



Cellular Analysis of Breast Cancer – Faster RCNN

Yujui Chen, Shihhsien Yang, Pohan Wu

Introduction

- Cancer histology can be predictive of treatment for patients, since it contains underlying molecular processes and disease progression.
- In this project, we use deep learning techniques to classify different types of cells including '**tumor**', '**stromal**', '**TIL**' and '**unknown**', from digital pathology images.



Table of Contents

01

Motivation

02

Pathology

03

**Deep Learning
Method**

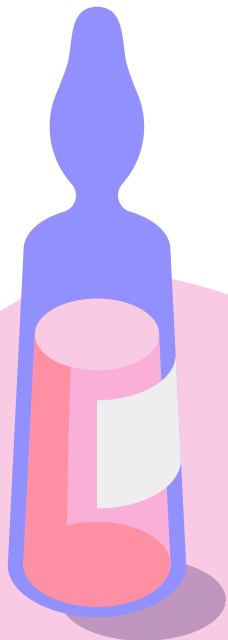
04

Result

05

Challenge





01

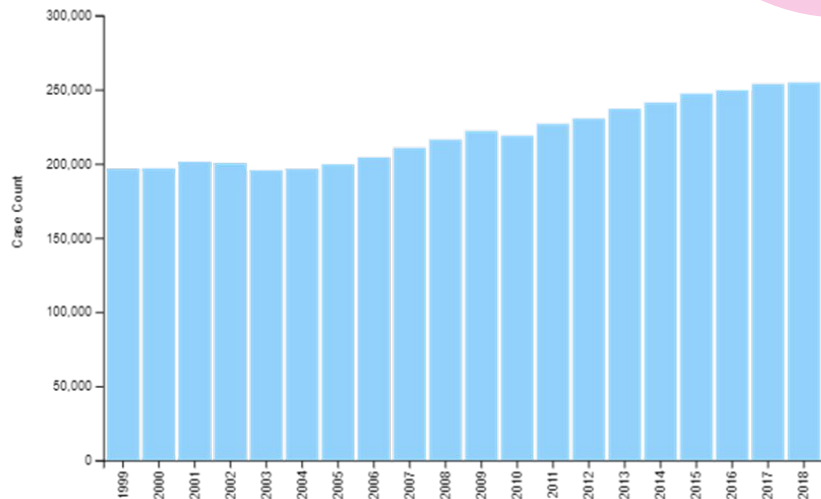
Motivation

- Statistic from CDC
- Method of Diagnose

Statistic from CDC

- Breast cancer is the second most common cancer among women in the United States.
- Deaths from breast cancer have declined over time, but breast cancer remains the second leading cause of cancer death among women overall
- Each year in the United States, about 255,000 cases of breast cancer are diagnosed in women and about 2,300 in men. About 42,000 women and 500 men in the U.S. die each year from breast cancer.

Annual Number of New Cancers, 1999-2018
Female Breast, United States



Method of Diagnose

Breast Ultrasound

Using a machine by sound waves to make pictures, called sonograms, of areas inside the breast.

MRI

A kind of body scan that uses a magnet linked to a computer. The MRI scan will make detailed pictures of areas inside the breast.

Diagnostic Mammogram

A more detailed X-ray of the breast.

