Cancerous and Non Cancerous Cell Detection

Mentor : Dr. Y. Radhika
Section : B5
Batch Number : 7

SHRUSHTI GUPTA, 121810305020
SHASHANK ALA, 121810305010
ROHITH REDDY, 121810305049
RASAMSETTI SAMPATH, 121810305058
CHINTA SUMANTH, 121810305028

Contents

- Abstract
- Introduction
- Literature review
- Problem Identification and Objectives
- Parameters
- System Methodology
- Project Methodology
- Methods and Technologies used
- Implementation
- Results, Analysis and discussions
- Conclusion and Future Scope

Abstract

- Image processing algorithms are commonly employed in various medical domains to increase disease detection capability. Cancer is one of the most dangerous diseases, and early detection is a challenging task in medicine. The project is based on breast cancer samples. After a comparative examination of commonly used methods in each category, an appropriate and efficient approach is adopted in each of the design processes of the proposed framework. Typically, the cancer detection method entails classifying the image biopsy as malignant or benign. Many abnormalities are observed and classified in microscopic biopsy image analysis by doctors and pathologists based on various characteristics of the cell.

Introduction

- For pathologists and healthcare professionals, cancer detection has always been a difficult task in terms of diagnosis and treatment planning.
- Manual cancer identification using microscopic biopsy images is subjective; findings can vary from expert to expert depending on their experience and other factors.
- The automated identification of malignant tissue from microscopic biopsy images helps to alleviate the problems outlined above and gives improved outcomes when biologically interpretable and clinically significant feature-based approaches are used for disease diagnosis.
- Histopathologists examine the particular traits in the cells and tissue architecture to discover and diagnose cancer from microscopic biopsy photos.
- Shape and size of cells, shape and size of cell nuclei, and cell distribution are frequent parameters used for cancer detection and diagnosis from microscopic biopsy images.

Literature Review

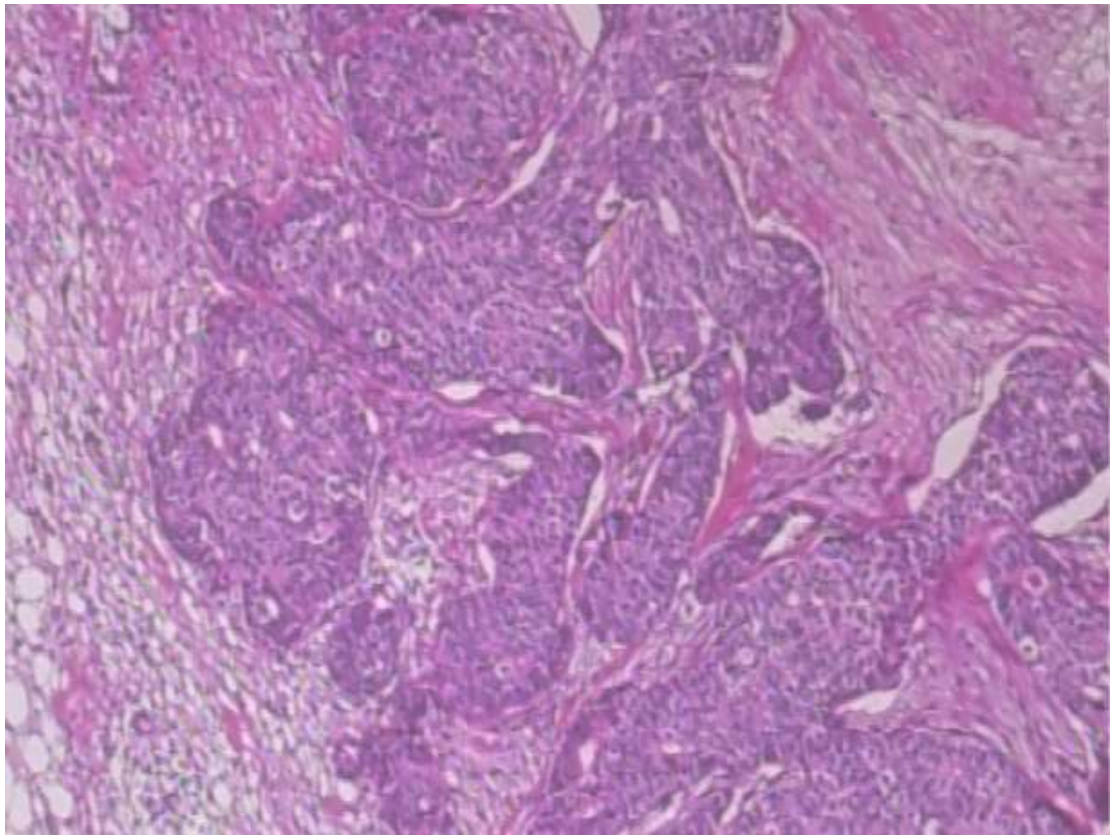
1. Boosting Breast Cancer Detection Using Convolutional Neural Network (Base Paper), 2021
2. Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features, 2015
3. Semantic Segmentation of Colon Glands with Deep Convolutional Neural Networks and Total Variation Segmentation, 2017
4. Lung Cancer Prediction and Detection Using Image Processing Mechanisms: An Overview, 2019
5. Automated classification of benign and malignant cells from lung cytological images using deep convolutional neural network, 2019

