

Histopathological Image Analysis using MIL and Transfer Learning

Team Members:

- Lokesh Vadlamudi
- Nupur Yadav
- Chetan Kulkarni

Introduction

- Classifying histopathological slides is challenging because of the two reasons:
 1. the histopathological images are high-resolution images that are meant to be examined under the microscope by the pathologists
 2. it is extremely difficult to obtain a large dataset of WSIs with detailed pixel-wise annotations that makes it challenging to use a supervised deep learning classifier
- In this project, we have used two techniques:
 1. Attention-based Deep Multiple Instance Learning which has proven to be successful in a wide range of medical imaging applications.
 2. A “DenseNet201” architecture with weights pre-trained on ImageNet.

Attention-based MIL Approach

- Multiple instance learning (MIL) is a variation of supervised learning where a single class label is assigned to a bag of instances.
- MIL uses the whole-slide labels instead of using pixel-wise annotations.
- In attention-based MIL approach, we have cropped an image in small square patches and made a bag of it. This bag of images will act like a batch.
- Passing it to the feature extractor which is basically a CNN block to generate instance level features.
- Then we are passing the Instance level features to the classifier for getting Instance level attention.
- Here we are getting the attention weights which we are further using for attention aggregation to get the bag level features.
- Then we are applying Dense layer for the classification of the Benign or Malignant