Biopsy

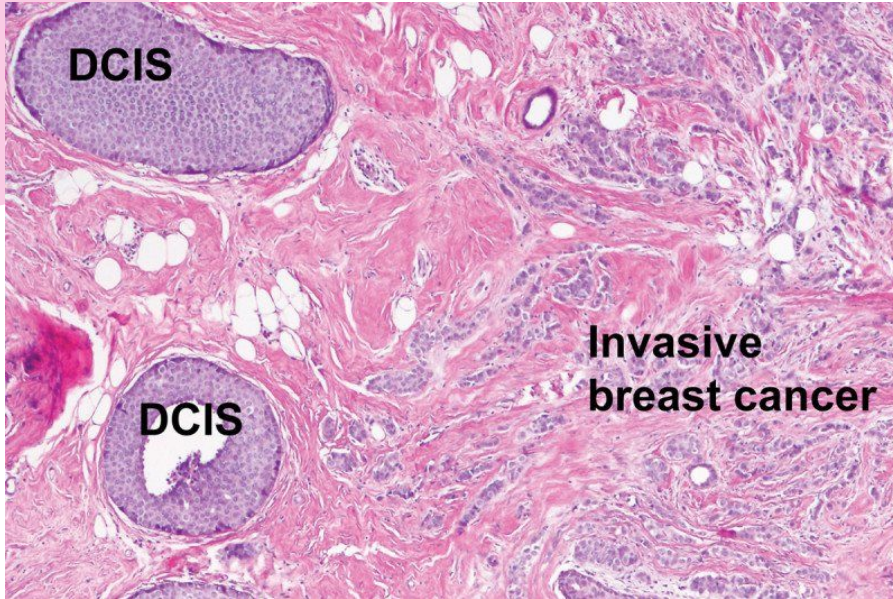
A test that removes tissue or fluid from the breast to be looked at under a microscope and do more testing.



02 Pathology



Introduction

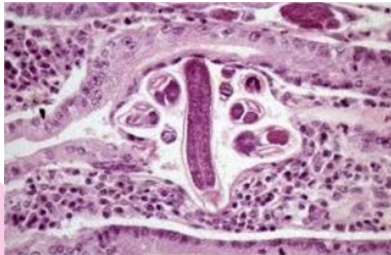


- Pathology is a medical science that involves the study and diagnosis of disease through the examination of surgically removed organs, tissues (biopsy samples), bodily fluids, and in some cases the whole body (autopsy).
- Pathologists specialize in a wide range of diseases including cancer and the vast majority of cancer diagnoses are made by pathologists.
- However, there are not enough pathologists to diagnose the breast cancer which we may applying Deep Learning method for helping pathologists.

Method of Pathology

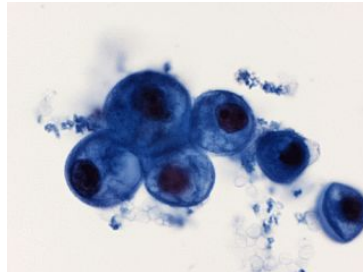
Surgical Pathology

The most significant and time consuming branch of pathology with a primary focus on examining tissues with the naked eye or under a microscope for definitive diagnosis of disease.



Cytopathology

A branch of pathology that studies and diagnoses diseases on the cellular level which generally used on samples of free cells or tissue fragments.