Problem Identification And Objectives

- To classify Cancerous and Non-cancerous cells using Image Processing and Machine Learning Techniques.
- To find out the accuracy of the classification with image processing and data augmentation.
- To compare the accuracy of various Machine Learning algorithms to check which one is the most accurate.

Parameters

- In histopathology the categorization of biopsy images is done into Cancerous cell & non - Cancerous cell.
- The tissues have been stained for better identification of nucleus and cytoplasm when watched with the naked eye.
- For cancerous cell, the volume of the nucleus is greater than the volume of the cytoplasm, and vice-versa for non-cancerous cell.



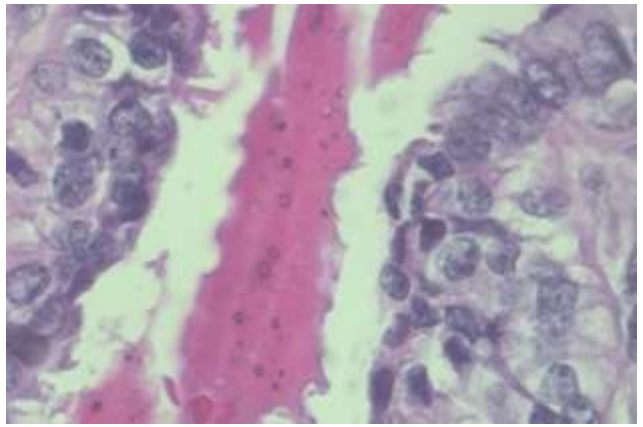
Cancerous tissue

System Methodology

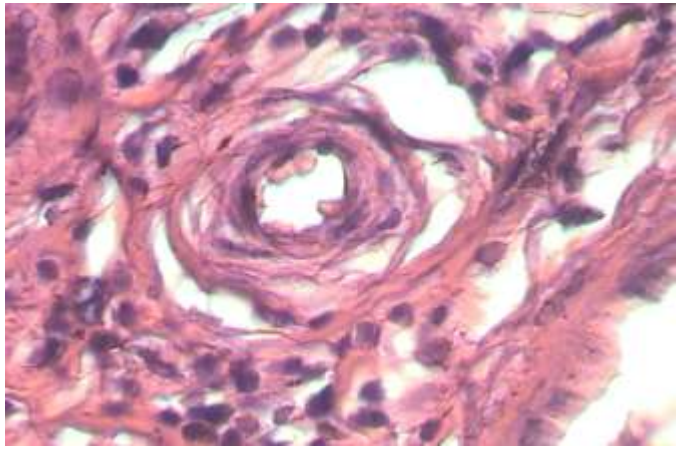
The Dataset : The dataset used in this project is microscopic biopsy images of breast cancer and microscopic biopsy images of non-cancerous cells derived from the Breast Cancer Histopathological Database (BreakHis). The Breast Cancer Histopathological Image Classification (BreakHis) is made up of 1693 microscopic images of breast tumor tissue collected from 82 patients, 547 of which are benign and 1147 of which are malignant (700X460 pixels, 3-channel RGB, 8-bit depth in each channel, PNG format).

https://drive.google.com/drive/folders/193mZ0QTkQj2B7OHR9Y_QwJjcxxO9X7QW?usp=sharing

