comparison with the case when the DWT is not used and the raw images are fed into the DCNN directly.

Some of the specific areas of comparison that we will look at in this performance trade-off study will include general classification accuracy with both benign vs. malignant classifications, as well as specific tumor classification (subcategories within benign and malignant), classification accuracy in the presence of added noise, and efficiency/speed of classification. We will compare how the different wavelet package selections perform with color data (RGB layers) vs. with only a single grayscale layer (do certain wavelet transforms work better for color vs. grayscale data?). We will also compare how they perform with different magnification levels (should the same waveform transform be used for 40x vs. 100x vs. 200x vs. 400x magnifications?).

References

[1] Fabio A. Spanhol, Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7):1455–1462, 2016.

[2] Amar Kumar Verma, Inturi Vamsi, Prerna Saurabh, Radhika Sudha, Sabareesh G.R., and Rajkumar S. Wavelet and deep learning-based detection of sars-ncov from thoracic x-ray images for rapid and efficient testing. *Expert Systems with Applications*, 185:115650, 2021.

[3] Teresa Araújo, Guilherme Aresta, Eduardo Castro, José Rouco, Paulo Aguiar, Catarina Eloy, António Polónia, and Aurélio Campilho. Classification of breast cancer histology images using convolutional neural networks. *PLOS ONE*, 12:e0177544, 06 2017.