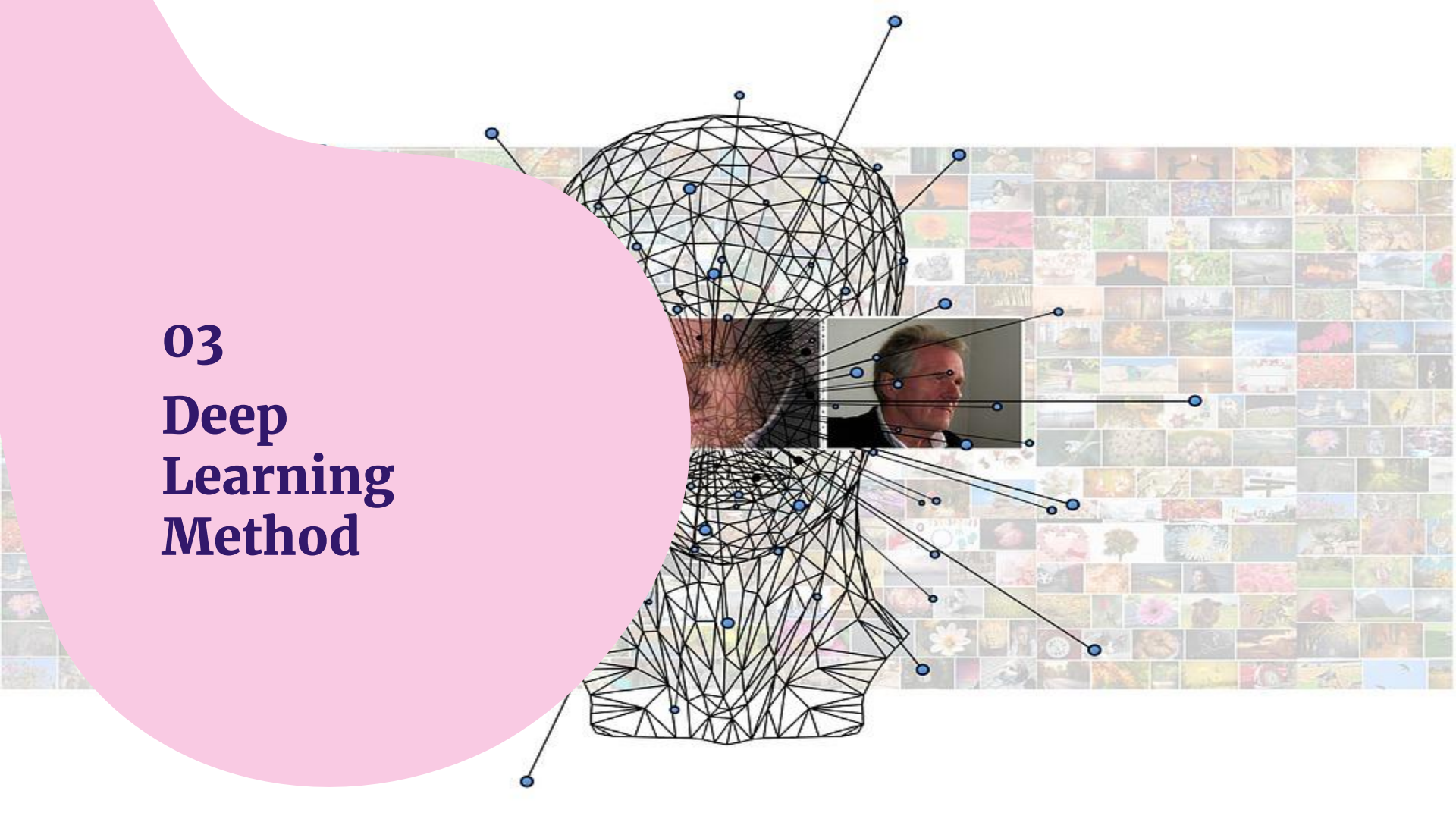


Molecular Pathology

Including the development of molecular and genetic approaches to the diagnosis and classification of human tumours which allows for the detection of very small tumours



03 Deep Learning Method



Structure

**Data
Augmentation**



Classification

Data Augmentation

**Random Flip
(Vertical &
Horizontal)**

**Random
Crop**

**Random
Rotation**

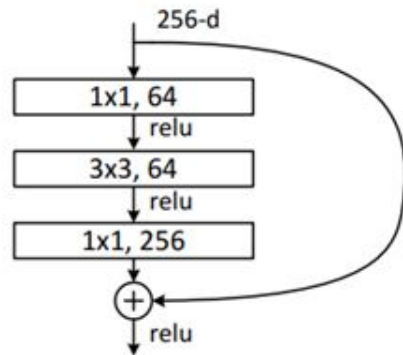
**Invert
Color**

Classification

ResNet50

`tf.keras.applications.ResNet50`

- Residual Learning Block
- Structure:
 - Conv1: 7×7 , 64, stride 2
 - Conv2: $[1 \times 1, 64; 3 \times 3, 64; 1 \times 1, 256] \times 3$
 - Conv3: $[1 \times 1, 128; 3 \times 3, 128; 1 \times 1, 512] \times 4$
 - Conv4: $[1 \times 1, 256; 3 \times 3, 256; 1 \times 1, 1024] \times 6$
 - Conv5: $[1 \times 1, 512; 3 \times 3, 512; 1 \times 1, 2048] \times 3$
 - Average pool, 1000-d FC, softmax
 - FLOPs: 3.8×10^9



Classification (Cont'd)

Optimizer

Adam optimization algorithm

```
tf.keras.optimizers.Adam()
```

Adam optimization is a stochastic gradient descent method that is based on adaptive estimation of first-order and second-order moments.