System Methodology



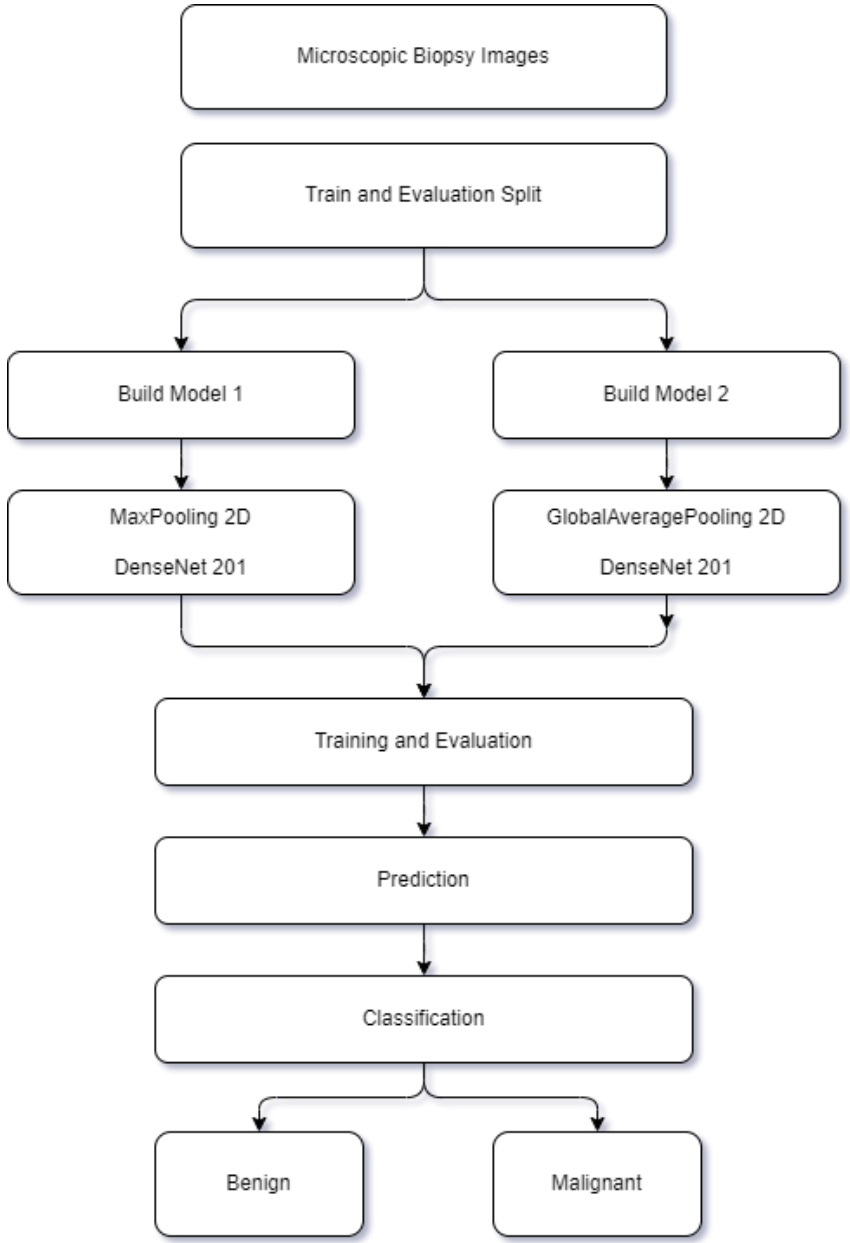
Cancerous tissue



Non-Cancerous tissue

Project Methodology

Project flow chart

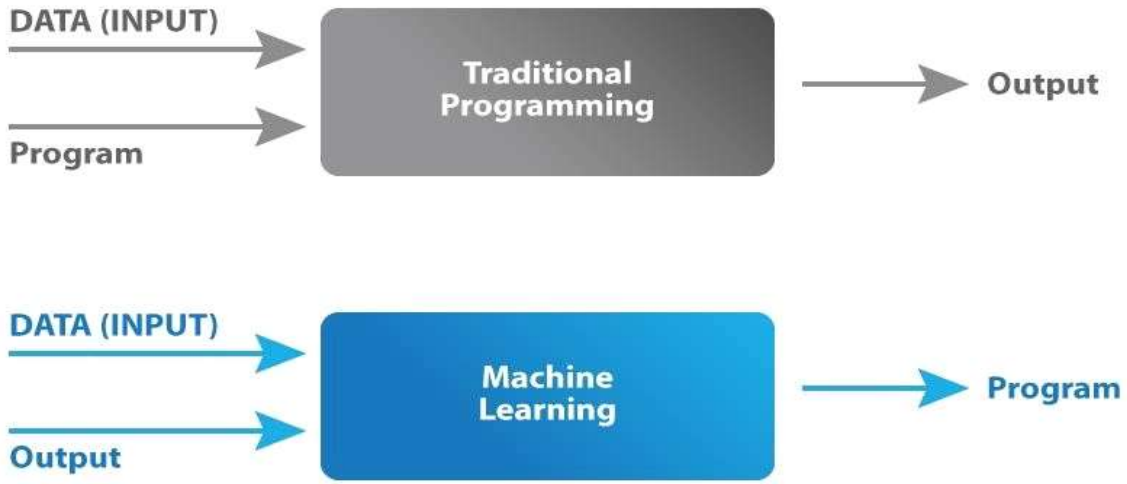


Methods and Technologies Used

- Machine Learning
- Image Processing
- Convolutional Neural Network

Machine Learning

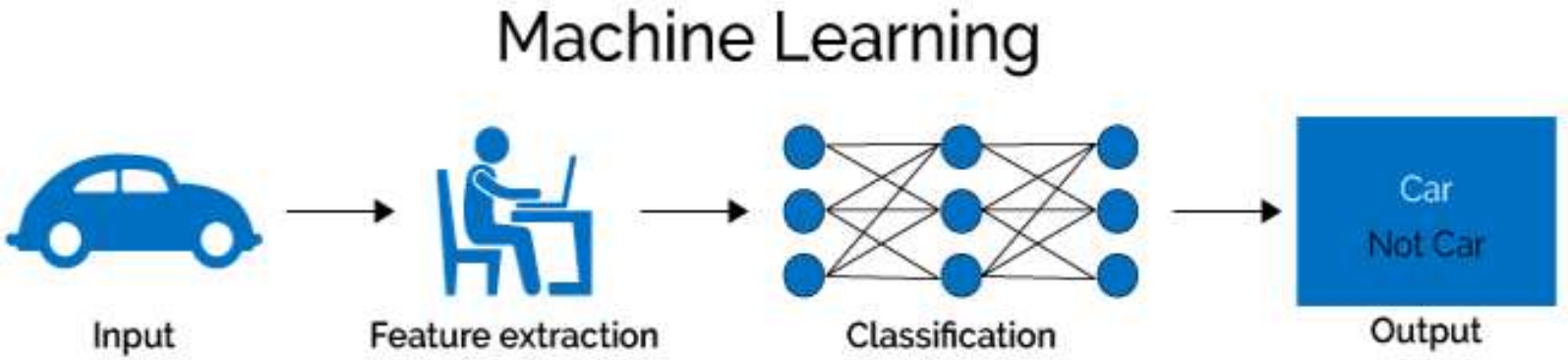
- Branch of artificial intelligence (AI)
- In order to forecast new output values, machine learning algorithms employ historical and statistical data as input.



Machine Learning

Machine Learning

- Inputting training data into the chosen algorithm is the first step in the Machine Learning process.
- New input data is fed into it to see if the machine learning algorithm is working correctly.
- After then, the prediction and the results are compared.
- If the prediction and the results do not match, the algorithm is re-trained until the data scientist achieves the desired result.