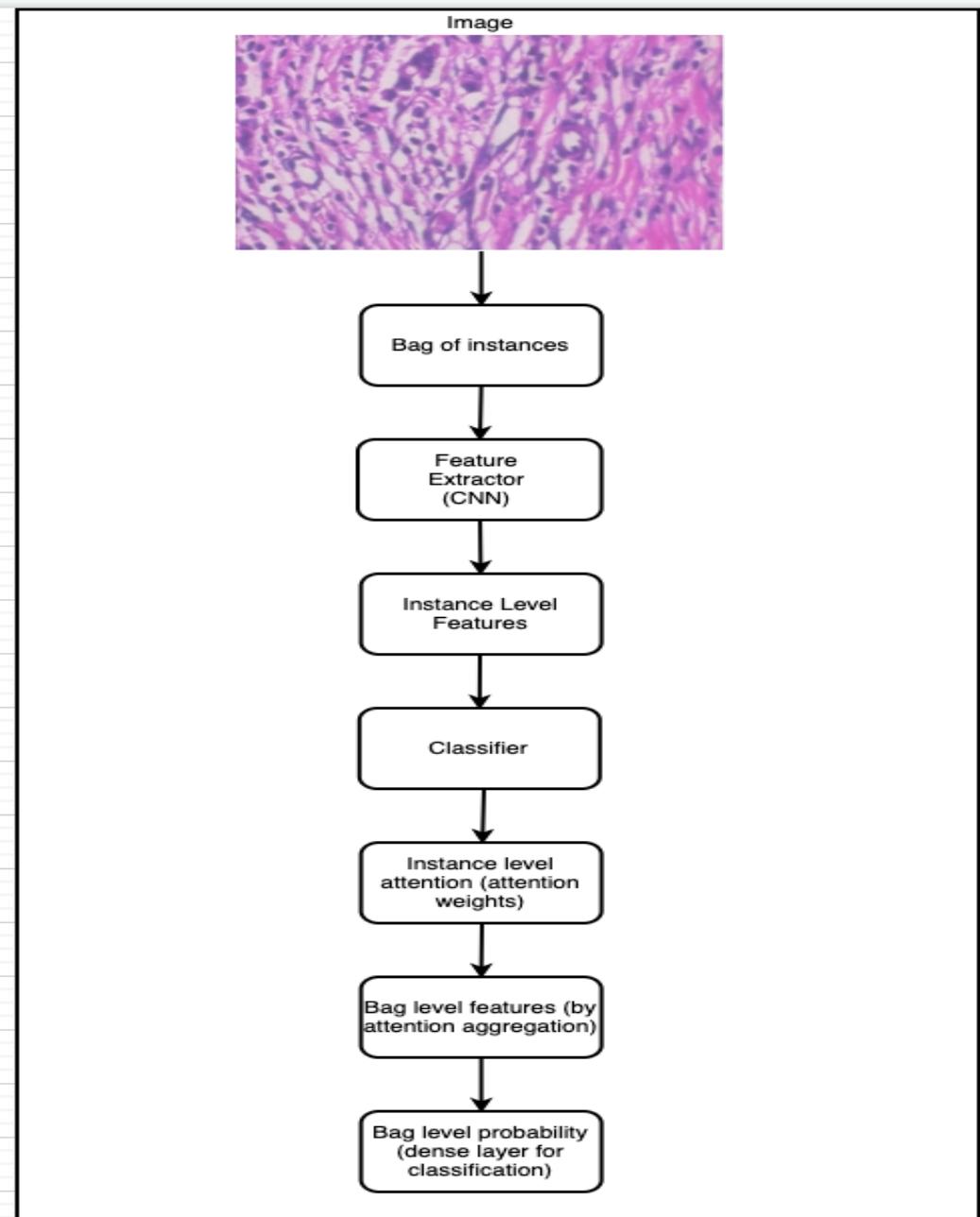


Fig 1. Attention-based MIL approach

Transfer Learning Approach

- Transfer learning allows us to train deep networks using significantly less data than we would need if we had to train from scratch.
- The idea is that the two tasks are not totally disjoint, and as such we can leverage whatever network parameters that model has learned through its extensive training, without having to do that training ourselves.
- Transfer learning has been consistently proven to boost model accuracy and reduce required training time. Less data, less time, more accuracy.
- In this approach, we have used “DenseNet201” architecture with weights pre-trained on ImageNet data.
- We have used ReduceLROnPlateau callback which reduces learning rate when a metric has stopped improving.
- Models often benefit from reducing the learning rate by a factor of 2-10 once learning stagnates. This callback monitors a quantity and if no improvement is seen for a 'patience' number of epochs, the learning rate is reduced.
- This model achieved a better accuracy than the attention-based MIL approach on breast cancer dataset (BreakHis).

Datasets

- We have used two datasets in our project. Both are taken from Kaggle.
1. [BreakHis \(Breast Cancer Histopathological Database\)](#) : This dataset is divided into two main groups: benign tumors and malignant tumors. It is composed of 9,109 microscopic images of breast tumor tissue collected from 82 patients using different magnifying factors (40X, 100X, 200X, and 400X). To date, it contains 2,480 benign and 5,429 malignant samples (700X460 pixels, 3-channel RGB, 8-bit depth in each channel, PNG format).
 2. [Breast Histopathology Images \(IDC\) dataset](#) : Invasive ductal carcinoma (IDC) is the most common form of breast cancer. The original dataset consisted of 162 whole mount slide images of Breast Cancer 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).

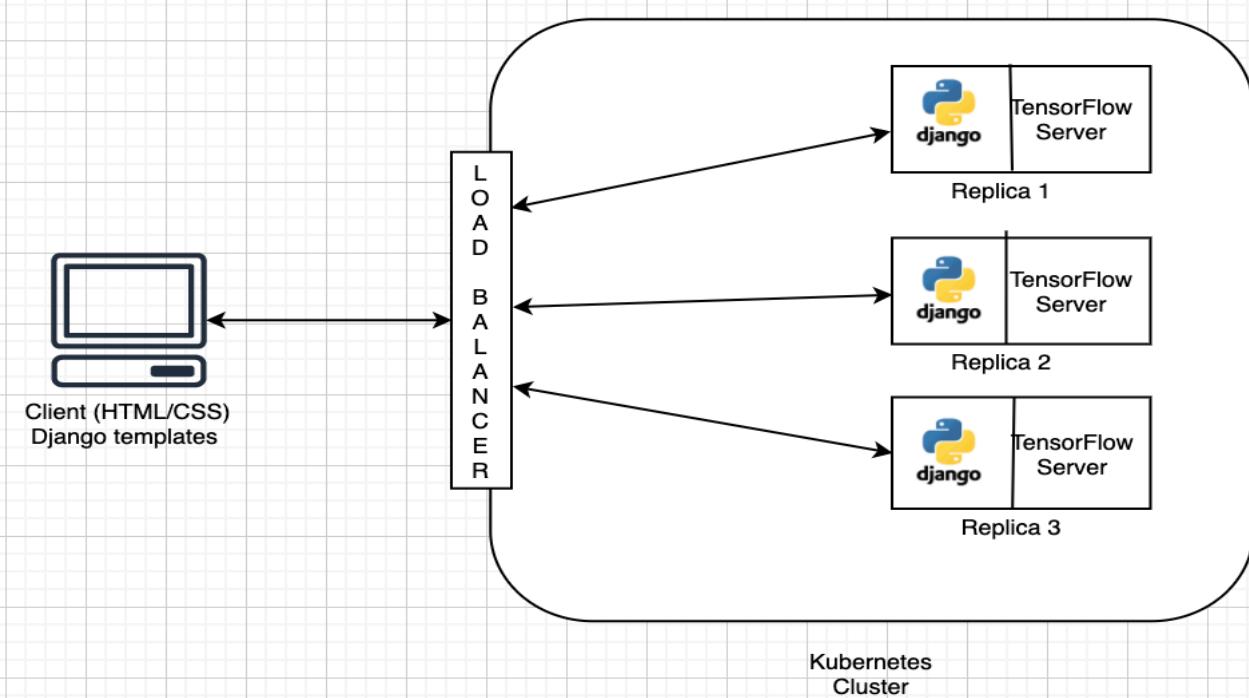
Observations

Technique	Accuracy
Attention-based MIL on BreakHis dataset (Pytorch implementation)	69%
Attention-based MIL on Breast Cancer histopathological images (different dataset, data augmentation and Keras implementation)	73%
Transfer Learning on BreakHis dataset (using DenseNet201)	94%