Loss

Binary Crossentropy

```
tf.keras.losses.BinaryCrossentropy()
```

Computes the cross-entropy loss between true labels and predicted labels.

Metrics

macro-average AUC

```
tf.keras.metrics.AUC()
```

The AUC of the ROC or PR curves are quality measures of binary classifiers which evaluate all the operational points of a model.

04

Result

- Original without data augmentation
- Random Flip (Vertical & Horizontal)
- Random Crop
- Invert Color
- Random Rotation

Accuracy of Model with original dataset volume

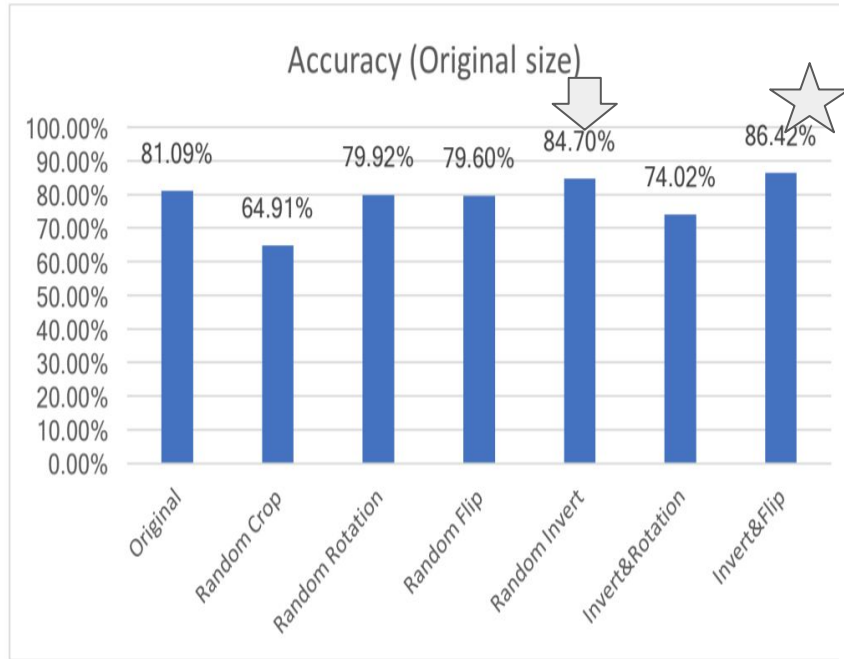


Fig 5 Accuracy of models in original dataset volume

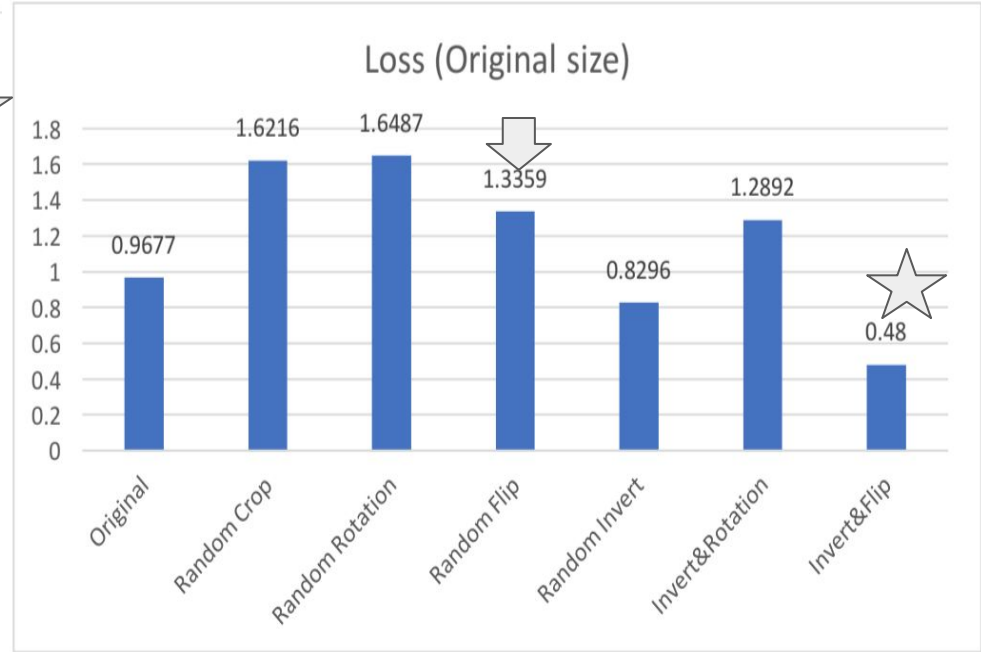


Fig 6 Loss of models in original dataset volume

Accuracy of Model with half dataset volume

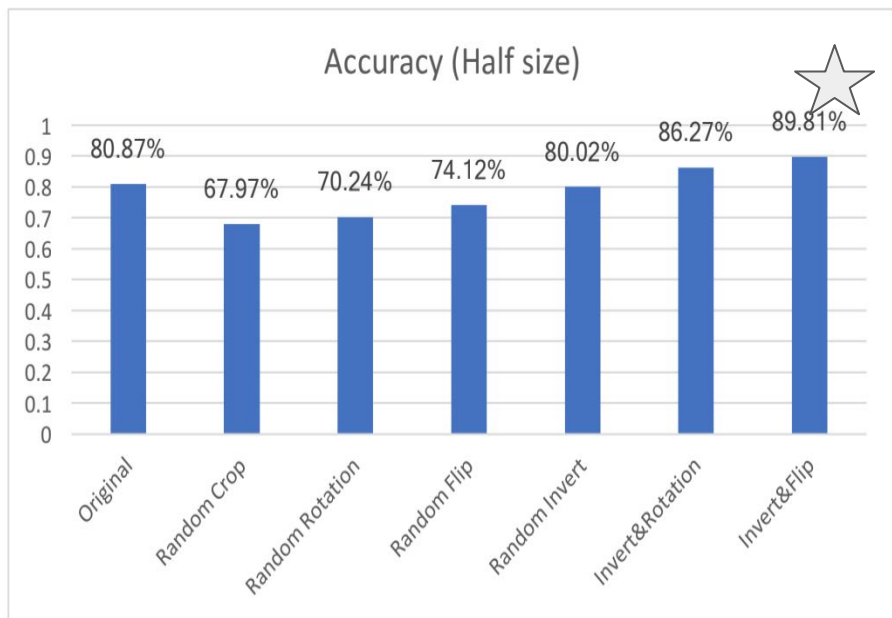


Fig 7 Accuracy of models in half of original dataset volume