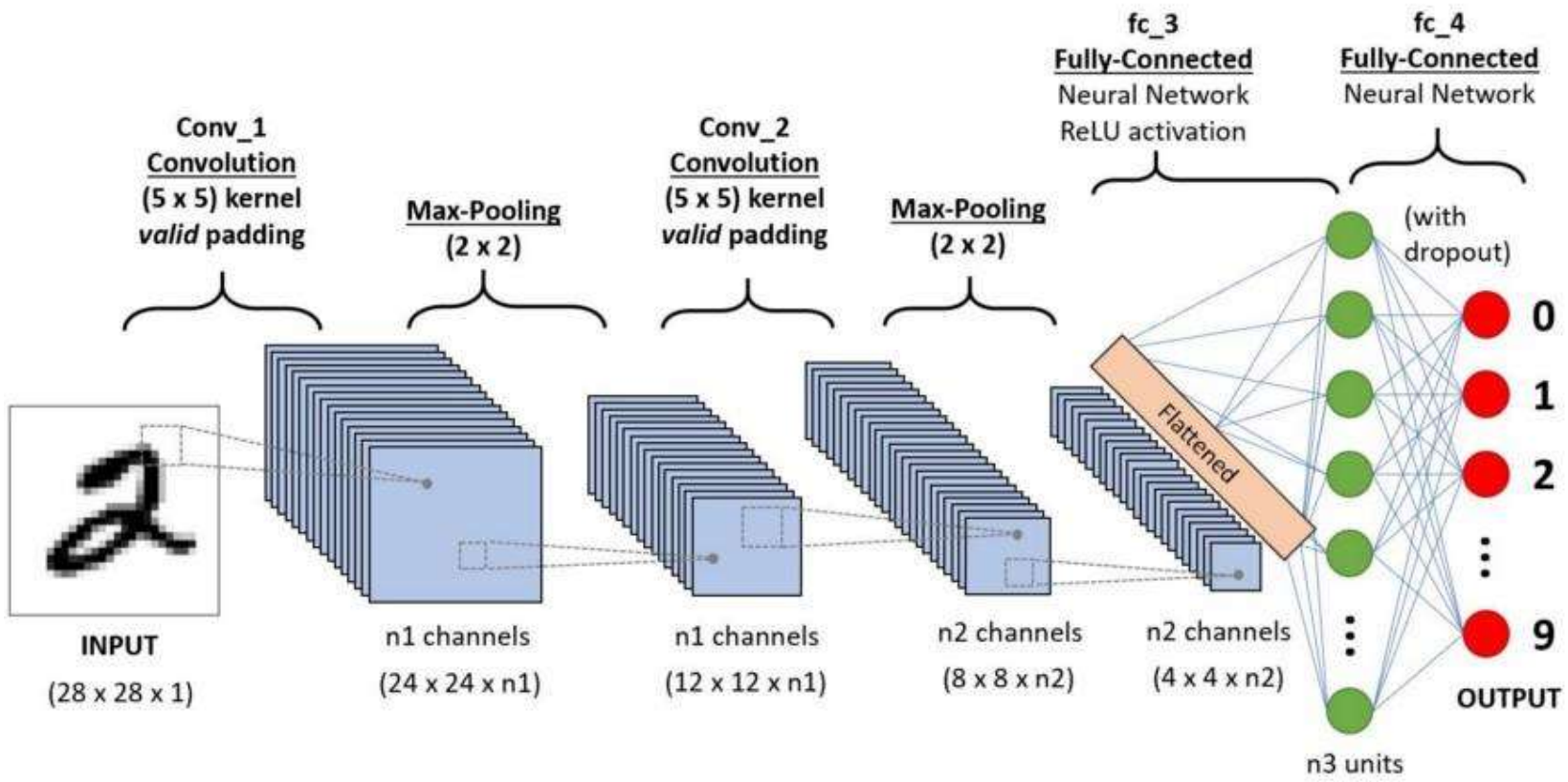


Machine Learning example

Convolutional Neural Network

- A CNN or Convolutional Neural Network is a deep learning neural network designed to analyze structured arrays of data-like representations. It employs a technique known as Convolution.
- CNN's are excellent at detecting unique features in input images, such as lines, gradient circles, and even eyes and faces.
- Because of this feature, convolutional neural networks are effective in computer vision.
- CNN does not require pre-processing and can run straight on an under-done image.
- Convolutional Neural Network strength stems from a layer known as the convolutional layer. CNN is made up of multiple convolutional layers stacked on top of each other, capable of identifying more complex structures.