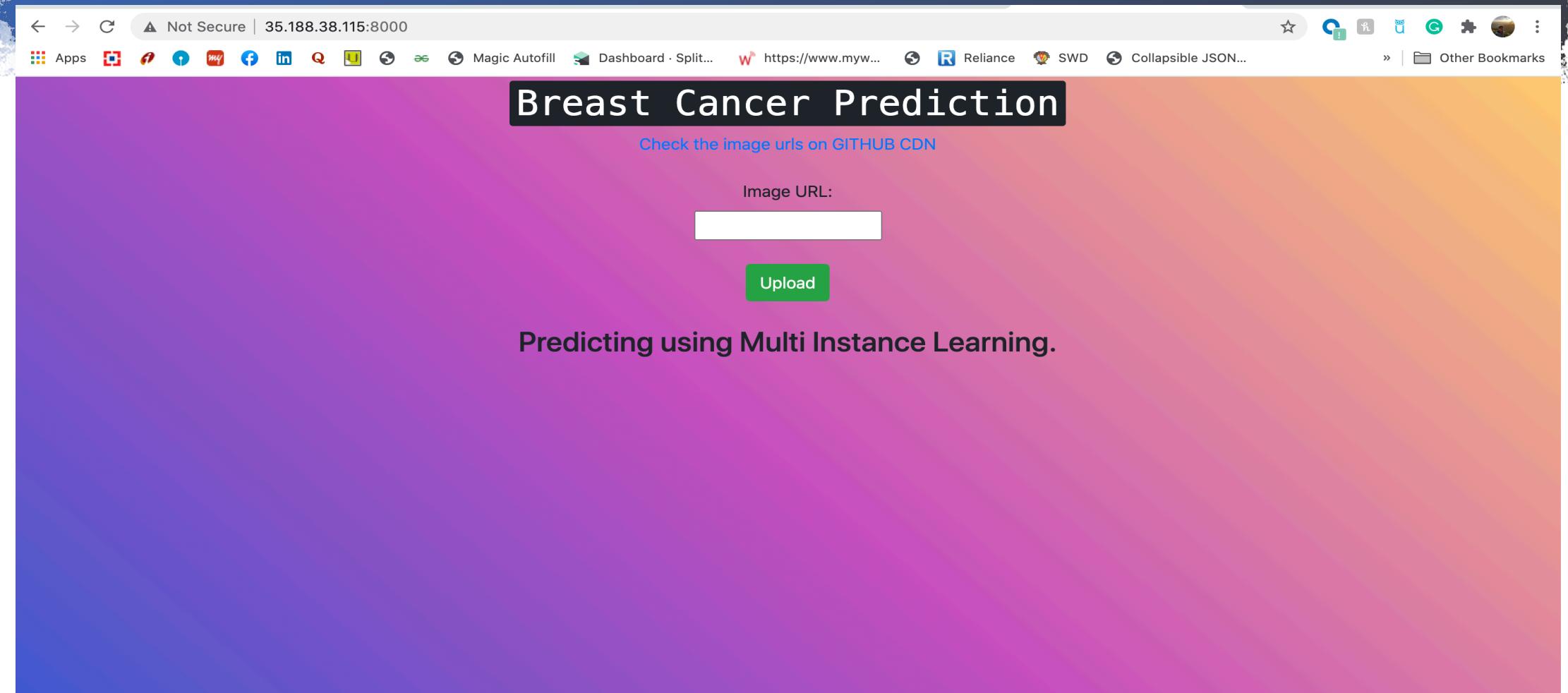
Hence, we used this transfer learning model for deployment and making future inferences.

Deployment architecture

We have used Django for building web application and TensorFlow serving, docker, Kubernetes for model deployment.



UI screenshots



UI screenshots

Not Secure | 35.188.38.115:8000/homer

Apps Dashboard · Split... SWD Collapsible JSON... Other Bookmarks

Breast Cancer Prediction

Check the image urls on GITHUB CDN

Image URL:

Upload

Predicting using Multi Instance Learning.

Original Image

Status : Malignant

Code Walkthrough and Demo..