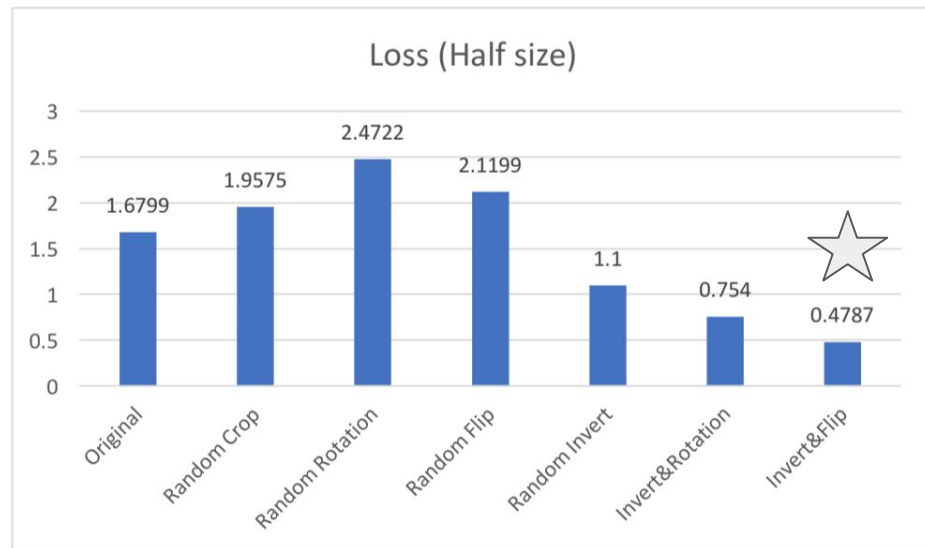


Fig 8 Loss of models in half of original dataset volume

Dataset saturation

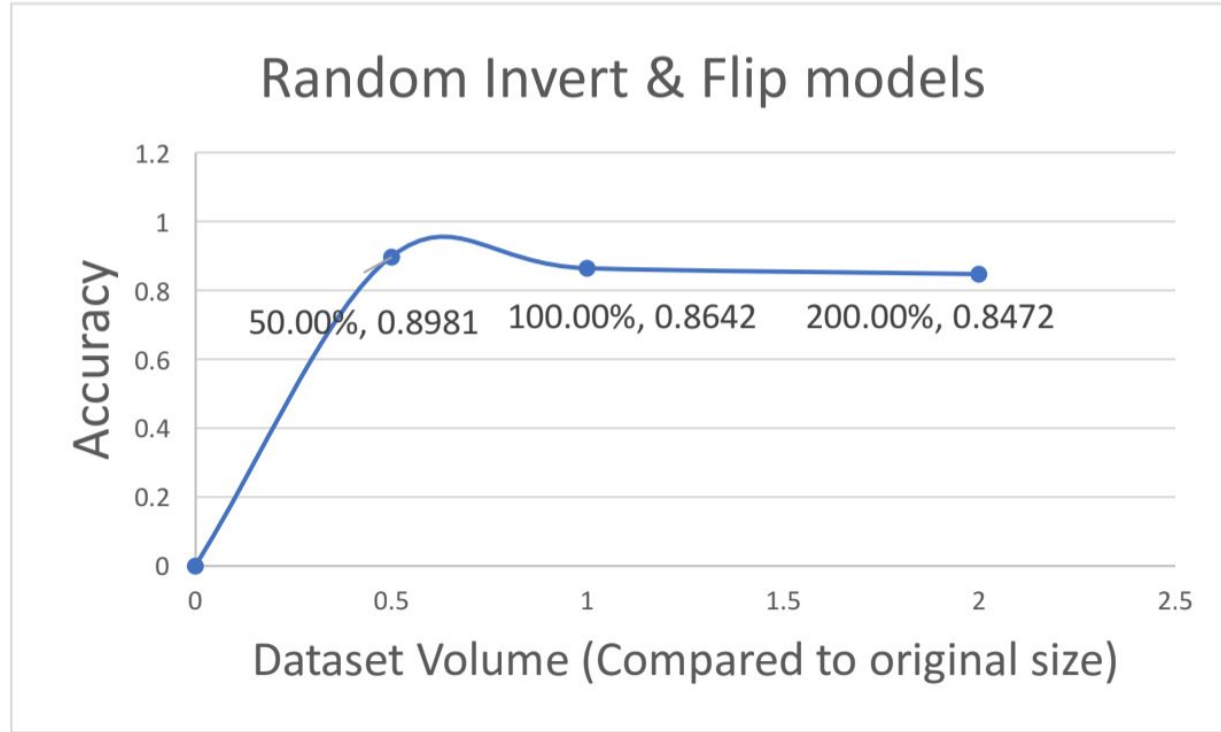
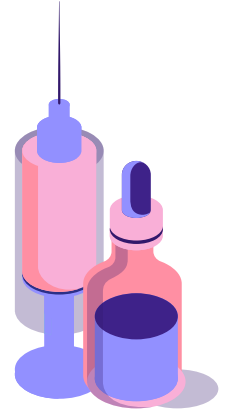


Fig 9 Accuracy versus Dataset volume in random invert&flip models



Future work

- 1 Different data augmentation methods, such as noise and grayscale filter
- 2 Adding more layer into network or trying different deep learning model like GAN