Convolutional Neural Network



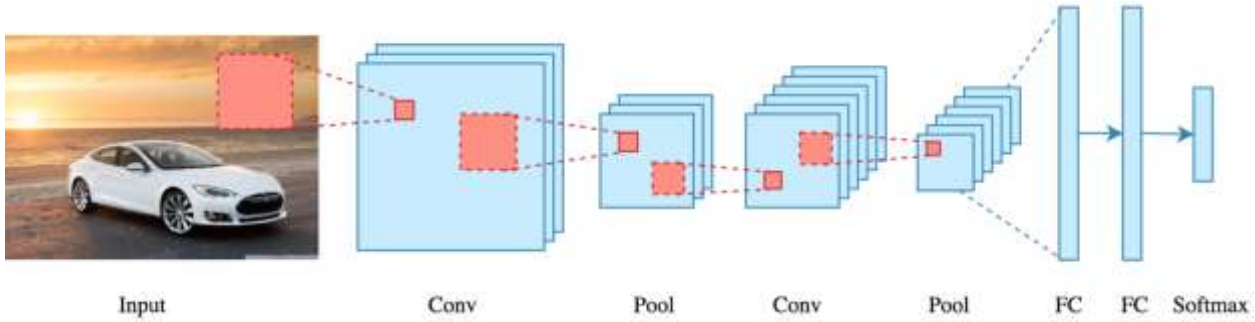
CNN model

Convolutional Neural Network

Convolutional layers, pooling layers, and fully-connected (FC) layers are the three types of layers that make up the CNN.

The following are the three types of pooling operations:

- Max pooling: The batch's maximum pixel value is considered
- Min pooling: The batch's minimum pixel value is considered.
- Average pooling: It selects the average value of all pixels in the batch.



CNN example

Convolutional Neural Network

- DenseNet-201 is a 201-layer Convolutional Neural Network (CNN) that uses dense conditions between layers via Dense Blocks, which allow all layers to be directly connected.
- A pre-trained version of the network, trained on millions of photos, can classify images into 1000 object categories, including keyboards, pencils, and many human faces and animals.
- In the end, ConvNet's job is to compress the images into a format that is easier to handle while preserving elements that are important for making a decent prediction.
- CNN delivers in-depth findings despite its power and resource complexity. It's all about spotting patterns and characteristics so small and insignificant that they go unnoticed by the human eye. However, when it comes to comprehending the contents of an image, it falls short.

Steps for Image Processing

Image processing is the application of operations to an image to improve it or derive helpful information from it.

The three phases that makeup image processing are as follows:

- Importing an image using image acquisition tools;
- Analyzing and altering the image;
- Producing an output that can be an altered image or a report based on image analysis.

