Research Project Proposal: Cancer Classification using Transfer Learning

Introduction:

The advance of deep learning has made huge changes in computer vision and produced various off-the-shelf trained models. Particularly, Convolutional Neural Network (CNN) has been widely used to build image classification model which allow researchers to transfer the pre-trained learning model for other classifications. We propose a transfer learning method to detect cancer using histopathology images based on Google's Inception v3 model which were initially trained for the classification of non-medical images. The pilot study shows the feasibility of transfer learning in the detection of cancer.

Recent research using transfer learning have obtained prominent results in image analysis. Transfer learning is a method that trains a pre-trained model, which is already learned in a specific domain, to another knowledge domain. Transfer learning method is known to be very useful when the data is not enough or training time and computing resources are restricted.

Background:

Cancer is one of the main causes of death worldwide. Biopsy is the only diagnostic procedure that can definitely determine if a suspected region is cancerous. A biopsy involves the extraction of sample cells or tissues for examination. Psychopathology refers to the examination of the specimens extracted. Diagnosis from analysis of histopathological images is the gold standard in diagnosing a considerable number of diseases, including almost all types of cancers

Computer-aided Diagnosis systems contribute to reduce the cost and increase the efficiency of this process. Conventional classification approaches rely on feature extraction methods designed for a specific problem based on field-knowledge. To overcome the many difficulties of the feature-based approaches, deep learning methods are becoming important alternatives. Especially, deep neural network has shown outperformance in image analysis due to the development of computing resources.

Data Sources:

1. Vision and Image Analysis Group/I-ELCAP Public Access Research Database.

2. Curated Breast Imaging Subset of DDSM Dataset

3. Bioimaging Challenge 2015 Breast Histology Dataset

Algorithms:

1. **Training Accuracy & Cross-Entropy:** https://rdipietro.github.io/friendly-intro-to-cross-entropy-loss/

2. **Optimizing Cut-Off:** Classification task to assist medical diagnosis has asymmetric misclassification cost, since the cost for missed detection of cancer (false negative) is higher than the false positive classification. Optimizing cut-off value method is used for such asymmetric misclassification cost.

3. **ROC Curve:** Receiver Operating Characteristic curve, i.e. ROC curve, is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings.

References:

I. **Convolution Neural Network – Wikipedia - https://en.wikipedia.org/wiki/Convolutional_neural_network**
II. **IEEE Library – Study on histopathology of cancer - http://ieeexplore.ieee.org/document/8210843/**
III. **Transfer Learning by Sebastian Ruder - http://ruder.io/transfer-learning/**
IV. **Classification of breast cancer histology images using CNN - http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0177544**