Project Synopsis

Histopathologic Cancer Detection

Prepared by

Ketan Anil Muddalkar

Rohan Prashant Dhere

Under the guidance of

Prof. Pravin Adivarekar

Bachelor of Engineering (BE)

Department of Computer Engineering

A. P. Shah Institute of Technology, Thane

University of Mumbai

2019-2020

Histopathologic Cancer Detection

Abstract

We are implementing an algorithm to identify metastatic cancer in small image patches taken from larger digital pathology scans. These images will be preprocessed and augmented before passing it to the neural network model in order to improve the training process. In this project we aim to classify cancer images with maximum accuracy using deep learning techniques like Convolutional Neural Network(CNN).

The dataset which we intend to use will contain 2 classes, class one will consist of images of cancer-free patients and class two will consist of images of patient having cancer. We expect to work with 2 or more models which will be ensembled to get better accuracy compared to a single model for this we will build a CNNs from scratch and also be using transfer learning methods. After achieving desired accuracy we intend to deploy the model on the web.

I. Introduction

Cancer is a group of diseases involving abnormal cell growth with the potential to invade or spread to other parts of the body. Following are the symptoms observed lump, abnormal bleeding, prolonged cough, unexplained weight loss.

Histopathology refers to the examination of a biopsy or surgical specimen by a pathologist, after the specimen has been processed and histological sections have been placed onto glass slides. This examination will be done by our deep learning model without human involvement. For any given input image of tissues, the model will try to classify it into two already predefined classes. The classes two being positive cancer detected or negative

.

II. Literature Review

In this project we aim to incorporate all the best methods in each step of creating a deep learning model. Research in this field has been in constant state of motion as cancer is a significantly serious disease in our community. There have been several studies utilising transfer learning, especially with CNNs to detect and classify cancerous regions in histopathology images. One of the transfer learning approach is to use an architecture that has done well in other tasks and to train it from scratch. And the other approach is to use a pre-trained model as a feature extractor and perform further classification with a seperate classifier.

Finally, the most prevalent way to perform transfer learning is to employ a pre-trained model and to fine tune it with data at hand. Several fine-tuning approaches can be utilized such as fine-tuning all the layers, freezing the initial layers and many more. Even though the domains are considered both visually and in nature very different, most of the transfer learning schemes are architectures or models are trained on ImageNet dataset. The performance of the deep learning network models depends on overall task at hand, dataset used, evaluation setup and much more.

III. Problem Statement

Cancer is a leading cause of death worldwide, accounting for an estimated 9.6 million deaths in 2018. Cancer arises from the transformation of normal cells into tumour cells in a multistage process that generally progresses from a precancerous lesion to a malignant tumour if undetected. Cancer cells are hard to detect for human eyes even for a professional pathologist.

We intend to focus on deep learning techniques which will be capable of identifying cancer cells upto micro level. Hence with the help of this deep learning method we intend to speed-up the process of cancer detection with high accuracy and hence contribute towards reduction in number of deaths caused due to cancer.

IV. Objective

To create Deep Learning model that will accurately detect cancer cells present in the tissues from the given histopathologic images. Cancer is a leading cause of death worldwide which can be cured if detected at an early stage. Thus it only makes detection of this disease our utmost priority to tackle this problem. To make this process smoother and fluent we are proposing a model which will help doctors in taking decision correctly & accurately.

We also aim to deploy the perfected model on the internet in the hopes that it will serve as a base for hospitals in the direction of fully digitizing their organization.

V. Scope

The scope of this project is to deliver a model, with world-class accuracy, which can detect cancer from histopathology images and produce top notch results. The project will involve gathering data from hospitals or websites. Preprocessing data as per requirements to make it suitable for our model. The model will be compiled on Keras with Tensorflow as its backend. We expect the model to output the class to which the given image will belong(i.e. Cancer detected or not). Furthermore, we are planning to deploy the model on a website which can act as an example for hospitals to create a similar portal for the benefits mentioned ahead.

VI. Benefits for environment

Training several large machine learning models with approximately 100 million parameters will have carbon footprint of roughly 600 kg also tuning them will require significant amounts of energy. To reduce this amount we'll be using an approach of transfer learning where pretrained models can be used to train on a new dataset. This method is computationally inexpensive which can save a lot of energy and help the environment.

VII. Benefits for Society

This project aims to provide world class accuracy in detecting cancer from histopathology images. We expect our project will have a positive impact on the society at the cost of low to almost zero resource consumption. This project can act as an asset to doctors who are working in hospitals with cancer patients. With the help of this project will speed up the process of cancer detection, and in-turn help the patient in getting the medical attention required at the earliest.

VIII. Application

Our model will be trained on images of patients having cancer. The same model, with minor tweaks, can be modified to work with MRI and CT scan images as well. Thus expanding its horizon in the medical field. These models can later be deployed on web. This will help the patient and doctors and all the system administrators to maintain a complete record of all the patients in one single go. Each hospital can maintain a portal by the help of which patients can access their reports and can consult to doctors immediately. In rural areas, hospitals can set-up their facilities and can send qualified doctors with such high-tech solutions and help the patients there.

IX. Technology Stack

A. Languages

a. Python for data preprocessing, data visualization, building deep learning models, Performance measurement graphs.
b. PHP/Python will be used as a backend of website and also to create medical report after diagnosingPHP/Python will be used as a backend of website and also to create medical report after diagnosing.
c. HTML CSS for frontend of website where user can interact with machine learning model.

B. Libraries

a. Tensorflow for machine learning applications such as neural networks.
b. Keras is a high-level neural networks API, running on top of Tensorflow.
c. Pandas,Seaborn for data visualization task.
d. Numpy library will make matrix operations faster compared to traditional matrix operations in python.
e. Scikit learn will be used to calculate confusion matrix which will determine the performance of model.
f. OpenCV library for preprocessing images.

C. Environments/ Tools

a. Apache server will be used to host the entire website.
b. Machine learning models will be trained on "NVIDIA DGX" supercomputer.

References

[1] Bardou, D., Zhang, K. and Ahmad, S.M., 2018. Classification of breast cancer based on histology images using convolutional neural networks. *IEEE Access*, *6*, pp.24680-24693.

[2] M. C. Chun. (2018). Breast Cancer: Symptoms, Risk Factors, and Treatment, Medical News Today. Accessed: Mar. 10, 2018.

[3] L. He, L. R. Long, S. Antani, and G. R. Thoma, ''Histology image analysis for carcinoma detection and grading,'' Comput. Methods Programs Biomed., vol. 107, no. 3, pp. 538–556, 2012.

[4] P. Filipczuk, T. Fevens, A. Krzyzak, and R. Monczak, ''Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies,'' IEEE Trans. Med. Imag., vol. 32, no. 12, pp. 2169–2178, Dec. 2013.

[5] M. Veta, J. Pluim, P. van Diest, and M. Viergever, ''Breast cancer histopathology image analysis: A review," IEEE Trans. Biomed. Eng., vol. 61, no. 5, pp. 1400–1411, May 2014.

Ketan Muddalkar

Rohan Dhere