Image Processing Sample

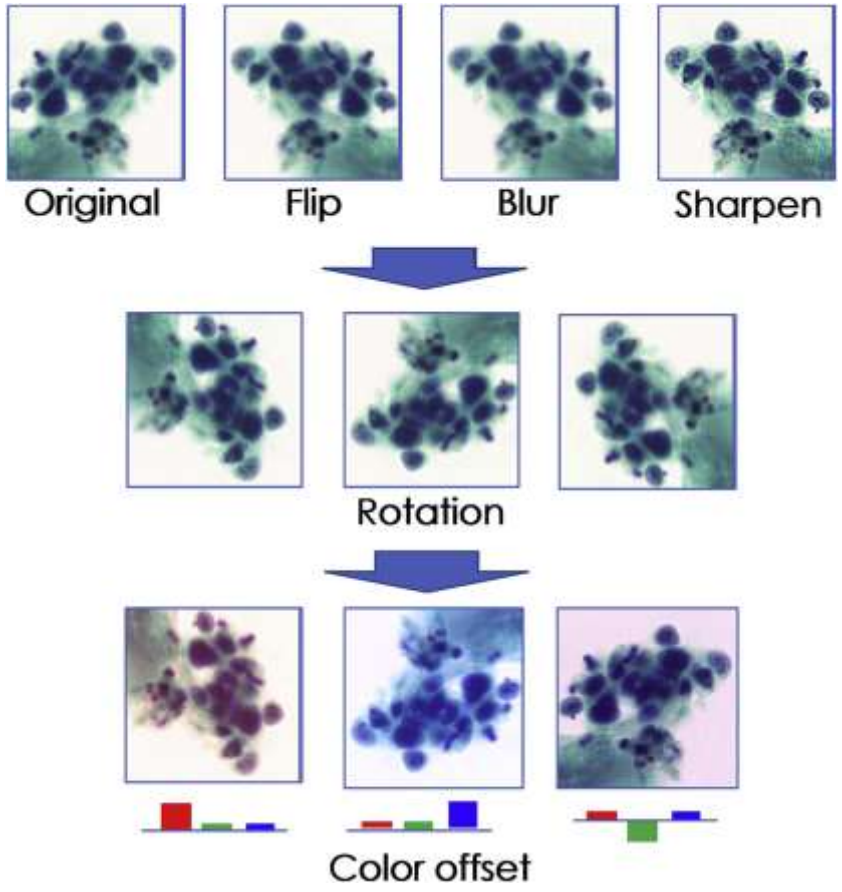


Image processing Sample

Libraries and Modules Used

- Numpy
- Matplotlib
- Scikit-image (skimage)
- Cv2
- Keras
- TensorFlow
- Scikit-learn
- Pandas
- Scipy
- JSON
- Math
- GC
- Intertool
- PIL
- Tqdm
- Functools
- Collection

Implementation

Coding :

- CNN Model 1 : https://colab.research.google.com/drive/1ZgLUbjubv1HIobjSh8oE1RGN23idT19e?usp=sharing#scrollTo=c0b8ae09

- CNN Model 2 : https://colab.research.google.com/drive/1mvklzcztrTt7qHPo3RyYY5hwfnWuPA2K?usp=sharing#scrollTo=31049657

Results, Analysis and Discussions

- 2 CNN models have been employed to determine the best accuracy score in order to detect the cancerous and non-cancerous tissues. Model 1 has an accuracy of 91%, whereas Model 2 has an accuracy of 98%.

	precision	recall	f1-score	support
0	0.97	0.73	0.83	176
1	0.89	0.99	0.93	369
accuracy			0.91	545
macro avg	0.93	0.86	0.88	545
weighted avg	0.91	0.91	0.90	545

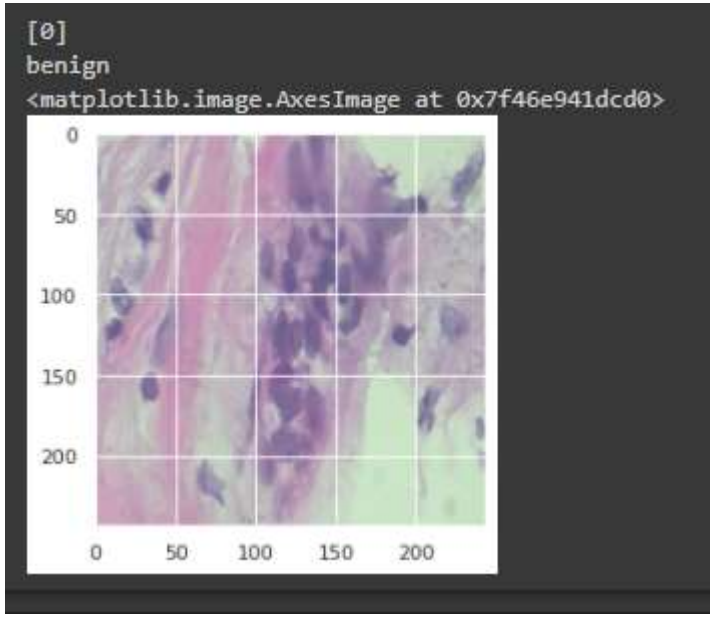
Classification Report for Model 1

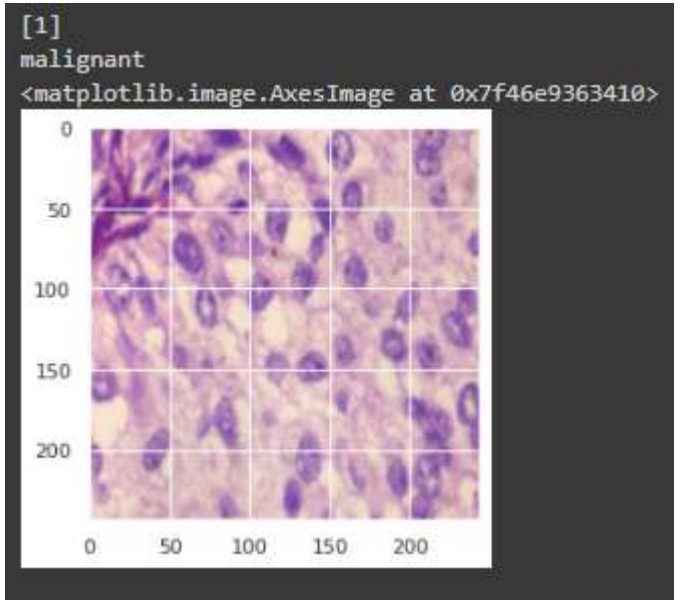
	precision	recall	f1-score	support
0	0.98	0.88	0.93	176
1	0.95	0.99	0.97	369
accuracy			0.96	545
macro avg	0.96	0.94	0.95	545
weighted avg	0.96	0.96	0.96	545

