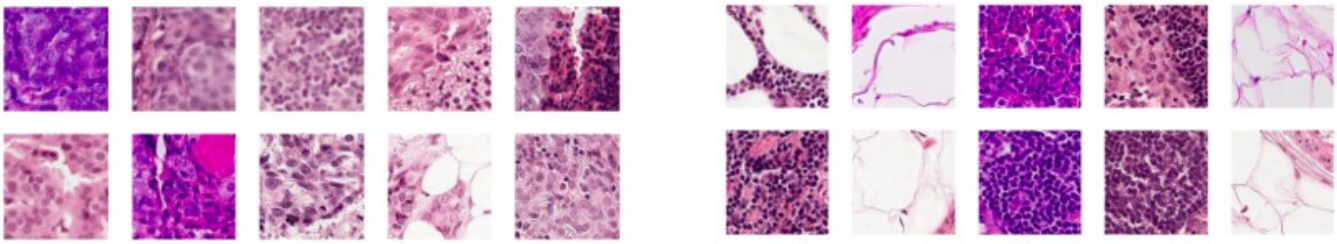
Histopathologic Cancer Detection

Abu Shahid *B20CS003*
Anirudh Srikanth *B20CS006*
Atharva Pandey *B20CS007*

Abstract- Modern medical image processing techniques work on histopathology images captured by a microscope, and then analyze them by using different algorithms and methods. Machine learning algorithms are now being used for processing medical imagery and pathological tools. Manual detection of a cancer cell is a tiresome task and involves human error, and hence computer-aided mechanisms are applied to obtain better results as compared with manual pathological detection systems.In this paper, we share our experience with the exploratory analysis on the data and trying out different models to give the best results.

I. Introduction

Advancing the fight against cancer requires early detection which can only be possible with an efficient detection system. Images are acquired by histopathology, which generally includes biopsy of the affected tissue. These microscopic images can be collected and used for developing computer-aided detection systems. Manual detection is a tedious, tiring task and most likely to comprise human error, as most parts of the cell are frequently part of irregular random and arbitrary visual angles. The goal is to identify whether a tumor is benign or of a malignant in nature, as malignant tumors are cancerous and should be treated as soon as possible to reduce and prevent further complications. In short, it is a binary classification problem and can be resolved by various machine learning methods



Dataset

histopathologic-cancer-detection.zip :The folder contains 2 folders and a csv file

- train: 220k images named with unique id
- test: 55k images
- train_labels.csv: containing ground truth of training set
    - Later a Random_Sample dataset was generated containing 5500 images.

Training dataset has 59.5 cancer positive images and 40.5 cancer negative ones.

II. Methodology

Overview