Classification Report for Model 2

Results, Analysis and Discussions

- Outputs of the Test Model using CNN Model 1

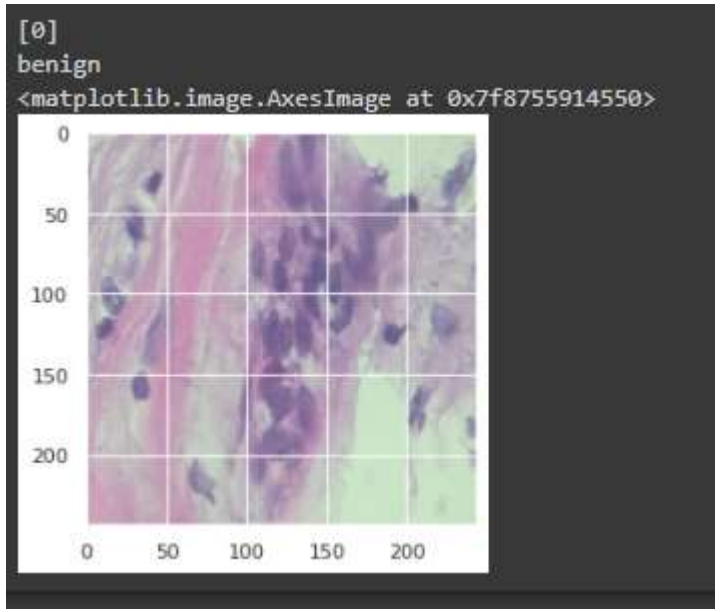


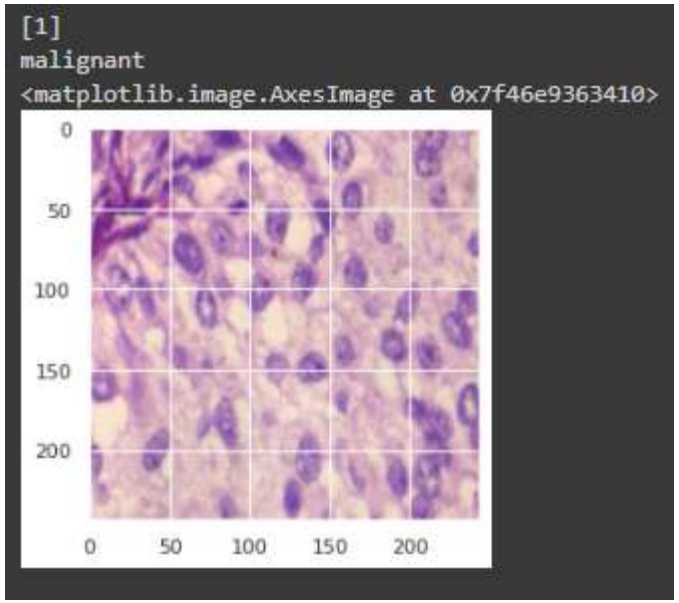


Test Models for Model 1

Results, Analysis and Discussions

- Outputs of the Test Model using CNN Model 2





Test Models for Model 2

Conclusions and Future Scope

- It's a difficult task to automate breast cancer screening in order to improve patient care. Above, two distinct CNN architectures were compared for the detection of breast cancer from the Breast Cancer Histopathological Database (BreakHis).
- The proposed system, which employs CNN Model 2, has a 98% accuracy rate in comparison to CNN Model 1, with an accuracy of 91%.
- The main scope of this project is for healthcare and oncologists to diagnose cancer accurately as early as possible and to reduce human mistakes in the diagnosis phase.
- In the future, the usage of AI and ML for efficient diagnosis should be implied to decrease human errors and help people fight cancer as early as possible.

THANK YOU